

# THESE

PRESENTEE A

L'UNIVERSITE DE PAU  
ET  
DES PAYS DE L'ADOUR

ECOLE DOCTORALE DES SCIENCES EXACTES ET DE LEURS  
APPLICATIONS

PAR

**Astrid JOURDAN**

POUR OBTENIR LE GRADE DE

DOCTEUR

**Spécialité :**  
**Mathématiques Appliquées**

*ANALYSE STATISTIQUE ET ECHANTILLONNAGE  
D'EXPERIENCES SIMULEES*

Soutenue le 27 juin 2000

Après avis de

M. DAUDIN Jean-Jacques, Professeur, INA-PG, PARIS ..... Rapporteur  
M. SAPORTA Gilbert, Professeur, CNAM, PARIS ..... Rapporteur

Devant la Commission d'examen formée des rapporteurs et de :

M. AMARA Mohamed, Professeur, Université de PAU ..... Président du Jury  
M. COLLOMBIER Dominique, Professeur, Université de STRASBOURG ..... Directeur de Thèse



# *Table des matières*

Introduction .....	1
<b><i>Première Partie : Deux approches statistiques classiques des expériences simulées</i></b>	
Introduction .....	7
1. Approche par résidu aléatoire.....	9
1.1. Modèle avec corrélations spatiales .....	10
1.2. Le meilleur prédicteur linéaire sans biais .....	11
1.2.1. Rappels sur l'estimation .....	11
1.2.2. Construction du prédicteur.....	12
1.2.3. Etude de l'erreur .....	14
1.2.4. Krigeage .....	18
1.3. La covariance du processus résiduel.....	20
1.3.1. Fonction de corrélation .....	20
1.3.2. Estimation des paramètres de la covariance.....	25
1.4. Les plans pour une approche par résidu aléatoire .....	27
1.4.1. L'erreur quadratique moyenne intégrée.....	27
1.4.2. L'entropie.....	28
1.4.3. La distance maximin .....	29
2. Approche par échantillon aléatoire .....	35
2.1. Modélisation et estimation .....	36
2.1.1. Le modèle .....	36
2.1.2. Estimation des paramètres du modèle .....	36
2.1.3. Estimation d'une intégrale : méthode de Monte Carlo .....	37
2.2. Méthodes de réduction de la variance .....	39
2.2.1. Echantillonnage à deux degrés .....	39
2.2.2. Décomposition de la variance .....	41
2.2.3. Méthode des hypercubes latins.....	42
2.2.4. Méthode des tableaux orthogonaux.....	44
2.3. Méthodes quasi-Monte Carlo.....	48
2.3.1. Méthodes quasi-Monte Carlo : discrédance .....	48
2.3.2. Les réseaux (lattice).....	50
2.3.3. Propriétés de projection des réseaux d'échantillonnage.....	55
2.4.1. Les nets (ou séquences) : méthode d'équidistribution.....	58
2.4.2. Les nets (ou séquences) : méthode de réduction de la variance.....	60
<b><i>Deuxième Partie : Une nouvelle approche statistique des expériences simulées</i></b>	
Introduction .....	69
3. Nouvelle approche : Le modèle .....	71
3.1. Présentation du modèle .....	72
3.2. Prédiction .....	74
3.3. Comportement du modèle .....	77
3.3.1. Comportement de la partie aléatoire du modèle .....	77
3.3.2. Contribution de la partie fixe du modèle .....	81
4. Régression linéaire trigonométrique ou modèle de Fourier.....	85
4.1. Définition de la régression linéaire trigonométrique .....	86
4.2. Construction d'une base orthonormée pour la fonction.....	89
4.2.1. Représentations linéaires des groupes finis .....	89
4.2.2. Base de Yates .....	91
4.3. Estimabilité et optimalité.....	95
<i>Optimalité des plans pour une régression linéaire complexe.....</i>	<i>95</i>
4.4. Régression trigonométrique dans l'approche par résidu aléatoire .....	100

5.	Les plans .....	105
5.1.	Fractions régulières .....	106
5.1.1.	Distribution marginale d'une fraction régulière .....	106
5.1.2.	Lien entre réseaux et fractions régulières .....	109
5.1.3.	Ajustement d'une régression trigonométrique .....	113
5.2.	Fractions régulières à éléments dans $GF(p)^d$ .....	114
5.2.1.	La résolution .....	116
5.2.2.	L'estimabilité .....	119
5.3.	Tableaux orthogonaux linéaires .....	121
5.4.	Tests numériques sur les tableaux orthogonaux linéaires .....	124
5.4.1.	Robustesse des meilleurs plans par rapport aux variations de $\theta$ .....	125
5.4.2.	Influence du paramètre $p$ et de la force du tableau sur l'erreur.....	127
5.4.3.	Test des critères de sélection des plans.....	129
<b>Troisième Partie : Construction de plans</b>		
	Introduction .....	135
6.	Construction de tableaux orthogonaux linéaires .....	137
6.1.	Construction d'un tableau orthogonal de force 2 par la méthode des différences.....	138
6.2.	Construction d'un tableau orthogonal linéaire de force $t$ .....	141
6.2.1.	Propriétés des matrices aux différences .....	141
6.2.2.	Lien entre tableaux orthogonaux lin. de force $t-1$ et matrices aux différences de force $t$ ..	145
6.2.3.	Construction d'un tableau orthogonal linéaire de force $t$ .....	149
6.3.	Construction d'un tableau orthogonal linéaire de force $t+m$ , $m \geq 0$ .....	154
6.3.1.	Lien entre tableau orthogonal linéaire de force $t+m$ et matrice aux différences.....	154
6.3.2.	Construction d'un tableau orthogonal linéaire de force $t+m$ , $m \geq 0$ .....	158
7.	Construction de fractions régulières dans $G = \prod_{i=1}^d (Z/p^{n_i})$ .....	163
7.1.	Première étape .....	165
7.2.	Passage de $P'$ à $B'$ dans $G$ .....	172
7.2.1.	Deuxième étape : Construction de $\langle x \rangle^\perp$ .....	172
7.2.2.	Troisième étape : Construction de $\langle x_1, \dots, x_r \rangle^\perp$ dans $G$ .....	177
8.	Fractions semi-régulières .....	185
8.1.	Définitions et généralités .....	186
8.1.1.	Définitions.....	186
8.1.2.	Construction de $T$ à partir de $(P,C)$ .....	187
8.2.	Estimabilité .....	189
8.3.	Propriétés de projection .....	192
8.3.1.	Répartition uniforme des marges de même dimension : résolution.....	192
8.3.2.	Répartition uniforme sur certaines marges de dimension $R$ .....	197
	Conclusion .....	203
	Perspectives .....	205
	Bibliographie .....	207

# Introduction

A l'heure actuelle, de nombreux phénomènes physiques sont modélisés par des équations mathématiques, puis étudiés à l'aide de simulateurs. Ces codes de calcul, malgré les progrès sans cesse croissants de l'informatique, restent souvent très coûteux, et sont donc difficiles à exploiter. Une possibilité pour résoudre à ce problème, consiste à effectuer une série de simulations afin d'obtenir un certain nombre d'informations sur le phénomène, c'est la phase d'expérimentation. On traite ensuite ces informations suivant l'objectif recherché, c'est la phase analyse.

Tous les domaines scientifiques sont concernés par le sujet. Nous pouvons citer quelques exemples tels que la simulation d'un système de refroidissement dans un procédé de moulage (Kenny (1998)), ou bien la résolution numérique d'un système d'équations différentielles issues de la mécanique des fluides pour modéliser la vitesse de flamme lors de la combustion du méthane (Sacks, Schiller et Welch (1989)), ou encore des simulations de circuits électroniques (Nassif, Strojwas et Director (1984)), *etc...*

Pour chaque simulation, l'utilisateur doit spécifier les variables d'entrée du code. Dans ce travail, seul le cas de variables continues est traité mais la plupart des résultats s'adaptent facilement au cas discret. Soit  $x \in \mathbb{R}^d$ , le *vecteur des paramètres* de simulation, représentant les variables d'entrée du code. Après une transformation linéaire classique, nous supposons que  $x$  appartient au cube unité  $[0,1]^d$ . On note alors la *réponse du simulateur*,

$$y(x), x \in [0,1]^d.$$

Afin de simplifier notre propos, nous supposons que  $y(x) \in \mathbb{R}$ , bien qu'en réalité le simulateur fourni souvent plusieurs réponses.

Dans ce contexte, l'expérimentation consiste à fixer plusieurs valeurs  $x_1, \dots, x_N$  du vecteur des paramètres d'entrée, et à récolter les réponses du simulateur en ces points,  $y(x_1), \dots, y(x_N)$ . Par opposition à l'expérimentation physique, une expérience menée à l'aide d'un code de calcul est dite *simulée*. Les objectifs de cette expérimentation sont très divers, par exemple

- prédire la réponse du simulateur,
- optimiser une fonction de la réponse,

- régler des paramètres du code pour que sa réponse coïncide avec les données réelles, *etc...*

Dans ce travail, seul le problème de la **prédiction** est envisagé. Deux questions se posent alors:

- Le problème du plan : " A quels points  $D=\{x_1, \dots, x_N\}$  du cube unité doit-on récolter les observations  $y(x_1), \dots, y(x_N)$  ?"
- Le problème de l'analyse : " Comment doit-on utiliser les observations  $y(x_1), \dots, y(x_N)$  pour prédire la réponse du simulateur ?"

Tout notre travail s'articule autour de ces deux questions. Depuis le début du siècle, des auteurs tel que Fisher tentent de répondre à ce problème. Cela a donné lieu à une branche spécifique des statistiques connue sous le nom de plans d'expériences. Pourquoi alors engager une recherche dans ce domaine si des outils performants existent déjà? Tout simplement parce que les simulateurs que nous étudions ici ne rentrent pas dans le contexte de la planification d'expériences classique. En effet, ces codes sont déterministes, autrement dit, si on effectue plusieurs simulations avec le même vecteur des paramètres fixé, on obtient des réponses identiques. Le simulateur n'étant pas bruité, pour tout  $x$  fixé,  $y(x)$  est reproductible puisqu'elle est entièrement déterminée par les paramètres de simulation. Il est alors évident dans ce contexte que l'approche statistique classique de la modélisation linéaire et des plans d'expériences n'est plus valable pour répondre aux deux questions posées.

Ce mémoire de thèse se décompose en trois parties.

Dans la première, nous allons présenter deux approches de ce problème, que l'on peut trouver dans la bibliographie. Chacune d'elle traite la question d'un point de vue statistique, elles ont donc pour origine l'interrogation suivante.

*Comment introduire de l'aléatoire dans un problème déterministe?*

Les deux méthodes se différencient par la réponse qu'elles apportent à cette question. Une première approche consiste à introduire l'aléa au travers du modèle proposé pour la réponse du simulateur. Nous l'appelons **approche par résidu aléatoire**. La deuxième méthode quant à elle, fait apparaître l'aléa au niveau du plan. Nous l'appelons **approche par échantillon aléatoire**. Ces méthodes répondent différemment aux deux questions posées, c'est-à-dire que chacune proposent un modèle pour  $y(x)$  construit à partir des observations  $y(x_1), \dots, y(x_N)$ , ainsi que des plans appropriés au modèle. Nous allons donc faire le bilan de ces approches de façon à mettre à jour leurs avantages mais aussi leurs inconvénients, afin de proposer une nouvelle méthode plus performante.

Dans la deuxième partie, nous proposons donc notre propre approche statistique du problème. Celle-ci associe les différents points forts que nous avons dégager des deux

méthodes existantes, tout en tentant d'en atténuer les inconvénients. Là encore, nous proposons un modèle pour la réponse du simulateur et des plans appropriés au modèle. L'aléatoire est ici introduit aux deux niveaux en utilisant les techniques de la première partie. Le modèle utilisé est celui de l'approche par résidu aléatoire auquel on ajoute un terme d'erreur et que l'on enrichit d'une **régression linéaire trigonométrique**. Les plans eux sont construits suivant les principes de l'approche par échantillon aléatoire. Nous utilisons pour cela une structure classique en plans d'expériences connue sous le nom de **fractions régulières**. Notre objectif est de montrer que ces plans sont adaptés au modèle proposé et qu'ils sont performants dans le contexte des expériences simulées.

La dernière partie est une extension afin de proposer des méthodes de construction des plans utilisés. En effet, par leur définition même, ces plans sont particulièrement bien adaptés au modèle que nous avons défini, mais très souvent nous ne savons pas les construire. Seules quelques techniques sont proposées dans la bibliographie concernant des cas particuliers peu intéressants pour notre problème. Pour terminer, nous étendons nos méthodes d'échantillonnage à une nouvelle catégorie de plans qui semble être prometteuse notamment dans le cas d'une régression trigonométrique.

## **Références**

Kenny Q.Y. (1998). Orthogonal column latin hypercubes and their application in computer experiments. *J. Amer. Statist. Ass.* **93**, 1430-1439.

Nassif S.R., Strojwas A.J. , Director S. W. (1984). FABRICS II : A statistically based IC fabrication process simulator. *IEEE Trans. Computer Aided Designs* **CAD-3**, 40-46.

Sacks J., Schiller S.B. , Welch W.J. (1989). Designs for computer experiments. *Technometrics* **31**, 41-47.



# Première partie

*Deux approches statistiques classiques des  
expériences simulées*



# Introduction

Dans cette première partie nous allons présenter les deux principales approches statistiques des expériences simulées, rencontrées dans la bibliographie.

Ces deux méthodes répondent aux questions posées en introduction, c'est-à-dire que les deux proposent,

- un modèle pour la réponse du simulateur, puis un prédicteur (ou estimateur) construit à partir des simulations  $y(x_1), \dots, y(x_N)$ , obtenues au moyen d'un plan d'échantillonnage,
- différents plans permettant d'analyser efficacement le modèle choisi.

Les modèles et les plans proposés ne sont pas ceux que l'on utilise habituellement en planification. En effet, ce travail se situe dans le cadre de simulations déterministes, pour traiter ce problème avec des outils statistiques, il est donc nécessaire d'introduire de l'aléatoire.

La première approche, par résidu aléatoire, consiste à introduire l'aléa au travers le modèle proposé pour la réponse du simulateur. Bien que déterministe, on considère cette dernière comme la réalisation d'une fonction aléatoire. Cette fonction se décompose en deux parties, la première est déterministe et correspond à un modèle linéaire classique fournissant une approximation de la réponse du simulateur en moyenne. L'aléatoire est introduit avec la seconde partie du modèle à l'aide d'un processus stochastique qui permet de prendre en compte la répartition spatiale des données, et donc les variations du résidu d'approximation.

Les plans utilisés dans cette première approche n'ont aucune structure particulière, ils sont construits de façon à optimiser des critères de qualité.

La deuxième approche, par échantillon aléatoire, propose une simple régression linéaire en tant que modèle. On suppose en fait que la réponse du simulateur peut se décomposer sur une base de fonctions, et que seuls les termes prédominants de cette décomposition suffisent à la modéliser.

Les plans proposés dans cette approche sont issus de techniques d'échantillonnage utilisées en intégration numérique. L'aléa est introduit ici en choisissant les points du plan au hasard sur le domaine d'échantillonnage, par une méthode de Monte Carlo. Cette technique est bien entendu améliorée en imposant des contraintes aux points du plan de façon à ce que ceux-ci

soient bien répartis dans le cube unité. On utilise pour cela une méthode d'échantillonnage à deux degrés, *i.e* à deux niveaux. Le premier degré impose les contraintes au plan et le deuxième degré permet d'obtenir un échantillon aléatoire.

On voit d'ores et déjà que la première approche privilégie le modèle aux plans utilisés alors qu'*a contrario* la deuxième approche utilise des techniques d'échantillonnage élaborées mais avec une modélisation sommaire. L'objectif principal de cette première partie est de présenter ces deux approches, tout en notant les avantages, inconvénients et particularités de chacune d'elles. Nous terminons chacun des chapitres par un bilan de l'approche présentée, afin de dégager les points positifs qu'il est intéressant de conserver dans la nouvelle approche que nous présentons dans la deuxième partie, ainsi que les points négatifs qu'il faut éviter.

En ce qui concerne les plans, un rapide bilan bibliographique nous a permis de regrouper en quatre points les principales qualités requises pour un plan.

1. Robustesse des plans : Kenny (1998) définit un bon plan comme un plan permettant d'analyser efficacement différents modèles.
2. Efficacité des plans : Bien entendu, un bon plan doit être efficace, c'est-à-dire minimiser une erreur. Celle-ci est définie différemment suivant l'approche envisagée.
3. Répartition des points du plan : Bates *et al.* (1996) pensent que la qualité première d'un plan est sa répartition dans le domaine d'échantillonnage. Il doit "remplir" de la façon la plus uniforme possible le cube unité ("space filling designs").
4. Construction des plans : Les problèmes engendrés par la construction des plans sont de deux sortes. Les premiers sont d'ordre numérique : le coût de calcul pour construire un plan doit rester raisonnable. Les deuxièmes sont d'ordre théorique : il faut tout simplement que ces plans existent et que l'on sache les construire.

Le deuxième objectif de ces deux chapitres est de mettre en place des outils que nous allons utiliser pour définir notre propre approche du problème. Certaines techniques ou notions présentées dans cette première partie n'appartiennent pas à la bibliographie sur les expériences simulées. Nous les définissons cependant ici afin de devancer nos besoins dans la deuxième partie.

## **Références**

Bates R.A., Buck R.J., Riccomagno E., Wynn H.P. (1996). Experimental design and observation for large systems. *J. R. Statist. Soc. B* **58**, 77-94.

Kenny Q.Y. (1998). Orthogonal column latin hypercubes and their application in computer experiments. *J of Am. Statist. Ass.* **93**, 1430-1439.

# 1. Approche par résidu aléatoire

La première approche envisagée dans la bibliographie consiste à introduire l'aléatoire au travers le modèle proposé pour la réponse du simulateur. Celui-ci se décompose en deux parties. La première est déterministe et correspond à un modèle linéaire classique fournissant une approximation de la réponse du simulateur en moyenne. La deuxième est aléatoire et est définie par un processus stochastique qui permet de prendre en compte le caractère spatial des données, et donc les variations du résidu d'approximation. C'est pourquoi nous l'appelons approche par résidu aléatoire. Ce type de modèle est couramment utilisé par les géostatisticiens.

Dans la première partie de ce chapitre, sont présentés le modèle et le meilleur prédicteur linéaire sans biais de la réponse du simulateur. Ensuite, on étudie le fonctionnement de ce modèle pour répondre par exemple aux questions suivantes.

*Quels sont ses particularités par rapport à un modèle linéaire classique?*

*Comment se comportent les deux parties du modèle?*

*Quels avantages et inconvénients cela présente-t-il? Etc ...*

La dernière partie du chapitre est consacrée à la présentation des plans utilisés pour ce type de modèle. Ils sont construits pour optimiser des critères de qualité qui dépendent tous du modèle fixé. Nous présentons les trois principaux critères rencontrés dans la bibliographie.

Le chapitre se termine par un bilan de cette première approche de façon à dégager les points positifs qu'il serait intéressant de conserver pour une nouvelle approche et les points négatifs qu'il faudrait éviter. Un bilan point par point est aussi effectué afin de savoir si les plans utilisés atteignent les quatre objectifs fixés en introduction.

## 1.1. Modèle avec corrélations spatiales

Dans cette première approche, l'aléatoire est introduit *via* le modèle, en considérant la réponse déterministe du simulateur comme la réalisation d'une fonction aléatoire,

$$Y(x) , x \in [0,1]^d,$$

où  $d$  représente le nombre de facteurs pris en compte par le code et  $[0,1]^d$  est le domaine d'échantillonnage. Cette méthode, qui est la plus couramment utilisée, a été proposée par Sacks, Welch, Mitchell et Wynn (1989) et Sacks Schiller et Welch (1989).

Pour analyser cette réponse il faut la modéliser et pour cela nous allons supposer que pour tout  $x \in [0,1]^d$ ,  $E[Y(x)]$  et  $\text{var}Y(x)$  existent. On peut alors décomposer  $Y(x)$  en deux parties,

$$Y(x) = Z(x)\beta + \Gamma(x) , x \in [0,1]^d.$$

- $Z(x)\beta = E[Y(x)]$  est la partie déterministe du modèle. Il s'agit en fait d'un modèle linéaire classique.

- $\Gamma(x)$  est la partie aléatoire du modèle. La méthode statistique traditionnelle consiste à considérer  $\Gamma(x)$  comme un bruit blanc correspondant à une erreur de mesure. Cependant, dans le cas d'un simulateur déterministe, les expériences sont nécessairement sans erreur,  $\Gamma(x)$  représente alors l'écart systématique entre la réponse du code et la régression linéaire présumée. Ce comportement est illustré dans le paragraphe 1.2.3.

### Définition 1.1

*Le modèle est défini par*

$$Y(x) = Z(x)\beta + \Gamma(x) , x \in [0,1]^d,$$

*avec*

- $Z(x)\beta$  la régression linéaire où  $Z(x)$  est le vecteur du modèle d'ordre  $1 \times m$  et  $\beta$  est le vecteur inconnu des paramètres du modèle d'ordre  $m \times 1$ ,
- $\Gamma(x)$  le processus résiduel tel que  $E[\Gamma(x)] = 0$  et  $\text{cov}(\Gamma(x), \Gamma(u)) = \sigma^2 R(x, u)$   $\forall x \in [0,1]^d, \forall u \in [0,1]^d$  où  $\sigma^2$  est la variance et  $R$  est la fonction de corrélation du processus.

Les propriétés du processus résiduel sont présentées plus en détail dans le paragraphe 1.3.

**Remarque :** Une approche bayésienne de ce modèle est aussi envisagée dans la bibliographie par Currin *et al.* (1988). Les composantes  $\beta_j$  du vecteur des paramètres  $\beta$  sont alors considérées comme des variables aléatoires auxquelles on attribue une distribution *a priori*.

Dans un premier temps, nous allons supposer que la forme de la fonction de corrélation est connue afin de construire un prédicteur de la réponse  $Y(x)$ . Ensuite nous verrons quelle forme donner à cette fonction et comment estimer ses paramètres.

## 1.2. Le meilleur prédicteur linéaire sans biais

Tout d'abord, revenons sur une question de vocabulaire. Nous parlons de prédiction pour des quantités aléatoires, par exemple  $Y(x)$ , et d'estimation pour des quantités fixes, par exemple  $\beta$ .

Voici des définitions de matrices et vecteurs construits à partir d'un plan d'échantillonnage  $D=\{x_1, \dots, x_N\}$  nécessaire à la prédiction.

- $Y$  :  $N \times 1$  **vecteur des observations** (réponses du simulateur) aux points du plan
- $Z = (Z(x_i))_{i=1, \dots, N}$  :  $N \times m$  **matrice du modèle** aux points du plan
- $R = (R(x_i, x_j))_{i, j=1, \dots, N}$  :  $N \times N$  **matrice de corrélation**,  $\text{cov}(Y) = \sigma^2 R$
- $r(x) = [R(x_1, x), \dots, R(x_N, x)]$  :  $N \times 1$  **vecteur de corrélation**,  $\text{cov}(Y(x), Y) = \sigma^2 r(x)$ .

La matrice de corrélation est supposée régulière.

### 1.2.1. Rappels sur l'estimation

Rappelons quelques définitions et résultats classiques sur l'estimation que l'on peut par exemple retrouver dans l'ouvrage de Christensen (1987).

#### Définitions 1.2

*On dit qu'une application linéaire  $K\beta$  est estimable si elle admet un estimateur linéaire sans biais de la forme  $AY$ , c'est-à-dire s'il existe  $A$  linéaire tel que*

$$E(AY) = K\beta.$$

*On appelle estimateur de Gauss-Markov de l'application linéaire  $K\beta$ , tout estimateur sans biais de  $K\beta$  de moindre dispersion (BLUE).*

#### Théorème 1.3 (Théorème de Gauss-Markov)

*Toute application linéaire  $K\beta$  estimable admet un estimateur de Gauss-Markov de la forme  $K\hat{\beta}$  où*

$$\hat{\beta} = (Z^T R^{-1} Z)^{-1} Z^T R^{-1} Y.$$

Tout plan d'échantillonnage pour lequel l'application linéaire est estimable est dit **adapté à cette estimation** (ou **adapté au modèle**). La proposition suivante permet de vérifier facilement si un plan possède cette propriété.

### Proposition 1.4

*Le vecteur des paramètres  $\beta$  est estimable si et seulement si les colonnes de la matrice du modèle  $Z$  sont indépendantes*

#### 1.2.2. Construction du prédicteur

Rappelons que le meilleur prédicteur linéaire sans biais  $f(x)$  (BLUP) de  $Y(x)$  doit vérifier

- $f(x)$  est linéaire en  $Y$
- $f(x)$  est sans biais :  $E[f(x)] = E[Y(x)] = Z(x)\beta$
- $f(x)$  minimise l'erreur quadratique moyenne :  $E[Y(x) - f(x)]^2$ .

Dans un premier temps, nous allons supposer le vecteur  $\beta$  connu et construire le BLUP de  $Y(x)$ .

### Proposition 1.5

*Pour  $\beta$  fixé,  $\forall x \in [0,1]^d$*

$$\bar{Y}(x) = Z(x)\beta + {}^t r(x)R^{-1}[Y - Z\beta]$$

*est le meilleur prédicteur linéaire sans biais de  $Y(x)$ .*

**Démonstration :** Sans perte de généralité, on peut écrire un prédicteur linéaire de  $Y(x)$  sous la forme

$$\underbrace{a}_{\in \mathbb{R}} + \underbrace{b}_{1 \times N}(Y - Z\beta).$$

1) Montrons dans un premier temps que le meilleur choix pour  $a$  est  $Z(x)\beta$ .

$$\begin{aligned} E[Y(x) - a - b(Y - Z\beta)]^2 &= E[(Y(x) - Z(x)\beta - b(Y - Z\beta)) + (Z(x)\beta - a)]^2 \\ &= E[(Y(x) - Z(x)\beta - b(Y - Z\beta))]^2 + E[(Z(x)\beta - a)]^2 \\ &\quad + 2E[(Y(x) - Z(x)\beta - b(Y - Z\beta))(Z(x)\beta - a)] \end{aligned}$$

$$\begin{aligned} \text{Or } E[(Y(x) - Z(x)\beta - b(Y - Z\beta))(Z(x)\beta - a)] &= (Z(x)\beta - a)E[Y(x) - Z(x)\beta - b(Y - Z\beta)] \\ &= (Z(x)\beta - a)[\underbrace{E[Y(x)]}_{=0} - Z(x)\beta + \underbrace{bE[Y]}_{=0} + bZ\beta] = 0. \quad \text{D} \end{aligned}$$

onc  $a=Z(x)\beta$  est le choix optimal pour  $a$  puisqu'il permet d'éliminer le terme  $[a-Z(x)\beta]^2$  et ainsi de minimiser l'erreur quadratique moyenne.

2) Montrons maintenant que le meilleur choix pour  $b$  est le vecteur  $\bar{b}$  solution de l'équation  $R^t b = r(x)$ .

$$\begin{aligned} E[Y(x) - Z(x)\beta - b(Y - Z\beta)]^2 &= E[Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta)]^2 + E[(\bar{b} - b)(Y - Z\beta)]^2 \\ &\quad + 2E[(Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta))(\bar{b} - b)(Y - Z\beta)]. \end{aligned}$$

$E[Y - Z\beta] = 0$  donc

$$\begin{aligned} E[(Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta))(\bar{b} - b)(Y - Z\beta)] &= (\bar{b} - b)\{E[Y(x)Y] - E[Y(x)Z\beta] - \sigma^2 \bar{b}R\} \\ &= (\bar{b} - b)\{\sigma^{2t}r(x) + E(Y)E[Y(x)] - Z\beta E[Y(x)] - \sigma^2 \bar{b}R\} \\ &= 0 \end{aligned}$$

puisque  $E(Y) = Z\beta$  et  $\bar{b}R = {}^t r(x)$ .

Il faut donc choisir  $b = \bar{b}$  afin d'éliminer le terme  $E[(\bar{b} - b)(Y - Z\beta)]^2$  et minimiser l'erreur quadratique moyenne.  $\square$

Le lemme suivant est directement obtenu à partir de la démonstration précédente.

### Lemme

*Soit  $a + b(Y - Z\beta)$  un prédicteur linéaire quelconque de  $Y(x)$ , alors*

$$E[Y(x) - (a + b(Y - Z\beta))]^2 = E[Y(x) - \bar{Y}(x)]^2 + E[\bar{Y}(x) - (a + b(Y - Z\beta))]^2.$$

Le prédicteur  $\bar{Y}(x)$  peut encore s'écrire sous la forme

$$\bar{Y}(x) = \tau + {}^t r(x)R^{-1}Y$$

où  $\tau = \tau(x) = K(x)\beta$  avec  $K(x) = [Z(x) - {}^t r(x)R^{-1}Z]$ .  $\beta$  étant le vecteur inconnu des paramètres du modèle,  $\bar{Y}(x)$  n'est pas connu. Dans le cas où  $\tau$  est estimable, une démarche logique consiste à le remplacer par un estimateur linéaire sans biais, c'est-à-dire de la forme

$$\hat{\tau} = A(x)Y \text{ où } E[A(x)Y] = \tau.$$

La proposition suivante donne l'erreur quadratique moyenne de ce nouveau prédicteur comme la somme de deux variances.

### Proposition 1.6

*Supposons le vecteur des paramètres estimable. Notons  $\hat{Y}(x)$  le prédicteur linéaire sans biais de  $Y(x)$ , obtenu en remplaçant  $\tau$  dans  $\bar{Y}(x)$  par un estimateur linéaire sans biais  $\hat{\tau}$ , alors  $\forall x \in [0,1]^d$*

$$E[Y(x) - \hat{Y}(x)]^2 = \text{var}(\hat{\tau}) + \text{var}({}^t r(x)R^{-1}Y - Y(x)).$$

**Démonstration :**  $\hat{Y}(x)$  est un prédicteur linéaire de  $Y(x)$  donc d'après le lemme

$$E[Y(x) - \hat{Y}(x)]^2 = E[Y(x) - \bar{Y}(x)]^2 + E[\bar{Y}(x) - \hat{Y}(x)]^2.$$

$$\begin{aligned}
E[Y(x) - \bar{Y}(x)]^2 &= E[Y(x) - \tau - {}^t r(x) R^{-1} Y]^2 \\
&= E[Y(x) - {}^t r(x) R^{-1} Y]^2 + \tau^2 - 2\tau E[Y(x) - \tau - {}^t r(x) R^{-1} Y] \\
&= E[Y(x) - {}^t r(x) R^{-1} Y]^2 - (E[Y(x) - \tau - {}^t r(x) R^{-1} Y])^2
\end{aligned}$$

puisque  $\tau = Z(x)\beta - {}^t r(x)R^{-1}Z\beta = E[Y(x) - {}^t r(x)R^{-1}Y]$ . Donc

$$E[Y(x) - \bar{Y}(x)]^2 = \text{var}(Y(x) - {}^t r(x)R^{-1}Y).$$

De plus  $E[\bar{Y}(x) - \hat{Y}(x)]^2 = E[\tau + {}^t r(x)R^{-1}Y - \hat{\tau} - {}^t r(x)R^{-1}Y]^2 = E[\tau - \hat{\tau}]^2 = \text{var}(\hat{\tau})$ .  $\square$

La deuxième variance étant indépendante du choix de l'estimateur de  $\tau$ , l'erreur quadratique moyenne est minimale si et seulement si la variance de  $\hat{\tau}$  est minimale. Or dans la classe des estimateurs linéaires sans biais, le meilleur estimateur de  $\tau$  (BLUE) est  $\hat{\tau} = K(x)\hat{\beta}$  où  $\hat{\beta}$  est l'estimateur de Gauss-Markov de  $\beta$  (définition 1.2), i.e  $\hat{\beta} = ({}^t Z R^{-1} Z)^{-1} {}^t Z R^{-1} Y$ . Nous pouvons alors énoncer le théorème suivant.

### Théorème 1.7

*Si le vecteur des paramètres est estimable alors le meilleur prédicteur linéaire sans biais (BLUP) de  $Y(x)$  est de la forme*

$$\hat{Y}(x) = Z(x)\hat{\beta} + {}^t r(x)R^{-1}[Y - Z\hat{\beta}], \forall x \in [0,1]^d$$

*avec  $\hat{\beta} = ({}^t Z R^{-1} Z)^{-1} {}^t Z R^{-1} Y$  l'estimateur de Gauss-Markov de  $\beta$ .*

**Remarque :** En général, un modèle bayésien conduit à des résultats différents. Cependant, dans le cas où le processus résiduel  $\Gamma$  est supposé gaussien et où les  $\beta_j$  suivent une loi normale dont la variance tend vers l'infini, alors les prédicteurs obtenus avec les deux modèles sont identiques (on peut par exemple se référer à Koehler et Owen (1996) pour une démonstration de ce résultat).

### 1.2.3. Etude de l'erreur

#### Interpolation

En premier lieu, nous pouvons montrer que le prédicteur passe par les réponses observées du simulateur.

#### Proposition 1.8

*Le prédicteur interpole les réponses du simulateur aux points du plan,  $\forall x_i \in D$*

$$\hat{Y}(x_i) = Y(x_i).$$

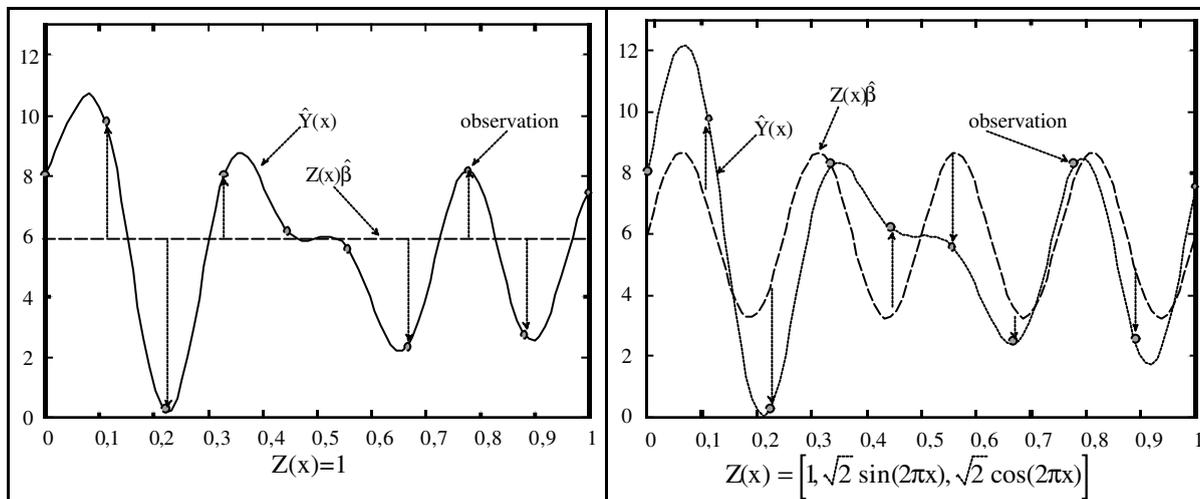
**Démonstration :** Soit  $x_i$  le  $i$ ème points du plan, alors  ${}^t r(x_i)R^{-1}$  est le vecteur ligne formé de  $N$  zéros sauf le  $i$ ème éléments qui est égal à 1, puisque  $R$  a pour  $i$ ème colonne  $r(x_i)$ . Donc

$$\hat{Y}(x_i) = Z(x_i)\hat{\beta} + (Y - Z\hat{\beta})_i = Z(x_i)\hat{\beta} + Y_i - Z(x_i)\hat{\beta} = Y_i. \quad \square$$

Que signifie cette interpolation concrètement? En fait, on peut remarquer que le prédicteur s'écrit comme la somme de l'estimateur de Gauss-Markov de  $E[Y(x)]$  plus un terme correctif

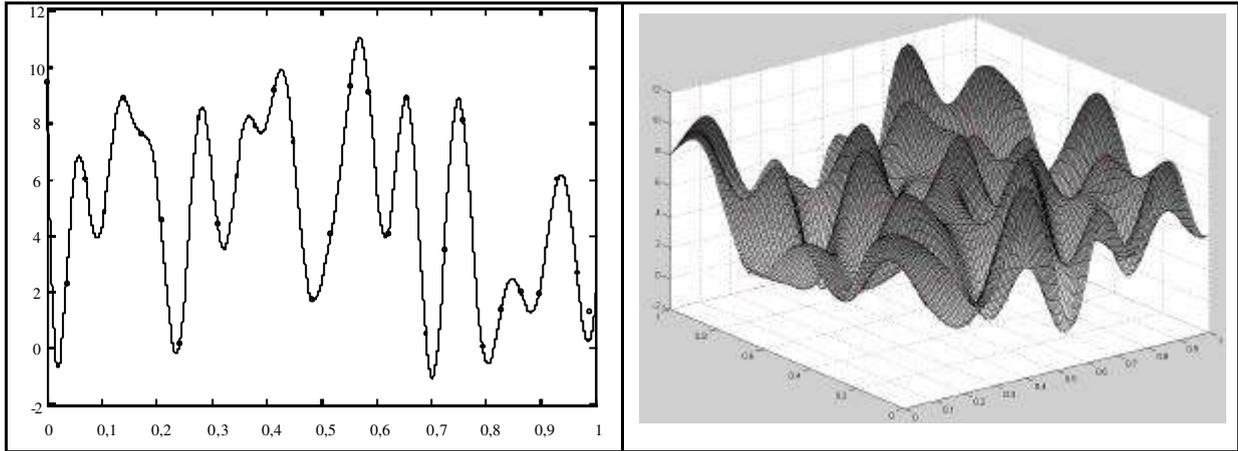
$$\hat{Y}(x) = \underbrace{Z(x)\hat{\beta}}_{\text{Gauss-Markov}} + \underbrace{r(x)R^{-1}[Y - Z\hat{\beta}]}_{\text{correction}}.$$

C'est ce dernier qui va corriger l'estimateur de la partie fixe du modèle de façon à interpoler les réponses observées du simulateur, même dans le cas extrême où cette partie est une constante. Les courbes ci-dessous permettent de mieux visualiser le phénomène. Elles représentent le prédicteur  $\hat{Y}(x)$  ainsi que  $Z(x)\hat{\beta}$  qui correspond au prédicteur moins le terme correctif, et ceci pour deux types de régressions linéaires, constante et sinusoidale. Le plan est ici constitué de 10 points également espacés sur  $[0,1]$  et les observations (réponses présumées d'un simulateur) sont choisies au hasard sur  $[0,10]$ .



**Remarque :** Par souci de clarté, la plupart des illustrations numériques présentées sont en dimension 1 ou 2. Mais les phénomènes mis en évidence restent les mêmes en dimension supérieure.

Le processus résiduel a donc un rôle primordial dans le modèle. Des auteurs tels que Welch *et al.* (1992) pensent que la partie fixe, *i.e* la régression, n'a qu'une contribution mineure et qu'il est suffisant de la choisir constante ( $Z(x)=1$ ). C'est bien sûr un avantage puisque cela simplifie la partie programmation, ainsi que le choix du plan d'échantillonnage, puisqu'il n'y a pas de contrainte quant à l'estimabilité du paramètre  $\beta$ . Mais c'est aussi l'inconvénient majeur de cette approche, car toute la prédiction repose sur la distribution *a priori* du processus résiduel, qui est fixée de façon subjective. De plus la contrainte d'interpolation est très forte et lorsque le nombre d'observations augmente,  $\hat{Y}(x)$  devient fortement irrégulier, comme on peut le voir sur ces graphes.



A gauche  $\hat{Y}$  en dimension 1 pour  $N=30$ , ( $\theta=150$ ), et à droite  $\hat{Y}$  en dimension 2 pour  $N=100$ , ( $\theta=10$ ), pour une régression constante.

### Erreur quadratique moyenne

D'après ce qui précède, nous savons déjà que l'erreur quadratique moyenne est nulle aux points du plan. Pour tout  $x \in [0,1]^d$ , nous avons le résultat suivant.

#### **Proposition 1.9**

*L'erreur quadratique moyenne du meilleur prédicteur linéaire sans biais est de la forme,  $\forall x \in [0,1]^d$*

$$MSE(x) = E[Y(x) - \hat{Y}(x)]^2 = \sigma^2 \left( 1 - {}^t r(x) R^{-1} r(x) + K(x) ({}^t Z R^{-1} Z)^{-1} {}^t K(x) \right),$$

où  $K(x) = [Z(x) - {}^t r(x) R^{-1} Z]$ .

On remarque que l'erreur quadratique moyenne dépend de  $\sigma^2$  que l'on ne connaît pas. Nous verrons comment l'estimer dans la troisième partie de ce chapitre.

**Démonstration :** Reprenons le résultat de la proposition 1.6 avec  $\hat{\tau}$  le BLUE de  $\tau$ .

$$\begin{aligned} E[Y(x) - \hat{Y}(x)]^2 &= \text{var}(\hat{\tau}) + \text{var}({}^t r(x) R^{-1} Y - Y(x)) \\ &= \text{var} \left[ K(x) ({}^t Z R^{-1} Z)^{-1} {}^t Z R^{-1} Y \right] + {}^t r(x) R^{-1} (\text{cov } Y) R^{-1} r(x) + \text{var}(Y(x)) \\ &\quad - {}^t r(x) R^{-1} \text{cov}(Y(x), Y) - \text{cov}(Y, Y(x)) R^{-1} r(x) \\ &= K(x) ({}^t Z R^{-1} Z)^{-1} {}^t Z R^{-1} (\sigma^2 R) R^{-1} Z ({}^t Z R^{-1} Z)^{-1} {}^t K(x) + {}^t r(x) R^{-1} (\sigma^2 R) R^{-1} r(x) \\ &\quad + \sigma^2 - {}^t r(x) R^{-1} (\sigma^2 r(x)) - (\sigma^2 {}^t r(x)) R^{-1} r(x) \\ &= \left( K(x) ({}^t Z R^{-1} Z)^{-1} {}^t K(x) + {}^t r(x) R^{-1} r(x) + 1 \right) \sigma^2. \quad \square \end{aligned}$$

**Notation matricielle :**  $MSE(x) = \sigma^2 \left( 1 - ({}^t Z(x), {}^t r(x)) \begin{pmatrix} 0 & {}^t Z \\ Z & R \end{pmatrix}^{-1} \begin{pmatrix} Z(x) \\ r(x) \end{pmatrix} \right)$

### Estimateur moindres carrés ordinaires

Dans la démonstration du théorème on a remplacé  $\tau$  par l'estimateur  $\hat{\tau} = K(x)\hat{\beta}$  où  $\hat{\beta}$  est l'estimateur de Gauss-Markov de  $\beta$ . Il est tout à fait possible de le remplacer par un estimateur plus simple

$$\tilde{\tau} = K(x)\tilde{\beta}$$

où par exemple  $\tilde{\beta} = ({}^tZZ)^{-1} {}^tZY$  est l'estimateur moindres carrés ordinaires de  $\beta$ . Le prédicteur de  $Y(x)$  est alors de la forme

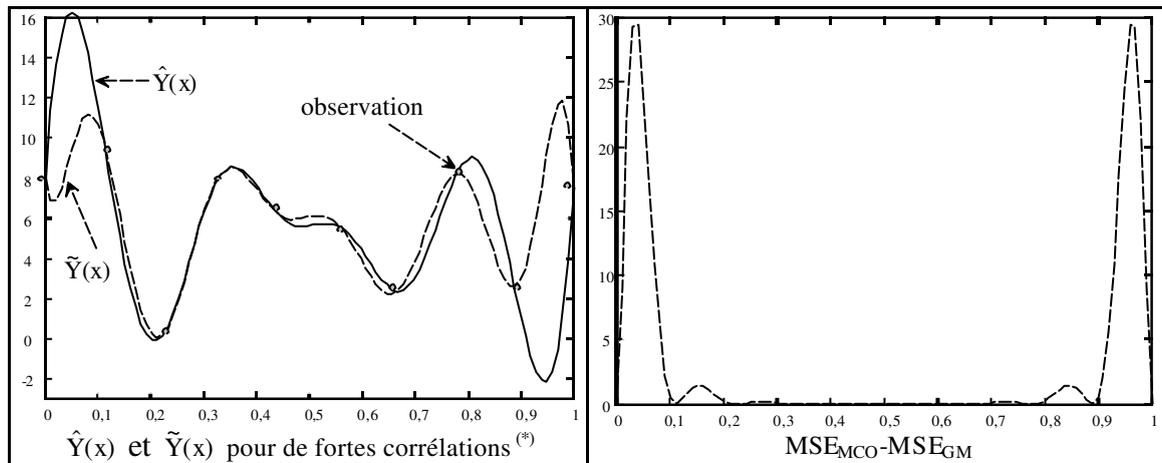
$$\tilde{Y}(x) = \tilde{\tau} + {}^t r(x)R^{-1}Y.$$

Ce n'est pas le meilleur prédicteur linéaire sans biais de  $Y(x)$  et la différence entre les deux erreurs quadratiques moyennes est égale à

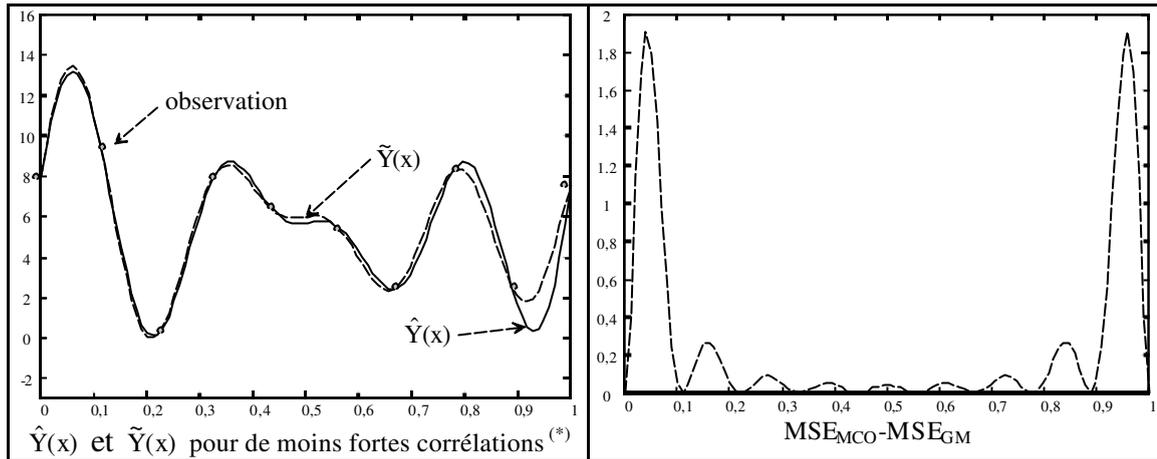
$$MSE_{MCO} - MSE_{GM} = K(x) \left[ ({}^tZZ)^{-1} - ({}^tZR^{-1}Z)^{-1} \right] {}^t K(x),$$

où  $MSE_{GM}$  (resp.  $MSE_{MCO}$ ) est l'erreur quadratique moyenne du prédicteur construit à partir de l'estimateur de Gauss-Markov (resp. moindres carrés ordinaires) de  $\beta$ .

Cependant, comme nous pouvons le constater sur les graphes ci-dessous, les deux prédicteurs ont sensiblement le même comportement, celui-ci diffère uniquement aux bornes de l'intervalle.



Graphes des prédicteurs construits à partir de l'estimateur de Gauss-Markov et moindres carrés ordinaires pour un modèle sinusoïdal,  $Z(x) = [1, \sqrt{2} \sin(2\pi x), \sqrt{2} \sin(2\pi x)]$ .



Graphes des prédicteurs construits à partir de l'estimateur de Gauss-Markov et moindres carrés ordinaires pour un modèle sinusoïdal,  $Z(x) = [1, \sqrt{2} \sin(2\pi x), \sqrt{2} \sin(2\pi x)]$ .

Ce nouveau prédicteur garde donc une précision très correcte surtout pour les faibles corrélations. De plus il semble être plus lisse que le BLUP, ce qui peut s'avérer intéressant quand il y a un grand nombre d'observations. Peut-être ce comportement vient-il du fait que seul le terme correctif du prédicteur dépend de la fonction de corrélation, ce qui lui donne plus de souplesse. Mais le principal avantage de ce prédicteur vient de la forme de  $\tilde{\beta}$ . En effet, nous verrons dans la deuxième partie qu'avec le modèle  $Z(x)$  que nous allons choisir (trigonométrique), il est possible de construire des plans tels que  ${}^tZZ=NI$ .

### 1.2.4. Krigeage

Les géostatisticiens ont une approche différente pour construire le meilleur prédicteur linéaire sans biais de  $Y(x)$ , qu'ils appellent *krigeage* (Matheron (1963)). Nous la présentons ici pour deux raisons. Tout d'abord parce que c'est principalement cette approche qui est exposée dans la bibliographie sur les expériences simulées. Ensuite parce que cette méthode est plus explicite que la précédente puisqu'elle est basée sur le caractère spatial des données. En effet, on cherche un prédicteur, combinaison linéaire des observations avec des poids qui dépendent de l'endroit où l'on se trouve dans le domaine, *i.e*

$$\hat{Y}(x) = {}^t c(x)Y$$

où les poids  $c(x) \in \mathbb{R}^N$  sont choisis tels que  $\hat{Y}(x)$  soit le BLUP. Autrement dit, il faut minimiser

$$MSE = E[Y(x) - \hat{Y}(x)]^2$$

sous la contrainte

$$E[{}^t c(x)Y] = E[Y(x)].$$

(\*) La force des corrélations sera fixée dans le § 1.3, mais on peut déjà signaler que plus les corrélations sont faibles et plus la matrice de corrélation est proche de la matrice identité, donc nécessairement plus  $\hat{Y}$  est proche de  $\tilde{Y}$ .

Cette dernière peut encore s'écrire sous la forme du système  ${}^tZc(x) = {}^tZ(x)$ , qui admet une solution si et seulement si les colonnes de  $Z$  sont indépendantes. On retrouve bien la condition d'estimabilité.

Puisque  $E[Y(x) - {}^t c(x)Y] = 0$ , l'erreur quadratique moyenne s'écrit

$$\begin{aligned} \text{MSE} &= \text{var}(Y(x) - {}^t c(x)Y) = \text{var}(Y(x)) - {}^t c(x)(\text{cov } Y)c(x) - \text{cov}(Y(x), Y)c(x) - {}^t c(x)\text{cov}(Y, Y(x)) \\ &= \sigma^2(1 - {}^t c(x)Rc(x) - 2{}^t c(x)r(x)). \end{aligned}$$

Trouver  $c(x)$  tel que le prédicteur soit le BLUP revient alors à minimiser la fonctionnelle quadratique sur  $\mathbb{R}^N$

$$J(c) = 1/2(Rc - r | c) - (r | c)$$

sous la contrainte

$${}^tZc = {}^tZ(x).$$

Si on note  $\varphi$  l'application sur  $\mathbb{R}^N$ ,  $\varphi(c) = {}^tZc - {}^tZ(x)$ , alors d'après le théorème sur les extrema liés (Ciarlet (1990)), ce problème admet comme solution, la solution  $(c, \lambda)$  du système suivant

$$\begin{pmatrix} 0 & {}^tX \\ X & R \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} {}^tX(x) \\ r(x) \end{pmatrix}$$

où  $\lambda = \lambda(x)$  est un multiplicateur de Lagrange.

L'inverse de la matrice par blocs ci-dessus est de la forme

$$\begin{pmatrix} 0 & {}^tZ \\ Z & R \end{pmatrix}^{-1} = \begin{pmatrix} -M & M{}^tZR^{-1} \\ R^{-1}ZM & R^{-1} - (R^{-1}ZM{}^tZR^{-1}) \end{pmatrix}$$

où  $M = ({}^tZR^{-1}Z)^{-1}$ , d'où  $c(x) = R^{-1}ZM{}^tZ(x) + [R^{-1} - R^{-1}ZM{}^tZR^{-1}]r(x)$ .

Finalement, on obtient la même expression que précédemment pour le meilleur prédicteur linéaire sans biais de  $Y(x)$ ,

$$\hat{Y}(x) = Z(x)\hat{\beta} + {}^t r(x)R^{-1}[Y - Z\hat{\beta}].$$

Nous avons préféré la première approche au krigeage pour construire le prédicteur, parce qu'elle permet plus de flexibilité. Par exemple, nous avons vu qu'il est possible de remplacer l'estimateur de Gauss-Markov par un autre estimateur linéaire sans biais. De plus elle permet de bien mettre en évidence les conditions sur la matrice du modèle pour que le paramètre  $\beta$  soit estimable. Bien souvent, cette condition n'est pas signalée dans les travaux qui utilisent une méthode de krigeage.

## 1.3. La covariance du processus résiduel

Le but de ce chapitre n'est pas de faire une analyse détaillée de la covariance du processus résiduel avec tous les problèmes que cela engendre : choix de la fonction de corrélation, estimation des paramètres, ... Comme cela est dit dans la partie concernant l'interpolation (§ 1.2.3), le résidu joue un rôle prépondérant dans l'ajustement, c'est pourquoi l'étude de sa covariance a donné lieu à une bibliographie très abondante. Il est donc préférable de consacrer cette troisième partie à la présentation de quelques résultats classiques, complétée par quelques références en fin de paragraphe pour de plus amples détails.

Dans tout ce qui précède nous avons supposé que la covariance du processus ( $\sigma^2$  et  $R$ ) était connue, ce n'est pas le cas bien évidemment. Nous allons donc préciser cette covariance mais pour cela nous sommes obligés d'émettre quelques hypothèses quant à la forme de la fonction de corrélation  $R$ .

### 1.3.1. Fonction de corrélation

La principale contrainte imposée à la fonction de corrélation est classique. Nous allons supposer que le processus résiduel est stationnaire,

$$R(x,u)=R(x-u),$$

et qu'il est fonction d'un vecteur  $\theta$  appelé **paramètre de corrélation**. Parmi cette classe de processus, on distingue plusieurs types de modèles pour la corrélation, des fonctions linéaires en  $\theta$ , des fonctions isotropiques (dépendent de la distance entre les points,  $R(x,u)=R(\|x-u\|)$ ), ou bien des fonctions particulières telles que les fonctions cubiques, exponentielles, Matèrn, .... L'ouvrage de Christensen (1990) et les articles de Sacks *et al.* (1989b) et de Koehler et Owen (1996) dressent une liste complète de toutes ces fonctions avec leurs avantages et inconvénients.

#### Forme de la fonction de corrélation

De façon classique, dans le cadre de ce travail, le processus est choisi gaussien avec

$$R(x,u) = \exp\left(\sum_{j=1}^d -\theta_j |x_j - u_j|^\alpha\right)$$

où  $0 < \alpha \leq 2$ .  $\alpha$  peut aussi varier avec  $j$  mais dans la plupart des cas il est égal à 2 de façon à avoir un processus ayant de bonnes propriétés. De même les  $\theta_j$  sont souvent choisis tous égaux et la fonction de corrélation s'écrit alors

$$R(x,u) = \exp\left(-\theta \sum_{j=1}^d |x_j - u_j|^2\right) = \exp\left(-\theta \|x - u\|^2\right).$$

Utiliser ce type de fonction de corrélation revient à supposer que si deux points du domaine d'échantillonnage sont proches, alors nécessairement les réponses du simulateurs en ces deux points sont fortement corrélées et *a contrario*, si deux points sont très éloignés les réponses sont considérées comme quasi indépendantes. Le degré de corrélation que l'on souhaite

introduire dans le modèle est fixé par  $\theta$ . Nous parlerons de **fortes corrélations** pour les petites valeurs de  $\theta$  et de **faibles corrélations** pour les grandes valeurs de  $\theta$ .

**Remarque :** Dans le cas où les corrélations sont fortes, *i.e*  $\theta$  petit, par développement limité on obtient pour modèle de covariance le cas linéaire simple

$$\text{cov}(Y(x), Y(u)) = \sigma^2(1 - \theta \|x - u\|^2).$$

On est donc ramené à un modèle linéaire à deux composantes de la variance.

**Remarques sur le paramètre de corrélation  $\theta$**

En choisissant les  $\theta_j$  tous égaux, on perd de l'information. En effet, il n'est alors plus possible de mettre en évidence l'importance de certains facteurs au travers la corrélation (avec des coefficients  $\theta_j$  plus petits pour ces facteurs que pour les autres, cf. Costa *et al.* (1999)).

A propos des valeurs que peut prendre  $\theta$ , on constate deux choses.

Il est impossible de descendre au dessous d'un certain seuil pour les valeurs de  $\theta$ . En effet, on remarque que plus  $\theta$  est petit et plus la matrice de corrélation est instable numériquement. On peut voir dans le tableau ci-dessous, qu'au delà d'une valeur minimale la dégradation du conditionnement de R est très rapide. Ce problème a été pris en compte lors des tests numériques en utilisant une décomposition QR de la matrice de corrélation.

$\theta$	0,1	0,5	1	5	10	100	1000
$\text{cond}(R)^{(*)}$	$6 \times 10^{17}$	$3 \times 10^{17}$	$3 \times 10^{14}$	$1,5 \times 10^8$	$3,6 \times 10^5$	3,4	1

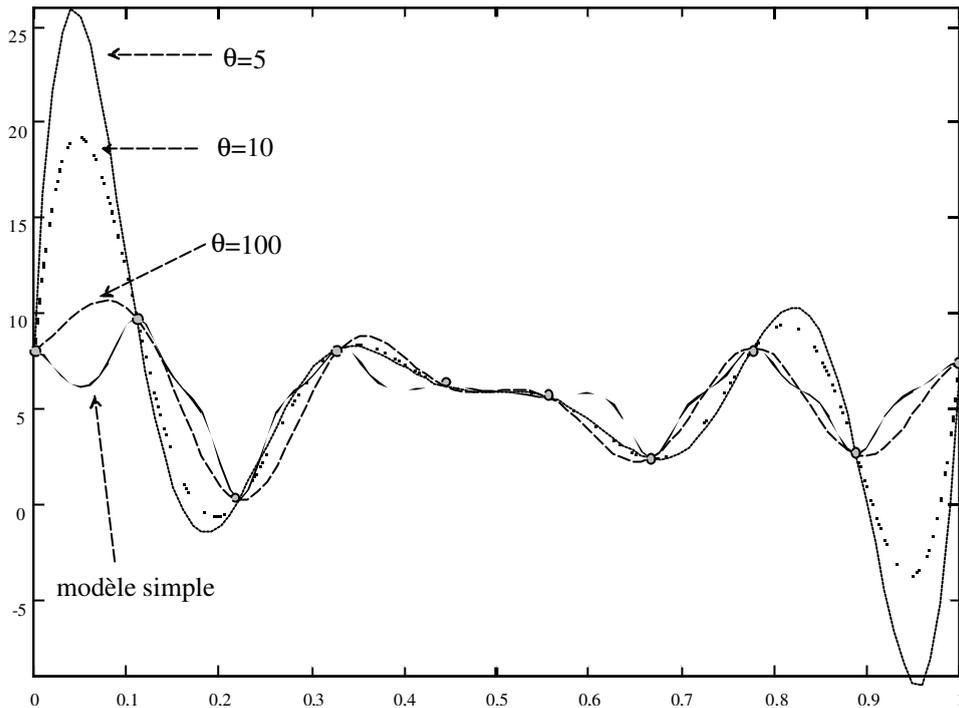
Exemple du conditionnement d'une matrice de corrélation en fonction de  $\theta$

*A contrario*, pour les grandes valeurs de  $\theta$  la matrice R est bien conditionnée, et à partir d'un seuil supérieur, elle devient proche de la matrice identité. Le modèle se comporte alors comme le modèle linéaire simple traditionnellement utilisé en plans d'expériences :

$$Y = Z\beta + e, \text{ avec } E(Y) = Z\beta \text{ et } \text{cov}Y = \sigma^2 I.$$

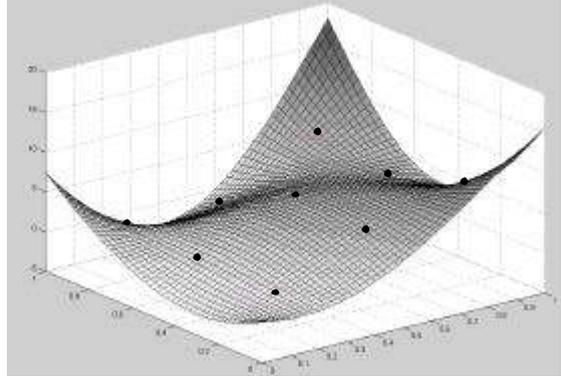
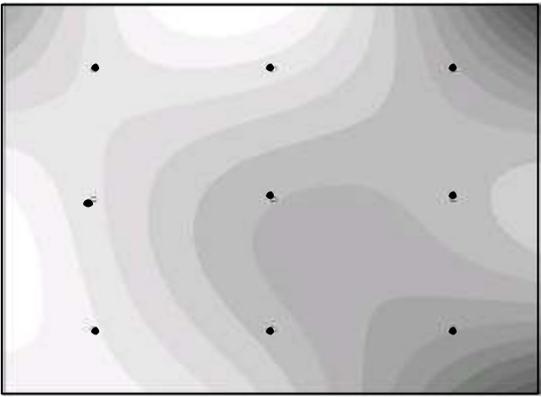
---

(\*) Le conditionnement du tableau correspond au conditionnement simple proposé par MATLAB, *i.e* le rapport de la plus grande valeur propre sur la plus petite. Plus le conditionnement s'éloigne de 1, et moins la matrice est régulière.

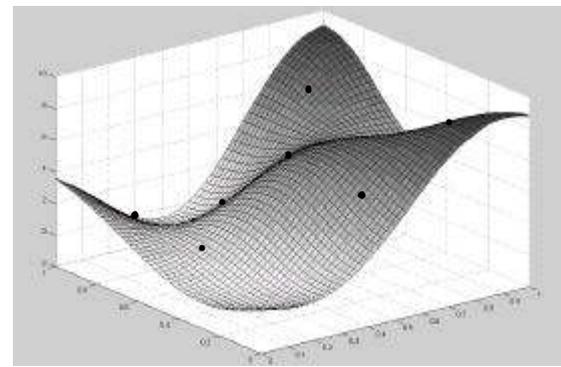
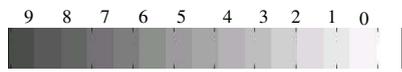
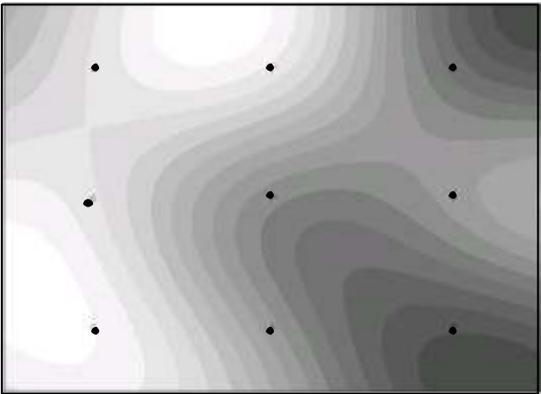


Graphes des prédicteurs  $\hat{Y}(x)$  pour différents coefficients de corrélations  $\theta$  ( $Z(x)=1$ ) et pour le vecteur des observations  $Y=[8; 9,8; 0,2; 8,2; 6,2; 5,6; 2,4; 8,2; 2,6; 7,5]$ .

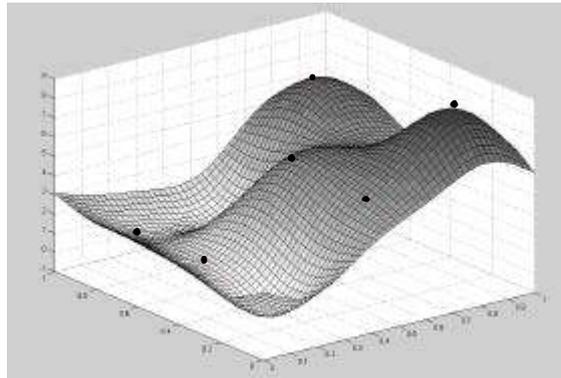
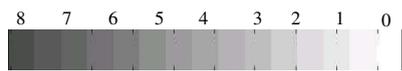
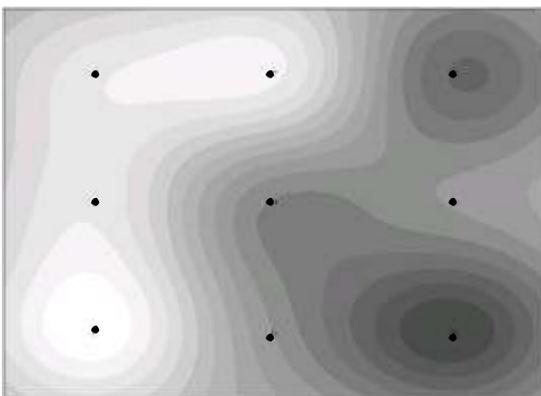
Regardons maintenant l'évolution du prédicteur  $\hat{Y}(x)$  lorsque que l'on passe de fortes corrélations à de faibles corrélations, en dimension 2. Soient un plan de 9 points également espacés dans  $[0,1]^2$  et le vecteur des observations  $Y$ , choisi au hasard entre 0 et 10. On suppose la régression constante ( $Z(x)=1$ ). La colonne de droite ci-dessous correspond à la surface de réponse du prédicteur construit à partir de ces éléments et calculé pour différentes valeurs de  $\theta$ . La colonne de gauche représente cette surface projetée sur  $[0,1]^2$ .



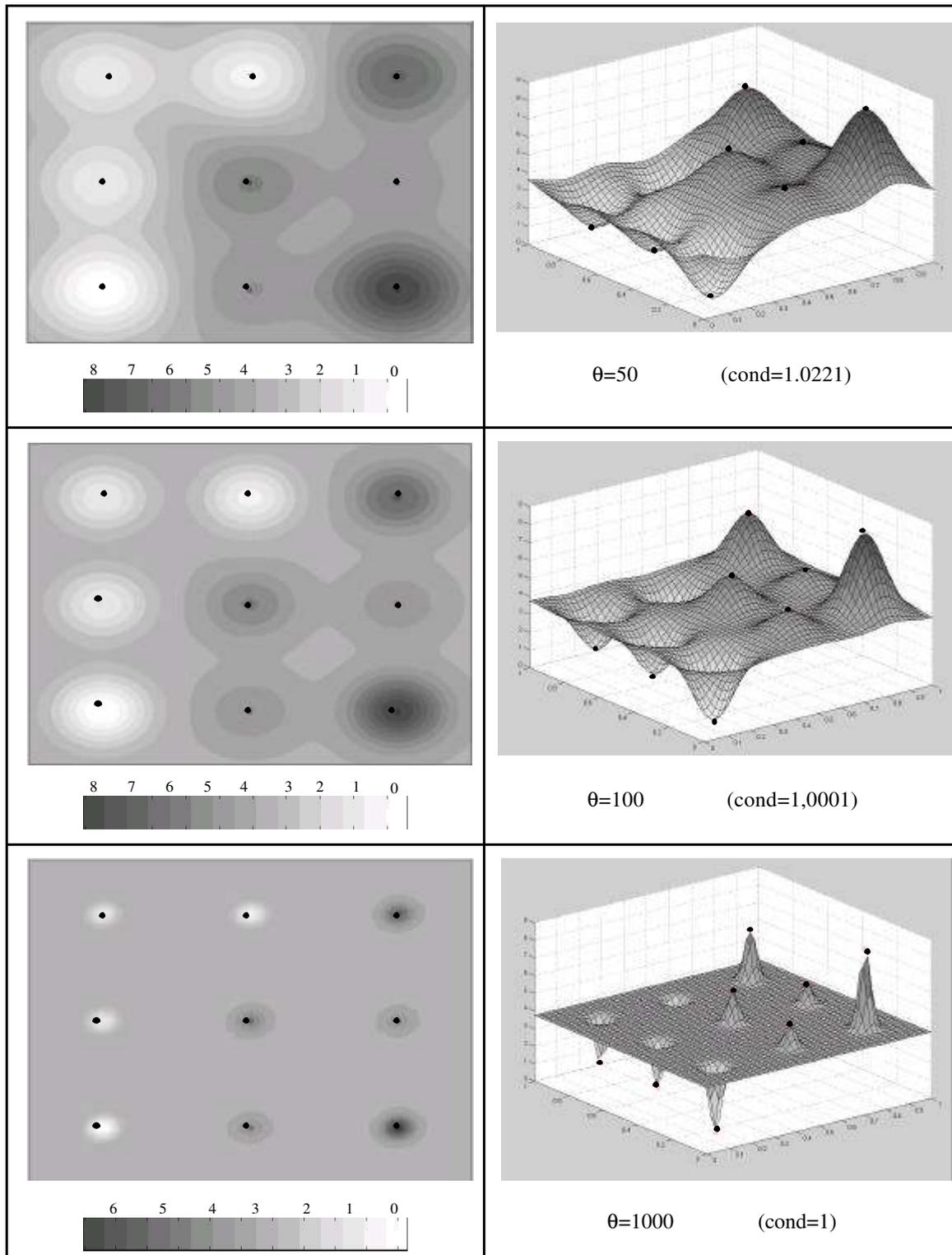
$\theta=1$  (cond= $3 \times 10^4$ )



$\theta=5$  (cond=60)



$\theta=20$  (cond=1,8)



Evolution du prédicteur en fonction de  $\theta$  pour  $Z(x)=1$  et  $N=9$ .

La première remarque que l'on peut faire à la suite de ces essais est que moins on introduit de corrélation dans le modèle et plus le prédicteur varie brusquement au voisinage des observations alors qu'il est très stable partout ailleurs. L'intérêt d'introduire des corrélations est donc évident, reste à savoir de quelle importance. En effet, rien ne dit que le

prédicteur obtenu pour  $\theta=1$  est meilleur que celui calculé pour  $\theta=20$  ou *vice-versa*, tout dépend du phénomène à prédire.

La deuxième remarque est que le prédicteur pour une régression constante est sensible aux variations de  $\theta$ . Dans le chapitre 4 nous définissons un modèle trigonométrique et nous montrons à partir d'illustrations numériques que plus cette régression est riche, plus le modèle est robuste aux variations de  $\theta$ . Ceci paraît être une propriété importante de la partie déterministe du modèle car, comme nous allons le voir dans le paragraphe suivant, l'estimation du paramètre  $\theta$  n'est pas toujours très facile.

