

Matrice $N \times N$ constituée de $-1, 0, 1$

Denis Cazor

Introduction :

Pendant le stage de Luminy 2017 (du 1 au 5 mai) nous avons discuté d'un problème que j'ai trouvé amusant et qui m'a occupé jusqu'à présent (8 décembre).

Problème :

On considère des matrices $N \times N$ remplies de $-1, 0$ ou 1 . On désire obtenir des matrices dont toutes les sommes des lignes et colonnes sont distinctes. Ces sommes constituent un ensemble de $2N$ valeurs distinctes prises dans l'intervalle $[-N, N]$.

Parmi ces sommes, il ne manque donc qu'une valeur. Soit X la valeur absente.

Conventions de calcul :

- 1) Les éléments de la matrice sont numérotés suivant les lignes des 0 à N^2-1 , en partant du haut à gauche.
- 2) Les lignes et les colonnes sont triées dans l'ordre des sommes décroissantes.

Constatations :

- 1) Je constate qu'il n'y a pas de solutions lorsque N est impair.
- 2) Lorsque N est pair, je constate que les sommes N et $-N$ ne peuvent coexister. X est donc égal à $-N$, puisque j'ai choisis de considérer une ligne complète de 1 .
- 3) $N/2$ colonnes sont strictement positives, $N/2$ colonnes sont négatives ou nulles.
- 4) $N/2$ lignes sont strictement positives, $N/2$ lignes sont négatives ou nulles.
- 5) Deux sous matrices $p \times p$ constituées de 1 et -1 .
- 6) Sommes 1 et 0 placés en p et $p+1$, sur les lignes ou bien les colonnes. Il y a beaucoup plus de solutions sur les colonnes médianes que sur les lignes médianes.
- 7) Dans de nombreux cas, les nombres de solutions sont identiques, pour des configurations différentes. Cela est lié aux possibilités de symétrie des solutions.

Algorithme :

- 1) La stratégie consiste à balayer toutes les valeurs possibles par backtracking. Pour $N = 6$, il faut explorer $1,5 \cdot 10^{17}$ cas différents. Il faut donc couper les branches de l'arbre de recherche le plus rapidement possible.
- 2) Noter que la valeur N ou bien $-N$ doit être présente. Il est possible de changer tous les signes de la matrice pour obtenir une nouvelle solution. On peut donc supposer que toute solution contient une somme égale à N , et donc une ligne de 1 , sans limitation de généralité.

3) La première ligne contient des 1 (somme N). Comme les solutions sont identiques par permutations, on peut trier la deuxième ligne et la première colonne. Pour $N = 6$, réduction de la recherche à $215800094376 \approx 2.15 \cdot 10^{11}$ cas.

4) On obtient de nouvelles solutions en effectuant des permutations des lignes et des colonnes, en multipliant par -1 , ou bien en transposant la matrice (pour $N = 30$, il existe $4(30!)^2 \sim 2,8 \cdot 10^{65}$ solutions différentes, construites à partir d'une solution ordonnée). Il faut donc identifier une solution et déterminer si elle n'est pas identique à l'une de celles déjà trouvées, par permutation.

5) Comme toutes les sommes sont différentes, je définis une "forme ordonnée" en triant les lignes et les colonnes en ordre décroissant. La première ligne est de somme N.

6) Pour identifier des solutions identiques rapidement, on peut calculer une clé de tri des solutions.

7) Après le tri (décroissant) des lignes et colonnes, on constate que le triangle supérieur gauche ne contient que des 1 et le triangle inférieur droit ne contient que des -1 . On peut donc accélérer la résolution en ne remplissant que des valeurs proches de la diagonale (paramètre L).

8) Pour éviter de nombreuses opérations, on effectue un pré-calcul des cases à remplir et des tests de somme à effectuer.

9) En ne générant que des matrices avec lignes et colonnes déjà triées, on coupe plus rapidement les branches de l'arbre de recherche.

10) La dernière colonne contient un 1, sa somme est donc au minimum égale à $2 - N$. La valeur $1 - N$ est donc située sur la dernière ligne (puisque $-N$ n'est pas une somme possible, voir la démonstration ci-après). On simplifie les dernières itérations en remarquant que la dernière ligne ne contient qu'un 0 et des valeurs -1 .

