Preuve du théorème d'incomplétude

Nous rassemblons tous les outils obtenus afin de prouver le premier théorème d'incomplétude de Gödel.

Remarque : C'est le bon moment pour relire le chapitre « Un aperçu du théorème d'incomplétude » et aussi le chapitre précédent « Diagonalisation ».

1. L'incomplétude de l'arithmétique de Peano

1.1. Énoncé

Théorème 1.
Si $\mathcal{PA}$ est cohérente, alors $\mathcal{PA}$ est incomplète, c'est-à-dire qu'il existe une formule fermée $\mathcal{G}$ telle que :

$$\mathcal{PA} \nvdash \mathcal{G} \qquad \text{et} \qquad \mathcal{PA} \nvdash \neg\mathcal{G}$$

Ainsi il existe des propositions vraies (dans le modèle standard $\mathbb{N}$) mais non démontrables.

- On rappelle que $\mathcal{PA}$ désigne la théorie de l'arithmétique de Peano, c'est-à-dire les axiomes usuels associés aux entiers naturels.
- L'incomplétude signifie qu'il existe une formule fermée $\mathcal{G}$ qui n'est pas démontrable à partir des axiomes de Peano, et telle que $\neg\mathcal{G}$ n'est pas non plus démontrable.
- Comme $\mathcal{G}$ est une formule fermée, alors dans toute interprétation, par exemple le modèle $\mathbb{N}$ ou un autre modèle, exactement l'une des deux possibilités est vérifiée : soit $\mathcal{G}$ est vraie, soit $\neg\mathcal{G}$ est vraie. L'une des deux est vraie et les deux sont indémontrables.
- Pour le modèle standard $\mathbb{N}$, et pour tout autre modèle, il existe donc des propositions vraies (dans ce modèle) et non démontrables (dans $\mathcal{PA}$).

1.2. Cohérence, incomplétude

Rappels sur la cohérence.

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

  $$T \vdash \mathcal{Q} \qquad \text{et} \qquad T \vdash \neg\mathcal{Q}.$$

  Elle est dite ***cohérente*** sinon.
- De façon équivalente, T est incohérente si $T \vdash \bot$ et alors T prouve n'importe quelle formule (et sa négation), par la règle d'explosion. Une théorie incohérente n'a donc aucun intérêt.
- Si T admet un modèle (c'est-à-dire $\models_{\mathcal{S}} T$ pour une certaine structure $\mathcal{S}$), alors T est cohérente via le théorème de validité.

Rappels sur la complétude.

- Une théorie T est ***incomplète*** s'il existe une formule fermée $\mathcal{Q}$ telle que :
$$T \nvdash \mathcal{Q} \qquad \text{et} \qquad T \nvdash \neg\mathcal{Q}.$$
- Elle est dite ***complète*** sinon, c'est-à-dire que pour toute formule fermée $\mathcal{Q}$ on a :
$$T \vdash \mathcal{Q} \qquad \text{ou} \qquad T \vdash \neg\mathcal{Q}.$$

Dans l'énoncé du théorème de Gödel, on part d'une théorie cohérente et on prouve qu'elle est incomplète.

1.3. Hypothèse de cohérence

Quand on avait énoncé le théorème d'incomplétude dans le chapitre « Un aperçu du théorème d'incomplétude », on n'avait pas énoncé l'hypothèse « $\mathcal{PA}$ est cohérente ». Pourquoi ? Il s'agit d'une des subtilités de l'énoncé.
Si on se place dans un monde où les entiers naturels $\mathbb{N}$ existent (par exemple dans $\mathcal{ZFC}$), alors $\mathbb{N}$ est un modèle de $\mathcal{PA}$ (tout est fait pour : les axiomes de $\mathcal{PA}$ correspondent aux propriétés des entiers), donc $\mathcal{PA}$ est cohérente, ce qui rend l'hypothèse inutile.
Mais si on souhaite rester dans le cadre strict du langage $\mathcal{L}_0$ et de la théorie $\mathcal{PA}$, alors l'existence de $\mathbb{N}$ n'est pas acquise, et par conséquent, la cohérence de $\mathcal{PA}$ non plus.
Ce point est d'ailleurs irrésoluble : le second théorème d'incomplétude affirmera qu'une théorie (suffisamment riche) ne peut pas prouver sa propre cohérence.

