Extrait du livre The Princeton companion to mathematics éditeurs Timothy Gowers, June Barrow-Green et Imre Leader, p. 715-717.

V.27. Problèmes et résultats en théorie additive des nombres

Tout nombre pair supérieur à 4 est-il la somme de deux nombres premiers impairs ? Existe-t-il une infinité de nombres premiers p tels que $p + 2$ soit également premier ? Tout entier positif suffisamment grand est-il la somme de quatre cubes ? Ces trois questions sont tous des problèmes non résolus célèbres en théorie des nombres : la première est appelée la conjecture de Goldbach, la seconde est la conjecture des nombres premiers jumeaux (discutée en détail dans THÉORIE ANALYTIQUE DES NOMBRES [IV.2]), et la troisième est un cas particulier du problème de Waring, que nous discuterons plus tard.

Ces trois problèmes appartiennent à un domaine des mathématiques connu sous le nom de théorie additive des nombres. Afin de décrire en termes généraux ce qu'est ce domaine, il est utile de fournir quelques définitions simples. Supposons que A soit un ensemble d'entiers positifs. Alors l'ensemble-somme de A, noté $A + A$, est l'ensemble de tous les $x + y$ tels que x et y (qui peuvent être égaux) appartiennent tous deux à A. Par exemple, si A est l'ensemble $\{1, 5, 9, 10, 13\}$, alors $A + A$ est l'ensemble $\{2, 6, 10, 11, 14, 15, 18, 19, 20, 22, 23, 26\}$. De même, l'ensemble-différence, noté $A - A$, est l'ensemble de tous les $x - y$ tels que x et y appartiennent tous deux à A. Dans l'exemple ci-dessus, $A - A = \{-12, -9, -8, -5, -4, -3, -1, 0, 1, 3, 4, 5, 8, 9, 12\}$.

En utilisant ce langage, nous pouvons énoncer deux de nos trois problèmes de manière très concise. Soit P l'ensemble de tous les nombres premiers impairs et soit C l'ensemble de tous les cubes. Alors la conjecture de Goldbach est l'affirmation que $P + P$ est l'ensemble $\{6, 8, 10, 12, \ldots\}$, et le cas particulier du problème de Waring demande si tout entier suffisamment grand appartient à $C + C + C + C$. La conjecture des nombres premiers jumeaux est légèrement plus compliquée : elle n'affirme pas seulement que 2 appartient à l'ensemble $P - P$, mais qu'il y appartient "une infinité de fois". (De manière analogue, si A est l'ensemble du paragraphe précédent, alors $A - A$ contient le nombre 4 trois fois.)

Ces problèmes sont notoirement difficiles. Cependant, il est remarquable qu'il existe certains problèmes étroitement liés, qui semblent tout aussi difficiles au premier abord, mais qui ont été résolus. Par exemple, le théorème des trois nombres premiers de Vinogradov est l'affirmation que tout entier impair suffisamment grand est la somme de trois nombres premiers impairs. Sans le "suffisamment grand", cela répondrait au problème de Goldbach ternaire, qui demande si tout nombre impair à partir de 9 est une somme de trois nombres premiers impairs. (À quel point "suffisamment grand" est-il grand ? Eh bien, jusqu'à récemment, il fallait que votre nombre ait environ 7 000 000 de chiffres, mais en 2002, ce seuil a été ramené à moins de 1500 chiffres.) Quant au problème de Waring, il est connu que tout entier positif suffisamment grand est une somme de sept cubes. Plus généralement, il semble probable que, pour tout k, tout entier suffisamment grand puisse s'écrire comme une somme d'au plus $100k$ puissances k-ièmes (où 100 est juste un nombre maximum choisi arbitrairement – il est possible que même $4k$ puissances k-ièmes suffisent), et bien qu'une démonstration de ce résultat soit bien au-delà de la technologie mathématique actuelle, on a montré qu'un

peu plus de $k \log k$ puissances k-ièmes suffisent. Puisque $\log k$ est une fonction à croissance très lente, ce résultat est, dans un certain sens, pas trop éloigné d'une solution du problème.

