Le problème de Prouhet

M. Rigo

4 septembre 2012

Résumé

Si on pense aux nombres, à leur théorie et à l'arithmétique, on fait rapidement face à de nombreuses questions simples à énoncer (elles ne font intervenir que des sommes, des produits ou des puissances de nombres entiers) mais leurs éventuelles solutions peuvent s'avérer redoutables. Dans ce texte relatif à mon intervention du 22 août 2012 au congrès annuel de la SBPMef, on s'intéressera à un problème accessible : partitionner l'ensemble $\{0, 1, 2, ..., 2^{n+1} - 1\}$ en deux sous-ensembles A et B de même taille de telle sorte que les sommes des éléments de A et B soient égales, les sommes des carrés des éléments de A et B soient égales, ..., les sommes des puissances n-ièmes des éléments de A et B soient égales. Par exemple, pour $n = 2$, on trouve $0 + 3 + 5 + 6 = 1 + 2 + 4 + 7$ et $0^2 + 3^2 + 5^2 + 6^2 = 1^2 + 2^2 + 4^2 + 7^2$. On en présentera une solution reposant de façon élégante sur les écritures en base 2 et on s'autorisera quelques digressions : produit de sinus, répétition et chevauchement, jeu d'échecs, composition musicale, ... Ce texte est construit pour être une balade arithmétique amusante et inattendue, pouvant montrer à des élèves ouverts, un peu comme le prétend André Deledicq, que les mathématiques peuvent être jubilatoires.

1 Introduction

Le thème du congrès annuel de la SBPMef "*Si le nombre m'était compté*" m'a décidé à présenter un de mes sujets préférés d'arithmétique en relation directe avec mes thèmes de recherche (la combinatoire des mots et la théorie des langages formels). De nature anodine, la question initiale permet de mettre en lumière une série de connexions intéressantes, dépassant probablement de loin, ce que l'on pourrait présenter devant une classe de terminale. Néanmoins, cette balade arithmétique, que nous espérons amusante et inattendue, peut montrer à des élèves ouverts des perspectives que l'on n'aurait *a priori* pas imaginées. Ce texte ne se veut nullement un article de recherche original et est largement inspiré de l'article [4] des mathématiciens contemporains Jean-Paul Allouche et Jeffrey Shallit qui ont largement œuvré à la dissémination du concept de suite automatique dont il sera notamment question ici.

Comme présenté dans [6], aux alentours de 1750, L. Euler et C. Goldbach avaient déjà observé que les ensembles d'entiers $A = \{a, b, c, a + b + c\}$ et $B = \{a + b, a + c, b + c\}$ sont tels que la somme des éléments de A est égale à celle des éléments de B et qu'il en va de même pour la somme des carrés des éléments de A et de B. Le ton de cet article est donné par le mémoire présenté en 1851 par Eugène Prouhet [18] lorsqu'il y résout le problème suivant : "*n et m étant*

deux nombres entiers quelconques, il existe une infinité de suites de n^m nombres, susceptibles de se partager en n groupes de n^{m-1} termes chacun et tels que la somme des puissances k des termes soit la même pour tous les groupes, k étant un nombre entier inférieur à m.".

Nous allons, dans un premier temps, considérer le problème suivant comme l'avaient étudié les mathématiciens Tarry et Escott au début du siècle passé. Il s'agit d'équations diophantiennes, autrement dit, d'équations pour lesquelles nous cherchons uniquement des solutions en nombres entiers.

Problème 1. Soient k, n deux entiers. Trouver deux ensembles *distincts* d'entiers $A = \{a_1, \ldots, a_n\}$ et $B = \{b_1, \ldots, b_n\}$ tels que $\{a_1, \ldots, a_n\} =_k \{b_1, \ldots, b_n\}$, i.e., tels que

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i, \quad \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} b_i^2, \quad \cdots, \ \sum_{i=1}^{n} a_i^k = \sum_{i=1}^{m} b_i^k.$$

On peut déjà remarquer que les ensembles A et B peuvent être supposés disjoints (c'est-à-dire d'intersection vide) car s'ils contenaient un élément commun, on pourrait supprimer cet élément des deux ensembles et conserver une solution au problème. C'est la raison pour laquelle on recherche le plus souvent une solution de *taille* n minimale pour un *degré* k donné, si tant est qu'une telle solution existe (cf. proposition 7).

Exemple 2. Pour le degré $k = 1$, il est facile de construire des solutions. En effet, en considérant 4 entiers consécutifs $4n, 4n+1, 4n+2$ et $4n+3$, il suffit de placer $4n$ et $4n+3$ dans un ensemble et $4n+1$ et $4n+2$ dans l'autre ensemble. On peut ainsi fournir des solutions de taille $2, 4, 6, \ldots$.

Exemple 3. Pour le degré $k = 2$, il faut ruser un peu plus. Si on remarque que la différence entre deux carrés consécutifs est de la forme $(2n+1)^2 - (2n)^2 = 4n+1$, alors on trouve une solution en considérant cette fois 8 entiers consécutifs $8n, \ldots, 8n+7$ et en plaçant $8n, 8n+3, 8n+5, 8n+6$ dans un ensemble et $8n+1, 8n+2, 8n+4, 8n+7$ dans l'autre. On peut ainsi fournir des solutions de taille $4, 8, 12, \ldots$.

Exemple 4. L'exemple que l'on trouve sur `Wikipedia` donne une solution de taille $n = 6$ pour un degré $k = 5$,

$$A = \{0, 5, 6, 16, 17, 22\} \text{ et } B = \{1, 2, 10, 12, 20, 21\}$$

et on vérifie que

$$\sum_{a\in A} a = 66 = \sum_{b\in B} b, \quad \sum_{a\in A} a^2 = 1090 = \sum_{b\in B} b^2, \quad \sum_{a\in A} a^3 = 19\,998 = \sum_{b\in B} b^3,$$

$$\sum_{a\in A} a^4 = 385\,234 = \sum_{b\in B} b^4, \quad \sum_{a\in A} a^5 = 7\,632\,966 = \sum_{b\in B} b^5.$$

2 Quelques propriétés

Maintenant que le problème qui va nous intéresser est bien posé, nous décrivons dans cette section quelques propriétés de ses solutions. La première d'entre elles nous montre que si l'on dispose d'une solution, on peut alors en construire une infinité par dilatation et translation.

