Transcription en LaTeX et traductions de petits articles courts de la revue Eureka des Archimédiens
ia deepseek pour Denise Vella-Chemla, septembre 2026

p, non-p ou ni l'un ni l'autre ?

George Raptis

Pour tous ceux qui n'ont jamais rencontré de systèmes logiques différents de la logique classique, la question ci-dessus semblera probablement absurde ou dénuée de sens. Dans ce court article, nous souhaitons initier le lecteur à ces logiques non classiques alternatives et, espérons-le, le motiver à approfondir le sujet par lui-même.

Il semble que lorsqu'on tente de parler de logiques non classiques, on se trouve souvent "obligé" d'expliquer les motifs justifiant l'introduction de telles idées. En un sens, cela est raisonnable si l'on pense à la longue tradition de la logique classique, à son influence sur les mathématiques, la philosophie et la science et, bien sûr, à son succès en mathématiques (par exemple, la complétude) et en science (par exemple, la physique classique). Ces éléments rendent difficile une révision de la logique classique, car elle exige des changements plus fondamentaux dans notre schéma conceptuel que, par exemple, une révision d'une théorie scientifique quelconque. Néanmoins, après avoir vu l'émergence et le développement de certaines théories modernes en philosophie et en physique (par exemple, la mécanique quantique, certaines nouvelles vues sur le déterminisme et la probabilité, l'intuitionnisme, etc.), le sujet semble être plus naturellement motivé ou, mieux, il semble moins absurde.

Habituellement, les motifs sous-jacents aux logiques non classiques sont de nature philosophique et fournissent donc une source infinie de discussion. Informellement, on peut dire que les systèmes logiques non classiques ont été motivés par la considération que la logique classique semblait trop permissive, en ce sens que l'implication (classique) admet trop d'énoncés comme vrais (en particulier, la propriété "paradoxale" de l'implication classique : $p \Rightarrow q$ est vrai pour tout q si p est faux, ce qui permet à certains arguments qui semblent absurdes ou hors de propos d'être néanmoins vrais), ou trop étroite, en ce sens que l'hypothèse de la logique classique selon laquelle les propositions sont soit vraies soit fausses (dichotomie) est trop restrictive. Cela a conduit à la création des calculs intuitionnistes, de la logique modale, des logiques multivalentes, de la logique floue et d'autres logiques non classiques qui diffèrent de la logique classique de différentes manières. Pour les logiciens qui trouvent la logique classique satisfaisante, les arguments soutenant les systèmes non classiques semblent parfois reposer trop sur des questions épistémologiques (ou métaphysiques).

Inutile de mentionner (ou quoi qu'il en soit), les logiques non classiques méritent d'être étudiées pour l'amour de la logique (mathématique) elle-même et nous pouvons toujours considérer cela comme un acte d'"expérimentation mathématique", qui pourrait s'avérer fructueux de manières que nous ne pouvons pas encore connaître.

La logique modale est une classe très utilisée de logiques non classiques. En logique modale, nous voulons que notre raisonnement implique des expressions telles que "il est possible que" ou "il est nécessaire que", avec tous les connecteurs usuels de la logique (propositionnelle) classique. Nous

introduisons donc un opérateur (modal) $\Box$, tel que $\Box p$ signifie "il est nécessaire que p". Ensuite, un autre opérateur $\Diamond$ peut être défini en termes de $\Box$ par $\Diamond p = \neg\Box\neg p$ (signifiant "il est possible que"), où $\neg$ est le connecteur propositionnel usuel pour la négation. À ce stade, il est difficile de savoir quelle devrait être la valeur de vérité (qui est nécessairement soit vraie soit fausse) de $\Box p$ lorsque p est donné comme vrai, de sorte que nous ne pouvons pas définir la validité en logique modale à l'aide de tables de vérité comme nous le faisions pour la logique propositionnelle classique. Cela s'exprime habituellement en disant que $\Box$ n'est pas un opérateur de vérité fonctionnel. Néanmoins, nous pouvons introduire une sémantique en logique modale par l'idée de "mondes possibles" (sémantique des mondes possibles). Selon cette approche, nous parlons de la validité d'une proposition p dans un monde possible. Une proposition p peut avoir différentes valeurs de vérité dans différents mondes. Plus précisément, une fonction de valuation donne toujours une valeur de vérité à une proposition p dans un monde w de W (= tous les mondes possibles), c'est-à-dire qu'on écrit $\mathrm{v}(p, w)$. Nous demandons alors que ce qui suit soit vérifié :

(i) $\mathrm{v}(p, w) = 1$ si et seulement si $\mathrm{v}(\neg p, w) = 0$,

(ii) $\mathrm{v}(p \Rightarrow q, w) = 1$ si et seulement si $\mathrm{v}(p, w) = 0$ ou $\mathrm{v}(q, w) = 1$ et

(iii) $\mathrm{v}(\Box p, w) = 1$ si et seulement si $\mathrm{v}(p, w') = 1$ pour tout w' dans W.

Maintenant, étant donné les valeurs d'un ensemble de propositions primitives P dans tous les mondes possibles, les valeurs de vérité des propositions dans le langage engendré par P peuvent être trouvées. Nous sommes maintenant prêts à définir la validité dans ce système. Un ensemble de propositions S implique sémantiquement une proposition p dans un ensemble donné W de mondes possibles si et seulement si toute valuation des propositions primitives qui attribue aux propositions de S la valeur 1 (vrai) dans un monde w de W attribue également à p la valeur 1 dans le même monde. Nous disons que l'argument dont les prémisses sont les propositions de S et la conclusion est p est valide si et seulement si S implique sémantiquement p pour tout ensemble non vide W de mondes possibles. La tâche suivante est d'introduire une notion de preuve qui crée un système qui est correct, c'est-à-dire que tous les énoncés prouvables sont valides, et, si possible, également complet, c'est-à-dire qu'une proposition est valide si et seulement s'il existe une preuve pour elle. Cela se fait (comme en logique classique) en prenant certains axiomes et règles d'inférence. Prenons les axiomes suivants :

(i) $\Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q)$,

(ii) $\Box p \Rightarrow p$ et

(iii) $\Diamond p \Rightarrow \Box\Diamond p$

avec la règle d'inférence :

(1) de p, inférer $\Box p$

et les axiomes et règles d'inférence usuels de la logique propositionnelle. Ce système (habituellement appelé S5) s'avère être complet.

Il existe également d'autres systèmes qui diffèrent de celui-ci soit par le choix des axiomes, soit par le choix et l'action des opérateurs (modaux), qui sont généralement encore appelés systèmes de logique modale. Pour clarifier cela, mentionnons ici qu'en logique modale, nous avons aussi une notion d'accessibilité, qui est formellement une relation binaire R sur W, de sorte que sRt exprime

le fait que t est accessible depuis s. Alors la valeur de vérité de $\Box p$ dans un monde w dépend uniquement des valeurs de vérité de p dans les mondes accessibles depuis w et la dépendance se fait de manière évidente. Dans le système S5, la relation R est triviale, c'est-à-dire que tout est accessible depuis tout le reste. Différentes relations binaires R nous fournissent différentes interprétations du langage, et les propriétés spéciales que R pourrait avoir (par exemple la transitivité) se reflètent dans l'interprétation des opérateurs modaux. C'est pourquoi le terme logique modale désigne une classe de systèmes logiques plutôt qu'un système spécifique. Peut-être la chose la plus importante concernant la logique modale que le lecteur devrait garder à l'esprit et sur laquelle réfléchir est l'idée de sémantique des mondes possibles - l'idée de valeurs de vérité dépendantes du cadre. Cette idée s'exprime aussi en disant qu'en logique modale nous distinguons deux types de vérité : la vérité nécessaire (lorsque $\mathrm{v}(\Box p, w) = 1$) et la vérité contingente (cela s'applique aux propositions qui, bien que vraies dans certains mondes, peuvent être fausses dans d'autres). Le lecteur voudra peut-être comparer cela avec l'interprétation des mondes possibles de la mécanique quantique.

Une autre classe de systèmes logiques (qui nous concernera dans le reste de l'article) est celle de la logique multivalente. La logique multivalente a été créée dans les années 1920 par le logicien et philosophe polonais J. Łukasiewicz. Le lecteur intéressé par les motifs philosophiques de la logique multivalente devrait rechercher les propres idées philosophiques de Łukasiewicz.

La logique multivalente diffère de la logique classique en ce qu'elle ne restreint pas le nombre de valeurs de vérité à deux. Les systèmes de logique multivalente peuvent avoir n'importe quel nombre fini (ou même infini dénombrable) de valeurs de vérité. Le problème de l'interprétation des degrés de vérité est principalement philosophique et il n'existe pas d'interprétation généralement acceptée. En général, il semble préférable de ne pas essayer de s'accorder sur une interprétation mais de laisser celle-ci dépendre du domaine dans lequel on applique la logique à chaque fois.

D'un point de vue épistémologique, il existe des arguments en faveur des logiques multivalentes, principalement dus à la mécanique quantique. Le lecteur est probablement familier avec certaines des difficultés qui surgissent en mécanique quantique lorsqu'on essaye de décrire des résultats expérimentaux ou lorsqu'on tente de comprendre des notions quantiques en termes de logique classique. Reichenbach a suggéré que si nous interprétons la mécanique quantique en termes de logique multivalente, il est possible que nous puissions nous débarrasser des problèmes conceptuels de la mécanique quantique. Dans une perspective philosophique, c'est très important : si cela est possible, cela remet en question le "dogme" de l'universalité de la logique classique et l'affirmation de son indépendance vis-à-vis de toute considération épistémologique ou métaphysique.

La manière standard d'introduire un système logique multivalent ressemble à ce que nous avons fait pour la logique formelle classique. Nous verrons comment cela peut être fait au niveau propositionnel uniquement. Nous fixons un ensemble W dont les valeurs sont les degrés de vérité et un ensemble de degrés de vérité désignés, qui est un sous-ensemble de l'ensemble des degrés de vérité et dont les éléments agissent comme substituts de la valeur de vérité traditionnelle "vrai".

La sémantique dans les logiques multivalentes est le plus souvent commodément définie par des tables de vérité comme en logique classique. Les tables de vérité montreront comment la fonction de valuation "interprète" les connecteurs propositionnels primitifs. La taille des tables de vérité

dépend du nombre de degrés de vérité dans la logique. Nous définissons la notion d'implication sémantique de manière similaire à la logique classique (où maintenant une proposition est vraie (valide) si et seulement si toutes les fonctions de valuation lui donnent un degré de vérité désigné).

L'étape suivante, comme en logique classique, est d'introduire une notion de preuve dans la logique. Nous pouvons le faire en fixant un ensemble d'axiomes avec un ensemble de règles d'inférence. Ensuite, nous définissons la preuve et l'implication syntaxique comme d'habitude. Dans ce qui suit, nous verrons comment nous pouvons appliquer cela pour définir les systèmes de Łukasiewicz.

Il est raisonnable de commencer avec 3 valeurs : 0, $\frac{1}{2}$, 1. Si nous réservons 1 pour vrai et 0 pour faux, comment pouvons-nous interpréter $\frac{1}{2}$? Si nous interprétons $\frac{1}{2}$ comme le degré de vérité des énoncés dépourvus de sens (mal formulés), alors nous ne gagnons pas grand-chose en introduisant ce nouveau système, puisque en logique classique nous restreignons de toute façon l'attention aux énoncés bien formés (formules). Nous devons donc comprendre ici que toute tentative de construire un système non classique doit d'abord s'engager dans un changement conceptuel concernant la validité ou la prouvabilité ou quelque chose d'autre d'également fondamental. De plus, nous devons mentionner ici que puisque nous utilisons l'anglais courant pour décrire les systèmes logiques multivalents (et pour énoncer des métathéorèmes à leur sujet), nous ne changeons pas la logique utilisée dans le métalangage.

Un système de Łukasiewicz est un système multivalent L_m, pour m un entier positif, qui a un ensemble de degrés de vérité de m éléments, que nous écrivons habituellement comme $W_m = \left\{\frac{k}{m-1} : 0 \leq k \leq m-1\right\}$. Dans ce système, le degré 1 est le seul degré de vérité désigné.

Les principaux connecteurs propositionnels, à partir desquels le langage est construit (comme dans le calcul propositionnel classique), sont une implication $\rightarrow$ ("implique") et une négation $\neg$ ("non"). Ainsi le système de Łukasiewicz a les mêmes formules bien formées que le système classique. Maintenant, nous pouvons définir comment la fonction de vérité (valuation) $\mathbf{v}$ s'applique à ces connecteurs propositionnels par les formules $\mathrm{v}(u \rightarrow v) = \min\{1, 1 - \mathrm{v}(u) + \mathrm{v}(v)\}$ et $\mathrm{v}(\neg u) = 1 - \mathrm{v}(u)$. Alors la valeur de toute formule bien formée dans le langage peut être déterminée étant données les valeurs de vérité des propositions primitives.

Nous voyons à partir de ces définitions que si nous restreignons notre attention aux deux seules valeurs de vérité 0 et 1, nous obtenons les tables de vérité des connecteurs classiques en logique bivalente. Cela montre que L_m "généralise" le calcul propositionnel classique (notons que L_2 est le système de la logique propositionnelle classique) et que l'ensemble des tautologies de L_m est un sous-ensemble propre ($(p \rightarrow \neg p) \rightarrow \neg p$ n'est pas toujours vrai dans L_m) de l'ensemble des tautologies de la logique classique. Par conséquent, ce système est moins permissif (a moins de propositions valides - tautologies) que la logique propositionnelle classique.