### 1.3.2. Estimation des paramètres de la covariance

Puisque le processus résiduel est supposé gaussien, nous allons estimer les paramètres par maximum de vraisemblance. La fonction de vraisemblance est égale à

$$L(\beta, R) = (2\pi)^{-N/2} \sigma^{-N} (\det R)^{1/2} \exp \left[ \frac{-1}{2\sigma^2} (Y - Z\beta) R^{-1} (Y - Z\beta) \right],$$

en passant par la fonction  $\ln$  on obtient

$$\ln(L(\beta, R)) = l(\beta, R) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \ln(\det R) - \frac{1}{2\sigma^2} (Y - Z\beta) R^{-1} (Y - Z\beta).$$

#### Estimation de $\sigma^2$

Il est possible d'estimer conjointement  $\beta$  et  $\sigma^2$  par maximum de vraisemblance. Si on dérive par rapport à  $\beta$  on a

$$\frac{\partial l(\beta, R)}{\partial \beta} = -\frac{1}{\sigma^2} ({}^t Z R^{-1} Y - {}^t Z R^{-1} Z \beta).$$

L'estimateur de  $\beta$  par maximum de vraisemblance est donc égal à  $\hat{\beta}$  l'estimateur de Gauss-Markov.

Si maintenant on dérive par rapport à  $\sigma^2$ ,

$$\frac{\partial l(\beta, R)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} (Y - Z\beta) R^{-1} (Y - Z\beta),$$

on obtient l'estimateur de  $\sigma^2$  par maximum de vraisemblance

$$\hat{\sigma}^2 = \frac{1}{N} (Y - Z\hat{\beta}) R^{-1} (Y - Z\hat{\beta})$$

#### Estimation du paramètre de corrélation $\theta$

De même que précédemment on peut estimer le paramètre de corrélation par maximum de vraisemblance. En dérivant par rapport à  $\theta_i$ , on obtient

$$\frac{\partial l(\beta, R)}{\partial \theta_i} = -\frac{1}{2} \text{tr} \left( R^{-1} \frac{\partial R}{\partial \theta_i} \right) + \frac{1}{2\sigma^2} (Y - Z\beta) R^{-1} \frac{\partial R}{\partial \theta_i} R^{-1} (Y - Z\beta)$$

Malheureusement cette dérivée ne permet pas en général d'avoir une solution analytique pour l'estimateur du maximum de vraisemblance de  $\theta$ . On a alors recours à des outils numériques

pour optimiser la fonction de vraisemblance (Marshall et Mardia (1984), Sacks *et al.* (1992)). Ces techniques sont souvent très coûteuses et engendrent parfois quelques problèmes de maximum local (Warnes et Ripley (1987)). Toutefois Costa *et al.* constatent après simulations qu'un maximum local est suffisant comme estimateur de  $\theta$ . Sacks, Schiller et Welch (1989) proposent une autre méthode que l'estimation pour choisir le paramètre de corrélation (étude de la robustesse de  $\theta$ ) que nous allons présenter dans le paragraphe suivant. Celle-ci ne pose pas de problème de maximum local mais paraît plus coûteuse. La détermination de ce paramètre reste à l'heure actuelle une question ouverte, et la difficulté majeure avec ce type de modèle.

Liste de références bibliographiques traitant différents problèmes quant à la fonction de corrélation :

- Présentation de différentes fonctions de corrélations avec leurs avantages et inconvénients : Sacks, Welch, Mitchell et Wynn. (1989+discussions), Koehler et Owen (1996), Christensen (1990).
- Propriétés asymptotiques de l'estimateur du maximum de vraisemblance des paramètres de la covariance : Ying (1993).
- Algorithmes pour maximiser la fonction de vraisemblance : Mardia et Marshall (1984) et Welch *et al.*(1992).
- Comment mettre en évidence au travers la covariance, l'importance de certains facteurs : Costa, Pronzato et Thierry. (1999).
- Difficulté qu'il y a à déterminer le maximum de vraisemblance numériquement : Warnes et Ripley (1987).
- Alternative au maximum de vraisemblance pour choisir les paramètres de la fonction de corrélation : Sacks, Schiller et Welch (1989).

## 1.4. Les plans pour une approche par résidu aléatoire

Le prédicteur dépend fortement des points de simulation (d'expériences), au travers les matrices  $X$ ,  $R$  et  $r$ , mais surtout  $Y_D$ , le vecteur des observations. Sélectionner un bon plan d'échantillonnage est donc la clef pour construire un prédicteur efficace. Pour cela il faut définir comment mesurer la qualité d'un plan.

Supposons la taille du plan  $N$  et la fonction de corrélation fixées. Dans cette quatrième partie, nous allons présenter trois critères utilisés en expériences simulées pour sélectionner des plans de bonne qualité : l'erreur quadratique moyenne intégrée, l'entropie et la distance maximin.

On note  $D$  l'ensemble des plans d'échantillonnage de taille  $N$ , adaptés à l'estimation du vecteur des paramètres du modèle  $\beta$  (cf. § 1.2.1).

### 1.4.1. L'erreur quadratique moyenne intégrée

Le modèle présenté dans ce chapitre est sans erreur de mesure, on peut donc espérer trouver un plan qui minimise la différence entre le réponse du simulateur et sa valeur prédite. Pour cela Sacks et Schiller (1988) et Sacks, Schiller et Welch (1989) proposent d'utiliser l'erreur quadratique moyenne intégrée.

#### Définition 1.10

$D^* \in D$  est un plan de IMSE (Integrated Mean Squared Error) minimale si

$$J(D^*) = \min\{J(D), D \in D\},$$

où

$$J(D) = \frac{1}{\sigma^2} \int_{[0,1]^d} E[Y(x) - \hat{Y}(x)]^2 dx.$$

**Remarque :** Il est possible d'inclure une fonction de poids dans l'intégrale de l'erreur quadratique moyenne.

La IMSE peut s'écrire sous la forme matricielle

$$\sigma^2 \left\{ 1 - \text{trace} \left\{ \begin{pmatrix} 0 & {}^tZ \\ Z & R \end{pmatrix}^{-1} \int \begin{pmatrix} {}^tZ(x)Z(x) & {}^tZ(x)r(x) \\ r(x)Z(x) & {}^tr(x)r(x) \end{pmatrix} dx \right\} \right\}.$$

Sacks, Schiller et Welch (1989) remarquent que si  $Z(x)$  et  $r(x)$  se décomposent en produit de fonctions à une seule variable, alors l'intégrale sur  $[0,1]^d$  se simplifie en produit d'intégrales à une dimension. Cette remarque est intéressante car minimiser la IMSE revient à optimiser numériquement une fonction à  $N \times d$  composantes, ce qui coûte très cher. Par exemple, Sacks, Welch, Mitchell et Wynn (1989) utilisent une technique d'optimisation quasi-Newton sur un ordinateur Cray X-MP/48. Ils mettent 11 minutes à minimiser la IMSE sur les plans à  $d=6$  facteurs de taille  $N=16$ . De plus, comme cela arrive souvent, ces méthodes numériques d'optimisation aboutissent à un minimum local.

On remarque aussi que ce critère dépend du modèle choisi. Autrement dit, un plan n'est optimal que pour un modèle donné et si l'on souhaite en changer, il est nécessaire de refaire les calculs pour trouver un nouveau plan. Bien entendu cela coûte cher, mais le principal inconvénient de ce critère vient du fait que le modèle dépend du paramètre de corrélation  $\theta$ . Or celui-ci n'est pas connu lors de la planification puisque généralement il est estimé par maximum de vraisemblance, c'est-à-dire à partir d'observations. Sacks, Schiller et Welch (1989) proposent une étude de robustesse pour pallier ce problème. Pour cela ils se fixent un ensemble  $\Theta$  de valeurs de  $\theta$  et calculent le coefficient d'efficacité du plan de IMSE minimale pour un  $\theta$  fixé par rapport aux plans de IMSE minimale calculée pour les autres valeurs  $\theta_i$  de  $\Theta$ ,

$$J_{\theta}(D_{\theta}^*) / J_{\theta}(D_{\theta_i}^*),$$

où  $J_{\theta}$  représente la IMSE calculée pour  $\theta$ . Cette technique coûte cher puisqu'elle nécessite l'optimisation de la IMSE pour différentes valeurs de  $\theta$ , mais elle permet non seulement de choisir le plan le plus robuste aux variations du paramètre de corrélation, mais aussi de fixer ce paramètre sans avoir à l'estimer par maximum de vraisemblance.

### 1.4.2. L'entropie

Dans son travail sur les plans bayésiens, Lindley (1956) propose un critère basé sur l'entropie d'un vecteur aléatoire définie par

$$\text{Ent}(Z) = -E_Z(\ln f(Z)) = -\int f(z) \ln f(z) dz,$$

où  $f$  est la densité de probabilité de  $Z$ . Celle-ci permet de mesurer la quantité d'information fournie par l'expérimentation.

Si l'on considère que le domaine d'échantillonnage  $E$  est discret, alors le critère d'entropie sélectionne le plan  $D$  qui minimise

$$E_{Y_D} \{ \text{Ent}(Y_{\bar{D}} | Y_D) \}$$

où  $\bar{D} = E \setminus D$ . Schwery et Wynn (1987) utilisent alors la décomposition classique de l'entropie

$$\text{Ent}(Y) = \text{Ent}(Y_D) + E_{Y_D} \{ \text{Ent}(Y_{\bar{D}} | Y_D) \}$$

pour montrer que ce critère est équivalent à maximiser l'entropie de  $Y_D$ , puisque l'entropie totale du système,  $\text{Ent}(Y)$ , est fixée.

#### Définition 1.11

$D^* \in \mathbf{D}$  est un plan d'entropie maximale si

$$\text{Ent}(Y_{D^*}) = \max \{ \text{Ent}(Y_D), D \in \mathbf{D} \}$$

Dans le cas particulier où  $Z$  est un vecteur aléatoire gaussien, l'entropie de  $Z$  est alors égale à  $\ln(\det(\text{cov}(Z)))$  aux constantes près, donc maximiser l'entropie de  $Y_D$  est équivalent à maximiser le déterminant de la matrice de corrélation.

### Proposition 1.12

Dans le cas gaussien,  $D^* \in \mathcal{D}$  est d'entropie maximale si et seulement si

$$\det(R_{D^*}) = \max \{ \det(R_D), D \in \mathcal{D} \},$$

où  $R_D$  est la matrice de corrélation aux points du plan  $D$ .

Tout comme pour le critère IMSE, sélectionner un plan par le critère d'entropie coûte très cher puisqu'il faut optimiser une fonction à  $N \times d$  composantes. Currin *et al.* (1991) proposent un algorithme adapté de celui de Mitchell (1974) DETMAX qui successivement supprime et ajoute des points pour améliorer le plan (DETMAX est un programme qui sert à la recherche empirique de plans d'expériences  $D$ -optimaux). Ici aussi, un plan n'est optimal que pour un modèle donné puisque le critère d'entropie dépend de la fonction de corrélation et de son paramètre  $\theta$ .

### 1.4.3. La distance maximin

Nous avons remarqué que les deux critères précédents dépendent du modèle au travers la fonction de corrélation. Or celle-ci est souvent choisie stationnaire de la forme

$$R(x,u) = g(d(x,y))$$

où  $g$  est une fonction sur  $\mathbb{R}$  et  $d(\cdot, \cdot)$  définit une distance sur  $[0,1]^d$ . Donc seuls les points les plus proches deux à deux contribuent de façon significative au déterminant de la matrice de corrélation. L'idée est donc de trouver un plan tel que la distance minimale entre deux points soit la plus grande possible. C'est ce que l'on appelle la distance maximin.

### Définition 1.13

$D^* \in \mathcal{D}$  est un plan de distance maximin si

$$\min_{x,y \in D^*} d(x,y) = \max \{ \min_{x,y \in D} d(x,y), D \in \mathcal{D} \}.$$

Ce critère a été proposé par Johnson *et al.* (1990) qui définissent aussi un critère minimax. Ils ont démontré que si la fonction  $g$  est décroissante par rapport à  $d(\cdot, \cdot)$  (par exemple, une corrélation gaussienne) alors un plan d'entropie maximale est, dans certain sens, un cas limite des plans de distance maximin (*i.e* pour  $R^k$  quand  $k$  tend vers l'infini).

Ce critère dépend encore du modèle puisque la distance doit être choisie en fonction de la forme de la fonction de corrélation. Cependant, un plan de distance maximin est maintenant indépendant du paramètre de corrélation  $\theta$ , qui rappelons-le, n'est pas connu *a priori*.

## Bilan de l'approche par résidu aléatoire

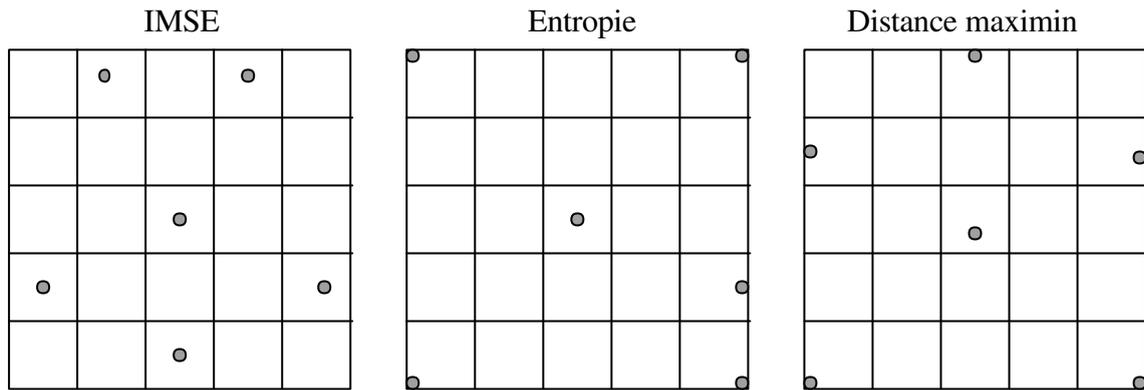
Cette approche est particulièrement bien adaptée au problème des simulateurs déterministes. En effet, rappelons que dans ce cas les simulations sont sans erreur de mesure, il paraît donc logique que le prédicteur passe par les observations. Cependant nous avons vu que cette particularité du modèle est intéressante uniquement dans le cas où le nombre d'observations n'est pas trop important, sinon on obtient vite un prédicteur très irrégulier.

L'idée d'introduire dans le modèle un processus permettant de tenir compte du caractère spatial des données est intéressante. Toutefois, nous avons vu - et certains auteurs ont déjà remarqué - que le processus joue un rôle prépondérant dans le modèle, et que la partie déterministe n'a qu'une contribution mineure. Ce qui pose quelques problèmes. En effet, est-il raisonnable de baser toute la prédiction sur la distribution que l'on impose au processus *a priori*? De plus, nous avons vu que le prédicteur est très peu robuste aux variations du paramètre de corrélation. Or nous savons qu'en pratique ce dernier n'est pas facile à déterminer.

En conclusion, l'idée d'utiliser un modèle en deux parties est intéressante si toutefois on arrive à équilibrer le rôle joué par chacune d'elles, trouver un bon compromis.

En ce qui concerne les plans, regardons point par point s'ils répondent aux principaux objectifs fixés en introduction.

1. Ces plans ne permettent pas d'analyser différents modèles avec cependant une nuance suivant les critères. La IMSE est très rigide puisque les plans dépendent à la fois de la partie déterministe du modèle et de la fonction de corrélation. L'entropie ne dépend plus que de la matrice de corrélation mais suppose que le paramètre  $\theta$  est fixé. Il semble donc que le critère le plus souple soit celui de distance maximin.
2. En ce qui concerne l'efficacité de ces plans, l'ordre de préférence des critères est inversé. En effet, pour la IMSE, on minimise directement la différence entre la réponse du simulateur et la réponse prédite, donc l'erreur est minimale. Le critère d'entropie étant considéré comme un cas limite du critère de distance maximin, il semble plus efficace que ce dernier.
3. Aucun des trois critères ne permet d'avoir des plans qui "remplissent" uniformément le domaine d'échantillonnage. L'entropie et la distance maximin ont tendance à privilégier les bords du domaine et à délaisser l'intérieur. Les points obtenus par IMSE semblent mieux répartis mais on peut remarquer que projetés sur des marges de dimension inférieure, la répartition est moins bonne puisqu'il y a des répétitions. Ceci peut poser des problèmes si la réponse du simulateur à prédire dépend en fait que d'un sous-ensemble des variables d'entrée.



Exemple issu de l'article de Koehler et Owen (1999) de plans de taille 6 sélectionnés par les 3 critères

4. Comme nous l'avons signalé ces critères coûtent très chers puisque les trois nécessitent l'optimisation numérique d'une fonction à  $N \times d$  variables. Certains auteurs ont envisagé diminuer le temps de calcul en limitant la recherche à certaines classes de plans que nous allons définir dans le chapitre suivant. Nous pouvons citer Bates, Buck, Riccomagno et Wynn (1996) qui définissent leurs plans dans la catégorie des réseaux de rang 1, *i.e* construits à partir des suites de faible discrédance (cf. 2.3). Park (1994) propose un algorithme pour trouver le plan vérifiant l'un des trois critères parmi des hypercubes latins (cf. 2.2.3). Tang (1994) construit des tableaux orthogonaux basés sur des hypercubes latins de distance maximin (cf. 2.2.4).

## Bibliographie

- Bates R.A., Buck R.J., Riccomagno E., Wynn H.P. (1996). Experimental design and observation for large systems. *J. R. Statist. Soc. B* **58**, 77-94.
- Christensen R. (1987). *Plane answers to complex questions : The theory of linear models*. Springer-Verlag.
- Christensen R. (1990). *Linear models for multivariate, time series, and spatial data*. Springer-Verlag.
- Ciarlet J.P. (1990). *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson.
- Costa J.P., Pronzato L., Thierry E. (1999). Nonlinear prediction by kriging, with application to noise cancellation. Soumis à *Signal Processing*.
- Currin C.T., Mitchell M., Morris M., Ylvisaker D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86**, 953-963.
- Johnson M.E., Moore L.M., Ylvisaker D. (1990). Minimax and maximin distance designs. *J. of Statist. Planning and Inference* **26**, 131-148.
- Koehler J.R., Owen A.B. (1996). Computer experiments. In *Handbook of statistics, 13 : Design and analysis of experiments* (Ghosh, Rao Eds.), North-Holland, 261-308.
- Lindley D.V. (1956). On a measure of information provided by an experiment. *Ann. of Math. Statist.* **27**, 986-1005.
- Mardia K.V., Marshall R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Matèrn B. (1947). Method of estimating the accuracy of line and sample plot surveys. *Medd. Skogsforskn Inst.* **36** (1).
- Matheron G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246-1266.
- Mitchell T. J. (1974). An algorithm for construction of 'D-optimal' experimental designs. *Technometrics* **16**, 203-210.
- Park J.S. (1994). Optimal Latin hypercube designs for computer experiments. *J. of Statist. Planning and Inference* **39**, 95-111.
- Sacks J., Schiller S. (1988). Spatial designs. In *Statistical Decision Theory and Related Topics IV : vol 2* (Gupta, Berger Eds.), Springer Verlag, New York, 385-399.
- Sacks J., Schiller S.B., Welch W.J. (1989). Designs for computer experiments. *Technometrics* **31**, 41-47.
- Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). Design and analysis of computer experiments. *Statist. Science* **4**, 409-435.
- Schwery M.C., Wynn H.P. (1988). Maximum entropy sampling with application to simulation code. In *Proc. 12<sup>th</sup> world congress on scientific computation, IMAC88* **2**, 517-519.
- Schwery M.C., Wynn H.P. (1987). Maximum entropy sampling. *J. of Appl. Statist.* **14**, 165-170.
- Tang B. (1994). A theorem for selecting OA-based Latin hypercubes using a distance criterion. *Comm. Statist. - Theory Meth.* **23**, 2047-2058.
- Warnes J.J., Ripley B.D. (1987). Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika* **74**, 640-642.

Welch W.J., Buck R.J, Sacks J., Wynn H.P., Mitchell T.J, Morris M.D. (1992). Screening, predicting and computer experiments. *Technometrics* **34**, 15-25.

Ying Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. of Statist.* **21**, 1567-1590.



## 2. Approche par échantillon aléatoire

La deuxième approche envisagée dans la bibliographie introduit l'aléatoire non plus au travers le modèle comme précédemment, mais *via* l'échantillon. Elle est basée sur des techniques utilisées en intégration numérique et a été proposées par Owen (1992) et Koehler et Owen (1996).

La partie modélisation et estimation des paramètres du modèle étant très simple, nous nous sommes intéressés aux différents plans utilisés en intégration numérique. La bibliographie sur ce sujet est particulièrement abondante. Nous avons recensé trois grandes méthodes d'échantillonnage, qui ont toutes pour principal objectif la répartition "uniforme" des points du plan dans le domaine expérimental. Elles se différencient par la façon de définir l'uniformité de la répartition. La première impose aux points du plan de représenter uniformément les marges de dimension fixée inférieure à  $d$  (*i.e* en projection). La deuxième méthode utilise une fonction de discrédance qui mesure la différence entre la répartition empirique des points du plan et la distribution uniforme. Et la troisième utilise des structures de points qui couplent les avantages des deux méthodes précédentes.

## 2.1. Modélisation et estimation

### 2.1.1. Le modèle

Owen (1992) et Koehler et Owen (1996) proposent de modifier le modèle du chapitre précédent, en ne prenant en compte que la partie déterministe, c'est-à-dire la régression linéaire.

#### Définition 2.1

*Le modèle est défini par*

$$Y(x) = Z(x)\beta, \quad \forall x \in [0,1]^d$$

*où  $Z(x)$  est le vecteur du modèle et  $\beta$  le vecteur inconnu des paramètres du modèle.*

On suppose en fait que la réponse du simulateur peut se décomposer sur une base de fonctions, et que seuls les termes prédominants de cette décomposition suffisent à la modéliser.

On peut bien évidemment envisager toute sorte de fonctions pour  $Z(x)$ , nous avons choisi d'utiliser une régression trigonométrique (où de Fourier) que nous définirons dans la deuxième partie. Celle-ci a pour avantage de former une base orthogonale de fonctions, ce qui simplifie l'estimation des paramètres du modèle, comme nous allons le voir maintenant.

### 2.1.2. Estimation des paramètres du modèle

Il existe plusieurs techniques pour déterminer le meilleur vecteur des paramètres  $\beta$ , mais la plus naturelle est d'utiliser le critère moindres carrés

$$\tilde{\beta} = \left( \int_{[0,1]^d} {}^tZ(x)Z(x) dF \right)^{-1} \int_{[0,1]^d} {}^tZ(x)Y(x) dF,$$

où  $F$  est une distribution quelconque sur  $[0,1]^d$ . Cependant, afin de simplifier les notations nous allons la supposer uniforme sur  $[0,1]^d$ , ce qui ne change en rien les résultats présentés.

Dans le cas où la base choisie pour décomposer  $Y(x)$  est orthogonale, alors l'expression de  $\tilde{\beta}$  se réduit à

$$\tilde{\beta} = \int_{[0,1]^d} {}^tZ(x)Y(x) dx.$$

Ce qui permet d'éviter le calcul d'une inversion de matrice.

Si on note  $m$  le nombre de termes de la régression, alors  ${}^tZ(x)Y(x)$  est un vecteur de  $m$  fonctions. L'ajustement du modèle se ramène donc à l'intégration simultanée de plusieurs

fonctions sur le cube unité.  $Y$  n'étant pas connue analytiquement, estimer les paramètres du modèle revient à évaluer l'intégrale d'une fonction simulée.

### 2.1.3. Estimation d'une intégrale : méthode de Monte Carlo

Les méthodes traditionnelles pour intégrer une fonction à une, deux ou trois variables sont très satisfaisantes, en revanche lorsque le nombre de facteurs augmente, ces méthodes deviennent extrêmement coûteuses. Elles ne peuvent donc pas être utilisées dans la problématique qui nous concerne, c'est-à-dire pour intégrer une fonction simulée dont la valeurs en un point du domaine d'intégration demande un temps de calcul considérable.

Dans ce contexte, les techniques employées pour évaluer une intégrale ont pour origine la méthode de Monte Carlo. Soit  $f$  une fonction simulée sur le domaine  $[0,1]^d$ , on estime la valeur de l'intégrale de  $f$ ,

$$If = \int_{[0,1]^d} f(x) dx ,$$

par la somme

$$\hat{If} = \frac{1}{N} \sum_{i=1}^N f(x_i) ,$$

où  $x_1, \dots, x_N$  sont des points choisis au hasard et indépendamment sur  $[0,1]^d$ . C'est donc au niveau de l'échantillonnage qu'intervient l'aléatoire, d'où le nom de cette approche.

**Remarque :** La méthode de Monte Carlo n'est en fait qu'un cas particulier des méthodes numériques d'approximation d'une intégrale, utilisant des formules du type  $\sum w_i f(x_i)$  où les  $w_i$  sont des nombres positifs de somme 1.

Cette méthode présente des avantages majeurs et c'est pourquoi elle a connu un tel engouement depuis son apparition au milieu du 20<sup>ème</sup> siècle.

*Pour toute fonction continue, la convergence de l'estimateur est assurée par la loi forte des grands nombres.*

*Pour toute fonction de carré intégrable, le théorème de la limite centrale permet d'établir que la vitesse de convergence est égale à  $\sigma N^{-1/2}$ , et que la variance est égale à  $\sigma^2 N^{-1}$  où  $\sigma^2$  est la variance de  $f(x)$  où  $x$  est un vecteur aléatoire à valeurs dans  $[0,1]^d$ .*

On remarque que ces résultats sont totalement indépendants du nombre de facteurs,  $d$ , ce qui n'est pas le cas des méthodes numériques traditionnelles. En revanche, il n'est pas possible d'obtenir une majoration de l'erreur mais uniquement un intervalle de confiance.

Si on applique cette méthode pour estimer les paramètres du modèle défini précédemment, on obtient

$$\hat{\beta} = \left( \int_{[0,1]^d} {}^t Z(x) Z(x) dx \right)^{-1} \frac{1}{N} \sum_{i=1}^N Z(x_i) Y(x_i) .$$

### Proposition 2.2

*L'opérateur de covariance de l'estimateur  $\hat{\beta}$  est de la forme*

$$\frac{1}{N} \left( \int {}^tZ(x)Z(x)dx \right)^{-1} \Sigma \left( \int {}^tZ(x)Z(x)dx \right)^{-1},$$

*où  $\Sigma$  est la matrice carrée d'ordre  $m$  de la covariance de  ${}^tZ(x)Y(x)$ .*

Dans le cas où la base est orthogonale, l'opérateur de covariance est réduit à  $\text{cov}({}^tZ(x)Y(x))$ .

Différentes modifications sur le choix des points ont été apportées à la méthode de Monte Carlo afin d'améliorer son efficacité. Dans ce chapitre nous allons faire un rapide état des lieux des trois principales méthodes que nous avons trouvées dans la bibliographie. La première technique dérivée de la méthode de Monte Carlo consiste à réduire la variance  $\sigma^2$ . On utilise pour cela des plans d'échantillonnage possédant une bonne répartition marginale de leurs points, notamment des hypercubes latins et tableaux orthogonaux, que l'on randomise. On suppose en fait que si le plan a une répartition uniforme de ses points sur les marges, alors il a des chances de bien représenter tout le domaine. La deuxième méthode a pour origine la constatation suivante. Les points d'échantillonnage  $x_1, \dots, x_N$  doivent former une suite de nombres aléatoires, or, en pratique, il est impossible de disposer d'une telle suite car, en informatique, le hasard est généré par des algorithmes déterministes. On parle alors de suites pseudo-aléatoires. Puisque les points ne sont pas réellement choisis au hasard, autant leur imposer certaines contraintes de façon à obtenir "mieux" qu'une suite aléatoire, c'est-à-dire obtenir une meilleure vitesse de convergence que  $N^{1/2}$ . Ces méthodes sont connues sous le nom de quasi-Monte Carlo et parmi celles-ci, nous étudierons en particulier la méthode des réseaux (lattice). Enfin, la troisième méthode est une méthode hybride des deux autres. Elle permet de réduire la variance en utilisant des structures de points ayant des propriétés de répartition marginale, tout en diminuant la vitesse de convergence grâce aux suites pseudo-aléatoires.

## 2.2. Méthodes de réduction de la variance

La vitesse de convergence de l'estimateur de l'intégrale ainsi que sa variance dépendent de  $\sigma^2$ , la variance de  $f(x)$  où  $x$  est un vecteur aléatoire de distribution uniforme sur  $[0,1]^d$ . Les méthodes que nous allons décrire maintenant ont pour but de réduire cette variance. L'article de référence sur ces techniques est celui de Mc Kay *et al.* (1979), mais on peut aussi citer Owen (1992/94) qui publie beaucoup sur le sujet, ainsi que Stein (1987), Loh (1993/96) et Tang (1993/94).

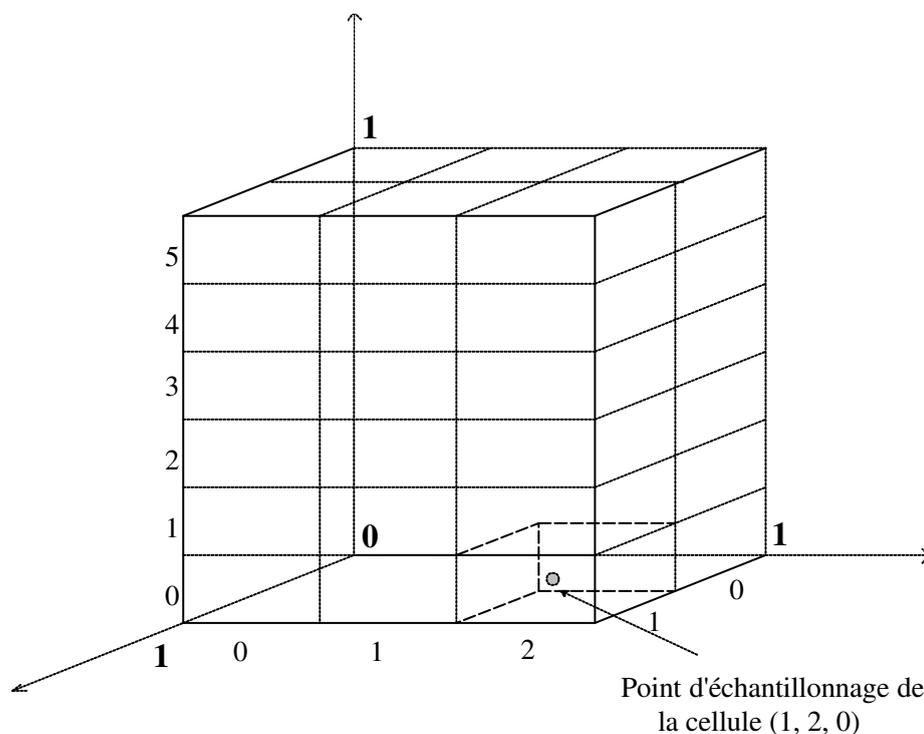
Tous les auteurs que nous venons de citer ont la même approche pour réduire  $\sigma^2$ . Ils utilisent une technique d'échantillonnage à plusieurs degrés de façon à obtenir des points ayant une bonne répartition lorsqu'on les projette en dimension inférieure à  $d$ .

### 2.2.1. Echantillonnage à deux degrés

Soit  $I=\{1,\dots,d\}$ , l'ensemble des index des facteurs. On découpe chaque arête du cube unité en un nombre fini de segments de même longueur que l'on numérote de 0 à  $q_i-1$  où  $i \in I$  désigne la  $i$ ème arête. On munit chaque ensemble  $\{0,\dots,q_i-1\}$  de la loi d'addition modulo  $q_i$ . On obtient ainsi une partition du cube unité en

$$\prod_{i \in I} q_i$$

cellules, indexées par les éléments du groupe abélien fini  $G = \times_{i \in I} (\mathbb{Z}/q_i)$ .



### Définition 2.3

*Un échantillonnage à deux degrés consiste à*

*1<sup>er</sup> degré : choisir les cellules du cube unité représentées par le plan*

*2<sup>ème</sup> degré : tirer au hasard un point dans chaque cellule sélectionnée au 1<sup>er</sup> degré.*

*Un point d'échantillonnage  $X$  est alors la réalisation d'un vecteur aléatoire de la forme*

$$X_j = \frac{c_j + \varepsilon_j}{q_j}, j=1, \dots, d$$

*où  $c=(c_1, \dots, c_d) \in G$  représente la cellule dans laquelle se trouve le point  $X$ , et  $\varepsilon=(\varepsilon_1, \dots, \varepsilon_d)$  est un vecteur aléatoire de distribution uniforme sur  $[0,1]^d$ , qui désigne à quel endroit le point se trouve dans la cellule.*

**Remarque :** Il est aussi possible de représenter la cellule par son centre, c'est-à-dire choisir  $c+1/2$  à la place de  $c$ .

On peut noter que la même cellule peut être représentée plusieurs fois dans le plan puisque le second degré de l'échantillonnage engendre des points différents.

Dans cette technique d'échantillonnage seul le premier degré est contrôlable. Le plan est donc choisi par les cellules qu'il représente. Comme nous l'avons signalé en introduction, ce choix se fait par rapport à des propriétés de projection que l'on définit à partir des  $J$ -marges.

### Définition 2.4

*On note  $E_i = \{ [0, 1/q_i[, \dots, [(q_i-1)/q_i, 1] \}$ ,  $i \in I$ , l'ensemble des segments de la  $i$ ème arête de  $[0,1]^d$ . Alors pour tout  $J \subseteq I$ , on appelle  $J$ -marge le produit cartésien*

$$\times_{i \in J} E_i.$$

*Les éléments de la  $J$ -marge s'appellent aussi des cellules.*

Les différentes méthodes que nous allons présenter imposent au plan de représenter toutes les cellules de toutes les  $J$ -marges pour  $|J|$  fixé. La difficulté pour parvenir à des résultats sur la variance avec ce type d'échantillonnage réside dans le fait que les contraintes marginales que l'on impose à l'échantillon sont croisées (*i.e* sur toutes les directions simultanément). Il est alors nécessaire d'utiliser la décomposition de la variance dite croisée que nous allons présenter.

## 2.2.2. Décomposition de la variance

Soit  $D=\{1,\dots,d\}$ , on note  $dx_j$  la mesure uniforme sur la  $\{j\}$ -marge,  $dx_j$  la mesure sur la  $J$ -marge où  $J=D\setminus\{j\}$ , et de façon plus générale, pour tout  $U\subseteq D$ ,  $dx_U$  est la mesure uniforme sur la  $U$ -marge et  $dx_{-U}$  sur la  $J=D\setminus U$ .

Soit  $X$  un vecteur aléatoire à valeurs dans  $[0,1]^d$ , on note

$$\mu = \int_{[0,1]^d} f(X) dX,$$

les effets de  $f$  sur les marges à une dimension sont décrits par

$$f_j(X_j) = \int_{[0,1]^d} (f - \mu) dX_{-j}, \quad 1 \leq j \leq d,$$

les effets sur les marges à deux dimensions sont décrits par

$$f_{ji}(X_j, X_i) = \int_{[0,1]^d} (f - f_i - f_j - \mu) dX_{-\{i,j\}}, \quad 1 \leq i < j \leq d,$$

et plus généralement les effets sur les  $U$ -marges pour tout  $U\subseteq D$

$$f_U = \int_{[0,1]^d} \left( f - \sum_{V\subseteq U} f_V \right) dx_{-U}.$$

D'après Efron et Stein (1981), si  $f$  est de carré intégrable alors on a la décomposition suivante de  $f(X)$ ,

$$f(X) = \mu + \sum_j f_j(X_j) + \sum_{i < j} f_{ij}(X_i, X_j) + \dots + f_{1\dots d}(X_1, \dots, X_d).$$

$\mu + \sum_j f_j(X_j)$  s'appelle la partie additive de la fonction. On note  $r$  le **reste d'ordre un** de la fonction, i.e

$$r(X) = f(X) - \left( \mu + \sum_j f_j(X_j) \right).$$

Une fonction est dite **additive** si sa décomposition se restreint à sa partie additive.

A partir de cette réécriture de  $f$  on déduit facilement la décomposition de la variance.

### Proposition 2.5

*Si  $f$  est de carré intégrable alors la variance de  $f(x)$  se décompose comme suit*

$$\sigma^2 = \sum_j \int_{[0,1]^d} f_j^2 dX + \sum_{i < j} \int_{[0,1]^d} f_{ij}^2 dX + \dots + \int_{[0,1]^d} f_{1\dots d}^2 dX.$$

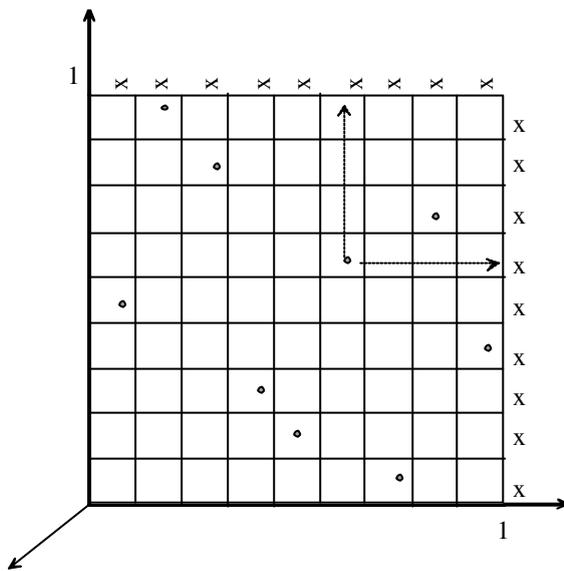
On remarque que la variance se décompose suivant les effets sur les marges à une, deux, ...,  $d$  dimensions. Nous allons voir avec la méthode d'échantillonnage suivante, que si l'on impose une répartition uniforme des points de l'échantillon sur toutes les marges à une dimension, on peut supprimer de la variance les effets de  $f$  sur ces marges, c'est-à-dire réduire la variance.

### 2.2.3. Méthode des hypercubes latins

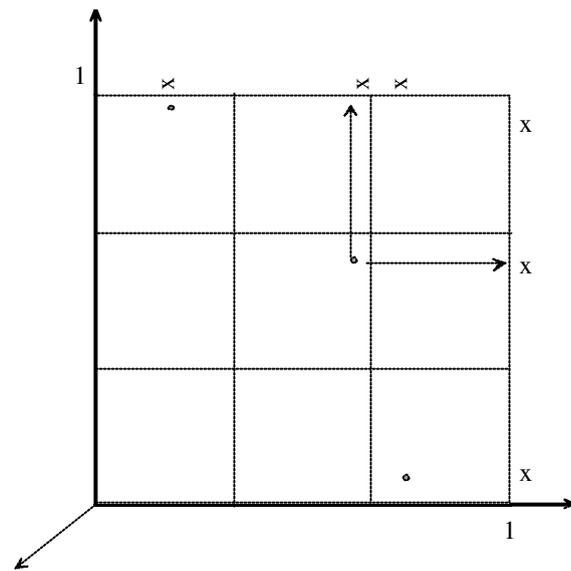
Pour cette méthode on découpe chaque arête du cube unité en un même nombre de segments de même longueur,  $q_1 = \dots = q_d = N$ , de façon à obtenir  $N^d$  cellules de  $[0, 1]^d$ . Les cellules représentées dans le plan sont alors choisies à partir d'un hypercube latin.

#### Définition 2.6

*Le tableau  $P=(p_{ij})$  d'ordre  $N \times d$  à éléments dans  $(\mathbb{Z}/N)$  est un hypercube latin si  $\forall J, |J|=1$  toutes les cellules de la  $J$ -marge sont représentées une fois chacune par les lignes de  $P$ .*



Plan d'échantillonnage issu d'un hypercube latin à éléments dans  $(\mathbb{Z}/9)$



Plan d'échantillonnage issu d'un hypercube latin à éléments dans  $(\mathbb{Z}/3)$

Un hypercube latin est une structure très simple à construire puisqu'en fait chaque colonne est une permutation de  $0, \dots, N-1$ .

#### Principe de la randomisation

Utiliser un échantillonnage à deux degrés conduit à un estimateur biaisé de l'intégrale, alors pour contourner ce problème, on randomise le premier degré. Cependant, la randomisation ne doit pas se faire n'importe comment, car on risquerait de perdre la répartition uniforme des points d'échantillonnage sur les  $J$ -marges. Une randomisation doit donc vérifier les deux principes suivants.

- Elle doit conserver la structure d'origine du premier degré (hypercube latin, tableau orthogonal,...).
- L'estimateur doit être sans biais.

Dans le cas d'un hypercube latin, la randomisation est très simple puisqu'il suffit d'effectuer une permutation aléatoire de  $0, \dots, N-1$  sur chaque colonne de  $P$ , ce qui ne change rien à sa structure. Un point d'échantillonnage  $X_i$  est alors défini par

$$X_{ij} = \frac{\pi_j(p_{ij}) + \varepsilon_{ij}}{N}, \quad 1 \leq j \leq d,$$

où chaque  $\pi_j$  est une permutation aléatoire de  $0, \dots, N-1$  uniformément distribuée parmi les  $N!$  permutations possibles. Ces permutations sont indépendantes entre elles.

L'idée de randomiser pour construire un échantillon vient de Patterson (1954) mais ce sont Mc Kay *et al.* (1979) qui ont utilisé en premier cette technique pour l'intégration numérique.

### **Etude de la variance et de la vitesse de convergence de l'estimateur**

On note  $\hat{I}_{f_{HL}}$  l'estimateur de l'intégrale utilisant un échantillon construit à partir d'un hypercube latin. Comme nous l'avons dit, cet estimateur est sans biais.

Dans un premier temps, Mc Kay *et al.* (1979) ont montré que sa variance pouvait se décomposer de la façon suivante

$$\text{var } \hat{I}_{f_{LH}} = N^{-1}\sigma^2 + N^{-1}(N-1) \text{cov}(f(X_1, X_2)),$$

puis, dans le cas particulier où  $f$  est monotone par rapport à chacun de ses arguments, que  $\text{cov}(f(X_1, X_2)) \leq 0$ , autrement dit que

$$\text{var } \hat{I}_{f_{LH}} \leq \text{var } \hat{I}_{f_{MC}},$$

où  $\hat{I}_{f_{MC}}$  est l'estimateur de l'intégrale par la méthode de Monte Carlo. Ce résultat est intéressant car c'est le seul que l'on ait à taille finie d'échantillon, toutefois la contrainte imposée à  $f$  est très forte.

Dans le cadre plus large des fonctions de carré intégrable, Stein (1987) donne un résultat asymptotique sur la variance de l'estimateur en fonction du reste d'ordre un de la fonction (p. 43)

$$\text{var } \hat{I}_{f_{LH}} = N^{-1} \int r^2(X) dX = \text{var } \hat{I}_{f_{MC}} - N^{-1} \sum_j \int f_j^2 dX + o(N^{-1}),$$

où, rappelons-le,  $\text{var } \hat{I}_{f_{MC}} = N^{-1}\sigma^2$ . Il montre ainsi qu'il y a réduction de la variance par rapport à la méthode de Monte Carlo même si la variance reste d'ordre  $O(N^{-1})$ , mais surtout que cette réduction ne se fait pas n'importe comment. On s'aperçoit en effet que ce sont **les effets de  $f$  sur les marges à une dimension qui sont supprimés de la variance**. Ainsi, si  $f$  est une fonction additive, la variance est de l'ordre de  $o(N^{-1})$  et non plus  $O(N^{-1})$ .

Cependant, ce résultat est asymptotique, il faut donc se poser le problème de la vitesse de convergence de l'estimateur vers  $\mu$ . Il existe pour cela deux théorèmes de la limite centrale, l'un démontré par Owen (1992) dans le cas des fonctions bornées, et l'autre démontré par Loh (1993) avec des conditions plus faibles sur  $f$ ,

$$N^{1/2}(\hat{I}_{f_{LH}} - \mu) \xrightarrow{d} \mathbf{N}(0, \int r^2(X) dX).$$

On retrouve la même vitesse de convergence que pour la méthode de Monte Carlo,  $N^{1/2}$ , mais avec une variance asymptotiquement plus petite.

Si on applique la méthode des hypercubes latins pour estimer les paramètres du modèle, alors on obtient la proposition suivante sur l'opérateur de covariance de l'estimateur.

**Proposition 2.7**

*La matrice de covariance de l'estimateur des paramètres du modèle pour la méthode des hypercubes latins est de la forme*

$$\Sigma = \begin{pmatrix} \int r_1^2 dx & \int r_1 r_j dx \\ \text{sym} & \dots & \int r_m^2 dx \end{pmatrix},$$

*où  $r_i$  est le reste d'ordre un de la  $i$ ème composante du vecteur des fonctions  ${}^tZ(x)Y(x)$ ,  $i=1, \dots, m$*

Actuellement, cette méthode d'échantillonnage basée sur les hypercubes latins est la plus (voir la seule) utilisée. En effet, non seulement elle améliore nettement la méthode de Monte Carlo, mais de plus, elle est très simple à mettre en place, puisque la construction d'un hypercube latin ne pose pas de problème. On peut aussi noter que les hypercubes latins sont souvent utilisés dans l'approche par résidu aléatoire, ce qui permet de restreindre la recherche d'un plan (par critère d'entropie, distance) à une classe de plans.

**2.2.4. Méthode des tableaux orthogonaux**

Un hypercube latin impose aux points d'échantillonnage des contraintes sur les marges de dimension 1. Nous avons vu que cela a pour conséquence de supprimer de la variance les effets de  $f$  sur ces marges. Nous allons voir maintenant que ce résultat se généralise aux marges de dimension quelconque. Pour cela on utilise des tableaux orthogonaux. Pour cette méthode on découpe chaque arête du cube unité en un même nombre de segments de même longueur,  $q_1 = \dots = q_d = q$ .

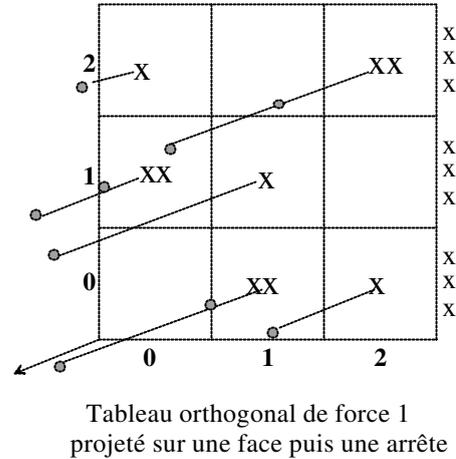
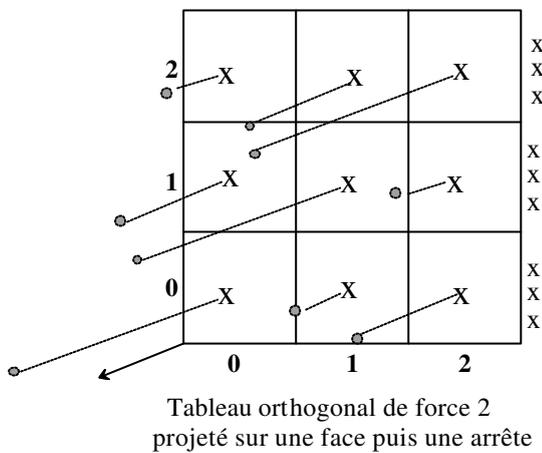
**Définition 2.8**

*Soit  $T$  un tableau d'ordre  $N \times d$  à éléments dans  $(Z/q)$ .  $T$  est un tableau orthogonal de force  $t$ ,  $1 \leq t \leq d$  à  $q$  symboles, et d'index  $\lambda$ , si et seulement si dans tout bloc formé de  $t$  colonnes de  $T$ , tous les éléments de  $(Z/q)^t$  figurent un même nombre de fois  $\lambda$  chacun. On le note  $OA_\lambda(N, d, q, t)$  où  $N = \lambda q^t$ .*

**Remarque :** Un hypercube latin est un tableau orthogonal de force 1.

## Propriété 2.9

*Si on effectue une permutation de  $0, \dots, q-1$  sur chaque colonne d'un tableau orthogonal de force  $t$ , on garde un tableau orthogonal de force  $t$ .*



On remarque sur les dessins ci-dessus que certaines parties du domaine d'échantillonnage ne sont pas représentées par le tableau de force 1 alors qu'elles le sont par le tableau de force 2. On peut alors supposer que plus un tableau est de force élevée, plus la répartition des points d'échantillonnage est uniforme dans  $[0, 1]^d$ .

Comme précédemment, on peut utiliser un tableau orthogonal randomisé pour construire le plan d'échantillonnage et obtenir un estimateur sans biais.

### Etude de la variance et de la vitesse de convergence de l'estimateur

On note  $\hat{f}_{OA}$  l'estimateur utilisant un échantillon construit à partir d'un tableau orthogonal de force  $t$ . Owen (1992) a démontré le résultat suivant

$$\text{var } \hat{f}_{OA} = \sum_{|u| \geq t} \int f_u^2 dX + o(N^{-1}).$$

Ce qui peut encore s'écrire sous la forme plus explicite suivante,

$$\text{var } \hat{f}_{OA} = \text{var } \hat{f}_{MC} - N^{-1} \sum_{1 \leq |u| \leq t} \int f_u^2 dX + o(N^{-1}).$$

Il y a réduction de la variance par rapport à la méthode de Monte Carlo, et de la même façon que précédemment, c'est-à-dire que les effets de  $f$  sur les marges à  $u$  dimensions pour  $u=1, \dots, t$  sont supprimés de la variance. Plus généralement, si on note  $\hat{f}_{OA(t)}$  l'estimateur construit avec

un tableau de force  $t$  et  $\hat{f}_{OA(s)}$  l'estimateur construit avec un tableau de force  $s < t$ , on a

$$\text{var } \hat{f}_{OA(t)} = \text{var } \hat{f}_{OA(s)} - N^{-1} \sum_{s+1 \leq |u| \leq t} \int f_u^2 dX + o(N^{-1}).$$

Donc on améliore aussi la méthode des hypercubes latins, en résumé, plus la force du tableau est élevée et plus on réduit la variance.

De même que précédemment, ce résultat est asymptotique, il faut donc se poser la question de la vitesse de convergence de l'estimateur vers  $\mu$ . A l'heure actuelle un seul théorème de la limite centrale a été établi par Loh (1994) mais uniquement dans le cas où le nombre de facteurs est égal à trois (avec des contraintes relativement faibles sur  $f$ ).

### Alternative à la randomisation des tableaux orthogonaux

Cette alternative est proposée par Tang (1993), elle consiste à construire de façon aléatoire un hypercube latin à partir d'un tableau orthogonal de force 2. Soit  $T=(t_{ij})$  un  $OA(q^2, d, q, 2)$  randomisé (permutation aléatoire de ses colonnes), pour chaque colonne du tableau on remplace le symbole  $k \in \{0, \dots, q-1\}$  par une permutation aléatoire de  $kq, kq+1, \dots, (k+1)q-1$ , pour obtenir un hypercube latin randomisé  $P$ . Réciproquement, à partir de l'hypercube latin, on récupère le tableau  $T$  en utilisant la relation

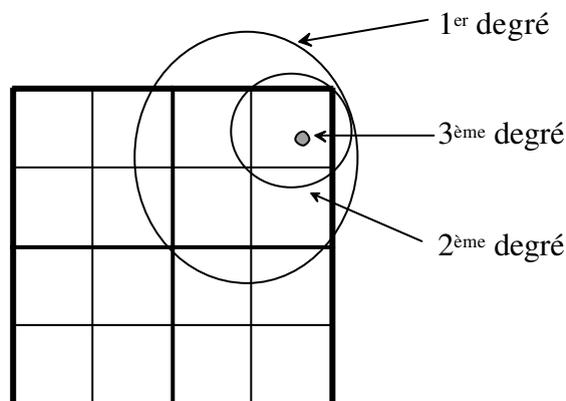
$$t_{ij} = [q_{ij} / q], \quad i=1, \dots, q^2,$$

où  $[x]$  est égal au plus petit entier supérieur ou égal à  $x$ . Par exemple, si  $T$  est un  $OA(4, 2, 2, 2)$  randomisé

$$T = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix} \quad \begin{matrix} 0 \rightarrow 0, 1 \\ 1 \rightarrow 3, 2 \end{matrix} \quad P = \begin{pmatrix} 0 & 1 & 3 & 2 \\ 0 & 3 & 1 & 2 \end{pmatrix}.$$

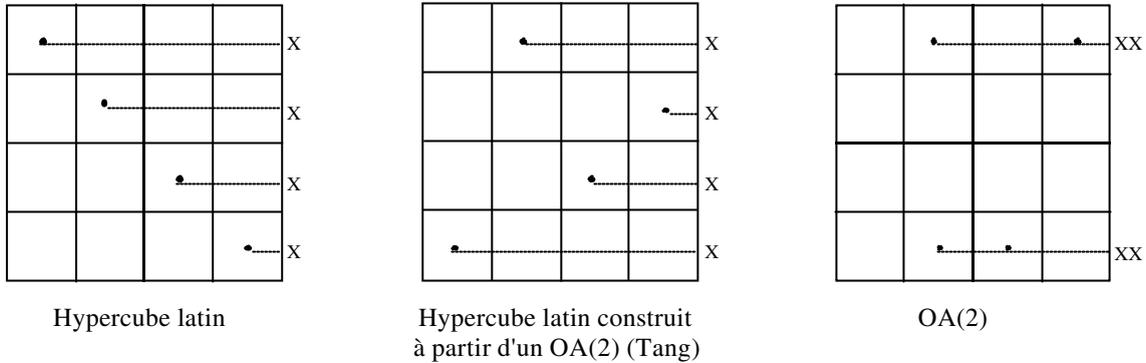
Cette technique est en fait une méthode d'échantillonnage à trois degrés. On partitionne une première fois le cube unité en  $q^d$  grandes cellules, puis on découpe chacune de celles-ci en  $q^d$  petites cellules. La méthode d'échantillonnage de Tang consiste alors à

- 1<sup>er</sup> degré : choisir des grandes cellules parmi les  $q^d$ , de façon à ce qu'elles forment un tableau de force 2.
- 2<sup>ème</sup> degré : choisir une petite cellule à l'intérieur de chaque cellule sélectionnée au 1<sup>er</sup> degré, de façon à ce qu'elles forment un hypercube latin.
- 3<sup>ème</sup> degré : choisir au hasard un point à l'intérieur de chaque petite cellule sélectionnée au 2<sup>ème</sup> degré.



Quel est l'intérêt d'utiliser un hypercube latin construit de cette façon? Tang (1993) a montré, indépendamment de Owen, que cela permettait de supprimer de la variance les effets

de  $f$  sur les marges à une et deux dimensions (car force 2). Il n'y a donc pas *a priori* d'amélioration de la méthode de Owen dans le cas général. Et pourtant les points d'échantillonnage sont mieux répartis dans le cube unité puisqu'on leur impose simultanément des contraintes sur les marges à une et deux dimensions.



Projection de différents plans sur une face du cube unité

Tang a tout de même démontré que dans le cas particulier où la fonction est additive, sa méthode s'avère plus efficace. En fait, l'intérêt de cette méthode ne vient pas des résultats obtenus mais de son principe de base, c'est-à-dire l'idée "d'emboîter un hypercube latin dans un tableau orthogonal". Cette technique a une généralisation très intéressante que nous présenterons brièvement avec la troisième méthode d'échantillonnage.

## 2.3. Méthodes quasi-Monte Carlo

Les méthodes quasi-Monte Carlo, encore appelées méthodes d'équidistribution, ont été abondamment étudiées depuis leur apparition dans les travaux de Koksma (1942/43) et Hlawka (1964). Elles ont pour origine des résultats issus de la théorie des nombres appliqués à l'analyse numérique. Les auteurs suivants ont publiés des ouvrages sur le sujet auxquels cette première partie se réfère : Fang et Wang (1993), Fang, Wang et Bentler (1994), Hua et Wang (1981), Wang et Fang (1990), Drmota et Tichy (1997), Sloan et Joe (1994). Cette liste n'est pas exhaustive.

### 2.3.1. Méthodes quasi-Monte Carlo : discrédance

#### Majoration de l'erreur d'intégration

L'idée de base des méthodes quasi-Monte Carlo est de remplacer l'échantillon aléatoire de la méthode de Monte Carlo par une suite déterministe de points "bien choisis". Nous parlons dans ce cas d'approximation de l'intégrale de façon à le différencier du cas de l'estimation de l'intégrale avec un plan d'échantillonnage à deux degrés. On note

$$Qf = \frac{1}{N} \sum_{i=1}^N f(x_i),$$

où  $x_1, \dots, x_N$  est une suite déterministe des points de  $[0,1]^d$ . Le critère de choix de ces points varie selon le problème traité. Dans le cas de l'intégration numérique, on utilise la fonction de discrédance. Cette fonction permet de mesurer la différence entre la répartition empirique des points et la distribution uniforme.

#### Définitions 2.10

*Soit  $x_1, \dots, x_N$  une suite de points de  $[0,1]^d$ . La fonction de répartition empirique de cette suite est définie par*

$$F_N(t) = \frac{1}{N} \sum_{i=1}^N I_{\{x_i \leq t\}},$$

*où  $I_A$  est la fonction indicatrice de  $A$ .*

*On appelle alors fonction de discrédance (ou discrédance) de la suite*

$$D(N) = \sup_{t \in [0,1]^d} |F_N(t) - U(t)|,$$

*où  $U(t)$  est la distribution uniforme sur  $[0,1]^d$ .*

*On dit qu'une suite de points est uniformément répartie sur  $[0,1]^d$  si*

$$\lim_{N \rightarrow \infty} D(N) = 0.$$

**Remarque :** Lorsqu'on parle de limite quand  $N$  tend vers l'infini, on suppose en fait que  $x_1, \dots, x_N$  fait partie d'une suite infinie de  $\mathbb{R}^d$ . En général, on a une suite  $(x_n)_{n \geq 1}$  de  $\mathbb{R}^d$ , et  $x_1, \dots, x_N$  correspondent aux  $\{x_n\}_{n \geq 1}$  où  $\{x\} = x \bmod 1$ .

D'après cette définition, une suite est uniformément répartie si asymptotiquement sa répartition empirique est très proche de la distribution uniforme. Dans le cadre de l'intégration numérique le résultat suivant, démontré par Drmota et Tichy (1996), justifie le choix de la discrédance comme mesure de l'uniformité. Une suite de points est uniformément répartie, c'est-à-dire que sa discrédance tend vers 0, si et seulement si

$$\lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N f(x_i) \right) = \int_{[0,1]^d} f(x) dx.$$

Il existe des variantes à la fonction de discrédance, discrédance étoile, discrédance  $L^p$ , ..., et discrédance isotropique. Cette dernière est souvent considérée comme la discrédance car plus facile à manipuler. Soit  $N(\gamma)$  le nombre de points d'une suite  $x_1, \dots, x_N$  de  $[0,1]^d$  vérifiant  $x_k \leq \gamma$   $k=1, \dots, N$ . La discrédance isotropique est définie par

$$D(N) = \sup_{\gamma \in [0,1]^d} \left| \frac{N(\gamma)}{N} - v([0, \gamma]) \right|,$$

où  $v([0, \gamma]) = \gamma_1 \dots \gamma_d$  est le volume du rectangle  $[0, \gamma]$ .

L'erreur d'intégration numérique est donnée par l'inégalité de Koksma-Hlawka suivante.

### **Théorème 2.11**

*Soient  $f$  une fonction à variations bornées au sens de Hardy et Krause et  $x_1, \dots, x_N$  une suite de discrédance  $D(N)$ . Alors*

$$|If - Qf| \leq V(f) D(N),$$

*où  $V(f)$  est la variation totale de  $f$ .*

Pour la définition d'une fonction à variations bornées, on peut se référer à l'ouvrage de Niederreiter (1992). Ce qui importe ici est que  $V(f)$  est indépendant de  $N$ , donc la vitesse de convergence de  $Qf$  est du même ordre que  $D(N)$ . Or comme nous l'avons dit en introduction, l'intérêt des méthodes quasi-Monte Carlo est d'améliorer la vitesse de convergence de la méthode de Monte Carlo qui est  $N^{-1/2}$ , d'où la définition suivante.

### Définition 2.12

*On dit qu'une suite de points a une bonne répartition uniforme sur  $[0,1]^d$  si sa discrédance est de l'ordre de  $N^{-1+\epsilon}$  avec  $0 < \epsilon < 1/2$ .*

#### Estimation de l'erreur d'intégration

On remarque que dans les méthodes d'équidistribution rien n'est aléatoire, et la majoration d'erreur obtenue est déterministe, contrairement à la méthode de Monte Carlo où la convergence est en loi. Toutefois, il n'est pas toujours possible de disposer de la majoration d'erreur. En effet, dans la plupart des cas, on n'a aucune idée de l'ordre de grandeur de la variation totale de la fonction puisque  $f$  représente la réponse d'un simulateur. On est alors obligé d'estimer l'erreur d'intégration et pour cela Fang et Zhu (1993) proposent d'introduire une perturbation aléatoire dans la suite en utilisant un échantillonnage à deux degrés. Soit  $a_1, \dots, a_N$  une suite de  $[0,1]^d$  de discrédance  $D(N)$ , on définit les points d'échantillonnage par

$$x_i = a_i + \frac{\epsilon_i}{N} \quad i = 1, \dots, N,$$

où les  $\epsilon_i$  sont des vecteurs aléatoires de distribution uniforme sur  $[0,1]^d$  indépendants (la suite  $a_1, \dots, a_N$  est choisie telle que  $a_i + \epsilon_i/N \in [0,1]^d$ ).

Il est impossible ici de randomiser le premier degré car sinon on perd la discrédance de la suite  $(a_i)$ . Donc, en général,  $\hat{I}_f$  n'est pas un estimateur sans biais de  $I_f$  mais uniquement asymptotiquement sans biais avec une vitesse de convergence de l'ordre de  $O(D(N))$  et une variance de l'ordre de  $O(D(N)^2)$  pour toute fonction à dérivées partielles bornées. Ce résultat dépend uniquement de la discrédance de la suite choisie, il permet de comparer les méthodes d'équidistribution avec les méthodes de réduction de la variance.

Il existe différents types de suites ayant une bonne répartition uniforme, par exemple les suites de Halton qui ont une discrédance de l'ordre de  $O(N^{-1}(\log N)^{d-1})$ ,  $d \geq 2$ , mais les plus répandues sont celles que l'on va étudier maintenant et que l'on appelle réseaux (lattice).

### 2.3.2. Les réseaux (lattice)

#### Définition algébrique des réseaux

Soit  $K$  l'un des corps  $\mathbb{R}$  ou  $\mathbb{Q}$  et  $V$  un  $K$ -espace vectoriel de dimension  $d$ .

### Définition 2.13

*Une partie  $L$  de  $V$  est un sous-réseau de  $V$  s'il existe une famille libre  $b=(b_1, \dots, b_d)$  de  $V$  telle que*

$$L = \mathbb{Z} b_1 + \dots + \mathbb{Z} b_d.$$

*On dit que  $b$  est une  $\mathbb{Z}$ -base de  $L$ .*

Soient  $L$  un sous-réseau de  $V$  et  $(b_1, \dots, b_t)$  une  $Z$ -base de  $L$ . Alors  $t$  est égal à la dimension du sous-espace engendré par les vecteurs de  $L$ ,  $t = \dim_{\mathbb{K}} \text{vect}(L)$ , il ne dépend pas de la  $Z$ -base choisie pour  $L$ .

**Définition 2.14**

*On dit que  $L$  est un réseau de  $V$  si et seulement si  $t = \dim(V) = d$ .*

Le réseau est souvent noté  $L$  puisqu'en anglais son nom est "lattice".

Supposons désormais que  $V = \mathbb{K}^d$ . Le résultat suivant est dû à Cassel (1959) et permet de caractériser les réseaux sans faire intervenir la notion de  $Z$ -base. Nous donnons ici cette caractérisation car dans la bibliographie sur l'intégration numérique, elle est souvent considérée comme la définition des réseaux.

**Proposition 2.15**

*Un sous-ensemble  $L$  de  $\mathbb{K}^d$  est un réseau si et seulement si il vérifie les 3 conditions suivantes*

- (i) si  $x$  et  $y$  appartiennent à  $L$  alors  $x+y$  et  $x-y$  aussi,*
- (ii)  $L$  contient  $d$  vecteurs linéairement indépendants,*
- (iii) il existe une sphère de centre  $0$  qui ne contient aucun point de  $L$  sauf  $0$ .*

On peut définir un certain nombre d'outils matriciels afin de travailler plus facilement avec le réseau. Soit  $(b_1, \dots, b_d)$  une  $Z$ -base d'un réseau  $L$  de  $V$ .

**Définitions 2.16**

*Les vecteurs de la  $Z$ -base sont appelés générateurs du réseau.*

*La matrice  $M$  carrée d'ordre  $d$  dont les lignes sont formées des vecteurs de la  $Z$ -base, s'appelle matrice génératrice du réseau.*

*On appelle déterminant du réseau la valeur absolue du déterminant de la matrice génératrice  $M$ ,  $\det L = | \det M |$ .*

On appelle cellule unité le parallélogramme défini par

$$\{\lambda_1 b_1 + \dots + \lambda_d b_d; 0 \leq \lambda_i \leq 1, i=1, \dots, d\}.$$

Alors Cassels (1971) a démontré le résultat suivant.

### Proposition 2.17

*Le déterminant du réseau est indépendant de la Z-base choisie. Il représente le volume de la cellule unité et  $(\det L)^{-1}$  le nombre moyen de points du réseau par unité de volume.*

Les réseaux utilisés pour l'intégration numérique sont des réseaux rationnels particuliers, dits réseaux d'échantillonnage.

### Réseaux d'échantillonnage

#### Définition 2.18

*Un réseau d'échantillonnage  $L$  est un réseau rationnel, i.e  $K=\mathbf{Q}$ , contenant  $\mathbf{Z}^d$ . Une Z-base de  $L$  est de la forme*

$$\left( \frac{a_1}{q_1}, \dots, \frac{a_d}{q_d} \right),$$

*où  $q_i$  est un entier et  $a_i$  un vecteur d'entiers  $i=1, \dots, d$ . On appelle points d'échantillonnage du réseau les points de  $L \cap [0,1]^d$ , il y en a  $N=(\det L)^{-1}$ .*

#### Définition et propriétés 2.19

*Soit  $L$  un réseau d'échantillonnage de matrice génératrice  $M$ . L'ensemble défini par*

$$L^\perp = \{x \in \mathbf{R}^d, (x | y) \in \mathbf{Z}, \forall y \in L\}$$

*s'appelle le dual du réseau  $L$ . Il vérifie les propriétés suivantes*

- (i)  $L^\perp$  est un réseau d'échantillonnage,*
- (ii)  $L^\perp \subseteq \mathbf{Z}^d \Rightarrow L^\perp = \{x \in \mathbf{Z}^d, (x | y) \in \mathbf{Z}, \forall y \in L\},$*
- (iii)  ${}^t M^{-1}$  est la matrice génératrice de  $L^\perp$  et  $\det L^\perp = (\det L)^{-1}$ .*

**Démonstration :** Soit  $(b_1, \dots, b_d)$  une Z-base de  $L$  (i.e les lignes de  $M$ ), on note  $c_1, \dots, c_d$  les colonnes de  $M^{-1}$ . Alors  $(b_i | c_j) = \delta_{ij}$   $1 \leq i, j \leq d$  donc les colonnes  $c_i$ ,  $i=1, \dots, d$ , appartiennent à  $L^\perp$ . De plus tout  $x \in L^\perp$  peut s'écrire sous la forme

$$x = M M^{-1} x = c_1 (b_1 | x) + \dots + c_d (b_d | x)$$

or  $(b_i | x) \in \mathbf{Z}$   $i=1, \dots, d$  par définition de  $L^\perp$ . Donc  $L^\perp$  est un réseau de Z-base  $(c_1, \dots, c_d)$ .

Montrons maintenant qu'il est inclus dans  $Z^d$ . Si on note  $e_i, i=1, \dots, d$  les vecteurs unités de  $Z^d$ , alors  $e_i \in L, i=1, \dots, d$  puisque  $L$  contient  $Z^d$ . Donc pour tout  $x \in L^\perp, (x|e_i) = x_i \in Z, i=1, \dots, d$ , c'est-à-dire  $x \in Z^d$ .  $\square$

Les réseaux d'échantillonnage sont difficiles à manipuler pour deux raisons. La première est que ce sont des ensembles de points infinis et non bornés (on remarque d'ailleurs que toute translation par un vecteur formé de deux points du réseau, laisse le réseau inchangé). Or seuls les points du réseau qui sont dans le cube unité, *i.e* les points d'échantillonnage, nous intéressent et nous savons qu'il y en a un nombre fini. La deuxième raison est qu'il n'existe pas une unique  $Z$ -base pour un réseau de déterminant donné, *i.e* pour un réseau dont le nombre de points d'échantillonnage est fixé. Le résultat suivant (Sloan et Joe (1994)) est très intéressant puisqu'il permet de caractériser de façon unique l'ensemble des points d'échantillonnage.

**Proposition 2.20**

*Soit  $L$  un réseau d'échantillonnage. Il existe un unique entier  $r$ , le rang du réseau, et une unique suite d'entiers  $q_1, \dots, q_r$ , les invariants du réseau, vérifiant*

$$q_{i+1} \text{ divise } q_i, i=1, \dots, r,$$

*tels que tout point d'échantillonnage s'écrit de façon unique*

$$\left\{ z_1 \frac{a_1}{q_1} + \dots + z_r \frac{a_r}{q_r} \right\} \text{ avec } \{x\} = x \pmod{1},$$

*où pour  $i=1, \dots, r, z_i \in (Z/q_i)$  et  $a_i$  sont des vecteurs d'entiers à éléments dans  $(Z/q_i)$  appelés générateurs du réseau.*

$\left( \frac{a_1}{q_1}, \dots, \frac{a_r}{q_r} \right)$  s'appelle *la Z-base canonique* de  $L$ . Pour avoir une  $Z$ -base de  $L$ , il suffit de compléter la  $Z$ -base canonique avec  $d-r$  vecteurs de  $Z^d$ . On remarque que

$$(\det L)^{-1} = N = \prod_{i=1}^r q_i$$

et que  $\left( \frac{a_1}{q_1}, \dots, \frac{a_r}{q_r} \right)$  est la  $Z$ -base canonique de  $L$  si et seulement si elle vérifie

$$\left\{ \sum_{i=1}^r z_i \frac{a_i}{q_i} \right\} = 0, \quad 0 \leq z_i < q_i \iff z_i = 0, i = 1, \dots, r.$$

On peut de plus vérifier aisément que si on note  $q = \text{PPCM}(q_i, i=1, \dots, r) = q_1$  alors pour tout point d'échantillonnage  $x$ , on a  $qx \in Z^d$ .

Cette décomposition s'obtient à partir du résultat classique sur la décomposition des groupes abéliens finis. Nous y revenons plus en détails dans la deuxième partie, lorsque nous faisons le lien entre réseaux et fractions régulières.