11) La matrice $[X]$ de somme x , de taille $p \times p$ ne contient que des 1. La matrice $[Y]$ de somme y , de taille $p \times p$ ne contient que des -1 (on suppose $N = 2p$, voir la démonstration $x = p^2$). Dans la suite on appelle U la matrice $[X]$, et $-U$ la matrice $[Y]$.

12) Le placement des sommes 1 et 0 s'effectue sur les colonnes ou lignes p et $p+1$.

Supposons que 0 soit placé sur la ligne d'indice $p+1$, alors cette ligne contient p fois 1 et p fois -1 . La somme de la colonne d'indice p est donc au minimum égale à $(p + 1) - (p - 1) = 2$. La valeur 1 n'est donc pas située sur une colonne, et se retrouve donc sur la ligne d'indice p .

Le raisonnement s'applique aussi aux colonnes. Les sommes 1 et 0 sont donc placés sur les deux colonnes médianes ou bien sur les deux lignes médianes.

13) Le remplissage de la dernière ligne (de somme $1-N$), et celui de la colonne ou de la ligne de somme 1 permet un découpage du problème tout en conservant la stratégie de backtracking. La recherche des solutions se découpe en $2(p^2 - p + 1)$ sous problèmes, ce qui permet de le traiter en multiprocesseur et hyper-threading.

14) Le placement de la somme -1 s'effectue sur la colonne ou la ligne p ou $p+1$. On améliore les possibilités de découpage du problème, et on obtient un plus grand nombre de sous problèmes.

15) Le placement de la somme $N-1$ s'effectue sur la deuxième ligne (cette ligne ne peut contenir qu'un seul 0) ou bien sur la première colonne (cette colonne ne peut pas contenir de -1).

Dans le premier cas, la deuxième ligne ne contient que des 1, et un 0.

Dans le deuxième cas, la première colonne ne contient que des 1 et un 0, la dernière ligne ne contient que des -1 et un 0. Le seul 0 est donc situé sur la première colonne et la dernière ligne.

On obtient encore un plus grand nombre de sous problèmes.

16) Utilisation des symétries, pour diviser le temps d'exécution par deux.

Résultats :

- 1) Pour $N = 4$, il existe $4 = 2^2$ solutions distinctes ($\times 2304$ par permutations).
- 2) Pour $N = 6$, il existe $39 = 3 \cdot 13$ solutions distinctes ($\times 2073600$ par permutations).
- 3) Pour $N = 8$, il existe $2260 = 2^2 \cdot 5 \cdot 113$ solutions distinctes ($\times 6502809600$ par permutations).
- 4) Pour $N = 10$, il existe $1338614 = 2 \cdot 17 \cdot 39371$ solutions distinctes ($\times 52672757760000$ en utilisant toutes les permutations).
- 5) Pour $N = 12$, il existe $8522456190 = 2 \cdot 5 \cdot 852265619$ solutions distinctes ($\times 917770131210240000$ en utilisant toutes les permutations).
- 6) Le problème reste de complexité exponentielle, et le temps de calcul devient rapidement trop grand.
Le cas $N = 10$ peut être traité en temps raisonnable (environ 220 secondes normalisé monoprocesseur).
Le cas $N = 12$ ne peut être traité en temps raisonnable (environ 10400 heures, 1.2 années normalisées monoprocesseur). Le traitement est 3^{11} fois plus long que le cas $N = 10$. Il faut envisager une solution multiprocesseur (environ 12 jours avec 43 instances du programme). En pratique, il a fallu environ 80 jours avec la vérification des résultats.
Le cas $N = 14$ semble définitivement hors de portée, avec ce programme.

Démonstration de François Coulombe

On suppose la matrice triée suivant les sommes de lignes et de colonnes décroissantes.

Comme $2N+1$ sommes sont possibles, et qu'il n'y en a que $2N$ distinctes, l'une des sommes est absente que nous appelons X . Nous supposons que $X \leq 0$, sans limiter la portée du raisonnement (puisque'on peut changer tous les signes d'une solution).

