Épistémologie mathématique

Henri Lombardi Stefan Neuwirth

version préliminaire de la deuxième édition

Table des matières

Avant-propos

L'épistémologie est la philosophie des sciences : elle étudie les conditions dans lesquelles une science produit des connaissances. L'épistémologie mathématique a pour but de réfléchir à ce qu'on fait vraiment quand on fait des mathématiques, et d'analyser le rapport entre cette pratique et la pratique des autres sciences. Les mathématiques ont une histoire, et leur histoire est toujours en cours. Aussi nous essaierons d'éclairer par l'histoire les questions soulevées.

Les objectifs généraux de ce cours d'épistémologie mathématique sont de comprendre, ou au moins de discuter les questions suivantes, qui sont quelques angles d'attaque pertinents.

— Qu'est-ce qu'un « objet mathématique » : un nombre entier, un nombre réel, une fonction réelle, un espace vectoriel, un espace de fonctions, les objets géométriques tels que point droite, plan, espace... ?

— Qu'est-ce qu'un « énoncé vrai » ? Qu'est-ce que cela signifie exactement lorsqu'on énonce : **Théorème**... ? Quelles méthodes de raisonnement sont-elles vraiment légitimes ?

— Quelle est la nature de l'infini mathématique ? Peut-on résoudre les paradoxes de Zénon ? Qu'est-ce exactement que le coup de force de Cantor, qui impose les ensembles infinis en mathématiques ? Quelle signification accorder au théorème de Cantor, qui affirme qu'il y a des infinis « plus grands » que d'autres ? Comment le retrouve-t-on dans le théorème de Turing, qui affirme qu'il n'y a pas de programme capable de sélectionner les bons programmes ?

— Qu'est-ce que la méthode formaliste en mathématiques ? Quelles limites le théorème d'incomplétude de Gödel impose-t-il au formalisme ?

Les méthodes de travail que nous utiliserons seront

— des cours sur ces questions : le lecteur ne doit pas s'attendre à y trouver le style usuel des textes mathématiques ;

— des analyses de preuve : pour un même énoncé, nous comparerons différentes preuves pour essayer de comprendre ce qu'elles nous disent, non seulement au sujet de la vérité, mais aussi au sujet de la nature des objets manipulés ;

— des commentaires de textes historiques.

La méthode de travail que nous vous conseillons est une lecture soigneuse de ce cours, crayon à la main pour accompagner les réflexions et les calculs du texte sur une feuille de brouillon.

Pour celles qui ont du temps, nous recommandons la lecture des trois ouvrages suivants [1] : *Les métamorphoses du calcul : une étonnante histoire de mathématiques* de Gilles Dowek (2007), *La naissance des objets mathématiques* d'Enrico Giusti (1999), *Preuves et réfutations* d'Imre Lakatos (1984).

Nous vous invitons aussi à consulter le site web http://hlombardi.free.fr.

1. La personne qui nous lit subit l'alternance des sexes, et sera donc légèrement surprise une fois sur deux.

Première partie

Épistémologie mathématique
Cours

Table des matières

Chapitre 1

La rigueur en mathématiques

Introduction

Ce chapitre introductif discute de la rigueur en mathématiques à travers quelques exemples puisés dans l'histoire. Évidemment, il s'agit surtout de situer les problèmes et non de les traiter de manière exhaustive. Certains sujets seront développés dans des chapitres ultérieurs.

Le problème de la rigueur est étroitement lié à la conception que l'on a des êtres mathématiques et des énoncés mathématiques. Par exemple :

1. Les êtres mathématiques :
   - en géométrie, qu'est-ce qu'un point, une droite ? que veut on dire lorsqu'on affirme que deux surfaces sont égales, etc. ?
   - en analyse, qu'est-ce qu'un nombre réel, une fonction continue, un infinitésimal, etc. ?
2. Les énoncés mathématiques :
   - qu'est ce qu'un énoncé qui a du sens ?
   - qu'est ce qu'un énoncé vrai ?
   - qu'est ce qu'un énoncé faux ?

C'est à la lumière de telles conceptions de base qu'on analyse une preuve et qu'on l'accepte ou non comme réellement convaincante. Inversement, l'analyse de certaines preuves peut éclairer des problèmes auparavant cachés et modifier nos conceptions de base. Toutes ces conceptions ont évolué à travers l'histoire, et le dernier mot n'est pas dit, et ne sera sans doute jamais dit.

1.1 La géométrie élémentaire

Euclide

Le traité des *Éléments* d'Euclide a très longtemps été considéré comme le modèle de la rigueur dans les raisonnements mathématiques. C'est le premier texte connu où est mise en place et utilisée de manière systématique ce qu'il est convenu d'appeler la méthode axiomatique. Une grande partie du traité est consacrée à ce que nous appelons aujourd'hui la géométrie élémentaire. La traduction la plus récente est due à Bernard Vitrac (Euclide d'Alexandrie 1990-2001).

Euclide ne prend pas position sur la nature exacte des objets géométriques qu'il manipule : points, (segments de) droites, surfaces planes ou courbes, volumes, etc. Implicitement il s'agit d'une idéalisation de notions communes, matérialisées par des figures bien concrètes. Quelques pseudo-définitions sont données, qui ressemblent plus à des commentaires qu'à des précisions utiles au raisonnement. Certains des premiers raisonnements utilisent l'intuition du déplacement des objets solides pour justifier les cas d'égalité des triangles. Mais ensuite on a un discours qui

n'utilise, au moins en apparence, que les axiomes et les cas d'égalité, et qui se développe selon les seules règles de la logique. La plupart des règles du raisonnement sont elles-mêmes implicites.

Un nombre important de propriétés remarquables sont démontrées, de manière apparemment irréprochable. D'autres auteurs de la même civilisation, tel Archimède, ont poursuivi la tradition.

Descartes

Descartes critique la géométrie grecque, non pour son absence de rigueur, mais pour son obscurité (il ne suffit pas de convaincre, il faut aussi expliquer) et pour l'absence d'une méthode générale qui serait susceptible d'attaquer tous les problèmes de géométrie de manière uniforme et efficace.

Il propose une telle méthode, le calcul algébrique sur les coordonnées. Tous les mystères doivent être résolus par de simples calculs mécaniques. C'est ce qu'on a appelé le *programme de Descartes*.

Aujourd'hui ce programme est en partie réalisé. D'une part les logiciels de géométrie dynamique sont capables de certifier sans se tromper un grand nombre de théorèmes de géométrie. Le théorème doit être d'un type particulier, quoique assez général. Par exemple il ne doit pas faire intervenir des inégalités ($a < b$), mais seulement des égalités. Le logiciel, à qui on a soumis une certaine situation avec des éléments variables, fait alors le calcul suivant : il choisit pour chacun des éléments variables quelques valeurs numériques au hasard. Au lieu de faire un calcul algébrique avec des lettres, il fait un calcul purement numérique, approché. Si dans tous les cas choisis au hasard, le théorème est vérifié (à la précision du calcul près), le logiciel déclare le théorème vrai. On ne l'a jamais vu se tromper. C'est extrêmement troublant, car on peut difficilement croire qu'il s'agit d'une preuve rigoureuse. Pourtant tout mathématicien, quoique professionnel de la rigueur, fait un jour ou l'autre une erreur de raisonnement, tandis que le logiciel, dont la méthode de preuve n'est pas rigoureuse, ne se trompe jamais.

D'autre part, Tarski a prouvé vers 1950 un résultat annoncé vers 1930 : si une question « purement algébrique » est posée concernant des nombres réels arbitraires, elle admet une réponse précise qui peut être calculée par un algorithme. Par exemple les théorèmes de la géométrie euclidienne usuelle rentrent tous dans ce cadre. Il faut ne considérer que des courbes, surfaces, etc., définis par des systèmes d'équations et d'inégalités algébriques. L'algorithme proposé par Tarski est en fait impraticable, même par des ordinateurs très puissants. Mais d'autres algorithmes sont régulièrement proposés et le champ des problèmes de géométrie qui peuvent être traités de manière automatique par une machine s'élargit chaque année.

Bolyai et Lobatchevski

Pendant des siècles les géomètres ont tenté de se débarrasser du cinquième postulat dans la géométrie d'Euclide.

Ce postulat qui, sous une forme équivalente, affirme que par un point extérieur à une droite passe une et une seule parallèle à cette droite, a toujours été considéré comme le seul axiome « non naturel » dans le système d'Euclide : il a une tête de théorème et non d'axiome.

À force de chercher à tirer des conséquences de la négation du cinquième postulat (dans le but de démontrer par l'absurde qu'il résultait des autres axiomes) ils ont fini par construire une géométrie « non euclidienne » tout à fait cohérente, dans laquelle seul le cinquième postulat est changé, en son contraire. Par exemple Lambert en était arrivé à la conclusion que la géométrie non euclidienne devait être celle d'une sphère de rayon imaginaire $\sqrt{-\mathrm{R}}$, où R dépend de l'unité de longueur choisie.

Le mérite principal revient cependant à Bolyai et Lobatchevski qui ont décidé (indépendamment et simultanément) de braver une fois pour toutes l'interdit et de développer la géométrie non euclidienne de manière vraiment systématique. On appelle *géométrie hyperbolique* la géométrie inventée par Bolyai et Lobatchevski.

Considérons une droite Δ et un point S extérieur à Δ, puis les sécantes (SM) à Δ où M est un point variable sur Δ. Une de ces sécantes est la droite (SH) orthogonale à Δ au point H. En admettant, comme notre intuition nous le dit, que la droite Δ peut être prolongée indéfiniment dans ses deux directions opposées, quelle est la position limite de (SM) lorsque M s'éloigne indéfiniment sur Δ dans chacune des deux directions ? Selon le cinquième postulat, les deux positions limites

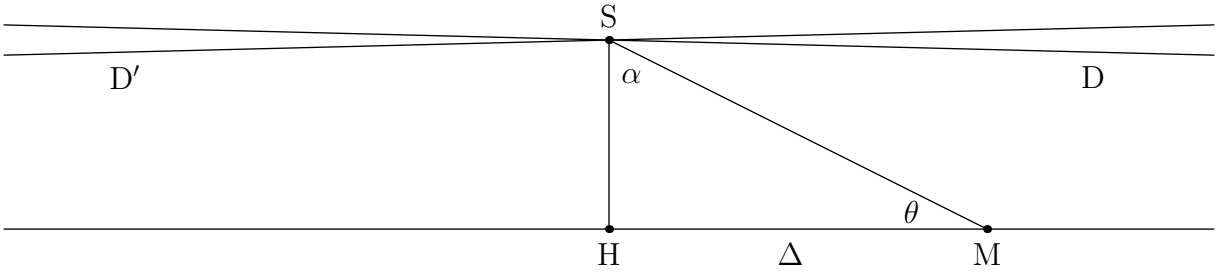


FIGURE 1.1.1 – Le contraire du postulat d'Euclide.
D et D′ sont les positions limites de (SM) lorsque
M part à l'infini sur Δ vers la droite ou vers la gauche.

de la droite (SM) sont les mêmes. Autrement dit, chacune de ces deux positions limites, qui sont symétriques par rapport à la droite (SH), forme un angle droit avec (SH) : lorsque θ tend vers l'angle nul, α tend vers un angle droit. Notons que l'angle droit en α est impossible pour une sécante (SM) car sinon, par raison de symétrie, la droite (SM) couperait aussi Δ en le point M′ symétrique de M par rapport à H, et il passerait deux droites distinctes par les points M et M′.

Dans le cas où les positions limites D et D′ de (SM) ne sont pas orthogonales à (SH) la valeur limite de l'angle α sera strictement inférieure à un droit et il y aura moyen d'insérer dans l'angle formé par les deux positions limites autant de droites que l'on veut ne coupant pas Δ et passant par S.

Ce qu'on réclame d'un plan hyperbolique ce sont les mêmes ingrédients de base que dans un plan euclidien : les points, les droites, les segments, les demi-droites – et la possibilité de *déplacer librement les figures géométriques*, ce qui correspond *grosso modo* aux deux premiers « cas d'égalités des triangles » (voir l'exercice 1.1.1).

Ce n'est que dans la deuxième moitié du 19^{e} siècle qu'on a dégagé le concept de *libre mobilité des figures* à partir de celui de congruence des figures. Il exprime que les déplacements ne changent pas les propriétés géométriques (longueurs et angles) des figures. Voici comment Hermann Helmholtz l'exprime en 1870.

> La base de toute démonstration, dans la méthode euclidienne, consiste à établir la congruence de lignes, d'angles, de figures planes, de solides, etc., etc. Pour rendre cette congruence évidente, on suppose qu'on applique les figures géométriques les unes sur les autres, sans changer bien entendu leurs formes ou leurs dimensions. La chose est possible en fait, nous l'avons toujours expérimenté depuis notre plus tendre enfance. Mais, quand nous voulons donner le caractère d'une nécessité logique à une proposition, en nous fondant sur la possibilité de transporter ainsi les figures, sans changer leur forme, dans toutes les parties de l'espace, nous devons rechercher si cette possibilité n'implique pas, au préalable, quelque proposition non encore démontrée. Nous verrons plus loin qu'il en est ainsi, et qu'il découle même de ce fait des conséquences très importantes. C'est pourquoi, dans ce cas, toute démonstration fondée sur la congruence reste appuyée sur un fait purement expérimental.
>
> . . . . . .
>
> En premier lieu, il faut supposer que la position d'un point quelconque A doit pouvoir être déterminée, par rapport à des éléments fixes, au moyen de mensurations opérées sur des grandeurs quelconques, lignes, angles, surfaces, etc. On sait que ces mensurations nécessaires pour déterminer

la position du point A s'appellent ses coordonnées. Le nombre des coordonnées nécesssaire pour fixer complètement la position d'un point détermine le nombre des dimensions de l'espace considéré. il faut supposer en outre que, dans le mouvement du point A, ses coordonnées varient d'une manière continue.

En second lieu, il faut donner la définition d'un corps solide par rapport à un système de points fixes, comme cela est nécessaire pour pouvoir entreprendre la comparaison des grandeurs par voie de congruence. Comme nous ne pouvons supposer encore ici aucune méthode spéciale pour mesurer les grandeurs, cette définition ne peut être donnée que de la manière suivante : Entre les coordonnées des points qui appartiennent à un corps solide, pris deux à deux, il doit exister une équation correspondant à la relation invariable qui subsiste entre les deux points pendant le mouvement du corps, et qui est la même pour tous les couples de points congruents. Les couples de points congruents sont ceux qui peuvent coïncider avec le même couple de points fixes dans l'espace. (Helmholtz 1877, pages 1198, 1202)

Ce concept exprime donc une propriété de l'espace euclidien, mais il est en fait caractéristique des espaces dont la "courbure" (définie par Gauss en 1829) est constante. Par exemple, les figures tracées sur une sphère de centre O ne changent pas lorsqu'on effectue des rotations d'axe passant par O ou des réflexions par rapport à un plan passant par O.

Quelle est la figure la plus simple qui permet d'exprimer cette propriété de libre mobilité ? En réfléchissant un peu, on se rend compte que cela revient à se demander combien de points d'un corps un déplacement doit fixer pour le déterminer. Cette propriété peut donc s'exprimer comme suit sans faire appel aux égalités de longueurs ou d'angles. On définit un « drapeau » comme

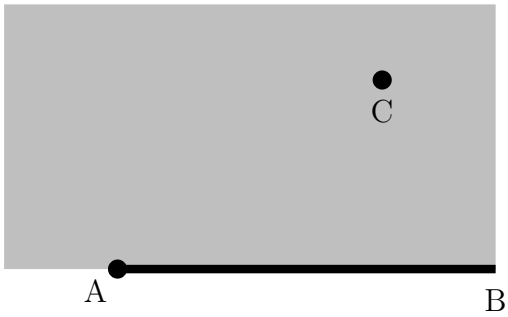


FIGURE 1.1.2 – Drapeau défini par la demi-droite [AB) et le demi-plan bordé par (AB) contenant le point C.

constitué par une demi-droite et un des deux demi-plans qui bordent la droite correspondante. Alors le groupe des isométries doit opérer simplement transitivement sur les drapeaux : autrement dit, étant donnés deux drapeaux, il y a exactement une isométrie qui envoie l'un sur l'autre. De même, le groupe des déplacements (ou isométries directes) doit opérer simplement transitivement sur les demi-droites.

Une manière de concevoir la géométrie intrinsèque est de postuler l'existence et l'unicité des transformations qui envoient une figure-drapeau sur une autre, puis de poser que deux segments sont égaux s'il existe une telle transformation qui envoie un segment sur l'autre. Il n'y a pas de cercle vicieux parce que la longueur est définie à l'aide des transformations ; on appelle celles-ci des "isométries" à cause de cette définition de la longueur. C'est très différent de ce qu'on fait lorsqu'on base la géométrie sur l'algèbre linéaire et les espaces vectoriels normés de deux ou trois dimensions.

Bolyai et Lobatchevski pensaient que l'on pouvait démontrer la cohérence de sa géométrie à partir de la cohérence des formules de trigonométrie du triangle qu'ils avaient établies dans le plan hyperbolique. Ces formules ressemblent beaucoup, avec de légères variations, aux formules de trigonométrie du triangle dans la géométrie sphérique.

Finalement la conviction que la géométrie hyperbolique était ni plus ni moins cohérente que la géométrie euclidienne se fit comme suit.

D'une part à l'intérieur de l'espace hyperbolique de dimension 3, Bolyai et Lobatchevski introduisirent de nouvelles surfaces, les *horosphères*, qu'on peut voir comme des « limites de sphères » : on prend la sphère passant par un point A de centre S situé sur une demi-droite Δ issue de A et on fait s'éloigner S à l'infini sur Δ ; les *horocycles* sont l'intersection des horosphères avec un plan contenant Δ : la figure 1.1.3 montre des cercles passant par A de centres S, S′, S″, S‴, S⁗ de plus en plus éloignés sur la droite Δ (voir l'exercice 1.1.2). Ils montrèrent que les points d'une

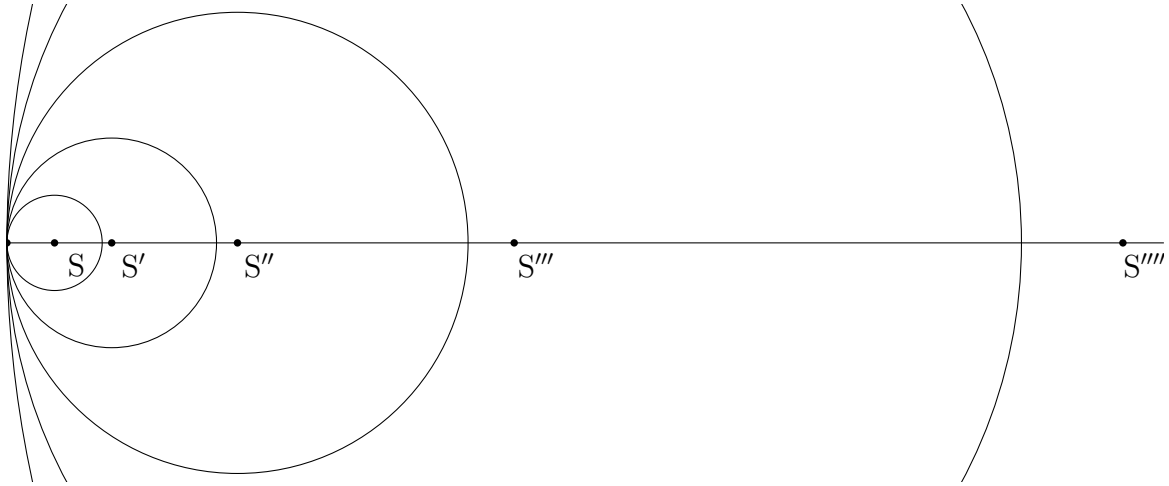


FIGURE 1.1.3 – Approximations de l'horocycle défini par A et Δ.

horosphère définissaient, avec les horocycles pris comme les « droites », la même géométrie que le plan euclidien usuel. Ceci donnait évidemment une grande crédibilité à la géométrie hyperbolique, sans toutefois prouver sa cohérence.

D'autre part Beltrami et Poincaré proposèrent des constructions inverses : des modèles de géométrie hyperbolique dans l'espace euclidien. Expliquons par exemple le modèle de Beltrami, dans lequel les droites du plan hyperbolique sont représentées par des segments ouverts de droite. On considère un disque ouvert D du plan euclidien et on note D_∞ le cercle qui borde ce disque ouvert. Les points du plan hyperbolique sont représentés par les points de D. Les droites du plan hyperbolique sont représentées par les segments ouverts de droites du type $\Delta =]a, b[$ avec a et b sur D_∞. Les points a et b peuvent être vus comme les deux points à l'infini qui définissent la droite Δ.

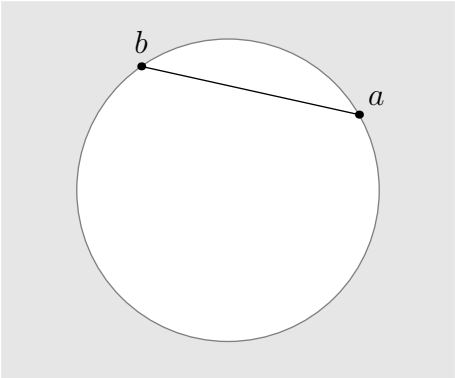


FIGURE 1.1.4 – Le modèle de Beltrami.

Hilbert

Après le tremblement de terre constitué par la découverte de la géométrie hyperbolique, ce ne peut plus être la géométrie qui justifie les nombres réels, mais bien les nombres réels qui, à travers le calcul algébrique sur les coordonnées, justifient la géométrie (que ce soit celle d'Euclide ou de Bolyai et Lobatchevski).

Hilbert en profite pour faire une relecture critique d'Euclide. À la suite de Moritz Pasch, il propose une axiomatique rénovée, qui s'appuie en bonne partie sur la théorie des ensembles alors en gestation, mais essaie d'éviter de faire appel directement aux nombres réels. Cela donne son fameux livre *Fondements de la géométrie* (Hilbert 1900), dont la première édition date de 1899.

Tout d'abord il découvre de nombreux « axiomes non dits » dans la géométrie d'Euclide. Ce sont des propriétés des figures qui sont utilisées sans même qu'on s'en aperçoive sur la foi de l'intuition immédiate. Il s'agit en général d'axiomes liés à la relation d'ordre sur $\mathbb{R}$ et transcrits sous forme géométrique. Par exemple une droite qui coupe un côté d'un triangle doit forcément passer par le sommet opposé ou couper l'un des deux autres côtés : c'est l'axiome de Pasch illustré par la figure 1.1.5 (voir l'exercice 2 du devoir n° 1). Ou encore : une droite qui a un point à l'intérieur d'un cercle coupe le cercle en deux points.

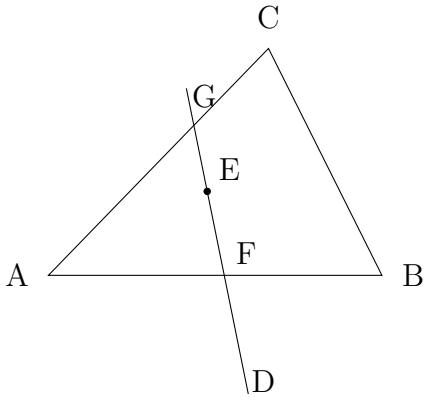


FIGURE 1.1.5 – Axiome de Pasch : si la droite DE coupe le côté AB d'un triangle ABC, alors elle coupe aussi le côté AC ou le côté BC, ou elle passe par C.

Pour éviter de parler des nombres réels dans le système axiomatique, Hilbert introduit un axiome de complétion, impossible à formuler sans la théorie des ensembles. Cela se lit : il est impossible de rajouter des points au plan euclidien sans détruire la cohérence du système axiomatique. L'analogue pour les nombres réels serait : $\mathbb{R}$ est un corps ordonné archimédien maximal (c.-à-d., qui ne possède pas de surcorps ordonné archimédien strictement plus gros). Un axiome de ce genre n'est pas « conforme » : au lieu d'imposer des propriétés aux objets de base (points, droites, segments) et à leurs relations de base (ici les relations d'incidence et de congruence [1]), il impose une propriété globale au « modèle » ; on dit dans ce cas que la propriété est "non prédicative". Dans les théories formelles que Hilbert proposera plus tard, seuls les axiomes du premier type seront autorisés : c'est ce qu'on appelle les théories formelles du premier ordre.

L'axiome « non conforme » de Hilbert pourrait aussi se reformuler comme suit : tout ensemble de nombres réels positifs possède une borne inférieure. Un tel axiome est dit « du second ordre » car il s'énonce avec un « pour toute partie de $\mathbb{R}\ldots$ » [2].

1. Le système de Hilbert utilise une relation primitive de « congruence » entre figures simples, dont la signification intuitive est qu'elles sont isométriques. Ainsi la distance entre deux points n'est pas une notion du système, car le système ne contient pas les nombres réels, mais l'égalité des longueurs AB et CD, appelée congruence, est une notion du système.

2. On peut comparer à ce sujet http://math.stackexchange.com/questions/1473770/does-there-exist-a-maximal-archimedean-ordered-field. La définition de $\mathbb{R}$ ne semble pas faire référence à elle-même : elle consiste à dire qu'on peut identifier un nombre rationnel r avec l'ensemble infini des nombres rationnels plus grands

Toute preuve de cohérence qu'on proposerait pour la géométrie euclidienne à la Hilbert fournirait une preuve de cohérence pour « le système des nombres réels » en tant que corps ordonné archimédien maximal. Un tel espoir parait aujourd'hui démesuré, vu l'axiome du second ordre.

En bref, en pensant clore le débat une fois pour toutes, Hilbert l'a en fait précisé et déplacé.

Aujourd'hui

Si on veut comprendre la géométrie euclidienne sans faire appel à ce que nous avons appelé le système des nombres réels, il est possible d'utiliser une théorie formelle du premier ordre dans laquelle seules les propriétés purement algébriques de $\mathbb{R}$ sont sous-jacentes. Ces propriétés sont concentrées dans l'affirmation suivante : *les nombres réels, avec l'addition, la multiplication et la relation d'ordre, forment un corps ordonné « réel clos »*. Le caractère réel clos signifie qu'une fonction polynôme qui change de signe aux extrémités d'un intervalle doit s'annuler sur l'intervalle. Cette propriété se formule comme une propriété des objets de base (nombres réels et opérations algébriques élémentaires)

N. B. : Ne considérer que les propriétés purement algébriques des nombres réels revient à ne considérer dans le plan que les courbes algébriques et à exclure les courbes transcendantes (la cycloïde par exemple). C'est dans ce cadre que Tarski a montré que tout problème géométrique posé en termes "purement algébriques"[3] admet une solution algorithmique.

Les *Fondements de la géométrie* de Hilbert revus à la Tarski correspondent à un système axiomatique dont on peut prouver la cohérence avec des méthodes « élémentaires ». Du point de vue de la cohérence, c'est donc très nettement préférable aux *Fondements de la géométrie* à la Hilbert.

La contrepartie, naturellement, est qu'on ne prouve rien quant à la cohérence du système des nombres réels lui-même, pourtant censé être le vrai système numérique implicite dans la géométrie euclidienne. Le système des nombres réels est autrement plus compliqué que celui des entiers naturels. Comme on sait qu'en un certain sens on n'aura jamais dit le dernier mot au sujet du

---

que r, puis à définir un nombre réel comme un ensemble infini de nombres rationnels vérifiant les trois propriétés suivantes :

— il contient un nombre rationnel ;

— il admet un minorant, c'est-à-dire un nombre rationnel inférieur ou égal à tous les éléments de l'ensemble ;

— s'il contient un nombre rationnel, il contient tout nombre rationnel plus grand que celui-ci.

C'est une version des coupures de Dedekind : par exemple, le nombre réel $\sqrt{2}$ est défini comme l'ensemble des nombres rationnels p/q tels que $pq > 0$ et $p^2 > 2q^2$. Le problème fondamental ici est de savoir dans quelle mesure on peut vraiment prétendre qu'un tel ensemble est bien défini.

Si on postule que les ensembles de rationnels existent d'avance et qu'on en détermine un avec ces trois propriétés, on passe ce problème sous silence : c'est ce que fait la théorie des ensembles.

Posons au contraire qu'un tel ensemble de rationnels est déterminé seulement si on fournit un procédé qui permet de déterminer si un nombre rationnel donné lui appartient ou non (comme pour l'exemple de $\sqrt{2}$ ci-dessus). Alors le problème devient apparent lorsqu'on examine la preuve que tout ensemble de nombres réels qui admet un minorant admet une borne inférieure. Cette borne inférieure est censée être un nombre réel, donc définie par un tel procédé ; or la preuve fournit un procédé qui est défini en faisant appel à ce procédé lui-même : il y a un cercle vicieux, à moins de dire que le procédé qui définit la borne inférieure ne définit pas un nombre réel parmi ceux qu'on a déjà définis, mais un nombre réel plus compliqué ; c'est dans ce sens qu'on parlera de nombre réel du 2e ordre.

Via les suites de Cauchy, le même problème apparait : si on pose qu'une suite de Cauchy (u_n) de nombres rationnels est donnée seulement si on propose un procédé pour calculer tous les termes u_n, et que les nombres réels sont donnés comme de tels procédés de calcul, la borne inférieure d'un ensemble de réels est un procédé de calcul différent de tous ces procédés. On commet un cercle vicieux en le comptant d'avance parmi ces procédés.

On peut continuer la réflexion à ce sujet avec l'exercice 5.5.1 page 200.

3. C'est-à-dire que sa forme logique est particulièrement simple : pour être précis, ce doit être une formule du "premier ordre", dans laquelle on quantifie uniquement sur les nombres entiers (quand on commence par dire "il existe un" ou "pour tout", le mot suivant est toujours "entier"). Une formule du second ordre est une formule dans laquelle on quantifie aussi sur les ensembles de nombres entiers.

système des entiers naturels, on doit être encore plus modeste avec le système des nombres réels.

1.2 La rigueur dans $\mathbb{N}$

La rigueur dans le raisonnement avec les entiers naturels semble ce qu'il y a de plus facile à maitriser, car les entiers naturels sont les objets mathématiques les plus simples.

Un entier naturel est codé par un mot écrit sur l'alphabet $0, 1, 2, \ldots, 9$ et la possibilité de raisonner avec des objets si élémentaires est directement reliée à la capacité de maitriser une langue. Parler une langue en se faisant comprendre semble au moins aussi compliqué que faire des calculs sans se tromper. Abstraire la notion de nombre entier à partir de ses réalisations concrètes (que sont les opérations de comptage) relève du même type d'abstraction que celui à l'œuvre dans le langage lorsqu'un même mot, table par exemple, est utilisé pour désigner des objets très différents les uns des autres. Et la combinatoire des mots à l'œuvre dans la construction des phrases est certainement aussi complexe que la combinatoire des chiffres dans une opération élémentaire du style $352 + 567 = 919$.

C'est pourquoi on considère souvent que les entiers naturels, en tant qu'objets de base, ne peuvent pas être définis à partir d'objets plus simples[4], mais seulement être commentés.

L'axiomatique de Peano

On peut cependant se poser la question d'une description précise des règles de base qui doivent être appliquées dans un raisonnement rigoureux avec les entiers naturels. Et ce n'est évidemment pas un hasard si c'est au 19^{e} siècle, quand la géométrie hyperbolique a fait douter des fondements, quand on s'est aperçu de la difficulté à fonder le calcul différentiel sur des bases solides qu'est apparu ce besoin. Peano a proposé une axiomatique pour cela. Elle est basée sur les concepts (non définis) de nombre (entier naturel) et de successeur. L'idée est de dire sous une forme axiomatique que les entiers sont engendrés par 0 et l'opération successeur, sans jamais boucler. En français, cela donne :

1. 0 est un nombre.
2. Le successeur d'un nombre est un nombre.
3. Deux nombres qui ont même successeur sont égaux.
4. 0 n'est le successeur d'aucun nombre.
5. Si une propriété concernant un nombre arbitraire est vraie pour 0, et si elle est vraie pour le successeur de n dès qu'elle est vraie pour n alors elle est vraie pour tous les nombres. (axiome de récurrence)

Notez qu'on obtient facilement par récurrence la propriété : Tout nombre différent de 0 est le successeur d'un nombre.

Maintenant, est-ce que les propriétés de base de l'addition et de la multiplication : commutativité, associativité, distributivité doivent être prouvées par récurrence (en suivant l'axiomatique de Peano) ? Dans ce cadre on considèrera que l'addition et la multiplication sont définies par récurrence au moyen des axiomes suivants, dans lesquels m' représente le successeur de m (on peut aussi noter $s(m)$ si on préfère).

1. Addition
   (a) $m + 0 = m$

4. Il est vrai que dans la théorie des ensembles, on a fini par s'écarter de ce point de vue, (qui était le point de vue initial) selon lequel les entiers sont des objets qui forment un ensemble mais ne sont pas eux-mêmes des ensembles. On trouve fréquemment aujourd'hui les entiers définis comme des ensembles, en posant $0 = \emptyset$, puis de proche en proche $n + 1 = \{n\} \cup n$. Mais cela n'éclaire en rien la question des entiers naturels, car cela revient à définir le simple par le compliqué.

(b) $m + (n') = (m + n)'$

2. Multiplication
   (a) $m \cdot 0 = 0$
   (b) $m \cdot (n') = (m \cdot n) + m$

Vous pouvez vous amuser à essayer de prouver les propriétés élémentaires de l'addition et de la multiplication par récurrence. C'était même au programme des classes de terminale dans les années 1970.

Des preuves plus intuitives

Mais les preuves plus traditionnelles par des arguments de comptage de collections finies sont-elles vraiment moins rigoureuses ? Dans cet autre cadre, plus naturel puisque les nombres ont d'abord été inventés pour compter, pas pour vérifier l'axiomatique de Peano, l'addition $m + n$ est définie par la juxtaposition de deux collections disjointes : m poires et n pommes font $m + n$ fruits. Démontrer $m + n = n + m$ ne semble même plus nécessaire, tant cela résulte de la définition de l'addition.

$$\underbrace{\{\bullet, \ldots, \bullet\}}_{m} \cup \underbrace{\{\times, \ldots\ldots, \times\}}_{n} = \underbrace{\{\bullet, \ldots, \bullet, \times, \ldots\ldots, \times\}}_{m+n}$$

Même chose concernant l'associativité de l'addition. La multiplication est définie à partir de la notion de couple. S'il y a m poires et n pommes, il y a $m \cdot n$ couples formés d'une pomme dans la première collection et d'une poire dans la seconde. Sous une forme plus visuelle on range les couples dans un tableau rectangle avec m lignes et n colonnes. On obtient, au choix m fois n cases, ou n fois m cases. L'associativité de la multiplication revient à dire qu'il y a deux manières de définir un triplet (a, b, c) avec des couples : $((a, b), c)$ ou $(a, (b, c))$. Enfin la distributivité correspond au dessin suivant.

$$\underbrace{\begin{array}{ccc}\bullet & \cdots & \bullet \\ \vdots & & \vdots \\ \bullet & \cdots & \bullet\end{array}}_{m \cdot n} \cup \underbrace{\begin{array}{ccc}\times & \cdots\cdots & \times \\ \vdots & & \vdots \\ \times & \cdots\cdots & \times\end{array}}_{m \cdot p} = \underbrace{\begin{array}{ccccc}\bullet & \cdots & \bullet & \times & \cdots\cdots & \times \\ \vdots & & \vdots & \vdots & & \vdots \\ \bullet & \cdots & \bullet & \times & \cdots\cdots & \times\end{array}}_{m \cdot (n+p)}$$

Ce qui se cache dans les arguments de comptage

Pourtant en réfléchissant bien, on verra que tous ces arguments de comptage reposent sur une propriété implicite, toujours admise comme intuitivement évidente, mais qui mérite peut-être une preuve. C'est la suivante.

Lorsqu'on compte de deux manières différentes une même collection finie d'objets, on aboutit toujours au même nombre.

Bien que cette propriété semble incontestable en tant que vérité d'expérience, elle n'est après tout pas évidente, surtout pour les nombres très grands. Aussi semble-t-il inévitable, pour rendre les preuves par comptage rigoureuses, de démontrer une bonne fois la propriété ci-dessus. Et sur ce point il semble qu'il n'y ait pas d'autres possibilités que le raisonnement par récurrence.

Si deux comptages d'une même collection finie aboutissent d'une part à m et d'autre part à n, nous allons montrer par récurrence sur n que $m = n$.

Notez que, en passant par la collection finie qui produit les deux comptages, on obtient une correspondance bijective entre les ensembles $[\![1, m]\!]$ et $[\![1, n]\!]$. Nous voulons montrer que l'existence d'une telle correspondance bijective force l'égalité $m = n$.

Pour $n = 0$ la chose est claire : la collection est vide donc $m = 0$. Si cela vous gêne d'initialiser une récurrence sur un problème de comptage dans le cas où il n'y a rien à compter, nous admettons

l'argument et nous vous donnons aussi la preuve directe pour $n = 1$. La collection est non vide, donc $m > 0$ et il s'écrit $p+1$. Par le processus même du comptage, p compte le nombre d'éléments dans la collection privée du dernier élément compté. Or cette collection est vide, donc $p = 0$. En fait vous allez voir que l'argument pour $n = 1$ serait produit à l'identique par la récurrence lorsqu'on l'initialise à 0.

Supposons maintenant la chose vraie pour n et montrons-la pour $n + 1$. Par hypothèse on a une correspondance bijective entre l'ensemble $[\![1, m]\!]$ et l'ensemble $[\![1, n+1]\!]$. Puisque la collection est non vide, on a $m \geq 1$ et on peut écrire $m = p + 1$. Notons $k \mapsto \sigma(k)$ cette correspondance bijective. Deux cas se présentent. Le premier est celui où $\sigma(p + 1) = n + 1$. Alors, puisque σ est bijective, elle établit par restriction une correspondance bijective entre les ensembles $[\![1, p]\!]$ et $[\![1, n]\!]$. Donc l'hypothèse de récurrence s'applique, $p = n$ et donc $p + 1 = n + 1$. Le deuxième cas est celui où $\sigma(p + 1) = r \neq n + 1$. En faisant suivre σ par la correspondance bijective de $[\![1, n + 1]\!]$ avec lui-même qui échange r et $n + 1$ sans toucher à rien d'autre, on obtient une correspondance bijective τ entre $[\![1, p+1]\!]$ et $[\![1, n+1]\!]$, correspondance qui vérifie cette fois-ci $\tau(p+1) = \tau(n+1)$. On est donc ramené au premier cas envisagé, et on peut conclure que $p + 1 = n + 1$.

Notons cependant que la preuve précédente peut encore être critiquée à cause de l'usage qu'elle fait de la collection $[\![1, m]\!]$ sans que celle-ci ait été au préalable clairement définie.

La collection $[\![1, m]\!]$ peut être définie sans « les trois petits points » en disant qu'il s'agit des entiers k qui vérifient $k \geq 1$ et $k \leq m$. Mais cette définition fait usage de la relation d'ordre. En fait nous avons implicitement utilisé cette autre propriété des entiers naturels, encore plus intuitivement évidente que celle que nous cherchions à montrer : le fait que deux entiers naturels peuvent être comparés. Dans la succession des entiers l'un des deux arrive avant l'autre. Cette intuition « temporelle » est-elle légitime *a priori* ou nécessite-t-elle une preuve ?

Précisément il y a une relation d'ordre total sur les entiers naturels qui peut être définie par : $m \leq n$ si et seulement si il existe un entier p tel que $n = p + m$. Montrer en toute rigueur cette dernière propriété [5] semble nécessiter une preuve par récurrence. Et celle-ci utilisera probablement la commutativité et l'associativité de l'addition.

Quelle conclusion tirer de tout ceci ?

Que la rigueur n'est certes pas une chose facile à cerner une fois pour toutes. Même dans le cas de propriétés simples des entiers naturels.

Le système axiomatique de Peano a son charme, celui du minimalisme. En fait il peut lui même être critiqué pour certaines imprécisions contenues dans l'axiome de récurrence :

— À quelles propriétés précisément s'applique cet axiome ?

— Quand nous avons défini l'addition « par récurrence », cet axiome nous autorisait-il à le faire, ou bien n'aurions nous pas dû en formuler un autre, qui autorise des constructions d'objets par récurrence ?

Par ailleurs rien n'oblige à utiliser un système aussi minimal. On pourrait aussi bien prendre comme axiomes les propriétés de base de l'addition, de la multiplication et de la relation d'ordre.

1.3 Le théorème fondamental de l'algèbre

L'histoire du théorème fondamental de l'algèbre est instructive. Rappelons une de ses formulations.

Théorème 1.3.1. *Si $f(\mathrm{X}) = \mathrm{X}^n + \sum_{k=0}^{n-1} a_k \mathrm{X}^k$ est un polynôme à coefficients complexes, il peut être décomposé en facteurs du premier degré : $f(\mathrm{X}) = \prod_{k=1}^{n} (\mathrm{X} - z_k)$.*

Un énoncé équivalent est que le polynôme admet au moins un zéro dans $\mathbb{C}$. En effet, une fois qu'on a trouvé un zéro z_n on peut diviser f par $\mathrm{X} - z_n$ et le polynôme g qui en résulte est de

5. À savoir, la relation en (m, n) définie par « $\exists p\ m = n + p$ » est une relation d'ordre total.

degré $n-1$. On peut conclure (par récurrence sur le degré) que f se décompose en un produit de facteurs du premier degré.

Ce théorème est réputé avoir été démontré rigoureusement pour la première fois par Gauss.

Euler avait démontré qu'un nombre complexe admet n racines $n^{\text{ièmes}}$ dans $\mathbb{C}$. Cela aurait suffi à démontrer le théorème fondamental de l'algèbre si les équations algébriques pouvaient toujours être résolues par radicaux, c'est-à-dire par un nombre fini d'opérations élémentaires sur les coefficients du polynôme et des calculs de racines $n^{\text{ièmes}}$, comme le sont les équations de degré ≤ 4. Mais Abel et Galois démontrèrent que cet espoir était vain et qu'à partir du degré 5 certaines équations n'ont aucune solution qui puisse s'exprimer en fonction des coefficients en utilisant uniquement les opérations arithmétiques usuelles et les extractions de racines.

Avant de discuter les preuves de Gauss, nous donnons d'abord une preuve intuitive qu'il est facile de comprendre dès qu'on a acquis une familiarité avec la représentation des nombres complexes dans le plan euclidien.

Une preuve intuitive

L'idée est d'identifier la « droite complexe » $\mathbb{C}$ avec le « plan réel » $\mathbb{R}^2$ et de considérer la fonction $z \mapsto f(z)$ comme une fonction qui associe des points de ce plan à des points de ce plan. Nous allons utiliser les coordonnées polaires, sachant que toute la problématique tourne autour de l'origine du plan, identifié avec le nombre complexe 0.

On recouvre alors ce plan par une famille de courbes C_ϱ, et on essaye d'analyser la famille des courbes images $\Gamma_\varrho = f(\mathrm{C}_\varrho)$. Une famille intéressante est la famille des cercles centrés à l'origine, indexés par leur rayon ϱ, pour laquelle on a le paramétrage classique

$$\mathrm{C}_\varrho = \{z \mid z = \varrho \exp(i\theta), \theta \in [0, 2\pi]\}$$

Ce qui se passe est la chose suivante.

— Pour $\varrho = 0$, la courbe Γ_ϱ est réduite au point a_0.

— Pour une valeur de ϱ petite mais non nulle, le terme $a_1 z = a_1 \varrho \exp(i\theta)$ dans la somme $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + z^n$ l'emporte sur les termes qui suivent avec un exposant plus grand $a_k z^k = a_k \varrho^k \exp(ik\theta)$ (du moins si $a_1 \neq 0$), et la courbe décrit à très peu près un petit cercle centré en a_0 et de rayon $|a_1|\varrho$.

— Pour une valeur de ϱ grande, le terme $z^n = \varrho^n \exp(in\theta)$ dans la somme l'emporte sur tous les autres et, vue de loin, la courbe ressemble à un grand cercle qui ferait n fois le tour de l'origine.

Comme la courbe fermée Γ_ϱ évolue continument en fonction du paramètre ϱ, qu'elle est concentrée sur le point a_0 quand $\varrho = 0$ et qu'elle fait n fois le tour de l'origine lorsque ϱ est très grand, il a bien fallu que pour au moins une valeur particulière de ϱ, elle passe par l'origine : en fait on « voit » que la famille de courbes balaye tout le plan complexe.

Voici par exemple ce qui se passe avec un polynôme de degré 5 : les figures 1.3.1 à 1.3.9 ci-dessous montrent les courbes $f(\mathrm{C}_\varrho) = \Gamma_\varrho$ obtenues pour des valeurs croissantes de ϱ (ϱ varie sur l'intervalle $[0, +\infty[$).

Cette preuve a le mérite d'être simple et de donner une information visuelle directement compréhensible. Naturellement, elle ne donne pas directement un moyen sûr de calculer un zéro du polynôme. En outre il y a un argument de continuité qui n'est pas facile à maitriser. Que signifie précisément qu'une courbe dépendant d'un paramètre évolue de manière continue en fonction du paramètre ?

Comment rendre cette preuve intuitive plus rigoureuse ?

Une possibilité est la suivante. On *compte le nombre de tours que la courbe fait autour de l'origine*. Si $g\colon [0, 2\pi] \to \mathbb{C}$ est une courbe fermée ($g(2\pi) = g(0)$) continument dérivable qui *ne*

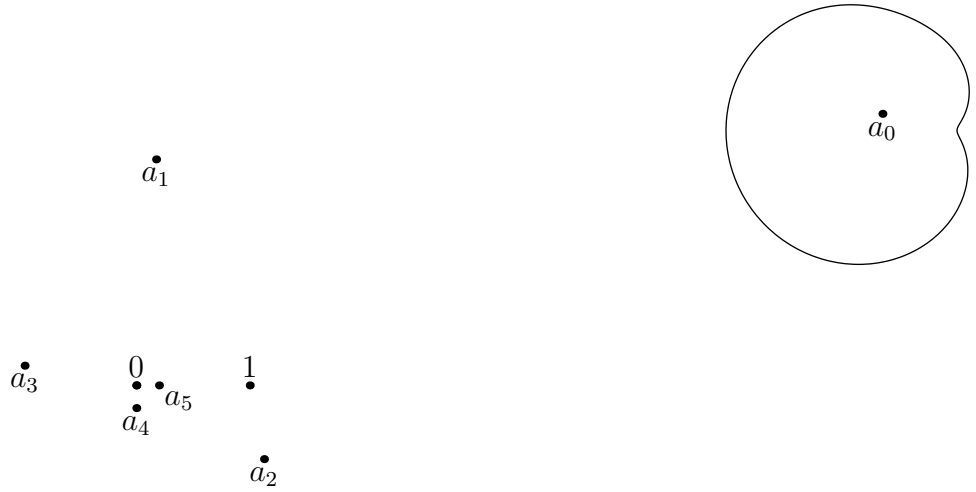


FIGURE 1.3.1 – Théorème fondamental de l'algèbre, $f(z) = \sum_{k=0}^{5} a_k z^k$, pour $|z| = \varrho = 0{,}5$

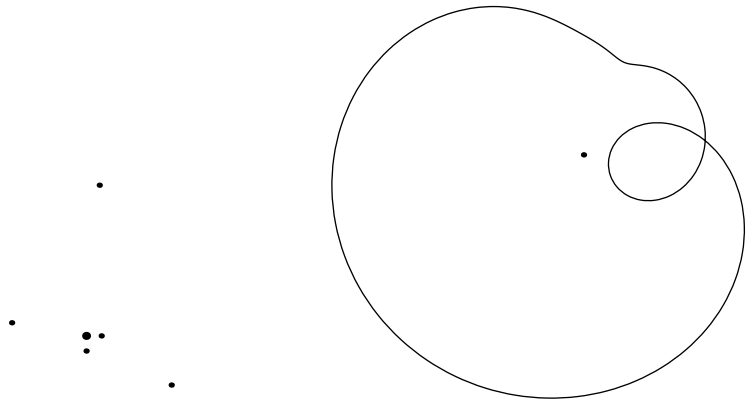


FIGURE 1.3.2 – Théorème fondamental de l'algèbre, $\varrho = 0{,}9$

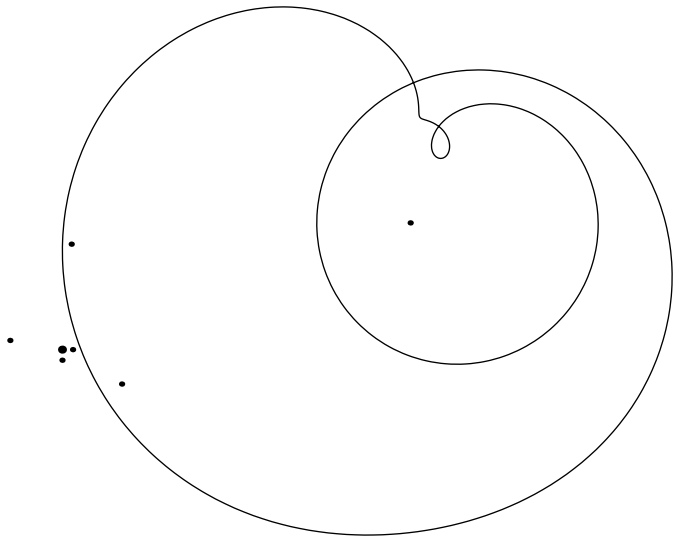


FIGURE 1.3.3 – Théorème fondamental de l'algèbre, $\varrho = 1{,}3$

passe pas par l'origine on peut calculer le nombre de tours qu'elle fait autour de l'origine lorsqu'on fait varier la variable t de 0 à 2π. Pour y arriver, on écrit $g(t)$ en coordonnées polaires :

$$g(t) = r(t)(\cos(\alpha(t)) + i \sin \alpha(t))) \text{ où } \begin{cases} r(t) = |g(t)| \\ \alpha(t) = \arg(g(t)) \text{ est déterminé à } 2\pi \text{ près,} \end{cases}$$

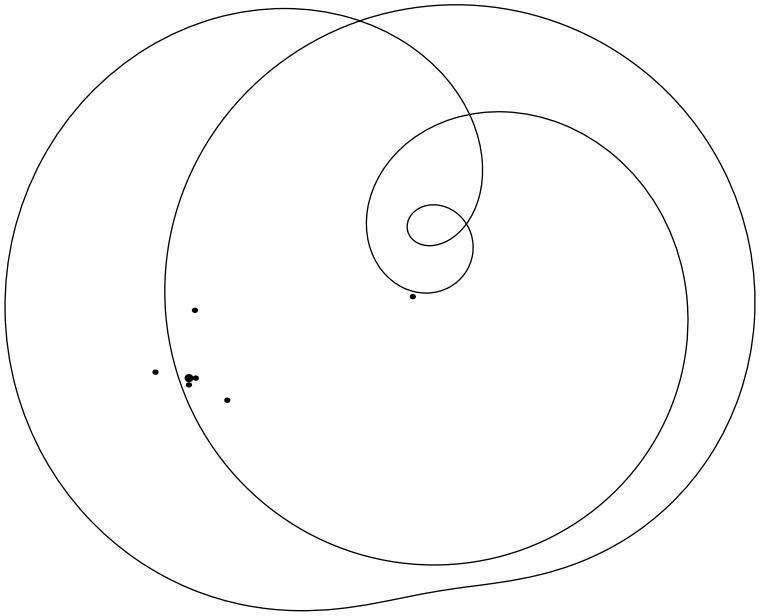


FIGURE 1.3.4 – Théorème fondamental de l'algèbre, $\varrho = 1{,}7$

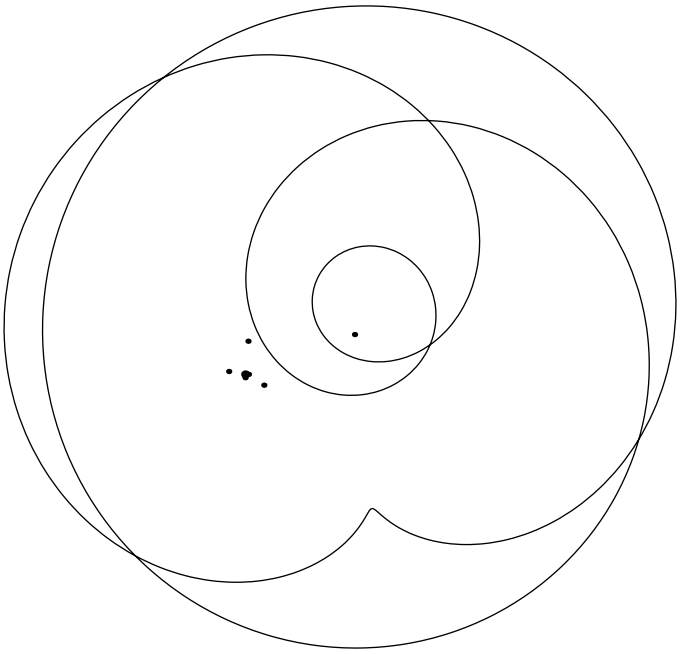


FIGURE 1.3.5 – Théorème fondamental de l'algèbre, $\varrho = 2{,}1$

et parce que la fonction r ne s'annule pas, on peut s'arranger pour *suivre* α *par continuité*, de sorte que les deux fonctions r et α sont continues et même continument dérivables. Alors non seulement $\alpha(t)$ exprime sous quel angle on voit $g(t)$ à partir de l'origine, mais $\alpha(t)$ augmente (resp. diminue) aussi de 2π à chaque fois que $g(t)$ fait un tour complet autour de l'origine dans le sens trigonométrique (resp. dans le sens des aiguilles d'une montre) : donc le nombre total de tours que fait $g(t)$ autour de l'origine lorsque t varie de 0 à 2π est égal à $\mathrm{N} = (\alpha(2\pi) - \alpha(0))/2\pi$. [6] Tout ce que nous venons d'écrire est assez intuitif si on a l'image de la courbe en tête, mais il n'est pas si

6. On peut exprimer N comme une intégrale : si $g(t) = x(t) + iy(t)$, alors on a $\tan\alpha = y/x$, ce qui donne $(1 + \tan^2\alpha)\,\mathrm{d}\alpha = (x\mathrm{d}y - y\mathrm{d}x)/x^2$ et donc $\mathrm{d}\alpha = (x\mathrm{d}y - y\mathrm{d}x)/(x^2 + y^2)$. En conséquence on obtient que le nombre de tours, comptés dans le sens trigonométrique, est égal à

$$\mathrm{N} = \frac{\alpha(2\pi) - \alpha(0)}{2\pi} = \frac{1}{2\pi}\int_{t=0}^{2\pi} (xy' - yx')/(x^2 + y^2)\,\mathrm{d}t.$$

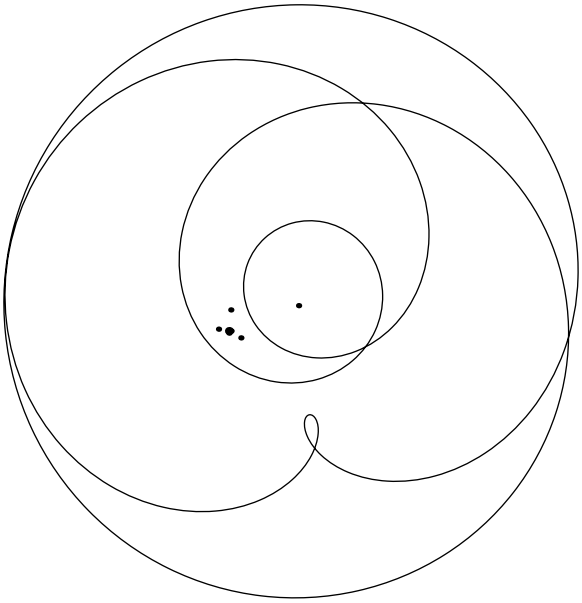


FIGURE 1.3.6 – Théorème fondamental de l'algèbre, $\varrho = 2{,}3$

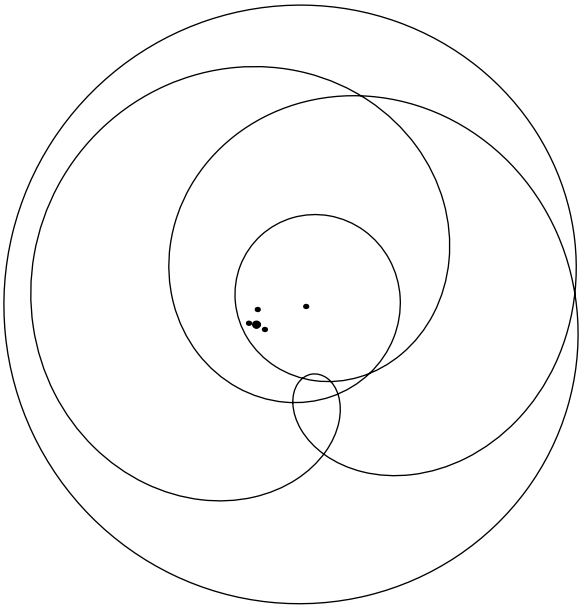


FIGURE 1.3.7 – Théorème fondamental de l'algèbre, $\varrho = 2{,}5$

facile de montrer que le résultat du calcul, à savoir N, est bien un nombre entier sans s'appuyer sur l'intuition géométrique. La clé de ce résultat est que le nombre N ne change pas si on change continument la fonction g tout en lui interdisant de passer par l'origine.

Mettons que ce travail soit fait proprement. On peut alors conclure comme suit.

Dans la famille de courbes Γ_ϱ, on montre que le nombre $\mathrm{N}(\varrho)$ dépend continument du paramètre ϱ tant que la courbe ne passe pas par l'origine. Comme $\mathrm{N}(0) = 0$ et $\mathrm{N}(\varrho) = n$ pour ϱ très grand, il est impossible que la fonction g_ϱ (dont l'image est la courbe Γ_ϱ) ne passe pas au moins une fois par l'origine pour une certaine valeur du paramètre ϱ : dans le cas contraire la fonction $\varrho \mapsto \mathrm{N}(\varrho)$ serait une fonction partout définie sur $[0, \infty[$, et continue, mais une fonction réelle définie et continue sur un intervalle, si elle ne prend que des valeurs entières, est constante sur cet intervalle.

On aura ainsi obtenu une preuve plus rigoureuse, au prix de grands efforts (qui donnent en fait des résultats très généraux utiles par la suite). Néanmoins on n'a toujours pas donné le moyen concret de calculer à coup sûr un zéro de la fonction polynôme f car la preuve se conclut par un raisonnement par l'absurde.

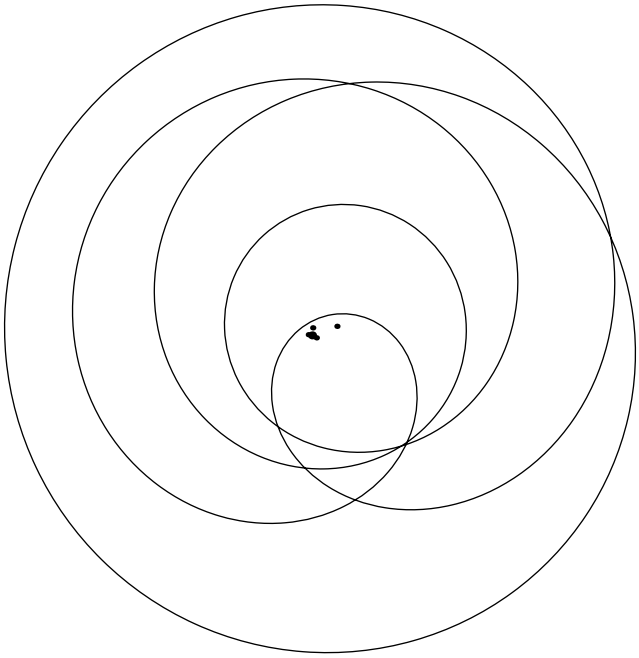


FIGURE 1.3.8 – Théorème fondamental de l'algèbre, $\varrho = 3$

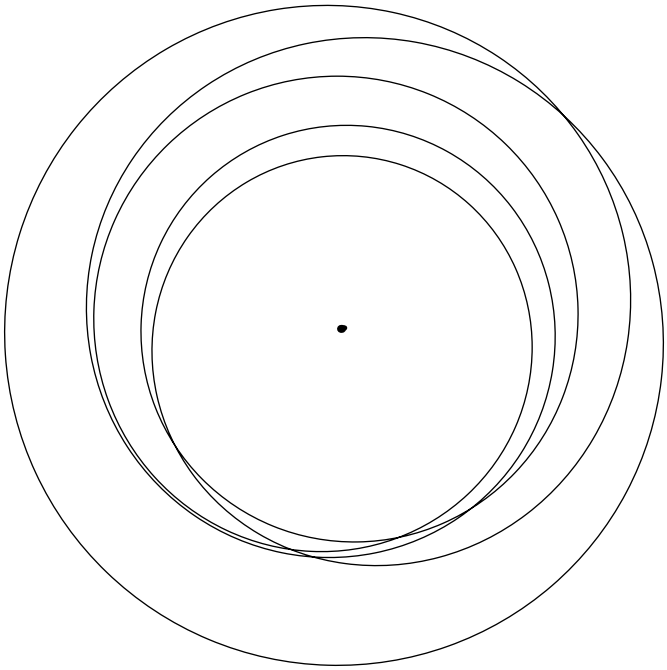


FIGURE 1.3.9 – Théorème fondamental de l'algèbre, $\varrho = 5$

Critique de Bolzano à la preuve de Gauss

La critique que Bolzano portait aux preuves par Gauss du théorème fondamental de l'algèbre (qui en fit plusieurs) était qu'elles étaient toutes basées sur une intuition géométrique. La meilleure de ces preuves, selon Bolzano, avait repoussé l'usage de l'intuition géométrique au seul théorème des valeurs intermédiaires : une fonction continue de $[a, b]$ dans $\mathbb{R}$ qui n'a pas le même signe en a et b doit s'annuler sur un point de l'intervalle.

C'est la principale motivation de Bolzano pour fournir une preuve irréprochable du théorème des valeurs intermédiaires, sans appel à l'intuition géométrique.

Critique de Brouwer

Kronecker non plus n'était pas très content des preuves de Gauss, car il n'avait pas une très grande confiance dans les calculs généraux impliquant des nombres réels arbitraires.

Il produisit une preuve du théorème fondamental de l'algèbre pour le cas où les coefficients du polynôme sont des *nombres complexes algébriques*. Ce sont les nombres complexes qui sont zéros d'un polynôme à coefficients entiers. On maitrise mieux ces nombres que les nombres complexes les plus généraux. Par exemple non seulement on peut les calculer avec une précision arbitraire, mais on sait tester s'ils sont nuls ou pas.

Brouwer, qui était au début du 20^{e} siècle le principal opposant à la théorie des ensembles de Cantor, fit une étude de la nature du continu qui le conduisit à rejeter les raisonnements par l'absurde dans le cas où ils servent à démontrer un théorème d'existence. Son point de vue sur le théorème fondamental de l'algèbre peut se résumer ainsi : pour un *vrai* nombre complexe $z = x + iy$ on connait les réels x et y avec une précision arbitrairement grande, mais finie. On ne peut donc pas argumenter dans une preuve en prétendant connaitre x et y avec une précision infinie. Par exemple, un raisonnement qui utilise les signes de x et y est *a priori* suspect, car décider si un nombre réel apparemment nul est ou n'est pas nul nécessite de le connaitre avec une précision infinie, au moins dans le cas où il est nul.

On ne peut être satisfait avec le théorème fondamental de l'algèbre que lorsqu'on a une preuve qui suppose que les coefficients du polynôme sont de *vrais* nombres complexes, et, sous cette seule hypothèse, calcule les zéros du polynôme sous forme de *vrais* nombres complexes.

Brouwer affirmait que le théorème des valeurs intermédiaires ne peut pas être prouvé en toute généralité (si on appliquait ses critères de rigueur), et, par contre il donna une preuve du théorème fondamental de l'algèbre conforme à ses critères de rigueur.

En conséquence, Brouwer montrait que Bolzano n'avait pas vraiment terminé la preuve de Gauss !

Aujourd'hui

L'histoire du calcul sur machine a dans une certaine mesure donné raison à Brouwer. Personne ne croit qu'on puisse implémenter sous forme d'un algorithme général le théorème des valeurs intermédiaires. Par contre le théorème fondamental de l'algèbre est maintenant implémenté selon les canons de la rigueur de Brouwer.

Le programme qui calcule les zéros d'un polynôme unitaire à coefficients complexes commence par demander trois choses :

— la précision souhaitée sur le résultat,

— le degré du polynôme (supposé unitaire),

— une majoration (du module) des coefficients du polynôme.

En fonction des réponses que vous lui fournissez, le programme vous demande de donner les coefficients du polynôme avec une certaine précision qu'il vous impose. Notons $\mathbb{D}_2$ l'ensemble des nombres rationnels qui ont une écriture finie en base 2. Vous donnez au programme ce qu'il vous demande, la liste des n valeurs approchées des coefficients, sous forme d'éléments de $\mathbb{D}_2 + i\mathbb{D}_2$, et le programme calcule une liste de n éléments de $\mathbb{D}_2 + i\mathbb{D}_2$, qui approche la liste des zéros, dans un certain ordre, avec la précision souhaitée.

Le seul inconvénient d'un tel algorithme, mais cet inconvénient est incontournable, est que l'ordre dans lequel sont données les « racines approchées » peut dépendre de la façon dont vous répondez à la dernière question « les coefficients avec une précision prescrite » : il n'y a pas de manière canonique de donner l'information correcte.

Chapitre 2

Analyse de preuves. Le pgcd

Introduction

Dans ce chapitre nous analysons un théorème non évident concernant les entiers naturels, qui sont les plus simples des objets mathématiques. Il s'agit du théorème qui affirme que deux entiers > 0 ont pour plus grand commun diviseur un élément qui vérifie deux propriétés *a priori* inattendues.

Les présupposés pour ce théorème sont la maitrise des opérations arithmétiques élémentaires : addition, multiplication, soustraction et division. Notamment : pour a, b entiers positifs, le quotient q et le reste r de la division de a par b sont deux entiers qui vérifient : $0 \leq r < b$ et $a = bq + r$.

Pour les commentaires et les preuves nous utilisons le langage algébrique aujourd'hui familier. Nous introduisons même des produits de matrices pour éclaircir la situation, sachant que nous nous adressons à des lectrices qui connaissent bien cela.

Euclide arrivait aux mêmes conclusions (au moins de manière implicite) sans le recours aux facilités algébriques que nous nous accordons.

2.1 L'anthyphérèse

Il semble utile de rappeler que chez Euclide, ce que nous appelons l'algorithme d'Euclide est aussi une méthode géométrique pour trouver une plus grande commune mesure, si elle existe, à deux grandeurs de même nature.

Considérons par exemple deux segments de droites, AB et CD qui admettent pour commune mesure un segment EF, ce qui signifie que l'unité EF est contenue un nombre entier de fois dans AB et CD. Supposant AB $>$ CD, on commence par retrancher autant de fois qu'il est possible le segment CD du segment AB. S'il ne reste rien, c'est que CD était une commune mesure, et c'est manifestement la plus grande possible. S'il reste quelque chose, que nous notons GH, alors toute commune mesure à AB et CD est aussi une commune mesure à CD et GH. On peut donc continuer le processus. Comme GH est strictement plus petit que CD, le nombre de fois que EF est contenu dans GH est inférieur au nombre de fois qu'il est contenu dans CD. Ainsi le processus s'arrête après un nombre fini d'étapes et fournit une commune mesure, nécessairement multiple entier de EF.

Voici donc un résultat qui n'était pas *a priori* évident : la commune mesure trouvée est multiple de toute autre commune mesure.

Si maintenant on raisonne avec les nombres entiers a et b qui mesurent AB et CD par rapport à l'unité EF, le processus devient un calcul avec des entiers positifs qui fournit un commun diviseur g de a et b, et tout autre diviseur commun de a et b divise g, donc est plus petit que g.

Ceci démontre que le plus grand commun diviseur de a et b est multiple de tout autre diviseur commun.

Ce processus de soustractions alternées (on retranche autant de fois qu'on peut CD de AB, puis autant de fois qu'on peut GH de CD, puis etc.) s'appelle l'*anthyphérèse* dans les textes anciens.

Plus que pour chercher une commune mesure quand il en existe une, ce procédé était surtout utilisé pour montrer l'impossibilité d'une commune mesure pour deux grandeurs données, lorsqu'on montre que l'anthyphérèse ne peut aboutir en un nombre fini d'étapes.

Voyons ceci sur l'exemple du côté et de la diagonale du carré (figure 2.1.1).

On démarre avec le côté et la diagonale d'un carré ABCD. Le cercle de centre A passant par B coupe la diagonale AC en E, de sorte que AB = AE. On retranche le côté AB de la diagonale AC et l'on obtient EC. On considère alors le carré EFGC. On a FB = FE = EC, la première égalité parce que FB et FE sont les deux tangentes au cercles menées depuis le point F. Ainsi une commune mesure à AB et AC est aussi une commune mesure à AB et EF, donc à BC et BF = EF, donc à FC et EF : la diagonale et le côté du carré EFGC. Le processus va se répéter à l'identique en remplaçant le carré ABCD par le carré EFGC. Comme le côté EF est moindre que la moitié du côté AB, les côtés des carrés successifs deviendront moindres que tout segment donné par avance (axiome d'Archimède), et ceci montre qu'une commune mesure à AB et AC est impossible.

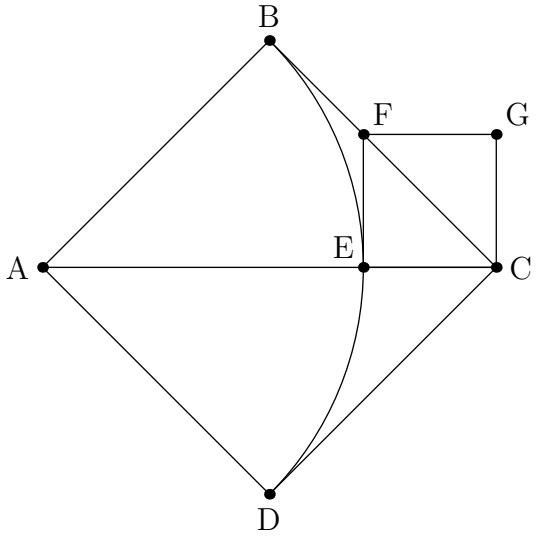


FIGURE 2.1.1 – L'anthyphérèse de la diagonale et du côté du carré

2.2 Le théorème du pgcd

Le théorème qui nous occupe est le suivant.

Théorème 2.2.1. *Si a et b sont deux entiers > 0, le plus grand diviseur commun $g > 0$ de a et b est multiple de tout autre diviseur commun.*

En fait on a découvert plus tard que le théorème 2.2.1 résulte du théorème suivant, qui concentre toute la difficulté du problème dans un énoncé simple, mais plus fort.

Théorème 2.2.2. *Si a et b sont deux entiers > 0, il existe un entier $g > 0$ de la forme $ua + vb$ avec $u, v \in \mathbb{Z}$, qui divise a et b.*

L'égalité $g = ua + vb$ s'appelle aussi une relation de Bézout entre a et b.

Démonstration que le théorème 2.2.2 implique le théorème 2.2.1. Soit h un diviseur commun de a et b : $a = ha_1$, $b = hb_1$ donc $g = au + bv = h(a_1u + b_1v)$. Ainsi g est à la fois supérieur ou égal à h et multiple de h. Le théorème 2.2.1 est bien satisfait. □

Le théorème doit donc être compris comme affirmant deux propriétés inattendues du plus grand des diviseurs communs à a et b.

— La première est que tout diviseur commun à a et b doit diviser g.

— La seconde est que g peut s'écrire sous la forme $ua + vb$ avec $u, v \in \mathbb{Z}$.

Remarque. En fait si l'on désigne par $x \preceq y$ la relation « x divise y », qui est une relation d'ordre partiel sur $\mathbb{N}^*$, on a l'équivalence

$$x \preceq g \iff x \preceq a \text{ et } x \preceq b.$$

Cela signifie : « g est le plus grand des minorants communs de a et b, au sens de la relation $\preceq$ ». Il s'agit donc, pour la relation de divisibilité, d'une *borne inférieure pour l'ensemble* $\{a, b\}$.

Ainsi, avec une relation d'ordre partiel, une borne inférieure d'une partie finie d'un ensemble peut exister sans appartenir à la partie. Pour une relation d'ordre total, ce phénomène ne se produit qu'avec des parties infinies.

Nous envisageons maintenant deux preuves du théorème 2.2.2.

2.3 Une preuve abstraite classique

Tout d'abord on établit le lemme suivant :

Lemme 2.3.1. *Toute partie non vide de* $\mathbb{N}$ *admet un plus petit élément.*

Preuve du lemme. Si une partie V de $\mathbb{N}$ contient un élément c alors le plus petit élément de l'ensemble fini $\mathrm{V} \cap [\![0, c]\!]$ est aussi le plus petit élément de V. □

Preuve abstraite du théorème 2.2.2. On considère l'ensemble

$$\mathrm{V}(a, b) = \mathrm{V} = (a\mathbb{Z} + b\mathbb{Z}) \cap \mathbb{N}^* = \{x > 0 \mid \exists u, v \in \mathbb{Z}, x = ua + vb\}.$$

L'ensemble V contient a et b : il est non vide. Soit $g = ua + vb$ le plus petit élément de V. Montrons par l'absurde qu'il divise a et b. Si ce n'était pas le cas, on aurait par exemple : g ne divise pas a. Alors en divisant a par g on obtiendrait un reste $r \in \{1, \ldots, g-1\}$ avec $a = gq + r$. On aurait alors

$$r = a - gq = a(1 - uq) + (-vq)b$$

et ceci montre que $r \in \mathrm{V}$, ce qui contredit le fait que g est le plus petit élément de V. □

2.4 Une preuve par algorithme

Algorithme 2.4.1. ***Algorithme de calcul du pgcd.***
Entrée: Deux entiers naturels a et b, > 0.
Sortie: Leur pgcd g.
Début
boucle
Tant que $b \neq 0$ **faire**
Remplacer a et b par : b et le reste de la division de a par b ;
fin tant que ;
fin de boucle
$g \leftarrow a$
Fin.

On considère l'algorithme 2.4.1 qui est à peu près celui exposé par Euclide.

Nous affirmons que pour n'importe quelles valeurs entrées pour a et b, l'algorithme fournit un élément g qui satisfait les propriétés requises dans le théorème 2.2.2.

Sous une forme un peu plus précise, et sans détruire le contenu des identificateurs a et b donnés en entrée, l'algorithme se réécrit comme dans l'encadré 2.4.1 *bis*.

Algorithme 2.4.1 *bis. Algorithme de calcul du pgcd, plus précis.*
Entrée: Deux entiers naturels a et b, > 0.
Sortie: Leur pgcd g.
Variables locales: a', b', b'' : entiers ≥ 0 ;
Début
initialisation
$a' \leftarrow a$; $b' \leftarrow b$;
boucle
Tant que $b' > 0$ **faire**
$b'' \leftarrow$ reste de la division de a' par b' ;
$a' \leftarrow b'$; $b' \leftarrow b''$;
fin tant que ;
fin de boucle
$g \leftarrow a'$
Fin.

Preuve de terminaison. Nous vérifions tout d'abord que l'algorithme 2.4.1 *bis* se *termine* bien après un nombre fini d'étapes. Cela tient à ce que, à chaque exécution de la boucle, b' est remplacé par un nombre strictement plus petit dans $\mathbb{N}$. Il atteint donc nécessairement la valeur 0 (ce qui est le seul moyen de sortir de la boucle) en un nombre fini d'étapes.

Il faut également noter que l'instruction « $b'' \leftarrow$ reste de la division de a' par b' » est toujours exécutée avec un $b' > 0$ et qu'il n'y aura donc pas d'erreur (la division par 0 n'est pas définie) lors de l'exécution du programme. □

Preuve de correction. Nous démontrons maintenant que l'algorithme fournit bien un résultat *correct*. Notons a_k et b_k $(k = 0, 1, \ldots, n)$ les valeurs successives prises par a' et b' chaque fois qu'on fait le test « $b' > 0$? » en début de boucle. On démarre avec $a_0 = a$ et $b_0 = b$. Si $b_k > 0$ la boucle est exécutée et l'on obtient $a_{k+1} = b_k$ et $b_{k+1} = a_k - q_k b_k$, en notant q_k le quotient dans la division. Alors on a par un calcul immédiat l'équivalence :

$$\forall x \in \mathbb{N}\ ((x \text{ divise } a_k \text{ et } b_k) \iff (x \text{ divise } a_{k+1} \text{ et } b_{k+1}))$$

On a donc pour tout $x \in \mathbb{N}$:

$$(x \text{ divise } a_0 \text{ et } b_0) \iff (x \text{ divise } a_1 \text{ et } b_1) \iff \cdots \iff (x \text{ divise } a_n \text{ et } b_n)$$

La dernière valeur du couple (a', b') est (a_n, b_n) avec $a_n > 0$ et $b_n = 0$. Or $a_n = b_{n-1}$ et $b_n = 0$ est le reste de la division de a_{n-1} par b_{n-1}. Cela signifie que a_n divise a_{n-1} et b_{n-1}, donc divise a et b. Par ailleurs on a de proche en proche le fait que a_k et b_k sont tous de la forme $u_k a + v_k b$ avec $u_k, v_k \in \mathbb{Z}$. Ainsi $g = a_n$, qui est la sortie donnée par l'algorithme, vérifie les conclusions du théorème 2.2.2. □

Notons qu'il est facile de compléter l'algorithme pour qu'il fournisse des entiers relatifs u et v vérifiant $ua + vb = g$, conformément à la preuve qui est donnée ci-dessus.

Pour $k = 0, 1, \ldots, n-1$ on peut en effet écrire :

$$\begin{bmatrix} a_{k+1} \\ b_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -q_k \end{bmatrix} \begin{bmatrix} a_k \\ b_k \end{bmatrix}$$

et cela donne :

$$\begin{bmatrix} a_k \\ b_k \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -q_{k-1} \end{bmatrix} \cdots \begin{bmatrix} 0 & 1 \\ 1 & -q_1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & -q_0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

On peut donc écrire :

$$\begin{bmatrix} a_k \\ b_k \end{bmatrix} = \begin{bmatrix} u_k & v_k \\ u_{k+1} & v_{k+1} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

avec la relation récurrente que l'on va utiliser dans l'algorithme :

$$\begin{bmatrix} u_k & v_k \\ u_{k+1} & v_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -q_{k-1} \end{bmatrix} \begin{bmatrix} u_{k-1} & v_{k-1} \\ u_k & v_k \end{bmatrix}.$$

Dans le nouvel algorithme, la matrice $\begin{bmatrix} u & v \\ u' & v' \end{bmatrix}$ prend les valeurs successives de $\begin{bmatrix} u_k & v_k \\ u_{k+1} & v_{k+1} \end{bmatrix}$.

Puisque $\begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ on a l'initialisation sans problème. On obtient :

Algorithme 2.4.2. ***Algorithme de calcul du pgcd et d'une relation de Bézout.***
Entrée: Deux entiers naturels a et b, > 0.
Sortie: Leur pgcd g ainsi que deux entiers relatifs u et v vérifiant $ua + vb = g$.
Variables locales: $a', b', b'', u, u', u'', v, v', v''$: entiers relatifs ;
Début
initialisation
$a' \leftarrow a$; $b' \leftarrow b$; $u \leftarrow 1$; $v \leftarrow 0$; $u' \leftarrow 0$; $v' \leftarrow 1$;
c'est-à-dire $\begin{bmatrix} a' \\ b' \end{bmatrix} \leftarrow \begin{bmatrix} a \\ b \end{bmatrix}$ et $\begin{bmatrix} u & v \\ u' & v' \end{bmatrix} \leftarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
boucle
Tant que $b' \neq 0$ **faire**
$(q, b'') \leftarrow$ quotient et reste de la division de a' par b' ;
$a' \leftarrow b'$; $b' \leftarrow b''$;
$u'' \leftarrow u - qu'$; $u \leftarrow u'$; $u' \leftarrow u''$;
$v'' \leftarrow v - qv'$; $v \leftarrow v'$; $v' \leftarrow v''$;
c'est-à-dire $\begin{bmatrix} a' \\ b' \end{bmatrix} \leftarrow \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} \begin{bmatrix} a' \\ b' \end{bmatrix}$ et $\begin{bmatrix} u & v \\ u' & v' \end{bmatrix} \leftarrow \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} \begin{bmatrix} u & v \\ u' & v' \end{bmatrix}$
fin tant que ;
fin de boucle
$g \leftarrow a'$
Fin.

2.5 Comparaison des deux preuves

On peut comparer les deux preuves proposées de différents points de vue. Élégance, rigueur, simplicité, facilité de compréhension, conviction de l'interlocuteur, effectivité du résultat annoncé.

Bien que ces critères, hormis le dernier, soient très subjectifs, et que les réponses dépendent beaucoup de notre éducation mathématique ils sont néanmoins très importants.

Une grande partie de l'activité de recherche en mathématiques consiste à essayer de simplifier ce qui a déjà été démontré. Souvent la première preuve trouvée pour un résultat important est obscure pour la plupart. La science mathématique ne pourrait pas progresser réellement sans l'activité continue de simplification des preuves.

La première preuve est appelée ici « classique » car c'est celle que l'on trouve désormais le plus souvent exposée [1].

Elle est particulièrement rapide et donne le sentiment d'aller droit au but. On peut la voir comme le résultat d'un effort de simplification de la preuve originelle, qui, dans la mesure où elle peut être pointée historiquement, ressemble beaucoup plus à la preuve par algorithme.

L'argument aussi est dans une certaine mesure plus simple que dans la deuxième preuve. Si l'on n'avait pas déjà la démonstration d'Euclide en tête, il aurait fallu de l'imagination pour trouver l'argument décisif :

$$\forall x \in \mathbb{N}\ ((x \text{ divise } a_k \text{ et } b_k) \iff (x \text{ divise } a_{k+1} \text{ et } b_{k+1})).$$

Par contre la première preuve ne semble pas fournir d'algorithme pour le calcul de u, v et g dont on affirme qu'ils existent. Certainement avant Cantor on n'aurait jamais écrit quelque chose qui ressemble à cela. Il faut une certaine audace pour dire « je les ai trouvés » quand on ne dit pas comment concrètement on peut les avoir. Il faut aussi un certain culot pour considérer l'ensemble infini $\mathrm{V}(a,b) \subset \mathbb{N}$ comme objet central de la preuve et raisonner avec son plus petit élément.

A priori, si l'on cherche l'algorithme sous-jacent à la preuve abstraite on est tenté de penser qu'il va falloir examiner tous les éléments de $\mathrm{V}(a,b) \cap [\![0,a]\!]$ et prendre le plus petit d'entre eux. Mais a-t-on une manière certaine d'énumérer en un temps fini $\mathrm{V}(a,b) \cap [\![0,a]\!]$? Un peu de réflexion montre que oui. En effet $ua + vb = (u - qb)a + (v - qa)b$ et le remplacement de u par $u - qb$ permet de se limiter aux u tels que $|u| \leq b/2$. Ensuite v peut être calculé par division euclidienne, de façon à rendre $ua + vb$ minimum pour cette valeur de u, ce qui donne $|ua + vb| \leq |ua|/2$. On obtient alors en fin de compte quelque chose d'horriblement compliqué et inélégant comme procédure de calcul du pgcd. Si ce dernier est égal à 1 et qu'on n'a pas de chance il faudra un nombre incroyablement grand d'essais (de l'ordre de $a/2$) avant de le trouver.

Par contre l'algorithme d'Euclide est très efficace. On peut montrer en effet qu'il est le plus lent lorsque tous les quotients successifs sont égaux à 1, et dans ce cas le nombre d'étapes est à très peu près égal à $\ln a / \ln \phi$ où ϕ est le nombre d'or $(1 + \sqrt{5})/2$. Avec des nombres à 100 chiffres, cela fera au maximum 500 étapes de calcul, autant dire rien pour un ordinateur. Tandis que 10^{100} étapes pour le mauvais algorithme ne seront jamais exécutées dans l'univers s'il est bien tel que nous supposons le connaitre : le soleil et même la galaxie auront disparu bien avant.

2.6 La preuve classique cache-t-elle un algorithme ?

Tout d'abord on peut remarquer que l'algorithme d'Euclide peut être interprété à la lumière de la preuve abstraite de la manière suivante : en considérant les valeurs successives prises par a' on voit que l'algorithme parcourt un chemin dans l'ensemble $\mathrm{V}(a,b)$, chemin strictement décroissant, qui s'arrête lorsqu'il ne peut plus descendre. Cependant, le test d'arrêt n'est pas directement lié à la certitude immédiate d'avoir atteint un diviseur commun à a et b. D'ailleurs, comme nous l'avons déjà remarqué la preuve de correction de l'algorithme, quoique assez simple, n'est pas absolument évidente et réclame un peu d'imagination.

Maintenant examinons la preuve classique avec cette idée de parcourir un chemin en descendant strictement dans l'ensemble $\mathrm{V}(a,b)$ avec pour objectif de réaliser le théorème 2.2.2 : trouver un diviseur commun à a et b dans $\mathrm{V}(a,b)$.

1. En fait, dans un langage plus abstrait on trouve en général l'énoncé suivant : *tout idéal de $\mathbb{Z}$ est principal*, et on en déduit ensuite le théorème 2.2.2. Nous avons préféré ici donner un résumé de cette preuve classique pour mieux faire ressortir son caractère fulgurant.

De ce point de vue, nous voyons que *l'argument « par l'absurde » qui termine la preuve classique, peut, comme très souvent, être vu comme le renversement d'un argument direct.*

L'argument direct est : je vous dis comment descendre dans $\mathrm{V}(a,b)$ tant que vous n'avez pas atteint votre objectif (un diviseur commun à a et b dans l'ensemble $\mathrm{V}(a,b)$).

Son renversement en une preuve par l'absurde donne : plaçons-nous au minimum de $\mathrm{V}(a,b)$ et montrons, par l'absurde, que notre objectif est atteint.

Cette preuve indirecte qui commence par un coup de force (plaçons-nous au minimum de V) et semble se terminer par l'absurde peut être remise sous forme d'une preuve directe qui donne le moyen de calculer g, comme suit. On démarre un chemin descendant dans $\mathrm{V}(a,b)$ à partir d'un point arbitraire de cet ensemble. Supposons être arrivés au point c. Si c divise a et b on est content. Sinon, si c ne divise pas a on peut le remplacer par le reste de la division de a par c, et de même, si c ne divise pas b on peut le remplacer par le reste de la division de b par c. On continue tant qu'on n'a pas atteint l'objectif.

Notez que l'on réalise ainsi de manière certaine l'objectif « trouver un diviseur commun à a et b dans l'ensemble $\mathrm{V}(a,b)$ » mais que l'on ne se soucie pas de savoir si le nombre g obtenu est le minimum de $\mathrm{V}(a,b)$. Cet « autre » objectif est atteint, comme dans l'algorithme d'Euclide, sans avoir été recherché en tant que tel.

En suivant l'idée informelle de l'algorithme il y a beaucoup de chemins qui s'ouvrent à nous pour descendre dans $\mathrm{V}(a,b)$: bien souvent, c ne divisera ni a ni b et l'on aura donc le choix entre les deux divisions. Par exemple cela pourrait donner l'algorithme 2.6.1.

Algorithme 2.6.1. ***Algorithme de calcul du pgcd, implicite dans la preuve classique.***
Entrée: Deux entiers naturels a et b, > 0.
Sortie: Leur pgcd g.
Variables locales: r, r' : entiers ≥ 0 ;
Début
initialisation
$g \leftarrow \inf(a,b)$;
boucle
Répéter
$r \leftarrow$ le reste de la division de a par g ;
Si $r > 0$ **alors** $g \leftarrow r$ **fin si** ;
$r' \leftarrow$ le reste de la division de b par g ;
Si $r' > 0$ **alors** $g \leftarrow r'$ **fin si** ;
jusqu'à ce que $r = r' = 0$ # fin de boucle
Fin.

Preuve de terminaison. Nous vérifions tout d'abord que l'algorithme s'arrête bien après un nombre fini d'étapes. Cela tient à ce que, à chaque exécution de la boucle, g est remplacé par un nombre strictement plus petit dans $\mathbb{N}^*$. On sortira donc nécessairement de la boucle après un nombre fini d'étapes.

Il faut également noter que les instructions

— « $r \leftarrow$ le reste de la division de a par g » et

— « $r' \leftarrow$ le reste de la division de b par g »

sont toujours exécutées avec un $g > 0$ et qu'il n'y aura donc pas d'erreur lors de l'exécution du programme. □

Preuve de correction. Au moment où l'on sort de la boucle, on a $g > 0$ qui divise a et b. En outre g reste dans l'ensemble $a\mathbb{Z} + b\mathbb{Z}$ chaque fois qu'il est réaffecté (calcul immédiat). L'algorithme réalise bien le théorème 2.2.2. □

Insistons sur le fait que la preuve de correction est plus simple que celle donnée pour l'algorithme d'Euclide, ce qui reflète la très grande simplicité de la preuve classique (de laquelle on a dérivé l'algorithme).

Enfin notons qu'on peut imaginer de nombreuses variantes, comme par exemple l'algorithme 2.6.2, où l'on essaie aléatoirement la division par a ou b.

Algorithme 2.6.2. ***Algorithme de calcul du pgcd, implicite dans la preuve classique, variante.***
Entrée: Deux entiers naturels a et b, > 0.
Sortie: Leur pgcd g.
Variables locales: r, r', c : entiers ≥ 0 ; $alea$: choisi à chaque fois au hasard entre a et b ;
Début
initialisation
$g \leftarrow alea$;
boucle
Répéter
$c \leftarrow alea$; $r \leftarrow$ le reste de la division de c par g ;
Si $r > 0$ **alors** $g \leftarrow r$
sinon $c \leftarrow a + b - c$; $r' \leftarrow$ le reste de la division de c par g ;
Si $r' > 0$ **alors** $g \leftarrow r'$ **fin si**
fin si ;
jusqu'à ce que $r = r' = 0$
fin de boucle
Fin.

Le lecteur est invité à faire tourner à la main l'algorithme d'Euclide et les deux algorithmes 2.6.1 et 2.6.2 issus de la preuve classique, par exemple avec les nombres 432 et 699 et à comparer leurs performances.

Sans doute en règle générale l'algorithme d'Euclide devrait être plus rapide car les divisions se font entre deux nombres qui diminuent simultanément, et sont donc de plus en plus rapides à exécuter. Mais dans certains cas un des deux autres algorithmes peut demander beaucoup moins d'étapes.

Il reste que la preuve basée sur l'algorithme 2.6.1 est en définitive « plus simple » que celle basée sur l'algorithme d'Euclide. Lorsque l'algorithme termine, il certifie par la même occasion que le nombre g obtenu divise bien a et b, ce qui n'était pas le cas pour l'algorithme d'Euclide.

Donnons pour terminer « une bonne rédaction » de la preuve classique. Cette rédaction évite tout aussi bien les difficultés inhérentes à l'emploi du lemme 2.3.1 que celles inhérentes aux preuves par l'absurde. Nous noterons $\mathrm{Rst}(x, y)$ le reste de la division euclidienne de x par y lorsque $x, y \in \mathbb{N}^*$.

Démonstration élégante et constructive. Remarquons que si nous trouvons un élément g dans $\mathrm{V}(a, b) = \mathrm{V}$ qui divise a et b, alors tout autre élément de V sera multiple de g. En effet, si $a = ga'$ et $b = gb'$ alors un z arbitraire de V s'écrit $z = ua + vb = g(ua' + vb')$. *A fortiori*, g doit être le plus petit élément de V.

Définissons alors une suite g_n par récurrence comme suit.

— $g_0 = \inf(a, b)$. On remarque que $g_0 \in \mathrm{V}$.

— Passage de n à $n+1$:
  — Si g_n divise a et b, on arrête la suite (ou, au choix, on pose $g_{n+1} = g_n$).
  — Si g_n ne divise pas a, on pose $g_{n+1} = \mathrm{Rst}(a, g_n)$. On remarque que $g_{n+1} \in \mathrm{V}$ et $g_{n+1} < g_n$.
  — Si g_n divise a mais ne divise pas b, on pose $g_{n+1} = \mathrm{Rst}(b, g_n)$. On remarque que $g_{n+1} \in \mathrm{V}$ et $g_{n+1} < g_n$.

Cette suite est bien définie, et tous ses termes sont dans V. Tant qu'on n'aboutit pas au premier cas, la suite est strictement décroissante, donc on est certain d'aboutir au premier cas. □

Chapitre 3

Les entiers naturels

Introduction

Dans ce chapitre nous essayons de mieux appréhender les êtres mathématiques les plus simples, à savoir les nombres entiers, aujourd'hui appelés « entiers naturels » par opposition aux « entiers relatifs ».

Nous écrivons $x \in \mathbb{N}$ comme abréviation du discours pour « x est un entier naturel », sans rien présupposer concernant la nature de « l'ensemble infini des entiers naturels ». Nous nous situons donc en quelque sorte avant la théorie des ensembles infinis.

Dans la section 3.1 nous donnons à lire un texte célèbre de Poincaré sur la nature du raisonnement mathématique. Il y décrit le raisonnement par récurrence comme le raisonnement mathématique par excellence, qui permet d'accéder à des vérités générales. Nous ferons quelques commentaires. Dans la section 3.2 nous donnons quelques exemples de raisonnements par récurrence et discutons leur pertinence. Dans la section 3.3 nous abordons brièvement la question de savoir si les preuves par algorithme sont fondamentalement de même nature que les preuves par récurrence. En particulier nous examinons dans quelle mesure les preuves données au chapitre 2 relèvent du raisonnement par récurrence.

Voici donc le chapitre premier de la première partie, *Le nombre et la grandeur* de l'ouvrage *La science et l'hypothèse* (Poincaré 1902, pages 9-28), cité selon l'édition de 1917 avec quelques amendements. Ce chapitre est une version corrigée et légèrement raccourcie d'un article de même titre (Poincaré 1894).

3.1 Henri Poincaré : *Sur la nature du raisonnement mathématique*

I

La possibilité même de la science mathématique semble une contradiction insoluble. Si cette science n'est déductive qu'en apparence, d'où lui vient cette parfaite rigueur que personne ne songe à mettre en doute ? Si, au contraire, toutes les propositions qu'elle énonce peuvent se tirer les unes des autres par les règles de la logique formelle, comment la mathématique ne se réduit-elle pas à une immense tautologie ? Le syllogisme ne peut rien nous apprendre d'essentiellement nouveau et, si tout devait sortir du principe d'identité, tout devrait aussi pouvoir s'y ramener. Admettra-t-on donc que les énoncés de tous ces théorèmes qui remplissent tant de volumes ne soient que des manières détournées de dire que A est A ?

Sans doute, on peut remonter aux axiomes qui sont à la source de tous les raisonnements. Si on juge qu'on ne peut les réduire au principe de contradiction, si on ne veut pas non plus y voir des faits expérimentaux qui ne pourraient participer à la nécessité mathématique, on a encore la ressource de les classer parmi les jugements synthétiques *a priori.* Ce n'est pas résoudre la difficulté, c'est seulement la baptiser ; et lors même que la nature des jugements synthétiques n'aurait plus pour nous de mystère, la contradiction ne se serait pas évanouie, elle n'aurait fait que reculer ; le raisonnement syllogistique reste incapable de rien ajouter aux données qu'on lui fournit ; ces données se réduisent à quelques axiomes et on ne devrait pas retrouver autre chose dans les conclusions.

Aucun théorème ne devrait être nouveau si dans sa démonstration n'intervenait un axiome nouveau ; le raisonnement ne pourrait nous rendre que les vérités immédiatement évidentes empruntées à l'intuition directe ; il ne serait plus qu'un intermédiaire parasite et dès lors n'aurait-on pas lieu de se demander si tout l'appareil syllogistique ne sert pas uniquement à dissimuler notre emprunt ?

La contradiction nous frappera davantage si nous ouvrons un livre quelconque de mathématiques ; à chaque page l'auteur annoncera l'intention de généraliser une proposition déjà connue. Est-ce donc que la méthode mathématique procède du particulier au général et comment alors peut-on l'appeler déductive ?

Si enfin la science du nombre était purement analytique, ou pouvait sortir analytiquement d'un petit nombre de jugements synthétiques, il semble qu'un esprit assez puissant pourrait d'un seul coup d'œil en apercevoir toutes les vérités ; que dis-je ! on pourrait même espérer qu'un jour on inventera pour les exprimer un langage assez simple pour qu'elles apparaissent ainsi immédiatement à une intelligence ordinaire.

Si l'on se refuse à admettre ces conséquences, il faut bien concéder que le raisonnement mathématique a par lui-même une sorte de vertu créatrice et par conséquent qu'il se distingue du syllogisme.

La différence doit même être profonde. Nous ne trouverons pas par exemple la clef du mystère dans l'usage fréquent de cette règle d'après laquelle une même opération uniforme appliquée à deux nombres égaux donnera des résultats identiques.

Tous ces modes de raisonnement, qu'ils soient ou non réductibles au syllogisme proprement dit, conservent le caractère analytique et sont par cela même impuissants.

II

Le débat est ancien ; déjà Leibniz cherchait à démontrer que 2 et 2 font 4 ; examinons un peu sa démonstration.

Je suppose que l'on ait défini le nombre 1 et l'opération $x+1$ qui consiste à ajouter l'unité à un nombre donné x.

Ces définitions, quelles qu'elles soient, n'interviendront pas dans la suite du raisonnement.

Je définis ensuite les nombres 2, 3 et 4 par les égalités

$$(1)\quad 1+1=2\,;\ (2)\quad 2+1=3\,;\ (3)\quad 3+1=4.$$

Je définis de même l'opération $x+2$ par la relation :

$$(4)\qquad\qquad x+2=(x+1)+1.$$

Cela posé nous avons :

$$\begin{aligned} 2+2&=(2+1)+1 && \text{(Définition 4)}\\ (2+1)+1&=3+1 && \text{(Définition 2)}\\ 3+1&=4 && \text{(Définition 3)} \end{aligned}$$

d'où

$$2+2=4 \qquad \text{C. Q. F. D.}$$

On ne saurait nier que ce raisonnement ne soit purement analytique. Mais interrogez un mathématicien quelconque : « Ce n'est pas une démonstration proprement dite, vous répondra-t-il, c'est une vérification ». On s'est borné à rapprocher l'une de l'autre deux définitions purement conventionnelles et on a constaté leur identité ; on n'a rien appris de nouveau. La *vérification* diffère précisément de la véritable démonstration, parce qu'elle est purement analytique et parce qu'elle est stérile. Elle est stérile parce que la conclusion n'est que la traduction des prémisses dans un autre langage. La démonstration véritable est féconde au contraire parce que la conclusion y est en un sens plus générale que les prémisses.

L'égalité $2 + 2 = 4$ n'a été ainsi susceptible d'une vérification que parce qu'elle est particulière. Tout énoncé particulier en mathématiques pourra toujours être vérifié de la sorte. Mais si la mathématique devait se réduire à une suite de pareilles vérifications, elle ne serait pas une science. Ainsi un joueur d'échecs, par exemple, ne crée pas une science en gagnant une partie. Il n'y a de science que du général.

On peut même dire que les sciences exactes ont précisément pour objet de nous dispenser de ces vérifications directes.

III

Voyons donc le géomètre à l'œuvre et cherchons à surprendre ses procédés.

La tâche n'est pas sans difficulté ; il ne suffit pas d'ouvrir un ouvrage au hasard et d'y analyser une démonstration quelconque.

Nous devons exclure d'abord la géométrie où la question se complique des problèmes ardus relatifs au rôle des postulats, à la nature et à l'origine de la notion d'espace. Pour des raisons analogues nous ne pouvons nous adresser à l'analyse infinitésimale. Il nous faut chercher la pensée mathématique là où elle est restée pure, c'est-à-dire en arithmétique.

Encore faut-il choisir ; dans les parties les plus élevées de la théorie des nombres, les notions mathématiques primitives ont déjà subi une élaboration si profonde, qu'il devient difficile de les analyser.

C'est donc au début de l'arithmétique que nous devons nous attendre à trouver l'explication que nous cherchons, mais il arrive justement que c'est dans la démonstration des théorèmes les plus élémentaires que les auteurs des traités classiques ont déployé le moins de précision et de rigueur. Il ne faut pas leur en faire un crime ; ils ont obéi à une nécessité ; les débutants ne sont pas préparés à la véritable rigueur mathématique ; ils n'y verraient que de vaines et fastidieuses subtilités ; on perdrait son temps à vouloir trop tôt les rendre plus exigeants ; il faut qu'ils refassent rapidement, mais sans bruler d'étapes, le chemin qu'ont parcouru lentement les fondateurs de la science.

Pourquoi une si longue préparation est-elle nécessaire pour s'habituer à cette rigueur parfaite, qui, semble-t-il, devrait s'imposer naturellement à tous les bons esprits ? C'est là un problème logique et psychologique bien digne d'être médité.

Mais nous ne nous y arrêterons pas ; il est étranger à notre objet ; tout ce que je veux retenir, c'est que, sous peine de manquer notre but, il nous faut refaire les démonstrations des théorèmes les plus élémentaires et leur donner non la forme grossière qu'on leur laisse pour ne pas lasser les débutants, mais celle qui peut satisfaire un géomètre exercé.

Définition de l'addition. — Je suppose qu'on ait défini préalablement l'opération $x + 1$ qui consiste à ajouter le nombre 1 à un nombre donné x.

Cette définition, quelle qu'elle soit d'ailleurs, ne jouera plus aucun rôle dans la suite des raisonnements.

Il s'agit maintenant de définir l'opération $x + a$, qui consiste à ajouter le nombre a à un nombre donné x.

Supposons que l'on ait défini l'opération

$$x + (a - 1),$$

l'opération $x + a$ sera définie par l'égalité :

$$x + a = [x + (a - 1)] + 1. \tag{1}$$

Nous saurons donc ce que c'est que $x + a$ quand nous saurons ce que c'est que $x + (a - 1)$, et comme j'ai supposé au début que l'on savait ce que c'est que $x + 1$, on pourra définir successivement et « par récurrence » les opérations $x + 2$, $x + 3$, etc.

Cette définition mérite un moment d'attention, elle est d'une nature particulière qui la distingue déjà de la définition purement logique ; l'égalité (1) contient en effet une infinité de définitions distinctes, chacune d'elles n'ayant un sens que quand on connait celle qui la précède.

Propriétés de l'addition. — *Associativité.* — Je dis que

$$a + (b + c) = (a + b) + c.$$

En effet le théorème est vrai pour $c = 1$; il s'écrit alors

$$a + (b + 1) = (a + b) + 1$$

ce qui n'est autre chose, à la différence des notations près, que l'égalité (1) par laquelle je viens de définir l'addition.

Supposons que le théorème soit vrai pour $c = \gamma$, je dis qu'il sera vrai pour $c = \gamma + 1$, soit en effet

$$(a + b) + \gamma = a + (b + \gamma),$$

on en déduira successivement :

$$[(a + b) + \gamma] + 1 = [a + (b + \gamma)] + 1,$$

ou en vertu de la définition (1)

$$(a + b) + (\gamma + 1) = a + (b + \gamma + 1) = a + [b + (\gamma + 1)],$$

ce qui montre, par une série de déductions purement analytiques, que le théorème est vrai pour $\gamma + 1$.

Étant vrai pour $c = 1$, on verrait ainsi successivement qu'il l'est pour $c = 2$, pour $c = 3$, etc.

Commutativité. — 1° Je dis que

$$a + 1 = 1 + a.$$

Le théorème est évidemment vrai pour $a = 1$, on pourrait *vérifier* par des raisonnements purement analytiques que s'il est vrai pour $a = \gamma$, il le sera pour $a = \gamma + 1$; or il l'est pour $a = 1$, il le sera donc pour $a = 2$, pour $a = 3$, etc. ; c'est ce qu'on exprime en disant que la proposition énoncée est démontrée par récurrence.

2° Je dis que

$$a + b = b + a.$$

Le théorème vient d'être démontré pour $b = 1$, on peut *vérifier* analytiquement que s'il est vrai pour $b = \beta$, il le sera pour $b = \beta + 1$.

La proposition est donc établie par récurrence.

Définition de la multiplication. — Nous définirons la multiplication par les égalités.

$$a \times 1 = a$$

$$a \times b = [a \times (b - 1)] + a. \tag{2}$$

L'égalité (2) renferme comme l'égalité (1) une infinité de définitions ; ayant défini $a \times 1$ elle permet de définir successivement : $a \times 2$, $a \times 3$, etc.

Propriétés de la multiplication. — *Distributivité.* — Je dis que

$$(a+b)\times c = (a\times c)+(b\times c).$$

On vérifie analytiquement que l'égalité est vraie pour $c=1$; puis que si le théorème est vrai pour $c=\gamma$ il sera vrai pour $c=\gamma+1$.

La proposition est encore démontrée par récurrence.

Commutativité. — 1° Je dis que

$$a\times 1 = 1\times a.$$

Le théorème est évident pour $a=1$.

On vérifie analytiquement que s'il est vrai pour $a=\alpha$ il sera vrai pour $a=\alpha+1$.

2° Je dis que

$$a\times b = b\times a.$$

Le théorème vient d'être démontré pour $b=1$. On vérifierait analytiquement que s'il est vrai pour $b=\beta$ il le sera pour $b=\beta+1$.

IV

J'arrête là cette série monotone de raisonnements. Mais cette monotonie même a mieux fait ressortir le procédé qui est uniforme et qu'on retrouve à chaque pas.

Ce procédé est la démonstration par récurrence. On établit d'abord un théorème pour $n=1$; on montre ensuite que s'il est vrai de $n-1$, il est vrai de n et on en conclut qu'il est vrai pour tous les nombres entiers.

On vient de voir comment on peut s'en servir pour démontrer les règles de l'addition et de la multiplication, c'est-à-dire les règles du calcul algébrique ; ce calcul est un instrument de transformation qui se prête à beaucoup plus de combinaisons diverses que le simple syllogisme ; mais c'est encore un instrument purement analytique et incapable de rien nous apprendre de nouveau. Si les mathématiques n'en avaient pas d'autre elles seraient donc tout de suite arrêtées dans leur développement ; mais elles ont de nouveau recours au même procédé, c'est-à-dire au raisonnement par récurrence et elles peuvent continuer leur marche en avant.

À chaque pas, si on y regarde bien, on retrouve ce mode de raisonnement, soit sous la forme simple que nous venons de lui donner, soit sous une forme plus ou moins modifiée.

C'est donc bien là le raisonnement mathématique par excellence et il nous faut l'examiner de plus près.

V

Le caractère essentiel du raisonnement par récurrence c'est qu'il contient, condensés pour ainsi dire en une formule unique, une infinité de syllogismes.

Pour qu'on s'en puisse mieux rendre compte, je vais énoncer les uns après les autres ces syllogismes qui sont, si l'on veut me passer l'expression, disposés en cascade.

Ce sont bien entendu des syllogismes hypothétiques.

Le théorème est vrai du nombre 1.

Or s'il est vrai de 1, il est vrai de 2.

Donc il est vrai de 2.

Or s'il est vrai de 2, il est vrai de 3.

Donc il est vrai de 3, et ainsi de suite.

On voit que la conclusion de chaque syllogisme sert de mineure au suivant.

De plus les majeures de tous nos syllogismes peuvent être ramenées à une formule unique.

Si le théorème est vrai de $n-1$, il l'est de n.

On voit donc que, dans les raisonnements par récurrence, on se borne à énoncer la mineure du premier syllogisme, et la formule générale qui contient comme cas particuliers toutes les majeures.

Cette suite de syllogismes qui ne finirait jamais se trouve ainsi réduite à une phrase de quelques lignes.

Il est facile maintenant de comprendre pourquoi toute conséquence particulière d'un théorème peut, comme je l'ai expliqué plus haut, être vérifiée par des procédés purement analytiques.

Si au lieu de montrer que notre théorème est vrai de tous les nombres, nous voulons seulement faire voir qu'il est vrai du nombre 6 par exemple, il nous suffira d'établir les 5 premiers syllogismes de notre cascade ; il nous en faudrait 9 si nous voulions démontrer le théorème pour le nombre 10 ; il nous en faudrait davantage encore pour un nombre plus grand ; mais quelque grand que soit ce nombre nous finirions toujours par l'atteindre, et la vérification analytique serait possible.

Et cependant, quelque loin que nous allions ainsi, nous ne nous élèverions jamais jusqu'au théorème général, applicable à tous les nombres, qui seul peut être objet de science. Pour y arriver, il faudrait une infinité de syllogismes, il faudrait franchir un abime que la patience de l'analyste, réduit aux seules ressources de la logique formelle, ne parviendra jamais à combler.

Je demandais au début pourquoi on ne saurait concevoir un esprit assez puissant pour apercevoir d'un seul coup d'œil l'ensemble des vérités mathématiques.

La réponse est aisée maintenant ; un joueur d'échecs peut combiner quatre coups, cinq coups d'avance, mais, si extraordinaire qu'on le suppose, il n'en préparera jamais qu'un nombre fini ; s'il applique ses facultés à l'arithmétique, il ne pourra en apercevoir les vérités générales d'une seule intuition directe ; pour parvenir au plus petit théorème, il ne pourra s'affranchir de l'aide du raisonnement par récurrence parce que c'est un instrument qui permet de passer du fini à l'infini.

Cet instrument est toujours utile, puisque, nous faisant franchir d'un bond autant d'étapes que nous le voulons, il nous dispense de vérifications longues, fastidieuses et monotones qui deviendraient rapidement impraticables. Mais il devient indispensable dès qu'on vise au théorème général, dont la vérification analytique nous rapprocherait sans cesse, sans nous permettre de l'atteindre.

Dans ce domaine de l'arithmétique, on peut se croire bien loin de l'analyse infinitésimale, et, cependant, nous venons de le voir, l'idée de l'infini mathématique joue déjà un rôle prépondérant, et sans elle il n'y aurait pas de science parce qu'il n'y aurait rien de général.

VI

Le jugement sur lequel repose le raisonnement par récurrence peut être mis sous d'autres formes ; on peut dire par exemple que dans une collection infinie de nombres entiers différents, il y en a toujours un qui est plus petit que tous les autres.

On pourra passer facilement d'un énoncé à l'autre et se donner ainsi l'illusion qu'on a démontré la légitimité du raisonnement par récurrence. Mais on sera toujours arrêté, on arrivera toujours à un axiome indémontrable qui ne sera au fond que la proposition à démontrer traduite dans un autre langage.

On ne peut donc se soustraire à cette conclusion que la règle du raisonnement par récurrence est irréductible au principe de contradiction.

Cette règle ne peut non plus nous venir de l'expérience ; ce que l'expérience pourrait nous apprendre, c'est que la règle est vraie pour les dix, pour les cent premiers nombres par exemple, elle ne peut atteindre la suite indéfinie des nombres, mais seulement une portion plus ou moins longue mais toujours limitée de cette suite.

Or, s'il ne s'agissait que de cela, le principe de contradiction suffirait, il nous permettrait toujours de développer autant de syllogismes que nous voudrions, c'est seulement quand il s'agit d'en enfermer une infinité dans une seule formule, c'est seulement devant l'infini que ce principe échoue, c'est également là que l'expérience devient impuissante. Cette règle, inaccessible à la démonstration analytique et à l'expérience, est le véritable type du jugement synthétique *a priori.* On ne saurait d'autre part songer à y voir une convention, comme pour quelques-uns des postulats de la géométrie.

Pourquoi donc ce jugement s'impose-t-il à nous avec une irrésistible évidence ? C'est qu'il n'est

que l'affirmation de la puissance de l'esprit qui se sait capable de concevoir la répétition indéfinie d'un même acte dès que cet acte est une fois possible. L'esprit a de cette puissance une intuition directe et l'expérience ne peut être pour lui qu'une occasion de s'en servir et par là d'en prendre conscience.

Mais, dira-t-on, si l'expérience brute ne peut légitimer le raisonnement par récurrence, en est-il de même de l'expérience aidée de l'induction ? Nous voyons successivement qu'un théorème est vrai du nombre 1, du nombre 2, du nombre 3 et ainsi de suite, *la loi est manifeste*, disons-nous, et elle l'est au même titre que toute loi physique appuyée sur des observations dont le nombre est très grand, mais limité.

On ne saurait méconnaitre qu'il y a là une analogie frappante avec les procédés habituels de l'induction. Mais une différence essentielle subsiste. L'induction, appliquée aux sciences physiques, est toujours incertaine, parce qu'elle repose sur la croyance à un ordre général de l'Univers, ordre qui est en dehors de nous. L'induction mathématique, c'est-à-dire la démonstration par récurrence, s'impose au contraire nécessairement, parce qu'elle n'est que l'affirmation d'une propriété de l'esprit lui-même.