### Cas particulier des réseaux de rang 1

A partir de la proposition précédente, on peut définir la classe des réseaux de rang 1, *i.e* engendrés par un seul générateur. En général un réseau de rang 1 de taille  $N$  est défini à partir d'un vecteur  $g=(g_1, \dots, g_d) \in (\mathbb{Z}/N)^d$  vérifiant les conditions suivantes.

$$\begin{cases} g_i = 1 \\ \text{PGCD}(g_i, N) = 1, \quad i = 1, \dots, d \\ g_i \neq g_j, \quad i \neq j \end{cases}$$

En pratique, ces réseaux sont les seuls (ou presque) à être utilisés. Les points d'échantillonnage sont en effet très simples à construire puisqu'ils sont de la forme

$$x_k = kg \pmod{N}, \quad k=0, \dots, N-1.$$

De plus le dual a aussi une expression simplifiée

$$L^\perp = \left\{ h \in \mathbb{Z}^d, \quad (h | g) = 0 \pmod{N} \right\}.$$

Ce qui s'avère intéressant pour déterminer le vecteur générateur que l'on doit choisir. En effet, la qualité du plan d'échantillonnage dépend uniquement de  $g$ . En intégration numérique, les vecteurs générateurs sont construits de façon à minimiser l'erreur de Koksma-Hawlka, qui dépend de la discrédance de la suite mais aussi de la variation totale de la fonction. Afin de simplifier cette construction, on se restreint à la classe des fonctions admettant un développement en série de Fourier absolument convergent,  $\forall x \in [0, 1)^d$

$$f(x) = \sum_{h \in \mathbb{Z}^d} \hat{f}(h) \exp\left[2\pi i(h | x)\right]$$

où  $\hat{f}(h)$  sont les coefficients de Fourier de la série,

$$\hat{f}(h) = \int_{[0,1]^d} \exp\left[-2\pi i(h | x)\right] f(x) dx, \quad h \in \mathbb{Z}^d.$$

L'erreur d'intégration pour ce type de fonctions est donnée par la proposition qui suit.

#### **Proposition 2.21**

*Soit  $L$  un réseau d'échantillonnage de rang 1. On note  $Qf$  l'approximation de l'intégrale de  $f$ ,  $If$ , sur les points d'échantillonnage de  $L$ . On a*

$$|Qf - If| = \sum_{h \in L^\perp - \{0\}} \hat{f}(h).$$

Les réseaux ainsi construits s'appellent les "good lattice points". La bibliographie les concernant est considérable, on peut trouver des tables de générateurs par exemple dans les ouvrages de Hua et Wang (1981), Wang et Fang (1994) ou Sloan et Joe (1994). Si les générateurs des réseaux de rang 1 sont construits à partir des fonctions admettant un développement en série de Fourier, on peut supposer que le modèle linéaire qu'il est préférable d'utiliser doit être de type trigonométrique.

La proposition suivante permet de faire le lien avec les hypercubes latins. Mais nous verrons dans la deuxième partie qu'il est possible de généraliser ce lien aux tableaux orthogonaux de force quelconque mais dans le cas où ceux-ci sont linéaires, *i.e* des cas particuliers de fractions régulières.

**Proposition 2.22**

*Le tableau  $P$  d'ordre  $N \times d$  dont les lignes sont définies par les points d'échantillonnage d'un réseau de rang 1 est un hypercube latin.*

**Démonstration :** Toutes les lignes de  $P$  sont de la forme  $kg \pmod{N}$ ,  $k=0, \dots, N-1$ . Montrons que chaque colonne de  $P$  est une permutation de  $0, \dots, N-1$ . Si l'on prend deux éléments distincts dans une même colonne de  $P$ , ils sont de la forme  $kg \pmod{N}$  et  $k'g \pmod{N}$  avec  $k \neq k' \pmod{N}$ . Supposons qu'ils soient égaux, alors  $kg = k'g \pmod{N} \Leftrightarrow (k-k')g = 0 \pmod{N} \Leftrightarrow k = k' \pmod{N}$  puisque  $\text{PGCD}(N, g) = 1$ . Donc tous les éléments de la colonne sont 2 à 2 différents et comme il y en a  $N$ , ils forment nécessairement une permutation de  $0, \dots, N-1$ .  $\square$

Les réseaux de rang supérieur à 1 ont été très peu étudiés et les résultats connus concernent uniquement le rang 2. Les auteurs qui se sont penchés sur ce problème laissent supposer que les réseaux de rang supérieur à 1 sont tout aussi compétitifs que les réseaux de rang 1. Cette hypothèse repose essentiellement sur des tests numériques effectués pour le rang 2 (Sloan et Walsh (1990)). Toutefois, Niederreiter (1992) a réussi à démontrer que les réseaux de rang 2 font aussi bien que les meilleurs réseaux de rang 1.

**2.3.3. Propriétés de projection des réseaux d'échantillonnage**

Certains auteurs se sont intéressés aux propriétés de projection des réseaux d'échantillonnage (Sloan et Lyness (1989), Sloan et Joe (1994)). Ces propriétés peuvent en effet jouer un rôle important dans le cas où, lors d'une estimation, on est amené à traiter un terme qui ne dépend que de  $s < d$  paramètres. Il est alors intéressant de savoir si les projections en dimension  $s$  du réseau utilisé gardent une bonne répartition de leurs points. Cette contrainte marginale que l'on impose aux réseaux ressemble beaucoup à la notion de force d'un tableau orthogonal. Nous verrons dans la deuxième partie qu'il existe effectivement des similitudes dans le cas où le tableau est linéaire.

Les propriétés suivantes permettent de connaître dans un premier temps, la structure de la projection.

## Propriétés 2.23

- 1) *Toute projection d'un réseau est un réseau.*  
 2) *Si le réseau d'origine est de rang  $r$  et d'invariants  $q_1, \dots, q_r$ , alors toute projection en dimension  $s < d$  est de rang  $r' \leq r$  et d'invariants  $q'_1, \dots, q'_{r'}$  où  $q'_k$  divise  $q_k$  pour  $1 \leq k \leq r'$ .*

La projection d'un réseau étant elle-même un réseau, tous ses points sont distincts. Cependant, on constate que si l'on projette deux points distincts d'un réseau en dimension  $1 \leq s \leq d$ , i.e si on omet  $(d-s)$  de leurs composantes, les deux points obtenus peuvent être confondus. Donc la taille du réseau projeté peut devenir très petite par rapport à la taille du réseau d'origine. La préoccupation des auteurs qui ont travaillé dans le domaine est de veiller à ce que les projections du réseau soient des réseaux d'un ordre aussi élevé que possible. C'est dans ce sens qu'ils ont défini les propriétés de projection qu'un réseau doit vérifier.

Soit un réseau d'échantillonnage de rang  $r$  et d'invariants  $q_1, \dots, q_r$ . On définit une Z-base en complétant la Z-base canonique

$$\left( \frac{a_1}{q_1}, \dots, \frac{a_d}{q_d} \right),$$

où  $a_{r+1}, \dots, a_d$  sont des vecteurs entiers quelconques et  $q_{r+1} = \dots = q_d = 1$ .

## Définition 2.24

*Soit  $s$ ,  $1 \leq s \leq d$ , une projection du réseau en dimension  $s$  est régulière si elle forme un réseau d'ordre  $q_1 \times \dots \times q_s$  et d'invariants  $q_1, \dots, q_s$ . On dit que le réseau est à projections régulières si les projections en dimension  $s$ , pour tout  $s$ ,  $1 \leq s \leq d$ , sont régulières.*

**Exemple 2.25 :** Le réseau d'échantillonnage de Z-base

$$\left( \frac{(1,2,1)}{6}, \frac{(1,1,1)}{2}, \frac{(0,0,1)}{1} \right)$$

est de rang 2, d'invariants 6,2,1. Ses projections en dimension 2 sont d'invariants 6,2 et celles en dimension 1 sont d'invariant 6. Il est donc à projections régulières.

En revanche, le réseau de Z-base

$$\left( \frac{(3,1,1)}{6}, \frac{(1,1,0)}{2}, \frac{(0,0,1)}{1} \right)$$

n'est pas à projections régulières. En effet, il est de rang 2, d'invariants 6,2,1 or sa projection en dimension 2 sur les deux premiers facteurs est uniquement d'ordre et d'invariant 6, alors qu'il faudrait qu'elle soit d'ordre 12 et d'invariants 6 et 2.

Un résultat sous forme matricielle (Sloan et Joe (1994)) nous permet de repérer plus rapidement si un réseau est oui ou non à projections régulières.

### Proposition 2.26

*Soit  $Q$  la matrice carrée d'ordre  $d$  dont les lignes sont formées des vecteurs  $a_1, \dots, a_d$  de la  $Z$ -base canonique complétée d'un réseau. Alors le réseau est à projections régulières si et seulement si la matrice  $Q$  peut s'écrire sous la forme d'une matrice triangulaire supérieure unitaire (i.e des 1 sur la diagonale).*

Il faut faire attention de ne pas confondre la matrice  $Q$  et la matrice  $M$  génératrice du réseau. Cette dernière est formée des vecteurs générateurs divisés par leurs invariants, ce qui n'est pas le cas pour la matrice  $Q$ .

### Exemples 2.27 :

1) Suite de l'exemple 2.25 : Pour le premier réseau, nous avons les vecteurs  $a_1=(1,2,1)$  associé à  $q_1=6$  et  $a_2=(1,1,1)$  associé à  $q_2=2$ . La matrice  $Q$  s'écrit

$$Q = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

En remplaçant  $a_2$  par  $a'_2=a_2-a_1$ , on ne change rien à la structure du réseau mais  $Q$  peut alors se mettre sous la forme d'une matrice triangulaire supérieure unitaire

$$Q = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Ce qui confirme bien le fait que le réseau est à projections régulières.

2) Tous les réseaux de rang 1 sont à projections régulières puisque la matrice  $Q$  s'écrit directement

$$Q = \left( \begin{array}{c|ccc} 1 & \mathbf{g}_2 & \dots & \mathbf{g}_d \\ \hline 0 & & \mathbf{I}_{d-1} & \end{array} \right).$$

## 2.4 Une méthode hybride : les nets et les séquences

La méthode hybride que nous présentons maintenant est basée sur l'utilisation des nets ou des séquences dont nous allons donner la définition. Elle permet de coupler les deux critères de qualité imposés par les méthodes précédentes, *i.e*

- Un échantillonnage à plusieurs degrés avec une répartition uniforme sur les différents degrés.
- Une suite de faible discrédance.

Cette méthode n'est pas utilisée dans la deuxième et troisième parties de ce travail mais nous en donnons un rapide aperçu car elle semble être une perspective intéressante.

### 2.4.1 Les nets (ou séquences) : méthode d'équidistribution

Cette méthode est considérée la plupart du temps comme une méthode d'équidistribution. Les points d'échantillonnage sont construits à partir d'un net ou d'une séquence dont nous allons donner brièvement la définition. Voir les travaux de Sobol' (1967), Faure (1982), Niederreiter (1987) et Mullen *et al.* (1995) pour plus de précisions sur l'existence et la construction de ces structures de points.

#### Définitions 2.28

*Soit  $b \geq 2$  un entier (en pratique  $b$  est toujours un nombre premier ou la puissance d'un nombre premier). On appelle pavé élémentaire en base  $b$ , un pavé de la forme*

$$\prod_{k=1}^d \left[ \frac{a_k}{b^k}, \frac{a_k + 1}{b^k} \right]$$

*avec  $a_k$  et  $l_k$  des entiers positifs ou nuls, et  $a_k < b^k$  pour tout  $k$ .*

*Soit  $t$  un entier positif ou nul. On dit qu'une suite finie de  $N=b^t$  points de  $[0,1]^d$  est un  $(0,t,d)$ -net en base  $b$ , si tout pavé élémentaire en base  $b$  de volume  $b^{-t}$  contient un seul terme de la suite.*

*Soit  $m \leq t$  un entier positif ou nul. Une suite finie de  $N=b^t$  points de  $[0,1]^d$  est un  $(m,t,d)$ -net en base  $b$ , si tout pavé élémentaire en base  $b$  de volume  $b^{m-t}$  contient exactement  $b^m$  points de la suite. Le cas où  $t=m$  revient à dire que tous les points de la suite sont dans  $[0,1]^d$ .*

*Pour tout  $m \geq 0$ , une suite infinie  $x=(x_i)_{i \geq 1}$  de points de  $[0,1]^d$  est une  $(m,d)$ -séquence en base  $b$ , si pour tout  $k \geq 0$  et pour tout  $t \geq m$ , la suite finie  $(x_{kb^t+1}, \dots, x_{(k+1)b^t})$  est un  $(m,t,d)$ -net en base  $b$ .*

L'intérêt de cette structure de points vient du fait qu'elle permet d'obtenir une répartition uniforme sur tous les  $r$ -uplets de coordonnées, pour  $1 \leq r \leq t$ . De plus cette propriété devient plus forte quand le nombre de points de la suite croît, c'est-à-dire quand  $t$  augmente.

Dans certains cas il est possible de faire le lien avec les structures de points vue précédemment.

$$\boxed{(0,1,d)\text{-net} \longleftrightarrow \text{hypercube latin}}$$

En effet, si  $(x_i)_{i=1,\dots,N}$  est une suite formée des lignes d'un hypercube latin de taille  $N$ , alors la suite  $(N^{-1}x_i)_{i=1,\dots,N}$  est une  $(0,1,d)$ -net en base  $N$ . Réciproquement, si  $(x_i)_{i=1,\dots,N}$  est un  $(0,1,d)$ -net en base  $N$  alors  $[Nx_i]$ ,  $i=1,\dots,N$  sont les lignes d'un hypercube latin de taille  $N$ .

$$\boxed{(m,t,d)\text{-net} \longrightarrow \text{OA}([t-m/q])}$$

Dans le cas des tableaux orthogonaux, il n'y a pas la réciproque. Si  $(x_i)_{i=1,\dots,N}$  est un  $(m,t,d)$ -net en base  $b$  alors, pour tout entier positif  $q \leq t-m$ ,  $[b^q x_i]$ ,  $i=1,\dots,N$  sont les lignes du tableau orthogonal  $\text{OA}(N,d,b^q,[t-m/q])$ .

$$\boxed{(0,2,d)\text{-net} \longleftrightarrow \text{hypercube latin de Tang}}$$

En revanche, si la suite  $(x_i)_{i=1,\dots,N}$  est formée des lignes d'un hypercube latin de taille  $N=b^2$  construit par la méthode de Tang à partir d'un  $\text{OA}(b^2,d,b,2)$  alors  $(b^{-2}x_i)_{i=1,\dots,N}$  est un  $(0,2,d)$ -net en base  $b$ , et réciproquement.

On peut aussi remarquer que plus  $m$  est petit, meilleure est la répartition. Faure (1982) donne la construction de  $(0,t,d)$ -net et  $(0,d)$ -séquence en base  $p$  où  $p \geq d$  est un nombre premier, et Niederreiter (1987) étend cette construction en base  $b$  où  $b \geq d$  est une puissance d'un nombre premier.

### Discrépance

Niederreiter (1987) fait un bilan des résultats établis sur la discrétion de ces suites, que nous résumons dans la proposition suivante.

#### **Proposition 2.29**

*Pour  $b \geq d$ , un  $(m,t,d)$ -net a une discrétion de l'ordre de  $O(N^{-1}(\log N)^{d-1})$ , et une  $(m,d)$ -séquence a une discrétion de l'ordre de  $O(N^{-1}(\log N)^d)$ .*

Ce résultat n'est pas surprenant, en effet, nous avons émis l'hypothèse que plus la force d'un tableau orthogonal est élevée, mieux ses points sont répartis dans le cube unité. On peut donc supposer que les contraintes marginales imposées aux nets ou séquences entraînent une répartition uniforme des points d'échantillonnage dans  $[0,1]^d$ , donc une faible fonction de discrédance. Ce qui est confirmé par la proposition.

## 2.4.2 Les nets (ou séquences) : méthode de réduction de la variance

Ce qu'il y a de tout à fait intéressant avec cette structure de points est que si l'on trouve une randomisation des points (*i.e* qui laisse la structure inchangée) alors d'après la proposition 2.29 la suite randomisée garde une discrédance de l'ordre de  $O(N^{-1}(\log N)^{d-1})$ . On peut donc majorer l'erreur d'intégration (de façon déterministe) grâce à l'inégalité de Koksma-Hlawka et l'estimer sans biais avec la randomisation.

### Randomisation

La randomisation a été proposée par Owen (1995). Soit  $(x_i)_i$  un  $(m,t,d)$ -net ou une  $(m,d)$ -séquence. Le  $i$ ème terme de la suite se note  $x_i=(x_{i1},\dots,x_{id})$  et chaque composante s'écrit en base  $b$

$$x_{ij} = \sum_{k=1}^{\infty} x_{ijk} b^{-k}$$

où  $0 \leq x_{ij} < b$ . Une randomisation de la suite  $(x_i)_i$  est une suite  $(X_i)_i$  d'éléments  $X_i=(X_{i1},\dots,X_{id})$  de

la forme  $X_{ij} = \sum_{k=1}^{\infty} X_{ijk} b^{-k}$  avec  $X_{ijk}$  une permutation aléatoire des  $x_{ijk}$  définie par

$$\begin{aligned} X_{ij1} &= \pi_j(x_{ij1}) \\ X_{ij2} &= \pi_{jx_{ij1}}(x_{ij2}) \\ &\dots \\ X_{ijk} &= \pi_{jx_{ij1}\dots x_{ijk-1}}(x_{ijk}) \end{aligned}$$

où chaque  $\pi$  est une permutation aléatoire de  $0,\dots,b-1$  uniformément distribuée parmi le  $b!$  permutations possibles. Elles sont indépendantes entre elles. Par exemple si on est en base 10 et si  $x_{ij}=0,123$ , on a  $x_{ij1}=1$ ,  $x_{ij2}=2$  et  $x_{ij3}=3$ , d'où

$$\begin{aligned} X_{ij1} &= \pi_j(x_{ij1}) = \pi_j(1) \\ X_{ij2} &= \pi_{jx_{ij1}}(x_{ij2}) = \pi_{j1}(2) \\ X_{ij3} &= \pi_{jx_{ij1}x_{ij2}}(x_{ij3}) = \pi_{j12}(3) \end{aligned}$$

Si l'on choisit les permutations  $\pi_j : \{0 \rightarrow 1 \rightarrow 2 \dots \rightarrow 9\}$ ;  $\pi_{j1} : \{0 \rightarrow 2, 1 \rightarrow 3, \dots, 9 \rightarrow 1\}$  et  $\pi_{j12} : \{0 \rightarrow 3, 1 \rightarrow 4, \dots, 9 \rightarrow 2\}$  alors on obtient  $X_{ij}=0,246$ .

### Proposition 2.30

*Si  $(x_i)_i$  est un  $(m,t,d)$ -net (resp. une  $(m,d)$ -séquence) en base  $b$  alors  $(X_i)_i$  est un  $(m,t,d)$ -net (resp. une  $(m,d)$ -séquence) en base  $b$  avec une probabilité de 1.*

*Soit  $x$  un point de  $[0,1]^d$  et  $X$  sa randomisation, alors  $X$  a une distribution uniforme sur  $[0,1]^d$ .*

Ces résultats ont été démontrés par Owen (1995). Il prouve que la randomisation est bien adaptée à l'intégration numérique puisqu'elle conserve la structure des points et l'estimateur est sans biais.

#### Etude de la variance

On note  $\hat{I}_N$  l'estimateur utilisant un échantillon construit à partir d'un net randomisé, et  $\hat{I}_S$  l'estimateur utilisant un échantillon construit à partir d'une séquence randomisée. Si la fonction est à dérivées partielles bornées, alors d'après la proposition précédente et le paragraphe 2.3.1, la variance est de l'ordre de  $O(N^{-2}(\log N)^{2d-2})$ . Dans le cas des fonctions de carré intégrable, Owen a démontré un résultat qui permet de comparer la variance de  $\hat{I}_N$  à celle de  $\hat{I}_{MC}$ . Pour cela, il n'utilise pas la décomposition croisée de la variance présentée dans le paragraphe 2.2.2, mais une décomposition qui tient compte de la structure en arbre des nets. En dimension un, dans le cas d'un  $(0,t,1)$ -net, on a

$$\text{var } \hat{I}_N \leq \text{var } \hat{I}_{MC},$$

et en dimension supérieure pour un  $(0,t,d)$ -net en base  $b \geq d$ ,

$$\text{var } \hat{I}_N \leq \left(\frac{b}{b-1}\right)^{\min(t,d-1)} \text{var } \hat{I}_{MC}$$

Il fait remarquer aussi que  $\text{var } \hat{I}_N$  tend vers 0 quand  $N$  tend vers l'infini. Ce comportement est bien meilleur que celui de la méthode de Monte Carlo où  $N \text{var } \hat{I}_{MC} = \sigma^2$ , ainsi que celui des méthodes basées sur les tableaux orthogonaux, qui tout en conservant une vitesse de convergence de  $N^{-1}$ , ne font que réduire  $\sigma^2$ .

Les nets et les séquences semblent être des structures ayant de bonnes propriétés pour construire des points d'échantillonnage, puisqu'ils allient répartition marginale uniforme et faible discrèpance. Bien que très peu utilisés jusqu'à présent, ils semblent offrir une perspective intéressante.

## Bilan de l'approche par échantillon aléatoire

Cette approche est assez peu réaliste en ce qui concerne le modèle. Bien sûr celui-ci est très simple et ne dépend pas d'une distribution *a priori* comme le précédent avec le processus résiduel. Mais, il n'est pas raisonnable de supposer que la réponse du simulateur peut se décomposer sur une base de fonctions et que seuls les termes prédominants de cette décomposition suffisent à la modéliser.

En revanche, la technique d'échantillonnage à deux degrés semble être la partie intéressante de cette approche. Elle permet d'imposer des conditions aux plans avec le premier degré tout en ayant un échantillon aléatoire. Les plans ainsi construits vérifient pratiquement les quatre principaux objectifs fixés en introduction.

1. Aucune contrainte n'est imposée aux points du plans pour qu'il y ait existence de l'estimateur. Donc les plans présentés dans ce chapitre permettent d'analyser n'importe quels modèles de cette approche. Cependant, comme nous l'avons signalé, ces modèles sont très pauvres, il serait donc plus intéressant de savoir si les plans utilisés en intégration numérique permettent d'analyser de façon efficace le modèle de l'approche par résidu aléatoire. Des tests numériques ont été effectués à cet effet. Ils sont présentés dans le chapitre 5 et ne concernent que les tableaux orthogonaux.
2. Les seules contraintes que l'on impose aux plans concernent leur distribution dans le domaine d'échantillonnage. On suppose que plus les points sont répartis uniformément, plus l'estimation est bonne. Cette hypothèse semble être confirmée par quelques résultats sur la variance de l'estimateur, malheureusement ceux-ci sont asymptotiques.
3. Ces plans sont construits de façon à ce que les points d'échantillonnage soient répartis uniformément dans le cube unité. C'est l'objectif principal des techniques d'échantillonnage utilisées en intégration numérique, seules les méthodes pour "mesurer" l'uniformité varient.
4. Ces plans pour la plupart sont construits à partir de générateurs. Numériquement, il n'y a pas de problème de coût de construction, notamment dans le cas des hypercubes latins et des réseaux de rang 1. Ce qui pose problème d'un point de vue théorique, est la détermination des générateurs. Dans le cas des réseaux, nous avons vu comment ils étaient construits. En ce qui concerne les tableaux orthogonaux, des techniques de construction sont étudiées dans la troisième partie.

## Bibliographie

- Bates R.A., Riccomagno E., Schwabe R, Wynn H.P. (1998). Lattices and dual lattices in optimal experimental design for Fourier models. *Comp. Statist. and data Analysis* **28** N°3, 283-296.
- Braaten E., Weller G. (1979). An improved low-discrepancy sequence for multidimensionnal quasi-Monte Carlo integration. *J. Comp. Physics* **33**, 249-258.
- Cassels J.W.S. (1971). *An introduction to the geommetry of numbers*. Springer Verlag, Berlin.
- Drmotá M., Tichý R.F. (1997). *Sequences, discrepancies and applications*. Lecture Notes in Mathematics 1651. Springer Verlag, Berlin.
- Efron B., Stein C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9**, 586-596.
- Fang K.T., Wang Y. (1993). *Number-theoretic methods in statistics*. Chapman and Hall, London.
- Fang K.T., Wang Y., Bentler P. (1994). Some applications of number-theoretic methods in statistics. *Statist. Sci.* **9**, 416-428.
- Fang K.T., Zhu L.X. (1993). On uniformly random design. *Technical report MATH-020*, Hong Kong Baptist College.
- Faure H. (1982). Discrépance de suites associées à un système de numération (en dimension  $s$ ). *Acta. Arith.* **41**, 337-351.
- Halton J.H. (1960). On efficiency of certain quasi-random sequences of points in evaluating multidimensionnal integrals. *Numerische Mathematik* **2**, 84-90.
- Hedayat A.S., Rao C.R., Stufken J. (1988). Sampling plans excluding contiguous units. *J. of Statist. Planning and Inference.* **19**, 159-170.
- Hlawka E. (1964). Discrepancy and uniform distribution of sequences. *Composition Math.* **16**, 83-91.
- Hua L.K., Wang Y. (1981). *Applications of number theoretic to numerical analysis*. Springer Verlag, Berlin.
- Jensen R.J. (1975). Square and cubic lattice sampling. *Biometrics* **31**, 449-471.
- Koehler JR., Owen A.B. (1996). Computer experiments. In Ghosh and C. Rao, editors, *Handbook of statistics, 13 : Design and analysis of experiments*.
- Koksma J.F. (1942/43). Een algemeene stelling uit de theorie der gelijkmatige verdeling modulo 1. *Mathematica B (Zutphen)* **11**, 7-11.
- Loh W.C. (1993). On latin hypercube sampling. *Technical Report N° 93-52*, dept. Of Stititics, Purdue University.
- Loh W.C. (1996). A combinatorial central limit theorem for randomized orthogonal array sampling designs. *Ann. Statist.* **24**, 1209-1224.
- Mc Kay M.D., Beckman R.J., Conover W.J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239-245.

- Mullen G.L., Mahalanabis A., Niederreiter H. (1995). Tables of (t,m,s)-net and (t,s)-sequence parameters. In Niederreiter, H. and Shiue, P. J.-S., editors, *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pp 299-317, New-York. Springer Verlag.
- Niederreiter H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.* **84**, 957-1041.
- Niederreiter H. (1987). Point sets and sequences with small discrepancy. *Monatsh. Math.* **104**, 273-337.
- Niederreiter H. (1992). *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF, SIAM, Philadelphia.
- Owen A.B. (1992). A central limit theorem for latin hypercube sampling. *J. Roy. Statist. Soc. B* **54**, 541-551.
- Owen A.B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2**, 439-452.
- Owen A.B. (1994). Lattice sampling revisited : Monte Carlo variance of means over randomized orthogonal arrays. *Ann. Statist.* **22**, 930-945.
- Owen A.B. (1995). Randomly permuted (t,m,s)-nets and (t,s)-sequences. In Niederreiter, H. and Shiue, P. J.-S., editors, *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pp 299-317, New-York. Springer Verlag.
- Owen A.B. (1997). Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis* **34(5)**. In press.
- Pagès G., Xiao Y.J. (1991). Sequence with low discrepancy and pseudo-random numbers. *Technical Report*, Laboratoire de mathématiques et de modélisation, Paris.
- Patterson H.D. (1954). The error of lattice sampling. *J. Roy. Statist. Soc. B* **16**, 140-149.
- Sloan I.H., Joe S. (1994). *Lattice methods for multiple integration*. Oxford science Publications.
- Riccomagno E., Schwabe R., Wynn H.P. (1997). Lattice-based D-optimum design for Fourier regression. *Ann. Statist.*
- Sloan I.H., Kachoyan P.J. (1987). Lattice methods for multiple integration : theory, error analysis and examples. *SIAM Journal on numerical Analysis* **24**, 116-128.
- Sloan I.H., Lyness J.N. (1989). The representation of lattice quadrature rules as multiple sums. *Math. of computation* **52**, 81-94.
- Sloan I.H., Lyness J.N. (1990). Lattice rules : projection regularity and unique representation. *Math. of computation* **54**, 649-660.
- Sloan I.H., Walsh L. (1990). A computer search of rank 2 lattice rules for multidimensional quadrature. *Math. of Computation* **54**, 281-302.
- Sobol' I.M. (1967). The distribution of points in a cube and the accurate evaluation of integrals (in russian). *Zh. Vychisl. Mat. i Mat. Phys.* **7**, 784-802.
- Stein M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics* **29**, 143-151.
- Tang B. (1993). Orthogonal array-based latin hypercubes. *J. Amer. Statist. Assoc.* **88**, 1392-1397.

- Tang B. (1994). A theorem for selecting OA-based latin hypercubes using a distance criterion. *Comm. Statist.-Theory Meth.* **23**, 2047-2058.
- Tocher K.D. (1963). *The art of simulation*. Princeton, N.J. D Van Nostrand. pp.106-107.
- Wang Y., Fang K.T. (1990a). Number theoretic methods applied statistics. *Chin. Ann. Math. Ser. B* **11**, 41-55.
- Wang Y., Fang K.T. (1990b). Number theoretic methods applied statistics (II). *Chin. Ann. Math. Ser. B* **11**, 859-914.
- Weber N.C. (1980). A martingale approach to central limit theorems for exchangeable random variables. *J. Appl. Prob.* **17**, 662-673.



# Deuxième partie

*Une nouvelle approche statistique des  
expériences simulées*



# Introduction

A l'issue de cette première partie, nous pouvons reconnaître que chacune des approches présente des qualités.

Il semble clair qu'en ce qui concerne la modélisation, il est préférable d'utiliser la méthode par résidu aléatoire. L'idée d'introduire un processus résiduel est intéressante, si toutefois on arrive à diminuer sa contribution dans le modèle, afin d'éviter les problèmes de robustesse du prédicteur.

En revanche, pour ce qui est de la planification, les méthodes d'échantillonnage de l'approche par échantillon aléatoire sont plus performantes. Nous avons notamment constaté que les critères de qualité des plans ne dépendent pas du modèle utilisé puisqu'ils sont basés sur la notion de répartition des points dans le cube unité.

De ces deux approches nous retenons les deux idées suivantes

- le processus résiduel pour la partie modélisation,
- l'échantillonnage à deux degrés pour la planification, avec une contrainte de bonne répartition des points pour le premier degré.

Dans le chapitre 3, nous allons proposer un nouveau modèle. Celui-ci doit donc tenir compte de la corrélation entre les points mais à la différence de l'approche par résidu aléatoire, il doit aussi

- permettre à la partie fixe du modèle de contre balancer le poids du processus résiduel,
- tenir compte de l'effet du deuxième degré de l'échantillonnage, puisque celui-ci est aléatoire.

Dans le chapitre 4, nous allons définir la partie fixe du modèle. En effet, celle-ci jouant un rôle plus important, il est nécessaire de la choisir avec soin. Nous proposons donc une régression trigonométrique ou modèle de Fourier.

Enfin dans le chapitre 5, nous définissons les plans utilisés pour le premier degré de l'échantillonnage. L'objectif étant toujours le même que dans le chapitre deux, c'est-à-dire une bonne répartition des points dans le cube unité. Toutefois nous allons enrichir la structure des plans de l'approche par échantillon aléatoire de façon à ce qu'ils soient bien adaptés à l'ajustement d'une régression trigonométrique.

En résumé, nous proposons une nouvelle méthode qui associe les avantages des approches par résidu et échantillon aléatoires, tout en atténuant leurs inconvénients.

# 3.Nouvelle approche :

## Le modèle

Dans le premier chapitre nous avons vu que le modèle proposé dans l'approche par résidu aléatoire est particulièrement bien adapté au problème des expériences simulées. Il semble que le processus résiduel soit la partie déterminante pour avoir une prédiction de qualité. Toutefois, lorsque le modèle repose uniquement sur le processus, cela engendre quelques problèmes, notamment de robustesse du prédicteur. Nous allons voir comment contre-balancer le rôle du processus résiduel de façon à ce que toutes les parties d modèle aient la même importance. Ce qui permet de ne plus baser entièrement la prédiction sur la distribution *a priori* que l'on donne au processus.

Dans le chapitre deux, nous avons défini une méthode d'échantillonnage à deux degrés qui permet d'imposer une bonne répartition des points du plan dans le cube unité. Nous allons utiliser cette technique dans l'approche que nous proposons. Il est donc nécessaire que le modèle tienne compte de l'effet du deuxième degré, qui rappelons-le, n'est pas contrôlable, puisqu'il correspond à une perturbation aléatoire.

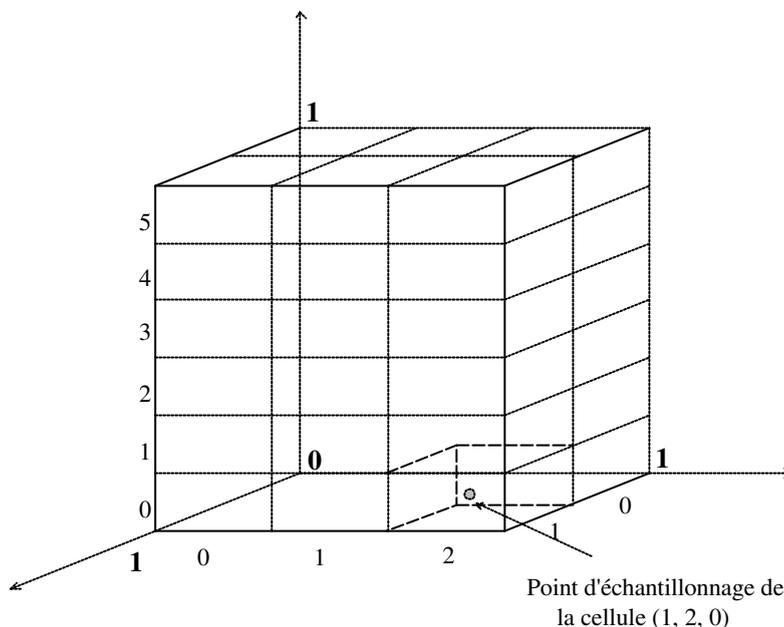
En résumé, le modèle que nous allons présenter, est construit de façon à

- Contre-balancer le poids du processus résiduel dans le modèle,
- tenir compte du deuxième degré de l'échantillonnage.

A la fin de ce chapitre, nous étudions le comportement de ce nouveau modèle, notamment le rôle joué par chaque parties, et la différence avec le modèle du chapitre un.

### 3.1. Présentation du modèle

Rappelons brièvement en quoi consiste l'échantillonnage à deux degrés défini dans le chapitre 2.2.



On découpe en  $q$  segments de même longueur chaque arête du cube unité, de façon à obtenir une partition du domaine d'échantillonnage en cellules, indexées par les éléments du groupe  $G = \times_{i \in I} (Z/q_i)$ .

**1<sup>er</sup> degré :** choix des cellules représentées dans le plan

**2<sup>ème</sup> degré :** tirage au hasard d'un point dans chaque cellule sélectionnée au 1<sup>er</sup> degré, ce point définissant les paramètres d'une simulation à réaliser.

Notons  $C_x$  toute cellule du premier degré, où  $x$  est le point de  $[0,1]^d$  représentant la cellule, et  $X$  le vecteur aléatoire appartenant à  $C_x$ , obtenu au deuxième degré. Les représentants des cellules sont choisis régulièrement espacés, soit au bord des cellules, soit au centre, *i.e*

$$x_j = \frac{g_j}{q_j} \text{ ou } x_j = \frac{g_j + 0,5}{q_j}, j=1, \dots, d,$$

où  $g=(g_1, \dots, g_d) \in G$ .

Le modèle doit donc tenir compte des deux degrés de l'échantillonnage bien que seul le premier degré soit contrôlable en fait. Une possibilité consiste à introduire artificiellement un processus indexé par les représentants des cellules, décomposé en trois termes

$$Y(x) = Z(x)\beta + \Gamma(x) + e(x).$$

- $Z(x)\beta = E[Y(x)]$  est la partie certaine du modèle, précisant l'approximation retenue pour la prédiction en moyenne.

- $\Gamma(x)$  est un processus aléatoire, introduit pour tenir compte du résidu de l'approximation. De même que dans le chapitre 1, il sert à corriger la prédiction en moyenne, il est donc à corrélations spatiales.
- $e(x)$  est un terme d'erreur, indépendant de  $\Gamma(x)$ , introduit pour prendre en compte l'effet du deuxième degré de l'échantillonnage.

### Définition 3.1

*Au point  $X \in C_x$ , la réponse du simulateur  $y(X)$  est modélisée par*

$$Y(x) = Z(x)\beta + \Gamma(x) + e(x).$$

•  $Z(x)\beta$  est la régression linéaire où  $Z(x)$  est le vecteur du modèle d'ordre  $1 \times m$  et  $\beta$  est le vecteur inconnu des paramètres du modèle d'ordre  $m \times 1$ .

•  $\Gamma(x)$  est le processus résiduel, tel que  $\forall x \in [0,1]^d, \forall u \in [0,1]^d$

$$E[\Gamma(x)] = 0 \text{ et } \text{cov}(Y(x), Y(u)) = \sigma^2 R(x, u).$$

•  $e(x)$  est le terme d'erreur indépendant de  $\Gamma(x)$ , tel que  $\forall x \in [0,1]^d, \forall u \in [0,1]^d$ .

$$E[e(x)] = 0 \text{ et } \text{cov}(e(x), e(u)) = \sigma_e^2 \delta_{xu}.$$

On raisonne donc ici comme si on procédait à un "recalage" des observations, c'est-à-dire en considérant la réponse du simulateur  $y(X)$ , au point  $X \in C_x$ , comme la réalisation de  $Y(x)$  bien que  $X \neq x$  presque sûrement.

Ceci se justifie par le fait que la plan d'échantillonnage consiste ici en un sous-ensemble de cellules, représentées par des points choisis *a priori*, mais qui ne fixe pas les paramètres des simulations à réaliser, puisque le deuxième degré n'est pas contrôlable. Il les contraint uniquement à appartenir à certaines cellules. Au demeurant, il est tout à fait possible d'envisager que seules soient connues pour la prédiction les cellules  $C_x$  et les réponses du simulateurs au point  $X \in C_x$ , mais pas les paramètres des simulations. Par exemple, dans le cas où les valeurs de certains facteurs ne peuvent pas être fixées avec précision.

C'est donc un modèle avec terme d'erreur qui intervient ici. Celui-ci n'est pas nouveau, puisqu'il est utilisé dans le cadre des expériences simulées, notamment par Sacks, Schiller et Welch (1989). Seulement leur motivation est tout autre que la nôtre, puisqu'elle vient de la nature de ce que l'on cherche à prédire. Un prédicteur sans terme d'erreur interpole les observations, il va donc prédire la réponse du simulateur et non pas le phénomène simulé. En effet, on peut considérer que cette réponse est entachée d'une erreur par rapport au phénomène simulé, due, d'une part aux arrondis machine, et d'autre part aux simplifications de la modélisation. Ils introduisent en fait une *erreur de mesure* pour prédire le phénomène simulé. De notre côté, le terme d'erreur représente l'effet du deuxième degré de l'échantillonnage, mais on peut imaginer qu'il englobe aussi une erreur de mesure.

## 3.2. Prédiction

Le prédicteur de  $Y(x)$  se construit très simplement en se ramenant au modèle du chapitre un, c'est-à-dire une fonction aléatoire en deux parties : une régression linéaire et un processus aléatoire. Pour cela on pose

$$\varepsilon(x) = \Gamma(x) + e(x), \quad \forall x \in [0,1]^d.$$

Le modèle s'écrit alors,

$$Y(x) = Z(x)\beta + \varepsilon(x),$$

où  $\varepsilon$  est un vecteur aléatoire centré, de variance  $\sigma^2 + \sigma_e^2$ , et de covariance

$$\text{cov}(Y(x), Y(u)) = \sigma_e^2 V(x, u) = \begin{cases} \sigma_e^2(1 + \eta^2 R(x, u)) & x = u \\ \sigma_e^2 \eta^2 R(x, u) & x \neq u \end{cases}.$$

$\eta^2 = \sigma^2 / \sigma_e^2$  est le coefficient de contribution relative de la corrélation.

Soit un plan  $D = \{x_1, \dots, x_N\}$ , les deux propositions suivantes donnent la forme du prédicteur de  $Y(x)$  ainsi que l'erreur quadratique moyenne.

### Proposition 3.2

*Si le vecteur des paramètres est estimable, alors le meilleur prédicteur linéaire sans biais de  $Y(x)$  est de la forme*

$$\hat{Y}(x) = X(x)\hat{\beta} + {}^t r(x) V^{-1} (Y - X\hat{\beta}),$$

*avec  $\hat{\beta} = ({}^t X V^{-1} X)^{-1} {}^t X V^{-1} Y$  l'estimateur de Gauss-Markov de  $\beta$ , où  $V$  est la matrice carrée d'ordre  $N$  à éléments  $V(x_i, x_j)$ .*

**Remarque :** La covariance entre  $Y(x)$  et le vecteur des observations est toujours égale à  ${}^t r(x)$  puisque pour tout  $x_i \in D$ ,  $e(x_i)$  est non corrélé avec  $Y(x)$  par hypothèse. Donc contrairement au modèle du chapitre un,  ${}^t r(x_i) V^{-1}$  ne se simplifie pas, autrement dit **le prédicteur ne passe pas par les observations**,  $\hat{Y}(x_i) \neq Y(x_i)$ .

**Démonstration :** Reprenons les démonstrations des propositions 1.5 et 1.6 mais cette fois avec

$$\begin{aligned} \text{cov}(Y) &= \sigma_e^2 V \\ \text{cov}(Y, Y(x)) &= \sigma^2 {}^t r(x) \end{aligned}$$

En effet, pour tout  $x_i \in D$ ,  $e(x_i)$  est non corrélé avec  $Y(x)$  par hypothèse.

1) Dans un premier temps supposons  $\beta$  connu. On peut écrire le meilleur prédicteur linéaire sans biais de  $Y(x)$  sous la forme

$$\bar{Y}(x) = \underbrace{a}_{\in \mathbb{R}} + \underbrace{b}_{1 \times N} (Y - Z\beta).$$

Le meilleur choix pour  $a$  se fait de la même façon que pour la proposition 1.5, i.e

$$a = Z(x)\beta.$$

Montrons que le meilleur choix pour  $b$  est le vecteur  $\bar{b}$  solution de  $bV = \eta^2 {}^t r(x)$ .

$$\begin{aligned} E[Y(x) - Z(x)\beta - b(Y - Z\beta)]^2 &= E[Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta)]^2 + E[(\bar{b} - b)(Y - Z\beta)]^2 \\ &\quad + 2E[(Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta))(\bar{b} - b)(Y - Z\beta)]. \end{aligned}$$

Or puisque  $E[Y - Z\beta] = 0$ , on a

$$\begin{aligned} E[(Y(x) - Z(x)\beta - \bar{b}(Y - Z\beta))(\bar{b} - b)(Y - Z\beta)] &= (\bar{b} - b)\{E[Y(x)Y] - E[Y(x)Z\beta] - \bar{b} \text{cov}(Y)\} \\ &= (\bar{b} - b)\{\text{cov}(Y, Y(x)) - E(Y)E[Y(x)] - Z\beta E[Y(x)] - \bar{b} \text{cov}(Y)\} \\ &= (\bar{b} - b)\{\text{cov}(Y, Y(x)) - \bar{b} \text{cov}(Y)\} \\ &= (\bar{b} - b)(\sigma^2 {}^t r(x) - \eta^2 \sigma_e^2 {}^t r(x)) = 0 \end{aligned}$$

Le choix de  $\bar{b}$  minimise bien l'erreur quadratique moyenne.

Ce prédicteur peut encore s'écrire sous la forme

$$\bar{Y}(x) = \tau + \eta^2 {}^t r(x)V^{-1}Y$$

où  $\tau = \tau(x) = K(x)\beta$  avec  $K(x) = [Z(x) - \eta^2 {}^t r(x)V^{-1}Z]$ .

2)  $\beta$  étant le vecteur inconnu des paramètres du modèle,  $\bar{Y}(x)$  n'est pas connu. Dans le cas où  $\tau$  est estimable, on le remplace par un estimateur linéaire sans biais,  $\hat{\tau}$  et on note  $\hat{Y}(x)$  le nouveau prédicteur. Alors on montre de la même façon que pour la proposition 1.6, que

$$E[Y(x) - \hat{Y}(x)]^2 = \text{var}(\hat{\tau}) + \text{var}(\eta^2 {}^t r(x)V^{-1}Y - Y(x)).$$

Autrement dit que le prédicteur linéaire sans biais  $\hat{Y}(x)$  minimise l'erreur quadratique moyenne si  $\hat{\tau}$  est de variance minimale, c'est-à-dire si  $\hat{\tau}$  est de la forme  $K(x)\hat{\beta}$  où  $\hat{\beta}$  est l'estimateur de Gauss-Markov de  $\beta$ .  $\square$

### Proposition 3.3

*L'erreur quadratique moyenne du meilleur prédicteur linéaire sans biais de  $Y(x)$  est de la forme,  $\forall [0, I]^d$*

$$MSE(x) = E[Y(x) - \hat{Y}(x)]^2 = \sigma_e^2 \left\{ 1 + K(x)({}^t ZV^{-1}Z)^{-1} {}^t K(x) + \eta^2 - \eta^4 {}^t r(x)V^{-1}r(x) \right\}$$

où  $K(x) = [Z(x) - \eta^2 {}^t r(x)V^{-1}Z]$ .

**Démonstration :** D'après la démonstration précédente

$$\begin{aligned}
 E[Y(x) - \hat{Y}(x)]^2 &= \text{var}(\hat{\tau}) + \text{var}(\eta^2 {}^t r(x) V^{-1} Y - Y(x)) \\
 &= K(x) ({}^t Z V^{-1} Z)^{-1} {}^t Z V^{-1} \text{cov}(Y) V^{-1} Z ({}^t Z V^{-1} Z)^{-1} K(x) + \eta^4 {}^t r(x) V^{-1} \text{cov}(Y) V^{-1} r(x) \\
 &\quad + \text{var}(Y(x)) - \eta^2 {}^t r(x) V^{-1} \text{cov}(Y(x), Y) - \eta^2 \text{cov}(Y, Y(x)) V^{-1} r(x) \\
 &= \sigma_e^2 K(x) ({}^t Z V^{-1} Z)^{-1} K(x) + \eta^4 \sigma_e^2 {}^t r(x) V^{-1} r(x) + (\sigma_e^2 + \sigma^2) - 2 \eta^2 \sigma^2 {}^t r(x) V^{-1} r(x) \\
 &= \sigma_e^2 \left\{ 1 + K(x) ({}^t Z V^{-1} Z)^{-1} K(x) + \eta^2 - \eta^4 {}^t r(x) V^{-1} r(x) \right\}
 \end{aligned}$$

### 3.3. Comportement du modèle

Le but de ce paragraphe est de savoir en quoi ce modèle se comporte différemment de celui du chapitre un et surtout quelle est la contribution de chacun des trois termes du modèle.

Il est illustré par des simulations numériques en dimension 1. Les points du plan utilisés sont obtenus par un échantillonnage à deux degrés, où le premier degré correspond à 30 points également espacés sur  $[0,1]$ . Les observations (réponses présumées d'un simulateur à un paramètre d'entrée) aux points du plan sont choisies de façon aléatoire entre 0 et 10, et ne correspondent à aucun phénomène particulier. Nous utilisons pratiquement toujours les mêmes plan, observations et régression linéaire constante ( $Z(x)=1$ ), dans le cas contraire, nous le précisons.

#### 3.3.1. Comportement de la partie aléatoire du modèle

La particularité de ce modèle vient de la matrice  $V$  qui fait intervenir conjointement le terme d'erreur et le processus résiduel dans la prédiction *via* la quantité  $\eta^2 R_\theta(\|x-u\|^2)$ .

$\eta^2 = \sigma^2 / \sigma_e^2$  représente la contribution de l'erreur dans le modèle, plus  $\eta^2$  est petit, plus l'erreur a un rôle important. On suppose toutefois que  $\eta > 1$ , c'est-à-dire que la contribution de l'erreur est relativement faible par rapport au processus résiduel.

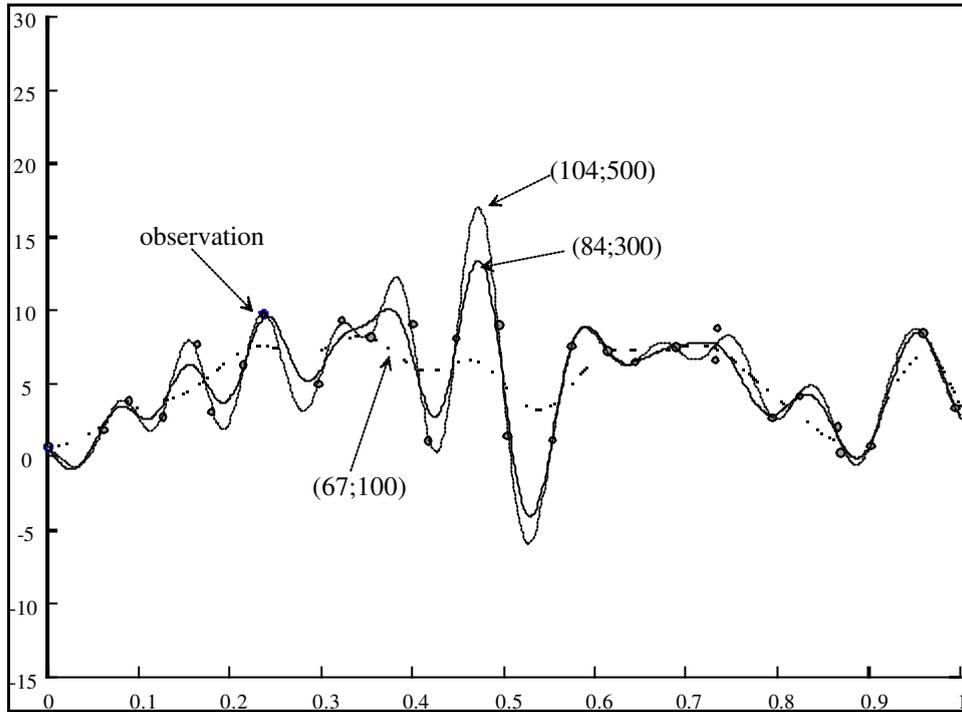
Rappelons que  $\theta$  est le coefficient de corrélation, plus il est petit, plus les corrélations sont fortes entre les points du plan.

#### Etude conjointe des deux paramètres

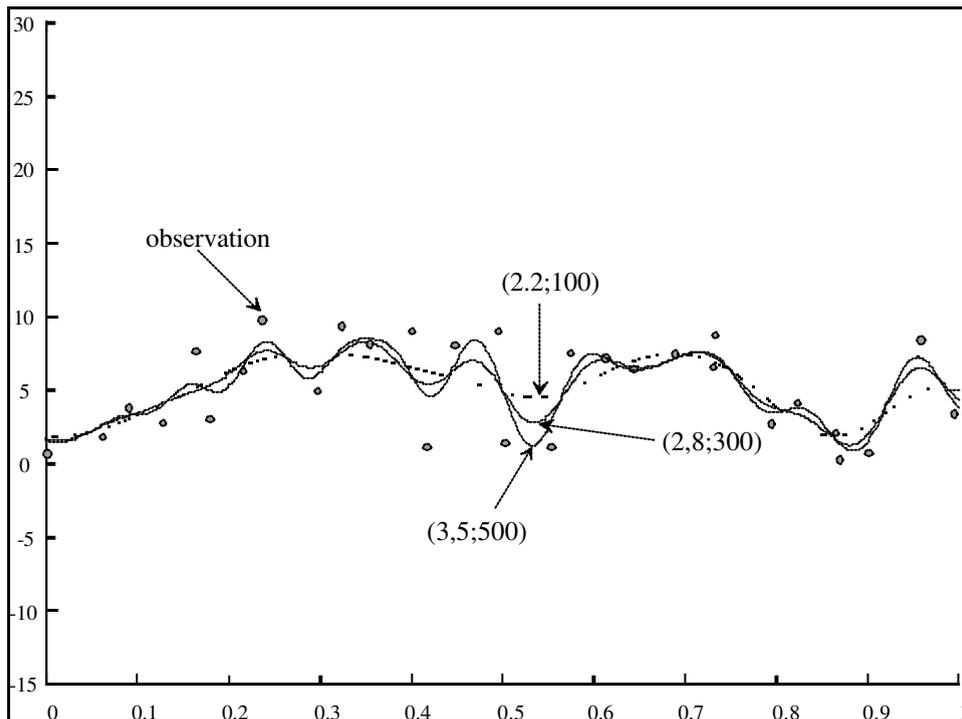
Pour cette étude, on fixe un ordre de grandeur pour  $\|x-u\|^2$ , par exemple  $N^2$ , la distance minimale de deux points du premier degré de l'échantillonnage. On note

$$g(\eta^2, \theta) = \eta^2 R_\theta\left(\frac{1}{N^2}\right).$$

La matrice  $V$  dépend uniquement de cette quantité. D'un point de vue calculatoire, on peut supposer que si celle-ci est fixée, le prédicteur aura le même comportement quelles que soient  $\theta$  et  $\eta^2$  utilisées pour obtenir la valeur fixée. Comme on peut le voir sur les graphes ci-dessous, ce n'est bien évidemment pas le cas. C'est la contribution respective des deux paramètres qui influence le comportement du prédicteur. Il faut donc les étudier séparément.



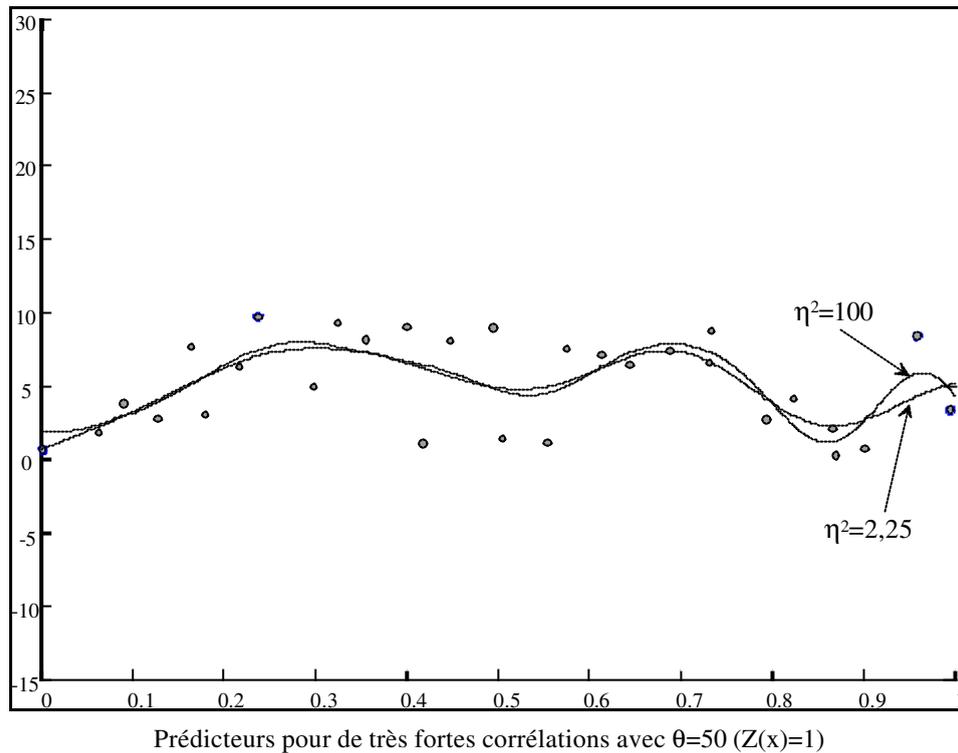
Prédicteurs pour  $g(\eta^2, \theta)=2$  avec différentes valeurs de  $(\eta^2, \theta)$  et forte contribution de l'erreur ( $Z(x)=1$ )



Prédicteurs pour  $g(\eta^2, \theta)=60$  avec différentes valeurs de  $(\eta^2, \theta)$  et faible contribution de l'erreur ( $Z(x)=1$ )

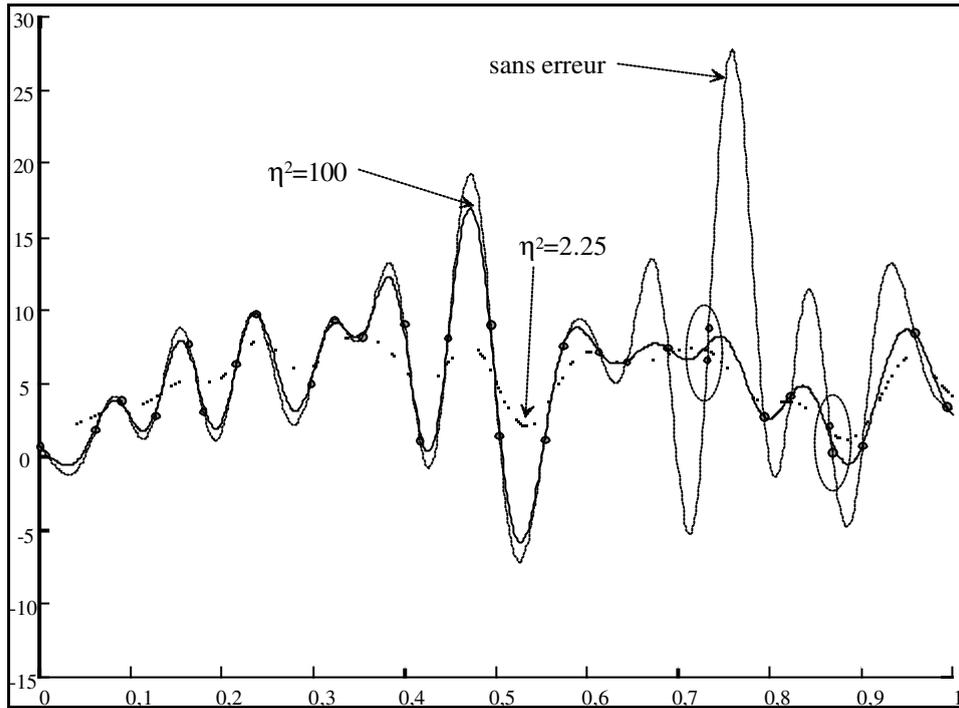
La matrice  $V$  est toujours bien conditionnée. Cela vient de sa forme qui n'est autre que la matrice  $\eta^2 R$  à laquelle on ajoute 1 sur la diagonale. Il est alors possible d'introduire dans le modèle de très fortes corrélations ( $\theta$  très petit) qui n'étaient pas numériquement envisageable avec le modèle sans terme d'erreur (cf. 1.3.1). En fait, on constate que même si c'est possible, le prédicteur n'est pas du tout efficace. Par exemple, ci-dessous, nous avons choisi  $\theta=50$ , il est alors numériquement impossible d'obtenir le prédicteur pour le modèle du chapitre un. En

revanche, la prédiction est envisageable dans le cas du modèle avec terme d'erreur, mais elle est peu intéressante.



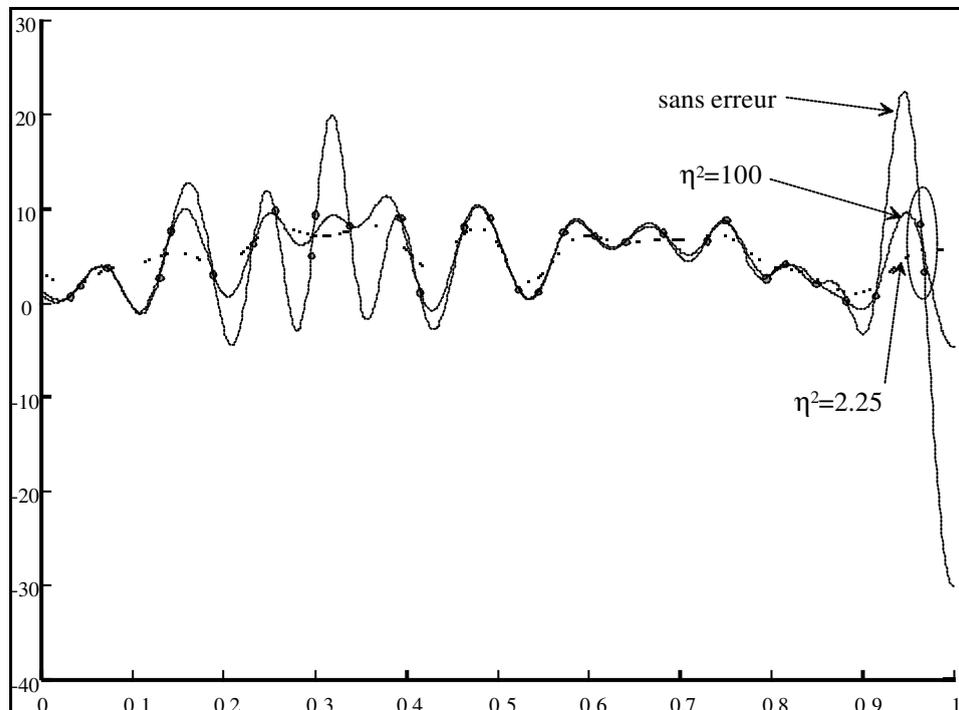
### Etude de $\eta^2$ pour $\theta$ fixé

Le principal effet du terme d'erreur sur le prédicteur est de lever la contrainte d'interpolation, c'est-à-dire qu'il ne passe plus obligatoirement par les observations. Bien évidemment plus  $\eta^2$  est petit, plus le prédicteur est loin des simulations, il se contente de donner l'allure générale de celles-ci, mais ne décrit pas les grandes variations qu'il peut y avoir entre deux observations. (voir ci-dessous pour  $\eta^2=2,25$ ). Dans le chapitre un, nous avons signalé que la contrainte d'interpolation engendrait un prédicteur très irrégulier. On constate ici qu'il suffit d'ajouter une erreur même de faible contribution pour lisser le prédicteur. Ceci est parfaitement illustré sur les deux dessins ci-dessous où le prédicteur pour un modèle sans terme d'erreur est particulièrement perturbé, alors que les observations sont obtenues à l'aide d'une loi uniforme sur  $[0,10]$ . En revanche le prédicteur construit avec une erreur faible de  $\eta^2=100$  est proche de l'interpolation tout en restant à peu près lisse.



Prédicteurs pour différentes erreurs  $\eta^2$  et  $\theta=500$  ( $Z(x)=1$ )

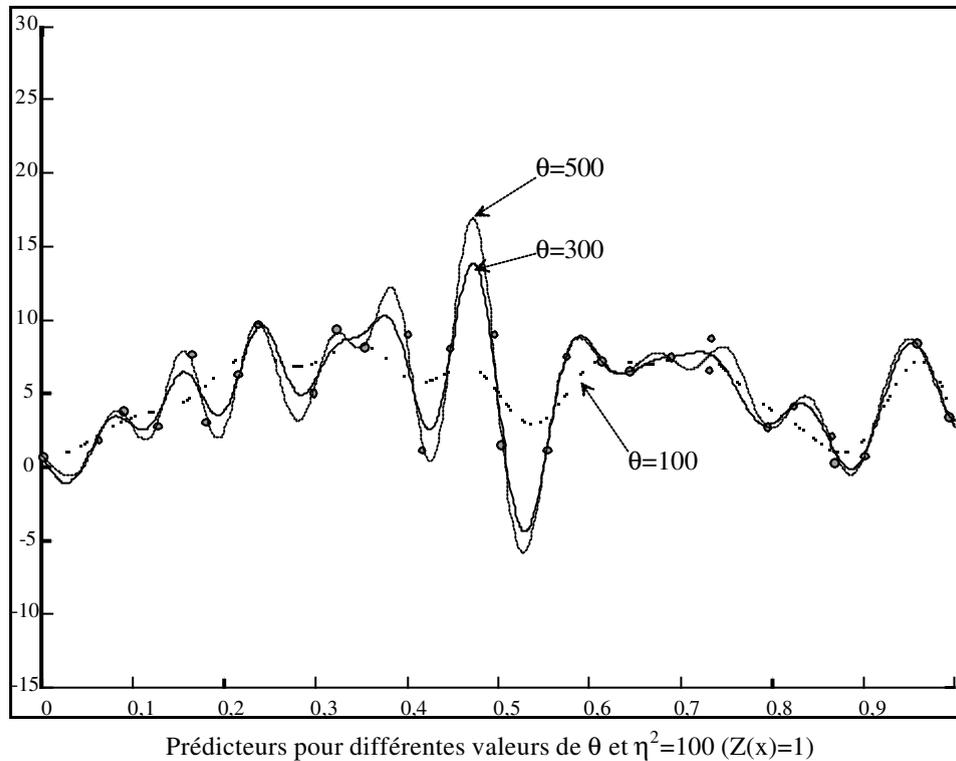
On remarque que les fortes oscillations du prédicteur apparaissent lorsque deux simulations sont effectuées en deux points très proche. En effet, le deuxième degré de l'échantillonnage fait que deux points du plan ne sont pas nécessairement espacés de  $1/30$ , or le modèle sans terme d'erreur ne tient pas compte de ce phénomène.



Prédicteurs pour différentes erreurs  $\eta^2$  et  $\theta=500$  avec un autre plan que précédemment ( $Z(x)=1$ )

### Etude de $\theta$ pour $\eta^2$ fixé

Plus  $\theta$  est petit, plus la corrélation entre les observations est forte. Nécessairement dès que l'on introduit un terme d'erreur dans le modèle, les contraintes entre les points deviennent difficiles à observer. Donc plus  $\theta$  est petit, et plus le prédicteur passe loin des observations et ceci pour une erreur fixée.



En conclusion, nous pouvons dire que la nécessité d'introduire un terme d'erreur est évidente pour deux raisons essentiellement.

1. Pour lisser le prédicteur
2. Pour tenir compte du deuxième degré de l'échantillonnage.

Toute la difficulté réside dans le bon compromis à trouver entre

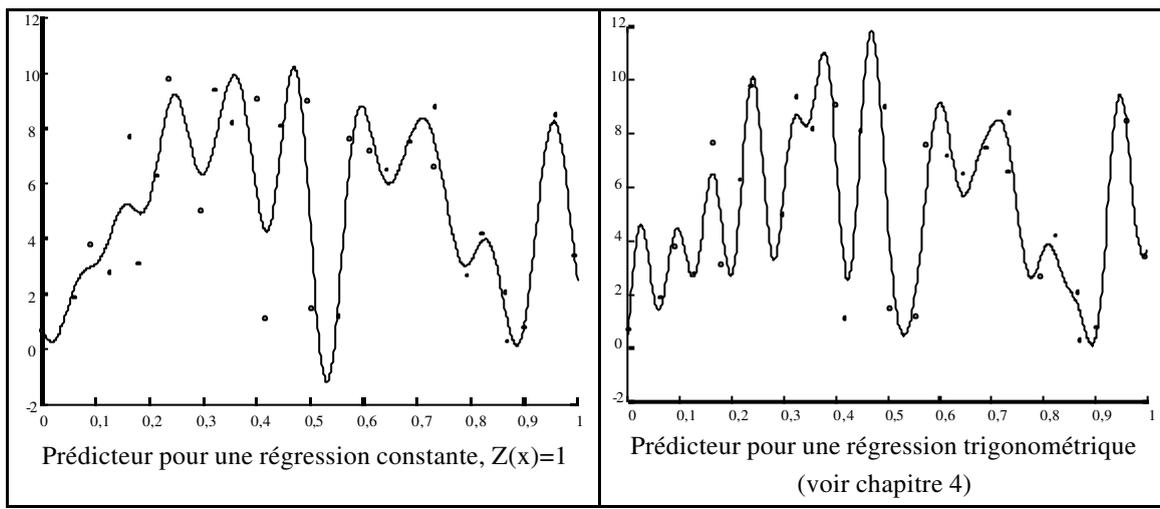
- l'importance du processus résiduel
- le poids relatif du terme d'erreur par rapport aux variances résiduelles.

Il est possible de déterminer le paramètre de corrélation et  $\eta^2$  en les estimant par maximum de vraisemblable (se référer à l'article de Costa *et al.* (1999) pour les formules des estimateurs). Cependant, cette estimation aboutit aux mêmes problèmes que dans le chapitre un, c'est-à-dire à des procédures numériques d'optimisation souvent très coûteuses et qui ne donnent pas toujours de bons résultats.

### **3.3.2. Contribution de la partie fixe du modèle**

Il semble clair, d'après ce que nous venons de voir sur la partie aléatoire du modèle, que le fait de lever la contrainte d'interpolation en introduisant un terme d'erreur, diminue le

rôle joué par le processus résiduel. Il est donc naturel de penser que maintenant la prédiction repose en partie sur l'ajustement en moyenne du modèle. Et comme on peut le voir ci-dessous, il est nécessaire de choisir une régression un peu plus riche qu'une constante pour obtenir un prédicteur efficace. C'est pourquoi dans le chapitre suivant, nous allons introduire une régression linéaire trigonométrique. Nous reviendrons aussi sur l'intérêt d'enrichir la partie fixe du modèle dans le cadre de cette nouvelle approche, mais aussi pour l'approche par résidu aléatoire. En effet, nous verrons que contrairement aux idées reçues, utiliser une régression trigonométrique permet d'avoir un meilleur ajustement en moyenne et rendre le prédicteur du chapitre un robuste aux variations de  $\theta$ .



Prédicteurs pour différentes régressions avec  $\theta=200$  et  $\eta^2=100$

Finalement, les trois termes du modèle semblent avoir chacun leur importance, et là encore, la difficulté réside dans le compromis à trouver entre

- la modélisation en espérance
- l'importance du processus résiduel
- le poids relatif du terme d'erreur par rapport aux variances résiduelles.

## Bibliographie

Christensen R. (1987). *Plane answers to complex question : The theory of linear models*. Springer-Verlag.

Christensen R. (1990). *Linear models for multivariate, time series, and spatial data*. Springer-Verlag.

Costa J.P., Pronzato L., Thierry E. (1999). Nonlinear prediction by kriging, with application to noise cancellation. Soumis à *Signal Processing*.

Robinson G.K. (1991). That BLUP is a good thing : the estimation of random effects. *Statist. Science* **6**, 15-51.

Sacks J., Schiller S.B., Welch W.J. (1989). Designs for computer experiments. *Technometrics* **31**, 41-47.