Tous les autres connecteurs propositionnels peuvent être définis en termes de $\rightarrow$ et $\neg$, par exemple, nous pouvons définir la disjonction $\vee$ de manière analogue à la logique classique comme $p \vee_1 q \equiv \neg p \rightarrow q$. Cependant, cette définition de la disjonction n'est pas satisfaisante car elle ne préserve pas les propriétés souhaitables de $\vee$ dans L_m. En particulier, $p \vee_1 p$ ne prend pas toujours la

même valeur de vérité que p. De plus, afin de rendre la définition de la disjonction intuitivement plus solide, nous souhaitons que la valeur de vérité de la disjonction de deux propositions soit le maximum des valeurs de vérité de ces propositions. À cette fin, nous définissons $p \vee_2 q \equiv ((p \to q) \to q)$, ce qui satisfait cette exigence et possède les propriétés que nous voulons. Ainsi, en logique multivalente, nous avons deux types de disjonction (parfois appelées faible et forte respectivement) et il en va de même, bien sûr, pour la conjonction $\wedge$.

Une propriété caractéristique de la logique multivalente est que l'énoncé "p est vrai si et seulement si $\neg p$ est faux" ne tient pas en général. Cependant, la négation d'une proposition désignée (vraie en logique multivalente) est non désignée (c'est-à-dire jamais désignée). Afin de définir un opérateur qui se comporte comme la négation en logique classique, Rosser et Turquette ont introduit les fonctions J. Le lecteur intéressé voudra peut-être consulter leur livre classique mais assez ancien sur les logiques multivalentes.

Il existe de nombreuses axiomatisations différentes pour le système de Łukasiewicz qui ont été prouvées complètes. Certains de ces systèmes ont des axiomatisations élégantes et des preuves de complétude difficiles, tandis que d'autres ne sont pas élégants mais facilitent la preuve de complétude. Écrire des preuves dans le système de Łukasiewicz (avec une certaine axiomatisation) est, en général, plus difficile qu'en logique classique. Une raison à cela est que le théorème de déduction n'est plus valide en logique multivalente (pourquoi ?).

Enfin, nous pourrions remarquer que la logique non classique est un système complexe de nombreux systèmes logiques différents et, peut-être, qu'il est plus intéressant de voir comment ils sont connectés les uns aux autres (et à la logique classique) que de discuter lequel est le meilleur ou lequel est le bon. Nous devons également dire qu'ici nous avons restreint l'attention au calcul propositionnel des systèmes logiques discutés ; les calculs des prédicats peuvent, bien sûr, être définis pour les logiques non classiques aussi, mais pour ce faire, nous aurions besoin d'un autre article.

Nombres premiers et particules
Jack Williams, Clare

La fonction zêta de Riemann, étudiée depuis longtemps et incroyablement insaisissable, déconcerte les théoriciens des nombres depuis des siècles. Profondément liée aux nombres premiers et à leur distribution parmi les entiers, son importance en théorie des nombres est bien connue. L'hypothèse de Riemann est peut-être le problème non résolu le plus célèbre en mathématiques, notamment parce que l'Institut Clay de mathématiques a offert 1 million de dollars pour une solution.

Cependant, l'intérêt pour cette fonction mystérieuse s'étend bien au-delà des résultats ésotériques en théorie des nombres. Avec des implications pour la physique, les probabilités et les statistiques, elle est plus qu'une simple curiosité théorique des nombres. La distribution sous-jacente des zéros le long de la ligne critique $\Re(s) = \frac{1}{2}$ pénètre de nombreuses branches des mathématiques et il y a un intérêt croissant pour les connexions obscures qu'elle révèle entre ces domaines.

Lorsqu'elle est définie comme

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

pour $\Re(s) > 1$, elle semble sans lien avec les nombres premiers. Mais Euler a d'abord remarqué la factorisation

$$\zeta(s) = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \dots\right)\left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \dots\right)$$
$$\left(1 + \frac{1}{5^s} + \frac{1}{5^{2s}} + \dots\right)\dots$$

qui tient par le théorème fondamental de l'arithmétique. En sommant les séries géométriques, on obtient

$$\zeta(s) = \prod_{p \text{ premier}} \frac{1}{1 - p^{-s}}$$

Cette factorisation, qui peut être rendue rigoureuse, révèle un lien profond avec les nombres premiers.

Comme illustration du rôle de la fonction zêta en théorie des nombres, nous donnons une preuve simple du théorème célèbre d'Euclide.

Théorème Il existe une infinité de nombres premiers.

Preuve : Supposons que l'ensemble des nombres premiers soit un ensemble fini. Alors en prenant $s \to 1^+$ dans l'identité,

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \equiv \prod_{p \text{ premier}} \frac{1}{1 - p^{-s}}$$

donne

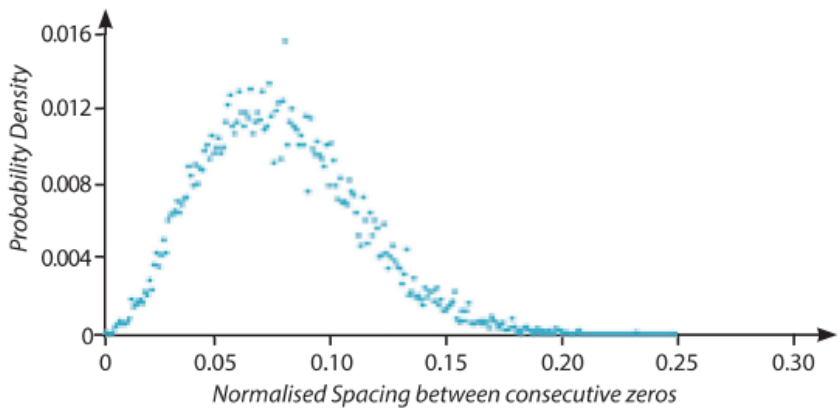
$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \text{ premier}} \frac{1}{1 - \frac{1}{p}}$$

qui doit être fini car il n'y a qu'un nombre fini de termes dans le membre de droite. Mais comme la série harmonique diverge, c'est une contradiction.

Dispersés imprudemment à travers les entiers, les nombres premiers ont été étudiés pendant des milliers d'années. Le théorème des nombres premiers de Gauss stipule que

$$\pi(n) \sim \frac{n}{\ln(n)}$$

où $\pi(n)$ est le nombre de nombres premiers inférieurs ou égaux à n. Cela signifie que la probabilité qu'un nombre soit premier est approximativement $\dfrac{n}{\ln(n)}$ pour n grand. Ici n n'a pas besoin d'être très grand pour une bonne approximation de $\pi(n)$. Pour $n = 100$, il estime environ 21 nombres premiers alors que la valeur correcte est 25.



◀ Probability density function for the normalised spacing between consecutive zeroes of the Riemann zeta-function

Fonction de densité de probabilité pour l'espacement normalisé entre zéros consécutifs de la fonction zêta de Riemann.

En 1859, Riemann publia un autre lien entre la fonction zêta et les nombres premiers, donnant une formule exacte pour $\pi(n)$, qui améliorait le théorème des nombres premiers. La formule de Riemann utilise l'emplacement des zéros dans le prolongement analytique de la fonction zêta de Riemann. Les théoriciens des nombres s'y intéressent car la distribution des zéros donne des informations sur la distribution des nombres premiers.

Alors que les nombres premiers sont plus rares parmi les grands entiers, les zéros de Riemann deviennent plus denses loin de la droite réelle. Il est possible de montrer que

$$n(T) \sim \frac{T\ln(T)}{2\pi}$$

où $n(T) = \left|\left\{y \in \mathbb{R} : \zeta\left(\frac{1}{2} + yi\right) = 0, \;\; 0 < y \leq T\right\}\right|$.

Cependant, ces zéros ne sont pas disposés régulièrement le long de la droite critique et cela a été étudié plus en profondeur de manière détaillée.

Une façon d'obtenir des informations sur cette distribution est d'étudier une distribution d'espacement, la distribution des distances entre zéros consécutifs. Ces distances peuvent être normalisées pour tenir compte de la distribution "globale" donnée par $n(T)$. La normalisation correcte est $\widehat{y}_j = \dfrac{y_j \ln(y_j)}{2\pi}$, ce qui donne la distribution de probabilité présentée sur le graphique ci-avant, estimée numériquement en utilisant les zéros numérotés de $10^{21} + 1$ à $10^{21} + 10000$.

De manière surprenante, cette distribution apparaît naturellement dans d'autres domaines des mathématiques et de la physique et semble fondamentale pour de nombreux problèmes apparemment sans rapport. Des expériences ont été réalisées dans lesquelles des neutrons sont diffusés par un noyau lourd. La section efficace résultante contient des pics, les résonances de diffusion, et des creux. Si l'énergie d'un neutron est proche de l'un des pics, il est repoussé par le noyau et s'il est proche d'un creux, il peut passer sans effort. Ce comportement étrange conduit à différentes diffusions. Bien que ces problèmes soient trop difficiles à résoudre analytiquement ou numériquement, des données empiriques peuvent être obtenues et une analyse similaire à la séparation des zéros peut être effectuée sur ces pics. Étonnamment, la densité de probabilité correspond à celle des zéros de Riemann. En fait, la densité ne dépend pas beaucoup du noyau particulier utilisé. Cette connexion remarquable est encore mal comprise.

La structure mathématique qui modélise les résonances de diffusion est aussi assez inattendue. Dans les années 1950, Eugene Wigner a proposé un modèle statistique basé sur des matrices aléatoires. Inspiré par la formulation de la mécanique quantique de Heisenberg, dans laquelle les propriétés d'un atome ou d'un noyau peuvent être décrites par une matrice hermitienne, il a avancé la conjecture des matrices aléatoires. Il a suggéré que ces pics suivent le même espacement que les valeurs propres de matrices hermitiennes dont les éléments sont choisis à partir d'une certaine distribution de probabilité, généralement normale. Les valeurs propres de la matrice correspondent aux niveaux d'énergie du spectre. En utilisant de telles matrices aléatoires, on peut obtenir des espacements statistiquement similaires à la fois aux données de diffusion de neutrons et à la distribution des zéros de la fonction zêta de Riemann.

Bien qu'il soit possible que les zéros jouent un rôle direct dans les systèmes physiques, la connexion inattendue entre les nombres premiers, les matrices et la diffusion de neutrons pourrait refléter un résultat plus profond, unissant de nombreux domaines apparemment disjoints des mathématiques. Des progrès ont été réalisés dans cette direction. Des théorèmes ont été prouvés montrant qu'il existe une unique distribution limite pour le spectre des valeurs propres provenant d'un grand ensemble de matrices aléatoires. Semblables d'esprit au théorème central limite, ces résultats reflètent l'universalité intrinsèque aux résultats qui connectent de nombreux domaines des mathématiques.

Le théorème fondamental de l'algèbre

Ian Stewart FRS, Professeur émérite de mathématiques, Université de Warwick

Le théorème fondamental de l'algèbre stipule que tout polynôme non constant $p(z)$ sur $\mathbb{C}$ a au moins un zéro ; c'est-à-dire $p(z_0) = 0$ pour un certain $z_0 \in \mathbb{C}$. Cela implique facilement que si le degré de p est n, alors il y a n zéros, à condition que les zéros multiples soient comptés correctement. En effet, $z - z_0$ divise $p(z)$, et le quotient a pour degré $n-1$, donc nous pouvons procéder par induction.

Ce résultat a été largement utilisé par Euler, Lagrange et d'autres, qui ont offert diverses preuves "faites à la main". La première preuve rigoureuse a été donnée par Gauss dans sa thèse de doctorat. Elle impliquait la manipulation de séries trigonométriques compliquées pour dériver une contradiction, et était loin d'être transparente. L'idée sous-jacente peut être reformulée en termes topologiques, impliquant le nombre d'enroulement d'une courbe autour d'un point, voir [1].

1. Preuves modernes

D'autres preuves classiques utilisent des résultats profonds d'analyse complexe, tels que le théorème de Liouville : une fonction bornée et analytique sur tout le plan complexe est constante. Cela dépend de la formule intégrale de Cauchy et prend la majeure partie d'un cours d'analyse complexe à prouver. Voir [3], 2.51. Une approche alternative utilise le théorème de Rouché, voir par exemple [3], 3.44. Une autre preuve - la première qu'on m'a montrée en tant qu'étudiant - utilise le théorème du maximum du module : si une fonction analytique n'est pas constante, alors la valeur maximale de son module sur un ensemble arbitraire se produit sur la frontière de cet ensemble.

Carl Friedrich Gauss (1777-1855), mathématicien et scientifique allemand. Parfois appelé Princeps mathematicorum (en latin, "le plus grand (le leader, le chef) des mathématiciens"), il est souvent considéré comme l'un des plus grands mathématiciens de l'histoire. On trouve dans une note célèbre écrite dans le journal de Gauss après qu'il eut prouvé que tout entier positif pouvait être exprimé comme la somme de trois nombres triangulaires la formule :

$$\text{“EYPHKA ! num} = \Delta + \Delta + \Delta\text{”}$$

Une variante utilise le théorème du minimum du module (la valeur minimale de son module sur un ensemble arbitraire est soit zéro, soit elle se produit sur la frontière de cet ensemble). Voir [2], Théorèmes 10.14, 10.15. L'approche d'Euler, qui annule les parties réelle et imaginaire de $p(z)$ et prouve que les courbes résultantes dans le plan doivent se croiser, peut être rendue rigoureuse. Clifford a donné une preuve basée sur l'induction sur la puissance de 2 qui divise le degré n, ce qui s'explique le plus facilement à l'aide de la théorie de Galois.