VII

Les mathématiciens, je l'ai dit plus haut, s'efforcent toujours de *généraliser* les propositions qu'ils ont obtenues, et pour ne pas chercher d'autre exemple, nous avons tout à l'heure démontré l'égalité :

$$a + 1 = 1 + a,$$

et nous nous en sommes servi ensuite pour établir l'égalité

$$a + b = b + a,$$

qui est manifestement plus générale.

Les mathématiques peuvent donc comme les autres sciences procéder du particulier au général.

Il y a là un fait qui nous aurait paru incompréhensible au début de cette étude, mais qui n'a plus pour nous rien de mystérieux, depuis que nous avons constaté les analogies de la démonstration par récurrence avec l'induction ordinaire.

Sans doute le raisonnement mathématique récurrent et le raisonnement physique inductif reposent sur des fondements différents, mais leur marche est parallèle, ils vont dans le même sens, c'est-à-dire du particulier au général.

Examinons la chose d'un peu plus près.

Pour démontrer l'égalité :

$$a + 2 = 2 + a, \tag{1}$$

il nous suffit d'appliquer deux fois la règle

$$a + 1 = 1 + a, \tag{2}$$

et d'écrire :

$$a + 2 = a + 1 + 1 = 1 + a + 1 = 1 + 1 + a = 2 + a.$$

L'égalité (1) ainsi déduite par voie purement analytique de l'égalité (2) n'en est pas cependant un simple cas particulier : elle est autre chose.

On ne peut donc même pas dire que dans la partie réellement analytique et déductive des raisonnements mathématiques, on procède du général au particulier, au sens ordinaire du mot.

Les deux membres de l'égalité (1) sont simplement des combinaisons plus compliquées que les deux membres de l'égalité (2) et l'analyse ne sert qu'à séparer les éléments qui entrent dans ces combinaisons et à en étudier les rapports.

Les mathématiciens procèdent donc « par construction », ils « construisent » des combinaisons de plus en plus compliquées. Revenant ensuite par l'analyse de ces combinaisons, de ces ensembles, pour ainsi dire, à leurs éléments primitifs, ils aperçoivent les rapports de ces éléments et en déduisent les rapports des ensembles eux-mêmes.

C'est là une marche purement analytique, mais ce n'est pas pourtant une marche du général au particulier, car les ensembles ne sauraient évidemment être regardés comme plus particuliers que leurs éléments.

On a attaché, et à juste titre, une grande importance à ce procédé de la « construction » et on a voulu y voir la condition nécessaire et suffisante des progrès des sciences exactes.

Nécessaire, sans doute, mais suffisante, non.

Pour qu'une construction puisse être utile, pour qu'elle ne soit pas une vaine fatigue pour l'esprit, pour qu'elle puisse servir de marchepied à qui veut s'élever plus haut, il faut d'abord qu'elle possède une sorte d'unité, qui permette d'y voir autre chose que la juxtaposition de ses éléments.

Ou plus exactement, il faut qu'on trouve quelque avantage à considérer la construction plutôt que ses éléments eux-mêmes.

Quel peut être cet avantage ?

Pourquoi raisonner sur un polygone par exemple, qui est toujours décomposable en triangles, et non sur les triangles élémentaires ?

C'est qu'il y a des propriétés que l'on peut démontrer pour les polygones d'un nombre quelconque de côtés et qu'on peut ensuite appliquer immédiatement à un polygone particulier quelconque.

Le plus souvent, au contraire, ce n'est qu'au prix des plus longs efforts qu'on pourrait les retrouver en étudiant directement les rapports des triangles élémentaires. La connaissance du théorème général nous épargne ces efforts.

Une construction ne devient donc intéressante que quand on peut la ranger à côté d'autres constructions analogues, formant les espèces d'un même genre.

Si le quadrilatère est autre chose que la juxtaposition de deux triangles, c'est qu'il appartient au genre polygone.

Encore faut-il qu'on puisse démontrer les propriétés du genre sans être forcé de les établir successivement pour chacune des espèces.

Pour y arriver, il faut nécessairement remonter du particulier au général, en gravissant un ou plusieurs échelons.

Le procédé analytique « par construction » ne nous oblige pas à en descendre, mais il nous laisse au même niveau.

Nous ne pouvons nous élever que par l'induction mathématique, qui seule peut nous apprendre quelque chose de nouveau. Sans l'aide de cette induction différente à certains égards de l'induction physique, mais féconde comme elle, la construction serait impuissante à créer la science.

Observons en terminant que cette induction n'est possible que si une même opération peut se répéter indéfiniment. C'est pour cela que la théorie du jeu d'échec ne pourra jamais devenir une science, puisque les différents coups d'une même partie ne se ressemblent pas.

Il est un peu vain de commenter un texte si bien écrit, nous essayons seulement d'en extraire quelques idées force.

Le syllogisme est une déduction formée de trois parties : deux prémisses (la mineure et la majeure) et une conclusion. En logique classique, on étudiait les différentes formes possibles des syllogismes. En mathématiques, Poincaré ne retient qu'une seule forme de syllogisme, celle qu'on appelle le *modus ponens* : dans celui-ci, la mineure est une proposition A, la majeure est la proposition que A implique une proposition B, et la conclusion est la proposition B. Il s'agit en fait exactement de la règle d'élimination de l'implication (notée ici $\to$) que nous verrons dans le chapitre 11 : $\dfrac{\mathrm{A} \qquad \mathrm{A} \to \mathrm{B}}{\mathrm{B}}$.

Le raisonnement par récurrence est le raisonnement mathématique par excellence car il permet d'accéder au général, à l'infini, ce que ne peuvent faire les raisonnements purement analytiques.

La certitude que confère ce raisonnement ne peut pas être réduite aux principes de la logique. Elle est quelque chose qui s'impose à l'esprit humain avec une force irrésistible, qui « voit » dans ce raisonnement un schéma de preuve qu'on peut répéter aussi loin que nécessaire pour obtenir une vérité générale concernant tous les entiers : d'une part P(1) est vraie, ensuite si P(1) est vraie alors P(2) est vraie, donc P(2) est vraie ; ensuite si P(2) est vraie alors P(3) est vraie, donc P(3) est vraie ; etc.

Notons que Poincaré ne présente pas cette étude comme une injonction à utiliser le raisonnement par récurrence chaque fois qu'on veut démontrer une propriété générale concernant les entiers naturels (parce qu'il serait lui seul rigoureux). Mais il affirme plutôt que, quant au fond, toute vérité générale concernant les entiers provient en dernière analyse de preuves qui se fondent sur la récurrence et non sur la seule logique (au sens de l'art général du raisonnement correct).

Poincaré affirme aussi que toute activité mathématique vraiment créative doit faire appel à ce type d'induction, nécessaire pour passer du particulier au général, et qui se distingue de l'induction en physique par son absence de toute référence au monde « extérieur à l'esprit humain ».

Considérons l'exemple de l'algèbre, dans laquelle l'activité mathématique a souvent pour fondement des familles d'identités algébriques.

Que nous apprend $(a+b)^2 = a^2 + b^2 + 2ab$? Rien, ou presque. Tout juste une égalité qui peut être parfois utilisée comme raccourci. Et en fait, on ne fait qu'appliquer sans réelle invention les axiomes des anneaux commutatifs.

Peut-on en dire autant de toute identité algébrique ? Pas vraiment puisqu'une identité « très grosse » ne peut pas être traitée sans la théorie générale qui nous dit que toute expression algébrique se ramène à une écriture normalisée sous forme d'un polynôme, écriture normalisée grâce à laquelle on peut réduire à un simple calcul mécanique la vérification de l'identité voulue. Maintenant cette théorie générale des expressions algébriques est fondée en dernière analyse sur un raisonnement par récurrence sur la taille de ces expressions. Et sans ce raisonnement on n'aurait pas pu concevoir les logiciels qui nous facilitent la vie en exécutant sans erreur les calculs algébriques trop gros pour nos feuilles de papier ou pour notre habileté.

Même une fois cette question réglée, nous voyons que l'algèbre ne peut pas se réduire à ce que nous avons mis dans les logiciels de calcul formel.

Le théorème de Cayley-Hamilton par exemple est un théorème général qui ne peut pas être démontré par un « logiciel de calcul algébrique ». Il revient certes à affirmer pour chaque taille de matrice carrée les identités algébriques correspondant à son énoncé : tel calcul aboutit à une matrice identiquement nulle. On peut donc « vérifier » par un calcul purement analytique le théorème pour les matrices de taille 3, 4, 5, etc., mais il faut nécessairement inventer quelque chose pour avoir accès au théorème général. Et cette invention, vraiment créatrice de neuf, reposera en dernière analyse sur quelque chose de même nature que le raisonnement par récurrence.

3.2 Exemples de raisonnements par récurrence

3.2.1 Des exemples (trop) simples

Il n'est pas si facile de trouver des exemples simples pour lesquels le raisonnement par récurrence s'impose comme le plus avantageux.

La somme des n premiers entiers

On propose souvent à titre d'exemple de démontrer par récurrence sur $n \geq 1$ que

$$1 + 2 + \cdots + n = n(n+1)/2.$$

Ce cas d'école n'est en fait pas très convaincant, car « comment a-t-on deviné le résultat ? ».

Pour établir cette égalité il vaut beaucoup mieux écrire, comme le fit Gauss pour $n = 100$:

$$\begin{array}{rcccccccccc} 2\times(1+\cdots+100) & = & 1 & + & 2 & + & \cdots & + & 99 & + & 100 \\ & + & 100 & + & 99 & + & \cdots & + & 2 & + & 1 \\ & = & 101 & + & 101 & + & \cdots & + & 101 & + & 101 \\ & = & 101 & \times & 100 & & & & & & \end{array}$$

Sommations finies

Dans le même style, on peut proposer de démontrer par récurrence que

$$1^3 + 2^3 + \cdots + n^3 = n^2(n+1)^2/4.$$

Mais si cette égalité est découverte empiriquement de façon naturelle, il n'en va pas de même pour une somme $\mathrm{P}(1) + \mathrm{P}(2) + \cdots + \mathrm{P}(n)$, où P est un polynôme arbitraire. Dans un tel cas, on ne peut pas deviner le résultat, et il est nécessaire de développer une « théorie générale » de ces sommes finies.

Pascal a montré sur un exemple une méthode qui peut s'appliquer dans tous les cas. En termes modernes on peut la présenter comme ceci.

Si le polynôme $\mathrm{P}(x)$ a son monôme de plus haut degré de la forme ax^k, c'est-à-dire si $\mathrm{P}(x) = ax^k + \mathrm{R}_1(x)$ avec $\deg \mathrm{R}_1 < k$, on veut trouver un polynôme $\mathrm{T}(x)$ qui vérifie $\mathrm{T}(n) - \mathrm{T}(n-1) = \mathrm{P}(n)$ et $\mathrm{T}(0) = 0$. On considère $\mathrm{Q}(x) = \dfrac{ax^{k+1}}{k+1}$. Alors

$$\mathrm{Q}(n) - \mathrm{Q}(n-1) = an^k + \mathrm{S}_1(n) = \mathrm{P}(n) + (\mathrm{S}_1 - \mathrm{R}_1)(n)$$

avec $\deg \mathrm{R}_1 < k$. On est donc ramené à résoudre le même problème avec le polynôme $\mathrm{P}_1 = \mathrm{S}_1 - \mathrm{R}_1$, qui est de degré $< k$.

Mais il est significatif que ce n'est pas pour la solution de ce problème que Pascal a mis en forme le raisonnement par récurrence moderne. L'exemple de sommation qu'il traitait avec un polynôme de faible degré était à ses yeux suffisamment général pour ne pas réclamer de preuve plus convaincante. C'est seulement à propos d'une propriété subtile du triangle arithmétique (appelé aujourd'hui triangle de Pascal) qu'il a senti la nécessité d'un nouveau type de raisonnement.

Signalons que la solution du problème des sommes finies peut d'ailleurs être améliorée du point de vue du calcul en considérant les polynômes

$$\mathrm{F}_0(x) = \binom{x}{0} = 1, \quad \mathrm{F}_1(x) = \binom{x}{1} = x, \quad \mathrm{F}_2(x) = \binom{x}{2} = \frac{x(x-1)}{2},$$
$$\mathrm{F}_k(x) = \binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!} \ (k \geq 3).$$

Le calcul donne

$$\mathrm{F}_k(p+1) - \mathrm{F}_k(p) = \mathrm{F}_{k-1}(p), \quad \text{et} \quad \mathrm{F}_k(0) = 0 \text{ pour } k > 0,$$

de sorte que pour $k > 0$

$$\begin{aligned} \sum_{n=0}^{p-1} \mathrm{F}_{k-1}(n) &= (\mathrm{F}_k(1) - \mathrm{F}_k(0)) + (\mathrm{F}_k(2) - \mathrm{F}_k(1)) + \cdots + (\mathrm{F}_k(p) - \mathrm{F}_k(p-1)) \\ &= \mathrm{F}_k(p). \end{aligned} \qquad (*)$$

Si par exemple un polynôme P de degré 5 s'écrit

$$\mathrm{P}(x) = a_0 + a_1\mathrm{F}_1(x) + \cdots + a_5\mathrm{F}_5(x)$$

on aura

$$\sum_{n=0}^{p-1} \mathrm{P}(n) = a_0\mathrm{F}_1(p) + a_1\mathrm{F}_2(p) + \cdots + a_5\mathrm{F}_6(p).$$

Quant à la question de trouver facilement les coefficients a_i à partir de P, elle a été résolue de manière élégante par ce qu'on appelle *le tableau des différences finies* (invention parfois attribuée à Newton : on ne prête qu'aux riches). Par exemple le tableau des différences finies pour $\mathrm{P}(n) = n^3 - 2n^2 + 3$ se présente comme suit :

	0		1		2		3		
		$\ddots$		$\ddots$		$\ddots$		$\ddots$	
deg = 3			**3**		**2**		**3**		**12**
deg = 2				**−1**		**1**		**9**	
deg = 1					**2**		**8**		
deg = 0						**6**			

Sur la première ligne « deg = k » du tableau des différences finies d'un polynôme P de degré k on écrit les valeurs de $\mathrm{P}(n)$, pour $n = 0, \ldots, k$. Chacune des lignes en dessous, du degré $k-1$ jusqu'au degré 0 se calcule par différence des deux termes au dessus, ce qui donne le triangle marqué en gras.

La lectrice se convaincra que pour un polynôme de degré 3 arbitraire qui s'écrit $\mathrm{P} = a_0\mathrm{F}_0 + a_1\mathrm{F}_1 + a_2\mathrm{F}_2 + a_3\mathrm{F}_3$, on trouve en descendant la diagonale au bord gauche du tableau (elle correspond à l'évaluation en 0) les coefficients $[a_0, a_1, a_2, a_3]$: cela se déduit du fait que pour F_3 on trouve $[0, 0, 0, 1]$, pour F_2 $[0, 0, 1]$, etc. Dans notre exemple, on a $\mathrm{P} = 3\mathrm{F}_0 - \mathrm{F}_1 + 2\mathrm{F}_2 + 6\mathrm{F}_3$.

Manque de rigueur ?

On trouve des gens qui contestent l'usage des trois petits points dans les preuves (mais sans doute pas sur leurs feuilles de brouillon) et qui prétendent par exemple que l'égalité $(*)$ plus haut ne peut être établie en toute rigueur que par récurrence.

En fait ces « puristes » se situent non pas dans le cadre intuitif des entiers naturels, mais dans le cadre d'un système formel (à préciser) où on développe une théorie dans laquelle l'usage intuitif des trois petits points est interdit *a priori*.

L'inconvénient est que la théorie du système formel lui même, nécessaire pour valider son utilisation, ne peut pas être développée sans une théorie naïve des entiers naturels intuitifs. En effet, même si les entiers naturels intuitifs n'apparaissent que sous une forme déguisée dans le système formel, la machinerie propre du système formel est au moins aussi compliquée que la machinerie des entiers naturels intuitifs. Et les raisonnements qui montrent que le système formel fait bien ce qu'on veut qu'il fasse utilisent de manière inévitable des preuves par récurrence sur la longueur des formules. En dernière analyse ces preuves reposent sur l'utilisation intuitive des trois petits points (ou ce qui revient au même sur l'usage du « et ainsi de suite ») : ce que Poincaré explique en disant que le raisonnement par récurrence ne peut pas être réduit au syllogisme. Insistons en citant de nouveau un passage du texte précédent, dans lequel nous soulignons ce qui nous intéresse.

> Nous voyons *successivement* qu'un théorème est vrai du nombre 1, du nombre 2, du nombre 3 *et ainsi de suite*, la *loi est manifeste*, disons-nous, et elle l'est au même titre que toute loi physique appuyée sur des observations dont le nombre est très grand, mais limité.
>
> On ne saurait méconnaitre qu'il y a là une analogie frappante avec les procédés habituels de l'induction. Mais une différence essentielle subsiste. L'induction, appliquée aux sciences physiques, est toujours incertaine, parce qu'elle repose sur la croyance à un ordre général de l'Univers, ordre qui est en dehors de nous. L'induction mathématique, c'est-à-dire la démonstration par récurrence,

s'impose au contraire nécessairement, parce qu'elle n'est que l'affirmation d'une propriété de l'esprit lui-même.

Nous terminons cette digression en renvoyant à un autre texte de Poincaré que nous citons dans l'annexe 1 du chapitre 5 : voir notamment la section III page 68 et la section VII page 70.

3.2.2 Un exemple plus difficile

Nous présentons maintenant un exemple plus sophistiqué où le raisonnement par récurrence semble incontournable. Il s'agit du problème suivant :

Problème 3.2.1. *Étudier les couples (a, b) d'entiers naturels tels que $1 + ab$ divise $a^2 + b^2$ et montrer que dans un tel cas le quotient $\dfrac{a^2+b^2}{1+ab}$ est un carré.*

On peut supposer $a \geq b$.

Si $b = 0$, alors tout $a > 0$ convient et $\dfrac{a^2+b^2}{1+ab} = a^2$

Si $a = b > 0$, alors $2a^2 = k + ka^2$, avec $k \geq 1$. Mais cette égalité implique aussi $k < 2$, donc $\dfrac{a^2+b^2}{1+ab} = \dfrac{2a^2}{1+a^2} = k = 1$ et $a = b = 1$.

Passés ces cas triviaux, il faut examiner le cas $a > b > 0$.

L'idée est de trouver un entier naturel $c < b$ tel que $\dfrac{a^2+b^2}{1+ab} = \dfrac{c^2+b^2}{1+cb}$. Si on y arrive, un processus s'enclenche $(a, b) \mapsto (b, c) \mapsto \cdots$ qui aboutit en un temps fini à un couple (u, v) avec $u > v = 0$, qui est correct d'après l'étude préliminaire.

Un tel processus s'appelle parfois (à tort) une « descente infinie », alors qu'il devrait plutôt s'appeler une « descente bornée ». Nous y reviendrons page 45.

Pour trouver un entier naturel $c < b$ tel que $\dfrac{a^2+b^2}{1+ab} = \dfrac{c^2+b^2}{1+cb}$, on note que l'équation

$$(a^2 + b^2)(1 + cb) = (c^2 + b^2)(1 + ab)$$

se réécrit $(a - c)(a + c + abc - b^3) = 0$. Avec $a \neq c$ cela équivaut à $a + c + abc = b^3$, ce qui se réécrit $c(1 + ab) = b^3 - a$, ou aussi bien $a(1 + cb) = b^3 - c$.

On note aussi que si

$$a^2 + b^2 \equiv 0 \mod 1 + ab$$

alors

$$a^2b + b^3 \equiv 0 \mod 1 + ab$$

et puisque

$$a^2b \equiv -a \mod 1 + ab$$

on a

$$b^3 - a \equiv 0 \mod 1 + ab.$$

Ainsi $c := \dfrac{b^3 - a}{1 + ab}$ définit bien un nombre entier c.

Si $a > b > 0$, on ne peut pas avoir $c < 0$ car sinon $1+cb \leq 0$, ce qui impliquerait $(c^2+b^2)(1+ab) \leq 0$. Par ailleurs de $(1 + ab)c = b^3 - a$, on déduit l'égalité $(1 + ab)(b - c) = a + b + b^2(a - b)$ qui montre que $b > c$. En conclusion $b > c \geq 0$.

Le contrat est donc rempli.

On obtient le résultat précis suivant : *Tout couple (a, b) solution du problème avec $a > b \geq 0$ est un élément d'une suite (u_n, v_n) construite par récurrence comme suit : $u_0 > 1$, $v_0 = 0$, et*

$(u_{n+1}, v_{n+1}) = \left(\dfrac{u_n^3 - v_n}{1 + u_n v_n}, u_n\right)$. *En outre* $\dfrac{u_k^2 + v_k^2}{1 + u_k v_k}$ *reste constant et est égal à* u_0^2. *En particulier* $\dfrac{a^2 + b^2}{1 + ab} = u_0^2$.

Remarque. On peut montrer que toute suite obtenue de cette manière donne bien des u_n croissants. Cela se déduit de l'égalité $(1+ab)(1+cb) = b^4+1$ qui se démontre comme suit : $(1+ab)(1+bc) = 1 + ab + b(1+ab)c = 1 + ab + b(b^3 - a) = 1 + b^4$.

3.3 Preuves par algorithme et preuves par récurrence

Intuitivement, les preuves par algorithme ressemblent à des raisonnements par récurrence. On démontre que quelque chose se conserve à chaque étape de l'algorithme, ce qui implique que l'algorithme réalise le but recherché lorsqu'il se termine. Si la terminaison de l'algorithme est prouvée en montrant qu'un entier $n \geq 0$ relié aux valeurs prises par les variables décroit à chaque exécution de la boucle, on peut sans doute transformer la preuve par algorithme en une preuve par récurrence sur l'entier n en question.

Toute preuve par algorithme semble donc structurellement très proche d'un raisonnement par récurrence.

3.3.1 Un exemple : le théorème du pgcd

Nous allons vérifier ceci sur l'exemple particulier du théorème 2.2.2 : *Si* a *et* b *sont deux entiers* > 0, *il existe un entier* $g > 0$ *de la forme* $ua + vb$ *avec* $u, v \in \mathbb{Z}$, *qui divise* a *et* b.

Preuve par récurrence correspondant à l'algorithme 2.4.2. Cet algorithme calcule à la fois le pgcd g et une relation de Bézout $ua + vb = g$.

Nous proposons tout d'abord une preuve par récurrence sur $a + b$, directement inspirée par l'algorithme. L'initialisation est un peu surprenante. Nous ferons un commentaire sur ce sujet plus loin.

Si $a + b = 0$ l'hypothèse est impossible, donc l'implication est juste.

Supposons la propriété démontrée pour toutes les valeurs de $a + b$ comprises entre 0 et n et montrons qu'elle est vraie pour $a + b = n + 1$.

Si $a \geq b$ divisons a par b : $a = bq + r$, $0 \leq r < b$. Si $r = 0$, on peut prendre $u = 0$, $v = 1$ et $g = b$: g divise a et b. Si $r > 0$, puisque $r < b \leq a$ on a $b + r < b + a$ et l'hypothèse de récurrence s'applique pour le couple (b, r). Il existe donc u, v, b_1, r_1, g tels que $ub + vr = g$, $gb_1 = b$ et $gr_1 = r$. On en déduit $g = ub + v(a - qb) = va + (u - vq)b$, $gb_1 = b$, $g(b_1 q + r_1) = bq + r = a$.

Si $a < b$ on échange les rôles de a et b dans le calcul précédent. □

Le lecteur remarquera que cette preuve par récurrence abstraite, si on la « met à plat » pour voir les calculs qu'elle implique de faire, donne essentiellement l'algorithme 2.4.2.

Preuve par récurrence correspondant à l'algorithme 2.6.1. Rappelons que cet algorithme était implicite dans la preuve classique, par l'absurde, donnée page 25.

On considère l'ensemble S des triplets $(a, b, c) \in (\mathbb{N}^*)^3$ tels que

$$c \in a\mathbb{Z} + b\mathbb{Z}. \tag{3.1}$$

On va démontrer par récurrence sur $a + b + c$ que *pour tout* $(a, b, c) \in$ S *il existe un triplet* $(a, b, g) \in$ S *tel que* g *divise* a *et* b.

Cela suffira à démontrer le théorème car pour tous $a, b > 0$ le triplet (a, b, b) est évidemment dans S.

Si $a + b + c = 0$ l'hypothèse est impossible, donc l'implication est juste.

Supposons la propriété démontrée pour toutes les valeurs de $a+b+c$ comprises entre 0 et n et montrons qu'elle est vraie pour $a+b+c=n+1$.

Si c divise a et b la propriété est vraie. Sinon, par exemple c ne divise pas a. On écrit $a=cq+r$ avec $0<r<c$. On a $r\in a\mathbb{Z}+c\mathbb{Z}\subseteq a\mathbb{Z}+b\mathbb{Z}$ (puisque $c\in a\mathbb{Z}+b\mathbb{Z}$). On applique l'hypothèse de récurrence avec le triplet (a,b,r) (pour lequel $a+b+r\leq n$). Il existe donc $g\in a\mathbb{Z}+b\mathbb{Z}$, tel que $g>0$ et g divise a et b. □

La lectrice remarquera que cette preuve par récurrence abstraite, si on la « met à plat » pour voir les calculs qu'elle implique de faire, donne essentiellement l'algorithme 2.6.1. En outre, si la propriété à vérifier par récurrence est un peu plus sophistiquée dans cette preuve, le calcul qui justifie le passage de n à $n+1$ est par contre « un peu plus simple ». Cela correspond au fait, déjà noté, que la preuve classique est « un peu plus simple » que la preuve par l'algorithme d'Euclide. Enfin, on note que toute trace de preuve par l'absurde a disparu sans changer l'essentiel de la preuve, qui reste une division de a ou b par c.

Le lecteur pourra aussi noter que la récurrence mise en place en regard de l'algorithme 2.6.1 est plus proche de l'algorithme que celle mise en place en regard de l'algorithme 2.4.2. Il aurait été préférable, si on veut suivre de plus près l'algorithme, de raisonner par récurrence sur $c+d$, où $(c,d)\in \mathrm{T}\subseteq\mathbb{N}^2$, T étant l'ensemble des couples tels que $c\mathbb{Z}+d\mathbb{Z}=a\mathbb{Z}+b\mathbb{Z}$.

Nous laissons en exercice la rédaction précise de la démonstration par récurrence.

3.3.2 Récurrence et descente infinie

C'est maintenant le moment de faire un commentaire sur la manière dont a été mise en place la récurrence dans les deux preuves précédentes. On présente usuellement le raisonnement par récurrence de la manière suivante :

> Pour prouver que, pour tout $n\geq 0$, on a $\mathrm{P}(n)$,
> on prouve $\mathrm{P}(0)$,
> puis, pour un $n\geq 0$ arbitraire, que $\mathrm{P}(n)$ implique $\mathrm{P}(n+1)$.

C'est ainsi qu'est apparu le raisonnement par récurrence dans sa formulation pascalienne.

Mais pour une propriété $\mathrm{P}(n)$ un peu compliquée, il peut arriver qu'on ait besoin d'utiliser $\mathrm{P}(k)$ pour une ou plusieurs valeurs de $k\leq n$ pour pouvoir prouver $\mathrm{P}(n+1)$. Dans ce cas on peut se ramener au raisonnement par récurrence usuel en considérant que la propriété à démontrer est la propriété $\mathrm{Q}(n)$ qui signifie « $\mathrm{P}(k)$ pour $k=0,\ldots,n$ ». C'est ce que nous avons fait dans le deux preuves précédentes.

On peut sans doute considérer que la première apparition de ce type de raisonnement par récurrence amélioré se trouve dans la « descente infinie » de Fermat. Fermat démontrait qu'il est impossible d'avoir $a^4+b^4=c^4$ avec $a,b,c\in\mathbb{N}^*$ en construisant, à partir d'une solution hypothétique (a,b,c), une autre solution, avec c strictement plus petit. La terminologie « descente infinie » tient à ce qu'on démontre une impossibilité : la solution ne peut exister car sinon s'enclencherait un processus de descente infinie dans $\mathbb{N}$, ce qui est impossible. En fait même pour aboutir à une contradiction à partir de l'hypothèse fausse, on n'a besoin que d'un nombre borné d'étapes du processus de descente.

Alors que Fermat utilisait la descente infinie pour montrer une impossibilité, on peut tout aussi bien l'utiliser pour démontrer l'existence d'un objet. La formulation positive de cette « descente bornée » est la récurrence généralisée suivante :

> Pour prouver que, pour tout $n\geq 0$, on a $\mathrm{P}(n)$,
> on prouve, pour un $n\geq 0$ arbitraire, que $[\forall k<n\ \mathrm{P}(k)]$ implique $\mathrm{P}(n)$.

Ici il semble qu'il y ait eu un petit tour de magie, car l'initialisation de la récurrence semble avoir disparu. En fait montrer « $[\forall k<0\ \mathrm{P}(k)]$ implique $\mathrm{P}(0)$ » n'est rien d'autre que montrer $\mathrm{P}(0)$: puisqu'il n'y a pas de $k<0$ dans $\mathbb{N}$, l'hypothèse « $[\forall k<n\ \mathrm{P}(k)]$ » est vraie.

À vrai dire, dans un raisonnement par récurrence généralisée, comme dans les deux exemples donnés ci-dessus, seules certaines valeurs de $k < n$ sont utilisées pour établir $\mathrm{P}(n)$. Dans les exemples que nous avons donnés, on ne passe jamais par $k = 0$ (d'où le caractère artificiel de l'initialisation), mais, à chaque étape du processus, ou bien la preuve se termine, ou bien k décroit strictement tout en restant dans $\mathbb{N}$.

Chapitre 4

Analyse de preuves. Espaces vectoriels et systèmes linéaires

Introduction

Les mathématiques sont réputées créer des idéalités abstraites qui permettent de mieux comprendre ce qu'elles étudient au premier chef, qui sont des réalités plus concrètes.

Par exemple pour comprendre dans quelles conditions une équation pouvait être résoluble par radicaux, Galois a mis en place une théorie générale des équations et a « inventé » la notion abstraite de groupe.

L'analyse de cet exemple paradigmatique nous pousserait trop loin et nous nous attachons dans ce chapitre à un exemple de nature beaucoup plus élémentaire, même si le concept abstrait correspondant, celui d'espace vectoriel, est apparu nettement après la théorie de Galois.

Nous voulons soutenir la thèse qu'une idéalité abstraite inventée par les mathématiciens est toujours basée en dernière analyse sur quelque chose d'assez concret. Abstraction oui, mais abstraction de quelque chose de plus concret.

La notion d'espace vectoriel s'est imposée en mathématiques à partir des deux exemples fondamentaux constitués par la géométrie et les systèmes d'équations linéaires.

Le but de ce chapitre est d'examiner les liens entre

— d'une part, la théorie « abstraite » des espaces vectoriels de dimension finie, et

— d'autre part, la théorie « concrète » des systèmes linéaires.

Nous appellerons $\mathbb{K}$ le corps de base (en pratique on pourra penser à $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$).

Nous allons nous soumettre à une gymnastique intellectuelle très particulière, absolument nécessaire à notre propos.

Pour comparer la théorie abstraite des espaces vectoriels de dimension finie et la théorie concrète de la résolution des systèmes linéaires il faudra faire semblant d'oublier tout ce que l'on a appris par la théorie abstraite. Ceci, et ceci seulement, permettra de comprendre quels résultats de la théorie abstraite découlent de façon immédiate de la théorie concrète.

Si la lectrice n'accepte pas cette mise en sommeil provisoire d'une partie de ses connaissances, elle ne pourra absolument pas comprendre ce qui se passe dans le chapitre présent.

4.1 Un texte classique sur la théorie « abstraite »

Le texte reproduit ci-après en encadré (extrait du manuel Lesieur et Joulain 1966, p. 121, 124-127) est un texte standard qui explique la théorie des espaces vectoriels de dimension finie.

Propriété 2. *Si n vecteurs sont linéairement indépendants, p quelconques d'entre eux sont également linéairement indépendants.*

. . . . . .

V – Espaces vectoriels de dimension finie

Dans le paragraphe précédent, nous avons défini une base d'un espace vectoriel, constituée de n vecteurs (base finie) ; mais un espace vectoriel quelconque ne possède pas toujours de telles bases. Nous allons étudier les espaces vectoriels dans lesquels il existe une base (finie) constituée de n vecteurs.

Théorème 1 (de la base incomplète)**.** *Si* E *est un espace vectoriel possédant une base* $e_1, e_2, \ldots, e_n$*, et si* $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$ $(p \leq n)$ *sont* p *vecteurs linéairement indépendants de* E*, on peut former une nouvelle base de* E *en adjoignant à* $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$*,* $n-p$ *des vecteurs* $e_1, e_2, \ldots, e_n$*, convenablement choisis.*

Nous démontrerons ce théorème par récurrence sur l'entier p ; démontrons-le d'abord dans le cas $p = 1$.

Soit V un vecteur linéairement indépendant, c'est-à-dire non nul (III ; remarque). Les vecteurs $e_1, e_2, \ldots, e_n$ formant une base de E, on a :

$$\mathrm{V} = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n \,; \tag{10}$$

V étant $\neq 0$, les scalaires x_i ne sont pas tous nuls ; on peut supposer $x_1 \neq 0$ (en changeant s'il y a lieu l'ordre des vecteurs $e_1, e_2, \ldots, e_n$). On déduit alors de (10) :

$$e_1 = \frac{1}{x_1}\,\mathrm{V} - \frac{x_2}{x_1} e_2 - \cdots - \frac{x_n}{x_1} e_n \,; \tag{11}$$

Il en résulte que toute combinaison linéaire de $e_1, e_2, \ldots, e_n$ est aussi une combinaison linéaire de $\mathrm{V}, e_2, \ldots, e_n$; autrement dit, les vecteurs $\mathrm{V}, e_2, \ldots, e_n$ engendrent E. Montrons que ces vecteurs sont linéairement indépendants. Soit :

$$\lambda_1 \mathrm{V} + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0 \,;$$

si λ_1 n'est pas nul, on a :

$$\mathrm{V} = -\frac{\lambda_2}{\lambda_1} e_2 - \cdots - \frac{\lambda_n}{\lambda_1} e_n,$$

ce qui est contraire à $x_1 \neq 0$ dans l'unique décomposition (10) du vecteur V par rapport à la base $e_1, e_2, \ldots, e_n$; il en résulte $\lambda_1 = 0$ et :

$$\lambda_2 e_2 + \cdots + \lambda_n e_n = 0 \,;$$

mais $e_2, \ldots, e_n$ sont linéairement indépendants (propriété 2), d'où :

$$\lambda_2 = \lambda_3 = \cdots = \lambda_n = 0.$$

V, $e_2, \ldots, e_n$ sont donc linéairement indépendants, et, puisqu'ils engendrent E, ils forment une base de E, ce qui établit le théorème pour $p = 1$.

On suppose maintenant le théorème vrai pour $p-1$ vecteurs. Soit p vecteurs linéairement indépendants $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$. On peut appliquer le théorème aux $p-1$ vecteurs $\mathrm{V}_1, \ldots, \mathrm{V}_{p-1}$ qui sont linéairement indépendants (propriété 2). Il existe $n-p+1$ vecteurs convenablement choisis dans la base $e_1, e_2, \ldots, e_n$ (on peut prendre $e_p, e_{p+1}, \ldots, e_n$ en changeant s'il y a lieu l'ordre des vecteurs e_i) tels que : $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_{p-1}, e_p, e_{p+1}, \ldots, e_n$ forment une base de E ; on a alors :

$$\mathrm{V}_p = x_1 \mathrm{V}_1 + \cdots + x_{p-1} \mathrm{V}_{p-1} + x_p e_p + x_{p+1} e_{p+1} + \cdots + x_n e_n. \tag{12}$$

$\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$ étant linéairement indépendants, l'un au moins des scalaires $x_p, x_{p+1}, \ldots, x_n$ n'est pas nul. On peut supposer $x_p \neq 0$ (en modifiant éventuellement l'ordre des vecteurs $e_p, e_{p+1}, \ldots, e_n$) ; on déduit alors de (12) :

$$e_p = -\frac{x_1}{x_p}\mathrm{V}_1 - \cdots - \frac{x_{p-1}}{x_p}\mathrm{V}_{p-1} + \frac{1}{x_p}\mathrm{V}_p - \frac{x_{p+1}}{x_p}e_{p+1} - \cdots - \frac{x_n}{x_p}e_n. \tag{13}$$

Toute combinaison de $\mathrm{V}_1, \ldots, \mathrm{V}_{p-1}, e_p, \ldots, e_n$ est donc aussi une combinaison linéaire de $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p, e_{p+1}, \ldots, e_n$, ce qui prouve que ces derniers vecteurs engendrent E ; démontrons qu'ils sont linéairement indépendants.

Soit : $\lambda_1\mathrm{V}_1 + \cdots + \lambda_p\mathrm{V}_p + \lambda_{p+1}e_{p+1} + \cdots + \lambda_n e_n = 0$; si λ_p n'est pas nul, on a :

$$\mathrm{V}_p = -\frac{\lambda_1}{\lambda_p}\mathrm{V}_1 - \cdots - \frac{\lambda_{p-1}}{\lambda_p}\mathrm{V}_{p-1} - \frac{\lambda_{p+1}}{\lambda_p}e_{p+1} - \cdots - \frac{\lambda_n}{\lambda_p}e_n,$$

ce qui est contraire à $x_p \neq 0$ dans l'unique décomposition (12) du vecteur V_p ; on a donc $\lambda_p = 0$. Il en résulte :

$$\lambda_1\mathrm{V}_1 + \cdots + \lambda_{p-1}\mathrm{V}_{p-1} + \lambda_{p+1}e_{p+1} + \cdots + \lambda_n e_n = 0\,;$$

or, $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_{p-1}, e_{p+1}, \ldots, e_n$ sont linéairement indépendants (propriété 2), d'où : $\lambda_1 = \cdots = \lambda_{p+1} = \cdots = \lambda_n = 0$.

Finalement $\mathrm{V}_1, \ldots, \mathrm{V}_p, e_{p+1}, \ldots, e_n$ engendrent E et sont linéairement indépendants, donc constituent une base de E, ce qui établit le théorème pour p vecteurs linéairement indépendants, avec $p \leq n$. □

On en déduit le théorème suivant :

Théorème 2 (de la dimension)**.** *S'il existe dans un espace vectoriel* E *une base composée de n vecteurs :*

1° *Tout système de n vecteurs linéairement indépendants est une base de* E.

2° *Tout système de vecteurs linéairement indépendants comprend au plus n vecteurs.*

3° *Toute autre base de* E *se compose aussi de n vecteurs.*

1° Soit $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_n$, n vecteurs linéairement indépendants ; d'après le théorème 1 appliqué au cas $p = n$, ces vecteurs forment une base de E.

2° Soit $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$, p vecteurs linéairement indépendants. Supposons $p > n$; alors $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_n$ sont linéairement indépendants (propriété 2) et ils forment une base de E d'après 1° ; on peut décomposer V_{n+1} suivant cette base :

$$\mathrm{V}_{n+1} = x_1\mathrm{V}_1 + \cdots + x_n\mathrm{V}_n,$$

ce qui est contraire à l'indépendance linéaire de $\mathrm{V}_1, \ldots, \mathrm{V}_n, \ldots, \mathrm{V}_p$. Il en résulte $p \leq n$.

3° Soit $f_1, f_2, \ldots, f_m$ une autre base de E. Les vecteurs $f_1, f_2, \ldots, f_m$ étant linéairement indépendants, on a, d'après 2°, $m \leq n$. Pour la même raison, en échangeant les rôles des deux bases $e_1, \ldots, e_n$ et $f_1, \ldots, f_m$, on a : $n \leq m$ et finalement $m = n$. □

Ce théorème conduit à la notion de dimension.

Définition 6. S'il existe un entier n positif tel que l'espace vectoriel E possède une base composée de n vecteurs, cet entier est unique et on l'appelle *dimension* de l'espace vectoriel E (on note $n = \dim \mathrm{E}$). L'espace réduit à $\{0\}$ est dit de dimension nulle.

La dimension est donc le nombre de vecteurs de toute base de E et aussi le nombre maximum d'éléments d'un système de vecteurs linéairement indépendants de E.

S'il n'existe pas de tel entier n, E est dit de *dimension infinie*.

Dans la suite, on se limitera aux espaces vectoriels de dimension finie.

. . . . . .

VI – Rang d'un système de vecteurs

Définition 7. On appelle *rang* d'un système de p vecteurs d'un espace vectoriel E, le nombre maximum r de vecteurs linéairement indépendants qu'on peut extraire de ce système.

Théorème 3. *Dans un espace vectoriel* E*, le sous-espace* E′ *engendré par les vecteurs* V_1*,* V_2*, ...,* V_p *est de dimension finie égale au rang* r *du système de vecteurs* $\mathrm{V}_1, \ldots, \mathrm{V}_p$ *et, de façon plus précise, tout système de* r *vecteurs linéairement indépendants extrait du système* $\mathrm{V}_1, \ldots, \mathrm{V}_p$ *est une base de* E′*.*

Soit $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$ p vecteurs de E et r le rang de ce système ; on peut supposer $\mathrm{V}_1, \ldots, \mathrm{V}_r$ linéairement indépendants (en modifiant, s'il y a lieu l'ordre des p vecteurs). Alors, pour tout entier h tel que $1 \leq h \leq p - r$ les vecteurs $\mathrm{V}_1, \ldots, \mathrm{V}_r, \mathrm{V}_{r+h}$ sont linéairement dépendants ; on a une relation à coefficients non tous nuls de la forme :

$$\lambda_1 \mathrm{V}_1 + \cdots + \lambda_r \mathrm{V}_r + \lambda_{r+h} \mathrm{V}_{r+h} = 0.$$

$\mathrm{V}_1, \ldots, \mathrm{V}_r$ étant linéairement indépendants on a $\lambda_{r+h} \neq 0$ et :

$$\mathrm{V}_{r+h} = -\frac{\lambda_1}{\lambda_{r+h}} \mathrm{V}_1 - \cdots - \frac{\lambda_r}{\lambda_{r+h}} \mathrm{V}_r.$$

Toute combinaison linéaire des vecteurs $\mathrm{V}_1, \ldots, \mathrm{V}_p$ se réduit donc à une combinaison linéaire de $\mathrm{V}_1, \ldots, \mathrm{V}_r$. Il en résulte que $\mathrm{V}_1, \ldots, \mathrm{V}_r$ engendrent E′ et par conséquent (puisqu'ils sont linéairement indépendants) forment une base de E′. Le théorème en résulte. □

. . . . . .

On déduit de ce théorème le résultat suivant qui complète le théorème 2 de la dimension.

Théorème 4. *Dans un espace vectoriel* E *de dimension finie* n*, tout système de générateurs de* E *comprend au moins* n *vecteurs et tout système de* n *générateurs de* E *est une base de* E*.*

Soit $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_p$ un système de générateurs de E. D'après le théorème 3 le rang r de ce système est égal à la dimension n de E ; or $r \leq p$ (par définition même du rang) ; il en résulte : $n \leq p$.

Si $p = n$, on a $r = p$ et les vecteurs $\mathrm{V}_1, \ldots, \mathrm{V}_p$ sont linéairement indépendants ; comme ils engendrent E, ils forment une base de E.

La dimension de E est donc aussi le nombre minimum d'éléments d'un système de générateurs. □

4.2 De la méthode du pivot à la théorie de la dimension

Dans toute cette section on oublie ce que l'on sait de la théorie de la dimension (en particulier on oublie qu'on connait les résultats démontrés dans le texte encadré) et on examine ce que la méthode du pivot nous apprend au sujet de cette théorie.

Dépendance linéaire dans $\mathbb{K}^n$ et systèmes linéaires

Dans ce paragraphe on explique la traduction du langage vectoriel « abstrait » dans le langage « concret » des systèmes linéaires, du moins dans le cas où l'espace vectoriel considéré est $\mathbb{K}^n$.

Soient $\mathrm{V}_1, \ldots, \mathrm{V}_p, \mathrm{V} \in \mathbb{K}^n$. Une relation de dépendance linéaire éventuelle

$$x_1 \mathrm{V}_1 + \cdots + x_p \mathrm{V}_p = \mathrm{V} \tag{4.1}$$

s'interprète en termes d'un système linéaire de n équations à p inconnues comme suit.

Nous notons

$$\mathrm{V}_j = \begin{bmatrix} v_{1,j} \\ \vdots \\ v_{n,j} \end{bmatrix}, (j = 1, \ldots, p), \mathrm{V} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

L'équation (4.1) peut se lire coordonnée par coordonnée et elle signifie alors exactement

$$\begin{array}{ccccccc} x_1 v_{1,1} & + & \cdots & + & x_p v_{1,p} & = & v_1 \\ \vdots & & & & \vdots & \vdots & \vdots \\ x_1 v_{n,1} & + & \cdots & + & x_p v_{n,p} & = & v_n \end{array} \tag{4.2}$$

Autrement dit, $\mathrm{X} = {}^{\mathrm{t}}(x_1, \ldots, x_p)$ est une solution du système linéaire $\mathrm{AX} = \mathrm{V}$ où A est la matrice ayant pour vecteurs colonnes les V_j.

Si maintenant on considère que V est un vecteur variable, alors que les V_j sont fixés, on a un système linéaire (4.2) dont le premier membre est fixé. La terminologie du langage vectoriel (indépendance linéaire, système générateur, base) a alors la signification suivante dans le langage des systèmes linéaires.

1. « Les vecteurs V_j sont linéairement indépendants (ou libres) » signifie : le système homogène (c.-à-d. sans second membre, ou plus exactement avec second membre nul) admet $0 = (0, \ldots, 0)$ pour unique solution.
2. « Les vecteurs V_j engendrent $\mathbb{K}^n$ » signifie : quel que soit le second membre, le système linéaire admet une solution.
3. « Les vecteurs V_j forment une base de $\mathbb{K}^n$ » signifie : quel que soit le second membre, le système linéaire admet une solution unique.

Rappels sur la méthode du pivot

Nous supposons le lecteur familier avec la méthode du pivot pour la résolution des systèmes linéaires. Nous ne prétendons donc pas exposer ici cette méthode, mais simplement la rappeler, et préciser la terminologie que nous utiliserons dans la suite.

Nous considérons la théorie des systèmes linéaires traitée par la méthode du pivot. Nous désirons la comparer à la théorie « abstraite » de la dimension donnée dans le texte encadré.

Méthode du pivot (version matricielle pour les systèmes linéaires)**.** *On considère un système linéaire de n équations à p inconnues écrit sous la forme* $\mathrm{AX} = \mathrm{B}$ *où* X *est le vecteur colonne des inconnues x_i. La méthode du pivot ramène le système à un système équivalent* $\mathrm{A'X'} = \mathrm{B'}$*, pour lequel les inconnues ont été éventuellement permutées, et où* $\mathrm{A'}, \mathrm{B'}$ *ont la forme standard suivante :*

$$\mathrm{A'} = \begin{bmatrix} 1 & \bullet & \cdots & \bullet & \bullet & \cdots & \bullet \\ 0 & 1 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \bullet & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & \bullet & \cdots & \bullet \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathrm{B'} = \begin{bmatrix} \bullet \\ \vdots \\ \vdots \\ \bullet \\ \bullet \\ \vdots \\ \bullet \end{bmatrix}, \quad \mathrm{X} = \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ \\ \\ \vdots \\ \vdots \\ x_p \end{bmatrix}, \quad \mathrm{X'} = \begin{bmatrix} x_{\sigma_1} \\ \vdots \\ \vdots \\ \\ \\ \vdots \\ \vdots \\ x_{\sigma_p} \end{bmatrix} \tag{4.3}$$

($(\sigma_1, \ldots, \sigma_p)$ est une permutation de $(1, \ldots, p)$).

La matrice A′ ci-dessus (à n lignes et p colonnes) s'écrit comme une matrice par blocs

$$\mathrm{A}' = \begin{bmatrix} \mathrm{T} & \mathrm{S} \\ 0_{n-r,r} & 0_{n-r,p-r} \end{bmatrix}$$

dans laquelle les lignes de 0 en dessous de T sont éventuellement absentes (si $r = n$), et pareillement pour les colonnes à droite de T (si $r = p$). La matrice T est triangulaire supérieure d'ordre r, avec des 1 sur la diagonale. Si on a effectué des échanges de colonnes lorsqu'on a traité la matrice, on a gardé la mémoire de la permutation σ qui en résulte et on sait que les colonnes $1, 2, \ldots, p$ du système linéaire final correspondent aux inconnues $x_{\sigma_1}, \ldots, x_{\sigma_p}$. Si $p > r$ les inconnues $x_{\sigma_{r+1}}, \ldots, x_{\sigma_p}$ sont appelées les inconnues auxiliaires, et $x_{\sigma_1}, \ldots, x_{\sigma_r}$ les inconnues principales.

La méthode traite la matrice $\mathrm{C} = \begin{bmatrix} \mathrm{A} & \mathrm{B} \end{bmatrix}$ par manipulations élémentaires de lignes et (éventuellement) échanges de colonnes. Elle procède par grandes étapes. L'étape n° k ($1 \leq k \leq \min(n, p)$) réalise les opérations élémentaires suivantes :

- si tous les coefficients de A dans les lignes d'indice $\geq k$ sont nuls, le calcul s'arrête ;
- sinon et si le coefficient $c_{k,k}$ est nul on procède à un échange de lignes et/ou un échange de colonnes pour mettre en position (k, k) un élément $c_{i,j}$ non nul trouvé dans la partie sud-est de la matrice A (c.-à-d., avec $k \leq i \leq p$, $k \leq j \leq n$) ;
- ensuite on multiplie la ligne n° k de C par l'inverse du coefficient « pivot », celui en position (k, k) ;
- enfin « on tue » les coefficients en dessous du pivot par manipulations élémentaires de lignes sur la matrice C.

La méthode du pivot utilise donc les manipulations élémentaires suivantes :

- échange de lignes (de C) ou de colonnes (de A),
- multiplication de la ligne (de C) où se trouve le pivot par une constante non nulle : $\mathrm{L}_k \leftarrow \alpha \mathrm{L}_k$ ($\alpha \neq 0$),
- ajout à une ligne en dessous de celle où se trouve le pivot un multiple de la ligne où se trouve le pivot : $\mathrm{L}_i \leftarrow \mathrm{L}_i + \beta \mathrm{L}_k$.

Il est clair que la méthode du pivot remplace un système linéaire par un système linéaire équivalent. Lors d'un échange de colonnes, il faut évidemment garder en mémoire à quelle inconnue correspond chaque colonne. Lors d'une manipulation de lignes du type $\mathrm{L} \leftarrow \mathrm{L} + \beta \mathrm{L}'$, la nouvelle équation est évidemment une conséquence des anciennes, et vice versa car on peut retrouver l'ancienne équation par la manipulation « opposée » $\mathrm{L} \leftarrow \mathrm{L} - \beta \mathrm{L}'$.

On a toujours $r \leq n$ et p, et donc se présentent *a priori* 4 cas possibles

1. Le cas $n = p = r$. La matrice A′ est alors *unitriangulaire supérieure*, c'est-à-dire qu'elle est triangulaire supérieure et que tous les coefficients sur sa diagonale principale sont égaux à 1.

$$\mathrm{A}' = \begin{bmatrix} 1 & \bullet & \cdots & \bullet \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bullet \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \quad \mathrm{B}' = \begin{bmatrix} \bullet \\ \vdots \\ \vdots \\ \bullet \end{bmatrix}$$

2. Le cas $n > p = r$.

$$\mathrm{A}' = \begin{bmatrix} 1 & \bullet & \cdots & \bullet \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bullet \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \quad \mathrm{B}' = \begin{bmatrix} \bullet \\ \vdots \\ \vdots \\ \bullet \\ \bullet \\ \vdots \\ \bullet \end{bmatrix}$$

3. Le cas $p > n = r$.

$$A' = \begin{bmatrix} 1 & \bullet & \cdots & \bullet & \bullet & \cdots & \bullet \\ 0 & 1 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \bullet & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & \bullet & \cdots & \bullet \end{bmatrix}, \quad B' = \begin{bmatrix} \bullet \\ \vdots \\ \vdots \\ \bullet \end{bmatrix}$$

4. Le cas $p > r$ et $n > r$ (voir l'équation (4.3)).

1) A priori *on a $r = p$ ou $r < p$. On examine ici la signification de ces deux éventualités.*

Si $r < p$ on a les inconnues principales $x_{\sigma_1}, \ldots, x_{\sigma_r}$ et une ou plusieurs inconnues auxiliaires $x_{\sigma_{r+1}}, \ldots, x_{\sigma_p}$. On considère le système homogène associé (c'est-à-dire le système linéaire avec 0 au second membre). On peut choisir arbitrairement la valeur de chaque inconnue auxiliaire. Ensuite le système se résout en commençant par x_{σ_r} et en continuant jusqu'à x_{σ_1}. La solution n'est donc pas unique. Et cela signifie que le système $V_1, \ldots, V_p$ n'est pas libre.

Si $r = p$ le système linéaire avec des 0 au second membre admet la seule solution 0 (on commence par $x_p = 0$ et on obtient les autres de proche en proche). Et cela signifie que le système $V_1, \ldots, V_p$ est libre.

2) A priori *on a $r = n$ ou $r < n$. On examine ici la signification de ces deux éventualités.*

Si $r < n$ on va montrer qu'il y a des seconds membres V pour lesquels le système linéaire n'a pas de solution. Cela signifie que le système $V_1, \ldots, V_p$ n'est pas générateur. Cela est clair pour le système linéaire équivalent (4.3) sous forme simplifiée que la méthode du pivot produit : il suffit de choisir un vecteur W au second membre dont la coordonnée numéro $r + 1$ est non nulle. Un tel vecteur W peut-il être obtenu à partir d'un certain vecteur V qui subit les manipulations de lignes que produit la méthode du pivot ? Oui, car toute manipulation de ligne admet une manipulation inverse qui permet de refaire ce qu'on a défait :

— l'échange de lignes $ML_{i,k} : L_i \leftrightarrow L_k$ est une manipulation égale à la manipulation inverse,

— la manipulation de lignes $ML_{i,k,\beta} : L_i \leftarrow L_i + \beta L_k$ admet pour manipulation inverse la manipulation de lignes $ML_{i,k,-\beta} : L_i \leftarrow L_i - \beta L_k$

— la manipulation de lignes $ML_{i,\alpha} : L_i \leftarrow \alpha L_i$ ($\alpha \neq 0$) admet pour manipulation inverse la manipulation de lignes $ML_{i,\alpha^{-1}} : L_i \leftarrow \alpha^{-1} L_i$

Ainsi, connaissant les manipulations de lignes M_1,..., M_t qui ont été opérées par la méthode du pivot, on fera subir à W les manipulations inverses dans l'ordre inverse pour connaitre le vecteur V qui convient.

Si $r = n$ le système linéaire avec un second membre arbitraire admet au moins une solution. S'il y a des inconnues auxiliaires on les prend toutes nulles. Ensuite il reste un système triangulaire qui se résout de proche en proche (en commençant avec x_p). Cela signifie que le système $V_1, \ldots, V_p$ est générateur.

Ainsi, la méthode du pivot, appliquée au système linéaire (4.2) correspondant à l'équation (4.1), prouve les résultats suivants :

1. On a $r = p$ si et seulement si les vecteurs V_i forment un système libre.
2. On a $r = n$ si et seulement si les vecteurs V_i forment un système générateur dans $\mathbb{K}^n$.
3. On a $r = n = p$ si et seulement si les vecteurs V_i forment une base de $\mathbb{K}^n$.
4. Tout système générateur dans $\mathbb{K}^n$ admet au moins n éléments.
5. Tout système libre dans $\mathbb{K}^n$ admet au plus n éléments.
6. Si $p = n$ les propriétés suivantes sont équivalentes :
   (a) Les V_i forment un système générateur.
   (b) Les V_i forment un système libre.
   (c) Les V_i forment une base.

En particulier les théorèmes 2 et 4 du texte encadré sont démontrés dans le cas de l'espace vectoriel $\mathbb{K}^n$.

Nous montrons dans le paragraphe qui suit que nous obtenons en fait nettement mieux.

Que peut-on déduire de la méthode du pivot pour la théorie de la dimension ?

Dans les théorèmes 1, 2, 4 du texte encadré on a affaire à un espace vectoriel E avec une base finie $\mathcal{B} = (e_1, \dots, e_n)$.

Expliquons pourquoi, si on a un théorème concernant un espace vectoriel E *qui possède une base* $\mathcal{B}$ *de* n *éléments, il suffit de démontrer le théorème dans le cas de l'espace vectoriel* $\mathbb{K}^n$ *en prenant pour* $\mathcal{B}$ *la base canonique.*

Comme tout élément de E s'écrit de manière unique sous forme $x_1e_1 + \cdots + x_ne_n$, on obtient une bijection $\varphi_{\mathcal{B}} \colon \mathbb{K}^n \to \mathrm{E}$, $(x_1, \dots, x_n) \mapsto x_1e_1 + \cdots + x_ne_n$.

Notons $\underline{x}$ pour $(x_1, \dots, x_n)$. Comme $\varphi_{\mathcal{B}}(\underline{x} + \underline{y}) = \varphi_{\mathcal{B}}(\underline{x}) + \varphi_{\mathcal{B}}(\underline{y})$ et $\varphi_{\mathcal{B}}(\alpha\underline{x}) = \alpha\varphi_{\mathcal{B}}(\underline{x})$ tous les résultats concernant l'espace vectoriel E ont une interprétation *via* $\varphi_{\mathcal{B}}$ en tant que résultats concernant l'espace vectoriel $\mathbb{K}^n$, et vice versa.

Par ailleurs $\mathbb{K}^n$ est un cas particulier d'espace vectoriel avec une base de n éléments : la base canonique $f_1 = (1, 0, \dots, 0), f_2 = (0, 1, \dots, 0), \dots, f_n = (0, \dots, 0, 1)$. □

Ainsi : *lorsque l'on considère un espace vectoriel muni d'une base de* n *éléments, on peut supposer sans perte de généralité que* $\mathrm{E} = \mathbb{K}^n$ *et que la base considérée au départ est la base canonique.*

On obtient donc comme conséquence de l'étude faite au paragraphe précédent les théorèmes 2 et 4 du texte encadré.

Par contre le théorème 1 semble donner un résultat plus précis dans le cadre d'un espace vectoriel dont on connait une base.

Et le théorème 3 nécessite la notion de rang d'un système dans un espace vectoriel sans que soit déjà connue une base de l'espace vectoriel, donc il ne semble pas pouvoir relever directement de ce que l'on a dit de la méthode du pivot. Nous traiterons cette question dans la section 4.3.

On examine maintenant dans quelles mesure la méthode du pivot fournit le théorème de la base incomplète (le théorème 1 du texte encadré).

On considère le système de vecteurs $\mathrm{V}_1, \dots, \mathrm{V}_p, e_1, \dots, e_n$ et le système linéaire associé dont le premier membre correspond à une matrice à n lignes et $p + n$ colonnes :

$$\mathrm{A} = \begin{bmatrix} v_{1,1} & \dots & v_{1,p} & 1 & 0 & \dots & 0 \\ v_{2,1} & \dots & v_{2,p} & 0 & 1 & & \vdots \\ \vdots & & \vdots & \vdots & & \ddots & 0 \\ v_{n,1} & \dots & v_{n,p} & 0 & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathrm{A}_1 & \mathbb{I}_n \end{bmatrix}$$

Lorsque l'on applique la méthode du pivot au système linéaire correspondant on trouvera nécessairement $r = n$ inconnues principales, car le système linéaire admet manifestement une solution quel que soit le second membre.

Par ailleurs lors du traitement des p premiers pivots on peut ne procéder à aucun échange de colonnes car la nécessité absolue de faire pour la première fois un échange de colonnes pour trouver le pivot numéro k apparait uniquement dans le cas que nous allons expliquer. On note $\mathrm{C}_1, \dots, \mathrm{C}_\ell$ les vecteurs colonnes au départ, après traitement des $k-1$ premiers pivots, les colonnes sont devenues $\mathrm{C}'_1, \dots, \mathrm{C}'_\ell$ (aucune n'a changé de place) et le système linéaire est remplacé par un système équivalent (les inconnues restant dans le même ordre). S'il fallait absolument procéder à un échange de colonnes pour le pivot numéro k, cela signifierait que si on avait traité uniquement la matrice $[\mathrm{C}_1 \cdots \mathrm{C}_k]$ on aurait abouti à $[\mathrm{C}'_1 \cdots \mathrm{C}'_k]$ avec $r = k - 1$, et donc, puisque $r < k$, que les vecteurs $\mathrm{C}_1, \dots, \mathrm{C}_k$ seraient linéairement dépendants.

Ainsi la méthode du pivot appliquée à la matrice $A = [V_1 \cdots V_p e_1 \cdots e_n]$ donne les n inconnues principales $x_1, \ldots x_p, x_{\sigma_{p+1}}, \ldots, x_{\sigma_n}$ (où les σ_j sont dans $\{p+1, \ldots, n\}$) ce qui signifie exactement qu'il y a une base qui commence par $V_1, \ldots, V_p$ et se termine par certains des e_i.

Récapitulons : les théorèmes 1, 2 et 4 du texte encadré peuvent être déduits facilement des résultats que donne la méthode du pivot pour la résolution des systèmes linéaires.

Interprétons maintenant la solution générale des systèmes linéaires donnée par la méthode du pivot dans le langage vectoriel.

Étant donnés $V_1, \ldots, V_p, V$ dans un espace vectoriel possédant une base de n éléments, on peut extraire de $V_1, \ldots, V_p$ un système libre (disons de r éléments) tel que les vecteurs restants s'expriment en fonction du système libre. En outre on sait déterminer si V est combinaison linéaire des V_i.

Commentaire. Le contenu proprement algorithmique de la méthode du pivot disparait un peu dans cette reformulation plus abstraite.

4.3 Retour sur la théorie abstraite de la dimension

À propos du théorème 3 dans le texte encadré

Dans ce paragraphe nous discutons les points suivants :

1. *Nous expliquons pourquoi dans le texte encadré le théorème 3 vient* après *le théorème 2.*
2. *Nous expliquons comment il aurait fallu modifier l'énoncé du théorème 3 pour énoncer ce que donne sa démonstration avant même d'énoncer les théorèmes 1 et 2.*
3. *Nous comparons les théorèmes 1 et 3 des deux points de vue suivants : le caractère plus ou moins général, le caractère plus ou moins explicite.*

1) Dans l'énoncé du théorème 3 on parle de la dimension du sous-espace vectoriel E′ de E. Il est donc nécessaire d'avoir défini ce qu'est la dimension d'un espace vectoriel, et le théorème 2 est nécessaire pour pouvoir donner cette définition.

2) Si on lit en détail la preuve du théorème 3 on voit qu'elle donne précisément le résultat suivant de manière totalement indépendante des théorèmes 1 et 2, avec une preuve extrêmement simple. *Si* $V_1, \ldots, V_p$ *sont des éléments d'un espace vectoriel* E, *si* $V_1, \ldots, V_r$ $(r < p)$ *sont linéairement indépendants et si tout système* $V_1, \ldots, V_r, V_{r+k}$ $(k \geq 1, r + k \leq p)$ *est linéairement dépendant, alors* $V_1, \ldots, V_r$ *est une base du sous-espace* E′ *engendré par* $V_1, \ldots, V_p$.

Ou encore : *si* $V_1, \ldots, V_p$ *sont des éléments d'un espace vectoriel* E, *tout système libre maximal extrait de* $V_1, \ldots, V_p$ *engendre le même sous-espace* E′ *que* $V_1, \ldots, V_p$.

3) Dans l'énoncé que l'on vient de donner, si un système extrait $V_1, \ldots, V_s$ est seulement supposé libre, on pourra toujours le compléter en un système libre maximal (en rajoutant des vecteurs tant qu'il est possible de garder un système libre) et on obtient donc aussi bien cette conclusion :

Théorème 0. *Si dans un espace vectoriel on a un système libre contenu dans un système générateur, alors il existe une base contenue dans le système générateur et contenant le système libre.*

Ceci semble à première vue être le théorème de la base incomplète, à moindres frais, avec une preuve plus élégante. Pourquoi diable alors les auteurs se fatiguent-ils avec la démonstration nettement plus longue et délicate du théorème 1 alors qu'ils auraient pu commencer par le théorème 0, qui ne suppose même pas que l'espace vectoriel ambiant E admet une base finie ?

En réalité, dans le théorème 1 on affirme que la base obtenue possède n éléments, c'est-à-dire le même nombre d'éléments que celui d'une base donnée *a priori*. En ce sens le théorème 1 est donc beaucoup plus précis : *il permet de prouver que toutes les bases ont le même nombre d'éléments.*

Par ailleurs, le théorème 3 (ou l'une de ses reformulations ci-dessus sans changer la preuve) est plus général puisqu'il ne suppose pas *a priori* que l'espace E′ admet une base.

Une autre différence importante entre ces deux théorèmes se situe du côté de l'effectivité. La démonstration du théorème 1 donne une procédure explicite pour calculer la nouvelle base. Il n'en est pas de même dans la démonstration du théorème 3, car *on ne dit pas comment, dans un espace vectoriel général, on peut décider si un système de vecteurs est libre* et, en cas de réponse négative, comment on peut fournir une relation de dépendance linéaire.

En d'autres termes, la démonstration du théorème 3 est peut-être trop élégante, car derrière son apparente simplicité se cache un difficile problème mathématique, proprement escamoté dans l'énoncé du théorème.

Commentaires. 1) Dans le cas d'un espace vectoriel réel le fait de disposer d'un produit scalaire (forme bilinéaire définie positive) sur l'espace ambiant E permet dans une certaine mesure de tester l'indépendance linéaire et de calculer des relations de dépendance linéaire, grâce à la méthode de Gram par exemple.

2) Le caractère explicite de la démonstration du théorème 1 n'est sans doute pas complètement clair pour qui n'a pas l'habitude de soumettre les démonstrations à ce genre d'examen (la démonstration est-elle explicite ou pas ?) Nous allons examiner cette question plus en détail dans le paragraphe suivant. On se convaincra du fait que la démonstration du théorème 1 revient à faire subir la méthode du pivot au système linéaire qui correspond aux hypothèses du théorème. Ceci ramène la question du caractère explicite (ou pas) de la démonstration du théorème 1 à la question analogue du caractère explicite (ou pas) de la méthode du pivot.

3) La méthode du pivot elle-même est complètement explicite *uniquement sous certaines hypothèses* qui apparaissent clairement lorsqu'on se propose d'en faire un algorithme sur machine. On voit alors qu'il faut (évidemment) être capable d'effectuer les opérations arithmétiques de base dans le corps des coefficients de la matrice, mais aussi qu'on a besoin d'un test d'égalité à zéro pour « la recherche du prochain pivot » qui est un élément non nul dans une partie précise de la matrice en cours de traitement. Par exemple en analyse numérique, on ne dispose pas d'un tel test : cela tient *a priori* à la nature des « réels flottants » en machine, mais c'est aussi un problème avec les « vrais nombres réels », dont nous rediscuterons plus avant. C'est pourquoi la méthode du pivot n'est pas considérée comme une méthode fiable. De même en analyse numérique la détermination du rang d'une matrice est un problème « instable », qui ne peut être traité que sous certaines hypothèses restrictives.

À propos du théorème de la base incomplète

Dans ce paragraphe nous expliquons ce que donne la démonstration du théorème de la base incomplète donnée dans le texte encadré en termes de la matrice ayant pour vecteurs colonnes $\mathrm{V}_1, \dots \mathrm{V}_p$ *(lorsque* $\mathrm{E} = \mathbb{K}^n$ *et lorsque la base* $e_1, \dots, e_n$ *est la base canonique).*

Nous reprenons la démonstration du théorème 1 sur un exemple et examinons quels calculs cette preuve demande de faire pour certifier qu'on a obtenu une nouvelle base de la forme voulue. Nous allons prendre $p = 3$ et $n = 5$. Les vecteurs $\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3$ supposés linéairement indépendants forment les 3 colonnes d'une matrice A :

$$\mathrm{A} := \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \\ a_4 & b_4 & c_4 \\ a_5 & b_5 & c_5 \end{bmatrix}$$

Dans la preuve on remarque que puisque $\mathrm{V}_1 \neq 0$ un des a_i est non nul. Sans perte de généralité on suppose que c'est a_1 et on exprime e_1 sous la forme

$$e_1 = \frac{1}{a_1}\mathrm{V}_1 - \frac{a_2}{a_1}e_2 - \frac{a_3}{a_1}e_3 - \frac{a_4}{a_1}e_4 - \frac{a_5}{a_1}e_5. \tag{4.4}$$

Ceci montre que V_1, e_2, e_3, e_4, e_5 est un système générateur. Un argument supplémentaire montre que le système est bien libre. Si on veut voir comment est obtenu précisément le fait que V_1, e_2, e_3, e_4, e_5 est un système générateur grâce au calcul (4.4), on considère un vecteur arbitraire $U = u_1e_1 + u_2e_2 + u_3e_3 + u_4e_4 + u_5e_5$ et on voit qu'il se réécrit :

$$U = u_1\left(\frac{1}{a_1}V_1 - \frac{a_2}{a_1}e_2 - \frac{a_3}{a_1}e_3 - \frac{a_4}{a_1}e_4 - \frac{a_5}{a_1}e_5\right) + u_2e_2 + u_3e_3 + u_4e_4 + u_5e_5$$

c'est-à-dire encore

$$U = \frac{u_1}{a_1}V_1 + \left(u_2 - \frac{u_1a_2}{a_1}\right)e_2 + \left(u_3 - \frac{u_1a_3}{a_1}\right)e_3 + \left(u_4 - \frac{u_1a_4}{a_1}\right)e_4 + \left(u_5 - \frac{u_1a_5}{a_1}\right)e_5$$

On reconnait là la résolution du système linéaire $x_1V_1 + x_2e_2 + x_3e_3 + x_4e_4 + x_5e_5 = U$ par la méthode du pivot. La système au départ est représenté par les matrices

$$\begin{bmatrix} a_1 & 0 & 0 & 0 & 0 \\ a_2 & 1 & 0 & 0 & 0 \\ a_3 & 0 & 1 & 0 & 0 \\ a_4 & 0 & 0 & 1 & 0 \\ a_5 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix},$$

et la méthode du pivot le transforme en le système équivalent

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u_1/a_1 \\ u_2 - a_2u_1/a_1 \\ u_3 - a_3u_1/a_1 \\ u_4 - a_4u_1/a_1 \\ u_5 - a_5u_1/a_1 \end{bmatrix}.$$

Passons à la deuxième étape dans la preuve du théorème 1. On sait que V_1, e_2, e_3, e_4, e_5 est une base, donc on exprime V_2 sur cette base. En fait on doit utiliser le calcul fait à l'étape précédente si on veut rendre les choses plus précises. On obtient

$$V_2 = b'_1V_1 + b'_2e_2 + b'_3e_3 + b'_4e_4 + b'_5e_5$$

où les b'_i peuvent être obtenus en appliquant la première étape de la méthode du pivot au système linéaire

$$\begin{bmatrix} a_1 & b_1 & 0 & 0 & 0 \\ a_2 & b_2 & 0 & 0 & 0 \\ a_3 & b_3 & 1 & 0 & 0 \\ a_4 & b_4 & 0 & 1 & 0 \\ a_5 & b_5 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}, \tag{4.5}$$

ce qui donne

$$\begin{bmatrix} 1 & b_1/a_1 & 0 & 0 & 0 \\ 0 & b_2 - a_2b_1/a_1 & 0 & 0 & 0 \\ 0 & b_3 - a_3b_1/a_1 & 1 & 0 & 0 \\ 0 & b_4 - a_4b_1/a_1 & 0 & 1 & 0 \\ 0 & b_5 - a_5b_1/a_1 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u_1/a_1 \\ u_2 - a_2u_1/a_1 \\ u_3 - a_3u_1/a_1 \\ u_4 - a_4u_1/a_1 \\ u_5 - a_5u_1/a_1 \end{bmatrix},$$

ou sous forme abrégée

$$\begin{bmatrix} 1 & b'_1 & 0 & 0 & 0 \\ 0 & b'_2 & 0 & 0 & 0 \\ 0 & b'_3 & 1 & 0 & 0 \\ 0 & b'_4 & 0 & 1 & 0 \\ 0 & b'_5 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u'_1 \\ u'_2 \\ u'_3 \\ u'_4 \\ u'_5 \end{bmatrix}. \tag{4.6}$$

Si par exemple $b'_2 \neq 0$ cela permet d'exprimer e_2 comme combinaison linéaire de $\mathrm{V}_1, \mathrm{V}_2, e_3, e_4, e_5$:

$$e_2 = -\frac{b'_1}{b'_2}\mathrm{V}_1 + \frac{1}{b'_2}\mathrm{V}_2 - \frac{b'_3}{b'_2}e_3 - \frac{b'_4}{b'_2}e_4 - \frac{b'_5}{b'_2}e_5 \tag{4.7}$$

puis ensuite n'importe quel vecteur U comme combinaison linéaire de $\mathrm{V}_1, \mathrm{V}_2, e_3, e_4, e_5$. Ceci revient à faire subir la deuxième étape de la méthode du pivot au système linéaire transformé (4.5). Cela donne

$$\begin{bmatrix} 1 & b'_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u'_1 \\ u'_2/b'_2 \\ u'_3 - b'_3u'_2/b'_2 \\ u'_4 - b'_4u'_2/b'_2 \\ u'_5 - b'_5u'_2/b'_2 \end{bmatrix},$$

ou sous forme abrégée

$$\begin{bmatrix} 1 & b'_1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} u'_1 \\ u''_2 \\ u''_3 \\ u''_4 \\ u''_5 \end{bmatrix} \tag{4.8}$$

En fait si on veut obtenir la solution complète, on devra faire une manipulation de lignes supplémentaire pour se ramener au système équivalent :

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u'_1 - b'_1u''_2 \\ u''_2 \\ u''_3 \\ u''_4 \\ u''_5 \end{bmatrix}. \tag{4.9}$$

Dans la troisième étape de la démonstration (c'est la dernière pour cet exemple), on sait que $\mathrm{V}_1, \mathrm{V}_2, e_3, e_4, e_5$ est une base, donc on exprime V_3 sur cette base. Là encore, si on veut clarifier les calculs que demande de faire la preuve, on verra que cela revient à faire subir au système linéaire suivant les transformations successives de la méthode du pivot pour le résoudre :

$$\begin{bmatrix} a_1 & b_1 & c_1 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 \\ a_3 & b_3 & c_3 & 0 & 0 \\ a_4 & b_4 & c_4 & 1 & 0 \\ a_5 & b_5 & c_5 & 0 & 1 \end{bmatrix} \quad \text{et} \quad \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}. \tag{4.10}$$

Commentaire final. Naturellement, comme la solution du système linéaire (4.10) est unique il n'est pas étonnant que la preuve abstraite donne le même résultat que la méthode du pivot. Ce qui l'est par contre, c'est qu'en fait les deux calculs sont essentiellement les mêmes : ce n'est pas parce qu'un problème admet une solution unique que deux méthodes de résolution différentes sont toujours essentiellement les mêmes.

Chapitre 5

Points de repères historiques sur l'infini en mathématiques

Introduction

La rupture introduite par Cantor repose sur une différence d'attitude à l'égard de l'infini. Le point de vue antérieur tient l'infini pour une abstraction de caractère « négatif » : l'ensemble des nombres entiers n'est jamais épuisé. Si on peut parler de tous les nombres entiers sans restriction, cela ne signifie pas qu'ils soient immédiatement disponibles dans leur totalité, on qualifiera donc l'infinité des entiers comme une simple possibilité, comme un *infini potentiel*.

Le point de vue cantorien estime l'existence « d'ensembles infinis en acte » comme assurée, au moins dans un univers mathématique idéal. En conséquence, on peut raisonner avec les ensembles infinis de la même manière qu'avec les ensembles finis, presque sans aucune précaution supplémentaire.

Le point de vue constructif contemporain, qui considère que l'on ne peut démontrer l'existence d'objets mathématiques qu'en donnant une construction de ceux-ci, retourne à une conception négative de l'infini comme un infini potentiel.

Les points de repère qui suivent se veulent tout sauf exhaustifs et objectifs. Tout en essayant de présenter les choses dans leur développement, ils sont écrits par un auteur qui soutient un point de vue constructif. Un tenant de l'infini en acte de Cantor ne donnerait sans doute pas la même description des faits, et ne sélectionnerait pas les mêmes points de repère.

Quelques-uns des sujets abordés ici seront repris dans la suite du cours.

5.1 L'infini chez les mathématiciens grecs

Chez Euclide, une droite n'existe jamais « en entier ». Il y a seulement des segments de droites, qu'on peut prolonger « à volonté ». Le mot « droite » chez Euclide signifie pour nous « segment de droite ». De manière générale aucun infini mathématique (au sens moderne) n'est jamais nommé. Et un segment de droite est un objet à lui tout seul, qui ne peut certainement pas être considéré comme un ensemble infini de points.

La totalité des nombres entiers ne peut pas être nommée, car on admet le principe de base « le tout est plus grand que la partie » qui est contredit par les totalités infinies. Plus tard Galilée reprendra l'argument en faisant remarquer qu'il y a autant d'entiers pairs que de nombres entiers et en concluant qu'il est impossible d'envisager les entiers comme une totalité existante, car sinon « le tout $\mathbb{N}$ devrait être considéré comme égal à la partie (les entiers pairs) » : voir page 76. En fait avant Cantor, les mathématiciens, les plus inconnus comme les plus prestigieux, tel Gauss, refusaient de considérer les ensembles infinis comme des totalités existantes pouvant servir de base au raisonnement mathématique : voir page 77. L'infini était une notion purement négative et ne pouvait être intégré comme un objet positif des mathématiques.

Pour les Grecs, le nombre $\sqrt{2}$ ne peut pas être *nommé*, il n'a pas le statut d'un nombre parce qu'il ne possède pas de description arithmétique finie (ce n'est pas une fraction).

Pour contourner la difficulté Euclide reprend chez Eudoxe une théorie de la comparaison des rapports de grandeurs qui permet de parler de ce que nous appelons aujourd'hui les nombres réels sans jamais les nommer comme « infinis en acte » : si A et B sont des grandeurs de même nature (par exemple deux segments) et C et D sont deux autres grandeurs de même nature (par exemple deux surfaces) on dira que A est à B comme C est à D lorsque, pour n'importe quels couples d'entiers m et n on a :

$$\begin{array}{lcllll} m\mathrm{A} & > & n\mathrm{B} & \Rightarrow & m\mathrm{C} & > n\mathrm{D} \\ m\mathrm{A} & = & n\mathrm{B} & \Rightarrow & m\mathrm{C} & = n\mathrm{D} \\ m\mathrm{A} & < & n\mathrm{B} & \Rightarrow & m\mathrm{C} & < n\mathrm{D} \end{array} \tag{5.1}$$

Il s'ensuit que l'égalité de deux rapports non rationnels ne peut pas se constater par un processus direct simple (il faut *a priori* envisager « n'importe quels couples d'entiers m et n »). D'où la technique de double réduction à l'absurde, qui peut être vue comme une paraphrase de la nécessité de montrer les deux implications extrêmes dans (5.1).

C'est Brouwer qui élucidera d'un point de vue constructif cette question en établissant une nette distinction entre énoncés de caractère positif, qui peuvent être testés par un simple calcul, et énoncés de caractère négatif, qui nécessitent une preuve, parce que l'approche naïve nécessiterait une infinité de calculs. Le fait qu'un rapport de grandeurs est strictement plus grand qu'un autre est un énoncé de caractère positif (il correspond à une constatation du type « $m\mathrm{A} > n\mathrm{B}$ et $m\mathrm{C} < n\mathrm{D}$ ») tandis que l'égalité de deux rapports est un énoncé de caractère négatif (au moins quand le rapport n'est pas rationnel).

Le paradoxe de Zénon sur la flèche qui ne peut atteindre son but car il lui faut toujours parcourir « la moitié du chemin qui lui reste à parcourir » avant d'arriver au but, nous parle aussi de l'infini.

Il peut être interprété comme une réfutation de la possibilité d'existence de l'infini en acte.

Sans doute ce questionnement se situe-t-il (au moins pour nous aujourd'hui) plutôt du côté de la nature du continu en physique. Nous pourrions par exemple interpréter le paradoxe en posant la question : « Y a-t-il vraiment une infinité d'instants en acte qui adviennent durant le parcours de la flèche ? ». La réponse n'est pas claire à la lumière des théories physiques contemporaines. Mais il semble qu'il soit impossible d'envisager une expérience de pensée qui permettrait de donner une réponse positive à la question, car une telle « expérience », censée consommer une énergie infinie dans un intervalle de temps et dans un espace limité, détruirait le dispositif expérimental avant de s'achever.

Pour la plupart des mathématiciens aujourd'hui, la question de savoir si l'infinité des nombres envisagés par Zénon dans ce paradoxe, à savoir $1/2, 3/4, 7/8, \ldots, (2^n-1)/2^n, \ldots$ existe bel et bien, disponible pour nous, ne se pose pas tant la réponse est évidemment « oui ». Et le paradoxe semble facile à réfuter, puisque :

$$1/2 + 1/4 + 1/8 + \cdots + 1/2^n + \cdots = 1$$

Donc l'infinité des petits laps de temps donne une somme finie : le moment où la flèche atteint son but. Dans cette réfutation du paradoxe, le mathématicien n'a pas l'air de se rendre compte qu'il croit résoudre un problème sur la nature de l'infini simplement en posant une définition : celle de la somme d'une série convergente. Mais une définition ne constitue pas un argument décisif dans un débat. Tout juste permet-elle de l'éclairer. Poser la définition est certes légitime, et fort utile. Justement parce que jamais personne n'écrira jamais la somme infinie en entier.

Une réponse plus acceptable serait donc : cette question ne nous concerne pas vraiment, car nous savons comment la contourner. Peu importe que les termes de la somme infinie soient immédiatement disponibles ou pas, puisque nous savons donner un sens, par une définition

judicieuse, à l'égalité qui pose *a priori* problème. Les trois petits points qui terminent la somme n'ont pas besoin de représenter des nombres effectivement présents pour qu'on puisse définir la somme, et raisonner avec cela de manière relativement sûre.

Ajoutons que la définition elle-même n'utilise qu'une infinité potentielle d'objets, et tout calcul mis en œuvre lorsqu'on utilise cette définition reste un calcul fini.

5.2 La crise des infinitésimaux

L'efficacité du calcul différentiel et de la méthode des infinitésimaux développés aux 17e et 18e siècles prépare le terrain pour la reconnaissance d'un statut « ordinaire » accordé à l'infini. C'est en raisonnant sur des infiniment grands et infiniment petits *comme si* c'était des quantités finies, c'est à dire en se débarrassant des scrupules imposés par la tradition grecque, qu'on obtient, par des méthodes uniformes et extraordinairement efficaces, des résultats qui auparavant étaient inaccessibles ou dans le meilleur des cas demandaient une ingéniosité toujours renouvelée. Les notions de nombre réel et de fonction s'avèrent des notions fondamentales. Cependant, on ne trouve pas de justification convaincante pour ces objets qui sont des sortes d'infinis en acte.

La crise sera résolue au 19e siècle en chassant les infinitésimaux au profit de la notion de limite, puis en décrétant légitimes les infinis en acte qui semblent nécessaires pour les nombres réels et les fonctions continues.

Il faudra attendre l'analyse non standard d'Abraham Robinson dans les années 1960 pour s'apercevoir que les infinitésimaux ne sont ni plus ni moins sulfureux que les nombres réels. Dans sa version « Internal set theory » [Théorie des ensembles internes], développée par Edward Nelson (1977), l'analyse non standard est aujourd'hui une théorie dont le langage se rapproche beaucoup plus de celui des physiciens que ne le font l'analyse constructive ou l'analyse classique : voir le livre de Diener et Reeb (1989).

5.3 La crise des géométries non euclidiennes

L'existence des géométries non euclidiennes, tout aussi cohérentes que la géométrie euclidienne, introduit un bouleversement conceptuel concernant la notion de vérité, et la place des mathématiques.

Avant les géométries non euclidiennes, on pouvait croire à l'existence d'une vérité géométrique « absolue », que l'esprit humain, par la seule force de son raisonnement, arrivait à maitriser.

Après, tout change. L'esprit humain ne fait plus que proposer des modèles mathématiques imparfaits pour une description proprement humaine d'une réalité extérieure fuyante et insaisissable dans sa totalité. Il va maintenant falloir mesurer la somme des angles de très grands triangles et regarder des étoiles cachées derrière le soleil pour connaitre la géométrie de l'univers dans lequel on vit. La vérité perd son caractère absolu pour devenir une construction proprement humaine et donc beaucoup plus relative. À tous points de vue, nous ne sommes plus au centre du monde, mais bel et bien perdus dans un coin de galaxie.

5.4 Cantor et l'avènement de l'infini en acte

On a du mal à imaginer à quel point Cantor heurtait de front les conceptions établies en introduisant la théorie des ensembles « infinis en acte ». C'était à l'époque un véritable acte de foi.

Des affirmations telles que « l'ensemble $\mathbb{N}$ est strictement plus petit que l'ensemble $\mathbb{R}$ » semblaient relever de la folie. Et Cantor lui-même avait du mal à admettre que $\mathbb{R}$ et $\mathbb{R}^2$ ont le même cardinal : le 29 juin 1877, il écrit à ce sujet à Richard Dedekind.

> Veuillez excuser mon zèle pour cette affaire, si je fais appel tellement souvent à votre bonté et à votre peine ; ce que je vous ai communiqué tout récemment est pour moi-même si inattendu, si nouveau, que je ne pourrai pour ainsi dire pas arriver à une certaine tranquillité d'esprit avant que je n'aie reçu, très honoré ami, votre jugement sur son exactitude. Tant que vous ne m'aurez pas approuvé, je ne puis que dire : Je le vois, mais je ne le crois pas [en français dans le texte]. (« Correspondance Cantor-Dedekind » 1962, page 211.)

Cependant, les infinis en acte de Cantor ont permis de donner les premières « constructions » en toute généralité des nombres réels. Admettre que les nombres réels formaient un ensemble parmi d'autres, une totalité objective et en acte, suffisait à décoincer une situation qui en avait bien besoin.

Que le continu (en mathématiques) soit l'ensemble de ses points et rien d'autre nous parait aujourd'hui une idée naturelle, tellement toute question sur la nature du continu est désormais évacuée du champ même de la réflexion par la simple logique du discours : $\mathbb{R}$ est un ensemble de points, point à la ligne.

La définition générale de la notion de fonction a semblé elle aussi grandement facilitée par la liberté qu'on se donnait de manipuler des ensembles infinis. Plutôt que se poser des questions à n'en plus finir sur ce qu'on fait exactement avec le calcul différentiel, il suffit de définir une fonction réelle par son graphe, qui n'est rien d'autre qu'un ensemble de points dans $\mathbb{R}^2$.

Et Cantor a eu rapidement des adeptes enthousiastes parmi les plus grand mathématiciens. En particulier Hilbert, qui fut sans doute le plus influent dans la première moitié du 20$^{\text{e}}$ siècle, fit la déclaration suivante : « *Du paradis que Cantor a créé pour nous, personne ne doit pouvoir nous chasser.* » (Hilbert 1926).

5.5 Les paradoxes de la théorie des ensembles

Les constructions d'infinis autorisés par la théorie des ensembles de Cantor sont *a priori* aberrantes parce qu'elles permettent de monter beaucoup trop loin dans l'échelle des infinis.

Tout ensemble est strictement plus petit que l'ensemble des parties. En partant de $\mathbb{N}$ vous obtenez donc déjà une suite infinie d'ensembles toujours plus gros, en prenant à chaque fois l'ensemble des parties du précédent : $\mathbb{N}, \mathfrak{P}(\mathbb{N}), \mathfrak{P}(\mathfrak{P}(\mathbb{N})), \ldots$

Mais il y a pire : tout ensemble S peut être muni d'une relation de bon ordre : une relation d'ordre total pour laquelle toute partie non vide de S possède un élément minimum. Une relation de bon ordre permet de faire des constructions par récurrence transfinie le long du bon ordre en question.

Imaginez par exemple que vous mettez une relation de bon ordre sur $\mathbb{R}$, puis que par récurrence transfinie sur ce bon ordre vous construisez à chaque étape l'ensemble des parties de la réunion des ensembles précédemment construits. En ayant démarré avec l'ensemble $\mathbb{N}$, le plus petit des infinis, au bout du compte vous obtenez quelque chose de vraiment très très très gros, et dont l'utilité mathématique est tout à fait contestable.

En outre, personne n'a jamais réussi à « faire voir » une relation de bon ordre sur $\mathbb{R}$.

Certains ensembles infinis s'avèrent de toute manière trop grands, trop infinis : par exemple

$$\mathrm{X} = \{x \mid x \text{ est un ensemble tel que } x \notin x\}\,.$$

C'est un ensemble pour lequel on a (comme conséquence directe de la définition) :

$$\mathrm{X} \in \mathrm{X} \;\Leftrightarrow\; \mathrm{X} \notin \mathrm{X}.$$

On se tirera d'affaire en décrétant que les ensembles trop infinis ne sont pas des ensembles,

mais des classes[1]. La théorie formelle correspondante n'a depuis donné lieu à aucune contradiction logique.

Par ailleurs Brouwer développe au début du 20e siècle une réfutation systématique de la notion d'infini en acte et met en pratique des mathématiques où le principe du tiers exclu n'est plus admis pour les affirmations correspondant à une infinité de vérifications. La logique sous-jacente sera formalisée par Heyting, sous le nom de *logique intuitionniste*.

Expliquons brièvement les difficultés que Brouwer voit dans le principe du tiers exclu. Par exemple lorsque l'on démontre que toute partie non vide A de $\mathbb{N}$ admet un plus petit élément on dit :

1. je considère un a dans A,
2. ou bien il n'y a pas de plus petit élément que a dans A, et A possède a comme plus petit élément,
3. ou bien il y a un b dans A plus petit que a. On recommence le même raisonnement avec b, et on aboutit en un temps fini au plus petit élément de A car en dessous de a il n'y a que a éléments dans $\mathbb{N}$.

Ce raisonnement est convaincant du point de vue de la logique du tiers exclu :

1. ou bien a est le plus petit élément de A, ou bien il n'est pas le plus petit !

Autrement dit : une affirmation mathématique ayant une signification claire est nécessairement vraie ou fausse « dans l'absolu » : il n'y a pas de troisième cas possible, le tiers cas est exclu.

Cependant si on considère une partie A de $\mathbb{N}$ définie de manière un peu compliquée, il est fréquent qu'il soit tout à fait improbable de savoir déterminer le plus petit élément de A, par exemple parce qu'il faudrait pour cela résoudre d'abord un problème mathématique très difficile.

Dans ce cas le : « ou bien oui, ou bien non » (principe du tiers exclu) n'est pas considéré comme légitime par Brouwer qui dit que l'existence d'un objet n'est assurée que si l'on donne une méthode sûre qui permet de le trouver. Il s'agit en fait d'une remise en cause de la notion de vérité absolue. Pour Brouwer, non seulement les objets usuels des mathématiques doivent être construits (ils ne peuvent être de pures idéalités abstraites), mais la vérité elle-même doit être construite.

Pour essayer d'y voir clair Hilbert propose une méthode formaliste : au lieu de comprendre exactement ce que sont les objets étranges qui apparaissent dans la théorie des ensembles à la Cantor, essayons de comprendre la théorie mathématique de ces objets étranges. Une théorie mathématique formelle est un objet assez simple qui peut être étudié, par des méthodes mathématiques, pour elle-même.

Une théorie mathématique formelle pour les ensembles de Cantor est proposée par Zermelo, puis étendue par Skolem et Fraenkel. On parlera de la théorie de Zermelo-Fraenkel, ou théorie **ZF**.

Poincaré développera (parallèlement à Brouwer mais pas exactement avec les mêmes arguments) une critique en règle de la théorie formelle de Zermelo : voici un extrait de *La logique de l'infini* (Poincaré 1909, réédité dans Poincaré 1913).

> M. Zermelo a voulu construire un système impeccable d'axiomes ; mais ces axiomes ne peuvent être regardés comme des décrets arbitraires, puisqu'il faudrait démontrer que ces décrets ne sont pas contradictoires, et qu'ayant fait entièrement table rase on n'a plus rien sur quoi l'on puisse appuyer une semblable démonstration. Il faut donc que ces axiomes soient évidents par eux-mêmes. Or quel est le mécanisme par lequel on les a construits ? on a pris les axiomes qui sont vrais des collections finies ; on ne pouvait les étendre tous aux collections infinies, on n'a fait cette extension que pour un certain nombre d'entre eux, choisis plus ou moins arbitrairement. À mon sens, d'ailleurs, ainsi

1. Un ensemble est une classe pas trop infinie. Il peut être élément d'une classe, mais les classes « trop infinies » ne sont éléments d'aucune classe.

que je l'ai dit plus haut, aucune proposition concernant les collections infinies ne peut être évidente par intuition.

. . . . . .

Quant à moi, je proposerais de s'en tenir aux règles suivantes :

1. Ne jamais envisager que des objets susceptibles d'être définis en un nombre fini de mots ;
2. Ne jamais perdre de vue que toute proposition sur l'infini doit être la traduction, l'énoncé abrégé de propositions sur le fini ;
3. Éviter les classifications et les définitions non-prédicatives.

Toutes les recherches dont nous avons parlé ont un caractère commun. On se propose d'enseigner les mathématiques à un élève qui ne sait pas encore la différence qu'il y a entre l'infini et le fini ; on ne se hâte pas de lui apprendre en quoi consiste cette différence ; on commence par lui montrer tout ce qu'on peut savoir de l'infini sans se préoccuper de cette distinction ; puis dans une région écartée du champ qu'on lui a fait parcourir, on lui découvre un petit coin où se cachent les nombres finis.

Cela me parait psychologiquement faux ; ce n'est pas ainsi que l'esprit humain procède naturellement, et quand même on devrait s'en tirer sans trop de mésaventures antinomiques, cela n'en serait pas moins une méthode contraire à toute saine psychologie.

M. Russell me dira sans doute qu'il ne s'agit pas de psychologie, mais de logique et d'épistémologie ; et moi, je serai conduit à répondre qu'il n'y a pas de logique et d'épistémologie indépendantes de la psychologie ; et cette profession de foi clora probablement la discussion parce qu'elle mettra en évidence une irrémédiable divergence de vues. (Poincaré 1909, pages 482-483.)

5.6 Les avatars de l'hypothèse du continu

La comparaison des infinis selon leur taille va réserver quelques surprises.

Cantor passera la fin de sa vie à essayer, sans succès, de démontrer « l'hypothèse du continu » : il n'y a pas de cardinal strictement compris entre celui de $\mathbb{N}$ et celui de $\mathbb{R}$.

Skolem démontrera que toute théorie mathématique formelle cohérente possède un modèle, c'est-à-dire une réalisation à l'égard de laquelle ses énoncés sont vrais, dont le cardinal est celui de $\mathbb{N}$. En conséquence si la théorie **ZF** est cohérente, dans un modèle à la Skolem de cette théorie, $\mathbb{R}$ n'admet pas « plus » d'éléments que $\mathbb{N}$, et il existe une bijection entre $\mathbb{N}$ et $\mathbb{R}$... mais elle ne fait pas partie du modèle. Cela commence à faire mal au crâne.

Plus tard K. Gödel et P. Cohen démontreront que l'hypothèse du continu est indépendante des axiomes de **ZF**. Face à ce genre de résultats, deux attitudes opposées sont développées.

Les uns, comme Skolem, estiment que la comparaison des cardinaux a un caractère très relatif. Pour eux, l'infini ne correspond à aucune réalité objective et est simplement une manière de parler.

Les autres, comme Gödel, s'en tiennent à un point de vue « réaliste » : un univers mathématique cantorien existe bel et bien, et dans cet univers, l'hypothèse du continu est nécessairement vraie ou fausse. En conséquence, ils estiment que de nouveaux axiomes raisonnables doivent être cherchés, qui décideront vraie ou fausse l'hypothèse du continu.

Du point de vue constructif développé par Brouwer, $\mathbb{R}$ est un infini potentiel *plus compliqué (et non pas plus grand)* que $\mathbb{N}$. Mais il y a aussi des parties de $\mathbb{N}$ qui sont plus compliquées que $\mathbb{N}$, et en particulier il n'est pas absurde de supposer que l'infini potentiel $\mathbb{R}$ est isomorphe à une partie compliquée de l'infini potentiel simple $\mathbb{N}$ (cette hypothèse est admise comme vraie par l'école constructive russe héritière de Markov).

5.7 Le programme de Hilbert

Hilbert tient les critiques de Brouwer pour sérieuses. Mais ce qui lui importe, c'est avant tout de faire des mathématiques. Or, dit-il, les mathématiques sont plus faciles à faire dans le paradis de Cantor que dans l'enfer de Brouwer.

Hilbert tient le langage suivant : finalement peu nous importe de savoir si les infinis en acte existent ou non, si l'hypothèse du continu a un sens ou si elle n'en a pas ; ce qui nous importe, c'est de savoir si, en utilisant la théorie des ensembles infinis, on est assuré de ne jamais démontrer des énoncés qui ont du sens mais qui seraient faux. C'est exactement la même attitude que vis à vis des nombres imaginaires qui servent à trouver la racine réelle d'une équation du troisième degré : l'important est avant tout que, une fois le calcul terminé, et les nombres imaginaires évaporés, le résultat soit juste. Autrement dit, si on pouvait réduire l'infini mathématique à n'être qu'une manière de parler, on serait pleinement satisfaits.

À défaut de pouvoir élucider « la sémantique » des ensembles infinis (quelle peut bien être leur signification objective ?), essayons au moins de comprendre « leur syntaxe » : comprendre ce qu'on fait exactement quand on les utilise en mathématiques.

Pour cela, il faut considérer une théorie purement formelle, dans laquelle on puisse librement utiliser les infinis de Cantor d'une part, et dans laquelle on puisse écrire les énoncés mathématiques qui « ont surement du sens », d'autre part.

Ensuite, il faut, par des méthodes convaincantes, démontrer que tout énoncé « ayant du sens » et démontré comme vrai dans la théorie formelle, est vrai dans la réalité.

La première théorie à tester est une théorie très simple (la théorie formelle dite *de Peano*) où les seuls objets qui interviennent à titre de variables sont des nombres entiers, et où on peut seulement écrire des énoncés relativement simples les concernant. Mais la logique du tiers exclu est admise, même pour des énoncés correspondant à une infinité de vérifications. Ensuite, il faudra passer à des théories plus sophistiquées, en espérant un jour traiter la théorie des ensembles elle-même.

En ce qui concerne les méthodes convaincantes auxquelles il vient d'être fait allusion, Hilbert les voulait vraiment très élémentaires (finitistes dans sa terminologie). En particulier, elles devaient être plus simples que les méthodes de preuves formalisées dans la théorie de Peano.

Gödel apportera deux réponses contradictoires [2] au problème posé.

La première réponse (dans l'ordre chronologique) est que, pour toute théorie formelle qui prétend décrire au moins $\mathbb{N}$, certains énoncés vrais dans la réalité sont indémontrables dans la théorie formelle (premier théorème d'incomplétude). Pire encore la cohérence de la théorie ne peut pas être démontrée avec les seuls moyens formalisés dans la théorie (deuxième théorème d'incomplétude).

A fortiori, la théorie des ensembles semble bien ne jamais pouvoir être prouvée consistante dans la mesure où on a incorporé dans son formalisme toutes les méthodes de démonstrations connues, mêmes celles qui sont le plus douteuses.

En fait, le théorème d'incomplétude de Gödel dit qu'il est impossible de capturer la complexité de $\mathbb{N}$, le plus simple des infinis, à l'intérieur d'une théorie formelle à la Hilbert.

La deuxième réponse est que, pour ce qui concerne la théorie de Peano, la logique *avec tiers exclu* n'introduit aucune contradiction. Autrement dit, si on pense que la théorie de Peano sans tiers exclu (l'arithmétique de Heyting) est consistante, on est assuré que la théorie de Peano l'est également.

Gödel donnera plusieurs interprétations constructives pour les théorèmes et pour les preuves de la théorie de Peano. Ces résultats, affinés depuis, constituent une réalisation partielle du programme de Hilbert, au moins si on admet comme pleinement convaincantes les méthodes constructives (et pas seulement les méthodes finitistes), puisqu'on s'accorde sur le fait que l'arithmétique de Heyting est une bonne formalisation de méthodes constructives appliquées à l'arithmétique.

2. Contradictoires seulement au premier regard, naturellement.

Néanmoins, on n'a pas réussi pour le moment à faire un travail analogue pour des théories formelles beaucoup plus sophistiquées que celle de Peano ; et cette dernière est trop pauvre pour pouvoir être prise comme base du travail mathématique ordinaire.

Une réponse d'une toute autre nature est apportée par Bishop (1967) avec son livre d'analyse constructive. Cantor n'est pas le paradis, et Brouwer n'est pas l'enfer. En ne mettant en œuvre que les idées les plus incontestables de Brouwer, on fait une mathématique, parfois un peu plus difficile[3], mais où tous les énoncés ont du sens (leur signification en dernier ressort est toujours qu'un certain calcul fini aboutit à un certain résultat), où tous les théorèmes sont incontestables, et où, simplement, on regarde un peu plus en détail la signification algorithmique réelle des énoncés classiques.

5.8 Le point de vue formaliste

L'introduction du point de vue formaliste en mathématiques a été un phénomène à double tranchant.

D'une part, cela a permis de développer le point de vue des « structures ». Prenons par exemple la théorie des corps. Jusqu'à l'aube du 20^{e} siècle, on ne connaissait que des corps particuliers : le corps des nombres rationnels, celui des nombres réels, celui des fonctions méromorphes (les quotients de fonctions entières), les corps finis (introduits par Galois)... Déjà l'idée de considérer le corps des racines d'un polynôme, et tous les sous-corps de ce corps, avait permis à Galois[4] de faire un progrès extraordinaire dans la théorie de la résolution des équations algébriques. La mise en place d'une théorie générale des corps, basée sur les axiomes aujourd'hui usuels[5] a permis de prendre encore du recul et de faire bénéficier de nombreux domaines des mathématiques des résultats généraux de cette théorie.

Cette méthode des structures, impulsée en grande partie par Hilbert, est un puissant outil de compréhension et d'unification en mathématiques.

Mais il y a un autre aspect de la méthode formaliste, qui est la volonté de figer une fois pour toutes, à l'intérieur d'un système formel précis, le cadre dans lequel devraient se développer les mathématiques. L'avantage bien illusoire d'une telle position est d'éviter les questions des fondements.

Au lieu de considérer un tel système formel comme un objet d'étude, à mettre en concurrence avec d'autres pour comparer leurs mérites respectifs, on le sacralise en décrétant que désormais « les règles du jeu sont fixées » et que personne ne doit s'aviser de les contester.

Il s'agit d'une attitude de repli frileux, bizarrement adoptée par Bourbaki dans l'introduction de son monumental Traité, et qui n'a en réalité que bien peu d'adeptes[6]. On peut la résumer par « courage, fuyons ! », fuyons la question du sens puisqu'elle nous embête.

Vous croyiez naïvement faire *des* mathématiques dans le but de construire des modèles abstraits pour le mouvement des planètes, la propagation de la chaleur où les éruptions volcaniques.

Vous vous trompiez, vous êtes en train de jouer à un nouveau jeu, fort intelligent, très sophistiqué, qui s'appelle *la* mathématique, et dont les règles ont été édictées par quelques très grands savants dans leur infinie sagesse (Bourbaki). Les situations (ou énoncés) gagnant(e)s de ce jeu subtil s'appellent « théorèmes ».

Amusez-vous bien. Plus les théorèmes démontrés auront l'air d'avoir du sens et plus ce sera amusant.

Mais ne discutez pas la question du sens. La seule signification indiscutable d'un théorème c'est qu'il a été démontré en respectant scrupuleusement les règles du jeu.

3. Mais ceci est sans doute pour l'essentiel dû au dépaysement.
4. Galois utilisait la belle terminologie de « champ de rationalité » pour désigner un sous-corps de $\mathbb{C}$.
5. La première mise en forme de la théorie est due à Steinitz.
6. La majorité des mathématiciens, y compris sans nul doute les rédacteurs de Bourbaki, est sur une position proche de celle de Gödel.

Un effet étonnamment pervers de ce point de vue formaliste « à outrance » est qu'il réduit toutes les mathématiques à n'être plus qu'une partie infime de ce qu'elles devraient être, puisque la recherche des théorèmes d'une théorie formelle est seulement l'étude des valeurs d'une fonction particulière, de l'ensemble $\mathbb{N}$ vers l'ensemble des énoncés bien formés de la théorie : la fonction qui énumère les théorèmes démontrés en respectant les règles du jeu.

5.9 Et demain ?

La situation actuelle semble dominée par le pragmatisme.

Dans la mesure où les résultats de nature effective sont de plus en plus recherchés, on peut penser que la méthodologie constructive va finir par s'imposer comme la méthodologie normale. Cette méthodologie, remarquablement mise en œuvre par Bishop (1967) dans son livre d'analyse constructive, est basée sur l'idée que l'infini ne peut pas être considéré comme une réalité achevée et que rien ne permet de croire en un univers cantorien idéal qui serait la garantie du fondement des mathématiques. Il faut être beaucoup plus modeste et n'affirmer que ce à quoi on est capable d'assigner une signification indiscutable. En renonçant à la logique du tiers exclu, qui obscurcit la question du sens dès que les énoncés manipulent des infinis, on fait des mathématiques plus sûres, avec lesquelles tout théorème a en dernière analyse la signification qu'un calcul aboutit au bout d'un temps fini à une certain type de résultat.

Néanmoins, les facilités offertes par le tiers exclu et l'axiome du choix sont grandes, et les fabricants de théorèmes ne sont pas prêts à lâcher la proie pour (ce qu'ils pensent n'être que) l'ombre.

En fait, si on adopte le point de vue constructif, le programme de Hilbert garde toute sa valeur mais il reste un énorme travail à faire pour le réaliser : démontrer que, si on délimite convenablement le cadre de travail, l'analyse et l'algèbre classiques, voire non standards, ne conduiront jamais à des résultats faux sur les énoncés qui ont du sens. Ceci est déjà fait pour une partie significative de l'analyse et de l'algèbre dans des livres de mathématiques constructives.

Annexe 1. Henri Poincaré, *Les mathématiques et la logique*, un texte qui critique le formalisme

Le texte ci-dessous est constitué d'extraits de l'article Poincaré 1905 tel qu'il a été repris dans le recueil Poincaré 1906, livre II, chapitre III, pages 152-171.

Introduction

Les mathématiques peuvent-elles être réduites la logique sans avoir à faire appel à des principes qui leur soient propres ? Il y a toute une école, pleine d'ardeur et de foi, qui s'efforce de l'établir. Elle a son langage spécial où il n'y a plus de mots et où on ne fait usage que de signes. Ce langage n'est compris que de quelques initiés, de sorte que les profanes sont disposés à s'incliner devant les affirmations tranchantes des adeptes. Il n'est peut-être pas inutile d'examiner ces affirmations d'un peu près, afin de voir si elles justifient le ton péremptoire avec lequel elles sont présentées.

Mais pour bien faire comprendre la nature de la question, il est nécessaire d'entrer dans quelques détails historiques et de rappeler en particulier le caractère des travaux de Cantor.

. . . . . .

De nombreux mathématiciens se sont lancés sur ses traces et se sont posé une série de questions de même genre. Ils se sont tellement familiarisés avec les nombres transfinis qu'ils en sont arrivés à faire dépendre la théorie des nombres finis de celle des nombres cardinaux de Cantor. À leurs yeux,

pour enseigner l'arithmétique d'une façon vraiment logique, on devrait commencer par établir les propriétés générales des nombres cardinaux transfinis, puis distinguer parmi eux une toute petite classe, celle des nombres entiers ordinaires. Grâce à ce détour on pourrait arriver à démontrer toutes les propositions relatives à cette petite classe (c'est-à-dire toute notre arithmétique et notre algèbre) sans se servir d'aucun principe étranger à la logique.

Cette méthode est évidemment contraire à toute saine psychologie ; ce n'est certainement pas comme cela que l'esprit humain a procédé pour construire les mathématiques ; aussi ses auteurs ne songent-ils pas, je pense, à l'introduire dans l'enseignement secondaire. Mais est-elle du moins logique, ou pour mieux dire est-elle correcte ? Il est permis d'en douter.

. . . . . .

II

Ce qui nous frappe d'abord dans la nouvelle mathématique, c'est son caractère purement formel : « Pensons, dit Hilbert, trois sortes de *choses* que nous appellerons points, droites et plans, convenons qu'une droite sera déterminée par deux points et qu'au lieu de dire que cette droite est déterminée par ces deux points, nous pourrons dire qu'elle passe par ces deux points ou que ces deux points sont situés sur cette droite ». Que sont ces *choses*, non seulement nous n'en savons rien, mais nous ne devons pas chercher à le savoir. Nous n'en avons pas besoin, et quelqu'un, qui n'aurait jamais vu ni point, ni droite, ni plan pourrait faire de la géométrie tout aussi bien que nous. Que le mot *passer par*, ou le mot *être situé sur* ne provoquent en nous aucune image, le premier est simplement synonyme de *être déterminé* et le second de *déterminer*.

. . . . . .

Ce que Hilbert avait fait pour la géométrie, d'autres ont voulu le faire pour l'arithmétique et pour l'analyse. Si même ils y avaient entièrement réussi, les Kantiens seraient-ils définitivement condamnés au silence ? Peut-être pas, car en réduisant la pensée mathématique à une forme vide, il est certain qu'on la mutile. Admettons même que l'on ait établi que tous les théorèmes peuvent se déduire par des procédés purement analytiques, par de simples combinaisons logiques d'un nombre fini d'axiomes, et que ces axiomes ne sont que des conventions. Le philosophe conserverait le droit de rechercher les origines de ces conventions, de voir pourquoi elles ont été jugées préférables aux conventions contraires.

Et puis la correction logique des raisonnements qui mènent des axiomes aux théorèmes n'est pas la seule chose dont nous devions nous préoccuper. Les règles de la parfaite logique sont-elles toute la mathématique ? Autant dire que tout l'art du joueur d'échecs se réduit aux règles de la marche des pièces. Parmi toutes les constructions que l'on peut combiner avec les matériaux fournis par la logique, il faut faire un choix ; le vrai géomètre fait ce choix judicieusement parce qu'il est guidé par un sûr instinct, ou par quelque vague conscience de je ne sais quelle géométrie plus profonde, et plus cachée, qui seule fait le prix de l'édifice construit.

Chercher l'origine de cet instinct, étudier les lois de cette géométrie profonde qui se sentent et ne s'énoncent pas, ce serait encore une belle tâche pour les philosophes qui ne veulent pas que la logique soit tout. Mais ce n'est pas à ce point de vue que je veux me placer, ce n'est pas ainsi que je veux poser la question. Cet instinct dont nous venons de parler est nécessaire à l'inventeur, mais il semble d'abord qu'on pourrait s'en passer pour étudier la science une fois créée. Eh bien, ce que je veux rechercher, c'est s'il est vrai qu'une fois admis les principes de la logique, on peut je ne dis pas découvrir, mais démontrer toutes les vérités mathématiques sans faire de nouveau appel à l'intuition.

III

À cette question, j'avais autrefois répondu que non (Voir *Science et Hypothèse*, chapitre I^{er}) ; notre réponse doit-elle être modifiée par les travaux récents ? Si j'avais répondu non, c'est parce

que « le principe d'induction complète » me paraissait à la fois nécessaire au mathématicien et irréductible à la logique. On sait quel est l'énoncé de ce principe :

« Si une propriété est vraie du nombre 1, et si l'on établit qu'elle est vraie de $n+1$ pourvu qu'elle le soit de n, elle sera vraie de tous les nombres entiers. » J'y voyais le raisonnement mathématique par excellence. Je ne voulais pas dire, comme on l'a cru, que tous les raisonnements mathématiques peuvent se réduire à une application de ce principe. En examinant ces raisonnements d'un peu près, on y verrait appliqués beaucoup d'autres principes analogues, présentant les mêmes caractères essentiels. Dans cette catégorie de principes, celui de l'induction complète est seulement le plus simple de tous et c'est pour cela que je l'ai choisi pour type.

Le nom de principe d'induction complète qui a prévalu n'est pas justifié. Ce mode de raisonnement n'en est pas moins une véritable induction mathématique qui ne diffère de l'induction ordinaire que par sa certitude.

IV Définitions et axiomes

. . . . . .

Eh bien, les logiciens admettent pour le principe d'induction complète, ce que j'admets pour le postulatum d'Euclide, ils ne veulent y voir qu'une définition déguisée.

Mais pour qu'on ait ce droit, il y a deux conditions à remplir. Stuart Mill disait que toute définition implique un axiome, celui par lequel on affirme l'existence de l'objet défini. À ce compte, ce ne serait plus l'axiome qui pourrait être une définition déguisée, ce serait au contraire la définition qui serait un axiome déguisé. Stuart Mill entendait le mot existence dans un sens matériel et empirique ; il voulait dire qu'en définissant le cercle, on affirme qu'il y a des choses rondes dans la nature.

