REVUE DE STATISTIQUE APPLIQUÉE

G. MOZZO

Plan quadratique gigogne

Revue de statistique appliquée, tome 38, nº 3 (1990), p. 23-34

<http://www.numdam.org/item?id=RSA_1990__38_3_23_0>

Rev. Statistique Appliquée, 1990, **XXXVIII** (3), 23-34

PLAN QUADRATIQUE GIGOGNE

G. MOZZO

ATOCHEM

Parmi les nombreux types de plans d'expériences mis en œuvre en vue d'optimiser les procédés industriels, ceux qui permettent à la fois pour les p variables explicatives de calculer leurs interactions d'ordre 2 et leurs effets quadratiques, sont considérés comme étant les plus intéressants.

Ils permettent d'étudier simultanément l'équation de réponse pour les q propriétés caractéristiques du procédé.

Pour chacune des réponses, le modèle mathématique s'écrit :

$$yk = bo + \Sigma biXi + \Sigma biiXi^2 + \Sigma bijXiXj$$

Il nécessite un nombre minimal d'essais égal au nombre c de coefficients du modèle :

$$c = 1 + 1,5p + 0,5p^2$$

Habituellement on fait appel aux 3 types de plan suivants, qui comportent tous 5 niveaux d'essai des variables explicatives :

- plan composite centré de BOX et WILSON (1) pour lequel la valeur de α (distance des points situés sur l'axe passant par les centres des faces opposées) doit dépendre du nombre no d'essais effectués au centre, si l'on veut annuler les corrélations entre les termes quadratiques.
- plan composite fractionnaire, bâti sur le plan $2^{(p-1)}$, que l'on peut utiliser pour $p \geq 5$ si l'on désire se limiter au modèle indiqué plus haut.
- plan hybrides de Roquemore (2), pour lesquels la corrélation entre les termes quadratiques n'est pas nulle.

Le tableau suivant montre l'intérêt de ces plans qui permettent une réduction importante du nombre d'essai, par rapport aux plans complets du type 3^p.

Nombre d'essais minimum par type de plan

Nombre de variables p	Nombre de coef. c	Plan fraction. 3^p	Plan composite centré minimal (no = 1)	Plan composite fraction. minimal (no = 1)	Plan hybride
3	10	27	15(α = 1.215)		11
4	15	81	25(α = 1.414)		16
5	21	243	43(α = 1.596)	27(α = 1.547)	non publié
6	28	729	77(α =1.761)	45(α = 1.724)	28
7	36	2187	143(α = 1.909)	79(α =1.885)	46

La réduction du nombre d'essais par rapport aux plan 3^p est certes du plus grand intérêt, mais outre le fait qu'elle implique une expérimentation à 5 niveaux, elle comporte néanmoins une conséquence fâcheuse pour l'optimisation des procédés industriels.

En effet, si l'on décide d'opter pour un plan composite ou pour un plan hybride, il est impératif de connaître, ou bien de décider a priori, le nombre de variables p ayant un effet non négligeable sur l'une quelconque de l'ensemble des différentes réponses q du procédé.

Ce n'est en général pas toujours le cas, puisque le but de l'expérimentation est précisemment de quantifier l'effet des variables explicatives, de sorte que dans la pratique, on est confronté au dilemme suivant :

- soit prendre en compte un nombre de variables supérieur à celui qui est strictement suffisant pour atteindre l'objectif, ce qui entraîne un surcroît d'essais.
- soit ne retenir qu'un nombre limité de variables pour résoudre le problème posé, ce qui, une fois le premier plan terminé et le résultat constaté comme insuffisant, nécessitera de mettre en jeu au moins une variable supplémentaire, avec le risque d'avoir encore à recommencer.

Dans ce deuxième cas le nombre total réel des essais, que l'on est bien obligé de cumuler, s'élève rapidement.

En effet, il n'est pas possible pour les plans composites et pour les plans hybrides, de poursuivre un plan à p variables par des essais complémentaires autorisant le calcul ultérieur de l'effet d'une $p+$ 1ème variable, tout en tenant compte de la totalité des essais précédents.

Une solution est offerte par les plans de DOELHERT (3), mais ces derniers nécessitent de mettre en oeuvre :

5 niveaux pour la première variable
7 niveaux pour les variables suivantes
3 niveaux pour la dernière variable.

ce qui n'est pas toujours commode, voire possible, en milieu industriel.