2. Une preuve élémentaire

Toutes ces preuves sont assez sophistiquées. Mais il y a une manière plus simple. Il y a quelques années, j'ai trouvé une preuve simple utilisant quelques idées de topologie élémentaire des ensembles de points et des estimations du type que nous rencontrons tôt dans tout cours d'analyse réelle. J'ai

rapidement découvert qu'elle était déjà connue des experts : on peut la trouver sur Wikipédia, par exemple. Mais elle mérite d'être plus largement connue, car elle est simple et va droit au cœur du sujet. Les faits nécessaires peuvent être prouvés directement par des moyens élémentaires, et ils auraient été considérés comme évidents avant que les mathématiciens ne commencent à se soucier de la rigueur en analyse, vers 1850.

L'idée peut être résumée en quelques lignes. Supposons par l'absurde que $p(z)$ ne soit jamais nul. Alors $|p(z)|^2$ a une valeur minimale non nulle et atteint ce minimum en un certain point $w \in \mathbb{C}$. Considérons des points v sur un petit cercle centré en w, et montrons que $|p(v)|^2$ doit être inférieur à $|p(w)|^2$ pour un certain v. Contradiction.

Voici les détails :

Théorème 1 : Si $p(z)$ est un polynôme non constant sur $\mathbb{C}$, alors il existe $z_0 \in \mathbb{C}$ tel que $p(z_0) = 0$.

Preuve : Supposons par l'absurde qu'un tel z_0 n'existe pas. Pour un certain $R > 0$, l'ensemble $D = \{z : |p(z)|^2 \leq R\}$ est non vide. L'application $\psi : \mathbb{C} \to \mathbb{R}^+$ définie par $\psi(z) = |p(z)|^2$ est continue, donc $D = \psi([0, R])$ est compact. Pour un sous-ensemble de $\mathbb{C}$, cela équivaut à être fermé et borné. Il s'ensuit que $|p(z)|^2$ atteint sa valeur minimale sur D. Par la définition de D, c'est aussi sa valeur minimale sur $\mathbb{C}$.

Supposons que ce minimum soit atteint en $w \in \mathbb{C}$. Alors $|p(z)|^2 \geq |p(w)|^2 \ \forall z \in \mathbb{C}$ et par hypothèse $p(w) \neq 0$.

Nous considérons maintenant $|p(z)|^2$ lorsque z parcourt un petit cercle centré en w, et nous en déduisons une contradiction.

Soit $h \in \mathbb{C}$. Développons $p(w + h)$ en puissances de h pour obtenir

$$p(w + h) = p_0 + p_1 h + p_2 h^2 + \cdots + p_n h^n \tag{1}$$

où n est le degré de p. Ici les p_j sont des nombres complexes spécifiques. Ce sont en fait les coefficients de la série de Taylor $p_j = \dfrac{p^{(j)}(w)}{j!}$ mais nous n'avons pas réellement besoin de l'utiliser, et (1) peut être prouvé algébriquement sans difficulté.

Clairement $p_0 = p(w)$, et nous supposons qu'il est non nul, donc $p_0 \neq 0$. Si $p_1 = p_2 = \ldots = p_n = 0$ alors $p(z) = p_0$ est constant, contrairement à l'hypothèse. Donc un certain $p_j \neq 0$. Soit m le plus petit entier ≥ 1 pour lequel $p_m \neq 0$. Dans (1), posons $h = \epsilon e^{i\theta}$ pour un petit $\epsilon > 0$.

$$p(w + \epsilon e^{i\theta}) = p_0 + p_m \epsilon^m e^{mi\theta} + \mathrm{O}(\epsilon^{m+1})$$

Par conséquent

$$|p(w + \epsilon e^{i\theta})|^2 = |p_0 + p_m \epsilon^m e^{mi\theta}|^2 + O(\epsilon^{m+1}) = p_0 \overline{p}_0 + \overline{p}_0 p_m \epsilon^m e^{mi\theta} + p_0 \overline{p}_m \epsilon^m e^{-mi\theta} + \mathrm{O}(\epsilon^{m+1})$$

Soit $p_0\overline{p}_m = re^{i\phi}$ pour $r \geq 0$. Puisque $p_0 \neq 0$ et $p_m \neq 0$ nous avons $r > 0$. En posant $h = 0$, nous voyons que $p_0\overline{p}_0 = |p(w)|^2$. Maintenant

$$\begin{aligned} |p(w + \epsilon e^{i\theta})|^2 &= p_0\overline{p}_0 + re^{i\phi}\epsilon^m e^{mi\theta} + re^{-i\phi}\epsilon^m e^{-mi\theta} + \mathrm{O}(\epsilon^{m+1}) \\ &= |p(w)|^2 + 2\epsilon^m r\cos(m\theta + \phi) + \mathrm{O}(\epsilon^{m+1}) \end{aligned}$$

Posons $\theta = \dfrac{1}{m}(\phi - \pi)$, de sorte que $\phi = \pi - m\theta$. Alors $\cos(m\theta + \phi) = \cos(\pi) = -1$, et

$$|p(w + \epsilon e^{i\theta})|^2 = |p(w)|^2 - 2\epsilon^m r + \mathrm{O}(\epsilon^{m+1})$$

Mais $\epsilon, r > 0$, donc pour ϵ suffisamment petit nous avons

$$|p(w + \epsilon e^{i\theta})|^2 < |p(w)|^2$$

contredisant la définition de w. Par conséquent, il existe $z_0 \in \mathbb{C}$ tel que $p(z_0) = 0$.

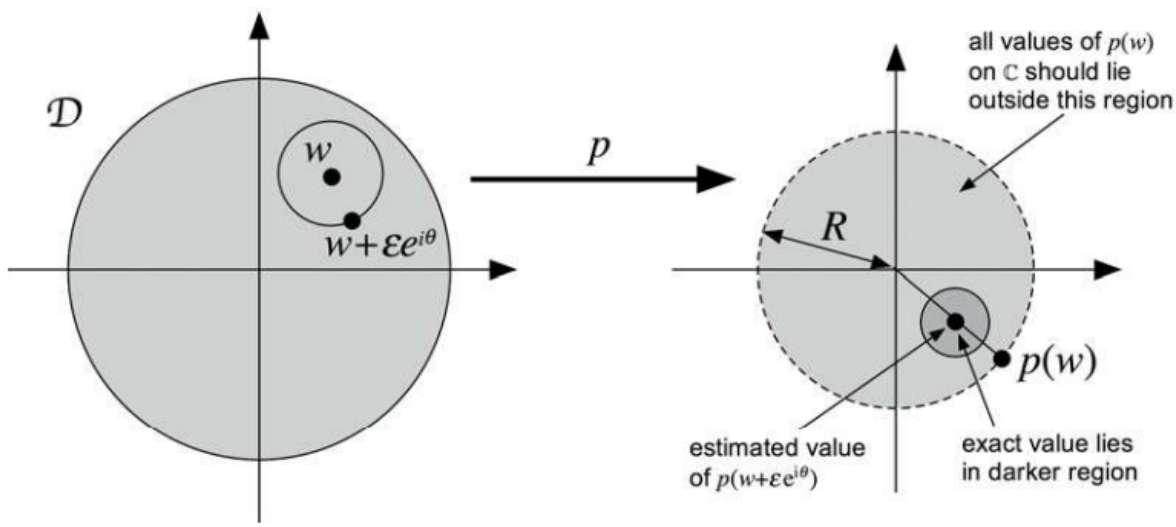


FIGURE 1 : Idée de la preuve

La même idée peut être adaptée pour donner une preuve tout aussi simple du théorème de Liouville :

Théorème 2 : Si $f(z)$ est analytique sur tout le plan complexe, et n'est pas constante, alors $f(z_0) = 0$ pour un certain $z_0 \in \mathbb{C}$.

La seule nouveauté dans la preuve est que le polynôme en (1) devient une série entière, et maintenant nous avons vraiment besoin du théorème de Taylor.

3. Références

[1] Ian Stewart ; 1977 ; Gauss ; Scientific American 237 122-131.

[2] Ian Stewart et David O. Tall ; 1983 ; Complex Analysis ; Cambridge University Press ; Cambridge.

[3] Edward C. Titchmarsh ; 1939 ; The Theory of Functions ; Clarendon press ; Oxford.

Sommes, produits et sommes-et-produits

1. Sommes

Supposons que nous partitionnions les nombres naturels en un nombre fini de classes. Pouvons-nous toujours trouver x et y tels que x, y et $x+y$ soient tous dans la même classe ? De manière équivalente, si nous "colorons de manière finie" les naturels, pouvons-nous trouver x, y, z de la même couleur avec $x+y=z$? C'est une question typique de la théorie de Ramsey, qui cherche à répondre à des questions de la forme "pouvons-nous trouver un certain ordre dans suffisamment de désordre ?" Dans ce cas, le désordre est la coloration inconnue, et la petite parcelle d'ordre est $x, y, x+y$. (Il vaut la peine de mentionner en passant que le lecteur peut déduire de la formulation de la question que nous ne considérons pas 0 comme un nombre naturel. Il serait plus intéressant de demander "sommes-nous autorisés à prendre $x=y$?", mais en fait il est assez facile de trouver des colorations qui garantissent que nous ne pouvons pas avoir $x=y$, donc en réalité cela ne fait aucune différence que nous insérions ou non le mot "distinct").

Par exemple, si nous colorons chaque nombre pair en rouge et chaque nombre impair en bleu, alors nous pourrions prendre $x=4$ et $y=6$. Si nous colorons chaque carré en rouge et chaque non-carré en bleu, nous pourrions prendre $x=5$ et $y=6$ (ou même $x=9$ et $y=16$). Dans tout exemple particulier, il peut être assez facile de trouver un x et un y convenables, mais est-ce toujours le cas ? Il s'avère que la réponse est oui. C'est ce qu'on appelle le théorème de Schur, datant de 1916. Le théorème de Schur n'est pas trop difficile à déduire du théorème de Ramsey - en fait, il apparaît souvent sur les feuilles d'exercices du cours de théorie des graphes de deuxième année, dans la section sur la théorie de Ramsey.

Qu'en est-il des sommes de plus de termes ? Qu'en est-il des sept sommes non nulles à partir de x, y, z ? Et en général, qu'en est-il de $FS(x_1, \ldots, x_n)$, c'est-à-dire l'ensemble de toutes les sommes $\Sigma_i x_i$, où I est un sous-ensemble non vide de l'ensemble d'indices ? C'est beaucoup plus difficile : on ne trouve un tel problème sur aucune feuille de deuxième année dont j'ai connaissance. Il est vrai, cependant, que (pour tout n) chaque fois que les naturels sont colorés de manière finie, il existe $x_1, \ldots, x_n$ tels que $FS(x_1, \ldots, x_n)$ soit d'une seule couleur. C'est ce qu'on appelle le théorème des sommes finies, et c'est un cas particulier d'un célèbre théorème de Rado de 1933.

Frank Ramsey (1903-1930) était un mathématicien, philosophe et économiste brillant. Le théorème de Ramsey fut en fait prouvé par Ramsey comme un lemme mineur sur le chemin d'un résultat en logique. Il souffrait de problèmes hépatiques chroniques et mourut de jaunisse à l'âge de seulement 26 ans.

2. Produits

Qu'en est-il des produits ? Pouvons-nous toujours trouver, dans une coloration finie des naturels, trois nombres x, y, z de la même couleur avec $xy=z$ (à part $x=y=z=1$, bien sûr) ? En fait, il s'avère que c'est une question stupide : cela découle immédiatement du théorème de Schur, simplement en restreignant notre attention aux puissances de 2. En effet, étant donné une coloration

finie des naturels, considérons une "nouvelle" coloration où la nouvelle couleur de x est la couleur originale de 2^x. Pour cette nouvelle coloration, le théorème de Schur nous dit que nous pouvons trouver $x, y, x+y$ de la même couleur, et cela se traduit par $2^x, 2^y, 2^{x+y}$ ayant la même couleur dans la coloration originale, comme requis.

Et de même, bien sûr, pour les produits de plus de facteurs. Nous pouvons toujours trouver un ensemble de la forme $FP(x_1, \ldots, x_n)$ (signifiant tous les produits non vides) qui est d'une seule couleur, par exactement le même argument.

3. Sommes et produits

Jusqu'ici tout va bien. Nous avons traité les sommes, et les produits. Qu'en est-il de leurs combinaisons ? Est-il donc vrai que chaque fois que les entiers naturels sont colorés de manière finie, nous pouvons toujours trouver x et y tels que les quatre nombres $x, y, x+y, xy$ soient de la même couleur ? Plus généralement, bien sûr, nous aimerions trouver (pour tout n donné) des nombres $x_1, \ldots, x_n$ tels que $FS(x_1, \ldots, x_n) \cup FP(x_1, \ldots, x_n)$ (en d'autres termes : toutes les sommes et tous les produits) soit d'une seule couleur. En fait, plus généralement encore, on aimerait pouvoir itérer la sommation et le produit (sans répéter un terme ou un facteur) : ainsi par exemple pour trois variables on aimerait vraiment avoir $x, y, z, x+y, x+z, y+z, xy, xz, yz, x+y+z, xyz$ et aussi $xy+z, xz+y, yz+x$ et $x(y+z), y(x+z), z(x+y)$.