On considère les sommes strictement positives des lignes et des colonnes, puis les sommes négatives ou nulles.

	p	q
r	x	u
s	v	y

Soit p le nombre de colonnes de somme strictement positives. Soit q le nombre de colonnes de somme négative ou nulle. Soit r le nombre de lignes de somme strictement positives. Soit s le nombre de lignes de somme négative ou nulle.

Soit x la somme de tous les éléments de la sous matrice (r, p) . Soit y la somme de tous les éléments de la sous matrice (s, q) . Soit u la somme de tous les éléments de la sous matrice (r, q) . Soit v la somme de tous les éléments de la sous matrice (s, p) .

Soit A la somme des lignes et colonnes de somme strictement positives, et B la somme des lignes et colonnes de somme négative ou nulle :

$$A = 2x + u + v = \sum_{k=1}^N k = \frac{N(N+1)}{2}, \quad B = 2y + u + v = \sum_{k=-N}^{-1} k - X = -\frac{N(N+1)}{2} - X.$$

Soit finalement : $A - B = 2(x - y) = N(N+1) + X.$

D'où l'on déduit que X est pair nécessairement.

En considérant le nombre de colonnes, le nombre de lignes, les sommes positives, les sommes négatives ou nulles (sans X), on obtient de plus :

$$p + q = N, \quad r + s = N, \quad p + r = N, \quad q + s = N.$$

Soit finalement : $s = p$ et $r = q.$

En considérant les termes des matrices (r, p) et (s, q) qui appartiennent tous à l'intervalle $[-1, 1]$, on obtient :

$$-pq \leq x \leq pq, \quad -pq \leq y \leq pq, \quad \text{et} \quad 2(x - y) = N(N+1) + X \leq 4pq.$$

Comme $N = p + q$, et que $|X| \leq N$, il vient :

$$0 \leq N + X \leq 4pq - N^2 = 4pq - (p + q)^2 = -(p - q)^2 \leq 0.$$

Soit finalement $p = q$, $X = -N$ et $N = 2p$ qui est donc pair nécessairement.

On constate de plus que :

$$-p^2 \leq x \leq p^2, \quad -p^2 \leq y \leq p^2, \quad \text{et} \quad x - y = N^2/2 = 2p^2.$$

On en déduit que $x = p^2$ et $y = -p^2$, et donc que la sous matrice $[X]$ ne contient que des 1, et la sous matrice $[Y]$ ne contient que des -1 . La somme $u+v$ est égale à p .

Construction d'une nouvelle solution par symétrie

Lorsqu'une solution M_N est obtenue à l'ordre N , sous forme ordonnée, on obtient une nouvelle solution, sous forme ordonnée, en complétant la première matrice. On appelle U la matrice $p \times p$ constituée de 1, A et B les deux sous matrice $p \times p$ à compléter.

1^{er} cas : sommes 0 et 1 sur les lignes

U	A
B	-U

$$L_i = l_i + p$$

$$L_j = l_j - p$$

$$C_i = c_i + p \quad C_j = c_j - p$$

U	B
A	-U

$$L'_i = l_j + p$$

$$L'_j = l_i - p$$

$$C'_i = c_j + p \quad C'_j = c_i - p$$

$$\begin{aligned} L'_i &= L_j + N, & L_j &\in [1 - N, 0], & l_j &\in [1 - p, p] & \Rightarrow & L'_i &\in [1, N] \\ C'_i &= C_j + N, & C_j &\in [1 - N, -1], & c_j &\in [1 - p, p - 1] & \Rightarrow & C'_i &\in [1, N - 1] \\ L'_j &= L_i - N, & L_i &\in [1, N], & l_i &\in [1 - p, p] & \Rightarrow & L'_j &\in [1 - N, 0] \\ C'_j &= C_i - N, & C_i &\in [1, N - 1], & c_i &\in [1 - p, p - 1] & \Rightarrow & C'_j &\in [1 - N, -1] \end{aligned}$$

Lignes et colonnes sont encore triées et distinctes. On associe deux solutions distinctes si les matrices A et B ne sont pas égales. Le programme vérifie cette propriété, par placement des sommes particulières.