2. ω-cohérence, être une preuve, être démontrable

2.1. ω-cohérence

On va prouver le théorème de Gödel avec une hypothèse un peu plus forte : l'ω-cohérence (c'est celle de l'article original de Gödel). On discutera un peu plus tard de la façon de s'en passer.

Définition.
Une théorie $\mathcal{T}$ est ***ω-incohérente*** s'il existe une formule $\mathcal{Q}(x)$ (ayant x pour seule variable libre) telle que :
$$\mathcal{T} \quad \vdash \quad \exists x\ \mathcal{Q}(x)$$
$$\text{et pour chaque } n \in \mathbb{N}, \qquad \mathcal{T} \quad \vdash \quad \neg\mathcal{Q}(\overline{n})$$
Sinon elle est dite ***ω-cohérente***.

- C'est donc une situation d'incohérence spécifique aux entiers naturels : on a $\neg\mathcal{Q}(0)$, $\neg\mathcal{Q}(1)$, $\neg\mathcal{Q}(2)$, etc., et pourtant, il existe x tel que $\mathcal{Q}(x)$.
- Dans une théorie ω-cohérente, une telle formule n'existe pas.
- Cette définition sort du langage $\mathcal{L}_0$ car elle fait intervenir les entiers naturels $n = 0$, $n = 1$, …
- Si $\mathcal{T}$ admet $\mathbb{N}$ comme modèle, alors $\mathcal{T}$ est ω-cohérente. Et donc, si on considère que $\mathbb{N}$ est un modèle de $\mathcal{PA}$, alors $\mathcal{PA}$ est ω-cohérente.

Voici le lien avec la cohérence classique :

- $\mathcal{T}$ ω-cohérente implique $\mathcal{T}$ cohérente.
- $\mathcal{T}$ incohérente implique $\mathcal{T}$ ω-incohérente.

Justification : si $\mathcal{T}$ est incohérente, alors $\mathcal{T}$ prouve n'importe quoi, donc prouve les formules de la définition de la ω-incohérence. L'autre sens s'obtient par contraposition.
Ces deux notions ne sont pas équivalentes, mais on verra un peu plus loin l'astuce de Rosser qui permet de remplacer l'hypothèse « ω-cohérente » de Gödel par l'hypothèse plus faible de « cohérente ».

2.2. Être une preuve, être démontrable

Nous revenons sur la distinction fondamentale entre vérifier qu'on a une preuve et décider si une formule est démontrable.

- On rappelle que est_preuve(N, n), qui décide si la suite des formules de super-nombre de Gödel N constitue une preuve de la formule dont le nombre de Gödel est n, est un prédicat récursif primitif.
- Donc par la représentabilité, il existe une formule correspondante dans $\mathcal{L}_0$:

$$\mathcal{E}\text{st-preuve}(y, x)$$

- Par contre, est_démontrable(n) n'est pas un prédicat récursif primitif, car la recherche d'une preuve doit se faire sur un ensemble d'entiers N non borné.
- Mais ce n'est pas grave car dans $\mathcal{L}_0$ il est tout à fait légitime de définir la formule :

$$\mathcal{E}\text{st-démontrable}(x) \ : \quad \exists y \ \ \mathcal{E}\text{st-preuve}(y, x)$$

  Noter que pour obtenir cette formule de $\mathcal{L}_0$, on est passé par l'intermédiaire de la représentabilité de la relation « être une preuve ».

Corollaire 1.
L'ensemble des formules démontrables est définissable.

En effet, par définition, une formule $\mathcal{P}$ est démontrable (dans $\mathcal{PA}$) si et seulement si $\mathcal{E}$st-démontrable($\ulcorner\mathcal{P}\urcorner$) est vraie (sur $\mathbb{N}$).
C'est à mettre en opposition avec la propriété « être vraie » qui n'est pas définissable (voir le théorème de Tarski du chapitre « Diagonalisation »).
Retenez donc ces deux différences cruciales :

- « être une preuve » et « être démontrable » sont des prédicats fondamentalement différents, l'un est facilement calculable, l'autre non.
- « être démontrable » et « être vraie » sont aussi deux concepts fondamentalement différents, l'un est définissable, l'autre pas.

3. Preuve du théorème d'incomplétude pour $\mathcal{PA}$

Voici l'énoncé que l'on va prouver :

Théorème 2.
Si $\mathcal{PA}$ est ω-cohérente, alors $\mathcal{PA}$ est incomplète.

