Un aperçu de l'incomplétude de Gödel

Avant de se lancer dans la construction des outils nécessaires à l'énoncé précis des théorèmes de Gödel et à leur preuve, on donne un aperçu d'ensemble des théorèmes d'incomplétude et du long chemin pour y parvenir.

1. Énoncés des théorèmes

1.1. Énoncé pour l'arithmétique des entiers

Voici une version encore informelle du premier théorème d'incomplétude.

Théorème 1.
Dans la théorie des entiers naturels, il existe des formules vraies mais non démontrables.

On note $\mathcal{PA}$ la théorie de l'arithmétique des entiers. C'est une liste de formules, appelées *axiomes de Peano* (par exemple « $x + 0 = x$ » et « $x \times 0 = 0$ »). Ils sont construits de sorte que les entiers $\mathbb{N}$ satisfont naturellement ces axiomes.
Gödel ne parlait en fait pas de vérité dans son énoncé afin d'éviter les discussions du type paradoxe du menteur qui affirme « Je suis un menteur ». Il construit plutôt une formule $\mathcal{G}$ telle que ni $\mathcal{G}$, ni $\neg\mathcal{G}$ n'est démontrable. On parle alors d'une théorie *incomplète* (d'où le nom du théorème).
Maintenant, si on se place sur le modèle $\mathbb{N}$ des entiers naturels, qui satisfont les axiomes de Peano, alors soit $\mathcal{G}$ est vraie, soit $\neg\mathcal{G}$ est vraie, et on a bien l'une des deux formules qui est vraie mais non démontrable.

1.2. Théorème général

On pourrait penser que c'est la faute des axiomes de Peano qui sont mal choisis ou pas assez nombreux. Mais le théorème de Gödel est en fait plus général.

Théorème 2.
Toute théorie $\mathcal{T}$ cohérente, contenant $\mathcal{PA}$ et ayant des axiomes récursifs est incomplète.

Oublions un moment les hypothèses et revenons à l'arithmétique de Peano. Puisque cette théorie est incomplète, il existe une formule $\mathcal{P}_0$ non démontrable dans $\mathcal{PA}$. Notons $\mathcal{T}_0 = \mathcal{PA}$ et $\mathcal{T}_1 = \mathcal{T}_0 \cup \{\mathcal{P}_0\}$ la nouvelle théorie obtenue en ajoutant à $\mathcal{T}_0$ la formule indémontrable $\mathcal{P}_0$.
On a maintenant $\mathcal{P}_0$ démontrable dans $\mathcal{T}_1$: $\mathcal{T}_1 \vdash \mathcal{P}_0$, car $\mathcal{P}_0$ est l'une des formules de $\mathcal{T}_1$. Malheureusement la version générale du théorème d'incomplétude affirme que cette nouvelle théorie $\mathcal{T}_1$ est aussi incomplète : il existe une formule $\mathcal{P}_1$ telle que ni $\mathcal{P}_1$, ni $\neg\mathcal{P}_1$ ne sont démontrables. On pourrait rajouter $\mathcal{P}_1$ (ou $\neg\mathcal{P}_1$) afin de construire une théorie $\mathcal{T}_2$,... mais on ne fait que repousser le problème à l'infini.
Ainsi, on aura beau rajouter une à une des formules indémontrables comme axiomes, il existera toujours des formules indémontrables. C'est l'incomplétude !

1.3. Notions en jeu

Revenons sur les définitions des termes des hypothèses et de la conclusion.

Complétude.

- Une théorie $\mathcal{T}$ est ***incomplète*** s'il existe une formule fermée $\mathcal{Q}$ telle que :

$$\mathcal{T} \nvdash \mathcal{Q} \quad \text{et} \quad \mathcal{T} \nvdash \neg\mathcal{Q}$$

- Si on se place dans un ***modèle*** de $\mathcal{T}$, c'est-à-dire une structure où toutes les formules de $\mathcal{T}$ sont satisfaites, alors dans ce cas, un des énoncés $\mathcal{Q}$ ou $\neg\mathcal{Q}$ est vrai mais pas démontrable.