Proposition 5. *Si* $\{a_1, \ldots, a_n\} =_k \{b_1, \ldots, b_n\}$*, alors, pour des entiers* $N \geq 1$ *et* M*, on a aussi* $\{Na_1 + M, \ldots, Na_n + M\} =_k \{Nb_1 + M, \ldots, Nb_n + M\}$*.*

La preuve de ce résultat ne nécessite que très peu de concepts (binôme de Newton, manipulation de symboles sommatoires et preuve par récurrence) et devrait donc être accessible, pour le moins, à des étudiants des sections les plus avancées en mathématique (et pourquoi pas comme exercice de renforcement).

Démonstration. Fixons deux entiers $N \geq 1$ et M. On procède par récurrence sur $k \geq 1$. Pour le cas de base, si $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$, il est alors clair que

$$\sum_{i=1}^{n}(Na_i + M) = \sum_{i=1}^{n}(Nb_i + M).$$

Il suffit en effet de distribuer et de mettre en évidence.

Supposons à présent qu'il existe $k \geq 1$ tel que, si $\{a_1, \ldots, a_n\} =_k \{b_1, \ldots, b_n\}$, alors $\{Na_1 + M, \ldots, Na_n + M\} =_k \{Nb_1 + M, \ldots, Nb_n + M\}$. Supposons en outre que $\{a_1, \ldots, a_n\} =_{k+1} \{b_1, \ldots, b_n\}$. Il nous faut montrer que

$$\{Na_1 + M, \ldots, Na_n + M\} =_{k+1} \{Nb_1 + M, \ldots, Nb_n + M\}.$$

On a déjà $\{Na_1 + M, \ldots, Na_n + M\} =_k \{Nb_1 + M, \ldots, Nb_n + M\}$, par hypothèse de récurrence. Il reste donc à vérifier que

$$\sum_{i=1}^{n}(Na_i + M)^{k+1} = \sum_{i=1}^{n}(Nb_i + M)^{k+1}.$$

Or, si on applique le binôme de Newton, il vient (en permutant les sommes à la dernière étape)

$$\sum_{i=1}^{n}(Na_i + M)^{k+1} = \sum_{i=1}^{n}\sum_{j=0}^{k+1} \mathrm{C}_{k+1}^{j}\, N^j\, a_i^j\, M^{k+1-j}$$

$$= \sum_{j=0}^{k+1} \mathrm{C}_{k+1}^{j}\, N^j\, M^{k+1-j} \sum_{i=1}^{n} a_i^j.$$

Or, puisque $\{a_1, \ldots, a_n\} =_{k+1} \{b_1, \ldots, b_n\}$, cette dernière expression est aussi égale à

$$\sum_{j=0}^{k+1} \mathrm{C}_{k+1}^{j}\, N^j\, M^{k+1-j} \sum_{i=1}^{n} b_i^j$$

et on retrouve donc l'expression attendue en procédant au même calcul. □

Voici deux autres propriétés des plus intéressantes car elles assurent l'existence d'une solution au problème quel que soit le degré k que l'on se fixe et montrent qu'une telle solution est de taille au moins $k + 1$.

Proposition 6. *Pour tout degré* k *fixé, si l'on dispose d'une solution de taille* n *pour le problème 1, alors* $n \geq k + 1$*.*

Proposition 7. *Pour tout degré* k *fixé, il existe une solution pour le problème 1. De plus, on peut trouver une solution de taille ne dépassant pas* $\frac{k(k+1)}{2} + 1$*.*

Sans pour autant reposer sur des résultats difficiles, nous avons pris le parti de ne pas démontrer cette dernière propriété ; les techniques sont similaires à celles développées plus bas. Le lecteur pourra se référer à [6] dont on trouve facilement une version en ligne.

La démonstration de la proposition 6 permet de rappeler qu'un polynôme est, à une constante multiplicative près, entièrement caractérisé par ses zéros et leur multiplicité [1]. Dès lors, un polynôme unitaire (i.e., dont le coefficient dominant vaut 1) peut aussi bien être défini par ses coefficients que par ses zéros et leur multiplicité. En guise d'introduction, on se souviendra des formules donnant la somme et le produit des zéros d'un polynôme de degré 2 à partir de ses coefficients :

$$\text{si } aX^2+bX+c = a(X-x_1)(X-x_2), \text{ alors } x_1+x_2 = -b/a \text{ et } x_1x_2 = c/a.$$

D'une manière générale, on dispose des formules de Viète qui permettent d'exprimer la somme des zéros d'un polynôme, la somme des produits 2 à 2 des zéros, etc. à partir de ses coefficients.

Proposition 8. *Soient $a_1, \ldots, a_n$ des nombres complexes. On a*

$$\prod_{i=1}^{n}(X-a_i) = \sum_{j=0}^{n}(-1)^j\ C_j\ X^{n-j} \quad \textit{avec}\ \ C_0 = 1$$

$$C_1 = a_1+\cdots+a_n, \quad C_2 = \sum_{1\le i<j\le n} a_ia_j,\ \ \ldots,\ C_n = a_1\cdots a_n.$$

Démonstration. Pour prouver ce résultat, il suffit de distribuer le produit de facteurs et d'identifier les coefficients. □

Dans le problème 1, nous nous intéressons à des sommes de puissances. Si les nombres complexes $a_1, \ldots, a_n$ introduits dans les formules de Viète sont sous-entendus, nous usons de la notation $S_j := a_1^j + \cdots + a_n^j$. Dès le milieu du 17e siècle, Girard (1629), puis de façon indépendante Newton (1666), s'intéressaient déjà à exprimer la somme des j-ièmes puissances des zéros d'un polynôme à partir des sommes des i-ièmes puissances, $i < j$, et des coefficients de celui-ci. Ainsi, on peut déterminer de façon effective la somme des j-ièmes puissances des zéros d'un polynôme sans pour autant connaître ces zéros. Avec un peu de théorie des polynômes symétriques et en utilisant les notations de la proposition 8, on peut montrer que l'on dispose des relations suivantes (d'ailleurs connues sous le nom d'*identités de Newton*). Pour une preuve, on pourra consulter [16] ou [12] reposant sur quelques résultats d'algèbre linéaire.

$$\begin{array}{ll}
S_1 = C_1 & C_1 = S_1 \\
S_2 = C_1S_1 - 2C_2 & C_2 = (-S_2 + C_1S_1)/2 \\
S_3 = C_1S_2 - C_2S_1 + 3C_3 & C_3 = (S_3 - C_1S_2 + C_2S_1)/3 \\
S_4 = C_1S_3 - C_2S_2 + C_3S_1 - 4C_4 & C_4 = (-S_4 + C_1S_3 - C_2S_2 + C_3S_1)/4 \\
\vdots & \vdots
\end{array}$$

1. Il est donc intéressant d'introduire les nombres complexes dans l'enseignement secondaire pour avoir au moins accès au théorème fondamental de l'algèbre stipulant qu'un polynôme de degré d possède exactement d zéros (complexes) comptés avec leur multiplicité.