# 4. Régression linéaire trigonométrique ou modèle de Fourier

Comme nous venons de le voir, la qualité de la prédiction en moyenne de la réponse du simulateur, joue maintenant un rôle prépondérant dans notre modèle. Nous verrons aussi que sa contribution n'est pas négligeable dans un modèle sans terme d'erreur, si on veut que la prédiction soit robuste. La régression linéaire doit donc être choisie avec soin. Dans ce travail, nous utilisons une régression trigonométrique (ou modèle de Fourier) à la place du traditionnel modèle polynomial. L'idée de choisir cette régression dans le cas des expériences simulées a été suggérée par Bates *et al.* (1996). Ils considèrent que c'est un modèle particulièrement bien adapté aux plans qu'ils utilisent, c'est-à-dire les réseaux de rang 1. En effet, nous avons vu dans le chapitre deux (proposition 2.19) que ces plans sont construits de façon à minimiser l'erreur d'intégration pour des fonctions admettant un développement de Fourier absolument convergent. Les points du premier degré de l'échantillonnage que nous utilisons ne sont pas des réseaux, bien que, comme nous le verrons dans le chapitre suivant, ils peuvent leur être associés très simplement. Nous avons cependant gardé l'idée d'une régression trigonométrique pour deux raisons essentiellement.

La première concerne l'estimabilité des paramètres du modèle. Cette question ne se posait pratiquement pas pour les deux approches de la première partie. Pour la méthode fréquentiste, aucune contrainte sur le plan est nécessaire pour calculer l'estimateur. Et dans le cas de l'approche bayésienne, nous avons vu qu'en pratique seule la régression linéaire constante est envisagée. Nous allons établir dans ce chapitre des conditions très simples d'estimabilité des paramètres, dans le cas où les points utilisés forment un sous-groupe de  $G$ . Or les plans que nous utilisons et qui sont définis dans le chapitre suivant sont basés sur cette structure de sous-groupe.

La deuxième raison concerne l'optimalité des plans. Nous allons démontrer que tout sous-groupe de  $G$  pour lequel les paramètres sont estimables est universellement optimal au sens de Kiefer, dans le cas d'un modèle linéaire trigonométrique sans corrélation spatiale.

## 4.1. Définition de la régression linéaire trigonométrique

L'idée d'utiliser, dans le cadre des expériences simulées, une régression linéaire trigonométrique (ou modèle de Fourier) à la place du traditionnel modèle polynomial, a été suggérée par Bates *et al.* (1996). Depuis elle a été étudiée dans les articles de Riccomagno *et al.* (1997) et Bates *et al.* (1998). Ils en donnent la définition suivante.

### Définition 4.1

*Soit  $A^+$  un ensemble de  $d$ -uples d'entiers tel que  $0 \notin A^+$  et  $h \in A^+$  implique  $-h \notin A^+$ . La régression trigonométrique définie par cet ensemble est de la forme,  $\forall x \in [0, 1]^d$ ,*

$$E[Y(x)] = f(x) = \beta_0 + \sqrt{2} \sum_{h \in A^+} [\beta_h \sin(2\pi(x|h)_R) + \delta_h \cos(2\pi(x|h)_R)],$$

*où  $(x|h)_R$  est le produit scalaire usuel de  $\mathbb{R}^d$ .*

*Les éléments de  $A^+$  s'appellent les fréquences du modèle, elle le définissent entièrement.*

On parle de régression trigonométrique ou bien de modèle de Fourier bien que la condition porte seulement sur l'espérance de  $Y(x)$ .

Puisque  $E[Y(x)] = Z_R(x)\beta$ , on a  $m = 2\text{card}(A^+) + 1$  et

$$\begin{cases} Z_R(x) = [1, \{\sqrt{2} \sin(2\pi(x|h)_R)\}_{h \in A^+}, \{\sqrt{2} \cos(2\pi(x|h)_R)\}_{h \in A^+}] \\ \beta = {}^t [\beta_0, \{\beta_h\}_{h \in A^+}, \{\delta_h\}_{h \in A^+}] \end{cases}$$

Le modèle de Fourier est entièrement déterminé par l'ensemble  $A^+$  de ses fréquences. Nous utilisons ici la terminologie suivante.

On appelle **effet simple** du  $i$ ème facteur, une fréquence de la forme  $h = (0, \dots, h_i, \dots, 0)$ .

On appelle **interaction** du  $i$ ème et du  $j$ ème facteur, une fréquence de la forme  $h = (0, \dots, h_i, \dots, h_j, \dots, 0)$ .  $h_i \neq 0$  représente l'**ordre** du  $i$ ème facteur dans la fréquence.

Nous reviendrons sur cette terminologie au cours de ce chapitre, ainsi que sur le comportement du prédicteur en fonction du choix des fréquences du modèle.

On parle de modèle de Fourier **complet** et on utilise la notation de Riccomagno (1997),  $F(d; m_1, \dots, m_d; M)$ , si  $\forall i \in \{1, \dots, d\} \exists m_i \in \mathbb{N}^*$  tel que toutes les fréquences  $h = (h_1, \dots, h_d)$  telles que

$$0 \leq h_i \leq m_i \quad i = 1, \dots, d \quad \text{et} \quad \text{wt}(h) \leq M,$$

où  $\text{wt}(h)$  est le poids de  $h$  *i.e* le nombre  $h_i$  non nuls, figurent dans le modèle.

On utilise le plus souvent des modèles complets faisant intervenir les facteurs uniquement à l'ordre un, par exemple  $F(d;1,\dots,1;1)$  pour les effets simples et  $F(d;1,\dots,1;2)$  pour les interactions de deux facteurs.

**Exemple 4.2 :** le modèle complet  $F(2;1,2;2)$  correspond à  $A^+ = \{(1,0), (0,1), (0,2), (1,1), (1,2)\}$ . Il représente l'effet simple d'ordre 1 du premier facteur, les effets simples d'ordre 1 et 2 du deuxième facteur, et les interactions d'ordre 1 et d'ordre 1 et 2 des deux facteurs. La partie sinus du modèle est de la forme

$$\beta_{(1,0)} \sin(2\pi x_1) + \beta_{(0,1)} \sin(2\pi x_2) + \beta_{(0,2)} \sin(4\pi x_2) + \beta_{(1,1)} \sin(2\pi(x_1 + x_2)) + \beta_{(1,2)} \sin(2\pi(x_1 + 2x_2)).$$

Le modèle peut s'écrire sous une **forme complexe**. Pour tout  $x$  de  $[0,1]^d$ ,

$$f(x) = Z_c(x)\alpha$$

où  $Z_c(x)$  et  $\alpha$  sont des vecteurs à éléments dans  $\mathbb{C}$  tels que

$$Z_R(x) = \left[ 1, \left\{ \exp(2\pi i(x|h)) \right\}_{h \in A^+}, \left\{ \exp(-2\pi i(x|h)) \right\}_{h \in A^+} \right]$$

$$\alpha = \left[ 1, \left\{ \alpha_h \right\}_{h \in A^+}, \left\{ \bar{\alpha}_h \right\}_{h \in A^+} \right]$$

$f(x)$  appartient à  $\mathbb{R}$  du fait des contraintes de conjugaison sur les éléments de  $\alpha$ . Les deux formes sont liées par les relations  $\alpha = H\beta$  et  $Z_c(x) = Z_R(x)H^{-1}$  (cf. Bates *et al.* (1998)) où  $H$  est la matrice

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -iJ & J \\ 0 & iJ & J \end{pmatrix} \text{ avec } J = \frac{1}{\sqrt{2}}I.$$

On remarque qu'il n'y a en fait que  $\text{card}A^+ + 1 < m$  paramètres à estimer.

Par la suite, nous utilisons la notation  $Z$  pour désigner la matrice du modèle sous sa forme complexe.

Rappelons qu'avec un échantillonnage à deux degrés (cf. chapitre 2), on partitionne le cube unité en cellules et que le point  $x$  de notre modèle est le représentant d'une de ces cellules. Ces représentants, qu'ils soient choisis au bord de la cellule ou au milieu, sont associés à un élément  $g$  du groupe abélien fini d'ordre  $\prod q_i$  et d'exposant  $q = \text{PPCM}\{q_i \mid i=0,\dots,d\}$

$$G = \times_{i=1}^d (Z/q_i).$$

Dans ce chapitre on désigne par plan le sous-ensemble de  $G$  auquel sont associés les représentants des cellules du premier degré de l'échantillonnage. C'est en effet le seul paramètre que l'on puisse contrôler pour effectuer les simulations.

Soit  $\mathbb{R}^G$  l'ensemble des fonctions à valeurs réelles sur  $G$ , alors  $f$  appartient à  $\mathbb{R}^G$ . Dans un premier temps nous allons montrer que toute fonction de  $\mathbb{R}^G$  peut se décomposer de façon unique sur une base orthonormée (base de Yates), et que le modèle sous sa forme complexe n'est autre qu'une troncature de cette décomposition.

## 4.2. Construction d'une base orthonormée pour la fonction

La fonction  $f$  appartient à  $\mathbb{R}^G$  que nous allons considérer comme un sous-ensemble de  $\mathbb{C}^G$ , l'ensemble des fonctions à valeurs complexes sur  $G$ .

Avant de construire la base sur laquelle  $f$  se décompose, quelques rappels d'algèbre s'imposent, en particulier sur les représentations linéaires des groupes finis.

### 4.2.1. Représentations linéaires des groupes finis

On retrouvera les résultats et définitions de cette partie ainsi de plus amples détails dans des ouvrages spécialisés comme celui de Serres (1967).

#### Définitions 4.3

*Soient  $V$  un  $\mathbb{C}$ -espace vectoriel de dimension  $n$  et  $G$  un groupe fini.*

*Une représentation linéaire de  $G$  dans  $V$  est un isomorphisme  $\rho$  de  $G$  dans le groupe linéaire sur  $V$ . Autrement dit, l'application*

$$\begin{aligned} \rho : G &\rightarrow GL(V) \\ g &\mapsto \rho(g) = \rho_g \end{aligned}$$

*est telle que  $\forall g, t \in G$ ,*

$$\left\{ \begin{array}{l} \rho_{gt} = \rho_g \circ \rho_t \\ \rho_1 = I \\ \rho_{g^{-1}} = \rho(g)^{-1} \in GL(V) \end{array} \right. ,$$

*où  $g^{-1}$  désigne l'inverse de  $g$  dans  $G$  et  $I$  est l'élément neutre de  $G$ .*

*$V$  est l'espace de représentation de  $G$  (ou la représentation de  $G$ ),  $n$  le degré de cette représentation.*

*On dit que  $\rho$  est irréductible si  $V$  n'est pas réduit à  $0$  et s'il n'existe pas de sous-espace  $W$  stable par  $\rho_g$ ,  $\forall g \in G$  (i.e  $x \in W \Rightarrow \rho_g(x) \in W$ ).*

*On appelle caractère d'une représentation linéaire  $\rho$ , la fonction  $\chi_\rho$  de  $\mathbb{C}^G$  définie par*

$$\chi_\rho(g) = \text{trace} \rho_g, \quad \forall g \in G.$$

*Enfin, on dit que deux représentations linéaires de  $G$ ,  $\rho^1: G \rightarrow GL(V_1)$  et  $\rho^2: G \rightarrow GL(V_2)$ , sont isomorphes s'il existe un isomorphisme linéaire  $\tau$  de  $V_1$  dans  $V_2$  tel que*

$$\tau \circ \rho_g^1 = \rho_g^2 \circ \tau, \quad \forall g \in G.$$

*Elles ont alors mêmes degré et caractère, nous pouvons donc les identifier.*

### Cas particulier des groupes commutatifs (abéliens)

Supposons maintenant que  $G$  est un groupe commutatif et munissons  $\mathbb{C}^G$  du produit scalaire

$$(\varphi, \psi) = \frac{1}{|G|} \sum_{g \in G} \varphi(g) \overline{\psi(g)}$$

#### Propriétés 4.4

*Les représentations irréductibles de  $G$  sont de degré 1. Autrement dit, une représentation de  $G$  n'est autre qu'un morphisme de  $G$  dans le groupe multiplicatif des complexes non nuls.*

*De plus tout élément de  $G$  est d'ordre fini, les caractères des représentations irréductibles de  $G$  ont donc pour valeurs des racines  $q$ -ième de l'unité. Ils vérifient aussi*

$$\chi(1) = 1 \text{ et } \chi(g^{-1}) = \overline{\chi(g)}, \forall g \in G.$$

*Les caractères des représentations irréductibles de  $G$  forment une base orthonormée de  $\mathbb{C}^G$ .*

#### Définition et propriétés 4.5

*Munis du produit d'Hadamard, ces caractères forment un groupe abélien fini isomorphe à  $G$ , appelé le dual de  $G$  et noté  $G^*$ . Donc tout élément de  $G$  peut être identifié à un élément de son dual et réciproquement.*

*Si on suppose en outre que  $G = G_1 \times G_2$  alors*

$$\chi_1 \in G_1^* \text{ et } \chi_2 \in G_2^* \Rightarrow \chi_1 \otimes \chi_2 \in G^*.$$

En utilisant les propriétés ci-dessus, nous allons pouvoir construire de façon explicite une base de  $\mathbb{C}^G$  dans le cas où  $G$  est le produit cartésien de groupes cycliques. Il suffit pour cela de connaître les caractères des représentations linéaires irréductibles des groupes cycliques.

### Caractères des représentations irréductibles des groupes cycliques

Soit  $G$  le groupe cyclique d'ordre  $p$ . Il est formé des puissances  $1, r, \dots, r^{p-1}$  d'un élément  $r$  tel que  $r^p = 1$ . D'après les propriétés 1.1,  $G$  étant un groupe abélien, ses représentations irréductibles sont de degré 1, de caractères tels que  $\xi(r^k) = w^k$ ,  $k = 0, \dots, p-1$  avec  $w \in \mathbb{C}$ . Puisque  $r^p = 1$ , on doit avoir  $w^p = 1$ , c'est-à-dire que  $w$  peut prendre les valeurs  $\exp(2\pi i h/p)$   $h = 0, \dots, p-1$ . On trouve ainsi  $p$  représentations irréductibles non isomorphes de  $G$  dont les caractères  $\xi^0, \dots, \xi^{p-1}$  sont donnés par les fonctions de  $\mathbb{C}^G$  définies par

$$\xi^h(g) = \exp(2\pi i h g/p), \forall g \in G.$$

Si on considère les caractères comme des vecteurs de  $\mathbb{C}^p$  et si on les place en colonnes, on obtient une matrice carrée d'ordre  $p$ ,  $F_p$ , dite matrice de Fourier,  $w = \exp(2\pi i/p)$

$$F_p = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{p-1} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w^{p-1} & w^{2(p-1)} & \dots & w^{(p-1)^2} \end{pmatrix}.$$

**Exemple 4.6 :** Soit  $G$  le groupe cyclique d'ordre 3,  $G=Z/3$ . Les caractères des trois représentations irréductibles non isomorphes de  $G$  sont définis par

$$\xi^h(g) = \exp(2\pi i hg/3), \forall g \in G.$$

En colonnes, on obtient la matrice de Fourier

$$F_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & w & w^2 \\ 1 & w^2 & w \end{pmatrix} \text{ où } w = e^{2i\pi/3}.$$

**Proposition 4.7**

*Soit  $x = \exp(2\pi i/p)$  la racine  $p$ ème de l'unité, alors*

$$\sum_{k=0}^{p-1} w^k = 0.$$

**4.2.2. Base de Yates**

Maintenant que tous les outils sont en place, nous pouvons déterminer une base de  $\mathbb{R}^G$ , où  $G$  est le produit cartésien de  $d$  groupes cycliques,  $G=(Z/q_1) \times \dots \times (Z/q_d)$ . Comme nous l'avons vu précédemment les éléments du dual  $G^*$  forment une base orthonormée de  $\mathbb{C}^G$ , or nous savons que ces éléments sont produit tensoriel des caractères de représentations irréductibles des  $Z/q_i$ ,  $i=1, \dots, d$ . Donc si on note ces caractères  $\xi_i^{h_i}$ ,  $h_i \in Z/q_i$ ,  $i=1, \dots, d$ , tout éléments  $\chi_h$  du dual  $G^*$  s'écrit  $\chi_h = \otimes_{j=1}^d \xi_j^{h_j}$  où  $h=(h_1, \dots, h_d) \in G$ . C'est-à-dire,  $\forall g \in G$

$$\chi_h(g) = \exp\left[ i \frac{2\pi}{q} (g|h) \right] \text{ où } (g|h) = \sum_{j=1}^d h_j g_j \frac{q}{q_j} \pmod{q}.$$

**Propriétés 4.8**

- 1) Les caractères sont soit conjugués deux à deux,  $\overline{\chi_h} = \chi_{-h}$ , soit à valeurs réelles.
- 2)  $\forall h \in G, \forall k \in G, \chi_{h+k} = \chi_h \chi_k$ .
- 3)  $\forall h \in G, \forall g \in G$  et  $\forall g' \in G, \chi_h(g+g') = \chi_h(g) \chi_h(g')$
- 4)  $\forall h \in G, \forall g \in G, |\chi_h(g)| = 1$  où  $|z|$  est le module de  $z \in \mathbb{C}$

Etant donné que ces caractères forment une base orthonormée de  $\mathbb{C}^G$ , toute fonction  $f$  de  $\mathbb{C}^G$  peut se décomposer dans cette base de façon classique,

$$f = \sum_{h \in G} \alpha_h \chi_h \quad \text{où } \alpha_h = (f, \chi_h)$$

**Proposition 4.9**

*f est à valeurs réelles si et seulement si  $\overline{\alpha}_h = \alpha_{-h}$ .*

**Démonstration :** On a

$$\overline{\alpha}_h = \overline{(f, \chi_h)} = (\chi_h, f) = \frac{1}{|G|} \sum_{g \in G} \chi_h(g) \overline{f(g)}$$

$$\text{et } \alpha_{-h} = (f, \chi_{-h}) = (f, \overline{\chi_h}) = \frac{1}{|G|} \sum_{g \in G} f(g) \chi_h(g).$$

Donc si f est à valeurs réelles, i.e.  $\overline{f(g)} = f(g), \forall g \in G$ , alors  $\overline{\alpha}_h = \alpha_{-h}$ .

Et réciproquement si  $\overline{\alpha}_h = \alpha_{-h}$  alors, du fait que les caractères forment une base de  $\mathbb{C}^G$ , on a  $\overline{f(g)} = f(g), \forall g \in G$ , donc f à valeurs réelles.  $\square$

**Remarque :** Soit  $x \in \mathbb{R}$  tel que  $x_i = g_i/q_i \quad i=1, \dots, d$ , alors

$$\exp\left[ i \frac{2\pi}{q} (g | h) \right] = \exp\left[ i 2\pi (x | h)_{\mathbb{R}} \right].$$

Cette relation nous permet de faire le lien entre le modèle présenté sous sa forme complexe au paragraphe 4.1 et les caractères de représentation irréductibles de G.

**Définition 4.10**

*La base des fonctions de  $\mathbb{F}^d$  ainsi définie s'appelle base de Yates. Si l'on considère ses éléments comme des vecteurs colonnes, alors ceux-ci correspondent aux colonnes de la matrice F obtenue par produit tensoriel des matrices de Fourier  $F_{q_1} \otimes \dots \otimes F_{q_d}$ . Cette matrice s'appelle la table des caractères des représentations irréductibles de G.*

**Remarque :** La matrice F est la matrice de passage de la base canonique de  $\mathbb{C}^G$  à la base de Yates. Ses lignes sont indexées par les éléments de G.

**Exemple 4.11 :** Soit  $G = \times_{i=1}^3 (\mathbb{Z}/3)$ . Les éléments du dual  $G^*$  s'écrivent

$$\chi_h(g) = \exp\left[ i \frac{2\pi}{3} (gh) \right], \quad g \in G.$$

Soit alors F la table des caractères des représentations irréductibles de  $\times_{i=1}^3 (\mathbb{Z}/3)$ ,  $F = F_3^{\otimes 3}$

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w & w^2 & 1 & w & w^2 & 1 & w & w^2 \\ 1 & w^2 & w & 1 & w^2 & w & 1 & w^2 & w \\ 1 & 1 & 1 & w & w & w & w^2 & w^2 & w^2 \\ 1 & w & w^2 & w & w^2 & 1 & w^2 & 1 & w \\ 1 & w^2 & w & w & 1 & w^2 & w^2 & w & 1 \\ 1 & 1 & 1 & w^2 & w^2 & w^2 & w & w & w \\ 1 & w & w^2 & w^2 & 1 & w & w & w^2 & 1 \\ 1 & w^2 & w & w^2 & w & 1 & w & 1 & w^2 \end{pmatrix} \text{ où } w=e^{i2\pi/3}$$

### Modèles linéaires complexes

Il est clair que le modèle linéaire complexe correspond à une troncature de la décomposition de  $f$  sur la base de Yates, puisque  $\forall g \in G$

$$f(g) = \sum_{h \in G} \alpha_h \exp\left(\frac{2\pi i}{q} (g|h)\right),$$

où  $\alpha_h$  vérifie la proposition 4.9.

L'idée de construire ce type de modèle à partir des caractères de représentations linéaires des groupes finis a été émise par Bailey (1982,1985), elle a ensuite été longuement développée par Kobilinsky (1985a,1985b,1990) et Kobilinsky et Monod (1991).

Nous allons maintenant montré à quoi correspondent les éléments de la décomposition, c'est-à-dire effets simples et interactions des facteurs (comme pour le modèle linéaire polynomial).

Soient  $\alpha_h$ ,  $h \in G$ , les coefficients de la décomposition de  $f$  dans la base de Yates. Nous pouvons remarquer que si  $h=(0,\dots,0)$  alors

$$\alpha_h = \alpha_0 = \frac{1}{|G|} \sum_{g \in G} f(g)$$

correspond à la valeur moyenne de  $f$  sur  $G$ . De même si  $h=(h_1,0,\dots,0)$  avec  $h_1 \neq 0$ , alors

$$\alpha_h = \frac{1}{|G|} \sum_{g_1} e^{i\frac{2\pi}{q} h_1 g_1} \sum_{g_2, \dots, g_d} f(g)$$

correspond à **un** effet simple du facteur 1 sur  $G$ , *etc...* De façon générale, la nature du paramètre  $\alpha_h$  se déduit des éléments non nuls du vecteur  $h$ . Par exemple, si  $h=(h_1, h_2, 0, \dots, 0)$  avec  $h_1 \neq 0$  et  $h_2 \neq 0$  alors  $\alpha_h$  correspond à **une** interaction des facteurs 1 et 2. Ainsi, si nous voulons tenir compte dans l'approximation de  $f$  uniquement de certains effets ou interactions, nous pouvons réduire la décomposition de  $f$  dans la base de Yates au modèle suivant. Pour tout  $g \in G$ ,

$$f(g) = \sum_{h \in A} \alpha_h \chi_h(g)$$

où  $A$  est un sous-ensemble de  $G$  qui ne prend en compte que les effets ou interactions désirés. D'après la proposition 4.9,  $A$  doit être stable par l'opération  $h \mapsto -h$  pour que  $f$  soit à valeurs réelles, nous pouvons donc le décomposer comme suit

$$A = \{0\} \cup A^+ \cup \{-A^+\}$$

où  $A^+$  correspond à l'ensemble des fréquences défini précédemment, *i.e.* tel que  $0 \notin A^+$  et  $h \in A^+$  implique  $-h \notin A^+$ . On retrouve alors le modèle sous sa forme complexe sur  $G$ ,  $\forall g \in G$

$$f(g) = Z(g)\alpha = \alpha_0 + \sum_{h \in A^+} \alpha_h \exp\left(i \frac{2\pi}{q}(h|g)\right) + \overline{\alpha}_h \exp\left(-i \frac{2\pi}{q}(h|g)\right).$$

**Exemple 4.12 :** Soit  $G=(Z/3)^2$ , le modèle analysant des effets simples des deux facteurs ainsi que la valeur moyenne est défini par l'ensemble des fréquences  $A^+=\{(1,0),(0,1)\}$ . Il est donc de la forme,  $g=(g_1, g_2) \in G$

$$f(g) = \alpha_0 + \alpha_{(1,0)} e^{i \frac{2\pi}{3} g_1} + \alpha_{(0,1)} e^{i \frac{2\pi}{3} g_2} + \overline{\alpha}_{(1,0)} e^{-i \frac{2\pi}{3} g_1} + \overline{\alpha}_{(0,1)} e^{-i \frac{2\pi}{3} g_2}.$$

Pour un plan  $D=\{(0,0),(1,1),(2,2)\}$ , la matrice du modèle est égale à

$$Z = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & w & w & w^2 & w^2 \\ 1 & w^2 & w^2 & w & w \end{pmatrix} \text{ avec } w = e^{2\pi i/3}.$$

### 4.3. Estimabilité et optimalité

Les conditions d'estimabilité, les estimateurs des paramètres et toute autre expression où intervient la matrice du modèle (prédicteur, MSE, ...) s'appliquent dans le cas d'une régression linéaire complexe. Il convient cependant de remplacer la transposée de cette matrice par son adjointe (c'est-à-dire la conjuguée de la transposée). Toutefois, le fait d'utiliser ce type de modèle confère certaines particularités quand à l'estimabilité des paramètres et l'optimalité des plans. Celles-ci découlent de la forme particulière de la matrice d'information  $Z^*Z$ , notamment du fait que tous ses éléments diagonaux sont égaux à  $N$ , la taille de l'échantillon.

#### Proposition 4.13

Soit un plan  $D=\{g_1, \dots, g_N\}$ . La matrice  $Z^*Z$  a une structure en blocs de la forme

$$Z^*Z = \begin{pmatrix} N & a & \bar{a} \\ a^* & B & C \\ {}^t a & \bar{C} & \bar{B} \end{pmatrix}$$

où pour  $h \in A^+$  et  $k \in A^+$

$$a_h = \sum_{i=1}^N \chi_h(g_i), \quad B_{hk} = \sum_{i=1}^N \chi_{(k-h)}(g_i) \quad \text{et} \quad C_{hk} = \sum_{i=1}^N \chi_{-(k+h)}(g_i)$$

On remarque qu'il n'y a que trois types de produits scalaires qui interviennent dans la matrice  $Z^*Z$ . Dans le bloc  $a$  sont représentés les produits scalaires de  $\mathbf{1}$  avec les colonnes de la matrice du modèle indexées par les fréquences de  $A^+$ , dans le bloc  $B$  il y a les produits scalaires de ces dernières entre elles, et dans le bloc  $C$  le produit entre ces colonnes et leurs conjuguées. On note de plus que  $C$  est une matrice symétrique et  $B$  une matrice hermitienne.

#### Optimalité des plans pour une régression linéaire complexe

Le premier à s'intéresser à l'optimalité des plans dans le cas d'une régression trigonométrique est Hoel (1965) en se basant sur les travaux de Kiefer et Wolfowitz. Il a notamment montré qu'en dimension un et deux, les points également espacés d'une grille forment un plan  $D$ -optimal pour un modèle trigonométrique sans processus résiduel, c'est-à-dire que le plan minimise la quantité  $\max_{x \in D} \text{var}(\hat{Y}(x))$  où  $\hat{Y}(x)$  est un estimateur de  $Y(x)$ . Plus récemment, nous pouvons citer les travaux de Bates *et al.* (1998) et Riccomagno *et al.* (1997) pour l'étude de l'optimalité des réseaux.

#### Définitions 4.14

*On dit qu'un plan d'échantillonnage de taille  $N$  est adapté orthogonalement (ou bien est orthogonal) au modèle si  $Z^*Z=NI$  où  $Z$  est la matrice du modèle aux points du plan.*

Le plan est alors universellement optimal au sens de Kiefer, dans le cas où l'échantillonnage est à un seul degré, c'est-à-dire si on choisit par exemple le centre des cellules, et si on considère un modèle linéaire simple, autrement dit en l'absence de corrélation.

**Remarque :** Si un plan est orthogonal à un modèle défini par son ensemble de fréquences  $A$ , alors il est orthogonal à tout modèle défini par un ensemble de fréquences  $A' \subset A$ .

On établit le résultat suivant à partir de la forme de la matrice  $Z^*Z$  donnée dans la proposition 4.13.

#### Proposition 4.15

*Un plan est adapté orthogonalement au modèle si et seulement si  $\forall h \in A^+, \forall k \in A^+$   
 $a_h=0, B_{hk}=0 \text{ } h \neq k \text{ et } C_{hk}=0.$*

#### Cas où le plan $D$ est un sous-groupe de $G$

Nous étudions ce cas pour deux raisons. La première est que la structure de sous-groupe est la plus simple que l'on puisse donner aux sous-ensembles  $D$  de  $G$ , et effectivement, tous les plans d'échantillonnage que nous utilisons dans le chapitre suivant sont basés sur des sous-groupes. La deuxième raison est que l'on peut démontrer un résultat fondamental sur les caractères de représentations irréductibles de  $G$  aux points d'un sous-groupe. Celui-ci permet d'établir que, soit un plan (un sous-groupe) est adapté orthogonalement au modèle, soit il ne permet pas d'analyser le modèle.

#### Définition 4.16

*On appelle contrastes de définition (ou annulateur) de  $S$  le sous-groupe de  $G^*$*

$$S^\circ = \{ \chi \in G^*, \chi(s) = 1 \quad \forall s \in S \}.$$

*On appelle dual de  $S$  le sous-groupe de  $G$  défini par*

$$S^\perp = \{ h \in G, (s|h)=0 \quad \forall s \in S \}.$$

On remarquera que  $S^\perp$  s'appelle le dual alors que c'est un sous-groupe de  $G$ . Cela vient du fait qu'il caractérise entièrement  $S^\circ$  qui lui appartient au dual de  $G$ . En effet, soit  $\chi_h$  un caractère de représentation irréductible de  $G$  appartenant à  $S^\circ$ , alors

$$\chi_h(s) = 1 \quad \forall s \in S \Leftrightarrow \exp\left[i \frac{2\pi}{q}(s|h)\right] = 1 \quad \forall s \in S \Leftrightarrow (s|h) = 0 \quad \forall s \in S \Leftrightarrow h \in S^\perp$$

Autrement dit,

$$S^\circ = \{\chi_h \in G^*, h \in S^\perp\}.$$

Nous avons alors les quelques propriétés suivantes.

**Propriétés 4.17**

- 1)  $\text{card}(S^\perp) = \text{card}(G) / \text{card}(S)$
- 2) Si  $T$  est un sous-groupe de  $S$  alors  $T^\perp$  est un sous-groupe de  $S^\perp$ .
- 3)  $\chi \in S^\circ \Leftrightarrow \bar{\chi} \in S^\circ$

**Exemple 4.18 :** Soit  $S$  le sous-groupe de  $G=(\mathbb{Z}/3)^3$  de cardinal 9 défini par

$$S = \{s \in G, s_1 + s_2 + s_3 = 0\}$$

Les éléments  $h$  de son dual doivent vérifier la relation  $(h|s) = 0 \quad \forall s \in S$

$$\Leftrightarrow s_1 h_1 + s_2 h_2 + s_3 h_3 = 0 \quad \forall s \in S \Leftrightarrow s_2 (h_2 - h_1) + s_3 (h_3 - h_1) = 0 \quad \forall s \in S \Leftrightarrow h_1 \equiv h_2 \equiv h_3$$

Le dual est donc égal à  $S^\perp = \{(0,0,0), (1,1,1), (2,2,2)\}$ .

L'orthogonal  $S^\circ$  est constitué de la fonction triviale sur  $S$ ,  $\mathbf{1}$ , de la fonction

$$\chi_{(1,1,1)} s : \mapsto \exp\left[i \frac{2\pi}{3} (s_1 + s_2 + s_3)\right]$$

et de sa fonction conjuguée  $\chi_{(2,2,2)} = \overline{\chi_{(1,1,1)}}$ .

Nous pouvons alors démontrer le résultat fondamental suivant.

**Théorème 4.19**

Soit  $S = \{s_1, \dots, s_N\}$  un sous-groupe de  $G$  d'ordre  $N$ . Alors  $\forall h \in G$ ,

$$\sum_{j=1}^N \chi_h(s_j) = \begin{cases} N & \text{si } h \in S^\perp \\ 0 & \text{sinon} \end{cases}.$$

**Démonstration :** Soient  $\chi_{h|_S}$  la restriction à  $S$  du caractère  $\chi_h$  et  $\mathbf{1}_S$  le caractère trivial sur  $S$ .

Etant donné que les caractères de représentations irréductible de  $G$  forment une base orthonormée de  $\mathbb{C}^G$ , nous avons

$$(\chi_{h|_S}, \mathbf{1}_S) = \begin{cases} 1 & \text{si } \chi_{h|_S} = \mathbf{1}_S \\ 0 & \text{sin on} \end{cases}$$

Or d'une part,

$$(\chi_{h|_S}, \mathbf{1}_S) = N^{-1} \sum_{s \in S} \chi_{h|_S}(s) \overline{\mathbf{1}_S(s)} = N^{-1} \sum_{j=1}^N \chi_h(s_j)$$

et d'autre part

$$\chi_{h|_S} = \mathbf{1}_S \Leftrightarrow \chi_{h|_S}(s) = 1 \quad \forall s \in S \Leftrightarrow h \in S^\perp.$$

D'où le résultat.  $\square$

Ce théorème a pour première conséquence la proposition qui suit.

**Proposition 4.20**

*Les colonnes de la matrice du modèle aux points d'un sous-groupe sont, soit orthogonales, soit égales, deux à deux.*

Donc soit le sous-groupe est orthogonalement adapté au modèle, soit il ne permet pas d'estimer les paramètres, puisque alors les colonnes de  $Z$  ne sont pas indépendantes. Si maintenant on applique ce résultat à la matrice  $\tilde{Z}Z$  de la proposition 4.13, on obtient des conditions entre le dual  $S^\perp$  et les fréquences du modèle pour qu'il y ait estimabilité.

**Définition 4.21**

*Soient  $h$  et  $k$  deux fréquences non nulles du modèle et  $S$  un sous-groupe de  $G$ . On dit que les fréquences sont mutuellement orthogonales par rapport à  $S^\perp$ , si*

$$\begin{cases} h \notin S^\perp, k \notin S^\perp \\ (h+k) \notin S^\perp \\ (h-k) \notin S^\perp \text{ pour } h \neq k \end{cases}$$

Si deux fréquences  $h$  et  $k$  ne sont pas mutuellement orthogonales, on dit que les effets qu'elles représentent sont confondus.

**Proposition 4.22**

*Un sous-groupe  $S$  de  $G$  est orthogonalement adapté à un modèle complexe défini par son ensemble de fréquences  $A^+$  si et seulement toutes les fréquences non nulles de  $A^+$  sont deux à deux mutuellement orthogonales par rapport à  $S^\perp$ .*

**Remarques :**

- 1)  $h \notin S^\perp$  est équivalent à  $(-h) \notin S^\perp$  car  $S^\perp$  est un sous-groupe de  $G$ . Donc  $(h-k) \notin A^+ \Leftrightarrow (k-h) \notin A^+$ .
- 2) Si  $q$  est impair et si  $h \notin A^+$ , alors la deuxième condition de la définition 4.21 devient pour  $k \neq h$   $(h+k) \notin S^\perp$ , puisque  $2h$  ne peut pas appartenir à  $S^\perp$ .
- 3) Soit  $A^{++}$  l'ensemble défini par  $A^{++} = A^+ \cup \{h-k, h \neq k \in A^+\} \cup \{h+k, h \in A^+, k \in A^+\}$ . Les fréquences de  $A^+$  sont toutes deux à deux mutuellement orthogonales par rapport à  $S^\perp$  si et seulement si

$$A^{++} \cap S^\perp = \emptyset.$$

**Démonstration** : D'après la proposition 4.15, un sous-groupe  $S=\{s_1, \dots, s_N\}$  est adapté orthogonalement à l'estimation des paramètres si et seulement si

$$\forall h \in A^+, a_h = 0 \Leftrightarrow \sum_{i=1}^N \chi_h(s_i) = 0 \Leftrightarrow h \notin S^\perp \text{ d'après le théorème 4.19}$$

$$\forall k \neq h \in A^+, B_{hk} = 0 \Leftrightarrow \sum_{i=1}^N \chi_{(k-h)}(s_i) = 0 \Leftrightarrow (k-h) \notin S^\perp \text{ d'après le théorème 4.19}$$

$$\forall k \in A^+, C_{hk} = 0 \Leftrightarrow \sum_{i=1}^N \chi_{-(k+h)}(s_i) = 0 \Leftrightarrow (k+h) \notin S^\perp \text{ d'après le théorème 4.19. } \square$$

**Exemple 4.18 (suite)** :  $S$  est le sous-groupe engendré par  $(1,1,1)$ . Le modèle est défini par l'ensemble  $A^+=\{(1,0,0), (0,1,1)\}$ ,  $\forall x \in G$

$$f(x) = \alpha_0 + \alpha_{(1,0,0)} e^{i\frac{2\pi}{3}x_1} + \alpha_{(0,1,1)} e^{i\frac{2\pi}{3}(x_2+x_3)} + \overline{\alpha_{(1,0,0)}} e^{-i\frac{2\pi}{3}x_1} + \overline{\alpha_{(0,1,1)}} e^{-i\frac{2\pi}{3}(x_2+x_3)}.$$

Alors les deux fréquences ne sont pas mutuellement orthogonales par rapport à  $S$  puisque  $(1,0,0)+(0,1,1) \in S^\perp$ . Les 4 colonnes complexes de la matrice  $Z$  du modèle sont alors identiques par paires,

$$Z = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & w & w & w & w^2 & w^2 & w^2 \\ 1 & 1 & 1 & w^2 & w^2 & w^2 & w & w & w \\ 1 & 1 & 1 & w^2 & w^2 & w^2 & w & w & w \\ 1 & 1 & 1 & w & w & w & w^2 & w^2 & w^2 \end{pmatrix} \text{ où } w = e^{2\pi i/3}.$$

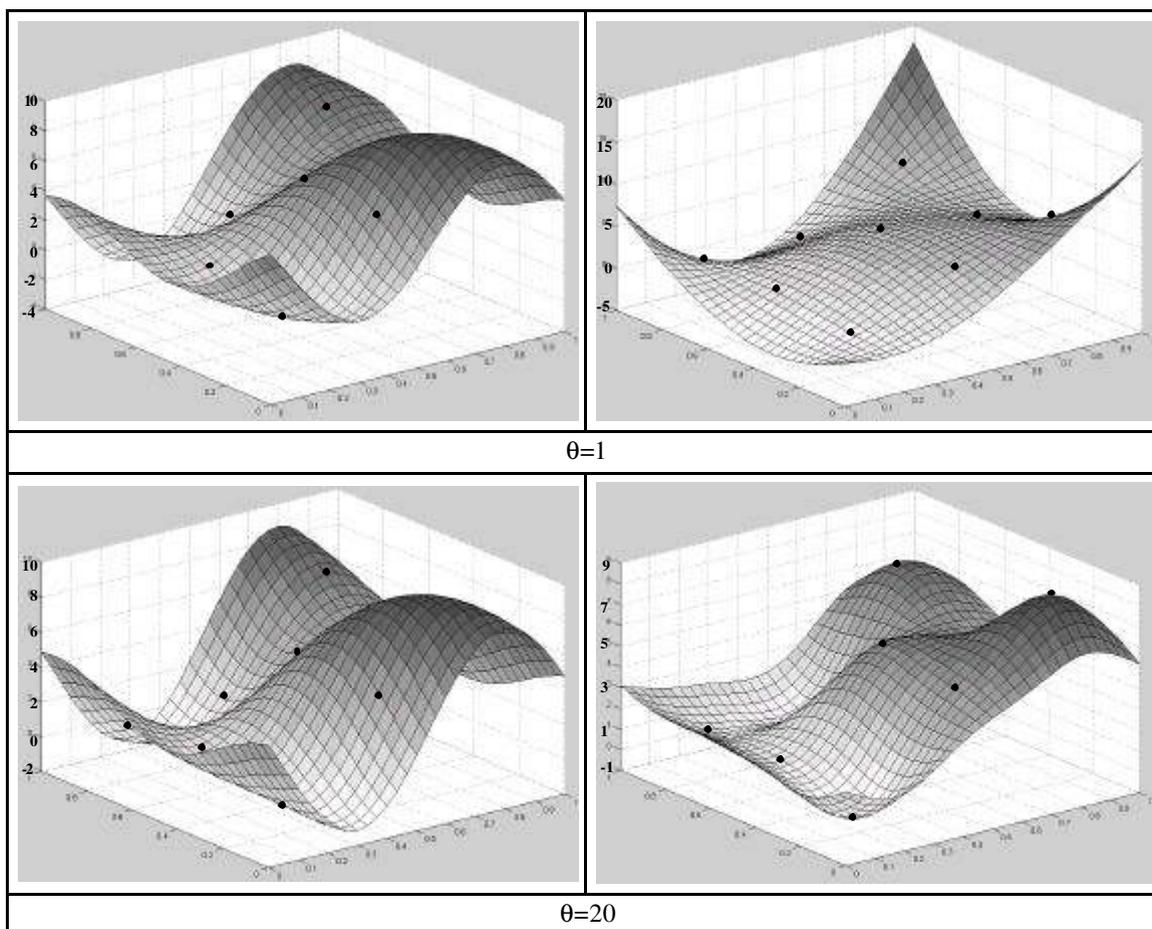
Il est donc impossible d'analyser l'effet simple  $(1,0,0)$  du facteur 1 et l'interaction  $(0,1,1)$  des facteurs 2 et 3 à l'aide de la fraction  $S$ .

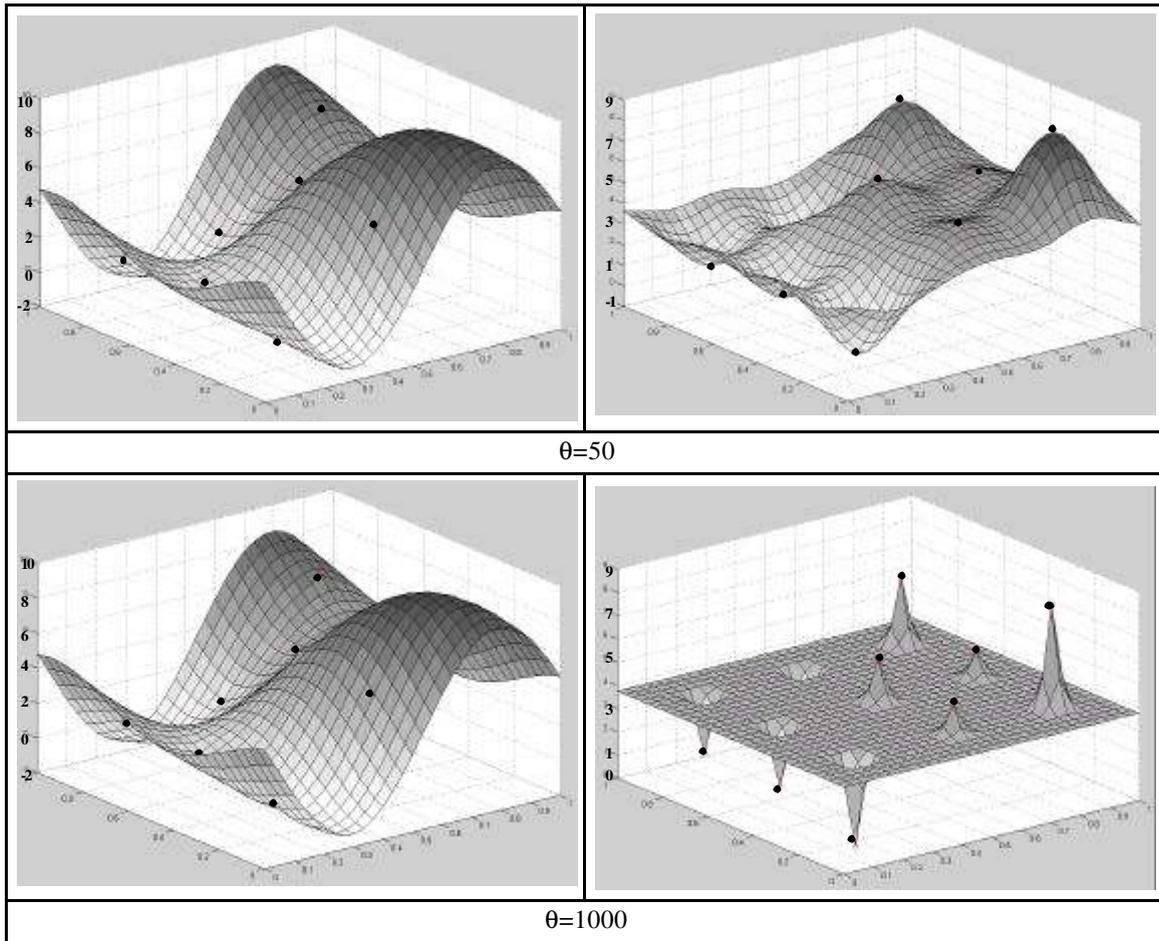
En revanche, si on choisit le modèle à effets simples,  $A^+=\{(1,0,0), (0,1,0), (0,0,1)\}$ , alors les fréquences sont toutes mutuellement orthogonales. Le sous-groupe  $S$  permet donc d'analyser ce modèle.

## 4.4. Régression trigonométrique dans l'approche par résidu aléatoire

Dans le chapitre un nous avons signalé que certains auteurs pensent que seul le processus résiduel joue un rôle important dans le modèle et que la régression n peut être réduite à une constante. Il semble au vue des quelques illustrations numériques présentées ici que cela ne soit pas le cas. Au contraire, plus la régression est riche, et plus le prédicteur est robuste aux variations de  $\theta$ .

Le tableau ci-dessous représente l'évolution du prédicteur construit dans le chapitre un, *i.e* pour un modèle sans terme d'erreur, en fonction de  $\theta$ . Dans la colonne de gauche le prédicteur est construit à partir de la régression trigonométrique définie par  $A^+ = \{(1,0), (0,1), (1,1)\}$  et dans la colonne de droite à partir d'une régression constante. Le plan utilisé correspond à  $3 \times 3$  points également espacés sur  $[0,1]^2$ , sans perturbation aléatoire. Les observations aux points du plan sont choisies de façon aléatoire entre 0 et 10. Elles ne correspondent à aucun phénomène en particulier, et sont les mêmes pour chaque simulation. Nous avons fait des essais avec d'autres observations et les résultats sont identiques.

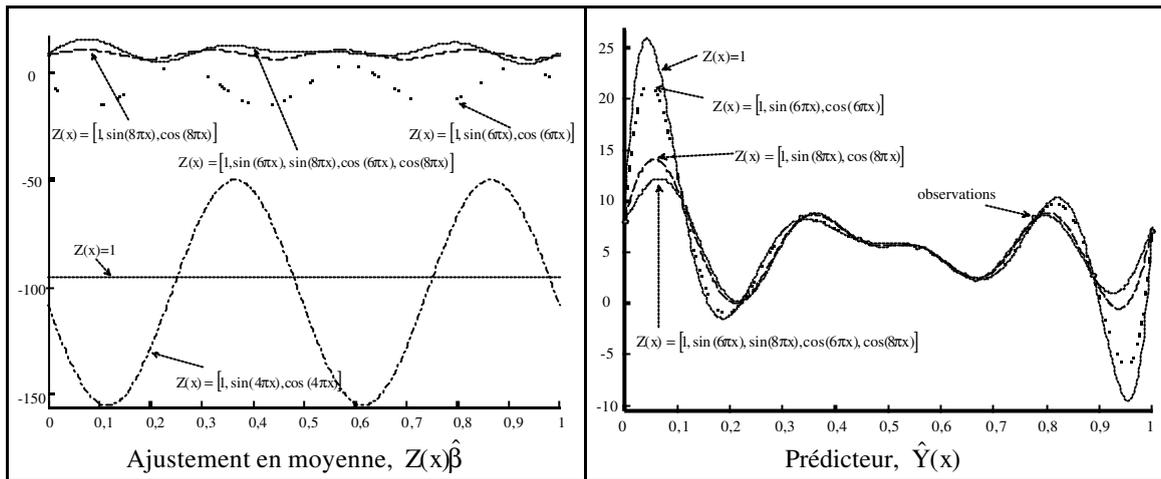




Evolution du prédicteur en fonction de  $\theta$ , à droite pour une régression constante, à gauche pour une régression trigonométrique

La stabilité du prédicteur de la colonne de gauche s'explique très bien par le fait que le rôle du processus résiduel est contre balancé par la modélisation en espérance. En effet, sur les graphiques ci-dessous nous avons représenté, à gauche l'ajustement en moyenne, *i.e*  $Z(x)\hat{\beta}$  pour différentes régressions trigonométriques, et à droite le prédicteur correspondant.

Il semble donc que plus la régression est riche, plus l'ajustement en moyenne est proche de la valeur réelle. Ce qui a pour conséquence de stabiliser le prédicteur et éviter de grandes variations. Par exemple, sur les graphes ci-dessous, les observations sont choisies au hasard entre 0 et 10, et pourtant, l'ajustement en moyenne pour une régression constante ou bien pour  $Z(x) = [1, \sin(4\pi x), \cos(4\pi x)]$  se situe aux alentours de  $-100$ . Les prédicteurs associés à ces deux régressions prennent alors des valeurs entre  $-10$  et  $25$  alors que les observations sont uniformément distribuées entre 0 et 10. En revanche l'ajustement en moyenne pour la régression  $Z(x) = [1, \sin(6\pi x), \sin(8\pi x), \cos(6\pi x), \cos(8\pi x)]$  est meilleur, ce qui entraîne un prédicteur plus lisse.



Le prédicteur ci-dessus est celui du chapitre un. Le plan comprend 10 points également espacés sur  $[0,1]$  sans deuxième degré. Et les observations aux points du plan sont choisies de façon aléatoire entre 0 et 10 et ne correspondent à aucun phénomène particulier. Sur cet exemple  $\theta=5$ , c'est-à-dire que la matrice de corrélation est mal conditionnée. Le fait que  $Z(x)$  soit plus riche permet de modérer l'erreur due à ce mauvais conditionnement

## Bibliographie

Bailey R.A. (1982). Dual abelian group in the design of experiments. In *Algebraic structures and applications* (P. Schultz *et al.*, eds.), New York, 45-54.

Bailey R.A. (1985). Factorial design and abelian groups. *Linear Algebra Appl.* **70**, 349-368.

Bailey R.A. (1990). Cyclic design and factorial designs. In *Proceedings of of the R.C. Bose Symposium, Delhi 1988 on probability, statistics and design of experiments* (R.R. Bahadur, K. Sen, Eds.), New York, 45-54.

Bates R.A., Buck R.J., Riccomagno E., Wynn H.P. (1996). Experimental design and observation for large systems. *J. R. Statist. Soc. B* **58**, 77-94.

Bates R.A., Riccomagno E., Schwabe R., Wynn H.P. (1998). Lattices and dual lattices in optimal experimental design for Fourier models. *Computational Statistics and Data Analysis* **28**, 283-296.

Collombier D. (1996). *Plans d'expérience factoriels*. Springer-Verlag, Berlin.

Hoel P.G. (1965). Minimax designs in two dimensional regression. *Ann. Math. Statist.* **36**, 1097-1106.

Kiefer J., Wolfowitz J. (1959). Optimum design in regression problems. *Ann. Math. Statist.* **30**, 271-294.

Kobilinsky A. (1985a). Orthogonal factorial designs for quantitative factors. *Statistics and Decisions, Supplement* **2**, 272-347.

Kobilinsky A. (1985b). Confounding in relation to duality of finite abelian groups. *Linear Algebra Appl.* **70**, 321-347.

Kobilinsky A. (1990). Complex linear models and cyclic designs. *Linear Algebra Appl.* **127**, 227-282.

Kobilinsky A., Monod H. (1991). Experimental design generated by group morphisms : An introduction. *Scand. J. Statist.* **18**, 119-134.

Riccomagno E., Schwabe R., Wynn H.P. (1997). Lattice-based D-optimum design for Fourier regression. *Ann. of statist.*

Serres J.P. (1967). *Représentations linéaires des groupes finis*. Herman, Paris.



## 5. Les plans

En ce qui concerne le modèle, nous avons suivi le principe de l'approche bayésienne, c'est-à-dire introduire des corrélations entre les points à l'aide du processus résiduel. En revanche, il semble clair à l'issue de la première partie que les plans présentant les meilleures qualités sont ceux de l'approche fréquentiste. Nous allons donc définir une catégorie de plans ayant les mêmes caractéristiques, autrement dit nous allons privilégier la répartition spatiale des points. Pour se faire, nous avons choisi des structures connues en plans d'expériences sous le nom de **fractions régulières**.

Ces dernières présentent deux caractéristiques principales qui ont retenues notre attention.

Tout d'abord, elles possèdent une bonne répartition de leurs points dans le domaine d'échantillonnage. Dans le chapitre deux, nous avons vu qu'il existe deux écoles à cette notion de répartition. La première concerne la représentation des marges du cube unité par le plan, avec les hypercubes latins et les tableaux orthogonaux. C'est principalement celle-ci que vérifient les fractions régulières. La répartition marginale est alors définie par la **résolution** de la fraction. La deuxième méthode utilise la fonction de discrédance pour mesurer la qualité de la répartition des points, notamment avec la structure de réseau. Nous verrons aussi que les fractions régulières peuvent être associées à cette catégorie de plans.

D'autre part, les fractions régulières sont particulièrement bien adaptée à l'ajustement d'une régression trigonométrique. Cela vient du fait que leur définition est basée sur la structure de sous-groupe. En effet, dans le chapitre précédent, nous avons vu que dans ce cas particulier, les conditions d'estimabilité des paramètres de la régression sont très simples et entraîne l'optimalité universelle dans le cas d'un modèle linéaire simple.

Nous allons montrer que les conditions concernant la résolution de la fraction régulière, ainsi que l'estimabilité des paramètres dépendent uniquement du dual de celle-ci. Cet ensemble jouant un rôle prépondérant, il est nécessaire de savoir le déterminer. Pour cela nous nous plaçons dans le cas où  $G$  est le groupe additif d'un corps de Galois. Il existe alors des techniques très simples pour passer de la matrice génératrice de la fraction à celle du dual. Nous étudions plus particulièrement une catégorie de plan présentant des propriétés intéressantes, appelés **tableaux orthogonaux linéaires**.

Dans une dernière partie nous allons montrer à partir d'un exemple numérique, en quoi des plans, doit l'emploi est tout à fait justifié dans le cas d'un modèle linéaire simple, conservent de bonne propriétés lorsqu'on ajoute un processus résiduel.

## 5.1. Fractions régulières

Soit le groupe abélien fini d'exposant  $q = \text{PPCM}(q_i, i=1, \dots, d)$

$$G = \prod_{i=1}^d (\mathbb{Z} / q_i).$$

Nous avons vu que tout plan d'échantillonnage est associé à un sous-ensemble de  $G$ . Dans le cadre de ce travail, nous avons choisi d'imposer une structure de classes latérales de sous-groupes de  $G$  aux plans. Cette forme particulière est bien connue en plans d'expériences sous le nom de **fractions régulières**.

### Définition 5.1

*On appelle fraction régulière tout sous-groupe  $S$  de  $G$ , ou plus généralement, toute classe latérale de ce sous-groupe, c'est-à-dire toute partie de  $G$  de la forme*

$$z+S$$

*où  $z \in G$  est fixé.*

*$S$  s'appelle alors la direction,  $S^\circ$  les contrastes de définition et  $S^\perp$  le dual de la fraction régulière, où  $S^\circ = \{\chi \in G^*, \chi(s) = 1 \ \forall s \in G\}$  et  $S^\perp = \{h \in G, (s|h) = 0 \ \forall s \in G\}$  sont les sous-groupes définis au paragraphe 4.3.*

Dans un premier temps nous allons présenter des propriétés de répartition spatiale des points d'une fraction régulière dans le cube unité. Ensuite, nous allons étudier l'ajustement d'une régression trigonométrique par une fraction régulière.

### 5.1.1. Distribution marginale d'une fraction régulière

En ce qui concerne les plans, dans cette nouvelle approche, nous avons choisi de privilégier la distribution marginale des points comme critère de répartition spatiale, avec la notion de résolution d'une fraction régulière.

### Définition 5.2

*On note  $E_i = \{[0, 1/q_i[, \dots, [(q_i-1)/q_i, 1]\}$ ,  $i \in I$ , l'ensemble des segments de la  $i$ ème arrête de  $[0, 1]^d$ . Alors pour tout  $J \subseteq I$ , on appelle  $J$ -marge le produit cartésien*

$$\prod_{i \in J} E_i.$$

*Les éléments de la  $J$ -marge s'appellent des cellules.*

*On dit que la  $J$ -marge est complète quand toutes ses cellules sont représentées au moins une fois dans le plan, et uniforme lorsque les cellules sont représentées uniformément par le plan.*

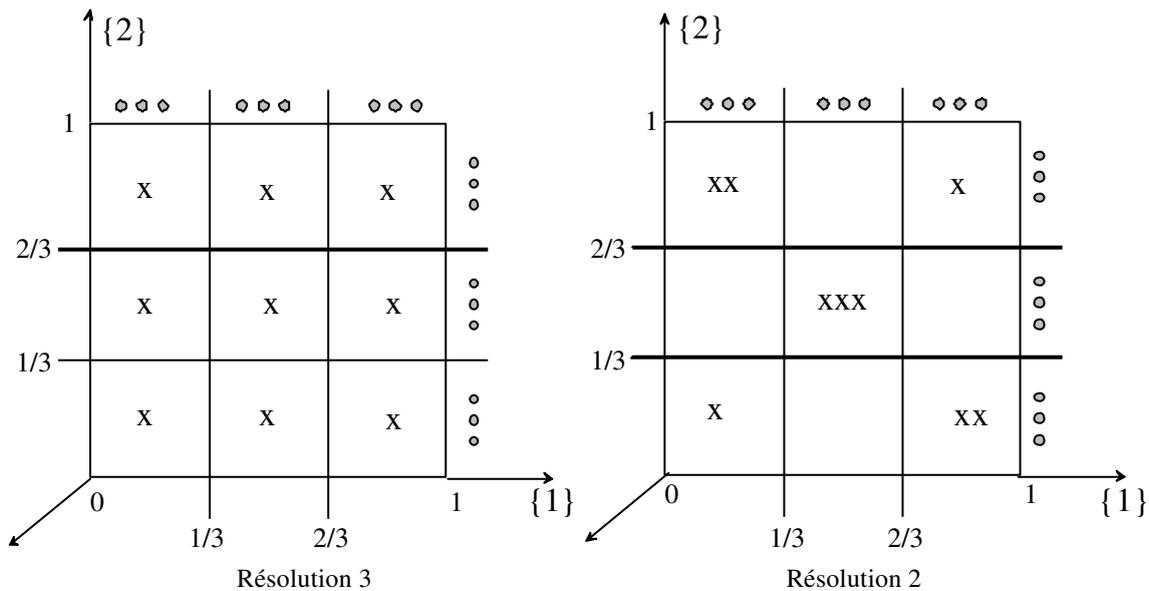
### Définition 5.3

*On dit qu'une fraction régulière est de résolution  $R$  si pour tout  $J \subseteq I$  tel que  $|J|=R+1$ , la  $J$ -marge est complète et uniforme.*

La définition ci-dessus de la résolution d'une fraction n'est pas la plus couramment utilisée. Elle n'est pas sans rappeler la notion de force d'un tableau orthogonal définie dans le chapitre deux. Et en effet, nous avons la relation suivante (cf. Collombier (1996) pour la démonstration)

**Une fraction régulière de résolution  $R$  est un tableau orthogonal de force  $R-1$ . En particulier une fraction régulière de résolution 2 est un hypercube latin.**

En fait ici, on impose en plus une structure de sous-groupe au plan par rapport au chapitre deux. Nous avons choisi cette définition de la résolution (équivalente uniquement dans le cas de fractions régulières) car nous nous intéressons essentiellement à la répartition marginale des points du plan. Les deux schémas ci-dessous permettent de mieux comprendre pourquoi nous pensons que la résolution est un bon indicateur d'une répartition spatiale de qualité.



Les croix correspondent aux cellules de la  $\{1,2\}$ -marge représentées dans le plan et les points aux cellules de le  $\{1\}$ -marge et  $\{2\}$ -marge. Les deux fractions sont de taille 9, celle de droite est de résolution 2 et celle de gauche de résolution 3. On remarque que certaines cellules de la  $\{1,2\}$ -marge ne sont pas représentées dans le plan pour la fraction régulière de résolution 2. Il paraît donc légitime de penser que plus la résolution est élevée, mieux le domaine d'échantillonnage est représenté par la fraction.

La proposition suivante, dont on peut trouver la démonstration dans l'article de Bailey (1985), permet de vérifier très simplement la résolution d'une fraction régulière à partir de son dual.

#### Définition 5.4

Soit  $v=(v_1, \dots, v_d)$  un vecteur de  $G$ .

On appelle support de  $v$ , et on note  $\text{supp}(v)$ , l'ensemble des indices des coordonnées non nulles de  $v$ .

On appelle poids de  $v$ , et on note  $\text{wt}(v)$ , le nombre de coordonnées non nulles de  $v$ ,  
 $\text{wt}(v)=\text{card}(\text{supp}(v))$ .

On appelle poids d'un sous-groupe de  $G$  le plus petit poids de ses vecteurs à l'exception de l'élément nul.

#### Proposition 5.5

Une fraction régulière de direction  $S$  est de résolution  $R$  si et seulement si son dual  $S^\perp$  est de poids supérieur ou égal à  $R$ .

Nous pouvons aussi énoncer ce résultat sous forme matricielle mais pour cela nous devons définir quelques notions de vocabulaire.

#### Définition 5.6

Soient  $x_1, \dots, x_k$  des éléments du  $\mathbb{Z}$ -module  $G$ , linéairement indépendants et  $o_1, \dots, o_k$  leur ordre respectifs. On dit que  $S$  est le sous-groupe de  $G$  engendré par  $x_1, \dots, x_k$  si

$$S = \bigoplus_{i=1}^k \langle x_i \rangle,$$

où  $\langle x_i \rangle$  est le sous-groupe cyclique de  $G$  d'ordre  $o_i$  engendré par  $x_i$ .

On dit qu'une matrice  $M$  est génératrice de  $S$  si les lignes de  $M$  sont linéairement indépendantes et engendrent  $S$ . Le poids de  $M$  est le poids du sous-groupe engendré.

La proposition précédente peut alors se formuler ainsi

#### Proposition 5.7

Une fraction régulière de direction  $S$  est de résolution  $R$  si et seulement si la matrice génératrice du dual  $P$  est de poids supérieur ou égal à  $R$ .

L'approche par échantillon aléatoire est celle qui propose les meilleurs plans. Nous avons vu qu'il existe deux écoles pour choisir un ensemble de points représentant au mieux toutes les parties du cube unité. Nous avons pris le parti de privilégier la distribution marginale des points comme critère de répartition spatiale. Nous venons d'expliquer ce choix d'un point de vue intuitif avec les illustrations précédentes, mais ce choix se justifie aussi d'un point de vue théorique grâce à un article très récent de Fang et Mukerjee (2000). En effet, ces auteurs font le lien entre une variante de la mesure de discrédance, appelée  $L_2$ -discrédance centrée et la distribution des poids dans le dual d'une fraction régulière à éléments dans  $\text{GF}(2)$ . D'après les propositions ci-dessus mettant en relation le dual et la résolution d'une fraction régulière, le lien entre répartition marginale et discrédance apparaît alors clairement. La  $L_2$ -discrédance

centrée a été introduite par Hickernell (1998) et permet en fait de considérer l'uniformité non seulement sur l'ensemble du cube unité, mais aussi des projections. Le résultat établi par Fang et Mukerjee confirme notre approche intuitive, c'est-à-dire un excellent comportement des fractions régulières de force élevée au regard de l'uniformité. Il est cependant à noter que leur travail ne concerne que les fractions régulières de GF(2) et que ce résultat paraît difficilement généralisable à GF(p) où p est un nombre premier. En effet, tout leur raisonnement repose sur le fait que la L<sub>2</sub>-discrépance centrée laisse les points du plan d'échantillonnage invariants par symétrie par rapport à tout plan passant par le centre du cube unité et parallèle à ses faces. On voit donc l'intérêt d'utiliser cette variante de la discrédance dans le cas où p=2.

### 5.1.2. Lien entre réseaux et fractions régulières

Dans notre travail, nous avons aussi établi un lien entre les deux écoles 'répartition marginale' et 'discrépance'. En effet, nous montrons dans ce paragraphe qu'à toute fraction régulière, on peut associer un réseau et *vice-versa*, où rappelons-le les réseaux forment une catégorie de plans de faible discrédance

#### Passage d'un réseau d'échantillonnage à une fraction régulière

Soit L un réseau d'échantillonnage de Z-base canonique

$$\left( \frac{a_1}{q_1}, \dots, \frac{a_r}{q_r} \right).$$

On note  $q = \text{PPCM}(q_i, i=1, \dots, r) = q_1$  et pour  $i=1, \dots, r$  on définit les vecteurs de  $Z^d$ ,

$$g_i = q \frac{a_i}{q_i}.$$

Ces vecteurs appartiennent à  $(Z/q)^d$  puisque par définition on a

$$0 \leq a_{ij} < q_i \Leftrightarrow 0 \leq q \frac{a_{ij}}{q_i} < q, j=1, \dots, d.$$

D'autre part  $g_i = q a_i$  où  $a_i$  est un vecteur d'entiers, d'où  $g_i = 0 \pmod{q}$ , donc  $g$  est un vecteur d'ordre  $q$ ,  $i=1, \dots, r$ . Soit  $D_i$  le sous-groupe cyclique de  $(Z/q)^d$ , d'ordre  $q$  engendré par  $g_i$ , montrons que ces sous-groupes sont disjoints. Supposons qu'il existe  $z_i \in (Z/q_i)$   $i=1, \dots, r$  tels que

$$\begin{aligned} \sum_{i=1}^r z_i g_i &= 0 \pmod{q} \Leftrightarrow \sum_{i=1}^r z_i g_i = qz \text{ où } z \in Z^d \\ \Leftrightarrow \sum_{i=1}^r z_i \frac{a_i}{q_i} &= z \Leftrightarrow \left\{ \sum_{i=1}^r z_i \frac{a_i}{q_i} \right\} = 0 \Leftrightarrow z_i = 0, \quad i = 1, \dots, r. \end{aligned}$$

Donc  $S = D_1 \oplus \dots \oplus D_r$  est un sous-groupe de  $(Z/q)^d$  d'ordre  $N = \prod q_i$ .

**Exemple 5.8 :** Soit L un réseau d'échantillonnage de Z-base canonique

$$\left( \frac{(1,2,1)}{6}, \frac{(1,1,1)}{2} \right).$$

On lui associe une fraction régulière de  $(Z/6)^3$ , d'ordre 12, de direction engendrée par les vecteurs (1,2,1) et (3,3,3).

**Passage d'une fraction régulière à un réseau d'échantillonnage**

Soit  $S$  la direction d'une fraction régulière de  $(\mathbb{Z}/q)^d$  où  $q$  est un entier quelconque.  $S$  est un groupe abélien fini, on peut donc l'écrire (de façon unique à un isomorphisme près) comme la somme directe de  $r$  groupes cycliques (décomposition canonique)

$$S = D_1 \oplus \dots \oplus D_r,$$

où chaque groupe cyclique  $D_i$  est engendré par un vecteur  $g_i$  de  $(\mathbb{Z}/q)^d$  d'ordre  $q_i$  tel que

$$q_{i+1} \text{ divise } q_i, \quad i=1, \dots, r \text{ et } q_1=q.$$

La fraction régulière est de taille  $N = \prod q_i$ .

On utilise la transformation inverse, *i.e* pour  $i=1, \dots, r$  on définit les vecteurs

$$a_i = q_i \frac{g_i}{q}.$$

Ces vecteurs appartiennent à  $\mathbb{Z}^d$ , car comme  $q_i$  est l'ordre de  $g_i$ , on a  $q_i g_i = 0 \pmod{q}$ , c'est-à-dire  $q_i g_i = qz$  où  $z \in \mathbb{Z}^d$ . D'autre part, on a

$$0 \leq g_{ij} < q \iff 0 \leq a_{ij} < q_i, \quad j=1, \dots, d.$$

De plus, s'il existe  $0 \leq z_i < q_i, i=1, \dots, r$  tels que

$$\left\{ \sum_{i=1}^r z_i \frac{a_i}{q_i} \right\} = 0 \iff \sum_{i=1}^r z_i \frac{a_i}{q_i} = z \in \mathbb{Z}^d$$

$$\iff \sum_{i=1}^r z_i g_i = 0 \pmod{q} \iff z_i = 0 \pmod{q_i}, \quad i = 1, \dots, r$$

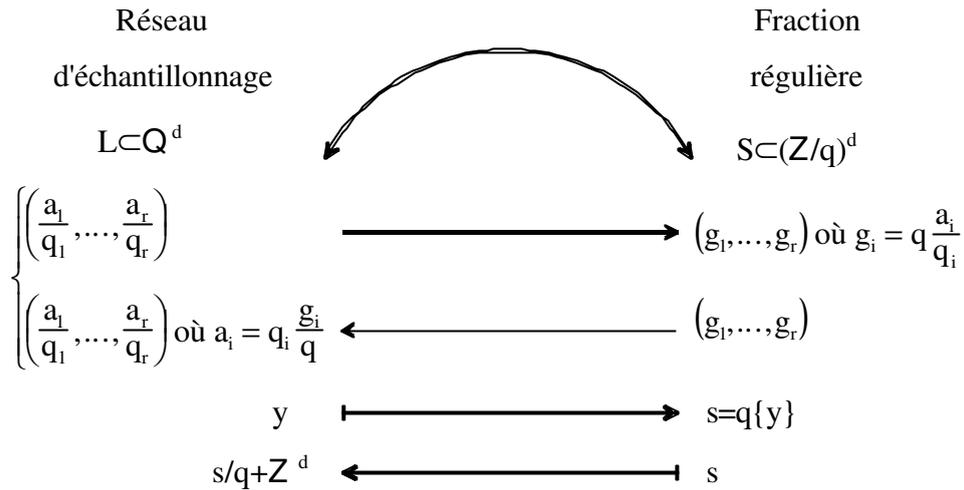
alors comme  $0 \leq z_i < q_i$ , on a  $z_i = 0, i=1, \dots, r$ . Finalement,

$$\left( \frac{a_1}{q_1}, \dots, \frac{a_r}{q_r} \right)$$

définit la  $\mathbb{Z}$ -base canonique d'un réseau d'échantillonnage de déterminant  $N^{-1}$ .