3.1. La formule à diagonaliser

On note $\mathcal{P}(x)$ la formule suivante :

$$\mathcal{P}(x) \ : \quad \neg\mathcal{E}\text{st-démontrable}(x)$$

C'est bien une formule de $\mathcal{L}_0$ d'après ce que l'on a dit précédemment.
On a donc :

$$\mathcal{E}\text{st-démontrable}(x) \ : \quad \exists y \ \ \mathcal{E}\text{st-preuve}(y, x)$$

donc

$$\neg\mathcal{E}\text{st-démontrable}(x) \ : \quad \forall y \ \ \neg\mathcal{E}\text{st-preuve}(y, x)$$

Par le lemme de diagonalisation, il existe une formule fermée $\mathcal{G}$, dite ***formule de Gödel***, telle que :

$$\mathcal{PA} \quad \vdash \quad \big(\mathcal{G} \leftrightarrow \mathcal{P}(\ulcorner\mathcal{G}\urcorner)\big)$$

(On note $\ulcorner\mathcal{G}\urcorner$ à la fois pour le nombre de Gödel de $\mathcal{G}$, mais aussi pour son numéral.)

Ainsi le lemme de diagonalisation fournit dans $\mathcal{PA}$ l'équivalence :

$$\mathcal{G} \leftrightarrow \neg\exists y \ \ \mathcal{E}\text{st-preuve}(y, \ulcorner\mathcal{G}\urcorner)$$

c'est-à-dire :

$$\neg\mathcal{G} \leftrightarrow \exists y \ \ \mathcal{E}\text{st-preuve}(y, \ulcorner\mathcal{G}\urcorner)$$

Autrement dit :

$$\mathcal{G} \leftrightarrow \neg\mathcal{E}\text{st-démontrable}(\ulcorner\mathcal{G}\urcorner)$$

Ainsi $\mathcal{G}$ signifie « Je ne suis pas démontrable ».
On va maintenant montrer que cette formule $\mathcal{G}$ n'est pas démontrable, et que $\neg\mathcal{G}$ ne l'est pas non plus.

3.2. Preuve de $\mathcal{PA} \nvdash \mathcal{G}$

Par l'absurde, si $\mathcal{PA} \vdash \mathcal{G}$, alors il existe une preuve de $\mathcal{G}$ à partir des axiomes de $\mathcal{PA}$. On note N le super-nombre de Gödel associé aux formules qui constituent cette preuve.
On a donc est_preuve$(N, \ulcorner\mathcal{G}\urcorner)$ (en tant que prédicat récursif primitif) donc :

$$\begin{array}{lll} & \mathcal{PA} \vdash & \mathcal{E}\text{st-preuve}(\overline{N}, \ulcorner\mathcal{G}\urcorner) \\ \text{donc} & \mathcal{PA} \vdash & \exists y \ \ \mathcal{E}\text{st-preuve}(y, \ulcorner\mathcal{G}\urcorner) \\ \text{donc} & \mathcal{PA} \vdash & \mathcal{E}\text{st-démontrable}(\ulcorner\mathcal{G}\urcorner) \\ \text{donc} & \mathcal{PA} \vdash & \neg\mathcal{G} \quad \text{par l'équivalence de la diagonalisation.} \end{array}$$

Conclusion : si $\mathcal{PA} \vdash \mathcal{G}$ alors $\mathcal{PA} \vdash \neg\mathcal{G}$, ce qui implique que $\mathcal{PA}$ est incohérente. Or on a supposé $\mathcal{PA}$ ω-cohérente, donc cohérente (cette hypothèse simple suffit pour cette première étape). On obtient une contradiction.

3.3. Preuve de $\mathcal{PA} \nvdash \neg\mathcal{G}$

Supposons par l'absurde que $\mathcal{PA} \vdash \neg\mathcal{G}$, alors par l'équivalence de la diagonalisation on a :

$$\mathcal{PA} \quad \vdash \quad \exists y \ \ \mathcal{E}\text{st-preuve}(y, \ulcorner\mathcal{G}\urcorner)$$