C'est pour cette raison pratique, qu'il nous semble souhaitable de retenir une démarche différente pour l'optimisation des procédés industriels :

- compte-tenu des connaissances des spécialistes du procédé, il convient de dresser la liste des variables potentiellement explicatives, par ordre décroissant d'intérêt ou d'intensité d'effet supposée.
- il faut ensuite mettre progressivement en jeu des plans d'essais successifs, qui permettent à chaque stade d'inclure la totalité des essais précédents pour établir les modèles quadratiques progressifs à 3 puis 4 et 5 variables ... etc.

Cette démarche est tout à fait praticable, en faisant appel à un plan d'essai particulier, que nous appellerons conventionnellement plan quadratique gigogne.

Ce plan comporte 3 niveaux seulement, ce qui représente une simplification par rapport aux plans évoqués plus haut, et son principe est le suivant : on effectue les essais de telle sorte que tous les points représentatifs des variables explicatives, situés tout d'abord sur une sphère, se retrouvent ensuite sur une hypersphère de dimension 4, puis sur chacune des hypersphères comportant une dimension supplémentaire.

Il suffit pour celà de retenir simultanément les conditions suivantes, exprimées en variables explicatives codées, pour des hypersphères de rayon $\sqrt{2}$:

$$Xi = -1; 0 \text{ ou } 1 \qquad \Sigma Xi^2 = 2$$

L'écriture du plan quadratique gigogne ne pose pas de difficulté majeure.

En se limitant à quatre variables, et en omettant les niveaux 0 dans le but de rendre le tableau plus immédiatement lisible, on aboutit à la représentation suivante :

$X1$	$X2$	$X3$	$X4$
−	−		
+	−		
−	+		
+	+		
	−	−	
−		−	
+		−	
	+	−	
	−	+	
−		+	
+		+	
	+	+	
		−	−
	−		−
−			−
+			−
	+		−
		+	−
		−	+
	−		+
−			+
+			+
	+		+
		+	+

On trouvera sur le tableau 1, le plan gigogne poursuivi jusqu'à 6 variables, et susceptible d'être complété par une écriture analogue : pour la p + 1ème variable à prendre en compte, il faut rajouter 4 p essais, pour moitié au niveau -1, et pour moitié au niveau 1, puis écrire les valeurs des niveaux des p variables précédentes en se remémorant la règle $\Sigma Xi^2 = 2$.

Un exemple en est donné par le tableau 2 qui indique comment introduire une 7ème variable.

Le nombre total d'essais à réaliser pour parvenir à l'estimation du modèle quadratique comportant également toutes les interactions d'ordre 2, a pour expression, dans le cas du plan quadratique gigogne :

$$\Sigma n = 2p(p-1)$$

Ce nombre total d'essais croît rapidement avec le nombre de variables indroduites dans le modèle représentatif, tout en demeurant inférieur à celui du plan composite centré.

Nombre total d'essais des plans gigognes

Nombre de variables p	Nombre de coefficients c	Plan quadratique gigogne	Essais complémentaires
3	10	12	
4	15	24	12
5	21	40	16
6	28	60	20
7	36	84	24

Dans la dernière colonne du tableau ci-dessus, nous avons fait figurer le nombre d'essais complémentaires permettant d'accéder à la connaissance des effets d'une p + 1ème variable, car dans la réalité c'est ce dernier chiffre qui sera déterminant pour la décision de poursuite des essais, dans la mesure où l'objectif visé n'a pas encore été atteint en se limitant à l'étude des p premières variables, que l'on désire conserver en totalité dans le modèle représentatif.

Pour le calcul de ce modèle représentatif, on doit faire intervenir des essais au centre ($\Sigma Xi^2 = 0$), qui en fait correspondent à la marche normale du procédé, et ne nécessitent donc pas d'expérimentation supplémentaire.

Leur nombre sera choisi en se rapportant au tableau suivant, qui indique la corrélation entre les termes Xi^2. Les autres corrélations sont toutes nulles.

Valeur de R Xi^2 pour les plans gigognes

n_0 \ p	3	4	5	6	7
0	−0,50	−0,33	−0,25	−0,20	−0,17
4	0	−0,17	−0,18	−0,16	−0,15
8	0,17	−0,07	−0,13	−0,13	−0,13
12	0,25	0	−0,08	−0,11	−0,11
N.B : la condition rigoureuse d'orthogonalité est donnée par $No = (p-1)(p-2)$					

Nous indiquons sur la figure 1 la position des points représentatifs pour le plan de départ qui comporte les 3 premières variables. Il s'agit d'un tétradécaèdre irrégulier comportant 6 carrés et 8 triangles équilatéraux.