**Exemple 5.9 :** Soit la fraction régulière de  $(\mathbb{Z}/9)^3$  dont la direction est engendrée par les vecteurs  $(8,1,0)$ ,  $(8,0,1)$  d'ordre 9 et  $(3,0,0)$  d'ordre 3. On lui associe un réseau d'échantillonnage de taille  $3 \times 9^2$  et de  $\mathbb{Z}$ -base canonique

$$\left( \frac{(8,1,0)}{9}, \frac{(8,0,1)}{9}, \frac{(1,0,0)}{3} \right).$$



Nous avons vu précédemment que le dual a un rôle important. La proposition suivante fait le lien entre les matrices génératrices d'un réseau d'échantillonnage et d'une fraction régulière

**Proposition 5.10**

Soit  $L$  un réseau d'échantillonnage de matrice génératrice  $M$  et soit  $S$  la fraction régulière de  $(\mathbb{Z}/q)^d$  que l'on associe à  $L$ . On note  $P$  la matrice génératrice de  $S^\perp$ . On a alors  $L^\perp = S^\perp \oplus q\mathbb{Z}^d$ , d'où

$$P = [{}^t M^{-1}] \pmod{q},$$

où  $[A]$  est la matrice formée des lignes de  $A$  non nulles modulo  $q$ .

**Exemple 5.11 :** Reprenons l'exemple 5.8 du passage d'un réseau à une fraction. Pour avoir une matrice génératrice du réseau il faut compléter la  $\mathbb{Z}$ -base canonique par un vecteur quelconque de  $\mathbb{Z}^d$ , ici on choisit  $(0,0,1)$ .

$$M = \frac{1}{6} \begin{pmatrix} 1 & 2 & 1 \\ 3 & 3 & 3 \\ 0 & 0 & 6 \end{pmatrix} \longrightarrow P = \begin{pmatrix} 4 & 4 & 0 \\ 5 & 0 & 1 \end{pmatrix}.$$

Si maintenant on reprend l'exemple 5.9, la matrice génératrice du dual  $P$  est de la forme  $(3,3,3)$  (cf. Chapitre 7) et on lui associe la matrice génératrice du réseau  $M$

$$P = (3 \ 3 \ 3) \longrightarrow M = \frac{1}{9} \begin{pmatrix} 3 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

On retrouve la matrice génératrice de la  $\mathbb{Z}$ -base qui est égale ici à la  $\mathbb{Z}$ -base canonique.

**Démonstration :** Soit  $h \in L^\perp$ , alors d'après la proposition 2.19  $h \in \mathbb{Z}^d$ , on peut alors l'écrire  $h = g + qz$  où  $g \in (\mathbb{Z}/q)^d$  et  $z \in \mathbb{Z}^d$ . On sait que  $\forall y \in L (h|y) \in \mathbb{Z} \Leftrightarrow (g|y) \in \mathbb{Z}$ . Or pour tout  $y \in L$ , il

existe  $z \in \mathbb{Z}^d$  tel que  $y = \{y\} + z$  donc  $(h|y) \in \mathbb{Z} \Leftrightarrow (g|\{y\}) = \alpha \in \mathbb{Z} \Leftrightarrow (g|q\{y\}) = q\alpha \Leftrightarrow (g|s) = 0 \pmod{q} \forall s \in S$ , d'où  $g \in S^\perp$ .  $\square$

Cette proposition permet d'utiliser les réseaux d'échantillonnage pour ajuster un modèle trigonométrique de la même façon que pour les fractions régulières, c'est-à-dire en utilisant la notion de plan orthogonal à un modèle. Et réciproquement, nous allons voir que l'erreur d'intégration numérique définie dans le chapitre deux proposition 1.9, dépend du réseau uniquement au travers le dual de la fraction régulière qui lui est associée.

**Proposition 5.12**

*Soit  $f$  une fonction admettant un développement de Fourier absolument convergent. Soient  $L$  un réseau d'échantillonnage et  $S$  le sous-groupe de  $(\mathbb{Z}/q)^d$  associé. On note  $Qf$  l'approximation de l'intégrale de  $f$ ,  $If$ , sur les points d'échantillonnage de  $L$ . On a*

$$|Qf - If| = \sum_{g \in S^\perp} \left( \sum_{\substack{z \in \mathbb{Z}^d; \\ g+qz \neq 0}} \hat{f}(g + qz) \right),$$

*où  $\hat{f}(\cdot)$  sont les coefficients de Fourier de la série.*

**Démonstration :** On note  $x_1, \dots, x_N$  les points d'échantillonnage de  $L$  et  $s_1, \dots, s_N$  les éléments de  $S$ . On remarque tout d'abord que  $\hat{f}(0) = If$ , donc

$$\begin{aligned} |Qf - If| &= \left| \sum_{h \in \mathbb{Z}^d} \hat{f}(h) Qe^{2\pi i(h|x)} - \hat{f}(0) \right| = \sum_{h \neq 0 \in \mathbb{Z}^d} \hat{f}(h) Qe^{2\pi i(h|x)} \\ &= \frac{1}{N} \sum_{h \neq 0 \in \mathbb{Z}^d} \hat{f}(h) \sum_{j=1}^N \exp \left[ 2\pi i(h | x_j) \right] = \frac{1}{N} \sum_{h \neq 0 \in \mathbb{Z}^d} \hat{f}(h) \sum_{j=1}^N \exp \left[ \frac{2\pi i}{q} (h | s_j) \right] \end{aligned}$$

Tout  $h \neq 0 \in \mathbb{Z}^d$  s'écrit  $h = g + qz$  où  $g \in (\mathbb{Z}/q)^d$  et  $z \in \mathbb{Z}^d$ , tel que  $g + qz \neq 0$ , donc

$$|Qf - If| = \frac{1}{N} \sum_{g \in (\mathbb{Z}/q)^d} \sum_{\substack{z \in \mathbb{Z}^d; \\ g+qz \neq 0}} \hat{f}(g + qz) \sum_{j=1}^N \exp \left[ \frac{2\pi i}{q} (g | s_j) \right] \times \underbrace{\exp \left[ 2\pi i (z | s_j) \right]}_{=1}$$

Puisque  $S$  est un sous-groupe de  $(\mathbb{Z}/q)^d$ , nous savons que

$$\sum_{j=1}^N \exp \left[ \frac{2\pi i}{q} (g | s_j) \right] = \begin{cases} N & \text{si } g \in S^\perp \\ 0 & \text{sinon} \end{cases}$$

d'où le résultat.  $\square$

La liaison entre réseaux et fractions régulières ne se fait que dans le cas où le découpage du cube unité est symétrique, i.e  $G = (\mathbb{Z}/q)^d$ .

Dans ce paragraphe nous avons établi le lien entre fractions régulières et réseaux, mais contrairement à l'article de Fang et Mukerjee (2000), aucun résultat n'a pu être démontré concernant la relation entre résolution et discrédance. Il ne nous est donc pas permis d'affirmer d'un point de vue théorique que plus la résolution est élevée, mieux les points sont répartis dans le cube unité, c'est-à-dire plus la discrédance du réseau associé à la fraction régulière est faible. Cependant, nous verrons à partir de tests numériques dans le paragraphe 5.4, qu'effectivement la résolution est un critère de qualité d'un plan.

### 5.1.3. Ajustement d'une régression trigonométrique

La structure de sous-groupe que l'on impose aux plans est la deuxième raison pour laquelle nous avons choisi d'utiliser des fractions régulières. En effet, nous avons vu dans le chapitre précédent que dans le cas d'un sous-groupe les conditions d'estimabilité des paramètres d'une régression trigonométrique sont particulièrement simples à vérifier. Ces conditions restent les mêmes dans le cas d'une classe latérale puisque si on note  $h \in G$ ,  $\chi_h$  un caractère de représentation irréductible de  $G$ , et  $f_j = z + s_j$  où  $s_j \in S$ ,  $j=1, \dots, N$ , les éléments d'une fraction régulière de taille  $N$ , alors étant donné que  $\chi_h(f_j) = \chi_h(z) \times \chi_h(s_j)$ , le théorème 4.20 devient

$$\sum_{j=1}^N \chi_h(f_j) = \begin{cases} [\chi_h(z)]^{-1} N & \text{si } h \in S^\perp \\ 0 & \text{sinon} \end{cases}.$$

D'où la proposition sur l'estimabilité des paramètres et l'optimalité universelle dans le cas d'un modèle linéaire simple (sans corrélation).

#### Proposition 5.13

*Une fraction régulière de direction  $S$  est adaptée à un modèle si et seulement si elle est orthogonalement adaptée au modèle, c'est-à-dire, si et seulement si toutes les fréquences non nulles du modèle sont 2 à 2 mutuellement orthogonales par rapport à  $S^\perp$ .*

Nous pouvons remarquer que le dual joue un rôle déterminant, aussi bien en ce qui concerne la résolution de la fraction régulière, que l'estimabilité des paramètres d'une régression trigonométrique. Dans le paragraphe suivant, nous allons voir comment déterminer sa matrice génératrice.

## 5.2. Fractions régulières à éléments dans $GF(p)^d$

Le dual joue un rôle déterminant en ce qui concerne la résolution de la fraction régulière et l'estimabilité des paramètres d'un modèle. Dans le cas où  $G$  est un groupe abélien fini quelconque, déterminer le dual n'est pas une chose facile. En revanche, si on suppose que  $G$  est le groupe additif d'un corps de Galois d'ordre  $p$ , noté  $GF(p)^d$ , c'est-à-dire si on découpe chaque arrête du cube unité en un même nombre de segments  $p$ , où  $p$  est premier, alors le passage de la fraction régulière à son dual est simple. En effet, dans ce cas les fractions régulières sont directement construites à partir de leur dual. Cette méthode de construction a été proposée par Bose (1947). Elle est couramment utilisée et nous pouvons par exemple citer les articles de Patterson (1976), Raktoc *et al.* (1969), Bailey *et al.* (1977) et Patterson et Bailey (1978). Le principe est simple et se déduit de la proposition suivante.

### Proposition 5.14

*Soient  $P$  une matrice d'ordre  $r \times d$ , de rang  $r$ , à éléments dans  $GF(p)$  et  $c$  un vecteur de  $GF(p)^{d-r}$ . Le sous-ensemble de  $GF(p)^d$  formé des  $p^{d-r}$  solutions du système*

$$Pt=c$$

*est une fraction régulière de direction  $S=Ker P$ .  $P$  est la matrice génératrice du dual  $S^\perp$ .*

Inversement, pour une fraction régulière de la forme  $S+z$ , où  $S$  est un sous-groupe de  $GF(p)^d$  et  $z$  un vecteur de  $GF(p)^d$ . Les éléments de la fractions sont les solutions du système  $Pt=c$ , où  $P$  est la matrice génératrice du dual et  $c$  est un vecteur tel que  $c=Pz$ .

Dans la classification proposée par Srivastava et Chopra (1973), les plans définis par les solutions d'un système  $Pt=c$  appartiennent à la catégorie des tableaux orthogonaux de type I.

**Démonstration :** Soit  $z$  une des  $p^{d-r}$  solutions du système  $Pt=c$ , alors quel que soit  $t$  une solution du système, on a  $t-z$  qui appartient à  $S=Ker P$  puisque  $Pt=Pz=c$ . Donc  $t$ , qui peut s'écrire sous la forme  $t=z+(t-z)$  appartient bien à la fraction régulière  $S+z$ .  $\square$

Ainsi, on peut construire simplement un fraction régulière en deux étapes,

- la première pour déterminer la direction de la fraction en résolvant le système  $Pt=0$ ,
- la deuxième pour trouver une solution particulière au système.

Dans le cas général où  $G$  est un groupe cyclique quelconque, les ressources du calcul matricielle sont difficilement utilisables puisqu'on doit alors se placer dans un anneau qui n'est pas intègre. Les équations du système ne peuvent pas être résolues par inversion (généralisée)

de la matrice  $P$ . En revanche, dans le contexte que l'on s'est fixé, c'est-à-dire le groupe additif de  $GF(p)^d$ , ce procédé devient particulièrement avantageux grâce à la proposition qui suit.

**Proposition 5.15**

*Soient deux entiers positifs  $d$  et  $r$  tels que  $d > r$  et  $P$  une matrice  $r \times d$  à éléments dans  $GF(p)$ .*

*Alors il existe une matrice  $M$  carrée régulière d'ordre  $r$  et une matrice de permutation  $\Sigma$  d'ordre  $d$  telles que*

$$MP\Sigma = \left( \begin{array}{c|c} I_k & Q \\ \hline 0 & 0 \end{array} \right)$$

*où  $0 \leq k \leq r$  est le rang de  $P$  et  $Q$  est d'ordre  $k \times (d-k)$ .*

*Si les lignes de  $P$  sont indépendantes, i.e.  $rgP=k=r$ , l'équation  $Px=0$  admet  $p^{d-r}$  solutions et le sous-espace  $ker P$  qu'elles constituent est engendré par les lignes de*

$$B = \left( -{}^t Q \mid I_{d-r} \right) {}^t \Sigma .$$

*Inversement  $By=0$  si et seulement si  $y \in Im {}^t P$ .*

Donc une fraction régulière engendrée par les lignes de  $B = (-{}^t Q \mid I_{d-r})$  a pour matrice génératrice du dual  $P = (I_r \mid Q)$

**Remarque :** On dit que  $MP\Sigma$  est la représentation sous forme échelon de  $P$  ou bien  $MP\Sigma$  est une matrice échelon de  $P$ .

**Démonstration :** La première partie est un résultat classique d'algèbre linéaire. Supposons, sans perte de généralité, que  $P$  soit de la forme  $P = [I_r \mid Q]$  où  $r$  est le rang de  $P$  et  $Q$  est une matrice d'ordre  $r \times (d-r)$ . Si on note  $P^-$  un inverse généralisé de  $P$  alors par définition  $P^-P$  est un projecteur sur  $Ker P$ , d'où

$$Ker P = Im(I - P^-P).$$

Une  $g$ -inverse de  $P$  est une matrice de la forme  $P^- = [T \mid W]$  où  $T$  est une matrice carrée d'ordre  $r$ ,  $W$  est une matrice d'ordre  $(d-r) \times r$ , telles que  $T + QW = I_r$ . Prenons par exemple  $T = I_r$  et  $W = 0_{(d-r) \times r}$ , alors l'espace  $Ker P$  est engendré par les colonnes de

$$I - \begin{bmatrix} I_r \\ 0 \end{bmatrix} [I_r \mid Q],$$

c'est-à-dire par les lignes de  $B = (-{}^t Q \mid I_r)$ .  $\square$

Outre la simplicité de cette méthode, nous pouvons remarquer deux autres avantages à ce procédé, l'un concernant la résolution de la fraction régulière et l'autre l'estimabilité des paramètres du modèle.

### 5.2.1. La résolution

Puisque la matrice  $P$  est de rang  $r$  et s'écrit sous la forme  $P=[I_r|Q]$ , alors le lemme suivant simplifie le calcul de la résolution de la fraction.

#### Lemme 5.16

*Soit  $P$  une matrice d'ordre  $r \times d$ , de rang  $r$ , de la forme  $P=[I_r | Q]$  et soit la relation*

*$(R_k)$  : toute combinaison linéaire de  $k$  lignes de  $Q$*

*est de poids supérieur ou égal à  $R-k$ .*

*Alors  $wt(P) \geq t$  si et seulement si,  $\forall k=1, \dots, \min(r, R-1)$ ,  $(R_k)$  est vérifiée.*

*De plus  $wt(P)=R$  si et seulement si une des combinaisons linéaires est de poids  $R-k$ .*

Supposons les relations  $(R_1), \dots, (R_{k-1})$  vérifiées et notons  $x_1, \dots, x_k$ ,  $k$  lignes de la matrice  $Q$ . Le lemme suivant permet de vérifier si l'espace  $\langle x_1, \dots, x_k \rangle$ , qui correspond aux combinaisons linéaires des  $k$  lignes de  $Q$ , est de poids  $R-k$ .

#### Lemme 5.17

*Soit  $I=I(x_1, \dots, x_k)$  l'ensemble des indices  $i$ ,  $1 \leq i \leq d$ , pour lesquels  $x_{ij}$  est non nul pour un unique indice  $j$ ,  $1 \leq j \leq k$ . On note  $\bar{I}$  son complémentaire dans  $\{1, \dots, d\}$ . Alors pour un vecteur  $x \in GF(p)^d$ ,  $x(\bar{I})$  désigne le vecteur formé des composantes de  $x$  d'indices dans  $\bar{I}$ .*

*Supposons  $I$  non vide.*

*(i) si  $|I| \geq R-k$  alors  $\langle x_1, \dots, x_k \rangle$  est de poids  $R-k$*

*(ii) si  $|I| < R-k$  alors  $\langle x_1, \dots, x_k \rangle$  est de poids  $R-k$  si et seulement si les vecteurs  $x_I(\bar{I}), \dots, x_k(\bar{I})$  engendrent un espace de poids  $R-k-|I|$ .*

#### Remarques :

- 1) Si la condition  $I$  non vide n'est pas vérifiée, on choisit  $i \in \text{supp}(x_1)$  et on note  $u_p = x_{pi}x_1 - x_{1i}x_p$ ,  $2 \leq p \leq k$ . On a alors  $\langle x_1, \dots, x_k \rangle = \langle x_1, u_2, \dots, u_k \rangle$  et  $J(x_1, u_2, \dots, u_k)$  non vide puisque  $u_{pi} = 0$ ,  $2 \leq p \leq k$ .
- 2) Si  $|I| = R-k-1$ , les vecteurs  $x_1(\bar{I}), \dots, x_k(\bar{I})$  engendrent un espace de poids  $R-k-|I|=1$  si et seulement si ils sont indépendants.

**Exemple 5.18 :** Soit  $P$  la matrice d'ordre  $2 \times 5$  à éléments dans  $GF(5)$  de la forme

$$P = (I_2 | Q) \text{ où } Q = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 3 \end{pmatrix}.$$

Montrons que cette matrice est de poids  $R=4$ .

La relation  $(R_1)$  est vérifiée. Pour le relation  $(R_2)$ , nous avons  $I=\emptyset$ , d'après le 1) de la remarque, on pose  $u=x_1y-y_1x=(4,2,3)-4(1,1,1)=(0,3,4)$ , on a bien  $wt(u)=2=R-2$ . Avec ce nouveau vecteur,  $I$  est égal à  $\{1\}$ , i.e  $|I|=1=R-2$ , donc, d'après le 2) de la remarque, il faut que les vecteurs  $x(\bar{I})=(3,2)$  et  $u(\bar{I})=(3,4)$  soient libres pour que  $P$  soit de poids  $R=4$ . Ce qui est le cas.

**Démonstration :** Soit  $T$  le tableau dont les lignes sont formées par les  $x_1, \dots, x_k$ . Un élément de  $\langle x_1, \dots, x_k \rangle$  est une combinaison linéaire des lignes de  $T$ . Par hypothèse, toute combinaison linéaire de moins de  $k$  lignes de  $T$  strictement est de poids supérieur ou égal à  $R-k$ . Soit alors  $z$  une combinaison linéaire de toutes les  $k$  lignes de  $T$ , alors

$$wt(z)=wt(z(I))+wt(z(\bar{I})).$$

Nous avons de plus  $wt(z(I))=|I|$ . En effet, les colonnes de  $T$  indexées par  $I$  possèdent un et un seul élément non nul, donc dans toute combinaison linéaire de  $k$  lignes de  $T$  la partie restreinte à ces colonnes,  $z(I)$ , est de poids  $|I|$ . D'où

- (i) Si  $|I| \geq R-k$  alors  $wt(z) \geq R-k$ .
- (ii) Si  $|I| \leq R-k-1$ . Supposons tout d'abord  $\langle x_1(\bar{I}), \dots, x_k(\bar{I}) \rangle$  de poids  $R-k-|I|$ .  $z(\bar{I}) \in \langle x_1(\bar{I}), \dots, x_k(\bar{I}) \rangle$ , donc  $wt(z(\bar{I})) \geq R-k-|I|$ , d'où  $wt(z) \geq R-k-|I|+|I|=R-k$ . Réciproquement, si  $wt(z) \geq R-k$  alors  $wt(z(\bar{I}))=wt(z)-wt(z(I)) \geq R-k-|I|$ , donc  $\langle x_1(\bar{I}), \dots, x_k(\bar{I}) \rangle$  est de poids  $R-k-|I|$ .  $\square$

### Cas particulier où $P$ est de rang $r=2$

Dans ce cas la matrice  $P$  est de la forme

$$P = (I_2 \mid Q) \text{ où } Q = \begin{pmatrix} x_1 & \dots & x_{d-2} \\ y_1 & \dots & y_{d-2} \end{pmatrix} \text{ est une matrice d'ordre } 2 \times (d-2).$$

On note  $\det(m,n)=x_m y_n - x_n y_m$  le déterminant des colonnes  $m$  et  $n$  de la matrice  $Q$ . Le lemme suivant propose des conditions suffisantes pour que la matrice  $P$  soit de poids  $R$ . Elles sont particulièrement simples à vérifier, notamment d'un point de vue algorithmique.

### **Lemme 5.19**

Soit  $R$  un entier tel que  $2 \leq R \leq \min(d-1, q)$ . Si  $x$  et  $y$  vérifient les conditions suivantes,

$$\left\{ \begin{array}{ll} x_i \neq 0 & i = 1, \dots, d-2 \quad (1) \\ y_i \neq 0 & i = 1, \dots, d-2 \quad (2) \\ \det(1, p) \neq 0 & p = 2, \dots, d-2 \quad (3) \\ \dots & \\ \det(R-3, p) \neq 0 & p = R-2, \dots, d-2 \quad (R-1) \\ \det(R-2, R-1) \neq 0 & (R) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \text{Pour } k=1, \dots, R-3 \\ \det(k, p) \neq 0 \quad p=k+1, \dots, d-2 \quad (k + \end{array} \right.$$

alors la matrice  $P$  est de poids  $R$ .

Ces conditions sont nécessaires uniquement dans le cas où  $R=d-1$ .

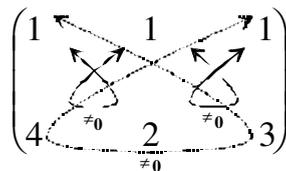
**Exemple 5.18 (suite) :** Soit  $P$  la matrice d'ordre  $2 \times 5$  à éléments dans  $GF(5)$  de la forme

$$P = (\mathbb{I}_2 \mid Q) \text{ où } Q = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 3 \end{pmatrix}.$$

Les conditions du lemme précédent sont nécessaires et suffisantes et s'écrivent

$$\begin{cases} x_i \neq 0 \text{ et } y_i \neq 0 & i = 1, 2, 3 \\ \det(1, 2) \neq 0 \text{ et } \det(1, 3) \neq 0 \\ \det(2, 3) \neq 0 \end{cases}$$

Elles sont rapidement vérifiées, ce qui confirme que la matrice  $P$  est de poids 4



**Démonstration :** 1) Dans un premier nous allons montrer par récurrence la relation suivante.

$\forall k=2, \dots, R-2$ , on note

$$\begin{cases} x'_k = (x_k, \dots, x_{d-2}) \\ y'_k = \left( \prod_{i=1}^{k-2} x_i \right) \times (\det(k-1, k), \dots, \det(k-1, d-2)) \\ w_k = R - k - 1 \end{cases}$$

alors  $\langle x'_k, y'_k \rangle$  est de poids  $w_k$ .

Au rang  $R-2$ , il faut montrer que  $\langle x'_{R-2}, y'_{R-2} \rangle$  est de poids 1, c'est-à-dire, d'après la remarque du lemme 5.17, que  $x'_{R-2}$  et  $y'_{R-2}$  sont linéairement indépendants. Remarquons que deux vecteurs  $x=(x_1, \dots, x_d)$  et  $y=(y_1, \dots, y_d)$  sont indépendants s'il existe  $m$  et  $n$  tels que  $\det(m, n) \neq 0$ . En effet, supposons que  $x$  et  $y$  soient liés alors il existe  $\alpha \in GF(p)$  tel que  $x_m = \alpha y_m$  et  $x_n = \alpha y_n$ , d'où  $\det(m, n) = x_m y_n - x_n y_m = \alpha y_m y_n - \alpha y_n y_m = 0$ . Dans notre cas nous avons

$$\begin{aligned} (x'_{R-2})_1 (y'_{R-2})_2 - (x'_{R-2})_2 (y'_{R-2})_1 &= \left( \prod_{i=1}^{R-4} x_i \right) \times (x_{R-2} \det(R-3, R-1) - x_{R-1} \det(R-3, R-2)) \\ &= \left( \prod_{i=1}^{R-3} x_i \right) \times \det(R-2, R-1) \neq 0 \text{ d'après les relations (1) et (R).} \end{aligned} \quad \text{Su}$$

supposons la relation vérifiée au rang  $k+1$  et montrons qu'elle est vérifiée au rang  $k$ , i.e  $\langle x'_k, y'_k \rangle$  est de poids  $R-k-1$ . D'après les conditions (1) et (k+2), nous avons  $I(x'_k, y'_k) = \emptyset$ , on pose donc

$$\begin{aligned} u_k &= (x'_k)_1 y'_k - (y'_k)_1 x'_k \\ &= \left( \prod_{i=1}^{k-2} x_i \right) \times (0, x_k \det(k-1, k+1) - \det(k-1, k) x_{k+1}, \dots, x_k \det(k-1, d-2) - \det(k-1, k) x_{d-2}) \\ &= \left( \prod_{i=1}^{k-1} x_i \right) \times (0, \det(k, k+1), \dots, \det(k, d-2)). \end{aligned}$$

d'après la condition  $(k+3)$ ,  $\text{wt}(u_k)=d-k-2 \geq w_k$  puisque  $R \leq d-1$ . D'après la remarque du lemme 5.17,  $\langle x'_k, u_k \rangle = \langle x'_k, y'_k \rangle$  et on a maintenant  $I(x'_k, u_k) = \{1\}$ .  $|I| = 1 \geq w_k - 1$  puisque  $2 \geq k \geq R-2$ , donc d'après le lemme 5.17  $\langle x'_k, y'_k \rangle$  est de poids  $w_k$  si et seulement si  $\langle x'_k(\bar{I}), y'_k(\bar{I}) \rangle$  est de poids  $w_k - 1$ . Or

$$\begin{aligned} x'_k(\bar{I}) &= (x_{k+1}, \dots, x_{d-2}) = x'_{k+1} \\ y'_k(\bar{I}) &= \left( \prod_{i=1}^{k-1} x_i \right) (\det(k, k+1), \dots, \det(k, d-2)) = y'_{k+1}, \end{aligned}$$

et par hypothèse la relation est vérifiée au rang  $k+1$ , i.e que  $\langle x'_{k+1}, y'_{k+1} \rangle$  est de poids  $w_{k+1} = w_k - 1$ .

2) Montrons maintenant que  $\langle x, y \rangle$  est de poids  $R-2$ . Il faut vérifier que les relations  $(R_1)$  et  $(R_2)$  du lemme 5.16. D'après les conditions (1) et (2),  $(R_1)$  est vérifiée. D'après les conditions (1) et (3),  $I(x, y) = \emptyset$ , on pose donc  $u = x_1 y - y_1 x = (0, \det(1, 2), \dots, \det(1, d-2))$ . D'après la condition (3),  $\text{wt}(u) = d-3 \geq R-2$  puisque  $R \geq d-1$ .  $\langle x, u \rangle = \langle x, y \rangle$  et  $I(x, u) = \{1\}$ , donc d'après le lemme 5.17,

Si  $R \leq 3$  alors  $|I| \geq R-2$  donc  $\langle x, y \rangle$  est de poids  $R-2$

Si  $R > 3$  alors  $|I| < R-2$  donc  $\langle x, y \rangle$  est de poids  $R-2$  si et seulement si  $\langle x(\bar{I}), y(\bar{I}) \rangle$  est de poids  $R-3$ . Or  $x(\bar{I}) = (x_2, \dots, x_{d-2}) = x'_2$  et  $y(\bar{I}) = (\det(1, 2), \dots, \det(1, d-2)) = y'_2$ , et d'après le 1)  $\langle x'_2, y'_2 \rangle$  est de poids  $w_2 = R-3$ .

### 5.2.2. L'estimabilité

Cette technique de construction a un deuxième avantage qui vient du fait que l'on peut rapidement vérifier si la fraction régulière est orthogonalement adaptée à un modèle donné. En effet, rappelons que toutes les fréquences du modèle doivent être deux à deux mutuellement orthogonales par rapport à  $S^\perp$ , le dual de la direction de la fraction régulière, ce qui se traduit par  $h \notin S^\perp$ ,  $m \notin S^\perp$ ,  $h+m \notin S^\perp$  et  $h-m \notin S^\perp$   $h \neq m$ ,  $\forall h \in A^+$ ,  $\forall m \in A^+$ . Dans le cas d'une fraction régulière à éléments dans  $\text{GF}(p)$ , il est nettement plus simple de vérifier si un vecteur appartient au dual  $S^\perp$ .

#### Proposition 5.20

*Soient  $h$  un vecteur de  $\text{GF}(p)^d$  et  $S^\perp$  le dual d'une fraction régulière engendré par une matrice  $P$  d'ordre  $r \times d$ , de rang  $r$ , de la forme  $P = [I_r \mid Q]$ . On décompose  $h$  en deux parties,  $h = [h_r \mid h_{d-r}]$  où  $h_r$  (resp.  $h_{d-r}$ ) est un vecteur d'ordre  $r$  (resp. d'ordre  $d-r$ ).*

*Pour que  $h$  n'appartienne pas à  $S^\perp$ , il faut et il suffit que*

$$h_r Q \neq h_{d-r}.$$

**Exemple 5.21 :** Soit la matrice  $P$  d'ordre  $2 \times 3$  à éléments dans  $\text{GF}(3)$

$$P = \left( \begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array} \right).$$

La matrice P est de poids 2, c'est-à-dire que la fraction S dont la direction est engendrée par la matrice  $B=(2 \ 2 \ 1)$  est de résolution 2. Soit  $h=(h_1,h_2,h_3) \in GF(3)^3$ , d'après la proposition précédente

$$h \notin S^\perp \Leftrightarrow h_1+h_2 \neq h_3.$$

On remarque alors que la fraction n'est pas très intéressante puisqu'elle ne permet d'analyser qu'une seule fréquence en plus de l'effet moyen, par exemple  $h=(1,0,0)$ , un effet simple du premier facteur. On vérifie de façon exhaustive que l'on ne peut pas ajouter d'autres fréquences. Supposons par exemple que l'on veuille analyser en plus les effets simples  $(0,1,0)$  et  $(0,0,1)$ , alors nous avons

$$(1,0,0)-(0,1,0)=(1,2,0) \in S^\perp \text{ puisque } h_1+h_2=h_3,$$

$$(1,0,0)+(0,0,1)=(1,0,1) \in S^\perp \text{ puisque } h_1+h_2=h_3.$$

Ces fréquences ne sont donc pas mutuellement orthogonales deux à deux.

**Démonstration :** La démonstration est triviale et se fait à partir soit de la matrice P, soit de la matrice B génératrice de la fraction.  $h$  appartient à  $S^\perp$  si et seulement si  $h$  est combinaison linéaire des lignes de P, or P est de la forme  $[I_r \mid Q]$ , donc nécessairement, il faut et il suffit que  $h_r Q = h_{d-r}$ .  $\square$

Les fractions régulières, à éléments dans  $GF(p)$ , a donné lieu à une abondante bibliographie. Leur construction a été largement développée, notamment dans le cas où  $p=2$ . Nous pouvons par exemple citer les travaux de Franklin et Bailey (1977), Franklin (1985) et Collombier (1996). C'est pourquoi nous ne nous attardons pas sur ce type de fractions en général, mais nous nous consacrons à une classe plus restreinte ayant des propriétés particulières, que l'on appelle tableaux orthogonaux linéaires.

### 5.3. Tableaux orthogonaux linéaires

Les tableaux orthogonaux linéaires forment une classe particulière de fractions régulières de  $GF(p)^d$ , définis de la façon suivante.

**Définition 5.22**

*Soit  $T$  un tableau d'ordre  $N \times d$ , à éléments dans un groupe abélien fini  $E$ . On dit qu'il est orthogonal de force  $t$  sur  $E^d$  où  $1 \leq t \leq d$ , si dans tout bloc formé de  $t$  colonnes de  $T$ , les éléments de  $E^t$  figurent un même nombre de fois  $\lambda$ . On le note  $OA_\lambda(E^d, t)$ .*

*Si de plus  $E$  est le groupe additif d'un corps fini  $K$  et si les lignes de  $T$  sont toutes distinctes et constituent un sous-espace vectoriel de  $K^d$ , on dit que  $T$  est un tableau orthogonal linéaire. Dans ce cas on le note  $OA_\lambda(K^d, t)$ .*

On suppose que le corps fini  $K$  est le corps de Galois d'ordre  $p$ ,  $GF(p)$ .

**Remarque :**  $\lambda$  est appelé l'index du tableau,  $p$  le nombre de symboles et  $N$  la taille. On a nécessairement  $N = \lambda p^t$ . On note aussi parfois le tableau  $OA(N, d, p, t)$ . Pour des raisons de coût et afin d'éviter les répétitions, nous choisirons de préférence des tableaux orthogonaux d'index  $\lambda = 1$ .

Là encore nous pouvons faire le lien avec les plans de l'approche par échantillon aléatoire. Tout d'abord, de part sa définition, un tableau orthogonal linéaire de force  $t$ , n'est autre qu'un tableau orthogonal de force  $t$  avec une structure de sous-espace vectoriel. Notamment, un tableau orthogonal linéaire de force 1 est un hypercube latin. D'autre part, un tableau orthogonal linéaire étant une fraction régulière, nous savons qu'il est possible de l'associer à un réseau. Mais dans ce cas particulier, nous pouvons préciser le type du réseau et généraliser la proposition 2.20 sur les réseaux de rang 1 et les hypercubes latins.

**Proposition 5.23**

*A tout tableau orthogonal linéaire  $T$ , de force  $t$ , d'index 1, à éléments dans  $GF(p)$ , on associe un réseau d'échantillonnage de rang  $t$  et d'invariants tous égaux à  $p$ .*

**Démonstration :** Un tableau orthogonal linéaire d'index 1 de taille  $N = p^t$  est fixé dès lors que sont donnés  $t$  vecteurs de  $GF(p)^d$ ,  $g_1, \dots, g_t$ , qui forment une base du sous-espace de  $GF(p)^d$ ,  $S$ , constitué des lignes de  $T$ . C'est à dire que tout vecteur de  $S$  se décompose de façon unique sur ces  $t$  vecteurs, donc

$$S = D_1 \oplus \dots \oplus D_t$$

où  $D_k = \langle g_k \rangle$  est le sous-groupe de  $S$  engendré par  $g_k$   $k=0, \dots, t$ .

Les  $D_k$  sont des sous-groupes cycliques de  $S$  d'ordre  $p$ . En effet, si on note  $d_k$  l'ordre de  $D_k$ , alors

$$d_k > 1 \text{ (sinon } D_k \text{ est réduit à l'élément neutre) et } \prod_{k=1}^t d_k = p^t.$$

Ce qui implique nécessairement, dans le cas où  $p$  est premier, que  $d_k = p$   $k=1, \dots, t$ .

Montrons alors que cette décomposition en groupes cycliques est minimale. Supposons que l'on additionne deux de ces groupes cycliques  $D = D_k \oplus D_{k'}$ , tout élément de  $D$  s'écrit  $g = \alpha_k g_k + \alpha_{k'} g_{k'}$  où  $\alpha_k \in GF(p)$  et  $\alpha_{k'} \in GF(p)$ . Le fait que  $D_k$  et  $D_{k'}$  soient d'ordre  $p$  entraîne que les deux vecteurs générateurs  $g_k$  et  $g_{k'}$  sont des éléments d'ordre  $p$ , d'où

$$pg = \alpha_k (pg_k) + \alpha_{k'} (pg_{k'}) = 0,$$

*i.e* tout élément de  $D$  est d'ordre  $p$ . Or  $D$  est d'ordre  $p^2$ , il ne peut donc pas être cyclique.

$S$  peut être décomposé en, au minimum,  $t$  groupes cycliques, chacun d'ordre  $p$ , donc d'après ce que nous venons de dire, on peut lui associer un réseau de rang  $t$  et d'invariants tous égaux à  $p$ .  $\square$

Dans le chapitre deux, nous nous sommes intéressés à la qualité de projection des réseaux en dimension inférieure à  $d$ . Nous pouvons considérer que la résolution d'une fraction régulière confère elle aussi, des propriétés de projection aux plans. Nous avons à ce propos le résultat suivant.

### Proposition 5.24

*Un tableau orthogonal linéaire d'index 1, de force quelconque, est associé à un réseau à projections régulières.*

**Démonstration :** Soient  $T$  un tableau orthogonal linéaire d'index 1, de force  $t$ , à éléments dans  $GF(p)$ , et  $L$  le réseau qui lui est associé.

Comme nous venons de le voir, ce réseau est de rang  $t$  et d'invariants  $q_1 = \dots = q_t = p$ . On peut étendre la  $Z$ -base canonique du réseau, en posant  $q_{t+1} = \dots = q_d = 1$  et  $a_{t+1}, \dots, a_d$  des vecteurs entiers quelconques.

Projeter le réseau en dimension  $s$  revient à sélectionner  $s$  colonnes du tableau orthogonal, soit  $T'$  le tableau formé de ces  $s$  colonnes.

Si  $t \leq s \leq d$ , alors les lignes de  $T'$  sont toutes distinctes car  $T$  est un tableau orthogonal de force  $t$  et d'index 1, donc tout élément de  $GF(p)^t$  est représenté une et une seule fois dans tout bloc de  $t$  colonnes de  $T$ . Le réseau obtenu par projection est alors d'ordre  $p^t$  et d'invariants  $q_1 = \dots = q_t = p$  et  $q_{t+1} = \dots = q_s = 1$ . La projection en dimension  $s$  est régulière.

Si  $1 \leq s \leq t$ , alors nous avons la propriété suivante. Un tableau orthogonal linéaire de force  $t$ , d'index 1 est un tableau orthogonal linéaire de force  $s$  et d'index  $\lambda = p^{t-s}$  (car  $N = \lambda p^s = p^t$ ). C'est-à-dire que les points de  $T'$  sont répétés  $p^{t-s}$  fois. Or étant donné que l'on omet les répétitions dans le réseau projection, celui-ci est obligatoirement d'ordre  $p^s$  et d'invariants  $q_1 = \dots = q_s = p$ . Là aussi, la projection en dimension  $s$  est régulière.  $\square$

Il semble que le critère de projection régulière proposé par Sloan et Joe (1994) soit un peu faible et ne permette pas de distinguer un tableau orthogonal de force 1 d'un tableau de force 2, alors que par définition les projections en dimension 2 sont de meilleure qualité pour la force 2.

### Tableaux orthogonaux linéaires de force 1

Un tableau orthogonal linéaire de force 1 est entièrement défini par un vecteur  $g=(g_1, \dots, g_d) \in GF(p)^d$ , tel que  $g_i \neq g_j$ ,  $i \neq j$ . Supposons sans perte de généralité que  $g_1=1$ . Ce tableau est de taille  $N=p$  et d'après la proposition 5.15 la matrice génératrice du dual  $S^\perp$  est de la forme

$$P = \left( \begin{array}{c|c} -g_2 & \\ \dots & \\ -g_d & \end{array} \middle| I_{d-1} \right).$$

La proposition 5.20 sur les conditions d'estimabilité des paramètres se simplifie par

$$h \notin S^\perp \iff (h | g) \neq 0.$$

On peut lui associer un réseau de rang 1 de matrice génératrice

$$M = \left( \begin{array}{cccc|c} 1/p & g_2/p & \dots & g_d/p & \\ \hline 0 & & & & I_{d-1} \end{array} \right)$$

Dans la troisième partie, nous proposons des méthodes de construction de tableaux orthogonaux linéaires de force  $t$ , puis de fractions régulières de  $G = \times(Z/q_i)$ , à partir de leur matrice génératrice du dual. Mais auparavant nous allons tester numériquement ces plans pour savoir comment ils se comportent par rapport aux critères de sélection proposés dans le chapitre un, *i.e* IMSE, entropie et distance.

## 5.4. Tests numériques sur les tableaux orthogonaux linéaires

Nous savons déjà que ces plans répondent aux principales qualités requises puisque nous pouvons les assimiler aux plans du chapitre deux. Ils représentent de façon uniforme toutes les parties du domaine d'échantillonnage. Ils peuvent être utilisés pour tous les modèles (sous réserve d'estimabilité des paramètres). Et quand ils existent, leur coût de construction est relativement faible.

Ce que nous voulons savoir maintenant, c'est la qualité de ces plans dans le cas d'une approche par résidu aléatoire, c'est-à-dire comment ils se comportent par rapport aux critères de IMSE, d'entropie et de distance. Pour cela nous avons choisi d'étudier un exemple numérique très simple

On se fixe une fonction connue

$$f(x) = \frac{\sum_{h \in A^+} \cos(h|x) + \sin(h|x)}{1 + \sum_{h \in A^+} (h|x)}, \quad \forall x \in [0,1]^d,$$

que l'on va prédire par une approche par résidu aléatoire, *i.e* un modèle sans terme d'erreur et un échantillonnage à un seul degré.

A l'aide des différents résultats de ce chapitre, nous avons construit numériquement tous les tableaux orthogonaux linéaires à 3, 4 et 5 facteurs avec  $p=3, 5$  ou  $7$ , en indiquant leur force respective. Nous allons étudier l'efficacité de ces plans pour prédire la fonction  $f$ . Pour la partie fixe du modèle, nous avons choisi une régression trigonométrique définie par l'ensemble des fréquences,

$$A^+ = \{(1,0,\dots,0), \dots, (0,\dots,0,1)\}.$$

Ce choix a été fait de façon à s'assurer que la partie fixe du modèle est appropriée à la prédiction de  $f$ , puisqu'on cherche principalement, à étudier le comportement des plans par rapport au processus résiduel. Celui-ci est choisi gaussien de la forme

$$R(x, u) = \exp\left(-\theta \|x - u\|^2\right), \quad \forall x \in [0,1]^d, u \in [0,1]^d.$$

Afin de mesurer la différence entre la fonction et le prédicteur, nous allons utiliser l'erreur quadratique moyenne intégrée empirique (EISE) sur une grille de  $10^d$  points de  $[0,1]^d$  (Sacks *et al.* (1989)) pour un  $\theta$  fixé,

$$\text{EISE}_\theta = \frac{1}{10^d} \sum_{i=1}^{10^d} \left( \hat{Y}(x_i) - f(x_i) \right)^2.$$

Pour un coût fixé,  $N$ , le ou les meilleurs plans sont alors ceux qui minimisent cette erreur. Les plans testés appartiennent tous à la classe des tableaux orthogonaux linéaires.

### 5.4.1. Robustesse des meilleurs plans par rapport aux variations de $\theta$

Des premiers tests ont montré que, sauf exception, il n'existe pas de plan qui minimisent l'erreur quel que soit  $\theta$ . En fait, on constate qu'il existe 3 types de plans, ceux qui minimisent l'erreur pour les faibles, les moyennes et les fortes corrélations. La force de la corrélation correspond à des valeurs de  $\theta$  qui dépendent de  $N=p^t$  et  $d$ . On note  $D_\theta^*$  le plan qui minimise  $EISE_\theta$  pour  $N$  fixé, où  $\theta$  représente l'une des 3 catégories, faibles, moyennes ou fortes corrélations. Il se peut que pour certains de ces paramètres  $p$  et  $t$ , il n'existe que deux catégories faibles-moyennes et fortes, voire même dans certains cas une seule catégorie, ce qui signifie alors que le meilleur plan  $D_\theta^*$  minimise l'erreur quel que soit  $\theta$ .

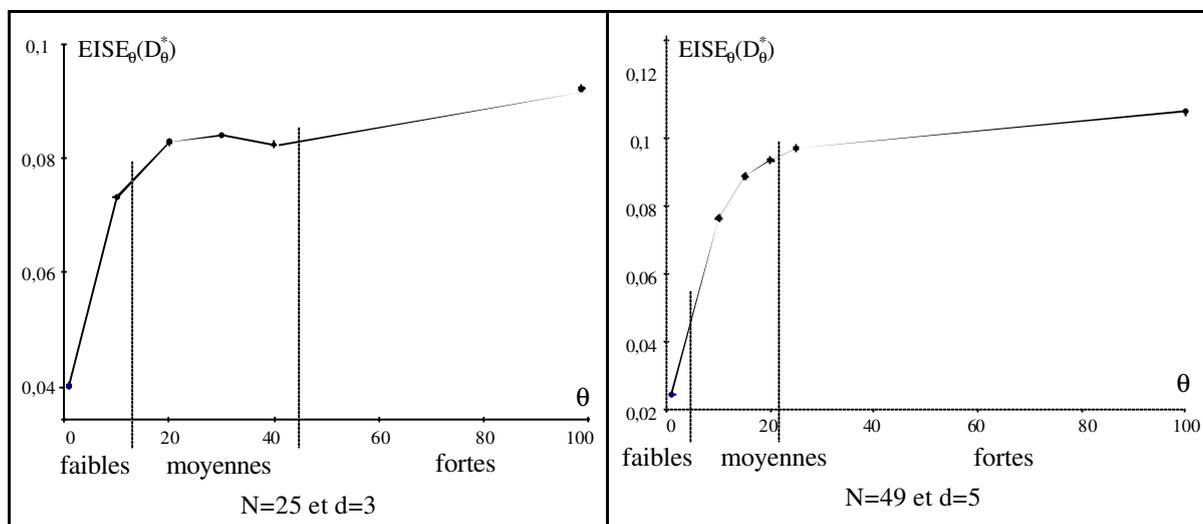


Fig. 1 Evolution de l'erreur en fonction de  $\theta$ .

Notre but n'est pas de déterminer quel paramètre de corrélation choisir, puisque, comme nous l'avons signalé dans le chapitre un, il existe des techniques d'estimation pour cela (Mardia et Marshall (1984), Christensen (1990)). On peut toutefois constater sur cet exemple, que l'erreur diminue lorsque les corrélations augmentent ( $\theta$  diminue), avec un seuil inférieur pour  $\theta$  (en général 1) en dessous duquel la matrice  $R$  est numériquement instable, et un seuil supérieur (en général 100) au dessus duquel  $R$  devient proche de la matrice identité, donc plus de corrélation.

Nous avons choisi ici d'étudier la robustesse des meilleurs plans des 3 catégories lorsque  $\theta$  varie d'une catégorie à l'autre. Ceci est une propriété non négligeable étant donné les problèmes qu'engendre l'estimation du paramètre de corrélation en pratique. Pour cela on calcule le coefficient d'efficacité (Sacks *et al.* (1989)),

$$(1) \quad EISE_\theta(D_\theta^*)/EISE_{\theta_i}(D_{\theta_i}^*),$$

où  $\theta_i$  appartient à une des 2 autres catégories que  $\theta$ .

$\theta$ $\theta_i$	Fortes	Moyenne s	Faibles
Fortes	100%	82%	80%
Moyenne s	95%	100%	80%
Faibles	7%	53%	100%

N=25 et d=3

$\theta$ $\theta_i$	Fortes	Moyenne s	Faibles
Fortes	100%	100%	72%
Moyenne s	95%	100%	80%
Faibles	22%	61%	100%

N=49 et d=4

Tab. 1 : Tableaux d'efficacité des meilleurs plans lorsque  $\theta_i$  varie

On remarque alors que  $D_{\theta}^*$ , le meilleur plan pour une forte ou moyenne corrélation, est très robuste aux changements de  $\theta$ . *A contrario*, l'efficacité de  $D_{\theta}^*$ , le meilleur plan pour une faible corrélation ( $\theta$  grand), devient catastrophique si on utilise une petite valeur de  $\theta$  (forte corrélation). Donc quand on ne connaît pas la valeur de  $\theta$  *a priori*, il est préférable d'utiliser les meilleurs plans des fortes et des moyennes corrélations. Les matrices génératrices de ces tableaux orthogonaux linéaires sont données ci-dessous pour  $d=3,4$  et 5. Il suffit alors de multiplier cette matrice par la matrice d'ordre  $p^d \times d$  formée des vecteurs de  $GF(p)^d$  pour obtenir le tableau orthogonal linéaire.

**Notation :**  $C_r = N/(2d+1)$  représente le coût relatif, *i.e* la taille du plan divisée par le nombre de paramètres à estimer dans la régression trigonométrique.

Caractéristiques				Fortes corrélations		Moyennes corrélations	
t	N	p	Cr	Efficacité	$D_{\theta}^*$	Efficacité	$D_{\theta}^*$
2	9	3	1.3	$\begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}^{(*)}$			
2	25	5	3.6	100% / Moy. 72% / Faibles	$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}$	95% / Fortes 80% / Faibles	$\begin{pmatrix} 3 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix}$
2	49	7	7	81% / Faibles	$\begin{pmatrix} 3 & 1 & 0 \\ 5 & 0 & 1 \end{pmatrix}$	Pas de moyenne corrélation	

Tab. 2 : Matrices génératrices des meilleurs plans à 3 facteurs pour les fortes et moyennes corrélations.

(\*) Pour ces paramètres on ne détecte qu'une seule catégorie, *i.e* que  $D_{\theta}^*$  minimise l'erreur quel que soit  $\theta$ . On ne peut donc pas calculer le coefficient d'efficacité (1).

Caractéristiques				Fortes corrélations		Moyennes corrélations	
t	N	p	Cr	Efficacité	$D_{\theta}^*$	Efficacité	$D_{\theta}^*$
2	25	5	2.8	99% / Moy. 93% / Faibles	$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix}$	50% / Fortes 96% / Faibles	$\begin{pmatrix} 2 & 3 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix}$ $\begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 & 0 \\ 3 & 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 3 & 0 & 1 \end{pmatrix}$
3	27	3	3	100% / Moy. 87% / Faibles	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}$	96% / Fortes 90% / Faibles	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{pmatrix}$
2	49	7	5.4	82% / Moy. 80% / Faibles	$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 5 & 6 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{pmatrix}$	95% / Fortes 80% / Faibles	$\begin{pmatrix} 4 & 2 & 1 & 0 \\ 4 & 6 & 0 & 1 \end{pmatrix}$
3	125	5	13.9	98% / Moy. 76% / Faibles	$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{pmatrix}$	61% / Fortes 82% / Faibles	$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}$

Tab. 3 : Matrices génératrices des meilleurs plans à 4 facteurs pour les fortes et moyennes corrélations.

Caractéristiques				Fortes corrélations		Moyennes corrélations	
t	N	p	Cr	Efficacité	$D_{\theta}^*$	Efficacité	$D_{\theta}^*$
2	25	5	2.3	87% / Faibles	$\begin{pmatrix} 3 & 2 & 3 & 1 & 0 \\ 3 & 4 & 4 & 0 & 1 \end{pmatrix}$	Pas de moyenne corrélation	
2	27	3	2.4	99% / Faibles	$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix}$	Pas de moyenne corrélation	
2	49	7	4.4	95% / Moy. 90% / Faibles	$\begin{pmatrix} 2 & 1 & 3 & 1 & 0 \\ 3 & 4 & 6 & 0 & 1 \end{pmatrix}$	80% / Fortes 89% / Faibles	$\begin{pmatrix} 2 & 1 & 2 & 1 & 0 \\ 3 & 4 & 6 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 3 & 1 & 0 \\ 2 & 3 & 5 & 0 & 1 \end{pmatrix}$
4	81	3	7.3	91% / Faibles	$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix}$	Pas de moyenne corrélation	
3	125	5	11.3	82% / moy. 82% / Faibles	$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 1 & 0 & 1 & 0 \\ 3 & 4 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 2 & 3 & 0 & 0 & 1 \end{pmatrix}$	82% / Fortes 95% / Faibles	$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 3 & 4 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 1 & 0 \\ 3 & 4 & 0 & 0 & 1 \end{pmatrix}$

Tab. 4 : Matrices génératrices des meilleurs plans à 5 facteurs pour les fortes et moyennes corrélations.

### 5.4.2. Influence du paramètre p et de la force du tableau sur l'erreur

Un tableau orthogonal linéaire est essentiellement déterminé par deux paramètres, sa force bien sûr, mais aussi par p, le découpage du cube unité. Nous allons étudier l'influence de ces deux paramètres sur la qualité du plan.

A l'issue de ces tests, on remarque que globalement l'erreur diminue lorsque le nombre de points du plan augmente, mais surtout que cette diminution dépend de la valeur de p. En effet, **plus le découpage du domaine expérimental est fin (p grand), plus l'erreur est petite et ceci à moindre coût.** Prenons par exemple le cas des plans à 5 facteurs,

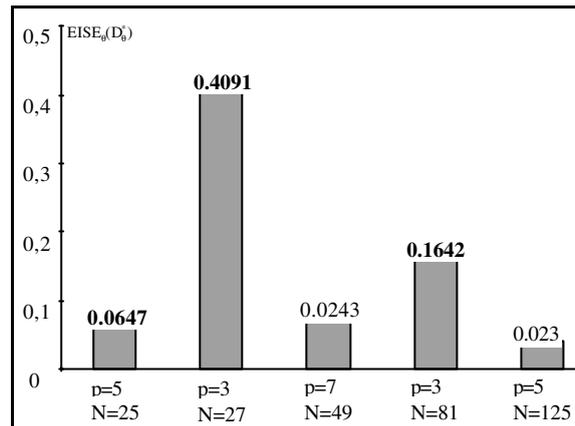


Fig. 2 : Evolution de l'erreur en fonction de p pour les plans à 5 facteurs

Pour  $N=25,27$  et  $81$ , nous avons les plus petites erreurs que l'on puisse trouver quel que soit  $\theta$ , ce qui veut dire qu'avec un plan de  $27$  points et  $p=3$ , on commet une erreur  $2,5$  fois plus grande que pour un plan pratiquement de même taille,  $N=25$ , mais avec  $p=5$ . Pour  $N=49$  il se peut que l'erreur soit inférieure à  $0,0243$  avec un  $\theta$  plus petit, mais cela ne change rien au fait qu'il est préférable d'utiliser ce plan que celui  $2$  fois plus grand, de taille  $81$  avec  $p=3$ .

Une fois que  $p$  est fixé, il reste à choisir  $t$ , la force du tableau orthogonal linéaire. On remarque alors que, **pour une taille fixée,  $N$ , les plans qui minimisent l'erreur sont toujours de force maximale et ceci quel que soit  $\theta$** . Attention toutefois, un tableau de force maximale n'est pas nécessairement d'erreur minimale.

Ces résultats confirment l'intuition que nous avons eu précédemment, c'est-à-dire qu'une bonne répartition marginale des points du plan est un critère de qualité.

La taille du plan est proportionnelle à  $p^t$ , il faut donc trouver un bon compromis entre ces deux paramètres de façon à garder un coût raisonnable. Nous conseillons alors la stratégie suivante

**Premièrement, privilégier la finesse de découpage du cube**

**Deuxièmement, choisir ensuite la force du tableau de façon à ce que la taille soit acceptable.**

En effet, comme on peut le voir sur la figure 2 et le tableau 4, il est préférable d'utiliser le tableau de taille  $49$ , de force  $2$ , au tableau de taille  $81$  et de force  $4$ .

A l'issue de ces tests numériques, il semblerait donc que les plans les plus appropriés soient les hypercubes latins (tableaux de force  $1$ ). Ce qui vient contre-dire l'hypothèse que nous avons émise sur le fait que, plus la force d'un tableau est élevée, mieux les points d'échantillonnage sont répartis dans le cube unité, et donc de meilleure qualité sera le tableau (hypothèse confirmée dans l'approche par échantillon aléatoire avec la variance de l'estimateur de l'intégrale qui diminue quand la force augmente §2.2.4). Il faut cependant nuancer ce résultat suivant le type de simulateur étudié. Il est possible que dans le cas d'un simulateur prenant en compte un grand nombre de paramètres d'entrée avec un temps de calcul moyennement long, il est préférable d'utiliser un hypercube latin, alors que dans le cas

d'un simulateur à peu de paramètres d'entrée mais un temps de calcul très long, un tableau orthogonal de force plus élevée semble plus indiqué.

### 5.4.3. Test des critères de sélection des plans

En pratique il est impossible de sélectionner un plan à l'aide de la EISE puisque l'on ne connaît pas la réponse du simulateur aux points de la grille. Même avec l'exemple que nous traitons, calculer l'erreur sur  $10^d$  points devient très long quand  $d > 5$ . Nous avons vu dans le chapitre un qu'il existe des critères pour sélectionner des plans de bonne qualité dans le cas d'une approche par résidu aléatoire.

Le critère d'entropie

max :  $\det(R)$

Le critère de distance

max :  $\min_{1 \leq i < j \leq N} \|x_i - x_j\|$ .

Nous avons testé l'efficacité et la particularité de ces critères avec les tableaux orthogonaux linéaires à 3, 4 et 5 facteurs. On remarque 2 choses.

Premièrement, il existe un lien évident entre la force du tableau et les critères. **Pour les plans de taille fixée, plus t est important et plus la distance ou le déterminant est grand.**

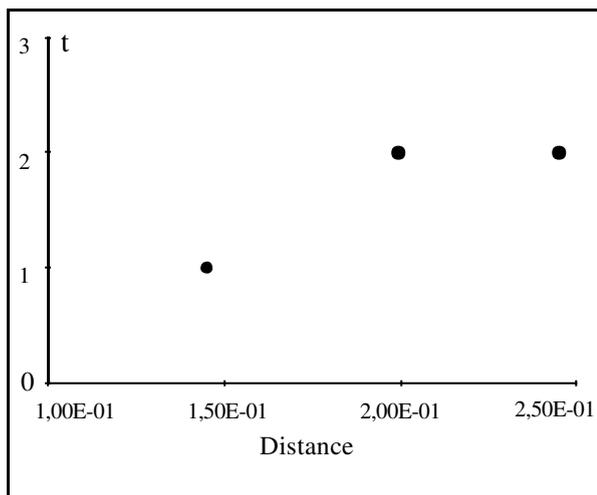


Fig. 3 : Force en fonction de la distance pour les plans de taille 49 à 3 facteurs.

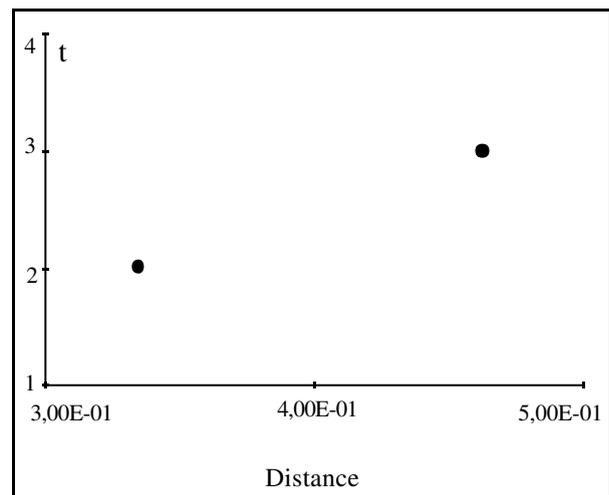


Fig. 4 : Force en fonction de la distance pour les plans de taille 27 à 4 facteurs.

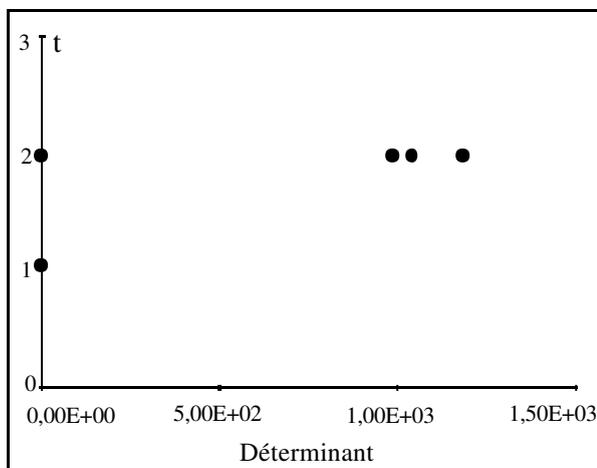


Fig. 5 : Force en fonction du déterminant pour les plans de taille 25 à 3 facteurs.

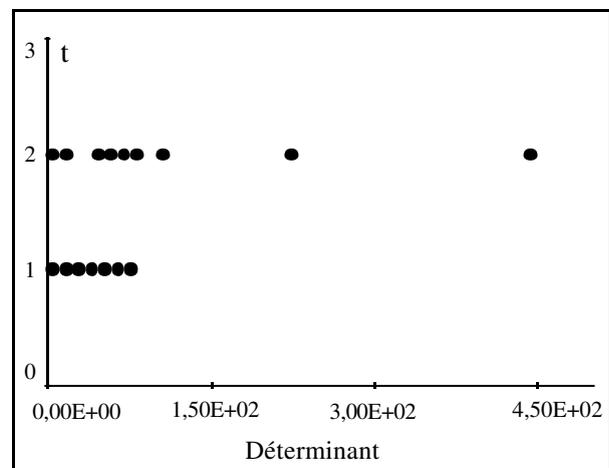


Fig. 6 : Force en fonction du déterminant pour les plans de taille 25 à 5 facteurs.

De plus, pour les tableaux de même taille et de force maximale, un plan de déterminant maximal et toujours de distance maximale.

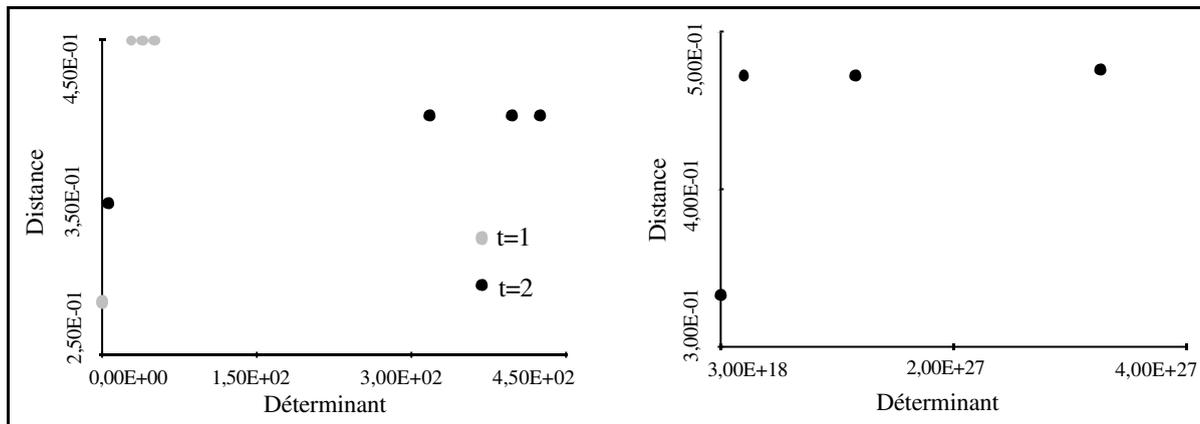


Fig. 7 : Distance en fonction du déterminant pour les plans de taille 25 à 4 facteurs.

Fig. 8 : Distance en fonction du déterminant pour les plans de taille 27 à 5 facteurs.

Deuxièmement, les plans sélectionnés par ces critères ne minimisent pas l'erreur EISE. On est donc en droit de se demander s'ils sont tout de même efficaces par rapport aux meilleurs plans. Les tableaux suivants donnent les matrices génératrices des tableaux orthogonaux linéaires de force maximale et de déterminant maximal, notés  $D_{det}^*$ , ainsi que leurs coefficients d'efficacité par rapport aux meilleurs plans des catégories fortes et faibles corrélations.

N	$D_{det}^*$	Fortes corrélations	Faibles corrélations
9	$\begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$	100% (**)	
25	$\begin{pmatrix} 2 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}$	95%	80%
49	$\begin{pmatrix} 3 & 1 & 0 \\ 5 & 0 & 1 \end{pmatrix}$	100%	81%

Tab. 5 : Tableau d'efficacité des plans à 3 facteurs de déterminant maximal.

N	$D_{det}^*$	Fortes corrélations	Faibles corrélations
25	$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix}$	100%	93%
27	Tous les plans ont même déterminant		
49	$\begin{pmatrix} 1 & 4 & 1 & 0 \\ 3 & 6 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 5 & 0 & 1 \end{pmatrix}$	59%	86%
125	$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}$	61%	82%

Tab. 6 : Tableau d'efficacité des plans à 4 facteurs de déterminant maximal.

(\*\*) Il n'y a qu'une seule catégorie de corrélation pour N=9 et d=3 (voir tab. 2)

N	$D_{\det}^*$	Fortes corrélations	Faibles corrélations
25	$\begin{pmatrix} 3 & 2 & 3 & 1 & 0 \\ 3 & 4 & 4 & 0 & 1 \end{pmatrix}$	100%	89%
27	$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 2 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	94%	100%
49	$\begin{pmatrix} 4 & 2 & 4 & 1 & 0 \\ 4 & 6 & 6 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 4 & 2 & 1 & 0 \\ 2 & 4 & 6 & 0 & 1 \end{pmatrix}$	91%	91%
81	Tous les plans ont même déterminant		
125	$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 \\ 3 & 4 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 1 & 0 \\ 3 & 4 & 0 & 0 & 1 \end{pmatrix}$	62%	93%

Tab. 7 : Tableau d'efficacité des plans à 5 facteurs de déterminant maximal.

On constate que même si le critère d'entropie ne donne pas les meilleurs plans, il est tout de même intéressant puisque les plans sélectionnés sont souvent très efficaces, et surtout, ils ne dépendent pas de la régression trigonométrique choisie. On peut donc espérer qu'ils seront aussi efficaces pour n'importe quel modèle en espérance.

Bien entendu ces conclusions sont issues d'un exemple très simple, mais on peut espérer les généraliser. En effet, les tableaux de force maximale sélectionnés par un critère d'entropie ne dépendent en aucune façon de la régression trigonométrique ou de la fonction choisie. Il est donc fort probable qu'ils sont tout aussi efficaces dans n'importe quelle situation et surtout quel que soit le modèle.

## Bibliographie

- Bailey R.A (1985). Factorial designs and abelian groups. *Lin. Algebra Appl.* **70**,349-363.
- Bailey R.A., Gilchrist F.H.L, Patterson H.D. (1977). Identification of effects and confounding patterns in factorial design. *Biometrika* **64**, 347-354.
- Bose R.C. (1947). Mathematical theory of symmetrical factorial design. *Sankhya* **8**, 107-166.
- Christensen R. (1987). *Plane answers to complex questions : The theory of linear models.* Springer-Verlag.
- Christensen R. (1990). *Linear models for multivariate, time series, and spatial data.* Springer-Verlag.
- Collombier D. (1996). *Plans d'expérience factoriels.* Springer-Verlag, Berlin.
- Fang K.T., Mukerjee R. (2000). Miscellanea. A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* **87**, 193-198.
- Franklin M.F. (1985). Selecting defining contrasts and confounded effects in  $p^{n-m}$  factorial experiments. *Technometrics* **27**, 165-172.
- Franklin M.F., Bailey R.A (1977). Selection of defining contrasts and confounded effects in two-level experiments. *Appl. Statist.* **26**, 321-326.
- Hickernell F.J. (1998). A generalized discrepancy and quadrature error bound. *Math. of Comp.* **67**, 299-322.
- Mardia K.V., Marshall R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Patterson H.D. (1976). Generation of factorial designs. *J. Of Roy Soc. B* **38**, 175-179.
- Patterson H.D., Bailey R.A. (1978). Design keys for factorial experiments. *Appl. Statist.* **27**, 335-343.
- Raktoe B.L. (1969). Combining elements from distinct finite fields in mixed factorials. *Ann. Math. Statist.* **40**,498-504.
- Sacks J., Schiller S.B., Welch W.J. (1989). Designs for computer experiments. *Technometrics* **31**, 41-47.
- Sloan I.H., Joe S. (1994). *Lattice methods for multiple integration.* Oxford science Publications.
- Srivastava J.N., Chopra D.V. (1973). Balanced arrays and orthogonal arrays. In *A survey of combinatorial theory* (Srivastava ed.) North Holland, 411-428.

# Troisième partie

*Construction de plans*



# Introduction

La troisième partie est consacrée aux aspects combinatoires de ce travail. Elle est consacrée à la construction des plans présentés dans le chapitre 5.

Dans un premier temps, nous allons nous intéresser aux tableaux orthogonaux linéaires. Nous avons vu avec les tests numériques que ceux-ci semblent former une catégorie de plans de bonne qualité pour les modèles avec corrélations que nous utilisons. Leur comportement est essentiellement dû à leur bonne répartition marginale qui est déterminée par leur force. Les méthodes de construction des tableaux orthogonaux de force  $t$  sont bien connues. Cependant ces techniques ne fournissent pas nécessairement des tableaux linéaires, or nous avons vu que pour l'ajustement d'une régression trigonométrique, les plans présentant une structure de sous-groupe possèdent des propriétés d'optimalité. Nous allons donc généraliser ces méthodes de façon à construire des tableaux orthogonaux de force  $t$  qui soient linéaires.

Les tableaux orthogonaux linéaires sont des sous-espaces de  $GF(p)^d$ , où  $GF(p)$  est le corps de Galois d'ordre  $p$  premier. Ce qui correspond à un découpage du cube unité symétrique et en un nombre premier de segments. Il peut être souhaitable de découper le domaine d'échantillonnage différemment. C'est pourquoi le deuxième chapitre de cette dernière partie est consacré à la construction de fractions régulières de  $G = \prod_{i=1}^d (Z/p^m)$ . La difficulté réside ici dans le fait que  $G$  n'est plus un espace vectoriel donc le calcul matriciel n'est plus envisageable. Le passage de la fraction régulière à son dual demande une technique particulière que nous allons développer. En effet, la connaissance du dual est indispensable pour l'ajustement d'une régression trigonométrique.

Dans un dernier chapitre, nous nous intéressons à une autre catégorie de plans qui ne sont, ni des sous-espaces vectoriels, ni des classes latérales de sous-groupes. On les appelle les fractions semi-régulières. Elles sont définies comme une superposition de fractions régulières. Nous n'allons pas aborder ici la question de la construction de ces plans, mais étudier, d'une part les conditions d'estimabilité des paramètres d'une régression trigonométrique. En effet, nous allons voir que contrairement aux fractions régulières, l'orthogonalité n'est pas une condition nécessaire à l'estimabilité. D'autre part, nous allons étudier la répartition marginale de ces plans avec la notion de résolution que nous avons définie précédemment.

## 6.

# Construction de tableaux orthogonaux linéaires

Les tableaux orthogonaux linéaires forment une classe de plans très intéressante pour les expériences simulées. En effet, nous avons vu qu'ils sont définis comme des fractions régulières à éléments dans  $GF(p)$ , où  $p$  est premier, présentant deux caractéristiques principales.

La première vient de leur structure de sous-espaces vectoriels qui en fait des plans particulièrement bien adaptés à l'ajustement d'une régression trigonométrique pour deux raisons. Tout d'abord, les tableaux orthogonaux linéaires sont entièrement définis par leur dual. Or nous avons vu que c'est cet ensemble qui permet de déterminer les effets confondus. De plus, l'estimabilité des paramètres pour un tableau orthogonal linéaire donné entraîne l'optimalité universelle dans le cas d'un modèle linéaire sans corrélation.