La réponse est : personne ne sait. C'est un problème ouvert. On y a beaucoup réfléchi, mais sans succès. Remarquablement, en laissant de côté tout ce "ce que nous voudrions vraiment" , c'est même un problème ouvert pour le cas de seulement deux termes ! En d'autres termes, on ne sait pas si, chaque fois que les naturels sont colorés de manière finie, il existe x et y tels que tous les $x, y, x+y, xy$ soient de la même couleur. Il semble tout à fait scandaleux que ce tout petit cas de départ reste non résolu. En fait, personne n'a même d'idées ou d'intuitions. Personne n'a jamais proposé d'"idée de preuve" ou de "schéma de preuve" qui semble, ne serait-ce qu'un instant, avoir la moindre chance de fonctionner. Dans l'autre sens, personne n'a jamais trouvé de coloration qui pourrait, même avec 5 secondes de réflexion, avoir le potentiel d'être un contre-exemple. Pour le dire autrement, pour toute coloration réelle qu'on vous donne, il est invariablement très facile de trouver de tels x et y.

Une partie de la raison du manque de progrès vers une preuve pourrait être ce qu'on pourrait appeler un "manque de portabilité". Si nous prouvons un énoncé sur les sommes (comme par exemple l'énoncé que si nous colorons les nombres de 1 à 100 avec 2 couleurs, alors nous pouvons trouver une solution à $x+y=z$ dans une classe de couleur), alors nous pouvons le répéter : si nous regardons les nombres pairs de 2 à 200, nous trouverons le même objet (en langage prétentieux, il y a plein d'homomorphismes du groupe $\mathbb{Z}$ vers lui-même). Cette caractéristique de "répétition" s'avère être d'une grande importance dans la plupart des preuves de résultats sur les sommes. Et de même pour les produits : si nous savons quelque chose sur les produits à l'intérieur des nombres de 1 à 100, alors nous savons le même énoncé sur l'ensemble des nombres carrés de 1^2 à 100^2 (là encore, il y a une manière prétentieuse de le dire). Mais, hélas, il n'y a aucun moyen de transférer les résultats sur les sommes-et-produits d'un endroit à un autre - il n'y a pas trop d'homomorphismes d'anneaux de $\mathbb{Z}$ vers lui-même !

Il est également possible de donner une raison pour laquelle les mauvaises colorations sont si difficiles à concevoir. Pour trouver une coloration sans ces structures de sommes-et-produits, il faudrait une coloration qui “s’harmonise bien” avec l’addition et la multiplication. Or, il y a beaucoup de colorations qui s’harmonisent bien avec l’addition (par exemple, la coloration par valeur modulo quelque chose, ou la coloration par le chiffre non nul le moins significatif en base quelque chose), et de même pour la multiplication - mais personne n’a trouvé de coloration qui s’harmonise avec les deux.

J’ai gardé le “manque de connaissance” le plus embarrassant pour la fin. Nous avons dit que, même dans le cas le plus simple possible de “deux termes”, c’est-à-dire $x, y, x + y, xy$, la réponse n’est pas connue. Mais que se passe-t-il si nous allons encore plus loin, et ignorons la couleur de x et y eux-mêmes ? La question devient alors : est-il vrai que, chaque fois que les entiers naturels sont colorés de manière finie, il existe x et y tels que les deux nombres $x+y$ et xy aient la même couleur l’un que l’autre ? (Nous excluons le cas $x = y = 2$, ou de manière équivalente nous insistons pour que x et y soient distincts).

Même ce cas super-spécial et minuscule est inconnu. Il peut être vérifié par ordinateur pour un petit nombre de couleurs (je pense que le résultat est connu pour 5 couleurs ou moins), mais ce n’est pas une grande preuve en sa faveur. C’est l’un des problèmes les plus exaspérants, et les plus alléchants, de toute la théorie de Ramsey !

Oh mon Dieu, c'est couvert de règles 30 !
Stephen Wolfram

On se reportera aux images très nombreuses de l'article en anglais dans le numéro 65 de la revue Eureka.

Il y a une semaine [21 mai 2017], une nouvelle gare, nommée "Cambridge North", a ouvert à Cambridge, au Royaume-Uni. Normalement, un tel événement serait bien en dehors de ma sphère de conscience. (Je pense que la dernière fois que j'ai pris un train pour Cambridge, c'était en 1975.) Mais la semaine dernière, les gens ont commencé à m'envoyer des photos de la nouvelle gare, se demandant si je pouvais identifier le motif qui s'y trouve.

Et, oui, cela ressemble en effet beaucoup aux motifs que j'ai passés des années à étudier - qui proviennent de programmes simples dans l'univers computationnel. Mes premiers - et toujours préférés - exemples de programmes simples sont les automates cellulaires unidimensionnels comme celui-ci.

Le système évolue ligne par ligne à partir du haut, déterminant la couleur de chaque cellule selon la règle en dessous. Cet automate cellulaire particulier, je l'ai appelé "règle 182", parce que le motif de bits dans la règle correspond au nombre 182 en binaire. Il y a en tout 256 automates cellulaires possibles comme celui-ci, et voici ce que tous font.

Beaucoup d'entre eux montrent un comportement assez simple. Mais l'énorme surprise que j'ai eue lorsque j'ai exécuté tous ces automates cellulaires au début des années 1980 est que, même si toutes les règles sont très simples à énoncer, certaines d'entre elles génèrent un comportement très complexe. Le premier dans la liste à faire cela - et toujours mon exemple préféré - est la règle 30.

Si on l'exécute pendant 400 étapes, on obtient ceci.

Et, oui, il est remarquable qu'en partant d'une seule cellule noire en haut, et en suivant simplement une règle simple de manière répétée, il soit possible d'obtenir toute cette complexité. Je pense que c'est en fait un exemple d'un phénomène extrêmement important, qui est central dans la façon dont la complexité se crée dans la nature, ainsi que dans la façon dont nous pouvons obtenir un nouveau niveau de technologie. Et en fait, je pense que c'est suffisamment important pour que j'aie passé plus d'une décennie à écrire un livre de 1200 pages (qui vient de célébrer son 15e anniversaire) basé sur cela.

Et pendant des années, j'ai en fait eu la règle 30 sur mes cartes de visite.

Mais revenons à la gare de Cambridge North. Son motif n'est manifestement pas complètement aléatoire. Mais s'il a été fait par une règle, quel genre de règle ? Pourrait-ce être un automate cellulaire ?

J'ai zoomé sur une photographie du motif.

Soudain, quelque chose m'a semblé terriblement familier : les triangles, les rayures, les formes en L. Attendez une minute... ne serait-ce pas en fait ma règle préférée de tous les temps, la règle 30 ?

De toute évidence, le motif est incliné à 45° par rapport à la façon dont je présenterais habituellement un automate cellulaire. Et il y a des triangles noirs sur la photographie, pas des blancs comme dans la règle 30. Mais si on inverse le noir et blanc de la règle (c'est donc maintenant la règle 135), on obtient ceci.

Et, oui, c'est le même genre de motif que sur la photographie ! Mais si c'est la règle 30 (ou la règle 135), quelle est sa condition initiale ? La règle 30 peut en fait être utilisée comme cryptosystème - parce qu'il peut être difficile (peut-être même NP-complet) de reconstruire sa condition initiale.

Mais, OK, si c'est ma règle préférée, je me suis demandé si c'était peut-être aussi ma condition initiale préférée - une seule cellule noire. Et, oui, c'est le cas ! Le motif de la gare provient exactement du bord droit (inversé) de mon motif préféré de la règle 30 !

Voici le code Wolfram Language. D'abord exécuter l'automate cellulaire, puis faire pivoter le motif.

C'est un peu plus délicat d'extraire précisément la section du motif qui est utilisée. Voici le code (le PlotRange est ce qui détermine la partie du motif qui est montrée).

OK, alors où ce motif est-il réellement utilisé à la gare ? Partout !

Il est fait d'aluminium perforé. On peut réellement regarder à travers, comme on le faisait jadis au travers d'une fenêtre à treillis. De l'intérieur, le motif est inversé gauche-droite - donc si c'est la règle 135 de l'extérieur, c'est la règle 149 de l'intérieur. Et la nuit, le motif est inversé noir-blanc, parce qu'il y a de la lumière venant de l'intérieur - donc de l'extérieur c'est "règle 135 le jour, et règle 30 la nuit".

Quels sont les faits connus au sujet du motif de la règle 30 ? Il est extrêmement difficile de prouver rigoureusement des choses à son sujet (et c'est intéressant en soi - et étroitement lié au phénomène fondamental de l'irréductibilité computationnelle). Mais, par exemple - comme, disons, les chiffres de π - de nombreux aspects semblent aléatoires. Et, par exemple, les carrés noirs et blancs semblent apparaître avec une fréquence égale - ce qui signifie qu'à la gare, les panneaux laissent passer environ 50% de la lumière extérieure.

Si l'on regarde des séquences de n cellules, il semble que toutes les 2^n configurations apparaîtront en moyenne avec une fréquence égale. Mais tout n'est pas aléatoire. Et ainsi, par exemple, si l'on regarde des blocs de 3×2 cellules, seulement 24 des 32 motifs possibles apparaissent. (Peut-être que certains personnes attendant leur train découvriront quels blocs manquent...).

Lorsque nous regardons le motif, notre système visuel repère particulièrement les triangles noirs. Et, oui, il semble que des triangles de n'importe quelle taille puissent finalement apparaître, bien qu'avec une fréquence décroissant exponentiellement avec la taille.

Si l'on regarde attentivement le bord droit du motif de la règle 30, on peut voir qu'il se répète. Cependant, la période de répétition semble augmenter exponentiellement à mesure qu'on s'éloigne du bord.

À la gare, il y a beaucoup de panneaux identiques. Mais la règle 30 est en fait une source inépuisable de nouveaux motifs. Alors que se passerait-il si l'on continuait simplement l'évolution, et qu'on la rendait sur des panneaux successifs ? Voici le résultat. C'est dommage pour la pointe de périodicité sur le bord droit, et le grand triangle sur le panneau 5 (qui pourrait être un problème de sécurité à la gare).

Quinze étapes de plus en partant du bord, il n'y a plus aucune trace de cela.

Qu'en est-il des autres conditions initiales ? Si les conditions initiales se répètent, alors le motif aussi. Mais sinon, pour autant qu'on puisse en juger, le motif semblera essentiellement le même qu'avec une condition initiale à cellule unique. On peut aussi essayer d'autres règles. En voici quelques-unes du même ensemble de 256 règles plus simples que la règle 30.

En s'éloignant plus profondément du bord, les résultats semblent un peu différents (pour les aficionados, la règle 89 est une version transformée de la règle 45, la règle 182 de la règle 90, et la règle 193 de la règle 110).

Et en partant de conditions initiales aléatoires, plutôt que d'une seule cellule noire, les choses semblent à nouveau différentes.

Et voici quelques règles supplémentaires, démarrées à partir de conditions initiales aléatoires.

Voici un site web (fait en quelques minutes avec un petit bout de code Wolfram Language) qui vous permet d'expérimenter (y compris avec des numéros de règle plus grands, basés sur des règles à plus longue portée). (Et si vous voulez explorer plus systématiquement, voici un Wolfram Notebook à essayer).

C'est incroyable ce qui existe dans l'univers computationnel des programmes possibles. Il y a une gamme infinie de motifs possibles. Mais c'est cool que la gare de Cambridge North utilise ma découverte préférée de tous les temps dans l'univers computationnel - la règle 30 ! Et cela rend très bien !

1. La vue d'ensemble

Il y a quelque chose de curieusement intemporel dans les formes générées algorithmiquement. Un dodécaèdre de l'Égypte ancienne semble encore net et moderne aujourd'hui. Tout comme les pavages périodiques - ou les formes imbriquées - même de siècles passés.

Mais peut-on générer des formes plus riches algorithmiquement ? Avant de découvrir la règle 30, j'avais toujours supposé que toute forme générée à partir de règles simples finirait toujours par être manifestement simple. Mais la règle 30 a été un grand choc pour mon intuition - et à partir de là, j'ai réalisé qu'en fait, dans l'univers computationnel de toutes les règles possibles, il est en réalité très facile d'obtenir un comportement riche et complexe, même à partir de règles sous-jacentes

simples.

Et de plus, les motifs qui sont générés ont souvent un intérêt visuel remarquable. En voici quelques-uns produits par des automates cellulaires (maintenant avec 3 couleurs possibles pour chaque cellule, plutôt que 2).

Il y a une diversité étonnante de formes. Et, oui, elles sont souvent compliquées. Mais parce qu'elles sont basées sur des règles sous-jacentes simples, elles ont toujours une certaine logique : en un sens, chacune raconte une "histoire algorithmique" définie.

Une chose notable à propos des formes que nous voyons dans l'univers computationnel est qu'elles ressemblent souvent beaucoup aux formes que nous voyons dans la nature. Et je ne pense pas que ce soit une coïncidence. Au contraire, je pense que ce qui se passe est que les règles de l'univers computationnel capturent l'essence des lois qui gouvernent de nombreux systèmes dans la nature - que ce soit en physique, en biologie ou ailleurs. Et peut-être qu'il y a une certaine familiarité ou un certain confort associé aux formes de l'univers computationnel qui vient de leur similitude avec les formes auxquelles nous sommes habitués dans la nature.

Mais ce que nous obtenons de l'univers computationnel est-il de l'art ? Lorsque nous choisissons quelque chose comme la règle 30 pour un usage particulier, ce que nous faisons est conceptuellement un peu comme de la photographie : nous ne créons pas les formes sous-jacentes, mais nous sélectionnons celles que nous choisissons d'utiliser.