2^{ème} cas : sommes 0 et 1 sur les colonnes

U	A
B	-U

$$L_i = l_i + p$$

$$L_j = l_j - p$$

$$C_i = c_i + p \quad C_j = c_j - p$$

U	^tA
^tB	-U

$$L'_i = c_j + p$$

$$L'_j = c_i - p$$

$$C'_i = l_j + p \quad C'_j = l_i - p$$

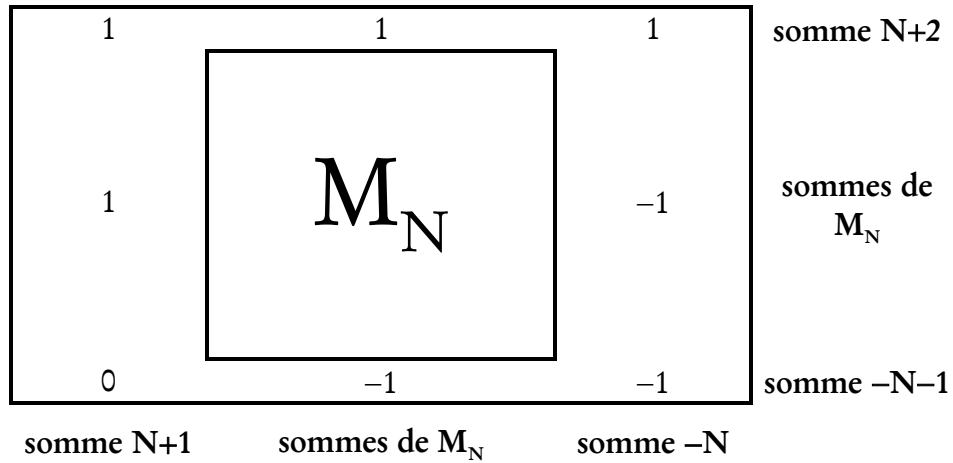
$$\begin{aligned} L'_i &= C_j + N, & C_j &\in [1 - N, 0], & c_j &\in [1 - p, p] & \Rightarrow & L'_i &\in [1, N] \\ C'_i &= L_j + N, & L_j &\in [1 - N, -1], & l_j &\in [1 - p, p - 1] & \Rightarrow & C'_i &\in [1, N - 1] \\ L'_j &= C_i - N, & C_i &\in [1, N - 1], & c_i &\in [1 - p, p - 1] & \Rightarrow & L'_j &\in [1 - N, -1] \\ C'_j &= L_i - N, & L_i &\in [1, N], & l_i &\in [1 - p, p] & \Rightarrow & C'_j &\in [1 - N, 0] \end{aligned}$$

Lignes et colonnes sont encore triées et distinctes. On associe deux solutions différentes si les matrices A et B ne sont pas toutes les deux symétriques. Le programme vérifie cette propriété, par placement des sommes particulières.

Ces propriétés permettent de ne calculer que la moitié des configurations environ (279 sur 458 pour $N = 8$, 1348 sur 2330 pour $N = 12$, 4189 sur 7474 pour $N = 16$). Les temps de calcul des symétriques peuvent être très différents suivant la configuration considérée.

Construction d'une nouvelle solution (passage de N à $N+2$)

Lorsqu'une solution M_N est obtenue à l'ordre N , sous forme ordonnée, on obtient une solution à l'ordre $N+2$, sous forme ordonnée, en complétant la première matrice suivant :



Lignes et colonnes sont encore triées. Les quatre valeurs de sommes absentes à l'ordre N sont présentes à l'ordre $N+2$, conformément à la forme ordonnée de la matrice.