Sous cette forme, son opinion est inadmissible. Les mathématiques sont indépendantes de l'existence des objets matériels ; en mathématiques le mot exister ne peut avoir qu'un sens, il signifie exempt de contradiction. Ainsi rectifiée, la pensée de Stuart Mill devient exacte ; en définissant un objet, on affirme que la définition n'implique pas contradiction.

Si nous avons donc un système de postulats, et si nous pouvons démontrer que ces postulats n'impliquent pas contradiction, nous aurons le droit de les considérer comme représentant la définition de l'une des notions qui y figurent. Si nous ne pouvons pas démontrer cela, il faut que nous l'admettions sans démonstration et cela sera alors un axiome ; de sorte que si nous voulions chercher la définition sous le postulat, nous retrouverions encore l'axiome sous la définition.

. . . . . .

Pour établir que les postulats n'impliquent pas contradiction, il faut alors envisager toutes les propositions que l'on peut déduire de ces postulats considérés comme prémisses et montrer que, parmi ces propositions, il n'y en a pas deux dont l'une soit la contradictoire de l'autre. Si ces propositions sont en nombre fini, une vérification directe est possible. Ce cas est peu fréquent et d'ailleurs peu intéressant.

Si ces propositions sont en nombre infini, on ne peut plus faire cette vérification directe ; il faut recourir à des procédés de démonstration où en général on sera forcé d'invoquer ce principe d'induction complète qu'il s'agit précisément de vérifier.

Nous venons d'expliquer l'une des conditions auxquelles les logiciens devaient satisfaire, *et nous verrons plus loin qu'ils ne l'ont pas fait.*

V

Il y en a une seconde. Quand nous donnons une définition, c'est pour nous en servir.

Nous retrouverons donc dans la suite du discours le mot défini ; avons-nous le droit d'affirmer, de l'objet représenté par ce mot, le postulat qui a servi de définition ? Oui, évidemment, si le mot a conservé son sens, si nous ne lui attribuons pas implicitement un sens différent. Or c'est ce qui arrive quelquefois et il est le plus souvent difficile de s'en apercevoir ; il faut voir comment ce mot

s'est introduit dans notre discours, et si la porte par laquelle il est entré n'implique pas en réalité une autre définition que celle qu'on a énoncée.

Cette difficulté se présente dans toutes les applications des mathématiques. La notion mathématique a reçu une définition très épurée et très rigoureuse ; et pour le mathématicien pur toute hésitation a disparu ; mais si on veut l'appliquer aux sciences physiques par exemple, ce n'est plus à cette notion pure que l'on a affaire, mais à un objet concret qui n'en est souvent qu'une image grossière. Dire que cet objet satisfait, au moins approximativement, à la définition, c'est énoncer une vérité nouvelle, que l'expérience peut seule mettre hors de doute, et qui n'a plus le caractère d'un postulat conventionnel.

Mais, sans sortir des mathématiques pures, on rencontre encore la même difficulté.

Vous donnez du nombre une définition subtile ; puis, une fois cette définition donnée, vous n'y pensez plus ; parce qu'en réalité, ce n'est pas elle qui vous a appris ce que c'était que le nombre, vous le saviez depuis longtemps, et quand le mot nombre se retrouve plus loin sous votre plume, vous y attachez le même sens que le premier venu ; pour savoir quel est ce sens et s'il est bien le même dans telle phrase ou dans telle autre, il faut voir comment vous avez été amené à parler de nombre et à introduire ce mot dans ces deux phrases. Je ne m'explique pas davantage sur ce point pour le moment, car nous aurons l'occasion d'y revenir.

Ainsi voici un mot dont nous avons donné explicitement une définition A ; nous en faisons ensuite dans le discours un usage qui suppose implicitement une autre définition B. Il est possible que ces deux définitions désignent un même objet. Mais qu'il en soit ainsi, c'est une vérité nouvelle, qu'il faut, ou bien démontrer, ou bien admettre comme un axiome indépendant.

Nous verrons plus loin que les logiciens n'ont pas mieux rempli la seconde condition que la première.

VI

Les définitions du nombre sont très nombreuses et très diverses ; je renonce à énumérer même les noms de leurs auteurs. Nous ne devons pas nous étonner qu'il y en ait tant. Si l'une d'elles était satisfaisante, on n'en donnerait plus de nouvelle. Si chaque nouveau philosophe qui s'est occupé de cette question a cru devoir en inventer une autre, c'est qu'il n'était pas satisfait de celles de ses devanciers, et s'il n'en était pas satisfait, c'est qu'il croyait y apercevoir une pétition de principe.

J'ai toujours éprouvé, en lisant les écrits consacrés à ce problème, un profond sentiment de malaise ; je m'attendais toujours à me heurter à une pétition de principe et, quand je ne l'apercevais pas tout de suite, j'avais la crainte d'avoir mal regardé.

C'est qu'il est impossible de donner une définition sans énoncer une phrase, et difficile d'énoncer une phrase sans y mettre un nom de nombre, ou au moins le mot plusieurs, ou au moins un mot au pluriel. Et alors la pente est glissante et à chaque instant on risque de tomber dans la pétition de principe.

Je ne m'attacherai dans la suite qu'à celles de ces définitions où la pétition de principe est le plus habilement dissimulée.

VII La pasigraphie

Le langage symbolique créé par M. Peano joue un très grand rôle dans ces nouvelles recherches. Il est susceptible de rendre de grands services, mais il me semble que M. Couturat y attache une importance exagérée et qui a dû étonner M. Peano lui-même.

L'élément essentiel de ce langage, ce sont certains signes algébriques qui représentent les différentes conjonctions : si, et, ou, donc. Que ces signes soient commodes, c'est possible ; mais qu'ils soient destinés à renouveler toute la philosophie, c'est une autre affaire. Il est difficile d'admettre que le mot si acquiert, quand on l'écrit ɔ, une vertu qu'il n'avait pas quand on l'écrivait si. Cette invention de M. Peano s'est appelée d'abord la pasigraphie, c'est-à-dire l'art d'écrire un traité de mathématiques sans employer un seul mot de la langue usuelle. Ce nom en définissait très exactement la portée.

Depuis on l'a élevée à une dignité plus éminente, en lui conférant le titre de *logistique*. Ce mot est, parait-il, employé à l'École de Guerre pour désigner l'art du maréchal des logis, l'art de faire marcher et de cantonner les troupes ; mais ici aucune confusion n'est à craindre et on voit tout de suite que ce nom nouveau implique le dessein de révolutionner la logique. Nous pouvons voir la nouvelle méthode à l'œuvre dans un mémoire mathématique de M. Burali-Forti, intitulé : *Una Questione sui numeri transfiniti*, et inséré dans le tome XI des *Rendiconti del circolo matematico di Palermo*. Je commence par dire que ce mémoire est très intéressant, et si je le prends ici pour exemple, c'est précisément parce qu'il est le plus important de tous ceux qui sont écrits dans le nouveau langage. D'ailleurs les profanes peuvent le lire grâce à une traduction interlinéaire italienne. Ce qui fait l'importance de ce mémoire, c'est qu'il a donné le premier exemple de ces antinomies que l'on rencontre dans l'étude des nombres transfinis et qui font depuis quelques années le désespoir des mathématiciens. Le but de cette note, dit M. Burali-Forti, c'est de montrer qu'il peut y avoir deux nombres transfinis (ordinaux), a et b, tel que a ne soit ni égal à b, ni plus grand, ni plus petit. Que le lecteur se rassure, pour comprendre les considérations qui vont suivre, il n'a pas besoin de savoir ce que c'est qu'un nombre ordinal transfini. Or Cantor avait précisément démontré qu'entre deux nombres transfinis, il ne peut y avoir d'autre relation que l'égalité, ou l'inégalité dans un sens ou dans l'autre. Mais ce n'est pas du fond de ce mémoire que je veux parler ici ; cela m'entrainerait beaucoup trop loin de mon sujet ; je veux seulement m'occuper de la forme, et précisément je me demande si cette forme lui fait beaucoup gagner en rigueur et si elle compense par là les efforts qu'elle impose à l'écrivain et au lecteur. Nous voyons d'abord M. Burali-Forti définir le nombre 1 de la manière suivante :

$$1 = \iota \mathrm{T}`\{\mathrm{Ko} \cap (u, h) \ \mathrm{\Im} \ (u \ \mathrm{\varepsilon} \ \mathrm{Un})\}$$

définition éminemment propre à donner une idée du nombre 1 aux personnes qui n'en auraient jamais entendu parler. J'entends trop mal le Péanien pour oser risquer une critique, mais je crains bien que cette définition ne contienne une pétition de principe, attendu que j'aperçois 1 en chiffre dans le premier membre et Un en toutes lettres dans le second. Quoi qu'il en soit, M. Burali-Forti part de cette définition et, après un court calcul, il arrive à l'équation :

$$1 \ \mathrm{\varepsilon} \ \mathrm{No} \tag{27}$$

qui nous apprend que Un est un nombre. Et puisque nous en sommes à ces définitions des premiers nombres, rappelons que M. Couturat a défini également 0 et 1. Qu'est-ce que zéro ? c'est le nombre des éléments de la classe nulle ; et qu'est-ce que la classe nulle ? c'est celle qui ne contient aucun élément. Définir zéro par nul, et nul par aucun, c'est vraiment abuser de la richesse de la langue française ; aussi M. Couturat a-t-il introduit un perfectionnement dans sa définition, en écrivant :

$$0 = \iota \Lambda : \varphi x = \Lambda \,.\, \mathrm{\Im} \,.\, \Lambda = (x \ \mathrm{\Im} \ \varphi x)$$

ce qui veut dire en français : zéro est le nombre des objets qui satisfont à une condition qui n'est jamais remplie.

Mais comme jamais signifie *en aucun cas* je ne vois pas que le progrès soit considérable.

Je me hâte d'ajouter que la définition que M. Couturat donne du nombre 1 est plus satisfaisante.

Un, dit-il en substance, est le nombre des éléments d'une classe dont deux éléments quelconques sont identiques. Elle est plus satisfaisante, ai-je dit, en ce sens que pour définir 1, il ne se sert pas du mot un ; en revanche, il se sert du mot deux. Mais j'ai peur que si on demandait à M. Couturat ce que c'est que deux, il ne soit obligé de se servir du mot un.

Une pétition de principe est un raisonnement dans lequel on prend pour point de départ ce qui doit être démontré. C'est exactement dans ce sens que Poincaré utilise cette expression : il écrit que lorsqu'il voit une tentative de définition du nombre, il s'attend à trouver dans cette définition une utilisation du concept de nombre antérieure à cette définition.

Annexe 2. *Archive d'histoire des mathématiques MacTutor : l'infini*, de J. J. O'Connor et E. F. Robertson

La version originale de ce texte se trouve sur http://www-history.mcs.st-andrews.ac.uk/HistTopics/Infinity.html. La traduction est de Vincent Hugot avec quelques amendements.

Écrire un article sur l'infini pour une archive d'histoire des mathématiques présente des difficultés particulières : doit-on se concentrer exclusivement sur les aspects mathématiques du sujet ou en considérer aussi les aspects philosophiques voire religieux ? Notre parti pris dans cet article est que l'on ne peut pas dissocier ces trois approches, tant la philosophie et la religion ont joué un rôle important dans le développement des idées mathématiques. Ceci était particulièrement vrai au temps de la Grèce antique, ainsi que l'écrit Knorr :

> L'interaction entre philosophie et mathématiques se révèle rarement aussi clairement que dans l'étude de l'infini par les Grecs anciens. Les énigmes dialectiques des éléates du cinquième siècle, perfectionnées par Platon et Aristote au quatrième, sont complémentées par l'invention de méthodes précises de limites, telles qu'appliquées par Eudoxe au quatrième siècle et par Euclide et Archimède au troisième. (Knorr 1982, page 112.)

Bien entendu, à partir du moment où les gens commencèrent à s'interroger sur le monde dans lequel ils vivaient surgirent des questions sur l'infini. Ces questions portaient sur le temps : le monde est-il né à un instant précis ou a-t-il toujours existé ? Existera-t-il toujours ou cessera-t-il d'être ? D'autres portaient sur l'espace : que se passe-t-il si l'on voyage, très longtemps, toujours dans la même direction ? Finit-on par atteindre le bord du monde ou peut-on marcher pour l'éternité ? Et au dessus de la Terre brillent les étoiles et les planètes, le Soleil et la Lune, mais cet espace est-il fini ou s'étend-il indéfiniment ?

Ces questions sont fondamentales et ont donné bien de la peine à tous les penseurs longtemps avant l'avènement de l'histoire écrite. D'autres questions sur l'infini, plus subtiles, furent posées quand les réflexions menées sur le monde devinrent plus profondes : que se passerait-il si quelqu'un se mettait en tête de couper un morceau de bois en deux, puis de couper l'un des morceaux en deux, et ainsi de suite. Cet obstiné pourrait-il bucheronner pour l'éternité ?

Il est bon de commencer notre compte rendu sur l'infini par l'éléate du cinquième siècle qu'est Zénon. Les Grecs anciens avaient buté sur le problème de l'infini très tôt dans leur développement des mathématiques et des sciences.

En étudiant le sujet, ils dégagèrent la question fondamentale : peut-on continuer à diviser la matière en morceaux de plus en plus petits, indéfiniment, ou finit-on par trouver un minuscule morceau, que rien ne peut briser ? Pythagore avait affirmé que « tout est nombre », et son univers était entièrement constitué de nombres entiers, finis. Puis vinrent les Atomistes, qui croyaient que la matière était faite d'un nombre infini d'indivisibles. Parménide et l'école éléatique, dont Zénon faisait partie, étaient en désaccord avec les Atomistes. Cependant les paradoxes de Zénon montrent que la croyance en l'infinie divisibilité de la matière et la théorie Atomiste mènent toutes deux à d'apparentes contradictions.

Bien entendu, ces paradoxes naissent de l'infini. Aristote, semble-t-il, n'avait pas apprécié à sa juste valeur l'importance des arguments de Zénon, mais l'infini le préoccupait malgré tout. Il introduisit une idée qui domina la pensée pendant deux milliers d'années et qui reste un argument convaincant de nos jours aux yeux de certains. Aristote était opposé à l'idée d'un infini « en acte » et, en lieu et place, envisageait un infini « potentiel ». Son idée était que l'on ne pourra jamais concevoir l'ensemble des nombres naturels comme une totalité achevée. Cependant ces derniers sont en nombre potentiellement infini en ce sens que pour toute collection finie il est possible d'en trouver une autre, encore plus grande.

Il n'est pas inopportun dans notre discussion de mentionner la remarquable avancée que nous devons aux Babyloniens, inventeurs du système de numérotation positionnel qui, pour la première fois, permettait une représentation concise des nombres sans pour autant en limiter la taille. En

dépit des systèmes de numérotation positionnels, l'argument d'Aristote est assez convaincant : seul un nombre fini d'entiers a jamais été écrit ou envisagé. Si n est le plus grand entier jamais envisagé à ce jour, alors je peux aller plus loin en écrivant $n+1$ ou n^2 mais cela ne change rien au fait que seul un nombre fini d'entiers a été écrit. Aristote a discuté de cela dans les chapitres 4–8 du troisième livre de *La Physique* (voir Spalt 1990), où il prétendait que nier l'existence de l'infini en acte et ne tolérer que l'infini potentiel ne serait pas un obstacle pour les mathématiciens :

> La théorie ne supprime pas les considérations des mathématiciens, en supprimant l'infini qui existerait en acte dans le sens de l'accroissement, considéré comme ne pouvant être parcouru ; car, en réalité, ils n'ont pas besoin et ne font point usage de l'infini, mais seulement de grandeurs aussi grandes qu'ils voudront, mais limitées [...] (Aristote 1926, 207b27-31.)

Cantor, près de deux mille ans plus tard, écrivit qu'Aristote faisait une distinction qui n'existait en fait que dans le choix de ses mots :

> [...] en vérité l'infini potentiel n'a qu'une réalité empruntée, dans le sens où le concept d'infini potentiel renvoie toujours à un concept d'infini en acte qui lui est logiquement antérieur, et dont son existence dépend. (Cantor 1932, V, page 404.)

Nous parlerons des idées de Cantor vers la fin de cet article, mais pour le moment, penchons-nous sur l'effet qu'Aristote eut sur les mathématiciens grecs en autorisant l'usage de l'infini potentiel, et plus particulièrement sur Euclide (voir par exemple Spalt 1990). Comment donc, peut-on se demander, Euclide put-il prouver que l'ensemble des nombres premiers est infini en 300 av. J.-C. ? La réponse est, tout simplement, qu'Euclide n'a pas prouvé cela dans les *Éléments*. Il s'agit simplement d'une reformulation moderne de ce qu'Euclide a en fait énoncé comme son théorème, qui est selon la traduction de Vitrac :

> Les nombres premiers sont plus nombreux que toute multitude de nombres premiers proposée. (Euclide d'Alexandrie 1990-2001, Proposition IX.20, volume 2, page 444.)

Ainsi, ce qu'Euclide a réellement prouvé est que l'ensemble des nombres premiers est potentiellement infini mais en pratique, bien entendu, cela revient au même. Sa preuve montre que pour toute collection finie de nombres premiers, il doit exister un nombre premier qui n'est pas dans la collection.

Nous devrions discuter d'autres aspects de l'infini qui jouent un rôle crucial dans les *Éléments*. Euclide y détaille la méthode « d'exhaustion » due à Eudoxe de Cnide. De nos jours, cette méthode est souvent comprise comme une approximation du cercle par une suite de polygones réguliers dont le nombre de côtés augmente indéfiniment. Nous nous devons cependant d'insister fortement sur le fait que ce n'est pas du tout ainsi que les anciens Grecs voyaient cette méthode ; il s'agissait pour eux d'un argument par réduction à l'absurde qui évitait l'usage de l'infini. Par exemple, pour prouver que deux aires A et B sont égales, la méthode consiste à supposer que A est plus petite que B, puis à en déduire une contradiction en un nombre fini d'étapes. De même supposer B plus petite que A conduit à une contradiction en un nombre fini d'étapes.

L'examen des textes a cependant suggéré, fort récemment, que tous les mathématiciens grecs anciens ne se sentaient pas obligés à se restreindre à l'infini potentiel. Netz, Saito et Tchernetska (2001) ont redécouvert dans *La Méthode* la manière remarquable qu'a Archimède de manipuler des collections infinies d'objets.

> [...] Archimède prend trois paires de grandeurs infinies en nombre et affirme qu'elles sont, respectivement, « égales en multitude » [...] Nous suspectons qu'il n'y a aucun autre endroit connu dans les mathématiques grecques – ni même dans le corpus de grec ancien – où des objets infinis en nombre sont dits « égaux en multitude ». [...] Cette suggestion même que certains objets, infinis en nombre, sont « égaux en multitude » à d'autres implique que tous les objets infinis en nombre ne sont pas égaux de cette façon. [...] Nous avons ici des objets infiniment nombreux – ayant des multitudes précises, et différentes (c'est-à-dire qu'elles ont presque un nombre) ; de telles multitudes sont manipulées d'une manière concrète, apparemment au moyen de quelque chose plutôt

> semblable à une bijection. [...] dans ce cas, Archimède étudie des infinis en acte presque comme s'ils avaient un nombre au sens usuel [...]

Même si la plupart des mathématiciens acceptèrent les arguments d'Aristote en ce qui concerne l'infini potentiel, d'autre défendirent l'infini en acte. Au premier siècle avant J.-C. Lucrèce écrivit le poème *De rerum natura* dans lequel il réfute l'idée d'un Univers limité dans l'espace. Son argument est très simple. Supposons que l'univers soit fini. Alors il faut bien qu'il ait une frontière. Maintenant, si l'on s'approchait de cette frontière, et lui lançait un objet, rien ne pourrait l'arrêter car ce qui l'arrêterait devrait se trouver au delà de la frontière, et par définition, il n'existe rien en dehors de l'Univers. Bien sûr nous savons aujourd'hui que l'argument de Lucrèce est erroné car l'espace peut très bien être fini sans pour autant avoir de frontière. Cependant l'argument de la frontière a dominé le débat concernant la finitude de l'espace pendant bien des siècles.

Ce furent principalement les théologiens qui militèrent en faveur de l'infini en acte. Par exemple saint Augustin, le philosophe chrétien qui introduisit la majeure partie de la philosophie de Platon dans le christianisme au début du 5e siècle de notre ère, affirmait que Dieu est infini, et capable de pensées infinies. Il écrivit dans son très célèbre ouvrage *La Cité de Dieu* :

> Et, quand ils disent que la science de Dieu ne peut appréhender l'infini, il ne leur reste plus qu'à oser dire, en se plongeant dans le gouffre d'une profonde impiété, que Dieu ne connait pas tous les nombres ! [...] Qui serait assez fou pour le dire ? [...] qui sommes-nous donc, nous misérables avortons, pour avoir l'audace de donner des limites à la science de Dieu [...] (Augustin 2000, livre XII, XIX, pages 497–498.)

Les mathématiciens indiens travaillèrent pour introduire le zéro dans leur système numérique pendant près de 500 ans. Brahmagupta, au 7e siècle, fut le premier. Faire en sorte que le zéro respecte les opérations usuelles de l'arithmétique était l'épineux problème qu'ils eurent à affronter... Bhāskarācārya écrit dans le *Bījaganita* :

> Il ne doit pas y avoir de changement pour lui, quand de grands [nombres] entrent ou sortent d'un [nombre] « qui a pour diviseur zéro », de même qu'il n'y a pas de changement pour l'Illimité quand, au moment de la destruction ou à celui de la création, des multitudes d'êtres entrent ou sortent de Viṣṇu. (Bhāskarācārya 2004, Bījaganita, strophe 16, page 215.)

C'était une tentative d'introduire l'infini, en même temps que zéro, dans le système numérique. Bien entendu ceci ne fonctionne pas car si on l'introduisait ainsi que Bhāskarācārya le suggère, alors 0 fois l'infini doit être égal à chaque nombre n, donc tous les nombres sont égaux.

Thomas d'Aquin, le théologien et philosophe chrétien, utilisait le fait qu'il n'existe pas de nombre qui puisse représenter l'infini comme une preuve que l'infini en acte n'existe pas. Dans *Summa theologia*, au 13e siècle, Thomas d'Aquin écrit :

> Mais cela [une multitude infinie en acte] est impossible. En effet, une multitude doit appartenir à une espèce donnée de multitude. Or les espèces de la multitude correspondent aux espèces du nombre. Mais nulle espèce de nombre n'est infinie, car le nombre se définit une multitude mesurée par l'unité. On doit donc dire que toute multitude infinie en acte est impossible, par soi ou par accident. (Thomas d'Aquin 1984, question 7, article 4, page 201.)

Cette objection est en effet très raisonnable et n'avait pas de réponse satisfaisante au temps de Thomas d'Aquin. Un ensemble infini en acte exige une mesure, et aucune mesure de cette sorte ne semblait possible pour Thomas d'Aquin. Il faut attendre Cantor et la fin du 19e siècle pour qu'une mesure satisfaisante soit trouvée pour les ensembles infinis. Davenport 1997 examine

> [...] les arguments mathématiques utilisés par deux théologiens du treizième siècle, Alexandre Nequam et Richard Fishacre, pour défendre la cohérence de l'infinité divine. La question suivante est soulevée par rapport à leurs arguments : pourquoi des théologiens jugèrent-ils approprié d'en appeler à des exemples mathématiques pour traiter un problème purement théologique ? (Davenport 1997, page 263.)

La récurrence mathématique commença à être utilisée des siècles avant qu'aucune formulation rigoureuse de la méthode ne fût énoncée. Elle donnait une technique pour prouver qu'une proposition est vraie pour un nombre infini de valeurs entières. Par exemple Al-Karaji utilise, vers l'an mil de notre ère, une forme non rigoureuse de récurrence mathématique dans ses arguments. À peu de chose près, il démontrait un argument pour $n = 1$, puis prouvait qu'il était vrai aussi pour $n = 2$ en utilisant le cas où $n = 1$, et ainsi de suite jusqu'à $n = 5$. Et là il s'arrêtait, en constatant que l'on pouvait répéter cette opération indéfiniment. Grâce à cette méthode il donna une belle description de la façon de générer les coefficients binomiaux en utilisant le triangle de Pascal.

Pascal n'avait pas connaissance des travaux d'Al-Karaji dans ce domaine, mais il savait que Maurolico avait utilisé une forme d'argument par récurrence au milieu du 17e siècle. Pascal, définissant son fameux triangle « de Pascal », écrit :

> Quoique cette proposition ait une infinité de cas, j'en donnerai une démonstration bien courte, en supposant 2 lemmes.
>
> Le 1er, qui est évident de soi-même, que cette proportion se rencontre dans la seconde base ; car il est bien visible que ϕ est à σ comme 1 à 1.
>
> Le 2e, que si cette proportion se trouve dans une base quelconque, elle se trouvera nécessairement dans la base suivante.
>
> D'où il se voit qu'elle est nécessairement dans toutes les bases : car elle est dans la seconde base par le premier lemme ; donc par le second elle est dans la troisième base, donc dans la quatrième, et à l'infini. (Pascal 1970b, page 1294.)

Nous avons bien avancé dans le temps en suivant les progrès de la récurrence. Revenons quelque peu sur nos pas pour examiner les discussions qui eurent lieu pendant ce temps au sujet du caractère infini de l'Univers.

Le modèle aristotélicien d'un univers fini, composé de neuf sphères célestes centrées sur la Terre, fit office de paradigme pendant bien longtemps. Quelques contradicteurs se firent entendre malgré tout, et nous avons déjà vu l'argument de Lucrèce en faveur d'un univers infini. Nicolas de Cues, au milieu du 15e siècle, était un brillant scientifique qui affirmait que l'univers était infini et que les étoiles étaient des soleils lointains. Durant le 16e siècle, l'Église catholique en Europe entreprit de pourchasser ce genre d'hérésies. Giordano Bruno n'était ni un scientifique ni un mathématicien, mais il se prononça vigoureusement en faveur d'un univers infini dans *De l'infini de l'univers et des mondes* (1584). Mené devant l'Inquisition, il fut torturé durant neuf ans, dans une vaine tentative de lui faire admettre que l'univers est fini. Obstiné, il fut brulé sur le bucher en 1600.

Galilée était tout à fait conscient du sort de Bruno aux mains de l'Inquisition, c'est donc avec la plus extrême prudence qu'il présenta ses idées. Il s'attaqua à la question de l'infini dans ses *Discours et démonstrations mathématiques concernant deux sciences nouvelles* (1638) où il étudie le problème suivant de deux cercles concentriques de centre O, le cercle A ayant un diamètre double de celui de B. Une formule mathématique familière donne alors la circonférence de A double de celle de B. Cependant, si l'on prend un point P de A, alors le segment [PO] coupe le cercle B en exactement un point. De même, si Q est un point de B, la demi-droite [OQ) coupe A en exactement un point. Finalement, bien que la circonférence de A soit double de celle de B, les deux cercles ont le même nombre de points. Galilée proposait d'ajouter un nombre infini de trous infiniment petits à la plus petite longueur pour la rendre égale à la plus grande tout en leur permettant d'avoir le même nombre de points. Il écrivit :

> Ces difficultés sont réelles, et ne sont pas les seules ; mais rappelons-nous que nous traitons d'infinis et d'indivisibles, inaccessibles à notre entendement fini, les premiers à cause de leur immensité, les seconds à cause de leur petitesse. Pourtant nous constatons que la raison humaine ne peut s'empêcher de sans cesse y revenir ; [...] (Galilei 1638, page 26.)

Cependant, Galilée explique que les difficultés viennent de ce que

> [...] nous discutons, avec notre esprit fini, des choses infinies, et leur attribuons les épithètes que nous utilisons pour les choses finies et limitées ; ce qui, à mon avis, est

> incorrect, car j'estime que des épithètes comme « plus grand », « plus petit » et « égal » ne conviennent pas aux grandeurs infinies, dont il est impossible de dire que l'une est plus grande, plus petite ou égale à une autre. (Galilei 1638, page 30.)

Il énonça alors un autre paradoxe, similaire à celui du cercle mais construit de telle sorte qu'aucun indivisible ne puisse être utilisé pour corriger la situation. Il établit la correspondance un-à-un entre les entiers positifs et leurs carrés. Il y a donc autant de carrés que d'entiers positifs. Cependant, la plupart des nombres ne sont pas des carrés. Galilée dit que cela signifie seulement que

> [...] l'ensemble des nombres est infini, que le nombre des carrés est infini, et le nombre de leurs racines pareillement ; que le total des nombres carrés n'est pas inférieur à l'ensemble des nombres, ni celui-ci supérieur à celui-là, et, finalement, que les attributs « égal », « plus grand » et « plus petit » n'ont pas de sens pour les quantités infinies, mais seulement pour les quantités finies. (Galilei 1638, page 31)

Knobloch (1999) s'intéresse à nouveau à ce travail de Galilée. Dans le même article, les très circonspectes définitions que Leibniz avait données de l'infinitésimal et de l'infini en termes de procédures limites sont aussi examinées. Le développement du calcul différentiel et intégral par Leibniz est fondé sur l'idée de l'infiniment petit qui était étudié depuis longtemps.

Cavalieri écrivit en 1635 *Geometria indivisibilibus continuorum*, dans lequel il pensait les lignes comme comprenant un nombre infini de points, et les surfaces comme formées d'une infinité de lignes. Il donna des méthodes assez rigoureuses permettant de comparer des aires, dont celle connue sous le nom de « Principe de Cavalieri » : si une droite est déplacée, parallèlement à elle-même, d'un bout à l'autre de deux aires, et si le rapport des longueurs des segments compris dans chaque aire est toujours $a : b$, alors le rapport des aires est $a : b$.

Roberval alla encore plus loin dans cette ligne de pensée ; il voyait les droites comme sommes d'une infinité de minuscules parties indivisibles. Il introduisit des méthodes pour comparer les tailles des indivisibles, de telle sorte que l'on pût définir des rapports de leurs grandeurs bien qu'ils en fussent dépourvus. C'était un grand pas en avant dans la manière de manipuler l'infini, car pour la première fois, il pouvait ignorer les grandeurs qui étaient négligeables comparées à d'autres. Cependant, il y a une différence entre être capable d'utiliser une méthode correctement, et écrire rigoureusement dans quelles conditions elle s'applique... Les paradoxes qui en découlèrent incitèrent certaines personnes à rejeter la méthode des indivisibles.

Le Collège romain rejeta les indivisibles et bannit leur enseignement des collèges jésuites en 1649. L'Église avait échoué à faire taire Bruno malgré son exécution, échoué à réduire Galilée au silence bien qu'elle lui interdît de quitter sa maison, et elle échoua tout autant dans son ambition de stopper les progrès qui menaient au calcul différentiel et intégral en interdisant l'enseignement des indivisibles. Au lieu de cela, elle ne fit que forcer les mathématiciens à s'astreindre à la plus grande rigueur, dans l'espoir de se prémunir contre les critiques.

Le symbole ∞ que nous utilisons aujourd'hui fut utilisé pour la première fois par John Wallis dans *De sectionibus conicis* en 1655 et à nouveau dans *Arithmetica infinitorum* en 1656. Il choisit ce symbole pour représenter le fait que l'on peut traverser la courbe un nombre infini de fois.

Trois ans plus tard, Fermat identifia une propriété importante des entiers positifs, à savoir qu'aucune suite dans $\mathbb{N}$ ne peut décroitre indéfiniment. Il énonça ce résultat en introduisant la méthode de la descente infinie, en 1659 :

> Et pour ce que les méthodes ordinaires, qui sont dans les Livres, étaient insuffisantes à démontrer des propositions si difficiles, je trouvai enfin une route tout à fait singulière pour y parvenir.
>
> J'appelai cette manière de démontrer la *descente infinie* ou *indéfinie*, etc. [...] (Fermat 1894, Lettre de Fermat à Carcavi, page 431.)

Le principe de cette méthode est de montrer que si une proposition est vraie pour un certain entier positif n, alors elle est vraie pour un entier positif moindre que n. Étant donné qu'aucune suite infiniment décroissante d'entiers positifs n'existe, une telle preuve débouche sur une contradiction Fermat utilisa sa méthode pour prouver qu'aucune solution en entiers positifs n'existait pour

l'équation

$$x^4 + y^4 = z^4$$

Newton préférait aux indivisibles ses « fluxions », qui étaient la mesure de la variation instantanée d'une quantité. Bien entendu, l'infini n'était pas évité pour autant, puisqu'il devait toujours considérer des incréments infiniment petits. C'était, dans un sens, sa réponse au paradoxe de la flèche de Zénon :

> Si, dit Zénon, toute chose est au repos lorsqu'elle occupe un espace égal à elle-même, et que l'objet mobile occupe à un instant un espace égal à lui-même, la flèche mobile est immobile. (Heath 1949, page 127.)

Les fluxions de Newton produisirent de merveilleux résultats mathématiques, mais nombreux étaient ceux qui se défiaient de son usage des incréments infiniment petits. La célèbre remarque de George Berkeley résume les objections de façon concise :

> Que sont ces fluxions ? Les vitesses d'incréments évanouissants, et que sont ces mêmes incréments évanouissants ? Ce ne sont ni des quantités finies, ni des quantités infiniment petites, ni pourtant rien. Ne pouvons-nous les appeler les fantômes des quantités défuntes ? (Berkeley 1987, § 35, page 313.)

Newton croyait que l'espace est en fait infini, et non pas seulement indéfiniment grand. Il affirmait qu'un tel infini peut être compris, en particulier en utilisant des arguments géométriques, mais qu'il restait inconcevable. C'est intéressant car, ainsi que nous le verrons, d'autres entendaient prouver l'inexistence de l'infini en acte en disant justement qu'il était inconcevable.

Le problème de l'infinie divisibilité du temps et de l'espace continua à troubler les gens. David Hume écrivit en 1739 dans son *Traité de la nature humaine* qu'il existe une plus petite taille perceptible :

> Faites une tache d'encre sur du papier, fixez votre regard sur cette tache et reculez jusqu'à ce qu'enfin vous la perdiez de vue : il est évident qu'à l'instant précédant sa disparition, l'image ou l'impression étaient parfaitement indivisibles. (Hume 1995, I, II, I, page 76.)

Emmanuel Kant écrivit en 1781 dans sa *Critique de la raison pure* que l'infini en acte ne peut exister car il ne peut être perçu.

> D'où il résulte que, pour se représenter comme un tout le monde qui remplit tous les espaces, il faudrait regarder comme achevée la synthèse successive des parties d'un monde infini, c'est-à-dire qu'il faudrait regarder un temps infini comme s'étant écoulé au fil de l'énumération de toutes les choses coexistantes – ce qui est impossible. (Kant 2006, Premier conflit des Idées transcendentales, page 430.)

Ceci nous amène à une question souvent posée par les philosophes : le monde existerait-il s'il n'y avait aucune intelligence capable d'en concevoir l'existence ? Kant dit que non [note du traducteur : en fait la position attribuée ici à Kant n'est pas la sienne, mais celle de Berkeley] ; nous sommes donc revenus à un argument donné au début de cet article, à savoir que la collection des entiers n'est pas infinie pour la simple raison qu'on ne peut jamais énumérer qu'un nombre fini d'entiers.

Peu de progrès avait été fait sur la question de l'infini en acte. Les mêmes arguments apparaissaient sans avancée définitive vers une meilleure compréhension de cette notion. Gauss s'oppose à l'usage de l'infini en acte dans une lettre à Schumacher en 1831 :

> Pour ce qui est maintenant de votre démonstration [...], je commencerai par protester contre l'usage que vous faites d'une grandeur infinie, en la traitant comme une quantité achevée [vollendete], ce qui n'est jamais permis en mathématiques. L'infini n'est qu'une *façon de parler*, parce qu'il s'agit en réalité de *limites*, dont certains rapports peuvent approcher d'autant que l'on voudra, tandis que d'autres sont susceptibles de croitre indéfiniment. (Sebestik 1992, page 435, note 2.)

Peut-être l'un des évènements les plus significatifs dans le développement du concept d'infini fut-il la publication en 1840 des *Paradoxes de l'infini* de Bernard Bolzano. Il y écrit que l'infini existe réellement, et son argument utilise la notion d'ensemble, définie pour la toute première fois :

> J'appelle *ensemble* une collection à laquelle nous imputons un concept tel que l'arrangement des parties soit indifférent (dans lequel rien d'essentiel n'est donc changé pour nous lorsque seul l'arrangement est modifié) [...] (Bolzano 1993, § 4, page 60.)

En quoi la notion d'ensemble contribue-t-elle à faire de l'infini en acte une réalité ? La réponse est simple. Une fois que l'on regarde les entiers comme formant un ensemble, il y a nécessairement une entité unique qui doit être infinie en acte. Aristote disait des entiers que l'on peut en trouver des collections arbitrairement grandes. Mais une fois que l'on dispose de la notion d'ensemble, celles-ci sont considérées comme des sous-ensembles de l'ensemble des entiers qui se doit donc d'être infini en acte. De façon peut-être surprenante, Bolzano n'utilise pas cet exemple d'un ensemble infini, mais au lieu de cela considère toutes les propositions vraies :

> Comme il est facile de le voir, l'ensemble des *propositions et vérités en soi* est infini. Si nous considérons, en effet, une vérité quelconque, par exemple la proposition : « il y a en général des vérités », ou toute autre proposition que je désigne par A, nous remarquons que la proposition : « A est vraie » est différente de A elle-même [...] (Bolzano 1993, § 13, pages 71-72.)

À ce stade, l'étude mathématique de l'infini migre vers la théorie des ensembles, et nous renvoyons le lecteur à l'article *Beginnings of set theory* [Débuts de la théorie des ensembles, http://www-history.mcs.st-andrews.ac.uk/HistTopics/Beginnings_of_set_theory.html] pour plus d'informations sur la contribution de Bolzano, ainsi que le traitement de l'infini par Cantor, qui construisit une théorie des différentes tailles de l'infini par ses définitions des nombres cardinaux et ordinaux.

Le problème des infinitésimaux fut formulé d'une manière mathématiquement rigoureuse par Robinson dans son fameux livre de 1966 sur l'analyse non standard. Kreisel a écrit dans sa recension aux Mathematical reviews :

> Ce livre, qui est paru exactement 250 ans après la mort de Leibniz, présente une théorie rigoureuse et efficace des infinitésimaux obéissant, ainsi que Leibniz l'entendait, aux mêmes lois que les nombres ordinaires.

Fenstad (1988) discute de l'infini et de l'analyse non-standard. Il examine aussi son utilisation pour la modélisation des phénomènes naturels.

Chapitre 6

À propos de Cauchy et de l'uniformité

Introduction

Nous analysons dans ce chapitre plusieurs définitions et preuves du cours d'analyse de Cauchy à l'École polytechnique, en relation avec la notion d'uniformité (fonction uniformément continue sur un intervalle, fonction uniformément dérivable sur un intervalle, suite uniformément convergente de fonctions).

La source est le *Cours d'analyse de l'École royale Polytechnique* (Cauchy 1821), qui se réduit à sa *1.re partie. Analyse algébrique* et a étéréimprimé dans *Œuvres complètes, série II, tome 3*, et son *Résumé des leçons sur le calcul infinitésimal, Calcul différentiel* (Cauchy 1823), réimprimé dans *Œuvres complètes, série II, tome 4*, ainsi qu'un compte rendu à l'Académie des Sciences.

Dans ce cours qui date de 1821, de nombreuses preuves de Cauchy sont réputées fautives, mais elles se révèlent correctes si on utilise une interprétation « uniforme » des définitions. En outre, les preuves sont particulièrement simples et claires. Comme les définitions de Cauchy contiennent une part d'ambigüité, nous sommes amenés à penser que les preuves de Cauchy ne sont pas fausses, mais plutôt « incomplètes » au regard de la rigueur moderne.

Cauchy eut le grand mérite de commencer à fonder l'analyse sur des bases simples, en fournissant des définitions relativement précises pour les notions de limite, de continuité, de dérivabilité, et surtout en élaborant des preuves pour des résultats considérés par les uns comme évidents et par les autres comme parfaitement obscurs. Même si certains de ses théorèmes « souffraient des exceptions », au moins des preuves relativement précises étaient-elles en place, qu'il suffirait d'examiner à la loupe pour faire évoluer définitions, énoncés des théorèmes, et interprétations sémantiques des résultats obtenus.

La lectrice pourra consulter *Preuves et réfutations* d'Imre Lakatos (1984) au sujet de la place centrale des preuves, plutôt que des théorèmes, dans l'activité mathématique, ainsi que sur le sujet plus précis des « théorèmes prouvés mais souffrant des exceptions » chez Cauchy.

Le contenu de ce chapitre reprend en bonne partie l'article *L'uniformité, un concept implicite efficace chez Cauchy* d'Henri Lombardi (1994). L'article *Gli "errori" di Cauchy e i fondamenti dell'analisi* d'Enrico Giusti (1984) est sur le même thème, paru bien avant, et plus complet. En conséquence nous nous baserons souvent sur ce dernier article, que nous recommandons vivement à ceux qui lisent l'italien.

6.1 Nombres, quantités, variables, infiniment petits

Dans cette section, nous citons quelques passages des *Préliminaires* du cours d'*Analyse algébrique*, pages 1-3 et 5-6, dont la lecture nous semble nécessaire pour comprendre le contexte de l'époque et le langage utilisé ensuite par Cauchy.

6.1.1 Nombres et quantités

Tout d'abord, il est frappant de constater qu'en 1821, la notion de « nombre » est une notion première, qu'on ne cherche pas à définir, et que le fait d'attribuer un signe à un nombre lui fait subir un changement de nature, qui nécessite un changement de vocabulaire, ce qui le transforme en une « quantité ».

Revue des diverses espèces de quantités réelles que l'on considère, soit en algèbre, soit en trigonométrie, et des notations à l'aide desquelles on les représente. [...]

Pour éviter toute espèce de confusion dans le langage et l'écriture algébriques, nous allons fixer dans ces préliminaires la valeur de plusieurs termes et de plusieurs notations que nous emprunterons soit à l'algèbre ordinaire, soit à la trigonométrie. Les explications que nous donnerons à ce sujet sont nécessaires, pour que nous ayons la certitude d'être parfaitement compris de ceux qui liront cet ouvrage. Nous allons indiquer d'abord quelle idée il nous parait convenable d'attacher à ces deux mots, *nombre* et *quantité.*

Nous prendrons toujours la dénomination de *nombres* dans le sens où on l'emploie en arithmétique, en faisant naitre les nombres de la mesure absolue des grandeurs ; et nous appliquerons uniquement la dénomination de *quantités* aux quantités *réelles positives* ou *négatives*, c'est-à-dire, aux nombres précédés des signes + ou −. De plus, nous regarderons les quantités comme destinées à exprimer des accroissements ou des diminutions ; en sorte qu'une grandeur donnée sera simplement représentée par un nombre, si l'on se contente de la comparer à une autre grandeur de même espèce prise pour unité, et par ce nombre précédé du signe + ou du signe −, si on la considère comme devant servir à l'accroissement ou à la diminution d'une grandeur fixe de la même espèce. Cela posé, le signe + ou − placé devant un nombre en modifiera la signification, à-peu-près comme un adjectif modifie celle du substantif. Nous appellerons *valeur numérique* d'une quantité le nombre qui en fait la base, quantités *égales* celles qui ont le même signe avec la même valeur numérique, et quantités *opposées* deux quantités égales quant à leurs valeurs numériques, mais affectées de signes contraires. En partant de ces principes, il est facile de rendre compte des diverses opérations que l'on peut faire subir aux quantités. Par exemple, deux quantités étant données, on pourra toujours en trouver une troisième qui, prise pour accroissement d'un nombre fixe, si elle est positive, et pour diminution dans le cas contraire, conduise au même résultat que les deux quantités données, employées l'une après l'autre à pareil usage. Cette troisième quantité, qui à elle seule produit le même effet que les deux autres, est ce qu'on appelle leur *somme*. Ainsi les deux quantités −10 et +7 ont pour somme −3, attendu qu'une diminution de 10 unités, jointe à une augmentation de 7 unités, équivaut à une diminution de 3 unités. *Ajouter* deux quantités, c'est former leur somme. La différence entre une première quantité et une seconde, c'est une troisième quantité qui, ajoutée à la seconde, reproduit la première. Enfin, on dit qu'une quantité est *plus grande* ou *plus petite* qu'une autre, suivant que la différence de la première à la seconde est positive ou négative. D'après cette définition, les quantités positives surpassent toujours les quantités négatives, et celles-ci doivent être considérées comme d'autant plus petites que leurs valeurs numériques sont plus grandes.

On appréciera la clarté pédagogique de l'exposé qui justifie comment doivent s'ajouter des quantités. On notera que cet effort n'était pas considéré comme inutile, même lorsqu'on s'adressait aux élèves ayant réussi au concours d'entrée de la prestigieuse École polytechnique. On a redécouvert récemment les vertus d'un exposé non formel, mais opératoire, des « nombres avec signes » pour faire « comprendre » (et non pas digérer de force) aux élèves du collège les règles du calcul algébrique avec signes.

En algèbre, on représente non-seulement les nombres, mais aussi les quantités, par des lettres. Comme on est convenu de ranger les nombres absolus dans la classe des quantités positives, on peut désigner la quantité positive qui a pour valeur numérique le nombre A, soit par +A, soit par A seulement, tandis que la quantité négative opposée se trouve représentée par −A. De même, dans le cas où la lettre a représente une quantité, on est convenu de regarder comme synonymes les deux expressions a et $+a$, et de représenter par $-a$ la quantité opposée à $+a$. Ces remarques suffisent pour établir ce qu'on appelle *la règle des signes* [voyez la note I.re].

. . . . . .

Une longueur, comptée sur une ligne droite ou courbe, peut être, comme toute espèce de grandeur, représentée soit par un nombre, soit par une quantité, savoir : par un nombre, lorsqu'on a simplement égard à la mesure de cette longueur, et par une quantité, c'est-à-dire, par un nombre précédé du signe + ou −, lorsque l'on considère la longueur dont il s'agit comme portée, à partir d'un point fixe, sur la ligne donnée dans un sens ou dans un autre, pour servir soit à l'augmentation soit à la diminution d'une autre longueur constante aboutissant à ce point fixe. Le point fixe dont il est ici question, et à partir duquel on doit porter les longueurs variables désignées par des quantités, est ce qu'on appelle *l'origine* de ces mêmes longueurs. Deux longueurs comptées à partir d'une origine commune, mais en sens contraires, doivent être représentées par des quantités de signes différents. On peut choisir à volonté le sens dans lequel on doit compter les longueurs désignées par des quantités positives ; mais, ce choix une fois fait, il faudra nécessairement compter dans le sens opposé les longueurs qui seront désignées par des quantités négatives.

6.1.2 Variables, infiniment petits, infiniment grands

Voici tout d'abord un passage repris des préliminaires, pages 4 et 5.

On nomme quantité *variable* celle que l'on considère comme devant recevoir successivement plusieurs valeurs différentes les unes des autres. On désigne une semblable quantité par une lettre prise ordinairement parmi les dernières de l'alphabet. On appelle au contraire quantité *constante*, et on désigne ordinairement par une des premières lettres de l'alphabet toute quantité qui reçoit une valeur fixe et déterminée. Lorsque les valeurs successivement attribuées à une même variable s'approchent indéfiniment d'une valeur fixe, de manière à finir par en différer aussi peu que l'on voudra, cette dernière est appelée la *limite* de toutes les autres. Ainsi, par exemple, un nombre irrationnel est la limite des diverses fractions qui en fournissent des valeurs de plus en plus approchées. En géométrie, la surface du cercle est la limite vers laquelle convergent les surfaces des polygones inscrits, tandis que le nombre de leurs côtés croît de plus en plus ; etc.. . .

Lorsque les valeurs numériques successives d'une même variable décroissent indéfiniment, de manière à s'abaisser au-dessous de tout nombre donné, cette variable devient ce qu'on nomme un *infiniment petit* ou une quantité *infiniment petite*. Une variable de cette espèce a zéro pour limite.

Lorsque les valeurs numériques successives d'une même variable croissent de plus en plus, de manière à s'élever au-dessus de tout nombre donné, on dit que cette variable a pour limite l'*infini positif,* indiqué par le signe ∞, s'il s'agit d'une variable positive, et l'*infini négatif,* indiqué par la notation $-\infty$, s'il s'agit d'une variable négative. Les infinis positif et négatif sont désignés conjointement sous le nom de *quantités infinies*.

Les commentaires de Giusti sur la notion de variable chez Cauchy sont les suivants. Cauchy se situe dans la tradition du siècle précédent, où tous les auteurs, aux exceptions notables d'Euler et Lagrange, introduisent les variables comme des quantités dépendant implicitement ou explicitement du temps qui s'écoule, en filiation directe avec les « fluentes » de Newton.

Cependant, tout en ne voulant pas heurter de front la tradition, Cauchy n'en opère pas moins un changement majeur. En disant qu'une quantité variable est celle que l'on considère comme devant

recevoir *successivement* plusieurs valeurs différentes les unes des autres, il change de paradigme, remplaçant l'écoulement continu du temps par une suite infinie mais discrète de valeurs.

Voici un passage du chapitre II du même cours, page 26, où Cauchy revient plus en détail sur la notion d'infiniment petit. Comme on peut le constater, il ne s'agit pas d'un infiniment petit en acte, d'une quantité infinitésimale, mais bien d'une « manière de parler » d'une quantité variable tendant vers 0. On peut comprendre ceci comme un avertissement : « quand je parlerai d'un infiniment petit, il faudra comprendre une suite de réels tendant vers 0. »

§. 1.er *Des quantités infiniment petites et infiniment grandes.*

On dit qu'une quantité variable devient *infiniment petite*, lorsque sa valeur numérique décroit indéfiniment de manière à converger vers la limite zéro. Il est bon de remarquer à ce sujet qu'on ne doit pas confondre un décroissement constant avec un décroissement indéfini. La surface d'un polygone régulier circonscrit à un cercle donné décroit constamment à mesure que le nombre des côtés augmente, mais non pas indéfiniment, puisqu'elle a pour limite la surface du cercle. De même encore, une variable qui n'admettrait pour valeurs successives que les différents termes de la suite

$$\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \frac{6}{5}, \dots$$

prolongée à l'infini, décroitrait constamment, mais non pas indéfiniment, puisque ses valeurs successives convergeraient vers la limite 1. Au contraire, une variable qui n'aurait pour valeurs successives que les différents termes de la suite

$$\frac{1}{4}, \frac{1}{3}, \frac{1}{6}, \frac{1}{5}, \frac{1}{8}, \frac{1}{7}, \dots,$$

prolongée a l'infini, ne décroitrait pas constamment, puisque la différence entre deux termes consécutifs de cette suite est alternativement positive et négative ; et, néanmoins, elle décroitrait indéfiniment, puisque sa valeur finirait par s'abaisser au-dessous de tout nombre donné.

Dans toute la suite nous verrons Cauchy mélanger avec brio le langage des infiniment petits (hérité du siècle précédent) et celui des suites qui convergent vers 0. Mais le vrai concept à l'œuvre dans les preuves est celui de suite convergente. Et ce n'est pas un hasard si c'est à lui que revient d'avoir inventé (découvert ?) cet admirable critère : le critère de Cauchy.

6.1.3 Le critère de Cauchy

Cet extrait provient du chapitre VI du même cours, pages 123-125.

§. 1.er *Considérations générales sur les Séries*

On appelle *série* une suite indéfinie de quantités

$$u_0, u_1, u_2, u_3, \text{etc.} \dots$$

qui dérivent les unes des autres suivant une loi déterminée. Ces quantités elles-mêmes sont les différents *termes* de la série que l'on considère.

Notez « une suite indéfinie de quantités [...] qui dérivent les unes des autres selon une loi déterminée. » L'idée communément acceptée aujourd'hui qu'une suite infinie de nombres réels

puisse être arbitraire et soumise à aucune loi est complètement hors du cadre de pensée au début du 19e siècle.

Cauchy définit ensuite la notion de série convergente et donne immédiatement son critère (qu'il n'expose pas pour les suites, mais pour les séries). Cela donne :

> D'après les principes ci-dessus établis, pour que la série
>
> (1) $$u_0, u_1, u_2 \dots u_n, u_{n+1}, \text{etc.} \dots$$
>
> soit convergente, il est nécessaire et il suffit que des valeurs croissantes de n fassent converger indéfiniment la somme
>
> $$s_n = u_0 + u_1 + u_2 + \text{etc.} \dots + u_{n-1}$$
>
> vers une limite fixe s : en d'autres termes, il est nécessaire et il suffit que, pour des valeurs infiniment grandes du nombre n, les sommes
>
> $$s_n, s_{n+1}, s_{n+2}, \text{etc.} \dots$$
>
> diffèrent de la limite s, et par conséquent entre elles, de quantités infiniment petites. D'ailleurs, les différences successives entre la première somme s_n et chacune des suivantes sont respectivement déterminées par les équations
>
> $$\begin{aligned} s_{n+1} - s_n &= u_n, \\ s_{n+2} - s_n &= u_n + u_{n+1}, \\ s_{n+3} - s_n &= u_n + u_{n+1} + u_{n+2}, \\ &\text{etc.} \dots \end{aligned}$$
>
> Donc, pour que la série (1) soit convergente, il est d'abord nécessaire que le terme général u_n décroisse indéfiniment, tandis que n augmente ; mais cette condition ne suffit pas, et il faut encore que, pour des valeurs croissantes de n, les différentes sommes
>
> $$\begin{aligned} &u_n + u_{n+1}, \\ &u_n + u_{n+1} + u_{n+2}, \\ &\text{etc.} \dots \end{aligned}$$
>
> c'est-à-dire, les sommes des quantités
>
> $$u_n, u_{n+1}, u_{n+2}, \text{etc.} \dots$$
>
> prises, à partir de la première, en tel nombre que l'on voudra, finissent par obtenir constamment des valeurs numériques inférieures à toute limite assignable. Réciproquement, lorsque ces diverses conditions sont remplies, la convergence de la série est assurée.

Suivent plusieurs exemples, puis le passage examiné dans la section 6.2.3.

Avec ce critère, Cauchy donne le moyen de savoir si une suite de réels est convergente *sans parler de sa limite*. Il s'agit d'un véritable tour de force. Désormais, plus besoin de s'appuyer sur l'intuition géométrique de la droite réelle pour dire que la droite réelle « n'a pas de trous ». Le critère de Cauchy est une caractérisation purement numérique de l'absence de trou.

On notera que Cauchy n'éprouve absolument pas le besoin de donner une preuve du critère. Et pour cause ! Aucune définition précise de la notion de nombre réel n'était encore clarifiée.

Nous voyons par contre aujourd'hui dans cet énoncé une clarification majeure concernant la nature des nombres réels, autrement dit une base sur laquelle peuvent être définis les nombres réels. C'est à Cantor que revient la mérite d'avoir construit une définition des nombres réels à partir des « suites de Cauchy de nombres rationnels ».

6.2 Continuité : globale, locale ou ponctuelle ?

6.2.1 Continuité des fonctions : une définition problématique

Nous entrons maintenant dans le vif de notre sujet, en examinant en détail le paragraphe un peu ambigu où sont proposées plusieurs définitions pour la notion de continuité d'une fonction (d'une variable réelle) définie sur un intervalle. Il apparait dans le chapitre II, pages 34-35.

§. 2.e *De la continuité des Fonctions.*

Parmi les objets qui se rattachent à la considération des infiniment petits, on doit placer les notions relatives à la continuité ou à la discontinuité des fonctions. Examinons d'abord sous ce point de vue les fonctions d'une seule variable.

Soit $f(x)$ une fonction de la variable x, et supposons que, pour chaque valeur de x intermédiaire entre deux limites données, cette fonction admette constamment une valeur unique et finie. Si, en partant d'une valeur de x comprise entre ces limites, on attribue à la variable x un accroissement infiniment petit α, la fonction elle-même recevra pour accroissement la différence

$$f(x+\alpha) - f(x),$$

qui dépendra en même temps de la nouvelle variable α et de la valeur de x. Cela posé, la fonction $f(x)$ sera, entre les deux limites assignées à la variable x, fonction *continue* de cette variable, si, pour chaque valeur de x intermédiaire entre ces limites, la valeur numérique de la différence

$$f(x+\alpha) - f(x)$$

décroît indéfiniment avec celle de α. En d'autres termes, *la fonction $f(x)$ restera continue par rapport à x entre les limites données, si, entre ces limites, un accroissement infiniment petit de la variable produit toujours un accroissement infiniment petit de la fonction elle-même.*

On dit encore que la fonction $f(x)$ est, dans le voisinage d'une valeur particulière attribuée à la variable x, fonction continue de cette variable, toutes les fois qu'elle est continue entre deux limites de x, même très rapprochées, qui renferment la valeur dont il s'agit.

Enfin, lorsqu'une fonction $f(x)$ cesse d'être continue dans le voisinage d'une valeur particulière de la variable x, on dit qu'elle devient alors *discontinue* et qu'il y a pour cette valeur particulière *solution de continuité.*

Dans ce que nous appellerons la *définition 1*, Cauchy s'attaque *pour commencer* à la définition de la continuité d'une fonction sur un intervalle $[x_0, x_1]$, c'est-à-dire selon ses termes, « entre deux limites assignées de la variable x » (on peut hésiter entre intervalle fermé ou ouvert, mais cela ne change pas pour l'essentiel la discussion qui suit).

Cela posé, la fonction $f(x)$ sera, entre les deux limites assignées à la variable x, fonction continue de cette variable, si, pour chaque valeur de x intermédiaire entre ces limites, la valeur numérique de la différence

$$f(x+\alpha) - f(x),$$

décroit indéfiniment avec celle de α.

Rappelons que *valeur numérique* signifie à l'époque ce que nous désignons aujourd'hui par *valeur absolue* (voir section 6.1.1).

Une traduction contemporaine fidèle de cette définition 1 semble être que si une suite (α_n) de réels tend vers 0, alors la suite $(f(x+\alpha_n) - f(x))$ tend aussi vers 0. Ce qui revient à la définition

actuelle de « fonction continue en tout point de l'intervalle », c'est-à-dire avec des quantificateurs portant sur des variables réelles :

$$\forall x \in [x_0, x_1]\ \forall \varepsilon > 0\ \exists \eta > 0\ \forall \alpha \in [-\eta, +\eta] \quad |f(x+\alpha) - f(x)| < \varepsilon$$

(pour ne pas alourdir encore cet énoncé, ni non plus l'éloigner trop de la formulation de Cauchy, nous n'insisterons pas plus que lui sur le fait que $x + \alpha$ doit encore être sur l'intervalle $[x_0, x_1]$).

Il y a évidement un certain effort à faire pour obtenir cette traduction. Et surtout elle nous laisse comme un gout amer dans la bouche. Car si Cauchy avait eu clairement cela en tête, pourquoi n'aurait-il pas commencé par définir la continuité en un point ? [1]

Résumant sa pensée, soulignant la mise en forme définitive par un passage en italique, Cauchy énonce ensuite ce que nous appellerons la *définition 1bis*, 1bis parce qu'elle est simplement censée répéter plus clairement la définition 1.

> En d'autres termes, *la fonction $f(x)$ restera continue par rapport à x entre les limites données, si, entre ces limites, un accroissement infiniment petit de la variable produit toujours un accroissement infiniment petit de la fonction elle-même.*

Mais là, la traduction la plus fidèle de cette phrase en langage moderne est celle de *fonction uniformément continue sur l'intervalle*. C'est-à-dire avec les quantificateurs :

$$\forall \varepsilon > 0\ \exists \eta > 0\ \forall x \in [x_0, x_1]\ \forall \alpha \in [-\eta, +\eta] \quad |f(x+\alpha) - f(x)| < \varepsilon.$$

Vient ensuite ce que nous appellerons la *définition 2*, qui est une définition locale de la continuité. Cette définition venant après la définition 1 (ou 1bis) nous laisse penser que cette première n'avait pas un caractère « local » dans l'esprit de Cauchy.

> On dit encore que la fonction $f(x)$ est, dans le voisinage d'une valeur particulière attribuée à la variable x, fonction continue de cette variable, toutes les fois qu'elle est continue entre deux limites de x, même très rapprochées, qui renferment la valeur dont il s'agit.

La traduction la plus fidèle en langage moderne semble ici être : la fonction sera dite continue au voisinage de x si on peut trouver un intervalle contenant x sur laquelle la fonction est *globalement* continue (c'est-à-dire uniformément continue si on accepte la traduction proposée de la définition 1bis).

Insistons sur le fait que cette définition serait parfaitement inutile si on avait une conception purement locale, ou pire, ponctuelle, de la continuité dans la définition 1. Par contre, elle devient indispensable si la première forme de continuité envisagée est globale et uniforme sur un intervalle, car il faut quand même pouvoir parler de la continuité d'une fonction comme $f(x) = 1/x$ sur l'intervalle $]0, 1]$ par exemple.

Notons aussi que n'apparait jamais dans ce texte la notion de continuité en un point, au sens où nous l'entendons aujourd'hui.

Ce que nous appelons la définition 3, enfin, concerne la discontinuité en un point. La phrase est particulièrement malaisée à interpréter :

> Enfin, lorsqu'une fonction $f(x)$ cesse d'être continue dans le voisinage d'une valeur particulière de la variable x, on dit qu'elle devient alors *discontinue* et qu'il y a pour cette valeur particulière *solution de continuité.*

1. Comme on le fait aujourd'hui sans jamais soulever le moindre problème à cet égard.

Notons déjà que le mot *solution* est ici employé dans son sens étymologique de destruction et de désagrégation. Un point de rupture de continuité semble donc être un point au voisinage duquel la fonction n'est plus continue, alors qu'elle est continue au voisinage de tout point voisin distinct, comme par exemple le point 0 pour la fonction $f(x) = 1/x$.

Des exemples de discontinuités plus perverses, fournis par des fonctions tarabiscotées comme : x si x est rationnel, 0 sinon, ne sont tout bonnement pas envisagés.

Et il semble peu probable que l'on puisse admettre dans le cadre fixé par Cauchy qu'une telle fonction soit continue au point 0. Bien au contraire, la continuité envisagée est toujours ou bien globale (sur un intervalle), ou bien locale (au voisinage de chaque point d'un intervalle).

Pour nous résumer, disons que les notions modernes de continuité qui nous semblent traduire le mieux les notions relativement floues de Cauchy sont celles de *continuité uniforme pour le cas d'un intervalle fermé borné, et celle de continuité localement uniforme pour un intervalle arbitraire.*

On peut se demander si l'interprétation « uniforme » que nous proposons pour la définition 1bis se trouve plutôt renforcée ou plutôt infirmée dans la suite du texte.

Les preuves de continuité qui sont données pour les fonctions usuelles, et que nous ne reproduisons pas ici, peuvent en fait être lues comme rigoureuses, aussi bien du point de vue de la continuité en tout point que de celui de la continuité uniforme ou localement uniforme.

C'est plutôt à l'occasion des preuves de théorèmes réputées fautives que nous pouvons constater à quel point l'uniformité semble présente en filigrane.

Comme dit Giusti : « l'erreur est fille de la nécessité : on se trompe parce qu'on ne peut pas faire autrement, et donc elle est plus que tout autre passage capable de mettre en évidence le mécanisme de la pensée. »

Nous allons voir qu'une dose suffisante d'uniformité rend les preuves fautives tout à fait correctes.

6.2.2 Continuité des fonctions de plusieurs variables

Le premier théorème « faux » que nous examinerons est le théorème suivant : une fonction de plusieurs variables qui est séparément continue par rapport à chaque variable est continue par rapport à l'ensemble des variables (le texte de Cauchy, extrait du chapitre II, pages 37-39, suit immédiatement).

> Soit maintenant
>
> $$f(x, y, z \dots)$$
>
> une fonction de plusieurs variables $x, y, z \dots$; et supposons que, dans le voisinage de valeurs particulières $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \dots$ attribuées à ces variables, $f(x, y, z \dots)$ soit à-la-fois fonction continue de x, fonction continue de y, fonction continue de z, etc. On prouvera aisément que, si l'on désigne par $\alpha, \beta, \gamma \dots$ des quantités infiniment petites, et si l'on attribue à $x, y, z \dots$ les valeurs $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \dots$, ou des valeurs très-voisines, la différence
>
> $$f(x + \alpha, y + \beta, z + \gamma \dots) - f(x, y, z \dots)$$
>
> sera elle-même infiniment petite. En effet, il est clair que, dans l'hypothèse précédente, les valeurs numériques des différences
>
> $$f(x + \alpha, y, z \dots) - f(x, y, z \dots),$$
> $$f(x + \alpha, y + \beta, z \dots) - f(x + \alpha, y, z \dots),$$
> $$f(x + \alpha, y + \beta, z + \gamma \dots) - f(x + \alpha, y + \beta, z \dots),$$
> etc. . . .

décroitront indéfiniment avec celles des quantités variables $\alpha, \beta, \gamma \ldots$, savoir, la valeur numérique de la première différence avec la valeur numérique de α, celle de la seconde différence avec la valeur numérique de β, celle de la troisième avec la valeur numérique de γ, et ainsi de suite. On doit en conclure que la somme de toutes ces différences, savoir

$$f(x+\alpha, y+\beta, z+\gamma \ldots) - f(x, y, z \ldots),$$

convergera vers la limite zéro, si $\alpha, \beta, \gamma \ldots$ convergent vers cette même limite. En d'autres termes,

$$f(x+\alpha, y+\beta, z+\gamma \ldots)$$

aura pour limite

$$f(x, y, z \ldots).$$

. . . . . .

1.er Théorème. *Si les variables $x, y, z \ldots$ ont pour limites respectives les quantités fixes et déterminées* $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \ldots$, *et que la fonction $f(x, y, z \ldots)$ soit continue par rapport à chacune des variables $x, y, z \ldots$ dans le voisinage du système des valeurs particulières*

$$x = \mathrm{X}, y = \mathrm{Y}, z = \mathrm{Z} \ldots,$$

$f(x, y, z \ldots)$ aura pour limite $f(\mathrm{X}, \mathrm{Y}, \mathrm{Z} \ldots)$.

Nous faisons deux remarques concernant la preuve fournie par Cauchy. D'abord, ce qui est en vue est la continuité locale et non pas la continuité ponctuelle. La preuve n'est pas écrite pour fonctionner en *un* point $(\mathrm{X}, \mathrm{Y}, \mathrm{Z} \ldots)$, au contraire il est explicitement dit que tout se passe de la même manière pour un $(x, y, z \ldots)$ suffisamment voisin de $(\mathrm{X}, \mathrm{Y}, \mathrm{Z} \ldots)$. Ce fait n'apparait d'ailleurs clairement que dans la preuve et non dans l'énoncé du théorème.

La deuxième remarque est que la preuve fonctionne si on comprend la continuité par rapport à chaque variable séparément, x, y, $z \ldots$ comme devant être à chaque fois une continuité dans laquelle le « $\forall \varepsilon > 0\ \exists \eta > 0$ » a une signification doublement uniforme : dans la continuité séparée par rapport à la variable y, seul y a le droit de varier mais le $\eta > 0$ ne doit dépendre que de $\varepsilon > 0$. C.-à-d. avec les quantificateurs, pour l'exemple de la continuité par rapport à la variable y, et en nous limitant à trois variables :

$$\forall \varepsilon > 0\ \exists \eta > 0\ \forall x, y, z\ \forall \beta \quad (|\beta| < \eta \Rightarrow |f(x, y+\beta, z) - f(x, y, z)| < \varepsilon)$$

où le domaine de variation de x, y, z est un voisinage de $\mathrm{X}, \mathrm{Y}, \mathrm{Z}$.

En langage moderne cela s'appellerait de l'équicontinuité localement uniforme.

6.2.3 Somme d'une série convergente de fonctions continues

Nous examinons maintenant un autre fameux théorème « faux » de Cauchy : *toute série convergente de fonctions continues converge vers une fonction continue,* selon un extrait du chapitre VI, pages 130-131.

La série

$$u_0, u_1, u_2, u_3, \text{etc.} \ldots$$

étant supposée convergente, si l'on désigne sa somme par s, et par s_n la somme de ses n premiers

termes, on trouvera

$$\begin{aligned} s &= u_0 + u_1 + u_2 + \cdots + u_{n-1} + u_n + u_{n+1} + \text{etc.} \ldots . \\ &= s_n + u_n + u_{n+1} + \text{etc.} \ldots , \end{aligned}$$

et par suite

$$s - s_n = u_n + u_{n+1} + \text{etc.} \ldots .$$

De cette dernière équation, il résulte que les quantités

$$u_n, u_{n+1}, u_{n+2}, \text{etc.} \ldots .$$

formeront une nouvelle série convergente dont la somme sera équivalente à $s - s_n$. Si l'on représente cette même somme par r_n, on aura

$$s = s_n + r_n;$$

et r_n sera ce qu'on appelle le *reste* de la série (1) à partir du $n.^{\text{me}}$ terme.

Lorsque, les termes de la série (1) renfermant une même variable x, cette série est convergente, et ses différents termes fonctions continues de x, dans le voisinage d'une valeur particulière attribuée à cette variable ;

$$s_n, \quad r_n \quad \text{et} \quad s$$

sont encore trois fonctions de la variable x, dont la première est évidemment continue par rapport à x dans le voisinage de la valeur particulière dont il s'agit. Cela posé, considérons les accroissements que reçoivent ces trois fonctions, lorsqu'on fait croitre x d'une quantité infiniment petite α. L'accroissement de s_n sera, pour toutes les valeurs possibles de n, une quantité infiniment petite ; et celui de r_n deviendra insensible en même temps que r_n, si l'on attribue à n une valeur très-considérable. Par suite, l'accroissement de la fonction s ne pourra être qu'une quantité infiniment petite. De cette remarque on déduit immédiatement la proposition suivante.

1.$^{\text{er}}$ Théorème. *Lorsque les différents termes de la série (1) sont des fonctions d'une même variable x, continues par rapport à cette variable dans le voisinage d'une valeur particulière pour laquelle la série est convergente, la somme s de la série est aussi, dans le voisinage de cette valeur particulière, fonction continue de x.*

On note que l'énoncé du théorème « faux » concerne la continuité locale, au voisinage du point x, et non pas ponctuelle, de la somme de la série.

Comme le signale Giusti il y a manifestement un *lapsus calami* lorsque Cauchy énonce dans le théorème que la série est convergente « en un point particulier ». Au contraire, dans la preuve qui précède, il est bien indiqué autre chose :

> Lorsque, les termes de la série (1) renfermant une même variable x, cette série est convergente, et ses différents termes fonctions continues de x, dans le voisinage d'une valeur particulière attribuée à cette variable ; [...]

Ceci dit si on lit l'énoncé avec nos lunettes modernes, il s'agit d'un théorème « qui souffre des exceptions », une des plus notables étant fournie par la série trigonométrique

$$\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \cdots .$$

Cette « exception » est signalée par exemple par Abel dans une note au bas de la page 316 de son mémoire au Journal für die reine und angewandte Mathematik (Crelle's Journal) 1, (1826), pages 311-339. Pour chaque valeur particulière de x la série converge : quand $x \in]-\pi, +\pi[$ la

limite est $x/2$, quand $x = \pm\pi$ la limite est 0, et la série est périodique de période 2π. Mais la somme admet une solution de continuité en chaque point $\pi \pm 2k\pi$. Il ne suffit donc pas que la série soit convergente ponctuellement pour que la limite soit continue.

L'analyse qui suit est entièrement prise dans l'article de Giusti.

Si Abel signale le premier l'exception, la série en question était connue déjà bien avant. Fourier en avait parlé dans son manuscrit *Théorie de la propagation de la chaleur dans les solides* présenté à l'Institut de France en 1807. Il en parlait comme d'une série dont la somme est bien connue. Et en effet Euler avait calculé cette somme et publié le résultat en 1783 (voir Euleri Opera Omnia, XV, p. 435-497). En outre ce résultat est reporté dans le traité de Lacroix *Traité du calcul différentiel et intégral*. Paris, Duprat, 1797, vol 1, p. 75. Il est donc plus que probable que Cauchy connaissait bien cette série trigonométrique en 1821, quand il écrit son cours. Il était en tout cas certainement au courant du contre-exemple en 1833, lorsqu'il réitère son théorème et sa démonstration sans en changer une virgule dans les *Résumés analytiques*.

Et lorsqu'en 1853, il finit par reconnaitre que son théorème a besoin de quelque précision dans une note aux Comptes rendus de l'Académie des sciences, c'est plus sur le ton de quelqu'un ennuyé de devoir revenir sur des détails minimes que sur celui de quelqu'un qui répare une erreur. La raison la plus probable en est que *Cauchy pensait que la série ne constituait pas vraiment un contre-exemple à son théorème.*

Quelle était la valeur à attribuer à la somme pour la valeur particulière π de la variable pour les prédécesseurs de Cauchy ? Euler suggère de prendre la somme continue à gauche. Tandis que Fourier considère le segment vertical $[-\pi/2, +\pi/2]$ comme faisant partie intégrante du graphe de la fonction. Et effectivement le graphe des sommes partielles converge bien vers le graphe décrit par Fourier.

Et Dirichlet a une conception similaire en 1837 quand il dit que lorsque se présente une solution de continuité la fonction a proprement deux valeurs (G. Lejeune Dirichlet's Werke, volume 1, page 156).

Il est donc possible que Cauchy considérait que la série n'était pas vraiment convergente au point π, et donc ne satisfaisait pas aux hypothèses du théorème, parce que la fonction y avait plusieurs valeurs.

Mais l'interprétation la plus simple semble que, pour Cauchy, la série ne satisfaisait pas aux hypothèses du théorème parce que, au voisinage de la valeur particulière π, la série ne convergeait pas « toujours » au sens correct de la chose, qui est le sens de la convergence uniforme.

C'est en tout cas l'option qu'il retient dans la note ci-dessous avec les hypothèses suivantes dans l'énoncé modifié du théorème :

> Si les différents termes de la série [...] sont des fonctions de la variable réelle x, continues, par rapport à cette variable, entre des limites données, si, d'ailleurs, la somme
>
> (3) $$u_n + u_{n+1} + \cdots + u_{n'-1}$$
>
> devient toujours infiniment petite pour des valeurs infiniment grandes des nombres entiers n et $n' > n$, [...]

Pour que la lectrice se fasse une opinion plus précise par elle-même, le mieux est de reproduire la note Cauchy 1853 en entier, parue dans les Comptes rendus hebdomadaires des séances de l'Académie des sciences du 14 mars 1853.

ANALYSE MATHÉMATIQUE. – *Note sur les séries convergentes dont les divers termes sont des fonctions continues d'une variable réelle ou imaginaire, entre des limites données ; par* **M. Augustin Cauchy**.

En établissant, dans mon *Analyse algébrique*, les règles générales relatives à la convergence des séries, j'ai, de plus, énoncé le théorème suivant :

Lorsque les différents termes de la série

$$u_0, u_1, u_2, \dots, u_n, u_{n+1}, \dots, \tag{1}$$

sont des fonctions d'une même variable x, continues par rapport à cette variable, dans le voisinage d'une valeur particulière pour laquelle la série est convergente, la somme s de la série est aussi, dans le voisinage de cette valeur particulière, fonction continue de x.

Comme l'ont remarqué MM. Bouquet et Briot, ce théorème se vérifie pour les séries ordonnées suivant les puissances ascendantes d'une variable. Mais, pour d'autres séries, il ne saurait être admis sans restriction. Ainsi, par exemple, il est bien vrai que la série

$$\sin x, \frac{\sin 2x}{2}, \frac{\sin 3x}{3}, \dots, \tag{2}$$

toujours convergente pour des valeurs réelles de x, a pour somme une fonction de x qui reste continue, tandis que x, supposée réelle, varie, dans le voisinage d'une valeur distincte d'un multiple $\pm 2n\pi$ de la circonférence 2π, et qui se réduit, en particulier, à $\frac{\pi - x}{2}$, entre les limites $x = 0$, $x = 2\pi$. Mais à ces limites mêmes, la somme s de la série (2) devient discontinue, et cette somme, considérée comme fonction de la variable réelle x, acquiert, à la place de la valeur

$$+\frac{\pi}{2} \text{ ou } -\frac{\pi}{2},$$

donnée par la formule

$$s = \frac{\pi - x}{2},$$

la valeur *singulière* $s = 0$, qui reparait encore quand on suppose

$$x = \pm 2n\pi,$$

n étant un nombre entier quelconque.

Au reste, il est facile de voir comment on doit modifier l'énoncé du théorème, pour qu'il n'y ait plus lieu à aucune exception. C'est ce que je vais expliquer en peu de mots.

D'après la définition proposée dans mon *Analyse algébrique*, et généralement adoptée aujourd'hui, une fonction u de la variable réelle x sera *continue*, entre deux limites données de x, si, cette fonction admettant pour chaque valeur intermédiaire de x une valeur unique et finie, un accroissement infiniment petit attribué à la variable produit toujours, entre les limites dont il s'agit, un accroissement infiniment petit de la fonction elle-même. Cela posé, concevons que la série (1) reste convergente, et que ses divers termes soient fonctions continues d'une variable réelle x, pour toutes les valeurs de x renfermées entre certaines limites. Soient alors

s la somme de la série ;

s_n la somme de ses n premiers termes ;

et $r_n = s - s_n = u_n + u_{n+1} + \dots$ le reste de la série indéfiniment prolongée à partir du terme général u_n.

Si l'on nomme n' un nombre entier supérieur à n, le reste r_n ne sera autre chose que la limite vers laquelle convergera, pour des valeurs croissantes de n', la différence

$$s'_n - s_n = u_n + u_{n+1} + \dots + u_{n'-1}. \tag{3}$$

Concevons, maintenant, qu'en attribuant à n une valeur suffisamment grande, on puisse rendre, pour toutes les valeurs de x comprises entre les limites données, le module de l'expression (3) (quel que soit n'), et, par suite, le module de r_n, inférieur à un nombre ε aussi petit que l'on voudra. Comme un accroissement attribué à x pourra encore être supposé assez rapproché de zéro pour que l'accroissement correspondant de s_n offre un module inférieur à un nombre aussi petit que l'on voudra, il est clair qu'il suffira d'attribuer au nombre n une valeur infiniment grande, et à l'accroissement de x une valeur infiniment petite, pour démontrer, entre les limites données, la continuité de la fonction

$$s = s_n + r_n.$$

Mais cette démonstration suppose évidemment que l'expression (3) remplit la condition ci-dessus énoncée, c'est-à-dire que cette expression devient infiniment petite pour une valeur infiniment grande attribuée au nombre entier n. D'ailleurs, si cette condition est remplie, la série (1) sera évidemment convergente. En conséquence, on peut énoncer le théorème suivant :

1er *Théorème.* Si les différents termes de la série

$$u_0, u_1, u_2, \ldots, u_n, u_{n+1}, \ldots, \tag{1}$$

sont des fonctions de la variable réelle x, continues, par rapport à cette variable, entre des limites données, si, d'ailleurs, la somme

$$u_n + u_{n+1} + \cdots + u_{n'-1} \tag{3}$$

devient toujours infiniment petite pour des valeurs infiniment grandes des nombres entiers n et $n' > n$, la série (1) sera convergente, et la somme s de la série (1) sera, entre les limites données, fonction continue de la variable x.

Terminons par une remarque sur l'énoncé modifié et sa preuve. Il ne s'agit pas de série uniformément convergente de fonctions ponctuellement continues dont la somme serait ponctuellement continue, mais bien de série uniformément convergente de fonctions uniformément continues (dans le voisinage d'un point) dont la somme est uniformément continue (dans ce même voisinage).

6.3 Fonction dérivée et théorème des accroissements finis

Nous étudions dans cette section la notion de fonction dérivable chez Cauchy.

Commençons par lire la définition de la notion de fonction dérivée dans son *Résumé des Leçons données à l'École royale Polytechnique sur le calcul infinitésimal* de 1823 (Cauchy 1823, page 9), réimprimé dans les *Œuvres complètes, série II, tome 4*, pages 22-23.

TROISIÈME LEÇON.

Dérivées des Fonctions d'une seule Variable

Lorsque la fonction $y = f(x)$ reste continue entre deux limites données de la variable x, et que l'on assigne à cette variable une valeur comprise entre les deux limites dont il s'agit, un accroissement infiniment petit, attribué à la variable, produit un accroissement infiniment petit de la fonction elle-même. Par conséquent, si l'on pose alors $\Delta x = i$, les deux termes du *rapport aux différences*

$$\frac{\Delta y}{\Delta x} = \frac{f(x+i) - f(x)}{i} \tag{1}$$

seront des quantités infiniment petites. Mais, tandis que ces deux termes s'approcheront indéfiniment et simultanément de la limite zéro, le rapport lui-même pourra converger vers une autre limite, soit positive, soit négative. Cette limite, lorsqu'elle existe, a une valeur déterminée, pour chaque valeur particulière de x ; mais elle varie avec x. Ainsi, par exemple, si l'on prend $f(x) = x^m$, m désignant un nombre entier, le rapport entre les différences infiniment petites sera

$$\frac{(x+i)^m - x^m}{i} = mx^{m-1} + \frac{m(m-1)}{1.2}x^{m-2}i + \cdots + i^{m-1}$$

et il aura pour limite la quantité mx^{m-1}, c'est-à-dire, une nouvelle fonction de la variable x. Il en sera de même en général ; seulement, la forme de la fonction nouvelle qui servira de limite au rapport $\frac{f(x+i)-f(x)}{i}$ dépendra de la forme de la fonction proposée $y = f(x)$. Pour indiquer cette dépendance, on donne à la nouvelle fonction le nom de *fonction dérivée*, et on la désigne, à l'aide d'un accent, par la notation

$$y' \text{ ou } f'(x).$$

Remarquons que la valeur de la dérivée en un point n'intéresse pas vraiment Cauchy, mais que c'est plutôt la notion de fonction dérivée qu'il cherche à définir. *A priori*, il y a donc au moins deux lectures modernes de cette définition, selon que l'on demande une convergence en tout point ou une convergence uniforme du taux d'accroissement moyen vers la fonction dérivée. Avec les quantificateurs, la définition « ponctuelle » s'écrit :

$$\forall x \in [x_0, x_1]\ \forall \varepsilon > 0\ \exists \eta > 0\ \forall \alpha \in [-\eta, +\eta] \quad \left| \frac{f(x+\alpha) - f(x)}{\alpha} - f'(x) \right| < \varepsilon.$$

C'est la définition habituellement donnée aujourd'hui. Quant à la définition « uniforme », elle s'écrit :

$$\forall \varepsilon > 0\ \exists \eta > 0\ \forall x \in [x_0, x_1]\ \forall \alpha \in [-\eta, +\eta] \quad \left| \frac{f(x+\alpha) - f(x)}{\alpha} - f'(x) \right| < \varepsilon.$$

Cette définition implique que la fonction dérivée est elle-même (uniformément) continue[2], comme limite uniforme de la suite de fonctions continues $v_n(x) := n\Big(f(x + \frac{1}{n}) - f(x)\Big)$.

Dans la page suivante du *Résumé* apparaissent plusieurs exemples de calculs de dérivées, dont

pour $y = \frac{a}{x}$, $\frac{\Delta y}{\Delta x} = \frac{\frac{a}{x+i} - \frac{a}{x}}{i} = -\frac{a}{x(x+i)}$, $y' = -\frac{a}{x^2}$;

donc Cauchy dérive une fonction qui n'est pas continue, puisqu'elle devient infinie en 0 ; quelques lignes plus loin, il dérive aussi la fonction $y = x^a$, qui est continue entre 0 et 1 sans que sa dérivée le soit si $a = 1/2$ par exemple. C'est la seule explication que j'ai trouvée pour expliquer la présence de l'hypothèse que $f'(x)$ soit continue dans le corollaire à la fin de ce paragraphe.

Remarquez aussi que la preuve donnée par Cauchy concernant la dérivée de la fonction puissance fonctionne parfaitement dans le cadre de la définition « uniforme » (sur un intervalle fermé borné).

Venons en maintenant au théorème des accroissements finis, pages 25-28.

2. On peut d'ailleurs montrer que cette définition équivaut au fait que la fonction admet une dérivée uniformément continue sur l'intervalle.

SEPTIÈME LEÇON.

Valeurs de quelques expressions qui se présentent sous les formes indéterminées $\frac{\infty}{\infty}$, ∞^0, *etc. Relation qui existe entre le rapport aux Différences finies et la Fonction dérivée.*

. . . . . .

Nous allons maintenant faire connaitre une relation digne de remarque* qui existe entre la dérivée $f'(x)$ d'une fonction quelconque $f(x)$, et le rapport aux différences finies $\frac{f(x+h)-f(x)}{h}$. Si dans ce rapport on attribue à x une valeur particulière x_0, et si l'on fait, en outre, $x_0 + h = \mathrm{X}$, il prendra la forme $\frac{f(\mathrm{X})-f(x_0)}{\mathrm{X}-x_0}$. Cela posé, on établira sans peine la proposition suivante.

Théorème. *Si, la fonction* $f(x)$ *étant continue entre les limites* $x = x_0, x = \mathrm{X}$, *on désigne par* A *la plus petite, et par* B *la plus grande des valeurs que la fonction dérivée* $f'(x)$ *reçoit dans cet intervalle, le rapport aux différences finies*

$$\frac{f(\mathrm{X}) - f(x_0)}{\mathrm{X} - x_0} \tag{4}$$

sera nécessairement compris entre A *et* B.

Démonstration. Désignons par δ, ε deux nombres très-petits, le premier étant choisi de telle sorte que, pour des valeurs numériques de i inférieures à δ, et pour une valeur quelconque de x comprise entre les limites x_0, X, le rapport

$$\frac{f(x+i) - f(x)}{i}$$

reste toujours supérieur à $f'(x) - \varepsilon$ et inférieur à $f'(x) + \varepsilon$. Si, entre les limites x_0, X, on interpose $n-1$ valeurs nouvelles de la variable x, savoir,

$$x_1, x_2, \ldots, x_{n-1},$$

de manière à diviser la différence $\mathrm{X} - x_0$ en éléments

$$x_1 - x_0, x_2 - x_1, \ldots, \mathrm{X} - x_{n-1}$$

qui, étant tous de même signe, aient des valeurs numériques inférieures à δ ; les fractions

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0}, \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \ldots, \frac{f(\mathrm{X}) - f(x_{n-1})}{\mathrm{X} - x_{n-1}}, \tag{5}$$

se trouvant comprises, la première entre les limites $f'(x_0) - \varepsilon$, $f'(x_0) + \varepsilon$, la seconde entre les limites $f'(x_1) - \varepsilon$, $f'(x_1) + \varepsilon$, etc.... seront toutes supérieures à la quantité $\mathrm{A} - \varepsilon$, et inférieures à la quantité $\mathrm{B} + \varepsilon$. D'ailleurs, les fractions (5) ayant des dénominateurs de même signe, si l'on divise la somme de leurs numérateurs par la somme de leurs dénominateurs, on obtiendra une fraction *moyenne*, c'est-à-dire, comprise entre la plus petite et la plus grande de celles que l'on considère [*voyez* l'Analyse algébrique, note II, 12.e théorème]. L'expression (4), avec laquelle cette moyenne coïncide, sera donc elle-même renfermée entre les limites $\mathrm{A} - \varepsilon$, $\mathrm{B} + \varepsilon$; et, comme cette conclusion subsiste, quelque petit que soit le nombre ε, on peut affirmer que l'expression (4) sera comprise entre A et B.

* On peut consulter sur ce sujet un mémoire de M. *Ampère*, inséré dans le 13.e cahier du *Journal de l'École polytechnique.*

La preuve du théorème des accroissements finis donnée par Cauchy a le mérite d'être simple et *naturelle*, contrairement aux preuves actuellement en vigueur dans les cours élémentaires de calcul différentiel. On constate sans difficulté que la définition « uniforme » pour la notion de fonction dérivée rend cette preuve élémentaire de Cauchy parfaitement rigoureuse (alors qu'elle est souvent

considérée comme incorrecte, parce qu'on se réfère à la définition « ponctuelle »).

En outre on constate également que les théorèmes usuels concernant la dérivée de fonctions élémentaires, ou sur la dérivée d'un produit, d'une somme, d'un quotient (sur un intervalle où le dénominateur reste de signe constant et en valeur absolue $> \eta$ strictement positif donné), sont de démonstration aussi facile en version « uniforme » ou « localement uniforme » (dérivabilité uniforme sur tout intervalle fermé borné contenu dans l'intervalle de définition) qu'en version « ponctuelle ».

Terminons cette section par un commentaire sur une deuxième version du théorème des accroissements finis, énoncée par Cauchy dans le corollaire donné dans l'encadré suivant, page 28.

Chez Cauchy, cette deuxième forme (rapport aux différences finies égal à une valeur de la dérivée sur l'intervalle) est prouvée en utilisant la continuité de la fonction dérivée. Comme nous l'avons déjà signalé, cette continuité est nécessaire lorsqu'on adopte la définition uniforme.

Il est bien connu que la première forme du théorème des accroissements finis se généralise aisément pour une fonction de plusieurs variables continument dérivable (avec essentiellement la même preuve que ci-dessus d'ailleurs), ce qui n'est pas le cas de la seconde forme.

Corollaire. Si la fonction dérivée $f'(x)$ est elle-même continue entre les limites $x = x_0$, $x = \mathrm{X}$, en passant d'une limite à l'autre, cette fonction variera de manière à rester toujours comprise entre les deux valeurs A et B, et à prendre successivement toutes les valeurs intermédiaires. Donc alors toute quantité moyenne entre A et B sera une valeur de $f'(x)$ correspondante à une valeur de x renfermée entre les limites x_0 et $\mathrm{X} = x_0 + h$ ou, ce qui revient au même, à une valeur de x de la forme

$$x_0 + \theta h = x_0 + \theta(\mathrm{X} - x_0),$$

θ désignant un nombre inférieur à l'unité. En appliquant cette remarque à l'expression (4), on en conclura qu'il existe, entre les limites 0 et 1, une valeur de θ propre à vérifier l'équation

$$\frac{f(\mathrm{X}) - f(x_0)}{\mathrm{X} - x_0} = f'[x_0 + \theta(\mathrm{X} - x_0)],$$

ou, ce qui revient au même, la suivante

$$\frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0 + \theta h). \tag{6}$$

Cette dernière formule devant subsister, quelle que soit la valeur de x représentée par x_0, pourvu que la fonction $f(x)$ et la dérivée $f'(x)$ restent continues entre les valeurs extrêmes $x = x_0$, $x = x_0 + h$, on aura généralement, sous cette condition,

$$\frac{f(x + h) - f(x)}{h} = f'(x + \theta h), \tag{7}$$

puis, en écrivant Δx au lieu de h, on en tirera

$$f(x + \Delta x) - f(x) = f'(x + \theta \Delta x).\Delta x. \tag{8}$$

Il est essentiel d'observer que, dans les équations (7) et (8), θ désigne toujours un nombre inconnu, mais inférieur à l'unité.

La démarche la plus courante aujourd'hui dans les cours de calcul différentiel utilise la deuxième forme pour prouver la première forme. La deuxième forme est prouvée à partir du théorème de Rolle. Mais le théorème de Rolle est lui-même prouvé en utilisant une technique « non opératoire » : considérer un point sur l'intervalle où la fonction atteint son maximum.

Or il n'y a pas d'algorithme général pour cette recherche, même si la dérivée est donnée comme fonction uniformément continue sur l'intervalle. La recherche d'un point où une fonction

uniformément continue atteint un maximum local sur un intervalle fermé borné ne peut être réalisé par un algorithme qu'en imposant des restrictions à la fonction étudiée (par exemple qu'elle ait un tableau de variation fini).

6.4 Conclusion

Le fait de savoir s'il faut lire Cauchy de manière uniforme (avec des preuves essentiellement justes) ou ponctuelle (avec de nombreuses preuves fautives) a plutôt été tranché en faveur de la seconde hypothèse par de nombreux historiens des mathématiques.

Mais on peut se demander s'ils ont vraiment fait l'effort de se plonger dans le contexte de l'époque.

Les quatre citations suivantes sont empruntées à l'article de Giusti.

En 1829, bien avant que la question « continuité ponctuelle/continuité uniforme » ait été clarifiée, Dirichlet, qui connait certainement le texte de Cauchy, écrit dans la note *Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données* :

> Désignons par h un nombre positif inférieur ou tout au plus égal à $\frac{\pi}{2}$ et par $f(\beta)$ une fonction de β qui reste continue entre les limites 0 et h ; j'entends par là une fonction qui a une valeur finie et déterminée pour toute valeur de β comprise entre 0 et h, et en outre telle que la différence $f(\beta+\varepsilon)-f(\beta)$ diminue sans limite lorsque ε devient de plus en plus petit. (Lejeune-Dirichlet 1829, page 159.)

Riemann écrit (dans une citation en note sans référence précise par H. Weber) :

> Par l'expression : la grandeur w varie d'une manière continue avec z entre les limites $z = a$, $z = b$, nous entendons ceci : Dans cet intervalle, à toute variation infiniment petite de z correspond une variation infiniment petite de w ; ou encore, en s'exprimant d'une manière plus détaillée : pour une grandeur donnée quelconque ε, 1'on peut toujours déterminer la grandeur α, de telle sorte que dans un intervalle relatif à z, plus petit que α, la différence entre deux valeurs de w ne soit jamais plus grande que ε. La continuité d'une fonction, même lorsque ce point n'est pas expressément énoncé, entraine d'après cela ce fait : la fonction est toujours finie. (Riemann 1876, page 56.)

Par ailleurs, Lebesgue, relisant Cauchy, parle de continuité au point x_0 tout en donnant une interprétation uniforme sur l'intervalle (a, b) dans ses *Leçons sur l'intégration et la recherche des fonctions primitives* :

> Pour Cauchy une fonction $f(x)$ est continue pour la valeur x_0 si, quel que soit le nombre positif ε, on peut trouver un nombre $\eta(\varepsilon)$ tel que l'inégalité $|h| \leq \eta(\varepsilon)$ entraine
>
> $$|f(x_0 + h) - f(x_0)| \leq \varepsilon;$$
>
> la fonction f est continue dans (a, b) si la correspondance entre ε et $\eta(\varepsilon)$ peut être choisie indépendamment du nombre x_0, quelconque dans (a, b). (Lebesgue 1904, pages 4-5.)

Enfin il faut citer le cours de J. Tannery (1906) :

La fonction $f(x)$ définie dans l'intervalle (a, a') est continue dans cet intervalle si à chaque nombre positif ε on peut faire correspondre un nombre positif ε' tel que l'on ait $|f(x_1) - f(x_0)| < \varepsilon$ sous la condition que x_0, x_1 appartiennent à l'intervalle (a, a') et que l'on ait $|x_1 - x_0| < \varepsilon'$. (Tannery 1906, page 61)

Commentaire personnel. Malgré la déferlante de la continuité ponctuelle au 20e siècle, on a vu réapparaitre la définition uniforme comme concept central lorsqu'on s'est posé des problèmes d'effectivité. Par exemple le livre d'analyse constructive de Bishop (1967) (*Foundations of constructive analysis*; voir aussi Bishop et Bridges 1985, *Constructive analysis*) définit la continuité sur un intervalle fermé borné comme étant la continuité uniforme, et la continuité sur un intervalle arbitraire comme étant la continuité sur tout sous intervalle fermé borné. De même la dérivabilité est définie sur un intervalle fermé borné comme étant la dérivabilité uniforme, et sur un intervalle arbitraire, comme étant la dérivabilité sur tout sous intervalle fermé borné.

Ceci n'a pas seulement l'avantage de rendre les théorèmes effectifs, mais aussi de simplifier de nombreuses preuves.

Chapitre 7

Nombres réels et fonctions continues : le théorème des valeurs intermédiaires

Introduction

Nous étudions dans les premières sections de ce chapitre le théorème des valeurs intermédiaires.

En examinant des preuves de ce théorème nous sommes amenés à discuter sa signification et la nature des objets mathématiques qu'il traite : les nombres réels d'une part, les fonctions continues d'autre part.

7.1 L'énoncé d'un théorème et sa signification intuitive

Théorème 7.1.1. *Une fonction réelle continue définie sur un intervalle* $[a, b]$ *prend sur cet intervalle toutes les valeurs entre* $f(a)$ *et* $f(b)$*.*

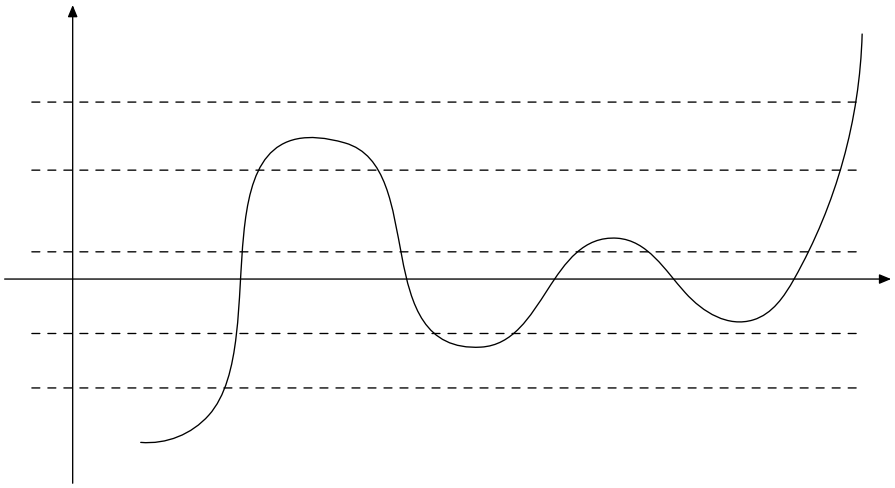


FIGURE 7.1.1 – Théorème des valeurs intermédiaires

Une signification intuitive est que « la droite réelle n'a pas de trou ».

Ce théorème est la version numérique de l'évidence géométrique suivante : une courbe continue qui joint un point situé à l'intérieur d'un triangle à un point extérieur doit couper un côté du triangle. Ou le même énoncé avec un cercle à la place d'un triangle. Ou encore : une droite partage le plan en deux parties, et une courbe qui joint un point d'un côté à un point de l'autre doit couper la droite.

Le paradoxe de la diagonale du carré, qui n'a pas de commune mesure avec le côté du carré, peut se reformuler comme suit en langage moderne :

Si tous les points du plan avaient pour coordonnées des nombres rationnels, le cercle centré au point $\mathrm{O} = (0,0)$ *et passant par* $\mathrm{B} = (1,1)$ *ne couperait pas la droite horizontale passant par* O *(la droite* OA *si on trace le carré* OABC*).*

Du point de vue du théorème des valeurs intermédiaires on peut donner l'énoncé équivalent :

La fonction $x \mapsto x^2 - 2$ *qui prend la valeur* -1 *en* 1 *et* $+2$ *en* 2 *doit prendre la valeur* 0 *en un* $x \in [1,2]$ *(qu'on a l'habitude d'écrire* $\sqrt{2}$*).*

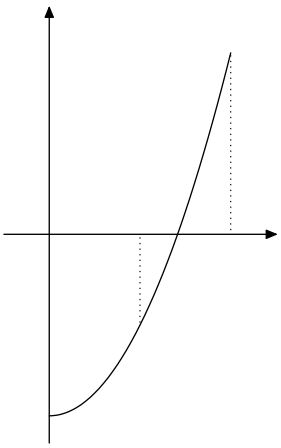


FIGURE 7.1.2 – La fonction $x \mapsto x^2 - 2$ sur l'intervalle $[0,2]$

Historiquement le théorème des valeurs intermédiaires, que nous abrégeons dans la suite en TVI, a eu une importance dans l'élaboration de la notion de fonction continue. Quand on a pris conscience qu'il faudrait donner une définition précise de cette notion si on voulait être sur un terrain plus ferme que celui offert par le calcul infinitésimal et ses intuitions, on a pu se demander si la propriété des valeurs intermédiaires ne devait pas être justement prise pour définition. Mais une telle définition semblait un peu trop ad hoc.

Parmi les fonctions usuelles, certaines ne vérifient pas le TVI, mais la raison est un discontinuité bien visible qui se manifeste par un saut. Par exemple la fonction $x \mapsto 1/x$, qui présente une forte discontinuité en 0 et qui change de signe entre -1 et $+1$ sans passer par la valeur 0. Un autre exemple est la fonction dessinée figure 7.1.3 (on n'a pas indiqué sur le dessin si la fonction est continue à droite ou à gauche au point de discontinuité). Ceci conduit à essayer de définir la continuité par l'absence de saut.

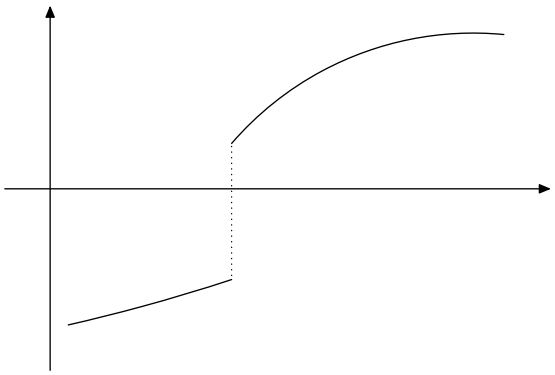


FIGURE 7.1.3 – Une fonction qui fait un saut

Informellement, on peut interdire à une fonction de sauter en imposant une condition du style : si $|x - x'|$ est très petit alors $|f(x) - f(x')|$ doit être aussi très petit. Cette idée naturelle peut cependant se décliner de différentes façons, comme nous l'avons vu dans l'étude des textes de Cauchy au chapitre 6.

Nous nous en contenterons pour le moment.

7.2 Deux preuves

Tout d'abord on simplifie un tout petit peu l'énonce en ramenant l'intervalle de définition de la fonction à $[0,1]$ (ce qui est toujours possible par un changement de variable $x \mapsto \alpha x + \beta$). Par ailleurs si le point c est sur l'intervalle ouvert d'extrémités $f(0)$ et $f(1)$, en considérant la fonction $x \mapsto f(x) - c$, on est ramené à démontrer le théorème suivant (remarquez que si on a $f(0) > 0 > f(1)$ on remplace f par $-f$ pour se ramener au cas envisagé ici).

Théorème 7.2.1. *Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue avec $f(0) < 0 < f(1)$. Alors il existe $x \in [0,1]$ tel que $f(x) = 0$.*

La première preuve que nous présentons est celle que l'on trouve dans de nombreux livres d'analyse aujourd'hui. Elle est assez proche de celle proposée par Bolzano. Bolzano l'avait élaborée pour compléter la preuve par Gauss du théorème fondamental de l'algèbre.

Preuve n° 1, aujourd'hui usuelle. On considère le sous ensemble suivant de l'intervalle unité : $\mathrm{A} = \{x \in [0,1] \mid f(x) > 0\}$. L'ensemble A contient le point 1 et il est minoré par 0. Soit alors c la borne inférieure de A. On a $c > 0$ parce que la fonction reste négative si on ne s'écarte pas trop de 0. Pour tous les points de l'intervalle $[0,c]$ la fonction est ≤ 0 car sinon A contiendrait un élément $< c$ et c ne serait pas sa borne inférieure. Par continuité on aura donc $f(c) \leq 0$. Donc c n'est pas dans A. Mais puisque c est la borne inférieure de A, pour tout $n \in \mathbb{N}$ l'intervalle $]c, c + 1/2^n] \cap [0,1]$ contient un élément de A, que nous notons x_n. La suite x_n tend vers c lorsque n tend vers l'infini, et puisque les $f(x_n)$ sont tous > 0, leur limite, qui est égale à $f(c)$ parce que la fonction est continue, est ≥ 0. En conclusion $f(c)$ est à la fois ≥ 0 et ≤ 0 donc égal à 0. □

Commentaire. On trouve au cœur de cette preuve l'idée que la droite réelle n'a pas de trou exprimée sous la forme suivante : *une partie non vide et minorée de $\mathbb{R}$ admet une borne inférieure.* Cette propriété est parfois prise comme axiome pour $\mathbb{R}$ (par exemple dans les *Foundations of modern analysis* de Dieudonné 1960). Par ailleurs la continuité de la fonction intervient sous la forme suivante : si une suite x_n de l'intervalle de définition de f tend vers une valeur c alors $f(x_n)$ tend vers $f(c)$.

La preuve suivante est plus simple et remonte à Cauchy. Cauchy la signalait en annexe dans son cours à l'École polytechnique. Ce cours est souvent considéré comme le premier effort systématique et réussi d'introduire la rigueur en analyse.

Preuve par dichotomie. On définit deux suites (a_n) et (b_n) de nombres qui admettent un développement fini en base 2 (on les appelle souvent les nombres 2-imaux, et on note $\mathbb{D}_2 = \left\{z/2^k \mid z \in \mathbb{Z}, k \in \mathbb{N}\right\}$) vérifiant pour tout n :

$$b_n = a_n + 1/2^n \text{ et } f(a_n) \leq 0 < f(b_n).$$

On démarre avec $a_0 = 0$ et $b_0 = 1$. Si a_n et b_n sont définis, on considère $c_n = (a_n + b_n)/2$. Si $f(c_n) > 0$ on prend $a_{n+1} = a_n$ et $b_{n+1} = c_n$, sinon on prend $a_{n+1} = c_n$ et $b_{n+1} = b_n$. Les propriétés annoncées sont clairement vérifiées par récurrence. Les deux suites a_n et b_n ont une limite commune c. Par continuité, puisque les $f(a_n)$ sont ≤ 0 on a $f(c) \leq 0$ et puisque les $f(b_n)$ sont > 0 on a $f(c) \geq 0$. □

Commentaire. Cette preuve est beaucoup plus simple que la précédente, et fournit quelque chose qui ressemble à un algorithme pour calculer un zéro de f. Au lieu de considérer toutes les valeurs de la fonction sur l'intervalle pour extraire l'ensemble A on se contente d'examiner des valeurs successives de la fonction en des points de $\mathbb{D}_2 \cap [0,1]$. Pour calculer un zéro de f en suivant la première preuve on a besoin d'avoir une connaissance d'emblée globale de la fonction. Dans la deuxième preuve on a seulement besoin de savoir l'évaluer en des points, ce qui *a priori* est plus facile.

En outre, l'idée que la droite réelle n'a pas de trou est exprimée ici de manière plus simple : si on a une suite d'intervalles emboités $[a_n, b_n]$ telle que $b_n - a_n = 1/2^n$ il y a un *et un seul* point réel commun à tous ces intervalles. Cette idée est reprise par Cauchy dans son fameux « critère de Cauchy » qui énonce une condition suffisante pour qu'une suite (a_n) de nombres réels converge vers une limite finie mais inconnue : la valeur $|a_n - a_p|$ peut être rendue arbitrairement petite, sous la seule contrainte que n et p soient plus grands qu'un entier donné.

Notons qu'un tel critère ne peut pas être démontré par Cauchy car il ne dispose pas d'une définition en bonne et due forme des nombres réels.

C'est Cantor qui reprendra l'idée du critère de Cauchy en proposant (*grosso modo*) qu'un nombre réel puisse toujours être défini comme limite d'une suite de Cauchy de nombres rationnels. Il reste ensuite à montrer qu'une suite de Cauchy de nombres réels converge vers un nombre réel, pour boucler la boucle et *démontrer* le critère de Cauchy.

7.3 Un algorithme pour le TVI ?

La preuve n° 2 se traduit directement par l'algorithme 7.3.1.

Algorithme 7.3.1. ***Théorème des valeurs intermédiaires : algorithme par dichotomie.***
Entrée: n : entier ; f : fonction continue $[0,1] \to \mathbb{R}$.
on suppose $f(0) < 0 < f(1)$
on demande un intervalle de longueur $1/2^n$ sur lequel soit situé le zéro de f
Sortie: a, b : nombre 2-imaux ;
$[a, b]$ contient un zéro z de f, $f(a) < 0 \leq f(b)$, $b - a = 1/2^n$
N. B. : le zéro z ne dépend pas de n.
Variables locales: c : nombre 2-imal.
Début
initialisation
$a \leftarrow 0$; $b \leftarrow 1$;
boucle
Pour i **de** 1 **à** n **faire**
$c \leftarrow (a + b)/2$;
Si $f(c) > 0$ **alors** $b \leftarrow c$ **sinon** $a \leftarrow c$ **fin si** ;
fin pour ;
fin de boucle
Retourner a, b
Fin.

On a contourné ici la difficulté *a priori* posée par le problème de manipuler des nombres réels à l'intérieur d'un algorithme en ne manipulant que des nombres dans $\mathbb{D}_2$ et en ne donnant en sortie qu'une approximation, dans $\mathbb{D}_2$, du nombre réel qu'on cherche.

Comme le zéro recherché de f est en général non unique il est très important que l'approximation avec la précision $1/2^{n+1}$ qui est donnée en sortie soit toujours l'approximation du même zéro. Ceci est assuré par la structure même de l'algorithme, qui construit toujours la même suite d'intervalles emboités $[a_k, b_k]$, sans autre influence de la donnée n que celle de savoir à quel moment on s'arrête.

Une grande interrogation qui se pose à la lecture de cet algorithme c'est la manière dont la fonction f intervient en tant que fonction continue lorsqu'on évalue son signe en un nombre 2-imal.

Plus généralement on doit pouvoir se faire une idée claire des calculs que l'on peut faire pour simuler sur machine une instruction $y \leftarrow f(x)$ qui interviendrait dans un algorithme utilisant des variables x, y de type « nombre réel » et une variable f de type « fonction continue ».

Les logiciels utilisés en analyse numérique utilisent des variables représentant des « nombres flottants » qui sont des nombres 2-imaux particuliers. Quant à l'évaluation des fonctions usuelles pour les flottants, elle se fait avec des erreurs d'arrondis inévitables dans la mesure où il est bien rare que $f(x) \in \mathbb{D}_2$ lorsque $x \in \mathbb{D}_2$. En fait même l'addition de deux flottants fournit en général seulement un résultat approché dans l'arithmétique des ordinateurs.

Donc un logiciel usuel, qui manipule des « nombres flottants », ne peut pas exécuter de manière absolument sûre l'algorithme.

Mais nous voudrions discuter ici une question plus fondamentale, à savoir : *la possibilité théorique d'exécution de l'algorithme lorsqu'on se permet une précision arbitraire, mais finie, dans les calculs.*

Nous discutons dans les deux sections qui suivent ce que peut signifier exactement « calculer avec des nombres réels » et « calculer avec des fonctions continues ». C'est une manière pertinente d'aborder la question de la nature exacte des objets mathématiques correspondants.

7.4 Calculer avec les nombres réels

Les nombres réels ont été inventés par les hommes pour boucher les trous qu'ils découvraient dans la « droite numérique rationnelle », formée par les seuls points ayant une abscisse rationnelle (une fois définies une origine, une orientation et une unité de longueur). Des trous ont été clairement identifiés, comme $\sqrt{2}$: il y avait là un point géométrique incontestable (intersection de la droite avec un cercle bien défini), mais pas de point rationnel.

Quand Cauchy, Bolzano et d'autres introduisent les premiers raisonnements « rigoureux » à l'intérieur du calcul infinitésimal, ils ne songent pas encore à définir les nombres réels, mais ils cherchent à n'utiliser qu'un petit nombre de propriétés fondamentales de base correspondant à l'intuition géométrique de la droite.

Ce qui ressort de la preuve du TVI par dichotomie, c'est qu'un nombre réel doit être considéré comme bien défini s'il se trouve à l'intersection d'une suite infinie d'intervalles emboités dont la longueur est diminuée de moitié à chaque étape. Cette propriété est ici formulée en langage très moderne : point à l'intersection d'une suite infinie d'intervalles... Nous la reformulons de manière plus simple sous forme d'un adage.

Adage pour les nombres réels. *Un nombre réel est (considéré comme) connu lorsqu'on est capable d'en donner des approximations arbitrairement précises dans $\mathbb{Q}$ (ou dans $\mathbb{D}_{10}$, ou dans $\mathbb{D}_2$).*

En fait ce n'est pas la première chose qui vient en général à l'esprit. Spontanément, beaucoup de gens interrogés auraient tendance à dire : un nombre réel est (considéré comme) connu lorsqu'on est capable de donner son développement décimal illimité.

Par exemple Turing, qui a défini précisément une notion de « suite infinie calculable de nombres entiers » et qui a inventé une machine abstraite (aujourd'hui appelée Machine de Turing) pour définir ce qu'est un « calcul mécanique », a d'abord proposé de définir un nombre réel calculable comme un nombre dont la suite des décimales est calculable. Il s'est ensuite ravisé après avoir remarqué qu'il n'y a pas de méthode algorithmique générale pour calculer les décimales d'un nombre réel pourtant connu avec une précision arbitraire.