Si on définit $\mathcal{Q}(y)$ par $\mathcal{E}$st-preuve$(y, \ulcorner\mathcal{G}\urcorner)$ cela s'écrit :

$$\mathcal{PA} \quad \vdash \quad \exists y \ \ \mathcal{Q}(y)$$

D'un autre côté, comme on suppose $\mathcal{PA} \vdash \neg\mathcal{G}$ et $\mathcal{PA}$ est cohérente (car ω-cohérente), alors $\mathcal{PA} \nvdash \mathcal{G}$ (ce qu'on a en fait déjà prouvé dans la section précédente).
Cela signifie que pour chaque entier $N = 0, 1, 2, \ldots$, il est faux que N soit une preuve de $\ulcorner\mathcal{G}\urcorner$, autrement dit, il est vrai que ¬est_preuve$(N, \ulcorner\mathcal{G}\urcorner)$. Par la (forte) représentabilité des prédicats récursifs primitifs, on a :

$$\mathcal{PA} \quad \vdash \quad \neg\mathcal{E}\text{st-preuve}(\overline{N}, \ulcorner\mathcal{G}\urcorner)$$

Cela signifie que pour $N = 0, 1, 2, \ldots$, on a

$$\mathcal{PA} \quad \vdash \quad \neg\mathcal{Q}(\overline{N})$$

Cela implique que $\mathcal{PA}$ est ω-incohérente, en contradiction avec notre hypothèse.

3.4. $\mathcal{G}$ est-elle vraie ?

Nous avons prouvé $\mathcal{PA} \nvdash \mathcal{G}$ et $\mathcal{PA} \nvdash \neg\mathcal{G}$. On se place dans un modèle de $\mathcal{PA}$, alors :

- soit la phrase $\mathcal{G}$ est vraie et elle n'est pas démontrable,
- soit la phrase $\neg\mathcal{G}$ est vraie et elle n'est pas non plus démontrable.

Dans tous les cas, il existe des vérités non démontrables !

Plaçons-nous dans la situation classique où $\mathbb{N}$ est considéré comme modèle de $\mathcal{PA}$ (par exemple on se place dans $\mathcal{ZFC}$). D'une part $\mathcal{PA}$ est ω-cohérente (donc cohérente). D'autre part $\mathcal{G}$ est vraie ! En effet $\mathcal{G}$ n'est pas démontrable, ce qui signifie que la formule $\neg\mathcal{E}$st-démontrable($\ulcorner\mathcal{G}\urcorner$) est vraie (dans $\mathbb{N}$). Donc, par l'équivalence de la diagonalisation, $\mathcal{G}$ est vraie.
Encore une fois, cela suppose l'hypothèse que $\mathbb{N}$ est un modèle de $\mathcal{PA}$ (qui sort du cadre de l'arithmétique de Peano). C'est pourquoi Gödel énonce son théorème avec une hypothèse de cohérence et sans notion de vérité afin de rester dans le cadre strict de $\mathcal{PA}$.

3.5. L'astuce de Rosser

Rosser (1936) remplace l'hypothèse d'ω-cohérence de Gödel en la simple hypothèse « $\mathcal{PA}$ est cohérente ».
Au lieu d'utiliser la phrase « Je ne suis pas démontrable » dans la preuve, il utilise la phrase : « Si je suis démontrable alors il existe une preuve plus courte de ma négation ».
Formellement la phrase $\mathcal{P}(x)$ à laquelle il applique le lemme de diagonalisation est :

$$\mathcal{P}(x) : \qquad \forall y \ \ \mathcal{E}\text{st-preuve}(y,x) \ \rightarrow \ \exists z \leqslant y \ \mathcal{E}\text{st-preuve}(z, \text{neg}(x))$$

Si $x = \ulcorner\mathcal{Q}\urcorner$, on a noté $\text{neg}(x) = \ulcorner\neg\mathcal{Q}\urcorner$.
Remarquer la quantification bornée « $\exists z \leqslant y$ », qui évite d'avoir à introduire la ω-cohérence et d'énumérer $N = 0, 1, 2, \ldots$ jusqu'à l'infini.

4. Le théorème d'incomplétude général

4.1. Énoncé

Théorème 3.
Soit $\mathcal{T}$ une théorie cohérente, ayant des axiomes récursifs et qui contient $\mathcal{PA}$. Alors $\mathcal{T}$ est incomplète.