La deuxième caractéristique de ces plans est leur répartition marginale uniforme qui est déterminée par la force du tableau. D'après les tests numériques du chapitre 5, il semble que cette particularité en fait des plans de bonne qualité pour les modèles avec corrélations.

Ce chapitre est donc consacré à la construction des tableaux orthogonaux linéaires. De part leur définition, ces plans peuvent être considérés soit comme des fractions régulières à éléments dans  $GF(p)$ , soit comme des tableaux orthogonaux. La bibliographie concernant la construction ce type de fractions régulières est très abondante, notamment dans le cas de facteurs à deux niveaux. Nous pouvons par exemple citer les travaux de Franklin et Bailey (1977), Franklin (1985) et Collombier (1996). Nous avons préféré ici utiliser les méthodes de construction des tableaux orthogonaux. Ces techniques sont bien connues pour les tableaux à deux symboles. En revanche lorsque  $p$  est supérieur à deux, il y a déjà moins de bibliographie et celle-ci concerne souvent les tableaux de force 2. On peut par exemple citer Raghavarao (1971), Wang et Wu (1991) et Hedayat *et al.* (1992). Très peu d'articles sont consacrés aux tableaux orthogonaux ayant un nombre de symbole et une force quelconques. De plus ceux-ci ne sont pas nécessairement linéaires, ce qui est indispensable dans notre cas. On retiendra cependant les travaux de Adhikary et Das (1992) et Hedayat *et al.* (1996).

La première partie de ce chapitre est un bref rappel d'une technique de construction des tableaux orthogonaux de force 2 couramment utilisée, que l'on appelle méthode des différences. Nous allons de plus montrer dans quels cas les tableaux ainsi construits sont linéaires. Plus précisément, la deuxième partie généralise cette technique à la construction de tableaux orthogonaux linéaires de force  $t$  à partir d'un tableau orthogonal de force  $t-1$ . La partie 3 est une extension de ce résultat aux tableaux orthogonaux linéaires de force supérieure à  $t$ .

## 6.1. Construction d'un tableau orthogonal de force 2 par la méthode des différences

La construction des tableaux orthogonaux de force 2 a donné lieu à une abondante bibliographie. Nous nous contentons de présenter ici la méthode la plus couramment utilisée, celle que Bose et Bush (1952) ont appelée la méthode des différences. Pour plus d'informations sur les autres méthodes de construction nous vous proposons de vous reporter à l'ouvrage de Collombier (1996).

Comme son nom l'indique, cette méthode fait appel à des matrices à éléments dans des groupes abéliens finis, dites aux différences.

### Définition 6.1

*Soient  $G$  un groupe abélien fini d'ordre  $q$  et  $D$  une matrice d'ordre  $\lambda q \times d$  à éléments dans  $G$ . On dit que  $D$  est une matrice aux différences de force  $t$ ,  $1 \leq t \leq d$  sur  $G$  si dans tout bloc constitué de  $t$  colonnes de  $D$ , les classes latérales du sous-groupe  $H = \{(g, \dots, g), \forall g \in G\} \subseteq G^t$  sont représentées  $\lambda$  fois chacune. On note  $D_\lambda(G^d, t)$  toute matrice aux différences de force  $t$  sur  $G$  à  $d$  colonnes et d'index  $\lambda$ .*

**Remarque :** Avec cette notation, l'ensemble dans lequel nous construisons nos plans est  $G^d$ . Il ne faut pas confondre avec les notations utilisées précédemment, où  $G$  désignait  $\prod_{i=1}^d (\mathbb{Z} / q_i)$ .

Rappelons que la différence entre tableau orthogonal et tableau orthogonal linéaire, vient de la structure de sous-espace vectoriel de ce dernier, ce qui impose de se placer dans le cas où  $G$  est le groupe additif d'un corps de Galois,  $GF(p)$  avec ici  $p$  premier. Cependant, tous les résultats présentés dans ce chapitre ne nécessitent pas la linéarité, notamment ceux concernant les tableaux orthogonaux et les matrices aux différences. Dans ce cas  $G$  est un groupe abélien fini quelconque.

**Remarque :** Pour la construction de tableaux orthogonaux de force 2, seules les matrices aux différences de force 2 sont nécessaires. La définition devient alors : la matrice  $D = (d_{ij})$  est aux différences de force 2 sur  $G$  lorsque pour tout couple  $(h, j)$  de colonnes de  $D$ , on obtient  $\lambda$  fois chaque élément de  $G$  quand on forme les différences

$$d_{ij} - d_{ih} \quad i=1, \dots, \lambda q.$$

C'est la définition proposée par Bose et Bush (1952), elle est généralisée ici au cas  $t$  quelconque.

Notons  $\oplus$  la somme de Kronecker de matrices, c'est-à-dire l'opération qui à tout couple de matrices  $A = (a_{ij})$  d'ordre  $n \times d$  et  $B$  d'ordre  $n' \times d'$  à éléments dans  $G$ , associe la matrice d'ordre  $(nn') \times (dd')$  à éléments dans  $G$ , telle que

$$A \oplus B = (a_{ij} + B \mid i=1, \dots, n \text{ et } j=1, \dots, d)$$

où  $a_{ij} + B$  veut dire que l'on ajoute  $a_{ij}$  à chaque élément de  $B$ .

Soient par exemple  $A = {}^t(a_1, a_2)$  et  $B$  d'ordre quelconque alors

$$A \oplus B = \begin{pmatrix} a_1 + B \\ a_2 + B \end{pmatrix}.$$

### Proposition 6.2

*La somme de Kronecker de deux matrices aux différences sur un même groupe est une matrice aux différences sur ce groupe.*

La méthode aux différences est fondée sur la propriété suivante, dont on trouvera une démonstration dans Collombier (1996). Il s'agit en fait d'une généralisation d'un résultat de Bose et Bush (1952).

### Proposition 6.3

*Soient deux groupes abéliens,  $G = G_1^{d_1} \times \dots \times G_k^{d_k}$  et  $G'$ . Soient  $B$  et  $B'$  deux tableaux orthogonaux de force 2 sur  $G$  et  $G'$  respectivement, tels que*

1)  $B = (B_1 \mid \dots \mid B_k)$  d'ordre  $N \times d$  où  $d = \sum d_i$

2)  $B'$  d'ordre  $N' \times c$

*et  $D = (D_1 \mid \dots \mid D_k)$ , d'ordre  $N' \times d'$ , tel que  $D_j$  soit une matrice aux différences de force 2 sur  $G_j$ ,  $j=1, \dots, k$ .*

*Alors le tableau*

$$T = (B \oplus D \mid 0_N \oplus B')$$

*où  $B \oplus D = (B_1 \oplus D_1 \mid \dots \mid B_k \oplus D_k)$  et  $0_N$  est le vecteur colonne formé de  $N$  répliques de l'élément neutre d'un groupe donné, est orthogonal de force 2 sur  $G \times G'$ .*

### Corollaire 6.4

*La proposition reste vraie quand on remplace  $B$  ou/et  $B'$  par un vecteur colonne à éléments dans  $G$  ou/et  $G'$ , à condition que ces éléments y figurent un même nombre de fois.*

La méthode des différences permet de construire des tableaux orthogonaux uniquement de force 2, de plus, ces tableaux ne sont pas nécessairement linéaires. Seuls quelques cas très particuliers garantissent la linéarité. Par exemple, si on utilise le corollaire avec  $D$  la table de multiplication sur le corps de Galois  $GF(p)$  et  $B$  et  $B'$  vecteurs colonnes choisis parmi les colonnes non nulles de  $D$ , alors on obtient un tableau orthogonal de force 2 linéaire.

**Exemple 6.5 :** Si  $p=3$  alors

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix}$$

On choisit  $B=B'={}^t(0,1,2)$  et on obtient le tableau orthogonal linéaire de force 2 suivant

$$T = \left( \begin{array}{c} \text{B} \oplus \text{D} \\ \hline \text{0}_3 \oplus \text{B}' \end{array} \right) = \left( \begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 0 & 1 & 2 & 1 & 2 & 0 & 2 & 0 & 1 \\ 0 & 2 & 1 & 1 & 0 & 2 & 2 & 1 & 0 \\ \hline 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \end{array} \right)$$

Nous allons maintenant voir comment généraliser cette technique afin de construire des tableaux orthogonaux de force  $t$  supérieure ou égale à 2 et linéaires. Citons pour références Hedayat *et al.* (1996) et Collombier (cours de DEA, Université de Pau, 1997/98).

## 6.2. Construction d'un tableau orthogonal linéaire de force $t$

Le plan de cette partie est le suivant. Dans un premier temps, nous allons présenter quelques propriétés des matrices aux différences ainsi que la notion de matrices associables. Cette notion nous permettra de généraliser la méthode des différences à la construction d'un tableau orthogonal de force  $t$ . Nous établirons ensuite le lien entre matrice aux différences et tableau orthogonal linéaire. En dernier lieu nous déterminerons quelles conditions imposer aux matrices associables afin que le tableau orthogonal de force  $t$  soit linéaire.

### 6.2.1. Propriétés des matrices aux différences

#### Propriétés 6.6

*Si  $D=(d_{ij})$  est une matrice  $N \times d$  aux différences de force  $t$  sur le sous-groupe abélien  $G$  d'ordre  $q$ , alors on a les propriétés suivantes.*

- 1)  $\lambda=N/q^{t-1}$  entier non nul.
- 2) Toute matrice déduite de  $D$  par permutation des lignes et/ou des colonnes est une matrice aux différences de force  $t$  sur  $G$ .
- 3)  $D$  est une matrice aux différences de force  $u$ ,  $\forall 1 \leq u \leq t$ .
- 4) Pour  $h \in \{1, \dots, d\}$  fixé, la matrice  $D^*=(d_{ij}^*)$ , où  $d_{ij}^*=d_{ij}-d_{ih}$ , est aux différences de force  $t$  sur  $G$ . De plus le bloc formé des colonnes de  $D$  hormis la colonne  $h$  est un tableau orthogonal de force  $t-1$
- 5) En ajoutant aux éléments de  $D$  un élément fixé  $g$  de  $G$ , on obtient une matrice, notée  $g+D$ , aux différences de force  $t$  sur  $G$ .
- 6) Si  $D_1, \dots, D_p$  sont des matrices aux différences de force  $t$  sur  $G$ , alors la matrice  $D$  juxtaposition verticale des matrices  $D_1, \dots, D_p$  est une matrice aux différences de force  $t$  sur  $G$ .

**Démonstration :** Il s'agit ici de propriétés classiques des matrices aux différences. Nous limitons donc notre démonstration à la propriété 4 2<sup>ème</sup> partie. D'après la 1<sup>ère</sup> partie de cette propriété 4, on peut supposer sans perte de généralité que la matrice aux différences  $D$  peut s'écrire sous la forme

$$D = \left( \begin{array}{c|c} 0 & \\ \dots & T \\ 0 & \end{array} \right)$$

où  $T$  est à la fois une matrice aux différences de force  $t$  et un tableau orthogonal de force  $t-1$ . En effet,  $T$  est une matrice aux différences de force  $t$  d'après la 1<sup>ère</sup> partie de la propriété 4, et

T est un tableau orthogonal de force  $t-1$  car si on choisit  $t$  colonnes dans D, la première et  $t-1$  dans T, alors chaque classe latérale de  $H=\{(g,\dots,g), g \in G\} \subseteq G^t$ , i.e.  $x+H$ , où  $x \in G^t$ , est représentée  $\lambda$  fois, et les représentants sont de la forme  $(0, x_1, \dots, x_{t-1})$  pour tout  $(x_1, \dots, x_{t-1}) \in G^{t-1}$ , donc dans tout bloc de  $t-1$  colonnes de T, tout élément de  $G^{t-1}$  est représenté  $\lambda$  fois.

**Proposition 6.7**

*Si D est une matrice aux différences de force t,*

$$T = \left( \begin{array}{c} \mathbf{g} + \mathbf{D} \\ \forall \mathbf{g} \in G \end{array} \right)$$

*est un tableau orthogonal de force t et réciproquement.*

**Démonstration :** Dans tout bloc de  $t$  colonnes de T chaque classe latérale du sous-groupe H, i.e.  $x+H$  avec  $x \in G^t$ , est représentée  $\lambda$  fois. Or d’après le 5) de la propriété 6.6, si D est une matrice aux différences de force  $t$ ,  $g+D$  avec  $g \in G$ , l’est aussi, donc quand on ajoute  $g$  aux  $\lambda$  représentants de la classe latérale  $x+H$ , on obtient  $\lambda$  autres représentants de cette classe. En faisant cela pour tout  $g \in G$ , on obtient tous les représentants de la classe et ceci  $\lambda$  fois chacun. Autrement dit, les éléments de  $G^t$  sont tous représentés  $\lambda$  fois chacun. La réciproque est triviale, il suffit de choisir  $\lambda$  représentants de chaque classe de  $G^t$  pour construire D.  $\square$

Dans le cas où la force est égale à 2, nous avons vu que l’on pouvait ajouter par la méthode des différences une ou plusieurs colonnes à D. Malheureusement, lorsque la force est supérieure à 2 cela n’est pas toujours possible. Il nous faut donc introduire la notion de matrices associables.

**Définition 6.8**

*Soient G un groupe abélien fini d’ordre q et les entiers positifs  $\lambda, d_1, d_2, t \leq d_1 + d_2, r = \lambda q^{t-1}$ . On note de plus,*

- 1)  $D=(d_{ij})$  une matrice aux différences de force  $t$  sur  $G^{d_1}$ , à  $r$  lignes,
- 2)  $B=(b_{ij})$  un tableau orthogonal de force  $\min(t, d_2)$  sur  $G^{d_2}$ , de taille  $r$ .

*Pour tout entier non nul  $v$ , tel que  $v \leq \min(t-1, d_2)$ , considérons  $J_v \subseteq \{1, \dots, d_2\}$  tel que  $|J_v| = v, z \in G^v$  et le bloc suivant de D*

$$D_{I_{B[J_v], z}} = (d_{ik} \mid i \in I_{B[J_v], z} \text{ et } k = 1, \dots, d_1)$$

*où  $I_{B[J_v], z} = \{i \in \{1, \dots, r\}; (b_{ij} \mid j \in J_v) = z\}$ .*

*On dit que B et D sont associables (pour construire un tableau orthogonal de force t sur  $G^{d_1+d_2}$ ) s’il existe une matrice de permutation S telle que, quel que soit v, un entier non nul tel que  $v \leq \min(t-1, d_2)$ ,  $J_v \subseteq \{1, \dots, d_2\} : |J_v| = v$ , et  $z \in G^v$ ,  $D_{I_{SB[J_v], z}}$  est une matrice aux différences de force  $t-v$  sur  $G^{d_1}$ .*



Il apparaît que Q est la juxtaposition verticale des  $q^{v+1}$  blocs de  $Q_{g,z}$  pour  $z \in G^v$  puis  $g \in G$ . B et D sont associables donc  $D_{I_{B[J_v],z}}^H$  est une matrice aux différences de force  $t-v$ . D'après la proposition 6.7, la matrice

$$T_z = \left( \frac{g + D_{I_{B[J_v],z}}}{\forall g \in G} \right)$$

est un tableau orthogonal de force  $t-v$  et d'index  $\lambda$  sur  $G^{t-v}$ . Or Q est la juxtaposition verticale des  $q^v$  blocs  $T_z$  pour  $z \in G^v$ . Par conséquent tout élément de  $G^t$  apparaît  $\lambda$  fois dans Q.  $\square$

**Remarque :** Il est souvent très difficile de trouver D et B associables. Nous pouvons tout de même remarquer que si B est un vecteur colonne dans lequel tous les éléments de G sont répétés  $\lambda q^{t-2}$  fois et si D est une matrice aux différences de force  $t$  d'ordre  $\lambda q^{t-1} \times d$  alors D et B sont associables pour construire un tableau orthogonal de force  $t$  à  $d+1$  colonnes. En effet, dans ce cas B est bien un tableau orthogonal de force 1 sur G de taille  $\lambda q^{t-1}$ ,  $v$  est obligatoirement égal à 1, donc  $\forall z \in G$ ,

$$\text{card}(I_{B[J_v],z} = \{i \in \{1, \dots, \lambda q^{t-1}\}; b_i = z\}) = \lambda q^{t-2},$$

autrement dit  $D_{I_{B[J_v],z}}$  correspond à  $\lambda q^{t-2}$  lignes de D. D est une matrice aux différences de force  $t$  donc elle est aussi de force  $t-1$ , c'est-à-dire que dans tout bloc de  $t-1$  colonnes de D les  $q^{t-2}$  classes latérales de  $H_{t-1} = \{(g, \dots, g), g \in G\} \subseteq G^{t-1}$  sont représentées  $\lambda q$  fois chacune. Il est alors possible de choisir un bloc de  $\lambda q^{t-2}$  lignes de D,  $D_{I_{B[J_v],z}}$ , de façon à ce que dans tout bloc de  $t-1$  colonnes de  $D_{I_{B[J_v],z}}$  les  $q^{t-2}$  classes latérales soient représentées  $\lambda$  fois chacune, ce qui veut dire que  $D_{I_{B[J_v],z}}$  est une matrice aux différences de force  $t-1$ .

**Exemple 6.10 :** Soient G le groupe cyclique d'ordre 3, D la matrice aux différences de force 3 d'index 1 sur  $G^3$  suivante

$$D = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 2 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

et B le vecteur colonne,  $B = {}^t(0,0,0,1,1,1,2,2,2)$ .

Alors le tableau

$$T = \left( \begin{array}{c|c|c} D & 1+D & 2+D \\ \hline B & B & B \end{array} \right)$$

est orthogonal de force 3 à 4 facteurs.

## 6.2.2. Lien entre tableaux orthogonaux linéaires de force t-1 et matrices aux différences de force t

La construction d'un tableau orthogonal de force t envisagée ici suppose qu'il existe une matrice aux différences de force t. Nous allons donc nous intéresser à la construction de ces dernières et pour cela nous allons exploiter le 4) de la propriété 6.6. C'est-à-dire que nous allons établir les conditions à imposer à la matrice T pour que celle-ci soit à la fois un tableau orthogonal linéaire de force t-1 et une matrice aux différences de force t.

Il est nécessaire ici de se placer dans le cas d'un corps de Galois que l'on note  $K=GF(p)$ , pour obtenir la linéarité des tableaux orthogonaux, *i.e* leur imposer une structure de sous-espace vectoriel.

### Théorème 6.11

*Soient P une matrice d'ordre  $r \times d$ , de rang r, à éléments dans  $GF(p)$  et t un entier compris entre 1 et d.*

*Considérons le tableau T d'ordre  $p^{d-r} \times d$  dont les lignes sont les solutions de l'équation  $Px=0$ .*

*Alors T est à la fois un tableau orthogonal linéaire de force t-1 et une matrice aux différences de force t sur  $GF(p)$  (d'index  $\lambda = p^{d-r}/p^{t-1}$ ) si et seulement si*

- 1)  $wt(P) \geq t$
- 2) *pour tout  $y=(y_1, \dots, y_d) \in Im {}^tP$  tel que  $wt(y)=t$ , on a  $y1_d \neq 0$ , où  $1_d$  est le vecteur d'ordre d formé de 1.*

**Remarque :** Pour déterminer le poids de la matrice P, il suffit d'utiliser les résultats du chapitre précédent (5.15, 5.16, 5.16).

**Exemple 6.12 :** Reprenons l'exemple des tableaux orthogonaux linéaires de force 1 définis dans le chapitre 5 (§ 5.3). La matrice génératrice du dual P est d'ordre  $(d-1) \times d$ , à éléments dans  $GF(p)$ , de la forme

$$P = \left( \begin{array}{c|c} -g_2 & \\ \dots & \\ -g_d & \end{array} \middle| I_{d-1} \right)$$

Nous savons déjà que ces tableaux sont de force 1 donc la condition 1 du théorème est vérifiée pour t=2. Vérifions maintenant la condition 2.

Soit  $u=(u_2, \dots, u_d)$  un vecteur non nul d'ordre  $1 \times (d-1)$ . Toute combinaison linéaire des lignes de P s'écrit

$$uP = \left( - \sum_{k=2}^d u_k g_k, u_2, \dots, u_d \right).$$

Nous savons par définition du vecteur g que pour tout  $k \in \{2, \dots, d\}$

$$u_k g_k \neq 0 \pmod{p}, \quad \forall u_k \in GF(p) \setminus \{0\}.$$

Supposons, sans perte de généralité, que  $u=(u_2,0,\dots,0)$ ,  $u_2 \neq 0$ , alors  $\text{wt}(uP)=2$  et dans ce cas  $(uP)\mathbf{1}_d=u_2(g_2+1) \neq 0 \pmod{p}$ , sauf si  $g_2=p-1$ . Il faut donc vérifier que toutes les coordonnées de  $g$  sont différentes de  $p-1$ .

Une autre possibilité pour avoir une combinaison linéaire des lignes de  $P$  de poids égal à 2 est que  $u$  ait 2 coordonnées non nulles, par exemple  $u_2$  et  $u_3$ , mais telles que  $u_2g_2+u_3g_3=0$ . Dans ce cas  $(uP)\mathbf{1}_d=u_2+u_3=0 \pmod{p}$  si et seulement si  $u_2=-u_3 \pmod{p}$ , mais alors  $u_2g_2+u_3g_3=u_2(g_2-g_3)=0$  d'où  $g_2=g_3$  ce qui est impossible par définition de  $g$ .

La deuxième condition du théorème est donc vérifiée (si toutes les coordonnées de  $g$  sont différentes de  $p-1$ ), le tableau orthogonal linéaire  $T$  de force 1 défini par  $P$  est aussi une matrice aux différences de force 2 sur  $\text{GF}(p)$ . D'après la propriété 6.6, la matrice  $(0|T)$  est aussi une matrice aux différences de force 2 à  $d+1$  colonnes. Il est possible ensuite, grâce à la remarque consécutive au théorème 6.9, de construire le tableau orthogonal de force 2 à  $(d+2)$  paramètres

$$T' = \left( \begin{array}{c|c} g+T & B \\ \hline \forall g \in \text{GF}(p) \end{array} \right) \text{ où } B = {}^t(0, \dots, p-1).$$

Par exemple, on choisit  $d=3$ ,  $g=(1,2,4)$  et  $p=5$  alors

$$T' = {}^t \left( \begin{array}{c|c|c|c|c} T & 1+T & 2+T & 3+T & 4+T \\ \hline B & B & B & B & B \end{array} \right) \text{ avec } T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 4 \\ 0 & 2 & 4 & 3 \\ 0 & 3 & 1 & 2 \\ 0 & 4 & 3 & 1 \end{pmatrix} \text{ et } B = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

est de force 2 à 5 paramètres de taille 25.

Ceci n'est pas la meilleure façon de construire un tableau orthogonal de force 2. Par exemple si on utilise la méthode des différences avec la table de multiplication de  $\text{GF}(5)$  comme dans l'exemple 6.5, on peut construire un tableau orthogonal de force 2 à 6 paramètres chacun à 5 niveaux de taille 25.

**Exemple 6.13 :** Soit  $P$  la matrice d'ordre  $1 \times 6$  dans le corps de Galois d'ordre 2,

$$P=[1,0,1,1,1,1].$$

Les conditions du théorème sont trivialement vérifiées pour  $t=5$ . Le tableau  $T$  d'ordre  $2^6 \times 6$  dont les lignes sont les solutions de l'équation  $Px=0$  est donc à la fois un tableau orthogonal linéaire de force 4 et une matrice aux différences de force 5 (d'index 4). Il est de la forme

$$P = \left( \begin{array}{c|c} P_0 & B_0 \\ \hline P_1 & B_1 \end{array} \right)$$

où  $P_0$  (resp.  $P_1$ ) est la matrice d'ordre  $32 \times 4$  formée des vecteurs de  $EA(2^4)$  dont la somme des première, troisième et quatrième coordonnées est nulle (resp. égale à 1), et  $B_0$  (resp.  $B_1$ ) est la matrice d'ordre  $32 \times 2$  dont les lignes sont formées des couples  $(0,0)$  et  $(1,1)$  (resp.  $(1,0)$  et  $(0,1)$ ) répétés 16 fois chacun.

D'après la propriété 6.6, si on ajoute une colonne de 0 à T, on garde une matrice aux différences de force 5 à 7 paramètres. Et en appliquant le théorème 6.9 avec B vecteur colonne d'ordre 64 dans lequel 0 et 1 apparaissent 32 fois chacun, on obtient un tableau orthogonal de force 5 à 8 paramètres de taille 64\*2 de la forme

$$T' = \left( \begin{array}{c|c|c|c} 0 & \frac{P_0}{P_1} & \frac{B_0}{B_1} & B \\ \hline & \frac{P_0}{P_1} & \frac{B_0}{B_1} & B \\ \hline 1 & \frac{P_0}{P_1} & \frac{B_0}{B_1} & B \end{array} \right)$$

puisque  $P_{0+1}=P_1$ ,  $B_{0+1}=B_0$  et  $B_{1+1}=B_1$ .

La démonstration du théorème 6.11 nécessite que l'on établisse le lemme suivant dont on peut trouver la démonstration dans Hedayat *et al.* (1996).

#### Lemme 6.14

*Soit T un tableau orthogonal linéaire d'ordre  $p^{d-r} \times d$ .*

*T est un tableau orthogonal linéaire de force  $t \leq d$  si et seulement si toute famille de t colonnes de T est libre.*

#### Démonstration du théorème 6.11:

1) Montrons d'abord que T est tableau orthogonal de force t-1 si et seulement si  $wt(P) \geq t$ .

*Condition nécessaire :* Soient v un vecteur de  $K^r$  et w le poids de (vP). Par définition du tableau, nous avons  $P^t T = 0$ , donc  $(vP)^t T = 0$ , i.e qu'il existe une combinaison linéaire de w colonnes de T qui est nulle. Mais par hypothèse, T est un tableau orthogonal de force t-1, donc d'après le lemme 6.14,  $w > t-1$ . Donc  $\forall v \in K^r$ ,  $wt(vP) \geq t$ , i.e  $wt(P) \geq t$ .

*Condition suffisante :* Soit  $y \in \ker T \setminus \{0\}$ , i.e  $T^t y$  forme une combinaison linéaire nulle de colonnes de T. Par construction  $\ker T = \text{Im}^t P$ , d'où par hypothèse  $wt(y) \geq t$ . Donc toute famille de t-1 colonnes de T est libre et d'après le lemme cela implique que T est un tableau orthogonal de force t-1.

2) D'après la propriété 6.6,  $T = (t_{ij})$  est une matrice aux différences de force t, si et seulement si  $D = [0 \mid t_{i1} - t_{ij} \mid j=2, \dots, d]$  est aussi une matrice aux différences de force t, c'est-à-dire si et seulement si  $T' = [t_{i1} - t_{ij} \mid j=2, \dots, d]$  est un tableau orthogonal de force t-1. Or d'après le lemme 6.14 ceci est vrai si et seulement si toute famille de t-1 colonnes de T' est libre.

Il nous reste à montrer qu'il en est ainsi si et seulement si pour tout  $y \in \text{Im}^t P$  tel que  $wt(y) = t$ ,  ${}^t y_{1d} \neq 0$ .

*Condition nécessaire :* Raisonnons par l'absurde et supposons qu'il existe  $y \in \text{Im}^t P$  tel que  $wt(y) = t$  et  ${}^t y_{1d} = 0$ . Sans perte de généralité, on peut supposer que  $y = (y_1, \dots, y_t, 0, \dots, 0)$  avec  $y_i \neq 0$   $i=1, \dots, t$ . Alors

$$({}^t y \mathbf{1}_d = 0 \Leftrightarrow y_1 = -\sum_{j=2}^t y_j) \Rightarrow y = \left( -\sum_{j=2}^t y_j, y_2, \dots, y_t, 0, \dots, 0 \right) = \sum_{j=2}^t y_j v_j$$

où  $v_j = (-1|\delta_i^j, i=2, \dots, t)$ .

Etant donné que  $y \in \text{Im}^t P$ , on a

$$T^t y = 0 \Leftrightarrow \sum_{j=2}^t y_j T^t v_j = 0 \Leftrightarrow \sum_{j=2}^t y_j T'_j = 0_{p^{d-r}}$$

où  $T'_j$  est la  $j$ ème colonne de  $T'$ . Autrement dit il existe  $t-1$  colonnes de  $T'$  dépendantes puisque  $y_i \neq 0$   $i=2, \dots, t$ , et d'après le lemme 6.14 ceci est impossible car  $T'$  est un tableau orthogonal de force  $t-1$ .

*Condition suffisante* : Raisonnons aussi par l'absurde, c'est-à-dire supposons sans perte de généralité que les  $t-1$  premières colonnes de  $T'$  sont dépendantes, *i.e* il existe des entiers de  $\text{GF}(p)$  non tous nuls  $y_2, \dots, y_t$  tels que

$$\begin{aligned} \sum_{j=2}^t y_j T'_j = 0_{p^{d-r}} &\Leftrightarrow \sum_{j=2}^t y_j (t_{ij} - t_{il}) = 0 \quad \forall i = 1, \dots, p^{d-r} \\ &\Leftrightarrow \left( -\sum_{j=2}^t y_j \right) t_{i1} + \sum_{j=2}^t y_j t_{ij} = 0, \quad \forall i = 0, \dots, p^{d-r} \Leftrightarrow T^t y = 0 \end{aligned}$$

avec  $y = \left( -\sum_{j=2}^t y_j, y_2, \dots, y_t, 0, \dots, 0 \right)$ .

D'où  $y \in \text{Im}^t P$ ,  $\text{wt}(y) \leq t$  et  ${}^t y \mathbf{1}_d = 0$ , ce qui est impossible par hypothèse.  $\square$

Le théorème 6.9 généralise la méthode des différences et permet de passer de la construction des tableaux orthogonaux de force 2 à la construction des tableaux orthogonaux de force quelconque. Malheureusement, cette technique n'est pas suffisante pour construire des tableaux orthogonaux linéaires. Nous allons voir maintenant comment on peut choisir les matrices associables  $D$  et  $B$  de façon à obtenir systématiquement un tableau orthogonal linéaire.

### 6.2.3. Construction d'un tableau orthogonal linéaire de force $t$

#### Théorème 6.15

Soient  $P$  une matrice d'ordre  $r \times d$ , de rang  $r$ , à éléments dans  $K$  et  $t$  un entier compris entre 1 et  $d$ .

Considérons le tableau  $T$  d'ordre  $p^{d-r} \times d$  dont les lignes sont les solutions de l'équation  $Px=0$ .

Supposons les conditions du théorème 6.11 vérifiées, i.e  $wt(y) \geq t$  et  $\forall y \in Im^t P$  tel que  $wt(y)=t \Rightarrow {}^t y I_d \neq 0$ .

Notons alors  $D=(0 | T)$  la matrice aux différences de force  $t$  sur  $K^{d+1}$  obtenue à partir des solutions de  $Px=0$ . Indexons les colonnes de  $D$  par  $0, \dots, d$ .

Soient  $s$  vecteurs de  $K^d$ ,  $v_1, \dots, v_s$ , et  $\forall J \subseteq \{1, \dots, s\}$ , la matrice d'ordre  $(r+|J|) \times d$ ,

$$Q_J = \left( \begin{array}{c|c} P & \\ \hline v_j, j \in J & \end{array} \right).$$

Supposons vérifiées par les matrices  $Q_J$  les conditions suivantes :

1)  $\forall J \subseteq \{1, \dots, s\}$ ,  $1 \leq |J| \leq \min(t-1, s)$

(i)  $wt(Q_J) \geq t - |J|$

(ii)  $\forall y \in Im^t Q_J$  tel que  $wt(y) = t - |J| \Rightarrow {}^t y I_d \neq 0$

2)  $wt(Q_J) \geq 1$  si  $s \geq t$ .

Considérons le tableau d'ordre  $p^{d-r} \times s$ ,  $B=TV$  où  $V=({}^t v_1, \dots, {}^t v_s)$  ( $b_{ij}=(x_i | v_j)$ , où  $x_i$  est une solution de l'équation  $Px=0$ ), alors  $B$  est un tableau orthogonal de taille  $p^{d-r}$ , de force  $\min(t, s)$  sur  $K^s$ , associable à  $D$ .

**Remarque :** La condition 1 équivaut à dire que quelles que soient les lignes que l'on ajoute à  $P$  pour former  $Q$ , le tableau formé des lignes de  $\ker Q$  reste à la fois une matrice aux différences de force  $t$  et un tableau orthogonal de force  $t-1$ .

#### Corollaire 6.16

On reprend les notations du théorème précédent et on suppose que les conditions 1 et 2 sont vérifiées. Le tableau d'ordre  $p^{d-r+1} \times (d+s+1)$  construit à partir des matrices associables  $D$  et  $B$  est défini par sa matrice génératrice du dual

$$P' = \left( \begin{array}{c|c|c} -PI_d & P & 0 \\ \hline -VI_d & V & -I_s \end{array} \right)$$

et est orthogonal linéaire de force  $t$  sur  $K^{d+s+1}$ .

**Exemple 6.12 (suite) :** Reprenons l'exemple sur les tableaux orthogonaux de force 1, définis par la matrice génératrice du dual

$$P = \left( \begin{array}{c|c} -g_2 & \\ \dots & \\ -g_d & \end{array} \middle| I_{d-1} \right).$$

Nous avons vu que la matrice P vérifie les conditions du théorème 6.11. Soit  $v = {}^t(1, 0, \dots, 0)$  et  $s=1$ , nécessairement  $J=\{1\}$  et  $Q_J = {}^t(P|v)$ . Etant donné que  $t=2$ , les conditions 1 et 2 du théorème 6.15 sont identiques à  $wt(Q_J) \geq 1$  et elles sont trivialement vérifiées. Le vecteur colonne  $B = Tv$  est alors associé à  $D = (0|T)$  où T est la tableau orthogonal de force 1 engendré par le vecteur  $(1, g_2, \dots, g_d)$ .

Le vecteur B est égal à  $(0, g_d, \dots, (p-1)g_d)$ , i.e une permutation de  $0, \dots, p-1$ .

On peut ajouter ici que le tableau T' ainsi formé

$$T' = \left( \begin{array}{c|c} h + D & B \\ \hline \forall h \in K & \end{array} \right),$$

est linéaire et sa matrice génératrice du dual est de la forme

$$P' = \left( \begin{array}{c|c|c|c|c} g_2 - 1 & -g_2 & & & 0 \\ \dots & \dots & & I_{d-1} & \dots \\ g_d - 1 & -g_d & & & 0 \\ \hline 1 & 1 & 0 & \dots & 0 & -1 \end{array} \right).$$

**Exemple 6.13 (suite) :** On ne peut pas appliquer le théorème 6.15 et son corollaire à l'exemple 6.13. En effet  $t=5$ , il faut donc trouver un vecteur v de poids supérieur ou égal à 4 tel que  ${}^t(P|v)$  soit aussi de poids supérieur ou égal à 4. Or il est facile de vérifier que si  $wt(v) \geq 4$ , alors  $wt({}^t(P|v)) = 2$  ou 3.

**Exemple 6.17 :** Soient P la matrice  $5 \times 8$  à éléments dans  $GF(3)$

$$P = (I_5 | Q) \text{ où } Q = \begin{pmatrix} 1 & 2 & 0 & 2 & 2 \\ 1 & 1 & 1 & 0 & 1 \\ 2 & 1 & 2 & 2 & 0 \end{pmatrix},$$

$t=3$  et  $V = {}^t v_1 = {}^t(0, 0, 0, 0, 0, 1, 1, 1)$ .

Montrons tout d'abord que P vérifie bien les conditions du théorème 6.11.

Chaque ligne de Q est de poids supérieur ou égal à 2 et toute combinaison linéaire de 2 lignes de Q est de poids supérieur ou égal à 1. D'après le lemme 5.16,  $wt(P) \geq 3$ . On vérifie de plus que toute combinaison linéaire de lignes de P de poids égal à 3 n'a pas la somme de ses coordonnées nulle. Les deux conditions du théorème 6.11 sont bien vérifiées.

Vérifions maintenant les conditions du théorème 6.15. Comme  $s=1$ ,  $J=\{1\}$  et  $Q_J = {}^t(P|v_1)$ . D'après le lemme 5.16,  $wt(Q_J) \geq t-1=2$  puisque

$${}^t \left( \begin{array}{cccc|c} & & & & 0 \\ & & & & 1 \\ \hline & & & & \\ & & & & \\ & & & & \\ & & & & \end{array} \right)_{Q_J} = {}^t \left( \begin{array}{cccccc|c} & & & & & & \\ & & & & & & \\ \hline & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \end{array} \right)$$

donc toute combinaison linéaire y des lignes de  $Q_J$  telle que  $wt(y)=2$  vérifie  $y\mathbf{1}_8=2$ .

Soit  $H$  la matrice  $27 \times 3$  qui a pour lignes les  $3^3$  vecteurs de  $GF(3)^3$ . D'après la proposition 5.15 on a

$$T=H(-{}^tQ | I_3),$$

d'où

$$D=(0 | -{}^tQ | I_3)$$

et

$$B=H\mathbf{1}_3$$

et d'après le théorème 6.15,  $B$  et  $D$  sont associables.

En appliquant le corollaire, on obtient le tableau orthogonal linéaire d'ordre  $3^4 \times 10$  de force 3 suivant

$$T' = {}^t \left( \begin{array}{c|c|c} D & 1+D & 2+D \\ \hline B & B & B \end{array} \right)$$

Sa matrice génératrice du dual est de la forme

$$P' = \left( \begin{array}{c|c|c} -P\mathbf{1}_8 & P & 0 \\ \hline -V\mathbf{1}_8 & V & -I_8 \end{array} \right) = \left( \begin{array}{c|c|c|c} -P\mathbf{1}_8 & I_5 & Q & 0 \\ \hline 0 & 0 & {}^t\mathbf{1}_3 & -1 \end{array} \right).$$

D'après Seiden (1955) il n'existe pas de tableau orthogonal de taille  $3^d$  et de force 3 sur  $K^d$  avec  $d > 10$  et même plus généralement de tableau orthogonal à 3 symboles à plus de 10 paramètres. On a donc trouvé le plus grand au sens nombre de paramètres.

**Exemple 6.18 :** Soient  $P$  la matrice d'ordre  $4 \times 8$  à éléments dans  $GF(3)$

$$P=(I_4 | Q) \text{ où } Q = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 2 & 0 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 2 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0_4 & | & 1 & 1 & 1 & 1 \\ 0_4 & | & 1 & 2 & 1 & 2 \end{pmatrix} \text{ et } t=4.$$

La matrice  $P$  vérifie les conditions du théorème 6.11 et les matrices  $Q_J$ ,  $J=\{1\}, \{2\}$  ou  $\{1,2\}$  vérifient la condition du théorème 6.15.

D'après la proposition 5.15, si on note  $H$  la matrice d'ordre  $81 \times 4$  dont les lignes sont les vecteurs de  $GF(3)^4$ , alors

$$T=H(-{}^tQ | I_4),$$

d'où, d'après le théorème 6.15, les matrices

$$D=(0 | -{}^tQ | I_4)$$

et

$$B=TV=H \begin{pmatrix} 1 & 1 & | & 1 & 1 \\ 1 & 2 & | & 1 & 2 \end{pmatrix}$$

avec B tableau orthogonal de force 2 d'index 9, sont associables.

En appliquant ensuite le corollaire, on construit le tableau orthogonal linéaire de force 4 à 11 paramètres de taille  $3^5$  de matrice génératrice du dual,

$$P' = \left( \begin{array}{c|c|c} -P\mathbf{1}_8 & P & 0 \\ \hline -V\mathbf{1}_8 & V & -I_s \end{array} \right) = \left( \begin{array}{ccc|c} \mathbf{1}_4 & I_4 & Q & 0_{2 \times 2} \\ \hline 2 & 0_4 & 1111 & 22 \\ 0 & 0_4 & 1212 & 22 \end{array} \right).$$

Cette fois encore, nous avons trouvé le plus grand tableau orthogonal au sens du nombre de paramètres, cf. Rao (1947).

**Démonstration du théorème 6.15 :** La démonstration se fait en deux parties. Il faut montrer premièrement que B est bien le tableau souhaité et ensuite que B et D sont associables.

1) Pour montrer que B est un tableau orthogonal de force  $w = \min(t, s)$ , il faut montrer que dans tout bloc de  $w$  colonnes de B tout élément de  $K^w$  apparaît  $p^{d-r-w}$  fois. Or étant donné que  $B = ((x_i | v_j))_{ij}$  avec  $Px_i = 0$ , cela revient à montrer que pour tout  $J \subseteq \{1, \dots, s\}$  tel que  $|J| = w$ , le système

$$Q_J x = \begin{pmatrix} 0 \\ z \end{pmatrix}$$

admet  $p^{d-r-w}$  solutions pour tout  $z \in K^w$ .

Si  $w < t$ , alors  $t - |J| > 1$  et d'après la condition 1 du théorème  $wt(Q_J) \geq 1$ , de même si  $w = t$  la condition 2 du théorème entraîne que  $wt(Q_J) \geq 1$ . Ce qui veut dire qu'aucune combinaison linéaire des lignes de  $Q_J$  n'est nulle, autrement dit, que toutes les lignes de  $Q_J$  sont indépendantes.  $Q_J$  est d'ordre  $p^{r+w}$ , le système admet donc  $p^{d-r-w}$  solutions.

2) D est la matrice d'ordre  $p^{d-r} \times (d+1)$ ,  $D = (0 | T)$ , sa  $i$ ème ligne s'écrit donc  $(0 | x_i)$  où  $Px_i = 0$ .

$\forall v$  un entier non nul,  $v \leq \min(t-1, s)$ ,  $\forall J \subseteq \{1, \dots, s\}$ ,  $|J| = v$  et  $\forall z = (z_1, \dots, z_v) \in K^v$

$$I_{B[J], z} = \left\{ i \in \{1, \dots, p^{d-r}\}; \left( x_i | v_j \right) = z_j \quad \forall j \in J \right\}$$

et

$$D_{I_{B[J], z}} = \left( d_{ik} \mid i \in I_{B[J], z} \text{ et } k = 0, \dots, d \right)$$

Si on note  $D'_{J, z}$  le tableau des solutions de l'équation

$$Q_J x = \begin{pmatrix} 0 \\ z \end{pmatrix},$$

alors  $D_{I_{B[J], z}} = (0 | D'_{J, z})$ .

Comme  $1 \leq |J| \leq \min(t-1, s)$ , on a d'après la condition 1  $wt(Q_J) \geq t - |J|$  et  $\forall y \in \text{Im}^t Q_J$  tel que  $wt(y) = t - |J| \Rightarrow y \mathbf{1} \neq 0$ .

Les conditions du théorème 6.11 sont vérifiées donc  $D'_{J, z}$  est une matrice aux différences de force  $t - |J|$  et *a fortiori*  $D_{I_{B[J], z}}$  aussi. Donc B et D sont associables.  $\square$

**Démonstration du corollaire 6.16 :** En utilisant le théorème 6.9 on construit à partir de B et D associables le tableau

$$T = \left( \begin{array}{c|c|c} \mathbf{g} \mathbf{1}_{p^{d-r}} & \mathbf{g} + \mathbf{D} & \mathbf{B} \\ \hline \forall \mathbf{g} \in \text{GF}(p) & & \end{array} \right)$$

où  $\mathbf{D}=(d_{ij})$  avec  $(d_{ij} \mid j=1, \dots, d) = {}^t x_i$  et  $\mathbf{B}=(x_i \mid v_j)_{ij}$  où  $P x_i = 0$ .  
Alors la matrice

$$P' = \left( \begin{array}{c|c|c} -P \mathbf{1}_d & P & 0 \\ \hline -V \mathbf{1}_d & V & -I_s \end{array} \right)$$

vérifie  $P'^t T' = 0$  car  $P^t T = 0$  et  $V^t D = {}^t B$ .

Donc les lignes de  $T'$  forment un sous-espace vectoriel de  $\ker P'$ .

De plus  $P'$  est d'ordre  $(r+s) \times (1+d+s)$ , or le bloc de  $P'$   $\left( \begin{array}{c|c} P & 0 \\ \hline V & -I_s \end{array} \right)$  est de rang  $(r+s)$  puisque le rang de  $P$  est  $r$  donc  $\text{rg } P' = r+s$ .

En outre,  $T'$  comporte  $p^{d-r+1}$  lignes distinctes car les lignes de  $D=(0 \mid T)$  sont distinctes étant donné que  $T$  est linéaire. Donc les lignes de  $T'$  sont les vecteurs de  $\ker P'$ .  $\square$

A partir d'un tableau orthogonal linéaire  $T$  de force  $t-1$  (ou d'une matrice aux différences de force  $t$ ), nous avons construit un tableau orthogonal linéaire de force  $t$  et cela en augmentant le nombre de paramètres. Nous allons maintenant nous intéresser à la construction d'un tableau orthogonal linéaire ayant le même nombre de paramètres que le tableau de départ  $T$  mais de force supérieure à  $t$ .

## 6.3. Construction d'un tableau orthogonal linéaire de force $t+m$ , $m \geq 0$

Ce paragraphe se décompose en deux sous-parties. La première correspond à la généralisation à la force  $t+m$  du théorème 6.11 et la deuxième à la généralisation du théorème 6.15.

### 6.3.1. Lien entre tableau orthogonal linéaire de force $t+m$ et matrice aux différences

Soit la matrice

$$T' = \left( \frac{g + T}{\forall g \in GF(p)} \right)$$

nous allons regarder, dans un premier temps, quelles conditions imposer à  $T$  pour que  $T'$  soit linéaire et de force  $t+m$ ,  $m \geq 0$ . Nous obtenons ainsi une généralisation des résultats de Adhikary et Das (1992).

#### Proposition 6.19

*Soit  $T$  un tableau orthogonal linéaire de force  $t-1$  sur  $K^d = GF(p)^d$  tel que  $rgT = d-r$ .*

*Le tableau*

$$T' = \left( \frac{g + T}{\forall g \in GF(p)} \right)$$

*est orthogonal linéaire de force  $t+m$ ,  $m \geq 0$  si la matrice  $M = \begin{pmatrix} T \\ \mathbf{1}_d \end{pmatrix}$  vérifie*

*1)  $rgM = d-r+1$*

*2) toute famille de  $t+m$  colonnes de  $M$  est libre.*

**Démonstration :** Tout vecteur ligne de  $T'$  appartient à  $\text{Im}({}^tT \mid \mathbf{1}_d)$ . Si  $rgM = d-r$ , alors une ligne de  $T$  est égale à  ${}^t\mathbf{1}_d$  et  $T'$  est alors formé de  $p$  répétitions de  $T$ . Il faut donc  $rgM = d-r+1$  pour que les lignes de  $T'$  forment un sous-espace vectoriel de  $K^d$ .

Si  $t+m$  colonnes de  $M$  forment une famille libre alors les  $t+m$  colonnes de  $T'$  correspondantes sont aussi libres (attention la réciproque est fausse).

Les deux conditions de la proposition nous permettent donc d'appliquer le lemme 6.14 et d'en déduire que  $T'$  est un tableau orthogonal linéaire de force  $t+m$ .  $\square$

Le théorème suivant nous permet d'établir, en quelque sorte, une réciproque de la proposition 6.19 en modifiant la deuxième condition.

## Théorème 6.20

*On reprend les notations de la proposition 6.19.*

*Pour tout  $J \subseteq \{1, \dots, d\}$ , notons  $P_J$  le projecteur orthogonal canonique de  $K^d$  sur  $K^J$  ( $P_J : x = (x_1, \dots, x_d) \in K^d \rightarrow (x_j)_{j \in J}$ ).*

*Soit par ailleurs  $\varepsilon_T$  le sous-espace de  $K^d$  formé des lignes de  $T$ .*

*Supposons  $\text{rg}M = d - r + 1$ .*

*Pour que  $T'$  soit un tableau orthogonal linéaire de force  $t+m$ ,  $m \geq 0$ , il faut et il suffit que*

*$\forall J \subseteq \{1, \dots, d\}, |J| = t+m$*

*ou bien  $\mathbf{1}_{t+m} \in \text{Im}_{P_J} \varepsilon_T$  et les  $t+m$  colonnes de  $T$  indexées par  $J$  sont libres.*

*ou bien  $\mathbf{1}_{t+m} \notin \text{Im}_{P_J} \varepsilon_T$  et les  $t+m$  colonnes de  $M$  indexées par  $J$  sont libres.*

**Démonstration :** Comme  $\text{rg}M = d - r + 1$ ,  $\varepsilon_T$ , l'espace vectoriel formé des lignes de  $T'$ , est un sous-espace vectoriel de  $K^d$ .

Pour tout  $J \subseteq \{1, \dots, d\}, |J| = t+m$

$$\text{Im}_{P_J} \varepsilon_{T'} = \begin{cases} \text{Im}_{P_J} \varepsilon_T & \text{si } \mathbf{1}_{t+m} \in \text{Im}_{P_J} \varepsilon_T \\ \text{Im}_{P_J} \varepsilon_T \oplus \langle \mathbf{1}_{t+m} \rangle & \text{sinon} \end{cases}$$

où  $\langle v \rangle$  désigne le sous-espace engendré par  $v$ .

Or d'après le lemme 6.14,  $T'$  est un tableau orthogonal linéaire de force  $t+m$  si et seulement si

$$K^J = \text{Im}_{P_J} \varepsilon_{T'}.$$

Donc si  $\mathbf{1}_{t+m} \in \text{Im}_{P_J} \varepsilon_T$ , alors  $K^J = \text{Im}_{P_J} \varepsilon_T$ . Ce qui équivaut à dire que  $\text{rg}(T_j, j \in J) = t+m$  où  $T_j$  est la  $j$ ème colonne de  $T$ , autrement dit, que les  $t+m$  colonnes de  $T$  indexées par  $J$  sont libres.

Si  $\mathbf{1}_{t+m} \notin \text{Im}_{P_J} \varepsilon_T$ , alors  $K^J = \text{Im}_{P_J} \varepsilon_T \oplus \langle \mathbf{1}_{t+m} \rangle = \text{Im}_{P_J} \varepsilon_M$ . Ce qui équivaut à dire que  $\text{rg}(M_j, j \in J) = t+m$  où  $T_j$  est la  $j$ ème colonne de  $T$ , autrement dit, que les  $t+m$  colonnes de  $M$  indexées par  $J$  sont libres.  $\square$

**Remarque :** Rappelons que si  $T'$  est un tableau orthogonal de force  $t+m$ , alors  $T$  est une matrice aux différences de force  $t+m$ .

On obtient un nouvel énoncé du théorème 6.20 en faisant intervenir la matrice génératrice du dual. Ce nouveau théorème est en fait une généralisation du théorème 6.11 à la force  $t+m$ .

## Théorème 6.21

Soit  $P$  un matrice d'ordre  $r \times d$  à éléments dans  $K=GF(q)$ . Supposons  $rgP=r$ , soit alors le tableau d'ordre  $p^{d-r} \times d$  dont les lignes sont les solutions de  $Px=0$ . Soient de plus les entiers  $t, 0 \leq t \leq d$  et  $m, 0 \leq m < d-t$ .

Pour tout  $J \subseteq \{1, \dots, d\}$ , on note  $v_J(P)$  le sous-espace de  $K^d$  formé des vecteurs de  $Im^t P$  dont le support est contenu dans  $J$ ,

$$v_J(P) = \{y \in Im^t P ; \text{supp}(y) \subseteq J\}.$$

$T$  est à la fois un tableau orthogonal linéaire de force  $t-1$  et une matrice aux différences de force  $t+m$  si et seulement si

1)  $wt(P) \geq t$

2) pour tout  $J \subseteq \{1, \dots, d\}, |J| = t+m$  ; soit  $v_J(P) = \emptyset$ , soit  $\dim v_J(P) = 1$  et  $v_J(P)$

n'est pas orthogonal à  $1_{|J|}$ .

**Démonstration :** Comme nous l'avons déjà vu, la condition 1 est nécessaire et suffisante pour que le tableau linéaire  $T$  soit orthogonal de force  $t-1$ , reste à démontrer la condition 2.

*Condition suffisante :* Montrons par l'absurde que la condition est suffisante. Supposons que  $T$  n'est pas un tableau de force  $t+m$ , i.e d'après la remarque ci-dessus, que  $T$  n'est pas une matrice aux différences de force  $t+m$ . Alors il existe  $J \subseteq \{1, \dots, d\}, |J| = t+m$  tel que les  $t+m$  colonnes de  $T$  indexées par  $J$  ne sont pas indépendantes. Soit  $T'_j$  (resp.  $T_j$ ) la jème colonne de  $T'$  (resp.  $T$ ), on a alors  $\sum_{j \in J} \lambda_j T'_j = 0$  avec  $\lambda_j, j \in J$  non tous nuls, d'où

$$\begin{cases} \sum_{j \in J} \lambda_j = 0 \\ \sum_{j \in J} \lambda_j T_j = 0 \end{cases}$$

Ainsi le vecteur  $y = (y_1, \dots, y_d) \in K^d$  défini par

$$\begin{cases} y_j = \lambda_j & j \in J \\ y_j = 0 & j \notin J \end{cases}$$

vérifie  $\text{supp}(y) \subseteq J$  et  $y \in Im^t P$  (car  $y \in \ker T$ ) donc  $y \in v_J$ . Or  $y \mathbf{1}_d = 0$  avec  $y \neq 0$  donc  $y$  est orthogonal à  $\mathbf{1}_d$ , la condition 2 n'est pas vérifiée.

*Condition nécessaire :* Montrons maintenant, toujours par l'absurde que la condition est nécessaire. Supposons que 1) soit  $\dim v_J(P) > 1$ , 2) soit  $\dim v_J(P) = 1$  et  $v_J(P) \perp \mathbf{1}_d$  pour  $J \subseteq \{1, \dots, d\}, |J| = t+m$ .

Montrons dans un premier temps que  $\forall J \subseteq \{1, \dots, d\}, |J| = t+m$

$$Im_{P_j} \varepsilon_T = \left( Im_{P_j} v_J(P) \right)^\perp.$$

Par définition nous avons  $V_J(P) \subseteq \varepsilon_T^\perp$ , donc  $(x|y)=0$ ,  $\forall x \in V_J(P)$  et  $\forall y \in \varepsilon_T$ . D'autre part, étant donné que pour tout  $y$  de  $V_J(P)$ ,  $\text{supp}(y) \subseteq J$ , nous avons  $(x|y)=(P_J(y)|P_J(x))$ . Donc  $\text{Im}_{P_J} \varepsilon_T$  est orthogonal à  $\text{Im}_{P_J} V_J(P)$ , autrement dit,  $\text{Im}_{P_J} \varepsilon_T \subseteq \left(\text{Im}_{P_J} V_J(P)\right)^\perp$ .

Regardons maintenant l'inclusion inverse. Supposons sans perte de généralité que  $J=\{1,\dots,t+m\}$  et notons  $\bar{J}=\{t+m+1,\dots,d\}$ ,  $\varepsilon_T^\perp$  peut alors s'écrire sous la forme  $V_J(P) \oplus V_{\bar{J}}(P)$ . Montrons que tout couple  $(x,y)$  de

$$\left( \left(\text{Im}_{P_J} V_J(P)\right)^{\perp_{K^{(t+m)}}}, \left(\text{Im}_{P_{\bar{J}}} V_{\bar{J}}(P)\right)^{\perp_{K^{(d-t-m)}}} \right)$$

appartient aussi à  $\varepsilon_T$ .

$\forall (a,b) \in \varepsilon_T^\perp$ ,  $(a,b) = (v,0) + (0,s)$  où  $(v,0) \in V_J(P)$  et  $(0,s) \in V_{\bar{J}}(P)$ , alors

$$\left( (x,y) \mid (a,b) \right) = \left( (x,y) \mid [(v,0) + (0,s)] \right) = (x|v) + (y|s).$$

Or  $(v,0) \in V_J(P) \Rightarrow v \in \text{Im}_{P_J} V_J(P)$  et  $x \in \left(\text{Im}_{P_J} V_J(P)\right)^{\perp_{K^{(t+m)}}}$  donc  $(x|v)=0$ , de même

$(0,s) \in V_{\bar{J}}(P) \Rightarrow s \in \text{Im}_{P_{\bar{J}}} V_{\bar{J}}(P)$  et  $y \in \left(\text{Im}_{P_{\bar{J}}} V_{\bar{J}}(P)\right)^{\perp_{K^{(d-t-m)}}}$  donc  $(y|s)=0$ . D'où  $\forall (a,b) \in \varepsilon_T^\perp$ ,

$\left( (x,y) \mid (a,b) \right) = 0$ , donc  $(x,y) \in \left(\varepsilon_T^\perp\right)^\perp = \varepsilon_T$ . Nous avons donc

$$\text{Im}_{P_J} \left( \left(\text{Im}_{P_J} V_J(P)\right)^{\perp_{K^{(t+m)}}}, \left(\text{Im}_{P_{\bar{J}}} V_{\bar{J}}(P)\right)^{\perp_{K^{(d-t-m)}}} \right) = \left(\text{Im}_{P_J} V_J(P)\right)^\perp \subseteq \text{Im}_{P_J} \varepsilon_T.$$

D'où l'égalité souhaitée.

Alors dans le cas 1)  $\dim V_J(P) > 1 \Rightarrow \dim(\text{Im}_{P_J} \varepsilon_T) < t+m-1$  et dans le cas 2)  $\dim V_J(P) = 1$  et  $V_J(P) \perp \mathbf{1}_d \Rightarrow \dim(\text{Im}_{P_J} \varepsilon_T) = t+m-1$  et  $\mathbf{1}_d \in \text{Im}_{P_J} \varepsilon_T$ . Par conséquent que 1) ou 2) soit vérifié, nous avons

$$\text{Im}_{P_J} \varepsilon_T = \text{Im}_{P_J} \varepsilon_T + \langle \mathbf{1}_d \rangle$$

qui est un sous-espace strict de  $K^J$  puisque  $|J|=t+m$ . Mais alors  $T'$  ne peut pas être un tableau orthogonal de force  $t+m$  d'après le lemme 6.14. Il s'ensuit par construction de  $T'$  que  $T$  ne peut pas être une matrice aux différences de force  $t+m$ .  $\square$

**Remarques :** 1) Si  $\text{wt}(P) \geq t$  alors  $\forall y \in \text{Im}^\dagger P$ ,  $|\text{supp}(y)| \geq t$ . Si de plus  $m=0$ , alors la condition 2 du théorème entraîne que pour  $y \in \text{Im}^\dagger P$  tel que  $\text{supp}(y) \subseteq J$  avec  $|J|=t$ ,  $\text{supp}(y)=J$ , i.e  $\text{wt}(y)=t$  et  $y_{1_d} \neq 0$ . On retrouve la condition 2 du théorème 6.11.

2) Si deux vecteurs de  $\text{Im}^\dagger P$  ont même support dans  $J$  sans être colinéaires, alors ils admettent une combinaison linéaire dont le support est strictement inclus dans  $J$ .

### 6.3.2. Construction d'un tableau orthogonal linéaire de force $t+m$ , $m \geq 0$ (généralisation du théorème 6.15)

Comme pour le théorème 6.15, nous allons utiliser la notion de matrices associables à laquelle nous intégrerons les conditions imposées par le théorème 6.21 afin que le tableau construit soit de force  $t+m$ ,  $m \geq 0$ .

#### Théorème 6.22

Soit  $P$  une matrice d'ordre  $r \times d$  à éléments dans le groupe additif de  $K=GF(p)$ . Supposons  $rgP=r$ , soit alors  $T$  le tableau d'ordre  $p^{d-r} \times d$  dont les lignes sont les solutions de  $Px=0$ . Soient de plus deux entiers  $t$  et  $m$  tels que  $0 \leq t \leq d$  et  $0 \leq m \leq d-t$ . Pour tout  $I \subseteq \{1, \dots, d\}$ , on note

$$V_I(P) = \{y \in \text{Im}^t P ; \text{supp}(y) \subseteq I\}.$$

Supposons les conditions du théorème 6.21 vérifiées par  $P$ , on note  $D=(0 \mid T)$  la matrice aux différences de force  $t+m$  sur  $K^{d+1}$ .

Soient de plus  $s$  vecteurs de  $K^d$ ,  $v_1, \dots, v_s$ , et pour tout  $J \subseteq \{1, \dots, s\}$ , la matrice d'ordre  $(r+ \mid J) \times d$

$$Q_J = \left( \begin{array}{c|c} P & \\ \hline v_j & j \in J \end{array} \right).$$

Supposons vérifiées par les matrices  $Q_J$  les conditions suivantes.

1)  $\forall J \subseteq \{1, \dots, s\}$  tel que  $1 \leq |J| \leq \min(t-1, s)$

(i)  $wt(Q_J) \geq t - |J|$

(ii)  $\forall I \subseteq \{1, \dots, d\}$  tel que  $|I| = t+m - |J|$  ; soit  $V_I(Q_J) = \emptyset$ , soit  $\dim V_I(Q_J) = 1$  et  $V_I(Q_J)$  n'est pas orthogonal à  $1$ .

2)  $\forall J \subseteq \{1, \dots, s\}$  tel que  $\min(t-1, s) < |J| \leq \min(t+m-1, s)$

(i)  $wt(Q_J) \geq t+m - |J|$

(ii)  $\forall y \in \text{Im}^t Q_J$  tel que  $wt(y) = t+m - |J| \Rightarrow y1_d \neq 0$

3)  $wt(Q_J) \geq 1$  pour  $s \geq t+m$ .

Alors le tableau d'ordre  $p^{d-r} \times s$ ,  $B=TV$  où  $V=(^t v_1, \dots, ^t v_s)$  est orthogonal de force  $\min(t+m, s)$  associable à  $D$ .

**Remarque :** Si  $m \geq t-1$  alors la condition 2-(i) entraîne la condition 1-(i).

Le corollaire 6.16 reste valable ici puisque sa démonstration repose sur la construction des matrices  $D$  et  $B$  qui reste la même. Nous pouvons donc dire que le tableau d'ordre  $p^{d-r+1} \times (d+s+1)$  construit à partir des matrices associables  $B$  et  $D$  est orthogonal linéaire de force  $t+m$ , de matrice génératrice du dual

$$P' = \left( \begin{array}{c|c|c} -P1_d & P & 0 \\ \hline -V1_d & V & -1_s \end{array} \right).$$

**Démonstration :**

1) Pour montrer que B est un tableau orthogonal de force  $w=\min(t+m,s)$ , on procède de la même façon que dans la démonstration du théorème 6.15. C'est-à-dire qu'il faut montrer que pour tout  $J \subseteq \{1, \dots, s\}$  tel que  $|J|=w$  le système

$$Q_J x = \begin{pmatrix} 0 \\ - \\ z \end{pmatrix}$$

admet  $p^{d-r-w}$  solutions pour tout  $z \in K^w$ .

Si  $w=s$ , alors  $|J|=s \leq t+m$ , donc  $|J| \leq \min(t+m-1, s)$ , alors soit  $|J| < t$  et d'après la condition 1-(i)  $\text{wt}(Q_J) \geq t - |J| > 1$ , soit  $t \leq |J| \leq \min(t+m-1, s)$  et d'après la condition 2-(i)  $\text{wt}(Q_J) \geq t+m - |J| \geq 1$ . Si maintenant  $w=t+m$  alors  $s \geq t+m$  et d'après la condition 3  $\text{wt}(Q_J) \geq 1$ . Donc  $\forall J \subseteq \{1, \dots, s\}$  tel que  $|J|=w$ , la matrice  $Q_J$  est de poids supérieur à 1, c'est-à-dire que toutes ses lignes sont indépendantes.  $Q_J$  étant d'ordre  $p^{r+w} \times d$  le système

$$Q_J x = \begin{pmatrix} 0 \\ - \\ z \end{pmatrix}$$

admet donc  $p^{d-r-w}$  solutions pour tout  $z \in K^w$ .

2) D est une matrice d'ordre  $p^{d-r} \times (d+1)$ ,  $D=(0 | T)$ , sa  $i$ ème ligne s'écrit donc  $(0 | x_i)$  où  $P_{x_i}=0$ .  $\forall v$  un entier non nul,  $1 \leq v \leq \min(t+m-1, s)$ ,  $\forall J \subseteq \{1, \dots, s\}$  tel que  $|J|=v$ ,  $\forall z=(z_1, \dots, z_v) \in K^v$ , on note

$$D_{I_{B[J],z}} = (d_{ik} | i \in I_{B[J],z} \text{ et } k = 0, \dots, d)$$

où  $I_{B[J],z} = \{i \in \{1, \dots, p^{d-r}\}; (x_i | v_j) = z_j \forall j \in J\}$ .

Pour montrer que D et B sont associables, il faut montrer que  $D_{I_{B[J],z}}$  est une matrice aux différences de force  $t+m - |J|$ .

Soit  $D'_{J,z}$  le tableau des solutions du système

$$Q_J x = \begin{pmatrix} 0 \\ - \\ z \end{pmatrix},$$

alors  $D_{I_{B[J],z}} = (0 | D'_{J,z})$ .

On a  $\min(t-1, s) \leq \min(t+m-1, s)$ , on peut donc découper l'intervalle en deux.  $\forall J \subseteq \{1, \dots, s\}$ ,  $1 \leq |J| \leq \min(t-1, s)$ , d'après la condition 1,  $Q_J$  vérifie les conditions du théorème 6.21 avec  $t - |J|$ , donc  $D'_{J,z}$  est à la fois un tableau orthogonal linéaire de force  $t - |J|$  et surtout une matrice aux différences de force  $t+m - |J|$ .

$\forall J \subseteq \{1, \dots, s\}$ ,  $\min(t-1, s) < |J| \leq \min(t+m-1, s)$ , d'après la condition 2,  $Q_J$  vérifie les conditions du théorème 6.21 avec  $t+m - |J|$  donc  $D'_{J,z}$  est à la fois un tableau orthogonal linéaire de force  $t+m - |J|$ .

Donc  $\forall J \subseteq \{1, \dots, s\}$ ,  $1 \leq |J| \leq \min(t+m-1, s)$ ,  $D'_{J,z}$  est une matrice aux différences de force  $t+m - |J|$ . D et B sont associables.  $\square$

**Exemple 6.23** :  $d=7, p=3, t=3, m=1$  et  $s=1$ .

Soit  $P$  la matrice d'ordre  $2 \times 7$  à éléments dans  $GF(3)$

$$P=(I_2 \mid M) \text{ où } M = \begin{pmatrix} 0 & 1 & 2 & 1 & 2 \\ 1 & 2 & 0 & 0 & 0 \end{pmatrix}.$$

On peut alors vérifier que  $wt(P)=t=3$  et que pour tout ensemble  $J$  de 4 coordonnées  $V_J(P)$  forme un espace de dimension 1 non orthogonal à  $\mathbf{1}$ .

Les conditions du théorème 6.21 sont donc vérifiées. Soit alors  $D=(0 \mid T)$  où  ${}^tT=\ker P$  la matrice aux différences de force 4 sur  $GF(3)^8$ . Soit maintenant le vecteur

$$v=(1,1,1,1,0,0,1).$$

Etant donné que  $\min(t+m-1,s)=\min(t-1,s)=s=1$  seule la condition 1 du théorème est à vérifier.

On a nécessairement  $J=\{1\}$  et on voit que  $Q={}^t({}^tP \mid {}^t v)$  est de poids supérieur à  $\epsilon_1=2$ . La condition 1-(i) est vérifiée. Soit  $I \subseteq \{1, \dots, 7\}$  tel que  $|I|=t+m-1=3$ ,

si  $I=\{2,3,4\}$  alors  $V_I(Q)=\{(0,1,1,2,0,0,0),(0,2,2,1,0,0,0)\}$ ,

si  $I=\{1,4,7\}$  alors  $V_I(Q)=\{(2,0,0,1,0,0,2),(1,0,0,2,0,0,1)\}$ ,

si  $I=\{1,5,6\}$  alors  $V_I(Q)=\{(1,0,0,0,1,2,0),(2,0,0,0,2,1,0)\}$ ,

dans les 3 cas  $\dim V_I(Q)=1$  et  $V_I(Q)$  n'est pas orthogonal à  $\mathbf{1}$ . Pour les autres ensembles  $I$  de cardinal 3  $V_I(Q)=\emptyset$ , la condition 1-(ii) est donc vérifiée.

Le tableau  $B=TV$  d'ordre  $3^5 \times 2$  est orthogonal de force 1 associable à  $D$ .

Si on note  $H$  la matrice dont les lignes sont les vecteurs de  $GF(3)^5$ , alors

$$T=H(-{}^tM \mid I_5),$$

d'où

$$D=(0 \mid -{}^tM \mid I_5)$$

et

$$B=H^t(2,1,2,1,0).$$

En appliquant le théorème, on construit le tableau orthogonal linéaire de force 4 à 9 paramètres de taille  $3^6$  défini par sa matrice génératrice du dual

$$P' = \left( \begin{array}{c|cc|c} 2 & & & \\ 2 & I_2 & M & 0 \\ \hline 1 & 1111001 & & 2 \end{array} \right).$$

## Bibliographie

- Adhikary B., Das P. (1992). Multiple fold-over and its application to orthogonal arrays. *Sankhya*, special volume **54**, 21-24.
- Bailey R.A (1985). Factorial designs and abelian groups. *Lin. Algebra Appl.* **70**,349-363.
- Bailey R.A., Gilchrist F.H.L, Patterson H.D. (1977). Identification of effects and confounding patterns in factorial designs. *Biometrika* **64**, 347-354.
- Bose R.C., Bush K.A. (1952). Orthogonal arrays of strength two and three. *Ann. Math. Statist.* **23**, 508-524.
- Collombier D. (1996). *Plans d'expériences factoriels*. Springer-Verlag, Berlin.
- Franklin M.F. (1985). Selecting defining contrasts and confounded effects in  $p^{n-m}$  factorial experiments. *Technometrics* **27**, 165-172.
- Franklin M.F., Bailey R.A (1977). Selection of defining contrasts and confounded effects in two-level experiments. *Appl. Statist.* **26**, 321-326.
- Hedayat A.S., Pu K., Stufken J. (1992). On the construction of asymmetrical orthogonal arrays. *Ann. Statist.* **20**, 2142-2152.
- Hedayat A.S., Sloane N.J.A, Stufken J. (1999). *Orthogonal Arrays : Theory and applications*. Springer-Verlag, New-York.
- Hedayat A.S., Stufken J., Su O. (1996). On differences schemes and orthogonal arrays of strength t. *J. of Statist. Planning and Inferences* **56**, 307-324.
- Patterson H.D. (1976). Generation of factorial designs. *J. of Roy Soc. B* **38**, 175-179.
- Patterson H.D., Bailey R.A. (1978). Design keys for factorial experiments. *Appl. Statist.* **27**, 335-343.
- Raghavarao D. (1971). *Construction and combinatorial problems in design of experiments*. Wiley, New York.
- Raktoe B.L. (1969). Combining elements from distinct finite fields in mixed factorials. *Ann. Math. Statist.* **40**,498-504.
- Rao C.R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc. (suppl.)* **9**, 128-139.
- Seiden E. (1954). On the problem of construction of orthogonal array. *Ann. Math. Statist.* **25**, 1451-156.
- Srivastava J.N., Chopra D.V. (1973). Balanced arrays and orthogonal arrays. In *A survey of combinatorial theory* (Srivastava ed.) North Holland, 411-428.
- Wang J.C., Wu C.F.J. (1991). An approach to the construction of asymmetrical orthogonal arrays. *J. Amer. Statist* **86**, 450-456.