Voici le cas d'école qui montre cette impossibilité. On montre deux nombres réels ayant chacun une suite de décimales connue mais dont la somme n'est plus du même type. Un premier nombre réel est $a = 1/9$ et la suite de ses décimales est $0,11111111111\ldots$. Un deuxième nombre réel est $b = 8/9 + \varepsilon$ avec ε nul ou très petit en valeur absolue. Plus précisément la suite des décimales de b commence par $0,88888888888\ldots$, et cette suite infinie est éventuellement perturbée, une seule fois,

par l'apparition d'un 7 ou d'un 9. Peu nous importe de savoir ce qui se passe dans la boite noire qui nous fournit les décimales de b. L'adage pour les nombres réels nous dit que b est un nombre bien défini. Et il n'est pas difficile de voir que $a+b$ est également bien défini en suivant notre adage, car la précision $1/2^n$ est obtenue pour $a+b$ dès que la précision $1/2^{n+1}$ est obtenue pour a et b. Par contre nous restons dans l'incertitude concernant *la première décimale* de $a+b$ tant que la valeur exacte de b ne nous est pas révélée (et elle ne le sera peut-être jamais si aucun 7 et aucun 9 n'apparait jamais), car si un 7 apparait, la suite des décimales de $a+b$, qui est alors égal à $1-1/10^{\mathrm{N}}$ (pour un certain N) commence par $0,999999999\ldots$ tandis que dans le cas contraire $a+b$ est égal à 1 ou $1+1/10^{\mathrm{N}}$, et la suite de ses décimales commence par $1,00000000\ldots$.

En résumé, si on adoptait le mauvais adage selon lequel un nombre réel est considéré comme connu uniquement lorsqu'on sait déterminer la suite infinie de ses décimales, on ne disposerait d'aucune méthode sûre pour calculer la somme de deux nombres réels « connus » et trouver un nombre réel « connu ».

Et Turing a modifié dans une note rectificative sa définition de « nombre réel calculable » de manière à ce qu'elle soit conforme au premier adage (Turing et Girard 1995, pages 101-102, nous avons rajouté les passages entre crochets pour une meilleure compréhension du texte) :

> [...] si des nombres réels calculables doivent satisfaire à des exigences intuitives, on doit avoir :
> (A) *Soient (a_n), (b_n) deux suites calculables de nombres rationnels. Si on a, pour tout n, $a_n \leq a_{n+1} < b_{n+1} \leq b_n$ et $b_n - a_n \leq 2^{-n}$, alors il existe un nombre réel calculable α tel que, pour tout n, $a_n \leq \alpha \leq b_n$.*
> Nous pouvons donner une démonstration de cette proposition [avec la définition qu'un nombre réel est calculable si la suite de ses décimales binaires est une suite calculable], valable suivant les standards mathématiques en vigueur, mais faisant intervenir le principe du tiers exclu. La proposition suivante par contre est fausse :
> (B) *Avec les mêmes hypothèses il existe une procédure générale qui permet de donner le programme d'une machine qui calcule [les décimales binaires du] le nombre réel α à partir des règles de construction [c'est-à-dire des programmes des machines qui calculent les] des suites (a_n) et (b_n).*
> [suit une justification précise de l'impossibilité] [...] Ainsi (A) nous indique qu'il existe une machine qui calcule, par exemple, la constante d'Euler, mais nous ne pouvons pas décrire à présent une telle machine, car nous ne savons pas encore si cette constante est un nombre de la forme $m/2^n$.
> Nous pouvons éviter cette situation désagréable en modifiant la manière dont les nombres réels calculables sont [codés] [...] [suit la proposition qu'un réel calculable α soit codé par un couple $(k,\beta) \in \mathbb{Z} \times \{0,1\}^{\mathbb{N}}$ avec $\alpha = k + \sum_{r=0}^{\infty}(2\beta(r)-1)(2/3)^r$]... Dans cette définition, l'unicité [du codage] est perdue [...]

Signalons que la machine de Turing est à l'origine des ordinateurs modernes, et que seuls des détails d'ordre technique distinguent un ordinateur d'une machine de Turing.

Le lecteur est invité à imaginer une procédure de calcul pour la multiplication de deux nombres réels répondant aux spécifications du premier adage.

Nous résumons maintenant la discussion précédente sur l'impossibilité d'une procédure générale qui décide de manière certaine si un nombre réel, pourtant connu de manière irréprochable en tant que nombre réel, est inférieur, supérieur ou égal à un nombre rationnel donné :

On ne peut pas tout savoir. *Ce n'est pas parce qu'un nombre réel est connu*[1] *qu'on peut pour autant déterminer son signe de manière sûre.*

Ce principe d'ignorance et de modestie peut se formuler de manière plus concrète. En particulier, on peut établir la proposition suivante :

1. Selon l'adage des nombres réels.

Une étrange suite de nombres réels. *On sait construire une suite calculable de nombres réels $(x^{(m)})_{m\in\mathbb{N}}$, tous rationnels de la forme 0 ou $1/2^k$ ($k \geq 1$) avec la propriété suivante : la suite $(u_n)_{n\in\mathbb{N}}$ des signes des $x^{(n)}$ ($u_n = 1$ si $x^{(n)} > 0$, 0 si $x^{(n)} = 0$) n'est pas calculable.*

Une conséquence étrange est que la suite n'est pas calculable en tant que suite de nombres rationnels.

7.5 Calculer avec une fonction continue

Ainsi la discussion du théorème des valeurs intermédiaires, l'un des premiers théorèmes d'analyse réelle nous a obligé à préciser ce que signifie « connaitre un nombre réel, pouvoir calculer avec lui sur machine », et nous a conduit à notre adage pour les nombres réels.

Outre l'importance théorique fondamentale de cette question, elle est désormais à l'ordre du jour de la recherche[2], sous diverses appellations. La plus populaire s'appelle « l'arithmétique d'intervalles ». Au lieu de travailler avec des flottants, on travaille avec des intervalles dans $\mathbb{D}_2$.

Ceci signifie qu'à chaque moment, chaque nombre réel intervenant dans l'algorithme est connu comme étant sur un intervalle dont la longueur peut être rendue toujours plus petite selon les besoins du calcul. L'intervalle dans $\mathbb{D}_2$ est par exemple codé en machine sous la forme $[(k, z, z')]$ où k, z et $z' > z$ sont dans $\mathbb{Z}$ (des « entiers longs ») et l'intervalle correspondant est $[z/2^k, z'/2^k]$.

On cherche alors un adage raisonnable pour les fonctions continues sur un intervalle $[a, b]$. Cela devrait ressembler à quelque chose du genre suivant.

Adage pour les fonctions continues, première tentative. *Une fonction continue est (considérée comme) connue si on est capable de calculer sa valeur avec n'importe quelle précision prescrite en n'importe quel point de l'intervalle.*

Mais l'interprétation de cette phrase n'est pas si simple. Il ne suffit pas de savoir calculer les valeurs $f(x)$ avec une précision arbitraire pour les $x \in \mathbb{D}_2 \cap [a, b]$ mais bien pour tous les points de l'intervalle. *Et même si en théorie une fonction continue est parfaitement déterminée par les valeurs prises sur $\mathbb{D}_2 \cap [a, b]$, il n'en va pas de même en pratique.* Si vous voulez connaitre la valeur en $\sqrt{2}$ par exemple, vous devrez prendre la limite des valeurs $f(x_n)$ pour une suite x_n dans $\mathbb{D}_2$ qui converge vers $\sqrt{2}$. Mais le calcul de cette limite est impossible si vous ne connaissez pas comment (avec quelle rapidité) la suite $f(x_n)$ converge vers $f(\sqrt{2})$.

Dans la définition usuelle d'une fonction continue, on écrit :

$$\forall x \in [a, b]\ \forall \epsilon > 0\ \exists \delta > 0 \ldots$$

On peut se dispenser dans une certaine mesure des quantificateurs portant sur ϵ et δ, qui n'ont pas besoin de représenter des nombres réels arbitraires. On peut prendre $\epsilon = 1/2^n$ et $\delta = 1/2^m$ de sorte qu'on obtient :

$$\forall x \in [a, b]\ \forall n \in \mathbb{N}\ \exists m \in \mathbb{N} \ldots$$

Mais il reste que même sous cette forme, m dépend de n et de x, qui lui, est bien un nombre réel arbitraire. Ainsi pour qu'une fonction continue soit calculable sur $[a, b]$ il faut non seulement savoir calculer ses valeurs sur $\mathbb{D}_2 \cap [a, b]$ avec une précision arbitraire, mais il faut aussi avoir une information sur la manière dont m dépend de n et de x, et ceci semble hors de portée si la dépendance en x doit être précisée pour *tous* les réels de l'intervalle.

Finalement, la seule façon envisageable de réaliser l'adage ci-dessus pour les fonctions continues est d'utiliser la *continuité uniforme* de la fonction sur l'intervalle $[a, b]$. Ceci conduit à un autre adage, convenablement modifié.

2. La nécessité de calculs numériques dont le résultat est certifié sans aucune possibilité d'erreur se fait jour pour certaines applications sensibles, comme le guidage des fusées qui lancent des satellites.

Adage pour les fonctions continues sur un intervalle compact. *Une fonction continue est (considérée comme) connue si on est capable de calculer une* approximation rationnelle uniforme *de cette fonction avec une précision arbitraire.*

Par approximation rationnelle uniforme, il faut entendre précisément deux choses.

La première est que l'approximation g de la fonction f est *uniforme* avec la précision prescrite sur l'intervalle, c'est-à-dire

$$\|f-g\|_\infty \stackrel{\text{def}}{=} \sup\{|f(x)-g(x)|; x\in[a,b]\} \leq 1/2^n.$$

La deuxième est que g est *rationnelle* au sens que son calcul n'utilise que des procédures élémentaires définies au niveau des nombres rationnels. Par exemple on n'utilisera qu'un petit nombre d'affectations de base pour le programme qui calcule l'approximation g sur $\mathbb{D}_2$ (ou sur $\mathbb{Q}$).

$$\begin{array}{lcll} x & \leftarrow & c & \text{avec } c\in\mathbb{D}_2 \\ x & \leftarrow & y+z & \\ x & \leftarrow & yz & \\ x & \leftarrow & -y & \\ x & \leftarrow & \min(y,z) & \\ x & \leftarrow & \max(y,z) & \end{array}$$

Si on remplace l'instruction $x \leftarrow yz$ par l'instruction $x \leftarrow y/2$, on obtiendra des fonctions affines par morceaux, rationnelles, qui suffisent pour approcher toutes les fonctions continues. On pourrait aussi supprimer les deux dernières affectations et on obtiendrait les fonctions polynômes à coefficients dans $\mathbb{D}_2$. Le théorème d'approximation de Weierstrass dit qu'elles suffisent pour approcher toutes les fonctions continues sur un intervalle fermé borné.

À propos de la continuité uniforme et de l'évaluation

Lorsqu'une fonction f est connue sur un intervalle $[a,b]$ conformément à l'adage des fonctions continues, on obtient de manière explicite son caractère uniformément continu.

Supposons en effet que, à partir d'un entier donné n, nous cherchions un entier m tel que l'on ait l'implication :

$$\forall x,x'\in[a,b]\ (|x-x'|\leq 1/2^m \Rightarrow |f(x)-f(x')|\leq 1/2^n).$$

Nous commençons par considérer une approximation rationnelle g de f avec la précision $1/2^{n+2}$: $\|f-g\|_\infty \leq 1/2^{n+2}$. L'approximation rationnelle g est, selon son expression explicite, lipschitzienne. Autrement dit on peut déterminer un entier k tel que :

$$\forall x,x'\in[a,b]\ |g(x)-g(x')|\leq 2^k|x-x'|.$$

Alors on obtient pour f :

$$\forall x,x'\in[a,b]\ (|x-x'|\leq 1/2^{n+k+1} \Rightarrow |f(x)-f(x')|\leq 1/2^n).$$

En effet si $|x-x'|\leq 1/2^{n+k+1}$, alors $|g(x)-g(x')|\leq 1/2^{n+1}$, et puisque $|g(x)-f(x)|\leq 1/2^{n+2}$ et $|g(x')-f(x')|\leq 1/2^{n+2}$ on obtient $|f(x)-f(x')|\leq 1/2^{n+1}+1/2^{n+2}+1/2^{n+2}=1/2^n$.

Ceci permet aussi d'évaluer la fonction f en un réel arbitraire x de l'intervalle $[a,b]$, comme suit. Le réel x est supposé connu conformément à l'adage des nombres réels. On a donc pour chaque n un élément $x_n\in\mathbb{D}_2\cap[a,b]$ tel que $x\in[x_n-1/2^n, x_n+1/2^n]$. On utilise une approximation rationnelle g de f avec la précision $1/2^{n+2}$, on détermine un entier k tel que : $\forall x,x'\in[a,b]\ |g(x)-g(x')|\leq 2^k|x-x'|$. On utilise alors une approximation rationnelle x_{n+k+1} de x avec la précision $1/2^{n+k+1}$, et on dispose alors explicitement de $g(x_{n+k+1})$ qui vérifie $|g(x_{n+k+1})-f(x)|\leq 1/2^n$.

On peut résumer ceci comme suit :

On peut évaluer une fonction continue connue en un nombre réel connu. *Si une fonction continue $f\colon [a,b] \to \mathbb{R}$ est connue selon l'adage des fonctions continues et $x \in [a,b]$ est connu selon l'adage des nombres réels, alors $f(x)$ est connu selon l'adage des nombres réels.*

Autrement dit la première tentative de formulation de l'adage pour les fonctions continues est satisfaite avec la deuxième formulation, plus précise.

7.6 Ne pas renoncer au théorème des valeurs intermédiaires

Il est à peu près clair que le TVI ne peut pas être traduit en un algorithme si les nombres réels qui y apparaissent sont conformes à l'adage page 101. En quelque sorte la vérité énoncée par le TVI est purement idéale, et ne peut pas être explicitée en toute généralité. Cela résulte du principe *on ne peut pas tout savoir*.

Il existe des formulations mathématiques précises pour cette impossibilité, mais nous sommes plutôt intéressés par les possibilités que les impossibilités. Et c'est cela que nous voudrions examiner dans cette section, même de manière succincte.

Dans toute cette section on suppose que la fonction est connue conformément à l'adage des fonctions continues. On peut par exemple imaginer que les approximations rationnelles de la fonction sont fournies par une boite noire dont on ignore totalement ce qui se passe à l'intérieur.

Tout d'abord nous commençons par donner un énoncé qui dit que le calcul d'un zéro de f est possible sous certaines conditions, et qui d'une certaine manière se rapproche de l'énoncé du théorème 7.2.1.

Théorème 7.6.1. *Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue avec $f(0) < 0 < f(1)$. Supposons que pour tout $x \in \mathbb{D}_2 \cap [0,1]$ on ait $f(x) \neq 0$. Alors on peut calculer un $x \in [0,1]$ tel que $f(x) = 0$.*

Cela tient à ce que l'algorithme 7.3.1 va pouvoir être exécuté : à partir du moment où un réel (le réel $f(c)$ pour un $c \in \mathbb{D}_2$ dans l'algorithme) peut être connu avec une précision arbitraire, s'il est non nul, on finira par connaitre son signe un jour. Naturellement, le temps d'exécution de l'algorithme est *a priori* impossible à majorer, en l'absence d'informations plus précises sur le fait que $f(x) \neq 0$ pour $x \in \mathbb{D}_2$.

Pour écrire un algorithme correspondant à ce théorème (algorithme 7.6.2), nous utilisons un nouveau type d'instruction, où une variable a dans $\mathbb{D}_2$ est affectée d'une valeur approchée d'un nombre réel y supposé connu selon l'adage des nombres réels. C'est l'affectation suivante :

$$a \leftarrow y \pm 1/2^n$$

qui signifie que $y \in [a-1/2^n, a+1/2^n]$. En fait on peut toujours réaliser ceci avec un a de la forme $k/2^n$ et $k \in \mathbb{Z}$, il faut pour cela situer d'abord y sur un intervalle défini dans $\mathbb{D}_2$ et de longueur $< 1/2^{n+1}$.

Algorithme 7.6.2. ***Théorème des valeurs intermédiaires : véritable algorithme par dichotomie, sous hypothèses restrictives.***
Entrée: n : entier ; f : fonction continue $[0,1] \to \mathbb{R}$.
f est connue selon l'adage des fonctions continues,
on suppose que $f(x) \neq 0$ pour tout $x \in \mathbb{D}_2 \cap [0,1]$, et que $f(0) < 0 < f(1)$.
on demande un intervalle de longueur $1/2^n$ sur lequel soit situé le zéro de f
Sortie: a, b : nombre 2-imaux ;

```
    # [a, b] contient un zéro z de f, f(a) < 0 < f(b), b − a = 1/2^n
    # N. B. : le zéro z ne dépend pas de n.
Variables locales: c, y : nombres 2-imaux, m entier.
Début
            # initialisation
  a ← 0 ; b ← 1 ; c ← 1/2 ;
            # boucle
  Pour i de 1 à n faire
    m ← 0 ; c ← (a + b)/2 ;
    Répéter
      m ← m + 1 ; y ← f(c) ± 1/2^m
    jusqu'à ce que |y| > 1/2^m ;
    Si y > 0 alors b ← c sinon a ← c fin si ;
  fin pour ;
            # fin de boucle
  Retourner a, b
Fin.
```

Comme conséquences du théorème 7.6.1 on a les deux corollaires suivants.

Corollaire 7.6.3. *Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue avec $f(0) < 0 < f(1)$. Il est absurde de supposer que $f(x)$ soit clairement $\neq 0$ pour tous les $x \in [0,1]$.*

Le corollaire précédent suffit, pour la majorité des mathématiciens, à assurer l'existence (idéale) d'un zéro de f. Cependant il n'affirme rien en positif. Le corollaire qui suit est nettement plus intéressant.

Corollaire 7.6.4. *Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue. Il existe une suite infinie $(y_n)_{n\in\mathbb{N}}$ dans $\mathbb{R}$ telle que pour tout y dans $[f(0), f(1)]$ distinct de tous les y_n il existe un $x \in [0,1]$ avec $f(x) = y$.*

Démonstration. Il suffit d'énumérer les éléments de $\mathbb{D}_2 \cap [0,1]$ en une suite $(x_n)_{n\in\mathbb{N}}$ et de définir $y_n = f(x_n)$ pour tout n. □

L'intérêt du corollaire précédent est donné par le théorème de Cantor, qui affirme qu'on peut trouver, sur n'importe quel intervalle ouvert non vide, *beaucoup* de nombres réels distincts de tous les termes d'une suite donnée de nombre réels (comme la suite $(y_n)_{n\in\mathbb{N}}$ décrite dans la preuve précédente).

Venons en à une version susceptible d'intéresser les physiciens et les numériciens dans la plupart des applications. En général, ce qui importe ce n'est pas que $f(c)$ soit exactement égal à 0, mais qu'il soit très petit en valeur absolue. De ce point de vue le premier algorithme s'avère parfois complètement inefficace, si la fonction subit des variations très brusques, c'est-à-dire si son graphe comporte des morceaux de courbe presque verticaux. On peut en effet alors trouver un c très proche d'un zéro de f sans que la valeur de f au point c soit petite.

Supposons qu'on veuille trouver un c tel que $|f(c)| < 1/2^n$. Si on a une approximation rationnelle g_n de la fonction f avec la précision $1/2^{n+1}$ il suffit de réaliser $|g_n(c)| < 1/2^{n+1}$. D'après le programme qui calcule g_n on peut déterminer un *module de Lipschitz* pour cette fonction : un entier k tel que $|g_n(x) - g_n(x')| < 2^k|x - x'|$ pour tous x, x' sur l'intervalle. Si on subdivise l'intervalle en petits segments de longueur $1/2^{n+1+k}$ l'écart entre les valeurs de g_n sur chaque segment ne dépasse jamais $1/2^{n+1}$. Donc sur un petit segment aux extrémités duquel g_n change de signe, on aura $|g_n(x)| \leq 1/2^{n+1}$. Nous avons démontré le théorème suivant.

Théorème 7.6.5. (théorème des valeurs intermédiaires approximatif)
Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue avec $f(0) < 0 < f(1)$ et $n \in \mathbb{N}$. Alors il existe un $x \in [0,1] \cap \mathbb{D}_2$ tel que $|f(x)| < 1/2^n$.

Un algorithme correspondant à ce théorème, et qui n'utilise que l'évaluation approchée de f en des points de $[0,1] \cap \mathbb{D}_2$, est donné dans l'encadré 7.6.6.

Algorithme 7.6.6. ***Algorithme pour le théorème des valeurs intermédiaires approximatif, par dichotomie.***
Entrée: n : entier ; # précision désirée sur le résultat
f : fonction continue $[0,1] \to \mathbb{R}$. # avec $f(0) < 0 < f(1)$
Sortie: c : nombre 2-imal. # avec $|f(c)| < 1/2^n$.
N. B. : le pseudo-zéro c peut sauter d'un point à un point très éloigné si n augmente d'une unité.
Variables locales: a, b, y : nombres 2-imaux ;
Début
initialisation
$a \leftarrow 0$; $b \leftarrow 1$;
boucle
Répéter
$c \leftarrow (a+b)/2$; $y \leftarrow f(c) \pm 1/2^{n+1}$;
Si $y > 1/2^{n+1}$ **alors** $b \leftarrow c$ **sinon si** $y < -1/2^{n+1}$ **alors** $a \leftarrow c$
fin si ;
jusqu'à ce que $|y| \leq 1/2^{n+1}$;
fin de boucle
Retourner c
Fin.

La preuve de terminaison de cet algorithme, pour l'entrée n, repose sur l'existence d'une approximation rationnelle de f avec la précision $1/2^{n+1}$, à partir de laquelle on peut majorer le nombre d'étapes.

Enfin, nous terminons avec un théorème « rassurant ».

Nous disons qu'une fonction $f\colon [a,b] \to \mathbb{R}$ est *localement non constante*, si pour tout intervalle $[c,d] \subset [a,b]$ il existe deux points x, x' de $[c,d]$ tels que $f(x) \neq f(x')$. C'est par exemple le cas d'une fonction polynôme non constante ou plus généralement d'une fonction continument dérivable dont la dérivée n'admet qu'un nombre fini de zéros sur l'intervalle $[a,b]$ (par exemple une fonction analytique dans un intervalle ouvert contenant $[a,b]$, et non constante).

Théorème 7.6.7. *Soit $f\colon [0,1] \to \mathbb{R}$ une fonction continue localement non constante avec $f(0) < 0 < f(1)$. Alors il existe un $x \in [0,1]$ tel que $f(x) = 0$ et on peut en calculer un (au sens de l'adage des nombres réels et de celui des fonctions continues).*

Pour prouver ce théorème, l'algorithme du théorème des valeurs intermédiaires par dichotomie doit être modifié (cela donne l'algorithme 7.6.8 page suivante) en ne prenant pas c exactement au milieu de l'intervalle $[a,b]$ défini lors de la boucle précédente mais sur un petit intervalle situé au milieu de $[a,b]$.

Notez que l'instruction « trouver $c \in \mathbb{D}_2$ sur l'intervalle $[a + 3d/8, b - 3d/8]$ tel que $f(c) \neq 0$ » réalise le certificat que la fonction est localement non constante : cette procédure permet de fournir, sur un intervalle spécifié $\mathrm{I} = [u,v]$, deux réels x, x' tels que $f(x) \neq f(x')$. Ces réels sont

en fait connus *via* des approximations rationnelles x_m et x'_m dans $\mathbb{D}_2$, suffisantes pour décider $f(x) \neq f(x')$. Ici l'intervalle est $\mathrm{I} = [a + 3d/8, b - 3d/8]$ et on prend $c = x_m$ ou x'_m de façon à ce que $f(c) \neq 0$.

Algorithme 7.6.8. ***Algorithme pour le théorème des valeurs intermédiaires, par dichotomie, cas des fonctions localement non constantes.***

Entrée: n : entier ; # précision désirée sur le résultat
f : fonction continue $[0,1] \to \mathbb{R}$. # localement non constante, avec $f(0) < 0 < f(1)$
Sortie: a, b : nombre 2-imaux ;
$[a,b]$ contient un zéro z de f, $f(a) < 0 < f(b)$, $b - a \leq 1/2^n$
N. B. : le zéro z ne dépend pas de n.
Variables locales: c, d : nombres 2-imaux ;
Début
initialisation
$a \leftarrow 0$; $b \leftarrow 1$; $d \leftarrow 1$;
boucle
Répéter
trouver $c \in \mathbb{D}_2$ sur l'intervalle $[a + 3d/8, b - 3d/8]$ tel que $f(c) \neq 0$;
Si $f(c) > 0$ **alors** $b \leftarrow c$ **sinon** $a \leftarrow c$ **fin si** ;
$d \leftarrow b - a$
jusqu'à ce que $d \leq 1/2^n$;
fin de boucle
Retourner a, b
Fin.

Chapitre 8

La structure du continu

Introduction

Le continu physique est un grand mystère.

Le continu mathématique est une invention humaine qui essaie de donner un cadre mathématique correspondant à notre intuition du continu physique.

Cette invention est le fruit d'un processus très long qui voit son aboutissement à la fin du 19^{e} siècle avec la *définition de l'ensemble des réels*, *via* les coupures de Dedekind ou *via* les suites de Cauchy de nombres rationnels.

Lorsque nous avons donné notre adage pour les nombres réels, nous n'avons fait que donner sous forme intuitive une description de l'invention des nombres réels *via* les suites de Cauchy de nombres rationnels. Notre formulation est dans le langage des mathématiques constructives dans la mesure où nous parlons des réels « connus », mais elle est également acceptable en mathématiques classiques où le mot « connu » est simplement pris au sens d'une existence abstraite. Cela se reformule alors comme suit :

> *Un nombre réel est défini par une suite d'approximations rationnelles* $(x_n)_{n\in\mathbb{N}}$ *avec la précision* $1/2^n$.

Notons que si on veut vraiment en faire une définition *stricto sensu*, on doit rajouter :

> *Cette suite doit vérifier la condition de cohérence donnée par le critère de Cauchy : ici* $|x_n - x_m| \leq 1/2^n + 1/2^m$ *pour tous* m, n.

Dans la section 8.1 nous donnons un court extrait du chapitre II de *La science et l'hypothèse* de H. Poincaré, chapitre dans lequel il discute la question du continu.

Dans les sous-sections suivantes, nous présentons un embryon d'étude sur la structure de la droite réelle définie en mathématiques. Naturellement, nous devons garder à l'esprit que cela ne nous éclaire pas vraiment sur la nature du continu physique, comme le dit Poincaré. Voir aussi à ce sujet notre discussion sur les paradoxes de Zénon dans la section 5.1.

8.1 Qu'est-ce que le continu ?

Extrait du chapitre II de *La science et l'hypothèse*

La grandeur mathématique et l'expérience.

Si l'on veut savoir ce que les mathématiciens entendent par un continu, ce n'est pas à la géométrie qu'il faut le demander. Le géomètre cherche toujours plus ou moins à se représenter les figures qu'il

étudie, mais ses représentations ne sont pour lui que des instruments ; il fait de la géométrie avec de l'étendue comme il en fait avec de la craie ; aussi doit-on prendre garde d'attacher trop d'importance à des accidents qui n'en ont souvent pas plus que la blancheur de la craie.

L'analyste pur n'a pas à craindre cet écueil. Il a dégagé la science mathématique de tous les éléments étrangers, et il peut répondre à notre question : Qu'est-ce au juste que ce continu sur lequel les mathématiciens raisonnent ? Beaucoup d'entre eux, qui savent réfléchir sur leur art, l'ont fait déjà ; M. Tannery, par exemple, dans son *Introduction à la théorie des Fonctions d'une variable.*

Partons de l'échelle des nombres entiers ; entre deux échelons consécutifs, intercalons un ou plusieurs échelons intermédiaires, puis entre ces échelons nouveaux d'autres encore, et ainsi de suite indéfiniment. Nous aurons ainsi un nombre illimité de termes, ce seront les nombres que l'on appelle fractionnaires, rationnels ou commensurables. Mais ce n'est pas assez encore ; entre ces termes qui sont pourtant déjà en nombre infini, il faut encore en intercaler d'autres, que l'on appelle irrationnels ou incommensurables.

Avant d'aller plus loin, faisons une première remarque. Le continu ainsi conçu n'est plus qu'une collection d'individus rangés dans un certain ordre, en nombre infini, il est vrai, mais *extérieurs* les uns aux autres. Ce n'est pas là la conception ordinaire, où l'on suppose entre les éléments du continu une sorte de lien intime qui en fait un tout, où le point ne préexiste pas à la ligne, mais la ligne au point. De la célèbre formule, le continu est l'unité dans la multiplicité, la multiplicité seule subsiste, l'unité a disparu. Les analystes n'en ont pas moins raison de définir leur continu comme ils le font, puisque c'est toujours sur celui-là qu'ils raisonnent depuis qu'ils se piquent de rigueur. Mais c'est assez pour nous avertir que le véritable continu mathématique est tout autre chose que celui des physiciens et celui des métaphysiciens.

Nous gardons à l'esprit que le continu mathématique, conçu comme un simple ensemble de points, présente *a priori* un grave défaut, celui de ne pas représenter fidèlement ce qu'il est censé représenter, et nous tentons une petite exploration de la structure de cet ensemble.

8.2 Le théorème de Cantor

Le théorème de Cantor affirme qu'une suite de nombres réels ne peut pas recouvrir la droite réelle. Il nous avertit que la structure de la droite réelle est intrinsèquement plus complexe que celle de l'ensemble des entiers naturels.

Cela semble bien banal aujourd'hui, mais cela n'est en fait pas du tout évident et ce fut une grande découverte. Par exemple on peut organiser l'ensemble de tous les nombres rationnels en une suite $(r_n)_{n\in\mathbb{N}}$. Naturellement on trouve de nombreux réels non rationnels, comme $\sqrt{2}$, ce qui montre que la suite des rationnels ne recouvre pas la droite réelle : il y a plein de trous entre les rationnels. Mais chaque fois qu'on trouve un trou, on peut le remplir. Il n'est donc nullement clair *a priori* qu'on ne puisse pas remplir tous les trous au moyen d'une suite de réels.

Une formulation précise du théorème de Cantor est la suivante.

Théorème 8.2.1. *Si $(u_n)_{n\in\mathbb{N}}$ est une suite de nombres réels connus, et si $[a,b]$ est un intervalle rationnel (avec $b>a$) on peut construire un $x\in[a,b]$ tel que pour tout n, $|x-u_n|>|b-a|/3^{n+2}$.*

Démonstration. Cette preuve est délicate. On recommande vivement à la lectrice de faire des petits dessins pour « voir » les intervalles successifs qui interviennent dans la preuve, avec les différents cas de figure possibles.

Posons $[a,b]=[a_1,b_1]$ et $\ell=b-a$. Nous sommes à l'étape 1. Puisque le réel u_0 est connu, on connait un intervalle rationnel $]e_1,f_1[$ de longueur $\ell/9$ sur lequel est situé u_0. On pose $c_1=e_1-\ell/9$, $d_1=f_1+\ell/9$ de sorte que $]c_1,d_1[$ est un intervalle rationnel de longueur $\ell/3$ et tout élément x de $[a_1,b_1]\setminus]c_1,d_1[$ vérifie $|x-u_0|\geq\ell/9$. L'ensemble $[a_1,b_1]\setminus]c_1,d_1[$ est formé de un ou deux intervalles rationnels, dont la longueur totale est $\geq 2\ell/3$ (il est recommandé de faire un dessin

pour voir les différents cas de figures). Il contient donc un intervalle rationnel $[a_2, b_2]$ de longueur $\ell/3$. Tout élément x de $[a_2, b_2]$ vérifie $|x - u_0| \geq \ell/9$.

Il reste à itérer le processus. Nous précisons l'hypothèse de récurrence comme suit.

Au début de l'étape n on dispose d'un intervalle $[a_n, b_n]$ de longueur $\ell/3^{n-1}$, avec la propriété que tout élément $x \in [a_n, b_n]$ vérifie $|x - u_k| \geq \ell/3^{k+2}$ pour $k = 0, \ldots, n-2$.

On connait un intervalle rationnel $]e_n, f_n[$ de longueur $\ell/3^{n+1}$ sur lequel est situé u_{n-1}. On pose $c_n = e_n - \ell/3^{n+1}$, $d_n = f_n + \ell/3^{n+1}$ de sorte que $]c_n, d_n[$ est un intervalle rationnel de longueur $\ell/3^n$ et tout élément x de $[a_n, b_n] \setminus]c_n, d_n[$ vérifie $|x - u_{n-1}| \geq \ell/3^{n+1}$. L'ensemble $[a_n, b_n] \setminus]c_n, d_n[$ est formé de un ou deux intervalles rationnels, dont la longueur totale est $\geq 2\ell/3^n$. Il contient donc un intervalle rationnel $[a_{n+1}, b_{n+1}]$ de longueur $\ell/3^n$. Et tout élément $x \in [a_{n+1}, b_{n+1}]$ vérifie $|x - u_k| \geq \ell/3^{k+2}$ pour $k = 0, \ldots, n-1$. La récurrence fonctionne bien.

Enfin la limite commune des suites (a_n) et (b_n) est un réel x qui appartient à chacun des intervalles $[a_k, b_k]$ parce qu'il s'agit d'une suite d'intervalles emboités. En conséquence on a $|x - u_k| \geq \ell/3^{k+2}$ pour tout k, ce qui était le but recherché. □

Dans la variante suivante du théorème de Cantor, on montre que non seulement une suite de nombres réels laisse des trous, mais qu'elle en laisse beaucoup.

Nous devons pour cela introduire l'*espace de Cantor*. C'est l'ensemble des suites $(\alpha_n)_{n\in\mathbb{N}}$ dont tous les termes sont égaux à 0 ou 1. C'est une *espace métrique* si on définit la distance entre deux suites distinctes $\alpha = (\alpha_n)_{n\in\mathbb{N}}$ et $\beta = (\beta_n)_{n\in\mathbb{N}}$ comme égale à $d(\alpha, \beta) = 1/2^k$, où k est le premier indice pour lequel $u_k \neq v_k$. On note cet espace $\mathcal{F}(\mathbb{N}, \{0, 1\})$ ou encore $\{0, 1\}^{\mathbb{N}}$.

L'application $\theta_3 \colon \{0, 1\}^{\mathbb{N}} \to [0, 1] \subset \mathbb{R}$ définie par $\theta_3(\alpha) = \sum_{n=0}^{\infty} 2\alpha_n/3^{n+1}$ a pour image une partie $\mathbb{K}$ de $[0, 1]$ qui est appelée « l'ensemble de Cantor ».

On peut obtenir $\mathbb{K}$ à partir de l'intervalle $[0, 1]$ en enlevant le tiers central, ce qui laisse $[0, 1/3] \cup [2/3, 1]$, puis en recommençant le processus avec chacun des deux intervalles restants. En poursuivant ainsi indéfiniment on obtient, comme intersection de tous les ensembles successifs, l'ensemble de Cantor $\mathbb{K}$. On vérifie que $d(\alpha, \beta) = 1/2^k$ implique $1/3^{k+1} \leq |\theta_3(\alpha) - \theta_3(\beta)| \leq 1/3^k$. Donc l'application $\theta_3 \colon \{0, 1\}^{\mathbb{N}} \to \mathbb{K}$ est une bijection uniformément continue dans les deux sens. Ainsi $\mathbb{K}$ peut être considéré comme une « photocopie » de l'espace de Cantor.

La variante du théorème de Cantor affirme qu'étant donnée une suite $(u_n)_{n\in\mathbb{N}}$ de réels dans l'intervalle $[a, b]$ on peut insérer dans $[a, b]$ une « photocopie » de l'espace de Cantor qui évite tous les termes de la suite.

Variante. *Si $(u_n)_{n\in\mathbb{N}}$ est une suite de nombres réels connus, et si $[a, b]$ est un intervalle rationnel (avec $b > a$), on peut construire pour chaque suite infinie $\alpha \in \{0, 1\}^{\mathbb{N}}$ un réel x_α de telle façon que :*

— *pour $\alpha \neq \beta$ on a $|x_\alpha - x_\beta| > 0$, et plus précisément si $d(\alpha, \beta) = 1/2^k$ on a :*

$$|b - a|/7^{k+1} \leq |x_\alpha - x_\beta| \leq |b - a|/7^k.$$

— *pour tout α et tout n, $|x_\alpha - u_n| \geq |b - a|/7^{n+1}$.*

Notez que l'on a ainsi défini une bijection $\alpha \mapsto x_\alpha$ de l'espace de Cantor sur son image qui est uniformément continue dans les deux sens, ce qui justifie le terme imagé de photocopie que nous avons utilisé auparavant.

Démonstration. Nous allons indexer certains intervalles rationnels contenus dans $[a, b]$ par des listes finies de 0 ou 1. Au départ, on indexe avec la liste vide. On pose donc $[a, b] = [a_{[]}, b_{[]}]$ et $\ell = b - a$.

Nous sommes à l'étape 1. Puisque le réel u_0 est connu, on connait un intervalle rationnel $]e_{[]}, f_{[]}[$ de longueur $\ell/7$ sur lequel est situé u_0. On pose $c_{[]} = e_{[]} - \ell/7$, $d_{[]} = f_{[]} + \ell/7$ de sorte que $]c_{[]}, d_{[]}[$ est un intervalle rationnel de longueur $3\ell/7$ et tout élément x de $[a_{[]}, b_{[]}] \setminus]c_{[]}, d_{[]}[$ vérifie $|x - u_0| \geq \ell/7$. L'ensemble $[a_{[]}, b_{[]}] \setminus]c_{[]}, d_{[]}[$ est formé de un ou deux intervalles rationnels, dont

la longueur totale est $\geq 4\ell/7$. Il contient donc deux intervalles $[a_{[0]}, b_{[0]}]$ et $[a_{[1]}, b_{[1]}]$ de longueurs $\ell/7$ avec $a_{[0]} < b_{[0]} < a_{[1]} < b_{[1]}$ et $b_{[0]} + \ell/7 \leq a_{[1]}$. Ainsi si $x \in [a_{[0]}, b_{[0]}]$ et $y \in [a_{[1]}, b_{[1]}]$, les trois distances $|x - u_0|$, $|y - u_0|$ et $|x - y|$ sont $\geq \ell/7$.

Il reste à itérer le processus.

Au début de l'étape n on dispose pour chaque liste finie γ de longueur $n-1$ formée de 0 et de 1 d'un intervalle $[a_\gamma, b_\gamma]$ de longueur $\ell/7^{n-1}$, avec la propriété que :

— tout élément $x \in [a_\gamma, b_\gamma]$ vérifie $|x - u_k| \geq \ell/7^{k+1}$ pour $k = 0, \ldots, n-2$.

— si $x \in [a_\gamma, b_\gamma]$, $y \in [a_\delta, b_\delta]$ et $\gamma(k) \neq \delta(k)$ alors $|x - y| \geq |b - a|/7^{k+1}$.

Chacun des 2^{n-1} intervalles $[a_\gamma, b_\gamma]$ est alors traité avec u_{n-1} comme l'a été l'intervalle $[a_{[\,]}, b_{[\,]}]$ avec u_0. Ceci donne lieu à deux intervalles contenus dans $[a_\gamma, b_\gamma]$, indexés respectivement par la liste γ prolongée par 0 et la liste γ prolongée par 1. Les vérifications que la récurrence se passe bien sont laissées au lecteur.

Enfin pour chaque $\alpha \in \mathcal{F}(\mathbb{N}, \{0,1\})$ on prend pour x_α la limite des $x_{[\alpha(1),\ldots,\alpha(n)]}$ lorsque n tend vers l'infini. □

8.3 Mesurer

Nous cherchons maintenant à comprendre et généraliser le théorème de Cantor dans le cadre de la théorie de la mesure.

Dans ce cadre le théorème de Cantor se comprend comme suit : une partie dénombrable A de $\mathbb{R}$ est de mesure nulle, « donc » on peut trouver des réels n'appartenant pas à A dans tout ensemble de mesure strictement positive. Mais pour justifier ce « donc » il faut développer une théorie de la mesure.

Pour ne pas entrer dans les arcanes de la théorie, nous allons considérer la situation relativement simple suivante, dans laquelle les questions de mesure sont présentes mais réduites au minimum.

On note $|\mathrm{I}|$ la longueur d'un intervalle rationnel borné I. On étend cette notation à la situation où L est une réunion finie d'intervalles rationnels bornés (ouverts, fermés ou semi-ouverts) : $|\mathrm{L}|$ est la mesure de L. Ceci est clairement défini si les intervalles sont rationnels, car alors L se réécrit comme réunion finie d'intervalles rationnels disjoints et la mesure est la somme des longueurs de ces intervalles disjoints.

En outre lorsque L et L′ sont du type ci-dessus, il en va de même pour $\mathrm{L} \cap \mathrm{L}'$, $\mathrm{L} \cup \mathrm{L}'$ et $\mathrm{L} \setminus \mathrm{L}'$, avec les égalités attendues concernant leurs mesures :

— $|\mathrm{L}| + |\mathrm{L}'| = |\mathrm{L} \cap \mathrm{L}'| + |\mathrm{L} \cup \mathrm{L}'|$ et

— $|\mathrm{L}| = |\mathrm{L} \cap \mathrm{L}'| + |\mathrm{L} \setminus \mathrm{L}'|$.

La théorie de la mesure sur $\mathbb{R}$ peut être considérée comme la recherche d'une manière raisonnable d'étendre la définition de la « mesure de longueur » $|\mathrm{L}|$ ci-dessus à une classe très large de parties de $\mathbb{R}$.

On considère maintenant la situation suivante.

Hypothèse. *1. On fixe un intervalle rationnel fermé borné* I. *Soit par ailleurs* $(\mathrm{J}_n)_{n\in\mathbb{N}}$ *une suite d'intervalles rationnels* ouverts *bornés. Pour chaque* n *on considère* $\mathrm{V}_n = \mathrm{J}_0 \cup \mathrm{J}_1 \cup \cdots \cup \mathrm{J}_n$. *Chaque* V_n *est un ouvert très simple : une réunion finie d'intervalles rationnels ouverts bornés. De même, chaque* $\mathrm{I} - (\mathrm{I} \cap \mathrm{V}_n)$ *est un fermé borné très simple : une réunion finie d'intervalles rationnels fermés bornés. En outre la suite* $\mathrm{I} - (\mathrm{I} \cap \mathrm{V}_n)$ *est décroissante.*

2. On suppose que la limite de la suite de nombres réels $(|\mathrm{V}_n \cap \mathrm{I}|)_{n\in\mathbb{N}}$ *existe. En mathématiques classiques c'est toujours vrai parce qu'il s'agit d'une suite croissante, majorée par* $|\mathrm{I}|$. *En mathématiques constructives cela signifie que l'on connait la vitesse de convergence de la suite : pour chaque entier* m *on connait un entier* N *tel que, pour tout* $\mathrm{N}' > \mathrm{N}$ *on a :* $0 \leq |\mathrm{V}_{\mathrm{N}'} \cap \mathrm{I}| - |\mathrm{V}_{\mathrm{N}} \cap \mathrm{I}| \leq 1/2^m$.

3. *On pose* $\ell = \lim_n |\mathrm{V}_n \cap \mathrm{I}|$, $\mathrm{V} = \bigcup_n \mathrm{V}_n$, *on a clairement* $\ell \leq |\mathrm{I}|$.
4. *On suppose* $\ell < |\mathrm{I}|$.

On veut démontrer le résultat suivant.

Conclusion. *Il existe* $x \in \mathrm{I}$ *tel que* $d(x, \mathrm{J}_n) > 0$ *pour tout* n.

La signification est la suivante : puisque la suite $\mathrm{I} - (\mathrm{I} \cap \mathrm{V}_n)$ est décroissante, l'ensemble compact $\mathrm{L} = \bigcap_n (\mathrm{I} - (\mathrm{I} \cap \mathrm{V}_n)) = \mathrm{I} - \mathrm{V}$ « doit » avoir pour mesure la limite des mesures des $\mathrm{I} - (\mathrm{I} \cap \mathrm{V}_n)$, c'est-à-dire $|\mathrm{I}| - \ell$. Si cette mesure est > 0 il faut que l'ensemble soit non vide. Autrement dit, démontrer la conclusion est une condition impérative pour qu'on puisse étendre valablement la notion de mesure à des parties telles que V ou L.

D'autre part, ce résultat généralise le théorème de Cantor puisqu'une suite de réels peut être couverte par une réunion $\mathrm{J} = \bigcup_n \mathrm{J}_n$ d'intervalles ouverts de mesure arbitrairement petite (il suffit de prendre $|\mathrm{J}_n| = \epsilon/2^{n+1}$ pour avoir $|\mathrm{J}| = \epsilon$).

Démonstration. Il suffit en fait de démontrer :

Conclusion *bis*. *Il existe* $x \in \mathrm{I}$ *tel que pour tout* $y \in \mathrm{V}$ *on a* $|y - x| > 0$.

En effet le deuxième énoncé est *a priori* un peu moins fort. Mais s'il est démontré, on peut remplacer chaque J_n par un J'_n qui déborde à droite et à gauche, tout en modifiant suffisamment peu les $|\mathrm{V}_n|$ pour que l'on ait encore $\ell < |\mathrm{I}|$ (on utilise la technique usuelle en théorie de la mesure : $\epsilon = \sum_{n=0}^{\infty} \epsilon/2^{n+1}$). Enfin on remarque que si $x \notin \bigcup_n \mathrm{J}'_n$ alors $d(x, \mathrm{J}_m) > 0$ pour tout m. □

Voici maintenant une preuve du deuxième énoncé en mathématiques classiques.

Démonstration en mathématiques classiques. Si la conclusion était fausse, quitte à remplacer chaque J_n par un J'_n qui déborde à droite et à gauche, tout en modifiant suffisamment peu les $|\mathrm{V}_n|$ pour que l'on ait encore $\ell < |\mathrm{I}|$ on peut supposer que les J_n recouvrent l'intervalle fermé borné I. Alors, d'après le théorème de Heine-Borel, on peut extraire un recouvrement fini (du recouvrement par les J_n). Donc un $\mathrm{V_N}$ recouvre I : on a $\mathrm{V} \cap \mathrm{I} = \mathrm{V_N} \cap \mathrm{I} = \mathrm{I}$, $|\mathrm{I}| \geq \ell \geq |\mathrm{V_N} \cap \mathrm{I}| = |\mathrm{I}|$, donc $\ell = |\mathrm{I}|$, ce qui est absurde. □

La preuve précédente est élégante, mais elle ne fournit pas explicitement le réel x annoncé dans l'énoncé. Cela tient au caractère non explicite du théorème de Heine-Borel, au moins dans son énoncé le plus général (de tout recouvrement ouvert d'un fermé borné de $\mathbb{R}$ on peut extraire un recouvrement fini).

Nous allons maintenant donner une preuve[1] du deuxième énoncé qui fournit un réel x de façon explicite.

Démonstration constructive. On construit une suite décroissante d'intervalles I_m fermés bornés par un procédé de dichotomie. On démarre avec $\mathrm{I}_1 = \mathrm{I}$. Au début de l'étape m, on dispose d'un intervalle I_m de longueur $u_m = |\mathrm{I}|/2^{m-1}$ tel que $\ell_m = \lim_{p\to\infty} |\mathrm{V}_p \cap \mathrm{I}_m|$ existe et $\ell_m < u_m$. Soit $\delta > 0$ tel que $\ell_m < u_m - 4\delta$. Soit N tel que $|\mathrm{V_N} \cap \mathrm{I}_m| > \ell_m - \delta$. L'intervalle I_m est coupé en 2 et on obtient deux intervalles I'_m et I''_m de longueur $u_{m+1} = u_m/2 = |\mathrm{I}|/2^m$. Puisque $|\mathrm{V_N} \cap \mathrm{I}_m| = |\mathrm{V_N} \cap \mathrm{I}'_m| + |\mathrm{V_N} \cap \mathrm{I}''_m|$, pour l'un de ces deux intervalles, qu'on appelle I_{m+1}, on a explicitement

$$|\mathrm{V_N} \cap \mathrm{I}_{m+1}| < \frac{|\mathrm{V_N} \cap \mathrm{I}_m|}{2} \leq \frac{\ell_m}{2} + \delta < u_{m+1} - 2\delta.$$

En outre $\ell_{m+1} = \lim_{p\to\infty} |\mathrm{V}_p \cap \mathrm{I}_{m+1}|$ existe, parce que

$$0 \leq |\mathrm{V}_{p+q} \cap \mathrm{I}_{m+1}| - |\mathrm{V}_p \cap \mathrm{I}_{m+1}| \leq |\mathrm{V}_{p+q} \cap \mathrm{I}_m| - |\mathrm{V}_p \cap \mathrm{I}_m| = \ell_m - |(\mathrm{V}_p \cap \mathrm{I}_m|.$$

1. Cette preuve est pour l'essentiel due à Newcomb Greenleaf : voir Bridges et Richman (1987, Chapter 3, Theorem 4.5). Merci à Peter Schuster qui nous a indiqué la référence.

En outre

$$\ell_{m+1} \leq |\mathrm{V_N} \cap \mathrm{I}_{m+1}| + (\ell_m - |\mathrm{V_N} \cap \mathrm{I}_m|) < u_{m+1} - 2\delta + \delta = u_{m+1} - \delta.$$

On a donc une récurrence qui fonctionne. L'intersection des intervalles I_m est réduite à un point x. Montrons que $x \notin \mathrm{J}_n$. Pour cela considérons un $y \in \mathrm{J}_n$ et un $r > 0$ tel que $[y - r, y + r] \subset \mathrm{J}_n$. Soit m tel que $|\mathrm{I}_m| < r/2$, puisque $|\mathrm{J}_n \cap \mathrm{I}_m| \leq |\mathrm{V}_n \cap \mathrm{I}_m| < |\mathrm{I}_m|$ il existe $z \in \mathrm{I}_m - (\mathrm{J}_n \cap \mathrm{I}_m)$. On a $|x - z| < r/2$ et $d(z, \mathrm{J}_m) > r$ donc $|x - y| > r/2$. □

8.4 Heine-Borel

Le théorème de Heine-Borel présente l'avantage de pouvoir être énoncé en parlant presque uniquement d'intervalles.

Dans la mesure où on pense que structure intuitive du continu est mieux comprise à travers l'idée d'intervalles qui se chevauchent plutôt qu'à travers celle d'un ensemble (totalement ordonné) de points, le théorème de Heine-Borel nous donnerait un meilleur accès à la structure du continu.

Une formulation simple du théorème de Heine-Borel est la suivante :

> *Si une famille d'intervalles ouverts recouvre l'intervalle* $[0, 1]$ *alors une sous-famille finie recouvre aussi l'intervalle* $[0, 1]$.

Sous une forme encore plus simple, on peut supposer qu'il s'agit d'une famille indexée par $\mathbb{N}$ $(]a_n, b_n[)_{n\in\mathbb{N}}$ d'intervalles rationnels ($a_n, b_n \in \mathbb{Q}$).

Que nous dit la preuve de ce théorème, comment fonctionne-t-elle ?

Notons

$$\mathrm{V}_k = \bigcup_{0\leq n\leq k}(]a_n, b_n[\cap [0, 1]).$$

Chaque V_k est une réunion finie d'intervalles de la forme $[0, \alpha[$ ou $]\beta, 1]$ ou $]\gamma, \delta[$. Il a donc une mesure bien définie, c'est un nombre rationnel que nous notons $|\mathrm{V}_k|$. Remarquons que si la limite de la suite $|\mathrm{V}_k|$ existe, elle doit être égale à 1 : ceci résulte de la section précédente. Car dans le cas contraire on est capable de trouver des réels en dehors de $\bigcup_k \mathrm{V}_k$. Mais nous voulons beaucoup plus, nous voulons que $|\mathrm{V}_k| = 1$ pour un certain k. Si cela se produit la différence entre V_k et $[0, 1]$ est constituée d'un nombre fini de points, qui seront recouverts par un certain V_ℓ et celui-ci sera égal à $[0, 1]$. Mais comment trouver un tel indice k ?

La démonstration usuelle procède par contradiction. Supposons qu'aucun V_k ne recouvre $[0, 1]$. Notons F_k le complémentaire de V_k dans $[0, 1]$. C'est une réunion finie d'intervalles rationnels fermés, et $\mathrm{F}_{k+1} \subseteq \mathrm{F}_k$ pour tout k. Tous les F_k sont donc supposés non vides. On construit maintenant des intervalles fermés emboités I_n contenus dans $[0, 1]$. Le premier intervalle est $\mathrm{I}_0 = [0, 1]$.

Construisons le second intervalle emboité, I_1. Considérons les deux moitiés de l'intervalle I_0 : $\mathrm{I}_{0,0} = [0, 1/2]$ et $\mathrm{I}_{0,1} = [1/2, 1]$. Ou bien tous les F_k coupent $\mathrm{I}_{0,0}$. Ou bien, dans le cas contraire ils sont, à partir d'un certain rang, tous contenus dans $\mathrm{I}_{0,1}$: puisque la suite F_k est décroissante, tous les F_k coupent $\mathrm{I}_{0,1}$. Dans le premier cas on prend $\mathrm{I}_1 = \mathrm{I}_{0,0}$, dans le second cas on prend $\mathrm{I}_1 = \mathrm{I}_{0,1}$. Dans les deux cas, tous les F_k coupent I_1.

Construisons le troisième intervalle emboité, I_2. Pour ceci considérons les deux moitiés $\mathrm{I}_{1,0}$ et $\mathrm{I}_{1,1}$ de I_1 (par exemple dans le premier cas $\mathrm{I}_{1,0} = [0, 1/4]$ et $\mathrm{I}_{1,1} = [1/4, 1/2]$). Ou bien tous les F_k coupent $\mathrm{I}_{1,0}$. Ou bien, dans le cas contraire ils sont, à partir d'un certain rang, tous contenus dans $\mathrm{I}_{1,1}$: puisque la suite F_k est décroissante, tous les F_k coupent $\mathrm{I}_{1,1}$. Dans le premier cas on prend $\mathrm{I}_2 = \mathrm{I}_{1,0}$, dans le second cas on prend $\mathrm{I}_2 = \mathrm{I}_{1,1}$. Dans les deux cas, tous les F_k coupent I_2.

Et ainsi de suite.

On a donc une suite d'intervalles emboités I_n avec la propriété que chaque F_k coupe chaque I_n.

Ayant en mains cette suite I_n, nous savons qu'un nombre réel x est à l'intersection de ces intervalles emboités. Mais ce nombre réel est aussi dans tous les F_k puisque, pour chaque k, x est

la limite d'une suite dans F_k : il suffit de prendre pour $x_{k,n}$ un point de l'intersection non vide $\mathrm{F}_k \cap \mathrm{I}_n$.

Nous pouvons être fiers de nous, ou plutôt de Borel, qui a trouvé la preuve.

Mais une fois que nous sommes convaincus qu'il existe, comment diable allons nous trouver l'indice ℓ tel que $\mathrm{I}_\ell = [0,1]$? Nous ne connaissons pas la réponse qu'aurait donnée Borel, mais cela pourrait être la suivante : « wait and see ». Autrement dit il suffit d'attendre. Puisque chaque V_k peut se réécrire comme une réunion finie d'intervalles rationnels disjoints, on constatera bien un jour que $\mathrm{V}_\ell = [0,1]$.

Commentaire. Nous nous trouvons ainsi dans une situation assez typique pour certains théorèmes d'analyse. L'hypothèse est à vrai dire un peu étrange : les intervalles rationnels donnés au départ recouvrent $[0,1]$. Comment certifier une telle hypothèse ? Il faudrait pour chaque réel x donner le moyen de trouver un indice m tel que $x \in]a_m, b_m[$. On a beau se creuser la tête, personne n'a jamais dit comment on pourrait être capable d'un tel tour de force : on démarre avec un nombre réel x arbitraire, c'est-à-dire en pratique avec une suite de Cauchy de nombres rationnels. Et à partir de cet x on doit trouver un indice m. Toute procédure que nous pouvons imaginer pour cela se doit d'être suffisamment uniforme pour que nous puissions démontrer $x \in]a_m, b_m[$ à partir d'une information finie concernant x.

Dans ce cas on aura donc un intervalle ouvert rationnel J_x contenant x pour lequel tous les éléments de l'intervalle sont justifiables du même certificat. Ce que nous sommes censés savoir par hypothèse, c'est que de tels intervalles recouvrent $[0,1]$. Mais pour que la procédure de certification donnée en hypothèse soit suffisamment uniforme, il faudrait que le théorème de Heine-Borel s'applique à cette nouvelle famille d'intervalles rationnels. Bref, certifier l'hypothèse sans avoir recours à la conclusion semble hors d'atteinte, car nous tournons en rond.

En résumé personne n'est capable de donner la moindre explication quant à la possibilité de certifier l'hypothèse du théorème, si ce n'est de certifier directement sa conclusion. Car la conclusion est de nature tout à fait élémentaire tandis que l'hypothèse est de nature extrêmement compliquée.

Par ailleurs la preuve que nous avons donnée est de nature hautement non explicite. D'une part il s'agit d'une preuve par contradiction, et les preuves d'existence par contradiction ne donnent pas le résultat sous forme explicite. D'autre part à l'intérieur même de la preuve par contradiction, nous avons une procédure hautement non explicite. En effet, par exemple pour déterminer quelle est la moitié de I_0 que nous allons prendre pour I_1, il nous faut décider si tous les F_k coupent ou ne coupent pas la première moitié de I_0. Ceci nécessite la connaissance *en son entier* de la suite des intervalles $]a_n, b_n[$ qui est donnée en entrée.

Conclusion. Nous espérions y voir plus clair sur la structure du continu avec le théorème de Heine-Borel, qui parle surtout d'intervalles. Mais il s'est avéré que les points de la droite réelle sont quand même au cœur du théorème lorsqu'on parle d'une famille d'intervalles ouverts qui *recouvrent* l'intervalle $[0,1]$. Notre analyse du théorème et de sa démonstration nous a plutôt conduit à un nouveau point d'interrogation sur la nature de certains théorèmes d'analyse.

Après tout nous n'avons donc pas perdu notre temps !

Chapitre 9

Cantor et l'infini en acte

Introduction

Nous présentons dans ce chapitre quelques résultats de Cantor concernant les cardinaux infinis. Nous discutons leur signification.

Nous introduisons le paradoxe de Russell concernant les « ensembles trop infinis » et discutons sa portée.

Nous décrivons comment les brèches ont été colmatées et nous discutons quelle appréciation on peut porter sur le résultat.

9.1 Grands résultats sur les petits infinis

Dans le titre un peu provocateur de cette section, on parle des « petits infinis » qui sont les infinis habituellement utilisés en mathématiques. L'infini dénombrable, l'infini des nombres réels (la puissance du continu) et parfois l'infini de toutes les parties de la droite réelle (en théorie de la mesure usuelle par exemple).

Il y a déjà beaucoup à faire avec ces petits infinis, et l'espoir initial de Cantor était une hiérarchie d'infinis plus fournie à la base des mathématiques. Par exemple, cela aurait été bien si les infinis $\mathbb{R}^n$ avaient formé une suite strictement croissante pour $n = 1, 2, 3, \ldots$.

Malheureusement, Cantor finit par démontrer que tous les $\mathbb{R}^n$ (pour $n \geq 1$) avaient la même puissance (le même cardinal) : "Je le vois, mais je ne le crois pas", avait-il écrit à Dedekind (voir page 62).

En fait la plupart des espaces qui interviennent en analyse fonctionnelle sont des espaces de Banach séparables, et ils ont le même cardinal que $\mathbb{R}$.

De ce point de vue, la théorie fut décevante : elle permit d'établir une nette distinction entre $\mathbb{N}$ et $\mathbb{R}$, mais guère plus. Il n'est pas bien clair de savoir si des ensembles de cardinalité plus grande sont *vraiment* utiles (autrement que pour le confort de certaines preuves). L'ensemble des parties de $\mathbb{R}$, noté $\mathfrak{P}(\mathbb{R})$, s'avère certes utile en théorie de la mesure, mais son cardinal n'intervient pas dans la théorie, et on est surtout intéressé par les sous-ensembles boréliens de $\mathbb{R}$, qui forment un ensemble de même cardinal que $\mathbb{R}$.

9.1.1 Définitions et propriétés de base

Qu'est-ce qu'un ensemble ?

La définition la plus fondamentale dans la théorie des ensembles de Cantor est naturellement celle d'ensemble.

Celle que proposait Cantor était de considérer qu'on pouvait prendre n'importe quelle propriété concernant des objets mathématiques, et que cela permettait *ipso facto* de définir l'ensemble des

objets vérifiant cette propriété. En démarrant avec un univers d'objets mathématiques comprenant les entiers naturels cela permettait de définir des ensembles à volonté.

Les ensembles infinis « en acte »

La définition extrêmement laxiste des « ensembles » par Cantor autorise au moins l'ensemble de tous les nombres entiers, en considérant la propriété, pour un objet mathématique, d'être un entier. Cela autorise donc des ensembles « infinis en acte ».

C'était là le principal « coup de force » contre la tradition grecque qui interdisait la considération de tels infinis. Pour justifier l'existence des ensembles infinis en acte, Cantor et ses supporters ont donné des arguments qui semblent aujourd'hui extrêmement étranges. Dedekind disait par exemple que l'ensemble des pensées est infini parce que si A est une pensée, la pensée que A est une pensée est une pensée distincte de A, ce qui enclenche un processus infini. Sous une forme un peu plus convaincante il aurait pu dire que l'ensemble des nombres entiers susceptibles d'être pensés est *a priori* non borné. Mais on voit alors qu'on revient ainsi à l'infini « potentiel » des Grecs, lui même basé sur le pari d'un futur infini.

Les ensembles en vrac

Une chose importante, qu'on a un peu oublié tant elle est devenue implicite et usuelle, est que la collection des objets constituant l'ensemble est considérée comme une pure collection, sans qu'aucune structure additionnelle soit prise en compte. Les éléments d'un ensemble sont là, en vrac, sans aucun ordre, sans autre relation entre eux que la relation d'inégalité.

Intuitivement cependant, on voit difficilement comment on pourrait considérer un ensemble à deux éléments sans les penser dans un certain ordre. Pour le faire il faut par exemple utiliser une ruse : l'ensemble des entiers impairs compris entre 4 et 8, par exemple. Mais bien qu'on l'ait un peu cachée, la structure ordonnée des entiers naturels est bien présente, et 5 vient avant 7 dans l'ordre naturel de l'énumération.

L'ensemble des parties d'un ensemble

Le coup de force de l'« ensemble infini en acte de tous les entiers naturels » semble *a posteriori* relativement acceptable. En effet on pourra sans doute systématiquement prendre l'expression $n \in \mathbb{N}$ comme un raccourci de « n est un entier naturel », et l'ensemble $\mathbb{N}$ n'interviendra dans la pratique mathématique qu'à travers le prédicat $n \in \mathbb{N}$. C'est ce qui se passe avec la version constructive de la théorie des ensembles développée par Bishop (1967) dans ses *Foundations of constructive analysis*.

Par exemple la preuve du théorème de Cantor donnée page 110 au chapitre 8 n'utilise aucun ensemble infini en acte, ni celui des entiers, ni celui des réels. Elle nous dit simplement comment construire, à partir de la donnée d'une suite de nombres réels, un nouveau nombre réel clairement distinct de tous les termes de la suite. Nul besoin pour cela de disposer de l'ensemble $\mathbb{N}$ en son entier, ni *a fortiori* de l'ensemble $\mathbb{R}$. Il suffit de savoir avec quelles règles on manipule les nombres entiers, les nombres réels, les suites de nombres réels.

Le coup de force de « l'ensemble des parties d'un ensemble » est nettement plus problématique.

Ce nouveau coup de force est caché dans la définition extrêmement générale de Cantor pour les ensembles. Cette définition autorise en effet « l'ensemble des parties d'un ensemble ». L'ensemble $\mathfrak{P}(\mathrm{E})$ correspond à la propriété suivante de l'objet X : X est un ensemble dont tout élément est un élément de E.

Dans la version constructive de la théorie des ensembles développée par Bishop, pour définir un ensemble (un type d'objets avec lesquels on désire travailler), la première chose à faire est de dire comment on construit un élément de l'ensemble. Qu'est-ce que pourrait être alors un élément de l'ensemble des parties de $\mathbb{N}$? Comment construit-on une partie de $\mathbb{N}$? On peut dire qu'une

partie X de $\mathbb{N}$ est définie par la propriété $P(n)$ qui définit X. Mais qu'est-ce qu'une propriété concernant les entiers ? Comment construit-on le type d'objet : « propriété » ? Il s'agit d'un objet beaucoup trop vague et fuyant pour pouvoir faire l'objet d'une construction systématique au sujet de laquelle on pourrait s'accorder.

Les cardinaux

Comme les ensembles n'ont pas d'autre structure que celle de « collection en vrac », pour comparer deux ensembles, il n'y a que « le nombre des objets » qui est pertinent. Cantor étend aux ensembles infinis la notion de *nombre d'éléments* de la manière suivante.

> *On dit que deux ensembles* E *et* F *ont* même puissance*, ou* même cardinal*, lorsqu'on peut les mettre en bijection l'un avec l'autre. Nous écrirons* $\mathrm{Card}(E) = \mathrm{Card}(F)$*. On dit encore que les ensembles sont* équipotents*.*

En particulier un ensemble est dit *dénombrable* lorsque son cardinal est égal à celui de $\mathbb{N}$. Et les ensembles qui ont même cardinal que $\mathbb{R}$ *ont la puissance du continu.* Au début de ses investigations, une préoccupation essentielle de Cantor était de comprendre la structure de $\mathbb{R}$, d'élucider le problème du continu.

Naturellement on est intéressé par une *relation d'ordre* entre les cardinaux, qui généralise la relation d'ordre des entiers naturels, lesquels correspondent aux cardinaux finis. Une définition raisonnable est de dire que $\mathrm{Card}(E) < \mathrm{Card}(F)$ *lorsque* E *n'est pas équipotent à* F *mais qu'il est équipotent à une partie de* F. À défaut d'avoir le tout plus grand que la partie, on a ainsi le tout supérieur ou égal à la partie.

Une première question qui se pose pour légitimer cette deuxième définition est de montrer l'antisymétrie au sens des relations d'ordre strict : les propriétés $\mathrm{Card}(E) < \mathrm{Card}(F)$ et $\mathrm{Card}(F) < \mathrm{Card}(E)$ sont incompatibles.

Autrement dit : *si* E *est équipotent à une partie de* F *et si* F *est équipotent à une partie de* E *alors* E *et* F *sont équipotents.*

Ce dernier résultat s'appelle le théorème de Cantor-Bernstein-Schröder.

L'axiome du choix

La question qui se pose naturellement ensuite c'est de savoir si deux infinis peuvent toujours être comparés selon leur taille :

> *deux ensembles* E *et* F *étant donnés, l'un est-il nécessairement équipotent à une partie de l'autre ?*

Cantor le pensait, mais la preuve nécessite l'axiome du choix : voyons pourquoi.

Un moyen de comparer les cardinaux s'offre avec les surjections : s'il existe une application surjective $\varphi\colon E \to F$ alors ce serait bien que le cardinal de F soit inférieur ou égal au cardinal de E. Autrement dit, on voudrait avoir une injection $\psi\colon F \to E$. Cela semble assez simple. Pour chaque $y \in F$ on considère l'un des $x \in E$ tels que $\varphi(x) = y$ (il en existe par hypothèse), et on pose $\psi(y) = x$. Mais cette construction qui semble anodine, et qui est difficilement contestable si F est fini ou dénombrable, est une des formes de l'axiome du choix.

Il faut peut être d'abord expliquer la terminologie. La fonction ψ choisit un élément $\psi(y)$ dans chacun des ensembles (non vides par hypothèse) $\varphi^{-1}(y)$. Le problème n'est pas de trouver, pour un y fixé, un élément dans l'ensemble $\varphi^{-1}(y)$, car cet ensemble est justement non vide par hypothèse, mais bien de faire ce choix de manière simultanée pour tous les $y \in F$ à la fois. Le nom « axiome du choix » est donc un raccourci pour « axiome de choix infini simultané ». Précisément :

Axiome du choix. *Pour toute application* $\varphi\colon E \to F$ *surjective, il existe* $\psi\colon F \to E$ *telle que* $\varphi \circ \psi = \mathrm{Id}_F$.

Dans la formulation qui vient d'être donnée l'axiome du choix semble effectivement anodin. Il semble qu'il s'agisse d'une extrapolation raisonnable aux ensembles infinis d'une propriété vraie des ensembles finis. Mais ce sont certaines conséquences de l'axiome du choix qui sont problématiques et qui mettent ce dernier sur la sellette.

Notons que l'argument selon lequel l'axiome du choix devrait être vrai par extrapolation du fini à l'infini est en soi un argument très faible. Car l'infini actuel a justement été *créé* par le coup de force consistant à dire que désormais on considèrerait comme irrémédiablement fausse la propriété, vraie dans le domaine des collections finies, selon laquelle il ne peut y avoir d'application bijective d'une partie stricte de E sur E lui-même (le tout est plus grand que la partie).

L'arithmétique des cardinaux

Une fonction f d'un ensemble E vers un ensemble F peut être définie par son graphe $\{\,(x, f(x)) \mid x \in \mathrm{E}\,\} \subseteq \mathrm{E} \times \mathrm{F}$. Il s'ensuit que selon le point de vue cantorien les fonctions de E vers F forment de nouveau un ensemble, noté $\mathcal{F}(\mathrm{E}, \mathrm{F})$, ou parfois F^{E}. Cette dernière notation est justifiée par le fait que si E est un ensemble fini à k éléments, alors l'ensemble $\mathcal{F}(\mathrm{E}, \mathrm{F})$ s'identifie à l'ensemble des $k$uples d'éléments de F, c'est-à-dire à F^k.

On a alors une *arithmétique des cardinaux* définie comme suit :

1. $\mathrm{Card}(\mathrm{E}) + \mathrm{Card}(\mathrm{F}) = \mathrm{Card}(\mathrm{E}_1 \cup \mathrm{F}_1)$, où E_1 et F_1 sont des copies disjointes de E et F. Par exemple E_1 et F_1 sont les deux parties de $(\mathrm{E} \cup \mathrm{F}) \times \{0, 1\}$ définies par $\mathrm{E}_1 = \{(x, 0) \mid x \in \mathrm{E}\}$ et $\mathrm{F}_1 = \{(x, 1) \mid x \in \mathrm{F}\}$.
2. $\mathrm{Card}(\mathrm{E}) \cdot \mathrm{Card}(\mathrm{F}) = \mathrm{Card}(\mathrm{E} \times \mathrm{F})$,
3. $\mathrm{Card}(\mathrm{E})^{\mathrm{Card}(\mathrm{F})} = \mathrm{Card}(\mathrm{E}^{\mathrm{F}})$,

Il est bon de noter que, d'un point de vue naïf, $\mathfrak{P}(\mathrm{E})$ est en bijection naturelle[1] avec $\{0, 1\}^{\mathrm{E}}$: à une partie X de E on fait correspondre sa fonction caractéristique ϕ_{X} définie par $\phi_{\mathrm{X}}(x) = 1$ si $x \in \mathrm{X}$ et $\phi_{\mathrm{X}}(x) = 0$ si $x \notin \mathrm{X}$.

9.1.2 Quelques ensembles dénombrables

Un premier résultat dans la théorie de Cantor est que $\mathbb{N}$ et $\mathbb{Q}$ ont même cardinal. D'autres collections nettement plus grandes que $\mathbb{Q}$ sont également dénombrables.

Par exemple l'ensemble $\mathbb{Z}[\mathrm{X}]$: on peut définir une application injective de $\mathbb{Z}[\mathrm{X}]$ dans $\mathbb{N}$ comme suit. On considère les entiers naturels qui interviennent dans les coefficients d'un polynôme $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$ comme écrits en base 2, avec les chiffres 0 et 1. Le signe $-$ sera représenté par le symbole 2 et un séparateur $|$ est représenté par le symbole 3. Un élément $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$ va alors être codé par un entier naturel écrit en base 4, avec les chiffres $0, 1, 2, 3$. On code le polynôme sous la forme de la liste de ses coefficients en commençant par le degré le plus haut et en s'arrêtant au coefficient constant. Par exemple $13\mathrm{X}^4 + \mathrm{X}^2 - 15$ est d'abord réécrit sous forme de la liste $13|0|1|0|-15$, puis les entiers sont écrits en base 2 ce qui donne $1101|0|1|0|-1111$, enfin les symboles $|$ et $-$ sont remplacés les symboles par 3 et 2 et on obtient 1101303130321111. Il suffit enfin de lire le mot obtenu comme un entier écrit en base 4. Le polynôme nul est représenté par 0, les constantes > 0 sont représentées « par elles mêmes », à ceci près qu'on les écrit en base 2 et qu'on les lit en base 4. On obtient bien ainsi un *codage* : autrement dit deux polynômes distincts sont codés de manières distinctes. En outre un polynôme admet un seul code. On a bien défini une application injective de $\mathbb{Z}[\mathrm{X}]$ dans $\mathbb{N}$.

Autre exemple : *l'ensemble des nombres réels algébriques est dénombrable.* Un réel algébrique est un réel ξ qui est racine d'un polynôme $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$. Cette fois-ci on va définir une application

1. Nous adoptons ici le point de vue naïf selon lequel toute propriété bien énoncée est automatiquement vraie ou fausse, ce qui nous permet de définir la fonction caractéristique d'une partie arbitraire de E. En fait cette affirmation sera mise en doute par Brouwer, qui contestera le principe du tiers exclu. Nous en reparlerons par ailleurs.

surjective d'une partie dénombrable de $\mathbb{N}$ vers l'ensemble des nombres réels algébriques, que nous noterons $\mathbf{R}$. Autrement dit, on va coder les éléments de $\mathbf{R}$ par des entiers naturels, mais un même réel algébrique aura plusieurs codes possibles. Si P est codé par l'entier k écrit en base 4 selon la procédure précédente, on décide de coder ξ en rajoutant au début du code un entier r écrit en base 2 suivi du séparateur 3. L'entier r est le numéro d'ordre de ξ parmi les racines réelles de P rangées en ordre croissant. Par exemple la première racine réelle du polynôme $13\mathrm{X}^4 - \mathrm{X}^3 + \mathrm{X} - 15$ sera codée par 1311013213031321111. En fait ce genre de codage est tout à fait « effectif ». Il existe en effet un algorithme (l'algorithme de Sturm) qui permet de calculer le nombre de racines réelles d'un polynôme à coefficients entiers. On peut alors certifier par une procédure explicite si un entier écrit en base 4 est ou n'est pas le code d'un réel algébrique. On peut également tester par une procédure explicite si deux codes de réels algébriques représentent le même réel ou deux réels distincts. En fait, non seulement la relation d'égalité, mais toute la structure algébrique de $\mathbf{R}$ peut être explicitée pour ce codage.

Les explications qui précèdent relèvent aujourd'hui de la routine pour les informaticiens, qui passent leur temps à coder des ensembles divers et variés au moyen de listes de 0 et 1 (la seule chose que sache manger une machine).

9.1.3 La puissance du continu

On a déjà établi le premier théorème de Cantor (de manière constructive) au chapitre 8. Sous forme condensée et appauvrie on le lit comme suit : l'ensemble des nombres réels n'est pas dénombrable. Nous allons voir maintenant comment on peut établir un résultat fondamental de même nature, qui valut la gloire à Cantor :

Théorème 9.1.1 (Théorème de Cantor, 2)**.** *En posant* $\omega = \mathrm{Card}(\mathbb{N})$ *on a* $\omega < 2^\omega = \mathrm{Card}(\mathbb{R})$.

Autrement dit le cardinal de $\mathbb{R}$ est strictement plus grand que celui de $\mathbb{N}$ et il est égal à celui de l'ensemble $\{0,1\}^{\mathbb{N}}$ des fonctions de $\mathbb{N}$ dans $\{0,1\}$.

On a tout d'abord le résultat général suivant.

Théorème 9.1.2 (Théorème de Cantor, 3)**.**

1. *Pour tout ensemble* E, $\mathrm{Card}(\mathrm{E}) < 2^{\mathrm{Card}(\mathrm{E})}$.
2. *Plus précisément il n'existe pas d'application surjective* $\mathrm{E} \to \{0,1\}^{\mathrm{E}}$.
3. *Plus précisément, si* $\varphi\colon \mathrm{E} \to \{0,1\}^{\mathrm{E}}$ *est une application arbitraire, l'élément* $\psi \in \{0,1\}^{\mathrm{E}}$ *défini par*
$$\psi\colon x \mapsto 1 - \varphi(x)(x)$$
n'est pas dans l'image de φ.

Démonstration. En effet si on avait $\psi = \varphi(a)$ on aurait en particulier $\psi(a) = \varphi(a)(a)$, ce qui n'est pas le cas. □

Cette preuve magistrale par sa simplicité n'a pas été découverte tout de suite. On l'appelle « la preuve diagonale de Cantor », parce qu'elle utilise astucieusement la diagonale $\{(a,a) \mid a \in \mathrm{E}\}$ du produit cartésien $\mathrm{E} \times \mathrm{E}$, en définissant ψ à partir de l'information donnée non pas par tous les $\varphi(a)(b)$, mais en se limitant à la diagonale.

Remarque. Beaucoup de gens pensent qu'il s'agit d'une preuve par l'absurde du point *1.* du théorème, $\mathrm{Card}(\mathrm{E}) < 2^{\mathrm{Card}(\mathrm{E})}$, parce qu'on dit dans le point *2.* « il n'existe pas d'application surjective… ». En fait il s'agit d'une preuve tout à fait constructive du point *3.* du théorème, qui implique trivialement les deux points précédents. La définition même de l'inégalité de deux cardinaux est négative : il n'y a pas de bijection… Prouver une négation c'est prouver une absurdité. Il s'agit donc d'une preuve *de* l'absurde (et non pas d'une preuve *par* l'absurde), *via* une preuve constructive d'un fait positif.

Lemme 9.1.3. *Il existe une application injective de* $\{0,1\}^{\mathbb{N}}$ *dans* $[0,1] \subset \mathbb{R}$.

Démonstration. On peut prendre $\theta_3 \colon (u_n)_{n\in\mathbb{N}} \mapsto \sum_{n=0}^{\infty} \frac{2u_n}{3^{n+1}}$. □

Remarque. L'ensemble des valeurs prises par θ_3 est l'ensemble $\mathbb{K}$ que nous avons appelé ensemble de Cantor page 111.

Lemme 9.1.4. *Il existe une application surjective de* $\{0,1\}^{\mathbb{N}}$ *sur* $[0,1] \subset \mathbb{R}$.

Démonstration. On peut prendre $\theta_2 \colon (u_n)_{n\in\mathbb{N}} \mapsto \sum_{n=0}^{\infty} \frac{u_n}{2^{n+1}}$. □

Lemme 9.1.5. *Il existe une application injective de* $[0,1]$ *dans* $\{0,1\}^{\mathbb{N}}$.

Démonstration. D'après le lemme 9.1.4 et l'axiome du choix. □

Proposition 9.1.6. *Il existe une bijection de* $\{0,1\}^{\mathbb{N}}$ *sur* $[0,1]$.

Première preuve. On utilise le théorème de Bernstein et les lemmes 9.1.3 et 9.1.5. □

Deuxième preuve. On se passe de l'axiome du choix, utilisé dans la preuve du lemme 9.1.5 comme suit. On considère l'application θ_2 utilisée dans la preuve du lemme 9.1.4. On remarque qu'elle est presque injective. Les seuls cas de non injectivité sont les nombres $z \in \mathbb{D}_2 \cap]0,1[$. Ces nombres ont exactement deux développements infinis en base 2. On énumère les éléments de $\mathbb{D}_2 \cap]0,1[$ dans une suite infinie[2] (a_n) et on modifie comme suit θ_2, pour obtenir une bijection θ_2' :

— si $\theta_2(u) \in \mathbb{D}_2 \cap]0,1[$, alors $\theta_2(u) = a_n$ pour un certain entier n et on pose : $\theta_2'(u) = a_{2n}$ si u se termine par des 0, et $\theta_2'(u) = a_{2n+1}$ si u se termine par des 1,

— si $\theta_2(u) \notin \mathbb{D}_2 \cap]0,1[$, on ne modifie pas θ_2 : $\theta_2'(u) = \theta_2(u)$. □

Corollaire 9.1.7. *Les ensembles* $\{0,1\}^{\mathbb{N}}$, $[0,1]$, $]0,1[$ *et* $\mathbb{R}$ *ont même cardinal.*

Démonstration. Il est facile de donner une bijection croissante continue entre $]0,1[$ et $\mathbb{R}$. Par exemple, pour $x \in]0,1/2]$, $f(x) = 4-1/x$ et pour $x \in [1/2,1[$, $f(x) = 1/(1-x)$. Ces deux ensembles ont donc même cardinal. Les inclusions $]0,1[\subset [0,1] \subset \mathbb{R}$ montrent alors que $\mathrm{Card}(]0,1[) = \mathrm{Card}([0,1]) = \mathrm{Card}(\mathbb{R})$. D'après la proposition 9.1.6 on a donc $\mathrm{Card}(\mathbb{R}) = 2^{\omega}$ où $\omega = \mathrm{Card}(\mathbb{N})$. □

Remarque. Ceci achève la démonstration du théorème 9.1.1.

Le premier théorème de Cantor, déjà traité au chapitre 8 fut démontré par Cantor avant les théorèmes 9.1.1 et 9.1.2. Nous en donnons ici une nouvelle preuve (non constructive cette fois-ci).

Théorème 9.1.8. *L'ensemble des réels n'est pas dénombrable. Aucune suite de réels ne recouvre l'intervalle* $[0,1]$.

Première preuve. D'après la proposition 9.1.6 et le théorème diagonal de Cantor 9.1.2. □

Deuxième preuve. On court-circuite la preuve précédente comme suit. Si $(x_n)_{n\geq 1}$ est une suite de réels dans $[0,1]$ on écrit chaque x_n sous forme de son développement en base 2, en prenant soin d'écrire les deux développements en base 2 lorsque $x_n \in \mathbb{D}_2 \cap]0,1[$. Cela nécessite de changer un petit peu la numérotation de la suite, puisque les éléments de la suite qui sont dans $\mathbb{D}_2 \cap]0,1[$ doivent être répétés deux fois. On a maintenant (avec la nouvelle numérotation) $x_n = \sum_{m=1}^{\infty} y_{n,m}/2^m$. Posons alors $z = \sum_{m=1}^{\infty}(1-y_{m,m})/2^m$. Alors le développement de z est différent de celui de tous les x_n, y compris ceux qui admettent deux développements. Donc z est distinct de tous les x_k. □

2. Par exemple on peut les ordonner par dénominateurs croissants (le numérateur étant impair), puis pour un dénominateur fixé, par ordre croissant.

Rappelons qu'un espace métrique est dit *séparable* s'il possède une partie dénombrable dense.

Corollaire 9.1.9. *Les ensembles* $\{0,1\}^{\mathbb{N}}$, $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{\mathbb{N}}$ *et tous les espaces de Banach séparables (non réduits à* 0*) ont même cardinal.*

Démonstration. Voici une bijection croissante continue entre $]0,1[$ et $\mathbb{R}$.

$$f\colon]0,1[\to \mathbb{R}, \quad x \mapsto \begin{cases} 4 - 1/x & \text{si } x \in]0,1/2], \\ 1/(1-x) & \text{si } x \in [1/2,1[. \end{cases}$$

Ces deux ensembles ont donc même cardinal.

Les inclusions $]0,1[\subset [0,1] \subset \mathbb{R}$ montrent alors que $\mathrm{Card}(]0,1[) = \mathrm{Card}([0,1]) = \mathrm{Card}(\mathbb{R})$. D'après la proposition 9.1.6 on a donc $\mathrm{Card}(\mathbb{R}) = 2^\omega$ où $\omega = \mathrm{Card}(\mathbb{N})$. □

9.1.4 Des preuves constructives

Dans cette section nous analysons les preuves données dans la section 9.1.3 en nous plaçant du point de vue constructif. Plus précisément nous voulons répondre à la question suivante : « lorsque les hypothèses sont supposées réalisées de manière explicite, en est-il de même pour les conclusions ? » Dans certains cas, ceci est réalisé directement par la preuve donnée dans le section 9.1.3.

Du point de vue constructif les ensembles infinis sont considérés comme des infinis potentiels. On n'utilise pas l'axiome du choix. Plus généralement les procédures pour construire des ensembles ou des applications entre ensembles doivent être explicites.

Preuve constructive du théorème 9.1.2. Il n'y a rien à modifier à la preuve qui a été donnée. □

L'interprétation intuitive du théorème 9.1.2 est que pour n'importe quel ensemble infini E *clairement défini, l'ensemble* $\{0,1\}^{\mathrm{E}}$ *est un infini potentiel de nature intrinsèquement plus compliquée que* E *lui-même.*

Preuve constructive du lemme 9.1.3. On a déjà vu page 111 que l'application θ_3 établit une bijection entre $\{0,1\}^{\mathbb{N}}$ et son image $\mathbb{K}$ (l'ensemble de Cantor). *A fortiori* elle est injective. □

Preuve constructive du lemme 9.1.4. La preuve qui a été donnée présuppose qu'on sache calculer le développement standard en base 2 d'un nombre réel arbitraire. Mais ceci n'est pas possible en général : il ne suffit pas de connaitre un nombre réel avec une précision finie arbitrairement grande pour pouvoir le comparer sans erreur à tous les éléments de $\mathbb{D}_2$ (ce qui est nécessaire pour calculer son développement en base 2).

Par contre on a vu par ailleurs qu'il est possible d'obtenir constructivement un développement illimité en base 2 à condition d'utiliser les trois chiffres $-1, 0, 1$. On peut décider de coder -1 par $(0,0)$, 0 par $(0,1)$ ou $(1,0)$ et 1 par $(1,1)$. Autrement dit, (a,b) code $a+b-1$. Ceci fournit une application explicitement surjective $\lambda\colon \{0,1\}^{\mathbb{N}} \to [0,1]$ que l'on peut définir comme suit :

$$\lambda((x_n)_{n\in\mathbb{N}}) = \sup\left(0, \inf\left(1, \frac{1}{2} + \sum_{k=1}^{\infty} \frac{x_{2k-1} + x_{2k} - 1}{2^k}\right)\right). \qquad \square$$

Preuve constructive du lemme 9.1.5 ? La preuve donnée ne peut pas être rendue algorithmique, car elle utilise l'axiome du choix. □

Preuve constructive de la proposition 9.1.6 ? Aucune des deux preuves données ne peut être rendue algorithmique, la première utilise l'axiome du choix, la seconde plus simple n'utilise que le principe du tiers exclu, elle nécessite de savoir calculer le développement en base 2 d'un réel arbitraire, et de savoir décider si un réel fait ou non partie de $\mathbb{D}_2$. □

Bien que le corollaire 9.1.9 soit impossible à obtenir constructivement on a néanmoins de façon entièrement explicite le résultat le plus important de Cantor : « $\mathbb{R}$ n'est pas dénombrable ».

Preuve constructive que $\mathbb{R}$ *n'est pas dénombrable.* Elle a déjà été donnée : c'était le théorème 8.2.1. □

Là encore l'interprétation intuitive du théorème est que l'ensemble $\mathbb{R}$ est un infini potentiel de nature intrinsèquement plus compliquée que $\mathbb{N}$.

9.2 Paradoxes et incertitudes en théorie des ensembles

9.2.1 Le paradoxe de Cantor-Russell-Skolem

L'Univers de Cantor : un ensemble trop infini

Un des premiers paradoxes qui apparut dans l'Univers ensembliste de Cantor était le suivant. Si l'Univers de tous les ensembles mathématiques est un ensemble U, alors, toute partie de U est aussi un ensemble, donc un élément de U. Cela signifie que $\mathfrak{P}(\mathrm{U}) \subseteq \mathrm{U}$, mais c'est impossible parce que cela impliquerait $\mathrm{Card}(\mathfrak{P}(\mathrm{U})) \leq \mathrm{Card}(\mathrm{U})$, contrairement au théorème de Cantor (théorème 9.1.2).

Cette découverte est due à Cantor lui-même, en 1899.

En analysant de près la preuve de cette contradiction on aboutit directement au paradoxe de Russell publié en 1903 (mais Zermelo avait indépendamment abouti à la même conclusion, voir Rang et Thomas 1981).

Paradoxe de Russell. *Si l'Univers de tous les ensembles mathématiques est un ensemble* U*, alors considérons la partie*

$$\mathrm{X} = \{\mathrm{Y} \in \mathrm{U} \mid \mathrm{Y} \notin \mathrm{Y}\}$$

On obtient l'équivalence $\mathrm{X} \in \mathrm{X} \Leftrightarrow \mathrm{X} \notin \mathrm{X}$ *ce qui est absurde.*

En effet de manière générale, d'une implication $\mathrm{A} \Rightarrow \neg\mathrm{A}$ on déduit $\neg\mathrm{A}$, car supposer A conduit à une absurdité. Donc de l'équivalence ci-dessus on déduit $\mathrm{X} \notin \mathrm{X}$ et $\neg(\mathrm{X} \notin \mathrm{X})$. On vient donc de démontrer deux choses contradictoires.

9.2.2 Zermelo et Fraenkel colmatent les brèches

Après l'apparition du paradoxe de Russell, l'impérieuse nécessité d'un éclaircissement conduit les mathématiciens et logiciens à tenter de définir un cadre axiomatique pour la théorie des ensembles, dans lequel l'univers de tous les ensembles ne puisse pas avoir droit de cité.

Ce sera l'œuvre de Zermelo (1908). Dans le système d'axiomes qu'il propose, des limites sont imposées aux ensembles pour les empêcher d'être « trop infinis ».

En particulier si on note α_n la suite strictement croissante des cardinaux définie par $\alpha_0 = \mathrm{Card}(\mathbb{N})$ et $\alpha_{n+1} = 2^{\alpha_n}$ pour tout n, le système de Zermelo ne permet pas de démontrer l'existence d'un ensemble dont le cardinal dépasse celui de tous les α_n.

Certains se sont sentis frustrés par cette limitation, et Skolem et Fraenkel (1922) ont proposé des axiomes supplémentaires pour pouvoir aller beaucoup plus loin dans l'échelle des infinis.

On note **ZF** le système d'axiomes ainsi mis au point. Depuis son invention, il tient bon, et personne n'a trouvé de paradoxe logique. La plupart des mathématicien(ne)s professionnel(le)s sont satisfaits de cette situation : en fait la définition de l'intégrale de Lebesgue et la théorie de la mesure qui se développent au début du 20$^{\mathrm{e}}$ siècle semblent au premier abord nécessiter une théorie comme **ZF**, et l'intégrale de Lebesgue est un outil trop précieux.

9.2.3 Le paradoxe de Banach-Tarski

En 1904, répondant à un problème posé par Hilbert, Zermelo montre que l'axiome du choix implique l'existence d'une *relation de bon ordre* sur tout ensemble (une relation de bon ordre sur un ensemble E est une relation d'ordre total telle que toute partie de E possède un plus petit élément).

Ceci commence à jeter un doute sur l'axiome du choix lui-même, car il semble manifestement impossible de donner explicitement un bon ordre sur des ensembles tels que $\mathbb{R}$ ou $\{0,1\}^{\mathbb{N}}$.

Mais le coup le plus dur contre l'axiome du choix est sans doute le paradoxe de Banach-Tarski, bien qu'il soit apparu nettement plus tard et qu'il ne constitue pas un paradoxe au sens fort (qui serait de rendre la théorie **ZF** incohérente).

C'est le suivant. Si on suppose vrai l'axiome du choix et si on se situe dans le cadre axiomatique de Zermelo-Fraenkel, alors il existe une partition finie de la boule unité B de $\mathbb{R}^3$, $\mathrm{B} = \mathrm{B}_1 \cup \mathrm{B}_2 \cup \cdots \cup \mathrm{B}_n$, des *isométries* $\varphi_1, \ldots, \varphi_n$ de $\mathbb{R}^3$ et un entier $r < n$ tels que :

— $\varphi_1(\mathrm{B}_1), \ldots, \varphi_r(\mathrm{B}_r)$ constituent une partition de B, et

— $\varphi_{r+1}(\mathrm{B}_{r+1}), \ldots, \varphi_n(\mathrm{B}_n)$ constituent une partition de B.

9.2.4 Hypothèse du continu et axiome du choix

Dans les 23 problèmes de Hilbert au Congrès International de Mathématiques qui se tient à Paris en 1900, le premier concerne la structure de $\mathbb{R}$. Deux questions sont posées :

— Y a-t-il un cardinal strictement compris entre ω et 2^ω ? autrement dit y a-t-il une partie infinie Y de $\mathbb{R}$ qui ne puisse être mise en bijection ni avec $\mathbb{N}$, ni avec $\mathbb{R}$?

— Peut-on mettre une relation de bon ordre sur $\mathbb{R}$?

Ces deux questions étaient considérées comme cruciales par Cantor, qui espérait une réponse négative à la première et positive à la seconde.

La réponse apportée par Zermelo à la seconde question est par nature ambigüe (voir paragraphe précédent).

On a appelé « hypothèse du continu » (**HC**) l'affirmation selon laquelle *il n'y a pas de cardinal strictement compris entre ω et 2^ω.*

K. Gödel et P. Cohen démontreront que l'hypothèse du continu et l'axiome du choix sont indépendants des axiomes de **ZF**, si toutefois cette théorie est cohérente.

Il est écrit sur la tombe de Hilbert « Nous devons savoir, nous saurons ». Les résultats de Gödel et Cohen concernant l'axiome du choix et l'hypothèse du continu ont semé un doute sérieux quant à la possibilité de réaliser ce souhait.

9.2.5 Le réalisme platonicien

Ce genre de résultats pose le problème de savoir si toutes les questions qu'on se pose en mathématiques ont une signification objective claire. Y a-t-il un « monde des idées » dans lequel les questions mathématiques bien posées admettent une réponse *a priori*, indépendamment des preuves qui pourront être fournies par *Homo sapiens*.

Accepter une réponse positive à cette question relève du « réalisme platonicien ». Une certaine dose de ce genre de réalisme semble inévitable quand on parle des objets les plus simples, les entiers naturels par exemple. Le théorème fondamental de l'arithmétique, qui affirme que tout nombre entier positif admet une décomposition en facteurs premiers unique (à l'ordre des facteurs près), n'est pas une évidence *a priori*, mais il est vrai absolument, indépendamment des preuves que nous pouvons en fournir. De même il semble difficilement contestable que le théorème de

Fermat [3] soit une vérité absolue du même genre, et que la preuve qui en a finalement été donnée nous a fait connaitre cette vérité, qui était seulement restée longtemps cachée.

Notons cependant que lorsqu'on est en présence d'un énoncé du même style mais pour lequel aucune preuve ni aucun contre-exemple n'a pu être fourni, il est possible de contester qu'il y ait une vérité objective *a priori* concernant cet énoncé. En effet, non seulement la preuve ou le contre-exemple peuvent être hors de portée de *Homo sapiens*, auquel cas la vérité nous serait à tout jamais inaccessible, mais rien n'interdit qu'en plus ils soient « hors de portée de l'Univers ». En particulier si ce dernier est fini, il semble plausible qu'il existe des problèmes concernant les entiers naturels dont l'énoncé soit structurellement simple (pas plus compliqué structurellement que le théorème de Fermat [4]) mais dont aucune solution ne puisse être explicitée dans l'Univers, par exemple parce qu'un contre-exemple éventuel serait trop gros.

L'attitude du réalisme platonicien devient en tout cas nettement plus sujette à caution lorsqu'on se frotte à des énoncés d'une grande complexité logique, et surtout lorsqu'on parle d'objets abstraits inventés pour les besoins du discours, mais dont la réalité objective est *a priori* douteuse.

Skolem ou Poincaré estimaient que la comparaison des cardinaux a un caractère très conventionnel et ne reflète pas des vérités absolues. Pour eux, l'infini ne correspond à aucune réalité objective. C'est simplement une manière de parler. Pour Poincaré les énoncés sur l'infini ne devraient être considérés que comme des raccourcis du langage pour des énoncés concernant uniquement des objets finis. Si la comparaison des cardinaux n'a pas un caractère objectif, le problème posé par l'hypothèse du continu est un problème mal posé, qui ne correspond à aucune réalité, même dans le monde des idées. Et il n'est pas étonnant qu'aucune voie d'accès solide ne s'offre à nous pour le résoudre.

Beaucoup d'autres mathématicien(ne)s, la plupart sans doute, comme Gödel, s'en tiennent au point de vue du réalisme platonicien concernant l'infini : un univers mathématique cantorien existe bel et bien, et dans cet univers, l'hypothèse du continu est nécessairement vraie ou fausse, même dans le cas où la réponse nous serait à tout jamais cachée. Gödel estimait quant à lui que de nouveaux axiomes raisonnables doivent être cherchés, qui décideront vraie ou fausse l'hypothèse du continu. De même pour l'axiome du choix.

Si les infinis actuels $\mathbb{N}$ et $\mathbb{R}$ existent réellement « au moins de manière idéale » alors l'hypothèse du continu devrait avoir « une signification claire ». En fait si on y regarde d'un peu plus près, la bijection qui doit exister entre d'un côté une partie infinie Y arbitraire de $\mathbb{R}$ et, de l'autre côté ou bien $\mathbb{N}$, ou bien $\mathbb{R}$, est définie par son graphe, lequel est une partie de $\mathbb{R} \times \mathbb{R}$, c'est-à-dire un élément de $\mathfrak{P}(\mathbb{R} \times \mathbb{R})$. Comme $\mathbb{R} \simeq \mathfrak{P}(\mathbb{N})$ et $\mathbb{R} \times \mathbb{R} \simeq \mathbb{R}$, l'énoncé **HC** se réécrit au moyen d'une formule utilisant des variables quantifiées parcourant $\mathfrak{P}(\mathbb{N})$ et $\mathfrak{P}(\mathfrak{P}(\mathbb{N}))$. Il suffit donc que l'ensemble $\mathfrak{P}(\mathfrak{P}(\mathbb{N}))$ ait une réalité objective dans le monde des idées pour que l'hypothèse du continu soit objectivement vraie ou fausse.

Que cette constatation renforce le point de vue de Gödel ou celui de Poincaré, nous laisserons la lectrice en juger.

3. Le théorème de Fermat affirme que pour trois entiers $x, y, z > 0$ et pour un entier $n \geq 3$ on n'a jamais $x^n + y^n = z^n$. Fermat pensait en avoir une preuve. Mais une preuve convaincante a dû attendre la fin du 20$^{\text{e}}$ siècle pour voir le jour.

4. La structure logique du théorème de Fermat est la suivante. On a une fonction explicitement donnée f de $\mathbb{N}$ dans $\mathbb{N}$ et on demande de démontrer que $f(m) \neq 0$ pour tous les $m \in \mathbb{N}$. Pour chaque valeur de m, le calcul de $f(m)$ est explicite (par exemple on peut le programmer sur machine). Toute la complexité logique de l'énoncé se concentre sur le fait qu'il y a une infinité de vérifications individuelles en jeu (avec un infini du type $\mathbb{N}$).

Chapitre 10

Les définitions

Introduction

Je propose de mener une enquête sur la définition mathématique à partir d'un exemple géométrique : l'angle droit. Ce nom d'angle droit peut être défini en bonne et due forme, mais nous allons nous apercevoir que ce n'est le cas d'aucun des termes utilisés pour cela.

Les dictionnaires ont pour principe de définir tous les mots : j'en profite pour consulter la définition des mots d'angle et de droite. C'est l'occasion de constater que l'approche formaliste des mathématiques a aussi eu un impact sur les dictionnaires, ainsi que de documenter les difficultés de définir la droite.

À la fin, j'étudie dans quelle mesure la méthode axiomatique de Moritz Pasch (1882) peut proposer une définition implicite des termes primitifs de la géométrie.

Nous serons guidés dans cette enquête par l'opuscule *De l'esprit géométrique* de Blaise Pascal (1655).

10.1 Un exemple : l'angle droit.

Les cours de mathématiques emploient le mot "définition" lorsqu'ils introduisent un terme nouveau en l'exprimant au moyen de termes connus.

Je vais traiter un exemple en détail : la définition 10 du premier livre des *Éléments* d'Euclide.

10.1.1 Définir l'angle droit.

La définition 10 a deux parties, dont la première concerne l'angle *droit* et la deuxième la droite *perpendiculaire*.

> 10. Et quand une droite, ayant été élevée sur une droite, fait les angles adjacents égaux entre eux, chacun de ces angles égaux est *droit*, et la droite qui a été élevée est appelée *perpendiculaire* à celle sur laquelle elle a été élevée. (Euclide d'Alexandrie 1990-2001, volume 1, page 160.)

Considérons la première partie de cette définition. Selon une terminologie bien établie, elle est composée d'un *definiendum* (gérondif du verbe latin *definire*, ce qui est à définir) et d'un *definiens* (participe présent, ce qui définit) :

definiendum	*definiens*
angle droit	angle d'une droite élevée sur une autre égal à l'angle adjacent.

L'angle *droit* est défini au moyen des termes *angle*, *droite*, *élever*, *égal*, *adjacent*.

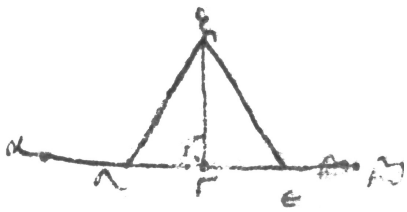
10.1.2 Invoquer la définition.

Cet exemple est particulier pour la raison suivante : le premier livre des *Éléments* contient 23 définitions, mais six seulement sont invoquées dans des démonstrations, et la définition 10 est la première dans ce cas. En fait, les autres sont inutiles dans les raisonnements et complètement absentes de la suite du traité.

La proposition 11 du premier livre, qui est un *problème* et non un théorème, c'est-à-dire qu'elle a « pour but de procurer, de rendre manifeste, de construire ce qui en un certain sens n'existe pas [encore] » (Proclus de Lycie, *Commentaire au premier livre des Éléments d'Euclide*, voir Caveing 1990, page 133), est la première à invoquer cette définition.

Voici ce problème avec la figure qui l'accompagne sur le folio 13 du manuscrit MS D'Orville 301 de l'an 888. Contrairement aux apparences, les lettres de la figure concordent bien avec les lettres de la démonstration, à ceci près qu'elles sont écrites en minuscule : les points de la base de la figure sont, de gauche à droite, les points A, Δ, Γ, E et B, et le point au sommet est le point Z. On y distingue le symbole de l'angle droit, ∟, pour l'angle sous ΔΓZ.

11. *Mener une ligne droite à angles droits avec une droite donnée, à partir d'un point donné sur celle-ci.*



Soit d'une part la droite donnée AB et d'autre part le point Γ donné sur elle.

Il faut alors mener, à partir du point Γ, une ligne droite à angles droits avec la droite AB.

Que soit pris au hasard le point Δ sur AΓ, et que soit placée ΓE égale à ΓΔ (Prop. 2). Que soit construit sur ΔE le triangle équilatéral ZΔE (Prop. 1), et que ZΓ soit jointe.

Je dis que la droite ZΓ est menée à angles droits avec la droite donnée AB à partir du point Γ donné sur celle-ci.

En effet puisque ΓΔ est égale à ΓE, que ΓZ est commune, alors les deux ΔΓ, ΓZ sont égales aux deux EΓ, ΓZ, chacune à chacune. Et la base ΔZ est égale à la base ZE (Df. 20). Donc l'angle sous ΔΓZ est égal à l'angle sous EΓZ (Prop. 8). Et ils sont adjacents.

Quand une droite, ayant été élevée sur une droite, fait des angles adjacents égaux entre eux, chacun de ces angles égaux est droit (Df. 10). Donc chacun des angles sous ΔΓZ, ZΓE est droit.

Donc la droite ΓZ a été menée à angles droits avec la droite donnée AB à partir du point Γ donné sur celle-ci. Ce qu'il fallait faire. (Euclide d'Alexandrie 1990-2001, volume 1, pages 217-218.)

La définition est invoquée de la manière suivante : la démonstration construit une droite sur la droite donnée, établit qu'elle fait les angles adjacents égaux, constate alors explicitement et mot pour mot le *definiens*, et conclut qu'on a bien le *definiendum*. C'est-à-dire que le *definiendum* abrège le *definiens* : dans ce sens, la définition est souvent qualifiée d'abréviation.

10.1.3 Établir la possibilité d'un concept.

Je conclus de la présence de cette proposition 11 que la définition 10 donne seulement une signification au nom d'angle *droit* sans s'exprimer sur l'existence d'un tel angle. C'est seulement la proposition 11 qui établit la possibilité qu'un angle soit droit.

Voici comment Blaise Pascal a exprimé en février 1648 le rôle spécifique de la définition vis-à-vis des propositions et postulats dans sa lettre à Jacques Le Pailleur.

D'où il est évident qu'il n'y a point de liaison nécessaire entre la définition d'une chose et l'assurance de son être ; et que l'on peut aussi bien définir une chose impossible qu'une véritable. Ainsi l'on

> peut appeler un triangle rectiligne, ou rectangle, celui qu'on s'imaginerait avoir deux angles droits, et montrer ensuite qu'un tel triangle est impossible. Ainsi Euclide définit d'abord les parallèles, et montre après qu'il y en peut avoir ; et la définition du cercle précède le postulat qui en propose la possibilité. (Pascal 1970a, page 563.)

En d'autres mots, la définition donne un nom à une chose, mais la possibilité ou l'impossibilité de la chose est indépendante de sa définition : elle doit être prouvée ou postulée.

10.2 La définition de nom.

10.2.1 Théorie.

On trouve dans l'opuscule *De l'esprit géométrique* de Blaise Pascal, daté des années 1655, la théorie de la définition telle que je viens de la présenter, appelée *définition de nom*.

> On ne reconnait en géométrie que les seules définitions que les logiciens appellent définitions de nom, c'est-à-dire que les seules impositions de nom aux choses qu'on a clairement désignées en termes parfaitement connus ; et je ne parle que de celles-là seulement.
>
> Leur utilité et leur usage est d'éclaircir et d'abréger le discours en exprimant, par le seul nom qu'on impose, ce qui ne se pourrait dire qu'en plusieurs termes ; en sorte néanmoins que le nom imposé demeure dénué de tout autre sens, s'il en a, pour n'avoir plus que celui auquel on le destine uniquement. En voici un exemple.
>
> Si l'on a besoin de distinguer dans les nombres ceux qui sont divisibles en deux également d'avec ceux qui ne le sont pas, pour éviter de répéter souvent cette condition, on lui donne un nom en cette sorte : j'appelle tout nombre divisible en deux également nombre pair.
>
> Voilà une définition géométrique, parce qu'après avoir clairement désigné une chose, savoir : tout nombre divisible en deux également, on lui donne un nom que l'on destitue de tout autre sens, s'il en a, pour lui donner celui de la chose désignée.
>
> D'où il parait que les définitions sont très libres, et qu'elles ne sont jamais sujettes à être contredites ; car il n'y a rien de plus permis que de donner à une chose qu'on a clairement désignée un nom tel qu'on voudra. (Pascal 1991, pages 393-394.)

La « géométrie » désigne ici toutes les mathématiques, arithmétique incluse, selon un usage assez commun jusqu'à Poincaré. Voici l'exemple donné par Pascal :

definiendum	*definiens*
nombre pair	nombre divisible en deux également.

10.2.2 Définition de nom et définition de chose.

Pascal écrit qu'il ne parle que de définitions de nom parce que les philosophes scolastiques du Moyen Âge (comme Guillaume d'Ockham, Jean Buridan et Pierre d'Espagne) en considéraient une autre : la définition *de chose*. Ce concept a été élaboré sur la base des *Topiques* d'Aristote, dont voici deux citations, et de ses *Seconds Analytiques* (voir Gomez-Lobo 1981).

Une définition est une formule qui exprime l'essentiel de l'essence d'un sujet.

. . . . . .

Une formule définitionnelle a pour composants un genre et des différences. (Aristote 1967, I, 2, 101 b 38 et I, 5, 103 b 15-16, pages 6 et 12.)

La formule dédoublée « l'essentiel de l'essence » proposée par le traducteur reprend le dédoublement du verbe être dans la formule d'Aristote τὸ τί ἦν εἶναι, qui se traduit littéralement par "le qu'est-ce que c'est qu'être" (un sujet) et est traduit traditionnellement par "la quiddité" (du sujet).

Les deux exemples de définitions que nous avons vus sont bien composées d'un genre et d'une différence.

	genre	différence
angle droit	angle de deux droites	égal à l'angle adjacent
nombre pair	nombre	divisible en deux également.

Nous nous rendons compte aussitôt que pour Aristote (et pour les philosophes en général), donner une définition est une entreprise très ambitieuse et qu'elle nécessite un long travail de réflexion. Le philosophe et mathématicien Giovanni Girolamo Saccheri l'a exprimé ainsi dans sa *Logique démonstrative* (1697).

De là s'entend que la définition quidditive ["qui explique la nature de la chose"] est le plus souvent le fruit d'une longue série de démonstrations sur un sujet donné. Elle ne peut pas, en fait, être établie sur le genre et la différence prochains si ce n'est à la suite d'un examen prolongé des propriétés ou des prédicats qui conviennent au sujet donné. J'ai dit *le plus souvent* parce qu'il peut arriver que la définition *de nom* posée en premier lieu soit elle-même quidditive ; dans tous les cas, qu'elle soit quidditive peut difficilement être établi dès le début, comme le révèle l'expérience. Pour cela, si ce n'est sur la base d'une cognition réfléchie, le fait qu'une définition soit quidditive est toujours le fruit de nombreuses démonstrations. (Saccherius 1701, page 120, ma traduction.)

10.2.3 Dans la Logique de Port-Royal.

On retrouve dans *La logique, ou l'art de penser* (première édition en 1662) d'Antoine Arnauld et Pierre Nicole l'opposition de la définition de nom à la définition de chose.

Le meilleur moyen pour éviter la confusion des mots qui se rencontrent dans les langues ordinaires, est de faire une nouvelle langue, et de nouveaux mots qui ne soient attachés qu'aux idées que nous voulons qu'ils représentent. [. . .]

C'est ce qu'on appelle la définition du nom, *definitio nominis*, dont les géomètres se servent si utilement, laquelle il faut bien distinguer de la définition de la chose, *definitio rei*.

Car dans la définition de la chose, comme peut-être celle-ci : *l'homme est un animal raisonnable* : *le temps est la mesure du mouvement*, on laisse au terme qu'on définit comme *homme* ou *temps* son idée ordinaire, dans laquelle on prétend que sont contenues d'autres idées, comme *animal raisonnable*, ou *mesure du mouvement* ; au lieu que dans la définition du nom, comme nous avons déjà dit, on ne regarde que le son, et ensuite on détermine ce son à être signe d'une idée que l'on désigne par d'autres mots. (Arnauld et Nicole 2011, I, XI, pages 232-233.)

En d'autres mots, une définition de chose n'est pas une définition au sens que lui donnent les cours de mathématiques : c'est une proposition qui semble seulement définir un terme en énonçant qu'il satisfait une certaine propriété, alors que ce terme signifie déjà son « idée ordinaire ».

Pour donner un exemple, je propose de considérer que la proposition 11 est, prise dans sa totalité, avec sa démonstration, une définition de chose de l'angle droit au sens suivant : la définition 10 propose simplement un point de départ, une « idée ordinaire » de l'angle droit, alors que la proposition 11 établit que c'est l'angle qu'on obtient en construisant d'abord un triangle équilatéral sur un segment de milieu donné puis la médiane issue de ce milieu ; il s'avère que cet angle est droit par symétrie. Elle assure ainsi l'existence d'une droite perpendiculaire à une droite donnée et menée à partir d'un point donné de cette droite. Cette définition de chose a un mérite supplémentaire : elle montre que l'angle droit ne relève pas de la théorie des parallèles et existe donc pareillement dans la géométrie non euclidienne (voir les sections 1.1 et 5.3).

Mais Arnauld et Nicole ne proposent pas de tel exemple. Pascal, dont on a vu à la section 10.1.3 qu'il est parfaitement conscient du rôle essentiel de telles propositions, choisit ses exemples en dehors des mathématiques et se borne à faire la mise en garde suivante.

> Combien y en a-t-il de même qui croient avoir défini le mouvement quand ils ont dit : *Motus nec simpliciter actus nec mera potentia est, sed actus entis in potentia* [*Le mouvement n'est ni simplement acte, ni pure puissance ; mais l'acte de ce qui est en puissance*, version latine libre de la définition donnée par Aristote, ἡ τοῦ δυνάμει ὄντος ἐντελέχεια, ᾗ τοιοῦτον, κίνησίς ἐστιν (*Physique* III, 1, 201 a 10-11)] ! Et cependant, s'ils laissent au mot de mouvement son sens ordinaire, comme ils font, ce n'est pas une définition, mais une proposition. Et ainsi, confondant les définitions qu'ils appellent définitions de nom, qui sont les véritables définitions libres, permises et géométriques, avec celles qu'ils appellent définitions de chose, qui sont proprement des propositions nullement libres, mais sujettes à contradiction, ils s'y donnent la liberté d'en former aussi bien que des autres ; et chacun définissant les mêmes choses à sa manière, par une liberté qui est aussi défendue dans ces sortes de définitions que permise dans les premières, ils embrouillent toutes choses et, perdant tout ordre et toute lumière, ils se perdent eux-mêmes et s'égarent dans des embarras inexplicables. (Pascal 1991, pages 399-400.)

En fait, de nombreux mathématiciens et philosophes attribuent à la définition un rôle créateur de la chose définie : voir la section 10.8.2. Cela explique pourquoi Proclus appelle les définitions les « hypothèses » d'une science (voir Caveing 1990, pages 122-123).