Autrement dit, dans n'importe quelle théorie suffisamment riche, il existe des formules vraies mais non démontrables.
Comme l'arithmétique de Peano est incomplète, on pourrait croire qu'il suffit d'ajouter une formule indémontrable (comme $\mathcal{G}$) à ses axiomes afin de la rendre immédiatement démontrable dans cette nouvelle théorie. Mais le théorème de Gödel prouve que cette nouvelle théorie sera aussi incomplète : il existe une nouvelle formule indémontrable. Il n'y a donc aucun moyen de s'en sortir : il existe toujours des formules vraies et non démontrables.
Revenons sur les hypothèses. On a déjà rappelé ce que signifie « $\mathcal{T}$ cohérente ». L'hypothèse « $\mathcal{T}$ contient $\mathcal{PA}$ » signifie exactement « $\mathcal{PA} \subset \mathcal{T}$ », c'est-à-dire que les axiomes de Peano font partie des axiomes de $\mathcal{T}$.
On dit de façon informelle que $\mathcal{T}$ est « suffisamment riche ». C'est cependant une exigence assez faible en mathématiques que de pouvoir faire des calculs sur les entiers.
Il reste l'hypothèse « les axiomes de $\mathcal{T}$ sont récursifs » que l'on va expliquer ci-dessous.

4.2. Fonctions récursives : la thèse de Church

Nous allons expliquer, sans la définir complètement, la notion de « récursive » qui est une version plus large de « récursive primitive ».
Définition informelle. Une fonction $f : \mathbb{N}^p \rightarrow \mathbb{N}$ est une fonction ***récursive*** si elle est calculable.
C'est donc un ensemble plus grand de fonctions que les fonctions récursives primitives, qui étaient les fonctions « facilement calculables ».
Il existe bien sûr une définition précise de ce qu'est une fonction récursive. C'est une variante plus technique de la notion de fonction récursive primitive avec, en plus, des opérations non bornées. Par

exemple, la fonction d'Ackermann (voir le chapitre « Fonctions récursives primitives ») est une fonction récursive, mais pas récursive primitive.
La « thèse de Church » est l'affirmation heuristique selon laquelle toutes les fonctions calculables sont récursives (et réciproquement).
Le résultat qui nous intéresse est que toute fonction récursive est représentable (comme pour les fonctions récursives primitives).
Terminons nos définitions :

- Un ensemble $A \subset \mathbb{N}$ est un ***ensemble récursif*** (resp. récursif primitif) si sa fonction caractéristique $\chi_A : \mathbb{N} \to \mathbb{N}$ est une fonction récursive (resp. fonction récursive primitive).
- Une théorie $\mathcal{T}$ a des ***axiomes récursifs*** (resp. récursifs primitifs) si $A = \{\ulcorner P \urcorner \mid P \text{ est un axiome de } \mathcal{T}\}$, l'ensemble des nombres de Gödel des axiomes de $\mathcal{T}$, est un ensemble récursif (resp. récursif primitif).

Exiger que l'ensemble des axiomes d'une théorie soit récursif signifie juste qu'on sait calculer si une formule donnée est un axiome ou pas. C'est une exigence extrêmement raisonnable : on doit pouvoir reconnaître les axiomes de notre théorie !

4.3. Preuve du théorème d'incomplétude

Comme on n'a pas donné la définition exacte d'une fonction récursive, on remplace cette hypothèse par l'hypothèse un peu plus forte : « les axiomes de $\mathcal{T}$ sont récursifs primitifs », ce qui fait que l'on sait calculer (facilement) si une formule donnée est un axiome ou pas.
Cela permet donc de construire la fonction prédicat est_preuve$(N, n) : \mathbb{N}^2 \to \mathbb{N}$ récursive primitive, puis de la représenter par une formule de $\mathcal{L}_0$, $\mathcal{E}$st-preuve(y, x). On remplace aussi l'hypothèse de cohérence par celle de ω-cohérence.
La démonstration est alors exactement la même que pour la théorie $\mathcal{PA}$: on construit une formule $\mathcal{G}$ par diagonalisation et on prouve que ni $\mathcal{G}$, ni $\neg\mathcal{G}$ ne sont démontrables.
Si $\mathbb{N}$ est un modèle de la théorie $\mathcal{T}$, alors $\mathcal{G}$ est vraie, bien que non démontrable.

Retour sur la complétude.
On sait par le théorème de complétude que si une formule est vraie dans tous les modèles, alors elle est démontrable. Ce n'est pas du tout une contradiction avec le théorème d'incomplétude, cela signifie juste ici qu'il existe des modèles dans lesquels $\mathcal{G}$ est fausse. Ces modèles sont appelés *modèles non standards* de $\mathcal{PA}$: ils respectent les axiomes de Peano mais ne sont pas égaux à $\mathbb{N}$ (par exemple, ils possèdent des entiers infiniment grands).