Exemple 9. Considérons le polynôme X^3-2X^2+X-1 possédant un zéro réel (non entier) et deux zéros complexes conjugués. Avec les notations ci-dessus, on a $C_1=2$, $C_2=1$ et $C_3=1$. La somme S_1 des trois zéros vaut $C_1=2$, la somme S_2 des carrés des zéros vaut $C_1S_1-2C_2=2.2-2.1=2$, la somme S_3 des cubes des zéros vaut $C_1S_2-C_2S_1+3C_3=2.2-1.2+3.1=5$. On peut continuer de la sorte (en posant ici $C_4=C_5=\cdots=0$), la somme S_4 des puissances 4-ièmes des zéros vaut $C_1S_3-C_2S_2+C_3S_1=2.5-1.2+1.2=10$.

Remarque 10. *En particulier, on remarque que le coefficient C_j du polynôme $\prod_{i=1}^{n}(X-a_i)=\sum_{j=0}^{n}(-1)^j\ C_j\ X^{n-j}$ dépend uniquement de $S_1,\ldots,S_j$ et des coefficients précédents C_i, $i<j$.*

Il est à présent très facile de démontrer le résultat technique suivant [6].

Lemme 11. *Soient n,k tels que $n\geq k$. On a $\{a_1,\ldots,a_n\}=_k\{b_1,\ldots,b_n\}$ si et seulement si*

$$\deg\left(\prod_{i=1}^{n}(X-a_i)-\prod_{i=1}^{n}(X-b_i)\right)\leq n-(k+1).$$

Démonstration. Avec les nombres $a_1,\ldots,a_n,b_1,\ldots,b_n$, on définit les quantités $S_1,\ldots,S_k,S'_1,\ldots,S'_k,C_0,\ldots,C_n,C'_0,\ldots,C'_n$ suivantes

$$S_j=a_1^j+\cdots+a_n^j,\quad P(X)=\prod_{i=1}^{n}(X-a_i)=\sum_{j=0}^{n}(-1)^j\,C_j\,X^{n-j},$$

$$S'_j=b_1^j+\cdots+b_n^j,\quad P'(X)=\prod_{i=1}^{n}(X-b_i)=\sum_{j=0}^{n}(-1)^j\,C'_j\,X^{n-j}.$$

De là, on trouve

$$\begin{array}{rcl}\{a_1,\ldots,a_n\}=_k\{b_1,\ldots,b_n\} & \Leftrightarrow & S_1=S'_1,\ldots,S_k=S'_k\\ & \Leftrightarrow & C_1=C'_1,\ldots,C_k=C'_k\end{array}$$

où la dernière équivalence découle de la remarque 10. En outre, si les deux polynômes de degré n ont les mêmes premiers coefficients dominants $C_0,\ldots,C_k$, leur différence $P-P'$ est un polynôme de degré au plus $n-(k+1)$. □

Nous pouvons à présent prouver la proposition 6.

Démonstration. Soit k fixé et supposons disposer d'une solution formée de deux ensembles distincts de taille $n\leq k$. Quitte à compléter (en ajoutant, si nécessaire, le même élément aux deux ensembles), on peut supposer avoir deux ensembles différents de taille exactement k tels que

$$\{a_1,\ldots,a_k\}=_k\{b_1,\ldots,b_k\}.$$

Si on applique le lemme précédent (quand $n=k$, le seul polynôme de degré négatif étant 0), les deux polynômes unitaires

$$\prod_{i=1}^{k}(X-a_i)=\sum_{j=0}^{k}(-1)^j\,C_j\,X^{k-j}\ \text{ et }\ \prod_{i=1}^{k}(X-b_i)=\sum_{j=0}^{k}(-1)^j\,C'_j\,X^{k-j}$$

ont exactement les mêmes coefficients. Ils sont donc égaux et ont les mêmes zéros. De là, on arrive à l'absurdité $\{a_1,\ldots,a_k\}=\{b_1,\ldots,b_k\}$. □

3 Une solution constructive

La proposition 7 garantit l'existence d'une solution, mais la preuve n'est pas constructive. Nous présentons, pour tout degré k, une solution de taille $n = 2^k$, avec la contrainte supplémentaire de partitionner l'ensemble $\{0, \ldots, 2^{k+1} - 1\}$.

Exemple 12. Pour $k = 1$, $n = 2$, $A \cup B = \{0, 1, 2, 3\}$, on prend

$$A = \{0, 3\} \text{ et } B = \{1, 2\},$$

$$0 + 3 = 1 + 2.$$

Exemple 13. Pour $k = 2$, $n = 4$, $A \cup B = \{0, 1, 2, 3, 4, 5, 6, 7\}$, on prend

$$A = \{0, 3, 5, 6\} \text{ et } B = \{1, 2, 4, 7\},$$

$$0 + 3 + 5 + 6 = 14 = 1 + 2 + 4 + 7,$$
$$0^2 + 3^2 + 5^2 + 6^2 = 70 = 1^2 + 2^2 + 4^2 + 7^2.$$

On peut vérifier, à l'aide de l'ordinateur, qu'il s'agit de l'unique solution partitionnant $\{0, \ldots, 7\}$.