Comment obtient-on des résultats tels que ceux-ci ? Certaines des démonstrations sont assez complexes, nous ne pouvons donc pas donner une réponse complète ici. Cependant, nous pouvons au moins expliquer une idée fondamentale pour beaucoup de ces arguments, à savoir l'utilisation des sommes exponentielles. Illustrons-la en examinant le début de la démonstration du théorème des trois nombres premiers de Vinogradov.

Imaginons donc que nous ayons un très grand entier impair n et que nous souhaitions démontrer qu'il est une somme de trois nombres premiers impairs. Voici un argument qui suggère fortement que notre tâche est impossible : si n est plus de trois fois plus grand que le plus grand nombre premier connu, ce qui peut très bien arriver, alors nous ne pouvons pas produire trois nombres premiers dont la somme est n sans trouver un nouveau nombre premier. En effet, nous pourrions prendre n astronomiquement grand, par exemple $10^{10^{100}}+1$, et alors $\frac{1}{3}n$ serait bien au-delà de tout nombre premier jamais découvert ou susceptible de l'être.

Cet argument est cependant erroné, et l'indice de ce qui ne va pas réside dans le mot "produire". Nous n'avons pas besoin de produire les trois nombres premiers pour montrer qu'ils existent, pas plus qu'Euclide n'avait besoin de spécifier une suite infinie de nombres premiers pour montrer qu'il en existe une infinité. (Pour une démonstration de leur existence, voir [IV.2 § 2].) Mais, pourrait-on demander, quelle alternative pourrait-il y avoir à trouver effectivement trois nombres premiers impairs dont la somme est n ?

Cette question a une réponse magnifiquement simple : nous tenterons de compter, ou plutôt d'estimer, le nombre de triplets p_1, p_2, p_3 de nombres premiers impairs tels que $p_1 + p_2 + p_3 = n$. Si l'estimation que nous parvenons à obtenir est assez grande, et si de plus nous pouvons montrer qu'elle est raisonnablement précise, alors le nombre réel de ces triplets doit également être assez grand. Cela impliquera qu'un tel triplet existe, et ne nous obligera pas à en "produire" un.

Cependant, notre réponse soulève immédiatement une question d'apparence difficile : comment estimer le nombre de tels triplets ? C'est là que les sommes exponentielles entrent en jeu. Nous utiliserons certaines propriétés de la FONCTION EXPONENTIELLE [III.25] pour reformuler notre problème de dénombrement en un problème d'estimation d'une certaine intégrale.

Comme il est d'usage dans ce domaine, écrivons $e(x)$ au lieu de $\mathrm{e}^{2\pi \mathrm{i} x}$. Les deux propriétés fondamentales que nous utiliserons de la fonction $e(x)$ sont que $e(x+y) = e(x)e(y)$ et que $\displaystyle\int_0^1 e(nx)\,\mathrm{d}x = 1$ si $n = 0$, et 0 si n est un autre entier quelconque. Adoptons également la convention que si nous écrivons $\displaystyle\sum_{p \le n}$, nous sommons sur tous les nombres premiers impairs inférieurs ou égaux à n. Définissons maintenant une fonction $F(x)$ par la formule $F(x) = \displaystyle\sum_{p \le n} e(px)$. C'est-à-dire,

$$F(x) = e(3x) + e(5x) + e(7x) + e(11x) + \cdots + e(qx),$$

où q est le plus grand nombre premier inférieur ou égal à n. C'est une somme d'exponentielles -

d'où l'expression "sommes exponentielles". Ensuite, nous considérons le cube de cette fonction :

$$F(x)^3 = (e(3x) + e(5x) + e(7x) + \cdots + e(qx))^3.$$

Lorsque nous développons le membre de droite, nous obtenons la somme de tous les termes de la forme $e(p_1x)e(p_2x)e(p_3x)$, où p_1, p_2 et p_3 sont des nombres premiers compris entre 3 et q.

