Andrei Kolmogorov, « Le concept de "quantité d'information" »

De la complexité

Andrei Nikolaevich Kolmogorov, né à Tambov en 1903 et mort à Moscou en 1987, est l'un des plus grands mathématiciens du XXe siècle. C'est un défi de choisir un seul texte parmi ses nombreux écrits fondateurs, qui ont marqué autant l'histoire des mathématiques que celle de la physique. On peut en effet considérer comme essentiels non seulement sa théorie de la turbulence* développée, connue sous le nom de K41 (pour Kolmogorov 1941), le théorème KAM (pour Kolmogorov-Arnold-Moser) décrivant l'effet de perturbations sur un système dynamique, l'axiomatisation de la théorie des probabilités* qu'il a proposée en 1933, ses travaux sur les processus stochastiques* (aléatoires), mais aussi la notion de complexité, qui porte aujourd'hui son nom et qu'il a développée dans cet article publié en 1965 dans le premier numéro de la revue *Problemy Peredachi Informatsii*.

CODER EN BINAIRE

La question est de définir la quantité d'information nécessaire pour spécifier un objet *y*. On peut toujours coder les objets en binaire, et se ramener ainsi à des suites de 0 et de 1 (Kolmogorov parle d'indexation de l'objet par une suite binaire). Les deux premières parties de l'article résument ce qui est connu en 1965 : spécifier une suite y revient à restreindre l'ensemble des suites possibles, et l'information à transmettre est définie comme la réduction correspondante d'entropie* (où l'entropie mesure l'indétermination moyenne d'une suite dans l'ensemble) ; l'information dépend donc des connaissances que l'on a initialement sur l'ensemble des suites possibles.

L'étape extraordinaire franchie par Kolmogorov est de considérer un objet individuel *y* (et non un ensemble), et de proposer une mesure intrinsèque, c'est-à-dire la longueur du plus court programme permettant de le décrire. Le théorème* fondamental de l'article montre que la grandeur *K(y)* ainsi définie ne dépend pas de la machine qui exécute le programme, à part une constante additive qui dépend de la machine, mais est identique pour tous les objets !

Cette notion s'inscrit dans la théorie de la calculabilité et des fonctions récursives*. Elle a inspiré au spécialiste de l'algorithmique* Gregory Chaitin (né en 1947) une démonstration en une page du théorème d'incomplétude de Gödel (*cf.* p. 66), suggérant que l'existence de propositions vraies mais non démontrables n'a rien d'exceptionnel. Elle a aussi permis de spécifier ce qu'est une suite individuelle aléatoire : une suite qui ne présente aucune régularité permettant de simplifier sa description, autrement dit dont la complexité est égale à la longueur. En pratique, cette grandeur a un inconvénient : elle n'est pas calculable. Mais on sait calculer des grandeurs qui la limitent supérieurement et qui, elles, peuvent être déterminées à l'aide d'algorithmes de compression, ceux-là même qui ont servi de base aux méthodes d'encodage de fichiers et d'images (le format jpeg, par exemple).

PROGRAMME DE TRAVAIL

Bien qu'on l'appelle parfois « complexité algorithmique », il ne faut pas confondre la complexité de Kolmogorov avec la complexité d'un algorithme, qui se définit par les ressources (temps de calcul, mémoire d'ordinateur) nécessaires pour l'exécuter. La première est une complexité de description, la seconde une complexité de calcul. Kolmogorov souligne déjà cette dualité et esquisse en conclusion un programme de travail pour articuler les deux notions. De fait, le terme de « complexité » qu'il emploie peut surprendre, car il est seulement question de quantités d'information. Mais ce court article a ouvert tout un champ de réflexion. On s'interroge ainsi aujourd'hui pour savoir comment mesurer la complexité de l'organisation d'un système : par exemple une ruche est radicalement plus complexe qu'un ensemble d'abeilles... Il ne suffit donc pas d'additionner les complexités individuelles. ●

Annick Lesne, chercheuse au CNRS,
a coordonné *L'Héritage de Kolmogorov en mathématiques* (Belin, 2004)
avec Éric Charpentier et Nikolaï Nikolski.