Exemple 14. Pour $k = 3$, $n = 8$, $A \cup B = \{0, \ldots, 15\}$, on prend

$$A = \{0, 3, 5, 6, 9, 10, 12, 15\} \text{ et } B = \{1, 2, 4, 7, 8, 11, 13, 14\},$$

$$0 + 3 + 5 + 6 + 9 + 10 + 12 + 15 = 60 = 1 + 2 + 4 + 7 + 8 + 11 + 13 + 14,$$
$$0^2 + 3^2 + 5^2 + 6^2 + 9^2 + 10^2 + 12^2 + 15^2 = 620 = 1^2 + 2^2 + 4^2 + \cdots + 14^2,$$
$$0^3 + 3^3 + 5^3 + 6^3 + 9^3 + 10^3 + 12^3 + 15^3 = 7200 = 1^3 + 2^3 + 4^3 + \cdots + 14^3.$$

Encore une fois, c'est l'unique solution partitionnant $\{0, \ldots, 15\}$.

Notons $r_2(n)$ l'écriture en base 2 de l'entier n.

n	1	2	3	4	5	6	7	8	9
$r_2(n)$	1	10	11	100	101	110	111	1000	1001

Si on reprend l'exemple ci-dessus, les écritures en base 2 des éléments de A sont

$$0, 11, 101, 110, 1001, 1010, 1100, 1111$$

et pour les éléments de B,

$$1, 10, 100, 111, 1000, 1011, 1101, 1110.$$

On peut remarquer que les éléments de A sont exactement ceux possédant une écriture en base 2 contenant un nombre pair de 1. Cela nous amène au théorème suivant.

Théorème 15. *Soit $k \geq 1$. Soient $A_k \subset \{0, \ldots, 2^{k+1}-1\}$ l'ensemble des entiers dont l'écriture en base 2 a un nombre pair de 1 et $B_k = \{0, \ldots, 2^{k+1} - 1\} \setminus A_k$. On a $A_k =_k B_k$.*

Remarque 16. *Les exemples 12, 13 et 14 pourraient à tort faire penser que A_k et B_k ainsi définis fournissent l'unique solution au problème pour l'ensemble $\{0,\dots,2^{k+1}-1\}$. Cependant, à partir des travaux de David Boyd [7], on trouve d'autres solutions que celles données par le théorème 15. Déjà pour $k=5$, $n=32$, $A\cup B=\{0,\dots,63\}$, on peut aussi prendre*

$$\begin{aligned} A &= \{0,1,6,7,9,12,14,15,16,17,18,21,24,27,30,31,32,33,36,39,42,\\ &\quad 45,46,47,48,49,51,54,56,57,62,63\},\\ B &= \{2,3,4,5,8,10,11,13,19,20,22,23,25,26,28,29,34,35,37,38,40,\\ &\quad 41,43,44,50,52,53,55,58,59,60,61\} \end{aligned}$$

et on vérifie que

$$\sum_{a\in A} a = 1\,008 = \sum_{b\in B} b,\quad \sum_{a\in A} a^2 = 42\,672 = \sum_{b\in B} b^2,\quad \sum_{a\in A} a^3 = 2\,032\,128 = \sum_{b\in B} b^3,$$

$$\sum_{a\in A} a^4 = 103\,223\,568 = \sum_{b\in B} b^4,\quad \sum_{a\in A} a^5 = 5\,461\,682\,688 = \sum_{b\in B} b^5.$$

Les ensembles

$$\begin{aligned} A' &= \{0,3,5,7,9,10,12,13,15,18,21,24,27,28,29,30,33,34,35,36,39,\\ &\quad 42,45,48,50,51,53,54,56,58,60,63\},\\ B' &= \{1,2,4,6,8,11,14,16,17,19,20,22,23,25,26,31,32,37,38,40,41,\\ &\quad 43,44,46,47,49,52,55,57,59,61,62\} \end{aligned}$$

fournissent exactement le même résultat.

Puisque la parité du nombre de 1 dans l'écriture est la clé du résultat, la notation suivante nous sera bien utile, $\sigma_2(n)$ désigne la somme des chiffres de l'écriture en base 2 de n. Ainsi, le tableau suivant reprend les premières valeurs de $\sigma_2(n)$

n	0	1	2	3	4	5	6	7	8	9
$r_2(n)$		1	10	11	100	101	110	111	1000	1001
$\sigma_2(n)$	0	1	1	2	1	2	2	3	1	2
$(-1)^{\sigma_2(n)}$	1	−1	−1	1	−1	1	1	−1	−1	1

Remarque 17. *Pour les entiers $0,\dots,2^{k+1}-1$, exactement la moitié (i.e., 2^k) des écritures en base 2, contiennent un nombre pair de 1. En effet, on peut procéder par récurrence sur k. Il suffit de remarquer que l'on obtient les écritures binaires des entiers $2^k,\dots,2^{k+1}-1$ en ajoutant un 1 "de tête" (suivi d'un nombre convenable de 0) aux écritures des entiers $0,\dots,2^k-1$.*

$$\begin{array}{cccc} 0 & 0 & \to & 4 & 100\\ 1 & 1 & \to & 5 & 101\\ 2 & 10 & \to & 6 & 110\\ 3 & 11 & \to & 7 & 111 \end{array}$$

La preuve suivante s'inspire largement de [19] (voir aussi [11]).

Démonstration. Soient $k \geq 1$ et A_k et B_k les ensembles définis dans l'énoncé du théorème. Nous devons montrer (et le lecteur attentif remarquera que l'on s'autorise à prendre $r = 0$) que

$$\forall r \in \{0, \ldots, k\},\ \sum_{a \in A_k} a^r = \sum_{b \in B_k} b^r.$$

Il est équivalent de montrer, pour tout $r \in \{0, \ldots, k\}$ que

$$\sum_{a \in A_k} a^r - \sum_{b \in B_k} b^r = 0$$

ou encore, en remarquant que chaque entier de l'ensemble $\{0, \ldots, 2^{k+1} - 1\}$ apparaît une unique fois soit dans A_k soit dans B_k et que les éléments de B_k ont précisément une écriture binaire contenant un nombre impair de 1, l'égalité précédente se réécrit

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n^r = 0.$$

Introduisons un polynôme $F_k(X)$ dont la factorisation découle précisément de l'écriture unique d'un entier en base 2 (c'est un passage délicat).

$$F_k(X) = \sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, X^n = \prod_{j=0}^{k} (1 - X^{2^j}).$$

Exemple 18. Pour $k = 1, 2$, ces polynômes sont donnés par

$$F_1(X) = 1 - X - X^2 + X^3 = (1 - X)(1 - X^2),$$

$$F_2(X) = 1 - X - X^2 + X^3 - X^4 + X^5 + X^6 - X^7 = (1 - X)(1 - X^2)(1 - X^4).$$

Par exemple, remarquez que $r_2(6) = 110$ car $6 = 4+2$ et que la seule façon d'obtenir un terme en X^6 dans $F_2(X)$ est précisément de multiplier $1.(-X^2).(-X^4)$.