Pour l'introduction des variables suivantes, on notera que leur ordre est indifférent pour le calcul du modèle final, de sorte que compte-tenu des premiers résultats, on pourra éventuellement modifier la hiérarchie préalablement retenue.

S'agissant enfin du choix du nombre total p de variables explicatives à faire intervenir pour l'optimisation d'un procédé dont on veut ajuster q propriétés, on tiendra compte de ce qui suit :

• si $p < q$

Il n'existe pas de solution rigoureuse, sauf cas très particulier, et seul un compromis est possible qui permettra de s'approcher au mieux des niveaux souhaités pour chacune des propriétés.

• si $p = q$

Il existe au moins une solution rigoureuse, qui peut être soit à l'intérieur, soit à l'extérieur du domaine technologiquement accessible. Dans le cas où la solution est située à l'extérieur, le compromis se localisera aux bornes du domaine étudié.

• si $p > q$

Il existe dans ce cas une infinité de solutions théoriques, dont un nombre fini peut se trouver à l'intérieur du domaine technologique.

On choisira dans ce dernier cas la solution la plus proche des conditions habituelles de fonctionnement.

L'intérêt du plan quadratique gigogne est de permettre pour q propriétés, d'augmenter progressivement le nombre p de variables de réglage du procédé industriel, et donc d'avoir la possibilité de traverser successivement les trois domaines évoqués plus haut :

- celui du compromis ($p < q$)
- celui de la solution rigoureuse ($p = q$)
- celui qui autorise le choix entre plusieurs solutions ($p > q$)

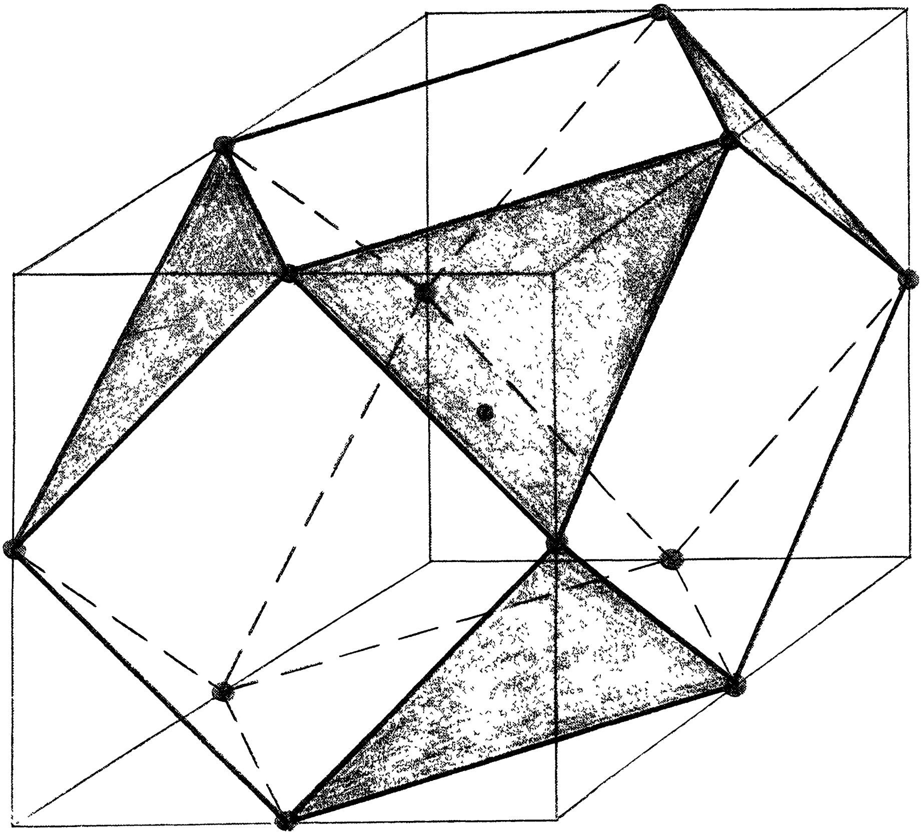

FIGURE 1
Les points représentent les essais à effectuer pour les 3 variables initiales du plan quadratique gigogne.

Par ailleurs, contrairement aux plans composites centrés et aux plans hybrides, on pourra au fur et à mesure de l'expérimentation, calculer l'effet des variables déjà expérimentées, sans avoir à attendre la fin complète du plan d'expériences.