10.3 Définir les termes qui définissent l'angle droit.

Pascal écrit encore dans *De l'esprit géométrique* qu'« une méthode encore plus éminente et plus accomplie, mais où les hommes ne sauraient jamais arriver [...] consisterait [...] à définir tous les termes ». Regardons donc l'exemple que nous avons choisi et étudions comment Euclide définit les termes *droite*, *élever*, *angle*, *adjacent*, *égal* au moyen desquels la définition 10 est formulée. Ils font l'objet d'un traitement très varié.

10.3.1 Définir *adjacent* et *élever*.

La construction *élever* est considérée comme parfaitement connue et ne fait pas l'objet d'une définition ou explication.

Il en est de même du prédicat *adjacent* ; en voici cependant une définition par Leibniz (1677).

> Des *angles* sont deux à deux *consécutifs* lorsqu'ils se situent de part et d'autre d'une unique droite et d'un seul côté d'une autre droite située dans le même plan. (Leibniz 1995, fragment II, page 65.)

Il arrive qu'Euclide pourvoie des définitions de relations de situation comme la suivante, issue du troisième livre.

> 9. Quand les droites contenant l'angle découpent une certaine circonférence, l'angle est dit *s'appuyer sur celle-ci*. (Euclide d'Alexandrie 1990-2001, volume 1, page 388.)

10.3.2 Définir *égal*.

L'usage du prédicat *égal* est décrit par la section des « Notions communes » du premier livre des *Éléments*, c'est-à-dire par une liste d'axiomes :

> 1. Les choses égales à une même chose sont égales entre elles.
> 2. Et si, à des choses égales, des choses égales sont ajoutées, les touts sont égaux.
> .....
> 7. Et les choses qui s'ajustent les unes sur les autres sont égales entre elles.
> 8. Et le tout est plus grand que la partie. (Euclide d'Alexandrie 1990-2001, volume 1, pages 178-179.)

C'est-à-dire qu'Euclide énonce les règles qui permettent d'établir l'égalité et l'inégalité de deux choses ; il invoque ces axiomes très souvent, plus de soixante fois dans le premier livre. Ces règles ne peuvent pas être interprétées comme une définition de nom, mais j'y vois la première occurrence d'une "définition implicite" de l'égalité. Voir l'exercice 10.3.2 pour deux autres occurrences. La définition implicite aura une place importante en axiomatique formelle et j'en parlerai plus longuement dans la section 10.8.

10.3.3 Définir *angle*.

L'angle et l'angle rectiligne sont l'objet des définitions 8 et 9.

> 8. Un *angle plan* est l'inclinaison, l'une sur l'autre, dans un plan, de deux lignes qui se touchent l'une l'autre et ne sont pas placées en ligne droite.
> 9. Et quand les lignes contenant l'angle sont droites, l'angle est appelé *rectiligne*. (Euclide d'Alexandrie 1990-2001, volume 1, page 158.)

À première vue, nous pourrions interpréter la définition 8 comme une définition de nom, mais elle repose d'une manière essentielle sur un terme, « inclinaison », qui ne fait l'objet d'aucune explication et ne me semble pas plus connu que le terme à définir. Néanmoins, cette définition précise les conditions sous lesquelles on parlera d'angle :

— il s'agit d'un rapport qu'entretiennent deux lignes ;

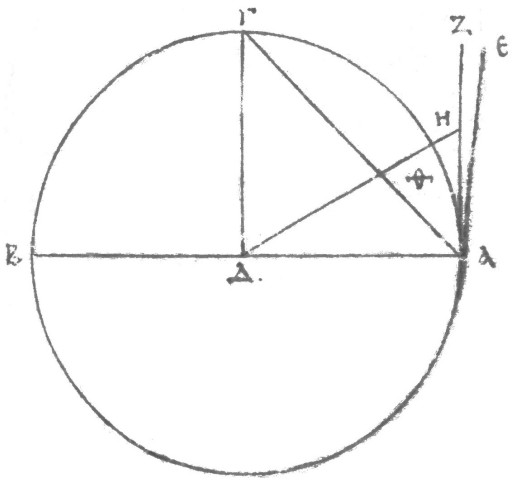
— elles doivent être planes, se rencontrer (mais pas nécessairement en un point qui limite ces lignes) ;

— l'une n'est pas une ligne droite que l'autre prolonge en ligne droite.

Cette dernière condition exclut ce que nous appelons aujourd'hui l'« angle plat » du genre des angles.

L'angle du demi-cercle.

En fait, la définition 8 autorise Euclide à considérer des angles non rectilignes dans la proposition 16 du troisième livre : « l'angle du demi-cercle, celui contenu par la droite BA et la circonférence ΓΘA » ainsi que « l'angle restant contenu par la circonférence ΓΘA et la droite AE », qui est « la droite menée à angles droits avec le diamètre du cercle », c'est-à-dire la demi-tangente au cercle en A (l'angle droit sous ΔAE de la figure ci-contre, extraite du folio 53 du manuscrit MS D'Orville 301, a l'air obtus ; les lettres de la figure concordent bien avec les lettres du texte).



Cette proposition 16 montre que l'inclinaison de deux lignes se constate au voisinage du sommet de l'angle : sa démonstration considère un angle aigu sous ΔAZ et la projection orthogonale H du centre du cercle Δ sur AZ (l'angle droit sous ΔHZ de la figure a l'air encore plus obtus) et établit qu'il est absurde que H soit à l'extérieur du cercle (notons que les figures géométriques absurdes sont plus choquantes que les raisonnements absurdes, mais néanmoins utiles pour suivre ce qui se passe). Or l'angle sous ΔAZ est l'angle sous ΔAH et il est donc plus petit que l'angle du demi-cercle. Dans le langage de la géométrie du 20e siècle, je dirais que l'angle contenu par deux lignes issues d'un point est le *germe*[1] de segments arbitrairement petits de ces deux lignes limités par ce point.

Le fait que le point H puisse se retrouver arbitrairement proche du sommet A a pu inspirer la critique suivante de Sextus Empiricus dans *Contre les géomètres* au 2e siècle.

> [...] ceux qui le décrivent, lorsqu'ils disent que « l'angle est la partie minimale sous l'inclinaison de deux droites non juxtaposées », soit définissent la partie minimale comme un corps sans partie, soit comme ce qu'ils appellent signe, c'est-à-dire le point. [...] de l'angle ils disent qu'il est soit plus grand soit plus petit ; mais rien n'est plus menu qu'un corps minimal, puisque c'est ce plus menu-ci et non ce moins menu-là qui sera minimal. Il reste donc à dire que ⟨l'angle⟩ est pour eux le point ; ce qui ouvre la voie là encore à des apories. (Giovacchini 2010, page 147.)

Si je reprends l'interprétation d'un angle comme germe, aucun des angles qui constituent le germe de cet angle n'est le plus petit, à moins de considérer le point lui-même comme deux segments, ce qui est absurde. Pourtant Proclus rapporte ainsi la définition de l'angle par Apollonius : « l'angle est la contraction [συναγωγή], en un point, d'une surface par une ligne brisée » (voir Caveing 1990, note 47, page 159).

Peut-on invoquer cette définition ?

Arnauld et Nicole montrent dans *La logique, ou l'Art de penser* que le terme *angle* n'est pas utilisé selon sa définition dans le premier livre en appliquant « un remède très sûr et très infaillible » de Pascal : « substituer mentalement la définition à la place du défini » (Pascal 1991, page 394).

> Euclide définit l'angle plan rectiligne, *la rencontre de deux lignes droites inclinées sur un même plan.* Si on considère cette définition comme une simple définition de mot, en sorte qu'on regarde le mot d'*angle* comme ayant été dépouillé de toute signification, pour n'avoir plus que celle de la rencontre de deux lignes, on n'y doit point trouver à redire. Car il a été permis à Euclide d'appeler du mot

1. L'idée du germe est la suivante : on veut considérer les angles au voisinage d'un point, et deux angles sont considérés égaux dès qu'ils coïncident dans un voisinage de ce point. On peut définir cette idée rigoureusement.

d'*angle* la rencontre des deux lignes. Mais il a été obligé de s'en souvenir, et de ne prendre plus le mot d'*angle* qu'en ce sens. Or pour juger s'il l'a fait, il ne faut que substituer toutes les fois qu'il parle de l'*angle*, au mot d'*angle* la définition qu'il a donnée, et si en substituant cette définition il se trouve quelque absurdité en ce qu'il dit de l'angle, il s'ensuivra qu'il n'est pas demeuré dans la même idée qu'il avait désignée ; mais qu'il est passé insensiblement à une autre, qui est celle de la nature. Il enseigne, par exemple, à diviser un angle en deux. Substituez sa définition. Qui ne voit que ce n'est point la rencontre de deux lignes qu'on divise en deux, que ce n'est point la rencontre de deux lignes qui a des côtés, et qui a une base ou soutendante ; mais que tout cela convient à l'espace compris entre les lignes, et non à la rencontre des lignes. (Arnauld et Nicole 2011, IV, IV, pages 538-539.)

En réalité, Euclide n'invoque jamais la définition 8. En particulier, la question de la possibilité de l'angle ne se pose pas.

Les *Commentaires* de Proclus rendent compte de la question embarrassante à quelle catégorie appartient l'angle : est-ce une relation, une qualité ou une quantité ? Sa conclusion est que l'angle est tout cela à la fois.

Je reviens sur la difficulté de définir l'angle à la section 10.5.2.

10.3.4 Définir *droite.*

La ligne droite est l'objet de la définition 4.

4. Une *ligne droite* est celle qui est placée de manière égale par rapport aux points qui sont sur elle. (Euclide d'Alexandrie 1990-2001, volume 1, page 154.)

Cette définition repose aussi de manière essentielle sur une formule, « placé de manière égale par rapport à », qui ne fait l'objet d'aucune explication et ne me semble pas plus connue que le terme à définir. En fait, même la forme logique de cette formule est ambigüe : est-ce un rapport de la droite à chacun de ses points (comme je le pense), ou est-ce un rapport de la droite à ses extrémités ? Nous allons revenir sur cette définition dans la section 10.6.

Euclide n'invoque jamais la définition 4, mais la possibilité des droites est postulée dans la section des « Demandes ».

1. Qu'il soit demandé de mener une ligne droite de tout point à tout point.
2. Et de prolonger continument en ligne droite une ligne droite limitée. (Euclide d'Alexandrie 1990-2001, volume 1, pages 167-168.)

L'usage moderne du terme de droite en fait un objet infini en extension, alors qu'Euclide ne considère que des segments de droite qu'il prolonge au besoin. L'adverbe « continument » apparait ici selon son seul usage dans les *Éléments* : il exprime que la droite que l'on prolonge "tient ensemble" avec le prolongement, de sorte que la limite au-delà de laquelle la droite a été prolongée devient un point ordinaire de la droite prolongée.

10.3.5 Le contexte de la définition de l'angle droit.

En étudiant les termes au moyen desquels l'angle droit est défini, nous avons découvert au fur et à mesure le contexte de sa définition, dans lequel je distingue une composante épistémologique et une composante mathématique.

Voici comment David Hilbert et Paul Bernays rendent compte du contexte épistémologique de la géométrie euclidienne dans leurs *Fondements des mathématiques* (1934).

> [...] l'axiomatique matérielle (inhaltlich) introduit ses concepts fondamentaux en référence à des évènements connus, et [...] ses principes fondamentaux sont, soit présentés comme des faits évidents qu'on peut discerner clairement, soit formulés comme un extrait d'un complexe empirique, donnant par là l'occasion d'exprimer la croyance qu'on est arrivé sur la trace de lois de la nature, tout en ouvrant la perspective d'appuyer cette croyance sur le succès de la théorie. (Hilbert et Bernays 2001, page 56.)

Le contexte mathématique est d'abord celui d'une géométrie où l'égalité se constate uniquement par décomposition et superposition (c'est-à-dire par puzzle, voir cependant le principe d'exhaustion page 73) et dont les fondements se passent du nombre. Puis c'est celui d'un agencement du texte des *Éléments* où les angles sont définis immédiatement après la ligne droite et avant le cercle. Je n'imagine pas d'autre définition de l'angle droit dans ce contexte !

Notons que l'angle droit fournit un étalon d'inclinaison, alors qu'un étalon de longueur ne peut être posé qu'arbitrairement. C'est peut-être une des motivations de la demande 4.

> 4. Et que tous les angles droits soient égaux entre eux. (Euclide d'Alexandrie 1990-2001, volume 1, page 173.)

Nous avons pu observer que c'est la description du contexte qui fournit les clés de la définition. Celle-ci devient claire dans la mesure où son contexte le devient.

10.4 La définition de mot.

Le fait que les définitions 4 et 8 ne sont jamais invoquées pose la question de leur sens : à quoi bon les formuler si elles ne servent pas en tant que telles ? En fait, je constate que ce ne sont pas des définitions mathématiques au sens que leur donnent les cours de mathématiques. Mais selon quelle signification du terme peut-on alors les qualifier de définition, et quelle est leur portée, leur pertinence ?

Notons déjà que les définitions 4 et 9 répondent aux exigences formelles d'Aristote dans la mesure où elles sont composées d'un genre et d'une différence, alors que la définition 8 introduit un genre primitif.

	genre	différence
droite	ligne	placée de manière égale par rapport aux points sur elle
angle rectiligne	angle	contenu par deux droites.

10.4.1 Définir *définition.*

Rappelons avec Bernard Vitrac ce que signifie « définition », le premier mot des *Éléments* d'Euclide :

> ὅρος désigne les bornes de pierre, les limites de propriété, les bornes hypothécaires, les frontières. Par analogie, le mot a pris le sens de « détermination » et de « définition ». C'est le nom des énoncés liminaires que nous sommes en train de commenter. (Euclide d'Alexandrie 1990-2001, volume 1, note 57, page 161.)

Le verbe définir a donc en grec le sens étymologique de délimiter, d'exprimer où commence et où s'arrête un terrain (sens propre) ou l'extension d'une notion (sens figuré). Il a été repris en latin dans ces deux sens (voir le *Dictionnaire latin-français* de Félix Gaffiot, http://micmap.org/dicfro/dictionary/gaffiot-dictionary/mImg/0000482.gif), et on les retrouve dans l'*Oxford English dictionary* et dans le *Dictionnaire du moyen français (1330-1500)*, http://cnrtl.fr/definition/dmf/D%C3%89FINIR2, ainsi que dans le *Trésor de la langue française*, http://cnrtl.fr/definition/d%C3%A9finir, mais le sens propre s'est peu à peu effacé au profit du sens figuré dans le français contemporain.

L'histoire de ce mot est en tous points parallèle à celui du mot "terme", dont voici le sens propre.

> **TERME.** n. m. Borne marquant une limite et faite d'un buste terminé en gaine, en souvenir du dieu Terme qui, chez les Romains, marquait et protégeait les limites des terres. *Planter des termes.* [...] (Académie française 1935, page 650.)

10.4.2 Théorie.

Les définitions 4 et 8 des termes de droite et d'angle ont disparu des ouvrages de géométrie et sont aujourd'hui reléguées dans les dictionnaires. Arnauld et Nicole font de telles définitions l'objet de l'avertissement suivant au chapitre XI du premier livre de leur *Logique* (de 1664).

> Il faut aussi prendre garde de ne pas confondre la définition de nom dont nous parlons ici, avec celle dont parlent quelques philosophes, qui entendent par là l'explication de ce qu'un mot signifie selon l'usage ordinaire d'une langue, ou selon son étymologie. (Arnauld et Nicole 2011, I, XI, page 233.)

Puis ils leur consacrent le dernier chapitre du premier livre.

> CHAPITRE XIII. D'une autre sorte de définitions de noms, par lesquels on marque ce qu'ils signifient dans l'usage.
>
> Tout ce que nous avons dit des définitions de noms ne se doit entendre que de celles où l'on définit les mots dont on se sert en particulier : et c'est ce qui les rend libres et arbitraires, parce qu'il est permis à chacun de se servir de tel son qu'il lui plait pour exprimer ses idées, pourvu qu'il en avertisse. Mais comme les hommes ne sont maitres que de leur langage, et non pas de celui des autres, chacun a bien droit de faire un dictionnaire pour soi ; mais on n'a pas droit d'en faire pour les autres, ni d'expliquer leurs paroles par les significations particulières qu'on aura attachées aux mots. C'est pourquoi quand on n'a pas dessein de faire connaitre simplement en quel sens on prend un mot, mais qu'on prétend expliquer celui auquel il est communément pris, les définitions qu'on en donne ne sont nullement arbitraires ; mais elles sont liées et astreintes à représenter non la vérité des choses, mais la vérité de l'usage, et on les doit estimer fausses, si elles n'expriment pas véritablement cet usage, c'est-à-dire si elles ne joignent pas aux sons les mêmes idées qui y sont jointes par l'usage ordinaire de ceux qui s'en servent. Et c'est ce qui fait voir aussi que ces définitions ne sont nullement exemptes d'être contestées, puisque l'on dispute tous les jours de la signification que l'usage donne aux termes.
>
> Or quoique ces sortes de définitions de mots semblent être le partage des grammairiens, puisque ce sont celles qui composent les dictionnaires, qui ne sont autre chose que l'explication des idées que les hommes sont convenus de lier à certains sons, néanmoins l'on peut faire sur ce sujet plusieurs réflexions très importantes pour l'exactitude de nos jugements.

> La première, qui sert de fondement aux autres, est que les hommes ne considèrent pas souvent toute la signification des mots, c'est-à-dire que les mots signifient souvent plus qu'il ne semble, et que lorsqu'on en veut expliquer la signification, on ne représente pas toute l'impression qu'ils font dans l'esprit. (Arnauld et Nicole 2011, I, XIII, pages 245-247.)

Les définitions de mot ont en commun avec les définitions de chose qu'elles ne sont pas arbitraires. Elles diffèrent par leur validité et leur mode de vérification : les premières rendent compte de l'usage alors que les deuxièmes recherchent le vrai, de sorte que les premières sont appuyées par des exemples et des citations, alors que les deuxièmes appellent à une justification : voir l'exercice 10.4.1

10.5 Définition du mot *angle* dans les dictionnaires.

Nous allons consulter la 1re, 8e et 9e édition du *Dictionnaire* de l'Académie française, datant respectivement de 1694, 1932 et 1992. Ce dictionnaire

> ne cite point, parce que plusieurs de nos plus célèbres Orateurs et de nos plus grands Poètes y ont travaillé, et qu'on a cru s'en devoir tenir à leurs sentiments. [...]
>
> [L'Académie] a donné la Définition de tous les mots communs de la Langue dont les Idées sont fort simples ; et cela est beaucoup plus malaisé que de définir les mots des Arts et des Sciences dont les Idées sont fort composées ; Car il est bien plus aisé, par exemple, de définir le mot de *Télescope*, qui est *une Lunette à voir de loin*, que de définir le mot de *voir* ; Et l'on éprouve même en définissant ces termes des Arts et des Sciences, que la Définition est toujours plus claire que la chose définie ; au lieu qu'en définissant les termes communs, la chose définie est toujours plus claire que la Définition. Ainsi quoique Aristote ait fait une définition excellente quand il a défini l'homme *Animal Raisonnable*, il est constant néanmoins que le mot *Homme* nous représente mieux ce qu'il signifie que cette définition. (Académie françoise 1694, préface.)

Ce dictionnaire pourvoit donc des exemples fabriqués par les Académiciens. Ils s'expriment sur leur utilité dans la 7e édition.

> C'est par des exemples nombreux et bien choisis que l'Académie, depuis qu'elle s'occupe du dictionnaire, s'est efforcée de remédier à cette nécessaire insuffisance des définitions. Les exemples, en plaçant successivement un mot sous tous ses jours, corrigent et rectifient ce que la définition a d'incertain et de trop vague dans ses termes généraux, et conduisent, en quelque sorte, naturellement l'esprit d'un sens au sens voisin par une gradation insensible. À un coup d'œil superficiel, on serait tenté de croire peut-être que l'Académie multiplie trop les exemples, tant ils semblent quelquefois différer peu les uns des autres ; un examen plus attentif fait revenir vite de cette erreur. Les exemples sont la vraie richesse et la partie la plus utile du dictionnaire. C'est là qu'avec un peu de patience le lecteur est toujours sûr de trouver ce qu'il cherche, soit qu'il ait des doutes sur la justesse et la propriété d'un terme, soit que le sens même d'une expression lui échappe. (Académie française 1878, page VII.)

Voici les définitions successives du mot *angle*.

ANGLE. s. m. Inclination de deux lignes qui aboutissent à un même point. *Angle droit. angle aigu. angle obtus. angle de tant de degrés. cette muraille fait un grand angle. angle saillant. angle rentrant. l'angle du centre. l'angle de la circonférence. une figure à plusieurs angles.* (Académie françoise 1694, page 40.)

ANGLE. n. m. T. de Géométrie. Ouverture de deux lignes qui se rencontrent en un point, degré d'inclinaison qu'elles ont l'une à l'égard de l'autre. *Angle droit. Angle aigu. Angle obtus. Angle de quarante-cinq degrés. Angle de cent degrés. Angle saillant. Angle rentrant. Angle rectiligne, curviligne. Une figure à plusieurs angles. Angle optique. Angle visuel. Angle de réflexion, de réfraction. Angle d'incidence. Sommet, côtés d'un angle.* [...] (Académie française 1932, page 57.)

ANGLE n. m. XII^e^ siècle. Du latin *angulus*, « angle, coin ».
1. GÉOM. Figure formée par deux demi-droites, deux courbes planes ou deux demi-plans qui se coupent. *Le sommet, les côtés d'un angle. Le grade, le degré, la minute, la seconde sont des unités d'angle. Dans le Système international, l'ouverture d'un angle plan se mesure en radians. Angle droit,* dont les côtés sont perpendiculaires entre eux. *Angle aigu,* plus petit que l'angle droit. *Angle obtus,* plus grand que l'angle droit. *Angles complémentaires,* dont la somme vaut un angle droit. *Angles supplémentaires,* dont la somme vaut deux angles droits. *Angles adjacents,* ayant même sommet et un côté commun. *Angle dièdre,* formé de deux demi-plans qui se coupent, limités à leur intersection. *Angle solide,* portion d'espace intérieure à un cône. *L'unité d'angle solide est le stéradian. Angle horaire d'un astre,* une des coordonnées qui caractérisent, à un instant donné, la position de l'astre par rapport au méridien du lieu. [...] (Académie française 1992, page 89.)

Cette dernière entrée renvoie successivement aux entrées suivantes pour la définition de l'angle droit.

PERPENDICULAIRE adj. XIV^e^ siècle, *perpendiculer* ; XV^e^ siècle, *perpendiculaire*. Emprunté du latin *perpendicularis*, de même sens, lui-même tiré de *perpendiculum*, « fil à plomb ».
1. GÉOM. Se dit de droites ou de plans qui forment entre eux un angle de 90°. (Académie française 2011, page 327.)

DEGRÉ n. m. XI^e^ siècle, au sens de « marche d'escalier ». Probablement composé latin tardif de la préposition *de*, marquant un mouvement de haut en bas, et *gradus*, « pas, marche (d'un escalier), hiérarchie, rang », de *gradi*, « marcher, s'avancer ».

. . . . . .

IV. Unité de mesures scientifiques. **1.** Chacune des divisions égales d'une circonférence ; unité de mesure des angles formés par les rayons d'un cercle (abréviation °). GÉOM. *Le degré centésimal* ou *grade est la quatre-centième partie de la circonférence ; le degré sexagésimal,* ou simplement *degré, en est la trois-cent-soixantième partie. 360° = 400 grades ou 2 π radians. Le quart de cercle comprend quatre-vingt-dix degrés.* [...] (Académie française 1992, page 622.)

Le *Dictionnaire* ne précise pas le rapport entre angle et circonférence de cercle (voir l'exercice 10.5.1) ; quant au « quart de cercle », il renvoie à une application du problème suivant, traité dans le troisième livre des *Éléments* d'Euclide.

> 30. Couper une circonférence donnée en deux parties égales.

Cependant, la démonstration de ce problème repose de manière essentielle sur le problème 11 du premier livre, que nous avons vu dans la section 10.1.2 !

10.5.1 Définition de mot ?

La 1re et la 8e édition s'inspirent de la définition 8 d'Euclide et en font une définition de mot. Les deux éditions montrent que comme Arnauld et Nicole (voir la section 10.3.3 et l'exercice 10.5.1), les Académiciens avaient pour souci de voir dans l'angle une quantité et non pas une qualité : voir la section 10.3.3.

10.5.2 Définition de nom ?

La 9e édition rompt avec la tradition euclidienne : j'interprète la définition qu'elle donne comme une définition de nom et je l'analyse comme suit :

definiendum	*definiens*
angle	figure formée par deux demi-droites, deux courbes planes ou deux demi-plans qui se coupent.

Consultons la définition du mot *figure* dans la même édition :

> **FIGURE** n. f. IXe siècle. Emprunté du latin *figura*, « forme, aspect, représentation sculptée, mode d'expression, manière d'être », dérivé de *fingere*, « façonner ».
>
> . . . . . .
>
> **III.** Combinaison d'éléments divers dessinant une forme, s'organisant en un motif. **1.** Représentation graphique de lignes, de surfaces, de volumes. *Le cercle, le trapèze sont des figures géométriques. Tracer la figure d'un cône.* (Académie française 2000, page 129.)

C'est précisément en ce sens que j'ai utilisé ce mot dans les sections 10.1.2 et 10.3.3 ; les Grecs utilisent dans ce cas le mot διάγραμμα.

Mais Arnaud et Nicole ne donnaient certainement pas ce sens au mot *figure* dans l'extrait cité dans la section 10.2.3 : je constate qu'il est incohérent avec l'usage du mot dans la définition d'*angle*. En effet, un terme géométrique ne peut être identifié à une de ses représentations graphiques.

Essayons d'appliquer la définition euclidienne du mot *figure*, qui correspond au mot grec σχῆμα.

> 13. Une *frontière* est ce qui est limite de quelque chose.
> 14. Une *figure* est ce qui est contenu par quelque ou quelques frontière(s). (Euclide d'Alexandrie 1990-2001, volume 1, page 161.)

Je constate que la « figure formée par deux demi-droites » est illimitée et que la définition de l'angle dans la 9e édition ne pourrait donc pas être insérée après la définition 14 des *Éléments*. J'objecte aussi que la notion d'angle est censée être plus élémentaire que celle de figure.

Cette définition ne permet pas non plus d'exprimer la proposition 16 du troisième livre puisque si on considère l'angle rectiligne aigu sous ΔAZ et l'angle du demi-cercle comme des figures, elles ne satisfont plus aucune relation de la partie au tout.

On pourrait considérer les « deux demi-droites » comme la figure elle-même et non comme une frontière. Mais alors cette définition ne permet plus une lecture quantitative, contrairement à la définition euclidienne (un angle est d'autant plus petit que deux lignes sont plus inclinées l'une vers l'autre), et oblige à définir séparément une relation d'ordre : voir la section 10.3.3.

10.6 Définition du mot *droite* dans les dictionnaires.

Voici à présent la définition du mot *droite* dans la 1re, 8e et 9e édition du *Dictionnaire* de l'Académie française.

DROIT, OITE. adj. Qui n'est pas courbé, qui ne penche ni de côté ni d'autre. *Ligne droite. droit comme un jonc. de droit fil. en droite ligne. il vient en droite ligne d'un tel. tenir la tête droite. petit garçon, tenez-vous droit.*

On dit, *Le droit chemin. le plus droit chemin*, pour dire, Le plus court.

On dit, qu'*Un homme est droit comme un cierge, comme un jonc*, pour dire, qu'Il se tient fort droit.

Il sign. quelquefois, Qui n'est pas couché, qui est debout. *Se tenir droit sur ses pieds. demeurer droit en son séant. cette statue paraitra bien plus belle quand elle sera droite, qu'à présent qu'elle est couchée.* [...] (Académie françoise 1694, page 350.)

DROIT, OITE. adj. Qui n'est pas courbe, qui va d'un point à un autre par le plus court chemin. *Cette rue est toute droite. De droit fil. Avoir la taille droite et bien prise. La rivière est droite depuis tel village jusqu'à telle ville. Voilà le droit chemin, le plus droit chemin.*

En termes de Géométrie, *Ligne droite*, ou par ellipse, comme nom féminin, *La droite*, Le plus court chemin d'un point à un autre.

Être droit comme un I, se dit de Quelqu'un qui se tient très droit.

Fig., *La droite voie*, en termes de Dévotion, La voie du salut. *La voie droite*, en termes de Morale, La voie du bien, de la vertu.

Il signifie aussi Qui est perpendiculaire à l'horizon, qui ne penche d'aucun côté. *Ce mur n'est pas droit, il penche.* [...] (Académie française 1932, page 422.)

DROIT, DROITE adj., adv. et n. XIe siècle, *dreit*, « juste, vrai, exact ». Du latin *directus*, « qui est en ligne droite, à angle droit » ; au figuré, « direct, sans détour, juste ».

I. Adj. Qui n'est ni courbe ni incliné. **1.** Qui s'étend sans déviation ni inflexion d'un point à un autre. *Un tronc d'arbre droit comme un mât. Une rue droite. Cette règle millimétrée n'est pas tout à fait droite. Pour qu'il soit parfaitement droit, l'alignement des peupliers a été tracé au cordeau. Être droit comme un I. Se tenir bien droit, très droit.* Spécialt. *Ligne droite. La ligne droite peut être matérialisée par un fil tendu entre deux points. La ligne droite est le plus court chemin d'un point à un autre.*

. . . . . .

III. N. **1.** N. f. GÉOM. Une droite, une ligne telle que deux des points par lesquels elle passe sont nécessaires et suffisants pour la définir. *Segment de droite*, portion de droite limitée par deux points. *Décrire une droite*, la tracer. *Deux droites parallèles ne se rencontrent jamais. Droites concourantes*, qui convergent vers un même point. (Académie française 1992, pages 739-740.)

10.6.1 Définition de mot ?

Je constate que la 1re et la 8e édition concordent sur la signification que l'usage donne au terme *droite* et qu'elles en donnent une définition de mot.

La 9e édition propose une définition de l'adjectif *droit* dans le même esprit (sauf qu'elle produit un *I* qui est incliné !), mais rajoute une définition du nom *droite* d'une nature complètement différente : on peut y reconnaitre la première demande citée dans la section 10.3.4 pour la suffisance et la notion commune 9 ci-dessous pour la nécessité.

> 9. Et deux droites ne contiennent pas une aire. (Euclide d'Alexandrie 1990-2001, volume 1, page 179.)

Cette 9e édition proposerait donc une définition à partir d'énoncés qu'Euclide a rangés parmi les postulats et axiomes ! Nous allons y revenir dans la section 10.8. Mais on peut y reconnaitre aussi la définition de la droite par Leibniz (voir Bkouche 2009, page 181) :

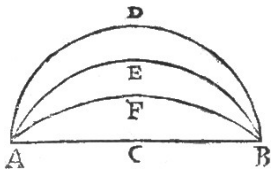
> Ce qui est déterminé par la donnée de deux points est l'extensum le plus simple passant par eux, que nous appellerons droite. (Leibniz 1995, fragment XIV, page 279.)

10.6.2 Commentaire de la définition euclidienne.

Toutes les éditions du *Dictionnaire* évitent la formule « de manière égale par rapport à » de la définition 4. On constate en fait que la définition de la droite par Euclide était déjà obscure pour Proclus au cinquième siècle de notre ère. Plutôt que de lire son commentaire, lisons celui de Christoph Clavius (1589, pages 30-31) dans la deuxième édition de sa traduction latine des *Éléments*, selon la version de Denis Henrion (1632). On y retrouve les références de Proclus : la définition "métrique" par Archimède (voir lignes 10-12), la définition "optique" par Platon (voir lignes 12-18) et la définition "mécanique" par Geminos comme « flux uniforme et privé d'inclinaison latérale d'un point » selon un « mouvement uniforme et minimal » (ma traduction de Proclus, voir lignes 19-26).

---

4. La ligne droite, est celle qui est également comprise et étendue entre ses points.



Les Mathématiciens ont de trois sortes de lignes, c'est à savoir la ligne droite, la ligne circulaire, qu'ils appellent aussi ligne courbe, et la ligne mixte : Euclide définit ici la droite, laquelle il dit être celle-là qui est également étendue entre ses points : ainsi la ligne ACB est dite ligne droite, pour ce que tous les points entremoyens d'icelle ligne, comme C, sont également posés entre les extrêmes A et B, l'un n'étant plus élevé ou abaissé que l'autre : ce qui n'advient aux trois autres lignes ADB, AEB, AFB, car il est manifeste que les points entremoyens D, E, F sont bien plus élevés que les extrêmes A et B. Quelques autres Auteurs ont diversement défini la ligne droite : car Campanus dit, que c'est le plus court chemin d'un point jusqu'à un autre : et, selon Archimède, la ligne droite est la plus courte de toutes celles qui ont mêmes extrémités. Mais Platon dit que c'est celle-là dont les points du milieu ombragent les extrêmes : comme par exemple, si en la ligne ACB, le point extrême A avait la vertu d'illuminer, et le point du milieu C la force de cacher : icelui point C empêcherait que le point extrême B fût illuminé de l'autre extrême A : Et aussi l'œil étant au point extrême A, il ne pourrait voir l'autre extrême B, à cause du point C posé entre iceux extrêmes : ce qui n'arriverait pas aux lignes non droites, comme le démontrent les lignes ADB, AEB, et AFB.

Or tout ainsi que les Mathématiciens conçoivent la ligne être décrite par le flux et mouvement imaginaire du point, ainsi aussi entendent-ils la qualité de la ligne décrite par

la qualité d'icelui mouvement : car si on entend que le point coule droit par le plus court chemin, ne se détournant çà ni là, la ligne ainsi décrite sera appelée ligne droite : mais si le point fluant vacille en son mouvement, et s'écarte çà et là, la ligne décrite sera appelée mixte : et finalement si le point fluant ne vacille en son mouvement, mais est porté en rond d'un certain mouvement uniforme et régulier, gardant toujours une égale distance à quelque certain point à l'entour duquel il est porté, cette ligne décrite sera appelée circulaire. Or Euclide ne traite ici que de deux simples lignes, savoir est de la droite et de la circulaire. Il a défini celle-là ci-dessus, et il définira cette-ci à la 15. déf. Mais quant à la mixte, il en omet la définition, pour ce qu'elle n'a aucun usage en ses éléments Géométriques : il y en a de plusieurs sortes, et d'icelles traitent amplement Apollonius Pergeus, Nicomède, Archimède et autres Auteurs. (Henrion 1632, pages 2-3.)

Je retiens de ce commentaire que la définition d'Euclide est une tentative désespérée de trouver une formulation positive pour "non courbé". Les mots « de manière égale » traduisent l'expression grecque ἐξ ἴσου, qui est en fait bien rendue par l'expression latine *ex aequo* : comme le dit Henrion, la ligne droite est la ligne décrite par un point « ne se détournant çà ni là », c'est-à-dire selon un « mouvement » qui s'écoule de manière égale.

Je conclus donc au titre d'une telle interprétation que la définition 4 réalise une définition de mot.

10.6.3 Définition de chose ?

Cependant, les mathématiques du 19e siècle permettent de prouver cette interprétation dans le cadre de la discipline qui traite des lignes et des surfaces courbes, c'est-à-dire la géométrie différentielle. En effet, on y prouve qu'une géodésique (c'est-à-dire une ligne de plus court chemin) ne se courbe que dans la mesure où se courbe la surface sur laquelle elle est décrite ; or la surface plane est de courbure nulle et la droite est une géodésique du plan par la proposition 20 du premier livre des *Éléments*. J'estime que cette preuve, jointe à la définition 4, permet de formuler une définition de chose de la droite !

Mais ce que montre l'histoire des géométries non euclidiennes, c'est que la véritable définition de chose de la droite dans la géométrie euclidienne résulte de la cinquième demande, c'est-à-dire du *postulat des parallèles* :

> 5. Et que, si une droite tombant sur deux droites fait les angles intérieurs et du même côté plus petits que deux droits, les deux droites, indéfiniment prolongées, se rencontrent du côté où sont les angles plus petits que deux droits. (Euclide d'Alexandrie 1990-2001, volume 1, page 175.)

En effet, il s'avère que c'est lui qui oblige deux droites à toujours garder leur inclinaison respective (définissons-la comme la différence entre deux angles droits et la somme des angles intérieurs et du même côté faits par une droite tombant sur elles) et donc à ne pas se courber l'une par rapport à l'autre : voir l'exercice 10.6.1.

10.7 L'impossibilité de tout définir.

Notre enquête sur la définition de l'angle droit nous a menés à la définition des cinq termes de la section 10.3, puis nous nous sommes éloignés du discours mathématique pour chercher conseil auprès des lexicographes. Or les dictionnaires ont pour vocation de définir *tous* les mots. Est-ce envisageable pour les définitions de nom ? Voici comment Pascal introduit son opuscule.

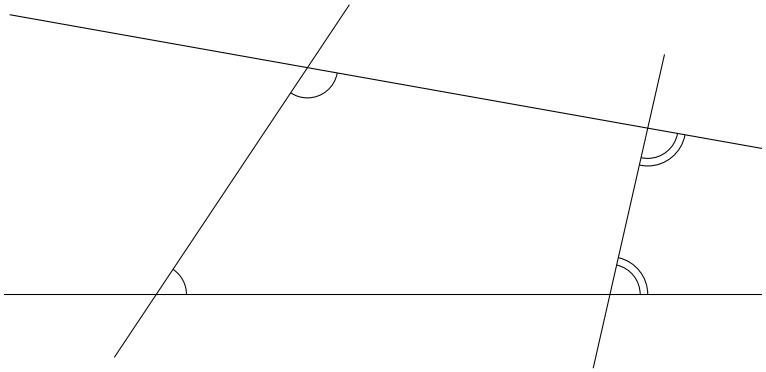


FIGURE 10.6.1 – Deux inclinaisons respectives de deux droites.

> Mais il faut auparavant que je donne l'idée d'une méthode encore plus éminente et plus accomplie, mais où les hommes ne sauraient jamais arriver : car ce qui passe la géométrie nous surpasse ; et néanmoins il est nécessaire d'en dire quelque chose, quoiqu'il soit impossible de le pratiquer.
>
> Cette véritable méthode, qui formerait les démonstrations dans la plus haute excellence, s'il était possible d'y arriver, consisterait en deux choses principales : l'une, de n'employer aucun terme dont on n'eût auparavant expliqué nettement le sens ; l'autre, de n'avancer jamais aucune proposition qu'on ne démontrât par des vérités déjà connues ; c'est-à-dire, en un mot, à définir tous les termes et à prouver toutes les propositions. (Pascal 1991, page 393.)

Pourquoi est-il impossible de tout définir ? Pascal l'explique après avoir donné sa théorie de la définition de nom.

> Certainement cette méthode serait belle, mais elle est absolument impossible : car il est évident que les premiers termes qu'on voudrait définir en supposeraient de précédents pour servir à leur explication, et que de même les premières propositions qu'on voudrait prouver en supposeraient d'autres qui les précédassent ; et ainsi il est clair qu'on n'arriverait jamais aux premières.
>
> Aussi, en poussant les recherches de plus en plus, on arrive nécessairement à des mots primitifs qu'on ne peut plus définir, et à des principes si clairs qu'on n'en trouve plus qui le soient davantage pour servir à leur preuve. (Pascal 1991, page 395.)

En d'autres mots, cet idéal est inatteignable parce qu'on s'engagerait dans une "régression à l'infini". Il y a donc des termes primitifs qui sont indéfinissables. Voici comment Pascal exprime les conséquences de ce constat.

> On trouvera peut-être étrange que la géométrie ne puisse définir aucune des choses qu'elle a pour principaux objets. Car elle ne peut définir ni le mouvement, ni les nombres, ni l'espace ; et cependant ces trois choses sont celles qu'elle considère particulièrement, et selon la recherche desquelles elle prend ces trois différents noms de mécanique, d'arithmétique, de géométrie, ce dernier nom appartenant au genre et à l'espèce.
>
> Mais on n'en sera pas surpris si l'on remarque que cette admirable science, ne s'attachant qu'aux choses les plus simples, cette même qualité qui les rend dignes d'être ses objets les rend incapables d'être définies ; de sorte que le manque de définition est plutôt une perfection qu'un défaut, parce qu'il ne vient pas de leur obscurité, mais au contraire de leur extrême évidence, qui est telle qu'encore qu'elle n'ait pas la conviction des démonstrations, elle en a toute la certitude. Elle suppose donc que l'on sait quelle est la chose qu'on entend par ces mots : *mouvement, nombre,*

> *espace* ; et, sans s'arrêter à les définir inutilement, elle en pénètre la nature, et en découvre les merveilleuses propriétés. (Pascal 1991, page 401.)

Pascal fait donc appel à une « extrême évidence » pour les termes primitifs : ils réfèrent pour tous au même concept. Il y reviendra dans les *Pensées* pour exprimer ce qu'elle a d'incertain : voici comment Antony McKenna en rend compte.

> Ainsi ces *idées*, dont la conformité chez les hommes n'est qu'une *puissante conjecture*, s'appuient sur des principes que l'esprit humain ne saurait saisir, parce que « l'âme est jetée dans le corps ». Notre incapacité de douter de ces principes découle ainsi du corps qui nous empêche de voir au-delà. (McKenna 2001, page 354.)

Chacun peut avoir une opinion différente sur un même concept : redonnons la parole à Pascal.

> On voit assez de là qu'il y a des mots incapables d'être définis ; et si la nature n'avait suppléé à ce défaut par une idée pareille qu'elle a donnée à tous les hommes, toutes nos expressions seraient confuses ; au lieu qu'on en use avec la même assurance et la même certitude que s'ils étaient expliqués d'une manière parfaitement exempte d'équivoques ; parce que la nature nous en a elle-même donné, sans paroles, une intelligence plus nette que celle que l'art nous acquiert par nos explications.
>
> Ce n'est pas que tous les hommes aient la même idée de l'essence des choses que je dis qu'il est impossible et inutile de définir.
>
> Car, par exemple, le temps est de cette sorte. Qui le pourra définir ? Et pourquoi l'entreprendre, puisque tous les hommes conçoivent ce qu'on veut dire en parlant de temps, sans qu'on le désigne davantage ? Cependant il y a bien différentes opinions touchant l'essence du temps. Les uns disent que c'est le mouvement d'une chose créée ; les autres, la mesure du mouvement, etc. Aussi ce n'est pas la nature de ces choses que je dis qui est commune à tous ; ce n'est simplement que le rapport entre le nom et la chose ; en sorte qu'à cette expression, temps, tous portent la pensée vers le même objet : ce qui suffit pour faire que ce terme n'ait point besoin d'être défini, quoique ensuite, en examinant ce que c'est que le temps, on vienne à différer de sentiment après s'être mis à y penser. Car les définitions ne sont faites que pour désigner les choses que l'on nomme, et non pas pour en montrer la nature. (Pascal 1991, pages 397-398.)

10.8 Définition et axiomatique formelle.

10.8.1 Moritz Pasch.

Je me suis cantonné jusqu'à présent au cadre épistémologique euclidien auquel David Hilbert et Paul Bernays ont donné le nom d'*axiomatique matérielle* : le discours des *Éléments* fait implicitement appel à l'évidence que la géométrie provient de l'espace physique. C'est particulièrement le cas pour tout ce qui concerne les relations d'ordre : être de part et d'autre d'un point donné sur une droite donnée, être de part et d'autre d'une droite donnée dans un plan donné, etc.

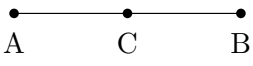
Vers la fin du 19^{e} siècle, des mathématiciens comme Moritz Pasch ont été amenés à envisager d'une nouvelle manière le rapport des mathématiques au réel. Celui-ci commence ainsi la préface à ses *Leçons de géométrie nouvelle* (1882).

> Les tentatives passées de mettre les parties fondamentales de la géométrie dans une forme qui répond aux exigences devenues strictes avec le temps n'ont pas dégagé l'origine empirique de la géométrie avec entière détermination. [...] en se restreignant d'avance au noyau empirique, on conserve à la géométrie le caractère de science naturelle, des autres parties de laquelle elle se distingue en ce qu'il ne lui faut prélever de l'expérience qu'un très petit nombre de concepts et de lois. (Pasch 1882, page III, ma traduction.)

Pasch isole ce noyau empirique par une analyse des tournures qui rendent compte de la géométrie dans la langue naturelle : l'évidence est explicitée et devient consciente. Il commence par la ligne droite et isole deux notions fondamentales.

> Nous allons d'abord nous occuper de la ligne droite. On dit : par deux points on peut tirer une ligne droite. Mais la ligne peut être diversement limitée ; l'indétermination de la ligne droite a mené à ce qu'on dit de la ligne qu'elle n'est pas limitée, qu'il faut se la "représenter" illimitée, d'étendue infinie. Cette demande ne correspond à aucun objet perceptible ; c'est plutôt seulement la ligne droite bien délimitée, le chemin droit entre deux points, le segment droit qui est appréhendé immédiatement à partir des perceptions. Nous voulons garder cette dernière expression et parlons
>
> 1) du segment droit tracé entre deux points,
> 2) de points situés à l'intérieur d'un segment droit.
>
> Toutes les tournures qui apparaissent dans ce paragraphe peuvent être ramenées à ces deux tournures données. [...] (Pasch 1882, page 4, ma traduction.)

Puis Pasch propose de formaliser quelques évidences.

> Les observations les plus primitives sur les segments droits et leurs points livrent une série de relations ; une partie d'entre elles constitue le contenu des axiomes de ce paragraphe. Quels que soient les points A et B supposés (dans les limites étudiées plus précisément à la fin de ce paragraphe), on peut toujours relier A avec B par un segment droit ; mais on ne peut l'effectuer de plusieurs manières. On peut supposer un point C à l'intérieur du segment. On peut tirer un segment droit de A à C ; celui-ci ne passe pas par B ; mais il tombe avec tous ses points dans le segment précédent. Donc, lorsqu'on relie A avec C et C avec B, on ne rencontre aucun point qu'on ne trouvait pas déjà dans le premier segment ; mais les points du premier segment seront eux aussi épuisés par ces deux segments-là. Les axiomes I.–V. restituent ces remarques.
>
> 
>
>
> Axiome **I.** – Entre deux points on peut toujours tirer un segment droit, et un seul.
>
> Par conséquent, l'indication des points terminaux suffit à désigner le segment. Le segment de A à B sera désigné par AB ou BA.
>
> Axiome **II.** – On peut toujours indiquer un point situé à l'intérieur d'un segment droit donné.
>
> Axiome **III.** – Si le point C est situé à l'intérieur du segment AB, alors le point A est situé hors du segment BC.
>
> De même, le point B est situé hors du segment AC.
>
> Axiome **IV.** – Si le point C est situé à l'intérieur du segment AB, alors tous les points du segment AC sont à la fois points du segment AB.
>
> Ou : si le point C est situé à l'intérieur du segment AB et le point D à l'intérieur du segment AC ou BC, alors D est situé aussi à l'intérieur du segment AB.
>
> Axiome **V.** – Si le point C est situé à l'intérieur du segment AB, alors un point qui n'appartient à aucun des segments AC et BC ne peut appartenir au segment AB.

Ou : si les points C et D sont situés à l'intérieur du segment AB et le point D hors du segment AC, alors le point D est situé à l'intérieur du segment BC.

. . . . . .

Formuler une assertion aussi triviale que celle contenue p. ex. dans le troisième axiome peut facilement être considéré comme inutile. Mais elle a été appliquée dans les démonstrations ci-dessus, et nous nous proposons de rendre compte de tous les arguments, même des plus insignifiants. (Pasch 1882, pages 5-6, ma traduction.)

Je constate que la démarche de Pasch est conforme aux conclusions de l'opuscule de Pascal : point et segment droit sont des termes primitifs et ne sont pas définis ; leurs propriétés, c'est-à-dire les relations qu'ils entretiennent, sont posées dans les axiomes. Toute référence à l'espace physique est proscrite après n'avoir « prélevé de l'expérience qu'un très petit nombre de concepts et de lois. »

La méthode de Pasch renvoie à deux nouvelles notions, celle de *type* d'objet (« point », « segment droit ») et celle de *relation* (point « terminal » d'un segment droit, point « à l'intérieur » d'un segment droit) qui ne sont pas définies tout en faisant partie des prérequis de sa méthode.

Les axiomes définissent donc les relations entre les termes primitifs. De là à affirmer qu'ils définissent le point et le segment droit, il y a un pas que Pasch ne franchit pas.

Dans la suite de son ouvrage, Pasch ne propose qu'une définition implicite de la ligne droite AB en introduisant le prédicat d'alignement avec A et B, c'est-à-dire d'« être situé dans la ligne droite des points A et B ».

10.8.2 David Hilbert.

Toutes les éditions des *Grundlagen der Geometrie* de David Hilbert concordent sur l'introduction de la relation "entre".

§ 3. **Die Axiomgruppe II : Axiome der Anordnung.**

Die Axiome dieser Gruppe definieren den Begriff "zwischen" und ermöglichen auf Grund dieses Begriffes die *Anordnung* der Punkte auf einer Geraden, in einer Ebene und im Raume.

Erklärung. Die Punkte einer Geraden stehen in gewissen Beziehungen zu einander, zu deren Beschreibung uns insbesondere das Wort "*zwischen*" dient. (Hilbert 1899, page 6.)

Cette introduction est suivie de la donnée d'axiomes qui simplifient les axiomes de Pasch. Léonce Laugel propose la traduction suivante de ce passage, en accord avec l'auteur qui en a révisé les épreuves (voir Hallett et Majer 2004, pages 413-414).

§ 3. **Le groupe d'axiomes II : Axiomes de distribution.**

Les axiomes de ce groupe définissent l'idée exprimée par le mot « ENTRE » et permettent, en se basant sur cette idée, d'effectuer la *distribution* des points sur une droite, dans un plan et dans l'espace.

CONVENTION. — Les points d'une droite ont entre eux une certaine relation qui s'exprime en particulier au moyen du mot « entre ». (Hilbert 1900, pages 106-107.)

Paul Rossier en propose une traduction très différente.

3. *Deuxième groupe d'axiomes : ordre.*

Les axiomes de ce groupe définissent le terme « entre » ; si l'on s'appuie sur la relation ainsi déterminée, ils permettent d'établir l'ordre des points alignés, coplanaires ou situés dans l'espace.

Définition. Entre les points d'une droite, il existe une relation dans la description de laquelle figure le mot « entre ». (Hilbert 1971, page 14.)

Le mot « Erklärung » est donc traduit successivement par « convention » et « définition ». En fait, dans la première édition, Hilbert réserve ce mot, que je traduirais littéralement par "déclaration", à l'introduction des termes primitifs de la géométrie et utilise le mot « Definition » pour les définitions de nom ; dès la deuxième édition, il abandonne cette distinction pour ne plus utiliser qu'« Erklärung » (voir Hallett et Majer 2004, page 421). Il ne prend par contre aucune précaution pour le verbe « definieren ». La première traduction montre l'embarras que cela cause à Laugel. Celui-ci se tire d'affaire en déplaçant l'acte de définition vers le domaine philosophique des idées : Hilbert ne met-il pas en exergue de son livre la phrase suivante de la *Critique de la raison pure*, « Toute science humaine commence par les intuitions, de là passe aux notions [Begriffe] et finit par les idées » ?

Gottlob Frege critique la démarche de Hilbert dès 1899.

Je commencerai par une formule de Thomae sur votre explication [Erklärung] du § 3. Il disait en substance : « Ce n'est pas une définition, car on ne donne pas de caractéristiques *(Merkmale)* permettant de reconnaitre si la relation « entre » est réalisée. » Je ne peux pas, moi non plus, la tenir pour une définition ; du reste vous ne la nommez pas ainsi, et vous parlez d'explication. Vous utilisez visiblement les deux expressions « explication » et « définition » pour désigner deux choses distinctes, mais la différence nous échappe. (Frege et Hilbert 1992, page 221.)

Nous sommes donc en droit de lire qu'Hilbert *définit* les points, les droites et les plans lorsqu'il écrit ceci.

Convention. — Concevons trois différents systèmes d'êtres : les êtres du PREMIER système, nous les nommerons *points* et nous les désignerons par A, B, C, ... ; les êtres du DEUXIÈME système, nous les nommerons droites et nous les désignerons par a, b, c, ... ; les êtres du TROISIÈME système, nous les nommerons *plans* et nous les désignerons par α, β, γ, ... ; les points seront aussi nommés *éléments de la Géométrie linéaire* ; les points et les droites, *éléments de la Géométrie plane* ; et les points, les droites et les plans, *éléments de la Géométrie de l'espace* ou *éléments de l'espace.*

Concevons que les points, droites et plans aient entre eux certaines relations mutuelles et désignons ces relations par des mots tels que : « SONT SITUÉS », « ENTRE », « PARALLÈLE », « CONGRUENT », « CONTINU » ; la description exacte et complète de ces relations a lieu au moyen des *axiomes de la Géométrie.* (Hilbert 1900, page 104.)

En fait, en réponse à Frege, il affirme ceci.

Vous dites que mon explication du § 3 n'est pas une définition du concept « entre », car les caractéristiques font défaut. Or ces caractéristiques sont données en détail dans les axiomes II1-II5. Néanmoins, si l'on tient à prendre le mot « définition » précisément dans le sens reçu, il suffit de dire :

« "Entre" est une relation entre points d'une droite, qui possède les caractéristiques suivantes : II1... II5. »

> Vous ajoutez : « Il en va tout autrement des explications du §1, où la référence des mots “points”, “droite”, . . . , n'est pas indiquée, mais supposée connue. » Le cœur du malentendu est ici. Je ne désire rien supposer connu ; je vois dans mon explication du § 1 la définition des concepts de point, de droite et de plan, si l'on prend comme caractéristiques tous les axiomes des groupes I-V. Si quelqu'un cherche d'autres définitions de « point », peut-être à l'aide de circonlocutions du type « sans extension », etc., il me faudra évidemment m'opposer à cette entreprise de la manière la plus nette ; on cherche là quelque chose qu'on ne pourra jamais trouver parce qu'à cet endroit il n'y a rien, et tout se perd, devient confus et vague, dégénère en un jeu de cache-cache. Si vous préférez dire que mes axiomes sont les caractéristiques des concepts que mes « explications » posent et amènent donc à l'existence, je n'ai vraiment rien à objecter, sinon peut-être que cela heurte les habitudes des mathématiciens et des physiciens. Naturellement, il faut que je puisse poser ces caractéristiques librement. Car sitôt que j'ai posé un axiome, il est là et il est « vrai ». (Frege et Hilbert 1992, page 226.)

Dans son cours du semestre d'été de 1902, Hilbert ira jusqu'à affirmer que « les choses dont les mathématiques s'occupent sont définies, *mises au monde* par les axiomes » (Hilbert 1902, page 47).

Les axiomes de la géométrie peuvent ainsi être conçus comme une définition de chose des objets de la géométrie et des relations qui les lient. Cette définition est implicite au sens de la section 10.3.2. Elle n'est pas arbitraire aux sens suivants :

— la légitimité des axiomes est mise à l'épreuve quant à leur simplicité et leur évidence lorsque nous les apprenons ;

— leur efficacité est mise à l'épreuve par leur usage dans les démonstrations ;

— de plus, Hilbert doit prouver leur cohérence : il estime qu'à présent c'est seulement la non-contradiction qui assure l'existence des objets mathématiques.

Voici comment Hilbert et Bernays circonscrivent le nouveau contexte épistémologique de l'*axiomatique formelle* dans leurs *Fondements des mathématiques* (1934).

> [. . .] parmi les matériaux objectifs de notre représentation et à partir desquels sont formés les concepts fondamentaux d'une théorie, la construction axiomatique de la théorie ne conserve que l'extrait formulé dans les axiomes, et fait abstraction de tout autre contenu (Inhalt). Un autre moment de cette axiomatique au sens le plus étroit est celui de la *forme existentielle*. [. . .] on travaille avec un système bien précis de choses (resp. plusieurs systèmes de cette sorte), système qui forme un *domaine* a priori *délimité de sujets*, pour tous les prédicats à partir desquels s'assemblent les énoncés de la théorie.
>
> Ainsi, on pose au départ une telle totalité du “domaine d'individus” ; or cette supposition est une hypothèse idéalisante – sauf dans les cas triviaux où une théorie porte seulement sur une collection finie, solidement délimitée – ; et cette hypothèse s'ajoute aux hypothèses formulées par les axiomes. (Hilbert et Bernays 2001, pages 55-56.)

Je rajoute que le but de cette section était non seulement de rendre compte de ce contexte qui permet de traiter les concepts fondamentaux d'une théorie exclusivement comme des « systèmes » de « sujets » (c'est-à-dire d'objets de certains types, comme les points, les droites, les plans, les cercles) de « prédicats » (c'est-à-dire de propriétés concernant ces objets, ou de relations liant ces objets, tels que “entre”) régis par des axiomes, mais encore de réfléchir au sens de ces concepts dans ce contexte. Hilbert et Bernays rajoutent ceci.

L'axiomatique formelle a besoin d'être complétée par l'axiomatique matérielle (inhaltlich) car celle-ci est seule à diriger le choix des formalismes ; de plus, pour une théorie formelle existante, elle est seule aussi à servir de guide pour son application à un domaine de la réalité. (Hilbert et Bernays 2001, page 57.)

Chapitre 11

Analyse logique de preuves. La déduction naturelle

Introduction

La logique classique telle qu'elle est présentée au lycée permet de représenter les assertions mathématiques par des formules, mais ne rend pas compte des étapes qui permettent de passer d'une assertion à l'autre, d'une formule à l'autre. C'est pourquoi Gerhard Gentzen (1935) a élaboré la *déduction naturelle*. Nous allons suivre la démarche de son article Gentzen 1936 pour la présenter : analyser la preuve euclidienne de l'infinité des nombres premiers et nous rendre compte des règles logiques utilisées explicitement ou tacitement.

11.1 L'arithmétique des *Éléments*

Lisons d'abord quatre définitions du septième livre des *Éléments* d'Euclide, le premier des livre arithmétiques.

> 1. Est *unité* ce selon quoi chacune des choses existantes est dite une.
> 2. Et un *nombre* est la multitude composée d'unités.
> .....
> 12. Un nombre *premier* est celui [qui est] mesuré par une seule unité.
> .....
> 14. Un nombre *composé* est celui [qui est] mesuré par un certain nombre.

On ne peut pas proprement appeler la définition 1 une définition mathématique : elle est placée au début des livres arithmétiques pour exprimer que l'unité, le 1, est le premier principe de l'arithmétique. La définition elle-même veut dire que l'unité est l'aspect d'une chose selon lequel celle-ci est indivisible. Voir l'exercice 11.1.1 pour réfléchir à ce type de définition.

Une conséquence des définitions 1 et 2 est que pour les Grecs, 1 n'est pas un nombre et que le plus petit nombre est la dyade, c'est-à-dire 2.

Le verbe « mesurer » veut dire « générer par additions répétées » tant en géométrie qu'en arithmétique : voir la section 2.1. Donc un nombre est premier si la seule manière de le générer comme addition répétée (au moins une fois) d'une chose est

$$1+1 \text{ ou } (1+1)+1 \text{ ou } ((1+1)+1)+1 \text{ ou } \ldots.$$

Par contre, un nombre a est composé s'il existe un nombre $b>1$ tel que a s'obtient en faisant

$$b+b \text{ ou } (b+b)+b \text{ ou } ((b+b)+b)+b \text{ ou } \ldots.$$

Notre présentation moderne de l'arithmétique utilise la multiplication pour exprimer la répétition de l'addition, mais pour Euclide, si la composition/juxtaposition/addition produit des grandeurs de même genre, il n'en est pas de même pour la multiplication : la multiplication de deux longueurs produit une aire, et Euclide représentera les nombres comme des longueurs et les produits de deux nombres comme des aires.

En regardant de plus près, on voit que le premier nombre premier est $1 + 1$, c'est-à-dire 2, que $(1 + 1) + 1$, c'est-à-dire 3, est aussi premier et que le premier nombre composé est $((1 + 1) + 1) + 1$, c'est-à-dire 4, puisqu'il est aussi généré par l'addition $(1 + 1) + (1 + 1)$.

11.2 Analyse logique de la proposition 31 du septième livre des *Éléments* d'Euclide

11.2.1 Les opérations logiques

Faire l'analyse logique consiste à déceler les règles logiques que nous appliquons lorsque nous lions des assertions pour en former de nouvelles, comme

- la combinaison de mots "si A est valide, alors B est valide", écrite formellement $\mathrm{A} \to \mathrm{B}$ (implication),
- "A est valide et B est valide", écrite $\mathrm{A} \wedge \mathrm{B}$ (conjonction),
- "A est valide ou B est valide", écrite $\mathrm{A} \vee \mathrm{B}$ (c'est-à-dire qu'au moins une des deux assertions est valide, disjonction),
- "A n'est pas valide", écrite $\neg\mathrm{A}$ (négation),
- "$\mathrm{A}(x)$ est valide pour tout x", écrite $\forall x\, \mathrm{A}(x)$ (quantification universelle),
- "il existe un x pour lequel $\mathrm{A}(x)$ est valide", écrite $\exists x\, \mathrm{A}(x)$ (quantification existentielle).

Nous sommes habitués à toujours spécifier le domaine des quantifications en utilisant le signe $\in$ de la théorie des ensembles, comme dans la formule $\forall x \in \mathbb{N}\ \mathrm{A}(x)$. Mais nous allons donner aux quantificateurs la signification qu'ils portent sur tout le domaine des objets, en l'espèce tous les nombres entiers, et nous interdire d'utiliser le signe $\in$ pour au moins deux raisons :

- les règles du raisonnement qui régissent l'usage des quantifications ne doivent pas tenir compte des théories particulières comme la théorie des ensembles ;
- cela nous évite de préciser les règles du signe $\in$, données dans la théorie des ensembles.

11.2.2 La proposition

Voici la proposition 31 du septième livre. Elle a l'air anodine, mais c'est la proposition des *Éléments* où la descente infinie apparait de la manière la plus nette, et elle établit un fait essentiel pour la preuve de l'infinité des nombres premiers.

> 31. *Tout nombre composé est mesuré par un certain nombre premier.*
>
> Soit un nombre composé a. Je dis que a est mesuré par un certain nombre premier.
>
> En effet, puisque a est composé, un certain nombre le mesurera. Qu'il le mesure et que ce soit b. Et si b est premier, ce qui était prescrit aura été fait. S'il est composé, un certain nombre le mesurera. Qu'il le mesure et que ce soit c. Et puisque c mesure b et que b mesure a, le [nombre] c mesure donc aussi a. Et, d'une part si c est premier, ce qui était prescrit aura été fait, d'autre part s'il est composé, un certain nombre le mesurera. Alors l'investigation étant poursuivie de cette façon, un certain nombre premier sera trouvé qui mesurera $[a]$. Car s'il ne s'en trouvait pas, des nombres en quantité illimitée mesureraient le nombre a, dont chacun serait plus petit que le précédent ; ce

qui est impossible dans les nombres. Donc un certain nombre premier sera trouvé qui mesurera le [nombre] précédent et qui mesurera aussi a.

Donc tout nombre composé est mesuré par un certain nombre premier. Ce qu'il fallait démontrer.

11.2.3 Reformuler la proposition

Dans notre analyse logique, nous allons adopter une terminologie plus moderne :

— remplacer le terme "mesurer" par "diviser proprement", c'est-à-dire diviser tout en étant différent (du dividende) ;

— compter le 1 parmi les nombres, mais pas parmi les nombres premiers, ni parmi les nombres composés ; nous exclurons cependant le 0 des nombres.

Cette démonstration considère donc un nombre quelconque a et elle suppose qu'il est composé. La démonstration procède par descente infinie : elle construit une suite strictement décroissante de diviseurs différents de l'unité et constate que cette suite est nécessairement finie.

Pour étudier les règles de déduction à l'œuvre dans cette démonstration, nous voulons d'abord transformer cette descente en récurrence généralisée sur l'entier a, comme nous l'avons vu dans la section 3.3.2, parce que la récurrence apparait comme un principe plus élémentaire que la descente infinie. Il faudra donc montrer ceci pour un nombre arbitraire a : « si tout nombre plus petit que a qui est composé admet un nombre premier qui le divise, alors, si a est composé, il admet aussi un nombre premier qui le divise ». Cela s'exprime ainsi formellement :

$$\{\forall y\ y < a \to [\text{composé}(y) \to \exists z\,(\text{Prem}(z) \wedge z|y)]\} \to [\text{composé}(a) \to \exists z\,(\text{Prem}(z) \wedge z|a)]$$

où nous avons introduit les prédicats "composé" et "Prem" ; notons que par définition on a composé(a) si a admet un diviseur plus grand que 1 et plus petit que lui-même, formellement

$$\exists x\,[x|a \wedge (x > 1 \wedge x < a)],$$

et nous allons utiliser le fait que si un nombre n'est pas premier et différent de 1 alors il est composé sans soumettre ce fait à une analyse logique.

La récurrence généralisée se démontre ainsi, en reformulant la preuve euclidienne :

— supposons que tout nombre plus petit que a qui est composé admet un nombre premier qui le divise ;

— si a est composé, il a un diviseur b tel que $b > 1$ et $b < a$;

— soit b est premier et on a trouvé un diviseur premier de a,

— soit il est composé ; mais dans ce cas, comme b est plus petit que a, il existe par notre hypothèse un nombre premier c qui divise b ;

— comme b divise a, c divise aussi a.

11.2.4 Analyse de la démonstration selon Gentzen

Notre démonstration consistera à établir la récurrence généralisée, puis à en déduire la proposition.

La démonstration de la récurrence se déroule ainsi. On va supposer que l'hypothèse

$$\forall y\ y < a \to [\text{composé}(y) \to \exists z\,(\text{Prem}(z) \wedge z|y)]$$

est valide.

Supposons que a est composé. On a par définition $\exists x\,[x|a \wedge (x > 1 \wedge x < a)]$. Soit b un tel nombre, et soit donc $b|a \wedge (b > 1 \wedge b < a)$. En particulier, on a alors $b|a$. On a b premier ou non : $\mathrm{Prem}(b) \vee \neg\,\mathrm{Prem}(b)$. C'est pourquoi nous pouvons distinguer deux cas.

Si b est premier, on a $\mathrm{Prem}(b) \wedge b|a$, d'où $\exists z\,(\mathrm{Prem}(z) \wedge z|a)$.

Supposons que b n'est pas premier. On a par l'hypothèse que $b < a \rightarrow [\text{composé}(b) \rightarrow \exists z\,(\mathrm{Prem}(z) \wedge z|b)]$. On a toujours $b|a \wedge (b > 1 \wedge b < a)$ et donc aussi $b > 1 \wedge b < a$ et donc aussi $b < a$. Par ce qui précède, on en conclut $\text{composé}(b) \rightarrow \exists z\,(\mathrm{Prem}(z) \wedge z|b)$; or b n'est pas premier et $b > 1$: donc b est composé et donc $\exists z\,(\mathrm{Prem}(z) \wedge z|b)$. Soit c un tel nombre, et soit donc $\mathrm{Prem}(c) \wedge c|b$. Donc $c|b$; or $b|a$ et la divisibilité est transitive : donc $c|a$. On a donc $\mathrm{Prem}(c) \wedge c|a$. D'où $\exists z\,(\mathrm{Prem}(z) \wedge z|a)$.

On a abouti à la même conclusion dans les deux cas, et on a donc établi $\exists z\,(\mathrm{Prem}(z) \wedge z|a)$ comme conséquence de l'hypothèse que a est composé. Donc

$$\text{composé}(a) \rightarrow \exists z\,(\mathrm{Prem}(z) \wedge z|a)$$

et nous avons établi la récurrence.

On en conclut par récurrence que $\text{composé}(a) \rightarrow \exists z\,(\mathrm{Prem}(z) \wedge z|a)$ est valide pour un nombre a quelconque. Cela montre

$$\forall x\ \text{composé}(x) \rightarrow \exists z\,(\mathrm{Prem}(z) \wedge z|x).$$

C'est le résultat final de la preuve euclidienne.

11.3 Les règles de déduction

On se rend compte que les règles de déduction qui régissent les termes logiques sont de deux sortes : celles par lesquelles les termes sont introduits, et celles par lesquelles ils sont éliminés.

Il s'agit vraiment de décrire ce qui se passe dans le discours mathématique lorsqu'apparait et lorsque disparait un “et”, un “non”, etc. Considérons l'explication des usages du “et”. Je vais considérer trois assertions :

— A ;

— B ;

— A et B.

Lorsqu'on constate que la première assertion, A, est valide, puis que la deuxième assertion, B, est valide, alors on en conclut que la troisième, A et B, l'est aussi.

Cette conclusion a permis d'introduire l'assertion A et B selon une certaine règle : on l'appelle la règle d'introduction du “et”.

Cela peut paraître confus si on n'insiste pas sur le fait que l'introduction consiste à créer une nouvelle assertion, A et B, qui contient le terme logique “et”.

L'élimination consiste à partir de l'assertion A et B et de voir ce qu'on peut en conclure. On peut en conclure A. On peut aussi en conclure B. On a ainsi deux règles d'élimination, l'une pour l'assertion à gauche du “et” et l'autre pour l'assertion à droite du “et”.

Le terme logique “et” est le plus simple et il ne se passe rien de passionnant quand on l'introduit et quand on l'élimine, et on a donc l'impression de tourner en rond, mais ce n'est pas le cas.

Donnons maintenant, comme exemple pour chaque cas d'introduction et d'élimination, une déduction de la preuve euclidienne.

Une **introduction de** $\forall$ a lieu à la fin de la preuve, ligne 20 : après avoir démontré que pour un nombre quelconque a on a composé$(a) \to \exists z\,(\text{Prem}(z) \wedge z|a)$, on en a conclu $\forall x$ composé$(x) \to \exists z\,(\text{Prem}(z) \wedge z|x)$.

Une **introduction de** $\wedge$ a lieu ligne 12 : les deux formules Prem(c) et $c|a$ ont donné ensemble Prem$(c) \wedge c|a$.

Une **introduction de** $\exists$ a lieu ligne 13 : de la formule Prem$(c) \wedge c|a$ on a déduit $\exists z\,(\text{Prem}(z) \wedge z|a)$.

Il n'y a aucune **introduction de** $\vee$ dans cette démonstration. Elle consiste à partir d'une formule valide A et à en conclure qu'une formule A $\vee$ B (respectivement B $\vee$ A) est aussi valide.

Une **introduction de** $\to$ a lieu ligne 16 : en partant de l'hypothèse composé(a), nous avons abouti au résultat $\exists z\,(\text{Prem}(z) \wedge z|a)$. Donc la formule composé$(a) \to \exists z\,(\text{Prem}(z) \wedge z|a)$ est valide.

Une **élimination de** $\forall$ a lieu ligne 8 : de $\forall y\ y < a \to [\text{composé}(y) \to \exists z\,(\text{Prem}(z) \wedge z|y)]$ on a conclu $b < a \to [\text{composé}(b) \to \exists z\,(\text{Prem}(z) \wedge z|b)]$.

Une **élimination de** $\wedge$ a lieu ligne 5 : de $b|a \wedge (b > 1 \wedge b < a)$ on a déduit $b|a$.

Une **élimination de** $\exists$ a lieu ligne 11 : en partant de la formule $\exists z\,(\text{Prem}(z) \wedge z|b)$ nous avons conclu Prem$(c) \wedge c|b$, où c est censé signifier un quelconque des nombres qui existent selon la formule.

Une **élimination de** $\vee$ commence ligne 6 : en partant de la formule Prem$(b) \vee \neg$ Prem(b) nous avons fait une distinction de cas : 1. b est premier ; 2. b n'est pas premier. Cette distinction de cas a été terminée en aboutissant finalement dans les deux cas à la même formule, c'est-à-dire seulement ligne 14.

Une **élimination de** $\to$ a lieu ligne 11 : de composé(b) et de composé$(b) \to \exists z\,(\text{Prem}(z) \wedge z|b)$ nous avons déduit $\exists z\,(\text{Prem}(z) \wedge z|b)$.

Pour la négation ($\neg$), la situation n'est pas aussi claire : cette preuve est pauvre en négations et sa seule apparition est dans la distinction de cas Prem$(b) \vee \neg$ Prem(b) qui a la forme d'une application de la règle du tiers exclu : voir la section 11.4.2.

11.4 La formalisation des règles de déduction

Nous allons décrire la forme générale des règles de déduction ci-dessus. Par exemple, on décrira la forme générale de la $\wedge$-élimination ainsi : si une formule A $\wedge$ B est prouvée (A et B étant des formules quelconques), alors A (respectivement B) est aussi valide. Mais il faudra prendre garde à ce qu'une démonstration mathématique n'est pas toujours construite de manière à procéder de formules valides en formules valides : il arrive au contraire qu'une formule est valide sous certaines hypothèses seulement, et que lorsqu'une autre formule en est déduite, sa validité dépend alors de la validité de ces hypothèses. On peut le voir ci-dessus pour la déduction par contradiction, la $\to$-introduction, et l'hérédité de la récurrence.

Pour exprimer complètement la signification de chacune des règles, il faut donc parfois préciser de quelles hypothèses dépendent les hypothèses.

11.4.1 La conjonction, la disjonction, les quantifications et l'implication

Pour ces cinq opérations, les règles de déduction se conçoivent de manière assez naturelle. Les voici. Elles se placent tacitement dans un contexte d'hypothèses données dont on exige qu'elles n'interfèrent pas avec la déduction en question.

$\wedge$**-introduction** : des assertions A et B se déduit l'assertion A $\wedge$ B.	$\wedge$**-élimination** : de A $\wedge$ B se déduit A (respectivement B).
$\vee$**-introduction** : de A se déduit A $\vee$ B (respectivement B $\vee$ A).	$\vee$**-élimination** : de l'assertion A $\vee$ B et de l'assertion C montrée sous l'hypothèse A ainsi que sous l'hypothèse B se déduit l'assertion C sans ces hypothèses.
$\forall$**-introduction** : de A(a) se déduit $\forall x\,$A(x), si on a pris garde que la variable a n'apparait pas dans les hypothèses de A(a).	$\forall$**-élimination** : de $\forall x\,$A(x) se déduit A(t).
$\exists$**-introduction** : de A(t) se déduit $\exists x\,$A(x).	$\exists$**-élimination** : de l'assertion $\exists x\,$A(x) et de l'assertion C montrée sous l'hypothèse A(a) se déduit l'assertion C sans cette hypothèse, si on a pris garde que la variable a n'apparait pas dans C ni dans les autres hypothèses de C.
$\rightarrow$**-introduction** : de l'assertion B montrée sous l'hypothèse A se déduit A $\rightarrow$ B sans cette hypothèse.	$\rightarrow$**-élimination** : des assertions A et A $\rightarrow$ B se déduit l'assertion B.

J'ai utilisé les expressions A(a), A(x), A(t) pour désigner des choses différentes.

— Les lettres a et x entre parenthèses sont des *variables* au sens où elles sont censées désigner un nombre indéterminé.

— De plus, ci-dessus, la lettre a n'est soumise à aucune contrainte : on dit que c'est une variable *libre*.

— La lettre t est un terme numérique, c'est-à-dire n'importe quelle expression numérique formée à partir de nombres et de variables à l'aide de fonctions numériques comme l'addition et la factorielle.

— L'expression A(x) désigne une formule dans laquelle j'attire l'attention sur la lettre x. Elle est toujours complétée ci-dessus par une quantification selon x, ce qui fera de ce x une variable *muette*, au sens où la lettre x n'aura plus d'importance en tant que telle : nous avons appris que si la lettre y n'apparait pas dans A(x), alors $\forall x\,$A(x) et $\forall y\,$A(y) ont la même signification.

— L'expression A(t) désigne la formule qu'on obtient en substituant un terme t à la variable x repérée dans l'expression A(x), comme nous avons pris l'habitude de le faire dans la notation fonctionnelle. Ce terme t peut être simplement une variable libre a : c'est ainsi qu'il faut comprendre A(a) dans les règles de $\forall$-introduction et de $\exists$-élimination.

Les mises en garde quant à la variable a dans la $\forall$-introduction et la $\exists$-élimination sont très naturelles dans la pratique mathématique :

— intuitivement, on sait bien que lorsqu'on a constaté une assertion de la forme A(a), par exemple $a + a = 2 \times a$, $\forall x\ x + x = 2 \times x$ a lieu parce qu'aucune hypothèse n'a été nécessaire pour obtenir $a + a = 2 \times a$;

— lorsqu'on utilise une assertion de la forme $\exists x\,$A(x) pour introduire un objet qui vérifie la propriété A, on sait bien qu'il faudra choisir une lettre pour cet objet qui n'a pas encore été utilisée, et si c'est le cas de a, on dira alors “Soit a tel que A(a)”.

Dans l'analyse de la preuve euclidienne, nous avons fait en sorte que les lettres a, b et c soient utilisées pour des variables libres, et les lettres x, y et z pour des variables muettes.

Maintenant que ces règles sont écrites, on se rend compte que $\wedge$ et $\forall$ sont apparentées : le deuxième signe est comme une généralisation du premier. De même pour $\vee$ et $\exists$.

Gentzen (1935) a proposé d'écrire ces règles sous la forme de figures qui peuvent s'agencer en arbre de preuve. Elles sont dans la table 11.1 et on peut les lire ainsi :

— une ou plusieurs formules sont juste au-dessus de la barre : ce sont les hypothèses de la déduction en question ;

Opération	règle(s) d'introduction	règle(s) d'élimination
$\wedge$	$\dfrac{\mathrm{A}\ \ \mathrm{B}}{\mathrm{A}\wedge\mathrm{B}}\ \wedge_{\mathrm{i}}$	$\dfrac{\mathrm{A}\wedge\mathrm{B}}{\mathrm{A}}\ \wedge_{\mathrm{e}}\ \dfrac{\mathrm{A}\wedge\mathrm{B}}{\mathrm{B}}\ \wedge_{\mathrm{e}}$
$\vee$	$\dfrac{\mathrm{A}}{\mathrm{A}\vee\mathrm{B}}\ \vee_{\mathrm{i}}\ \dfrac{\mathrm{A}}{\mathrm{B}\vee\mathrm{A}}\ \vee_{\mathrm{i}}$	$\dfrac{\mathrm{A}\vee\mathrm{B}\quad \begin{matrix}[\mathrm{A}]\\ \vdots\\ \mathrm{C}\end{matrix}\quad \begin{matrix}[\mathrm{B}]\\ \vdots\\ \mathrm{C}\end{matrix}}{\mathrm{C}}\ \vee_{\mathrm{e}}$
$\forall$	$\dfrac{\mathrm{A}(a)}{\forall x\,\mathrm{A}(x)}\ \forall_{\mathrm{i}}$	$\dfrac{\forall x\,\mathrm{A}(x)}{\mathrm{A}(t)}\ \forall_{\mathrm{e}}$
$\exists$	$\dfrac{\mathrm{A}(t)}{\exists x\,\mathrm{A}(x)}\ \exists_{\mathrm{i}}$	$\dfrac{\exists x\,\mathrm{A}(x)\quad \begin{matrix}[\mathrm{A}(a)]\\ \vdots\\ \mathrm{C}\end{matrix}}{\mathrm{C}}\ \exists_{\mathrm{e}}$
$\rightarrow$	$\dfrac{\begin{matrix}[\mathrm{A}]\\ \vdots\\ \mathrm{B}\end{matrix}}{\mathrm{A}\rightarrow\mathrm{B}}\ \rightarrow_{\mathrm{i}}$	$\dfrac{\mathrm{A}\ \ \mathrm{A}\rightarrow\mathrm{B}}{\mathrm{B}}\ \rightarrow_{\mathrm{e}}$

TABLE 11.1 – Figures de déduction

— la formule en-dessous de la barre est la conclusion de la déduction ;

— à droite de la barre est écrit le nom que j'ai choisi pour la règle ;

— une hypothèse est surmontée de points de suspension verticaux lorsqu'il est nécessaire d'indiquer qu'elle résulte d'une preuve (qui peut être plus ou moins longue) dans laquelle on a supposé que la formule qui figure au-dessus de ces points de suspension est valide (c'est donc l'hypothèse de l'hypothèse).

— une formule figure entre crochets si la déduction libère la conclusion de l'hypothèse de cette formule, c'est-à-dire que même si une des hypothèses de la déduction n'est valide que sous l'hypothèse de cette formule, la conclusion n'en dépend plus. Par exemple, dans la $\rightarrow$-introduction, la déduction consiste à partir de la validité d'une formule B sous une hypothèse A pour aboutir à la validité sans l'hypothèse A de la formule $\mathrm{A}\rightarrow\mathrm{B}$.

Nous avons présenté ces figures d'une manière qui les rend intuitives en les lisant de haut en bas : c'est la direction naturelle au cours d'une preuve. Mais on peut aussi les lire de bas en haut : elles indiquent alors que pour établir la conclusion, il suffit de prouver les hypothèses.

Voici le premier exemple de Gentzen (1935) dans sa recherche du chemin le plus naturel pour établir la validité d'une formule : la distributivité de la disjonction par rapport à la conjonction : $(\mathrm{X}\vee(\mathrm{Y}\wedge\mathrm{Z}))\rightarrow((\mathrm{X}\vee\mathrm{Y})\wedge(\mathrm{X}\vee\mathrm{Z}))$.

> On argumentera de la façon suivante. Supposons que X soit valide ou bien que $\mathrm{Y}\wedge\mathrm{Z}$ soit valide. Nous distinguons les deux cas : 1. X est valide ; 2. $\mathrm{Y}\wedge\mathrm{Z}$ est valide.
>
> Dans le premier cas, $\mathrm{X}\vee\mathrm{Y}$ est également valide, de même que $\mathrm{X}\vee\mathrm{Z}$, et par conséquent $(\mathrm{X}\vee\mathrm{Y})\wedge(\mathrm{X}\vee\mathrm{Z})$ est valide.
>
> Dans le second cas, $\mathrm{Y}\wedge\mathrm{Z}$ est valide : Y est donc valide et Z également. Si on a Y, on a $\mathrm{X}\vee\mathrm{Y}$, et si on a Z, on a $\mathrm{X}\vee\mathrm{Z}$. Donc ici également $(\mathrm{X}\vee\mathrm{Y})\wedge(\mathrm{X}\vee\mathrm{Z})$ est valide. Cette formule se trouve ainsi dérivée de $\mathrm{X}\vee(\mathrm{Y}\wedge\mathrm{Z})$, c'est-à-dire que l'on a :
>
> $$(\mathrm{X}\vee(\mathrm{Y}\wedge\mathrm{Z}))\rightarrow((\mathrm{X}\vee\mathrm{Y})\wedge(\mathrm{X}\vee\mathrm{Z})).$$
>
> (Gentzen 1955, p. 17-18.)

On peut alors formuler ce raisonnement avec les figures de déduction de la table 11.1, en précisant avec le chiffre 1 (respectivement 2) écrit au-dessus de l'hypothèse libérée que c'est la règle suivie du chiffre 1 (respectivement 2) qui la libère.

$$\dfrac{\dfrac{\overset{2}{[X \vee (Y \wedge Z)]} \quad \dfrac{\dfrac{\overset{1}{[X]}}{X \vee Y} \vee_i \quad \dfrac{\overset{1}{[X]}}{X \vee Z} \vee_i}{(X \vee Y) \wedge (X \vee Z)} \wedge_i \quad \dfrac{\dfrac{\overset{1}{[Y \wedge Z]}}{Y} \wedge_e}{X \vee Y} \vee_i \quad \dfrac{\dfrac{\overset{1}{[Y \wedge Z]}}{Z} \wedge_e}{X \vee Z} \vee_i}{(X \vee Y) \wedge (X \vee Z)} \wedge_i}{(X \vee Y) \wedge (X \vee Z)} \vee_e 1}{(X \vee (Y \wedge Z)) \to ((X \vee Y) \wedge (X \vee Z))} \to_i 2$$

Selon Gentzen, « la disposition en arbre généalogique pourrait paraitre un peu artificielle dans la mesure où la succession de la distinction de cas X, $Y \wedge Z$ après la constatation de $X \vee (Y \wedge Z)$ y disparait ».

11.4.2 La négation

Il y a une difficulté à donner une signification à l'acte de nier : qu'est-ce que cela veut dire que d'affirmer qu'une chose n'a pas lieu ? En poussant la réflexion, on se rend compte que dire $\neg A$, c'est dire que si on suppose A, on arrive à une contradiction : en introduisant le signe $\perp$ pour le faux (plutôt que de choisir arbitrairement une assertion manifestement fausse comme $1 + 1 = 1$), on peut poser la définition $\neg A$ par $A \to \perp$ puis utiliser les règles de $\to$ ci-dessus avec B le faux pour introduire et éliminer la négation :

$\neg$-introduction : du faux montré sous l'hypothèse A se déduit l'assertion $\neg A$ sans cette hypothèse.

$\neg$-élimination : des assertions A et $\neg A$ se déduit le faux.

Opération	règle d'introduction	règle d'élimination
$\neg$	$\dfrac{\begin{array}{c}[A] \\ \vdots \\ \perp\end{array}}{\neg A}\ \neg_i$	$\dfrac{A \quad \neg A}{\perp}\ \neg_e$

Table 11.2 – Figures de déduction : la négation

Mais ces deux règles ne rendent pas compte de tous les usages de la négation. Il faut déjà préciser l'usage du faux par la règle de l'*absurdité intuitionniste*, appelée aussi *ex falso sequitur quodlibet [du faux on peut déduire ce qu'on veut]* :

— du faux se déduit toute assertion A,

$$\dfrac{\perp}{A}\ \text{e.f.q.}$$

Si on ne veut pas retoucher la démonstration euclidienne, il faut rajouter cette règle très problématique de l'*absurdité classique* :

— du faux montré sous l'hypothèse $\neg A$ se déduit l'assertion A,

$$\dfrac{\begin{array}{c}[\neg A] \\ \vdots \\ \perp\end{array}}{A}\ \text{abs.cl.}$$

Cette règle est à la base de la règle tout aussi problématique du *tiers exclu*,

$$A \vee \neg A,$$

qui est la forme du seul usage de la négation dans la preuve euclidienne. Elle se déduit ainsi des autres règles de déduction :

— supposons que $A \vee \neg A$ n'est pas valide ;

— alors $\neg A$ est valide puisque si A était valide, $A \vee \neg A$ le serait, ce qui est absurde ;

— mais alors $A \vee \neg A$ est valide : c'est absurde !

— Donc $A \vee \neg A$ est valide.

Cela s'écrit ainsi avec des figures de déduction, où je précise toujours avec le chiffre 1 (respectivement 2) écrit au-dessus de l'hypothèse libérée que c'est la règle suivie du chiffre 1 (respectivement 2) qui la libère.

$$\dfrac{\dfrac{\dfrac{}{\dfrac{\dfrac{}{\dfrac{[\neg(A \vee \neg A)]^{2} \quad \dfrac{[A]^{1}}{A \vee \neg A}\,\vee_i}{\bot}\,\neg_e}{\neg A}\,\neg_i 1}{A \vee \neg A}\,\vee_i}{\bot}\,\neg_e}{A \vee \neg A}\,\text{abs.cl.}2}$$

Chapitre 12

La calculabilité mécanique

Introduction

La volonté de comprendre les mathématiques par l'étude de théories formelles rendant compte de « ce qui se passe quand on écrit des mathématiques » est une tentative de comprendre les mathématiques en termes purement calculatoires.

Il était donc naturel que nombre de ceux qui s'occupaient de logique formelle aient senti la nécessité de mettre au clair ce qu'est exactement « un calcul mécanisé ».

Un calcul est une manipulation de « mots » selon des règles précises et objectives.

Une fois qu'un alphabet a été fixé, les mots peuvent être codés par des entiers naturels, en choisissant par exemple un système de numération en base b, où b est supérieur au nombre de lettres dans l'alphabet.

Via ce type de codage un calcul peut être vu comme une fonction de $\mathbb{N}^k$ vers $\mathbb{N}$ et « comprendre ce qu'est un calcul » revient à « comprendre ce qu'est une fonction calculable de $\mathbb{N}^k$ vers $\mathbb{N}$ ».

Lorsque dans les années 1930 des mathématiciens et logiciens ont réfléchi à la manière de décrire en termes précis ce qu'est un calcul algorithmique, ils ont abouti à des résultats assez variés quant à la forme, mais identiques quant au fond. Tous les modèles élaborés ont abouti à la même notion de « fonction calculable de $\mathbb{N}$ vers $\mathbb{N}$ ». Citons notamment Gödel, Church, Post et Turing.

Cependant, c'est Alan Turing qui a emporté la conviction par la simplicité de son modèle, et par le caractère vraiment mécanique de ce dernier.

Nous donnons dans la section 12.1 quelques indications sur le modèle de calcul qu'il a proposé.

Un trait intéressant est que la taille des objets à manipuler et le temps d'exécution peuvent être pris en compte de manière naturelle dans ce modèle.

La section 12.2 explore le paradoxe apparent que constitue le fait de pouvoir définir en termes purement mécaniques ce qu'est un calcul mécanique. *A priori* ceci devrait être interdit par le théorème de Cantor. Le paradoxe se résout en montrant (par la méthode diagonale de Cantor) qu'il n'y a pas de machine de Turing qui puisse décider sur quelles entrées une machine de Turing universelle entre dans une boucle infinie. C'est un exemple, historiquement très important, de problème bien posé qui, bien que de nature concrète élémentaire, n'admet pas de solution mécanique.

Après le théorème d'incomplétude de Gödel, c'est une nouvelle apparition précise du principe (vague) « on ne peut pas tout savoir ».

La section 12.3 introduit le modèle de calculabilité donné par les fonctions récursives et discute brièvement la thèse de Church.

Le chapitre 13 examinera quelques conséquences spectaculaires du théorème d'indécidabilité de Turing.

12.1 Machines de Turing

La machine de Turing abstraite

Turing est parti de l'idée qu'un calcul doit pouvoir être exécuté par une machine idéale qui, à l'instar d'un calculateur humain, dispose d'une feuille de papier et d'un crayon, et procède selon une suite d'opérations élémentaires bien répertoriées une fois pour toutes, exécutées conformément à un plan de travail détaillé ne laissant place à aucune ambigüité. Ce modèle est basé sur la notion d'opération élémentaire. Une telle opération doit être suffisamment simple pour ne consommer qu'une quantité fixe de temps et d'énergie. On imagine donc que la machine dispose d'un alphabet fini fixé une fois pour toutes, et qu'une opération élémentaire consiste à lire, écrire ou effacer une lettre à un endroit précis (la feuille de papier doit être divisée en cases, par exemple on prend du papier quadrillé), ou encore à se déplacer vers une case voisine sur la feuille de papier. Naturellement on n'autorise qu'un nombre fini de lettres distinctes.

Dans le premier modèle qu'il a proposé, Turing utilise une feuille de papier idéale constituée d'une simple succession de cases sur une seule ligne potentiellement infinie : ce qu'on appelle la bande de la machine de Turing.

Par la suite, il a semblé plus naturel d'utiliser pour modèle une machine de Turing qui utilise plusieurs bandes pour son travail. Quant au crayon (muni d'une gomme), il est représenté par ce qu'il est convenu d'appeler une tête de lecture [1] qui se déplace le long de la bande. Il y a une tête de lecture pour chacune des bandes. Au départ, certaines bandes doivent contenir l'entrée de l'algorithme (convenablement codée), tandis que les autres sont entièrement vides. Lorsque la machine s'arrête, on lit le résultat à un endroit convenu.

Une tête de lecture est capable de reconnaitre si la case lue est vide, et si elle n'est pas vide de lire la lettre qui s'y trouve. Selon les instructions qu'elle reçoit elle est capable de modifier le contenu de la case, et de se déplacer éventuellement vers une case immédiatement voisine.

Un exemple : l'addition de deux entiers en base 3

Nous savons depuis l'école élémentaire comment on additionne deux entiers en base 10, selon une procédure entièrement mécanique. Supposons que nous confions cette tâche à une machine selon les principes mentionnés précédemment.

Nous prenons l'addition en base 3 pour simplifier l'exposé.

Chacun des deux nombres à additionner sera écrit, en base 3, sur une bande en respectant des conventions prédéterminées. Nous demandons qu'ils soient écrits de gauche à droite, le premier chiffre étant situé sur la première case libre de la bande correspondante. La convention est que la tête de lecture est placée sur le chiffre des unités, c'est-à-dire sur la dernière case écrite de la bande.

Nous introduisons 5 symboles, c'est-à-dire que nous allons utiliser un alphabet à 5 lettres. Tout d'abord les symboles `$` pour le début de la bande et `∅` pour une case vide ; enfin les symboles `0, 1, 2` pour les chiffres en base 3.

Les bandes sur lesquelles sont écrites les entrées seront notées `E1` et `E2`. Une bande de travail notée `T1` sera utilisée. Enfin la sortie sera écrite sur une bande notée `S1`.

Supposons que nous voulions additionner les nombres `21100221` et `211220102`. Au début du calcul la configuration de la machine peut être représentée comme suit. Nous avons mis des crochets autour de la case en face de laquelle se trouve la tête de lecture sur la bande considérée.

```
E1: |$|2|1|1|0|0|2|2[1]  |  |  |  |  |  | ...
E2: |$|2|1|1|2|2|0|1|0[2]  |  |  |  |  |  | ...
T1: |$[ ]  |  |  |  |  |  |  |  |  |  |  |  |  | ...
S1: |$[ ]  |  |  |  |  |  |  |  |  |  |  |  |  | ...
```

1. Il serait plus correct mais plus lourd de parler d'une tête de lecture/effaçage/écriture.

La convention est que la somme des deux nombres doit être écrite, lorsque la machine a terminé son travail, de gauche à droite, sur la bande de sortie, avec la tête de lecture en face du chiffre des unités.

Deux secondes de réflexion montrent qu'on est obligé de commencer par écrire le résultat de l'addition sur la bande de travail, de droite à gauche, au fur et à mesure que les chiffres sont calculés. Quand on a terminé l'addition, il restera à « dépiler » la bande `T1` pour « l'empiler » dans la bande `S1`.

Dans la première phase du calcul on peut se trouver à chaque moment dans l'une des deux situations suivantes : ou bien il y a une retenue (de 1) à prendre en compte, ou bien il n'y a pas de retenue.

Il y a par suite 4 états dans lesquels puisse se trouver la machine. Nous les noterons `R1` (première phase, avec retenue), `R0` (première phase, sans retenue), `DE` (deuxième phase, dépiler et empiler) et `F` (fin, le travail est terminé).

Pour une étape élémentaire du calcul la machine, selon l'état où elle se trouve et selon ce qu'elle lit sur les différentes bandes, exécute sur chaque bande une opération élémentaire et passe dans un nouvel état. Par opération élémentaire sur une bande on entend : modifier éventuellement le contenu de la case lue, se déplacer éventuellement d'un cran à droite ou à gauche.

Une opération élémentaire sur une bande sera symbolisée par un couple `(X,x)`, où :

— `X` est ou bien une lettre de l'alphabet utilisé, pour signifier que la machine écrit la lettre en question dans la case lue, ou bien la lettre `E` pour signifier l'effacement de la case, on bien la lettre `I` pour indiquer que la case est gardée intacte.

— `x` est l'une des lettres `g`, `d`, `i` et représente le déplacement de la tête de lecture (à gauche, à droite, ou immobile).

Si nous supposons que toutes les bandes doivent être rendues vierges sauf la sortie, les configurations successives de la machine seront alors les suivantes.

Étape 1. Dans la configuration initiale, la machine est dans l'état `R0`. Sans retenue, 1 et 2 font 3, je pose 0 et je retiens 1 (on est en base 3, et donc 3 s'écrit 10). Donc la machine passe dans l'état `R1` et la configuration des bandes est la suivante. Nous n'indiquons pas l'état de la bande `S1` dans la première phase, car il n'est pas modifié.

```
E1: |$|2|1|1|0|0|2[2]  |  |  |  |  |  |  ...
E2: |$|2|1|1|2|2|0|1[0]  |  |  |  |  |  |  ...
T1: |$|0[ ]  |  |  |  |  |  |  |  |  |  |  |  ...
```

Étape 2. Avec retenue, 2 et 0 font 3, je pose 0 et je retiens 1. Donc la machine reste dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: |$|2|1|1|0|0[2]  |  |  |  |  |  |  |  ...
E2: |$|2|1|1|2|2|0[1]  |  |  |  |  |  |  |  ...
T1: |$|0|0[ ]  |  |  |  |  |  |  |  |  |  |  ...
```

Étape 3. Avec retenue, 2 et 1 font 4, je pose 1 et je retiens 1. Donc la machine reste dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: |$|2|1|1|0[0]  |  |  |  |  |  |  |  |  ...
E2: |$|2|1|1|2|2[0]  |  |  |  |  |  |  |  ...
T1: |$|0|0|1[ ]  |  |  |  |  |  |  |  |  |  ...
```

Étape 4. Avec retenue, 0 et 0 font 1, je pose 1. Donc la machine passe dans l'état `R0` et la configuration des bandes est la suivante.

```
E1: |$|2|1|1[0]  |  |  |  |  |  |  |  |  |  ...
E2: |$|2|1|1|2[2]  |  |  |  |  |  |  |  |  ...
T1: |$|0|0|1|1[ ]  |  |  |  |  |  |  |  |  ...
```

Étape 5. Sans retenue, 0 et 2 font 2, je pose 2. Donc la machine reste dans l'état `R0` et la configuration des bandes est la suivante.

```
E1: |$|2|1[1] | | | | | | | | | | ...
E2: |$|2|1|1[2] | | | | | | | | | ...
T1: |$|0|0|1|1|2[ ] | | | | | | | ...
```

Étape 6. Sans retenue, 1 et 2 font 3, je pose 0 et je retiens 1. Donc la machine passe dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: |$|2[1] | | | | | | | | | | | ...
E2: |$|2|1[1] | | | | | | | | | | ...
T1: |$|0|0|1|1|2|0[ ] | | | | | | ...
```

Étape 7. Avec retenue, 1 et 1 font 3, je pose 0 et je retiens 1. Donc la machine reste dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: |$[2] | | | | | | | | | | | | ...
E2: |$|2[1] | | | | | | | | | | | ...
T1: |$|0|0|1|1|2|0|0[ ] | | | | | | ...
```

Étape 8. Avec retenue, 2 et 1 font 4, je pose 1 et je retiens 1. Donc la machine reste dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: [$] | | | | | | | | | | | | | ...
E2: |$[2] | | | | | | | | | | | | ...
T1: |$|0|0|1|1|2|0|0|1[ ] | | | | | ...
```

Étape 9. Avec retenue, rien et 2 font 3, je pose 0 et je retiens 1. Donc la machine reste dans l'état `R1` et la configuration des bandes est la suivante.

```
E1: [$] | | | | | | | | | | | | | ...
E2: [$] | | | | | | | | | | | | | ...
T1: |$|0|0|1|1|2|0|0|1|0[ ] | | | | ...
```

Étape 10. Avec retenue, rien et rien font 1, je pose 1. Donc la machine passe dans l'état `DE` et la configuration des bandes est la suivante.

```
E1: |$[ ] | | | | | | | | | | | | ...
E2: |$[ ] | | | | | | | | | | | | ...
T1: |$|0|0|1|1|2|0|0|1|0[1] | | | | ...
S1: |$[ ] | | | | | | | | | | | | ...
```

Étape 11. Désormais les bandes `E1` et `E2` ne subiront plus de modifications. La deuxième phase commence. Nous indiquons juste la première et la dernière étape de cette deuxième phase. Dans cette phase, tant que la tête de lecture sur `T1` lit un chiffre la machine reste dans l'état `DE`. La configuration des bandes `T1` et `S1` devient la suivante.

```
T1: |$|0|0|1|1|2|0|0|1[0] | | | | | ...
S1: |$|1[ ] | | | | | | | | | | | ...
```

Étape 19. On reste dans l'état `DE`. La configuration des bandes `T1` et `S1` devient la suivante.

```
T1: [$] | | | | | | | | | | | | | ...
S1: |$|1|0|1|0|0|2|1|1|0|0[ ] | | | ...
```

Étape 20. La case lue sur `T1` est `$`. La machine passe dans l'état `F` et la configuration finale des bandes est la suivante.

```
E1: |$[ ] | | | | | | | | | | | | ...
E2: |$[ ] | | | | | | | | | | | | ...
T1: |$[ ] | | | | | | | | | | | | ...
S1: |$|1|0|1|0|0|2|1|1|0[0] | | | | ...
```

On voit que le comportement de la machine peut être décrit par une sorte de programme qui indique précisément ce qu'elle doit faire en fonction de son état et des cases lues.

Ce programme peut par exemple être décrit comme suit. Le premier état indiqué, `R0`, est l'état de la machine au démarrage.

`R0` Les cases lues sur `E1` et `E2` sont indiquées avant les deux points. Après on indique successivement le nouvel état, puis les actions sur `E1`, sur `E2`, sur `T1` (on aurait pu rajouter `(I,i)` pour l'action sur la bande `S1`).