Dans l'univers computationnel, cependant, nous pouvons être plus systématiques. Étant donné un critère esthétique, nous pouvons automatiquement parcourir peut-être même des millions ou des milliards de règles possibles pour trouver les optimales : en un sens, "découvrir automatiquement de l'art" dans l'univers computationnel.

Nous avons fait une expérience sur ce sujet pour la musique en 2007 : WolframTones. Et ce qui est remarquable, c'est que même en échantillonnant un assez petit nombre de règles (des automates cellulaires, en l'occurrence), nous sommes capables de produire toutes sortes de courts morceaux de musique intéressants - qui semblent souvent remarquablement "créatifs" et "inventifs".

D'un point de vue pratique, la découverte automatique dans l'univers computationnel est importante car elle permet la personnalisation de masse. Elle facilite le fait d'être "original" (et "créatif") - et de trouver quelque chose de différent à chaque fois, ou de s'adapter à des contraintes qui n'ont jamais été vues auparavant (par exemple, un motif dans une région géométrique compliquée).

La gare de Cambridge North utilise une règle particulière de l'univers computationnel pour faire ce qui équivaut à un motif ornemental. Mais on peut aussi utiliser des règles de l'univers computationnel pour d'autres choses en architecture. Et on peut même imaginer un bâtiment dans lequel tout - de la masse globale aux détails des moulures - est complètement déterminé par quelque chose de proche d'une règle unique.

On pourrait supposer qu'un tel bâtiment serait en quelque sorte minimaliste et stérile. Mais le fait

remarquable est que cela n'a pas besoin d'être vrai - et qu'au contraire, il y a plein de formes riches, presque "organiques" à "extraire" de l'univers computationnel.

Depuis que j'ai commencé à écrire sur les automates cellulaires unidimensionnels au début des années 1980, toutes sortes d'œuvres d'art intéressantes ont été réalisées avec eux. Beaucoup de règles différentes ont été utilisées. Parfois, c'étaient ce que j'appelais des règles de "classe 4" qui ont un aspect particulièrement organique. Mais souvent, c'étaient d'autres règles - et la règle 30 a certainement fait sa part d'apparitions - que ce soit sur des sols, des chemises, des couvre-théières, des installations cinétiques, ou, récemment, des écharpes personnalisées en masse (avec la machine à tricoter exécutant réellement l'automate cellulaire).

Mais aujourd'hui, nous célébrons une manifestation nouvelle et différente de la règle 30. Formée de panneaux d'aluminium permanents, dans une ancienne ville universitaire, un merveilleux coin de l'univers computationnel orne l'une des structures les plus pratiques : une petite gare. Mes compliments aux architectes. Puisse ce qu'ils ont fait donner à des générations de voyageurs ferroviaires un petit aperçu des merveilles de l'univers computationnel. Et peut-être que quelques-uns, faisant écho aux derniers mots attribués au voyageur dans le film 2001 : L'Odyssée de l'espace, s'exclameront "oh mon Dieu, c'est couvert de règles 30 !".

Pourquoi les séries sont-elles musicales ?
Blanche Descartes

La plupart des mathématiciens connaissent la théorie du jeu de Nim, décrite dans les livres de récréations mathématiques. Mais peu semblent connaître la remarquable généralisation du Dr P. M. Grundy, publiée dans Eureka, 2, 6-8 (1939). Considérons un jeu Γ dans lequel 2 joueurs jouent alternativement, et le dernier joueur gagne (en se déplaçant vers une "position terminale"). Définissons inductivement une fonction $G(P)$ de la position P ($\Omega(P)$ dans l'article de Grundy) comme suit :

Si P est terminale, $G(P) = 0$. S'il y a des coups permis de P vers Q, de P vers R, de P vers S, et ainsi de suite, alors $G(P)$ est le plus petit entier non négatif différent de tous les $G(Q), G(R), G(S)$, etc.

Il s'ensuit que si $0 \leq r < G(P)$, il y a un coup de P vers un certain R avec $G(R) = r$, mais aucun coup vers une position U avec $G(U) = G(P)$. Si les positions P avec $G(P) = 0$ sont appelées "sûres", la stratégie gagnante est de se déplacer toujours vers une position sûre : soit celle-ci est terminale, et gagne immédiatement, soit l'adversaire se déplace vers une position non sûre et le cycle se répète.

Imaginons maintenant les joueurs s'engageant dans une "démonstration simultanée" de k jeux $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$ de ce type, la règle étant que chaque joueur à son tour fait un coup dans un et un seul jeu, ou s'il ne peut pas jouer dans aucun jeu, il perd. Soient $P_1, P_2, \ldots, P_k$ les positions dans les jeux respectifs $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$. Alors le théorème de Grundy stipule que :

1. cette position combinée est sûre si et seulement si k tas de $G(P_1), G(P_2), \ldots, G(P_k)$ jetons respectivement forment une combinaison sûre au Nim ;
2. plus généralement, la fonction G de la position combinée est la "nim-somme" des $G(P_s)$ séparés (c'est-à-dire obtenue en écrivant les $G(P_s)$ en base 2 et en additionnant les colonnes modulo 2).

Car aucun joueur ne peut tirer un avantage en jouant de manière à augmenter un $G(P_s)$, puisque l'adversaire peut rétablir le statu quo. Et si l'on ne considère que les diminutions de $G(P_s)$, le jeu est identique au Nim, ce qui prouve l'assertion (1). Par conséquent, $G(P) = g$ si et seulement si la position combinée (P, P') est sûre, où $G(P') = g$. De là, (2) s'ensuit assez facilement.

Il s'ensuit que nous pouvons analyser complètement tout jeu combiné de ce type, à condition de pouvoir trouver les $G(P_s)$ pour les positions composantes. Le Nim est un exemple ; un tas H_x de x jetons constitue une position composante, puisque chaque joueur à son tour ne modifie qu'un seul tas, et $G(H_x) = x$. De nombreuses variantes du Nim sont analysées de la même manière. Moins trivial est le jeu de Grundy, dans lequel tout tas est divisé en deux parties inégales (non vides). Ainsi les tas de 1, 2 sont terminaux, avec $G = 0$, un tas de 3 ne peut être divisé qu'en $2 + 1$, qui est terminal, donc $G(H_3) = 1$. En général, $G(H_x)$ dans le jeu de Grundy est le plus petit entier ≥ 0 différent de toutes les nim-sommes de $G(H_y)$ et $G(H_{x-y})$, $0 < y < x/2$. La série donne :

$$x = 0123456789101112,$$

$$G(H_x) = 0001021021021,$$

continuant avec 3, 2, 1, 3, 2, 4, 3, 0, 4, 3, 0, 4, 3, 0, 4, 1, 2, 3, 1, 2, 4, 1, 2, 4, 1, 2, ... Cette curieuse "série quelque peu périodique" semble essayer d'avoir une période 3, mais avec des sauts

qui se produisent continuellement. M. Richard K. Guy a confirmé cela jusqu'à $x = 300$. Il a suggéré qu'elle pourrait être jouée au piano, en prenant 0 pour le do du milieu, 1 = ré, 2 = mi, etc. Le sens profond devint alors évident :

Guy a également travaillé avec des rangées R_x de x jetons, dans lesquelles certains ensembles de jetons consécutifs pouvaient être extraits (laissant ainsi éventuellement deux rangées plus courtes, une de chaque côté de l'ensemble extrait). Dans son jeu "6", n'importe quel jeton peut être retiré, sauf un R_1 (= un seul jeton isolé). La série $G(R_x)$ $(x = 1, 2, \ldots)$ est une valse (N.B. certaines notes s'étendent sur deux mesures) :

Mais à ce stade, la mélodie s'est complètement effondrée. J'ai demandé à Guy s'il pouvait penser à une raison à cela : il a dit : "Oui, une erreur que j'ai faite dans le calcul." Après correction, la valse se poursuit :

Celle-ci essaie d'être périodique avec une période 26, mais des sauts continuent d'apparaître. Beaucoup d'autres jeux de ce type donnent des séries mélodieuses, quelque peu périodiques, sans raison évidente. Guy a découvert deux curieuses exceptions : son jeu "4", retirer 1 jeton pas à l'extrémité d'une rangée, a une période exacte de 34 pour $x \geq 54$, et Kayles, retirer 1 ou 2 jetons adjacents, a une période exacte de 12 pour $x \geq 71$. Ainsi ces jeux ont une analyse complète. Mais en général, il pourrait être utile de faire appel à un musicien professionnel pour étudier la théorie des nombres. Peut-être qu'une étude approfondie du dernier théorème de Fermat mettra au jour l'Accord perdu. Après tout, pourquoi pas ?

Une brève histoire des courbes de remplissage d'espace
Ian Stewart, Professeur émérite de mathématiques, Université de Warwick

On se reportera aux images très nombreuses de l'article en anglais dans le numéro 65 de la revue Eureka.

En 1890, les mathématiciens ont entrevu les complications topologiques des courbes continues du plan. Giuseppe Peano a donné une construction d'une courbe continue passant par chaque point du carré unité. Aujourd'hui, nous voyons cette courbe comme un précurseur de la notion de fractal de Benoît Mandelbrot [3, 10]. À l'époque, les contre-exemples de ce genre général étaient appelés "courbes pathologiques", et de nombreux mathématiciens y réagissaient comme nous réagissons habituellement à la pathologie - avec peur et dégoût. Plus tard, la profession s'y est habituée et a absorbé les profondes leçons topologiques qu'elles nous enseignent.

Je vais raconter quelques-unes de leurs histoires, vous montrer les constructions assez intriquées que les premiers pionniers ont utilisées pour prouver leur existence, puis essayer de vous convaincre qu'il existe une manière beaucoup plus simple.

1. Un infini plus grand

Tout a commencé avec Georg Cantor. Cantor est célèbre pour son introduction des cardinaux transfinis - comment compter le nombre d'éléments dans un ensemble infini. Il a célèbrement prouvé que certains infinis sont plus grands que d'autres. Plus précisément, il n'y a pas de correspondance biunivoque (bijection) entre les entiers et les nombres réels. Cherchant un cardinal transfini plus grand que celui des réels, il s'est convaincu pendant un temps que le cardinal du plan devait être supérieur à celui d'une ligne. Le 5 janvier 1874, il écrivit à Richard Dedekind :

Une surface (disons un carré qui inclut la frontière) peut-elle être mise en correspondance biunivoque avec une ligne (disons un segment de droite qui inclut les extrémités) de sorte que pour chaque point de la surface il y ait un point correspondant de la ligne et, réciproquement, pour chaque point de la ligne il y ait un point correspondant de la surface ? Je pense que répondre à cette question ne serait pas une tâche facile, malgré le fait que la réponse semble si clairement être "non" que la preuve paraît presque inutile.

Trois ans plus tard, il écrivit de nouveau à Dedekind pour dire qu'il s'était trompé. Il avait trouvé une correspondance biunivoque entre l'intervalle unité $[0, 1] \subseteq \mathbb{R}$ et l'espace à n dimensions $\mathbb{R}^n$ pour tout n fini. "Je le vois, mais je ne le crois pas !" écrivit-il. L'idée principale est simple : étant donné deux points $x, y \in [0, 1]$, écrivons-les en décimales comme

$$x = 0 \cdot x_1 x_2 x_3 x_4 \ldots,$$

$$y = 0 \cdot y_1 y_2 y_3 y_4 \ldots,$$

et associons-leur

$$0 \cdot x_1 y_1 x_2 y_2 x_3 y_3 x_4 y_4 \ldots,$$

en entrelaçant les décimales. Il faut prendre quelques précautions avec la représentation non unique d'un nombre réel en décimal, comme $0 \cdot 500000 \cdots = 0 \cdot 499999 \ldots$, mais c'est facile. Ensuite, il faut établir une bijection entre $[0, 1]$ et $\mathbb{R}$, ce qui est aussi facile.

En 1878, Cantor publia certains de ces résultats dans le Journal fur die Reine und Angewandte Mathematik, connu informellement sous le nom de "Journal de Crelle" d'après son célèbre éditeur et fondateur August Crelle, dans un article qui formulait la notion de correspondance biunivoque. Il étudia les ensembles dénombrables, et les ensembles qui sont en correspondance biunivoque les uns avec les autres. Il discuta de la "dimension", nota que la correspondance biunivoque entre $[0, 1]$ et le carré unité ne préserve pas la dimension, et souligna que l'application qu'il avait construite n'est pas continue.

Les idées de Cantor étaient controversées : certains mathématiciens éminents les considéraient comme des absurdités, probablement parce qu'elles étaient si originales qu'il fallait de l'imagination et un esprit ouvert pour les apprécier. D'autres, notamment David Hilbert, déclarèrent que le nouveau domaine que Cantor avait ouvert était un "paradis". La pleine reconnaissance de l'importance du travail de Cantor ne vint qu'après sa mort.

2. Courbes de remplissage d'espace

Introduisons la notation $\mathbb{U} = [0, 1]$. Alors $\mathbb{U}^2$ est le carré unité, et plus généralement $\mathbb{U}^n$ est l'hypercube dans $\mathbb{R}^n$. La bijection de Cantor $\mathbb{U} \to \mathbb{U}^2$ montre que les bijections ne préservent pas la dimension, mais comme il l'a souligné, son application est discontinue, donc il n'y a pas de raison particulière de s'attendre à ce que la dimension soit préservée.