LE TEXTE

« Un programme court au prix de calculs d'une durée totalement irréaliste »

§ 3. Une approche algorithmique

[...] En pratique, nous sommes le plus souvent intéressés par la quantité d'information « transmise par un objet individuel x sur un objet individuel y ». Comme nous l'avons déjà noté, il est vrai qu'une telle estimation quantitative de l'information pour des objets individuels n'a de sens que si la quantité d'information est suffisamment importante. Par exemple, cela n'a pas de sens de s'interroger sur la quantité d'information transmise par la suite 0 1 1 0 sur la suite 1 1 0 0.

Mais si on prend une table de nombres aléatoires comme celles couramment utilisées en statistiques, et que nous remplaçons chaque chiffre par le chiffre des unités de son carré, suivant la correspondance

0123456789
0149656941,

alors la nouvelle table va contenir environ $(\log_2 10 - \frac{8}{10})n$ bits d'information sur la table initiale (où n est le nombre de chiffres dans cette table).

[...]

La « complexité relative » d'un objet y, étant donné x, sera calculée comme la longueur minimale $l(p)$ d'un « programme » p permettant d'obtenir y à partir de x. La définition ainsi formulée dépend de la « méthode de programmation » qui n'est rien d'autre que la fonction

$$\varphi(p, x) = y$$

associant un objet y à un programme p et un objet x. En accord avec les idées universellement acceptées en logique mathématique moderne, nous devons supposer que la fonction est partielle récursive*. Pour une telle fonction, nous avons

$$K_\varphi(y/x) = \begin{cases} \min \varphi(p,x) = y^{l(p)} \\ \infty \text{ s'il n'y a pas de programme } p \text{ tel que } \varphi(p,x) = y \end{cases}$$

Dans ce contexte, une fonction $v = \varphi(u)$ de variable $u \in X$ et à valeurs $v \in X$ est dite partielle récursive si elle engendre une fonction partielle récursive décrivant la transformation des indices $n(v) = \Psi[n(u)]$ – [$n(u)$ et $n(v)$ étant des suites finies de 0 et 1, commençant par 1, et indexant respectivement u et v]. Pour comprendre cette définition, il faut noter qu'en général, les fonctions partielles récursives ne sont pas définies partout, et qu'il n'y a pas de méthode fixe pour déterminer si l'application d'un programme p à un objet x va donner un résultat ou non. Par conséquent, la fonction $K_\varphi(y/x)$ n'est pas calculable (c'est-à-dire que ce n'est pas une fonction totale récursive) même quand on sait qu'elle est finie pour tous x et y.

Théorème fondamental. Il existe une fonction partielle récursive $A(p, x)$ telle que pour toute autre fonction partielle récursive $\varphi(p, x)$ on ait l'inégalité

$$K_A(y/x) \leq K_\varphi(y/x) + C_\varphi$$

où la constante ne dépend ni de x ni de y. [...]

§ 4. Remarques finales

Le concept discuté au paragraphe précédent présente une importante lacune : il ne tient pas compte de la difficulté de mettre en œuvre un programme p permettant de passer de x à y. En introduisant les définitions appropriées, il est possible de démontrer des propositions mathématiques rigoureuses qui peuvent être légitimement interprétées comme une indication qu'il existe des cas où un objet x pouvant être décrit par un très court programme, c'est-à-dire un objet x de très petite complexité $K(x)$, ne peut être reconstruit par un programme court qu'au prix de calculs d'une durée totalement irréaliste. Dans un autre article, j'ai l'intention d'étudier la relation entre la complexité $K^t(x)$ du programme et la difficulté t de sa mise œuvre. La complexité $K(x)$ obtenue au § 3 est dans ce cas le minimum de $K^t(x)$ si l'on enlève toute contrainte sur t.

A. N. Kolmogorov, *Trois Approches de la définition du concept de « quantité d'information »*, 1965, traduction originale