

# Plans d'expériences optimaux : un exposé didactique

Jean-Pierre Gauchi<sup>1</sup>

INRA

Unité de Mathématiques et Informatique Appliquées (MIA)

Jouy-en-Josas, France

jean-pierre.gauchi@jouy.inra.fr

## Résumé

Cet article est un exposé didactique sur les plans d'expériences optimaux, essentiellement pour les modèles linéaires. Son objectif est de donner les définitions et les modes de construction des principaux plans optimaux rencontrés, avec un éclairage basé sur des exemples très simples. Le public visé est d'abord celui des personnels de recherche qui souhaitent améliorer leur démarche expérimentale, face à des problèmes expérimentaux caractérisés par des contraintes de divers types.

## Abstract

This paper is a didactic article on the optimal experimental designs, primarily for the linear models. Its objective is to give the definitions and the methods of construction of the principal optimal plans met, with a lighting based on very simple examples. The public concerned is initially that of the engineers of research who wish to improve their experimental approach, relatively to experimental problems characterized by constraints of various types.

**Mots-clés :** Plans d'expériences, optimalité, régression, didactique.

## 1 Introduction

Cet article est un exposé didactique, élémentaire même, sur les plans d'expériences optimaux. Il reprend certains éléments, en les simplifiant, de Gauchi (1997), publication où le lecteur pourra compléter ses connaissances sur ce sujet, ainsi que dans l'excellent ouvrage de Atkinson et Donev (1992). L'objectif de cet article est de définir précisément les critères d'optimalité les plus fréquemment rencontrés, et de montrer quelques algorithmes de construction de base, à l'aide d'exemples très simples. Le public visé est d'abord celui des personnels de recherche qui souhaitent améliorer leur démarche expérimentale, face à des problèmes caractérisés par des contraintes de divers types, tout en gérant au mieux la variabilité expérimentale. Le deuxième paragraphe illustre ces types de contraintes fréquemment rencontrées dans quelques situations classiques. Les outils nécessaires sont à l'intersection de la statistique, de l'analyse convexe et de l'optimisation. Le pré-requis pour la lecture de cet article est une formation élémentaire en statistique et en calcul matriciel d'une part, et une connaissance des plans d'expériences factoriels usuels d'autre part.

Cette conception de l'optimalité dans le domaine des plans d'expériences, outils eux-mêmes fondés par le statisticien Fisher et ses "disciples" dans les 1925 à 1935, est due aux travaux de Kiefer à partir de la fin des années 1950, puis à ceux de Fedorov à partir de la fin des années 1960. Plus tard, plusieurs plans classiques ont pu apparaître comme des plans optimaux pour des critères d'optimalité proposés par Kiefer. Si les objectifs des plans classiques et des plans optimaux sont essentiellement les mêmes, les méthodes pour les construire étaient au départ fondamentalement différentes, essentiellement d'essence algébrique pour les premiers et algorithmique pour les seconds. Toutefois, on a pu voir que certains types de plans classiques (recherche de plans en blocs équilibrés par exemple) nécessitaient également l'usage d'algorithmes performants pour suppléer aux méthodes algébriques. Et de façon analogue, il faut aussi rappeler que les algorithmes utilisés pour construire les plans optimaux sont très souvent améliorés ou qualifiés grâce à des résultats importants de l'analyse convexe et de l'optimisation (Fedorov, 1980, Silvey, 1980). Finalement, aujourd'hui, la dichotomie quant aux méthodes de construction des plans n'est plus aussi marquée, et on soulignera par ailleurs que la construction de plans par des méthodes algébriques reste toujours un thème de recherche fructueuse conduisant à des plans particulièrement adaptés aux facteurs explicatifs qualitatifs.

Bien que cette théorie de l'optimalité puisse prendre en compte plusieurs formes de modèle de la réponse, par exemple jusqu'à des systèmes d'équations différentielles, nous limiterons notre exposé à des modèles de forme analytique explicite : les modèles de régression ou d'analyse de la variance, linéaires dans les paramètres, et les modèles de régression (intrinsèquement) non linéaire dans les paramètres. Enfin, nous supposerons que les facteurs expérimentaux étudiés (discrets ou continus) sont, par nature, contrôlables, et l'erreur faite dans leur réglage est négligeable au regard de celle de l'expérience (au sens de sa reproductibilité ou de la répétabilité des mesures réalisées). Dans le langage usuel des statisticiens on parle de modèles à effets "fixes".

La structure de cet article est la suivante : après une section consacrée à l'évocation de quelques situations expérimentales réelles à but de sensibilisation, deux sections suivent, l'une pour les modèles linéaires et l'autre pour les modèles de régression non linéaire. Une conclusion clôture l'article.

## 2 Quelques situations expérimentales réelles

Pour entrer rapidement dans le vif du sujet nous présentons maintenant quelques exemples de situations expérimentales particulières mais concrètes et fréquentes (déjà présentées dans Gauchi (1997)). Celles-ci sont difficiles ou même impossibles à appréhender avec des plans classiques, notamment les plans factoriels (complets et fractionnaires) ou les plans de surface de réponse, à cause des contraintes qui les caractérisent.