Puisque $1 - X^n = (1 - X)(1 + X + X^2 + \cdots + X^{n-1})$, on en conclut que

$$F_k(X) = (1 - X)^{k+1}\, G_k(X) \text{ avec } G_k \in \mathbb{Z}[X].$$

De cette factorisation (autrement dit, puisque 1 est zéro de $F_k(X)$ de multiplicité au moins $k + 1$), pour tout $j \in \{0, \ldots, k\}$, en dérivant j fois, on a

$$(\mathrm{D}^j F_k)(1) = 0$$

c'est-à-dire, puisque $\mathrm{D}^j X^n = n(n-1)\cdots(n-j+1)X^{n-j}$, l'évaluation de la dérivée en 1 donne

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n(n-1)\cdots(n-j+1) = 0. \tag{1}$$

Pour rappel, il faut démontrer, pour tout $r \in \{0, \ldots, k\}$,

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n^r = 0.$$

On procède alors par récurrence sur r. Le cas $r = 0$ découle directement de la remarque 17. Supposons que, pour tout $r \in \{0, \dots, j-1\}$,

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n^r = 0.$$

Il reste à vérifier cette relation pour $r = j$. Vu (1), on sait que

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)} \underbrace{n(n-1)\cdots(n-j+1)}_{=n^j+\sum_{t=0}^{j-1}\alpha_t\, n^t} = 0$$

pour certains coefficients $\alpha_0, \dots, \alpha_{j-1}$ obtenus en distribuant le produit de facteurs. De là, on trouve

$$\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n^j + \sum_{t=0}^{j-1} \alpha_t \underbrace{\sum_{n=0}^{2^{k+1}-1} (-1)^{\sigma_2(n)}\, n^t}_{=0} = 0$$

et il suffit d'appliquer l'hypothèse de récurrence aux sommes des puissances t-ièmes d'entiers, $t < j$. □

Corollaire 19. *Avec les notations du théorème précédent. Pour tout polynôme P de degré au plus k, on a*

$$\sum_{a \in A_k} P(a) = \sum_{b \in B_k} P(b).$$

De façon informelle, un automate (fini déterministe) est une machine — un algorithme — lisant séquentiellement des mots en entrée, une lettre à la fois de gauche à droite, et dont le but est d'accepter ou de rejeter le mot fourni (pour plus de détails, voir par exemple [21, 22]). La figure 1 décrit à quoi ressemble l'automate acceptant les mots formés de 0 et de 1 contenant exactement un nombre pair de 1. On débute la lecture d'un mot dans l'état marqué d'une flèche entrante et on bascule d'un état à l'autre à la lecture des lettres successives du mot d'entrée tout en respectant les transitions données par les arcs. Autrement dit, cette machine permet de décider si un entier écrit en base 2 appartient à l'ensemble A_k ou B_k du théorème 15.

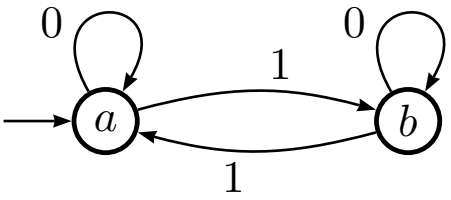

FIGURE 1 – Un automate fini déterministe.

En particulier, il n'y a ici aucune contrainte sur l'entier fourni en entrée. On définit donc une suite $(x_n)_{n\geq 0}$ telle que $x_n = a$ si $r_2(n)$ contient un nombre pair de 1 (resp. $x_n = b$ si $r_2(n)$ contient un nombre pair de 1),

$$(x_n)_{n\geq 0} = abba|baab|baababba|baababbaabbabaab|\cdots. \tag{2}$$

Les 2^{k+1} premiers termes de $(x_n)_{n\geq 0}$ (ou encore son préfixe) fournit la partition de l'ensemble $\{0,\ldots,2^{k+1}-1\}$ du théorème 15. Cette suite infinie, on parle aussi de mot infini sur l'alphabet $\{a,b\}$, est généralement appelée *mot de Thue–Morse*, ou à juste titre *suite de Prouhet–Thue–Morse*. On la baptise ainsi car elle fut (re)découverte respectivement au début du vingtième siècle et dans les années 1920 par les mathématiciens Axel Thue (1863–1922) et Marston Morse (1892–1977) dans des contextes très différents [24, 25, 17]. Cette suite obtenue à partir d'un automate fini auquel on fournit des écritures en base entière est un célèbre exemple de ce que l'on appelle *suite automatique* [5].

Remarque 20. *Lorsque l'on dispose de l'écriture binaire de n, on obtient facilement celle de $2n$ ou de $2n+1$ en ajoutant simplement un 0 ou un 1 à la droite de $r_2(n)$. Ainsi, $r_2(2n) = r_2(n)0$ et $r_2(2n+1) = r_2(n)1$. Au vu de la définition du mot de Thue–Morse* (2)*, on a directement les relations suivantes*

$$\textit{si } x_n = a, \ \textit{alors }, x_{2n} = a, \ x_{2n+1} = b,$$

$$\textit{si } x_n = b, \ \textit{alors }, x_{2n} = b, \ x_{2n+1} = a.$$

La suite de Thue-Morse ne peut donc pas être (purement) périodique. Il n'existe pas d'entier $p > 0$ tel que pour tout $n \geq 0$, $x_n = x_{n+p}$. En effet, supposons que $m+1$ soit une hypothétique période du mot, dès lors, $x_m = x_{2m+1}$ ce qui est impossible au vu des relations décrites ci-dessus.