```
(0,0)                    : R0, (E,g), (E,g), (0,d)
(0,1) ou (1,0)           : R0, (E,g), (E,g), (1,d)
(1,1) ou (2,0) ou (0,2): R0, (E,g), (E,g), (2,d)
(2,1) ou (1,2)           : R1, (E,g), (E,g), (0,d)
(2,2)                    : R1, (E,g), (E,g), (1,d)
($,0)                    : R0, (I,i), (E,g), (0,d)
($,1)                    : R0, (I,i), (E,g), (1,d)
($,2)                    : R0, (I,i), (E,g), (2,d)
(0,$)                    : R0, (E,g), (I,i), (0,d)
(1,$)                    : R0, (E,g), (I,i), (1,d)
(2,$)                    : R0, (E,g), (I,i), (2,d)
($,$)                    : DE, (I,d), (I,d), (I,g)
```

`R1` Selon les cases lues sur `E1` et `E2` on indique successivement le nouvel état, puis les actions sur `E1`, sur `E2`, sur `T1` :

```
(0,0)                    : R0, (E,g), (E,g), (1,d)
(0,1) ou (1,0)           : R0, (E,g), (E,g), (2,d)
(1,1) ou (2,0) ou (0,2): R1, (E,g), (E,g), (0,d)
(2,1) ou (1,2)           : R1, (E,g), (E,g), (1,d)
(2,2)                    : R1, (E,g), (E,g), (2,d)
($,0)                    : R0, (I,i), (E,g), (1,d)
($,1)                    : R0, (I,i), (E,g), (2,d)
($,2)                    : R1, (I,i), (E,g), (0,d)
(0,$)                    : R0, (E,g), (I,i), (1,d)
(1,$)                    : R0, (E,g), (I,i), (2,d)
(2,$)                    : R1, (E,g), (I,i), (0,d)
($,$)                    : DE, (I,d), (I,d), (1,i)
```

`DE` Selon la case lue sur `T1` on indique successivement le nouvel état, puis les actions sur `T1`, sur `S1` :

```
 0                   : DE, (E,g), (0,d)
 1                   : DE, (E,g), (1,d)
 2                   : DE, (E,g), (2,d)
 $                   : F , (I,d), (I,g)
```

`F` Fin

Fonctions calculables par une machine de Turing

Dans la suite nous appelons *fonction calculable par une machine de Turing* ou encore *fonction mécaniquement calculable* une fonction de $\mathbb{N}^k$ vers $\mathbb{N}$ qui peut être calculée par une machine de Turing, avec la convention suivante : les k variables entières sont des nombres écrits en binaire sur k bandes d'entrée, et la sortie doit être écrite en binaire sur la bande de sortie.

Rappelons que les têtes de lecture sont au départ placées à l'extrémité droite de chacune des entrées.

Toute autre convention naturelle pour le codage des entiers en entrée et sortie donnerait les mêmes fonctions mécaniquement calculables.

Notez que la machine de Turing ne fait preuve d'aucune « intelligence » lorsqu'elle exécute son programme. Tout lui est fourni de l'extérieur : le programme, et l'initialisation des bandes. L'interprétation du résultat du calcul est elle aussi l'objet d'une action « intelligente » extérieure à la machine. Le caractère purement automatique, mécanique du calcul proprement dit est traduit par la terminologie « mécaniquement calculable ».

Évidemment une machine de Turing peut servir à calculer d'autres choses que des fonctions de $\mathbb{N}^k$ vers $\mathbb{N}$. Mais nous ne prendrons pas la peine de donner les définitions précises correspondantes, qui sont naturelles.

Ce qui n'est pas explicite dans la définition des fonctions calculables par machine de Turing

Une fonction calculable par machine de Turing est décrite de manière précise et sans aucune ambigüité par le « programme » de la machine de Turing qui la calcule.

Cependant il y a quelque chose qui reste non explicite. La part de « non explicite » dans la définition des fonctions calculables par machine de Turing se trouve dans le fait que, étant donnée une machine de Turing avec son programme, on ne sait pas *a priori* si cela définit bien une fonction, c'est-à-dire si la machine aboutit à l'instruction « Fin » pour n'importe quelles valeurs données aux entrées.

Complexité d'un calcul

Le caractère très élémentaire du fonctionnement de la machine de Turing en a fait un candidat naturel, non seulement pour les questions de calculabilité théorique, mais également pour les questions de complexité, et en particulier pour la question de l'appréciation du temps et de l'espace nécessaires à l'exécution d'un algorithme. Une fois l'algorithme traduit dans le modèle de la machine de Turing, le *temps d'exécution* est simplement mesuré par le nombre d'opérations élémentaires qui sont effectuées avant d'aboutir à l'arrêt. L'*espace nécessaire à l'exécution* est représenté par le nombre de cases réellement utilisées sur les différentes bandes de la machine.

Machines de Turing concrètes

Un modèle de calcul extrêmement proche de la machine de Turing est donné par la notion de calcul programmable, dans un langage de programmation usuel (par exemple Pascal, ou Lisp, ou Maple).

La machine de Turing abstraite avec ses bandes le long desquelles se déplacent les têtes de lecture, n'est pas facile à réaliser sous forme d'un mécanisme concret, même électronique. En fait les ordinateurs, bien que directement inspirés des machines de Turing sont d'une conception légèrement différente. Les informations à traiter sont dans la mémoire (cela correspond aux cases sur les bandes de la machine de Turing). À chaque étape du calcul les informations sont transférées

de la mémoire vers le microprocesseur (qui correspond à la tête de lecture) pour être traitées (conformément au programme) et renvoyées à la mémoire.

Cela ne change pas fondamentalement les choses par rapport aux machines de Turing, en particulier cela ne change pas les fonctions calculables, mais cela modifie légèrement la mesure de la complexité des calculs.

On obtient en particulier :

Théorème 12.1.1. *Une fonction $f : \mathbb{N} \to \mathbb{N}$ est mécaniquement calculable si et seulement si il existe un programme en Pascal, utilisant uniquement des variables de type entier, qui, pour tout entier n, calcule $f(n)$ sur l'entrée n.*

12.2 Machine de Turing universelle

Un ordinateur qui ne serait soumis à aucune limitation physique de temps et d'espace serait une *machine universelle* en ce sens qu'il est capable d'exécuter n'importe quel programme qu'on lui soumet.

Un des résultats les plus importants d'Alan Turing est *l'existence d'une machine de Turing universelle* : il s'agit d'une machine de Turing U qui prend en entrée le programme d'une machine de Turing arbitraire M et qui simule le fonctionnement de M.

Une conséquence importante de l'existence d'une machine de Turing universelle est, *via* le processus diagonal de Cantor, l'existence de problèmes bien posés (pour les machines de Turing) mais qui ne pourront être résolus par aucun procédé mécanique du type machine de Turing : l'ensemble des (codes de) fonctions mécaniquement calculables de $\mathbb{N}$ vers $\mathbb{N}$ ne peut pas être énuméré par une fonction mécaniquement calculable.

Nous allons donner dans cette section quelques explications sur ce fait, très important.

12.2.1 Suites effectives et suites mécaniquement calculables

Nous notons $\mathcal{E}(\mathbb{N}, \mathbb{N})$ la classe des suites d'entiers effectives, c'est-à-dire les fonctions de $\mathbb{N}$ vers $\mathbb{N}$ qui peuvent être effectivement calculées.

Nous ne prétendons pas que $\mathcal{E}(\mathbb{N}, \mathbb{N})$ coïncide nécessairement avec la classe des fonctions mécaniquement calculables, que nous notons $\mathbf{Tu}(\mathbb{N}, \mathbb{N})$.

En effet tout d'abord la notion d'effectivité est plutôt une notion première (au même titre que la notion d'entier naturel). Intuitivement, cette notion se ramène à celle de construction : le résultat est effectif si on sait le construire à partir des données. Mais la notion de construction est aussi sujette à discussions. Certaines opérations sont clairement des constructions, et d'autres n'en sont clairement pas. Mais il y a le domaine inconnu des mathématiques à venir, sur lequel tout pari est hasardeux. Ainsi l'effectivité ne peut pas être définie *stricto sensu*, et elle n'a aucune raison de devoir être la même notion que celle d'effectivité « mécanique ».

D'autre part, la définition des suites mécaniquement calculables fait appel à la notion d'effectivité de manière apparemment incontournable, au moins si on se place d'un point de vue constructif.

L'argumentation du point de vue constructif

La part de *non mécaniquement calculable* dans la définition d'une fonction mécaniquement calculable ne peut pas être contournée. Lorsqu'on dit que pour toute entrée $m \in \mathbb{N}$, la machine de Turing T ne produit pas de boucle infinie, on dit que la machine aboutira « effectivement » à l'état final après un certain nombre d'étapes élémentaires de calcul. Autrement dit, sans le concept de base (non défini) d'effectivité, on ne peut comprendre de manière vraiment rationnelle la notion de fonction mécaniquement calculable.

En résumé, non seulement on a une inclusion $\mathbf{Tu}(\mathbb{N},\mathbb{N}) \subset \mathcal{E}(\mathbb{N},\mathbb{N})$ dont il semble impossible de démontrer qu'il s'agit d'une égalité, mais la définition même de la classe $\mathbf{Tu}(\mathbb{N},\mathbb{N})$ présuppose $\mathcal{E}(\mathbb{N},\mathbb{N})$, qui est une notion première et ne peut pas être définie.

L'argumentation du point de vue classique (réalisme platonicien)

L'affirmation « pour tout entrée m il existe un nombre d'étapes élémentaires de calcul, n, au bout duquel la machine est arrivée à l'état final », dont le mathématicien constructif semble mettre en doute qu'elle ait une signification objective claire, est certainement vraie, ou fausse, dans l'absolu, car l'ensemble des entiers naturels est un infini en acte au moins de manière idéale, c'est-à-dire au moins dans un monde des idées qui nous dépasse. Chaque fois qu'on démontre un théorème au sujet des entiers naturels, on ne fait que découvrir une vérité qui existait déjà de toute éternité, ou plutôt qui existe hors du temps et de l'espace, dans le monde des idées.

Nous reviendrons sur la question « $\mathbf{Tu}(\mathbb{N},\mathbb{N}) = \mathcal{E}(\mathbb{N},\mathbb{N})$? » dans la section 12.3.2.

12.2.2 Le théorème de Cantor

Théorème de Cantor pour les suites d'entiers effectives

Tout d'abord en nous appuyant sur ce que nous savons de l'effectivité nous pouvons affirmer le principe suivant.

Principe 12.2.1. *Donner une suite effective dans $\mathcal{E}(\mathbb{N},\mathbb{N})$ est la même chose que donner un élément $\psi \in \mathcal{E}(\mathbb{N}^2,\mathbb{N})$.*

Autrement dit en langage plus formalisé

$$\mathcal{E}(\mathbb{N},\mathcal{E}(\mathbb{N},\mathbb{N})) \simeq \mathcal{E}(\mathbb{N}^2,\mathbb{N}).$$

L'« isomorphisme » en question associe à l'élément $\varphi \in \mathcal{E}(\mathbb{N},\mathcal{E}(\mathbb{N},\mathbb{N}))$ l'élément $\psi \in \mathcal{E}(\mathbb{N}^2,\mathbb{N})$ donné par $\psi(m,n) = \varphi(m)(n)$.

Discussion. Ce principe ne fait que traduire une propriété intuitive claire de la notion de fonction effective. Ceci relève du même genre de conviction que ce qui nous permet d'accepter le raisonnement par récurrence pour les entiers naturels. Ce genre de principe de base pourrait d'un point de vue strictement formel être pris pour un axiome. Mais cela serait méconnaitre l'aspect sémantique des choses. Le principe 12.2.1 est une vérité de base beaucoup plus qu'une hypothèse commode pour élaborer un système formel.

On a alors de façon immédiate le théorème de Cantor pour les suites effectives :

Théorème 12.2.2. *Il n'existe pas d'énumération effective de $\mathcal{E}(\mathbb{N},\mathbb{N})$. Plus précisément si $\varphi \in \mathcal{E}(\mathbb{N}^2,\mathbb{N})$, la suite effective $\psi\colon \mathbb{N} \to \mathbb{N}$ définie par $\psi(n) = \varphi(n,n)+1$ n'est égale à aucune des fonctions $\varphi_n\colon m \mapsto \varphi(n,m)$.*

On a aussi la variante suivante : *Il n'existe pas d'énumération effective de $\mathcal{E}(\mathbb{N},\{0,1\})$. Plus précisément, si $\varphi \in \mathcal{E}(\mathbb{N}^2,\{0,1\})$, la suite effective $\psi\colon \mathbb{N} \to \{0,1\}$ définie par $\psi(n) = 1-\varphi(n,n)$ n'est égale à aucune des fonctions $\varphi_n\colon m \mapsto \varphi(n,m)$.*

La signification de ce magnifique théorème (magnifique, parce que sa preuve est évidente et sa signification profonde) est que *produire des suites effectives d'entiers est intrinsèquement une affaire nettement plus compliquée que de produire des entiers.*

Ce théorème met en avant une part de non décidabilité inhérente à la notion d'effectivité.

La discussion de ce point délicat est éclairée par le cas des fonctions mécaniquement calculables, qui obéissent à une définition précise, une fois admise la notion première d'effectivité.

Théorème de Cantor pour les suites mécaniquement calculables

On a évidemment une version mécaniquement calculable du théorème de Cantor.
Cela nécessite d'abord une définition.

Définition 12.2.3. *Un élément φ de $\mathcal{E}(\mathbb{N}, \mathbf{Tu}(\mathbb{N},\mathbb{N}))$ est appelé une suite mécaniquement calculable dans $\mathbf{Tu}(\mathbb{N},\mathbb{N})$ si la suite double $(n,m) \mapsto \varphi(n)(m)$ est mécaniquement calculable.*

Autrement dit on définit $\mathbf{Tu}(\mathbb{N}, \mathbf{Tu}(\mathbb{N},\mathbb{N}))$ *via* $\mathbf{Tu}(\mathbb{N}^2,\mathbb{N})$.
À la suite de ceci, le théorème de Cantor est toujours aussi évident.

Théorème 12.2.4. *Il n'existe pas d'énumération mécaniquement calculable de $\mathbf{Tu}(\mathbb{N},\mathbb{N})$. Plus précisément, si $\varphi \in \mathbf{Tu}(\mathbb{N}^2,\mathbb{N})$, la fonction mécaniquement calculable $\psi\colon \mathbb{N} \to \mathbb{N}$ définie par $\psi(n) = \varphi(n,n)+1$ n'est égale à aucune des fonctions $\varphi_n\colon m \mapsto \varphi(n,m)$.*

12.2.3 Une machine de Turing universelle

Du théorème 12.2.4 on va déduire qu'il n'existe pas de machine de Turing capable de « déboguer » toutes les machines de Turing.

Pour cela nous avons besoin de décrire ce que fait une machine de Turing universelle.

Nous commençons par des remarques assez informelles, qui pourraient faire l'objet de calculs et d'affirmations plus précises.

Si on limite une fois pour toutes l'alphabet utilisé par les machines de Turing sur les cases des bandes,

— d'une part on ne change pas les fonctions mécaniquement calculables de $\mathbb{N}$ dans $\mathbb{N}$,

— d'autre part, on peut coder n'importe quel programme de n'importe quelle machine de Turing sur un autre alphabet fini fixé.

Par exemple imaginons que les cases non vides reçoivent uniquement les lettres `0` ou `1`, alors on pourra coder le programme d'une machine de Turing en utilisant l'alphabet

`A, B, F, E, I, d, g, i, 0, 1, $, ∅, /`

de la manière suivante.

Les différents états de la machine seront codés par `A0, A1, A10, A11,...` et `F` pour l'état final. Il est convenu que l'état de départ est `A0`.

Les différents bandes seront codés par `B0, B1, B10, B11,...`.

Le code du programme est alors un mot (très long) écrit au moyen de l'alphabet ci-dessus.

Le début du mot indique quelle(s) bande(s) il faut lire en entrée et quelle(s) bande(s) il faut lire en sortie. Par exemple si le mot commence par

`B0B1/B100/...`

cela signifie qu'il y a deux entrées, sur les bandes `B0` et `B1` et une sortie, sur la bande `B100`. Le mot contient ensuite ce que doit faire la machine, dans chacun de ses états, en fonction des cases lues sur certaines bandes. Par exemple

`A0B0B1/B0B1B100/00Eg1i0dA10/.../$∅Eg0g1dF/...`

signifiera : dans l'état `A0` on lit les bandes `B0` et `B1` et on agit au niveau des bandes `B0`, `B1` et `B100`. Si on lit `0` et `0` on traite `B0` selon le code `(E,g)` (comme indiqué page 160), on traite `B1` selon le code `(1,i)`, on traite `B100` selon le code `(0,d)`, enfin l'état suivant sera l'état `A10`. `....` Si on lit `$` et ∅ (c'est-à-dire si on se trouve à l'extrémité gauche de la bande `B0` et si la case lue sur `B1` est vide) on traite les trois bandes selon les codes `(E,g)`, `(0,g)` et `(1,d)` et on passe dans l'état final.

Il est en outre convenu que ce qui doit être lu sur une bande en entrée ou en sortie est le mot se trouvant immédiatement à gauche de la tête de lecture sur cette bande, (pour une entrée, au moment du début du calcul, pour une sortie, au moment où la machine arrête son calcul).

Maintenant que nous avons décrit le codage d'une machine de Turing, nous pouvons expliquer ce que fait une machine de Turing universelle. Elle fonctionne avec l'alphabet {`A, B, F, E, I, d, g, i, 0, 1, $, ∅, /`} pour les lettres figurant sur les cases non vides de ses bandes.

Elle a trois entrées et une sortie. La première entrée est le code t d'une machine de Turing censée calculer une fonction de $\mathbb{N}$ dans $\mathbb{N}$. C'est donc un mot t sur l'alphabet {`A, B, F, E, I, d, g, i, 0, 1, $, ∅, /`}. La deuxième entrée est un entier naturel n, c'est un mot sur l'alphabet {`0, 1`}. Il représente l'entier donné en entrée à la machine de Turing codée par t. Enfin la troisième entrée est un entier naturel p, c'est un mot sur l'alphabet {`0, 1`}. Il représente le nombre d'étapes élémentaires pendant lequel on fait travailler la machine. La sortie représente une « description instantanée » de la machine t après qu'elle ait exécuté p étapes élémentaires de calcul sur l'entrée n.

Par description instantanée de la machine, on comprend : un mot q dans lequel est consigné sous forme codée :

— ce qu'il y a écrit sur chacune des bandes de la machine (jusqu'à la dernière des cases non vides),

— où se trouvent les têtes de lecture,

— l'état dans lequel se trouve la machine.

On convient que si la machine a abouti à l'état final en $k < p$ étapes élémentaires, sa description instantanée reste la même pour p que ce qu'elle était pour k.

On convient également que si l'exécution du programme se révèle impossible (en raison d'une erreur de syntaxe dans le programme codé par le mot t), la description instantanée est remplacée par celle de la machine ayant abouti à l'état final avec 0 écrit sur la bande de sortie.

On conçoit facilement que la fonction $\mathrm{U}' \colon (t, n, p) \mapsto q$ soit mécaniquement calculable. On peut le vérifier par un travail méticuleux. On peut enfin coder t et q par des entiers naturels, et on obtient une fonction U de $\mathbb{N}^3$ dans $\mathbb{N}$ qui a les deux caractéristiques suivantes, en apparence contradictoires :

— U n'est pas terriblement compliquée,

— U synthétise à elle seule toutes les fonctions mécaniquement calculables de $\mathbb{N}$ dans $\mathbb{N}$.

C'est ce qui va nous permettre de mettre en route la machinerie diagonale de Cantor.

12.2.4 Le théorème d'indécidabilité de Turing

Théorème 12.2.5 (le problème de la halte est indécidable)**.** *Le problème consistant à savoir si une machine de Turing (de code t) censée calculer une fonction de $\mathbb{N}$ dans $\mathbb{N}$ aboutit bien à l'état final sur une entrée n ne peut pas être décidé par une machine de Turing (qui prend en entrées t et n).*

Démonstration. La fonction $\mathrm{U} \colon \mathbb{N}^3 \to \mathbb{N}$ est la fonction universelle décrite au paragraphe précédent. Notons $\mathrm{V} \colon \mathbb{N}^3 \to \{0, 1\}$ la fonction qui prend la valeur 1 (pour l'entrée (u, n, p)) si le code $\mathrm{U}(u, n, p)$ indique que l'état final a été atteint [2], et 0 sinon. Pour u et n fixés, la fonction $\mathrm{V}(u, n, \bullet)$ qui ne prend que les valeurs 0 ou 1 est croissante.

Notons $\mathrm{F} \colon \mathbb{N}^3 \to \mathbb{N}$ la fonction qui prend la valeur $g \in \mathbb{N}$ suivante : g est l'entier à lire sur la bande de sortie dans la description instantanée $\mathrm{U}(n, n, p)$, ou 0 dans le cas où ce que contient la bande de sortie à gauche de la tête de lecture n'est pas le code d'un entier naturel en base 2.

Le problème affirmé indécidable dans le théorème est le test pour la question suivante :

$$\exists p \in \mathbb{N} \quad \mathrm{V}(u, n, p) = 1 \,? \tag{12.1}$$

Nous allons réduire à l'absurde l'hypothèse selon laquelle un tel test pourrait être réalisé par une machine de Turing.

Supposons donc qu'une machine de Turing $\mathcal{M}$ calcule une fonction $h \colon \mathbb{N}^2 \to \{0, 1\}$ telle que :

2. par la machine de Turing codée par u, pour l'entrée n, après p étapes élémentaires de calcul.

— $h(u,n) = 0$ lorsque la réponse à la question (12.1) est positive,

— $h(u,n) = 1$ dans le cas contraire.

Alors nous pourrions énumérer toutes les fonctions mécaniquement calculables de $\mathbb{N}$ dans $\mathbb{N}$ de la façon suivante.
On définit la fonction mécaniquement calculable $\mathrm{W}\colon \mathbb{N}^2 \to \mathbb{N} : (t,n) \mapsto v$ comme suit :

— si $h(u,n) = 0$ on calcule les valeurs successives de $\mathrm{U}(u,n,p)$ jusqu'à ce que $\mathrm{V}(u,n,p) = 1$ et on pose $\mathrm{W}(u,n) = \mathrm{F}(u,n,p)$.

— si $h(u,n) = 1$ on pose $\mathrm{W}(u,n) = 0$.

Notre hypothèse est que $\mathcal{M}$ calcule toujours $h(u,n)$, sans jamais se tromper. Alors il est clair que toute fonction mécaniquement calculable ψ de $\mathbb{N}$ dans $\mathbb{N}$ est une fonction $n \mapsto \mathrm{W}(u,n)$ pour un certain entier u : le code de la machine de Turing qui calcule ψ.

Il est clair également que la fonction W est partout bien définie et qu'elle est mécaniquement calculable.

Comme le théorème de Cantor pour les fonctions mécaniquement calculables interdit une telle énumération mécaniquement calculable cela signifie que $\mathcal{M}$ se trompe certainement au moins une fois en calculant h, ou bien qu'elle omet de donner une réponse.

En fait on peut préciser, en regardant la preuve du théorème de Cantor, que si u_0 est le code de la machine de Turing qui calcule la fonction

$$u \mapsto \mathrm{W}(u,u) + 1$$

alors $\mathcal{M}$ donne certainement une réponse erronée sur le couple (u_0,u_0), ou bien ne donne aucune réponse pour cette entrée (car elle est dans une boucle infinie) : en tout cas la valeur $\mathrm{W}(u_0,u_0)$ n'est pas calculée, soit parce que $h(u_0,u_0)$ n'est pas calculée, soit parce que $h(u_0,u_0)$ est égale à 0 alors que la réponse à la question (12.1) est négative. En effet, dans le cas contraire on aurait $\mathrm{W}(u_0,u_0) = \mathrm{W}(u_0,u_0) + 1$. □

Ainsi nous voyons que la preuve d'impossibilité donne un résultat plus précis : elle indique précisément un endroit où le programme de débogage ou bien se trompe, ou bien ne donne pas de réponse.

12.3 Autres modèles de calcul équivalents

12.3.1 Le modèle de calcul imaginé par Gödel

Nous indiquons ici rapidement le modèle de calcul imaginé par Gödel, d'après une suggestion de Herbrand.

Fonctions récursives générales

Commençons par l'exemple de la fonction $\exp\colon \mathbb{N}^2 \to \mathbb{N}, (m,n) \mapsto m^n$, que nous voulons décrire en utilisant uniquement la fonction successeur, notée s. Nous introduisons 3 noms de fonctions : add, mul et exp. Ces fonctions sont soumises à des contraintes de type universel :

$$\begin{array}{rcl} \mathrm{add}(m,0) = m \quad , & \mathrm{add}(m,s(n)) = s(\mathrm{add}(m,n)) & , \\ \mathrm{mul}(m,0) = 0 \quad , & \mathrm{mul}(m,s(n)) = \mathrm{add}(\mathrm{mul}(m,n),m) & , \\ \exp(m,0) = s(0) \quad , & \exp(m,s(n)) = \mathrm{mul}(\exp(m,n),m) & . \end{array} \tag{12.2}$$

Et il est clair que ces contraintes caractérisent complètement ces fonctions : pour n'importe quels entiers a, b, la valeur $\exp(a,b)$ peut être calculée uniquement à partir des égalités ci-dessus prises pour des valeurs particulières des entiers m et n, et en utilisant les règles élémentaires suivantes :

— la transitivité et la symétrie de l'égalité

— la substitution d'une expression par sa valeur (au sein d'une autre expression)

Pour exprimer que les valeurs prises par les fonctions présentes dans un système d'égalités universelles E, peuvent être calculées à partir des égalités en utilisant les deux règles ci-dessus, on dit que le système est *complet*.

En outre le système d'équations universelles (12.2) est *cohérent* en ce sens qu'on ne peut pas en déduire une égalité $a = b$ pour deux entiers distincts a et b.

En suivant Gödel on définit alors :

Définition 12.3.1. *Une fonction* récursive *est une fonction* $\mathbb{N}^k \to \mathbb{N}$ *qui est décrite, conjointement avec d'autres fonctions récursives, par un système complet et cohérent d'équations universelles.*

Dans la littérature, on trouve souvent l'expression « fonction récursive générale » ou « fonction récursive totale » comme synonyme de fonction récursive.

Cette définition comporte une part de « non décidable »

— parce que lorsqu'on a un système d'équations universelles, on ne sait pas *a priori* s'il est cohérent,

— et parce que dans le cas où il est cohérent on ne sait pas a priori s'il est complet.

Fonctions primitives récursives

Une classe importante de fonctions récursives est la classe des fonctions *primitives récursives* pour lesquelles les problèmes qu'on vient de soulever ne se posent pas.

Définition 12.3.2. *Les fonctions primitives récursives sont celles qui peuvent être définies (à partir de la fonction nulle et de la fonction successeur) en utilisant uniquement les définitions*

— *par* composition,

— cas par cas, *et*

— *par* récurrence simple.

Cette définition mérite des précisions.

Pour la *composition* nous l'illustrons sur un exemple : la fonction $f\colon \mathbb{N}^3 \to \mathbb{N} : (m, n, p) \mapsto (m + n^p)(n + p)$ peut être définie par $f(m, n, p) = \mathrm{mul}(g(m, n, p), \mathrm{add}(n, p))$ où $g(m, n, p) = \mathrm{add}(m, \exp(n, p))$.

Une fonction $f\colon \mathbb{N}^k \to \mathbb{N}$ est définie *cas par cas* à partir des fonctions $g, h, t\colon \mathbb{N}^k \to \mathbb{N}$ préalablement définies, si elle vérifie (dans ces égalités $\underline{m}$ représente un élément de $\mathbb{N}^k$) :

$$f(\underline{m}) = \begin{cases} g(\underline{m}) \text{ si } t(\underline{m}) = 0 \\ h(\underline{m}) \text{ sinon} \end{cases}$$

Et pour la *récurrence simple* : la fonction $f\colon \mathbb{N}^{k+1} \to \mathbb{N}$ est dite définie par récurrence simple à partir de $g\colon \mathbb{N}^k \to \mathbb{N}$ et $h\colon \mathbb{N}^{k+2} \to \mathbb{N}$ si on a les égalités universelles :

$$f(\underline{m}, 0) = g(\underline{m}) \text{ et } f(\underline{m}, s(n)) = h(\underline{m}, n, f(\underline{m}, n)).$$

Le fait que la construction d'une fonction par récurrence simple produise une fonction bien définie à partir de fonctions bien définies est une évidence intuitive du même ordre que celle qui nous conduit à admettre le raisonnement par récurrence comme valide pour les entiers naturels [3].

3. On peut se poser la question de prouver par récurrence que le système d'égalités universelles correspondant à la définition d'une fonction primitive récursive est complet et cohérent. La réponse est qu'une telle preuve par récurrence est possible. Mais il ne semble pas que la conviction qui en résulte, quant à la vérité du résultat, soit plus forte : construire un objet mathématique par récurrence est *a priori* ni plus ni moins légitime que construire la vérité d'une assertion par récurrence.

Notons que la définition cas par cas peut se ramener à la composition après avoir défini (par récurrence simple) une fonction de sélection $sl\colon \mathbb{N}^3 \to \mathbb{N}$ qui vérifie $sl(a,b,t) = a$ si $t = 0$ et $sl(a,b,t) = b$ sinon.

Nous noterons $\mathbf{Rec}(\mathbb{N}^k, \mathbb{N})$ la classe des fonctions récursives de $\mathbb{N}^k$ vers $\mathbb{N}$, et $\mathbf{PRc}(\mathbb{N}^k, \mathbb{N})$ la classe des fonctions primitives récursives.

On démontre sans trop de difficulté le résultat suivant.

Lemme 12.3.3. *La fonction universelle* $\mathrm{U}\colon \mathbb{N}^3 \to \mathbb{N}$ *attachée à une machine de Turing universelle (voir page 167) est primitive récursive. Il en va de même pour les fonctions* V *et* $\mathrm{F}\colon \mathbb{N}^3 \to \mathbb{N}$ *définies dans la preuve du théorème 12.2.5.*

Les programmes qui calculent les fonctions primitives récursives

On montre sans difficulté que les fonctions primitives récursives sont celles qui peuvent se calculer par un programme

- qui utilise uniquement des variables entières,
- dont les instructions de base sont uniquement des affectations du type $a \leftarrow 0$ ou $a \leftarrow 1$ ou $a \leftarrow b$ ou $a \leftarrow b + c$
- et qui utilise les deux constructeurs d'instructions suivants :
  - **Si** $a = 0$ **alors** ... **sinon** ... **fin si**
  - **Pour** i **de** 1 **à** n **faire** ... **fin pour**

Donnons par exemple un tel programme (algorithme 12.3.4) pour calculer la fonction $\exp\colon \mathbb{N} \to \mathbb{N}$ (notez d'ailleurs qu'une affectation $a \leftarrow a + b$ peut elle-même être réalisée au moyen d'une boucle **pour** et de la seule affectation $a \leftarrow a + 1$).

Algorithme 12.3.4. ***Calcul de la fonction*** exp**.**
Entrée: Deux entiers naturels m et $n \geq 0$
Sortie: $p = m^n$.
Variables locales: a, b, c : entiers ≥ 0 ;
Début
$p \leftarrow 1$;
boucle pour exp
Pour a **de** 1 **à** n **faire**
boucle pour $p \leftarrow p \times m$
$c \leftarrow 0$;
Pour b **de** 1 **à** m **faire**
$c \leftarrow c + p$
fin pour ;
$p \leftarrow c$
fin du calcul de $p \leftarrow p \times m$
fin pour ;
fin de la boucle pour exp
Fin.

Énumération récursive des fonctions primitives récursives

Le caractère complètement contrôlé des fonctions primitives récursives donne le résultat suivant.

Proposition 12.3.5. *Il existe une fonction récursive* $\mathrm{A}\colon \mathbb{N}^2 \to \mathbb{N}$ *qui énumère* $\mathbf{PRc}(\mathbb{N},\mathbb{N})$. *Autrement dit, pour toute* $\phi \in \mathbf{PRc}(\mathbb{N},\mathbb{N})$ *il existe un entier* m *(qui code la définition de* ϕ*) tel que :* $\forall n \in \mathbb{N}\ \phi(n) = \mathrm{A}(m,n)$.

Naturellement, la fonction A n'est pas primitive récursive car la fonction $n \mapsto \mathrm{A}(n,n)$ ne peut pas l'être, d'après la preuve diagonale de Cantor.

Une manière simple de définir une fonction A convenable est de considérer que m est le code d'un programme « primitif récursif » (voir paragraphe précédent) écrit en Pascal, avec une seule entrée. Alors $\mathrm{A}(m,n)$ est la valeur calculée par le programme de code m pour l'entrée n (si m est le code d'un mot qui n'est pas le texte d'un programme conforme à la syntaxe des programmes primitifs récursifs, on convient que $\mathrm{A}(m,n) = 0$ pour tout n).

Nous l'avons notée A en pensant à Ackermann, qui donna une fonction « définie par récurrence double » qui croissait plus vite que toute fonction primitive récursive.

Le procédé de minisation

Si on regarde la relation qui existe entre la fonction universelle $\mathrm{U}(u,n,p)$ et la fonction calculée par la machine de Turing dont le programme est codé par u (dans le cas où elle calcule bien un entier pour toute valeur entière proposée en entrée), on tombe sur le procédé de *minisation*, qui est une manière d'aller au delà des fonctions primitives récursives.

Lemme 12.3.6. *Si* $g\colon \mathbb{N}^2 \to \{0,1\}$ *est une fonction dans* $\mathcal{E}(\mathbb{N}^2,\{0,1\})$ *et si on a*

$$\forall n \in \mathbb{N}\ \exists p \in \mathbb{N} \quad g(n,p) = 1, \tag{12.3}$$

alors il existe une unique fonction $f \in \mathcal{E}(\mathbb{N},\mathbb{N})$ *qui vérifie les conditions* (12.4) *ci-dessous*

$$\forall n \in \mathbb{N}\ (g(n,f(n)) = 1 \text{ et } \forall r < f(n)\ g(n,r) = 0). \tag{12.4}$$

Si g *est récursive,* f *l'est également.*

Définition 12.3.7. *Dans les conditions du lemme 12.3.6 on dit que la fonction* f *est obtenue par* minisation *à partir de la fonction* g *et* on la note $f = \mu(g)$. *On note encore* $f(n) = \mu(g,n)$. *Plus généralement on dira que* $\mu(g,n)$ est définie *chaque fois que* $\exists p \in \mathbb{N}\ g(n,p) = 1$, *et dans ce cas* $\mu(g,n)$ *désigne la plus petite valeur* p *telle que* $g(n,p) = 1$.

On a alors le résultat technique important suivant, directement relié au lemme 12.3.3.

Théorème 12.3.8. *Pour toute fonction mécaniquement calculable* $f\colon \mathbb{N} \to \mathbb{N}$ *il existe deux fonctions primitives récursives* $g\colon \mathbb{N}^2 \to \{0,1\}$ *et* $h\colon \mathbb{N}^2 \to \mathbb{N}$ *telles que :* $\mu(g)$ *est partout définie et pour tout entier* n $f(n) = h(n,\mu(g,n))$.

Démonstration. Nous reprenons les notations de la preuve du théorème 12.2.5. Supposons que la fonction f soit calculée par une machine de Turing dont le code est u_0. Définissons $g\colon \mathbb{N}^2 \to \{0,1\}$ en posant $g(n,p) = \mathrm{V}(u_0,n,p)$ et $h\colon \mathbb{N}^2 \to \mathbb{N}$ en posant $h(n,p) = \mathrm{F}(u_0,n,p)$. Alors $\mu(g)$ est partout définie et $f(n) = h(n,\mu(g,n))$ pour tout n. □

Ceci permet de montrer le théorème suivant :

Théorème 12.3.9. *Les fonctions récursives* $\mathbb{N}^k \to \mathbb{N}$ *sont exactement les fonctions calculables par une machine de Turing :* $\mathbf{Rec}(\mathbb{N}^k,\mathbb{N}) = \mathbf{Tu}(\mathbb{N}^k,\mathbb{N})$.

En conclusion : la part de « non décidable » dans la définition des fonctions récursives peut être concentrée dans le processus de minisation.

Du point de vue des programmes, le théorème 12.3.8 revient à dire toute une fonction récursive peut être calculée par un programme qui a la structure suivante.

— **Début**
— $a \leftarrow 0$
— **Tant que** $a = 0$ **faire**
— un programme primitif récursif
— **fin tant que**
— un programme primitif récursif
— **Fin.**

12.3.2 La thèse de Church

Lorsque Church propose (à peu près au même moment que Turing) une définition pour les fonctions mécaniquement calculables, il prend le pari suivant, qu'on appelle **la thèse de Church** : *on ne trouvera jamais d'autres fonctions calculables* (que celles qu'il avait définies, et qui se sont avérées être les mêmes que celles définies par Turing, Gödel et Post).

La thèse de Church est généralement admise comme une vérité empirique, mais elle ne peut pas être démontrée.

Le fait que toute procédure algorithmique découverte jusqu'à maintenant pour résoudre n'importe quelle question mathématique se ramène toujours à une fonction mécaniquement calculable ne signifie pas que ce sera toujours le cas.

Et, *plus important encore*, même si c'est chaque fois le cas, on ne peut pas décrire un processus purement mécanique qui permette de passer de l'algorithme imaginé par le (ou la) mathématicien(ne) au programme qui réalisera cet algorithme. Ce passage qui consiste à transformer un algorithme intuitif en un programme ne peut certainement pas être réalisé par une machine, mais seulement par un être intelligent, pour qui les mathématiques ont « une signification ». Le rapport entre la sémantique (la signification) des mathématiques et leur syntaxe (leur mise en forme digérable par une machine) est un mystère qui ne relève pas de la thèse de Church mais de la compréhension de ce qu'est un collectif d'êtres intelligents.

Chapitre 13

On ne peut pas tout savoir

Introduction

Nous examinons dans ce chapitre quelques conséquences du théorème d'indécidabilité de Turing (théorème 12.2.5).

Nous indiquons les idées essentielles des démonstrations pour faire voir que les théorèmes « on ne peut pas tout savoir » ainsi obtenus sont tous plus ou moins des frères jumeaux.

13.1 Impossibilités liées aux suites calculables d'entiers

Proposition 13.1.1. *Soit* $\mathrm{A}\colon \mathbb{N}^2 \to \mathbb{N}$ *une fonction récursive qui énumère* $\mathbf{PRc}(\mathbb{N},\mathbb{N})$ *(voir proposition 12.3.5). Alors le test « $\mathrm{A}(n,\bullet) = 0$? » ne peut pas être réalisé par une fonction récursive de n. De même le test « $\mathrm{A}(n,\bullet) = \mathrm{A}(n',\bullet)$? » ne peut pas être réalisé par une fonction récursive de (n,n').*

La preuve est reportée après celle de la proposition suivante qui donne un résultat d'indécidabilité un peu plus fort.

Proposition 13.1.2. *Il existe une fonction* $v\colon \mathbb{N}^2 \to \{0,1\}$ *primitive récursive telle que le test « $v(m,\bullet) = 0$? » ne peut pas être réalisé par une fonction récursive de m.*

Démonstration. On utilise la fonction $\mathrm{V}(u,n,p)$ définie dans la preuve du théorème 12.2.5. On code le couple d'entiers (u,n) par un entier m et on pose $v(m,p) = \mathrm{V}(u,n,p)$. On a vu que le test « $v(m,\bullet) = 0$? » ne peut pas être réalisé par une machine de Turing. □

Preuve de la proposition 13.1.1. Cela résulte intuitivement de ce que A est nécessairement plus compliquée que v. Précisément, en s'appuyant sur la manière dont on a construit la fonction A on montre que pour toute fonction primitive récursive $a\colon \mathbb{N}^2 \to \mathbb{N}$, on peut définir une fonction primitive récursive $b\colon \mathbb{N} \to \mathbb{N}$ telle que, pour tout m, $a(m,\bullet) = \mathrm{A}(b(m),\bullet)$. En appliquant ceci avec $a = v$, puisque le test « $v(m,\bullet) = 0$? » ne peut pas être réalisé par une machine de Turing, il en va de même pour le test « $\mathrm{A}(n,\bullet) = 0$? » □

On en déduit la proposition suivante.

Proposition 13.1.3. *Il existe une suite* $u\colon \mathbb{N} \to \mathbb{N}$ *injective et primitive récursive telle que le test « $\exists m\, u(m) = n$? » ne peut pas être réalisé par une fonction récursive de n.*

Démonstration. La fonction v définie dans la preuve le la proposition 13.1.2 a la propriété suivante : pour m fixé, $v(m,\bullet)$ est une fonction croissante qui ne prend que les valeurs 0 ou/et 1. Considérons la fonction $w\colon \mathbb{N}^2 \to \mathbb{N}^2$ définie par :

— $w(m,p) = (m,0)$ si $v(m,p) = 1$ et p est le plus petit entier tel que $v(m,p) = 1$).

— $w(m,p) = (m, 1+p)$ sinon.

On montre facilement que w est primitive récursive : $w(m,0) = v(m,0)$ et $w(m,p+1) = 0$ si et seulement si $v(m,p) = 0 \neq v(m,p+1)$. En outre $(m,0)$ est dans l'image de w exactement lorsque $v(m,\bullet) \neq 0$. Ainsi w est injective, primitive récursive, et son image ne peut pas être testée par une fonction récursive. Il reste à coder $\mathbb{N}^2$ par $\mathbb{N}$, ceci remplace w par une fonction $u\colon \mathbb{N} \to \mathbb{N}$ vérifiant les conditions requises. □

13.1.1 Structure des ensembles infinis dénombrables

La proposition 13.1.1 indique clairement que l'ensemble $\mathbf{PRc}(\mathbb{N},\mathbb{N})$, qui peut être « facilement » énuméré, est cependant un infini de nature plus compliquée que $\mathbb{N}$ en raison de la difficulté du test d'égalité dans $\mathbf{PRc}(\mathbb{N},\mathbb{N})$: la relation d'égalité est partie intégrante de la « structure d'ensemble ». Et l'égalité peut être plus ou moins compliquée selon les ensembles. On voit ici s'évanouir l'illusion selon laquelle les ensembles infinis pourraient être considérés comme de pures collections en vrac, sans structure, auxquels « on ajouterait de la structure » pour en faire au choix des groupes, des ensembles ordonnés, des espaces de Banach, etc.

De même dans la proposition 13.1.3 la bijection explicite w établit un isomorphisme entre les ensembles $\mathbb{N}$ et $w(\mathbb{N})$, mais ce dernier est, *en tant que partie de* $\mathbb{N}$, assez compliqué, nettement plus compliqué par exemple que la partie $2\mathbb{N}$ formée des entiers pairs. Précisons cette idée. La partie X complémentaire de $w(\mathbb{N})$ dans $\mathbb{N}$ est une partie infinie de $\mathbb{N}$ et est donc considérée en mathématiques classiques comme étant dénombrable. Mais il n'existe pas d'énumération récursive de X, car sinon on pourrait tester l'appartenance $x \in w(\mathbb{N})$ en énumérant X et $w(\mathbb{N})$ jusqu'au moment où on obtient x dans l'une des deux parties. Ainsi, si on tient compte de l'effectivité, les couples (Y,Y') de parties infinies complémentaires dans $\mathbb{N}$ sont très loin d'être tous équivalents. Cela ne s'explique que parce que $\mathbb{N}$ n'est pas un ensemble « en vrac ». Bien au contraire il arrive, dès son entrée en scène, avec la structure que lui confère la fonction successeur, sans laquelle il est impossible à concevoir.

13.1.2 Importance de $\mathbf{PRc}(\mathbb{N},\mathbb{N})$

Un très grand nombre de conjectures mathématiques célèbres peuvent être ramenées à la forme : « montrer qu'une certaine suite primitive récursive d'entiers est identiquement nulle ». La proposition 13.1.1 nous dit donc qu'il n'existe pas de méthode automatique (en tout cas pas de méthode programmable sur ordinateur) pour résoudre systématiquement ce genre de conjectures : on ne peut pas tout savoir, il restera toujours du travail pour inventer de nouvelles méthodes de preuves pour résoudre ces conjectures.

Un autre fait qui souligne l'importance de l'ensemble $\mathbf{PRc}(\mathbb{N},\mathbb{N})$ est donné par l'analyse des systèmes formels (par exemple la théorie des ensembles formalisée dans le système axiomatique **ZF**). Nous y reviendrons dans la section 13.4.

13.2 Impossibilités liées aux nombres réels

Des résultats de même nature concernant les nombres réels sont peut-être plus spectaculaires.

L'ensemble $\mathbb{Q}$ des rationnels peut être numéroté explicitement par divers procédés naturels, tous équivalents du point de vue de la calculabilité (on passe d'une numérotation à une autre *via* des bijections récursives de $\mathbb{N}$ dans $\mathbb{N}$). Cela confère à $\mathbb{Q}$ une structure naturelle de « calculabilité » du même type que celle de $\mathbb{N}$. Ainsi une suite $(x_n)_{n\in\mathbb{N}}$ dans $\mathbb{Q}$ est mécaniquement calculable s'il existe une machine de Turing qui la calcule (*via* un codage naturel de $\mathbb{Q}$), ou encore de manière équivalente, si, *via* une numérotation naturelle de $\mathbb{Q}$, la suite correspond à une fonction récursive de $\mathbb{N}$ dans $\mathbb{N}$.

Définition 13.2.1. *Un nombre réel x est dit* mécaniquement calculable *ou encore* récursif *(resp.* primitif récursif*) s'il existe une suite mécaniquement calculable (resp. primitive récursive) $(x_n)_{n\in\mathbb{N}}$ dans $\mathbb{Q}$ vérifiant $|x - x_n| \leq 1/2^n$ pour tout n.*

Un suite de nombres réels $(x^{(m)})_{m\in\mathbb{N}}$ est dite mécaniquement calculable *ou encore* récursive *(resp.* primitive récursive*) s'il existe une suite double mécaniquement calculable (resp. primitive récursive) $(x_n^{(m)})_{m,n\in\mathbb{N}}$ dans $\mathbb{Q}$ vérifiant $|x^{(m)} - x_n^{(m)}| \leq 1/2^n$ pour tout n.*

On obtient alors la proposition suivante.

Proposition 13.2.2. *Il existe une suite primitive récursive de nombres réels $(x^{(m)})_{m\in\mathbb{N}}$, tous rationnels*[1] *de la forme 0 ou $1/2^k$ ($k \geq 1$) dont la suite des signes n'est pas récursive. En particulier la suite n'est pas récursive en tant que suite dans $\mathbb{Q}$.*

De même il existe une suite primitive récursive de nombres réels telle que la suite des développements en base 2 (qui est une suite double d'entiers) ne peut pas être calculée par une machine de Turing.

Démonstration. On considère la fonction primitive récursive $v\colon \mathbb{N}^2 \to \mathbb{N}$ de la proposition 13.1.2. On pose $x_n^{(m)} = 1/2^{k+1}$ si il existe $k \leq n$ tel que $v(m,k) = 1$ et $v(m,r) = 0$ pour tout $r < k$. Dans le cas contraire on pose $x_n^{(m)} = 0$. Soit $x^{(m)}$ la limite de $x_n^{(m)}$. On obtient ainsi une suite primitive récursive de nombres réels. On a alors $x^{(m)} = 0$ si et seulement si $v(m,\bullet)$ est identiquement nulle, ce qui ne peut pas être testé récursivement en fonction de m.

Dans le cas où $v(m,\bullet)$ n'est pas identiquement nulle, le réel $x^{(m)}$ est de la forme $1/2^k$.

Maintenant on considère la suite $(1-x^{(m)})_m$. Le réel $1-x^{(m)}$ est égal à 1 lorsque $v(m,\bullet) = 0$ et, dans le cas contraire, il s'écrit $0,111\ldots10000\ldots$ en base 2. Le premier chiffre du développement en base 2 ne dépend donc pas récursivement de m. □

Commentaire. Cette proposition montre que l'inclusion $\mathbb{Q} \subset \mathbb{R}$ est une chose infiniment plus compliquée qu'une inclusion entre ensembles finis. La suite $x^{(m)}$ est formée de réels qui sont rationnels en vertu d'un raisonnement incomplet, non constructif : si le raisonnement était complet, on pourrait identifier le rationnel en tant que nombre rationnel. Or les nombres rationnels donnés comme tels ont toujours un signe -1, 0 ou $+1$ parfaitement clair.

Proposition 13.2.3 (la suite de Specker)**.** *Il existe une suite primitive récursive croissante de nombres rationnels $(x_m)_{m\in\mathbb{N}}$ dans l'intervalle $[0,1]$ telle qu'aucun nombre réel récursif ne soit limite de la suite (x_m).*

Démonstration. On considère la fonction $u\colon \mathbb{N} \to \mathbb{N}$ dans la proposition 13.1.3. On pose $x^{(0)} = 0$ et $x_{m+1} = x_m + 1/4^{1+u(m)}$. Si on était capable de connaitre la limite (supposée exister) x de x_m avec une précision arbitraire $1/2^p$ sous forme d'une suite récursive $p \mapsto a_p \in \mathbb{Q}$ on serait aussi capable de tester récursivement si un entier k est dans l'image de u ou non. Par exemple on aura $x \geq 1/4$ si 0 est une valeur prise par u et $x \leq 1/12$ sinon. Situer x sur un intervalle rationnel de longueur $< 1/6$ permet donc de savoir si 0 est une valeur prise par u. □

13.3 Impossibilité de résolution systématique des problèmes diophantiens

Une conséquence du théorème d'indécidabilité est la réponse négative au 10$^{\text{e}}$ problème de Hilbert.

Ce problème s'énonce comme suit : trouver un algorithme pour décider si un polynôme à coefficients entiers (à n variables $x_1,\ldots,x_n$) prend la valeur 0 pour au moins un $x_1,\ldots,x_n \in \mathbb{Z}^n$.

1. Voir le commentaire qui suit la démonstration.

À la suite de nombreux travaux d'autres auteurs, Matiassevitch a montré en 1970 qu'une telle solution algorithmique des problèmes diophantiens est impossible (en supposant la thèse de Church, c'est-à-dire en supposant que tout algorithme correspond à une fonction récursive).

En fait Matiassevitch a montré que l'image de toute suite primitive récursive dans $\mathbb{N}$ est égale à l'ensemble des valeurs ≥ 0 prises par un certain polynôme de $\mathbb{Z}[X_1, \ldots, X_n]$ (qui dépend de la suite primitive récursive en question). Si P est le polynôme correspondant à la suite u de la proposition 13.1.3, pour savoir si $m \geq 0$ est dans l'image de u, il suffit de savoir si le polynôme $P - m$ prend la valeur 0.

Ce théorème signifie que si la thèse de Church reste vraie, il restera toujours du travail en théorie des nombres. Toujours de nouveaux théorèmes à découvrir en utilisant des méthodes de preuves radicalement nouvelles à chaque fois.

13.4 Impossibilités liées aux systèmes de preuves formalisés

13.4.1 Théorèmes d'incomplétude de Gödel

Considérons un système de preuves formalisé qui permet de parler des entiers naturels et dans lequel la fonction primitive récursive $v\colon \mathbb{N}^2 \to \{0,1\}$ donnée à la proposition 13.1.2 puisse être évaluée pour tout couple d'entiers en entrée.

Ce système sera capable de prouver l'assertion « $P_m : \exists p\, v(m,p) = 1$ » pour tout entier concret m pour lequel cela se produit, simplement en évaluant $v(m,p)$ pour le bon entier p. Appelons « mauvaise valeur de m » les entiers pour lesquels ceci se produit.

Concernant « les bonnes valeurs de m », c'est-à-dire les entiers m pour lesquels l'assertion « $\neg P_m : \forall p\ v(m,p) = 0$ » est vérifiée, considérons celles pour lesquelles le système formel est capable de prouver $\neg P_m$. Une chose que nous savons *a priori* est que, vu le fonctionnement complètement automatique des preuves dans un système formalisé, les entiers m en question parcourront l'ensemble des valeurs prises par une suite primitive récursive.

Mais si toutes les bonnes valeurs m pouvaient être énumérées par une suite primitive récursive, on aurait un algorithme récursif pour tester les bonnes valeurs de m : en énumérant à la fois les mauvaises valeurs et les bonnes valeurs, on tomberait un jour sur la valeur m_0 que l'on désire tester. Comme ce n'est pas le cas, il est impossible que le système formel prouve (pour chaque entier m) P_m quand P_m est vraie et $\neg P_m$ quand P_m est fausse, à moins qu'il soit inconsistant.

Notez que dans le théorème d'indécidabilité de Turing on est capable de déterminer précisément un endroit où le programme candidat à déboguer tous les programmes se trompe. Pour la conséquence présente de ce théorème il en ira de même : on sera capable de détecter une bonne valeur de m pour laquelle le système formel ne donnera pas de réponse (c'est-à-dire ne prouvera pas l'assertion $\neg P_m$), ou même donnera une réponse erronée (il prouvera qu'elle est fausse).

C'est essentiellement le

Premier théorème d'incomplétude de Gödel. *On considère un système formel dans lequel on est capable de décrire les fonctions primitives récursives d'entiers*[2] *et de les évaluer. Si ce système formel prouve sans jamais se tromper des théorèmes du type* P_m *et du type* $\neg P_m$*, on pourra déterminer un entier* m_0 *tel que* $\neg P_{m_0}$ *sera vrai mais improuvable.*

Comme on l'a remarqué auparavant, chaque fois que P_m est vrai, le système formel le prouve « par un simple calcul d'évaluation ». Mais un système formel pourrait être cohérent et néanmoins démontrer P_m pour certains entiers m alors que c'est faux. C'est la raison pour laquelle, dans

2. On entend précisément par là que lorsque $(a_n)_{n\in\mathbb{N}}$ est une suite primitive récursive, on est capable de donner une formule $A(n,p)$ pour laquelle : primo, le système prouve $A(n,p)$ pour chaque paire (n,p) telle que $p = a_n$, secundo, le système prouve $\forall n\ \exists! p\ A(n,p)$.

le premier théorème d'incomplétude, il y a une hypothèse plus forte que la consistance. Cette hypothèse, formulée de façon un peu plus générale, a reçu le nom technique de ω-consistance : si une suite primitive récursive $(a_n)_{n\in\mathbb{N}}$ ne prend jamais la valeur 0, le système formel ne doit pas prouver $\exists n\, a_n = 0$.

Notez que la méthode diagonale de Cantor à l'œuvre ici nous donne accès à de nouvelles vérités, que le système formel est incapable de démontrer, mais que *nous* sommes capables de démontrer. Cela peut paraitre un tour de force incroyable de l'esprit humain, immédiatement capable de se surpasser dès qu'il a fait le point sur les connaissances déjà acquises.

En fait, nous devons un petit peu tempérer notre autoglorification. Nous avons accès à une nouvelle vérité uniquement sur la base du fait que le système formel que nous avons mis au point prouve, concernant les théorèmes du type $\exists p\ v(m,p) = 1$ et du type $\forall p\ v(m,p) = 0$, uniquement des théorèmes vrais. Et bien que les systèmes formels que nous mettons au point nous semblent rentrer dans ce cadre, notre conviction à ce sujet peut être discutée. C'est ce que suggère le

Deuxième théorème d'incomplétude de Gödel. *Pour tout système formel qui est capable de décrire les fonctions primitives récursives et de les évaluer, la consistance du système formel est un théorème d'arithmétique (du type $\forall p\ a_p = 0$ pour une certaine suite primitive récursive $(a_p)_{p\in\mathbb{N}}$) qui est improuvable par le système formel, sauf s'il est inconsistant*[3].

Précisément, si on désigne par Consis(**F**) le théorème d'arithmétique qui signifie que le système formel **F** est consistant, et si on dispose d'une preuve de Consis(**F**) dans **F**, alors on a aussi une preuve de $0 = 1$ dans **F**.

Ce deuxième théorème d'incomplétude vient en fait confirmer ce que disait Poincaré sur les limites inhérentes à la démarche formaliste, vingt-cinq ans avant Gödel. Voir page 69 les paragraphes qui terminent la section IV et notamment : « [Pour établir que les postulats n'impliquent pas contradiction] ... il faut recourir à des procédés de démonstration où en général on sera forcé d'invoquer ce principe d'induction complète qu'il s'agit précisément de vérifier. »

13.4.2 Arithmétisation des mathématiques

Lorsque Dedekind a *défini* les nombres réels par les coupures[4], on a pu penser que toutes les mathématiques pouvaient se ramener aux entiers naturels, *via* un minimum de théorie des ensembles.

Ceci aboutit en définitive au programme formaliste de Hilbert : à défaut de pouvoir élucider « la sémantique » des ensembles infinis (quelle peut bien être leur signification objective ?), essayons au moins de comprendre « leur syntaxe » : comprendre ce qu'on fait exactement quand on les utilise en mathématiques. Pour cela, il faut considérer une théorie purement formelle, dans laquelle on puisse librement utiliser les infinis de Cantor d'une part, et dans laquelle on puisse écrire les énoncés mathématiques qui « ont sûrement du sens », d'autre part.

On se rend compte que les théorèmes d'indécidabilité de Turing et d'incomplétude de Gödel ont porté un coup sérieux au programme de Hilbert, en marquant des limites assez claires à notre capacité à organiser notre compréhension des mathématiques, ne serait-ce que la compréhension de $\mathbf{PRc}(\mathbb{N},\mathbb{N})$, dans un système formel fixé une fois pour toutes.

Néanmoins, il est indéniable que la considération des systèmes formels nous permet de développer une analyse pertinente de points cruciaux tels que :

— ce que c'est que l'activité mathématique,

— la sémantique des êtres mathématiques,

3. Sous cette forme, où la consistance remplace la ω-consistance, le théorème est dû à Rosser.

4. À un nombre réel intuitif x on peut associer la « coupure » $(\{r \in \mathbb{Q} \mid r < x\}, \{r' \in \mathbb{Q} \mid r' > x\})$. S'il n'y a pas d'infinitésimaux leibniziens, la coupure associée à x caractérise complètement x. Dedekind propose donc de définir un nombre réel *abstrait* comme donné par une coupure, qui est un couple de parties de $\mathbb{Q}$ vérifiant certaines contraintes bien précises.

— le degré de certitude que nous pouvons attribuer à tel ou tel énoncé.

Il est par contre un autre aspect du programme de Hilbert qui a été complètement invalidé, et qui curieusement ne fait l'objet de presque aucune critique rétrospective. C'est le fait de croire qu'avec un système formel particulier (en général, **ZFC**), désormais « les règles du jeu sont fixées » et qu'il n'y a donc plus à se préoccuper des questions de fondements.

Or l'étude de **ZFC**, si on veut bien voir les choses dans leur froide réalité, c'est seulement l'étude des valeurs d'une fonction primitive récursive particulière, de l'ensemble $\mathbb{N}$ vers l'ensemble des énoncés bien formés de la théorie : la fonction qui énumère les théorèmes démontrés en respectant les règles du « jeu formel **ZFC** ».

Et malgré l'importance cruciale de la compréhension de $\mathbf{PRc}(\mathbb{N}, \mathbb{N})$, on ne saurait quand même croire sérieusement que les mathématiques se réduisent à l'étude détaillée de l'ensemble des valeurs prises par une fonction primitive récursive particulière (c'est-à-dire un élément particulier l'ensemble $\mathbf{PRc}(\mathbb{N}, \mathbb{N})$).

Terminons en donnant quatre références pour celles qui veulent approfondir la question.

— Gilles Dowek : *La logique* (1995).

— Gilles Dowek : *Les métamorphoses du calcul : une étonnante histoire de mathématiques* (2007).

— René David, Karim Nour et Christophe Raffalli : *Introduction à la logique : théorie de la démonstration : cours et exercices corrigés* (2004).

— Alan Turing et Jean-Yves Girard : *La machine de Turing* (1995).

Deuxième partie

Épistémologie mathématique
Exercices

Exercices pour le chapitre 1

1.1

L'exercice qui suit est énoncé dans un style très intuitif et informel, et je vous propose de répondre dans le même style.

Exercice 1.1.1 (libre mobilité et les deux premiers cas d'égalité des triangles)**.** Voici les propositions 4 et 26 du premier livre des *Éléments* d'Euclide, qui correspondent aux deux premiers cas d'égalité des triangles, dans la traduction de Bernard Vitrac :

> **Proposition I.4** Si deux triangles ont deux côtés égaux à deux côtés, chacun à chacun, et s'ils ont un angle égal à un angle, celui contenu par les droites égales, ils auront aussi la base égale à la base, les triangles seront égaux et les angles restants seront égaux aux angles restants, chacun à chacun, c'est-à-dire ceux que les côtés égaux sous-tendent.
> **Proposition I.26** Si deux triangles ont deux angles égaux à deux angles, chacun à chacun, et un côté égal à un côté, soit celui des angles égaux, soit celui sous-tendant l'un des angles égaux, ils auront aussi les côtés restants égaux aux côtés restants, {chacun à chacun}, et l'angle restant égal à l'angle restant. (Euclide d'Alexandrie 1990-2001)

Le premier cas est le suivant : si les deux triangles ABC et A′B′C′ ont un angle égal ($\widehat{\mathrm{BAC}} = \widehat{\mathrm{B'A'C'}}$) compris entre des côtés deux à deux égaux ($\mathrm{AB} = \mathrm{A'B'}$ et $\mathrm{AC} = \mathrm{A'C'}$), ils sont égaux.

Dans les trois occurrences le mot « égalité » doit être compris comme « sont égales deux figures qui peuvent être emmenées à se superposer exactement l'une sur l'autre au moyen d'un déplacement ». Ainsi la notion de segments égaux est antérieure à la notion d'égalité des longueurs comme mesures.

Le deuxième cas contient le suivant : si les deux triangles ABC et A′B′C′ ont un côté égal ($\mathrm{AB} = \mathrm{A'B'}$) compris entre des angles deux à deux égaux ($\widehat{\mathrm{BAC}} = \widehat{\mathrm{B'A'C'}}$ et $\widehat{\mathrm{ABC}} = \widehat{\mathrm{A'B'C'}}$), ils sont égaux.

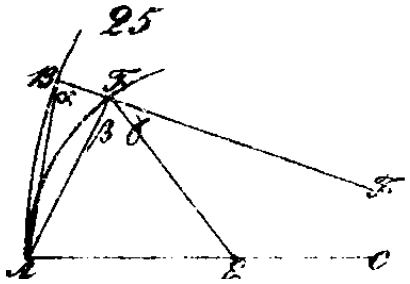
1. Dans quelle mesure ces cas d'égalité peuvent-ils être démontrés au moyen de la propriété de libre mobilité ?
2. Dans quelle mesure impliquent-ils la propriété de libre mobilité ?

Exercice 1.1.2. Voici comment Lobatchevski introduit les horocycles dans ses *Études géométriques sur la théorie des parallèles* de 1840. La proposition 30 est le résultat crucial et difficile.

> 30. Les perpendiculaires élevées aux milieux des côtés d'un triangle rectiligne seront toutes les trois parallèles entre elles, toutes les fois que l'on en supposera deux parallèles.
>
> . . . . . .
>
> 31. Nous appellerons COURBE-LIMITE (horicycle) la ligne courbe, située dans un plan, et telle que toutes les perpendiculaires élevées sur les milieux de ses cordes soient parallèles entre elles.

. . . . . .

32. Un cercle dont le rayon va en croissant se change en une courbe-limite.



Soit AB (*fig.* 25 [tirée du fac-similé de l'édition originale]) une corde de la courbe-limite. Par les extrémités A et B de la corde, menons deux axes, AC, BD [sur la figure, le point D est par erreur noté F], qui feront avec la corde des angles égaux BAC = ABD = α [prop. 31]. Sur un de ces axes AC, prenons un point quelconque E comme centre d'un cercle, et menons l'arc de cercle AF depuis l'origine A de l'axe AC, jusqu'à sa rencontre en F avec l'autre axe BD. Le rayon FE du cercle, correspondant au point F, formera d'un côté, avec la corde AF, l'angle AFE = β, et de l'autre côté, avec l'axe BD, l'angle EFD = γ. L'angle compris entre les deux cordes BAF = $\alpha - \beta < \beta + \gamma - \alpha$ [prop. 22], d'où résulte $\alpha - \beta < \frac{1}{2}\gamma$. Or, comme l'angle γ peut décroitre jusqu'à zéro, soit lorsque le point E se meut dans la direction AC, F restant fixe [prop. 21], soit encore lorsque F s'approche de B sur l'axe BF, le centre E conservant sa position [prop. 22] ; il s'ensuit que, l'angle γ décroissant ainsi, l'angle $\alpha - \beta$, ou l'inclinaison mutuelle des deux cordes AB, AF, et par suite aussi la distance du point B de la courbe-limite au point F du cercle, tendront vers zéro. Donc on peut appeler la courbe-limite *un cercle de rayon infiniment grand.* (Lobatchevski 1840, pages 19-21 de la traduction)

Dans quel sens la « courbe-limite » est-elle ici une limite de cercles ?

1.2

Exercice 1.2.1. Essayez de prouver trois propriétés élémentaires de l'addition et de la multiplication par récurrence en suivant l'axiomatique de Peano. Vous pouvez vous inspirer du texte de Poincaré reproduit dans la section 3.1.

Exercices pour le chapitre 2

2.1

Exercice 2.1.1. Rédiger une démonstration complète de l'impossibilité d'une commune mesure pour le côté et la diagonale du carré. Voici comment l'axiome d'Archimède apparait dans les *Éléments* d'Euclide.

> **Définition V.4.** Des grandeurs sont dites *avoir un rapport l'une relativement à l'autre* quand elles sont capables, étant multipliées, de se dépasser l'une l'autre.

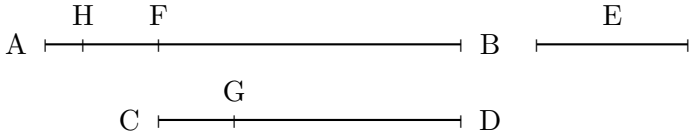
Un tel axiome est-il nécessaire pour conclure ?

2.4

Exercice 2.4.1. Rappelons d'abord que pour Euclide, l'unité n'est pas un nombre :

> **Définition VII.1.** Est *unité* ce selon quoi chacune des choses existantes est dite une.
> **Définition VII.2.** Et un *nombre* est la multitude composée d'unités.

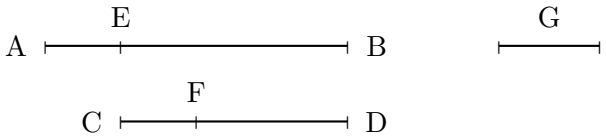
Voici comment l'anthyphérèse arithmétique apparait dans les *Éléments* d'Euclide.

> **Proposition VII.1.** *Deux nombres inégaux étant proposés et le plus petit étant retranché du plus grand de façon réitérée et en alternance, si le reste ne mesure jamais [le reste] précédent jusqu'à ce qu'il reste une unité, les nombres initiaux seront premiers entre eux.*
>
> En effet, que de deux nombres {inégaux} AB, CD, le plus petit étant retranché du plus grand de façon réitérée et en alternance, le reste ne mesure jamais le [reste] précédent jusqu'à ce qu'il reste une unité.
>
> 
>
>
> Je dis que AB, CD sont premiers entre eux, c'est-à-dire qu'une seule unité mesure AB, CD.
>
> Car si AB, CD ne sont pas premiers entre eux, un certain nombre les mesurera. Qu'il les mesure et que ce soit E.
>
> Et d'une part que, CD mesurant BF, il reste FA, plus petit que lui, et d'autre part que, AF mesurant DG, il reste GC, plus petit que lui, et que GC mesurant FH, il reste une unité HA.

Or, puisque E mesure CD et que CD mesure BF, E mesure donc aussi BF. Mais il mesure aussi BA tout entier ; donc il mesurera aussi le reste AF. Et AF mesure DG ; donc E mesure aussi DG. Mais il mesure aussi CD tout entier ; donc il mesurera aussi le reste GC. Et GC mesure FH ; donc E mesure aussi FH. Mais il mesure aussi AF tout entier ; donc il mesurera aussi l'unité restante AH, quoiqu'étant un nombre ; ce qui est impossible. Donc aucun nombre ne mesurera les nombres AB, CD ; donc AB, CD sont premiers entre eux. Ce qu'il fallait démontrer.

Proposition VII.2. *Étant donnés deux nombres non premiers entre eux, trouver leur plus grande commune mesure.*

Soient AB, CD les deux nombres non premiers entre eux donnés. Il faut alors trouver la plus grande commune mesure de AB, CD.



Or, d'une part si CD mesure AB, comme il se mesure aussi lui-même, CD est donc une commune mesure de AB, CD. Et il est évident que c'est aussi la plus grande ; car aucun [nombre] plus grand que CD ne mesurera CD.

Si d'autre part CD ne mesure pas AB, le plus petit des [nombres] AB, CD étant retranché du plus grand de façon réitérée et en alternance, il restera un certain nombre qui mesurera le [reste] précédent. Car il ne restera pas une unité : autrement AB, CD seraient premiers entre eux (VII.1), ce qui n'est pas l'hypothèse. Il restera donc un certain nombre qui mesurera le [reste] précédent.

Et que CD mesurant BE il reste EA, plus petit que lui, que EA mesurant DF il reste FC, plus petit que lui, et que CF mesure AE.

Or puisque CF mesure AE et que AE mesure DF, [le nombre] CF mesurera donc aussi DF ; mais il se mesure aussi lui-même ; donc il mesurera aussi CD tout entier. Et CD mesure BE ; et donc CF mesure BE ; mais il mesure aussi EA ; donc il mesurera aussi BA tout entier ; et il mesure aussi CD ; donc CF mesure AB, CD. Donc CF est une commune mesure de AB, CD. Je dis de plus que c'est aussi la plus grande.

Car si CF n'est pas la plus grande commune mesure de AB, CD, un certain nombre mesurera les nombres AB, CD, tout en étant plus grand que CF. Qu'il les mesure et que ce soit G.

Et puisque G mesure CD et que CD mesure BE, G mesure donc aussi BE ; mais il mesure aussi BA tout entier ; donc il mesurera aussi le reste AE. Et AE mesure DF ; donc G mesurera aussi DF ; mais il mesure aussi CD tout entier ; donc il mesurera aussi le reste CF, le plus grand, le plus petit ; ce qui est impossible. Donc aucun nombre ne mesurera les nombres AB, CD tout en étant plus grand que CF. Donc CF est la plus grande commune mesure des [nombres] AB, CD.

1. Proposez une reformulation de la proposition VII.2 dans un langage que vous maitrisez.
2. Commentez le rapport logique entre les propositions VII.1 et VII.2, en particulier tentez d'expliquer pourquoi la démonstration de la proposition VII.2 fait appel à la proposition VII.1.
3. Comparez l'anthyphérèse géométrique de l'exercice 3 du devoir n° 1 et l'anthyphérèse arithmétique.
4. Reconnaissez-vous l'algorithme 2.4.1 dans la proposition VII.2 ? Motivez votre réponse.

2.6

Exercice 2.6.1. Élaborez un algorithme qui calcule le pgcd et une relation de Bézout à partir de l'algorithme 2.6.1 implicite dans la preuve classique du théorème du pgcd.

Exercice 2.6.2. Si l'on prend deux nombres < 121 par exemple 93 et 63, il est facile de trouver leurs décompositions en produit de facteurs premiers : un nombre inférieur à 121 qui n'est pas

divisible par 2, 3, 5 et 7 est forcément premier. On trouve donc facilement $93 = 3 \times 31$, $63 = 3^2 \times 7$ et $\text{pgcd}(93, 63) = 3$. Pour chaque facteur premier figurant dans les deux décompositions on prend l'exposant minimum, et cela nous donne le pgcd. En outre cela s'applique aussi bien pour le pgcd de 3 entiers par exemple. Quels problèmes voyez vous derrière la simplicité apparente de cet algorithme ?

Exercices pour le chapitre 3

3.1

Exercice 3.1.1. Voici deux extraits de l'introduction à la *Critique de la raison pure* d'Emmanuel Kant dans la traduction d'Alain Renaut, Aubier, 1997.

Introduction

I. De la différence entre la connaissance pure et la connaissance empirique

Que toute notre connaissance commence avec l'expérience, il n'y a là absolument aucun doute ; car par quoi le pouvoir de connaitre devrait-il être éveillé et mis en exercice, si cela ne se produisait pas par l'intermédiaire d'objets qui affectent nos sens et qui, pour une part, produisent d'eux-mêmes des représentations, tandis que, pour une autre part, ils mettent en mouvement l'activité de notre entendement pour comparer ces représentations, les relier ou les séparer, et élaborer ainsi la matière brute des impressions sensibles en une connaissance des objets, qui s'appelle expérience ? En ce sens, *d'un point de vue chronologique*, nulle connaissance ne précède en nous l'expérience, et c'est avec celle-ci que toute connaissance commence.

Cela dit, bien que toute notre connaissance s'amorce *avec* l'expérience, il n'en résulte pas pour autant qu'elle dérive dans sa totalité *de* l'expérience. Car il pourrait bien se produire que même notre connaissance d'expérience soit un composé de ce que nous recevons par des impressions et de ce que notre propre pouvoir de connaitre (simplement provoqué par des impressions sensibles) produit de lui-même — ajout que nous ne distinguons pas de cette matière première avant qu'un long exercice nous y ait rendus attentifs et nous ait donné la capacité de l'isoler.

C'est par conséquent, pour le moins, une question qui requiert d'être examinée de plus près, et dont on ne saurait se débarrasser en la renvoyant à ce qui apparait d'emblée, que celle de savoir s'il y a une telle connaissance, indépendante de l'expérience et même de toutes les impressions des sens. On nomme de semblables connaissances *a priori*, et on les distingue des connaissances *empiriques*, lesquelles possèdent leur source *a posteriori*, c'est-à-dire dans l'expérience.

Cette expression n'est toutefois pas encore déterminée de façon assez précise pour désigner adéquatement tout le sens de la question proposée. Car, de fait, on a coutume de dire, à propos de maintes connaissances dérivées de sources se trouvant dans l'expérience, que nous en sommes capables ou que nous y avons accès *a priori*, parce que nous les dérivons, non pas immédiatement de l'expérience, mais d'une règle universelle que cependant nous avons empruntée elle-même à l'expérience. Ainsi dit-on de quelqu'un qui a miné les fondations de sa maison qu'il pouvait savoir *a priori* qu'elle s'effondrerait, c'est-à-dire qu'il n'avait pas besoin d'attendre l'expérience de son effondrement effectif. Reste qu'il ne pouvait pourtant pas non plus le savoir totalement *a priori*.

Car que les corps sont pesants et que par conséquent, si on leur retire ce sur quoi ils reposent, ils tombent, il fallait bel et bien que cela fût connu de lui auparavant par l'intermédiaire de l'expérience.

Nous entendrons donc par connaissances *a priori*, dans la suite de cet ouvrage, non pas des connaissances qui adviennent indépendamment de telle ou telle expérience, mais celles qui interviennent d'une manière *absolument* indépendante de toute expérience. Leur sont opposées des connaissances empiriques, autrement dit celles qui ne sont possibles qu'*a posteriori*, c'est-à-dire par expérience. Mais, dans les connaissances *a priori*, sont appelées *pures* celles auxquelles absolument rien d'empirique n'est mêlé. Ainsi, par exemple, la proposition : tout changement a sa cause est-elle une proposition *a priori*, mais non point pure, étant donné que le changement est un concept qui ne peut être tiré que de l'expérience.

. . . . . .

IV. De la différence des jugements analytiques et des jugements synthétiques

Dans tous les jugements où le rapport d'un sujet au prédicat se trouve pensé (si j'examine uniquement les jugements affirmatifs, car l'application aux jugements négatifs, ensuite, est facile), ce rapport est possible de deux manières. Ou bien le prédicat B appartient au sujet A comme quelque chose qui est contenu dans ce concept A (de façon implicite) ; ou bien B est tout à fait extérieur au concept A, bien qu'il soit tout de même en connexion avec lui. Dans le premier cas, j'appelle le jugement *analytique*, dans l'autre *synthétique*. Analytiques (pour ce qui est des jugements affirmatifs) sont donc les jugements dans lesquels la connexion du prédicat avec le sujet est pensée par identité, tandis que ceux dans lesquels cette connexion est pensée sans identité se doivent appeler jugements synthétiques. Les premiers, on pourrait les appeler aussi jugements *explicitatifs*, et les autres jugements *extensifs*, parce que les premiers, par le prédicat, n'ajoutent rien au concept du sujet, mais le décomposent seulement par analyse en ses concepts partiels qui étaient déjà pensés en lui (bien que confusément), alors qu'au contraire les seconds ajoutent au concept du sujet un prédicat qui n'était nullement pensé en lui et n'aurait pu en être tiré par aucune analyse de celui-ci. Par exemple, quand je dis : tous les corps sont étendus, c'est un jugement analytique. Car je n'ai pas besoin de sortir au-delà du concept que je relie au mot « corps » pour trouver que l'étendue lui est associée, mais il me suffit d'analyser ce concept, c'est-à-dire de prendre conscience du divers que je pense toujours en lui, pour y rencontrer ce prédicat : c'est donc un jugement analytique. En revanche, quand je dis : tous les corps sont pesants, le prédicat est quelque chose de tout à fait autre que ce que je pense dans le simple concept d'un corps en général. L'ajout d'un tel prédicat donne donc un jugement synthétique.

Les jugements d'expérience, comme tels, sont tous synthétiques. Car il serait insensé de fonder un jugement analytique sur l'expérience, étant donné que je n'ai nullement besoin de sortir de mon concept pour formuler le jugement et que nul témoignage de l'expérience ne m'est donc nécessaire pour cela. Qu'un corps soit étendu, c'est une proposition qui trouve sa consistance *a priori*, et non pas un jugement d'expérience. En effet, avant d'aller à l'expérience, je possède déjà toutes les conditions requises pour mon jugement dans le concept, dont je peux me borner à extraire le prédicat conformément au principe de contradiction, en prenant par là même conscience, en même temps, de la nécessité du jugement, que l'expérience ne m'enseignerait jamais. En revanche, bien que, dans le concept d'un corps en général, je n'inclue nullement le prédicat de la pesanteur, ce concept désigne néanmoins un objet de l'expérience par une partie de celle-ci, à laquelle je peux donc ajouter encore d'autres parties de la même expérience que celles qui appartenaient à ce concept. Je peux connaitre *analytiquement*, par avance, le concept du corps par les caractères de l'étendue, de l'impénétrabilité, de la figure, etc., qui tous sont pensés dans ce concept. Mais maintenant j'élargis ma connaissance et, en reportant mon regard sur l'expérience d'où j'avais tiré ce concept du corps, je trouve aussi la pesanteur toujours associée aux caractères indiqués — et je l'ajoute donc *synthétiquement*, comme prédicat, à ce concept. Ainsi est-ce sur l'expérience que se fonde la possibilité de la synthèse du

prédicat de la pesanteur avec le concept du corps, parce que les deux concepts, bien que l'un ne soit pas contenu dans l'autre, appartiennent pourtant l'un à l'autre, quoique de façon seulement contingente, comme parties d'un tout, à savoir l'expérience, qui elle-même est une liaison synthétique des intuitions.

Simplement, pour ce qui est des jugements synthétiques *a priori*, cette ressource fait totalement défaut. Si je dois aller au-delà du concept A, pour en connaitre un autre, B, comme lui étant lié, quel est ce sur quoi je prends appui et par quoi la synthèse devient possible, alors qu'ici je n'ai pas l'avantage de m'orienter pour cela dans le champ de l'expérience ? Prenons la proposition : tout ce qui arrive possède sa cause. Dans le concept de quelque chose qui arrive, je pense certes une existence qu'un temps précède, etc., et il s'en laisse dégager des jugements analytiques. Mais le concept d'une cause est totalement en dehors de ce concept, et il indique quelque chose de distinct de ce qui arrive : il n'est donc nullement contenu dans cette dernière représentation. Comment puis-je donc en venir à dire de ce qui arrive en général quelque chose de tout à fait distinct et à acquérir la connaissance que le concept de la cause, bien que ne s'y trouvant pas contenu, lui appartient cependant, et même avec nécessité ? Quel est ici l'inconnu $= x$ sur quoi l'entendement prend appui quand il croit découvrir en dehors du concept de A un prédicat B étranger à lui, qu'il estime cependant lié à ce concept ? Ce ne peut être l'expérience, dans la mesure où le principe mentionné a ajouté la deuxième représentation à la première, non seulement avec une généralité plus grande que celle que l'expérience peut fournir, mais même avec l'expression de la nécessité, par conséquent entièrement *a priori* et à partir de simples concepts. Or, c'est sur de tels principes synthétiques, c'est-à-dire extensifs, que repose dans son intégralité la visée finale de notre connaissance spéculative *a priori* ; car les principes analytiques sont certes extrêmement importants et nécessaires, mais uniquement afin d'accéder à cette clarté des concepts qui est requise pour une synthèse assurée et étendue, prenant la forme d'une acquisition effectivement nouvelle.

V. Dans toutes les sciences théoriques de la raison sont contenus des jugements synthétiques *a priori* faisant fonction de principes

1. Les jugements mathématiques sont tous synthétiques. Cette proposition semble avoir jusqu'ici échappé aux observations de ceux qui ont analysé la raison humaine, et s'opposer même directement à toutes leurs conjectures, bien qu'elle soit irréfutablement certaine et très importante pour la suite. En effet, parce que l'on trouvait que les raisonnements des mathématiciens procédaient tous conformément au principe de contradiction (ce que requiert la nature de toute certitude apodictique), l'on se persuada que les propositions fondamentales elles aussi étaient connues à partir du principe de contradiction — ce en quoi les auteurs de ces analyses se trompaient ; car une proposition synthétique peut certes être saisie d'après le principe de contradiction, toutefois uniquement en ceci que l'on suppose une autre proposition synthétique dont on peut la déduire, mais jamais en elle-même.

Avant tout, il faut remarquer que des propositions proprement mathématiques sont toujours des jugements *a priori* et ne sont pas empiriques, parce qu'elles apportent avec elles une nécessité qui ne peut être tirée de l'expérience. Mais si l'on ne veut pas accorder ce point, soit ! Je limite alors ma proposition à la mathématique pure, dont le concept implique déjà qu'elle ne contienne pas de connaissance empirique, mais seulement une connaissance pure *a priori*.

On devrait certes, au premier abord, penser que la proposition $7 + 5 = 12$ est une proposition simplement analytique qui résulte du concept d'une somme de 7 et de 5 d'après le principe de contradiction. Simplement, si l'on y regarde de plus près, on trouve que le concept de la somme de 7 et de 5 ne contient rien de plus que la réunion de deux nombres en un seul, ce par quoi l'on ne pense aucunement quel est ce nombre unique qui les rassemble tous les deux. Le concept de 12 n'est en aucune manière déjà pensé du fait que je pense simplement cette réunion de 7 et de 5, et

je peux bien décomposer analytiquement aussi loin qu'on voudra mon concept d'une telle somme possible : je n'y rencontrerai pourtant pas le nombre 12. Il faut sortir de ces concepts en s'aidant de l'intuition qui correspond à l'un des deux, par exemple ses cinq doigts ou (comme Segner dans son arithmétique) cinq points, et ainsi ajouter l'une après l'autre les unités du cinq donné dans l'intuition au concept du sept. Car je prends d'abord le nombre 7, et en me servant, pour le concept de 5, des doigts de ma main comme d'une intuition, j'ajoute alors, à la faveur de cette image que j'en ai, peu à peu au nombre 7 les unités qu'auparavant je prenais ensemble pour constituer le nombre 5, et je vois ainsi surgir le nombre 12. C'est dire que la proposition arithmétique est toujours synthétique, ce dont on devient d'autant plus clairement conscient que l'on prend des nombres un peu plus grands, étant donné qu'il apparait alors clairement, de fait, que nous pouvons bien tourner et retourner nos concepts comme nous le voulons : sans utiliser l'aide de l'intuition, nous ne pourrions jamais trouver la somme par la seule décomposition analytique de nos concepts.

1. Résumez en quelques mots ce qu'est un jugement *a priori*.
2. Résumez en quelques mots ce qu'est un jugement *synthétique*.
3. Donnez six exemples de jugements : sont-ils *a priori* ? synthétiques ? Motivez votre réponse à l'aide du texte de Kant.
4. Selon vous, Poincaré utilise-t-il ces mots exactement de la même manière que Kant, ou y a-t-il une différence dans leur usage de ces mots ?

3.2

Exercice 3.2.1. Proposez deux exemples de raisonnement par récurrence. Expliquez pour ces exemples si vous estimez que la récurrence y est juste une formalisation des points de suspension ou si elle vous semble incontournable.

3.3

Exercice 3.3.1. On note G_3 l'ensemble des matrices 3×3 à coefficients dans $\mathbb{Z}$ et de déterminant 1. Notez que si $\mathrm{A}, \mathrm{B} \in \mathrm{G}_3$ alors AB et A^{-1} appartiennent aussi à G_3 puisque $\det(\mathrm{AB}) = \det(\mathrm{A})\det(\mathrm{B})$ et $\mathrm{A}^{-1} = \widetilde{\mathrm{A}}/\det(\mathrm{A}) = \widetilde{\mathrm{A}}$ (où $\widetilde{\mathrm{A}}$ désigne la transposée de la matrice des cofacteurs). On s'intéresse au théorème suivant :

Théorème. *Si* $(a,b,c) \in \mathbb{N}^3$ *avec* $(a,b,c) \neq (0,0,0)$*, il existe* $\mathrm{A} \in \mathrm{G}_3$ *et* $g \in \mathbb{N}$ *tels que*

$$\begin{bmatrix} g \\ 0 \\ 0 \end{bmatrix} = \mathrm{A} \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

N. B. : Puisque $\mathrm{A}^{-1} \begin{bmatrix} g \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, on voit que g divise a, b et c. L'entier g doit donc être le pgcd de a, b, c. Mais ceci ne sera pas utile par la suite.

Voici une preuve abstraite du théorème.

Démonstration. On considère l'ensemble $\mathrm{E} \subseteq \mathbb{N}^3$ des triplets (a', b', c') tels qu'il existe une matrice $\mathrm{B} \in \mathrm{G}_3$ vérifiant

$$\begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} = \mathrm{B} \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Soit dans E un élément (a_0, b_0, c_0) tel que $a_0+b_0+c_0$ soit le plus petit possible et soit B_0 une matrice correspondante dans G_3. Montrons que deux des trois éléments a_0, b_0, c_0 sont nécessairement nuls. Sinon par exemple $b_0 \geq a_0 > 0$. Si $b_1 = b_0 - a_0 q$ est le reste de la division de b_0 par a_0, on obtient :

$$\begin{bmatrix} a_0 \\ b_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -q & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -q & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} B_0 \begin{bmatrix} a \\ b \\ c \end{bmatrix} = B_1 \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

avec $B_1 \in G_3$. Et $a_0 + b_1 + c_0 < a_0 + a_0 + c_0 \leq a_0 + b_0 + c_0$, ce qui est absurde.

Ainsi un seul des trois éléments a_0, b_0, c_0 est non nul. Si ce n'est pas a_0 on le ramène en première position par une matrice convenable de déterminant 1. □

Questions :

1. La preuve par l'absurde ci-dessus donne en fait l'idée d'un algorithme pour réaliser le théorème, c'est-à-dire pour calculer A et g à partir de a, b, c. Montrer ceci sur l'exemple suivant : $(a, b, c) = (90, 126, 210)$.
2. Écrire un algorithme précis pour réaliser le théorème.
3. Écrire une preuve par récurrence sur $m = a + b + c$ pour démontrer le théorème.

Exercice 3.3.2. Voici un extrait de la lettre de Fermat à Carcavi datée d'aout 1659, dans laquelle l'expression de descente infinie apparait pour la première fois.

Relation des nouvelles découvertes en la science des nombres.

... 1. Et pour ce que les méthodes ordinaires, qui sont dans les Livres, étaient insuffisantes à démontrer des propositions si difficiles, je trouvai enfin une route tout à fait singulière pour y parvenir.

J'appelai cette manière de démontrer la *descente infinie* ou *indéfinie*, etc. ; je ne m'en servis au commencement que pour démontrer les propositions négatives, comme, par exemple :

Qu'il n'y a aucun nombre, moindre de l'unité qu'un multiple de 3, qui soit composé d'un carré et du triple d'un autre carré ;

Qu'il n'y a aucun triangle rectangle en nombres dont l'aire soit un nombre carré.

La preuve se fait par ἀπαγωγὴν εἰς ἀδύνατον en cette manière :

S'il y avait aucun triangle rectangle en nombres entiers qui eût son aire égale à un carré, il y aurait un autre triangle moindre que celui-là qui aurait la même propriété. S'il y en avait un second, moindre que le premier, qui eût la même propriété, il y en aurait, par un pareil raisonnement, un troisième, moindre que ce second, qui aurait la même propriété, et enfin un quatrième, un cinquième, etc. à l'infini en descendant. Or est-il qu'étant donné un nombre, il n'y en a point infinis en descendant moindres que celui-là (j'entends parler toujours des nombres entiers). D'où on conclut qu'il est donc impossible qu'il y ait aucun triangle rectangle dont l'aire soit carrée.

On infère de là qu'il n'y en a non plus en fractions dont l'aire soit carrée ; car, s'il y en avait en fractions, il y en aurait en nombres entiers, ce qui ne peut pas être, comme il se peut prouver par la *descente*.

Je n'ajoute pas la raison d'où j'infère que, s'il y avait un triangle rectangle de cette nature, il y en aurait un autre de même nature, moindre que le premier, parce que le discours en seroit trop long et que c'est là tout le mystère de ma méthode. Je serai bien aise que les Pascal et les Roberval et tant d'autres savants la cherchent sur mon indication.

Je fus longtemps sans pouvoir appliquer ma méthode aux questions affirmatives, parce que le tour et le biais pour y venir est beaucoup plus malaisé que celui dont je me sers aux négatives. De sorte que, lorsqu'il me fallut démontrer que *tout nombre premier, qui surpasse de l'unité un multiple de 4, est composé de deux carrés*, je me trouvai en belle peine. Mais enfin une méditation diverses fois réitérée me donna les lumières qui me manquaient, et les questions affirmatives passèrent par ma méthode, à l'aide de quelques nouveaux principes qu'il y fallut joindre par nécessité. Ce progrès

de mon raisonnement en ces questions affirmatives est tel : si un nombre premier pris à discrétion, qui surpasse de l'unité un multiple de 4, n'est point composé de deux carrés, il y aura un nombre premier de même nature, moindre que le donné, et ensuite un troisième encore moindre, etc. en descendant à l'infini jusques à ce que vous arriviez au nombre 5, qui est le moindre de tous ceux de cette nature, lequel il s'ensuivrait n'être pas composé de deux carrés, ce qu'il est pourtant. D'où on doit inférer, par la déduction à l'impossible, que tous ceux de cette nature sont par conséquent composés de deux carrés.

Traduisez les théorèmes dans un langage que vous maitrisez, détaillez les descentes infinies décrites par Fermat et précisez à chaque fois le cœur de la descente infinie, c'est-à-dire l'énoncé qui permet de la réaliser. Vous pouvez consulter https://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_de_Fermat_sur_les_triangles_rectangles et https://fr.wikipedia.org/wiki/Th%C3%A9or%C3%A8me_des_deux_carr%C3%A9s_de_Fermat pour la preuve de l'énoncé en question, éludée par Fermat.

Exercices pour le chapitre 4

4.3

Exercice 4.3.1. Dans le manuel dont nous avons reproduit des extraits dans la section 4.1, on peut trouver le texte ci-dessous à la suite de la définition 6 de la dimension.

> [...] dans un espace vectoriel de dimension finie tous les sous-espaces sont de dimension finie, et plus précisément :
>
> **Propriété 4.** *Tout sous-espace* E' *d'un espace vectoriel* E *de dimension finie* n *est de dimension finie* $n' \leq n$*, et l'égalité* $n' = n$ *entraîne* $\mathrm{E}' = \mathrm{E}$*.*
>
> p vecteurs linéairement indépendants de E' sont également des vecteurs linéairement indépendants de E et on a : $p \leq n$. Le nombre d'éléments d'un système de vecteurs linéairement indépendants de E' étant ainsi majoré par n, il existe un nombre maximum n' d'éléments d'un système de vecteurs linéairement indépendants de E'.
>
> Alors, si $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_{n'}$ sont n' vecteurs linéairement indépendants de E', pour tout vecteur V de E' on a :
>
> $$\lambda \mathrm{V} + \lambda_1 \mathrm{V}_1 + \cdots + \lambda_{n'} \mathrm{V}_{n'} = 0.$$
>
> où les scalaires $\lambda, \lambda_1, \ldots, \lambda_{n'}$ ne sont pas tous nuls. $\mathrm{V}_1 \ldots, \mathrm{V}_{n'}$ étant linéairement indépendants, λ ne peut être nul, d'où :
>
> $$\mathrm{V} = -\frac{\lambda_1}{\lambda}\mathrm{V}_1 - \cdots - \frac{\lambda_{n'}}{\lambda}\mathrm{V}_{n'},$$
>
> ce qui montre que $\mathrm{V}_1, \ldots, \mathrm{V}_{n'}$ forment une base de E' et par conséquent $\dim \mathrm{E}' = n' \leq n$.
>
> Si $n' = n$, $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_n$ constituent une base de E ; tout vecteur de E s'écrivant $x_1\mathrm{V}_1 + \cdots + x_n\mathrm{V}_n$ est un vecteur de E' ; il en résulte $\mathrm{E}' = \mathrm{E}$.
>
> EXERCICES : 19, 20, 21, 22.

Voici les exercices annoncés ci-dessus.

Exercices

. . . . . .

19. Dans l'espace $\mathrm{E} = \mathbb{C}^2$, montrer que les vecteurs $\mathrm{V} = (\mathrm{X}, \mathrm{Y})$ tels que : $\mathrm{X} + \mathrm{Y} = 0$, forment un sous-espace E' de dimension un ; en donner une base.

20. Dans l'espace $\mathrm{E} = \mathbb{C}^3$, montrer que les vecteurs $\mathrm{V} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ tels que : $\mathrm{X} + \mathrm{Y} + \mathrm{Z} = 0$, forment un sous-espace E' de dimension deux ; en donner une base.

21. *a*) Dans l'espace $\mathrm{E} = \mathbb{R}^4$ montrer que les vecteurs $\mathrm{V} = (x_1, x_2, x_3, x_4)$ tels que : $x_1 + x_3 = x_2 + x_4 = 0$, forment un sous-espace E'.
b) Déterminer la dimension de E' et en trouver une base.

22. Dans un espace vectoriel E de dimension quatre sur le corps $\mathbb{K}$, on donne un sous-espace E' et une base $\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3$ de E'. Montrer que, si V est un vecteur de E n'appartenant pas à E', les vecteurs $\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3, \mathrm{V}$ forment une base de E.

Les exercices sont suivis de « résultats ou indications ». Nous les reproduisons, ainsi qu'une définition donnée quelques pages auparavant dans ce manuel pour que vous sachiez ce que sont « les axiomes (6) et (7) ».

Définition 1. On appelle *sous-espace vectoriel* d'un espace vectoriel E sur le corps $\mathbb{K}$ ou en abrégé *sous-espace*, un sous-ensemble E' de E satisfaisant aux deux propriétés :

(6) *si* $\mathrm{V} \in \mathrm{E}'$ *et* $\mathrm{V}' \in \mathrm{E}'$ *:* $\mathrm{V} + \mathrm{V}' \in \mathrm{E}'$ *;*

(7) *si* $\mathrm{V} \in \mathrm{E}'$ *et* $\lambda \in \mathbb{K}$ *:* $\lambda \mathrm{V} \in \mathrm{E}'$*.*

. . . . . .

Résultats ou indications

. . . . . .

19. Vérifier les axiomes (6) et (7). $\{0\} \neq \mathrm{E}' \subset \mathrm{E} \Rightarrow 0 < \dim \mathrm{E}' < 2$ (propriété 4, $\Rightarrow \dim \mathrm{E}' = 1$). Base : $\mathrm{U} = (1, -1)$.

20. Vérifier les axiomes (6) et (7). $\{0\} \neq \mathrm{E}' \subset \mathrm{E} \Rightarrow 0 < \dim \mathrm{E}' < 3$ (propriété 4). $\mathrm{V}_1 = (1, -1, 0) \in \mathrm{E}'$, $\mathrm{V}_2 = (1, 0, -1) \in \mathrm{E}'$ sont linéairement indépendants $\Rightarrow 2 \leq \dim \mathrm{E}'$ (théorème de la dimension) ; $\dim \mathrm{E}' = 2$, base de E' : $\mathrm{V}_1, \mathrm{V}_2$.

21. *a*) Vérifier les axiomes (6) et (7).
b) $\mathrm{U}_1 = (1, 0, -1, 0) \in \mathrm{E}'$ et $\mathrm{U}_2 = (0, 1, 0, -1) \in \mathrm{E}'$ sont linéairement indépendants. De plus, tout vecteur $\mathrm{V} = (x_1, x_2, x_3, x_4) \in \mathrm{E}'$ s'écrit :

$$\mathrm{V} = x_1 \mathrm{U}_1 + x_2 \mathrm{U}_2 \,;$$

$\dim \mathrm{E}' = 2$, base : $\mathrm{U}_1, \mathrm{U}_2$.

22. Écrire

$$\lambda_1 \mathrm{V}_1 + \lambda_2 \mathrm{V}_2 + \lambda_3 \mathrm{V}_3 + \lambda \mathrm{V} = 0. \qquad \mathrm{V} \notin \mathrm{E}' \Rightarrow \lambda = 0 \Rightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0.$$

$\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3, \mathrm{V}$ sont linéairement indépendants ; appliquer le théorème de la dimension.

Proposez des solutions alternatives de ces exercices qui se placent dans le cadre concret de la résolution des systèmes linéaires. L'appel à la propriété 4 s'avère-t-il nécessaire ? Proposez une discussion de la démonstration de cette propriété à la lumière du cours et de vos solutions alternatives des exercices. Sur quel fondement peut-on y affirmer l'existence d'« un nombre maximum n' d'éléments d'un système de vecteurs linéairement indépendants de E' » ?

Exercices pour le chapitre 5

5.1

Exercice 5.1.1 (Les rapports de grandeurs)**.** Dans cet exercice, nous allons étudier la définition eudoxienne du rapport de deux grandeurs introduite dans la section 4.1 du cours avec l'aide d'extraits d'un livre d'Augustus De Morgan (1836), *The connexion of number and magnitude : an attempt to explain the fifth book of Euclid*, qui est aussi téléchargeable. Il interprète la notion de rapport comme la donnée d'une *échelle relative de multiples*. J'ai découvert ce livre dans un article de Sébastien Gandon (2009), téléchargeable, dont j'ai repris et complété les traductions.

Dans tout cet exercice, on peut penser une grandeur comme un nombre réel strictement positif. Cependant, on commet alors un anachronisme : les nombres réels ont été créés quelques décennies plus tard par Richard Dedekind comme une arithmétisation des grandeurs, c'est-à-dire comme une manière de rendre compte des grandeurs par les nombres. Euclide et De Morgan, pour leur part, distinguent rigoureusement

— les nombres qui pour eux sont les multiples d'une unité, c'est-à-dire les entiers ;

— les grandeurs qui pour eux sont d'un genre donné, par exemple du genre des longueurs, et dont il suffit de savoir qu'on peut les ajouter et qu'on peut leur retirer une grandeur plus petite : on peut procéder aussi ainsi avec les grandeurs physiques comme la masse ;

— les rapports de grandeurs.

(1) Premier extrait (De Morgan 1836, pages 3–14) : l'algorithme d'Euclide. Étant données deux grandeurs A et B, l'algorithme construit des nombres entiers p et q tels que $p\mathrm{A} - q\mathrm{B}$ ou $q\mathrm{B} - p\mathrm{A}$ est une grandeur arbitrairement petite.

> Que A représente une grandeur – non pas comme en algèbre le nombre d'unités qu'elle contient, mais la grandeur elle-même – de sorte que si c'est, par exemple, de poids que nous parlons, A n'est pas un nombre de kilogrammes, mais le poids lui-même. Soit B une autre grandeur du même genre ; nous pouvons alors obtenir une troisième grandeur, soit en mettant les deux grandeurs ensemble, soit en enlevant de la plus grande une grandeur égale à la plus petite. Que celles-ci soient représentées par $\mathrm{A} + \mathrm{B}$ et $\mathrm{A} - \mathrm{B}$, A étant supposée la plus grande. Nous pouvons aussi construire d'autres grandeurs en prenant un nombre de grandeurs chacune égale à A, et en en mettant un nombre quelconque ensemble. Ainsi nous avons
>
> | | | |
> |---|---|---|
> | $\mathrm{A} + \mathrm{A}$ | que nous abrégeons en | 2A |
> | $\mathrm{A} + \mathrm{A} + \mathrm{A}$ | ..... | 3A |
> | $\mathrm{A} + \mathrm{A} + \mathrm{A} + \mathrm{A}$ | ..... | 4A |
>
> et ainsi de suite. Nous avons ainsi un ensemble de grandeurs, dépendant de A, et toutes connues si A est connue ; c'est-à-dire
>
> A 2A 3A 4A 5A etc.

que nous pouvons mener aussi loin qu'il nous plait. Celles-ci (sauf la première) sont distinguées de toutes les autres grandeurs par le nom de *multiples* de A ; et il est évident qu'elles croissent en continu. Que ces dernières soient appelées l'*échelle des multiples* de A.

. . . . . .

Proposition VI. Que soient données deux grandeurs du même genre, A et B, et que soient formées les échelles des multiples

$$\text{A}, \quad 2\text{A}, \quad 3\text{A}, \quad \text{etc.} \qquad \text{B}, \quad 2\text{B}, \quad 3\text{B}, \quad \text{etc.},$$

alors une de ces deux choses doit être vraie ; OU BIEN il y a des multiples dans la première échelle qui sont égaux à des multiples de la seconde échelle ; OU il y a des multiples dans la première échelle qui sont aussi près que nous voulons de multiples (peut-être pas les mêmes) dans le second ensemble : c'est-à-dire, nous pouvons en trouver un du premier ensemble, disons mA, qui sera soit égal à un autre du second ensemble, disons nB, ou dépassera ou manquera celui-ci d'une quantité inférieure à une quantité donnée Z, que nous pouvons prendre aussi petite qu'il nous plait.

. . . . . .

Cette proposition ne prouve rien pour une seule grandeur, mais elle établit deux relations apparemment très différentes entre grandeurs considérées en paires. Il peut y avoir des cas où la première alternative est finalement établie ; et il peut y avoir des cas où elle elle n'est jamais établie. Nous allons d'abord considérer le cas où la première alternative est établie.

. . . . . .

Le terme *mesure* est utilisé inversement à multiple, ainsi : si A est un multiple de M, M est dit être une mesure de A. Donc, dans le cas que nous sommes en train de considérer maintenant, A et B ont une *commune mesure*, et sont dits être *commensurables*.

. . . . . .

Il reste, ensuite, seul le second cas à considérer, qui comme il est maintenant évident contient ces grandeurs (s'il y en a de telles) qui n'ont pas de commune mesure, quelle qu'elle soit.

. . . . . .

Nous allons maintenant examiner l'effet de substitutions successives à partir du début, en faisant d'abord la remarque suivante : S'il y a deux quantités incommensurables quelconques A et B, dont A est la plus grande, alors il suit un ensemble interminable de nombres entiers, $\beta, \beta', \beta'', \ldots$ qui ne sont sujets à aucune loi particulière, mais peut peuvent être trouvés quand A et B sont donnés ; et un ensemble interminable de quantités $\text{A}, \text{B}, \text{B}', \text{B}'', \ldots$ liées aux premières par cette loi que A contient B entre β et $\beta + 1$ fois ; B contient B$'$ entre β' et $\beta' + 1$ fois, et ainsi de suite.

Nous avons $\text{B}' = \text{A} - \beta\text{B}$

$$\text{B}'' = \text{B} - \beta'\text{B}'$$

ou $$\text{B}'' = (\beta\beta' + 1)\text{B} - \beta'\text{A}$$

$$\begin{aligned}\text{B}''' &= \text{B}' - \beta''\text{B}'' = \text{A} - \beta\text{B} - (\beta\beta' + 1)\beta''\text{B} + \beta'\beta''\text{A} \\ &= (\beta'\beta'' + 1)\text{A} - (\beta\beta'\beta'' + \beta + \beta'')\text{B}\end{aligned}$$

et ainsi nous continuons de représenter les restes en alternance, dans la forme $p\text{A} - q\text{B}$ et $q\text{B} - p\text{A}$. Nous pouvons facilement trouver la loi des coefficients, comme suit :

. . . . . .

En fait, De Morgan avait décrit auparavant les substitutions successives évoquées dans le cas de deux grandeurs commensurables. Plutôt que de vous traduire ce passage, je vous propose de lire la première définition et la deuxième proposition du livre X des *Éléments* d'Euclide, qui implémente une version géométrique de l'algorithme d'Euclide, appelée « anthyphérèse », ce qui veut littéralement dire « soustraction alternée » en grec. Par grandeur, il faut ici entendre toute sorte de grandeur, comme des segments et des surfaces, qui sont infiniment divisibles contrairement aux nombres.

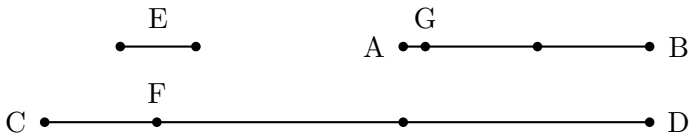
Définition X.1. Sont dites grandeurs *commensurables* celles qui sont mesurées par la même mesure, et *incommensurables*, celles dont aucune commune mesure ne peut être produite.

Proposition X.2. *Si, de deux grandeurs inégales proposées la plus petite étant retranchée de la plus grande de façon réitérée et en alternance, le dernier reste ne mesure jamais le reste précédent, les grandeurs seront incommensurables.*

En effet AB, CD étant deux grandeurs inégales et AB la plus petite, la plus petite étant retranchée de la plus grande de façon réitérée et en alternance, que le reste ne mesure jamais le précédent. Je dis que les grandeurs AB, CD sont incommensurables.

En effet si elles sont commensurables une certaine grandeur les mesurera. Qu'elle les mesure, si c'est possible, et que ce soit E.

Et d'une part que AB mesurant FD, il reste CF plus petite qu'elle, d'autre part que CF mesurant BG, il reste AG plus petite qu'elle et que ceci soit toujours poursuivi jusqu'à ce qu'il reste une certaine grandeur plus petite que E. Que ceci se produise et qu'il reste AG plus petite que E.



Et puisque E mesure AB, mais que AB mesure DF, E mesurera donc aussi DF. Or elle mesure aussi le tout CD ; et donc elle mesurera le reste CF. Mais CF mesure BG ; et donc E mesure BG. Or elle mesure aussi le tout AB ; et donc elle mesurera le reste AG, la plus grande, le plus petit. Ce qui est impossible. Donc aucune grandeur ne mesurera les grandeurs AB, CD ; les grandeurs AB, CD sont donc incommensurables (Df. X.1).

Donc si de deux grandeurs inégales…, etc.

1. Reformulez d'abord cette proposition X.2 et sa démonstration dans un langage que vous maitrisez. Vous satisfait-elle ? Puis comparez-la aux substitutions successives évoquées dans l'extrait du livre de De Morgan.
2. Déterminez maintenant la loi des coefficients p et q dans la représentation de B′, B″, B‴, … sous la forme $p\mathrm{A} - q\mathrm{B}$ ou $q\mathrm{B} - p\mathrm{A}$:
   - notez cette suite plutôt $\mathrm{B}^{(1)}, \mathrm{B}^{(2)}, \mathrm{B}^{(3)}, \ldots$;
   - tirez du texte p_1, q_1, p_2, q_2 tels que $\mathrm{B}^{(1)} = p_1\mathrm{A} - q_1\mathrm{B}$ et $\mathrm{B}^{(2)} = q_2\mathrm{B} - p_2\mathrm{A}$, puis déterminez p_3 en fonction de p_1, p_2, β'', et q_3 en fonction de q_1, q_2, β'' tels que $\mathrm{B}^{(3)} = p_3\mathrm{A} - q_3\mathrm{B}$;
   - puis faites de même pour déterminer p_4, q_4 tels que $\mathrm{B}^{(4)} = q_4\mathrm{B} - p_4\mathrm{A}$;
   - pouvez-vous maintenant écrire comment on obtient les coefficients p_n, q_n par récurrence ?

(2) Deuxième extrait (De Morgan 1836, pages 15–16) : propriétés des coefficients p et q.

$$
\begin{array}{rcl}
p_1\mathrm{A} \text{ est plus grand que } & q_1\mathrm{B} & \text{mais plus petit que } (q_1+1)\mathrm{B} \\
p_2\mathrm{A} \text{ est plus petit que } & q_2\mathrm{B} & \text{mais plus grand que } (q_2-1)\mathrm{B} \\
p_3\mathrm{A} \text{ est plus grand que } & q_3\mathrm{B} & \text{mais plus petit que } (q_3+1)\mathrm{B} \\
\text{etc.} & \text{etc.} & \text{etc.}
\end{array}
$$

Par conséquent, il apparait que

$$
\begin{array}{rl}
\text{A est plus grand que} & \dfrac{q_1}{p_1}\mathrm{B} \\
\text{plus petit que} & \dfrac{q_2}{p_2}\mathrm{B} \\
\text{plus grand que} & \dfrac{q_3}{p_3}\mathrm{B} \\
\text{plus petit que} & \dfrac{q_4}{p_4}\mathrm{B} \text{ etc. à l'infini.}
\end{array}
$$

. . . . . .

2. [. . .] Par conséquent, pour arranger toutes les fractions ainsi considérées, nous devons les écrire ainsi,

$$\frac{q_1}{p_1}\quad\frac{q_3}{p_3}\quad\frac{q_5}{p_5}\quad\ldots\quad\ldots\quad\frac{q_6}{p_6}\quad\frac{q_4}{p_4}\quad\frac{q_2}{p_2}$$

3. Nous pouvons ainsi rapprocher deux fractions autant qu'il nous plait :

. . . . . .

Pouvez-vous prouver les deux assertions qui précèdent les points 2 et 3 ? Ces deux points 2 et 3 peuvent aussi être prouvés : essayez, mais il faut travailler un peu davantage. Si vous n'y arrivez pas, ne soyez pas déçu et admettez-les.

(3) Troisième extrait (De Morgan 1836, pages 19–20) : un exemple. Euclide dit qu'un segment A est « coupé en extrême et moyenne raison » s'il est composé de deux parties B et B′ telles que ci-dessous. Tout ceci est lié au nombre d'or.

Supposons, comme un exemple, que nous ayons deux grandeurs A et B qui, soumises au processus ci-dessus, donnent

$$\mathrm{A} = \mathrm{B} + \mathrm{B}', \quad \mathrm{B} = \mathrm{B}' + \mathrm{B}'', \quad \mathrm{B}' = \mathrm{B}'' + \mathrm{B}''', \quad \text{etc. à l'infini,}$$

ou supposons que $\beta = 1, \quad \beta' = 1, \quad \beta'' = 1, \quad$ etc. à l'infini.

Par conséquent les différentes valeurs de p et q sont $p_1 = 1$, $p_2 = 1$, $p_3 = 2$, etc. comme dans ce tableau,

	1	2	3	4	5	6	7	8	9	10	11	12	
p	1	1	2	3	5	8	13	21	34	55	89	144	etc.
q	1	2	3	5	8	13	21	34	55	89	144	233	etc.

ou $\quad \mathrm{A} > \mathrm{B} < 2\mathrm{B} > \frac{3}{2}\mathrm{B} < \frac{5}{3}\mathrm{B} > \frac{8}{5}\mathrm{B} < \frac{13}{8}\mathrm{B} \quad$ etc.

Par conséquent

A	se situe entre	B	et	2B	
2A	. . . . .	3B	. .	4B	
3A	. . . . .	4B	. .	5B	
4A	. . . . .	6B	. .	7B	
5A	. . . . .	8B	. .	9B	etc.

. . . . . .

Nous pouvons donc former ce que nous pouvons appeler une échelle *relative* de multiples en notant les multiples de A, et en y insérant les multiples de B à leurs places propres ; ou *vice versa*. Dans le cas considéré à l'instant, le commencement de cette échelle est

$$\mathrm{B}, \mathrm{A}, 2\mathrm{B}, 3\mathrm{B}, 2\mathrm{A}, 4\mathrm{B}, 3\mathrm{A}, 5\mathrm{B}, 6\mathrm{B}, 4\mathrm{A}, 7\mathrm{B}, 8\mathrm{B}, 5\mathrm{A}, 9\mathrm{B}, \text{etc.}$$

que nous pouvons continuer aussi loin qu'il nous plait par simple arithmétique.

1. Expliquez comment il convient de lire le tableau des p et q.
2. Comment convient-il de comprendre la ligne $\mathrm{A} > \mathrm{B} < 2\mathrm{B} > \frac{3}{2}\mathrm{B} < \frac{5}{3}\mathrm{B} > \frac{8}{5}\mathrm{B} < \frac{13}{8}\mathrm{B}$ que l'on n'oserait plus écrire ainsi aujourd'hui ? Comment en tire-t-on que A est entre B et 2B, que 2A est entre 3B et 4B, etc. ?
3. Entre quels multiples successifs de B peut-on placer 100A ?
4. Illustrez le commencement de l'échelle relative des multiples de A et de B en traçant une droite sur laquelle vous désignerez les quatorze premiers multiples de B par une succession de points équidistants et en plaçant les huit premiers multiples de A entre les bons multiples de B.

(4) Quatrième extrait (De Morgan 1836, pages 23 et 51) : échelle relative et rapport.

> Qu'est-ce donc, alors, qui *est* donné lorsque l'échelle est donnée ? Non pas les grandeurs elles-mêmes ; car si l'échelle appartient à A et B, elle appartient aussi à chacun des cas infinis de $\frac{p}{q}$A et $\frac{p}{q}$B. L'échelle, par conséquent, définit seulement une relation entre les grandeurs telle qu'elle appartient à 2A et 2B, 3A et 3B, etc., autant qu'à A et B. Il est usuel d'appeler cette relation la *proportion* entre les deux quantités dans la vie courante, et leur *rapport* en mathématiques ; dans Euclide le terme est λόγος [logos].
>
> . . . . . .
>
> *Récapitulation.* Par le *rapport* de A à B, nous entendons (sans autre spécification pour l'instant) une relation entre les grandeurs A et B, déterminée par la façon dont les multiples de A sont distribués, si chacun est écrit entre les plus proches multiples de B en grandeur. C'est-à-dire que si B, 2B, 3B, etc., et A, 2A, 3A, etc., sont formés, et si A se trouve entre B et 2B, 2A entre 2B et 4B, et ainsi de suite, l'échelle relative
>
> B, A, 2B, 3B, 2A, 4B, etc.
>
> doit être le seul élément déterminant du rapport, de sorte qu'il n'y ait rien que l'ordre de cette échelle dont le rapport dépende.

Voici la définition euclidienne du rapport de deux grandeurs dans le livre V des *Éléments*.

> **Définition V.5.** Des grandeurs sont dites *être dans le même rapport*, une première relativement à une deuxième et une troisième relativement à une quatrième quand des équimultiples de la première et de la troisième ou simultanément dépassent, ou sont simultanément égaux ou simultanément inférieurs à des équimultiples de la deuxième et de la quatrième, selon n'importe quelle multiplication, chacun à chacun, et pris de manière correspondante.

1. Vérifiez que cette définition correspond bien à la définition donnée dans la première section du chapitre 4 du cours. Est-ce qu'elle est bien équivalente à la définition de De Morgan ? Motivez votre réponse.
2. Utilisez la notion d'échelle relative pour décrire dans quelle mesure la définition du rapport nécessite l'infini.
3. L'approche de De Morgan vous a-t-elle éclairé sur la notion de rapport ? Dites pourquoi et précisez, le cas échéant, ce qui reste obscur pour vous.

Exercice 5.1.2. Voici un extrait de la traduction du second livre du commentaire d'al-Khayyām sur les *Éléments* d'Euclide, *Exposé sur le rapport et la notion de proportionnalité, et sur leur véritable nature*. Ce commentaire date de 1077. Les mots entre chevrons sont ceux que le traducteur a rajoutés pour la clarté de la phrase.

> § 2. [. . .] En effet, chaque fois que l'on a deux grandeurs homogènes, ou bien elles sont égales, ou bien elles sont différentes. Et la différence a des limites et des divisions. C'est-à-dire que la plus petite est ou bien une partie de la plus grande – en effet elle la mesure et la remplit exhaustivement lors de la mise en relation –, ou bien elle est des parties, ou bien elle est selon une autre manière. Une des propriétés de la quantité est de <pouvoir> y considérer l'égalité et l'inégalité. Le rapport est donc cette considération elle-même lorsque l'on met en relation les <grandeurs> homogènes, ainsi que la considération d'une autre chose qui est jointe à elle, à savoir, la grandeur de ce rapport en tant qu'il est rapport entre grandeurs.

§ 3. Cela est particulièrement clair dans le cas des choses numériques ; et le premier lieu où l'on trouve cette notion, je veux dire le rapport, c'est dans les choses numériques. C'est-à-dire que l'on considéra les nombres rapportés les uns aux autres, et l'on trouva qu'ils étaient ou bien égaux entre eux, ou bien inégaux. (Cela fait partie des propriétés de la quantité.) On considéra ensuite l'inégalité, et l'on trouva

— ou bien que la plus petite mesurait la plus grande – par exemple trois relativement à neuf. On rechercha alors la quantité de la mesure de neuf par trois, et l'on trouva que c'était trois. Donc trois mesure neuf trois fois. On dériva alors de cette notion un nom conformément aux langues, et l'on dit : *C'est le tiers. Donc le rapport entre trois et neuf est le fait d'être un tiers.* Et c'est là la considération de l'égalité et de l'inégalité jointe à une autre considération, comme nous l'avons expliqué. Et le rapport entre neuf et trois est le trois multiplicatif. Mais l'on ne dériva pas de nom pour cela, et l'on se contenta du premier. Cela dépend de celui qui a posé <les règles> du langage.

— ou bien qu'elle ne mesurait pas la plus grande – par exemple le rapport de deux à sept. On les divisa alors en parties qui mesuraient tout ensemble sept et deux, et l'on ne trouva pas un autre nombre : au contraire, on trouva l'unité. On dit alors : *Le rapport de deux à sept est deux septièmes.*

On démontra ensuite que les nombres les plus petits étaient soit une partie, soit des parties des plus grands.

§ 4. Et lorsque l'on vit que le nombre était de même genre que la grandeur du fait qu'ils se subdivisaient tous deux sous le genre de la quantité, on rechercha aussi cette notion dans les grandeurs. Et l'on y trouva en plus de ces deux divisions-là une autre division. C'est-à-dire que les grandeurs ne sont pas composées de parties indivisibles, et il n'y a pas de limite définie à leur division comme pour le nombre.

§ 5. Car le nombre est composé de parties indivisibles, i. e. les unités. Chaque fois que l'on a deux nombres différents, on retranche du plus grand tous les multiples du plus petit, et l'on arrivera à un reste plus petit que le plus petit nombre ; on retranche ensuite du plus petit tous les multiples du reste, et l'on arrivera à un reste plus petit que le premier reste ; et l'on ne cesse de procéder ainsi ; on arrivera alors nécessairement à un reste qui mesure le reste qui le précède, ou bien à l'unité. C'est-à-dire que les deux nombres sont finis et donnés, et sont composés d'unités indivisibles. Et quand nous disons *composé* dans la définition du nombre, c'est de par la nécessité de l'expression. Car *combinaison*, *multiplicité*, *collection*, et *nombre* ont tous le même sens. Il a déjà mentionné cela en partie au début du septième <Livre> de son ouvrage. Et il te sera possible de le reconnaître avec un minimum de réflexion.

§ 6. Quant aux grandeurs, elles ne sont pas composées de parties indivisibles, et il n'y a pas de limite déterminée à leur division. Ainsi, cette notion ne s'ensuit pas pour tous les cas ; et il n'est pas nécessaire que l'on parvienne inévitablement à l'unité (puisqu'il n'y a pas dans leur cas d'unité), ni à un reste qui mesure celui qui le précède. Et ce n'est que par une démonstration que l'on saura si cette notion y existe. Euclide en a déjà longuement parlé dans le dixième <Livre> de son ouvrage, mais nous n'en avons absolument pas besoin dans cette explication.

§ 7. Et puisqu'il en est ainsi, il ne s'ensuivra pas nécessairement, chaque fois que l'on aura deux grandeurs, que la plus petite soit ou bien une partie de la plus grande, ou bien des parties. Il sera au contraire possible qu'elle soit selon une autre espèce qui n'est pas numérique mais propre aux grandeurs. Et si quelqu'un dit : Cette troisième division n'existe absolument pas ; au contraire, elle fait partie des deux divisions numériques ! nous lui répondrons et nous lui dirons : Il n'y aura aucun mal à ce que nous considérions les lois du rapport et de la proportionnalité dans le cas des grandeurs de ces trois points de vue. Si ensuite la division est abrogée par une démonstration, il n'y aura aucun reproche à nous faire. Mais si elle n'est abrogée, nous aurons alors avancé et épuisé toutes les divisions. C'est là un mystère d'où l'on découvrira des mystères de logique extrêmement profonds. Comprends-le donc.

§ 8. Il mentionne ensuite la proportionnalité en disant : *C'est la similitude des rapports.* C'est là, quant au langage, des propos excellents, si ce n'est qu'il s'est complètement écarté de la véritable nature de la proportionnalité en commentant cette expression. C'est-à-dire qu'il a dit : *Si l'on a quatre grandeurs homogènes, que l'on prenne à l'infini des équimultiples de la première et de la*

troisième, et des équimultiples de la deuxième et de la quatrième, quels qu'il soient, et qu'on les compare ; et que, lorsque le multiple de la première excède le multiple de la deuxième, le multiple de la troisième excède le multiple de la quatrième, et que lorsqu'il lui est égal, il lui est aussi égal, et que lorsqu'il est plus petit que lui, il est plus petit que lui, si on les compare successivement ; on dira alors que le rapport de la première à la deuxième est égal au rapport de la troisième à la quatrième. Et qu'elles soient appelées proportionnelles.

§ 9. Mais cela n'est pas construit à partir de la proportionnalité véritable. Ne vois-tu pas que si quelqu'un pose une question, disant : Quatre grandeurs sont proportionnelles selon la proportionnalité euclidienne, et la première est la moitié de la deuxième ; est-ce que la troisième sera alors égale à la moitié de la quatrième, ou non ? Comment sera-t-il alors possible de démontrer que la troisième est aussi égale à la moitié de la quatrième par la méthode d'Euclide ? Et si l'on répond en disant : Il faut, si la première est égale à la moitié de la deuxième, que la troisième soit égale à la moitié de la quatrième, en raison de l'existence de la proportionnalité ; quelle est la démonstration que l'on a pour dire que ce qu'Euclide a expliqué appartient aux conséquences nécessaires de la véritable proportionnalité ?

. . . . . .

§ 12. Je dis : J'ai conçu la véritable nature du rapport entre grandeurs. C'est-à-dire que chaque fois que l'on a deux grandeurs, ou bien l'une d'elles est égale à l'autre, ou bien elle ne l'est pas. Et celle qui est inégale sera ou bien une partie de l'autre, ou bien des parties. (Ces trois constituent le rapport numérique.) Ou bien elle sera selon une autre sorte propre à la géométrie, comme nous l'avons déjà expliqué précédemment.

§ 13. Et si l'on a quatre grandeurs, et que la première est égale à la deuxième, et la troisième égale à la quatrième ; ou que la première est une partie de la deuxième, et la troisième cette même partie de la quatrième ; ou que la première est des parties de la deuxième, et la troisième ces mêmes parties de la quatrième ; alors le rapport de la première à la deuxième sera inévitablement égal au rapport de la troisième à la quatrième. Ce rapport est numérique.

§ 14. Et s'il n'en est pas selon ces trois manières, mais, que l'on retranche de la deuxième tous les multiples de la première jusqu'à ce que l'on arrive à un reste plus petit que la première, et que de même l'on retranche de la quatrième tous les multiples de la troisième jusqu'à ce que l'on arrive à un reste plus petit que la troisième, et que le nombre des multiples de la première contenus dans la deuxième soit égal au nombre des multiples de la troisième contenus dans la quatrième ; et que l'on retranche de la première tous les multiples du reste de la deuxième jusqu'à ce que l'on arrive à un reste plus petit que le reste de la deuxième, et que de même l'on retranche de la troisième tous les multiples du reste de la quatrième jusqu'à ce que l'on arrive à un reste plus petit que le reste de la quatrième, et que le nombre des multiples du reste de la deuxième soit égal au nombre des multiples du reste de la quatrième ; et que de même l'on retranche du reste de la deuxième tous les multiples du reste de la première, et que l'on retranche du reste de la quatrième tous les multiples du reste de la troisième, et que leur nombre soit égal ; et que de même l'on retranche successivement tous les multiples des restes les uns des autres comme nous l'avons expliqué, et que le nombre de chaque reste de la première et de la deuxième soit indéfiniment égal au nombre du reste correspondant de la troisième et de la quatrième ; alors, le rapport de la première à la deuxième sera inévitablement égal au rapport de la troisième à la quatrième. Voilà la véritable proportionnalité relative au type géométrique.

Commentez ce texte en répondant aux questions suivantes.

1. Selon al-Khayyām, quels sont les rapports possibles entre deux nombres ?
2. Comment comprenez-vous le terme « grandeur » ? En quoi les grandeurs se distinguent-elles des nombres ?
3. Que décrit al-Khayyām dans le § 5 ?
4. Quel nom donne-t-on aujourd'hui à « cette troisième division » du § 7 dont sont capables deux grandeurs ? Donnez des exemples !

5. Comment comprenez-vous la critique qu'al-Khayyām adresse dans le § 9 à la définition d'Euclide rappelée dans le § 8 ?
6. Pourquoi al-Khayyām dit-il à la fin du § 13 que « Ce rapport est numérique » ?
7. Commentez la définition de la « véritable proportionnalité relative au type géométrique » donnée au § 14. L'avez-vous déjà rencontrée ?

Exercice 5.1.3. Voici la traduction par Jean-Paul Dumont d'un extrait du sixième livre de la *Physique* d'Aristote qui présente les paradoxes de Zénon connus sous le nom de la *dichotomie* et de l'*Achille*.

> Les arguments de Zénon contre le mouvement sont au nombre de quatre ; ils causent beaucoup de soucis à ceux qui veulent les résoudre. Le premier argument porte sur l'inexistence du « se mouvoir », compte tenu du fait que le mobile doit d'abord parvenir à la moitié avant d'atteindre le terme de son trajet, argument que nous avons déjà discuté auparavant.
>
> Le second argument est celui qu'on appelle l'*Achille*. Il consiste à dire que le plus lent à la course ne peut pas être rattrapé par le plus rapide, étant donné que le poursuivant doit nécessairement atteindre le point d'où le poursuivi est parti, de telle sorte que le plus lent doit sans cesse avoir une certaine avance. Cet argument est identique à celui de la dichotomie, à cette différence près que ce n'est pas en deux que se trouve divisée la grandeur restante. (*Physique* VI, IX, 239*b*9-20, *Les Présocratiques* 1988, pages 287-288.)

Reformulez ces deux paradoxes dans un langage que vous maitrisez. Quel rapport ont-ils selon vous avec la présentation faite dans le cours ? Voici la critique d'Aristote de ces paradoxes.

> C'est pourquoi l'argument de Zénon admet une prémisse fausse : qu'il n'est pas possible que les grandeurs illimitées soient chacune parcourue ou touchée une par une par les grandeurs illimitées en un temps limité. En effet, *illimité*, rapporté à la longueur et au temps, se dit en deux sens, de même que rapporté, plus généralement, à tout ce qui est continu : car on peut considérer soit l'infini selon la division, soit l'infini selon les extrémités. Alors qu'il n'est pas possible qu'une chose entre en contact dans un temps limité avec des grandeurs illimitées en quantité, cela est possible si ces grandeurs sont illimitées en division. En effet, du point de vue de la divisibilité, le temps lui-même est illimité. Il en résulte que c'est dans un temps illimité, et non pas dans un temps limité, que s'effectue le parcours de l'illimité et que le contact avec les grandeurs illimitées se fait par des grandeurs illimitées, et non pas par des grandeurs limitées. (*Physique* VI, II, 233*a*21-31, *Les Présocratiques* 1988, page 287.)

Commentez cette critique.

5.5

Exercice 5.5.1. Voici une lettre que Jules Richard a adressée à la *Revue générale des sciences pures et appliquées*, qui l'a publiée.

> Dans son numéro du 30 mars 1905, la *Revue* signale certaines contradictions qu'on rencontre dans la théorie générale des ensembles.
>
> Il n'est pas nécessaire d'aller jusqu'à la théorie des nombres ordinaux pour trouver de telles contradictions. En voici une qui s'offre dès l'étude du continu, et à laquelle plusieurs autres se ramèneraient probablement :

Je vais définir un certain ensemble de nombres, que je nommerai l'ensemble E, à l'aide des considérations suivantes :

Écrivons tous les arrangements deux à deux des vingt-six lettres de l'alphabet français, en rangeant ces arrangements par ordre alphabétique, puis, à la suite, tous les arrangements trois à trois, rangés par ordre alphabétique, puis, à la suite, ceux quatre à quatre, etc. Ces arrangements peuvent contenir la même lettre répétée plusieurs fois, ce sont des arrangements avec répétition.

Quel que soit l'entier p, tout arrangement des vingt-six lettres p à p se trouvera dans ce tableau, et comme tout ce qui peut s'écrire avec un nombre fini de mots est un arrangement de lettres, tout ce qui peut s'écrire se trouvera dans le tableau dont nous venons d'indiquer le mode de formation.

La définition d'un nombre se faisant avec des mots, et ceux-ci avec des lettres, certains de ces arrangements seront des définitions de nombres. Biffons de nos arrangements tous ceux qui ne sont pas des définitions de nombres.

Soit u_1 le premier nombre défini par un arrangement, u_2 le second, u_3 le troisième, etc.

On a ainsi, rangés dans un ordre déterminé, *tous les nombres définis à l'aide d'un nombre fini de mots.*

Donc : tous les nombres qu'on peut définir à l'aide d'un nombre fini de mots forment un ensemble dénombrable.

Voici maintenant où est la contradiction. On peut former un nombre n'appartenant pas à cet ensemble. « Soit p, la $n^{\text{ième}}$ décimale du $n^{\text{ième}}$ nombre de l'ensemble E ; formons un nombre ayant zéro pour partie entière, et pour $n^{\text{ième}}$ décimale $p+1$, si p n'est égal ni à 8, ni à 9, et l'unité dans le cas contraire ». Ce nombre N n'appartient pas à l'ensemble E. S'il était le $n^{\text{ième}}$ nombre de l'ensemble E, son $n^{\text{ième}}$ chiffre serait le $n^{\text{ième}}$ chiffre décimal de ce nombre, ce qui n'est pas.

Je nomme G le groupe de lettres entre guillemets.

Le nombre N est défini par les mots du groupe G, c'est-à-dire par un nombre fini de mots ; il devrait donc appartenir à l'ensemble E. Or, on a vu qu'il n'y appartient pas.

Telle est la contradiction.

Montrons que cette contradiction n'est qu'apparente. Revenons à nos arrangements. Le groupe de lettres G est un de ces arrangements ; il existera dans mon tableau. Mais à la place qu'il occupe, il n'a pas de sens. Il y est question de l'ensemble E, et celui-ci n'est pas encore défini. Je devrai donc le biffer. Le groupe G n'a de sens que si l'ensemble E est totalement défini, et celui-ci ne l'est que par un nombre infini de mots. *Il n'y a donc pas de contradiction.*

On peut encore remarquer ceci : l'ensemble de l'ensemble E et du nombre N forme un autre ensemble. Le second ensemble est dénombrable. Le nombre N peut être intercalé à un certain rang k dans l'ensemble E, en reculant d'un rang tous les autres nombres de rang supérieur à k. Continuons à appeler E l'ensemble ainsi modifié. Alors le groupe de mots G définira un nombre N′ *différent de* N, puisque le nombre N occupe maintenant le rang k, et que le $k^{\text{ième}}$ chiffre de N′ n'est pas égal au $k^{\text{ième}}$ chiffre du $k^{\text{ième}}$ nombre de l'ensemble E. (Richard 1905)

— Décrivez le paradoxe de Richard.

— Que pensez-vous de l'explication de Jules Richard qu'il n'y a en fait pas de contradiction.

— Quel rapport voyez-vous entre ce paradoxe et le paradoxe de Russell ?

— Faites le rapport avec la règle 3 de Poincaré page 64.

Exercices pour le chapitre 7

7.5

Le premier exercice, dont nous donnons une solution détaillée, s'occupe de la question de l'écriture décimale d'un nombre réel. Pour mieux saisir l'essence du problème nous prenons la base 2 au lieu de la base 10.

Calculer l'écriture binaire d'un nombre réel x revient à déterminer une entier $a_0 \in \mathbb{Z}$ et la suite des chiffres après la virgule, $(a_n)_{n>0}$, tous égaux à 0 ou 1, de sorte que $x = a_0 + \sum_{n=1}^{\infty} \frac{a_n}{2^n}$.

Pour obtenir les a_j à partir de x, on remarque que a_0 doit être la partie entière de x, $2a_0 + a_1$ la partie entière de $2x$, $4a_0 + 2a_1 + a_2$ la partie entière de $4x$, et ainsi de suite.

Ce calcul se heurte cependant à l'obstacle de savoir déterminer de manière sûre la partie entière d'un nombre réel (cf., la note rectificative de Turing page 102). On propose donc dans l'exercice un autre calcul, qui surmonte l'obstacle précédent, mais le prix à payer est d'utiliser trois chiffres $-1, 0, 1$, au lieu de deux. En conséquence la représentation de x comme développement binaire illimité n'est plus unique.

Exercice 7.5.1 (avec solution détaillée)**.** Montrer que si x est un nombre réel connu *via* des approximations rationnelles arbitrairement précises (adage page 101) on peut l'écrire sous la forme :

$$x = \lim x_n \text{ avec } x_n = k_n/2^n \ (n \in \mathbb{N}, k_n \in \mathbb{Z})$$

et pour tout $\in \mathbb{N}$, $k_{n+1} = 2k_n + u_{n+1}, u_{n+1} \in \{-1, 0, 1\}$, c'est-à-dire encore

$$x = k_0 + \sum_{n=1}^{\infty} \frac{u_n}{2^n} \qquad (k_0 \in \mathbb{Z}, u_n \in \{-1, 0, 1\})$$

Exercice 7.5.2. De même montrer que x peut s'écrire sous la forme

$$x = k + \sum_{n=1}^{\infty} \frac{v_n}{4^n} \quad \text{avec } k \in \mathbb{Z}, v_n \in \{-2, -1, 0, 1, 2\}$$

Exercice 7.5.3. Montrer que si $f\colon [a, b] \to [c, d]$ et $g\colon [c, d] \to \mathbb{R}$ sont des fonctions continues connues selon l'adage page 104, il en va de même pour $g \circ f$.

Exercice 7.5.4. Soit une fonction continue $f\colon [a, b] \to \mathbb{R}$ pour laquelle

— on connait une suite $\mu\colon \mathbb{N} \to \mathbb{N}$ qui contrôle la continuité uniforme de f au sens suivant :

$$\forall x, x' \in [a, b] \ \forall n \in \mathbb{N} \quad \left(|x - x'| \leq 1/2^{\mu(n)} \Rightarrow |f(x) - f(x')| \leq 1/2^n\right) ;$$

— pour tout $x \in \mathbb{D}_2 \cap [a, b]$ on sait évaluer $f(x)$ avec une précision arbitraire.

Montrer qu'alors la fonction f est connue selon l'adage page 104. Précisément, un entier n étant fixé, déterminer une suite finie croissante $x_1, \ldots, x_k$ dans $\mathbb{D}_2 \cap [a, b]$ et des éléments $y_1, \ldots, y_k$ dans $\mathbb{D}_2$ tels que la fonction g dont le graphe est dessiné figure 7.5.1 vérifie $\|f - g\|_\infty \leq 1/2^n$.

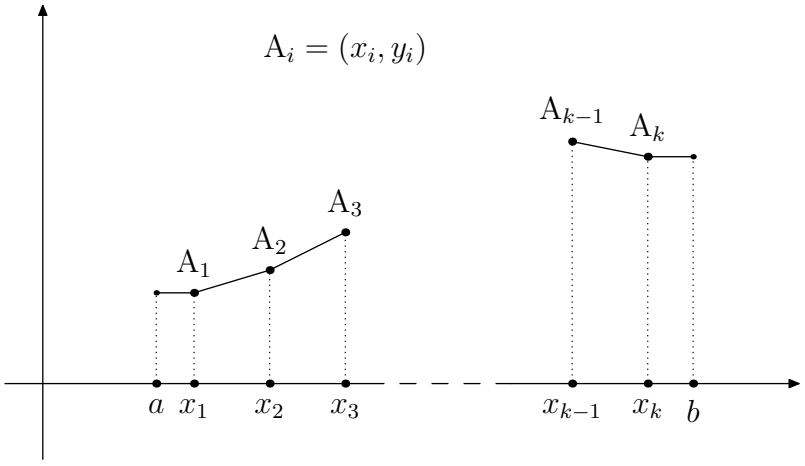


FIGURE 7.5.1 – Fonction affine par morceaux

Exercices pour le chapitre 8

8.4

Exercice 8.4.1. Commentez le texte suivant, extrait du livre *Leçons sur la théorie des fonctions* d'Émile Borel (1898). Ne vous laissez pas égarer par la notation bizarre pour l'intervalle $[a_k, b_k]$.

> Nous allons démontrer [...] le théorème suivant, dans l'énoncé duquel il est *expressément* entendu que les mots *intérieur à un intervalle* excluent les extrémités.
>
> *Si l'on a sur un segment limité de droite une infinité dénombrable d'intervalles partiels, tels que tout point de la droite soit intérieur à l'un au moins des intervalles, il existe* [1] *un nombre limité d'intervalles choisis parmi les intervalles donnés et ayant la même propriété (tout point de la droite est intérieur à, au moins, l'un d'eux).* Numérotons nos intervalles d'après une loi quelconque, mais déterminée ; je dis qu'il existe un nombre N, tel que tout point de la droite soit à l'intérieur d'un intervalle dont le rang ne dépasse pas N. En effet, nier l'existence du nombre N, c'est affirmer que, quel que soit le nombre donné n, il existe sur la droite un point tel que tous les intervalles qui le renferment ont un numéro supérieur à n. Il est clair d'ailleurs que, si l'on divise le segment de droite en deux segments égaux, l'un au moins de ces segments aura la même propriété ; car, si pour chacun de ces segments il existait un nombre N, soient N' et N'' ces deux nombres, il suffirait de prendre pour N le plus grand des deux. Si nous continuons à diviser le segment en deux parties égales et si nous conservons toujours le segment pour lequel il n'existe pas de nombre N (ou l'un d'eux, s'il y en a plusieurs), nous obtiendrons des segments de plus en plus petits, renfermés les uns dans les autres et ayant la propriété suivante : *quel que soit le nombre n, chacun d'eux contient au moins un point qui n'est renfermé dans aucun intervalle de rang inférieur à n.* Mais ces segments emboités les uns dans les autres et dont chacun est égal à la moitié du précédent ont un point limite α ; ce point α est, par hypothèse, *à l'intérieur* d'un intervalle de rang déterminé k, puisque nous avons supposé dénombrable l'ensemble de nos intervalles ; les extrémités a_k, b_k de cet intervalle ne coïncident d'ailleurs pas avec α (à cause du sens restreint que nous attachons au mot *intérieur*) ; donc, cet intervalle a_k-b_k comprend tout entier l'un des segments qui ont pour limite α, ce qui est absurde, puisque les points de ce segment seraient ainsi tous compris à l'intérieur de cet intervalle a_k-b_k dont le rang est un nombre fixe. L'existence du nombre N est donc établie.
>
> [1] On trouvera dans ma Thèse une autre démonstration de ce théorème, démonstration qui donne un moyen au moins théorique de déterminer effectivement les intervalles en nombre limité dont il est question.

Exercices pour le chapitre 9

9.1

Exercice 9.1.1. Donner une preuve intuitive du fait que tout ensemble infini contient une partie dénombrable (équipotente à $\mathbb{N}$).

Exercice 9.1.2. Faites un dessin pour visualiser la seconde preuve du théorème 9.1.8, et voyez apparaitre la « diagonale » d'un carré infini.

Exercices pour le chapitre 10

10.2

Exercice 10.2.1. Voici comment Guillaume d'Ockham distingue dans sa *Somme logique* de 1323 définition de chose (*quid rei*) et définition de nom (*quid nominis*). Les trois premiers paragraphes de cette citation proposent une classification très précise de la définition de chose, et le dernier est consacré à la définition de nom.

> La définition se comprend de deux façons. Elle peut exprimer ce qu'il en est de la chose ou ce qu'il en est du nom. La définition exprimant ce qu'il en est de la chose se comprend elle-même de deux façons. Au sens large, elle inclut la définition prise au sens strict et la définition descriptive. Au sens strict, c'est une formule abrégée exprimant toute la nature de la chose définie et ne manifestant rien qui lui soit extrinsèque.
>
> Cela peut se faire de deux façons. Parfois, dans un tel énoncé se trouvent des cas obliques exprimant des parties essentielles de la chose, comme lorsque je définis l'homme en disant : "l'homme est une substance composée d'un corps et d'une âme intellective" ; en effet, les noms qui sont à un cas oblique, "corps" et "âme intellective", expriment des parties de la chose. Cette définition peut être appelée définition naturelle.
>
> Autre est la définition dans laquelle ne se trouve aucun cas oblique, mais où le genre se trouve au nominatif, de même que la différence, et où les différences expriment les parties de la chose définie, à la manière dont "blanc" exprime la blancheur. Et pour cette raison, de même que "blanc", bien qu'il exprime la blancheur, ne suppose pas pour la blancheur mais seulement pour le sujet de cette blancheur, de même ces différences, bien qu'elles expriment les parties de la chose, ne supposent pas pour ces parties mais pour le tout qui en est composé. Telle est cette définition-ci de l'homme : "animal rationnel", ou celle-là : "substance animée, sensible, rationnelle". Car les différences "animée", "sensible" et "rationnelle" supposent pour l'homme, puisque l'homme est rationnel, animé et sensible ; pourtant ces termes renvoient à une partie de l'homme, tout comme les abstraits qui leur correspondent renvoient à une partie ou à des parties de l'homme, mais sur un mode différent. Cette définition peut être appelée définition métaphysique, puisque c'est ainsi que le métaphysicien définirait l'homme.
>
> . . . . . .
>
> La définition nominale, quant à elle, est une phrase manifestant explicitement ce à quoi renvoie un mot, comme lorsque quelqu'un, voulant apprendre à un autre ce que signifie le nom "blanc", dit que cela signifie la même chose que l'expression "quelque chose possédant la blancheur". [. . .] (Guillaume d'Ockham 1988, pages 88-89 et 92.)

1. Recherchez ce que Guillaume d'Ockham entend par « cas oblique ». Comment comprenez-vous la classification de la définition de chose ci-dessus ? En particulier, quelle différence voyez-vous entre définition « naturelle » et définition « métaphysique » ?
2. Pourquoi seule la blancheur est susceptible d'une définition de chose, mais pas « blanc » ?

Exercice 10.2.2. Arnauld et Nicole donnent ci-dessous un contre-exemple de définition de chose.

> Et de là il s'ensuit, 1. Que les définitions de noms sont arbitraires, et que celles des choses ne le sont point. Car chaque son étant indifférent de soi-même et par sa nature à signifier toutes sortes d'idées, il m'est permis pour mon usage particulier, et pourvu que j'en avertisse les autres, de déterminer un son à signifier précisément une certaine chose, sans mélange d'aucune autre. Mais il en est tout autrement de la définition des choses. Car il ne dépend point de la volonté des hommes que les idées comprennent ce qu'ils voudraient qu'elles comprissent ; de sorte que si en les voulant définir nous attribuons à ces idées quelque chose qu'elles ne contiennent pas, nous tombons nécessairement dans l'erreur.
>
> Ainsi pour donner un exemple de l'un et de l'autre, si dépouillant le mot *parallélogramme* de toute signification je l'applique à signifier un triangle, cela m'est permis, et je ne commets en cela aucune erreur, pourvu que je ne le prenne qu'en cette sorte ; et je pourrai dire alors qu'un parallélogramme a trois angles égaux à deux droits ; mais si laissant à ce mot sa signification et son idée ordinaire, qui est de signifier une figure dont les côtés sont parallèles, je venais à dire que le parallélogramme est une figure à trois lignes, parce que ce serait alors une définition de chose, elle serait très fausse, étant impossible qu'une figure à trois lignes ait ses côtés parallèles. (Arnauld et Nicole 2011, I, XI, pages 234-235.)

Proposez un commentaire de ce contre-exemple. Cela vous aide-t-il à comprendre la différence entre définition de nom et définition de chose ? pourquoi ?

10.3

Exercice 10.3.1. Maurice Caveing (1990, note 320, page 131) fait la remarque suivante.

> Il faut noter une autre sorte de termes non-définis appartenant à un lexique technique composé de verbes, comme « mener », « couper », « diviser », « tomber sur », « être élevé sur », « toucher », « se rencontrer », « prolonger », « s'ajuster », « se briser », etc., qui décrivent soit des opérations du géomètre, soit des relations de position entre lignes ou figures. Cet arsenal descriptif est emprunté à la langue commune, avec toutefois des exceptions notables, pour lesquelles une définition est donnée [...] (Caveing 1990, note 320, page 131.)

Comment selon vous Euclide fait-il pour se passer de définition de ces termes ? Essayez de les définir !

Exercice 10.3.2. Le but de cet exercice est d'approfondir la notion de définition implicite.

1. Dans la section 10.3.2, nous avons vu une "définition implicite" de *égal*. Voici deux autres exemples de définitions implicites dans les *Éléments*.

— Dans le cinquième livre, Euclide propose ainsi une définition implicite du *rapport*.

> 4. Des grandeurs sont dites *avoir un rapport l'une relativement à l'autre* quand elles sont capables, étant multipliées, de se dépasser l'une l'autre.
> 5. Des grandeurs sont dites *être dans le même rapport*, une première relativement à une deuxième et une troisième relativement à une quatrième quand des équimultiples de la première et de la troisième ou simultanément dépassent, ou sont simultanément égaux ou simultanément inférieurs à des équimultiples de la deuxième et de la quatrième, selon n'importe quelle multiplication, chacun à chacun, [et] pris de manière correspondante. (Euclide d'Alexandrie 1990-2001, volume 2, pages 38 et 41.)

— Dans le onzième livre, il procède ainsi pour l'angle *dièdre*.

> 6. L'*inclinaison d'un plan relativement à un plan* est l'angle aigu contenu par les [droites] menées à angles droits avec la section commune, au même point, dans chacun des plans.
> 7. Un plan, relativement à un plan, est dit *être incliné de la même manière* qu'un autre, relativement à un autre, quand lesdits angles des inclinaisons sont égaux l'un à l'autre. (Euclide d'Alexandrie 1990-2001, volume 4, pages 77-78.)

— Comparez ces trois définitions.

1. On fait remonter l'expression « définition implicite » à Joseph Diez Gergonne (1818).

> On conçoit fort bien, en effet, que, si une phrase contient un seul mot dont la signification nous soit inconnue, l'énoncé de cette phrase pourra souvent suffire pour nous en révéler la valeur. Si, par exemple, on dit à quelqu'un qui connaît bien les mots *triangle* et *quadrilatère*, mais qui n'a jamais entendu prononcer le mot *diagonale*, que *chacune des deux diagonales d'un quadrilatère le divise en deux triangles*, il concevra sur-le-champ ce que c'est qu'une diagonale, et le concevra d'autant mieux que c'est ici la seule ligne qui puisse diviser le quadrilatère en triangles.
>
> Ces sortes de phrases, qui donnent ainsi l'intelligence de l'un des mots dont elles se composent, au moyen de la signification connue des autres, pourraient être appelées *définitions implicites*, par opposition aux définitions ordinaires qu'on appellerait *définitions explicites* ; et on voit qu'il y aurait entre les unes et les autres la même différence qui existe entre les équations résolues et les équations non résolues. On conçoit aussi que, de même que deux équations entre deux inconnues les déterminent l'une et l'autre, deux phrases qui contiennent deux mots nouveaux, combinés avec des mots connus, peuvent souvent en déterminer le sens ; et on peut en dire autant d'un plus grand nombre de mots nouveaux combinés avec des mots connus, dans un pareil nombre de phrases ; mais il y a ici à exécuter une sorte d'élimination qui peut devenir d'autant plus pénible que le nombre des mots dont il s'agit est lui-même plus considérable. (Gergonne 1818, page 23.)

— Commentez ce passage. Est-ce que ces explications s'appliquent aux définitions d'Euclide ci-dessus ?

Exercice 10.3.3. Jacques Peletier écrit ceci dans *Les six premiers livres des Éléments géométriques d'Euclide* (1557, pages 10-12 de la traduction française par Jean II de Tournes de 1611, voir Barbin 1994, Loget 2002).

> Mais il y a bien plus grande difficulté en la forme et constitution de l'angle, quel il est, et en quoi il consiste. Car ce que quelques-uns ont dit, que l'angle est partie de la superficie, n'est pas probable. Car ainsi on ferait de l'angle le Triangle, en tirant une troisième ligne : ce qui ne vient pas à propos. Les autres veulent qu'il soit ensemble et au point, et en la ligne, et en la superficie : mais cela aussi engendre une plus grande recherche : car quantième partie d'icelui sera au point ? quantième à la ligne ? quantième à la superficie ? Que si ceci s'éloigne de la vérité, quelle sera la situation de l'angle ? Car si nous disons qu'elle consiste seulement au point, ou tous les angles seront égaux, ou il y aura inégalité entre les points. Le premier répugne ouvertement à la vérité, le second ne semble pas s'accorder avec la raison. De même ne pourra-t-on dire, que l'angle soit seulement en la ligne, ou en la superficie.
>
> Mais on pourra ainsi soudre ce doute à mon avis. Il est bien vrai que l'angle consiste en un point, mais c'est l'inclination qui le fait plus grand ou plus petit. La ligne coupant la ligne fait bien l'angle : mais pourtant l'angle n'est pas partie de la ligne : comme aussi les lignes ne sont pas parties de la superficie, combien que la superficie ne puisse être sans lignes qui la terminent. Partant l'angle

ne sera pas portion de la superficie, pour ce qu'il la clôt. En quoi certes on peut voir, que le point de la section est pressé et comme rendu plus étroit, par la mesure de l'inclination. Le point sera-t-il donc quantité ? Nenni. Car ce qu'une fois l'intellect a reçu et arrêté être très petit, il ne le peut plus diviser : mais cela n'empêche pas qu'il ne le puisse presser et contraindre. Et afin qu'on ne pense pas que nous disions choses répugnantes, il faut penser qu'en la Géométrie le point ne se considère pas comme un rien, mais bien comme quelque chose. Et comme nous menuisons l'unité en l'Arithmétique, ainsi faisons-nous le point en la quantité continue : afin que ce dont tout est produit nous puisse aussi bailler la représentation et l'image de tout, savoir est, du droit, de l'oblique, du long, du large, et du profond. Puis donc que la Géométrie nous représente la nature, comme aussi elle en est le miroir, pensons, que comme en l'angle physique, deux lignes, quelques déliées qu'elles soient, ne se peuvent entrecouper, si ce n'est que l'une s'encline sur l'autre au point de la décussation, ainsi en la section droite des lignes mathématiques, le point est aucunement carré : en l'obtuse il est plus mousse : et en l'aigüe, plus pressé et plus étroit. Ces choses sont comprises par l'intellect, lequel ne s'arrête jamais sinon avec la nature : tellement qu'il ne cesse d'amoindrir le point, jusqu'à ce que la ligne tombant soit faite une avec la couchée. Quand donc l'intellect présuppose que le point ne se peut aucunement partir, il entend qu'il ne soit ni ligne, ni superficie, ni corps. Mais quand il est parvenu à l'angle, lequel a toute autre considération que les autres quantités, alors il veut partir ce qu'auparavant il avait jugé indivisible : à savoir, (comme nous avons dit un peu auparavant) afin que ce dont est produite la quantité, se ressente aussi de la nature de la quantité.

Commentez ce passage.

10.4

Exercice 10.4.1. Voici comment Paul Imbs (1971) décrit la différence entre définition de mot et définition de chose dans la préface du *Trésor de la langue française*.

C'est pourquoi la définition lexicographique peut sans difficulté assumer la conception aristotélicienne de la définition, conçue comme un énoncé indiquant d'abord le genre prochain (*banc* a pour genre prochain « siège avec ou sans dossier ou bras, assez large pour servir à plusieurs personnes », genre prochain qu'il partage par exemple avec *canapé*), puis la différence spécifique (par rapport à *canapé*, *banc* a pour différence spécifique d'être moins profond et plus dur et de ne pas figurer dans le mobilier de salon). Par le genre prochain la définition insiste sur la substance sémique en orientant le mot vers l'objet trans- et extralinguistique (le *référé*, ou, selon la terminologie habituelle, le *référent*) qu'elle *montre* (c'est la fonction désignative ou déictique de la définition) ; par la différence spécifique elle délimite le mot par rapport à ses voisins et lui sert en quelque sorte de guide et de garde-fou dans son cheminement vers le référé. Toutes les définitions lexicographiques n'ont cependant pas cette forme classique : elles peuvent remplacer l'indication du genre prochain par celle du genre éloigné (« action de… ») ou par celle de l'ensemble (cf. supra [les classificateurs collectifs]) qui groupe en un singulier une pluralité d'entités identiques ; ou encore par celle d'une similitude (« espèce de…, sorte de… ») ou d'une privation (« absence de…, cessation de… »), c'est-à-dire par des déterminations sémantiques du premier mot de la définition, souvent réduite dans ce cas à un synonyme.

Mais à la différence de la définition aristotélicienne qui cherchait à dire le vrai, c'est-à-dire ce qui existe réellement en dehors et indépendamment de nos représentations de la réalité, la définition lexicographique ne vise qu'à appréhender celles-ci sans avoir à se préoccuper de leur vérité. Elle n'a donc pas à connaitre la différence entre définitions de choses (c'est-à-dire de réalités non créées par notre esprit) et définitions de mots (comme par exemple celles qu'appellent les concepts mathématiques, œuvres conventionnelles de notre esprit sans relations substantielles avec le réel). En revanche, du fait qu'elles ne visent pas à saisir la réalité, mais des vues sur la réalité ou des créations

> imaginaires, les définitions linguistiques sont toujours sujettes à variation, liées qu'elles sont à la situation contingente des sujets parlants qui ont élaboré ces vues ou ces créations imaginaires : celles-ci sont fragiles comme tout ce qui est historique et culturel, et comme la langue elle-même, création et recréation continue d'un donné malléable parce que maniable dans et pour des situations seulement historiques et naturellement non éternelles. Elle n'a pas davantage à s'inquiéter outre mesure de la distinction entre ce qui est linguistique et ce qui est encyclopédique, les informations encyclopédiques allant par nature au-delà des *traits déictiques et différenciateurs* que relève l'analyse componentielle et qui, comme on l'a rappelé plus haut, constituent l'essence même de la définition en tant qu'elle a un contenu : celui-ci n'est que le contenu *utile* pour le fonctionnement correct du langage, et non pas le contenu nécessaire pour la connaissance *exhaustive* du référé. Elle reflète le statut même de la langue, qui dans sa structure interne n'est ni *physis [nature]* ni *thesis [théorie]*, mais *poiesis [création]*. (Imbs 1971, page XXXVIII.)

1. Imitez la démarche de Paul Imbs pour affiner la définition de la droite, de l'angle rectiligne, de l'angle droit et du nombre pair.
2. Quelles similitudes et différences voyez-vous entre définition de mot et définition de chose ?

10.5

Exercice 10.5.1. Antoine Arnauld définit lui aussi dans ses *Nouveaux éléments de géométrie* (1667) que « toute circonférence se conçoit divisée en 360 parties égales qui s'appellent *degrés* » (livre cinquième, XX) et que les angles rectilignes se mesurent par « la partie proportionnelle d'une circonférence dont le centre est au point où ces lignes se joignent » (livre huitième, II), mais les droites perpendiculaires sont définies séparément (voir Arnauld 2009, pages 363, 454 et 366-367). Quels sont selon vous les avantages et inconvénients de cette démarche ?

10.6

Exercice 10.6.1. Dans la section 10.6.3, j'ai proposé une définition de l'inclinaison respective de deux droites qui ne se coupent pas nécessairement. Est-elle cohérente avec la définition de l'angle de deux droites lorsqu'elles se coupent ? Rappelez la démonstration du théorème sur la somme des angles d'un triangle puis montrez qu'en géométrie euclidienne l'inclinaison respective de deux droites ne dépend pas de la droite tombant sur elles.

Exercice 10.6.2. Voici une critique de D'Alembert de toutes les tentatives passées de définir la ligne droite dans un de ses *Éclaircissements* de 1767 consacré aux définitions mathématiques, à commencer par la définition des parallèles.

> On parviendrait peut-être plus facilement à la trouver, [la définition des parallèles,] si on avait une bonne définition de la ligne droite ; par malheur cette définition nous manque. Il ne parait pas possible d'en donner une autre que celle dont presque tous les Mathématiciens font usage ; mais cette définition, comme nous l'avons dit ailleurs, exprime plutôt une propriété de la ligne droite, que sa notion primitive. Ce n'est pas que je veuille, avec quelques Géomètres, chercher cette notion dans l'idée que la vision nous donne de la ligne droite, en nous apprenant que les points de cette ligne se couvrent les uns les autres lorsque l'œil se trouve placé dans son prolongement. Cette notion de la ligne droite serait très-peu géométrique, 1°. parce qu'il y a des lignes droites pour un aveugle, et que l'illustre Saunderson entre autres en avait une idée très-distincte sans en avoir jamais vu ; 2°. parce qu'il serait impossible de savoir que la lumière se répand en ligne droite, si, pour connaitre la

rectitude d'une ligne, nous n'avions d'autre moyen que d'examiner si les points de cette ligne se cachent les uns les autres quand l'œil est placé dans son prolongement. Si la lumière se propageait en suivant une ligne circulaire d'une courbure déterminée, et que l'œil fût placé sur la circonférence d'un tel cercle, tous les points de ce cercle se cacheraient les uns les autres, et cependant la ligne sur laquelle ils seraient placés ne serait pas droite.

On ne définirait pas mieux la ligne droite, en disant avec d'autres Auteurs que c'est une ligne dont tous les points sont dans la même direction. Car qu'est-ce que *direction ?* Et comment en peut-on avoir l'idée, si on n'a déjà celle de ligne droite ?

On est donc comme forcé d'en revenir à la définition ordinaire, que la ligne droite est celle qui est la plus courte d'un point à un autre. Mais il est aisé de sentir que cette définition n'est pas telle qu'on pourrait le désirer. En premier lieu, d'où sait-on que d'un point à un autre, il n'y a qu'un seul chemin qui soit le plus court ? Pourquoi ne pourrait-il pas y en avoir plusieurs, tous différents, tous égaux, et tous les plus courts ? On n'est persuadé de la vérité contraire, et on ne la suppose dans la définition de la ligne droite, que parce qu'on a déjà dans l'esprit ou plutôt dans les sens, si je puis parler de la sorte, une notion de la ligne droite qui renferme implicitement cette vérité. C'est cette notion qu'il faudrait exprimer ; mais les termes, et peut-être les idées, nous manquent pour cela. *Hoc opus, hic labor est [voilà l'obstacle, voilà l'épreuve].*

En second lieu, supposons qu'en effet la ligne droite soit le plus court chemin d'un point à un autre, que ce plus court chemin soit unique, et qu'il n'y en ait pas deux égaux ; je vois clairement comment on peut conclure de là, que si on veut mener une ligne droite d'un point à un autre, tous les points par lesquels doit passer cette ligne, sont nécessairement donnés, et que la ligne qui joint deux quelconques de ces points, est aussi la plus courte qu'on puisse mener ou imaginer de l'un à l'autre. Mais je ne vois pas avec la même évidence, en partant de la définition supposée, qu'une ligne droite tirée par deux points ne puisse être prolongée que d'une seule manière, ou ce qui revient au même, que deux lignes droites, tirées d'un même point à deux autres points, ne puissent pas avoir une partie commune : je ne dis pas que cela ne soit évident, je dis (et je me flatte qu'on en conviendra après y avoir fait attention) que cela ne suit pas évidemment de la définition supposée, mais d'une notion primitive de la ligne droite que nous avons dans l'esprit, sans pouvoir en quelque façon la rendre par des expressions ; idée dont la définition supposée n'est que la suite.

La définition et les propriétés de la ligne droite, ainsi que des lignes parallèles, sont donc l'écueil, et, pour ainsi dire, le scandale des éléments de Géométrie. ([D'Alembert] 1767, §. XI. *Sur les éléments de Géométrie*, pages 203-207.)

Commentez ce passage à la lumière de la section 10.6 du cours.

Exercices pour le chapitre 11

11.1

Exercice 11.1.1. Voici quatre extraits des *Seconds analytiques* d'Aristote, dans lesquels ce philosophe de l'Antiquité étudie comment les sciences procèdent pour élaborer des connaissances.

> Il y a deux manières selon lesquelles il est nécessaire d'avoir une pré-connaissance. Pour certaines choses, il est nécessaire de saisir à l'avance qu'elles sont, pour d'autres il faut comprendre ce qu'est la chose dont on parle, pour d'autres il faut les deux ; par exemple, que de toute chose il soit vrai de l'affirmer ou de la nier, il faut savoir à l'avance que c'est le cas ; pour « triangle », qu'il signifie ceci ; pour l'unité c'est les deux, ce qu'elle signifie et qu'elle est. Car ce n'est pas de la même façon que chacune de ces choses est claire pour nous.
>
> . . . . . .
>
> De la thèse, une espèce est celle qui admet n'importe laquelle des sortes de l'énonciation, je veux dire par exemple que quelque chose est ou que quelque chose n'est pas, et c'est une hypothèse ; une autre espèce, sans cela, est une définition. En effet, la définition est une thèse, car l'arithméticien pose que l'unité c'est l'indivisible du point de vue de la quantité ; or ce n'est pas une hypothèse. En effet, ce qu'est une unité et qu'elle est, ce n'est pas la même chose.
>
> . . . . . .
>
> J'appelle « principes » dans chaque genre ceux dont on ne peut pas prouver qu'ils sont le cas. Ce que, donc, signifient aussi bien les notions premières que celles qui en viennent, on l'admet, mais qu'ils sont, pour les principes il est nécessaire de l'admettre, et pour les autres de les prouver. Par exemple, ce que signifie l'unité, ou ce que signifient le droit ou le triangle il est nécessaire de l'admettre, d'admettre que l'unité et la grandeur sont, et pour les autres choses de les prouver. (Aristote 2005, 71a11-17, 72a18-24, 76a31-36, pages 61, 71, 113.)

1. Dans ces extraits, Aristote utilise plusieurs fois le verbe être sans complément, par exemple quand il écrit : "il est nécessaire de saisir à l'avance qu'elles sont". Comment interprétez-vous cet usage du verbe ? Comment le diriez-vous aujourd'hui ? Repérez toutes les occurrences de cet usage.
2. Explicitez et comparez les trois exemples du premier extrait.
3. Les sortes de l'énonciation sont l'affirmation et la négation et le deuxième extrait indique qu'une définition ne peut consister en une négation. Comment comprenez-vous la dernière phrase du deuxième extrait ?
4. Dans le premier extrait, Aristote constate qu'une science ne peut pas partir de rien. De quelle manière l'unité, le 1, est-il clair pour vous ?
5. Dans le deuxième et le troisième extrait, Aristote distingue le rôle des définitions et des axiomes dans l'élaboration des connaissances. Comparez-les avec la citation de Pascal dans la section 10.1.3.

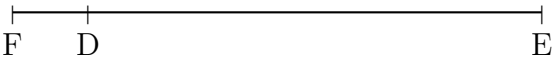
11.2

Exercice 11.2.1. Faire le lien entre la preuve originale d'Euclide et sa réécriture au cours de l'analyse logique.

Exercice 11.2.2 (la proposition 20 du neuvième livre des *Éléments* d'Euclide)**.** Voici à présent la preuve euclidienne de l'infinité des nombres premiers, donnée dans le théorème 20 du neuvième livre des *Éléments*.

> 20. *Les nombres premiers sont plus nombreux que toute multitude de nombres premiers proposée.*
> Soient les nombres premiers proposés a, b, c. Je dis que les nombres premiers sont plus nombreux que a, b, c.
> En effet, que soit pris le plus petit [nombre] mesuré par a, b, c, et que ce soit DE et que l'unité DF soit ajoutée à DE. Alors ou bien EF est premier ou bien non. D'abord qu'il soit premier ; donc sont trouvés les nombres premiers a, b, c, EF, plus nombreux que a, b, c.
> Mais alors que EF ne soit pas premier ; il est donc mesuré par un certain nombre premier (VII. 32). Qu'il soit mesuré par le [nombre] premier g. Je dis que g n'est pas le même que l'un quelconque des a, b, c. En effet, si c'est possible, qu'il le soit. Or a, b, c mesurent DE ; donc g mesurera aussi DE. Mais il mesure aussi EF ; il mesurera aussi l'unité DF restante tout en étant un nombre ; ce qui est absurde. g n'est donc pas le même que l'un des a, b, c. Et il est supposé premier. Donc sont trouvés les nombres premiers a, b, c, g, plus nombreux que la multitude proposée des a, b, c. Ce qu'il fallait démontrer.

Euclide ne connait pas le signe + et utilise l'intuition géométrique pour additionner des nombres : il représente un nombre par un segment DE, prolonge ce segment au-delà de D d'un segment unité DF jusqu'en F, et alors le segment EF représente la somme du nombre et de l'unité.



La preuve procède par construction d'un nombre premier supplémentaire à partir de tout ensemble fini de nombres premiers : il considère ce que nous appellerions aujourd'hui le produit de tous ces nombres (puisqu'ils sont tous premiers, ce produit est égal à leur ppcm), lui ajoute 1, applique la proposition 31 du septième livre [1] pour lui trouver un facteur premier et constate que ce facteur ne peut pas être dans l'ensemble dont on était parti.

La formalisation de cette preuve pose un problème qui est tout à fait surmontable : elle raisonne sur les ensembles finis d'entiers et non sur les entiers. D'un côté, il est possible de réinterpréter toute assertion qui traite d'ensembles finis d'entiers en une assertion qui traite d'entiers ; d'un autre côté, on peut considérer, et c'est que nous allons faire, que les ensembles finis non vides d'entiers $\{b_1, \ldots, b_k\}$ sont des objets aussi élémentaires de l'arithmétique que les entiers a et que l'expression $a \in \{b_1, \ldots, b_k\}$ a la signification intuitive que a est un des nombres $b_1, \ldots, b_k$. Nous allons aussi supposée construite la fonction ppcm qui à un ensemble fini non vide $\{b_1, \ldots, b_k\}$ associe le plus petit multiple commun des nombres dans $\{b_1, \ldots, b_k\}$; dans cette preuve, nous avons seulement besoin de savoir que si $a \in \{b_1, \ldots, b_k\}$, alors $a | \text{ppcm}\{b_1, \ldots, b_k\}$.

Faire l'analyse logique de cette preuve.

1. Selon l'édition de Heiberg, la démonstration invoque la proposition 32, mais c'est discutable. Voici cette proposition et sa preuve.

« 32. *Tout nombre est soit premier soit mesuré par un certain nombre premier.*

« Soit un nombre a. Je dis que a est soit premier, soit mesuré par un certain nombre premier.

« Or d'une part si a est premier, ce qui était prescrit aura été fait. D'autre part s'il est composé, un certain nombre premier le mesurera (VII. 31). Donc tout nombre est soit premier soit mesuré par un certain nombre premier. Ce qu'il fallait démontrer. »

11.3

Exercice 11.3.1. Donner des exemples de preuves fausses parce qu'on n'a pas pris garde aux variables dans la ∀-introduction ou la ∃-élimination.

Exercice 11.3.2 (le deuxième exemple de déduction naturelle traité par Gentzen)**.** Validez la formule $(\exists x\, \forall y\, \mathrm{A}(x,y)) \to (\forall y\, \exists x\, \mathrm{A}(x,y))$ en proposant un raisonnement qui utilise uniquement les règles de déduction de la section 11.3.

Exercice 11.3.3. Dans son article de 1936, Gentzen (1936) propose une preuve directe du théorème de l'infinité des nombres premiers.

[...] je choisis comme « énoncé d'induction » la proposition suivante, en l'exprimant pour un nombre m : « Ou bien il y a parmi les nombres de 1 à m un nombre premier plus grand que a, ou bien tous ces nombres, sauf 1, ne divisent pas $a! + 1$ ». Formellement :

$$\{\exists z\,[z \leq m \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq m) \to \neg\, y|(a! + 1)].$$

La démonstration se déroule alors ainsi :

4.4 1. D'abord il faut démontrer l'énoncé d'induction pour $m = 1$. En ce cas son second membre est satisfait automatiquement puisqu'il n'y a pas du tout de nombre qui soit plus grand que 1 et plus petit ou égal à 1. En détail : pour un c quelconque, on a $\neg\,(c > 1 \wedge c \leq 1)$; c'est ce que nous supposons déjà connu. Cela étant, on a aussi $(c > 1 \wedge c \leq 1) \to \neg\, c|(a! + 1)$, et aussi, puisque c était quelconque, $\forall y\,[(y > 1 \wedge y \leq 1) \to \neg\, y|(a! + 1)]$. Il s'ensuit en outre, en vertu de la signification de $\vee$ (3.1 2), que l'énoncé d'induction pris en son entier est vrai pour $m = 1$, *i. e.* :

$$\{\exists z\,[z \leq 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq 1) \to \neg\, y|(a! + 1)].$$

4.4 2. On passe à « l'étape inductive », *i. e.* nous admettons que l'énoncé d'induction a déjà été démontré pour un nombre quelconque n, et qu'on a donc :

$$\{\exists z\,[z \leq n \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n) \to \neg\, y|(a! + 1)]\,;$$

il faut alors montrer qu'il est vrai pour $n + 1$. Cela se fait ainsi : sur la base de l'hypothèse d'induction, deux cas sont possibles :

1. $\exists z\,[z \leq n \wedge (\mathrm{Prem}(z) \wedge z > a)]$,
2. $\forall y\,[(y > 1 \wedge y \leq n) \to \neg\, y|(a! + 1)]$.

Dans le premier cas on obtient directement $\exists z\,[z \leq n + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]$, que je donne sans développer. Du même coup dans ce cas l'énoncé d'induction est démontré pour $n + 1$, autrement dit

$$\{\exists z\,[z \leq n + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n + 1) \to \neg\, y|(a! + 1)]$$

est démontré. Traitons donc le second cas :

$$\forall y\,[(y > 1 \wedge y \leq n) \to \neg\, y|(a! + 1)].$$

On a $(n + 1)|(a! + 1) \vee \neg\,(n + 1)|(a! + 1)$. En conséquence de quoi nous pouvons distinguer deux sous-cas :

1. Sous-cas 1 : $(n + 1)|(a! + 1)$. Il en résulte $\mathrm{Prem}(n + 1) \wedge (n + 1) > a$, ce que je me borne à indiquer rapidement, car n'y interviennent que des formes d'inférence dont les autres parties de la démonstration nous ont déjà offert des exemples : $(n + 1)$ est un nombre premier ; en effet s'il avait un diviseur différent de 1 et de lui-même, celui-ci serait plus petit que $n + 1$ et diviserait aussi $a! + 1$, ce qui contredirait notre hypothèse : $\forall y\,[(y > 1 \wedge y \leq n) \to \neg\, y|(a! + 1)]$. De plus $n + 1$ est plus grand que a, car les nombres compris entre 2 et a ne divisent pas $a! + 1$, puisque cette division donne toujours

le reste 1. Donc on a bien $\mathrm{Prem}(n+1) \wedge (n+1) > a$; de plus on a $(n+1) \leq (n+1)$ et donc aussi $(n+1) \leq (n+1) \wedge (\mathrm{Prem}(n+1) \wedge (n+1) > a)$, donc encore :

$$\exists z\,[z \leq n+1 \wedge (\mathrm{Prem}(z) \wedge z > a)],$$

et par là :

$$\{\exists z\,[z \leq n+1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n+1) \rightarrow \neg\, y|(a!+1)].$$

2. Sous-cas 2 : $\neg\,(n+1)|(a!+1)$. Soit d un nombre quelconque tel que $d > 1 \wedge d \leq n+1$. De $d \leq n+1$, suit, on le supposera connu, $d \leq n \vee d = n+1$. Soit d'abord $d \leq n$; alors on a : $\forall y\,[(y > 1 \wedge y \leq n) \rightarrow \neg\, y|(a!+1)]$, donc en particulier $(d > 1 \wedge d \leq n) \rightarrow \neg\, d|(a!+1)$. De $d > 1$ avec $d \leq n$, on obtient $d > 1 \wedge d \leq n$, ce qui, avec ce qui précède, donne $\neg\, d|(a!+1)$.

Si au contraire $d = n+1$, compte tenu de $\neg\,(n+1)|(a!+1)$, il suit encore $\neg\, d|(a!+1)$.

Ainsi on a dans tous les cas $\neg\, d|(a!+1)$, comme conséquence de l'hypothèse $d > 1 \wedge d \leq n+1$. Partant nous pouvons écrire : $(d > 1 \wedge d \leq n+1) \rightarrow \neg\, d|(a!+1)$, puis, comme d était un nombre quelconque,

$$\forall y\,[(y > 1 \wedge y \leq n+1) \rightarrow \neg\, y|(a!+1)],$$

et on a de nouveau :

$$\{\exists z\,[z \leq n+1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n+1) \rightarrow \neg\, y|(a!+1)].$$

Ainsi avons-nous obtenu pour tous les cas l'énoncé d'induction pour $n+1$, ce qui termine l'étape inductive.

4.4 3. Ensuite la démonstration s'achève rapidement : l'induction complète établit la validité de l'énoncé d'induction pour des nombres arbitraires. Nous n'avons besoin de cet énoncé que pour le nombre $a!+1$:

$$\{\exists z\,[z \leq a!+1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq a!+1) \rightarrow \neg\, y|(a!+1)].$$

Le second cas donne en particulier :

$$(a!+1 > 1 \wedge a!+1 \leq a!+1) \rightarrow \neg\,(a!+1)|(a!+1).$$

Or $a!+1 > 1 \wedge a!+1 \leq a!+1$ est vrai, comme nous le supposons connu ; on conclut donc $\neg\,(a!+1)|(a!+1)$. Mais par ailleurs il va de soi que $(a!+1)|(a!+1)$ est vrai, et nous obtenons une contradiction, *i. e.* le second cas ne peut pas se présenter. Formellement

$$\neg\,\forall y\,[(y > 1 \wedge y \leq a!+1) \rightarrow \neg\, y|(a!+1)].$$

Ne reste que le premier cas, *i. e.* $\exists z\,[z \leq a!+1 \wedge (\mathrm{Prem}(z) \wedge z > a)]$. Soit l un tel nombre, c'est-à-dire un nombre pour lequel $l \leq a!+1 \wedge (\mathrm{Prem}(l) \wedge l > a)$ est vrai. En particulier on a alors $\mathrm{Prem}(l) \wedge l > a$, d'où $\exists z\,(\mathrm{Prem}(z) \wedge z > a)$. Or a étant un nombre naturel absolument quelconque, cela vaut pour tous les nombres naturels, *i. e.* : $\forall y\,\exists z\,(\mathrm{Prem}(z) \wedge z > y)$. Telle est la conclusion de la démonstration d'Euclide. (Gentzen 1936, p. 507-509, traduit dans Largeault 1992, p. 302-305)

Trouvez toutes les occurrences de règles d'introduction et d'élimination de $\forall$, $\wedge$, $\exists$, $\vee$ et $\rightarrow$ dans cette analyse logique.

11.4

Exercice 11.4.1. 1. Complétez la preuve de la distributivité de la disjonction par rapport à la conjonction en détaillant la réciproque $((\mathrm{X} \vee \mathrm{Y}) \wedge (\mathrm{X} \vee \mathrm{Z})) \rightarrow (\mathrm{X} \vee (\mathrm{Y} \wedge \mathrm{Z}))$ en règles d'introduction et d'élimination. Formulez le raisonnement avec les figures de déduction de la table 11.1.

2. Traitez de même la formule duale $((\mathrm{X} \wedge \mathrm{Y}) \vee (\mathrm{X} \wedge \mathrm{Z})) \rightarrow (\mathrm{X} \wedge (\mathrm{Y} \vee \mathrm{Z}))$.

Exercice 11.4.2 (suite de l'exercice 11.3.2)**.**

1. Formulez le raisonnement de l'exercice 11.3.2 avec les figures de déduction de la table 11.1.
2. Que peut-on dire de la réciproque $(\forall y\, \exists x\, \mathrm{A}(x, y)) \to (\exists x\, \forall y\, \mathrm{A}(x, y))$?

Exercice 11.4.3.

1. Validez la formule $\mathrm{X} \vee (\mathrm{Y} \wedge (\mathrm{Z} \wedge \mathrm{W})) \to (\mathrm{X} \vee \mathrm{Y}) \wedge ((\mathrm{X} \vee \mathrm{Z}) \wedge (\mathrm{X} \vee \mathrm{W}))$ en proposant un raisonnement qui utilise uniquement les règles d'introduction et d'élimination des opérations logiques. Formulez ce raisonnement avec les figures de déduction des tables 11.1 et 11.2.
2. Traitez de même la formule duale $\mathrm{X} \wedge (\mathrm{Y} \vee (\mathrm{Z} \vee \mathrm{W})) \to (\mathrm{X} \wedge \mathrm{Y}) \vee ((\mathrm{X} \wedge \mathrm{Z}) \vee (\mathrm{X} \wedge \mathrm{W}))$.

Exercice 11.4.4. Montrez que les formules $(\mathrm{A} \vee \mathrm{B}) \to \mathrm{C}$ et $(\mathrm{A} \to \mathrm{C}) \wedge (\mathrm{B} \to \mathrm{C})$ sont équivalentes en construisant un arbre de preuve qui va de la première formule à la deuxième et *vice versa*.

Exercice 11.4.5. Validez la formule $(\neg\, \exists x\, \mathrm{A}(x)) \to (\forall y\, \neg \mathrm{F}(y))$ en proposant un raisonnement qui utilise uniquement les règles d'introduction et d'élimination des opérations logiques. Formulez ce raisonnement avec les figures de déduction des tables 11.1 et 11.2.

Exercice 11.4.6 (suite de l'exercice 11.3.3)**.** Décrivez en détail les règles appliquées pour l'opération de négation dans le raisonnement analysé dans l'exercice 11.3.3.

1. Reconnaissez-vous parmi ces règles une ou plusieurs des règles décrites dans la section 11.4.2 ?
2. Essayez de dériver celles que vous n'avez pas reconnues à partir des règles décrites dans les sections 11.4.1 et 11.4.2, en évitant, si c'est possible, la règle de l'absurdité classique et du tiers exclu.

Exercices pour le chapitre 12

12.1

On utilisera le même type de conventions que dans la section 12.1 du cours.

Les exercices consistent à écrire des « programmes » pour une tâche précise qu'on confie à une Machine de Turing. Avant d'écrire le programme proprement dit, il faut expliquer en français « ce qui se passe » et pourquoi on pense que le programme réalise l'objectif fixé.

Exercice 12.1.1. Un mot est appelé un palindrome lorsqu'on peut le lire indifféremment de gauche à droite ou de droite à gauche. Par exemple `kayak` est un palindrome.

Décrivez le « programme » d'une machine de Turing qui reconnait les palindromes. Les mots sont écrits avec les trois lettres `a`, `k`, `y`. En voici d'autres exemples : `kaykaykay`, `kya`, `yak`, `yaka`. Un mot est écrit sur la bande d'entrée. La machine écrit sa réponse sur la bande de sortie sous la forme suivante : elle écrit `y` si le mot est un palindrome et `a` sinon.

Expliquez d'abord en français le fonctionnement du programme. Donnez ensuite le programme sous forme précise.

Exercice 12.1.2. Décrire un programme qui multiplie deux entiers naturels écrits en base 2.

Exercice 12.1.3. Décrire un programme qui compare des entiers naturels écrits en base 3 sur une unique bande en entrée. Les entiers sont séparés les uns des autres par une case vide. La machine doit écrire sur la bande de sortie le plus grand entier qu'elle a trouvé dans la liste donnée au départ.

12.3

Exercice 12.3.1. Vérifier pour $\mathrm{mul}(3,5)$ et $\exp(2,3)$ que les valeurs prises peuvent être obtenues à partir du système (12.2) en utilisant les deux règles élémentaires de la transitivité et symétrie de l'égalité et de la substitution d'une expression par sa valeur.

Exercices pour le chapitre 13

13.4

Exercice 13.4.1. Discuter le passage suivant tiré de l'introduction du mémoire *Continuité et nombres irrationnels* de Richard Dedekind.

> Les considérations qui font l'objet de cet opuscule remontent à l'automne 1858. Alors professeur à l'École Polytechnique Fédérale de Zürich, je me trouvai pour la première fois dans l'obligation d'enseigner les éléments du Calcul différentiel, et à cette occasion, je ressentis plus vivement encore qu'auparavant combien l'Arithmétique manque d'un fondement réellement scientifique. À propos du concept de grandeur variable approchant une valeur limite fixe, et notamment pour démontrer le théorème selon lequel toute grandeur constamment croissante, mais pas au-delà de toute limite, approche certainement et nécessairement une valeur limite, je cherchai refuge dans les évidences géométriques. Aujourd'hui encore, un tel appel à l'intuition géométrique dans les premières leçons de Calcul différentiel me semble extrêmement utile du point de vue didactique, et même indispensable si l'on ne veut pas perdre trop de temps. Mais personne ne le niera, cette manière d'introduire au Calcul différentiel ne peut prétendre à la scientificité. Mon sentiment d'insatisfaction me dominait si fort que je pris la ferme résolution de réfléchir jusqu'à ce que j'aie trouvé un fondement purement arithmétique et parfaitement rigoureux aux principes du Calcul infinitésimal. L'on dit si souvent que le Calcul différentiel traite des grandeurs continues, et pourtant nulle part n'est donnée une définition de cette continuité ; et même les présentations [*Darstellungen*] les plus rigoureuses du Calcul différentiel ne fondent pas leurs démonstrations sur la continuité. Soit elles font plus ou moins consciemment appel à la géométrie, ou à des représentations [*Vorstellungen*] inspirées de la géométrie, soit elles s'appuient sur des théorèmes qui ne sont eux-mêmes jamais démontrés de manière purement arithmétique. C'est le cas, par exemple, du théorème cité plus haut, et une recherche plus précise m'a convaincu que ce théorème, ou tout autre théorème équivalent, pouvait dans une certaine mesure constituer un fondement suffisant de l'Analyse infinitésimale. Il restait seulement alors à en découvrir l'origine propre dans les éléments de l'Arithmétique et à obtenir ainsi une véritable définition [*Definition*] de l'essence de la continuité. (Dedekind 2008, pages 59-60)

Troisième partie

Épistémologie mathématique
Corrigé des exercices

Corrigé des exercices pour le chapitre 1

1.1

Corrigé de l'exercice 1.1.2. Commençons par trois remarques préalables.

a. Une réponse à ce type de questions est rigoureuse si elle reste précise tout au long du raisonnement proposé quant au rapport qu'on cherche à établir entre libre mobilité et cas d'égalité des triangles. Cela n'a pas nécessairement à voir avec une validité des déductions que l'on fait, parce que cette validité dépend de tous les présupposés (les "axiomes") que l'on garde sur la géométrie, et ceux que l'on suspend. C'est-à-dire que libre mobilité et cas d'égalité des triangles ont chacun leur signification, mais comme ils sont profondément ancrés dans la géométrie euclidienne, il est très facile de croire qu'on a déduit une propriété de l'autre alors qu'on l'a utilisée implicitement. On peut cependant accéder au lien profond entre ces propriétés en se plaçant dans un cadre où ils ne sont plus nécessairement vrais. Par ailleurs, les cas d'égalité des triangles peuvent être vrais dans des géométries non euclidiennes, comme en géométrie sphérique et en géométrie hyperbolique.

b. On peut par exemple se placer dans le cadre plus général de la géométrie des surfaces courbes. Sur une surface courbe, par exemple sur un œuf, on peut encore tracer des figures et les segments ont encore un sens précis en considérant les lignes de plus court chemin entre leurs extrémités ; on peut encore parler d'égalités de segments et d'angles parce que la surface de l'œuf est plate au premier ordre dans toutes les directions. Cependant, si on trace un triangle sur le petit bout ou sur le gros bout de l'œuf, ils ne seront jamais égaux : si on trace sur un œuf deux segments de 1 cm issus d'un meme point et inclinés l'un sur l'autre de 60°, ces deux segments non seulement ne se complèteront plus en un triangle équilatéral, mais le troisième côté aura une longueur qui dépend de où on a tracé les deux segments. Le but de l'exercice est de comprendre cela par le fait que la figure du triangle ne peut pas être déplacée du petit bout au gros bout sans changer sa géométrie (les six paramètres de ses trois côtés et ses trois angles).

c. Quand on parle d'isométries, on pense à des transformations qui préservent les longueurs, mais il faut se souvenir que c'est équivalent à la préservation des angles comme le montre la formule de polarisation dans le cadre vectoriel : $4\langle\overrightarrow{\mathrm{AB}},\overrightarrow{\mathrm{AC}}\rangle = \|\overrightarrow{\mathrm{AB}}+\overrightarrow{\mathrm{AC}}\|^2 - \|\overrightarrow{\mathrm{AB}}-\overrightarrow{\mathrm{AC}}\|^2$.

La réponse intuitive aux questions posées est que dans la libre mobilité, on pense tant à l'existence qu'à l'unicité des transformations qui emmènent un drapeau sur un autre, mais

— que l'existence est un fait qui relève de la possibilité générale de faire de la géométrie, pas seulement euclidienne (d'ailleurs, on est un peu embêtés sur un œuf parce qu'on ne peut pas y parler de demi-plan)

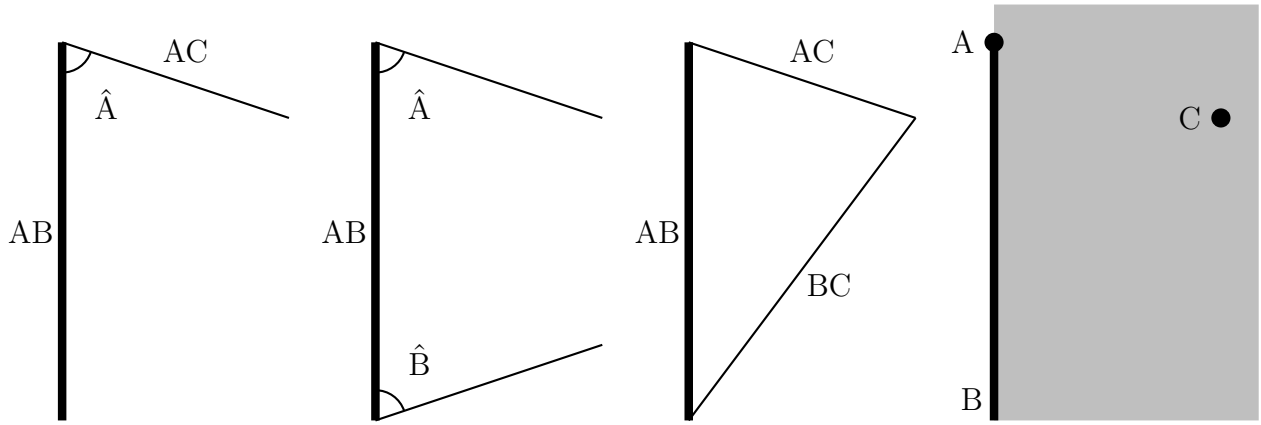
— et que c'est l'unicité qui a à voir avec les cas d'égalité des triangles.

En fait, chacun des trois cas d'égalité des triangles propose une variante du drapeau.

1. Le premier considère que deux côtés et un angle sont donnés : il propose la figure-drapeau composée d'un segment et d'un deuxième segment issu d'une extrémité du premier.
2. Le deuxième considère qu'un côté et deux angles sont donnés : il propose la figure-drapeau composée d'un segment et de deux demi-droites issues de ses extrémités dirigées du même côté du segment.

3. Le troisième considère que trois côtés sont donnés : il propose la figure-drapeau d'un triangle construit sur un segment.

Dessinons chacune de ces trois figures-drapeaux, avec le premier segment en trait épais considéré comme la hampe, et rajoutons en quatrième le drapeau défini par le point A, la demi-droite [AB) et le demi-plan bordé par (AB) contenant le point C, noté (AB; C).



Chaque cas d'égalité exprime alors que ses figures-drapeaux sont uniques, c'est-à-dire qu'elles ne dépendent pas du lieu où on les trace. Ce qu'exprime la libre mobilité par le drapeau est moins clair !

En se replaçant dans le cadre général des surfaces courbes, notons que la direction d'un segment issu d'un point donné est déjà déterminé dans le voisinage de ce point ; en particulier, les angles sont déjà déterminés au voisinage de leur sommet et fournissent seulement une information locale. Les deux premiers cas d'égalité affirment que ces informations locales permettent néanmoins d'obtenir des triangles qui sont égaux globalement.

1. Voici une manière de répondre plus rigoureusement à la question en mettant les drapeaux en rapport avec les figures-drapeaux.

Le premier cas d'égalité des triangles a pour hypothèses que

— le côté AB se superpose au côté A′B′,
— l'angle $\widehat{\mathrm{BAC}}$ se superpose à l'angle $\widehat{\mathrm{B'A'C'}}$,
— le côté AC se superpose au côté A′C′.

Considérons alors les trois isométries suivantes :

— la première hypothèse donne une isométrie qui envoie le segment [AB] sur [A′B′]. On peut supposer en toute généralité qu'elle envoie A sur A′, B sur B′ et le point C dans le demi-plan (A′B′; C′) : c'est intuitivement évident, mais la justification est complexe : il faut considérer les réflexions par rapport à la médiatrice du segment et par rapport à la droite.
— la deuxième hypothèse donne une isométrie qui envoie l'angle $\widehat{\mathrm{BAC}}$ sur $\widehat{\mathrm{B'A'C'}}$. Elle envoie A sur A′. On peut supposer en toute généralité qu'elle envoie [AB) sur [A′B′) et [AC) sur [A′C′) : c'est aussi intuitivement évident, mais il faut considérer la réflexion par rapport à la bissectrice de l'angle.
— la troisième hypothèse donne une isométrie qui envoie le segment [AC] sur [A′C′]. On peut supposer en toute généralité qu'elle envoie A sur A′, C sur C′ et le point B dans le demi-plan (A′C′; B′).

Notons alors que les deux premières isométries envoient en fait tout le demi-plan (AB; C) dans le demi-plan (A′B′; C′), et que les deux dernières isométries envoient tout le demi-plan (AC; B) dans le demi-plan (A′C′; B′) : cela demande aussi une justification, en lien avec le concept qu'une droite sépare le plan en deux demi-plans.

La libre mobilité exprime que de telles isométries sont uniques. Donc les deux premières coïncident, et les deux dernières également. Elles sont donc toutes égales. En conclusion, les trois sommets A, B, C sont bien envoyés sur les trois sommets A′, B′, C′.

Le deuxième cas d'égalité des triangles a pour hypothèses que

— l'angle $\widehat{\mathrm{BAC}}$ se superpose à l'angle $\widehat{\mathrm{B'A'C'}}$,
— le côté AB se superpose au côté A′B′,
— l'angle $\widehat{\mathrm{ABC}}$ se superpose à l'angle $\widehat{\mathrm{A'B'C'}}$.

Reprenons les deux premières isométries listées ci-dessus, et remplaçons la troisième ainsi :

— la troisième hypothèse donne une isométrie qui envoie l'angle $\widehat{\mathrm{ABC}}$ sur $\widehat{\mathrm{A'B'C'}}$: elle envoie B sur B′ et on peut supposer en toute généralité qu'elle envoie [BA) sur [B′A′) et [BC) sur [B′C′).

On conclut à nouveau par la libre mobilité que les trois isométries sont toutes égales. En particulier, elles envoient l'intersection C des demi-droites [AC) et [BC) sur l'intersection C′ des demi-droites [A′C′) et [B′C′).

Angles orientés ou géométriques ? En fait, nous considérons depuis le début des transformations qui ne sont pas nécessairement des déplacements au sens technique du mot : si les angles $\widehat{\mathrm{BAC}}$ et $\widehat{\mathrm{B'A'C'}}$ sont égaux en tant qu'angles géométriques mais opposés en tant qu'angles orientés, il faut d'abord en retourner un par une symétrie par rapport à une droite pour pouvoir les superposer par un déplacement, c'est-à-dire qu'ils se superposent seulement par un antidéplacement. En particulier, l'énoncé de l'exercice devait évoquer des placements (c'est-à-dire des antidéplacements en plus des déplacements) pour rester cohérent avec Euclide qui ne suppose pas les angles orientés ; mais l'énoncé reste correct si on suppose que les angles sont orientés.

2. Voyons maintenant comment déduire partiellement la libre mobilité, c'est-à-dire l'unicité de l'isométrie qui envoie un drapeau sur un autre, des cas d'égalité des triangles. Considérons un premier drapeau défini par le point A, la demi-droite [AB) et le demi-plan (AB; C), et un autre drapeau défini par le point A′, la demi-droite [A′B′) et le demi-plan (A′B′; C′). Soient i et j deux isométries qui envoient le premier drapeau sur le second, c'est-à-dire A sur A′, [AB) sur [A′B′) et (AB; C) sur (A′B′; C′). Alors $i(\mathrm{A}) = j(\mathrm{A}) = \mathrm{A'}$; comme i et j sont des isométries, $i(\mathrm{B})$ et $j(\mathrm{B})$ sont des points de [A′B′) tels que $\mathrm{A'}i(\mathrm{B}) = \mathrm{B'}j(\mathrm{B})$ et donc $i(\mathrm{B}) = j(\mathrm{B})$.

Déduisons maintenant que $i(\mathrm{C}) = j(\mathrm{C})$ du premier cas d'égalité des triangles : en effet, comme i et j sont des isométries, on a $\widehat{\mathrm{B'A'}i(\mathrm{C})} = \widehat{\mathrm{B'A'}j(\mathrm{C})}$ et $\mathrm{A'}i(\mathrm{C}) = \mathrm{A'}j(\mathrm{C})$.

Déduisons maintenant que $i(\mathrm{C}) = j(\mathrm{C})$ du deuxième cas d'égalité des triangles : en effet, comme i et j sont des isométries, on a $\widehat{\mathrm{B'A'}i(\mathrm{C})} = \widehat{\mathrm{B'A'}j(\mathrm{C})}$ et $\widehat{\mathrm{A'B'}i(\mathrm{C})} = \widehat{\mathrm{A'B'}j(\mathrm{C})}$.

Mais en fait, dans ce raisonnement, on peut remplacer B par n'importe quel point de la demi-droite [AB) et C par n'importe quel point du demi-plan (AB; C). Donc i et j coïncident sur cette demi-droite et ce demi-plan. De plus, i et j envoient aussi la demi-droite issue de A ne contenant pas B sur la demi-droite issue de A′ ne contenant pas B′, ainsi que le demi-plan bordé par (AB) ne contenant pas C sur le demi-plan bordé par (A′B′) ne contenant pas C′. On en conclut que $i = j$.

1.2

Corrigé de l'exercice 1.2.1. Prouvons l'associativité de l'addition. Pour cela, rappelons que l'addition est définie par les deux axiomes suivants pour un entier arbitraire m : $m + 0 = m$ et $m + (n') = (m + n)'$. C'est-à-dire que la suite (u_n) des sommes $m + n$ est définie par récurrence par : $u_0 = m$, $u_{n'} = u'_n$. Considérons la propriété suivante d'un entier naturel p : que, pour deux entiers naturels arbitraires m et n, on ait $(m + n) + p = m + (n + p)$.

— Elle est vraie pour $p = 0$ parce que $(0 + 0) + 0 = 0 + 0 = 0 + (0 + 0)$.
— Supposons qu'elle soit vraie pour un p. On a $(m+n)+(p') = ((m+n)+p)' = (m+(n+p))' = m + ((n + p)') = m + (n + (p'))$.

Par l'axiome de récurrence, cette propriété est vraie pour tout nombre p.

Prouvons la commutativité de l'addition. Pour cela, considérons la propriété suivante d'un entier naturel n : que, pour un entier naturel arbitraire m, on ait $m + n = n + m$.

— Nous allons prouver par récurrence qu'elle est vraie pour $n = 0$, c'est-à-dire que nous considérons la propriété suivante d'un entier naturel m : que $m + 0 = 0 + m$.

  — Elle est vraie pour $m = 0$ parce que $0 + 0 = 0$.

  — Supposons qu'elle soit vraie pour un m. On a $0 + (m') = (0 + m)' = (m + 0)' = m' = (m') + 0$.

  Par l'axiome de récurrence, cette propriété est vraie pour tout nombre m. Cette preuve montre aussi que 0 est **élément neutre pour l'addition**.

— Supposons qu'elle soit vraie pour un n. On a $m + (n') = (m + n)' = (n + m)' = n + (m')$. Prouvons par récurrence la propriété de l'entier naturel m que $n + (m') = (n') + m$, ce qui fera qu'on a $m + (n') = (n') + m$.

  — Elle est vraie pour $m = 0$ parce que $n + (0') = (n + 0)' = n' = (n') + 0$.

  — Supposons qu'elle soit vraie pour un m. On a $n + (m')' = (n + (m'))' = ((n') + m)' = (n') + (m')$.

  Par l'axiome de récurrence, cette propriété est vraie pour tout nombre m.

Par l'axiome de récurrence, cette propriété est vraie pour tout nombre n.

Nous pouvons en conclure que l'associativité de l'addition apparait ici comme une propriété beaucoup plus élémentaire que sa commutativité !

Prouvons la distributivité de la multiplication par rapport à l'addition. Pour cela, rappelons que la multiplication est définie par les deux axiomes suivants pour un entier arbitraire m : $m \cdot 0 = 0$ et $m \cdot (n') = (m \cdot n) + m$. C'est-à-dire que la suite (u_n) des produits $m \cdot n$ est définie par récurrence par : $u_0 = 0$, $u_{n'} = u_n + m$. Considérons la propriété suivante d'un entier naturel p : que, pour deux entiers naturels arbitraires m et n, on ait $m \cdot (n + p) = (m \cdot n) + (m \cdot p)$.

— Elle est vraie pour $p = 0$ parce que $m \cdot (n + 0) = m \cdot n = (m \cdot n) + 0 = (m \cdot n) + (m \cdot 0)$.

— Supposons qu'elle soit vraie pour un p. On a $m \cdot (n + (p')) = m \cdot ((n + p)') = (m \cdot (n + p)) + m = ((m \cdot n) + (m \cdot p)) + m$. Comme l'addition est associative, ceci est égal à $(m \cdot n) + ((m \cdot p) + m) = (m \cdot n) + (m \cdot (p'))$.

Par l'axiome de récurrence, cette propriété est vraie pour tout nombre p.

Voici d'autres propriétés que l'on peut montrer par récurrence :

— que 1 est **élément neutre pour la multiplication** ;

— que la multiplication est **associative** et **commutative** ;

— que l'addition est **simplifiable** : si $a + b = a + c$, alors $b = c$.

Je ne me suis pas inspiré du texte de Poincaré, mais la comparaison est intéressante, en particulier pour la démonstration de la commutativité.

Corrigé des exercices pour le chapitre 2

2.1

Corrigé de l'exercice 2.1.2. Soit ABCD un carré : ses quatre côtés sont égaux, ainsi que ses deux diagonales. Pour démontrer que le côté AB est incommensurable à la diagonale AC, nous allons considérer un segment I quelconque et étudier ce que veut dire que le segment I mesure les segments AB et AC, c'est-à-dire que I est contenu un nombre entier de fois dans AB et AC, en procédant par soustractions alternées (anthyphérèse).

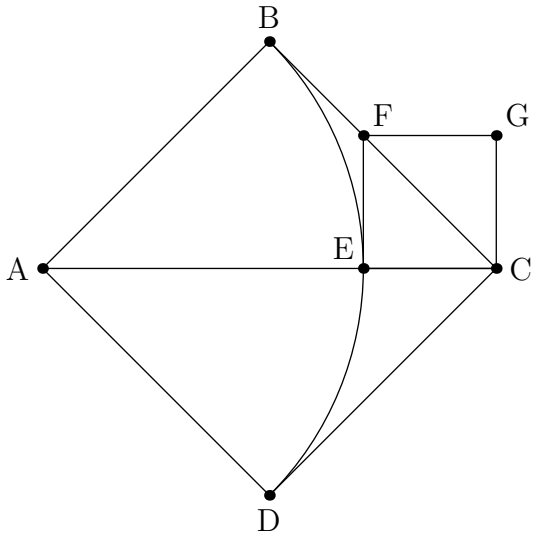


FIGURE 2.1.1 – L'anthyphérèse de la diagonale et du côté du carré

Par la définition V.4 des *Éléments* d'Euclide, on sait déjà que le segment I, étant multiplié, doit dépasser le segment AB.

Pour retrancher le segment AB au segment AC, traçons un arc du cercle de centre A passant par le point B jusqu'à la diagonale AC : il la coupe en un point E tel que AE = AB. Alors EC < AB et si le segment I mesure AB et AC, alors aussi le segment EC.

Pour retrancher le segment EC du segment AB, ou du segment BC, notons que la perpendiculaire au segment AC en le point E coupe le segment BC en un point F tel que EF = EC parce que les angles EFC et ECF sont égaux. De plus on a FB = FE : on peut le voir d'au moins trois manières différentes :

— les droites FB et FE sont les deux tangentes au cercle de centre A passant par les points B et E, menées depuis le point F ;

— les triangles ABF et AEF sont rectangles en B et E et ils ont les côtés AB, AF égaux aux côtés AE, AF, chacun à chacun : alors $AF^2 - AB^2 = AF^2 - AE^2$ et par le théorème de Pythagore on a $FB^2 = FE^2$;

— le triangle BAE est isocèle en A et donc les angles EBA et BAE sont égaux ; comme les angles FBA et FEA sont égaux puisque droits, les angles EBF et BEF sont égaux et donc le triangle BFE est isocèle en F.

Donc, si le segment I mesure AB et AC, alors aussi le segment FC. De plus, EC < FC et donc EC < $\frac{1}{2}$AB.

Soit G le symétrique du point E par rapport à la droite FC : EFGC est un carré. Donc, si le segment I mesure le côté et la diagonale du carré ABCD, il mesure aussi le côté et la diagonale du carré EFGC ; en particulier, il est plus petit que le segment EC et donc plus petit que la moitié du côté du carré ABCD. En réitérant cette construction, on montre que si le segment I mesure AB et AC, il est plus petit qu'un quart, huitième, seizième, etc., du segment AB : le segment I, multiplié par quatre, huit, seize, etc., ne dépasse pas AB. C'est une contradiction et donc I ne peut pas mesurer à la fois AB et AC.

On démontre en fait que si côté et diagonale du carré étaient commensurables, l'axiome d'Archimède serait en défaut. Mais l'axiome d'Archimède est-il réellement nécessaire pour mener la démonstration ? En fait, on peut démontrer qu'il est absurde de supposer que les segments AB et AC aient une commune mesure :

— supposer que le segment I mesure les segments AB et AC et expliciter qu'alors il existe un entier n tel que AB = nI ;

— avec la même démonstration, montrer que 2^kI < AB pour $k = 1, 2, 3$, etc. ;

— conclure que c'est absurde.

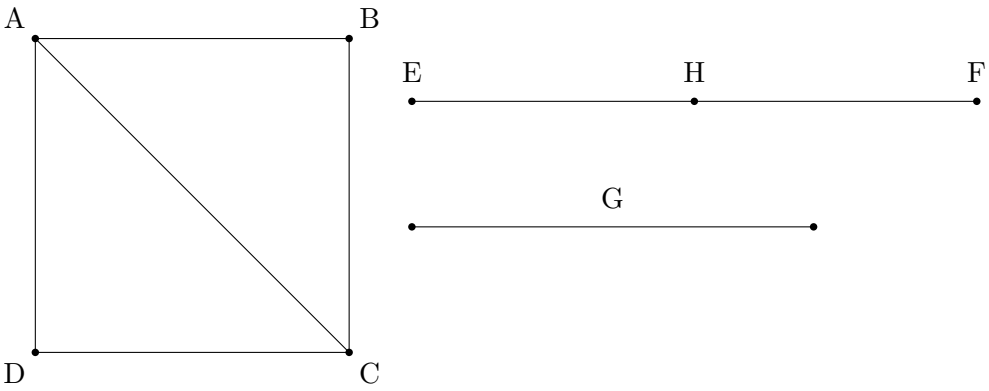
C'est une preuve de l'absurde bien plus qu'une preuve par l'absurde comme on dit souvent trop vite (voir Lombardi 1997). Donc l'axiome d'Archimède n'est pas nécessaire pour conclure. [1]

Cependant, notons que Jacques Harthong et Georges Reeb ont défini un modèle discret non standard de la droite réelle dans lequel l'axiome d'Archimède est en défaut et $\sqrt{2}$ est un nombre rationnel (dont le numérateur et le dénominateur sont infiniment grands) !

Voici en supplément une version « expurgée » par Bernard Vitrac (2012) de la proposition 117vulgo du livre X des *Éléments* d'Euclide. Cette proposition est considérée comme inauthentique et n'apparait donc pas dans les éditions modernes, mais elle remonte à l'Antiquité. Elle démontre l'absurde par la méthode du pair et de l'impair.

> Qu'il nous soit proposé de démontrer que la diagonale des figures carrées est incommensurable en longueur avec le côté.
>
> Soit le carré ABCD et AC sa diagonale ; je dis que CA est incommensurable en longueur avec AB.
>
> Si, en effet, c'est possible, qu'elle soit commensurable. Et puisque CA est commensurable avec AB, la [droite] CA, relativement à AB, a comme rapport celui d'un nombre relativement à un nombre. Que ce soit celui de EF relativement à G et que EF, G soient les plus petits parmi ceux ayant le même rapport qu'eux.

1. Les nombres sont définis comme étant une multiplicité d'unités et ils vérifient donc l'axiome d'Archimède par construction. C'est pourquoi on invoque l'axiome d'Archimède seulement pour des grandeurs d'un autre genre, comme les longueurs par exemple, qui ne sont pas construites comme une multiplicité.



Et puisque comme CA [est] relativement à AB, ainsi est EF relativement à G, donc aussi comme le [carré] sur CA [est] relativement à celui sur AB, ainsi [est] celui sur EF relativement à celui sur G. Or celui sur CA est double de celui sur AB ; donc celui sur EF est aussi double de celui sur G ; celui sur EF est donc pair ; de sorte que EF lui-même est aussi pair. Qu'il soit coupé en deux [parties égales] en H. Et puisque les [nombres] EF, G sont les plus petits parmi ceux qui ont le même rapport qu'eux, ils sont premiers entre eux. Et EF est pair ; donc G est impair. Et puisque EF est double de EH, le [carré] sur EF est quadruple donc de celui sur EH. Or celui sur EF est double de celui sur G ; donc celui sur G est double de celui sur EH ; donc celui sur G est pair. G est donc pair ; mais aussi impair ; ce qui est impossible. Ce n'est pas le cas que CA soit commensurable en longueur avec AB.

[Elle est donc incommensurable] ; ce qu'il fallait démontrer. (Vitrac 2012, page 49.)

2.4

Corrigé de l'exercice 2.4.1. Faisons d'abord deux remarques préliminaires.

a. La rédaction euclidienne représente les nombres par des segments. Il ne faut pas en déduire qu'Euclide postule que les nombres sont des longueurs. Il va seulement utiliser le langage géométrique pour représenter l'addition d'un nombre à un autre comme un prolongement de segment, et la soustraction d'un nombre à un autre comme un retranchement de segment. Dans notre reformulation, nous allons profiter de l'algèbre élémentaire de l'addition et de la multiplication des nombres et abandonner cette représentation.

b. Euclide décrit le procédé d'anthyphérèse de manière très compacte : "retrancher le plus petit du plus grand de façon réitérée et en alternance jusqu'à ce que le reste mesure le reste précédent". Nous ressentons là naturellement le besoin de préciser explicitement un algorithme. Dans les deux démonstrations, Euclide traite cependant un cas particulier où l'anthyphérèse s'arrête après quelques étapes, et ce cas particulier contient tous les arguments nécessaires pour conduire un raisonnement par récurrence. On peut exprimer cela aussi comme suit : Euclide comprend intuitivement le concept d'algorithme, ainsi que le principe de récurrence qui consiste à concevoir qu'un raisonnement peut être réitéré à l'infini, mais n'a pas l'idée de les isoler et de les exprimer en tant que tels, c'est-à-dire de les abstraire de leur application. Dans notre reformulation, nous restons fidèles à Euclide.

1. Reformulation de la proposition VII.2 : *Étant donnés deux nombres non premiers entre eux, trouver leur pgcd.*

Soient a, b les deux nombres non premiers entre eux donnés. Il faut trouver le pgcd de a, b. On peut supposer en toute généralité $a > b$.

Si b divise a, alors b est un commun diviseur de a, b. Et il est évident que c'est aussi le plus grand.

Si b ne divise pas a, faisons la division de a par b, et, aussi longtemps que le reste r ne divise

pas b, recommençons avec a prenant la valeur de b et b celle de r. Cet algorithme s'arrête parce que les restes successifs décroissent (strictement). Le dernier reste r sera différent de 1 parce que sinon les a, b de départ seraient premiers entre eux (proposition VII.1).

Supposons pour simplifier que l'algorithme aboutit au bout de trois divisions euclidiennes :

— $a = q_1 b + r_1$ avec $r_1 < b$ reste non nul ;

— $b = q_2 r_1 + r_2$ avec $r_2 < r_1$ reste non nul ;

— $r_1 = q_3 r_2$.

Or puisque r_2 divise r_1, r_2 divisera aussi b ; et donc r_2 divisera aussi a ; donc r_2 divise a, b. Donc r_2 est un commun diviseur de a, b. Montrons que c'est aussi le plus grand.

Car si r_2 n'est pas le pgcd de a, b, soit s un certain nombre qui divise a, b tout en étant plus grand que r_2.

Et puisque s divise a, b, s divise aussi le reste r_1. Puisque s divise b, r_1, s divise aussi le reste r_2. Or s est plus grand que r_2 : c'est impossible.

2. Le rapport logique apparent est le suivant. La proposition VII.2 fait appel à la proposition VII.1 pour constater que le dernier reste est un nombre et non pas une unité. Elle renforce la contraposée de VII.1 qui s'énonce qu'"étant donnés deux nombres non premiers entre eux, soit ils sont égaux, soit le processus d'anthyphérèse n'aboutit pas à une unité", parce que VII.2 rajoute que ce processus aboutit, et qu'il aboutit au pgcd des deux nombres.

Cependant, nous lecteurs modernes comptons l'unité parmi les nombres et sommes donc tentés de réunir les deux énoncés en enlevant

— de l'énoncé de VII.2 l'hypothèse que les nombres sont non premiers entre eux, et

— de sa démonstration la distinction s'il restera une unité ou non à l'issue de l'anthyphérèse.

Alors l'énoncé et la démonstration sont encore valides, et on prouve en passant VII.1 : si l'anthyphérèse aboutit à une unité, le pgcd vaut 1, c'est-à-dire que les nombres dont nous sommes partis sont premiers entre eux.

3. Il y a un parallélisme entre les propositions X.2 et X.3 d'une part et les propositions VII.1 et VII.2 d'autre part.

— Pour les grandeurs géométriques, l'anthyphérèse peut ne jamais terminer : alors la proposition X.2 prouve qu'elles sont incommensurables. Sinon, la proposition X.3 construit la plus grande commune mesure.

— Pour les grandeurs numériques, l'anthyphérèse se termine toujours ; si elle se termine sur une unité, alors la proposition VII.1 prouve qu'elles sont premières entre elles. Sinon, la proposition VII.2 construit la plus grande commune mesure (le pgcd).

On est tenté d'appeler les nombres premiers entre eux nombres "incommensurables", mais Euclide ne le fait pas. Les démonstrations des propositions respectent aussi ce parallélisme.

4. On retrouve bien l'algorithme 2.4.1 dans la proposition VII.1, d'abord dans la description de l'anthyphérèse : "et le plus petit étant retranché du plus grand de façon réitérée et en alternance" ; cette description est précisée dans la démonstration qu'on peut reformuler ainsi :

$$\begin{aligned} \mathrm{AB} &= q \times \mathrm{CD} + \mathrm{FA} \text{ pour un certain } q \\ \mathrm{CD} &= q' \times \mathrm{FA} + \mathrm{GC} \text{ pour un certain } q' \\ \mathrm{FA} &= q'' \times \mathrm{GC} + 1 \text{ pour un certain } q''. \end{aligned}$$

Il n'y a par contre pas de preuve que l'anthyphérèse se termine : cela figure déjà dans l'hypothèse. Dans la proposition VII.2, cette terminaison est annoncée, mais la raison n'est pas très explicite : c'est parce que chaque reste est (strictement) plus petit que le reste précédent.

2.6

Corrigé de l'exercice 2.6.1. Voici un algorithme qui calcule le pgcd et une relation de Bézout à partir de l'algorithme 2.6.1 implicite dans la preuve classique du théorème du pgcd.

Algorithme 2.6.1. ***Algorithme implicite dans la preuve classique, étendu.***
Entrée: Deux entiers naturels a et b, $a > b > 0$.
Sortie: Leur pgcd g ainsi que deux entiers relatifs u et v
vérifiant $ua + vb = g$.
Variables locales: r, r', q : entiers ≥ 0 ;
Début
initialisation
$g \leftarrow b$; $u \leftarrow 0$; $v \leftarrow 1$;
boucle
Répéter
$(q, r) \leftarrow$ quotient et reste de la division de a par g ;
Si $r > 0$ **alors** $g \leftarrow r$; $u \leftarrow 1 - qu$; $v \leftarrow -qv$ **fin si** ;
$(q, r') \leftarrow$ quotient et reste de la division de b par g ;
Si $r' > 0$ **alors** $g \leftarrow r'$; $u \leftarrow -qu$; $v \leftarrow 1 - qv$ **fin si** ;
jusqu'à ce que $r = r' = 0$
fin de boucle
Fin.

Corrigé de l'exercice 2.6.2. *Le premier problème* est d'ordre théorique : pourquoi diable un nombre entier admet-il une seule décomposition en facteurs premiers ? C'est bien sûr un fait d'expérience pour les petits entiers, mais pour les grands ?

Pour voir que la réponse n'est pas si facile, nous remplaçons l'anneau $\mathbb{Z}$ par le sous-anneau $\mathbf{A}$ de $\mathbb{Q}[\mathrm{X}]$ formé par les polynômes P de la forme $a + \mathrm{X}^2\mathrm{Q}(\mathrm{X})$ ($a \in \mathbb{Q}$, $\mathrm{Q}(\mathrm{X}) \in \mathbb{Q}[\mathrm{X}]$). On peut aussi dire : les polynômes P tels que $\mathrm{P}'(0) = 0$. Ou encore $\mathbf{A} = \mathbb{Q}[\mathrm{X}^2, \mathrm{X}^3]$. Le lecteur se convaincra facilement que $\mathrm{P}_2 := \mathrm{X}^2$ et $\mathrm{P}_3 := \mathrm{X}^3$ sont irréductibles dans $\mathbf{A}$ (le seul facteur unitaire strict de P_2 dans $\mathbb{Q}[\mathrm{X}]$ est X, mais X n'appartient pas à $\mathbf{A}$). Maintenant on voit le phénomène étrange suivant : le polynôme $\mathrm{X}^6 \in \mathbf{A}$ admet deux décompositions distinctes en produit de facteurs irréductibles dans $\mathbf{A}$:

$$\mathrm{X}^6 = {\mathrm{P}_2}^3 = {\mathrm{P}_3}^2.$$

Quant aux polynômes X^5 et X^6, leurs diviseurs communs (unitaires) sont 1, P_2 et P_3, et donc il n'y a pas de diviseur commun qui soit multiple de tous les autres diviseurs communs. Donc la propriété inattendue du pgcd de deux entiers donnée par le théorème 2.2.1 n'a rien qui lui corresponde dans $\mathbf{A}$.

En bref la divisibilité dans des anneaux un peu plus compliqués que $\mathbb{Z}$ offre des surprises désagréables.

Alors d'où vient l'unicité de la décomposition en facteurs premiers pour les entiers ? Elle vient de l'algorithme du pgcd, ou plus précisément du théorème 2.2.2, duquel on tire facilement que si un nombre premier p divise un produit ab il doit diviser l'un des facteurs. Cela s'appelle le lemme d'Euclide, mais c'est Gauss qui en a apparemment le premier donné une démonstration correcte.

La voici : si p ne divise pas a, puisque p n'a que 1 comme facteur strict, le seul facteur commun à p et a est égal à 1, d'où par la relation de Bézout une égalité $up + va = 1$. Alors si $pq = ab$, cela donne $b = (up + va)b = upb + vab = upb + vpq = p(ub + vq)$: p divise b.

Le deuxième problème est d'ordre pratique : s'il est vrai que pour les petits entiers l'algorithme consistant à les décomposer en facteurs premiers pour trouver leur pgcd est plus rapide que l'algorithme d'Euclide, cela ne va plus du tout pour les grands entiers. Une machine calcule en une fraction de seconde la relation de Bézout entre deux entiers à 1000 chiffres. Elle est en général tout à fait incapable (en l'état actuel de la science) par contre de trouver pour un tel entier les facteurs premiers qui auraient 300 chiffres.

Corrigé des exercices pour le chapitre 3

3.1

Corrigé de l'exercice 3.1.1. 1. Un jugement *a priori* est un jugement indépendant de l'expérience et même de toutes les impressions des sens : il advient non seulement indépendamment de telle ou telle expérience, mais d'une manière absolument indépendante de toute expérience.

2. Un jugement de la forme « A est B », c'est-à-dire « le sujet A a le prédicat B » est synthétique si le prédicat B est extérieur au concept A, bien qu'il soit tout de même en connexion avec lui : le jugement synthétique ajoute au concept du sujet A un prédicat B qui n'était pas pensé en lui et n'aurait pas pu en être tiré par une analyse de celui-ci.

3. Voici les jugements qui apparaissent dans les extraits cités :

   (a) Qui mine les fondations de sa maison doit s'attendre à ce qu'elle s'effondre.

   (b) — Tout changement a sa cause.
   — Tout ce qui arrive possède sa cause.

   (c) — Tous les corps sont étendus.
   — Tous les corps sont impénétrables.
   — Tous les corps ont une figure.

   (d) Tous les corps sont pesants.

   (e) $7 + 5 = 12$.

   Les jugements (a) et (d) ne sont pas *a priori* parce que seule l'expérience peut nous enseigner l'effet de la gravitation qui fait s'effondrer la maison et chuter les corps.

   Les jugements (a), (d), (e) et (b) sont synthétiques : la chute n'est pas pensée dans les corps ou la maison, le prédicat 12 n'est pas pensé dans la somme de 7 et de 5, la cause n'est pas pensée dans l'évènement.

   Les jugements (e) et (b) sont *a priori* puisqu'ils sont indépendants de toute expérience.

   Les jugements (c) sont *a priori* parce qu'ils sont analytiques : les prédicats de l'étendue, de l'impénétrabilité, de la figure sont déjà pensés dans le concept de corps et ne résultent pas de l'expérience.

   Voici six autres jugements :

   (a) Ce stylo est blanc.

   (b) Il neige.

   (c) Socrate est un homme.

   (d) Toute partie non vide de $\mathbb{N}$ admet un plus petit élément.

   (e) La vitesse de la lumière est finie.

   (f) Il [La Palisse] n'eût pas eu son pareil,
   s'il avait été seul au monde.

(Le jugement (f) est tiré de la *Chanson de La Palisse* de Bernard de la Monnoye.)

Les jugements (a), (b) et (e) sont *a posteriori* parce que seule l'expérience peut nous renseigner sur la couleur du stylo, la météorologie, la vitesse de la lumière. Ils sont tous synthétiques comme jugements d'expérience.

Le jugement (d) est synthétique *a priori* puisque c'est un jugement mathématique.

Les jugements (c) et (f) sont *a priori* parce qu'ils sont analytiques.

4. Poincaré adapte la classification des jugements par Kant à ses besoins d'épistémologie mathématique. Pour lui, les jugements mathématiques sont analytiques ou synthétiques selon qu'ils se ramènent ou non au principe d'identité (A est A) et au principe de contradiction (ce n'est pas le cas que A soit à la fois B et non B). Ces deux principes sont à la base du raisonnement syllogistique qu'il évoque au paragraphe I. En fait, il appelle analytiques les jugements qui peuvent être établis par simple vérification et pose la question provocante s'il y en a d'autres en mathématiques.

   Le paragraphe II explique que $2+2=4$ est un jugement analytique, et l'argument s'applique également à $7+5=12$. Cette différence avec Kant apparait déjà dans le paragraphe I : il ne cherche pas à faire la différence entre jugement analytique et jugement qui peut « sortir analytiquement d'un petit nombre de jugements synthétiques ». Dans ce cas précis, le jugement synthétique nécessaire pour démontrer que $2+2=4$ est l'égalité $x+2=(x+1)+1$, associativité qui n'est nullement pensée dans le concept de somme selon Kant : « Il faut sortir de ces concepts en s'aidant de l'intuition qui correspond à l'un des deux, par exemple ses cinq doigts [...], et ainsi ajouter l'une après l'autre les unités du cinq donné dans l'intuition au concept du sept. » Nous vous incitons à relire à ce sujet la section 1.2 du cours et en particulier le paragraphe *Ce qui se cache dans les arguments de comptage* !

   Pour Poincaré, le raisonnement par récurrence « est le véritable type du jugement synthétique *a priori* » et « il n'est que l'affirmation de la puissance de l'esprit qui se sait capable de concevoir la répétition indéfinie d'un même acte dès que cet acte est une fois possible ». Il permet de passer du fini à l'infini et produit ainsi des jugements qui ne peuvent être établis par simple vérification.

3.2

Corrigé de l'exercice 3.2.1. Voici des exemples tirés de copies d'étudiants.

1. Soit (u_n) la suite définie par récurrence par $u_0=0$ et $u_{n+1}=2u_n+1$. Alors $u_n=2^n-1$ pour tout n.

— C'est vrai si $n=0$.

— Supposons que ce soit vrai à un rang n donné. Alors

$$u_{n+1}=2u_n+1=2(2^n-1)+1=2^{n+1}-1$$

et donc c'est vrai aussi au rang $n+1$.

Voici une preuve avec points de suspension : on a $u_{n+1}+1=2(u_n+1)$ pour tout n et donc

$$u_n+1=2(u_{n-1}+1)=\underbrace{2\cdots 2}_{n \text{ fois}}(u_0+1)=2^n.$$

2. Soit (u_n) la suite de Fibonacci, définie par $u_0=0$, $u_1=1$ et $u_{n+2}=u_{n+1}+u_n$. Alors (u_n) est croissante. Comme $u_{n+2}-u_{n+1}=u_n$, il suffit de constater que $u_1\geq u_0$ et de montrer que la suite (u_n) est positive. Nous allons montrer que $u_n, u_{n+1}\geq 0$ pour tout n.

— On a bien $u_0, u_1\geq 0$.

— Supposons que ce soit vrai à un rang n donné. Alors $u_{n+2}=u_{n+1}+u_n\geq 0$ et donc $u_{n+1}, u_{n+2}\geq 0$: c'est vrai aussi au rang $n+1$.

Voici une preuve avec points de suspension : on a

$$\begin{aligned} u_{n+2} = u_{n+1} + u_n = u_n + u_{n-1} + u_n &= 2u_n + u_{n-1} \\ &= 2(u_{n-1} + u_{n-2}) + u_{n-1} = 3u_{n-1} + 2u_{n-2} \\ &= \cdots = \mathrm{A}u_1 + \mathrm{B}u_0 \end{aligned}$$

pour certains entiers positifs A et B. Comme $u_0, u_1 \geq 0$, $u_n \geq 0$ pour tout n.

3. Soient a, b deux nombres réels. Alors $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ pour tout n.

— C'est vrai pour $n = 0$: $(a+b)^0 = 1 = \binom{0}{0} \times 1 \times 1$.

— Supposons que ce soit vrai à un rang n donné. Alors

$$\begin{aligned} (a+b)^{n+1} &= (a+b) \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k+1} \\ &= \sum_{k=1}^{n+1} \binom{n}{k-1} (a^k b^{n-k+1}) + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k+1} \\ &= \binom{n}{n} a^{n+1} + \binom{n}{0} b^{n+1} + \sum_{k=1}^{n} \left(\binom{n}{k-1} + \binom{n}{k} \right) a^k b^{n+1-k} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} (a^k b^{n+1-k}) \end{aligned}$$

en appliquant les identités $\binom{n}{n} = \binom{n+1}{n+1}$, $\binom{n}{0} = \binom{n+1}{0}$ et $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$. Donc c'est vrai aussi au rang $n+1$.

Voici une preuve avec des points de suspension : en développant $(a+b)^n$, on obtient la somme de tous les produits $x_1 x_2 \cdots x_n$, où chacun des x_i parcourt les valeurs a et b. Chacun de ces produits est égal à $a^k b^{n-k}$ pour un k donné, et il y a autant de tels produits qu'il y a de cas où k parmi les x_i ont choisi la valeur a : il y en a donc $\binom{n}{k}$.

4. Considérons la suite des nombres de Fermat définie par $\mathrm{F}_n = 2^{2^n} + 1$. Alors on a $\mathrm{F}_0 \cdots \mathrm{F}_{n-1} = \mathrm{F}^n - 2$.

— C'est vrai si $n = 0$ parce qu'alors le produit n'a pas de terme et est donc égal à 1, et $\mathrm{F}_0 - 2 = 3 - 2 = 1$.

— Supposons que ce soit vrai à un rang n donné. Alors

$$\mathrm{F}_0 \cdots \mathrm{F}_n = (\mathrm{F}_n - 2)\mathrm{F}_n = (2^{2^n} - 1)(2^{2^n} + 1) = 2^{2 \times 2^n} - 1 = \mathrm{F}_{n+1} - 2$$

et donc c'est vrai aussi au rang $n+1$.

Voici une preuve avec des points de suspension :

$$\begin{aligned} \mathrm{F}_n - 2 = 2^{2^n} - 1 = (2^{2^{n-1}} + 1)(2^{2^{n-1}} - 1) &= \mathrm{F}_{n-1}(2^{2^{n-2}} + 1)(2^{2^{n-2}} - 1) \\ &= \cdots = \mathrm{F}_{n-1}\mathrm{F}_{n-2} \cdots \mathrm{F}_0 (2^{2^0} - 1). \end{aligned}$$

Il me semble qu'on comprend toujours mieux les démonstrations avec des points de suspension, parce que l'endroit précis où on réitère devient apparent, de même que les raisons qui expliquent pourquoi le résultat est vrai (sauf peut-être pour la suite de Fibonacci). Mais dans l'enseignement, beaucoup préfèrent (voire exigent) la rédaction précise du raisonnement par récurrence, en particulier parce qu'elle évite au lecteur de la démonstration de chercher le sens des points de suspension.

3.3

Corrigé de l'exercice 3.3.1. *Question 1.* Bien que la démonstration donnée semble procéder par l'absurde, elle contient une indication claire sur les calculs à faire pour obtenir le théorème sous forme explicite.

En effet considérons l'ensemble $\mathcal{A}$ des couples (A, V) tels que

— $\mathrm{A} \in \mathrm{G}_3$,

— V est un vecteur colonne dans $\mathbb{N}^3$ et

— $\mathrm{V} = \mathrm{A}\begin{bmatrix} a \\ b \\ c \end{bmatrix}$

Cet ensemble est non vide (prendre $\mathrm{A} = \mathrm{I}_3$) On voudrait obtenir un $(\mathrm{A}, \mathrm{V}) \in \mathcal{A}$, ${}^{\mathrm{t}}\begin{bmatrix} u & v & w \end{bmatrix}$ tel que $|\mathrm{V}| \overset{\text{def}}{=} u + v + w$ soit le plus petit possible. La démonstration nous donne une « recette » : comment faire pour diminuer $|\mathrm{V}|$ si celui-ci n'a pas deux coefficients nuls. On peut donc utiliser la recette proposée de manière itérée jusqu'à obtenir deux coefficients nuls. Le procédé s'arrête parce que, à chaque itération $|\mathrm{V}|$ diminue strictement.

On termine en ramenant le coefficient non nul en première position si ce n'est pas la cas.

En fait on réalise ainsi directement le but fixé par le théorème et c'est seulement ensuite, si on le désire, qu'on peut montrer que le nombre $|\mathrm{V}|$ obtenu in fine est bien le minimum possible (cela tient à ce que g est nécessairement le pgcd de (a, b, c)).

On peut faire les opérations suivantes qui conduisent au pgcd des 3 nombres, et à la transformation du triplet de départ en $(\mathrm{pgcd}, 0, 0)$.

1. $210 - 2 \times 90 = 30 : (90, 126, 210) \mapsto (90, 126, 30)$
2. $126 - 4 \times 30 = 6$, $90 - 3 \times 30 = 0 : (90, 126, 30) \mapsto (0, 6, 30)$
3. $30 - 5 \times 6 = 0 : (0, 6, 30) \mapsto (0, 6, 0)$
4. Échange des deux premiers : $(0, 6, 0) \mapsto (6, 0, 0)$

En termes matriciels, en initialisant C_0 avec I_3, cela donne la chose suivante (les cases vides sont des 0) :

1. $\mathrm{A}_1\mathrm{C}_0 = \begin{bmatrix} 1 & & \\ & 1 & \\ -2 & & 1 \end{bmatrix}\begin{bmatrix} 90 \\ 126 \\ 210 \end{bmatrix} = \begin{bmatrix} 90 \\ 126 \\ 30 \end{bmatrix} = \mathrm{C}_1$
2. $\mathrm{A}_2\mathrm{C}_1 = \begin{bmatrix} 1 & & -3 \\ & 1 & -4 \\ & & 1 \end{bmatrix}\begin{bmatrix} 90 \\ 126 \\ 30 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 30 \end{bmatrix} = \mathrm{C}_1$
3. $\mathrm{A}_3\mathrm{C}_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -5 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 6 \\ 30 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix} = \mathrm{C}_3$
4. $\mathrm{A}_4\mathrm{C}_3 = \begin{bmatrix} & 1 & \\ -1 & & \\ & & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \mathrm{C}_4$

Les matrices unitriangulaires A_1, A_2, A_3 sont évidemment de déterminant 1, ainsi que la matrice de transposition avec signe $-$. En définitive :

$$\mathrm{A}_4\mathrm{A}_3\mathrm{A}_2\mathrm{A}_1 = \mathrm{C}_4$$

avec $\det(\mathrm{A}_4\mathrm{A}_3\mathrm{A}_2\mathrm{A}_1) = 1$, ce qui donne bien le résultat cherché.

Notons qu'une multiplication à gauche par une matrice élémentaire revient à une manipulation de lignes. Ceci permet de calculer facilement le produit des 4 matrices.

Pour faire le produit A_2A_1 on retranche dans A_1 la 3e ligne 4 fois de la 2e et 3 fois de la première :

$$A_2A_1 = \begin{bmatrix}1 & 0 & -3\\ 0 & 1 & -4\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -2 & 0 & 1\end{bmatrix} = \begin{bmatrix}7 & 0 & -3\\ 8 & 1 & -4\\ -2 & 0 & 1\end{bmatrix}$$

Pour faire le produit $A_3A_2A_1$ on retranche dans A_2A_1 la 2e ligne 5 fois de la 3e :

$$A_3A_2A_1 = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -5 & 1\end{bmatrix}\begin{bmatrix}7 & 0 & -3\\ 8 & 1 & -4\\ -2 & 0 & 1\end{bmatrix} = \begin{bmatrix}7 & 0 & -3\\ 8 & 1 & -4\\ -42 & -5 & 21\end{bmatrix}$$

La dernière manipulation consiste à échanger les lignes 1 et 2, puis multiplier la ligne 2 par -1 :

$$A_4A_3A_2A_1 = \begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}7 & 0 & -3\\ 8 & 1 & -4\\ -42 & -5 & 21\end{bmatrix} = \begin{bmatrix}8 & 1 & -4\\ -7 & 0 & 3\\ -42 & -5 & 21\end{bmatrix}$$

Voici comment traiter la question 1 avec Maple.