Cohérence.

- Une théorie $\mathcal{T}$ est ***incohérente*** s'il existe une formule fermée $\mathcal{Q}$ telle que :

$$\mathcal{T} \vdash \mathcal{Q} \quad \text{et} \quad \mathcal{T} \vdash \neg\mathcal{Q}$$

  Elle est dite ***cohérente*** sinon.
- Notez bien que la définition de « incohérente » n'est pas la négation de « incomplète » car la négation d'un « et » est un « ou ».
- Ainsi, une théorie incohérente prouve $\mathcal{T} \vdash (\mathcal{Q} \wedge \neg\mathcal{Q})$ donc $\mathcal{T} \vdash \bot$. Alors par la règle d'explosion, $\mathcal{T} \vdash \mathcal{Q}$ (quelle que soit la formule $\mathcal{Q}$) donc on a aussi $\mathcal{T} \vdash \neg\mathcal{Q}$.
- Une théorie incohérente n'a donc pas d'intérêt puisqu'elle ne peut pas avoir de modèle. En effet, si elle possédait un modèle, et comme $\mathcal{T} \vdash \mathcal{Q}$ et $\mathcal{T} \vdash \neg\mathcal{Q}$, par le théorème de validité, alors $\mathcal{Q}$ et $\neg\mathcal{Q}$ devraient être toutes les deux vraies, ce qui fournit une contradiction.
- Autrement dit si $\mathcal{T}$ admet un modèle alors $\mathcal{T}$ est cohérente.

Contenir $\mathcal{PA}$.

- Une théorie est un ensemble de formules, appelées axiomes.
- $\mathcal{PA}$ est une théorie spécifique dont les formules sont les axiomes de Peano qui modélisent les entiers naturels $\mathbb{N}$.
- L'hypothèse « $\mathcal{T}$ contient $\mathcal{PA}$ » signifie que tous les axiomes de Peano doivent être des axiomes de $\mathcal{T}$.
- On dit alors que $\mathcal{T}$ est « suffisamment riche », c'est-à-dire qu'on doit pouvoir travailler avec les entiers naturels. C'est une exigence relativement faible !

Axiomes récursifs

- L'hypothèse que $\mathcal{T}$ a des axiomes récursifs (le terme courant est « récursivement axiomatisable ») est plus technique. Cela signifie que l'on doit pouvoir reconnaître les axiomes.
- Plus précisément pour une formule donnée, on doit pouvoir décider par un calcul (ou un algorithme) si c'est un axiome de $\mathcal{T}$ ou pas.
- Encore une fois, c'est une hypothèse raisonnable : pour décider quels théorèmes on peut prouver il faut pouvoir identifier les axiomes qui servent d'hypothèse !

1.4. Second théorème

Théorème 3.

Une théorie ne peut prouver sa propre cohérence.

Cela signifie concrètement qu'on ne peut espérer vérifier les hypothèses du théorème d'incomplétude en restant dans la théorie $\mathcal{T}$, il faudrait se placer dans un cadre plus large.

Par exemple, on considère généralement que $\mathcal{PA}$ est une théorie qui a un modèle. En réalité, on a accepté implicitement l'existence des entiers naturels $\mathbb{N}$. Pour avoir une construction rigoureuse de $\mathbb{N}$, il faut se placer dans un cadre plus large, par exemple la théorie $\mathcal{ZFC}$ (voir le chapitre « Arithmétiques de Peano »).