L'intégrale que nous allons considérer est $\int_0^1 F(x)^3 e(-nx)\,\mathrm{d}x$. D'après notre discussion du paragraphe précédent, nous savons qu'il s'agit de la somme de toutes les intégrales de la forme $\int_0^1 e(p_1x)e(p_2x)e(p_3x)e(-nx)\,\mathrm{d}x$. Or, la première propriété fondamentale de $e(x)$ nous dit que cette dernière intégrale est égale à $\int_0^1 e((p_1+p_2+p_3-n)x)\,\mathrm{d}x$, et la seconde nous dit alors qu'elle vaut 1 si $p_1+p_2+p_3 = n$ et 0 sinon. Par conséquent, lorsque nous sommons sur tous les triplets possibles p_1, p_2, p_3 de nombres premiers impairs inférieurs ou égaux à n, nous obtenons une contribution de 1 pour chaque triplet dont la somme vaut n et 0 pour tous les autres triplets. En d'autres termes, l'intégrale $\int_0^1 F(x)^3 e(-nx)\,\mathrm{d}x$ est exactement égale au nombre de façons d'écrire n comme une somme de trois nombres premiers impairs.

Cela "réduit" notre problème à l'estimation de l'intégrale $\int_0^1 F(x)^3 e(-nx)\,\mathrm{d}x$. Mais la fonction $F(x)$ semble assez difficile à analyser. Est-il vraiment possible d'estimer une expression telle que $\sum_{p\le n} e(px)$, qui mêle nombres premiers et exponentielles ?

Étonnamment, oui. Les détails sont complexes, mais le fait que cela puisse être fait devient moins mystérieux après avoir réfléchi un instant aux sommes exponentielles que nous savons définitivement estimer. Existe-t-il au moins certains ensembles A d'entiers pour lesquels nous pouvons traiter des sommes de la forme $\sum_{a\in A} e(ax)$? Oui : les progressions arithmétiques. Supposons que A soit l'ensemble $\{s, s+d, s+2d, \ldots, s+(m-1)d\}$: c'est-à-dire la progression arithmétique de longueur m et de différence commune d commençant à s. Alors, en utilisant les propriétés fondamentales de $e(x)$, nous trouvons que

$$\begin{aligned}\sum_{a\in A} e(ax) &= e(sx) + e((s+d)x) + \cdots + e((s+(m-1)d)x)\\ &= e(sx) + e(dx)e(sx) + \cdots + e((m-1)dx)e(sx)\\ &= e(sx)\Big(1 + e(dx) + e(dx)^2 + \cdots + e(dx)^{m-1}\Big). \quad (1)\end{aligned}$$

Cette dernière expression est la somme d'une progression géométrique qui commence à $e(sx)$ et a pour raison $e(dx)$. En utilisant la formule standard et les propriétés fondamentales de $e(x)$, nous en déduisons que

$$\sum_{a\in A} e(ax) = \frac{e(sx) - e((s+dm)x)}{1 - e(dx)}.$$

De telles expressions sont utiles car on peut souvent montrer qu'elles sont petites. Supposons, par exemple, que $|1 - e(dx)|$ soit au moins aussi grand qu'une certaine constante c. Nous savons que $|e(sx) - e((s+dm)x)| \le 2$, donc le module du membre de droite est au plus $2/c$. Si c n'est pas trop petit, cela montre qu'il y a une énorme annulation dans la somme $\sum_{a\in A} e(ax)$: nous avons additionné

m nombres de module 1 et obtenu un nombre de module au plus $2/c$.