En 1879, Eugen Netto répondit à une question évidente en prouvant qu'il n'existe pas de bijection continue entre ces ensembles, ce qui est plus délicat qu'il n'y paraît.

La percée la plus significative survint en 1890, lorsque Giuseppe Peano jeta un pavé dans la mare en montrant qu'il existe une application continue de $\overline{\mathbb{U}}$ sur $\overline{\mathbb{U}^2}$. C'est-à-dire : il existe une courbe continue dont l'image remplit le carré unité. La courbe n'est pas seulement dense dans $\overline{\mathbb{U}^2}$: elle passe par chaque point. Nous appelons maintenant de tels objets des courbes de remplissage d'espace. Elles montrent que notre image mentale par défaut d'une courbe continue peut être nettement trompeuse. L'article de Peano ne contient pas d'images. Il définit la courbe en utilisant des développements en base 3 des points de l'intervalle unité. Sagan [18] affirme que la construction est équivalente à celle géométrique montrée sur la figure 2 (à gauche). Pour une explication, voir Jaffer [8]. En 1891, David Hilbert publia un autre exemple de courbe de remplissage d'espace, et inclut une image, essentiellement comme celle de la figure 2 (à droite).

Les deux constructions sont assez compliquées : les figures montrent un stade précoce d'un processus récursif dans lequel des polygones simples sont remplacés de manière répétée par des polygones plus élaborés. Ce qui importe vraiment ici n'est pas visible en tant que tel dans l'image : comment les courbes sont paramétrées par $t \in [0, 1]$. Par défaut, on utilise la longueur d'arc, mise à l'échelle pour que la longueur totale soit 1. Une courbe dans un espace X est alors définie comme une application continue $\gamma : [0, 1] \to X$.

Lorsque nous passons à la limite infinie des polygones finis qui sont construits, nous prenons la limite de ces applications. Dessiner seulement l'image $\gamma([0,1])$ peut être trompeur.

De nombreuses autres courbes de remplissage d'espace ont été trouvées depuis. Moore [13] a inventé une modification de la construction de Hilbert qui donne une courbe de remplissage d'espace fermée - une qui revient à son point de départ. Sierpinski en a conçu une qui remplit un triangle, figure 1. Pólya en a trouvé une autre [17]. On pourra consister au sujet de ces courbes les livres de Bader [1] et Sagan [19].

Une construction plus simple, également attribuée à Peano, est montrée à la figure 3. Telle qu'elle est dessinée, on a du mal à imaginer comment tracer cette courbe. Il y a plusieurs façons de dessiner le premier stade comme un polygone continu. Peitgen, Jurgens et Saupe [16] page 384 remarquent que n'importe laquelle d'entre elles suffira. Au stade suivant de la construction, chaque segment de droite est remplacé par une version plus petite du même polygone bouclé, tracé de manière continue, et ainsi de suite. Les images de ces courbes ressemblent toutes à la figure 3, mais les chemins tracés à travers elles varient.

Les courbes de remplissage d'espace ont des applications en informatique, comme le stockage et la récupération de données multidimensionnelles (voir Lawder [9]). L'idée de base est que nous pouvons parcourir un tableau multidimensionnel en suivant une approximation d'une courbe de remplissage d'espace, réduisant les problèmes au cas unidimensionnel. Une autre application donne une solution rapide et approximative du problème du voyageur de commerce : trouver le chemin le plus court le long d'un réseau routier qui visite une liste spécifiée de villes. L'idée est de faire passer une approximation finie d'une courbe de remplissage d'espace à travers la région contenant les villes, de les ordonner en utilisant le paramètre de cette courbe, puis de les visiter dans cet ordre en utilisant le chemin de liaison le plus court. Cela produit un itinéraire qui n'est généralement pas plus de 25% plus long que l'optimal (voir Bartholdi [2]). Cette approche heuristique a été utilisée pour trouver un itinéraire efficace pour amener les repas à domicile à Atlanta, en Géorgie ; pour acheminer la livraison de sang aux hôpitaux ; pour cibler un laser spatial dans l'Initiative de défense stratégique hypothétique de Ronald Reagan ("Guerre des étoiles") ; et pour contrôler une table traçante XY pour l'infographie, où le temps passé à déplacer le stylo a été réduit d'un facteur vingt. (Les graphiques d'aujourd'hui n'utilisent pas de tables traçantes mécaniques, bien sûr, mais des gadgets similaires existent pour d'autres applications, y compris la broderie).

3. Une condition suffisante

Ces constructions sont très largement citées, mais elles sont assez compliquées. La plus simple est la figure 3. Mandelbrot [10] page 64 attribue une approche plus simple à Ernesto Cesàro, mais même sa méthode alterne deux étapes de construction différentes et remplit un triangle. Cependant, il existe une construction encore plus simple, qui est très naturelle une fois que l'on abandonne l'idée que les étapes successives de la courbe devraient paraître de plus en plus sinueuses. Vouloir faire cela confond en effet la courbe, en tant que chemin paramétré par $t \in \mathbb{U}$, avec son image. L'image peut paraître assez simple, même si le chemin que trace la courbe est très sinueux. Mais les gens

ont tendance à rechercher des images compliquées.

Je n'ai pas pu trouver cette construction dans la littérature, mais elle est si simple et naturelle qu'elle doit sûrement être connue des experts. Chaque étape produit une courbe dont l'image est une grille carrée terne et ennuyeuse. Ce n'est pas aussi joli que toutes ces courbes de remplissage d'espace plus populaires, et là encore il est impossible de déduire la façon dont la courbe ondule à partir d'une image de son image. Mais les images sont de toute façon des approximations successives de la courbe finale. Leur forme sinueuse peut offrir une certaine intuition sur la raison pour laquelle la courbe limite remplit l'espace, mais cette courbe va droit au but, obtenant le résultat désiré avec un minimum d'histoires et de tracas.

Toutes les constructions ci-dessus définissent la courbe de remplissage d'espace γ comme la limite d'une suite de courbes $\gamma_n : \mathbb{U} \to X \subseteq \mathbb{R}^2$. Un théorème de base simple donne des conditions suffisantes pour qu'une telle limite soit de remplissage d'espace, énoncées en termes des approximations γ_n.

Un point crucial est que la suite doit être uniformément convergente. C'est-à-dire, étant donné tout $\epsilon > 0$, il existe N tel que

$$\|\gamma_m(x) - \gamma_n(x)\| < \epsilon \text{ pour tout } m, n \geq N \text{ et } x \in \mathbb{U},$$

où $\|\cdot\|$ est la norme euclidienne. (En fait, n'importe quelle norme fera l'affaire ici). Cette condition implique que pour tout $x \in \mathbb{U}$ la limite

$$f(x) = \lim_{n\to\infty} f_n(x),$$

existe, et que f est continue. Nous pouvons alors nous appuyer sur le résultat purement topologique suivant, prouvé dans la plupart des cours d'analyse réelle.

Théorème : Soit $\gamma_n : \mathbb{U} \to \mathbb{U}^2$ une suite de courbes continues, pour $n \in \mathbb{N}$. Supposons que :

(i) La suite de fonctions (γ_n) est uniformément convergente.

(ii) L'union des images de tous les γ_n est dense dans $\mathbb{U}^2$.

Alors

$$\gamma = \lim_{n\to\infty} \gamma_n,$$

est continue, et son image est $\mathbb{U}^2$. Donc γ est une courbe de remplissage d'espace.

Preuve : Par la condition (i), la limite γ existe et est continue. Soit $z \in \mathbb{U}^2$. Par la condition (ii), pour chaque $m > 0$, $m \in \mathbb{N}$, il existe $t_m \in [0, 1]$ et $n_m \in \mathbb{N}$ tels que

$$\|\gamma_{n_m}(t_m) - z\| < \frac{1}{m}.$$

La sous-suite γ_{n_m} tend aussi uniformément vers γ. La suite (t_m) est dans $[0, 1]$ qui est fermé et borné, donc elle a une sous-suite convergente, de limite t_0. Par continuité uniforme, si m est assez grand

$$\|\gamma_{n_m}(t_0) - \gamma(t_0)\| < \frac{1}{m}.$$

Donc

$$\|\gamma(t_0) - z\| < \frac{2}{m},$$

pour tout m, donc $\gamma(t_0) = z$ et γ est surjective.

4. Une construction plus simple

Je vais maintenant construire une telle suite γ_n ; appelons sa limite γ la courbe en T. Je vais la prescrire géométriquement, mais rendre la construction précise est routinier.

Définissons γ_0 de sorte que son image soit la frontière de $\mathbb{U}^2$ décrite dans le sens antihoraire, paramétrée proportionnellement à la longueur d'arc ; voir en haut à gauche de la figure 4.

Pour obtenir γ_1, remplaçons chaque segment de droite de γ_0 par un polygone en forme de T, comme montré en bas de la figure 4. La partie supplémentaire du T s'étend toujours vers la gauche du segment, tel qu'orienté par la flèche. Le chemin γ_1 suit le chemin original jusqu'à son milieu, tourne à gauche, fait demi-tour pour revenir au milieu, puis continue le long du segment original jusqu'à son extrémité. Si nous faisons cela pour chacun des quatre segments de γ_0 et ajoutons les chemins en forme de T (dont les extrémités et les débuts s'emboîtent correctement), nous obtenons le chemin γ_1 en haut à droite de la figure 4. Pour plus de clarté, le polygone gris montre la direction dans laquelle t parcourt γ_1. Ce chemin visite le centre du carré quatre fois, mais tous les segments sont traités de la même manière, ce qui facilite la définition inductive des γ_n. Nous répétons maintenant la même construction sur chaque segment de droite de γ_1 pour obtenir γ_2, et continuons inductivement pour définir γ_n pour tout $n \in \mathbb{N}$.

L'image de γ_n est une grille carrée de lignes séparées par une distance 2^{-n}. Clairement, l'union de ces images est dense dans $\mathbb{U}^2$: elle consiste en tous les (x, y) pour lesquels soit x soit y est de la forme $m2^{-n}$ pour $0 \leq m \leq 2^n$, où $m \in \mathbb{N}$. Donc la condition (i) est vérifiée, et nous pouvons définir γ comme la limite des γ_n. En considérant l'étape de construction en bas de la figure 4, il est facile de prouver que

$$|\gamma_n(t) - \gamma_{n+1}(t)| \leq \sqrt{2} \cdot 2^{-n},$$

d'où la condition (i), la convergence uniforme, découle en sommant une série géométrique. La condition (ii) est évidente.

Par conséquent, γ est une courbe de remplissage d'espace. Comme celle de Moore, elle est fermée.

Bien que l'image de γ_n soit une grille, la structure réelle de γ devient de plus en plus compliquée en raison de la façon dont elle zigzague d'avant en arrière. Si nous écrivons (x_n, y_n), où $x, y : \mathbb{U} \to \mathbb{U}$, sont elles-mêmes continues, alors la figure 5 montre les graphiques de x_0, y_0 et x_1, y_1.

Les pentes des parties de ces graphiques qui ne sont pas horizontales deviennent de plus en plus raides à mesure que n augmente. Néanmoins, la limite est continue - en raison de la convergence uniforme, et plus intuitivement, parce que les intervalles sur lesquels ces pentes se produisent deviennent arbitrairement petits (bien qu'il y ait de plus en plus de tels intervalles).

5. Structure fractale et dimension

Ces constructions récursives de courbes de remplissage d'espace sont des exemples de fractales, Mandelbrot [10]. C'est-à-dire qu'elles ont une structure à toutes les échelles de grossissement. Elles sont (très proches d'être) auto-similaires, c'est-à-dire constituées de copies plus petites d'elles-mêmes.

Par exemple, la courbe en T peut être divisée en quatre morceaux, paramétrés par t dans les intervalles $[0, 1/4], [1/4, 1/2]$, $[1/2, 3/4]$, et $[3/4, 1]$. Chaque morceau remplit un triangle rectangle isocèle, joignant un côté du carré unité au milieu. Et chaque morceau est fait de quatre copies de lui-même, de taille la moitié de la sienne. La figure 6 tente d'illustrer cela en déformant légèrement le triangle pour laisser de petits espaces. La courbe visite les points numérotés dans l'ordre des numéros ; notons que 3, 6 et 9 sont tous le même point.

Je ne prétends pas simplement que le triangle est auto-similaire. Je prétends que le chemin que trace la courbe, avec toutes ses ondulations, est auto-similaire.

En géométrie fractale, il y a une notion de dimension fractale, plus précisément la dimension de Hausdorff-Besicovich. Je ne la définirai pas ici car elle est compliquée ; voir Falconer [3]. Dans le cas auto-similaire, l'idée est plus simple. Il faut 2 copies de l'intervalle unité pour faire un intervalle de longueur 2. Il faut $4 = 2^2$ copies du carré unité pour faire un carré de côté 2. Il faut $8 = 2^3$ copies du cube unité pour faire un cube de côté 2. En général, il faut 2^d copies de l'hypercube unité de dimension d pour faire un hypercube de côté 2, et il faut k^d copies de l'hypercube unité de dimension d pour faire un hypercube.

Les fractales ont souvent une dimension non entière. Mais dans ce cas, il faut 4 copies de la courbe en T pour en faire une deux fois plus grande, donc $4 = 2^d$ et $d = 2$. Ce n'est pas une grande surprise, car une courbe de remplissage d'espace remplit l'espace, et là l'espace a pour dimension 2. Néanmoins, il est rassurant que la numérologie fonctionne correctement.