```
> with(linalg):
```

```
> M0:=matrix(3,4,[[1,0,0,90],[0,1,0,126],[0,0,1,210]]);
```

$$M0 := \begin{bmatrix}1 & 0 & 0 & 90\\ 0 & 1 & 0 & 126\\ 0 & 0 & 1 & 210\end{bmatrix}$$

```
> M1:=addrow(M0,1,3,-2);
```

$$M1 := \begin{bmatrix}1 & 0 & 0 & 90\\ 0 & 1 & 0 & 126\\ -2 & 0 & 1 & 30\end{bmatrix}$$

```
> M2:=addrow(M1,3,1,-3); M3:=addrow(M2,3,2,-4);
```

$$M2 := \begin{bmatrix}7 & 0 & -3 & 0\\ 0 & 1 & 0 & 126\\ -2 & 0 & 1 & 30\end{bmatrix}$$

$$M3 := \begin{bmatrix}7 & 0 & -3 & 0\\ 8 & 1 & -4 & 6\\ -2 & 0 & 1 & 30\end{bmatrix}$$

```
> M4:=addrow(M3,2,3,-5);
```

$$M4 := \begin{bmatrix}7 & 0 & -3 & 0\\ 8 & 1 & -4 & 6\\ -42 & -5 & 21 & 0\end{bmatrix}$$