2. Conséquences

2.1. Conséquences mathématiques

Autour de 1900, Hilbert souhaite définir un cadre rigoureux pour les fondations de la logique. Dans les années 1800, on a déjà formalisé beaucoup de notions mathématiques : les réels, les limites, la continuité,... Il est temps d'aller plus loin et d'axiomatiser complètement les mathématiques.
L'objectif est de faire abstraction du monde physique (la continuité c'est tracer le graphe sans lever le crayon) et de n'avoir à écrire que des formules logiques. Cependant, on se retrouve vite confronté à des problèmes, comme le paradoxe de Russell : il n'est pas facile de définir ce qu'est un ensemble.
Le premier et le second théorèmes de Gödel font exploser le programme de Hilbert. Après Gödel, il n'y a plus aucune chance d'obtenir une axiomatisation parfaite :

- d'une part, si on se place dans une théorie (suffisamment riche et cohérente), il y aura des formules vraies mais non démontrables,
- en plus on ne peut même pas prouver que la théorie est cohérente afin de garantir qu'on ne puisse pas prouver une chose et son contraire.

C'est donc fatal pour le programme de Hilbert.

Malgré ce choc, les théorèmes de Gödel ont eu des conséquences positives profondes bien au-delà d'un résultat concernant un petit cercle de logiciens.
Tout d'abord, les outils développés par Gödel sont précurseurs de la théorie de l'informatique. Par exemple, l'arithmétisation de la syntaxe (coder toute formule par un nombre) est une idée naturelle dans le monde numérique dans lequel nous vivons aujourd'hui, mais ce n'était pas le cas en 1930.
Plus généralement, Gödel s'intéresse à savoir ce qui est calculable, bien avant la notion d'algorithme et les travaux de Turing (1936) sur un ordinateur théorique, la machine de Turing, et l'indécidabilité.

L'hypothèse du continu.
Même si Gödel construit une formule $\mathcal{G}$ explicite qui n'est pas démontrable, cette formule a pour seul intérêt de prouver le théorème.
L'*hypothèse du continu* est un énoncé beaucoup plus concret qui est, lui aussi, un énoncé indécidable (ni $\mathcal{HC}$ ni $\neg\mathcal{HC}$ ne sont démontrables dans $\mathcal{ZFC}$).
Cela a été prouvé par Cohen (en 1963), résolvant ainsi le premier de la liste des 23 problèmes de Hilbert.
Expliquons ce problème.

- Les ensembles $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ sont dénombrables (en bijection avec $\mathbb{N}$), ils ont donc le même cardinal, même si ce sont des ensembles infinis. On note ce cardinal $\aleph_0$ (*aleph* est la première lettre de l'alphabet hébreu).
- Les ensembles $\mathbb{R}$, $[0,1]$, $\mathbb{R}^2$, $\mathbb{C}$ sont indénombrables, on note $2^{\aleph_0}$ leur cardinal (voir plus loin pour l'explication de cette notation).
- L'hypothèse du continu affirme qu'il n'y a pas d'autre type d'infini entre $\aleph_0$ et $2^{\aleph_0}$.
- Autrement dit, si on note $\aleph_1$ le premier type d'infini strictement plus grand que $\aleph_0$, alors on a $\aleph_1 = 2^{\aleph_0}$.
- Cohen prouve donc qu'on ne peut pas répondre à cette question, ce qui fait qu'on peut choisir d'accepter ou de refuser $\mathcal{HC}$ indépendamment du reste.

Revenons sur cette notation $2^{\aleph_0}$ pour le cardinal de $\mathbb{R}$.
Pour E, F deux ensembles, on note parfois F^E l'ensemble des fonctions de $f : E \to F$. Cela se justifie par le fait que si E et F sont finis, le nombre de telles fonctions est $\mathrm{Card}(F)^{\mathrm{Card}(E)}$. Par exemple, si $F = \{0,1\}$ et $\mathrm{Card}(E) = n$, alors le nombre de fonctions $f : E \to \{0,1\}$ est 2^n.
Expliquons comment les réels de $[0,1]$ peuvent être décrits par des fonctions de $\mathbb{N} \to \{0,1\}$. Pour $x \in [0,1]$, on calcule l'écriture binaire : $x = 0.100110101\ldots = \sum_{n=1}^{\infty} a_n 2^{-n}$ où $a_n \in \{0,1\}$. C'est comme