4 Un exercice d'étude de signe

Dans un séminaire donné en 1999 à l'ULB, pour introduire son sujet, Jean-Paul Allouche proposait un petit exercice d'étude de signe se rapportant à un produit de sinus. Pour plus sur le sujet, consulter [3]. Bien que disposant d'une formule simple pour le produit d'itérations de duplication de cosinus,

$$\cos(x)\cos(2x)\cos(4x)\cdots\cos(2^n x) = \frac{\sin(2^{n+1}x)}{2^{n+1}\sin(x)},$$

on ne disposait pas d'analogue dans le cas de sinus. Ainsi, Allouche propose en particulier d'étudier le signe, sur l'intervalle $[0,\pi]$ de la fonction

$$F_n(x) = \sin(x)\sin(2x)\sin(4x)\cdots\sin(2^n x).$$

Pour $n = 4$, on a le graphique de la Figure 2. Les distributions des signes sur l'intervalle $[0,\pi]$, pour les premières valeurs de n, sont données dans le tableau suivant.

$n=0$	$+$							
$n=1$	$+$	$-$						
$n=2$	$+$	$-$	$-$	$+$				
$n=3$	$+$	$-$	$-$	$+$	$-$	$+$	$+$	$-$

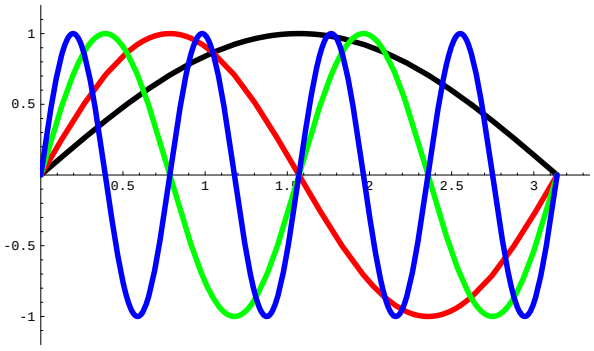

FIGURE 2 – Le produit $\sin(x)\sin(2x)\sin(4x)\sin(8x)$.

Connaissant la distribution des signes à l'étape n, il est facile d'en déduire la distribution à l'étape $n+1$. La fonction $\sin(2^{n+1}x)$ a une période $\pi/2^n$ et sur une période, cette fonction est d'abord positive, puis négative. De plus $F_{n+1}(x)$ étant égal à $\sin(2^{n+1}x)\,F_n(x)$, alors si $F_n(x)$ est positif (resp. négatif) sur un intervalle de longueur $\pi/2^n$, alors $F_{n+1}(x)$ sera positif puis négatif (resp. négatif puis positif).

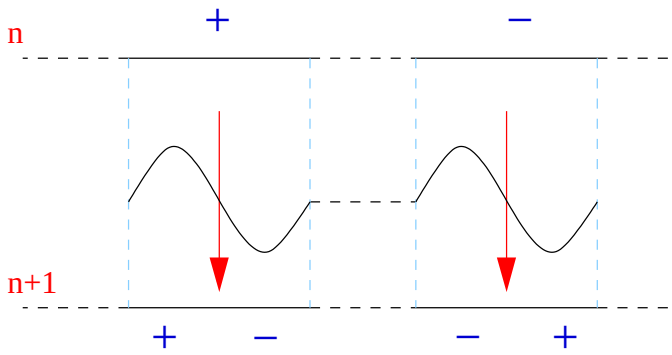

FIGURE 3 – Comment passer de F_n à F_{n+1}.

On dispose dès lors d'une règle, d'une fonction, facile à itérer :

$$\begin{cases} + & \mapsto & +- \\ - & \mapsto & -+ \end{cases}.$$

Remplaçons les signes par des lettres a et b. On peut alors définir une suite en *itérant un morphisme* (le produit est ici la concaténation de mots)

$$f : a \mapsto ab, \quad b \mapsto ba.$$

En itérant f à partir de a, on obtient la suite de mots finis suivantes,

$$\begin{array}{rcl} f(a) & = & ab \\ f^2(a) & = & abba \\ f^3(a) & = & abbabaab \\ f^4(a) & = & abbabaabbaababba \\ f^5(a) & = & abbabaabbaababbabaababbaabbabaab. \end{array}$$

Si l'on définit une distance sur les mots infinis traduisant le fait que deux mots infinis sont d'autant plus proches qu'ils ont un long préfixe commun (et en complétant les mots finis par la répétition d'une nouvelle lettre pour en faire des mots infinis), on peut alors donner un sens tout à fait rigoureux au fait que la suite de mots finis $(f^n(a))_{n\geq 0}$ converge vers un mot infini limite. Ici, chaque mot $f^n(a)$ est préfixe du suivant $f^{n+1}(a)$.

Le lecteur attentif ne manquera pas d'observer que le mot infini qui apparaît ici (même si on n'en découvre que les premières lettres) ressemble étrangement au mot de Thue–Morse (2) introduit à la section précédente. Ceci n'est bien évidemment pas une surprise ! Alan Cobham a démontré en 1972 [8], sans entrer dans les détails techniques, que les deux procédés de construction sont équivalents.

5 Un peu de combinatoire des mots

La combinatoire des mots est un thème de recherche actuelle en mathématiques. On lui consacre annuellement des conférences internationales. Pour résumer ce dont il s'agit, voici la description que j'en donne aux étudiants de Master en Sciences Mathématiques : D'une manière générale, la combinatoire est la branche des mathématiques qui étudie les "configurations" d'un ensemble "discret" (et généralement fini). On s'intéresse alors au dénombrement ou à l'énumération effective des objets de cet ensemble, à la structure (éventuellement algébrique) de celui-ci ou encore à ses propriétés extrémales (éléments maximaux, ...). Comme son nom l'indique, la combinatoire des mots s'intéresse plus spécialement aux mots (finis ou infinis), i.e., aux suites de symboles ou de lettres appartenant à un alphabet fini.

La combinatoire des mots possède par exemple des applications en bioinformatique ; une séquence génétique n'étant rien d'autre qu'un long mot sur un alphabet de 4 symboles.

Voici le résultat le plus simple que l'on puisse imaginer. Un *carré* est la répétition immédiate de deux facteurs identiques. Ainsi, les mots `bonbon` et `baba` sont des carrés et le mot `tartare` débute par un carré.