```
> mulrow(swaprow(M4,1,2),2,-1);
```

$$\begin{bmatrix}8 & 1 & -4 & 6\\ -7 & 0 & 3 & 0\\ -42 & -5 & 21 & 0\end{bmatrix}$$

Question 2. Ici il faut essayer de faire une rédaction courte mais convaincante. Voici une tentative.

Dans le calcul sur l'exemple, on s'aperçoit qu'il y a beaucoup de choix possibles pour déterminer quelles divisions on doit faire.

La matrice souhaitée A doit être initialisée à I_3.

Les divisions successives sont faites à l'intérieur d'une boucle.

Au début de chaque étape, nous choisissons de chercher parmi les trois coefficients de la colonne C un coefficient non nul le plus petit possible. Notons k l'indice de la ligne correspondante. Si $\mathrm{C}[k]$ est le seul coefficient non nul on sort de la boucle. Sinon, on effectue pour $i \neq k$ et $\mathrm{C}[i] \neq 0$ la division de $\mathrm{C}[i]$ par $\mathrm{C}[k]$. On effectue les manipulations de lignes correspondantes sur les matrices A et C.

Algorithme 3.3.1.

Entrée: Trois entiers naturels a et b, c, non tous nuls. On note $\mathrm{C} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

Sortie: Une matrice A à coefficients dans $\mathbb{Z}$, de déterminant 1 telle que $\mathrm{AC} = \begin{bmatrix} g \\ 0 \\ 0 \end{bmatrix}$.

Variables locales: i, k : numéros de lignes $(1, 2$ ou $3)$; q, r : entiers ≥ 0.

Début

initialisation

$\mathrm{A} \leftarrow \mathrm{I}_3$;

début de boucle

Boucle

Trouver un numéro de ligne k où se trouve la plus petite valeur non nulle $\mathrm{C}[k]$;

Si un seul coefficient $\mathrm{C}[k]$ est non nul **alors sortir de la boucle**

sinon

Trouver un indice $i \neq k$ tel que $\mathrm{C}[i]$ est non nul ;

Calculer le quotient q et le reste r de la division de $\mathrm{C}[i]$ par $\mathrm{C}[k]$;

Retrancher q fois la ligne k à la ligne i, dans C et dans A ;

fin si

fin de boucle

fin de boucle

Si $k \neq 1$ **alors**

Échanger les lignes k et 1 (dans A et dans C) et multiplier la ligne k par -1 dans A.

fin si

Fin.

En sortant de la boucle, il faut éventuellement échanger la ligne où se trouve le coefficient non nul avec la ligne 1. Il faut prendre garde cependant qu'un échange de lignes correspond au produit par une matrice de déterminant -1. En conséquence, on doit en plus multiplier une ligne (qui ne doit pas être la première) par -1.

N. B. : selon le logiciel utilisé, il y aura des variations, par exemple en Maple on fabriquera une seule matrice en accolant C à droite de A. Mais ce genre de détail n'a pas d'intérêt pour la compréhension de l'algorithme.

Un algorithme assez proche de tourner sur machine est indiqué en encadré.

Pour l'initialisation de la récurrence on remarque que la plus petite valeur prise par $a + b + c$ est 1. On peut initialiser la récurrence à 1 ou à 0, cela n'a pas d'importance. L'avantage d'initialiser à 0, c'est que, comme le cas ne se produit jamais, il n'y a rien à faire.

Traitons maintenant l'étape de récurrence.

Soit $k \geq 1$ et supposons donc qu'on sache résoudre le problème posé pour toutes les valeurs de a, b, c telles que $a+b+c < k$. Nous devons résoudre le problème avec a, b, c tels que $a+b+c = k$.

Deux cas de figure se présentent.

1. Ou bien 2 des 3 coefficients sont nuls. Alors le problème est résolu avec la matrice I_3 s'il s'agit du coefficient en position 1, et avec une des deux matrices $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ou $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$ dans le cas contraire.
2. Ou bien deux coefficients au moins sont non nuls. On divise le plus grand par le plus petit (ou on retranche le premier du second s'ils sont égaux). Si q est le quotient ceci correspond à une manipulation élémentaire de lignes $\mathrm{L}_i \leftarrow \mathrm{L}_i - q\mathrm{L}_k$, ce qui revient à multiplier à gauche la colonne $\mathrm{C} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ par une matrice B de déterminant 1 (B est unitriangulaire). La nouvelle colonne $\mathrm{C}' = \mathrm{AC}$ tombe sous le coup de l'hypothèse de récurrence. Il existe donc une matrice A de déterminant 1 telle que AC' a la forme voulue. Et la matrice AB résout le problème pour C puisque $\det(\mathrm{AB}) = \det(\mathrm{A})\det(\mathrm{B})$.

Question 3. Pour l'initialisation de la récurrence on remarque que la plus petite valeur prise par $a+b+c$ est 1. On peut initialiser la récurrence à 1 ou à 0, cela n'a pas d'importance. L'avantage d'initialiser à 0, c'est que, comme le cas ne se produit jamais, il n'y a rien à faire.

Traitons maintenant l'étape de récurrence.

Soit $k \geq 1$ et supposons donc qu'on sache résoudre le problème posé pour toutes les valeurs de a, b, c telles que $a+b+c < k$. Nous devons résoudre le problème avec a, b, c tels que $a+b+c = k$.

Deux cas de figure se présentent.

1. Ou bien 2 des 3 coefficients sont nuls. Alors le problème est résolu avec la matrice I_3 s'il s'agit du coefficient en position 1, et avec une des deux matrices $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ou $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$ dans le cas contraire.
2. Ou bien deux coefficients au moins sont non nuls. On divise le plus grand par le plus petit (ou on retranche le premier du second s'ils sont égaux). Si q est le quotient ceci correspond à une manipulation élémentaire de lignes $\mathrm{L}_i \leftarrow \mathrm{L}_i - q\mathrm{L}_k$, ce qui revient à multiplier à gauche la colonne $\mathrm{C} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ par une matrice B de déterminant 1 (B est unitriangulaire). La nouvelle colonne $\mathrm{C}' = \mathrm{AC}$ tombe sous le coup de l'hypothèse de récurrence. Il existe donc une matrice A de déterminant 1 telle que AC' a la forme voulue. Et la matrice AB résout le problème pour C puisque $\det(\mathrm{AB}) = \det(\mathrm{A})\det(\mathrm{B})$.

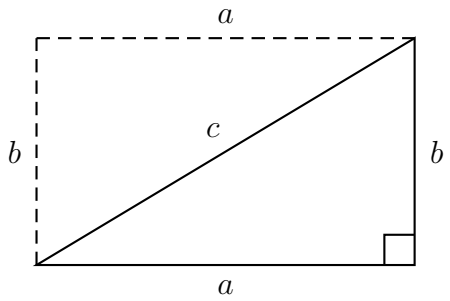
Corrigé de l'exercice 3.3.2. Fermat raconte à Carcavi qu'il a développé une nouvelle méthode, qu'il appelle *descente infinie*, pour montrer certains théorèmes négatifs, comme :

— l'équation $3a - 1 = b^2 + 3c^2$ n'a pas de solution avec a, b, c entiers naturels strictements positifs ;

— le système d'équations $\begin{cases} a^2 + b^2 = c^2 \\ \dfrac{a \cdot b}{2} = d^2 \end{cases}$ n'a pas de solution avec a, b, c, d entiers naturels strictements positifs.

(En effet, être moindre de l'unité qu'un multiple de 3, c'est être de la forme $3a - 1$; un triangle rectangle en nombres est un triangle dont les côtés sont des multiples a, b, c d'une unité de longueur

tels que $a^2 + b^2 = c^2$, et ce triangle est alors de même aire que la moitié du rectangle de côtés a, b fois cette unité de longueur.)



Fermat dit que la preuve se fait par réduction à l'absurde. Par exemple, il explique que pour prouver le deuxième théorème, il suffit de procéder ainsi : supposer que le système aurait une solution avec a, b, c, d entiers naturels strictements positifs et en déduire qu'alors il y aurait une nouvelle solution telle que le nouveau d serait strictement plus petit que l'ancien. C'est le cœur de la descente infinie. En effet, le dire une fois revient à le dire sans cesse, et on trouverait successivement des solutions $a = a^{(n)}, b = b^{(n)}, c = c^{(n)}, d = d^{(n)}$ avec $a^{(n)}, b^{(n)}, c^{(n)}, d^{(n)}$ entiers naturels strictements positifs tels que $d^{(1)} > d^{(2)} > d^{(3)} > \cdots$. Or il n'y a pas de descente infinie dans les entiers naturels strictements positifs. C'est donc absurde.

Fermat précise que la déduction de la nouvelle solution avec un d strictement plus petit que l'ancien est très astucieuse.

Puis Fermat discute un théorème positif : que tout nombre premier de la forme $p = 4n + 1$ s'écrit $p = a^2 + b^2$ avec a, b entiers naturels strictements positifs. Il en propose une preuve par réduction à l'absurde : il suppose qu'un nombre premier $p = 4n + 1$ n'admettrait pas d'écriture de la forme $p = a^2 + b^2$ et montre alors comment on trouverait un nouveau p de la même forme, strictement plus petit, et n'admettant pas non plus d'écriture de la forme $p = a^2 + b^2$. C'est le cœur de la descente infinie. En effet, le dire une fois revient à le dire sans cesse, et on trouverait successivement des nombres premiers $p = p^{(k)}$ de la forme $4n + 1$ n'admettant pas d'écriture de la forme $p = a^2 + b^2$ avec $p^{(1)} > p^{(2)} > p^{(3)} > \cdots$. Or il n'y a pas de descente infinie dans les entiers naturels strictements positifs. C'est donc absurde.

Fermat n'a pas publié de preuve de cet énoncé.

Corrigé des exercices pour le chapitre 5

5.1

Corrigé de l'exercice 5.1.1 (Les rapports de grandeurs)**.** Il fallait lire des extraits d'un livre d'Augustus De Morgan qui interprète la notion de rapport de deux grandeurs chez Euclide d'Alexandrie comme la donnée d'une échelle relative de multiples de ces deux grandeurs.

(1) Premier extrait : l'algorithme d'Euclide.

Dans cet extrait, De Morgan s'inscrit dans une tradition du commentaire du cinquième livre d'Euclide qui remonte au Moyen Âge, dont le plus ancien qui nous soit parvenu est du IX^e^ siècle et dû à al-Māhānī, et dont le plus connu est *Sur certaines prémisses problématiques du Livre d'Euclide* d'al-Khayyām, du XII^e^ siècle : voir à ce sujet Bernard Vitrac (2002).

1. Les substitutions successives de De Morgan correspondent exactement à l'algorithme d'Euclide : seule la notation change. Je reformule donc la proposition X.2 avec les notations de De Morgan.
   Si on prend deux grandeurs inégales, qu'on soustrait la plus petite de la plus grande aussi souvent que possible (c'est-à-dire "de façon réitérée"), puis qu'on recommence en prenant la plus petite grandeur et le reste de la soustraction (c'est-à-dire "en alternance"), et qu'on peut continuer ainsi indéfiniment, alors les deux grandeurs dont on est parti sont incommensurables.
   Soient A et B deux grandeurs avec $\mathrm{B} < \mathrm{A}$ telles qu'en soustrayant B à A un nombre β de fois il reste une grandeur $\mathrm{B}' = \mathrm{A} - \beta\mathrm{B}$ inférieure à B (strictement positive !), et telles qu'en recommençant avec B et B′ à la place de A et B et ainsi toujours, apparaissent des grandeurs restantes B″, B‴, etc. (strictement positives ! s'il arrivait que la première grandeur fût un multiple de la deuxième, cela interromprait la construction).
   Montrons qu'il est absurde de supposer que A et B aient une commune mesure E.
   En effet, si on a une grandeur E qui mesure A et B, E mesure aussi B′, et donc aussi B″, B‴, etc. En particulier, on a $\mathrm{E} < \mathrm{B}''$, B‴, etc. Or on a $\mathrm{A} > \mathrm{B} > \mathrm{B}' > \mathrm{B}'' > \mathrm{B}''' >$ etc., et Euclide prétend que ces grandeurs deviennent aussi petites que l'on veut, en particulier plus petites que E. C'est absurde.
   Cette démonstration ne me satisfait pas tout à fait parce qu'Euclide ne précise pas comment les grandeurs $\mathrm{A}, \mathrm{B}, \mathrm{B}', \mathrm{B}'', \mathrm{B}'''$ décroissent. En regardant de plus près, on observe que $\mathrm{A} = \mathrm{B} + \beta\mathrm{B}' > \mathrm{B}' + \mathrm{B}'$ et donc $\mathrm{B}' < \frac{1}{2}\mathrm{A}$. De même, $\mathrm{B}''' < \frac{1}{2}\mathrm{B}'$, etc. Ainsi B′, B‴, etc. sont plus petits que la moitié, le quart, etc. de A.
2. — On a $\mathrm{B}^{(1)} = \mathrm{A} - \beta\mathrm{B}$ et donc $p_1 = 1$ et $q_1 = \beta$ conviennent.
   — On a $\mathrm{B}^{(2)} = (\beta\beta' + 1)\mathrm{B} - \beta'\mathrm{A}$ et donc $p_2 = \beta'$ et $q_2 = \beta\beta' + 1$ conviennent.
   — Alors
   $$\begin{aligned}\mathrm{B}^{(3)} &= \mathrm{B}^{(1)} - \beta''\mathrm{B}^{(2)}\\ &= p_1\mathrm{A} - q_1\mathrm{B} - \beta''(q_2\mathrm{B} - p_2\mathrm{A})\\ &= (p_1 + \beta'' p_2)\mathrm{A} - (q_1 + \beta'' q_2)\mathrm{B}\end{aligned}$$

et donc $p_3 = p_1 + \beta'' p_2$ et $q_3 = q_1 + \beta'' q_2$ conviennent.

— De même

$$\begin{aligned} \mathrm{B}^{(4)} &= \mathrm{B}^{(2)} - \beta'''\mathrm{B}^{(3)} \\ &= q_2\mathrm{B} - p_2\mathrm{A} - \beta'''(p_3\mathrm{A} - q_3\mathrm{B}) \\ &= (q_2 + \beta''' q_3)\mathrm{B} - (p_2 + \beta''' p_3)\mathrm{A} \end{aligned}$$

et donc $p_4 = p_2 + \beta''' p_3$ et $q_4 = q_2 + \beta''' q_3$ conviennent.

3. On constate par ces calculs qu'on peut définir les coefficients p_n, q_n par récurrence en posant

$$\begin{aligned} p_1 &= 1 & q_1 &= \beta \\ p_2 &= \beta' & q_2 &= \beta\beta' + 1 \\ p_{n+1} &= p_{n-1} + \beta^{(n)} p_n & q_{n+1} &= q_{n-1} + \beta^{(n)} q_n \end{aligned}$$

(plutôt que de spécifier p_2 et q_2, on aurait aussi pu poser $p_0 = 0$ et $q_0 = 1$).

(2) Deuxième extrait : propriétés des coefficients p et q.

Comme $\mathrm{B}^{(1)}$, $\mathrm{B}^{(2)}$, $\mathrm{B}^{(3)}$, etc. sont des grandeurs, $p_1\mathrm{A} - q_1\mathrm{B}$, $q_2\mathrm{B} - p_2\mathrm{A}$, $p_3\mathrm{A} - q_3\mathrm{B}$, etc. sont des grandeurs (strictement positives !) et donc $p_1\mathrm{A} > q_1\mathrm{B}$, $q_2\mathrm{B} > p_2\mathrm{A}$, $p_3\mathrm{A} > q_3\mathrm{B}$, etc. De plus, ces grandeurs sont toutes inférieures à B : donc $p_1\mathrm{A} < (q_1 + 1)\mathrm{B}$, $(q_2 - 1)\mathrm{B} < p_2\mathrm{A}$, $p_3\mathrm{A} < (q_3 + 1)\mathrm{B}$, etc.

Pour les points 2 et 3, il faut d'abord se rendre compte que la définition par récurrence des suites des coefficients p et q permet d'affirmer qu'elles sont strictement croissantes et qu'elles tendent donc vers l'infini, puis étudier le sens de variation des fractions $\dfrac{q_n}{p_n}$ avec n en considérant

$$\frac{q_{n+1}}{p_{n+1}} - \frac{q_n}{p_n} = \frac{q_{n+1}p_n - p_{n+1}q_n}{p_{n+1}p_n}$$

et en notant que

$$\begin{aligned} q_{n+1}p_n - p_{n+1}q_n &= (q_{n-1} + \cancel{\beta^{(n)}q_n})p_n - (p_{n-1} + \cancel{\beta^{(n)}p_n})q_n \\ &= -(q_n p_{n-1} - p_n q_{n-1}). \end{aligned}$$

Or $q_1p_0 - p_1q_0 = -1$: donc $q_2p_1 - p_2q_1 = 1$, $q_3p_2 - p_3q_2 = -1$, etc. Ceci prouve déjà que $\dfrac{q_{n+1}}{p_{n+1}} - \dfrac{q_n}{p_n}$ devient arbitrairement petit : c'est le point 3. Ceci montre aussi que la suite des fractions $\dfrac{q_n}{p_n}$ n'est pas monotone. Mais

$$\frac{q_{n+1}}{p_{n+1}} - \frac{q_{n-1}}{p_{n-1}} = \frac{q_{n+1}p_{n-1} - p_{n+1}q_{n-1}}{p_{n+1}p_{n-1}}$$

et

$$\begin{aligned} q_{n+1}p_{n-1} - p_{n+1}q_{n-1} &= (\cancel{q_{n-1}} + \beta^{(n)}q_n)p_{n-1} - (\cancel{p_{n-1}} + \beta^{(n)}p_n)q_{n-1} \\ &= \beta^{(n)}(q_n p_{n-1} - p_n q_{n-1}) \\ &= (-1)^n \beta^{(n)}, \end{aligned}$$

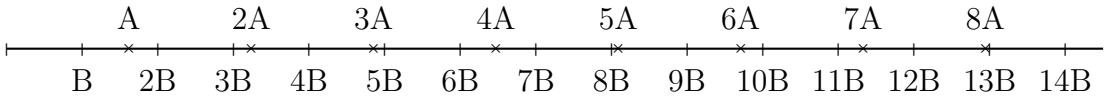
ce qui prouve que les $\dfrac{q_1}{p_1}$, $\dfrac{q_3}{p_3}$, etc. vont en ordre croissant et que les $\dfrac{q_2}{p_2}$, $\dfrac{q_4}{p_4}$, etc. vont en ordre décroissant : c'est le point 2.

(3) Troisième extrait : un exemple. Euclide dit qu'un segment A est « coupé en extrême et moyenne raison » s'il est composé de deux parties B et B′ telles que dans l'extrait. (Cette expression étrange s'explique ainsi : on peut vérifier à l'aide de la question suivante qu'alors les grandeurs A et B sont dans le même rapport que les grandeurs B et B′ ; Euclide appelle alors A et B′ les grandeurs *extrêmes* de cette égalité de rapports (ou *raisons*), et B la grandeur *moyenne*. Ce rapport est aujourd'hui appelé le « nombre d'or ».)

1. La première ligne du tableau est la suite des indices n, la deuxième celle des dénominateurs p_n en fonction de n, et la troisième celle des numérateurs q_n en fonction de n. Par exemple, $p_9 = 34$ et $q_9 = 55$.
2. Il faut lire « A est
   (a) plus grand que B mais plus petit que 2B,
   (b) plus grand que $\frac{3}{2}$B mais plus petit que $\frac{5}{3}$B,
   (c) plus grand que $\frac{8}{5}$B mais plus petit que $\frac{13}{8}$B. »

   Chacune de ces trois assertions donne un encadrement de A, 2A, 3A, etc. en termes de multiples de B, mais on sait que c'est la (c) qui est la plus précise. Elle donne que
   — 2A est entre $\frac{16}{5}$B et $\frac{13}{4}$B, et *a fortiori* entre 3B et 4B ;
   — 3A est entre $\frac{24}{5}$B et $\frac{39}{8}$B, et *a fortiori* entre 4B et 5B ;
   — 4A est entre $\frac{32}{5}$B et $\frac{13}{2}$B, et *a fortiori* entre 6B et 7B ;
   — 5A est entre 8B et $\frac{65}{8}$B, et *a fortiori* entre 8B et 9B.

   En fait, l'assertion (b) suffit à encadrer 2A, 3A et 4A entre deux multiples successifs de B.
3. Le tableau donne que A est plus grand que $\frac{144}{89}$B mais plus petit que $\frac{233}{144}$B. Regardons si cela suffit pour encadrer 100A entre deux multiples successifs de B : 100A est entre $\frac{14400}{89}$B et $\frac{5825}{36}$B, or $\frac{14400}{89} = 161 + \frac{71}{89}$ et $\frac{5825}{36} = 161 + \frac{29}{36}$. Donc 100A est entre 161B et 162B. En fait, cet encadrement est déjà conséquence du fait que A est plus grand que $\frac{21}{13}$B mais plus petit que $\frac{34}{21}$B !
4. Voici le commencement de l'échelle des multiples :



(4) Quatrième extrait : échelle relative et rapport.

1. Reformulons la définition d'Euclide : Une grandeur A est dans le même rapport à une grandeur B qu'une grandeur C à une grandeur D si on a pour des équimultiples mA et mC et des équimultiples nB et nD soit en même temps $m\text{A} > n\text{B}$ et $m\text{C} > n\text{D}$, soit en même temps $m\text{A} = n\text{B}$ et $m\text{C} = n\text{D}$, soit en même temps $m\text{A} < n\text{B}$ et $m\text{C} < n\text{D}$.

   Donc, connaitre le rapport d'une grandeur A à une grandeur B, c'est pouvoir décider, pour tous les couples d'entiers (strictement positifs) m et n, si on a $m\text{A} > n\text{B}$ ou $m\text{A} = n\text{B}$ ou $m\text{A} < n\text{B}$. C'est connaitre l'échelle relative des multiples de A et de B.
2. Pour connaitre le rapport d'une grandeur A à une grandeur B, il ne suffit pas de connaitre le début de l'échelle relative des multiples. Plus on poussera cette échelle loin et mieux on connaitra ce rapport, mais on devra se contenter malgré tout d'une connaissance finie, à moins que l'on découvre que cette échelle obéit à une certaine loi : c'est le cas si le rapport est commensurable et c'est aussi le cas du nombre d'or étudié en (3).

3. L'approche de De Morgan pourvoit un support visuel intuitif pour la notion de rapport, l'échelle des multiples. De plus, ce support contient tout ce qu'on peut connaitre d'*un* rapport donné, alors qu'Euclide ne définit pas *un* rapport mais l'*égalité* de *deux* rapports. Le rôle des nombres, qui sont présents dans la définition euclidienne à travers les équimultiples, apparait de manière plus évidente dans la définition de De Morgan. Elle révèle aussi l'infini qui se cache dans toute définition du rapport : il est simplement dissimulé lorsqu'on définit le rapport comme le nombre réel limite de la suite des nombres rationnels p_n/q_n.

Corrigé de l'exercice 5.1.2. 1. En ce qui concerne le rapport entre deux nombres, al-Khayyām distingue d'abord entre le rapport d'égalité et d'inégalité, puis, si les nombres sont inégaux, il constate qu'il y a deux possibilités :

   — soit le plus petit divise le plus grand (par exemple, 3 divise 9 : $3+3+3=9$),

   — soit il ne le divise pas, et alors ils ont un diviseur commun (par exemple, 1 divise 2 et 7 : $1+1=2$ et $1+1+1+1+1+1+1=7$).

2. Pour moi, on peut parler de grandeur d'un certain genre à chaque fois que l'on sait

   — additionner deux grandeurs de ce genre,

   — décider quand une grandeur est partie d'une autre grandeur du même genre,

   — et si c'est le cas, soustraire la première grandeur de la deuxième.

   La géométrie fournit des exemples de grandeurs : la ligne, la surface, le volume, l'angle. La physique en fournit beaucoup d'autres : la masse, la durée, etc.

   Les grandeurs se distinguent des nombres en ce qu'elles sont indéfiniment divisibles.

3. Al-Khayyām décrit l'algorithme d'Euclide pour les nombres et explique qu'il aboutit parce que les nombres sont composés d'unités indivisibles.

4. Cette « troisième division » correspond à l'incommensurabilité étudiée dans le dixième livre des *Éléments* d'Euclide. Aujourd'hui on utilise davantage le terme d'irrationalité. Les premiers exemples en sont le rapport du côté et de la diagonale d'un carré et le rapport du côté et de la diagonale d'un pentagone régulier. Un exemple plus difficile est le rapport d'un disque et d'un carré exinscrit.

5. Étudions l'exemple proposé par al-Khayyām : quatre grandeurs g_1, g_2, g_3 et g_4 sont données telles que

$$\text{pour tous entiers positifs } m \text{ et } n, \text{ on a } \begin{cases} \text{si } mg_1 > ng_2, \text{ alors } mg_3 > ng_4, \\ \text{si } mg_1 = ng_2, \text{ alors } mg_3 = ng_4, \\ \text{si } mg_1 < ng_2, \text{ alors } mg_3 < ng_4, \end{cases}$$

   $g_1 = \dfrac{g_2}{2}$.

   Peut-on en déduire que $g_3 = \dfrac{g_4}{2}$? En d'autres mots, la proportionnalité euclidienne implique-t-elle la proportionnalité véritable ? Il s'avère que c'est bien le cas, mais la démonstration n'est pas aussi naturelle qu'elle devrait l'être, même si elle est très simple : il faut constater que l'on a $2g_1 = g_2$ et que donc on a $2g_3 = g_4$ en appliquant la deuxième ligne de l'accolade avec $m = 2$ et $n = 1$.

6. Dans les trois cas considérés au § 13, le rapport de la première grandeur à la deuxième est égal au rapport de la troisième grandeur à la quatrième selon des rapports respectivement de la forme $\dfrac{1}{1}$, $\dfrac{1}{n}$ pour un entier n, $\dfrac{m}{n}$ pour deux entiers m et n : ce sont bien des rapports numériques.

7. La définition proposée par al-Khayyām consiste à appliquer simultanément une anthyphérèse de la première et de la deuxième grandeur et une anthyphérèse de la troisième et de la quatrième grandeur, et de comparer à chaque étape les quotients qui apparaissent dans chacune des anthyphérèses. Si ces quotients sont égaux à toutes les étapes, alors les rapports recherchés sont égaux : mis en formules, cela donne

$$g_2 = \boxed{q_1} \cdot g_1 + r_1 \text{ avec } r_1 < g_1, \quad g_4 = \boxed{q_1} \cdot g_3 + r'_1 \text{ avec } r'_1 < g_3,$$
$$g_1 = \boxed{q_2} \cdot r_1 + r_2 \text{ avec } r_2 < r_1, \quad g_3 = \boxed{q_2} \cdot r'_1 + r'_2 \text{ avec } r'_2 < r'_1,$$
$$r_1 = \boxed{q_3} \cdot r_2 + r_3 \text{ avec } r_3 < r_2, \quad r'_1 = \boxed{q_3} \cdot r'_2 + r'_3 \text{ avec } r'_3 < r'_2, \quad \text{etc.}$$

Nous avons rencontré cette définition dans le texte de De Morgan étudié dans l'exercice précédent.

Corrigé de l'exercice 5.1.3. Le premier argument est appelé *dichotomie* parce que le trajet que le mobile doit parcourir est coupé en deux (« -tomie » coupure, « dicho- » en deux). Comment cela peut-il fournir un argument contre le mouvement ? C'est que le dire une fois revient à le dire sans cesse : le mobile doit, avant d'atteindre le but de son trajet, parvenir à la moitié du trajet, au quart, au huitième, au seizième, etc., et tout cela est le cas avant même qu'il se soit élancé pour entreprendre son mouvement. Aristote constate effectivement que le mobile doit parvenir à une infinité de lieux avant d'atteindre son but. Comment est-ce possible ?

Le second argument, l'*Achille*, est connu comme « paradoxe d'Achille et de la tortue ». Dans celui-ci, deux protagonistes font la course, mais le plus rapide laisse une avance au plus lent. Aristote propose de décrire la course ainsi :

— au temps t_0, le plus rapide est en A_0 et le plus lent en T_0, et la course commence ;
— au temps t_1, le plus rapide arrive en $\mathrm{A}_1 = \mathrm{T}_0$ et le plus lent est arrivé entretemps en T_1 ;
— au temps t_2, le plus rapide arrive en $\mathrm{A}_2 = \mathrm{T}_1$ et le plus lent est arrivé entretemps en T_2 ;
— etc.

Donc, aux temps successifs t_0, t_1, t_2, etc., le plus lent garde une avance sur le plus rapide.

Dans le cours, on propose un paradoxe qui ressemple à la *dichotomie*, mais les moitiés successives sont celles du trajet *restant* à parcourir, et non du trajet initial. Cela permet de considérer successivement la moitié, les trois quarts, les sept huitièmes, les quinze seizièmes, etc. du trajet, et d'exprimer qu'ils sont tous strictement plus petits que le trajet initial. Que signifie alors la limite $\dfrac{2^n - 1}{2^n} \xrightarrow[n\to\infty]{} 1$?

Dans sa critique, Aristote considère deux grandeurs et le temps ; les deux grandeurs en question sont le trajet et le mobile. Les paradoxes exploitent le fait que le trajet, une longueur, est illimité en division. Partant, le mobile lui-même est illimité dans la mesure où il passe par tous les points issus de la division illimitée. C'est alors qu'Aristote dénonce que Zénon fait une « prémisse fausse » : qu'il conçoit que le temps du trajet et le temps de la course jusqu'à ce que le plus rapide rattrape le plus lent sont limités. En fait, ces temps sont « finis selon les extrémités », c'est-à-dire des intervalles de temps bornés, mais ils sont par contre illimités selon la division. Il n'y a donc pas de paradoxe.

N.B. : il y a un seul mot en grec pour dire « illimité » et « infini » : c'est ἄπειρον (apeiron) qui veut dire « sans limite ». Jean-Paul Dumont a utilisé les deux mots pour sacrifier à la tradition qui traduit « infini selon la division », « infini selon les extrémités ».

5.5

Corrigé de l'exercice 5.5.1. — Le paradoxe de Richard est le suivant. Il considère l'ensemble E des nombres réels qui peuvent être définis par une expression de longueur finie

(un « arrangement ») écrite avec les lettres d'un alphabet fini. Ces expressions peuvent être rangées sous la forme d'une suite, en les rangeant d'abord selon leur longueur et puis, à longueur fixée, selon l'ordre lexicographique défini à partir d'un ordre des lettres de l'alphabet. Puis il propose de définir un nombre réel différent de tous les nombres réels de E en utilisant l'argument diagonal de Cantor. Il constate que ce nombre réel est lui-même défini par une expression G de longueur finie et qu'elle figure par conséquent dans la suite ci-dessus. C'est paradoxal.

— Richard cherche à comprendre ce que désigne vraiment l'expression G. Il s'agit d'une définition qui contient une référence à l'ensemble E. Il faut donc que l'ensemble E soit défini en amont de la définition de G. Or Richard note que réciproquement l'expression G, du fait qu'elle est de longueur finie, rentre dans la définition de l'ensemble E : cela crée un cercle vicieux. Il propose donc de concevoir que l'ensemble E est *construit* par la donnée successive d'expressions de longueur finie, et qu'au moment où l'expression G apparaît dans cette *construction*, elle n'a pas de sens parce que E est en train d'être construit et que la définition de E n'est pas achevée. Je trouve que cette explication est très convaincante et qu'elle nous oblige à réfléchir à ce que veut dire qu'un objet mathématique est défini.

— Le paradoxe de Russell consiste à considérer l'ensemble X de tous les ensembles qui ne se contiennent pas eux-mêmes et à se demander si l'ensemble X est élément de lui-même : si oui, alors non, et sinon, alors oui ! Le rapport que je vois entre les deux paradoxes est que la définition de l'ensemble X est donnée en disant que $x \in \mathrm{X}$ si $x \notin x$, et que pour définir X, il faut aussi décider si on a $x \notin x$ pour $x = \mathrm{X}$ lui-même : dans ce sens, il y a un cercle vicieux.

— Justement, la règle 3 de Poincaré consiste à éviter toute définition d'un objet mathématique dans laquelle cet objet lui-même intervient : en d'autres mots, les termes qui définissent un objet doivent déjà être définis en amont de la définition de l'objet ; dans ce sens, l'objet défini est plus "compliqué" que ces termes définitionnels. La résolution du paradoxe par Richard consiste justement à dénoncer l'imprédicativité de la définition initiale de E : il constate que le paradoxe est levé dès qu'on substitue à la définition initiale une définition constructive.

N. B. : Cependant, la majorité des mathématiciens d'aujourd'hui revendiquent l'usge des définitions imprédicatives : ils ont pris l'habitude d'objets imprédicatifs comme la borne supérieure d'un ensemble borné de nombres réels. Cela implique qu'ils doivent résoudre les paradoxes de Richard et de Russell d'une autre manière qu'en imposant la prédicativité : ils le font en proposant des théories plus formelles dans lesquelles ces paradoxes ne peuvent plus être énoncés.

Corrigé des exercices pour le chapitre 7

7.5

Corrigé de l'exercice 7.5.1.

Quelques commentaires. Le fait qu'on ne sache pas *a priori* déterminer le signe d'un nombre réel connu avec une précision arbitraire empêche qu'on puisse calculer à coup sûr le développement en base 2 usuel de x (par exemple pour déterminer la partie entière de x il faut le comparer à un entier k proche et donc déterminer le signe de $x-k$).

C'est la raison pour laquelle on rajoute un chiffre. Avec les chiffres $-1, 0, 1$ les nombres réels qu'on peut obtenir sous la forme $k_0 + \sum_{n=1}^{\infty} u_n/2^n$ ($k_0 \in \mathbb{Z}$) sont ceux de l'intervalle $[k_0-1, k_0+1]$. Ainsi les intervalles possibles $[k_0-1, k_0+1]$ se chevauchent les uns les autres et cela nous donne un peu de large pour travailler.

En effet tout intervalle rationnel $[a, b]$ de longueur ≤ 1 peut à coup sûr être situé sur un intervalle $[k-1, k+1]$ avec $k \in \mathbb{Z}$. Il suffit de prendre pour k la partie entière de b puisqu'on a $k \leq b < k+1$ et $b - a \leq 1$, et donc $k \leq a+1$.

Ainsi si on demande x avec une précision « suffisante », en tout cas meilleure que $1/2$, on situe x sur un intervalle rationnel $[a_0, b_0]$ de longueur ≤ 1, donc contenu dans un intervalle $[k-1, k+1]$ pour un $k \in \mathbb{Z}$, et on peut prendre cette valeur de k pour $k_0 = x_0$.

Ensuite on double la précision et on peut espérer que le processus fonctionne, c'est-à-dire qu'on aura $[a_1, b_1] \subset [x_1 - \frac{1}{2}, x_1 + \frac{1}{2}]$, avec x_1 pris parmi $x_0 - \frac{1}{2}$, x_0, $x_0 + \frac{1}{2}$.

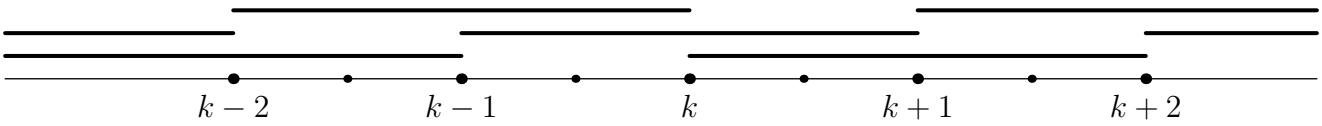


FIGURE 7.5.1 – Intervalles de longueur 2, centrés sur les entiers, couvrant la droite réelle.

Un petit dessin montre qu'on risque d'avoir des problèmes si l'intervalle $[a_1, b_1]$ obtenu lorsqu'on a doublé la précision n'est pas inclus dans l'intervalle $[a_0, b_0]$.

Mais puisque x est sur chacun des deux intervalles, il est aussi sur leur intersection.

Notre procédure va donc fonctionner comme suit.

Preuve proprement dite. À l'étape n on demande x avec la précision $1/2^{n+1}$, on obtient un intervalle $[a_n, b_n]$. L'intersection de tous les $[a_j, b_j]$ pour $j = 0, \ldots, n$ est un intervalle $[c_n, d_n]$ (en fait $c_n = \sup_{i=0}^{n} a_i$ et $d_n = \inf_{i=0}^{n} b_i$). De sorte que les intervalles successifs $[c_n, d_n]$ sont emboités les uns dans les autres et que la longueur de $[c_n, d_n]$ est toujours $\leq 1/2^n$.

Prenons l'étape 1. À l'étape 0 on a trouvé $k_0 \in \mathbb{Z}$ tel que $[c_0, d_0] \subset [k_0-1, k_0+1]$. Cet intervalle est recouvert par les 3 intervalles

$$\mathrm{J}_{-1} = [k_0 - 1, k_0], \mathrm{J}_0 = [k_0 - 1/2, k_0 + 1/2], \mathrm{J}_{+1} = [k_0, k_0 + 1],$$

tous trois de longueur 1 et respectivement centrés en $k_0 - 1/2$, k_0 et $k_0 + 1/2$.

Comme l'intervalle $[c_1, d_1]$ est contenu dans la réunion des J_i et comme il est de longueur $\leq 1/2$ il est forcément contenu dans l'un des J_i, de sorte qu'on a bien réalisé $[c_1, d_1] \subset \left[\dfrac{k_1 - 1}{2}, \dfrac{k_1 + 1}{2}\right]$, avec

$$k_1 = \begin{cases} 2k_0 - 1 & \text{si} \quad [c_1, d_1] \subset \mathrm{J}_{-1} \\ 2k_0 & \text{si} \quad [c_1, d_1] \subset \mathrm{J}_0 \\ 2k_0 + 1 & \text{si} \quad [c_1, d_1] \subset \mathrm{J}_{+1} \end{cases}$$

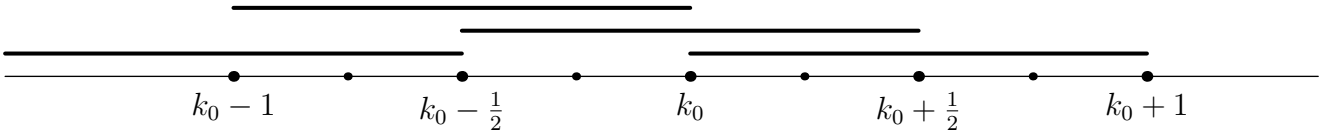


FIGURE 7.5.2 – L'intervalle $[k_0 - 1, k_0 + 1]$ est recouvert par les 3 intervalles de longueur 1 centrés en $k_0 - \frac{1}{2}$, k_0, $k_0 + \frac{1}{2}$. On peut remarquer que le dessin est « le même » au changement d'échelle près, que celui de la figure 7.5.1

Les étapes suivantes fonctionnent de la même manière en divisant toutes les longueurs par 2 à chaque étape. On peut le vérifier précisément en faisant une preuve par récurrence en bonne et due forme. L'hypothèse de récurrence est qu'on a

$$[c_n, d_n] \subset \mathrm{I}_n = \left[\frac{k_n - 1}{2^n}, \frac{k_n + 1}{2^n}\right] = \left[x_n - \frac{1}{2^n}, x_n + \frac{1}{2^n}\right].$$

Cet intervalle I_n est la réunion des 3 intervalles

$$\mathrm{I}_n^{(-1)} = \left[x_n - \frac{1}{2^n}, x_n\right], \mathrm{I}_n^{(0)} = \left[x_n - \frac{1}{2^{n+1}}, x_n + \frac{1}{2^{n+1}}\right], \mathrm{I}_n^{(+1)} = \left[x_n, x_n + \frac{1}{2^n}\right],$$

tous trois de longueur $1/2^n$ et respectivement centrés en $x_n - \dfrac{1}{2^{n+1}}$, x_n et $x_n + \dfrac{1}{2^{n+1}}$.

Comme l'intervalle $[c_{n+1}, d_{n+1}]$ est contenu dans la réunion des $\mathrm{I}_n^{(i)}$ et comme il est de longueur $\leq 1/2^{n+1}$ il est forcément contenu dans l'un des $\mathrm{I}_n^{(i)}$, de sorte qu'on a bien réalisé

$$[c_{n+1}, d_{n+1}] \subset \left[\frac{k_{n+1} - 1}{2^{n+1}}, \frac{k_{n+1} + 1}{2^{n+1}}\right] = \left[x_{n+1} - \frac{1}{2^{n+1}}, x_{n+1} + \frac{1}{2^{n+1}}\right],$$

avec

$$k_{n+1} = \begin{cases} 2k_n - 1 & \text{si } [c_{n+1}, d_{n+1}] \subset \mathrm{I}_n^{(-1)} \\ 2k_n & \text{si } [c_{n+1}, d_{n+1}] \subset \mathrm{I}_n^{(0)} \\ 2k_n + 1 & \text{si } [c_{n+1}, d_{n+1}] \subset \mathrm{I}_n^{(+1)} \end{cases}$$

□

Remarques.

1. La précision demandée de $1/2^{n+1}$ à l'étape n est ce qui est nécessaire pour assurer que l'intervalle obtenu soit à coup sûr sur un intervalle $\left[\dfrac{k-1}{2^n}, \dfrac{k+1}{2^n}\right]$ avec $k \in \mathbb{Z}$, car deux de ces intervalles successifs se chevauchent selon un intervalle de longueur $\dfrac{1}{2^n}$. Mais il n'était pas évident, sans aller voir de près ce qui se passe, que l'on peut bien réaliser $k_{n+1} = 2k_n - 1$ ou $k_{n+1} = 2k_n$ ou $k_{n+1} = 2k_n + 1$.
2. Le fait de devoir recourir à la précision $1/2$ pour calculer sans se tromper un k_0 convenable, qui ne donne x qu'avec la précision 1 est le prix à payer pour le caractère « normalisé » de la représentation en base 2 avec trois chiffres. Ce n'est cependant pas très cher payer en comparaison de la précision infinie nécessaire pour calculer une représentation en base 2 avec deux chiffres.

3. La procédure de calcul met en avant la « vraie structure » du continu. Intuitivement au moins, le continu est mieux décrit comme formé d'intervalles qui se chevauchent que comme constitué de points. Autrement dit, le fait de prendre les nombres réels comme objets de calculs plutôt que comme idéalités abstraites rétablit dans une certaine mesure la structure intuitive du continu, détruite par la révolution de Cantor.
4. La représentation de x que l'on obtient dépend de la suite des intervalles $[c_n, d_n]$, et pas seulement de x. Par exemple même si x est très proche de $x_0 + 1$ il se peut que le centre de l'intervalle $[c_0, d_0]$ soit plus proche de x_0 que de $x_0 + 1$, et par contre que le centre de l'intervalle $[c_1, d_1]$ soit plus proche $x_1 + 1/2 = x_0 + 1$ que de x_1. Les choses seraient plus simples si à chaque étape on pouvait prendre pour x_n l'élément de la forme $k/2^n$ ($k \in \mathbb{Z}$) le plus proche du milieu de l'intervalle $[c_n, d_n]$. Ceci va faire l'objet d'une deuxième solution, qui suit.

Une deuxième solution. En augmentant la précision requise à chaque étape, on va s'arranger pour arriver à prendre pour x_n l'élément de la forme $k/2^n$ ($k \in \mathbb{Z}$) le plus proche du milieu de l'intervalle $[c_n, d_n]$. Ceci n'assure pas pour autant l'unicité du développement obtenu, qui dépend, non pas uniquement de x, mais de la suite des approximations $[c_n, d_n]$.

Pour que le milieu de $[c_1, d_1]$ ne puisse pas être plus proche de $x_0 + 1$ que de $x_0 + 1/2$ il suffit d'avoir obligé $[c_0, d_0]$ à être contenu dans l'intervalle $[x_0 - 3/4, x_0 + 3/4]$. Or deux intervalles successifs $[k - 3/4, k + 3/4]$ avec $k \in \mathbb{Z}$ se chevauchent selon un intervalle de longueur $1/2$.

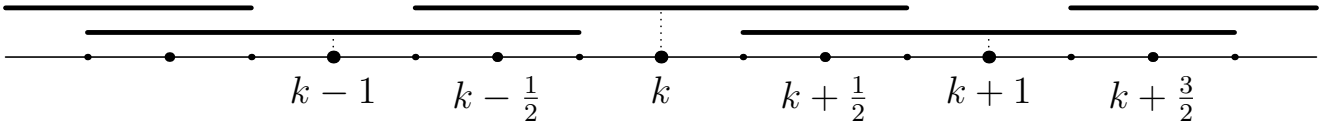


FIGURE 7.5.3 – Intervalles de longueur $3/2$, centrés sur les entiers, couvrant la droite réelle.

On demande donc à l'étape n de connaitre x avec la précision $1/2^{n+2}$, ce qui permet de situer x sur un intervalle $[c_n, d_n]$ de longueur $1/2^{n+1}$. En outre, en considérant l'intersection des intervalles d'approximation déjà donnés, on peut supposer que pour tout $n \geq 1$ on a $[c_n, d_n] \subset [c_{n-1}, d_{n-1}]$. On va noter $e_n = \dfrac{c_n + d_n}{2}$ le milieu de l'intervalle ainsi défini.

À l'étape 0 si k_0 est l'entier le plus proche de e_0, on a donc $[c_0, d_0] \subset \left[k_0 - \frac{3}{4}, k_0 + \frac{3}{4}\right]$ et $|k_0 - e_0| \leq 3/4$ (si e_0 est un demi entier, n'importe lequel des deux entiers les plus proches peut faire l'affaire).

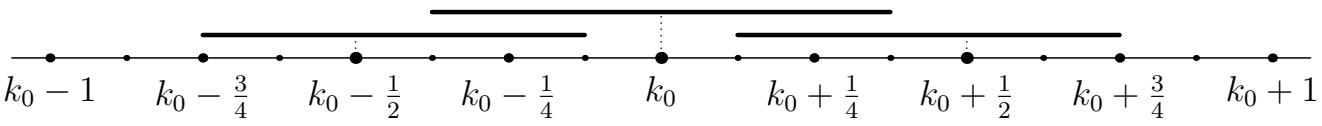


FIGURE 7.5.4 – Recouvrement de $\left[k_0 - \frac{3}{4}, k_0 + \frac{3}{4}\right]$ par 3 intervalles de longueur $\leq \frac{3}{4}$. On peut remarquer que les deux dessins sont presque « les mêmes » au changement d'échelle près.

L'intervalle $\left[k_0 - \frac{3}{4}, k_0 + \frac{3}{4}\right]$ est recouvert par les 3 intervalles $\left[k_0 - \frac{3}{4}, k_0 - \frac{1}{8}\right]$, $\left[k_0 - \frac{3}{8}, k_0 + \frac{3}{8}\right]$, $\left[k_0 + \frac{1}{8}, k_0 + \frac{3}{4}\right]$, (respectivement contenus dans $\left[k_0 - \frac{1}{2} - \frac{3}{8}, k_0 - \frac{1}{2} + \frac{3}{8}\right]$, $\left[k_0 - \frac{3}{8}, k_0 + \frac{3}{8}\right]$, et $\left[k_0 + \frac{1}{2} - \frac{3}{8}, k_0 + \frac{1}{2} + \frac{3}{8}\right]$). Deux de ces intervalles successifs se chevauchent suivant un intervalle de longueur $1/4$ et donc $[c_1, d_1]$ est contenu dans l'un de ces trois intervalles. En outre l'entier le plus proche de $2e_1$ est l'un des trois entiers $2k_0 - 1$, $2k_0$ ou $2k_0 + 1$.

Il est clair que le passage de l'étape n à l'étape $n + 1$ sera la même chose que le passage de l'étape 0 à l'étape 1 au changement d'échelle près.

Néanmoins, si on n'est pas convaincu, et si on n'aime pas les dessins on peut faire un raisonnement par récurrence purement calculatoire. Supposons par hypothèse de récurrence qu'à

l'étape n on a obtenu $[c_n, d_n] \subset \left[x_n - \frac{3}{2^{n+2}}, x_n + \frac{3}{2^{n+2}}\right]$ et $|x_n - e_n| \leq 1/2^{n+1}$ (i. e. x_n est le nombre de la forme $k/2^n$ le plus proche de e_n), autrement dit $2^n x_n = k_n$ est l'entier le plus proche de $2^n e_n$. Puisque $[c_{n+1}, d_{n+1}] \subset [c_n, d_n]$, et que ce dernier intervalle est de longueur $\leq 1/2^{n+1}$, on a $|e_n - e_{n+1}| \leq 1/2^{n+2}$ donc $|2^n e_n - 2^n e_{n+1}| \leq 1/4$ et puisque $|2^n e_n - k_n| \leq 1/2$ cela donne $|2^{n+1} e_{n+1} - 2k_n| \leq 3/2$. Ceci implique que l'entier le plus proche de $2^{n+1} e_{n+1}$ est l'un des trois entiers $2k_n - 1$, $2k_n$ ou $2k_n + 1$. □

Une troisième solution. On peut raisonner « sans considérer l'intersection des intervalles d'approximation successifs » comme suit, mais la solution obtenue est un peu moins optimale. Un avantage est qu'on fait un calcul direct sans utiliser de raisonnement par récurrence. Mais ceci se paye : la précision requise à l'étape n est plus grande.

On pose *a priori* qu'à l'étape n on demande un $v_n \in \mathbb{D}_2$ tel que $|x - v_n| \leq \alpha/2^n$ où α reste à déterminer. On prend alors pour k_n l'entier le plus proche de $2^n v_n$ (si $2^n v_n$ est un demi-entier, n'importe lequel des deux entiers les plus proches). Et on se pose la question : est-ce qu'une valeur convenable de α garantit que pour tout n on aura k_{n+1} égal à l'un des trois entiers $2k_n - 1$, $2k_n$ ou $2k_n + 1$?

Le calcul est le suivant. On a par hypothèse

$$|2^n v_n - k_n| \leq 1/2, \quad |2^{n+1} v_{n+1} - k_{n+1}| \leq 1/2, \quad |x - v_n| \leq \alpha/2^n \quad \text{et} \quad |x - v_{n+1}| \leq \alpha/2^{n+1}.$$

Donc $|v_n - v_{n+1}| \leq 3\alpha/2^{n+1}$ et

$$|2k_n - k_{n+1}| \leq 3\alpha + 3/2.$$

Or ce qu'on veut c'est $|2k_n - k_{n+1}| < 2$, car cela permet de conclure, puisque ce sont des entiers, que $|2k_n - k_{n+1}| \leq 1$. On prendra donc $\alpha < 1/6$ (par exemple $\alpha = 1/8$, de sorte qu'on demandera que $|x - v_n| \leq 1/2^{n+3}$, ou alors on demandera que $|x - v_n| < 1/(6 \times 2^n)$). □

Corrigé des exercices pour le chapitre 9

9.1

Corrigé de l'exercice 9.1.2. On part de la définition suivante donnée en théorie des ensembles pour le mot « infini ». Un ensemble X est dit infini s'il a même cardinal qu'une partie Y de X, distincte de X. On considère alors la bijection $f\colon \mathrm{X} \to \mathrm{Y}$ (qui traduit le fait que Y et X ont même cardinal) et un élément $a \in \mathrm{X}$ qui n'est pas dans Y.
À partir de cela on veut construire une bijection h entre $\mathbb{N}$ et une partie N de X. On procède par récurrence. On définit $h(0) = a$, et en supposant que $h(0), \ldots, h(n)$ ont déjà été construits, on définit $h(n+1)$ en posant $h(n+1) = f(h(n))$. Intuitivement au moins, ceci définit bien une fonction h de $\mathbb{N}$ vers X. On note N l'ensemble des valeurs prises par h. Il reste à voir que $h : \mathbb{N} \to \mathrm{N}$ est injective.

On montre par récurrence sur n la propriété suivante :

« Pour tout $m > n$, $h(m) \neq h(n)$ ».

Initialisation : $n = 0$. Soit $m > 0$, on pose $m = p + 1$, on a $h(m) = f(h(p)) \in \mathrm{Y}$ mais $h(0) = a \notin \mathrm{Y}$ donc $h(m) \neq h(0)$.

Passage de n à $n+1$. On suppose la propriété vraie pour n. Si $m > n+1$, on pose $m = p+1$. Alors $p > n$, $h(n+1) = f(h(n))$ et $h(p+1) = f(h(p))$. Par hypothèse de récurrence $h(p) \neq h(n)$. Puisque f est injective, cela implique $f(h(p)) \neq f(h(n))$, c'est-à-dire $h(p+1) \neq h(n+1)$.

Corrigé de l'exercice 9.1.3. Il suffit de montrer qu'une suite de réels ne peut pas recouvrir l'intervalle $[0, 1[$.

Si $(x_n)_{n\geq 1}$ est une suite de réels dans $[0, 1[$, on écrit chaque x_n sous forme de son développement en base 3 standard (i. e., on refuse les développements qui se terminent par 222. . .) :

$$x_n = \sum_{m=1}^{\infty} y_{n,m}/3^m.$$

Définissons $\mu\colon \{0,1,2\} \to \{0,1,2\}$ par $\mu(0) = 1$, $\mu(1) = 0$, $\mu(2) = 1$, et posons

$$z = \sum_{m=1}^{\infty} \mu(y_{m,m})/3^m.$$

Alors le développement de z est un développement standard, différent de celui de tous les x_n, donc $z \neq x_n$.

Supposons par exemple qu'on ait, en base 3 et en écriture standard :

$$\begin{array}{lcl}
x_0 & = & 0\,,\ \underline{1}\ 2\ 0\ 1\ 1\ 2\ 1\ 0\ 2\ 1\ 0\ 2\ 0\ 1\ 1\ 0\ 0\ 0\ 2\ 1\ \ldots\ldots \\
x_1 & = & 0\,,\ 0\ \underline{2}\ 1\ 0\ 2\ 0\ 1\ 1\ 0\ 0\ 1\ 2\ 0\ 1\ 1\ 2\ 1\ 0\ 2\ 1\ \ldots\ldots \\
x_2 & = & 0\,,\ 0\ 0\ \underline{0}\ 2\ 1\ 1\ 2\ 0\ 1\ 1\ 2\ 1\ 0\ 2\ 1\ 0\ 2\ 0\ 1\ 1\ \ldots\ldots \\
x_3 & = & 0\,,\ 0\ 1\ 2\ \underline{0}\ 1\ 1\ 2\ 0\ 0\ 2\ 1\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 2\ 1\ \ldots\ldots \\
x_4 & = & 0\,,\ 0\ 0\ 2\ 1\ \underline{0}\ 1\ 1\ 0\ 0\ 0\ 2\ 1\ 0\ 2\ 1\ 1\ 2\ 0\ 1\ 1\ \ldots\ldots \\
x_5 & = & 0\,,\ 1\ 2\ 0\ 1\ 1\ \underline{2}\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 2\ 1\ 0\ 2\ 1\ \ldots\ldots \\
\vdots & \vdots & \qquad\qquad\qquad\vdots
\end{array}$$

On lit les chiffres soulignés sur la diagonale : `0 , 1 2 0 0 0 2 . . . .`

L'élément z qui apparait dans la preuve est : `0 , 0 1 1 1 1 1 . . . .`

Par construction le n^{e} chiffre de z diffère du n^{e} chiffre de x_n. Or l'écriture de z est standard (elle ne se termine pas par une suite infinie de 2) puisqu'elle ne contient pas de 2. Et l'écriture de x_n est standard par hypothèse. Deux nombres écrits en écriture standard ne peuvent être égaux que s'ils ont tous leurs chiffres égaux. Ainsi z est bien distinct de tous les x_n.

Corrigé des exercices pour le chapitre 10

10.3

Corrigé de l'exercice 10.3.2. 1. *Les définitions euclidiennes.* Les définitions de *égal* et de *rapport* données par Euclide sont implicites au sens où elles n'expriment pas ce que c'est qu'égal et ce que c'est qu'un rapport. Pour l'*angle dièdre*, les choses sont moins claires : la deuxième des deux définitions propose d'approcher l'inclinaison de deux plans de manière implicite, par l'égalité, alors que la première fournit une définition de nom de cette inclinaison !

Il y a une raison à cela : pour Euclide, chacun de ces trois mots renvoie à un certain emploi dans la langue et non à un objet matériel.

- Le mot *égal* renvoie à une relation entre objets et pas à un objet en lui-même.
- Alors que nous nous permettons aujourd'hui de penser que l'*angle dièdre* de deux plans est un des quatre espaces délimités par ces deux plans, cela n'est pas pensable pour Euclide parce que ces espaces sont infinis, c'est-à-dire illimités (le même problème se pose pour l'angle de deux droites).
- Alors que pour nous, un rapport renvoie à un réel (par exemple le rapport de deux côtés d'un carré renvoie au réel 1 et le rapport de la circonférence d'un cercle à son diamètre renvoie au réel π), ce n'est pas le cas pour Euclide.

Cependant, ces trois définitions procèdent de manière très différente.

- La définition 6 du onzième livre est un définition de nom en bonne et due forme, mais le *definiens*, « l'angle aigu contenu par les [droites] » n'est pas un objet : nous avons vu que l'angle est défini comme une inclinaison, c'est-à-dire comme une relation entre lesdites droites. C'est cela que rappelle en fait la définition 7 suivante.
- La définition 5 du cinquième livre cherche à généraliser la notion de rapport numérique (d'un entier positif n à un entier positif m) au rapport de deux grandeurs quelconques. La définition exprime qu'on connait un tel rapport de grandeurs si on sait le comparer à tous les rapports numériques : la propriété que $m \times \mathrm{A}$ et $m \times \mathrm{C}$ « ou simultanément dépassent, ou sont simultanément égaux ou simultanément inférieurs à » $n \times \mathrm{B}$ et $m \times \mathrm{D}$ pour n'importe quels m et n est la seule manière qu'on a d'exprimer que le rapport de A à B et le rapport de C à D « ou simultanément dépassent, ou sont simultanément égaux ou simultanément inférieurs au » rapport numérique de n à m : si A, B, C et D étaient des réels positifs, cela s'écrirait aujourd'hui

$$\left(\frac{\mathrm{A}}{\mathrm{B}} > \frac{n}{m} \text{ et } \frac{\mathrm{C}}{\mathrm{D}} > \frac{n}{m}\right) \text{ ou } \left(\frac{\mathrm{A}}{\mathrm{B}} = \frac{n}{m} \text{ et } \frac{\mathrm{C}}{\mathrm{D}} = \frac{n}{m}\right) \text{ ou } \left(\frac{\mathrm{A}}{\mathrm{B}} < \frac{n}{m} \text{ et } \frac{\mathrm{C}}{\mathrm{D}} < \frac{n}{m}\right).$$

- En fait, Euclide ne prétend pas définir *égal* : il ne propose pas une définition, mais huit « notions communes ». Elles postulent qu'une égalité ou une inégalité a lieu sous

certaines hypothèses. Il s'agit en fait de règles, et si on utilise les figures de Gentzen pour écrire ces règles, on obtient

$$\frac{\mathrm{A}=\mathrm{B}\ \ \mathrm{C}=\mathrm{B}}{\mathrm{A}=\mathrm{C}}\ 1\quad \frac{\mathrm{A}=\mathrm{B}}{\mathrm{A}+\mathrm{C}=\mathrm{B}+\mathrm{C}}\ 2\quad \frac{\text{A s'ajuste sur B}}{\mathrm{A}=\mathrm{B}}\ 7\quad \frac{\text{B partie de A}}{\mathrm{A}>\mathrm{B}}\ 8\,.$$

Dans les premières règles, les hypothèses sont elles-mêmes des égalités, alors que seule la règle 7 permet de déduire une égalité d'une assertion géométrique, et seule la règle 8 permet de déduire une inégalité d'une assertion géométrique : cela assure à ces deux règles un rôle important dans les démonstrations.

Pour conclure, notons que les règles qui définissent l'égalité diffèrent des deux autres définitions en ce que certaines d'entre elles (les notions communes 1 et 2) ont un aspect circulaire : une égalité s'y déduit d'une égalité ; or la circularité est considérée comme un écueil à éviter absolument en matière de définitions. La définition du rapport et de l'angle dièdre sont proches dans leur démarche, mais la première est nettement plus compliquée parce qu'on y considère « n'importe quelle multiplication », c'est-à-dire une infinité de cas.

2. *Commentaire de l'« Essai sur la théorie des définitions » de Gergonne.* Dans son *Essai*, Gergonne tente de rendre compte d'un usage très courant de la définition hors des sciences : le contexte donne un sens à nos assertions sans que nous ayons besoin d'introduire tous les termes de ces assertions (et parfois il nous semblerait difficile de le faire). Mais en mathématiques, un tel usage est malvenu, et pour reprendre l'exemple de Gergonne, on demandera à celui qui dit que « chacune des deux diagonales d'un quadrilatère le divise en deux triangles » de définir proprement la diagonale, c'est-à-dire de résoudre l'« équation non résolue ».

   Euclide a fait ce travail pour les notions de rapport et d'angle dièdre, et les explications de Gergonne ne s'appliquent pas à leur définition.

   Mais pour ce qui est de *égal*, on peut parfaitement concevoir les notions communes 1, 2, 7 et 8 comme des équations non résolues qui ont pour finalité de décrire exhaustivement les règles selon lesquelles des égalités et des inégalités seront établies dans les *Éléments*.

10.6

Corrigé de l'exercice 10.6.2. D'Alembert considère que la définition de la ligne droite reste un défi pour les mathématiciens. Il passe en revue plusieurs tentatives et les critique. Il conclut qu'elles renvoient à « une notion primitive de la ligne droite que nous avons dans l'esprit », que nous restons « sans pouvoir en quelque façon [...] rendre par des expressions ». Par là, il veut dire que les définitions proposées jusqu'à ce jour ont été formulées suivant une intuition qui reste à dévoiler, et qu'elles ne constituent que des corolaires ou des propriétés de cette intuition.

On retrouve dans les tentatives passées en revue les définitions présentées dans le commentaire de Clavius de la définition de la ligne droite dans les *Éléments* d'Euclide (section 6.6.2 du cours). Pour chacune d'elles, D'Alembert a des réflexions intéressantes à proposer.

1. La première est la définition optique qu'on trouve déjà chez Platon. Elle est contestable parce qu'elle fait dériver le droit du phénomène de la lumière ; or les aveugles savent ce que c'est sans voir (par exemple, ils savent marcher droit devant eux) ; de plus, comment juger que la lumière avance tout droit si on n'a pas d'idée préalable du droit ?
2. La deuxième définition, en termes de direction, est celle qui est la plus proche de la définition « mécanique » qu'on trouve déjà chez Geminos. Elle est contestable parce qu'elle donne lieu à un cercle vicieux : en gros, on définit la ligne droite comme celle qui va tout droit ! Cela n'avance en rien.

3. La troisième définition, qu'il considère être la définition « ordinaire », est la définition « métrique » qu'on trouve déjà chez Archimède. Elle est contestable pour au moins deux raisons : il se pourrait qu'il y ait plusieurs lignes de plus court chemin entre deux points, comme c'est le cas pour les méridiens de la Terre qui relient tous le pôle Nord et le pôle Sud ; d'un autre côté, il se pourrait que la ligne de plus court chemin entre deux points donnés soit unique, mais que lorsqu'on cherche à la prolonger en une ligne toujours de plus court chemin, on se retrouve face à une bifurcation : c'est le cas dans les déplacements sur un réseau routier lorsqu'on arrive à un carrefour où on peut aller tout droit mais aussi à gauche ou à droite.

En ce qui me concerne, ces critiques de D'Alembert sont de nature à me convaincre qu'avec la ligne droite, nous sommes face à un de ces concepts premiers qui n'admettent pas de définition, et pour lesquels toute tentative de définition n'a pour effet que d'obscurcir leur « extrême évidence », comme le dit Pascal.

Je me permets d'ajouter trois remarques à ce commentaire.

— Les réflexions de D'Alembert sur la définition « optique » font penser à la théorie de la relativité générale. selon laquelle la lumière est déviée par les masses. Comment comprendre cette déviation autrement que par rapport à une idée préalable du droit ?

— Il se peut que la définition de la ligne droite comme celle de plus court chemin soit celle qui pour D'Alembert est ordinaire parce qu'elle attribue une finalité à la ligne droite, celle de relier deux points donnés par « le » plus court chemin ; dans ce sens, elle résulte d'un principe de moindre action formulé par Maupertuis en 1744 et très discuté depuis lors.

— Les critiques de D'Alembert par rapport à cette dernière définition résonnent avec une préoccupation naissante à son époque dans le domaine des équations différentielles : celle de l'unicité de leur solution.

Corrigé des exercices pour le chapitre 11

11.3

Corrigé de l'exercice 11.3.3. Voici un tableau de toutes les introductions et éliminations des opérations logiques dans la preuve rédigée par Gentzen.

Une **introduction de** $\forall$ a lieu ligne 10, ligne 44 et aussi à la fin de la preuve, ligne 62 : après avoir démontré que pour un nombre quelconque a on a $\exists z\,(\mathrm{Prem}(z) \wedge z > a)$, on en a conclu $\forall y\,\exists z\,(\mathrm{Prem}(z) \wedge z > y)$.

Une **élimination de** $\forall$ a lieu ligne 40, puis ligne 54 : de $\forall y\,[(y > 1 \wedge y \leq a! + 1) \rightarrow \neg\, y|(a! + 1)]$ on a conclu $(a! + 1 > 1 \wedge a! + 1 \leq a! + 1) \rightarrow \neg\, a! + 1|(a! + 1)$.

Une **introduction de** $\wedge$ a lieu lignes 33, 34 et aussi ligne 40 : les deux formules $d > 1$ et $d \leq n$ ont donné ensemble $d > 1 \wedge d \leq n$.

Une **élimination de** $\wedge$ a lieu ligne 60 : de $l \leq a \wedge (\mathrm{Prem}(l) \wedge l > a)$ on a déduit $\mathrm{Prem}(l) \wedge l > a$.

Une **introduction de** $\exists$ a lieu lignes 35 et 61 : de $\mathrm{Prem}(l) \wedge l > a$ on a déduit $\exists z\,(\mathrm{Prem}(z) \wedge z > a)$.

Une **élimination de** $\exists$ a lieu lignes 59-60 : en partant de la formule $\exists z\,[z \leq a! + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]$ nous avons conclu $l \leq a! + 1 \wedge (\mathrm{Prem}(l) \wedge l > a)$, où l est censé signifier un quelconque des nombres qui existent selon la formule.

Une **introduction de** $\vee$ a lieu lignes 12, 22, 37 et aussi 47 : de $\forall y\,[(y > 1 \wedge y \leq n + 1) \rightarrow \neg\, y|(a! + 1)]$ nous avons conclu $\exists z\,[z \leq n + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n + 1) \rightarrow \neg\, y|(a! + 1)]$.

Une **élimination de** $\vee$ commence ligne 16 : en partant de la formule $\{\exists z\,[z \leq n \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq n) \rightarrow \neg\, y|(a! + 1)]$ nous avons fait une distinction de cas : 1. $\exists z\,[z \leq n \wedge (\mathrm{Prem}(z) \wedge z > a)]\}$; 2. $\forall y\,[(y > 1 \wedge y \leq n) \rightarrow \neg\, y|(a! + 1)]$. Cette distinction de cas a été terminée en aboutissant finalement dans les deux cas à la même formule, c.-à-d. seulement ligne 47. Deux autres ont lieu lignes 25-47 et lignes 39-43.

Une **introduction de** $\rightarrow$ a lieu ligne 44 : en partant de l'hypothèse $d > 1 \wedge d \leq n + 1$, nous avons abouti au résultat $\neg\, d|(a! + 1)$. Donc la formule $(d > 1 \wedge d \leq n + 1) \rightarrow \neg\, d|(a! + 1)$ est valide.

Une **élimination de** $\rightarrow$ a lieu ligne 41 : de $d > 1 \wedge d \leq n$ et $(d > 1 \wedge d \leq n) \rightarrow \neg\, d|(a! + 1)$ nous avons déduit $\neg\, d|(a! + 1)$; puis ligne 55.

11.4

Corrigé de l'exercice 11.4.3. 1. Pour montrer que l'hypothèse $\mathrm{X} \vee (\mathrm{Y} \wedge (\mathrm{Z} \wedge \mathrm{W}))$ implique la conclusion $(\mathrm{X} \vee \mathrm{Y}) \wedge ((\mathrm{X} \vee \mathrm{Z}) \wedge (\mathrm{X} \vee \mathrm{W}))$, supposons que l'hypothèse soit valide, c'est-à-dire que $\mathrm{X} \vee (\mathrm{Y} \wedge (\mathrm{Z} \wedge \mathrm{W}))$. Comme cette hypothèse est de la forme "A ou B" avec A l'assertion X et B

l'assertion $Y \wedge (Z \wedge W)$, il faudra que je montre que la conclusion est valide dans le cas où A est valide ainsi que dans le cas ou B est valide.

Cas A. Si X est valide, alors $X \vee Z$ et $X \vee W$ aussi, et donc $(X \vee Z) \wedge (X \vee W)$. L'assertion $X \vee Y$ est valide aussi, et donc aussi $(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))$.

Cas B. Si $Y \wedge (Z \wedge W)$ est valide, alors Y et aussi Z et W sont valides. Donc $X \vee Y$ et aussi $X \vee Z$ et $X \vee W$ sont valides. Donc $(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))$ est valide.

J'ai obtenu la conclusion dans chacun des cas, ce qu'il fallait démontrer.

Dans la figure complète de la déduction, j'ai utilisé trois couleurs. Le rouge correspond à la partie de la déduction qui introduit le "implique", que j'ai affublée de l'étiquette "2".

$$\dfrac{\begin{array}{c}2\\ [X \vee (Y \wedge (Z \wedge W))]\\ \vdots\\ (X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))\end{array}}{(X \vee (Y \wedge (Z \wedge W))) \to ((X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W)))}\ {\to_i}2$$

Le bleu correspond à la partie de la déduction qui élimine le "ou", que j'ai affublée de l'étiquette "1".

$$\dfrac{X \vee (Y \wedge (Z \wedge W)) \quad \begin{array}{c}1\\ [X]\\ \vdots\\ (X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))\end{array} \quad \begin{array}{c}1\\ [Y \wedge (Z \wedge W)]\\ \vdots\\ (X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))\end{array}}{(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))}\ {\vee_e}1$$

Le noir correspond aux déductions intermédiaires faites dans chacun des cas A et B.

$$\dfrac{\begin{array}{c}2\\ [X \vee (Y \wedge (Z \wedge W))]\end{array} \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}1\\ [X]\end{array}}{X \vee Y}{\vee_i} \quad \dfrac{\dfrac{\begin{array}{c}1\\ [X]\end{array}}{X \vee Z}{\vee_i} \quad \dfrac{\begin{array}{c}1\\ [X]\end{array}}{X \vee W}{\vee_i}}{(X \vee Z) \wedge (X \vee W)}{\wedge_i}}{(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))}{\wedge_i} \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}1\\ [Y \wedge (Z \wedge W)]\end{array}}{Y}{\wedge_e}}{X \vee Y}{\vee_i} \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}1\\ [Y \wedge (Z \wedge W)]\end{array}}{Z \wedge W}{\wedge_e}}{Z}{\wedge_e}}{X \vee Z}{\vee_i} \quad \dfrac{\dfrac{\dfrac{\begin{array}{c}1\\ [Y \wedge (Z \wedge W)]\end{array}}{Z \wedge W}{\wedge_e}}{W}{\wedge_e}}{X \vee W}{\vee_i}}{(X \vee Z) \wedge (X \vee W)}{\wedge_i}}{(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))}{\wedge_i}}{(X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W))}\ {\vee_e}1}{(X \vee (Y \wedge (Z \wedge W))) \to ((X \vee Y) \wedge ((X \vee Z) \wedge (X \vee W)))}\ {\to_i}2$$

Corrigé de l'exercice 11.4.6. 1. On constate un ensemble de règles de déductions différentes qui ne se distinguent pas aisément en introductions et éliminations de $\neg$. Voici les quatre exemples issus de la démonstration euclidienne.

(a) La constatation d'une contradiction ligne 56 à partir de $(a! + 1)|(a! + 1)$ et $\neg\,(a! + 1)|(a! + 1)$.

(b) Une déduction "par contradiction" ligne 58 : $\neg\,\forall y\,[(y > 1 \wedge y \leq a) \to \neg\, y|(a! + 1)]$ a été déduit de ce que l'hypothèse $\forall y\,[(y > 1 \wedge y \leq a) \to \neg\, y|(a! + 1)]$ a abouti à une contradiction.

(c) Ligne 52, nous avons constaté la disjonction $\{\exists z\,[z \leq a! + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]\} \vee \forall y\,[(y > 1 \wedge y \leq a! + 1) \to \neg\, y|(a! + 1)]$, puis ligne 58, la négation de la deuxième formule de cette disjonction, $\neg\,\forall y\,[(y > 1 \wedge y \leq a! + 1) \to \neg\, y|(a! + 1)]$, et nous en avons déduit que la première formule, $\exists z\,[z \leq a! + 1 \wedge (\mathrm{Prem}(z) \wedge z > a)]$, est valide.

(d) Ligne 9, de la validité de $\neg\,(c > 1 \wedge c \leq 1)$ nous avons déduit $(c > 1 \wedge c \leq 1) \to \neg\, c|(a! + 1)$.

(e) Ligne 25, nous avons déduit $(n+1)|(a!+1) \vee \neg\,(n+1)|(a!+1)$ sans aucune hypothèse.

Le premier exemple d'usage de la négation dans la preuve euclidienne est une application de la règle d'élimination de la négation. Le deuxième exemple est une application de la règle d'introduction de la négation. Le cinquième exemple est une application de la règle du tiers exclu.

2. La règle de l'absurdité intuitionniste est à la base du quatrième exemple d'usage de la négation dans la preuve euclidienne : de l'assertion $\neg$A se déduit A $\to$ B.

$$\dfrac{\dfrac{[\mathrm{A}]\ \neg\mathrm{A}}{\dfrac{\bot}{\mathrm{B}}}}{\mathrm{A} \to \mathrm{B}}.$$

Cette règle explique aussi notre troisième exemple d'usage de la négation : des assertions A $\vee$ B et $\neg$A se déduit B.

$$\dfrac{\mathrm{A} \vee \mathrm{B} \qquad \dfrac{[\mathrm{A}]\ \neg\mathrm{A}}{\dfrac{\bot}{\mathrm{B}}} \qquad \dfrac{[\mathrm{B}]}{\mathrm{B}}}{\mathrm{B}}\ .$$

Corrigé des exercices pour le chapitre 12

12.1

Corrigé de l'exercice 12.1.2. On reconnait un palindrome en recopiant le mot à l'envers et en comparant une à une les lettres du mot d'origine et du mot à l'envers. Il est suffisant de mener cette comparaison sur la première moitié des deux mots, mais je ne vais pas chercher à l'exploiter parce que cela compliquerait beaucoup le programme sans gain véritable.

Pour réaliser une machine de Turing qui effectue cette tâche, on suppose que le mot est écrit sur la bande d'entrée `T0`, on rajoute une bande de travail `T1` sur laquelle la machine recopie le mot à l'envers dans un état initial `S0`, puis la machine se met dans un état `S1` dans lequel la tête de lecture de `T0` se replace sur la dernière lettre du mot d'origine, puis la machine se met dans un état `S2` dans lequel elle compare les mots écrits sur `T0` et sur `T1` lettre par lettre : dès qu'elle constate une différence, elle écrit `a` sur la bande de sortie `T2` et termine ; si elle arrive au début des deux mots sans avoir constaté de différence, elle écrit `y` sur `T2` et termine.

Pour décrire précisément ce programme, on va convenir que le début de la bande `T0` est marqué par le symbole `$`, que le mot à étudier y est écrit immédiatement après avec les lettres `a`, `k` et `y` et suivi du symbole `¤`, et que la tête de lecture est initialement placée sur la dernière lettre de ce mot. On va aussi veiller à ce que la tête de lecture de la bande `T1` sera aussi sur la dernière lettre du mot recopié lorsque la machine rentrera dans l'état `S2`.

Voici maintenant une description précise des instructions exécutées en fonction de l'état de la machine et des cases lues sur certaines bandes. Une action sur une bande est écrite `(I,g)` et `(I,d)` si la bande est laissée intacte et que sa tête de lecture se déplace respectivement vers la gauche et vers la droite, et elle est écrite `(`l`,d)` si elle écrit une lettre l parmi `a`, `k` et `y` et que sa tête de lecture se déplace vers la droite. La description a trois colonnes : les bandes concernées par la lecture, le nouvel état de la machine suite à cette lecture, les bandes concernées par les actions suite à cette lecture. La première et la troisième colonne ont autant de sous-colonnes que de bandes concernées.

Programme de l'état `S0`

lecture	état	actions	
`T0`		`T0`	`T1`
l	`S0`	`(I,g)`	`(`l`,d)`
`$`	`S1`	`(I,d)`	`(I,g)`

Programme de l'état `S1`

lecture	état	actions
`T0`		`T0`
l	`S1`	`(I,d)`
`¤`	`S2`	`(I,g)`

Dans l'écriture de ces deux programmes, on utilise la variable l pour ne pas avoir à écrire les trois lignes correspondant aux trois valeurs possibles `a`, `k` et `y` de l.

Programme de l'état S2

lecture			état	actions		
T0	T1			T0	T1	T2
l	m	avec $l \neq m$	Fin			(a,d)
l	m	avec $l = m$	S2	(I,g)	(I,g)	
$			Fin			(y,d)

Dans l'écriture de ce programme, on utilise les variables l et m pour ne pas avoir à écrire les six lignes correspondant aux six valeurs possibles ak, ay, ka, ky, ya et yk de $l \neq m$, ni les trois lignes correspondant aux trois valeurs possibles aa, kk et yy de $l = m$. La case de la colonne de lecture de T1 dans la dernière ligne est vide pour exprimer que ce que la tête de lecture de T1 lit n'a pas d'incidence ; j'aurais pu convenir que le début de la bande T1 est aussi marqué par $ et mettre ce symbole dans cette case.

Corrigé de l'exercice 12.1.3. Pour construire une machine de Turing qui calcule le produit de deux entiers naturels écrits en base 2, je vais m'inspirer de l'exemple de la machine de Turing qui calcule leur somme vu en cours. Commençons par décrire ce que l'on fait quand on multiplie deux entiers en base 2, par exemple 11 (c'est-à-dire 3 en base 10) et 1101 (c'est-à-dire 13 en base 10) :

$$\begin{array}{r} 11 \\ 1101 \\ \hline 11 \\ 11\cdot\cdot \\ {\scriptstyle 1\,1}11\cdot\cdot\cdot \\ \hline 100111 \end{array}$$

En fait, on calcule la somme de décalés à gauche du premier nombre, 11, un décalé par chiffre 1 du deuxième nombre, 1101, et on obtient 100111 (c'est-à-dire 39 en base 10).

Dans cet exemple, j'ai d'abord écrit ces décalés puis effectué leur addition et rencontré le problème des retenues. Pour la machine de Turing, on ne va pas employer autant de bandes qu'il y a de 1 dans le deuxième nombre, mais calculer successivement les sommes de ces décalés, en commençant avec 0 : on voudra donc calculer successivement $0 + 11 = 11$, $11 + 11\cdot\cdot = 1111$, $1111 + 11\cdot\cdot\cdot = 100111$.

Pour cette machine, je vais utiliser deux bandes pour les entrées, E1 et E2, et une bande de travail T1 sur laquelle j'écrirai les sommes successives, mais à l'envers parce que je ne sais pas *a priori* combien le produit aura de chiffres. À la fin, cette bande contiendra le produit des entrées ; il sera écrit à l'envers et je m'en contenterai, mais il serait facile de programmer, comme dans le cours, une deuxième phase avec un état DE qui dépilerait T1 et l'empilerait à l'endroit sur une bande de sortie S1.

J'utiliserai quatre symboles sur ces bandes : le symbole $ pour le début de la bande, le symbole $\emptyset$ pour une case vide et les symboles 0 et 1. Je demanderai que les deux nombres à multiplier soient écrits directement à la suite du symbole $ sur les bandes E1 et E2, de gauche à droite, et que les têtes de lecture soient sur la dernière case écrite de la bande, qui correspond au chiffre des unités. La bande T1 ne contient que le symbole $, et la tête de lecture est sur la case à droite de celle du $.

Par rapport à la machine de Turing pour l'addition, j'introduis deux états de plus. Le premier, MU, qui sera aussi l'état de la machine au démarrage, est l'état où la multiplication du premier nombre par un chiffre donné du deuxième nombre est entreprise :

— si ce chiffre est 1, la machine se met dans l'état R0 pour lancer l'addition du premier nombre au nombre écrit sur la bande de travail T1 ;

— si ce chiffre est 0, la machine décale juste la tête de lecture sur T1 vers la droite ;

— si ce n'est pas un chiffre, c'est-à-dire si la machine lit `$`, alors la multiplication est terminée.

Le deuxième état de plus est un état `RB` de rembobinage qui, une fois l'addition effectuée, remet la tête de lecture de `E1` sur le dernier chiffre et déplace également la tête de lecture de `T1` de telle sorte qu'elle se trouve exactement une case plus à droite qu'avant le début de l'addition.

Je vais maintenant décrire chacun des états par des tableaux à lire de gauche à droite : j'indique d'abord les cases lues par la machine sur certaines bandes ; puis j'indique l'action de la machine sur certaines bandes ; dans la dernière colonne, j'indique le nouvel état de la machine.

— `MU` lit les bandes `E2` et `T1`, déplace leur tête de lecture à gauche et à droite respectivement, et écrit sur `T1` uniquement pour remplacer, le cas échéant, une case vide par un `0`.

Programme de l'état `MU`

lecture		actions		état
`E2`	`T1`	`E2`	`T1`	
`0`	∅	à gauche	écrire `0` puis à droite	`MU`
`0`	`0` ou `1`	à gauche	à droite	`MU`
`1`	∅ ou `0` ou `1`	à gauche		`R0`
`$`	`0` ou `1`		à droite	`MU`
`$`	∅		à gauche	`Fin`

— `R0` et `R1` lisent les bandes `E1` et `T1`, écrivent la somme des chiffres lus sans et avec retenue respectivement sur `T1`, déplacent leur tête de lecture à gauche et à droite respectivement.

Programme de l'état `R0`

lecture		actions		état
`E1`	`T1`	`E1`	`T1`	
`0`	∅ ou `0`	à gauche	écrire `0` puis à droite	`R0`
`0`	`1`	à gauche	écrire `1` puis à droite	`R0`
`1`	∅ ou `0`	à gauche	écrire `1` puis à droite	`R0`
`1`	`1`	à gauche	écrire `0` puis à droite	`R1`
`$`	∅ ou `0` ou `1`	à droite		`RB`

Programme de l'état `R1`

lecture		actions		état
`E1`	`T1`	`E1`	`T1`	
`0`	∅ ou `0`	à gauche	écrire `1` puis à droite	`R0`
`0`	`1`	à gauche	écrire `0` puis à droite	`R1`
`1`	∅ ou `0`	à gauche	écrire `0` puis à droite	`R1`
`1`	`1`	à gauche	écrire `1` puis à droite	`R1`
`$`	∅ ou `0` ou `1`	à droite	écrire `1`	`RB`

— `RB` lit la bande `E1` et déplace les têtes de lecture de `E1` et de `T1` à droite et à gauche respectivement.

Programme de l'état `RB`

lecture	actions		état
`E1`	`E1`	`T1`	
`0` ou `1`	à droite	à gauche	`RB`
∅	à gauche	droite	`MU`

N.B. : après l'addition, la tête de lecture de T1 se trouve exactement une case plus à droite qu'avant, parce que, lorsque la machine est dans l'état R0 ou R1 et qu'elle lit $ sur E1, la tête de lecture de E1 va à droite alors que celle de T1 reste immobile.

Corrigé de l'exercice 12.1.4. Comme d'habitude, on va dire « tête de lecture » pour « tête de lecture/effaçage/écriture ».

La convention est que les nombres sont écrits sur la bande d'entrée comme des mots sur l'alphabet {0, 1, 2}, de gauche à droite, immédiatement après le symbole $ qui marque le début de la bande.

En outre, lorsque la machine commence son travail, la tête de lecture sur la bande d'entrée est placée à l'extrémité droite du mot, en face de la dernière lettre. De même les têtes de lecture sur les bandes de travail et de sortie, entièrement vides, sont situées sur la case immédiatement à droite du symbole $.

Nous aurons besoin de deux bandes de travail.

Nous notons la bande d'entrée T0. Les deux bandes de travail seront notées T1, T2. Enfin la bande de sortie sera notée T3.

Les états de la machine seront notés S1, S2, etc.

Dans l'état initial S1 on lit le premier entier de droite à gauche et on le recopie au fur et à mesure, de gauche à droite, sur la bande T1.

On a ensuite deux types de travail distincts.

Le premier type de travail consiste à recopier sur la bande de travail vierge le prochain entier disponible sur la bande T0.

Le deuxième type de travail consiste à comparer les deux entiers disponibles sur T1 et T2, puis à effacer le plus petit des deux.

En faisant alterner les deux types de travail, on finit par épuiser la bande d'entrée. Il reste alors à recopier sur la bande de sortie l'entier qui a été sélectionné.

Pour le premier type de travail nous avons deux états possibles S21 ou S22, selon que l'on doit recopier l'entier (lu sur l'entrée) sur la bande T1 ou T2.

Pour le deuxième type de travail, on lit de droite à gauche les deux entiers qui sont écrits « à l'envers » sur les deux bandes. *A priori* on sélectionne comme plus grand l'entier qui correspond à un chiffre lu plus grand, la premiere fois que cela se produit. Tant que cela ne s'est pas produit, on reste dans un état, noté S30, où les deux entiers sont présumés égaux.

On a ensuite deux états S31 et S32, selon que le nombre présumé plus grand se trouve sur la bande T1 ou T2.

Ceci dit, on conclut de manière sûre seulement lorsqu'on arrive en fin de lecture pour au moins l'un des deux entiers, car si les deux nombres n'ont pas la même longueur, c'est le plus long qui doit être retenu comme plus grand.

Une fois qu'on a l'assurance de savoir quel est le plus grand, ce qui se produit lorsque on lit $ sur au moins l'une des deux bandes de travail, on doit effacer la bande où se trouve l'entier le plus petit et repositionner la tête de lecture à la droite de l'entier le plus grand. Il y a ici deux états de la machine S33 et S34, selon que le nombre le plus grand se trouve sur la bande T1 ou T2.

Nous passons maintenant à la description précise des instructions exécutées en fonction de l'état de la machine et des cases lues sur certaines bandes.

Nous prenons un état après l'autre.

Un action sur une bande est codée par (I,g) si la bande est laissée intacte et la tête déplacée vers la gauche, (2,d) si elle écrit 2 et la tête est déplacée vers la droite, (E,i) si elle efface la case et ne déplace pas la tête.

Nous indiquons sur la première ligne, les bandes concernées en lecture, puis les noms des bandes où la machine travaille.

Sur les lignes suivantes :

— Ce qui est lu est indiqué entre parenthèses.

— Juste après les deux points, on indique le nouvel état.

— Ensuite on indique les actions sur les différentes bandes.

L'état initial est `S1`. Voici son programme.

```
   Programme de l'état S1
lecture            actions
T0       :        T0     T1
(0)      : S1 , (I,g), (0,d)
(1)      : S1 , (I,g), (1,d)
(2)      : S1 , (I,g), (2,d)
( )      : S22, (I,g), (I,g)
($)      : Stop
```

On note que s'il y a un seul entier écrit sur la bande, on arrête le programme sans rien écrire sur `T3`. On peut interpréter ceci comme un message d'erreur : la donnée n'est pas conforme à la syntaxe prévue.

Voici le programme de l'état `S21`.

```
  Programme de l'état S21
lecture            actions
T0       :       T0     T1
(0)      : S21, (I,g), (0,d)
(1)      : S21, (I,g), (1,d)
(2)      : S21, (I,g), (2,d)
( )      : S31, (I,g), (I,g)
($)      : S40, (I,i), (I,g)
```

Notons qu'on a introduit un état `S40` qui correspond au fait que la bande d'entrée a été lue en entier. Dans les états `S40`, `S41` et `S42` la machine fera la même chose que dans les états `S30`, `S31` et `S32`, à ceci près qu'après la comparaison des deux entiers on pourra passer à la phase terminale.

Voici le programme de l'état `S22`.

```
  Programme de l'état S22
lecture            actions
T0       :       T0     T2
(0)      : S22, (I,g), (0,d)
(1)      : S22, (I,g), (1,d)
(2)      : S22, (I,g), (2,d)
( )      : S31, (I,g), (I,g)
($)      : S40, (I,i), (I,g)
```

Voici le programme de l'état `S30`.

Programme de l'état S30

lecture	actions
T1 , T2 :	T1 T2
(m,n) avec m = n :	S30, (I,g), (I,g)
(m,n) avec m < n :	S31, (I,g), (I,g)
(m,n) avec m > n :	S32, (I,g), (I,g)
($,$) :	S33, (I,d), (I,d)
(m,$) :	S33, (I,d), (I,d)
($,m) :	S34, (I,d), (I,d)
sinon :	Stop

La première ligne est une abréviation correspondant aux trois possibilités (0,0), (1,1), (2,2).

De même dans toute la suite m et n constituent des abvréviations pour les cases lues 0 ou 1 ou 2.

La dernière ligne a été mise par pure précaution, pour le cas où une case lue serait vide, ce qui *a priori* ne peut pas arriver.

Voici les programmes des états S31 et S32.

Programme de l'état S31

lecture	actions
T1 , T2 :	T1 T2
(m,n) :	S31, (I,g), (I,g)
($,$) :	S33, (I,d), (I,d)
(m,$) :	S33, (I,d), (I,d)
($,m) :	S34, (I,d), (I,d)
sinon :	Stop

Programme de l'état S32

lecture	actions
T1 , T2 :	T1 T2
(m,n) :	S32, (I,g), (I,g)
($,$) :	S34, (I,d), (I,d)
(m,$) :	S33, (I,d), (I,d)
($,m) :	S34, (I,d), (I,d)
sinon :	Stop

Les programmes des états S40, S41 et S42 sont identiques à ceux des états S30, S31 et S32 en y remplaçant partout S30, S31, S32, S33, S34 respectivement par S40, S41, S42, S43, S44.

Nous passons maintenant à la procédure d'effaçage de la bande où se trouve le plus petit des deux entiers, en même temps qu'on replace la tête de lecture du plus grand entier à la droite, sur le premier chiffre. Notons que lorsqu'on atteint l'extrémité droite, on lit deux cases vides, on retourne d'un cran en arrière sur les deux bandes. Mais il reste à faire revenir en début de bande la tête de lecture sur la bande qui vient d'être effacée.

Cela donne les programmes S33, S34 avec leurs variantes S43, S44.

Voici les programmes des états S33 et S34.

Programme de l'état S33

lecture	actions
T1 , T2 :	T1 T2
(m,n) :	S33, (I,d), (E,d)
(,) :	S33, (I,g), (I,g)
(m,) :	S33, (I,i), (I,g)
(m,$) :	S21, (I,i), (I,d)
sinon :	Stop

Programme de l'état S34

lecture	actions
T1 , T2 :	T1 T2
(m,n) :	S34, (E,d), (I,d)
(,) :	S34, (I,g), (I,g)
(,m) :	S34, (I,g), (I,i)
($,m) :	S22, (I,d), (I,i)
sinon :	Stop

Les programmes des états S43 et S44 sont identiques à ceux des états S33 et S34 en y remplaçant partout S33, S34, S21, S22, respectivement par S43, S44, S51, S52.

Il reste à décrire la phase finale, dans laquelle on recopie l'entier le plus grand (qui se retrouvera écrit à l'endroit) sur la bande de sortie. Ces programmes S51, S52 sont tout à fait analogues, *mutatis mutandis*, aux programmes S21, S22.

Programme de l'état S51

lecture		actions		
T1	:	T1	T3	
(0)	: S51,	(I,g),	(0,d)	
(1)	: S51,	(I,g),	(1,d)	
(2)	: S51,	(I,g),	(2,d)	
($)	: Stop,	(I,d),	(I,g)	
sinon	: Stop			

Programme de l'état S52

lecture		actions		
T2	:	T2	T3	
(0)	: S52,	(I,g),	(0,d)	
(1)	: S52,	(I,g),	(1,d)	
(2)	: S52,	(I,g),	(2,d)	
($)	: Stop,	(I,d),	(I,g)	
sinon	: Stop			

Bibliographie

[D'Alembert] (1767). « Mélanges de littérature, d'histoire, et de philosophie ». In : t. cinquième. Ouvrage paru anonymement. Amsterdam : chez Zacharie Chatelain et Fils. Chap. Éclaircissemens sur différens endroits des Élémens de Philosophie, p. 1-272 (voir page 211).

Académie française (1878). *Dictionnaire*. 7e éd. T. premier : A-H. Paris : Firmin-Didot et Cie (voir page 136).

— (1932). *Dictionnaire*. 8e éd. T. premier : A-G. Paris : Hachette (voir pages 137, 139).

— (1935). *Dictionnaire*. 8e éd. T. second : H-Z. Paris : Hachette (voir page 135).

— (1992). *Dictionnaire*. 9e éd. T. 1 : A-Enz. Paris : Arthème Fayard et Imprimerie nationale (voir pages 137, 139).

— (2000). *Dictionnaire*. 9e éd. T. 2 : Éoc-Map. Paris : Arthème Fayard et Imprimerie nationale (voir page 138).

— (2011). *Dictionnaire*. 9e éd. T. 3 : Maq-Quo. Paris : Arthème Fayard et Imprimerie nationale (voir page 137).

Académie françoise (1694). *Dictionnaire*. T. premier : A-L. Paris : chez la veuve de Jean Baptiste Coignard : et chez Jean Baptiste Coignard (voir pages 136, 137, 139).

Aristote (1926). *Physique*. T. second, I–IV. Paris : Belles Lettres. Texte établi et traduit par Henri Carteron, disponible à la BU Lettres (voir page 73).

— (1967). *Topiques : tome I, livres I-IV*. Collection des universités de France. Texte établi et traduit par Jacques Brunschwig. Paris : Les Belles Lettres (voir page 129).

— (2005). *Seconds analytiques*. G. F. bilingue 1186. Paris : Flammarion. Introduction, traduction, notes, bibliographie et index par Pierre Pellegrin (voir page 212).

Arnauld, Antoine (2009). « Nouveaux éléments de géométrie ». In : *Géométries de Port-Royal*. Dir. Dominique Descotes. Sources classiques 100. Paris : Honoré Champion, p. 91-798 (voir page 210).

Arnauld, Antoine et Pierre Nicole (2011). *La logique, ou l'Art de penser*. Sources classiques 108. Édition critique par D. Descotes. Paris : Honoré Champion (voir pages 129, 133, 135, 136, 207).

Augustin, saint (2000). *Œuvres. II, La Cité de Dieu*. Paris : Gallimard. Édition publiée sous la direction de Lucien Jerphagnon, disponible à la BU Lettres (voir page 74).

Barbin, Évelyne (1994). « Sur la conception des savoirs géométriques dans les *Éléments de géométrie* ». In : *Cahiers de didactique des mathématiques (Thessalonique)* 14-15, p. 135-158 (voir page 208).

Berkeley, George (1987). « L'analyste ». In : *Œuvres*. T. 2. Traduction et annotations par Michel Blay. Paris : Presses universitaires de France, p. 257-332. Disponible à la BU Lettres (voir page 77).

Bhāskarācārya (2004). *Le Siddhāntaśiromaṇi I-II*. T. 2, Traduction. Éd., trad. et commentaire par François Patte. Genève : Droz. Disponible à la bibliothèque de mathématiques (voir page 74).

Bishop, Errett (1967). *Foundations of constructive analysis*. New York : McGraw-Hill. Disponible à la bibliothèque de mathématiques (voir pages 66, 67, 96, 117).

Bishop, Errett et Douglas Bridges (1985). *Constructive analysis*. T. 279. Grundlehren der mathematischen Wissenschaften. Berlin : Springer. http://dx.doi.org/10.1007/978-3-642-61667-9. Disponible à la bibliothèque de mathématiques (voir page 96).

Bkouche, Rudolf (2009). « Qu'est-ce qu'une ligne droite ? » In : *Histoire du calcul de la géométrie à l'algèbre*. Dir. Luc Sinègre. Rouen : Vuibert, p. 173-186. http://michel.delord.free.fr/rb/rb-lignedroite.pdf (voir page 140).

Bolzano, Bernard (1993). *Les paradoxes de l'infini*. Introduction, traduction de l'allemand et notes par Hourya Sinaceur. Paris : Éditions du Seuil, p. 193. Disponible à la BU Lettres (voir page 78).

Borel, Émile (1898). *Leçons sur la théorie des fonctions*. Paris : Gauthier-Villars et fils. https://archive.org/details/leconstheoriefon00borerich (voir page 204).
Bridges, Douglas et Fred Richman (1987). *Varieties of constructive mathematics*. T. 97. London Mathematical Society Lecture Note Series. Cambridge : Cambridge University Press, p. x+149. http://dx.doi.org/10.1017/CBO9780511565663. Disponible à la bibliothèque de mathématiques (voir page 113).
Cantor, Georg (1932). « Mitteilungen zur Lehre vom Transfiniten ». In : *Gesammelte Abhandlungen*. Berlin : Springer Verlag, p. 378-439 (voir page 73).
Cauchy, Augustin-Louis (1821). *Cours d'analyse de l'École royale polytechnique. I.re partie. Analyse algébrique*. Imprimerie royale chez Debure frères. Disponible à la BU Sciences (voir page 79).
— (1823). *Résumé des leçons données à l'École royale polytechnique sur le calcul infinitésimal. Tome premier*. Imprimerie royale chez Debure frères. Disponible à la BU Sciences (voir pages 79, 91).
— (1853). « Note sur les séries convergentes dont les divers termes sont des fonctions continues d'une variable réelle ou imaginaire, entre des limites données ». In : *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 36, p. 454-459. Réimprimé comme extrait no 518 dans les *Œuvres complètes*, série I, tome 12, pages 30-35 (voir page 89).
Caveing, Maurice (1990). « Introduction générale ». In : *Euclide d'Alexandrie : les Éléments*. Éd. Bernard Vitrac. T. 1 : livres I-IV : géométrie plane. Bibliothèque d'histoire des sciences. Paris : Presses universitaires de France, p. 13-148 (voir pages 127, 130, 132, 207).
Clavius, Christophorus (1589). *Euclidis Elementorvm lib[ri] XV*. Rome : Apud Bartholomaeum Grassium. http://hdl.handle.net/10481/9882 (voir page 140).
« Correspondance Cantor-Dedekind » (1962). In : Cavaillès, Jean. *Philosophie mathématique*. Histoire de la pensée 6. Paris : Hermann, p. 173-251. Traduction de Charles Ehresmann (voir page 62).
Davenport, Anne A. (1997). « The Catholics, the Cathars, and the concept of infinity in the thirteenth century ». In : *Isis* 88.2, p. 263-295. http://dx.doi.org/10.1086/383692 (voir page 74).
David, René, Karim Nour et Christophe Raffalli (2004). *Introduction à la logique : théorie de la démonstration : cours et exercices corrigés*. 2e édition. Paris : Dunod (voir page 178).
De Morgan, Augustus (1836). *The connexion of number and magnitude : an attempt to explain the fifth book of Euclid*. Londres : Taylor et Walton. http://archive.org/details/connexionofnumbe00demorich (voir pages 193, 195-197).
Dedekind, Richard (2008). « Continuité et nombres irrationnels ». In : *La création des nombres*. J. Vrin, p. 57-89. Introduction, traduction et notes par Hourya Benis Sinaceur, disponible à la BU Sciences (voir page 218).
Diener, Francine et Georges Reeb (1989). *Analyse non standard*. T. 40. Collection Enseignement des Sciences. Paris : Hermann. Disponible à la bibliothèque de mathématiques (voir page 61).
Dieudonné, J. (1960). *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. X. New York : Academic press. Traduction de Denise Huet : *Fondements de l'analyse moderne*, Gauthier-Villars, Paris, 1960 (voir page 99).
Dowek, Gilles (1995). *La logique*. Paris : Flammarion (voir page 178).
— (2007). *Les métamorphoses du calcul : une étonnante histoire de mathématiques*. Paris : Le Pommier (voir pages C, 178).
Euclide d'Alexandrie (1990-2001). *Les Éléments*. Bibliothèque d'histoire des sciences. 4 volumes. Livres I-IV : géométrie plane. Livres V-VI : proportions et similitude. Livres VII-IX : arithmétique. Livre X : grandeurs commensurables et incommensurables. Classification des lignes irrationnelles. Livres XI-XIII : géométrie des solides. Traduction du texte de Heiberg et commentaires par Bernard Vitrac. Paris : Presses universitaires de France. Disponible à la bibliothèque de mathématiques. Disponible à la BU Sciences (voir pages 7, 73, 126, 127, 131, 133, 134, 138, 140, 141, 180, 207, 208).
Fenstad, Jens Erik (1988). « Infinities in mathematics and the natural sciences ». In : *Methods and applications of mathematical logic (Campinas, 1985)*. T. 69. Contemp. Math. Providence : American Mathematical Society, p. 79-92 (voir page 78).
Fermat, Pierre de (1894). *Œuvres. 2, Correspondance*. Gauthier-Villars et fils. Publiées par les soins de MM. Paul Tannery et Charles Henry (voir page 76).

Frege, Gottlob et David Hilbert (1992). « Correspondance ». In : *Logique et fondements des mathématiques : anthologie (1850-1914)*. Dir. l'Institut d'histoire et de philosophie des sciences et des techniques. Traduction et introduction de J. Dubucs. Paris : Payot, p. 215-235 (voir pages 146, 147).
Galilei, Galileo (1638). *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze, attenenti alla mecanica & i movimenti locali*. Leiden : Elzevier. Traduction, notes et index par Maurice Clavelin : *Discours et démonstrations mathématiques concernant deux sciences nouvelles*, Presses universitaires de France, Paris, 1995, disponible à la BU Sciences (voir pages 75, 76).
Gandon, Sébastien (2009). « La théorie des rapports chez Augustus De Morgan ». In : *Revue d'histoire des sciences* 62.1, p. 285-311. http://www.cairn.info/revue-d-histoire-des-sciences-2009-1-page-285.htm (voir page 193).
Gentzen, Gerhard (1935). « Untersuchungen über das logische Schließen. I ». In : *Mathematische Zeitschrift* 39.1, p. 176-210. http://dx.doi.org.scd1.univ-fcomte.fr/10.1007/BF01201353. Traduit dans Gentzen 1955, p. 3-28, 41-76 (voir pages 149, 154, 155).
— (1936). « Die Widerspruchsfreiheit der reinen Zahlentheorie ». In : *Mathematische Annalen* 112, p. 493-565. Traduit dans Largeault 1992, p. 288-357 (voir pages 149, 214, 215).
— (1955). *Recherches sur la déduction logique*. Philosophie de la matière 5. Traduction et commentaire par Robert Feys et Jean Ladrière (voir pages 155, 265).
Gergonne, Joseph Diez (1818). « Essai sur la théorie des définitions ». In : *Annales de mathématiques pures et appliquées* IX.1, p. 1-35. https://eudml.org/doc/79733 (voir pages 208, 251).
Giovacchini, Julie (2010). « L'angle et l'atome dans la physique épicurienne : réflexions sur un témoignage de Sextus Empiricus ». In : *Philosophie et mathématiques*. Philosophie antique 10. Presses universitaires du Septentrion, Villeneuve d'Ascq, p. 139-166 (voir page 132).
Giusti, Enrico (1984). « Gli "errori" di Cauchy e i fondamenti dell'analisi ». In : *Boll. Storia Sci. Mat.* 4.2, p. 24-54 (voir page 79).
— (1999). *Ipotesi sulla natura degli oggetti matematici*. Turin : Bollati Boringhieri. Traduction de Georges Barthélemy : *La naissance des objets mathématiques*, Ellipses, Paris, 2000, disponible à la BU Sciences (voir page C).
Gomez-Lobo, Alfonso (1981). « Definitions in Aristotle's *Posterior Analytics* ». In : *Studies in Aristotle*. Dir. Dominic J. O'Meara. Studies in philosophy and the history of philosophy 9. Washington : The Catholic university of America press, p. 25-46 (voir page 128).
Guillaume d'Ockham (1988). *Somme de logique*. T.E.R. bilingue. Traduction, introduction et notes de Joël Biard. Mauvezin : Trans-Europ-Repress (voir page 206).
Hallett, Michael et Ulrich Majer, éd. (2004). *David Hilbert's Lectures on the foundations of geometry, 1891-1902*. Foundational Lectures 1. Berlin : Springer (voir pages 145, 146, 265).
Heath, Thomas (1949). *Mathematics in Aristotle*. Oxford : Clarendon Press (voir page 77).
Helmholtz, Hermann (16 juin 1877). « Les axiomes de la géométrie. Leur origine et leur signification ». In : *La revue scientifique de la France et de l'étranger. Revue des cours scientifiques (2e série)*. 3e sér. 6e année.51. Traduction de *Über den Ursprung und die Bedeutung der geometrischen Axiome*, Vortrag im Docentenverein, Heidelberg, 1870 (voir page 10).
Henrion, D. (1632). *Les quinze livres des Éléments géométriques d'Euclide*. Paris : Imprimerie d'Isaac Dedin. http://n2t.net/ark:/47881/m6cf9n9x (voir pages 140, 141).
Hilbert, David (1899). « Grundlagen der Geometrie ». In : *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*. 92 pages. Leipzig : B. G. Teubner (voir page 145).
— (1900). « Les Principes fondamentaux de la géométrie ». In : *Annales scientifiques de l'École Normale Supérieure (3)* 17. Traduit par Léonce Laugel, p. 103-209. https://eudml.org/doc/81144 (voir pages 12, 145, 146).
— (1902). *Grundlagen der Geometrie*. Lesesaal, Mathematisches Institut. Publié dans Hallett et Majer (2004, pages 531-606). Georg-August-Universität Göttingen (voir page 147).
— (1926). « Über das Unendliche ». In : *Mathematische Annalen* 95. Traduction par André Weil : « Sur l'infini », *Acta Mathematica* 48:91–122, 1926, doi:10.1007/BF02629757. Il y a aussi la traduction dans Largeault 1972, p. 220-245, disponible à la BU Lettres, p. 161-190. https://eudml.org/doc/159124 (voir page 62).
— (1971). *Les fondements de la géométrie*. Monographies universitaires de mathématiques. Édition critique avec introduction et compléments préparée par Paul Rossier. Paris : Dunod (voir page 146).

Hilbert, David et P. Bernays (2001). *Fondements des mathématiques 1*. Traduction de l'ouvrage *Grundlagen der Mathematik 1*, 2e édition (1968) avec les passages parallèles de la 1re édition (1934) par François Gaillard et Marcel Guillaume. Paris : L'Harmattan (voir pages 134, 147, 148).
Hume, David (1995). *Traité de la nature humaine. Livre I et appendice, L'entendement*. Traduction de Philippe Baranger et Philippe Saltel. Flammarion. Disponible à la BU Lettres (voir page 77).
Imbs, Paul (1971). « Préface ». In : *Trésor de la langue française : dictionnaire de la langue du XIXe et du XXe siècle (1789-1960)*. Dir. le Centre de recherche pour un Trésor de la langue française (Nancy). T. premier : A-affiner. Paris : Éditions du Centre national de la recherche scientifique, p. IX-XLVII. http://www.atilf.fr/IMG/pdf/La_Preface_originale_du_TLF.pdf (voir pages 209, 210).
Kant, Emmanuel (2006). *Critique de la raison pure*. Troisième édition. Paris : Flammarion. Traduction, présentation et notes par Alain Renaut (voir page 77).
Knobloch, Eberhard (1999). « Galileo and Leibniz : different approaches to infinity ». In : *Arch. Hist. Exact Sci.* 54.2, p. 87-99. http://dx.doi.org/10.1007/s004070050035 (voir page 76).
Knorr, Wilbur R. (1982). « Infinity and continuity : The interaction of mathematics and philosophy in antiquity ». In : *Infinity and continuity in ancient and medieval thought*. Dir. Norman Kretzmann. Ithaca : Cornell University Press, p. 112-145 (voir page 72).
Lakatos, Imre (1984). *Preuves et réfutations : essai sur la logique de la découverte mathématique*. Paris : Hermann. Traduction de Nicolas Balacheff et Jean-Marie Laborde, disponible à la BU Sciences (voir pages C, 79).
Largeault, Jean, éd. (1972). *Logique mathématique : textes*. Épistemologie. Paris : Armand Colin (voir page 265).
— éd. (1992). *Intuitionisme et théorie de la démonstration : textes de Bernays, Brouwer, Gentzen, Gödel, Hilbert, Kreisel, Weyl : réunis, traduits et présentés*. Paris : J. Vrin (voir pages 215, 265).
Lebesgue, Henri (1904). *Leçons sur l'intégration et la recherche des fonctions primitives*. Paris : Gauthier-Villars (voir page 95).
Leibniz, G. W. (1995). *La caractéristique géométrique*. Mathesis. Traduction du latin par Javier Echeverría. Paris : J. Vrin (voir pages 130, 140).
Lejeune-Dirichlet (1829). « Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données ». In : *Journal für die reine und angewandte Mathematik* 4.2, p. 157-169. https://eudml.org/doc/183134 (voir page 95).
Les Présocratiques (1988). Éd. Jean-Paul Dumont, Daniel Delattre et Jean-Louis Poirier. Paris : Gallimard (voir page 200).
Lesieur, L. et Cl. Joulain (1966). *Mathématiques. P. C. 1ère année et spéciales B. Tome 1 : algèbre et géométrie*. Seconde édition revue et corrigée. Armand Colin. Disponible à la bibliothèque de mathématiques (voir page 47).
Lobatchevski, Nikolaï (1840). *Geometrische Untersuchungen zur Theorie der Parallellinien*. Berlin : G. Fincke. Traduction de Jules Hoüel : *Études géométriques sur la théorie des parallèles*, Paris : Gauthier-Villars, 1866, disponible à la BU Sciences (voir page 181).
Loget, François (2002). « Jacques Peletier du Mans, mathématicien : l'angle de contact ». In : *Nouvelle revue du XVIe siècle* 20.2, p. 37-55 (voir page 208).
Lombardi, Henri (oct. 1991). « L'uniformité, un concept implicite efficace chez Cauchy ». In : *Repères IREM* 5, p. 112-126. http://www.univ-irem.fr/exemple/reperes/articles/5_article_30.pdf (voir page 266).
— (1994). « L'uniformité, un concept implicite efficace chez Cauchy ». In : *Quatrième Université d'été d'histoire des mathématiques, Lille, juillet 1990*. Dir. Commission inter-IREM Epistémologie et histoire des mathématiques (France). Lille : IREM. repris dans Lombardi 1991 (voir page 79).
— (oct. 1997). « Le raisonnement par l'absurde ». In : *Repères IREM* 29, p. 27-42. http://www.univ-irem.fr/exemple/reperes/articles/29_article_196.pdf (voir page 225).
McKenna, Antony (2001). « Les *Pensées* de Pascal : une ébauche d'apologie sceptique ». In : *Le retour des philosophes antiques à l'âge classique. Tome II, Le scepticisme au XVIe et au XVIIe siècle*. Dir. Pierre-François Moreau. Bibliothèque Albin Michel Idées. Albin Michel, p. 348-361 (voir page 143).
Nelson, Edward (1977). « Internal set theory : a new approach to nonstandard analysis ». In : *Bull. Amer. Math. Soc.* 83.6, p. 1165-1198. Traduction : *Théorie des ensembles internes : une nouvelle approche*

de l'analyse non standard dans *La mathématique non standard*, CNRS Éditions (1989), pages 355-399, disponible à la bibliothèque de mathématiques (voir page 61).
Netz, Reviel, Ken Saito et Natalie Tchernetska (2001). « A new reading of *Method* Proposition 14 : preliminary evidence from the Archimedes palimpsest (Part 1) ». In : *SCIAMVS* 2, p. 9-29 (voir page 73).
Pascal, Blaise (1970a). « Lettre à Le Pailleur ». In : *Œuvres complètes II*. Bibliothèque européenne. Texte établi, présenté et annoté par Jean Mesnard. Paris : Desclée de Brouwer, p. 556-576 (voir page 128).
— (1970b). *Œuvres complètes II*. Bibliothèque européenne. Texte établi, présenté et annoté par Jean Mesnard. Paris : Desclée de Brouwer (voir page 75).
— (1991). « De l'esprit géométrique ». In : *Œuvres complètes III*. Bibliothèque européenne. Texte établi, présenté et annoté par Jean Mesnard. Paris : Desclée de Brouwer, p. 360-428 (voir pages 128, 130, 132, 142, 143).
Pasch, Moritz (1882). *Vorlesungen über neuere Geometrie*. Leipzig : B. G. Teubner (voir pages 144, 145).
Poincaré, Henri (1894). « Sur la nature du raisonnement mathématique ». In : *Revue de métaphysique et de morale* 2.4, p. 371-384 (voir page 32).
— (1902). *La science et l'hypothèse*. Paris : Flammarion (voir page 32).
— (1905). « Les mathématiques et la logique ». In : *Revue de Métaphysique et de Morale* 13.6, p. 815-835. http://gallica.bnf.fr/ark:/12148/bpt6k110592/f821. Disponible à la BU Lettres (voir pages 67, 267).
— (1906). « Les mathématiques et la logique ». In : *Science et Méthode*. Paris : Flammarion, p. 152-171. Reprise partielle de Poincaré 1905 (voir page 67).
— (1909). « La logique de l'infini ». In : *Revue de Métaphysique et de Morale* 17.4, p. 461-482. http://gallica.bnf.fr/ark:/12148/bpt6k110999. Disponible à la BU Lettres (voir pages 63, 64, 267).
— (1913). « La logique de l'infini ». In : *Dernières pensées*. Paris : Flammarion, p. 101-139. Paru auparavant dans Poincaré 1909 (voir page 63).
Rang, Bernhard et Wolfgang Thomas (1981). « Zermelo's discovery of the "Russell Paradox" ». In : *Historia Mathematica* 8.1, p. 15-22. http://www.sciencedirect.com/science/article/pii/0315086081900021 (voir page 123).
Richard, Jules (30 juin 1905). « Les principes des mathématiques et le problème des ensembles ». In : *Revue générale des sciences pures et appliquées* 16e année.12, p. 541. http://gallica.bnf.fr/ark:/12148/bpt6k17080b/f545 (voir page 201).
Riemann, Bernhard (1876). *Gesammelte mathematische Werke und wissenschaftlicher Nachlass*. Leipzig : B. G. Teubner. Traduction de L. Laugel : *Œuvres mathématiques*, Gauthier-Villars et fils, 1898 (voir page 95).
Saccheri, Gerolamo (2012). *Logica dimostrativa*. Mathematica italiana 3. Par les soins de Massimo Mugnai et Massimo Girondino. Pise : Edizioni della Normale (voir page 267).
Saccherius, Hieronymus (1701). *Logica demonstrativa*. Reproduit en facsimilé dans Saccheri (2012). Pavie : Typis hæredum Caroli Francisci Magrij Impressorum Civit. (voir page 129).
Sebestik, Jan (1992). *Logique et mathématique chez Bernard Bolzano*. Paris : Librairie philosophique J. Vrin (voir page 77).
Spalt, Detlef D. (1990). « Die Unendlichkeiten bei Bernard Bolzano ». In : *Konzepte des mathematisch Unendlichen im 19. Jahrhundert*. T. 5. Stud. Wiss. Soz. Bildungsgesch. Math. Göttingen : Vandenhoeck & Ruprecht, p. 189-218 (voir page 73).
Tannery, Jules (1906). *Leçons d'algèbre et d'analyse à l'usage des élèves des classes de mathématiques spéciales*. T. second. Paris : Gauthier-Villars (voir pages 95, 96).
Thomas d'Aquin, saint (1984). *Somme théologique*. T. 1. Éditions du Cerf. Coordonné par Albert Raulin, disponible à la BU Lettres (voir page 74).
Turing, Alan et Jean-Yves Girard (1995). *La machine de Turing*. Paris : Éditions du Seuil. Disponible à la BU Sciences (voir pages 102, 178).
Vitrac, Bernard (2002). « 'Umar al Khayyam et l'anthyphérèse : étude du deuxième Livre de son commentaire *« Sur certaines prémisses problématiques du Livre d'Euclide »* ». In : *Farhang : quarterly journal of humanities and cultural studies* 14, p. 137-192. http://hal.archives-ouvertes.fr/hal-00174930 (voir page 238).

Vitrac, Bernard (2012). « Les démonstrations par l'absurde dans les *Éléments* d'Euclide : inventaire, formulation, usages ». http://hal.archives-ouvertes.fr/hal-00496748 (voir pages 225, 226).