l'écriture décimale à virgule mais avec seulement les chiffres 0 et 1. Ainsi, à chaque écriture binaire, on peut associer une fonction $f : \mathbb{N} \to \{0,1\}$ définie par $f(n) = a_n$. Cette correspondance est (à peu près) bijective.
Ainsi le cardinal de $[0,1]$ est celui de $\{0,1\}^{\mathbb{N}}$ qui est $2^{\aleph_0}$ (où $\aleph_0$ désigne le cardinal de $\mathbb{N}$).
C'est aussi l'occasion de rappeler le théorème de Cantor : $\text{Card}(\mathcal{P}(E)) > \text{Card}(E)$ quel que soit l'ensemble E. Il est facile de voir que $\mathcal{P}(\mathbb{N})$ correspond à $\{0,1\}^{\mathbb{N}}$. En effet, $A \subset \mathbb{N}$ définit naturellement un réel $x = 0.0110001\ldots$ où on a 1 à la n-ème place exactement lorsque $n \in A$. Ainsi $[0,1]$ a le même cardinal que $\mathcal{P}(\mathbb{N})$: $2^{\aleph_0}$ et ce cardinal est strictement plus grand que $\aleph_0$.

2.2. Conséquence philosophique

Les théorèmes d'incomplétude sont parmi les rares résultats mathématiques modernes connus par les non-mathématiciens pour leur influence sur certains courants de pensée.
Ces théorèmes rendent indéniable la différence fondamentale entre ce qui est vrai et ce qui est démontrable. Cela renforce le point de vue des platoniciens : la vérité est une réalité absolue, indépendante du monde qui nous entoure.
Les théorèmes de Gödel prouvent aussi que les mathématiques ne sont pas qu'un jeu d'écriture sur des symboles abstraits, en montrant que la syntaxe (ce que l'on peut démontrer) ne correspond pas forcément à toute la sémantique (ce qui est vrai).
Ils questionnent aussi la différence entre humain et machine. Si l'on modélise une machine par une théorie (les axiomes correspondant à des instructions), la machine ne pourra pas sortir de son cadre pour prouver sa propre cohérence. Par exemple, pour la phrase $\mathcal{G}$ de Gödel, qui dit « Je ne suis pas démontrable dans $\mathcal{PA}$ », il faut se placer dans un cadre plus large, par exemple $\mathcal{ZFC}$, pour pouvoir dire que cette phrase est vraie. Est-ce qu'une machine peut englober toute la complexité des raisonnements humains ?

2.3. Ce que ne sont pas les théorèmes d'incomplétude

Complétude et incomplétude.
Les théorèmes de complétude et les théorèmes d'incomplétude sont des théorèmes vrais, prouvés par Gödel, et ne sont pas contradictoires.
On rappelle la complétude, pour toute formule $\mathcal{Q}$:

$$\mathcal{T} \models \mathcal{Q} \quad \text{ssi} \quad \mathcal{T} \vdash \mathcal{Q}$$

Plaçons-nous maintenant dans le cadre d'une théorie $\mathcal{T} = \mathcal{PA}$ (ou contenant $\mathcal{PA}$) ayant pour modèle $\mathbb{N}$.
Le théorème d'incomplétude affirme qu'il existe une formule $\mathcal{Q}$ telle que :

$$\mathcal{T} \not\vdash \mathcal{Q} \quad \text{et} \quad \mathcal{T} \not\vdash \neg\mathcal{Q}$$

Il concerne des théories spécifiques (modélisant $\mathbb{N}$), alors que la complétude vaut quelle que soit la structure $\mathcal{S}$.
L'incomplétude alliée à la complétude prouve que $\mathcal{Q}$ ne peut être vraie pour toutes les structures (car sinon, par la complétude, $\mathcal{Q}$ serait démontrable).
Il existe donc des structures $\mathcal{S}$, différentes de $\mathbb{N}$, dans lesquelles $\mathcal{Q}$ est fausse (donc on n'a pas $\mathcal{T} \models \mathcal{Q}$) même si $\mathcal{S}$ satisfait les axiomes de Peano. On parle de modèle non standard des entiers.