Cette possibilité est du plus grand intérêt, car l'expérimentateur est toujours rebuté par l'impossibilité de pouvoir prendre une décision avant la fin des essais d'un plan classique à plusieurs variables, surtout si les essais sont longs et coûteux.

C'est la raison pour laquelle les stratégies séquentielles sont préférables. On notera à ce sujet qu'il est par exemple possible de n'effectuer au départ que 4 essais, en prenant en compte soit 2, soit 3 variables.

La figure 2 indique, dans le cas de 2 variables, quels sont les 4 essais de départ.

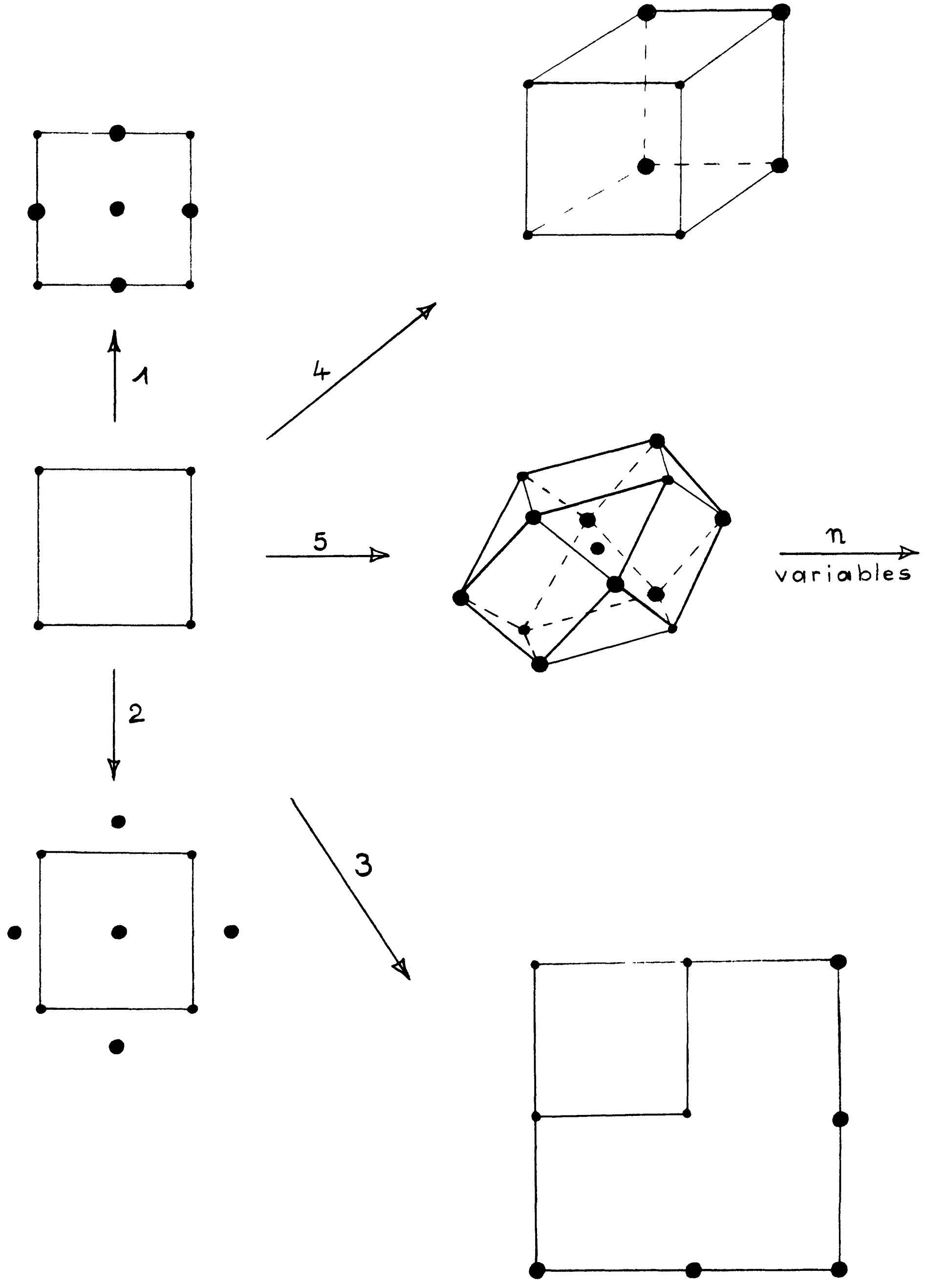

FIGURE 2

Démarche séquentielle prenant en compte 2 variables au départ.
Les points de petite dimension indiquent la position des 4 premiers essais.
Les points de grande dimension localisent les essais suivants.