Théorème 21. *Avec un alphabet de deux lettres, tout mot de longueur au moins* 4 *contient un carré.*

Démonstration. Sur l'alphabet $\{a, b\}$, essayons de construire un mot le plus long possible sans carré. Supposons que ce mot débute par a (ce n'est pas une restriction, on peut mener le même raisonnement avec b). Pour ne pas avoir le carré aa, on considère alors le mot ab. Ce mot ne peut pas être complété par un b sinon, on y retrouverait le carré bb. On a donc le mot aba. On s'aperçoit que ce mot ne peut plus être complété sans y faire apparaître soit le carré aa soit le carré $abab$. □

Face à un tel résultat, le mathématicien peut se poser deux types de question (qui sont largement étudiées dans des recherches récentes) :

- Sur un alphabet de 3 lettres, peut-on éviter les carrés ? Par exemple, un mot comme $abacabcacb$ ne contient aucun carré.
- Sur 2 lettres, quelles sont les configurations ou motifs qui sont (in)évitables ?

D'une manière générale, quels liens existe-t-il entre la taille de l'alphabet et les configurations que l'on peut éviter ou non.

Un *chevauchement* est la répétition immédiate de deux facteurs identiques suivie de la première lettre de cette répétition. Autrement dit, un chevauchement est un mot de la forme $cvcvc$ ou c est une lettre et v est un mot. Par exemple, `ananas` débute par un chevauchement. De façon informelle, un chevauchement est "un carré plus la première lettre du carré". Nous admettrons le résultat suivant (dont la preuve est longue, mais pas difficile), voir par exemple [14, 21].

Théorème 22. *Le mot de Thue–Morse est sans chevauchement. Donc, en particulier, il est sans cube et n'est pas (ultimement) périodique.*

De ce résultat, on déduit facilement qu'avec 3 lettres, on peut construire un mot infini sans carré. En effet, puisque le mot de Thue–Morse est sans chevauchement, il ne contient pas de cube (donc jamais de facteur bbb) et il se factorise de manière unique à l'aide de $\{a, ab, abb\}$:

$$abb|ab|a|abb|a|ab|abb|ab|a|ab|abb|a|abb|ab|a|\ a\cdots$$

Il suffit alors de considérer le codage de ces trois types de facteurs

$$g : \begin{cases} 1 \mapsto abb \\ 2 \mapsto ab \\ 3 \mapsto a \end{cases}$$

pour obtenir le mot infini $123132123213123\cdots$. Ce mot ne peut contenir de carré car s'il en contenait un, le mot de Thue–Morse contiendrait un chevauchement contredisant le théorème précédent. En effet, supposons que ce mot contienne le facteur 112, mais les images du codage g débutant toutes par a, le mot de Thue–Morse contiendrait le facteur $g(1)g(1)a = abbabba$; ce qui est absurde.

6 Une dernière définition

Une troisième définition de la suite de Thue–Morse est décrite à l'aide d'une opération de "conjugaison" inversant les deux lettres a et b. On construit une suite de mots finis par récurrence comme suit :

$$X_0 = a, \quad X_{n+1} = X_n\overline{X_n}$$

où $\overline{abbaa} = baabb$. Les premiers termes de la suite sont

$$\begin{aligned} X_0 &= a \\ X_1 &= ab \\ X_2 &= abba \\ X_3 &= abbabaab \\ X_4 &= abbabaabbaababba \\ X_5 &= abbabaabbaababbabaababbaabbabaab \end{aligned}$$

Il est facile de se convaincre (par récurrence) que cette suite converge une fois encore vers le mot de Thue–Morse (2). Pour la construction de X_{n+1}, si X_n correspond aux entiers $\{0, 1, \ldots, 2^n - 1\}$, alors $\overline{X_n}$ correspond aux entiers $\{2^n, \ldots, 2^{n+1} - 1\}$, mais si l'on se souvient de la remarque 17, puisqu'on ajoute un 1 de tête aux écritures binaires des éléments de $\{0, 1, \ldots, 2^n - 1\}$ pour obtenir celles de $\{2^n, \ldots, 2^{n+1} - 1\}$, la parité du nombre de 1 change. Ceci explique la conjugaison.

7 Miscellanées

7.1 Structure de fractal

Le mot de Thue–Morse possède ce que l'on pourrait appeler une structure de fractal ; on retrouve la suite à différentes échelles. En effet, des relations obtenues à la remarque 20, on s'aperçoit que la suite constituée uniquement des termes d'indices pairs (resp. impairs) est exactement la suite de Thue–Morse originelle (resp. le conjugué de la suite de Thue–Morse). Ainsi, d'une manière générale, en construisant des sous-suites dont les termes sont espacés d'une même puissance de 2, on ne retrouve jamais que la suite de Thue–Morse ou son conjugué,

$$\#\left\{(x_{2^t n+r})_{n\geq 0} \mid t \geq 0, r < 2^t\right\} = 2.$$

7.2 Musique

Le compositeur danois Per Nørgård, `http://pernoergaard.dk/`, s'est inspiré de la suite de Thue–Morse dans certaines de ses compositions. Sans pouvoir garantir la pérennité du lien, on en trouvera une adaptation assez libre sur `www.youtube.com/watch?v=FA7m7anh2xg`. Le compositeur Tom Johnson a aussi employé la suite dans certaines de ses compositions.

7.3 Jeu d'échecs

Le grand maître d'échecs Machgielis Euwe (1901–1981) redécouvrit la suite de Thue–Morse pour mettre en défaut un règlement du jeu d'échecs devant éviter l'existence de parties infinies. Une partie est déclarée nulle, si la même séquence de coups apparaît trois fois d'affilée ("German Rule"). Ainsi Max Euwe utilisa en 1929 la suite de Thue–Morse (que l'on sait être sans cube) pour montrer qu'avec une telle règle, une partie infinie peut encore exister. Il suffit de remplacer les symboles a et b de la suite par les coups suivants

$$a \mapsto Ng1-f3, Ng8-f6, Nf3-g1, Nf6-g8,$$

$$b \mapsto Nb1-c3, Nb8-c6, Nc3-b1, Nc6-b8.$$

8 Encore un peu de théorie des nombres

Passons en revue quelques autres situations dans lesquelles apparaît une fois encore la suite de Thue–Morse. Peut-on partitionner $\mathbb{N}$ en deux ensembles disjoints A et B de telle sorte que tout entier possède le même nombre de décompositions comme somme de deux éléments distincts de A et comme somme de deux éléments distincts de B ? J. Lambek et L. Moser montrent que la suite de Thue–Morse est l'unique solution à ce problème [13]. Ainsi, avec

$$A = \{0, 3, 5, 6, 9, 10, 12, 15, \ldots\} \text{ et } B = \{1, 2, 4, 7, 8, 11, 13, 14, \ldots\}$$

on a, par exemple, $12 = 0 + 12 = 3 + 9$ et $12 = 1 + 11 = 4 + 8$.