## 7. Construction de fractions régulières dans

$$G = \times_{i=1}^d (Z / p^{n_i})$$

Dans le chapitre précédent nous avons construit des tableaux orthogonaux linéaires, c'est-à-dire des cas particuliers de fractions régulières de  $GF(p)^d$ . Le fait de se placer dans un corps de Galois d'ordre  $p$ , implique deux contraintes en ce qui concerne le découpage du domaine d'échantillonnage.

1. Chaque arête du cube unité est découpée en un nombre  $p$  premier de segments.
2. Les arêtes du cubes unité sont découpées en un même nombre de segments, *i.e* que le découpage est symétrique.

Ces contraintes sont très astreignantes car, d'une part, le fait que le nombre de segments soit premier ne laisse pas beaucoup de possibilités,  $p=2, 3, 5$  ou  $7$ , ensuite le nombre de cellules augmentent considérablement. D'autre part, il peut être souhaitable dans certains cas de découper une arête du cube unité plus finement que les autres parce que le facteur qu'elle représente demande une meilleure répartition.

L'objectif de ce chapitre est donc de construire des fractions de

$$G = \times_{i=1}^d (Z / q_i)$$

où  $q_i$  est quelconque et représente le nombre de segments de la  $i$ ème arête. Dans le cadre de notre travail, on souhaite utiliser ces fractions en tant que plan d'échantillonnage pour l'ajustement d'une régression trigonométrique. D'après les résultats du chapitre 4, il est donc nécessaire que

- les fractions soient régulières, c'est-à-dire construites à partir de sous-groupes de  $G$ ,
- le dual du sous-groupe soit connu afin de déterminer les effets confondus.

Nous savons d'après Bailey (1985) et El Mossadeq *et al.* (1985) que la recherche du dual d'un sous-groupe peut être simplifiée en utilisant le fait que  $G$  est somme directe de ses  $p$ -sous-groupes de Sylow que l'on appelle ses composantes primaires.

### **Proposition**

On note  $q$  l'exposant de  $G$ , *i.e*  $q = \text{PPCM}(q_i, i=1, \dots, d)$ . Soient  $P$  l'ensemble des diviseurs premiers de  $q$ ,  $G = \oplus \{G_p, p \in P\}$  est la décomposition primaire de  $G$  et  $S = \oplus \{S_p, p \in P\}$  est un sous-groupe de  $G$ , avec  $S_p \subset G_p$  pour  $p \in P$ .

Si pour tout  $p \in P$ ,  $S_p^\perp$  est le dual de  $S_p$  dans  $G_p$ , alors

$$S^\perp = \oplus \{S_p^\perp, p \in P\}.$$

Ainsi la recherche du dual d'un sous-groupe de  $G$  peut être fractionnée, le problème à résoudre étant celui de la construction du dual d'un sous-groupe d'un  $p$ -groupe. C'est-à-dire que nous nous ramenons et limitons dans ce chapitre au cas où

$$G = \prod_{i=1}^d (Z / p^{n_i})$$

avec  $p$  premier.

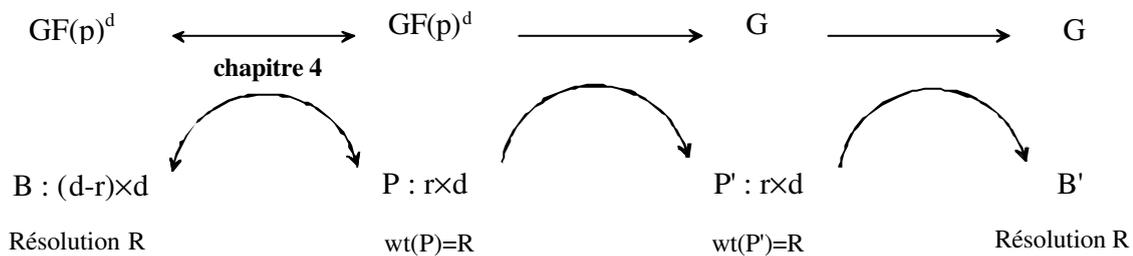
Dans le chapitre précédent?, nous avons vu que dans le cas où  $G$  est le groupe additif de  $GF(p)^d$ , le passage d'un sous-groupe à son dual ne pose pas de problème, puisqu'il s'effectue à l'aide du noyau des matrices génératrices. En revanche, pour  $G$  défini comme ci-dessus, le calcul matriciel n'est plus envisageable puisque  $G$  n'est plus un espace vectoriel. On utilise alors une technique de construction du dual tout à fait différente, proposée par El Mossadeq *et al.* (1985). Nous remarquerons cependant à la fin de ce chapitre qu'une généralisation de la méthode des noyaux au cas où le découpage du domaine est symétrique a été établie en théorie des codes.

La connaissance du dual du sous-groupe permet d'étudier l'estimabilité des paramètres de la régression trigonométrique, mais aussi de déterminer la résolution de la fraction régulière. Nous allons imposer à ce niveau une contrainte à la méthode de construction. Nous souhaitons déduire les fractions régulières de  $G$  à partir des tableaux orthogonaux linéaires construits au chapitre précédent, de façon à ce que les résolutions soient identiques. La démarche que nous allons suivre est alors la suivante.

Soit  $P$  une matrice de poids  $R$ , génératrice du dual d'un tableau orthogonal linéaire à éléments dans  $GF(p)$ . Nous allons tout d'abord construire à partir de  $P$ , une matrice  $P'$  de poids  $R$  dont les lignes appartiennent à  $G$ .

Cette matrice  $P'$  peut être considérée comme la matrice génératrice du dual  $S'^{\perp}$  d'un sous-groupe  $S'$  de  $G$ . Il suffit ensuite d'appliquer à  $S'^{\perp}$  la méthode de construction du dual d'un sous-groupe proposée par El Mossadeq *et al.* (1985) pour avoir une fraction régulière  $S'$  de résolution  $R$ . En effet

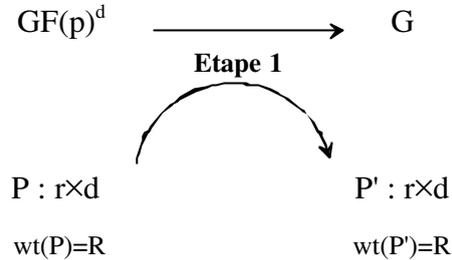
$$(S'^{\perp})^{\perp} = S' \text{ et } wt(S')=R \text{ par construction.}$$



où  $B'$  est la matrice génératrice de la fraction régulière  $S'$ .  $G$  n'étant pas un espace vectoriel, le terme matrice génératrice, signifie que le sous-groupe est somme directe des sous-groupes cycliques engendrés par les lignes de la matrice.

## 7.1. Première étape

Dans le chapitre précédent, nous avons vu comment construire des matrices  $P$  de poids  $R$  à éléments dans  $GF(p)$ . Dans cette première étape, nous allons montrer comment passer de  $P$  à une matrice  $P'$  dont les lignes appartiennent à  $G$ , de même poids que  $P$ .



On note  $l_1, \dots, l_r$  les lignes de la matrice  $P$ , où  $r$  est le rang de  $P$ .

Rappelons que la difficulté de cette construction réside dans le fait que l'on passe d'un espace vectoriel,  $GF(p)^d$ , à un  $Z$ -module,  $G$ . Afin de contourner ce problème on introduit un espace intermédiaire

$$G_p = \{x \in G, px=0\},$$

qui a pour caractéristique d'être à la fois un sous-ensemble de  $G$  et d'avoir une structure de  $GF(p)$ -espace vectoriel. Nous suivons ainsi la démarche de El Mossadeq (1996) qui consiste à décomposer le passage de  $P$  à  $P'$  en trois sous-étapes.

Dans un premier temps, nous allons montrer qu'il existe un isomorphisme d'espace vectoriel  $\Phi$  entre  $GF(p)^d$  et  $G_p$  qui conserve le poids. C'est-à-dire qu'à tout sous-espace  $S$  de  $GF(p)^d$  de poids  $R$ , on peut associer un sous-groupe  $S_p$  de  $G_p$  de même poids

$$\begin{aligned}
 \Phi : \quad GF(p)^d &\rightarrow G_p \\
 S; wt(S) = R &\mapsto S_p; wt(S_p) = R.
 \end{aligned}$$

Ensuite, nous allons montrer que s'il existe un sous-groupe  $S'$  de  $G$  vérifiant une condition par rapport à  $S_p$ , alors  $S'$  est aussi de poids  $R$ . Pour finir nous allons expliquer comment construire les générateurs de ce sous-groupe  $S'$  à partir des générateurs du sous-groupe  $S$ , c'est-à-dire comment construire la matrice  $P'$  à partir de  $P$  pour qu'elle soit de poids  $R$ .

### Proposition 7.1

*Soit  $\Phi$  l'application canonique*

$$\begin{aligned}
 \Phi : \quad GF(p)^d &\rightarrow G_p \\
 x = (x_1, \dots, x_d) &\mapsto x_p = (x_1 p^{n_1-1}, \dots, x_d p^{n_d-1})
 \end{aligned}$$

*Soit  $S$  un sous-groupe de  $GF(p)^d$ , de poids  $R$ , engendré par les vecteurs  $l_1, \dots, l_r$ . Alors le sous-ensemble de  $G_p$ ,*

$$S_p = \Phi(S),$$

*est un sous-espace vectoriel de  $G_p$  de poids  $R$ , engendré par les vecteurs  $\Phi(l_1), \dots, \Phi(l_r)$ .*

**Exemple 7.2 :** Soit  $P$  la matrice génératrice du dual d'une fraction régulière de  $GF(3)^4$ , de résolution 3,

$$P = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

Examinons le cas où le découpage du domaine est symétrique, par exemple  $G=(Z/9)^4$ , i.e  $n_1=\dots=n_4=2$ .

L'application  $\Phi$  est alors définie par

$$\Phi(x)=3x.$$

Le sous-espace vectoriel  $S_p$  est engendré par  $\Phi(l_1)=(3,0,6,1)$  et  $\Phi(l_2)=(0,3,3,3)$ . Il est de poids 3 comme  $P$ , et d'ordre 9 puisque  $\Phi(l_1)$  et  $\Phi(l_2)$  sont chacun d'ordre 3.

**Exemple 7.3 :** Soit  $P$  la matrice génératrice du dual d'une fraction régulière de  $GF(2)^5$ , de résolution 3,

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Soit maintenant le cas non symétrique, par exemple  $n_1=4, n_2=3$  et  $n_3=n_4=n_5=2$ , i.e  $G=(Z/16) \times (Z/8) \times (Z/4) \times (Z/4) \times (Z/4)$ .

L'application  $\Phi$  est alors définie par

$$\Phi(x)=(2^3 x_1, 2^2 x_2, 2x_3, 2x_4, 2x_5).$$

Le sous-espace vectoriel  $S_p$  est engendré par  $\Phi(l_1)=(8,0,2,2,2)$  et  $\Phi(l_2)=(0,4,2,2,0)$ . Il est de poids 3 comme  $P$ , et d'ordre 4 puisque  $\Phi(l_1)$  et  $\Phi(l_2)$  sont chacun d'ordre 2,

$$S_p = \{(0,0,0,0,0), (8,0,2,2,2), (0,4,2,2,0), (8,4,0,0,2)\}.$$

**Démonstration :** 1) Montrons tout d'abord que  $S_p$  est un sous-espace vectoriel de  $G_p$  engendré par  $\Phi(l_1), \dots, \Phi(l_r)$ .

Puisque  $S$  et  $G_p$  sont des espaces vectoriels, il suffit de montrer que  $\Phi$  est un isomorphisme d'espaces vectoriels.

- $\Phi$  est une application linéaire puisque  $\forall x \in GF(p)^d, \forall y \in GF(p)^d$  et  $\forall \lambda \in GF(p)$ ,

$$\Phi(x + y) = ((x_1 + y_1)p^{n_1-1}, \dots, (x_d + y_d)p^{n_d-1})$$

$$= (x_1 p^{n_1-1}, \dots, x_d p^{n_d-1}) + (y_1 p^{n_1-1}, \dots, y_d p^{n_d-1}) = \Phi(x) + \Phi(y)$$

$$\Phi(\lambda x) = (\lambda x_1 p^{n_1-1}, \dots, \lambda x_d p^{n_d-1}) = \lambda (x_1 p^{n_1-1}, \dots, x_d p^{n_d-1}) = \lambda \Phi(x).$$

- $\Phi$  est une application injective car si  $\Phi(x)=0$  alors  $x_i p^{n_i-1} = 0 \pmod{p^{n_i}}$   $i=1, \dots, d$ , d'où  $x_i = \alpha p$ ,  $\alpha \in \mathbb{N}$ , c'est-à-dire  $x=0 \pmod{p}$ .
- $\Phi$  est une application surjective puisqu'à chaque  $x_p = (x_1 p^{n_1-1}, \dots, x_d p^{n_d-1}) \in G_p$  on associe trivialement  $x=(x_1, \dots, x_d) \in GF(p)^d$ .

2) Montrons maintenant que  $S$  et  $S_p$  sont de même poids.

Soit  $x=(x_1, \dots, x_d) \in S$ ,  $x$  est de poids supérieur ou égal à  $R$ . Supposons sans perte de généralité que  $x_i \neq 0 \pmod{p}$   $i=1, \dots, R$ , alors  $x_i p^{n_i-1} \neq 0 \pmod{p^{n_i}}$   $i=1, \dots, R$  puisque  $0 < x_i < p$ . Donc  $\phi(x)$  possède au moins  $R$  composantes non nulles.  $\square$

**Proposition 7.4**

Soit  $S_p = \Phi(S)$  où  $S$  est le sous-groupe de poids  $R$  de  $GF(p)^d$ , engendré par  $l_1, \dots, l_r$ . Supposons qu'il existe un sous-groupe  $S'$  de  $G$  tel que  $S' \cap G_p$  soit un sous-groupe de  $S_p$ , i.e

$$S' \cap G_p \leq S_p.$$

Alors  $S'$  est de poids  $R$ .

**Exemple 7.3 (suite) :** Soit  $S'$  le sous-groupe de  $G$  engendré par les vecteurs  $(4,0,1,1,1)$  et  $(8,0,2,2,2)$ . Alors on peut vérifier que  $S' \cap G_p \leq S_p$ , donc que  $S'$  est de poids 3.

**Démonstration :** Soit  $x' \in S'$ , notons  $p^k$  son ordre. Alors d'une part nous avons  $p^{k-1}x' \in S'$ , et d'autre part,

$$p^k x' = 0 \Rightarrow p(p^{k-1}x') = 0 \Rightarrow p^{k-1}x' \in G_p,$$

d'où  $p^{k-1}x' \in S' \cap G_p \leq S_p$ . Par conséquent  $p^{k-1}x'$  est de poids supérieur ou égal à  $R$ , puisque  $Q_p$  est de poids  $R$ . Donc nécessairement  $x'$  est de poids supérieur ou égal à  $R$ .  $\square$

Tout le problème maintenant est de construire le sous-groupe  $S'$  vérifiant  $S' \cap G_p \leq S_p$ . Pour cela nous avons besoin des définitions suivantes.

**Définitions 7.5**

Soient  $l_1, \dots, l_r$  les lignes d'une matrice  $P$  de rang  $r$  et de poids  $R$ , et  $G$  le groupe abélien fini  $G = \times (Z/q_i)$  où  $q_i = p^{n_i} \forall i=1, \dots, d$ .

On note  $n_{min} = \min\{n_k \mid k \in \text{supp}(l_i)\}$ , puis pour  $m=1, \dots, n_{min}$   $C_m(l_i)$  l'ensemble des  $y=(y_1, \dots, y_d)$  de  $G$  tels que

$$\begin{cases} y_j = l_{ij} p^{n_j-m} \pmod{p^{n_j-m+1}} & \text{si } j \in \text{supp}(l_i) \\ y_j = 0 \pmod{p^{n_j-m+1}} & \text{si } j \notin \text{supp}(l_i) \text{ et } n_j \geq m. \\ y_j = a_j \text{ quelconque } \pmod{p^{n_j}} & \text{si } j \notin \text{supp}(l_i) \text{ et } n_j < m \end{cases}$$

On définit alors le sous-ensemble de  $G$ ,

$$C(l_i) = \bigcup_{m=1}^{n_{min}} C_m(l_i).$$

L'ensemble  $C(l_i)$  permet d'associer à un vecteur de  $GF(p)^d$  (une ligne de  $P$ ) plusieurs éléments de  $G$ . Nous allons voir dans la proposition suivante comment construire le sous-groupe  $S'$  tel que  $S' \cap G_p \leq S_p$ , à partir des ensembles  $C(l_i)$ ,  $i=1, \dots, d$ , mais auparavant notons une particularité des éléments de  $C(l_i)$ .

**Lemme 7.6**

*Soit  $l_i$  une ligne de la matrice  $P$ , alors tout élément de  $C_m(l_i)$  est d'ordre  $p^m$ , pour  $m=1, \dots, n_{\min}$ .*

**Remarque :** Dans le cas où  $G$  est symétrique, i.e  $G = \times (Z/p^n)$ , on a nécessairement  $n_{\min} = n$ . Donc un élément de  $C(l_i)$  peut être au maximum d'ordre  $p^n$ .

**Démonstration :** Soit  $y \in C_m(l_i)$ , Montrons alors que l'ordre de  $y$  divise  $p^m$ , i.e que  $p^m y = 0 \Leftrightarrow p^m y_j = 0 \pmod{p^{n_j}}$   $j=1, \dots, d$ . Pour cela on utilise la définition de  $C_m(l_i)$ ,

Si  $j \in \text{supp}(l_i)$ , alors  $y_j = l_{ij} p^{n_j - m} \pmod{p^{n_j - m + 1}} \Rightarrow p^m y_j = l_{ij} p^{n_j} \pmod{p^{n_j + 1}} = 0 \pmod{p^{n_j}}$ .

Si  $j \notin \text{supp}(x)$  et  $n_j \geq m$ , alors  $y_j = 0 \pmod{p^{n_j - m + 1}} \Rightarrow p^m y_j = 0 \pmod{p^{n_j + 1}} = 0 \pmod{p^{n_j}}$ .

Si  $j \notin \text{supp}(x)$  et  $n_j < m$ , alors  $y_j = a_j \pmod{p^{n_j}} \Rightarrow p^m y_j = p^{m - n_j} \times p^{n_j} \times a_j = 0 \pmod{p^{n_j}}$  car  $m - n_j > 0$ .

Ainsi  $p^m y = 0$ . Pour montrer que  $y$  est exactement d'ordre  $p^m$ , il suffit de vérifier qu'il n'est pas d'ordre  $p^{m-1}$ . Supposons le contraire, i.e  $p^{m-1} y = 0$ , alors pour  $j \in \text{supp}(l_i)$ , on a

$$p^{m-1} y_j = l_{ij} p^{n_j - 1} \pmod{p^{n_j}} = 0 \pmod{p^{n_j}},$$

d'où  $l_{ij} = 0 \pmod{p}$ . Or  $l_i$  est une ligne de  $P$ , donc  $l_{ij} \in (Z/p)$ . Nécessairement  $l_{ij} = 0$ , ce qui n'est pas possible puisque  $j \in \text{supp}(l_i)$ .  $\square$

**Exemple 7.3 (suite) :** Rappelons que  $G = (Z/16) \times (Z/8) \times (Z/4) \times (Z/4) \times (Z/4)$ . Nous allons construire  $C(l_1)$  où  $l_1 = (1, 0, 1, 1, 1)$ .

$\text{Supp}(l_1) = \{1, 3, 4, 5\}$ , donc  $n_{\min} = 2$ .

Pour  $m=1$

	Z/16	Z/8	Z/4	Z/4	Z/4
$l_{ij}$	1	0	1	1	1
$n_j - m$	3	2	1	1	1
$y_j$	$2^3 \pmod{16}$	$0 \pmod{8}$	$1 \pmod{4}$	$1 \pmod{4}$	$1 \pmod{4}$

donc  $C_1(l_1) = \{(8, 0, 1, 1, 1)\}$ . On remarque que le seul élément de  $C_1(l_1)$  est d'ordre  $p^m = 2$ .

Pour  $m=2$

	$\mathbb{Z}/16$	$\mathbb{Z}/8$	$\mathbb{Z}/4$	$\mathbb{Z}/4$	$\mathbb{Z}/4$
$l_j$	1	0	1	1	1
$n_j - m$	2	1	0	0	0
$y_j$	$2^2 \pmod{8}$	$0 \pmod{4}$	$1 \pmod{2}$	$1 \pmod{2}$	$1 \pmod{2}$

donc  $C_2(l_1) = \{(y_1, y_2, y_3, y_4, y_5) \mid y_1 \in \{4, 12\}, y_2 \in \{0, 4\}, y_3, y_4 \text{ et } y_5 \in \{1, 3\}\}$ . On remarque que tous les éléments de  $C_2(l_1)$  sont d'ordre  $p^m=4$ .

Finalement  $C(l_1) = C_1(l_1) \cup C_2(l_1)$ , donc à la première ligne de  $P$ , on peut associer  $2^5+1$  éléments de  $G$ .

### Proposition 7.7

*Soient  $y_1, \dots, y_r$  un  $r$ -uplet de  $C(l_1) \times \dots \times C(l_r)$  et  $S'$  le sous-groupe de  $G$  engendré par  $y_1, \dots, y_r$ . Alors  $S'$  est tel que*

$$S' \cap G_p \leq S_p.$$

**Démonstration :** 1) Supposons dans un premier temps que  $r=1$ , i.e que  $S'$  est un groupe cyclique engendré par  $y \in C(l_1)$ .

$y$  appartient à  $C(l_1)$  donc il existe  $n \leq n_{\min}$  tel que  $y$  appartienne à  $C_n(l_1)$ . Alors d'après le lemme précédent,  $y$  est d'ordre  $p^m$ .  $S' = \langle y \rangle$ , donc les éléments d'ordre  $p$  de  $S'$  appartiennent au sous-groupe cyclique engendré par  $p^{m-1}y$ , i.e  $S' \cap G_p = \langle p^{m-1}y \rangle$ .

Montrons maintenant que les éléments de ce sous-groupe appartiennent à  $S_p$ , i.e  $S' \cap G_p = \langle p^{m-1}y \rangle \leq S_p$ .

$$\text{Si } j \in \text{supp}(l_1), \text{ alors } y_j = l_j p^{n_j - m} \pmod{p^{n_j - m + 1}} \Rightarrow p^{m-1}y_j = l_j p^{n_j - 1} \pmod{p^{n_j}}.$$

$$\text{Si } j \notin \text{supp}(x), \text{ alors } l_j = 0 \pmod{p^{n_j}}$$

$$\text{si } n_j \geq m \text{ alors } y_j = 0 \pmod{p^{n_j - m + 1}} \Rightarrow p^{m-1}y_j = 0 \pmod{p^{n_j}}.$$

$$\text{si } n_j < m \text{ alors } y_j = a_j \pmod{p^{n_j}} \Rightarrow p^{m-1}y_j = p^{m-n_j-1} p^{n_j} \times a_j = 0 \pmod{p^{n_j}} \text{ car } m-n_j-1 \geq 0.$$

$$\text{Ce qui peut encore s'écrire } p^{m-1}y_j = l_j p^{n_j - 1} \pmod{p^{n_j}}.$$

Ainsi  $p^{m-1}y$  est de la forme  $(l_{11}p^{n_1-1}, \dots, l_{1d}p^{n_d-1})$ , il appartient donc à  $S_p$ . Par conséquent  $\langle p^{m-1}y \rangle \leq S_p$ .

2) Supposons maintenant  $r$  quelconque. Pour  $i=1, \dots, r$ , on note  $m_i$  l'entier tel que  $y_i \in C_{m_i}(l_i)$ . D'une part nous savons d'après le lemme précédent que  $y_i$  est d'ordre  $p^{m_i}$  et d'autre part d'après ce qui précède, nous pouvons écrire

$$p^{m_i-1}y_i = (l_{i1}p^{n_1-1}, \dots, l_{id}p^{n_d-1}).$$

S' est le sous-groupe engendré par  $y_1, \dots, y_r$ , donc tout élément  $s'$  de S' peut s'écrire

$$s' = \alpha_1 y_1 + \dots + \alpha_r y_r \text{ où } 0 \leq \alpha_i < p^{m_i}, i=1, \dots, r.$$

On note  $m = \max\{m_i, i=1, \dots, r\}$ , alors nous allons montrer ici aussi que les éléments d'ordre  $p$  de S' sont de la forme  $p^{m-1}s'$  et que  $p^{m-1}s' \in S_p$ .

$$p^m s' = \sum_{i=1}^r p^m \alpha_i y_i = \sum_{i=1}^r \alpha_i p^{m-m_i} (p^{m_i} y_i) = 0.$$

Donc les éléments de S' d'ordre  $p$  sont de la forme  $p^{m-1}s'$  or

$$\begin{aligned} p^{m-1}s' &= \sum_{i=1}^r \alpha_i p^{m-m_i} (p^{m_i-1} y_i) = \sum_{i=1}^r \alpha_i p^{m-m_i} (l_{i1} p^{n_1-1}, \dots, l_{id} p^{n_d-1}) \\ &= \sum_{i=1}^r \beta_i \Phi(l_i) \text{ avec } 0 \leq \beta_i < p^m \end{aligned}$$

D'où d'après la proposition 7.8  $p^{m-1}s' \in S_p$ . Par conséquent  $S' \cap G_p \leq S_p$ .  $\square$

Le théorème suivant résulte immédiatement des propositions.

### Théorème 7.9

*Soit P une matrice d'ordre  $r \times d$ , de rang  $r$ , à éléments dans  $GF(p)$  de poids  $R$ . On note  $l_1, \dots, l_r$  les lignes de P.*

*Toute matrice P', dont les lignes forment un  $r$ -uplet de  $C(l_1) \times \dots \times C(l_r) \subset G$  est de poids  $R$ .*

**Remarque :** Donc à chaque fraction régulière de  $GF(p)^d$  de résolution  $R$  définie par sa matrice génératrice du dual  $P$ , on associe plusieurs matrices  $P'$  de poids  $R$ , dont les lignes appartiennent à  $G$ . Il n'y a que dans le cas où  $n_{\min}=1$  pour toutes les lignes de  $P$  que la matrice  $P'$  est unique.

**Exemple 7.2 (suite) :** Rappelons que  $P$  est la matrice à éléments dans  $GF(3)$ , de poids 3,

$$P = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

et que  $G = (Z/9)^4$ . Dans ce cas  $n_j=2, j=1, \dots, 4$ , donc nécessairement  $n_{\min}=2$  pour toutes les lignes de  $P$ , d'où

$l_i$	$\text{supp}(l_i)$	$C_1(l_i)$	$C_2(l_i)$
(1,0,2,1)	{1,3,4}	(3,0,6,3)	$y_1$ et $y_4 \in \{1,4,7\}$ ; $y_2 \in \{0,3,6\}$ ; $y_3 \in \{2,5,8\}$
(0,1,1,1)	{2,3,4}	(0,3,3,3)	$y_1 \in \{0,3,6\}$ ; $y_2, y_3$ et $y_4 \in \{1,4,7\}$

Nous pouvons donc construire dans  $G$ ,  $(3^4+1)^2$  matrices  $P'$  de poids 3 dont la matrice  $P$  elle-même.

**Exemple 7.3 (suite) :** Rappelons que  $P$  est la matrice à éléments dans  $GF(2)$ , de poids 3,

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix},$$

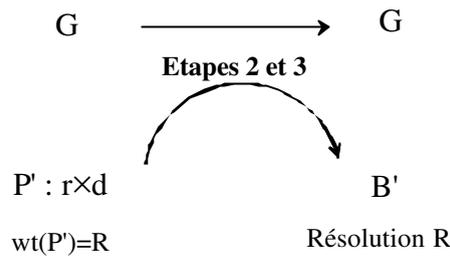
$k_i$	$\text{supp}(k_i)$	$n_{\min}$	$C_1(k_i)$	$C_2(k_i)$
(1,0,1,1,1)	{1,3,4,5}	2	(8,0,2,2,2)	$y_1 \in \{4,12\}; y_2 \in \{0,4\}; y_3, y_4 \text{ et } y_5 \in \{1,3\}$
(0,1,1,1,0)	{2,3,4}	2	(0,4,2,2,0)	$y_1 \in \{0,8\}; y_2 \in \{2,6\}; y_3 \text{ et } y_4 \in \{1,3\}; y_5 \in \{0,2\}$

et que  $G = (\mathbb{Z}/16) \times (\mathbb{Z}/8) \times (\mathbb{Z}/4) \times (\mathbb{Z}/4) \times (\mathbb{Z}/4)$ .

Nous pouvons donc construire dans  $G$ ,  $(2^5+1)^2$  matrices  $P'$  de poids 3.

## 7.2. Passage de P' à B' dans G

Dans la première étape nous avons construit une matrice P' de poids R dont les lignes appartiennent à G. On peut alors définir le dual S<sup>⊥</sup> d'une fraction régulière de G de résolution R, comme somme directe des sous-groupes cycliques engendrés par les lignes de P'. Le problème maintenant est de savoir comment reconstruire la fraction régulière S à partir de son dual S<sup>⊥</sup>.



où B' est la matrice génératrice de S, i.e que la fraction régulière est somme directe des sous-groupes cycliques engendrés par les lignes de B'.

Dans le cas où G est un corps de Galois, nous avons vu dans le chapitre 5 que le passage de l'un à l'autre est simple à l'aide du noyau des matrices génératrices (proposition 5.15). En revanche dans le contexte de ce chapitre, le calcul matriciel n'est plus envisageable puisque nous sommes dans le cas d'un anneau qui n'est pas intègre. Dans un article de 1985, El Mossadeq *et al.* proposent une méthode pour construire le dual d'un ensemble engendré par  $x_1, \dots, x_r$  éléments de G. Nous allons présenter ici cette technique et l'appliquer aux lignes de P'. On obtient alors une fraction régulière de G de résolution R puisque

$$S = (S^\perp)^\perp = \langle x_1, \dots, x_r \rangle^\perp.$$

Rappelons que S<sup>⊥</sup> est défini par  $\{y \in G; (y | x) = 0 \ x \in G\}$  où  $(y | x) = \sum_i x_i y_i q / q_i \pmod{q}$ .

Contrairement à l'article de El Mossadeq *et al.* (1985), les générateurs  $x_1, \dots, x_r$  que nous utilisons ne sont pas quelconques puisqu'ils représentent les lignes d'une matrice P' construite à l'étape 1. Nous allons donc nous intéresser aux conséquences de cette particularité sur la taille de la fraction régulière S.

La construction de S à partir de S<sup>⊥</sup> se décompose en deux étapes. Pour la première, nous supposons que S<sup>⊥</sup> est engendré par un seul élément x de G.

### 7.2.1. Deuxième étape : Construction de $\langle x \rangle^\perp$

Dans cette deuxième étape, nous allons construire le dual de  $\langle x \rangle$  dans G. x représente une ligne de P' construit à l'étape 1, or nous avons vu qu'il existe plusieurs choix possibles pour cet élément. Nous allons montrer comment choisir x pour que  $\langle x \rangle^\perp$  soit de taille minimale.

Soit  $x=(x_1,\dots,x_d)\in G$  et  $o_1,\dots,o_d$  les ordres respectifs des composantes de  $x$ . Supposons sans perte de généralité que  $x$  s'écrit sous la forme

$$x_i = h_i \frac{q_i}{o_i}, \quad 1 \leq i \leq k \quad \text{et} \quad x_i = 0, \quad k < i \leq d,$$

avec la suite  $o_1,\dots,o_k$  ordonnée par valeurs décroissantes.

### **Théorème 7.10**

*On note les vecteurs  $e_i=(\delta_{ij} \mid j=1,\dots,d)$  avec  $\delta_{ij}$  le symbole de Kronecker. Soit  $I$  l'ensemble formé des indices  $I$  et  $i \in \{2,\dots,k\}$  tels que*

$$\frac{p^{n_i}}{o_i} < \frac{p^{n_j}}{o_j} \quad \text{pour } j < i.$$

*Alors  $\langle x \rangle^\perp$  est somme directe des sous-groupes cycliques engendrés par  $y_1,\dots,y_d$ , définis par*

$$\begin{cases} y_1 = o_m e_m \\ y_i = h_j e_i - \frac{h_i o_j}{o_i} e_j & 2 \leq i \leq k, \\ y_i = e_i & i > k \end{cases}$$

*où  $m = \max\{i, i \in I\}$  et  $j < i$  est l'indice tel que*

$$\begin{cases} \frac{p^{n_i}}{o_i} \geq \frac{p^{n_j}}{o_j} & \text{si } i \notin I \\ j \text{ précède } i \text{ dans } I & \text{si } i \in I \end{cases}.$$

La démonstration de ce théorème se trouve dans l'article de El Mossadeq *et al.* (1985) et ne présente pas de difficulté particulière, c'est pourquoi nous ne la reprenons pas ici.

Avant d'appliquer ce résultat à un exemple, intéressons nous tout d'abord à la taille  $N$  du sous-groupe construit. Par définition, celle-ci est égale à  $|G|/|x|$ , or avec les notations ci-dessus  $o_1$  est l'ordre de  $x$ , d'où

$$\frac{1}{o_1} \times \prod_{i=1}^d p^{n_i}.$$

Donc le sous-groupe est de taille minimale si  $x$  est d'ordre maximal.

### **Proposition 7.11**

*Soient  $i \in \{1,\dots,r\}$  tel que  $x \in C(l_i)$  et  $n_{\min} = \min\{n_k, k \in \text{supp}(l_i)\}$  alors  $\langle x \rangle^\perp$  est d'ordre  $N$  minimal si et seulement si  $x \in C_{n_{\min}}(l_i)$  et alors*

$$N = \frac{1}{p^{n_{\min}}} \prod_{i=1}^d p^{n_i}.$$

*Dans le cas symétrique, i.e  $G \cong (Z/p^n)$ ,  $\langle x \rangle^\perp$  est d'ordre minimal si et seulement si*

$$x \in C_n(l_i) \text{ et alors}$$

$$N = q^{d-1} \text{ où } q = p^n.$$

**Démonstration :**  $x \in G$  est construit à l'étape 1, donc il existe  $i \in \{1, \dots, r\}$  et  $m \in \{1, \dots, n_{\min}\}$  tels que  $x \in C_m(l_i)$ . D'après le lemme 7.6,  $x$  est alors d'ordre  $p^m$ , donc pour que  $\langle x \rangle^\perp$  soit de taille minimale, il faut choisir  $x \in C_{n_{\min}}(l_i)$ . Dans le cas où  $G$  est symétrique, nous avons fait remarquer que nécessairement  $n_{\min} = n$ , d'où le résultat.

**Exemple 7.12 :** Dans cet exemple nous allons appliquer l'étape 1 ainsi que le théorème 7.10 afin de construire une fraction régulière de  $G$  de résolution 3, à partir de la matrice  $P$  de poids 3 à éléments dans  $GF(3)$ ,

$$P = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

Nous avons vu à l'étape 1 qu'il est possible de construire plusieurs matrices  $P'$  de poids 3 à partir de  $P$ . Nous allons regarder comment se comporte la fraction construite suivant différents choix pour  $P'$ . Celle-ci étant de résolution 3, nous allons nous intéresser aux représentations des  $J$ -marges où  $|J| = 2$ , c'est-à-dire aux projections en dimension 2 de la fraction.

### Cas symétrique

Appliquons l'étape 1 à  $P$  afin d'obtenir des matrices  $P'$  de poids 3 dont les lignes appartiennent à  $G = (Z/9)^3$ . Puisque  $n_1 = n_2 = n_3$ , nécessairement  $n_{\min} = 2$ .

Pour  $m=1$

	$Z/9$	$Z/9$	$Z/9$
$l_j$	1	1	1
$n_j - m$	1	1	1
$x_j$	3 (mod 9)	3 (mod 9)	3 (mod 9)

$$\Rightarrow C_1(l_1) = \{(3,3,3)\}$$

Pour  $m=2$

	$Z/9$	$Z/9$	$Z/9$
$l_j$	1	1	1
$n_j - m$	0	0	0
$x_j$	1 (mod 3)	1 (mod 3)	1 (mod 3)

$$\Rightarrow C_2(l_1) = \{x, x_i \in \{1,4,7\}\}$$

Supposons que l'on choisisse pour  $P'$  l'élément de  $C_1(l_1)$ ,  $P' = \begin{pmatrix} 3 & 3 & 3 \end{pmatrix}$ .

Pour  $i=1, \dots, 3$  les  $x_i$  sont d'ordre  $o_i=3$  et  $h_i=1$ , donc  $I = \{1\}$

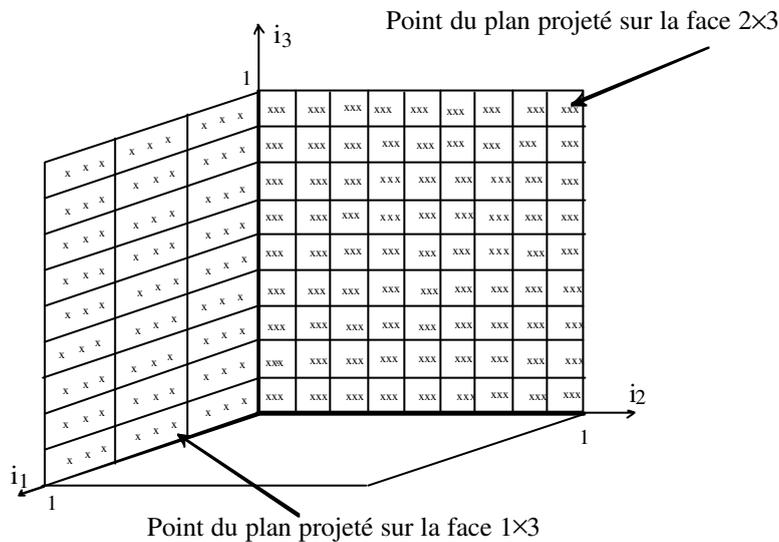
$$\begin{cases} y_1 = o_1 e_1 = (3,0,0) \\ y_2 = e_2 - e_1 = (-1,1,0) \\ y_3 = e_3 - e_1 = (-1,0,1) \end{cases}$$

D'après le théorème 7.10 la fraction régulière de  $(\mathbb{Z}/9)^3$  de matrice génératrice du dual  $P'$ , est somme directe des sous-groupes cycliques engendrés par  $y_1, y_2, y_3$ , c'est-à-dire qu'elle est engendrée par la matrice

$$B' = \begin{pmatrix} 3 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{matrix} \text{ordre} \\ 3 \\ 9 \\ 9 \end{matrix} \cdot$$

$\mathbb{Z}/9 \quad \mathbb{Z}/9 \quad \mathbb{Z}/9$

Cette fraction est de taille  $3 \times 81$ , ce qui n'est pas la taille minimale puisque  $x \in C_1(I_1)$  alors que  $n=2$ .  $x$  est donc uniquement d'ordre 3. La conséquence est que tous les couples de  $(\mathbb{Z}/9) \times (\mathbb{Z}/9)$  sont représentés 3 fois chacun. Cette répétition engendre un coût supplémentaire qui n'est pas nécessaire, même dans le cas d'un échantillonnage à deux degrés.



Supposons maintenant que  $P'$  soit un élément  $x$  de  $C_2(I_1)$ , par exemple

$$P' = (1 \ 1 \ 1).$$

Pour  $i=1, \dots, 3$  les  $x_i$  sont d'ordre  $o_i=9$  et  $h_i=1$ , donc  $I=\{1\}$

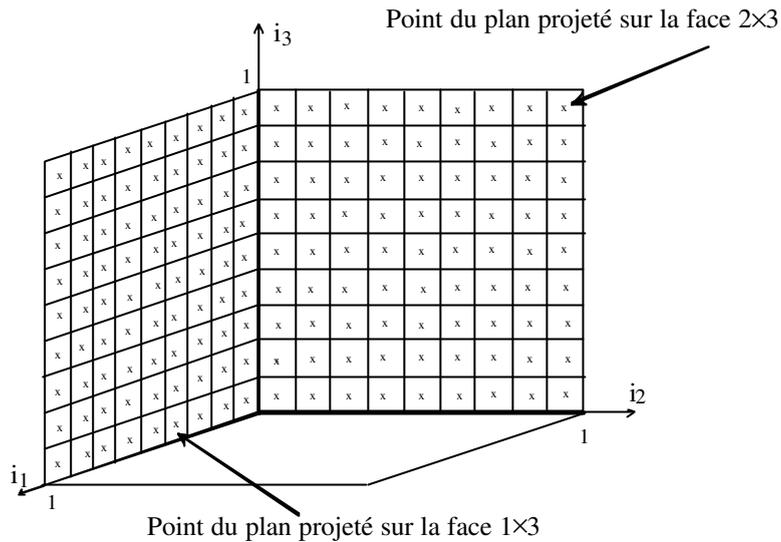
$$\begin{cases} y_1 = o_1 e_1 = (9,0,0) = 0 & \text{dans } G \\ y_2 = e_2 - e_1 = (-1,1,0) \\ y_3 = e_3 - e_1 = (-1,0,1) \end{cases}$$

D'après le théorème 7.10 la fraction régulière de  $(\mathbb{Z}/9)^3$  de matrice génératrice du dual  $P'$ , est somme directe des sous-groupes cycliques engendrés par  $y_1, y_2, y_3$ , c'est-à-dire qu'elle est engendrée par la matrice

$$B' = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{matrix} \text{ordre} \\ 9 \\ 9 \end{matrix} \cdot$$

$\mathbb{Z}/9 \quad \mathbb{Z}/9 \quad \mathbb{Z}/9$

Cette fois la fraction est bien de taille minimale  $N=81$ , puisque  $x \in C_n(I_1)$ , donc tous les couples de  $(\mathbb{Z}/9) \times (\mathbb{Z}/9)$  sont représentés une seule fois chacun.



### Cas non symétrique

Appliquons maintenant l'étape 1 afin d'obtenir une matrice  $P'$  de poids 3 dont les lignes cette fois appartiennent à  $G=(Z/3) \times (Z/9) \times (Z/9)$ . Dans ce cas  $n_{\min}=1$ , d'où

	$Z/3$	$Z/9$	$Z/9$
$l_{ij}$	1	1	1
$n_j - n_{\min}$	0	1	1
$x_j$	1 (mod 9)	3 (mod 9)	3 (mod 9)

Dans ce cas  $P'$  est unique et est formé de l'élément  $(1,3,3)$  de  $G$ .

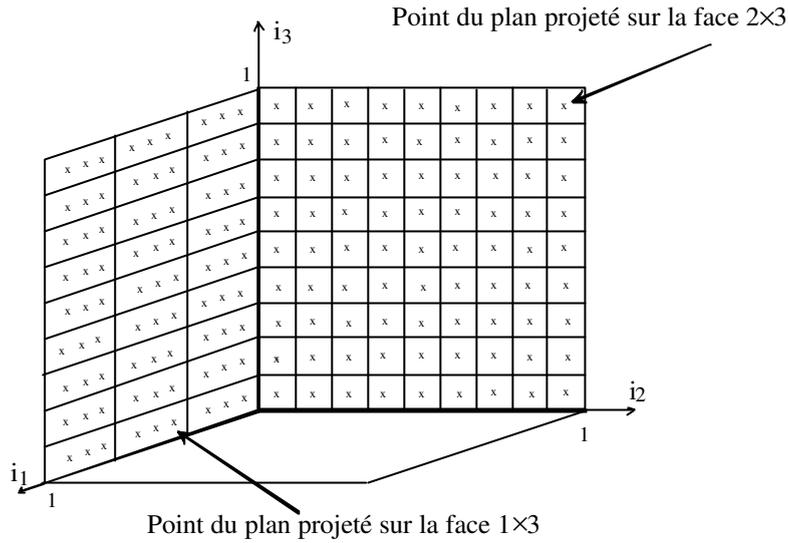
Pour  $i=1, \dots, 3$  les  $x_i$  sont d'ordre  $o_i=3$  et  $h_i=1$ , donc  $I=\{1\}$

$$\begin{cases} y_1 = o_1 e_1 = (3,0,0) = 0 & \text{dans } G \\ y_2 = e_2 - e_1 = (2,1,0) \\ y_3 = e_3 - e_1 = (2,0,1) \end{cases}$$

D'après le théorème 7.10 la fraction régulière de  $(Z/3) \times (Z/9) \times (Z/9)$  de matrice génératrice du dual  $P'$ , est somme directe des sous-groupes cycliques engendrés par  $y_1, y_2, y_3$ , c'est-à-dire qu'elle est engendrée par la matrice

$$B' = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{matrix} \text{ordre} \\ 9 \\ 9 \end{matrix} \cdot \begin{matrix} Z/3 & Z/9 & Z/9 \end{matrix}$$

Elle est de taille minimale 81, puisque  $x \in C_{n_{\min}}(1)$ . Donc tous les couples de  $(Z/9) \times (Z/9)$  sont bien représentés une seule fois chacun, mais puisque  $G$  n'est pas symétrique, les couples de  $(Z/3) \times (Z/9)$  sont ici représentés trois fois chacun.



### 7.2.2. Troisième étape : Construction de $\langle x_1, \dots, x_r \rangle^\perp$ dans $G$

L'étape 2 permet de construire une fraction régulière de  $G$  à partir de sa matrice génératrice du dual dans le cas où celle-ci est de rang 1. Nous allons maintenant utiliser l'étape 2 de façon itérative afin de généraliser la construction au cas où  $P'$  est de rang  $r$  quelconque.

Soient  $x_1, \dots, x_r$ , des éléments indépendants de  $G$ . On note  $n_{\max} = \max\{n_i, i=1, \dots, d\}$ , l'exposant de  $G$  est donc égal à  $q = p^{n_{\max}}$ , soit alors  $D_G$  la matrice  $\text{diag}(q/p^{n_i}, i=1, \dots, d)$ .

#### Théorème 7.13

Soit  $M_1$  la matrice dont les colonnes sont constituées des vecteurs générateurs de  $\langle x_1 \rangle^\perp$ . On définit par récurrence la suite de matrices

$$M_{k+1} = M_k B_{k+1}, \quad k=1, \dots, r-1,$$

où  $B_{k+1}$  est la matrice dont les colonnes sont les générateurs de  $\langle v_{k+1} \rangle^\perp$ .  $v_{k+1}$  est défini par  $v_{k+1} = M_k^* x_{k+1}$  avec  $M_k^* = D_k^{-1} M_k D_G$  et  $D_k = \text{diag}(q/s_j, j=1, \dots, n)$  où  $s_j$  est l'ordre de la  $j^{\text{ème}}$  colonne de  $M_k$ .

Alors  $\langle x_1, \dots, x_r \rangle^\perp$  est somme directe des sous-groupes cycliques engendrés par les colonnes de  $M_r$ .

Ce théorème a aussi été démontré par El Mossadeq *et al.* (1985).

Si  $x_1, \dots, x_r$  représentent les lignes d'une matrice  $P'$  de rang  $r$  et de poids  $R$ , alors  $\langle x_1, \dots, x_r \rangle^\perp$  est une fraction régulière de  $G$  de résolution  $R$ , engendrée par la matrice

$$B' = {}^t M_r.$$

L'intérêt de cette procédure de construction est que l'on connaît la matrice génératrice de la fraction mais aussi celle de son dual  $P'$ , ce qui permet d'ajuster une régression trigonométrique.

La taille de cette fraction est égale à  $N = |G| / |\langle x_1, \dots, x_r \rangle|$ . Si on note  $o_1, \dots, o_r$  les ordres respectifs de  $x_1, \dots, x_r$ , alors puisque  $\langle x_1, \dots, x_r \rangle$  est somme directe des sous-groupes cycliques engendrés par  $x_1, \dots, x_r$ , nous avons

$$N = \frac{\prod_{i=1}^r p^{n_i}}{\prod_{i=1}^r o_i}.$$

Là encore la fraction est de taille minimale si les  $x_i$ ,  $i=1, \dots, r$  sont tous d'un ordre aussi élevé que possible.

Soient  $P$  une matrice de rang  $r$  à éléments dans  $GF(p)$ , et  $l_1, \dots, l_r$  ses lignes. Par l'étape 1, on lui associe une matrice  $P'$  de rang  $r$  dont les lignes  $x_1, \dots, x_r$  appartiennent à  $G$ , telles que

$$x_i \in C(l_i), \quad i=1, \dots, r.$$

Soit  $B'$  la matrice obtenue en appliquant le théorème 7.13 à  $x_1, \dots, x_r$ . Alors la fraction régulière engendrée par  $B'$  est de taille minimale si et seulement si pour tout  $i=1, \dots, r$ ,  $x_i \in C_{n_{\min}}(l_i)$ , d'après le lemme 7.6. Dans le cadre de la simulation déterministe, nous nous intéressons plus particulièrement au cas où  $G$  est symétrique. Nous pouvons alors énoncer la proposition suivante.

**Proposition 7.14**

*Dans le cas symétrique, i.e  $G = \times(Z/p^n)$ , la fraction régulière engendrée par  $B'$  est de taille minimale si et seulement si  $x_i \in C_n(l_i)$   $i=1, \dots, r$  et alors*

$$N = q^{d-r} \text{ où } q=p^n.$$

Appliquons ce résultat aux exemples de la première étape.

**Exemple 7.2 (suite) :** Rappelons qu'à partir de la matrice  $P$  génératrice du dual d'une fraction régulière de  $GF(3)^4$  de résolution 3,

$$P = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},$$

on cherche à construire une fraction régulière  $S'$  de  $G=(Z/9)^4$  de résolution 3. A l'issue de l'étape 1, on dispose de  $(3^4+1)^2$  matrices  $P'$  de poids 3, génératrices du dual de  $S'$ . A l'aide du théorème 7.13, nous allons construire plusieurs fractions régulières de  $G$  afin de mettre en évidence l'importance du choix des lignes de  $P'$ . Puisque la résolution est 3, nous allons nous intéresser aux projections en dimension 2 de la fraction.

**Premier choix** : Supposons tout d'abord que les deux lignes de  $P'$  appartiennent aux ensembles  $C_1$ ,

$$P' = \begin{pmatrix} 3 & 0 & 6 & 3 \\ 0 & 3 & 3 & 3 \end{pmatrix}.$$

On cherche à construire  $\langle x_1, x_2 \rangle^\perp$  où  $x_1 = (3, 0, 6, 3)$  et  $x_2 = (0, 3, 3, 3)$ . Il nous faut tout d'abord déterminer  $\langle x_1 \rangle^\perp$  à l'aide de l'étape 2. Pour cela on permute les 2<sup>ème</sup> et 4<sup>ème</sup> composantes de  $x_1$ , de façon à ce que les ordres de chacune d'elles soient ordonnés par valeurs décroissantes

$$\tilde{x}_1 = (3, 6, 3, 0) .$$

ordre : 3 3 3 /

Par le théorème 7.10, on obtient

$$\begin{cases} \tilde{y}_{11} = 0_1 e_1 = (3, 0, 0, 0) \\ \tilde{y}_{12} = h_1 e_2 - h_2 e_1 = (-1, 1, 0, 0) \\ \tilde{y}_{13} = h_1 e_3 - h_3 e_1 = (-2, 0, 1, 0) \\ \tilde{y}_{14} = e_4 \end{cases} \Rightarrow \begin{cases} y_{11} = (3, 0, 0, 0) \\ y_{12} = (-1, 0, 0, 1) \\ y_{13} = (-2, 0, 1, 0) \\ y_{14} = (0, 1, 0, 0) \end{cases}.$$

D'où la matrice  $M_1$  avec l'ordre de chacune de ses colonnes

$$M_1 = \begin{pmatrix} \overset{Z/9}{3} & \overset{Z/9}{-1} & \overset{Z/9}{-2} & \overset{Z/9}{0} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

$s_j$     3    9    9    9

Par récurrence, on a  $M_2 = M_1 B_2$ . La matrice  $B_2$  est définie par les générateurs de  $v_2$ , avec

$$v_2 = D_1^{-1t} M_1 D_G x_2,$$

où  $D_G = I_4$  puisque  $G$  est symétrique et  $D_1 = \text{diag}(9/s_j, j=1, \dots, 4) = \text{diag}(3, 1, 1, 1)$ . Donc  $v_2 = (0, 3, 3, 3) \in G' = (Z/3) \times (Z/9) \times (Z/9) \times (Z/9)$ . On applique de nouveau l'étape 2 pour  $v_2$  dans  $G'$  et on trouve

$$B_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 3 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Finalement

$$B'^t M_2 = {}^t(M_1 B_2) = \begin{pmatrix} 0 & 3 & 0 & 0 \\ -1 & -1 & 0 & 1 \\ -2 & -1 & 1 & 0 \\ 3 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} \text{ordre} \\ 3 \\ 9 \\ 9 \\ 3 \end{matrix}$$

est la matrice génératrice d'une fraction régulière de  $(Z/9)^4$  de résolution 3, de taille  $9 \times 81$  et de matrice génératrice du dual  $P'$ . Tous les couples de  $(Z/9) \times (Z/9)$  sont représentés 9 fois chacun.

**Deuxième choix** : Supposons maintenant que la première ligne de  $P'$  appartienne à  $C_1$  et la deuxième à  $C_2$ . Soit par exemple

$$P' = \begin{pmatrix} 3 & 0 & 6 & 3 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

Les matrices  $M_1$ ,  $D_G$  et  $D_1$  sont les mêmes que précédemment d'où

$$v_2 = {}^t M_1 x_2 = (-3, 1, 1, 1) = (0, 1, 1, 1) \in G' = (Z/3) \times (Z/9) \times (Z/9) \times (Z/9).$$

En appliquant de nouveau l'étape 2 à  $v_2$  dans  $G'$ , on obtient

$$B_2 = \begin{pmatrix} 0 & 0 & 1 \\ -1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Donc

$$B' = {}^t M_2 = {}^t (M_1 B_2) = \begin{pmatrix} -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \\ 3 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} \text{ordre} \\ 9 \\ 9 \\ 3 \end{matrix}$$

est la matrice génératrice d'une fraction régulière de  $(Z/9)^4$  de résolution 3, de taille  $3 \times 81$  et de matrice génératrice du dual  $P'$ . Il n'y a plus que 3 répétitions des couples de  $(Z/9) \times (Z/9)$ .

**Troisième choix** : Supposons enfin que les deux lignes de  $P'$  appartiennent aux ensembles  $C_2$ . Soit par exemple

$$P' = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}.$$

En appliquant l'étape 2 à  $x_1$  on trouve

$$M_1 = \begin{pmatrix} -2 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

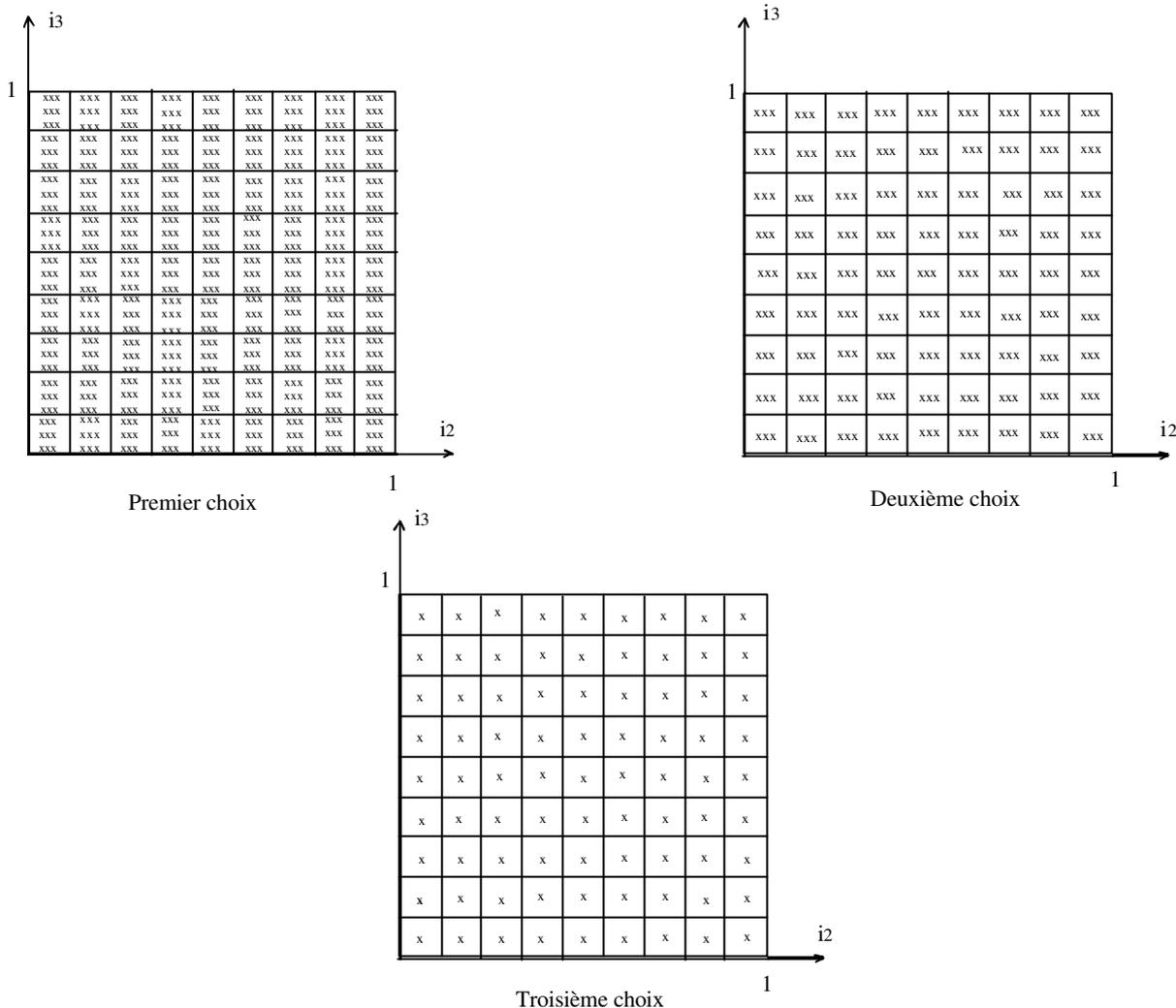
On a  $D_G = I_4$  et  $D_1 = I_3$  donc  $v_2 = {}^t M_1 x_2 = (1, 1, 1) \in (Z/9)^3$ . On applique de nouveau l'étape 2 à  $x_2$ , et on obtient

$$B_2 = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Finalement,

$$B' = {}^t M_2 = {}^t (M_1 B_2) = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 2 & 1 & -1 & 0 \end{pmatrix} \begin{matrix} \text{ordre} \\ 9 \\ 9 \end{matrix}$$

est la matrice génératrice d'une fraction régulière de  $(Z/9)^4$  de résolution 3, de taille minimale 81 et de matrice génératrice du dual  $P'$ . Tous les couples de  $(Z/9) \times (Z/9)$  sont représentés une seule fois chacun.



Points des fractions régulières construites projetés sur la face 2×3

**Exemple 7.3 (suite) :** Rappelons qu'à partir d'une matrice P génératrice du dual d'une fraction régulière de  $GF(2)^5$  de résolution 3,

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix},$$

on cherche à construire une fraction régulière S' de  $G=(Z/16) \times (Z/8) \times (Z/4) \times (Z/4) \times (Z/4)$  de résolution 3. A l'issue de la première étape, on dispose de  $(2^5+1)^2$  matrice P' de poids 3, génératrices du dual de S'. Nous allons choisir les deux lignes de P' dans les ensembles  $C_2$  de façon à ce que S' soit de taille minimale, par exemple

$$P' = \begin{pmatrix} 4 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 0 \end{pmatrix}.$$

A l'aide du théorème 7.13 on va construire  $\langle x_1, x_2 \rangle^\perp$  où  $x_1=(4,0,1,1,1)$  et  $x_2=(0,2,1,1,0)$ . Il nous faut tout d'abord déterminer  $\langle x_1 \rangle^\perp$  à l'aide de l'étape 1 et on obtient

$$M_1 = \begin{array}{cccc|cc} & & & & q_i & q/q_i \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} & & & & \begin{matrix} 16 \\ 8 \\ 4 \\ 4 \\ 4 \end{matrix} & \begin{matrix} 1 \\ 2 \\ 4 \\ 4 \\ 4 \end{matrix} \\ \hline s_j & 16 & 4 & 4 & 8 & \\ \hline q/s_j & 1 & 4 & 4 & 2 & \end{array}$$

$D_G = \text{diag}(1,2,4,4,4)$  et  $D_1 = \text{diag}(4,1,1,1,2)$  donc

$$v_2 = D_1^{-1} M_1 D_G x_2 = (4, 0, -1, 2) \in G' = (\mathbb{Z}/16) \times (\mathbb{Z}/4) \times (\mathbb{Z}/4) \times (\mathbb{Z}/8).$$

On applique de nouveau l'étape 2 pour déterminer  $\langle x_2 \rangle^\perp$  dans  $G'$  et on obtient

$$B_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

d'où

$$B' = {}^t M_2 = {}^t (M_1 B_2) = \begin{pmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 \end{pmatrix} \begin{array}{c} \text{ordre} \\ 16 \\ 8 \\ 4 \end{array}$$

On obtient la matrice génératrice d'une fraction régulière de  $(\mathbb{Z}/16) \times (\mathbb{Z}/8) \times (\mathbb{Z}/4) \times (\mathbb{Z}/4) \times (\mathbb{Z}/4)$  de résolution 3 et de taille  $4 \times 16 \times 8$ . Les couples de  $(\mathbb{Z}/16) \times (\mathbb{Z}/8)$  sont représentés 4 fois chacun.

### **Remarque dans le cas symétrique**

Dans le chapitre 5 (proposition 5.15), nous avons vu que si la matrice  $P$  est à éléments dans  $GF(p)$ , on peut l'écrire sous la forme  $P = (I_r \mid Q)$  et alors le noyau de  $P$ , *i.e* la matrice génératrice de la fraction régulière s'écrit  $(-{}^t Q \mid I_{d-r})$ . Zain et Rajan (1997), ont démontré que l'on peut généraliser ce résultat à une matrice  $P'$  à éléments dans un groupe cyclique d'ordre quelconque.

Cette généralisation peut s'appliquer ici dans le cas symétrique, *i.e* quand  $G$  est produit d'un même groupe cyclique d'ordre  $p^n$ , où  $p$  est premier. Dans l'exemple 7.2,  $G = (\mathbb{Z}/9)^4$  est symétrique et nous avons en troisième choix

$$P' = \left( \begin{array}{cc|cc} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{array} \right).$$

Elle s'écrit bien sous la forme  $P' = (I_r \mid Q)$ , donc par la méthode de Zain et Rajan, on obtient

$$B' = \left( \begin{array}{cc|cc} -2 & -1 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{array} \right).$$

A une combinaison des lignes près, cette matrice est bien celle que nous avons trouvée à l'issue de l'étape 3.

Cette technique est beaucoup plus simple à manipuler mais se restreint au cas où  $G$  est symétrique.

## Bibliographie

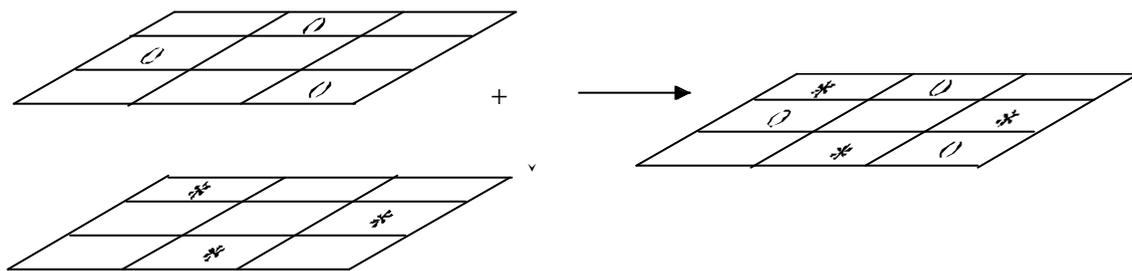
- Bailey R.A (1985). Factorial designs and abelian groups. *Lin. Algebra Appl.* **70**,349-363.
- El Mossadeq A. (1996). Fractions de résolution  $\omega$  et groupes abéliens finis. *Linear algebra Appl.* **237/238**, 539-547.
- El Mossadeq A., Kobilinsky A., Collombier D. (1985). Construction de tableaux orthogonaux dans les groupes abélien finis et confusions d'effets dans les plans factoriels. *Linear Algebra Appl.* **70**, 303-320.
- Zain A.A., Rajan B.S. (1997). Dual codes of systematic group codes over abelian groups. *AAECC* **8**, 71-83.



## 8. Fractions semi-régulières

Dans les chapitres précédents, nous avons vu que les plans construits à partir d'une structure de sous-groupes, présentent un avantage majeur pour l'ajustement d'une régression trigonométrique. En effet, dans ce cas, soit les colonnes de la matrice du modèle sont orthogonales, soit elles sont identiques deux à deux. Cela simplifie les calculs et confère des propriétés d'optimalité dans le cas d'un modèle linéaire simple. Cependant cette particularité peut devenir un inconvénient lorsqu'on souhaite analyser un grand nombre de fréquences. En effet, nous avons vu que pour qu'il y ait estimabilité des paramètres du modèle avec une fraction régulière, il faut que les colonnes soient mutuellement orthogonales. Cette contrainte est très forte et laisse peu de possibilités sur le choix des fréquences à analyser.

Dans le cas où le plan ne possède pas une structure de sous-groupe, l'orthogonalité n'est pas une condition nécessaire à l'estimabilité des paramètres du modèle. C'est pourquoi nous allons étudier dans ce chapitre des plans d'une autre nature que l'on appelle fractions semi-régulières ou bien en anglais 'parallel flats designs'. Ce nom est très significatif et représente bien la structure de ces plans qui sont construits en superposant un nombre fini de fractions régulières.



Superposition de 2 fractions régulières

Les fractions semi-régulières sont à éléments dans le groupe additif de  $GF(p^n)$  où  $p$  est un nombre premier. Ce qui correspond à un découpage symétrique du domaine d'échantillonnage.

Dans un premier temps, nous allons revenir sur les conditions d'estimabilité des paramètres dans le cas d'un modèle trigonométrique. Notamment, nous allons voir à l'aide d'un exemple, que l'orthogonalité n'est pas une condition nécessaire.

Ensuite nous nous intéressons aux propriétés de projection des fractions semi-régulières, c'est-à-dire à la répartition uniforme et complète des points du plan sur les marges de même dimension  $R-1$ . Certaines distributions marginales d'ordre  $R-1$  présentent réellement un intérêt, nous allons alors voir comment obtenir l'uniformité pour ces seules marges.

## 8.1. Définitions et généralités

Les fractions semi-régulières forment une classe importante des plans utilisés en planification d'expériences. Elles ont fait leur première apparition dans les travaux de Connor et Young (1959) et John (1961-62) pour le cas des facteurs à 2 ou 3 niveaux. Elles ont ensuite été étudiées par de nombreux auteurs, en particulier Srivastava (1965-84-87-90-96).

### 8.1.1. Définitions

#### Définition 8.1

*Soit  $q$  un nombre premier ou puissance d'un nombre premier. Soient  $P$  une matrice d'ordre  $r \times d$  et de rang  $r$ , à éléments dans le groupe additif de  $GF(q)$ , et  $c_k, k=1, \dots, f$  des vecteurs distincts d'ordre  $r \times 1$ . On note  $C$  la matrice d'ordre  $r \times f$  formée de ces vecteurs. On appelle fraction semi-régulière définie par le couple  $(P, C)$ , le sous-ensemble  $F$  de  $GF(q)^d$  formé des  $N=fq^{d-r}$  solutions des équations sur  $GF(q)$*

$$Px = c_k, k=1, \dots, f.$$

**Notation :** Soit une fraction semi-régulière définie par un couple  $(P, C)$ . On note  $T$  le tableau d'ordre  $N \times d$  dont les lignes sont les éléments de la fraction et on l'écrit sous la forme

$$T = \begin{pmatrix} T_1 \\ \dots \\ T_r \end{pmatrix},$$

où  $T_k$  est la matrice d'ordre  $q^{d-r} \times d$  dont les lignes sont les solutions de l'équation  $Px=c_k$ ,  $k=1, \dots, f$ .

Le lien avec les fractions régulières est immédiat puisque toute fraction semi-régulière  $F$  est une union finie de classes latérales d'un même sous-groupe  $S$  de  $GF(q)^d$ ,

$$F = \bigcup_{k=1}^f (z_k + S) = S + \Sigma$$

où  $\Sigma = \{z_1, \dots, z_f\} \subset GF(q)^d$ .

Donc une fraction semi-régulière est entièrement déterminée par le couple  $(S, \Sigma)$  où  $S$  s'appelle la **direction de la fraction** et  $\Sigma$  est l'**ensemble des solutions particulières**. Nous avons alors les relations suivantes entre les deux définitions

$$\begin{cases} S = \{s \in GF(q)^d, Ps = 0\} \\ c_k = Pz_k, \quad k = 1, \dots, f \end{cases}.$$

$P$  est donc la matrice génératrice du dual de  $S$ .

**Remarque :** Pour un vecteur  $c$  fixé, il existe  $q^{d-r}$  vecteurs  $z$  différents mais équivalents en ce qui concerne la définition de la fraction semi-régulière. C'est-à-dire que pour  $z_1$  tel que  $Pz_1=c$  et  $z_2 \neq z_1$  tel que  $Pz_2=c$ , on obtient la même classe latérale de  $S$ ,  $(z_1+S)=(z_2+S)$ . En effet pour  $x \in (z_1+S)$ , il existe  $s \in S$  tel que  $x=z_1+s=z_2+(z_1-z_2+s)$ , or  $P(z_1-z_2)=c-c=0$  donc  $z_1-z_2 \in S$ . Etant donné que  $S$  est un sous-groupe de  $G$ , on a  $s'=z_1-z_2+s$  qui appartient à  $S$ . Finalement,  $x$  peut s'écrire sous la forme  $x=z_2+s' \in (z_2+S)$ .