Pour certaines valeurs de x, nous pouvons utiliser cette observation simple pour nous aider à estimer la somme $\sum_{p\in P} e(px)$. Ce que nous devons faire, c'est exprimer la somme sur P comme une combinaison de sommes sur des progressions arithmétiques, et c'est une chose très naturelle à faire, puisque P est l'ensemble de tous les entiers jusqu'à n qui ne se trouvent pas dans certaines progressions arithmétiques (comme 14, 21, 28, 35, 42, ...). Nous pouvons donc commencer par prendre la somme $\sum_{t=1}^{n} e(tx)$. De là, nous devons soustraire la contribution de tous les entiers pairs, qui est $\sum_{t\le n/2} e(2tx)$. Nous devons également soustraire la contribution des multiples de 3, sauf 3 lui-même. Cette contribution est $\sum_{1<t\le n/3} e(3tx)$. Nous constatons alors que nous avons soustrait deux fois la contribution des multiples de 6, donc nous corrigeons cela en ajoutant $\sum_{t\le n/6} e(6tx)$.

Ce processus peut être poursuivi et conduit à une manière de décomposer la somme sur les nombres premiers en une combinaison de sommes sur des progressions géométriques. Si x n'est pas proche d'un rationnel de petit dénominateur, alors la plupart des raisons communes sont éloignées de 1, donc la plupart des sommes sur les progressions sont petites. Malheureusement, il y en a trop pour que cet argument simple conduise à une estimation utile. Cependant, il existe un argument plus sophistiqué d'une saveur similaire qui fonctionne.

Que se passe-t-il si x est proche d'un rationnel de petit dénominateur ? Par exemple, que pouvons-nous dire de la somme $\sum_{p\le n} e(p/3)$? Ici, nous utilisons des méthodes plus directes : on sait qu'environ la moitié de tous les nombres premiers sont 1 (mod 3) et l'autre moitié sont 2 (mod 3) (voir [IV.2 § 4]), ce qui nous dit que cette somme est approximativement $(|P|/2)(e(p/3) + e(2p/3))$, où $|P|$ désigne la taille de l'ensemble P.

Pour des raisons très similaires, dans le problème de Waring, on se trouve à vouloir connaître des sommes exponentielles telles que $G(x) = \sum_{t=0}^{m} e(t^k x)$. Là encore, on peut parfois les estimer en les réduisant à des sommes de progressions géométriques. C'est plus facile à montrer dans le cas $k = 2$. L'idée est de regarder non pas $G(x)$ directement mais $|G(x)|^2$, dont un rapide calcul montre qu'il est égal à $\sum_{t=0}^{m}\sum_{u=0}^{m} e((t^2-u^2)x)$. Or, $t^2-u^2 = (t+u)(t-u)$, donc nous pouvons changer de variables, en posant $v = t+u$ et $w = t-u$. Cela nous donne la somme $\sum_{(v,w)\in V} e(vwx)$, où V est l'ensemble de tous les (v, w) tels que $(v + w)/2$ et $(v - w)/2$ (qui sont égaux à t et u respectivement) sont tous deux compris entre 0 et m. Pour chaque v, l'ensemble des valeurs possibles de w est une progression arithmétique, donc nous avons décomposé $|G(x)|^2$ en une somme de sommes de progressions géométriques, une pour chaque v.

Jusqu'à présent, nous avons examiné les problèmes dits directs en théorie additive des nombres. Ce sont des problèmes où l'on spécifie un ensemble puis on essaie de comprendre son ensemble-somme

ou son ensemble-différence. Nous n'avons fait qu'effleurer la surface du sujet : d'autres résultats et techniques connexes sont discutés dans [IV.2] (voir en particulier les sections 7, 9 et 11).

Les problèmes directs ont une longue histoire, mais ces dernières années, une autre classe de problèmes, appelés problèmes inverses, est également devenue un centre de recherche important. Ils concernent la vaste question suivante : si l'on vous donne des informations sur un ensemble-somme ou un ensemble-différence, que pouvez-vous déduire sur l'ensemble original ? Nous terminons en décrivant l'un des points culminants de ce type de théorie additive des nombres, appelé théorème de Freiman.