Profondeur P de l'arbre de recherche dans le cas $N = 8, L = 4$

Numéro de tâche	Profondeur
1 → 13	12
14	13
15 → 30	11
31 → 174	13
175 → 190	11
191 → 202	12
203 → 458	8

$$3^5 = 243 \quad \& \quad 3^{13} \ll 3^{64} \quad \text{soit} \quad 1.59 \cdot 10^6 \ll 3.43 \cdot 10^{30}$$

Profondeur P de l'arbre de recherche dans le cas $N = 10, L = 5$

Numéro de tâche	Profondeur
1 → 21	24
22	25
23 → 47	23
48 → 447	25
448 → 472	23
473 → 492	24
493 → 1117	20

$$3^5 = 243 \quad \& \quad 3^{25} \ll 3^{100} \quad \text{soit} \quad 8.47 \cdot 10^{11} \ll 5.15 \cdot 10^{47}$$

Profondeur P de l'arbre de recherche dans le cas N = 12, L = 6

Numéro de tâche	Profondeur
1 → 31	40
32	41
33 → 68	39
69 → 968	41
969 → 1004	39
1005 → 1034	40
1035 → 2330	36

$$3^5 = 243 \quad \& \quad 3^{41} \ll 3^{144} \quad \text{soit} \quad 3.65 \cdot 10^{19} \ll 5.08 \cdot 10^{68}$$

Profondeur P de l'arbre de recherche dans le cas N = 14, L = 7

Numéro de tâche	Profondeur
1 → 43	60
44	61
45 → 93	59
94 → 1857	61
1858 → 1906	59
1907 → 1948	60
1949 → 4349	56

$$3^{20} \approx 3.5 \cdot 10^9$$

Profondeur P de l'arbre de recherche dans le cas N = 16, L = 8

Numéro de tâche	Profondeur
1 → 57	84
58	85
59 → 122	83
123 → 3258	85
3259 → 3322	83
3323 → 3378	84
3379 → 7474	80

$$3^{44} \approx 9.8 \cdot 10^{20}$$

La profondeur de l'arbre de recherche augmente de 12, 16, 20, puis 24.

Le temps de calcul passe de 72 ms (N=8, P=13), à 220 s (N=10, P=25) puis 10400 h (N=12, P=41).

La profondeur de l'arbre de recherche est augmentée de 12 puis de 16.

Le temps de calcul est multiplié par 3^7 puis par 3^{11} .

On pourrait en déduire que le temps de calcul est augmenté d'un facteur 3^{15} pour passer à N=14, puis d'un facteur 3^{19} pour passer à N=16, soit respectivement 17 millions d'années et 1.5 millions de fois l'âge de l'univers.

La profondeur maximale de l'arbre P_N (N=2p) est égale à $2p^2 - 6p + 5$.

Le temps de calcul T_N est alors de l'ordre de $3^{2p^2 - 11p + 12} T_8$.

Calculs effectués sur mes 8 machines (43 tâches indépendantes)

Étalonnage pour le problème 12-6, avec NT = 53 (la tâche la plus rapide)

Initiales	Machines	Processeur	Fréquence MHz	Année	Étalonnage heures	Efficacité mono-tâche	Nombre tâches		Efficacité globale
PA	Paris Anne	Core i5 3570	3400	2013	10.47	0.99	4	3.96	1.27
PE	Paris Ecp	Core i7 3820	3600	2013	13.05	0.80	8	6.40	2.04
PG	Paris Gauche	Core i7 950	3067	2011	13.28	0.78	7	5.46	1.76
PS	Paris Sam	Core i5 3570	3400	2014	10.40	1.00	4	4.00	1.28
SA	Sévérac Aline	Core 2 Quad Q9550	2833	2007	12.21	0.85	4	3.40	1.09
SE	Sévérac Elisa	Core 2 Quad Q9300	2500	2007	13.32	0.78	4	3.12	1.00
SJ	Sévérac Juliette	Core 2 Quad Q9550	2833	2008	11.71	0.89	4	3.56	1.14
SL	Sévérac Laurent	Core i7 930	2940	2010	16.01	0.65	8	5.20	1.66
Totaux							43	35.1	11.24

43 processeurs équivalents à 35 fois le processeur le plus efficace.
Les durées normalisées sont recalculées sur la base du processeur le plus rapide.

Calculs effectués avec Caml et Ocaml (traductions de ZUM.PAS)

Problème 8-4 / Calcul complet (2260 solutions) #
Caml light 3.75 s # Ocaml 3.25 s # Turbo Pascal 1.00 s
soient les rapports 15 13 4.