6. Remplir d'autres formes

Quelles autres formes une courbe peut-elle remplir ? Il est facile de trouver une bijection continue de $\mathbb{U}^2$ vers un disque circulaire plein, et en composant avec celle-ci, nous obtenons une courbe qui remplit le disque. Le même truc construit une courbe qui remplit tout espace topologiquement équivalent à $\mathbb{U}^2$, et ceux-ci peuvent être assez compliqués. Cependant, il y a beaucoup d'autres possibilités. Par exemple, deux carrés pleins joints par une courbe peuvent aussi être remplis en reliant deux courbes de remplissage d'espace ensemble.

La construction de Hilbert s'étend à trois dimensions, donnant une courbe qui remplit le cube unité $\mathbb{U}^3$. Mais là encore, il y a une manière plus simple de prouver qu'une telle courbe existe. L'existence d'une courbe de remplissage d'espace signifie que nous pouvons paramétrer $\mathbb{U}^2$ par $t \in \mathbb{U}$. Donc nous pouvons paramétrer $\mathbb{U}^3 = \mathbb{U}^2 \times \mathbb{U}$ par $\mathbb{U} \times \mathbb{U}$. Mais c'est $\mathbb{U}^2$ qui peut être paramétré par $\mathbb{U}$. Si nous posons $\gamma(t) = (x(t), y(t))$, et définissons ensuite

$$\theta(t) = (x(x(t)), y(x(t)), y(t)),$$

nous obtenons une courbe θ qui remplit $\mathbb{U}^3$. En continuant avec ce truc, la courbe

$$\phi(t) = (x(x(x(t))), y(x(x(t))),$$
$$y(x(t)), y(t)),$$

remplit $\mathbb{U}^4$, et ainsi de suite. Par induction, il existe une application continue de $\mathbb{U}$ sur $\mathbb{U}^n$ pour tout n fini. En effet, si $m, n > 0$ sont des entiers, il existe une application continue de $\mathbb{U}^m$ sur $\mathbb{U}^n$, même lorsque $m < n$.

Le dernier mot à ajouter concernant de tels résultats est le théorème de Hahn-Mazurkiewicz, prouvé par Hahn [4, 5] et Mazurkiewicz [11, 12]. Il stipule qu'un ensemble S est l'image d'une application continue de $\mathbb{U}$ vers un espace topologique de Hausdorff si et seulement si S est compact, connexe, localement connexe et métrique. Voir Hocking et Young [7] page 129. Curieusement, la preuve fait appel à une autre idée de Cantor : l'ensemble de Cantor. Mais c'est une autre histoire.

7. Références

Bader M (2013) Space-Filling Curves. Texts in Computational Science and Engineering 9. Springer, Heidelberg.

Bartholdi J. Some combinatorial applications of space-filling curves.

Falconer K (1990) Fractal Geometry. Wiley, Chichester.

Hahn H (1914) Uber die allgemeinste ebene Punktmenge, dies stetiges Bild einer Strecke ist. J Bericht Deutsch Math-Verein 23 :318-22.

Hahn H (1914) Mengentheoretische Charakterisierung der stetigen Kurven. S B Kaiserl Akad Wiss Wien 123 :2433-87.

Hilbert D (1891) Uber die stetige Abbildung einer Linie auf ein Flachenstuck. Math Ann 38 :459-60.

Hocking JG, Young GS (1961) Topology. Addison-Wesley, Reading MA.

Jaffer A. Peano space-filling curves.

Lawder J (1999) The application of space-filling curves to the storage and retrieval of multi-dimensional data. PhD Thesis, Birkbeck College, London.

Mandelbrot B (1982) The Fractal Geometry of Nature. W H Freeman, San Francisco.

Mazurkiewicz S (1913) O aritmetzacji kontinuow. C R Soc Sc Varsovie 6 :305-11.

Mazurkiewicz S (1913) O aritmetzacji kontinuow II. C R Soc Sc Varsovie 6 :941-5.

Moore EH (1900) On certain crinkly curves. Trans Amer Math Soc 1 :72-90.

Netto E (1879) Beitrag zur Mannigfaltigkeitslehre. J Reine Angew Math 86 :263-8.

Peano G (1890) Sur une courbe qui remplit toute une aire plane. Math Ann 36 :157-60.

Peitgen HO, Jurgens H, Saupe D (1992) Chaos and Fractals. Springer, New York. Pólya G (1913) Über eine Peanoscbe Kurve. Bull Acad Sci de Cracovie (sci math et nat Serie A) 1-9.

Sagan H (1991) Some reflections on the emergence of space-filling curves : the way it could have happened and should have happened, but did not happen. J Franklin Inst 328 :419-30.

Sagan H (1994) Space-Filling Curves. Springer, New York. Reissued 2013 in Universitext series.

Sierpinski W (1912) Sur une nouvelle courbe continue qui remplit toute une aire plane. Bull Acad Sci de Cracovie (sci math et nat Serie A) 193-219.

1. L'hypothèse de Riemann

L'hypothèse de Riemann est un excellent point de départ pour notre aventure explorant la jungle indomptée des Problèmes du Millénaire, car c'est le seul des 23 problèmes originels de Hilbert de 1900 qui reste non résolu (si vous n'avez aucune idée de ce dont je parle, vous feriez mieux de revenir en arrière et de lire mon premier article ici - il est génial, croyez-moi).

L'hypothèse de Riemann est à la fois un classique et un bon cru. Elle a été formulée pour la première fois en 1859 par Bernhard Riemann et elle déconcerte encore les mathématiciens plus de 150 ans plus tard. L'hypothèse porte sur les nombres premiers - ce sont ceux qui ne peuvent être obtenus qu'en se multipliant par un et qui n'ont aucun autre facteur. Prenons le nombre 42, aussi connu comme la réponse à l'univers et à tout ce qu'il contient. 42 a beaucoup de facteurs, 7×6, 21×2, 42×1, d'accord il a six facteurs, mais surtout il en a plus de deux. Maintenant, qu'en est-il du nombre à côté, 41. La seule façon de l'obtenir est de multiplier 41 par 1 - c'est tout. Cela fait de 41 un nombre premier.

Les nombres premiers sont très spéciaux en mathématiques parce que tous les nombres peuvent être construits à partir d'eux. Si je vous donne n'importe quel nombre, vous pouvez le décomposer en morceaux de plus en plus petits, chacun étant un nombre premier. Revenons à notre exemple de 42. 7×6 est la même chose que $7 \times (3 \times 2)$. 21×2 est la même chose que $(7 \times 3) \times 2$. 42×1 est aussi $7 \times 3 \times 2$. Vous remarquerez qu'il y a un motif ici, les briques pour construire 42 sont 7, 3 et 2 qui sont tous des nombres premiers. Ce sera vrai pour tout nombre auquel vous pouvez penser - essayez : écrivez les facteurs, puis continuez à décomposer chaque facteur en facteurs de plus en plus petits et finalement vous arrivez à des nombres premiers. Si l'ADN donne les briques de la vie, les nombres premiers donnent les briques des mathématiques. Alors que nous comprenons maintenant assez bien l'ADN - nous pouvons le modifier et provoquer toutes sortes de choses cool, on ne peut pas en dire autant des nombres premiers. Nous savons ce qu'ils sont et que tous les nombres peuvent être écrits en fonction d'eux, mais c'est à peu près toute l'histoire. Nous ne les comprenons vraiment pas très bien, ce qui nous ramène très joliment à notre Problème du Millénaire.

L'hypothèse de Riemann parle d'une fonction très compliquée appelée la fonction zêta de Riemann. Je vais vous montrer à quoi elle ressemble parce que c'est cool - et si vous voulez impressionner vos amis, dites simplement "je sais à quoi ressemble la fonction zêta de Riemann" et écrivez la formule. Avertissement : le succès de cette tactique peut ou non dépendre de l'opinion de vos amis sur les mathématiques d'apparence cool. Si vous me faisiez ça, je serais au moins impressionné. Quoi qu'il en soit, voici la formule

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1}}{e^x - 1} dx.$$

Au-delà de son apparence cool une fois écrite, nous n'avons pas vraiment besoin d'en savoir plus à son sujet. L'hypothèse de Riemann dit essentiellement : voici quelques propriétés de cette fonction de Riemann : sont-elles vraies ou non ? Si elles le sont, et la plupart des mathématiciens pensent qu'elles le sont, alors cela signifiera que nous comprendrons beaucoup mieux l'emplacement des nombres premiers et, surtout, que nous saurons où chercher pour les trouver.

Alors pourquoi nous soucions-nous de savoir où se trouvent les nombres premiers sur la droite numérique ? C'est la question à un million de dollars, littéralement. Lorsque vous utilisez votre carte de crédit en ligne, les détails sont cryptés avec un code composé d'un nombre d'environ 200 chiffres pour vous protéger des pirates. Même s'ils interceptent le signal, cela ressemble à du charabia. La seule façon de décrypter le code est de savoir quels deux nombres premiers multipliés ensemble donnent le nombre à 200 chiffres. Décomposer un nombre en facteurs premiers, comme nous l'avons fait ci-dessus pour 42, n'était pas trop difficile, parce que 42 est un petit nombre. Une fois que les nombres deviennent plus grands, le jeu change. Même les ordinateurs les plus puissants ne peuvent calculer les facteurs premiers de nombres allant jusqu'à 100 chiffres - donc avec 200 chiffres, les détails de votre carte de crédit sont "aussi sûrs que des maisons".

Les méthodes actuelles de sécurité en ligne fonctionnent précisément parce que nous ne comprenons pas les nombres premiers. Si vous résolviez l'hypothèse de Riemann, vous n'obtiendriez pas seulement le prix d'un million de dollars, vous pourriez aussi voler les détails de carte de crédit de tout le monde sur Internet... Je ne dis absolument pas que vous devriez faire cela, mais vous savez, vous le pourriez probablement si vous aviez envie de le faire.

2. P versus NP

Le numéro deux sur la liste est un autre problème de gros calibre du monde des mathématiques. Il est si célèbre qu'il est même apparu dans plusieurs films hollywoodiens alors que les réalisateurs tentent d'ajouter de la crédibilité à leurs histoires impliquant un robot-apocalypse - oui, je pense à toi iRobot. Le problème s'appelle P versus NP.

C'est un ajout relativement récent au monde des mathématiques dans sa forme actuelle, mais le problème qu'il tente de résoudre remonte à la préhistoire. Imaginez la scène : vous vous réveillez dans votre caverne et vous avez faim, vous avez soif et vous auriez probablement besoin de vous laver car votre voisin sympathique de la caverne d'à côté vient plus tard pour un moment de qualité "netflix and chill". C'est un bon point en fait, vous avez aussi besoin de quelques fleurs et de la dernière peinture rupestre à succès. Cela fait cinq choses, ce qui signifie cinq endroits différents que vous devez visiter. La question est : quel est le chemin le plus rapide que vous pouvez prendre qui passe par les cinq endroits et se termine avec vous de retour chez vous. Aujourd'hui, nous utilisons simplement Google, mais en tant que vous en mode préhistorique, vous n'êtes pas si chanceux, essayons un peu de mathématiques.

Il y a 5 endroits à visiter, ce qui signifie que vous avez un choix parmi 5 pour le premier endroit où aller. Une fois que vous avez visité le premier arrêt, vous avez alors un choix de 4 endroits à visiter ensuite. Puis un choix de 3, puis 2 et enfin aucun choix pour le dernier car c'est le seul qui reste. Donc au total cela fait $5 \times 4 \times 3 \times 2 \times 1 = 120$ itinéraires différents. Nous pouvons probablement rayer quelques itinéraires de la liste en utilisant notre connaissance de la région locale - peut-être la moitié d'entre eux traversent les marais abritant M. T. Rex, ce qui non seulement vous ralentit sévèrement, mais donne aussi une probabilité élevée de mort.

Lorsque vous demandez à Google de calculer l'itinéraire le plus rapide, il vérifiera en fait les 120 possibilités et vous donnera ensuite le plus rapide. Pas vraiment un problème aujourd'hui, les ordi-

nateurs sont rapides et vous pouvez donc probablement obtenir l'itinéraire le plus rapide en moins d'une seconde. Mais que se passe-t-il si vous voulez visiter 1000 endroits différents - peut-être que vous êtes très populaire dans votre village préhistorique et avez de nombreux partenaires à visiter - cela signifie $1000 \times 999 \times 998 \times \ldots \times 2 \times 1$ itinéraires différents à vérifier. Je ne vais pas écrire ce que ce nombre représente, mais je peux vous dire que pour qu'un ordinateur aujourd'hui vérifie autant d'itinéraires différents, il faudrait plus de temps que l'âge de l'univers.

C'est à peu près le problème P versus NP en un mot : si nous pouvons programmer un ordinateur pour qu'il soit capable de reconnaître si une réponse est une solution d'un problème, pouvons-nous aussi le programmer pour trouver rapidement cette solution ? Avec notre exemple, votre ordinateur peut vous dire si votre itinéraire suggéré visitera les 1000 arrêts en quelques millisecondes, mais l'enfer gèlera littéralement avant qu'il puisse vous dire si c'est le plus rapide de tous les itinéraires possibles. P versus NP demande s'il existe ou non un raccourci qui n'implique pas de vérifier chaque combinaison possible.