Compte-tenu des premiers résultats obtenus avec un modèle simple :

$$y = bo + b1X1 + b2X2 + b12X1X2$$

on aura le choix entre 5 possibilités :

1) explorer un modèle quadratique sans modifier les bornes du domaine.
2) élargir les bornes du domaine pour parvenir également à un modèle quadratique.
3) modifier les bornes et le centre du domaine, en se dirigeant vers la meilleure solution entrevue, tout en ayant accès au modèle quadratique.
4) rajouter une troisième variable excentrée en conservant un modèle simple, analogue à celui de départ.
5) poursuivre par un plan quadratique gigogne, qui autorisera, si celà est approprié, d'introduire ultérieurement d'autres variables supplémentaires.

Dans le cas de 3 variables, la figure 3 illustre la démarche séquentielle :

On effectue au départ 4 essais selon un plan factoriel fractionnaire, qui correspond au modèle :

$$y = bo + b1X1 + b2X2 + b3X3$$

On peut ensuite poursuivre les essais selon deux modalités :

1) compléter le plan factoriel 2^3 pour avoir accès aux interactions, et éventuellement terminer par un composite centré qui permettra le calcul des effets carrés.
2) se diriger vers la meilleure solution entrevue en poursuivant par un plan quadratique gigogne, dont le centre va correspondre au meilleur des 4 premiers essais. Là encore on aura la latitude de rajouter par la suite d'autres variables, si celà s'avère souhaitable.

Bibliographie

[1] BOX G.E.P., WILSON K.B J., *Roy. Stat.Soc.*, B 13, 1 (1951).

[2] ROQUEMORE K. G., *Technometrics*, 18, 4 (1976).

[3] DOELHERT D.H. *Appl. Statis.*, 19, 3 (1970).