Problème 10-5 / Calcul partiel P n°1/8 (206222 solutions) #
Caml light 5021 s # Ocaml 4426 s # Turbo Pascal 316 s
soient les rapports 16 14 1.

Gain Ocaml/Caml light 12%.

Calculs effectués avec Turbo Pascal (NMU.PAS et NMU_SYM.PAS / problème 12-6)

Vérification sans affichage des matrices (économie de temps normalisé 10.5 %)
Calcul initial effectué en 1099 jours soit 26373 heures.
Calcul initial effectuée en 944 jours soit 22654 heures normalisées.

Vérification effectuée (sans affichage des solutions) en 1042 jours soit 25011 heures.
Vérification effectuée (sans affichage des solutions) en 844 jours soit 20256 heures normalisées.

Avec l'utilisation des symétries (économie de temps normalisé 48.7 %)
Vérification (non réalisée) effectuée en 433 jours soit 10388 heures normalisées.
Fluctuation des temps de calcul normalisés pour une tâche donnée -46% +84%.
228 solutions trouvées par seconde, par processeur,
8000 solutions trouvées par seconde, pour l'ensemble des processeurs.

Évolution du programme de recherche de toutes les solutions

Traitement du problème 6-3

Programme	Durée d'exécution (en secondes)	Nombre de solutions détectées	Améliorations	Gain de vitesse
MATRICE.PAS V1	216784.0	39	deuxième ligne & première colonne triées.	1
MATRICE.PAS V3	1360.0	2	génération pseudo-aléatoire.	159
MATRICE.PAS V2	537.4	39	contrôle lignes et colonnes.	403
MATRICE.PAS V4	41.922	39	emplacements lignes colonnes pré-calculés.	5171
MATRICE.PAS V5	39.009	39	génération sur la diagonale.	5557
MATRICE.PAS V6	23.466	39	remplissage diagonale seule.	9238
D.PAS	14.836	39	remplissage décroissant diagonal en ligne, options de compilation {F-} {A+} (gain 33%).	14612
DD.PAS	6.125	39	remplissage diagonal descendant.	35393
T.PAS	0.968	19	lignes & colonnes triées.	223950
LL.PAS	0.678	39	placement de la somme 1-N, décomposition en sous problèmes indépendants.	319740
ZU.PAS	0.01682	39	placement des sommes 0 & 1.	12888466
ZT.PAS	0.01656	39	remplissage des sous matrices U & -U.	13090821
ZUM.PAS	0.015295	39	placement de la somme -1.	14173521
NMU.PAS	0.003844	39	placement de la somme N-1.	56395421
NMU_SYM.PAS	0.011608	39	utilisation des symétries.	18675396
NMU_SYM.PAS	0.001968	39	utilisation des symétries.	110154472

Estimation asymptotique basée sur la résolution du problème 12-6.

Option {F-} appel court (NEAR CALL) / Option {A+} alignement des octets et des mots.
Le gain de NMU_SYM.PAS, par rapport à MATRICE.PAS V1, est de l'ordre de $1.1 \cdot 10^8$.

Solution N°6 855400

```
1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 0 -1
1 1 1 1 1 1 1 1 1 1 -1 -1
1 1 1 1 1 1 1 1 0 -1 -1 -1
1 1 1 1 1 1 1 1 -1 -1 -1 -1
1 1 1 1 1 1 0 -1 -1 -1 -1 -1
1 1 1 1 1 1 -1 -1 -1 -1 -1 -1
1 1 1 1 0 -1 -1 -1 -1 -1 -1 -1
1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1
1 1 0 -1 -1 -1 -1 -1 -1 -1 -1 -1
1 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
0 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
```

```
12 9 8 5 4 1 0 -3 -4 -7 -8-11
11 10 7 6 3 2 -1 -2 -5 -6 -9-10
```

Nombre de solutions : 6

Mémoire disponible : 18094

Durée d'exécution 2 s 438 ms

Durée d'exécution 406 ms

N = 12, L = 1, Remplissage avec sommes égales à -1, 0, 1, 1-N & N-1 [N]

P n°1/1 10:15:34 12/12/17 10:15:36 12/12/17