On trouvera encore des traces de la suite dans divers problèmes comme celui proposé par D. R. Woods [23] à propos de la convergence d'une étrange suite de

rationnels ou encore dans [10, p. 227, ex. 35] où la suite de Thue–Morse devient un simple exercice pour développer des stratégies de résolution de problèmes !

F. M. Dekking [9] (il est aussi intéressant de voir [2]) montre, comme l'avait fait avant lui Mahler, que la constante de Thue–Morse dont le développement binaire est régi par la suite,

$$\frac{0}{2}+\frac{1}{4}+\frac{1}{8}+\frac{0}{16}+\frac{1}{32}+\frac{0}{64}+\frac{0}{128}+\frac{1}{256}+\cdots=\sum_{n=0}^{\infty}\frac{\sigma_2(n)\bmod 2}{2^{n+1}}\simeq 0,4124540336\cdots$$

est un nombre transcendant.

Soient $a, b > 0$ des entiers. Dans le même ordre d'idées, le réel dont le développement en fractions continues est dirigé par la suite de Thue–Morse,

$$x = a + \cfrac{1}{b + \cfrac{1}{b + \cfrac{1}{a + \ddots}}}$$

est également transcendant [20, 1].

Enfin, pour répondre à un problème de Gelfond, C. Mauduit et J. Rivat montrent le résultat suivant s'intéressant aux écritures binaires des nombres premiers [15]. Il signifie qu'en moyenne il y a autant de nombres premiers dont la somme des chiffres est paire que de nombres premiers dont la somme des chiffres est impaire.

$$\lim_{N\to+\infty}\frac{\#\{n\in\mathcal{P}\mid n\leq N \text{ and } \sigma_2(n)\equiv 0 \bmod 2\}}{\#\{n\in\mathcal{P}\mid n\leq N\}}=\frac{1}{2}.$$

Remerciements

Merci à Jean-Paul Allouche et Jeffrey Shallit pour leurs réactions positives et les compléments fournis (en particulier, la remarque 16 concernant la non-unicité des solutions) après une lecture rapide d'une première version de ce texte.

Références

[1] B. Adamczewski, Y. Bugeaud, A short proof of the transcendence of Thue-Morse continued fractions, *Amer. Math. Monthly* **114** (2007), 536-–540.

[2] J.-P. Allouche, M. Cosnard, The Komornik-Loreti constant is transcendental, *Amer. Math. Monthly* **107** (2000), 448-–449.

[3] J.-P. Allouche, M. Mendès France, Euler, Pisot, Prouhet-Thue-Morse, Wallis and the duplication of sines, *Monatsh. Math.* **155** (2008), 301-–315.

[4] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, *Sequences and their applications* (Singapore, 1998), 1–16, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999.

[5] J.-P. Allouche, J. Shallit, *Automatic Sequences : Theory, Applications, Generalizations*, Cambridge University Press, (2003).

[6] P. Borwein, C. Ingalls, The Prouhet-Tarry-Escott problem revisited , *Enseign. Math.* **40** (1994), 3–27. `http://retro.seals.ch/digbib/view?rid=ensmat-001:1994:40::10`

[7] D. W. Boyd, On a problem of Byrnes concerning polynomials with restricted coefficients, *Math. Comp.* **66** (1997), 1697-–1703.

[8] A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.

[9] F. M. Dekking, Transcendance du nombre de Thue-Morse, *C. R. Acad. Sci. Paris* **285** (1977), 157–160.

[10] A. Engel, *Problem-Solving Strategies*, Springer (1997).

[11] R. Honsberger, *Mathematical Diamonds*, The Math. Association of America (2003).

[12] D. Kalman, A Matrix Proof of Newton's Identities, *Math. Mag.* **73** (2000), 313–315.

[13] J. Lambek, L. Moser, On some two way classifications of integers, *Canad. Math. Bull.* **2** (1959), 85–89.

[14] M. Lothaire, *Combinatorics on Words*, Encyclopedia of Mathematics and Its Applications **17**, Addison-Wesley (1983).

[15] C. Mauduit, J. Rivat, Sur un problème de Gelfond : la somme des chiffres des nombres premiers, *Ann. of Math.* **171** (2010), 1591–1646.

[16] D. G. Mead, Newton's identities, *Amer. Math. Monthly* **99** (1992), 749–751.

[17] M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22** (1921) 84–100.

[18] E. Prouhet, Mémoire sur quelques relations entre les puissances de nombres, *C. R. Acad. Sci. Paris Sér. I* **33** (1851), 225.

[19] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lect. Notes in Math. **1794**, V. Berthé *et al.* Eds, Springer (2002).

[20] M. Queffélec, Transcendance des fractions continues de Thue-Morse, *J. Number Theory* **73** (1998), 201-–211.

[21] M. Rigo, *Théorie des automates et langages formels*, notes de cours, Université de Liège (2009–2010). `http://www.discmath.ulg.ac.be/notes.html`

[22] M. Rigo, *Automates et numération*, Bull. Soc. Roy. Sci. Liège **74** (2005) 249–262. `http://orbi.ulg.ac.be/handle/2268/15920`

[23] D. Robbins, Solution to Problem E2692, *Am. Math. Monthly* **86** (1979), 394–395.

[24] A. Thue, Über unendliche Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906), 1–22. Reprinted in T. Nagell, ed., *Selected Mathematical Papers of Axel Thue*, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[25] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in T. Nagell, ed., *Selected Mathematical Papers of Axel Thue*, Universitetsforlaget, Oslo, 1977, pp. 413–478.