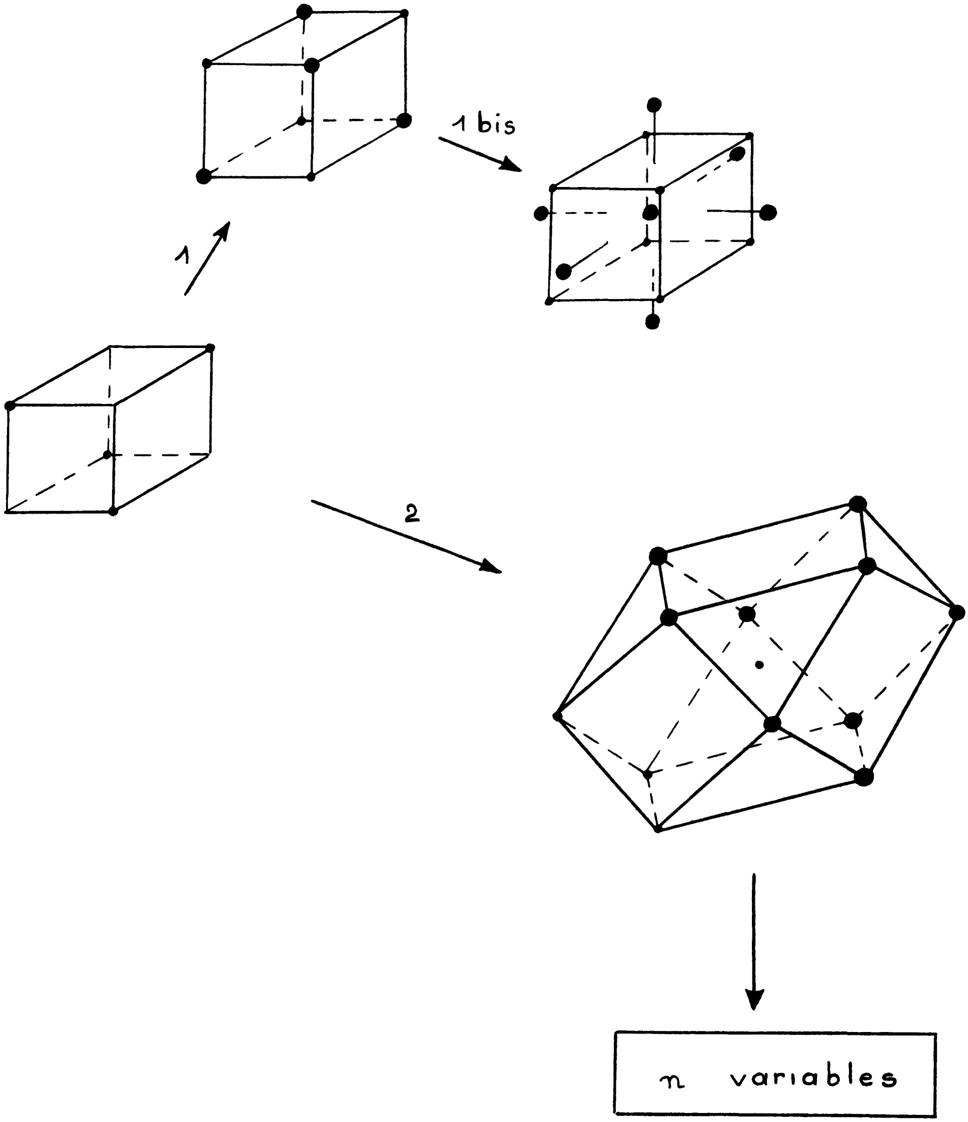

FIGURE 3
Démarche séquentielle prenant en compte 3 variables au départ.
Les 4 premiers essais sont représentés par des points de petite dimension.
Les essais complémentaires sont caractérisés par des points de plus grande dimension.

TABLEAU 1
Plan quadratique gigogne pour 6 variables

N°	X1	X2	X3	X4	X5	X6
1	−1	−1	0	0	0	0
2	1	−1	0	0	0	0
3	−1	1	0	0	0	0
4	1	1	0	0	0	0
5	0	−1	−1	0	0	0
6	−1	0	−1	0	0	0
7	1	0	−1	0	0	0
8	0	1	−1	0	0	0
9	0	−1	1	0	0	0
10	−1	0	1	0	0	0
11	1	0	1	0	0	0
12	0	1	1	0	0	0
13	0	0	−1	−1	0	0
14	0	−1	0	−1	0	0
15	−1	0	0	−1	0	0
16	1	0	0	−1	0	0
17	0	1	0	−1	0	0
18	0	0	1	−1	0	0
19	0	0	−1	1	0	0
20	0	−1	0	1	0	0
21	−1	0	0	1	0	0
22	1	0	0	1	0	0
23	0	1	0	1	0	0
24	0	0	1	1	0	0
25	0	0	0	−1	−1	0
26	0	0	−1	0	−1	0
27	0	−1	0	0	−1	0
28	−1	0	0	0	−1	0
29	1	0	0	0	−1	0
30	0	1	0	0	−1	0

31	0	0	1	0	−1	0
32	0	0	0	1	−1	0
33	0	0	0	−1	1	0
34	0	0	−1	0	1	0
35	0	−1	0	0	1	0
36	−1	0	0	0	1	0
37	1	0	0	0	1	0
38	0	1	0	0	1	0
39	0	0	1	0	1	0
40	0	0	0	1	1	0
41	0	0	0	0	−1	−1
42	0	0	0	−1	0	−1
43	0	0	−1	0	0	−1
44	0	−1	0	0	0	−1
45	−1	0	0	0	0	−1
46	1	0	0	0	0	−1
47	0	1	0	0	0	−1
48	0	0	1	0	0	−1
49	0	0	0	1	0	−1
50	0	0	0	0	1	−1
51	0	0	0	0	−1	1
52	0	0	0	−1	0	1
53	0	0	−1	0	0	1
54	0	−1	0	0	0	1
55	−1	0	0	0	0	1
56	1	0	0	0	0	1
57	0	1	0	0	0	1
58	0	0	1	0	0	1
59	0	0	0	1	0	1
60	0	0	0	0	1	1

TABLEAU 2
Introduction d'une 7ème variable. Essais complémentaires
Les niveaux 0 correspondent aux cases vides, pour permettre une lecture aisée

n	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	$X7$
61						−	−
62					−		−
63				−			−
64			−				−
65		−					−
66	−						−
67	+						−
68		+					−
69			+				−
70				+			−
71					+		−
72						+	−
73						−	+
74					−		+
75				−			+
76			−				+
77		−					+
78	−						+
79	+						+
80		+					+
81			+				+
82				+			+
83					+		+
84						+	+