Les mathématiques et la logique sont-elles incohérentes ?
Le second théorème dit qu'une théorie ne peut prouver sa propre cohérence. Cela signifie qu'en restant dans le cadre strict de cette théorie, on ne sera jamais sûr qu'on ne peut pas prouver à la fois une formule $\mathcal{Q}$ et sa négation $\neg\mathcal{Q}$.

Cependant, si on se place dans un cadre plus grand, on peut s'en sortir. Par exemple, on prouve la cohérence de l'arithmétique à partir de la théorie $\mathcal{ZFC}$. Apparaît alors un problème de régression infinie : on ne peut pas être sûr de la cohérence de $\mathcal{ZFC}$. Il faudrait se placer dans une théorie encore plus grande...
Il n'y a donc pas de cadre logique unique et universel au fondement des mathématiques. Le consensus empirique des mathématiques modernes est d'utiliser le cadre $\mathcal{ZFC}$ dans lequel personne n'a trouvé d'incohérence.

3. Plan de travail

Prouver le premier théorème d'incomplétude est un gros travail, assez technique avec quelques idées lumineuses (par exemple la fonction β qui encode une liste quelconque de nombres avec seulement deux entiers !). Même cent ans après, cela reste difficile à aborder. On ne peut être qu'admiratif qu'une seule personne ait fait tout cela.
On présente un par un les prochains chapitres.

3.1. Arithmétique de Peano

On y décrit les 7 axiomes de Peano et le schéma d'axiomes de la récurrence qui régissent le comportement des entiers naturels. On décrira aussi d'autres théories plus faibles (de Presburger, de Robinson) et une théorie plus forte, la théorie de Zermelo-Fraenkel éventuellement augmentée de l'axiome du choix $\mathcal{ZFC}$.

3.2. Codage de Gödel

C'est la partie « arithmétisation de la syntaxe » : il faut coder chaque formule, et même chaque preuve, par un seul entier. On pourra alors donner une preuve informelle, mais qui reste dans l'esprit du premier théorème de Gödel, par un argument de diagonalisation. On en profitera pour motiver cette méthode de diagonalisation via le paradoxe de Russell et l'argument diagonal de Cantor.

3.3. Fonctions récursives primitives

Il s'agit d'un des chapitres techniques. Les fonctions récursives primitives sont les fonctions $f : \mathbb{N}^p \to \mathbb{N}$ « facilement calculables ». Le but est de voir que les opérations sur les formules, par exemple vérifier qu'une suite de formules est bien une preuve, sont facilement calculables.
On abordera la construction de preuves à la Hilbert, plus adaptée que les preuves en déduction naturelle pour un traitement algorithmique.

3.4. Fonctions représentables

Encore un chapitre technique qui fait le passage entre l'univers mathématique des entiers et le langage $\mathcal{L}_0$ de l'arithmétique de Peano. On prouvera que toute fonction récursive primitive peut être traduite par une formule de $\mathcal{L}_0$. On obtient ainsi une formule qui décide si une preuve est valide ou non, ainsi qu'une formule exprimant si un énoncé est démontrable ou pas. C'est dans ce chapitre qu'on expliquera la fonction β de Gödel.

3.5. Diagonalisation

On y expliquera comment obtenir une formule $\mathcal{G}$ qui parle d'elle-même.
On y parlera aussi du paradoxe du menteur et d'un théorème de Tarski qui dit qu'il n'existe pas de formule qui définisse les formules vraies (alors qu'on a vu qu'on sait détecter les formules démontrables).

3.6. Preuve du théorème d'incomplétude

On énoncera et prouvera le premier théorème d'incomplétude d'abord dans le cadre de l'arithmétique de Peano, puis dans le cadre des théories plus larges. La preuve est basée sur le lemme de diagonalisation afin d'obtenir une phrase qui dit « Je ne suis pas démontrable ».

3.7. Le second théorème d'incomplétude

On terminera par l'énoncé du second théorème et on en donnera une preuve sans entrer dans les détails.