Il n'est pas difficile de démontrer que si A est un ensemble quelconque d'entiers de taille n, alors la taille de $A + A$ doit être comprise entre $2n - 1$ et $n(n + 1)/2$. (La première borne est atteinte si A est une progression arithmétique et la seconde si toutes les sommes que l'on peut former sont distinctes.) Que pouvons-nous dire de A si la taille de $A+A$ est au plus $100n$, ou, plus généralement, au plus Cn pour une constante C qui reste fixée lorsque n tend vers l'infini ?

Supposons que nous puissions trouver une progression arithmétique P de taille au plus $50n$ telle que A soit un sous-ensemble de P. Alors $A + A$ est un sous-ensemble de $P + P$, qui a une taille au plus $100n - 1$. Donc, si A représente deux pour cent d'une progression arithmétique, alors $A + A$ a une taille au plus $100n$. Cependant, il existe d'autres façons de produire de tels ensembles. Supposons, par exemple, que A soit constitué de tous les nombres ayant jusqu'à sept chiffres tels que les troisième, quatrième et cinquième chiffres à partir de la fin soient 0 : c'est-à-dire des nombres comme 35 000 26 ou 99 000 90. Il y en a $100 \times 100 = 10\,000$. Si nous en additionnons deux, nous obtenons un nombre comme 138 00162 ou 141 00068, qui est composé d'un nombre entre 0 et 198, suivi de deux 0, suivi d'un second nombre entre 0 et 198 (écrit avec des 0 devant si nécessaire pour faire trois chiffres). Il y a 199×199 de ces nombres, soit moins de 40 000. Par conséquent, la taille de $A + A$ est inférieure à quatre fois la taille de A. Cependant, A ne remplit pas deux pour cent d'une progression arithmétique P quelconque : une telle progression devrait avoir une différence commune de 1 et inclure à la fois les nombres 0 et 99 000 99, et 10 000 n'est en rien comparable à deux pour cent de 9 900 100.

Cependant, A est un ensemble très structuré : c'est un exemple de progression arithmétique bidimensionnelle. En gros, une progression arithmétique ordinaire, ou unidimensionnelle, est une progression que l'on construit en partant d'un nombre s et en ajoutant répétitivement un autre nombre d, appelé la différence commune. On construit une progression arithmétique bidimensionnelle en utilisant deux "différences communes" d_1 et d_2. C'est-à-dire que l'on a un nombre de départ s et l'on considère les nombres de la forme $s + ad_1 + bd_2$, en précisant que a doit être compris entre 0 et $m_1 - 1$ et b entre 0 et $m_2 - 1$. Notre ensemble A est une progression bidimensionnelle avec $s = 0$, $d_1 = 1$, $d_2 = 100\,000$, et $m_1 = m_2 = 100$.

De manière analogue, on peut définir des progressions de dimension supérieure. Il n'est pas difficile de montrer que si P est une progression de dimension r, alors la taille de $P + P$ est inférieure à 2^r fois la taille de P. Par conséquent, si A est un sous-ensemble de P et que la taille de P est au plus C fois la taille de A, alors la taille de $A + A$ est au plus la taille de $P + P$, qui est au plus $2^r C$ fois la taille de A.

Cela nous dit que si A est un grand sous-ensemble d'une progression arithmétique de basse dimension, alors A a un petit ensemble-somme. Le théorème de Freiman est l'affirmation remarquable que ce sont les seuls ensembles avec de petits ensembles-sommes. C'est-à-dire que si $A+A$ n'est pas beaucoup plus grand que A, alors il doit exister une progression arithmétique de basse dimension P qui contient A et n'est pas beaucoup plus grande que A. Les sommes exponentielles sont également essentielles pour la démonstration de ce théorème. Le théorème de Freiman a eu de nombreuses applications et est susceptible d'en avoir beaucoup d'autres.