La plupart des mathématiciens croient que P n'est pas égal à NP et en fait si les deux s'avéraient égaux, le monde tel que nous le connaissons prendrait probablement fin. P = NP signifie que nous pouvons instantanément casser n'importe quel code qui ait jamais existé, prédire les tendances boursières avant qu'elles ne se produisent et, plus inquiétant encore, construire des robots avec une intelligence artificielle supérieure à quiconque sur terre. Mieux vaut mettre Will Smith sur la composition rapide au cas où iRobot n'aurait finalement pas été si loin de la vérité...

Le lemme de Yoneda dans la catégorie des matrices
auteur anonyme

On se reportera aux images très nombreuses de l'article en anglais dans le numéro 67 de la revue Eureka.

1. Introduction

Lorsque vous arrivez pour la première fois à l'université et étudiez les mathématiques, vous commencez par regarder les matrices. Elles vous permettent d'empaqueter des systèmes d'équations linéaires dans une seule matrice. Et au début, vous apprenez l'élimination de Gauss. C'est similaire à la façon dont on traiterait normalement des équations simultanées à l'école - vous pouvez implémenter l'addition et la multiplication d'équations comme des opérations de ligne sur la matrice correspondante, et la réduire à une certaine forme canonique, souvent appelée "forme échelonnée" ou "forme échelonnée réduite".

Et plus tard, vous pourriez apprendre un fait intéressant qui relie ces opérations de ligne à la multiplication matricielle plus standard. Toute opération de ligne sur une matrice A peut être implémentée en multipliant simplement à gauche par une matrice fixe M - et la façon de trouver M est simplement d'appliquer l'opération à la matrice identité.

Par exemple, si nous prenons $\begin{pmatrix} 1 & 2 \\ 5 & 7 \end{pmatrix}$, et que nous voulons implémenter l'opération consistant à soustraire 5 fois la première ligne de la deuxième ligne, pour obtenir $\begin{pmatrix} 1 & 2 \\ 0 & -3 \end{pmatrix}$, alors nous pouvons utiliser la procédure suivante. Appliquez la même opération à la matrice identité, donc $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ devient $\begin{pmatrix} 1 & 0 \\ -5 & 1 \end{pmatrix}$, et on peut vérifier que $\begin{pmatrix} 1 & 0 \\ -5 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 5 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & -3 \end{pmatrix}$. Ainsi, lorsque nous résolvons un système linéaire $A\vec{x} = \vec{b}$ en utilisant des opérations de ligne, ce que nous faisons réellement est de multiplier à gauche par une série de matrices, $R_1, R_2, \ldots$, jusqu'à ce que nous réduisions le système à quelque chose de plus traitable (par exemple une matrice triangulaire) !

Mais - pourquoi est-ce vrai ? Il s'avère que c'est un cas d'un fait beaucoup plus général appelé le lemme de Yoneda, dont on dit qu'il est le théorème fondamental de la théorie des catégories. Maintenant, je n'ai pas le temps de passer en revue toutes les définitions nécessaires pour énoncer le lemme de Yoneda, et ce qu'est la théorie des catégories - au lieu de cela, je vais aller droit au but et vous dire ce que fait la théorie des catégories, et pourquoi vous devriez vous y intéresser. Pour comprendre, nous devons comprendre l'un de ses concepts fondamentaux - la naturalité.

2. Qu'est-ce que la naturalité ?

2.1. Invariance

Les mathématiciens utilisent souvent des mots comme "naturel" et "canonique", et l'un des grands succès de la théorie des catégories a été de préciser exactement ce que cela peut signifier. Mais lorsque vous entendez le mot naturel, la première chose qui peut venir à l'esprit est "nombre naturel". Commençons donc par une question... plus facile ? Qu'est-ce qu'un nombre ?

Eh bien, voici d'où je pense que les nombres proviennent. Vous avez un tas de collections différentes de choses - 2 pommes, 2 oranges, 2 portes, 2 voitures, 2 mains. Et vous demandez - qu'ont-elles toutes en commun ? Quel est l'invariant ?

Eh bien, elles ont toutes 2 objets ! Nous avons une fonction, "taille" ou "cardinalité", qui assigne, à chaque collection, un nombre.

Mais il y a quelque chose de plus intéressant ici - si nous avons une façon de relier les entrées, donc qu'elles sont en correspondance bijective, alors nous avons une façon de relier les sorties résultantes - elles sont égales ! C'est le type le plus simple de naturalité - l'invariance.

Remarquez que "taille" n'est pas le seul invariant que nous aurions pu considérer. Un autre invariant est la capacité de diviser la collection en deux moitiés égales - c'est-à-dire la "parité". La "primalité" est aussi un autre invariant. Il y a même certaines cultures qui n'ont pas de noms pour les nombres plus grands que 5, et qui ont juste un seul mot pour "beaucoup" - et cela fonctionne comme un système d'invariants.

Mais il y a un sens dans lequel les nombres sont spéciaux - tous les autres invariants en découlent ! Donc, si vous voulez étudier les invariants des ensembles finis, vous pouvez décomposer cela en déterminant d'abord leur taille, ce que nous appelons "comptage" ou "combinatoire", puis en voyant comment les nombres déterminent tous les autres invariants, ce que nous appelons "théorie des nombres".

Ok, alors qu'est-ce qui vient ensuite - je pense que la plupart des gens peuvent gérer le concept d'invariance. Eh bien, après avoir appris à compter, la chose suivante que vous apprenez probablement est comment additionner ! Et cela nous amène au deuxième type de naturalité - la covariance.

2.2. Covariance

Disons que vous avez 2 cerises et 3 bananes. Vous pouvez les combiner en une seule collection de 5 fruits en prenant leur "union", puis compter combien il y en a, en donnant leur nombre.

Ou, vous pouvez prendre leurs nombres individuellement, puis... additionner les nombres ensemble ! En effet, c'est essentiellement ce qui définit l'addition, et nous donne aussi ce que nous appelons la covariance.

L'idée est - nous avons une fonction, nombre, qui relie les entrées aux sorties. C'est ce que font toutes les fonctions. Cette fonction est "naturelle" si ce qui suit est vrai - chaque fois que nous avons une façon de relier les entrées, dans ce cas l'union, il y a une façon correspondante de relier les sorties !

La multiplication est un autre exemple - dans ce cas, nous pouvons prendre le produit cartésien et faire une grille de tous les choix possibles, puis les compter, ou nous pouvons prendre les tailles individuellement, puis multiplier les nombres ensemble !

2.3. Contravariance

En fait, la construction de l'ensemble des parties nous donne un exemple du troisième type de naturalité - la contravariance ! À savoir, une façon de relier X à Y, donc une fonction $f : X \to Y$, nous donne une façon de relier $P(Y)$ à $P(X)$ - l'image réciproque !

Mais de plus, c'est naturel - si vous avez une façon de relier les entrées, donc une fonction $f : X \to Y$, alors il y a une façon correspondante de relier les sorties - à savoir, étant donné tout sous-ensemble de X, vous pouvez prendre l'image sous f pour obtenir un sous-ensemble de Y. C'est ce dont il s'agit en covariance - élever une fonction entre ensembles en une fonction entre graphes. Il s'agit de reconnaître que vos entrées et sorties ne sont pas juste des objets séparés, mais sont liées de certaines façons. Elles sont un peu comme des "homomorphismes de graphes".

Il y a en fait une jolie description de cela juste en termes de composition de fonctions - vous pouvez identifier les sous-ensembles avec leurs fonctions indicatrices, donc vous pouvez penser à $P(X)$ comme la collection des fonctions de $X \to \{0, 1\}$. Alors, l'image réciproque est juste la composition de fonctions pour les fonctions indicatrices !

Vous avez peut-être aussi des souvenirs de l'école, lorsque vous deviez travailler avec des transformations de graphes - vous avez une fonction de x à y, et un changement de coordonnées x'. Vous devez inverser le changement pour obtenir la flèche de x' à x - en effet, c'est un principe général que les fonctions sont covariantes en sortie, mais contravariantes en entrée.

2.4. Matrices

Il y a aussi un exemple plus simple et plus concret de contravariance - les unités ! Disons que vous mesurez un bâton. Étant donné une unité de mesure, vous obtenez une valeur numérique - donc, par exemple pour votre bâton, l'unité cm donne 50. Et étant donné une autre unité, vous obtenez une valeur différente - l'unité m pour votre bâton donne 0,5. Il y a une façon de relier ces entrées - un mètre fait 100 centimètres. Et la façon de relier les sorties "va à l'envers" - 50 est 100 fois 0,5.

Bien sûr, ici vous pouvez simplement le voir comme un exemple de covariance puisque la flèche de sortie est réversible, alors que ce n'est pas vrai pour disons l'image réciproque. Mais je pense qu'il est utile de souligner la relation "inverse" sur les sorties - en fait, le même raisonnement explique pourquoi les composantes vectorielles sont transformées par P^{-1}. Imaginez une fonction assignant, à chaque base, les composantes du vecteur dans cette base. De plus, ce qui est souvent utile, c'est que la covariance et la contravariance peuvent se combiner pour donner l'invariance. C'est pourquoi

nous écrivons toujours les mesures avec à la fois l'unité et le nombre - l'unité change de manière covariante, le nombre change de manière contravariante, et ensemble ils donnent une notion invariante.

Donc c'était un tour éclair de la naturalité. Revenons aux matrices !

Il s'avère qu'il sera un peu plus facile de considérer d'abord les opérations de colonne - nous pouvons toujours transposer notre résultat pour obtenir la version pour les opérations de ligne.

Donc, une opération de colonne prend en entrée une matrice A avec un nombre fixe de colonnes, et produit une autre matrice $C(A)$. Et elle prend en entrée une autre matrice B et produit une matrice $C(B)$.

Maintenant, mon affirmation est que les opérations de colonne sont naturelles ! Donc, s'il y a une façon de relier A avec B, spécifiquement $B = MA$ (c'est dans ce sens parce que A et B ont le même nombre de colonnes), alors il y a une façon de relier les sorties, donc $C(B) = MC(A)$.

Pourquoi ? Eh bien, chaque colonne de $C(A)$ est de la forme $\sum \lambda_i A_i$ pour A_i la i-ème colonne de A. Alors, la j-ème colonne de $MC(A)$ est $M \sum \lambda_{ij} A_i = \sum \lambda_{ij} MA_i = C(MA) = C(B)$. C'est juste la linéarité !

Comment cela nous aide-t-il ? Eh bien, maintenant que nous comprenons la naturalité, nous sommes prêts à employer un truc - nous pouvons écrire $A = AI$ pour I la matrice identité. Alors $C(A) = C(AI) = AC(I)$. Et cela donne notre résultat - l'opération de colonne est juste une multiplication à droite par une certaine matrice, et nous la trouvons en l'appliquant à la matrice identité !

Nous pouvons aussi tout transposer et obtenir les résultats analogues pour les opérations de ligne :

- Les opérations de ligne sont naturelles, donc si $B = AM$, alors $R(B) = R(AM) = R(A)M$.
- Ensuite, vous pouvez écrire toute matrice A comme IA pour I la matrice identité appropriée.
- Donc, $R(A) = R(IA) = R(I)A$. Donc vous multipliez à gauche par une matrice fixe.

Il y a quelques choses utiles que nous obtenons immédiatement de ce résultat :

- Cela nous donne une façon rapide de composer des opérations de colonne - multipliez les matrices correspondantes ! Ou, appliquez-les successivement à la matrice identité.
- En prenant les déterminants, nous voyons que toute opération de colonne ne fait que multiplier le déterminant de la matrice par un facteur constant, $\det(C(I))$, lorsqu'elle agit sur des matrices carrées.
- Une opération de colonne est inversible si et seulement si la matrice correspondante est inversible.

Je veux souligner que - la preuve n'était vraiment pas difficile, du tout ! Le point était plutôt de créer une définition utile, la naturalité, qui rendait le problème simple et transparent - ce qui est souvent ce que la théorie des catégories donne comme impression. Une grande partie des mathématiques est l'étude des fonctions, des choses qui relient les entrées aux sorties. La naturalité vous demande de considérer la possibilité que les entrées soient liées les unes aux autres - plutôt que d'être des individus disparates et solitaires, elles forment une sorte de communauté, un graphe, un

réseau. La fonction est naturelle si, chaque fois qu'il y a une façon de relier les entrées, il y a une façon correspondante de relier les sorties. C'est un motif dans la façon dont les mathématiciens semblent penser et faire des mathématiques.

Enfin, je manquerais à mon devoir si je ne traçais pas le lien avec le théorème de Cayley de la théorie des groupes, dont le lemme de Yoneda est souvent considéré comme une généralisation. Nous avons vu que chaque matrice M implémente une opération de ligne, $A \to MA$. Cette même application apparaît dans le théorème de Cayley, où vous remplacez un élément de groupe g par l'application $k \to gk$, qui satisfait une condition de naturalité similaire à celle pour les matrices.

En général, ce que le lemme de Yoneda vous dit est que "ce qu'un objet est" est isomorphe à "ce qu'un objet fait", comment il *agit sur* ou *se relie à*[1] d'autres objets.

Et c'est souvent une perspective utile à adopter - vous pouvez spécifier ce que vous voulez que votre objet fasse, comment vous voulez l'utiliser, et Yoneda garantira que cela spécifie l'objet essentiellement de manière unique. Souvent appelé une "propriété universelle". En général, la théorie des catégories se soucie beaucoup plus de ce qu'un objet fait que de ce qu'un objet est, et Yoneda vous dit que ces deux choses sont équivalentes.

1. Je me suis ici exceptionnellement permis d'ajouter des guillemets et de l'italique pour essayer de rendre la phrase encore plus pédagogique que ce qu'elle n'est déjà. dvc