### 8.1.2. Construction de $T$ à partir de $(P,C)$

Supposons que la matrice  $P$  soit de la forme  $P=(I_r | Q)$ , où  $r$  est le rang de  $P$ . Nous avons vu dans les chapitres 5 (proposition 5.15) et 7 (dernière remarque) que dans  $GF(q)^d$  pour  $q$  premier ou puissance d'un nombre premier,  $S$  est engendré par les lignes de

$$B=(-{}^tQ | I_{d-r}).$$

Si on note  $T_0$  le tableau d'ordre  $q^{d-r} \times d$  dont les lignes sont les éléments de  $S$ , alors le tableau  $T$  de la fraction semi-régulière définie par le couple  $(P,C)$  est de la forme

$$T = \begin{pmatrix} T_0 + z_1 \\ \dots \\ T_0 + z_f \end{pmatrix},$$

où  $z_1, \dots, z_k$  sont des solutions particulières de  $Px=c_k$ ,  $k=1, \dots, f$ . Pour obtenir simplement ces solutions, il suffit de poser  $z_k=(z_{k,r} | z_{k,d-r})$  où  $z_{k,r} \in GF(q)^r$  et  $z_{k,d-r} \in GF(q)^{d-r}$ ,  $k=1, \dots, f$  et de choisir

$$\begin{cases} z_{k,z} = {}^t c_k \\ z_{k,d-r} = 0 \end{cases}.$$

**Exemple 8.2 :** Soient  $P$  et  $C$  les matrices d'ordre  $2 \times 3$  à éléments dans  $GF(3)$ ,

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix}.$$

Les lignes de  $T_0$  sont engendrées par le vecteur  $(2,2,1)$  et on choisit comme solution particulières  $z_1=0$ ,  $z_2=(1,2,0)$  et  $z_3=(1,1,0)$ . Alors  $T$  le tableau défini par  $(P,C)$  est égal à

$$T = \begin{pmatrix} 0 & 2 & 1 & | & 1 & 0 & 2 & | & 1 & 2 & 0 \\ 0 & 2 & 1 & | & 2 & 1 & 0 & | & 1 & 0 & 2 \\ 0 & 1 & 2 & | & 0 & 1 & 2 & | & 0 & 2 & 1 \end{pmatrix}.$$

$z_2$  est solution de l'équation  $Px=c_2$ , on aurait pu choisir comme autre solution particulière  $z'_2=(0,0,1)$ . On peut vérifier que  $(z_2+S)=(z'_2+S)$ , en effet

$$T_0 + z_2 = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} + (1 \ 1 \ 0) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 2 \end{pmatrix}$$

$$T_0 + z'_2 = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} + (0 \ 0 \ 1) = \begin{pmatrix} 0 & 0 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 0 \end{pmatrix}.$$

**Exemple 8.3 :** Soit P la matrice d'ordre  $2 \times 4$ , à éléments dans  $GF(9)$  de l'exemple 7.2, et soit la matrice C

$$P = \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 5 & 1 \\ 0 & 3 & 1 \end{pmatrix}.$$

D'après le chapitre précédent, les lignes de  $T_0$  sont engendrées par la matrice

$$B = \begin{pmatrix} -2 & -1 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{pmatrix}.$$

On choisit comme solutions particulières  $z_1=0$ ,  $z_2=(5,3,0,0)$  et  $z_3=(1,1,0,0)$ , alors T est d'ordre  $3 \times 81$  et est de la forme

$$T = \left( T_0 \mid T_0 + (5,3,0,0) \mid T_0 + (1,1,0,0) \right).$$

Pour le reste du chapitre nous allons exclure le cas où  $\Sigma$  est un sous-groupe de  $GF(q)^d$ . En effet, d'après le lemme qui suit, nous pouvons alors assimiler la fraction semi-régulière définie par le couple  $(S, \Sigma)$  à une fraction régulière de direction  $(S + \Sigma)$ .

**Lemme 8.4**

*Si S et  $\Sigma$  sont des sous-groupes de  $GF(q)^d$  alors  $S + \Sigma$  est un sous-groupe de  $GF(q)^d$  et*

$$(S + \Sigma)^\perp = S^\perp \cap \Sigma^\perp.$$

**Démonstration :**

1) Pour tout  $\varepsilon \in (S + \Sigma)$  et  $\varepsilon' \in (S + \Sigma)$ , il existe  $s \in S$ ,  $s' \in S$ ,  $z \in \Sigma$  et  $z' \in \Sigma$  tels que  $\varepsilon = s + z$  et  $\varepsilon' = s' + z'$ .  
Donc

$$\varepsilon + \varepsilon' = (s + s') + (z + z') = s'' + z'' \in (S + \Sigma),$$

puisque S et  $\Sigma$  sont des sous-groupes.

2) D'une part nous avons

$$\Sigma^\perp \cap S^\perp = \left\{ x \in GF(q)^d, \quad (x \mid s) = (x \mid z) = 0 \quad \forall s \in S, \forall z \in \Sigma \right\}$$

$$(S + \Sigma)^\perp = \left\{ x \in GF(q)^d, \quad (x \mid s + z) = 0 \quad \forall s \in S, \forall z \in \Sigma \right\}$$

donc de façon triviale on obtient,  $\Sigma^\perp \cap S^\perp \subset (S + \Sigma)^\perp$ . Et d'autre part on a

$$\left. \begin{array}{l} S \subset S + \Sigma \Rightarrow (S + \Sigma)^\perp \subset S^\perp \\ \Sigma \subset S + \Sigma \Rightarrow (S + \Sigma)^\perp \subset \Sigma^\perp \end{array} \right\} \Rightarrow (S + \Sigma)^\perp \subset \Sigma^\perp \cap S^\perp. \quad \square$$

## 8.2. Estimabilité

Le problème de l'estimabilité avec les fractions semi-régulières a été étudié notamment par les auteurs cités précédemment, mais uniquement dans le cas d'un modèle polynomial. Seuls Kobilinsky et Monod (1995) se sont penchés sur l'utilisation des fractions semi-régulières pour des modèles complexes.

Dans un premier temps, nous allons établir des conditions nécessaires et suffisantes pour qu'une fraction semi-régulière soit orthogonalement adaptée à un modèle trigonométrique. Ensuite nous verrons à l'aide d'un exemple que l'orthogonalité ( $Z^*Z=NI$ ) n'est pas une condition nécessaire à l'estimabilité des paramètres du modèle, contrairement aux fractions régulières.

### Théorème 8.5

*Soit  $F$  la fraction semi-régulière définie par le couple  $(S, \Sigma)$ , où  $S$  d'ordre  $q^{d-r}$  est la direction de la fraction et  $\Sigma = \{z_1, \dots, z_f\}$  est l'ensemble des solutions particulières. Soit un modèle défini par son ensemble de fréquences  $A^+$ . On note  $A^{++} = A^+ \cup \{h - m, h \in A^+, m \neq h \in A^+\} \cup \{h + m, h \in A^+, m \in A^+\}$ .*

- *Si  $f \neq \lambda q$  alors la fraction semi-régulière est orthogonale au modèle si et seulement si les fréquences de  $A^+$  sont mutuellement orthogonales par rapport à  $S$  (i.e  $\forall h \in A^{++}, h \notin S^\perp$ ).*
- *Si  $f = \lambda q$  alors la fraction semi-régulière est orthogonale au modèle si et seulement si pour tout  $h \in A^{++}$ , soit  $h \notin S^\perp$ , soit  $(h | z_k)_{k=1, \dots, f} = \pi(q)$  est une permutation de  $0, \dots, q-1$ .*

**Démonstration :** On note  $f_i, i=1, \dots, N$  un élément de la fraction semi-régulière  $F$ , alors  $\exists j \in \{1, \dots, q^{d-r}\}$  et  $\exists k \in \{1, \dots, f\}$  tels que  $f_i = s_j + z_k$ . On a alors

$$\sum_{i=1}^N \chi_h(f_i) = \sum_{j=1}^{q^{d-r}} \sum_{k=1}^f \chi_h(s_j + z_k) = \sum_{k=1}^f \chi_h(z_k) \times \sum_{j=1}^{q^{d-r}} \chi_h(s_j).$$

Dans le chapitre 4 nous avons vu que

$$\sum_{j=1}^{q^{d-r}} \chi_h(s_j) = 0 \text{ si et seulement si } h \notin S^\perp \text{ puisque } S \text{ est un sous-groupe}$$

$$\sum_{k=1}^f \chi_h(z_k) = 0 \text{ si et seulement si } (h | z_k)_{k=1, \dots, f} = \pi(q) \text{ est une permutation de } 0, \dots, q-1, \text{ ce}$$

qui implique que  $f$  est un multiple de  $q$

Le résultat est alors immédiat.  $\square$

Dans le cas où le nombre de solutions particulières  $f$  est un multiple de  $q$ , on remarque que l'on peut analyser plus de fréquences avec la fraction semi-régulière définie par le couple  $(S, \Sigma)$  qu'avec une fraction régulière de direction  $S$ , tout en conservant l'orthogonalité. Par exemple, si on choisit la matrice  $P$  de l'exemple 7.2 qui est d'ordre  $2 \times 4$ , à éléments dans  $GF(3)$ , génératrice du dual de  $S$

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}$$

alors la fréquence  $h=(1,0,1,1)$  et l'effet moyen ne peuvent pas être analysés à l'aide d'une fraction régulière de direction  $S$  puisque  $h \in S^\perp$ . En effet, dans ce cas on a

$$Z^*Z = \left( \begin{array}{c|cc|cc} 9 & 6+3w^2 & 6+3w & & & \\ & 9 & 6+3w^2 & & & \\ \hline & & & 9 & & \end{array} \right).$$

En revanche, si on pose  $z_1=0$ ,  $z_2=(1,0,0)$  et  $z_3=(0,2,0)$  alors  $(h \mid z_k)_{k=1,2,3}=(0,1,2)$ , donc la fraction semi-régulière définie par  $(S, \Sigma)$  est orthogonalement adaptée à l'analyse de la fréquence  $h$  et de l'effet moyen.

Dans le cas où  $f$  n'est pas un multiple de  $q$ , il n'est pas possible d'analyser plus de fréquences et pourtant la taille du plan est multipliée par  $f$ . Cependant, contrairement aux fractions régulières, l'orthogonalité n'est pas ici une condition nécessaire à l'estimabilité. En effet, il est possible d'analyser deux fréquences à l'aide d'une fraction semi-régulière sans que les deux colonnes correspondantes dans la matrice du modèle soient orthogonales. Prenons par exemple le cas de la matrice  $P$  d'ordre  $2 \times 3$  à éléments dans  $GF(3)$ ,

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Il est évident qu'une fraction régulière  $S$  de matrice génératrice du dual  $P$ , ne permet d'analyser qu'un seul effet simple en plus de l'effet moyen.

En revanche si on construit une fraction semi-régulière de direction  $S$  avec par exemple  $z_1=0$  et  $z_2=(1,0,0)$ , alors on peut analyser deux effets simples en plus de l'effet moyen, sans pour autant qu'il y ait orthogonalité. En effet, si on choisit les fréquences  $h_1=(1,0,0)$  et  $h_2=(0,0,1)$ , on obtient la matrice d'information

$$Z^*Z = \left( \begin{array}{c|ccc|cc} 6 & 0 & 0 & 0 & 0 & \\ \hline & 6 & 0 & 0 & -3w & \\ & & 6 & -3w & 0 & \\ \hline & & & 6 & 0 & \\ & & & & & 6 \end{array} \right)$$

qui est inversible. On peut donc analyser ces deux fréquences avec une fraction semi-régulière de taille 6 définie par le couple  $(S, \Sigma)$  où  $\Sigma=\{z_1, z_2\}$ , sans qu'il y ait orthogonalité. Si on ajoute à  $\Sigma$   $z_3=(0,1,0)$  alors on obtient la matrice d'information

$$Z^*Z = \left( \begin{array}{c|c|ccc} 9 & 0 & & & \\ \hline & & 0 & 0 & 6+3w^2 \\ \hline & 9I_3 & 0 & 0 & 6+3w^2 \\ \hline & & 6+3w^2 & 6+3w^2 & 0 \\ \hline & & & & 9I_3 \end{array} \right).$$

Elle aussi est inversible, on peut donc analyser les effets simples des trois facteurs avec cette fois une fraction semi-régulière de taille 9 sans qu'il y ait orthogonalité.

## 8.3. Propriétés de projection

Nous avons défini la résolution d'une fraction régulière de façon restrictive comme la répartition uniforme et complète des points du plan sur les marges de même dimension. Cette définition nous permet d'étendre la notion de résolution aux fractions semi-régulières. Le résultat de Bailey (1985) (proposition 5.5) qui consiste à déterminer la résolution de la fraction en fonction du poids de son dual ne peut plus s'appliquer ici puisque son champ d'application se limite aux sous-groupes. Il faut alors se référer aux travaux de Srivastava et Throop (1990) pour trouver une condition nécessaire et suffisante permettant d'obtenir la résolution d'une fraction semi-régulière.

Dans un premier temps, nous allons présenter le résultat de Srivastava et Throop et établir des conditions pour obtenir une fraction semi-régulière de résolution maximale. Ensuite nous allons proposer un théorème qui pour  $R$  fixé permet de déterminer quelles sont les marges de dimension  $R-1$  uniformément représentées par la fraction semi-régulière, sans que celle-ci soit pour autant de résolution  $R$ . Ce résultat peut s'avérer intéressant si l'on désire privilégier la répartition des points sur certains facteurs uniquement, ou bien lorsque la fraction semi-régulière de résolution  $R$  n'existe pas.

Srivastava et Throop ne s'intéressent aux fractions semi-régulières qu'au travers la construction de tableaux orthogonaux, ils se limitent donc à l'étude de l'orthogonalité. Avec le dernier paragraphe nous allons voir que l'intérêt des fractions semi-régulières réside dans le fait qu'elles limitent le coût des plans au prix de défauts limités et parfaitement contrôlés d'orthogonalité.

### 8.3.1. Répartition uniforme des marges de même dimension : résolution

#### Définition 8.6

*Soient la fraction semi-régulière définie par le couple  $(P,C)$  et  $T$  le tableau d'ordre  $(fq^{d-r}) \times d$  correspondant. On dit que la fraction semi-régulière est de résolution  $R$  si le tableau  $T$  est orthogonal de force  $R-1$ .*

Le tableau  $T$  est orthogonal mais pas linéaire, seul  $T_0$  est un tableau orthogonal linéaire.

Le théorème suivant est de Srivastava et Throop (1990), il donne une condition nécessaire et suffisante pour qu'une fraction semi-régulière soit de résolution  $R$ .

### Théorème 8.7

*Soient la fraction semi-régulière  $F$  définie par le couple  $(P,C)$  et  $T$  le tableau d'ordre  $(fq^{d-r}) \times d$  correspondant. Alors  $T$  est de force  $R-1$  (c'est-à-dire  $\varepsilon$  est de résolution  $R$ ) si et seulement si*

*pour tout vecteur  $y$  de  $GF(q)^r$  tel que  $yP \neq 0$  et  $wt(yP) \leq R-1$ , le vecteur d'ordre  $f$ ,  ${}^t(yC)$ , est un tableau orthogonal de force 1.*

Deux cas se présentent alors.

Soit  $wt(P) \geq R$ , alors la condition du théorème est automatiquement vérifiée. De plus comme nous l'avons vu dans le chapitre 6 sur la construction des tableaux orthogonaux linéaires,  $T_0$  est un tableau de force  $R-1$  (d'index 1) et  $T$  est un tableau de force  $R-1$  d'index  $f$ . Donc tous les éléments de  $GF(q)^{R-1}$  sont représentés  $f$  fois chacun dans tout bloc de  $R-1$  colonnes de  $T$ .

Soit il existe  $y$  tel que  $0 < wt(yP) \leq R-1$ , dans ce cas  $(yC)$  doit être un tableau de force 1, c'est-à-dire  $(yC)$  doit être égal à une ou des permutations de  $0, \dots, q-1$ . Or le vecteur  $(yC)$  est d'ordre  $f$ , nous sommes donc obligés de supposer que  $f$  est un multiple de  $q$ . Etant donné que  $T$  est de taille  $fq^{d-r}$ , nous choisirons  $f=q$  pour garder une taille raisonnable. La condition du théorème s'écrit alors

$$\forall \alpha \in GF(q), wt(\alpha {}^t(yC)) = f-1 = q-1.$$

Dans la plus part des cas nous choisirons  $q=3$ . Ce choix paraît approprié si l'on veut conserver une fraction semi-régulière de petite taille bien qu'alors le découpage du domaine d'échantillonnage soit assez grossier.

**Exemple 8.8 :** Soit  $P$  la matrice d'ordre  $2 \times 7$ , de rang 2, à éléments dans  $GF(5)$ ,

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & 3 & 4 \end{pmatrix}.$$

On cherche la matrice  $C$  d'ordre  $2 \times 5$  telle que la fraction semi-régulière définie par le couple  $(P,C)$ , i.e formée des solutions de  $Px=C$ , soit de résolution 6.

Seul le vecteur  $y=(0,1)$  est tel que  $wt(yP) \leq 5$ , il faut donc choisir la matrice  $C$  avec pour deuxième ligne une permutation de  $0, \dots, 4$  et un première ligne quelconque, par exemple

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}$$

On évitera cependant de choisir une première ligne avec des éléments tous égaux car alors  $\Sigma$  est un sous-groupe et la fraction est régulière. En effet,  $C$  est une partie de  $GF(5)^2$  et dans ce cas, tout ensemble de 5 éléments dont la 1<sup>ère</sup> coordonnées est nulle constitue un sous-groupe cyclique.

Etant donné que nous avons choisi  $f=q$ , la fraction semi-régulière cherchée est de taille  $q^{d-r+1}$ . Soit  $n=d-r$ , au mieux s'agissant de la résolution, nous pouvons donc espérer avoir un

tableau orthogonal de force  $n+1$  (et d'index 1) et donc une fraction semi-régulière de résolution  $n+2$ . Supposons les matrices  $P$  et  $C$  de la forme

$$P=(I_r \mid Q) \text{ et } C=(c_1 \mid \dots \mid c_q)$$

où  $c_i \in GF(q)^r$  et  $c_{ik} \neq c_{jk}$   $i \neq j=1, \dots, q$  et  $k=1, \dots, r$  (c'est-à-dire les lignes de  $C$  sont des permutations de  $0, \dots, q-1$ ). Nous allons voir comment choisir la matrice  $Q$  et les vecteurs  $c_i$ ,  $i=1, \dots, q$  de façon à ce que le couple  $(P, C)$  définisse une fraction semi-régulière de résolution  $n+2$ .

**Notations :** Soient  $M$  une matrice d'ordre  $n \times n$ ,  $v$  un vecteur d'ordre  $m$ . Pour  $I \subseteq \{1, \dots, m\}$ , et  $J \subseteq \{1, \dots, n\}$ , on note  $M(I, J)$  la sous-matrice de  $M$  d'ordre  $|I| \times |J|$  formée des lignes indexées par  $I$  et des colonnes indexées par  $J$ ,  $M(., J)$  désigne le bloc formé des colonnes de  $M$  indexées par  $J$ , et  $v(I)$  le vecteur d'ordre  $|I|$  formé des éléments de  $v$  indexés par  $I$ .

### Théorème 8.9

Soient les matrices  $P$  d'ordre  $r \times d$  et  $C$  d'ordre  $r \times q$  définies ci-dessus à éléments dans  $GF(p)$ . On note  $C^*$  l'ensemble des vecteurs d'ordre  $r \times 1$  défini par

$$C^* = \{c_i - c_j, 1 \leq i < j \leq q\}.$$

La fraction semi-régulière définie par le couple  $(P, C)$  est de force  $n+1$  si et seulement si pour tout  $m=2, \dots, \min(r, n+1)$ ,  $\forall U_m \subseteq \{1, \dots, r\} : |U_m| = m$ ,  $\forall W_{m-1} \subseteq \{1, \dots, d-r\} ; |W_{m-1}| = m-1$

$$1) \operatorname{rg}(Q(U_m, W_{m-1})) = m - 1$$

$$2) \det \left( [Q(U_m, W_{m-1}) \mid a(U_m)] \right) \neq 0, \quad \forall a \in C^*.$$

**Exemple 8.10 :** Soient  $P$  la matrice d'ordre  $2 \times 4$  de rang 2, à éléments dans  $GF(3)$  et  $C$  la matrice d'ordre  $2 \times 3$  suivantes

$$P = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \Rightarrow C^* = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Dans ce cas  $\min(r, n+1)=2$  et pour  $m=2$ ,  $U_2=\{1, 2\}$ .

Si  $W_1=\{1\}$ ,  $Q(U_2, W_1)=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ , si  $W_1=\{2\}$ ,  $Q(U_2, W_1)=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ . Dans les deux cas on peut constater que, tout vecteur de  $C^*$  est indépendant de  $Q(U_2, W_1)$ . Donc le couple  $(P, C)$  définit une fraction semi-régulière de taille 27 et de résolution maximale 4.

**Exemple 8.11 :** Soient  $P$  la matrice d'ordre  $4 \times 6$ , de rang 4, à éléments dans  $GF(5)$  et  $C$  la matrice d'ordre  $4 \times 5$  suivantes

$$P = \left( I_4 \mid \begin{array}{cc} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{array} \right) \text{ et } C = \begin{pmatrix} 0 & 2 & 4 & 1 & 3 \\ 0 & 3 & 1 & 4 & 2 \\ 0 & 3 & 1 & 4 & 2 \\ 1 & 0 & 4 & 3 & 2 \end{pmatrix} \Rightarrow C^* = \begin{pmatrix} 3 & 1 & 4 & 2 \\ 2 & 4 & 1 & 3 \\ 2 & 4 & 1 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Dans ce cas  $n+1=3$  et on peut vérifier que pour  $m=2,3$  les conditions 1) et 2) du théorème sont satisfaites. Donc le couple  $(P,C)$  définit une fraction semi-régulière de taille  $5^3$  et de résolution maximale 3.

**Exemple 8.12 :** Soit  $P$  la matrice d'ordre  $3 \times 5$ , de rang 3, à éléments dans  $GF(3)$

$$P = \left( I_3 \mid \begin{array}{cc} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{array} \right).$$

Dans ce cas,  $n+1=3$  et pour  $m=2$  les conditions du théorème imposent

$$\begin{pmatrix} a_i \\ a_j \end{pmatrix} \neq \alpha \mathbf{1}, 1 \leq i < j \leq 3 \text{ et } \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} \neq \beta \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ où } \alpha, \beta \in GF(3) \text{ et } a = (a_1, a_2, a_3) \in C^*.$$

On peut vérifier qu'il n'y a pas de solution à ce problème, il n'existe pas de fraction semi-régulière de  $GF(3)^5$  de résolution 4.

Ainsi l'exemple précédent prouve que pour  $d$  et  $N$  fixés, on ne peut pas toujours trouver un couple  $(P,C)$  définissant une fraction semi-régulière de résolution  $n+2$ , notamment dans le cas où  $q=3$ .

Dans le cas où  $r$  est grand mais  $\min(r, d-r+1) = d-r+1$  est petit (2 ou 3), les conditions du théorème sont plus faciles à vérifier que les conditions du théorème de Srivastava et Throop (1990). En effet, pour  $r$  grand il est difficile d'envisager toutes les combinaisons linéaires des lignes de  $P$ .

**Démonstration :** On note  $T_0$  (resp.  $T_i, i=1, \dots, q$ ) le tableau d'ordre  $q^n \times d$  dont les lignes représentent les solutions de l'équation  $Pt=0$  (resp.  $Pt=c_i, i=1, \dots, q$ ). Soit  $T$  le tableau dont les lignes sont les éléments de la fraction semi-régulière. Par définition  $T$  est orthogonal de force  $n+1$  et d'index 1, si et seulement si dans tout bloc de  $n+1$  colonnes de  $T$ , les éléments de  $GF(q)^{n+1}$  sont représentés une seule fois chacun. Ce qui peut encore s'écrire,  $T$  est un tableau orthogonal de force  $n+1$  si et seulement si pour tout  $K \subseteq \{1, \dots, d\}; |K|=n+1$

- 1) toutes les lignes de  $T_i(.,K)$  sont distinctes pour  $1 \leq i \leq q$ ,
- 2) les lignes de  $T_i(.,K)$  sont distinctes des lignes de  $T_j(.,K)$  pour tout  $0 \leq i < j \leq q$ .

1) Montrons tout d'abord que la condition 1) du théorème est nécessaire et suffisante pour que les  $q^n$  lignes de  $T_0(.,K)$  soient distinctes.

Soit  $y \in GF(q)^d$  une ligne de  $T_0$ , alors  $y$  est de la forme

$$y = x(-{}^tQ \mid I_n)$$

avec  $x = (x_{r+1}, \dots, x_d) \in GF(q)^n$  puisqu'ici  $n=d-r$  est le nombre de colonnes de  $Q$ . Nous remarquons que toute ligne de  $T_0$  s'écrit en 2 parties, l'une engendrée par  $-{}^tQ$  et l'autre par  $I_n$ . Nous allons alors décomposer  $K$  en 2 sous-ensembles d'index correspondants à chacune des ces parties.

Soit  $m \in \{1, \dots, \min(r, n+1)\}$ , supposons que parmi les  $n+1$  colonnes indexées par  $K$ , on choisisse  $m$  correspondantes aux lignes de  $T_0$  engendrées par  $-{}^tQ$ . On note  $U_m \subseteq \{1, \dots, r\}; |U_m|=m$ , l'ensemble des indices de ces colonnes. Pour compléter l'ensemble  $K$ , il reste alors à

choisir  $k=n+1-m$  colonnes de  $T_0$  parmi celles qui correspondent aux lignes engendrées par  $I_n$ . On note  $J_k \subseteq \{1, \dots, d-r\}$ ;  $|J_k|=k$ , l'ensemble des indices de ces colonnes. Finalement, on a

$$K = U_m \cup J_k, \text{ où } n+1=m+k \text{ et}$$

$$y = x \left( \begin{array}{c|c} \text{---}^t Q \text{---} & \text{---} I_{d-r} \text{---} \\ \hline U_m & J_k \end{array} \right)$$

De plus on note  $\bar{U}_m = \{1, \dots, r\} \setminus U_m$  et  $\bar{J}_k = \{1, \dots, d-r\} \setminus J_k = W_{m-1}$ .

Afin de simplifier les notations et sans perte de généralité supposons que

$$U_m = \{1, \dots, m\}$$

et  $J_k = \{1, \dots, k\}$ .

Alors la matrice  ${}^t Q$  s'écrit en blocs sous la forme suivante

$${}^t Q = \left( \begin{array}{c|c} {}^t Q(J_k, U_m) & {}^t Q(J_k, \bar{U}_m) \\ \hline {}^t Q(\bar{J}_k, U_m) & {}^t Q(\bar{J}_k, \bar{U}_m) \end{array} \right).$$

Or les lignes de  $T_0(., K)$  sont identiques si et seulement si il existe une combinaison linéaire non nulle,  $x \in GF(q)^n$ , telle que

$$(x(-{}^t Q \mid I_n))(K) = 0.$$

$$\begin{aligned} x(-{}^t Q \mid I_n) &= (x(J_k) \mid x(\bar{J}_k)) \left( \begin{array}{c|c|c|c} -{}^t Q(J_k, U_m) & -{}^t Q(J_k, \bar{U}_m) & I_k & 0 \\ \hline -{}^t Q(\bar{J}_k, U_m) & -{}^t Q(\bar{J}_k, \bar{U}_m) & 0 & I_{n-k} \end{array} \right) \\ &= (-x(J_k){}^t Q(J_k, U_m) - x(\bar{J}_k){}^t Q(\bar{J}_k, U_m) \mid \dots \mid x(J_k) \mid x(\bar{J}_k)) \end{aligned}$$

$$\text{Donc } (x(-{}^t Q \mid I_n))(K) = 0 \Leftrightarrow \begin{cases} x(J_k) = 0 \\ x(\bar{J}_k){}^t Q(\bar{J}_k, U_m) = 0 \end{cases}$$

Si  $m=1$   $x(J_k)=x=0$ . Sinon, la matrice  ${}^t Q(\bar{J}_k, U_m)$  est d'ordre  $(m-1) \times m$ , donc ce système admet une solution non nulle si et seulement si  $\text{rg}({}^t Q(\bar{J}_k, U_m)) < m-1$ . Donc en notant  $W_{m-1} = \bar{J}_{k-1}$  toutes les lignes de  $T_0(., K)$  sont distinctes si et seulement si la condition 1) est vérifiée. Il en va de même pour les lignes de  $T_i(., K)$  puisque  $T_i = T_0 + z_i$ ,  $i=1, \dots, q$ .

2) Montrons maintenant que la condition 2) du théorème est nécessaire et suffisante pour que toute ligne  $T_i(., K)$  soit distinctes de toute ligne de  $T_j(., K) \forall j > i$ . Pour  $i=1, \dots, q$ ,  $T_i = T_0 + z_i$ , et comme nous l'avons vu plus haut, on peut toujours choisir  $z_i$  égal à  $({}^t c_i \mid 0)$ , donc toute ligne  $y_i$  de  $T_i$  s'écrit

$$y_i = ({}^t c_i \mid -x {}^t Q \mid x I_n) \text{ où } x \in GF(q)^n.$$

Soit de plus  $y_j$  une ligne de  $T_j$ ,  $y_j = ({}^t c_j \mid -x' {}^t Q \mid x' I_n)$ ,  $x' \in GF(q)^n$ . Alors  $y_i = y_j \Leftrightarrow$

$$\begin{aligned} &({}^t c_i(U_m) - x(J_k){}^t Q(J_k, U_m) - x(\bar{J}_k){}^t Q(\bar{J}_k, U_m) \mid \dots \mid x(J_k) \mid x(\bar{J}_k)) \\ &= ({}^t c_j(U_m) - x'(J_k){}^t Q(J_k, U_m) - x'(\bar{J}_k){}^t Q(\bar{J}_k, U_m) \mid \dots \mid x'(J_k) \mid x'(\bar{J}_k)) \end{aligned}$$

$$\text{Donc } y_i(I)=y_j(I) \Leftrightarrow \begin{cases} (x'-x)(J_k) = 0 \\ (x'-x)(\bar{J}_k)'Q(\bar{J}_k, U_m) = {}^t(c_i - c_j)(U_m) \end{cases}$$

Si  $m=1$  ce système n'admet pas de solution puisque  $c_{ik} \neq c_{jk}$   $k=1, \dots, r$ . Sinon le système admet une solution si et seulement si

$$\det({}^tQ(\bar{J}_k, U_m) \mid a(U_m)) = \det(Q(U_m, \bar{J}_k) \mid a(U_m)) = 0 \text{ où } a=c_i-c_j \in C^*. \quad \square$$

### 8.3.2. Répartition uniforme sur certaines marges de dimension R

Il n'existe pas toujours de fraction semi-régulière de résolution R désirée pour  $q, r$  et  $d$  fixés, ou bien, quand elles existent, elles engendrent des plans souvent coûteux. Or certaines distributions marginales d'ordre  $R-1$  présentent réellement un intérêt, il est donc souhaitable d'obtenir l'uniformité sur ces seules marges. Nous allons voir comment étendre le résultat de Srivastava et Throop pour établir sous quelles conditions une marge de dimension  $R-1$  est complète et uniforme sans que la fraction semi-régulière soit de résolution R

Soit F la fraction semi-régulière de  $GF(q)^d$  définie par le couple  $(P, C)$  où P est une matrice d'ordre  $r \times d$ , de rang r et  $C=(c_0 \mid \dots \mid c_{q-1})$  est une matrice d'ordre  $r \times q$ . D'où  $f=q$ , i.e  $N=q^{n+1}$  où  $n=d-r$ .

Soient t un entier inférieur ou égal à  $n+1$  et  $J_t \subseteq \{1, \dots, d\}$  tel que  $|J_t|=t$ .

#### Définition 8.13

*On dit que le couple  $(P, C)$  vérifie la propriété  $R(J_t)$  si et seulement si pour tout vecteur non nul  $y$  d'ordre  $1 \times r$*

$$\text{supp}(yP) \subseteq J_t \Rightarrow (yC) \text{ est une permutation de } 0, \dots, q-1.$$

#### Théorème 8.14

*On suppose que le rang de la matrice P, r, est inférieur ou égal à q.*

*Le couple  $(P, C)$  vérifie la propriété  $R(J_t)$  si et seulement si les éléments de  $GF(q)^t$  sont représentés  $q^{n+1-t}$  fois chacun dans  $T(., J_t)$ .*

Ce résultat est une extension directe du théorème de Srivastava et Throop.

**Exemple 8.12 (suite) :** Les matrices P d'ordre  $5 \times 3$  et C d'ordre  $3 \times 3$  à éléments dans  $GF(3)$  sont définies par

$$P = \left( \begin{array}{c|cc} & 1 & 1 \\ \hline I_3 & 1 & 1 \\ & 1 & 0 \end{array} \right) \text{ et } C = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{pmatrix}.$$

Nous savons d'après l'exemple 3, que le couple  $(P, C)$  ne définit pas une fraction régulière de résolution maximale 3. Cependant toutes les marges de dimension 3 sont représentées

uniformément (sans répétition) sauf la marge définie par  $J_3=\{2,3,5\}$ . En effet, pour  $y=(0,1,2)$ , nous avons  $\text{supp}(yP)=J_3$  et  $yC=0$ , donc le couple  $(P,C)$  ne vérifie pas  $R(J_3)$ .

Cette fraction, même si elle n'est pas de résolution maximale, en est très proche, et garde une très bonne répartition de ses points.

**Exemple 8.15 :** Dans cet exemple nous allons voir que le choix de la matrice  $P$  est important. Soient les matrices à éléments dans  $GF(3)$

$$P = \left( I_3 \left| \begin{array}{cc} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{array} \right. \right) \text{ et } C = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 1 & 2 & 0 \end{pmatrix}.$$

Alors le couple  $(P,C)$  ne vérifie pas  $R(J_3)$  pour  $J_3=\{1,2,5\}$ , car pour  $y=(1,2,0)$   $\text{supp}(yP)=J_3$  et  $yC=1$ , et pour  $J_3=\{1,3,4\}$ , car pour  $y=(1,0,1)$   $\text{supp}(yP)=J_3$  et  $yC=2 \times 1$ . On ne peut pas trouver de meilleure matrice  $C$ , c'est-à-dire qu'il y a au minimum deux marges de dimension trois qui ne sont pas représentées uniformément, ce qui est moins bien que dans l'exemple ci-dessus.

**Exemple 8.16 :** Soient  $P$  la matrice d'ordre  $3 \times 6$ , de rang 3 à éléments dans  $GF(3)$  et la matrice  $C$  d'ordre  $3 \times 3$

$$P = \left( I_3 \left| \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right. \right) \text{ et } C = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 0 & 2 & 1 \end{pmatrix}.$$

Nous avons  $n+1=4$ , le couple  $(P,C)$  ne vérifie pas la relation  $R(J_4)$  uniquement pour  $J_4=\{1,2,5,6\}$ . En effet pour  $y=(1,2,0)$ , nous avons  $\text{supp}(yP)=J_4$  et  $yC=2 \times 1$ . Donc une seule marge fait défaut pour avoir la résolution 5.

### Démonstration :

1) Dans un premier temps montrons l'implication : si les éléments de  $GF(q)^t$  sont représentés une et une seule fois dans  $T(.,J_t)$  alors le couple  $(P,C)$  vérifie la propriété  $R(J_t)$ .

On suppose que pour  $y$  d'ordre  $1 \times r$ ,  $\text{supp}(yP)=J_t$ , ce qui entraîne que pour  $x \in GF(q)^d$

$$(1) \quad yPx = \sum_{j \in J_t} \theta_j x_j \quad \text{où } \theta_j \neq 0 \in GF(q).$$

Pour  $\alpha=0, \dots, q-1$ , on considère les équations (à  $t$  inconnues)  $yPx=\alpha$ . Chaque équation possède  $q^{t-1}$  solutions et d'après (1) chaque solution a son support égal à  $J_t$ . Soit  $M_\alpha$  la matrice d'ordre  $q^{t-1} \times t$  dont les lignes sont formées des solutions des équations  $yPx=\alpha$  restreintes à leur support. On remarque alors la relation suivante. On note  $T_k$  ( $k=0, \dots, q-1$ ) le tableau dont les lignes sont les solutions de l'équation  $Px=c_k$ .

Toute ligne de  $T_k(.,J_t)$  est une ligne de  $M_\alpha$  si et seulement si  $yc_k=\alpha$ .

En effet, si  $y$  est une ligne de  $T_k$  alors  $yPx=yc_k$  donc  $x(J_t)$  est une ligne de  $M_\alpha$  si et seulement si  $\alpha=yc_k$ .

$(yC)=(yc_0, \dots, yc_{q-1})$ , on note  $\mu_k$ ,  $k=0, \dots, q-1$  le nombre de fois où  $yc_k$  apparaît dans  $(yC)$  (remarque :  $yc_k=yc_k'$  n'entraîne pas  $q_k=c_k'$ ). Il existe donc  $\mu_k$  valeurs de  $\epsilon \in \{0, \dots, q-1\}$  pour

lesquelles les lignes de  $T_i$  (qui vérifient  $Px=c_i$ ) vérifient  $yPx=yc_k$ , c'est-à-dire pour lesquelles les vecteurs de  $T_i(.,J_t)$  sont aussi des vecteurs de  $M_{yck}$ . Ce qui veut dire que  $T(.,J_t)$  possède exactement  $\mu_k q^n$  lignes de  $M_{yck}$  (puisque  $T$  est la superposition verticale des  $T_k$ ). Or par hypothèse, tous les éléments de  $GF(q)^t$  apparaissent  $q^{n+1-t}$  fois chacun dans  $T(.,J_t)$ , donc celui-ci contient  $q^{n+1-t} q^{t-1} = q^n$  lignes de  $M_{yck}$  pour  $k=0, \dots, q-1$ . D'où

$$q^n = \mu_0 q^n = \dots = \mu_{q-1} q^n \Leftrightarrow \mu_0 = \dots = \mu_{q-1} = 1.$$

Les  $q$  éléments de  $(yC)$  apparaissent une seule fois chacun, donc  $(yC)$  est une permutation de  $0, \dots, q-1$ .

2) Montrons maintenant la réciproque. Nous savons par hypothèse que :  $\forall y \neq 0 \in GF(q)^r$  ;  $\text{supp}(yP) \subseteq J_t \Rightarrow (yC) = \pi_y(q)$  où  $\pi_y(q)$  est une permutation de  $0, \dots, q-1$ . Et nous voulons montrer que tous les éléments de  $GF(q)^t$  sont représentés  $q^{n+1-t}$  fois chacun dans  $T(.,J_t)$ .

Nous pouvons sans perte de généralité, supposer que  $I = \{1, \dots, t\}$ . En effet, si on note  $P^*$  (resp.  $x^*$ ) la matrice (resp. le vecteur) obtenue en permutant les colonnes (resp. les éléments) de  $P$  (resp. de  $x$ ) indexées par  $I$  en première position, alors  $Px = P^* x^*$ . Donc si  $T^*$  désigne le tableau défini par le couple  $(P^*, C)$ , les blocs de  $T^*(., \{1, \dots, t\})$  et  $T(., J_t)$  sont égaux à une permutation près.

Soit  $s \in GF(q)^t$ , on note  $\lambda$  le nombre de fois où  $s$  est une ligne de  $T(., J_t)$ .  $\lambda$  est en fait égal au nombre de solutions des équations

$$(*) \quad Px' = c_k, \quad k=0, \dots, q-1,$$

avec  $x' = {}^t[s \mid x]$  où  $x \in GF(q)^{d-t}$ .

Si on écrit  $P = [P_1 \mid P_2]$  où  $P_1$  est une matrice d'ordre  $r \times t$  et  $P_2$  est une matrice d'ordre  $r \times (d-t)$ , alors (\*) est équivalent à

$$P_2 x = (c_k - P_1 s), \quad k=0, \dots, q-1.$$

Si  $\text{rg}(P_2) = r$ , alors chaque équation possède  $q^{d-t+r}$  solutions, donc  $\lambda = q^{n+1-t}$ .

Si  $\text{rg}(P_2) = r-1$ . On décompose  $P$  de la façon suivante

$$P = \left[ \begin{array}{c|c} P_{11} & P_{12} \\ \hline p_1 & p_2 \end{array} \right]$$

avec  $P_{11}$  matrice d'ordre  $(r-1) \times t$ ,  $P_{12}$  matrice d'ordre  $(r-1) \times (d-t)$ ,  $p_1$  vecteur d'ordre  $1 \times t$  et  $p_2$  vecteur d'ordre  $1 \times (d-t)$ . Alors il existe un vecteur  $y$  d'ordre  $1 \times (r-1)$  tel que

$$yP_{12} = p_2.$$

Donc si on note  $\alpha_k = {}^t[c_{k,1} \mid \alpha_{k,2}]$  où  $\alpha_{k,1}$  est un vecteur d'ordre  $(r-1) \times 1$  et  $\alpha_{k,2} \in GF(q)$ , (\*) est équivalent au système

$$\begin{cases} P_{12}x = c_{k,1} - P_{11}s \\ yP_{12}x = c_{k,2} - p_1s \end{cases}$$

Soit  $A$  la matrice carrée d'ordre  $r$

$$A = \left[ \begin{array}{c|c} I_{r-1} & 0 \\ \hline y & 1 \end{array} \right],$$

alors (\*)  $\Leftrightarrow$

$$\begin{aligned} A \begin{bmatrix} P_{12}X \\ 0 \end{bmatrix} &= \begin{bmatrix} c_{k,1} - P_{11}s \\ c_{k,2} - p_1s \end{bmatrix} \\ \Leftrightarrow \begin{bmatrix} P_{12}X \\ 0 \end{bmatrix} &= A^{-1} \begin{bmatrix} c_{k,1} - P_{11}s \\ c_{k,2} - p_1s \end{bmatrix} \\ \Leftrightarrow \begin{cases} P_{12}X = c_{k,1} - P_{11}s & (1) \\ y'c_k = yP_{11} - p_1s & (2) \end{cases} \end{aligned}$$

où  $y' = [y \mid -1]$ .

$P_{12}$  est une matrice d'ordre  $(r-1) \times (d-t)$  et de rang  $(r-1)$  donc pour  $k$  fixé, l'équation (1) admet  $q^{d-t-r+1}$  solutions.

Il reste donc à montrer qu'il existe un unique  $k \in \{0, \dots, q-1\}$  tel que l'équation (2) soit vérifiée.

On note  $(yP_{11} - p_1s) = h(s) \in GF(q)$ .  $\text{supp}(y'P) \subseteq J_t$  puisque

$$y'P = [y \mid -1]P = [yP_{11} - p_1 \mid yP_{12} - p_2] = [yP_{11} - p_1 \mid 0].$$

Alors par hypothèse  $y'C$  est une permutation de  $0, \dots, q-1$ , donc nécessairement, il existe un unique  $k \in \{0, \dots, q-1\}$  tel que  $y'C = h(s)$ .

Montrons maintenant par l'absurde que le rang de  $P_2$  ne peut pas être inférieur strictement à  $r-1$ . On suppose et on note  $m = r - \text{rg}(P_2) > 1$ . On cherche alors s'il existe une matrice  $C$  d'ordre  $r \times q$  telle que  $\forall y \neq 0 \in GF(q)^r$  tel que  $\text{supp}(yP) \subseteq J_t$ , i.e  $yP_2 = 0$ ,  $yC = \pi_y(q)$ . On remarque que si  $yC = \pi_y(q)$  et  $y' = \alpha y$ ,  $\alpha \in GF(q)$  alors  $y'C = \pi_{y'}(q)$ , donc deux vecteurs linéairement dépendants imposent la même contrainte sur  $C$ . *A contrario*, si  $yC = \pi_y(q)$  et  $y'C = \pi_{y'}(q)$  alors on n'a pas nécessairement  $(y+y')C = \pi_{(y+y')}(q)$ . Donc le nombre de contraintes que l'on impose à la matrice  $C$  est égal au nombre de vecteurs  $y$  d'ordre  $1 \times r$  linéairement indépendants deux à deux tels que  $yP_2 = 0$ , c'est-à-dire

$$\frac{1 - q^m}{1 - q}.$$

Nous obtenons alors un système à  $rq$  inconnues (les éléments de  $C$ ) et  $q(q^m - 1)/(q - 1)$  équations (le nombre de contraintes sur  $C$  que multiplie le nombre de colonnes de  $C$ ). Soit la différence

$$d = \frac{q^m - 1}{q - 1} q - rq = \frac{q(q^m - rq + r - 1)}{q - 1}.$$

Etant donné que  $r \leq q$  et  $m \geq 2$ ,

$$d \geq \frac{q}{q-1} (r^m - r^2 + r - 1) \geq r - 1.$$

Si  $r > 1$  alors  $d > 0$  et si  $r = 1$ ,  $d = q(q^m - q)/(q - 1) > 0$ , autrement dit, le nombre d'équations est supérieur au nombre d'inconnues, le système n'admet pas de solution, i.e  $C$  n'existe pas.  $\square$

## Bibliographie

- BuHamra S., Anderson D.A. (1997). Prime power parallel flats fractions. Representation and properties. *Utilitas Mathematica* **51**, 238-254.
- BuHamra S., Anderson D. (1994). The spectrum of the information matrix for  $s^n$  parallel flats fractions when  $s$  is prime. *J. of Statist. Planning and Inference* **47**, 333-345.
- Connor W.S., Young S. (1959). *Fractional factorial designs for experiments with factors at two or three levels*. Appl. Math. Ser.. National bureau of standards U.S. government printing office, Washington.
- John P.N.M. (1961). Three-quarter replicates of  $2^4$  and  $2^5$  designs. *Biometrics* **17**, 319-321.
- John P.N.M. (1962). Three-quarter replicates of  $2^n$  designs. *Biometrics* **18**, 172-184..
- Kobilinsky A., Monod H. (1995). Juxtaposition of regular factorial designs and the complex linear model. *Scand. J. Statist.* **22**, 223-254.
- Srivastava J.N, Li J. (1996). Orthogonal designs of parallel flats type. *J. of statist. Planning and Inference* **53**,261-283.
- Srivastava J.N. (1965). Theory of optimal nonsingular semi-regular fractional designs. *35<sup>th</sup> of the international statistical Institute*.
- Srivastava J.N. (1987). Advances in the general theory of factorial designs based on partial pencils in euclidean  $n$ -space. *Utilitas Mathematica* **32**, 75-94.
- Srivastava J.N., Anderson D., Mardekian J. (1984). Theory of factorial designs of the parallel flats type I : the coefficient matrix. *J. of Statist. Planning and Inference* **9**,229-252.
- Srivastava J.N., Chopra D.V. (1973). Balanced arrays and orthogonal arrays. In *A survey of combinatorial theory* (Srivastava ed.) North Holland, 411-428.
- Srivastava J.N., Throop D. (1990). Orthogonal arrays obtainable as solutions to linear equations over finite fields. *Linear Algebra Appl.* **127**, 283-300.



# Conclusion

Dans la première partie du mémoire nous avons vu que les approches statistiques existantes dans la bibliographie présentent quelques idées intéressantes pour traiter les expériences simulées. L'approche par résidu aléatoire propose un modèle qui permet de prendre en compte le caractère déterministe du simulateur grâce au processus résiduel en introduisant une corrélation spatiale entre les réponses du simulateur. L'approche par échantillon aléatoire, quant à elle, détermine des méthodes d'échantillonnage utilisant des plans qui ont pour principale qualité de bien 'remplir' l'espace du domaine d'échantillonnage.

Cependant comme nous l'avons signalé, séparément chacune de ces deux approches présentent aussi quelques faiblesses. D'une part le prédicteur construit à partir du modèle de l'approche par résidu aléatoire est irrégulier et peu robuste aux variations du paramètre de corrélation. Ce qui est dû au rôle trop prépondérant du processus résiduel dans le modèle. D'autre part, les plans construits à partir de critère de qualité tels que l'entropie ou la distance maximin, ne représentent pas de façon uniforme le domaine d'échantillonnage et laissent des parties du cube unité non testées par les simulations. De plus ces plans dépendent entièrement du modèle choisi, notamment du paramètre de corrélation.

C'est pourquoi dans la deuxième partie du mémoire nous avons proposé une nouvelle approche statistique des expériences simulées qui permet d'intégrer les points forts des approches de la première partie et qui a pour but d'en supprimer les inconvénients. Les objectifs de cette nouvelle approche ont été atteints aussi bien au niveau du modèle que des plans d'échantillonnage.

En ce qui concerne le nouveau modèle proposé pour la réponse du simulateur, nous avons vu que celui-ci permet :

- D'une part de modérer le rôle joué par le processus résiduel de façon à ce que toute la prédiction ne repose pas sur l'*a priori* que l'on fait sur la fonction de corrélation, comme cela était le cas dans l'approche par résidu aléatoire.
- D'autre part, de prendre en compte l'effet du second degré de l'échantillonnage puisque seul le premier degré est contrôlable par l'utilisateur.

Ce nouveau modèle permet donc d'obtenir un prédicteur plus lisse que le prédicteur de l'approche par résidu aléatoire et robuste aux variations de  $\theta$ . C'est-à-dire que les deux points faibles du modèle du chapitre 1 sont maintenant supprimés.

Pour ce qui est des plans d'échantillonnage, ceux-ci remplissent bien les deux principales fonctions qui leur sont demandées, *i.e* :

- Une bonne répartition des points d'échantillonnage dans le cube unité afin que toutes les parties du domaine d'échantillonnage soient testées par les simulations,
- L'ajustement d'une régression linéaire trigonométrique.

Les techniques d'échantillonnage proposées permettent donc de construire des plans de bonne qualité quel que soit le modèle choisi et qui, de fait, sont robustes aux variations de  $\theta$ . L'utilisation de plans classiquement utilisés en plans d'expériences tels que les tableaux orthogonaux linéaires ou les fractions régulières est donc parfaitement appropriée et performante dans le contexte des expériences simulées

La difficulté avec le type de plans que nous utilisons réside dans le fait que très peu de méthodes de construction de ces plans sont proposées dans la bibliographie. C'est pourquoi la troisième partie de ce mémoire est consacrée à la mise en place de techniques de construction de tableaux orthogonaux linéaires et de fractions régulières. Dans cette dernière partie nous avons aussi vu qu'une nouvelle catégorie de plans, les fractions semi-régulières, offre des perspectives intéressantes dans le cadre de l'ajustement d'une régression trigonométrique. En effet, cette structure de plans permet de s'affranchir de la contrainte d'orthogonalité quant à l'estimabilité du paramètre  $\beta$  de la régression trigonométrique. Ce qui permet d'ajouter des fréquences à la régression et ainsi enrichir le modèle en espérance. Or nous avons vu que, lorsque l'ajustement en moyenne est de bonne qualité, le prédicteur est plus robuste aux variations du paramètre de corrélation.

Le travail réalisé dans ce mémoire traite le problème des expériences simulées de façon généraliste. Il permet à la fois de faire une synthèse de différentes approches envisagées dans la bibliographie, et à la fois de proposer des outils bien fondés pour aborder les expériences simulées. Ces outils sont à adapter et à développer suivant le type de simulateur étudié, ce qui offre de nouvelles perspectives.

# Perspectives

L'étude exposée dans ce mémoire est un travail exploratoire qui aborde des domaines aussi divers que par exemple, les méthodes d'intégration numérique, les processus spatiaux, les plans d'expériences, *etc* ... Les perspectives de ce travail sont donc très nombreuses et variées, dont voici quelques exemples.

## Application à un cas réel

L'approche statistique proposée dans ce mémoire aborde le problème des expériences simulées d'un point de vue généraliste. Il semble maintenant nécessaire d'appliquer cette théorie à un cas réel de façon à pouvoir aborder des questions circonstanciées telles que :

- L'influence et la sensibilité des différents paramètres d'entrée du simulateur. C'est en effet essentiellement pour débattre de cette question que l'utilisateur souhaite disposer d'un modèle résumant la réponse du simulateur.
- Le coût de simulation suivant le type de simulateur étudié. Est-il effectivement préférable d'utiliser un hypercube latin dans le cas où le simulateur prend en compte beaucoup de facteurs mais a un temps de calcul moyennement long, et un tableau orthogonal de force plus élevée si le simulateur a peu de paramètres d'entrée mais un temps de calcul très long ?
- Les bornes du domaine d'échantillonnage. Dans ce mémoire, nous avons réduit notre étude au cube unité mais en pratique la délimitation du domaine d'échantillonnage pose problème. En effet, dans le cas des expériences réelles les valeurs entre lesquelles varient les facteurs sont déterminées à la fois par des lois physiques, et à la fois par des contraintes pratiques dues à l'expérimentation. Pour les expériences simulées seules les lois physiques interviennent, ce qui laisse beaucoup de possibilités pour faire varier les paramètres d'entrée du simulateur. Déterminer le domaine d'échantillonnage devient alors un problème complexe.

## Régression linéaire polynomiale

Dans ce travail, la régression linéaire envisagée pour la modélisation en espérance est trigonométrique. Nous avons vu que ce type de modèle est d'une part bien adapté aux

fractions régulières utilisées pour l'échantillonnage, puisque l'une comme l'autre s'exprime simplement grâce à la théorie des groupes finis. D'autre part, il est clair à l'issue des illustrations numériques du chapitre 4 que le fait d'ajouter une régression trigonométrique dans le modèle permet de stabiliser le prédicteur et de le rendre robuste aux variations du paramètre de corrélation  $\theta$ . Il serait intéressant de savoir si un modèle plus couramment utilisé telle qu'une régression polynomiale ou spline, aboutirait au même résultat. Autrement dit, si la stabilisation du prédicteur n'est pas due uniquement au fait que l'ajustement en moyenne est plus riche. Et si cela est le cas, il faudrait aussi vérifier que les plans d'échantillonnage que nous avons proposés restent de bonne qualité. Ces plans étant construits de façon indépendante du modèle, il semble *a priori* qu'il n'y ait aucune raison pour qu'ils ne restent pas efficaces. C'est l'avantage et la raison pour lesquelles nous avons proposé ce type d'échantillonnage.

### **Perspectives plus théoriques**

De façon plus théorique, il faudrait certainement creuser des questions abordées dans le mémoire telles que :

- Déterminer la fonction de discrédance d'un tableau orthogonal linéaire de façon à démontrer, qu'effectivement, plus la force du tableau est élevée, plus la fonction de répartition empirique des points du tableau est proche de la distribution uniforme, *i.e* mieux les points répartis dans le domaine d'échantillonnage. Nous avons vu dans le chapitre 5 que Fang et Mukerjee (2000) ont effectué une première démarche en ce sens mais ils se limitent aux tableaux à deux niveaux. Il reste donc à trouver la bonne fonction de discrédance pour généraliser ce résultat à un nombre de niveaux quelconque.
- Etudier plus en avant les propriétés d'estimabilité et d'optimalité des fractions semi-régulières dans le cadre d'une régression linéaire trigonométrique, ainsi que définir des méthodes de construction pour ce type de plans.
- Envisager l'utilisation des nets ou des séquences pour l'échantillonnage dans le cas d'un découpage du cube unité sélectif, c'est-à-dire un maillage plus ou moins fin à certains endroits du domaine d'échantillonnage suivant le comportement de la réponse du simulateur.

Ainsi s'achève ce travail de thèse qui, comme nous pouvons le constater, est une première approche des expériences simulées, qui a pour ambition de répertorier les différents outils qui sont à notre disposition pour traiter cette question, et répondre à quelques problèmes soulevés dans la bibliographie grâce à la nouvelle approche statistique que nous avons proposée. L'étude des expériences simulées ne fait en fait que commencer.

# Bibliographie générale

- Adhikary B., Das P. (1992). Multiple fold-over and its application to orthogonal arrays. *Sankhya*, special volume **54**, 21-24.
- Bailey R.A (1985). Factorial designs and abelian groups. *Lin. Algebra Appl.* **70**,349-363.
- Bailey R.A. (1982). Dual abelian group in the design of experiments. In *Algebraic structures and applications* (P. Schultz *et al.*, eds.), New York, 45-54.
- Bailey R.A. (1990). Cyclic design and factorial designs. In *Proceedings of of the R.C. Bose Symposium, Delhi 1988 on probability, statistics and design of experiments* (R.R. Bahadur, K. Sen, Eds.), New York, 45-54.
- Bailey R.A., Gilchrist F.H.L, Patterson H.D. (1977). Identification of effects and confounding patterns in factorial design. *Biometrika* **64**, 347-354.
- Bates R.A., Buck R.J., Riccomagno E., Wynn H.P. (1996). Experimental design and observation for large systems. *J. R. Statist. Soc. B* **58**, 77-94.
- Bates R.A., Riccomagno E., Schwabe R., Wynn H.P. (1998). Lattices and dual lattices in optimal experimental design for Fourier models. *Computational Statistics and Data Analysis* **28**, 283-296.
- Bose R.C. (1947). Mathematical theory of symmetrical factorial design. *Sankhya* **8**, 107-166.
- Bose R.C., Bush K.A. (1952). Orthogonal arrays of strength two and three. *Ann. Math. Statist.* **23**, 508-524.
- Braaten E., Weller G. (1979). An improved low-discrepancy sequence for multidimensionnal quasi-Monte Carlo integration. *J. Comp. Physics* **33**, 249-258.
- BuHamra S., Anderson D. (1994). The spectrum of the information matrix for  $s^n$  parallel flats fractions when  $s$  is prime. *J. of Statist. Planning and Inference* **47**, 333-345.
- BuHamra S., Anderson D.A. (1997). Prime power parallel flats fractions. Representation and properties. *Utilitas Mathematica* **51**, 238-254.
- Cassels J.W.S. (1971). *An introduction to the geometry of numbers*. Springer Verlag, Berlin.
- Christensen R. (1987). *Plane answers to complex questions : The theory of linear models*. Springer-Verlag.
- Christensen R. (1990). *Linear models for multivariate, time series, and spatial data*. Springer-Verlag.
- Ciarlet J.P. (1990). *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson.

- Collombier D. (1996). *Plans d'expérience factoriels*. Springer-Verlag, Berlin.
- Connor W.S., Young S. (1959). *Fractional factorial designs for experiments with factors at two or three levels*. Appl. Math. Ser.. National bureau of standards U.S. government printing office, Washington.
- Costa J.P., Pronzato L., Thierry E. (1999). Nonlinear prediction by kriging, with application to noise cancellation. Soumis à *Signal Processing*.
- Currin C.T., Mitchell M., Morris M., Ylvisaker D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86**, 953-963.
- Drmosta M., Tichy R.F. (1997). *Sequences, discrepancies and applications*. Lecture Notes in Mathematics 1651. Springer Verlag, Berlin.
- Efron B., Stein C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9**, 586-596.
- El Mossadeq A. (1996). Fractions de résolution  $\omega$  et groupes abéliens finis. *Linear algebra Appl.* **237/238**, 539-547.
- El Mossadeq A., Kobilinsky A., Collombier D. (1985). Construction de tableaux orthogonaux dans les groupes abélien finis et confusions d'effets dans les plans factoriels. *Linear Algebra Appl.* **70**, 303-320.
- Fang K.T., Mukerjee R. (2000). Miscellanea. A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* **87**, 193-198.
- Fang K.T., Wang Y. (1993). *Number-theoretic methods in statistics*. Chapman and Hall, London.
- Fang K.T., Wang Y., Bentler P. (1994). Some applications of number-theoretic methods in statistics. *Statist. Sci.* **9**, 416-428.
- Fang K.T., Zhu L.X. (1993). On uniformly random design. *Technical report MATH-020*, Hong Kong Baptist College.
- Faure H. (1982). Discrépance de suites associées à un système de numération (en dimension  $s$ ). *Acta. Arith.* **41**, 337-351.
- Franklin M.F. (1985). Selecting defining contrasts and confounded effects in  $p^{n-m}$  factorial experiments. *Technometrics* **27**, 165-172.
- Franklin M.F., Bailey R.A (1977). Selection of defining contrasts and confounded effects in two-level experiments. *Appl. Statist.* **26**, 321-326.
- Halton J.H. (1960). On efficiency of certain quasi-random sequences of points in evaluating multidimensionnal integrals. *Numerische Mathematik* **2**, 84-90.
- Hedayat A., Pu K., Stufken J. (1992). On the construction of asymmetrical orthogonal arrays. *Ann. Statist.* **20**, 2142-2152.
- Hedayat A.S., Rao C.R., Stufken J. (1988). Sampling plans excluding contiguous units. *J. Statist. Plann. Inf.* **19**, 159-170.
- Hedayat A.S., Sloane N.J.A, Stufken J. (1999). *Orthogonal Arrays : Theory and applications*. Springer-Verlag, New-York.
- Hedayat A.S., Stufken J., Su O. (1996). On differences schemes and orthogonal arrays of strength  $t$ . *J. of Statist. Planning and Inferences* **56**, 307-324.

- Hickernell F.J. (1998). A generalized discrepancy and quadrature error bound. *Math. of Comp.* **67**, 299-322.
- Hlawka E. (1964). Discrepancy and uniform distribution of sequences. *Compositio Math.* **16**, 83-91.
- Hoel P.G. (1965). Minimax designs in two dimensional regression. *Ann. Math. Statist.* **36**, 1097-1106.
- Hua L.K., Wang Y. (1981). *Applications of number theoretic to numerical analysis*. Springer Verlag, Berlin.
- Jensen R.J. (1975). Square and cubic lattice sampling. *Biometrics* **31**, 449-471.
- John P.N.M. (1961). Three-quarter replicates of  $2^4$  and  $2^5$  designs. *Biometrics* **17**, 319-321.
- John P.N.M. (1962). Three-quarter replicates of  $2^n$  designs. *Biometrics* **18**, 172-184..
- Johnson M.E., Moore L.M., Ylvisaker D. (1990). Minimax and maximin distance designs. *J. of Statist. Planning and Inference* **26**, 131-148.
- Kiefer J., Wolfowitz J. (1959). Optimum design in regression problems. *Ann. Math. Statist.* **30**, 271-294.
- Kobilinsky A. (1985b). Confounding in relation to duality of finite abelian groups. *Linear Algebra Appl.* **70**, 321-347.
- Kobilinsky A. (1985a). Orthogonal factorial designs for quantitative factors. *Statistics and Decisions, Supplement* **2**, 272-347.
- Kobilinsky A. (1990). Complex linear models and cyclic designs. *Linear Algebra Appl.* **127**, 227-282.
- Kobilinsky A., Monod H. (1991). Experimental design generated by group morphisms : An introduction. *Scand. J. Statist.* **18**, 119-134.
- Kobilinsky A., Monod H. (1995). Juxtaposition of regular factorial designs and the complex linear model. *Scand. J. Statist.* **22**, 223-254.
- Koehler J.R., Owen A.B. (1996). Computer experiments. In *Handbook of statistics, 13 : Design and analysis of experiments* (Ghosh, Rao Eds.), North-Holland, 261-308.
- Koksma J.F. (1942/43). Een algemeene stelling uit de theorie der gelijkmatige verdeling modulo 1. *Mathematica B (Zutphen)* **11**, 7-11.
- Lindley D.V. (1956). On a measure of information provided by an experiment. *Ann. of Math. Statist.* **27**, 986-1005.
- Loh W.C. (1993). On latin hypercube sampling. *Technical Report* N° 93-52, dept. of Statistics, Purdue University.
- Loh W.C. (1996). A combinatorial central limit theorem for randomized orthogonal array sampling designs. *Ann. Statist.* **24**, 1209-1224.
- Mardia K.V., Marshall R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Matèrn B. (1947). Methods of estimating the accuracy of line and sample plot surveys. *Medd. Skogsforskn Inst.* **36** (1).
- Matheron G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246-1266.

- Mc Kay M.D., Beckman R.J., Conover W.J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239-245.
- Mitchell T. J. (1974). An algorithm for construction of 'D-optimal' experimental designs. *Technometrics* **16**, 203-210.
- Mullen G.L., Mahalanabis A., Niederreiter H. (1995). Tables of (t,m,s)-net and (t,s)-sequence parameters. In Niederreiter, H. and Shiue, P. J.-S., editors, *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pp 299-317, New-York. Springer Verlag.
- Niederreiter H. (1978). Quasi-Monte Carlo methods and pseudo-random numbers. *Bull. Amer. Math. Soc.* **84**, 957-1041.
- Niederreiter H. (1987). Point sets and sequences with small discrepancy. *Monatsh. Math.* **104**, 273-337.
- Niederreiter H. (1992). *Random number generation and quasi-Monte Carlo methods*. CBMS-NSF, SIAM, Philadelphia.
- Owen A.B. (1992). A central limit theorem for latin hypercube sampling. *J. Roy. Statist. Soc. B* **54**, 541-551.
- Owen A.B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2**, 439-452.
- Owen A.B. (1994). Lattice sampling revisited : Monte Carlo variance of means over randomized orthogonal arrays. *Ann. Statist.* **22**, 930-945.
- Owen A.B. (1995). Randomly permuted (t,m,s)-nets and (t,s)-sequences. In Niederreiter, H. and Shiue, P. J.-S., editors, *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pp 299-317, New-York. Springer Verlag.
- Owen A.B. (1997). Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis* **34(5)**. In press.
- Pagès G., Xiao Y.J. (1991). Sequence with low discrepancy and pseudo-random numbers. *Technical Report*, Laboratoire de mathématiques et de modélisation, Paris.
- Park J.S. (1994). Optimal Latin hypercube designs for computer experiments. *J. of Statist. Planning and Inference* **39**, 95-111.
- Patterson H.D. (1954). The error of lattice sampling. *J. Roy. Statist. Soc. B* **16**, 140-149.
- Patterson H.D. (1976). Generation of factorial designs. *J. of Roy Soc. B* **38**, 175-179.
- Patterson H.D., Bailey R.A. (1978). Design keys for factorial experiments. *Appl. Statist.* **27**, 335-343.
- Raghavarao D. (1971). *Construction and combinatorial problems in design of experiments*. Wiley, New York.
- Raktoe B.L. (1969). Combining elements from distinct finite fields in mixed factorials. *Ann. Math. Statist.* **40**, 498-504.
- Rao C.R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *J. Roy. Statist. Soc. (suppl.)* **9**, 128-139.
- Riccomagno E., Schwabe R., Wynn H.P. (1997). Lattice-based D-optimum design for Fourier regression. *Ann. of statist.*

- Robinson G.K. (1991). That BLUP is a goob thing : the estimation of random effects. *Statist. Science* **6**, 15-51.
- Sacks J., Schiller S. (1988). Spatial designs. In *Statistical Decision Theory and Related Topics IV : vol 2* (Gupta, Berger Eds.), Springer Verlag, New York, 385-399.
- Sacks J., Schiller S.B., Welch W.J. (1989). Designs for computer experiments. *Technometrics* **31**, 41-47.
- Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). Design and analysis of computer experiments. *Statist. Science* **4**, 409-435.
- Schwery M.C., Wynn H.P. (1988). Maximum entropy sampling with application to simulation code. In Proc. 12<sup>th</sup> world congress on scientific computation, IMAC88 **2**, 517-519.
- Seiden E. (1954). On the problem of construction of orthogonal array. *Ann. Math. Statist.* **25**, 1451-156.
- Serres J.P. (1967). *Représentations linéaires des groupes finis*. Herman, Paris.
- Shwery M.C., Wynn H.P. (1987). Maximum entropy sampling. *J. of Appl. Statist.* **14**, 165-170.
- Sloan I.H., Joe S. (1994). *Lattice methods for multiple integration*. Oxford science Publications.
- Sloan I.H., Kachoyan P.J. (1987). Lattice methods for multiple integration : theory, error analysis and examples. *SIAM Journal on numerical Analysis* **24**, 116-128.
- Sloan I.H., Lyness J.N. (1989). The representation of lattice quadrature rules as multiple sums. *Math. of computation* **52**, 81-94.
- Sloan I.H., Lyness J.N. (1990). Lattice rules : projection regularity and unique representation. *Math. of computation* **54**, 649-660.
- Sloan I.H., Walsh L. (1990). A computer search of rank 2 lattice rules for multidimensionnal quadrature. *Math. of Computation* **54**, 281-302.
- Sobol' I.M. (1967). The distribution of points in a cube and the accurate evaluation of integrals (in russian). *Zh. Vychisl. Mat. i Mat. Phys.* **7**, 784-802.
- Srivastava J.N, Li J. (1996). Orthogonal designs of parallel flats type. *J. of statist. Planning and Inference* **53**, 261-283.
- Srivastava J.N. (1965). Theory of optimal nonsingular semi-regular fractional designs. *35<sup>th</sup> of the international statistical Institute*.
- Srivastava J.N. (1987). Advances in the general theory of factorial designs based on partial pencils in euclidean n-space. *Utilitas Mathematica* **32**, 75-94.
- Srivastava J.N., Anderson D., Mardekian J. (1984). Theory of factorial designs of the parallel flats type I : the coefficient matrix. *J. of Statist. Planning and Inference* **9**, 229-252.
- Srivastava J.N., Chopra D.V. (1973). Balanced arrays and orthogonal arrays. In *A survey of combinatorial theory* (Srivastava ed.) North Holland, 411-428.
- Srivastava J.N., Throop D. (1990). Orthogonal arrays obtainable as solutions to linear equations over finite fields. *Linear Algebra Appl.* **127**, 283-300.
- Stein M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics* **29**, 143-151.

- Tang B. (1993). Orthogonal array-based latin hypercubes. *J. Amer.Statist. Assoc.* **88**, 1392-1397.
- Tang B. (1994). A theorem for selecting OA-based Latin hypercubes using a distance criterion. *Comm. Statist. - Theory Meth.* **23**, 2047-2058.
- Tocher K.D. (1963). *The art of simulation*. Princeton, N.J. D Van Nostrand. pp.106-107.
- Wang J.C., Wu C.F.J. (1991). An approach to the construction of asymmetrical orthogonal arrays. *J. Amer. Statist* **86**, 450-456.
- Wang Y., Fang K.T. (1990a). Number theoretic methods applied statistics. *Chin. Ann. Math. Ser. B* **11**, 41-55.
- Wang Y., Fang K.T. (1990b). Number theoretic methods applied statistics (II). *Chin. Ann. Math. Ser. B* **11**, 859-914.
- Warnes J.J., Ripley B.D. (1987). Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika* **74**, 640-642.
- Weber N.C. (1980). A martingale approach to central limit theorems for exchangeable random variables. *J. Appl. Prob.* **17**, 662-673.
- Welch W.J., Buck R.J, Sacks J., Wynn H.P., Mitchell T.J, Morris M.D. (1992). Screening, predicting and computer experiments. *Technometrics* **34**, 15-25.
- Ying Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. of Statist.* **21**, 1567-1590.
- Zain A.A., Rajan B.S. (1997). Dual codes of systematic group codes over abelian groups. *AAECC* **8**, 71-83.