### 2.1 Domaine expérimental sous contraintes

On cherche à modéliser à l'aide d'un polynôme le rendement d'une réaction chimique en fonction de deux facteurs continus, la température du milieu réactionnel et la pression dans le réacteur, pour disposer *in fine* à la fois d'estimations fiables des paramètres traduisant l'effet de ces facteurs, et d'un modèle prédictif de bonne qualité sur le domaine expérimental exploré. On ne dispose pas de modèle théorique pour la réponse "rende-

ment” qui lie les réactifs en présence. On se contente donc, en première approximation, du polynôme du second degré :

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1x_2 + \varepsilon \quad (1)$$

où  $y$  est la réponse, et les coefficients  $\beta_i$ ,  $\beta_{ii}$  et  $\beta_{12}$  représentent respectivement les effets principaux, les effets quadratiques et l’interaction double des deux facteurs continus, paramètres que l’on cherche à estimer ;  $\varepsilon$  est un terme d’erreur, de distribution inconnue mais supposé de variance constante sur le domaine que l’on cherche à explorer.

Supposons que les échelles des deux facteurs  $x_1$  et  $x_2$  soient codées de telle façon que leurs niveaux extrêmes minimum et maximum soient  $-1$  et  $+1$ . Puisque ce sont des facteurs continus ils définissent donc un domaine expérimental de dimension 2 représenté par l’intérieur (compact) d’un carré dont les coordonnées des sommets sont  $(-1, +1)$ ,  $(+1, +1)$ ,  $(-1, -1)$ ,  $(+1, -1)$ . Pour le modèle (1) et ce domaine on sait qu’un plan d’expériences possible, orthonormé si on transforme les termes quadratiques par codage avec les polynômes orthonormés (Kobilinsky, 1988), pourrait être constitué de répétitions d’expériences en les neuf points de support d’un plan factoriel  $3^2$  (voir Gauchi, 1997, page 327). Pour des raisons de faisabilité chimique imaginons maintenant que l’on ne puisse pas réaliser d’expériences dans certaines zones du domaine expérimental telles que trois expériences parmi les neuf deviennent interdites. Le domaine expérimental carré est devenu un polygone quelconque (voir Tableau II pour un exemple). La qualité du plan se dégrade, notamment il y a perte d’orthogonalité. Avec le plan formé de répétitions d’expériences sur les six points de support restants, il sera difficile de distinguer les influences marginales de ces deux facteurs, et la précision des estimateurs que l’on obtiendra sera beaucoup plus faible. La question qui se pose est donc de trouver d’autres expériences dans ce domaine sous contraintes pour permettre une estimation optimisée de ce modèle.

## 2.2 Modèle polynomial particulier

Un cas également fréquent est celui où certains termes sont absents du développement limité habituel. On peut donner, à titre d’illustration, la forme polynomiale incomplète pour quatre  $x_i$  continus :

$$y = \beta_0 + \sum_{i=1}^4 \beta_i x_i + \beta_{12} x_1 x_2 + \beta_{14} x_1 x_4 + \beta_{11} x_1^2 + \beta_{33} x_3^2 + \varepsilon$$

Avec ce modèle particulier un plan de surface de réponse usuel ne sera pas optimal.

## 2.3 Modèle en Structure-Activité

Les études de structure-activité ou de sélection de molécules constituent un domaine d’application pour lequel l’approche plan classique est particulièrement inadaptée. Le but est de construire un modèle quantitatif de la réponse ”activité biologique” en fonction des descripteurs moléculaires physico-chimiques (facteurs explicatifs) à partir de mesures sur un sous-ensemble de molécules. L’objectif est donc de sélectionner des molécules dont on va étudier l’activité. On cherche à attribuer à chacune des molécules une valeur numérique (une mesure en fait) traduisant leur importance dans l’établissement d’un modèle prédictif de l’activité biologique de la meilleure qualité possible, c’est-à-dire dont les prédictions seront de variance minimale, en moyenne.

## 2.4 Discrimination de modèles

C'est une situation où on cherche un plan optimal pour à la fois discriminer des modèles et faire de l'estimation de paramètres. C'est notamment la situation où le chercheur hésite entre plusieurs modèles possibles pour expliquer ou représenter son phénomène. Les modèles sont linéaires ou non, peuvent être emboîtés ou non, mais les solutions en termes de critères d'optimalité ne couvrent pas toutes ces situations dont certaines sont encore du domaine de la recherche en Statistique.

## 2.5 Modèle non linéaire en cinétique chimique

En cinétique chimique on étudie souvent l'évolution de la concentration d'un composé  $B$  issu d'un composé  $A$  et redonnant lui-même un composé  $C$ . C'est une double réaction du premier ordre, irréversible, symbolisée par :  $A \rightarrow B \rightarrow C$ . Le modèle exponentiel à deux compartiments

$$\eta(t, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} [\exp(-\theta_2 t) - \exp(-\theta_1 t)] + \varepsilon \quad (2)$$

solution d'un système de deux équations différentielles linéaires du premier ordre, traduit bien cette évolution. On a :  $\eta(t, \theta)$  la concentration du composé  $B$  dépendant du temps  $t$  et des paramètres  $\theta_1$  et  $\theta_2$  constantes de vitesse des première et seconde décompositions, et les conditions initiales (concentrations)  $[A]_0 = 1, [B]_0 = 1, [C]_0 = 1$ . La question est de trouver les temps optimaux auxquels faire les mesures de vitesse pour aboutir à une estimation optimisée des paramètres.

## 3 Plans optimaux pour modèles linéaires

La théorie de l'optimalité en plans d'expériences se fonde historiquement dans le cadre du modèle linéaire – régression et analyse de la variance – et débute avec les travaux de Kiefer (1959, 1961, 1974), Kiefer et Wolfowitz (1959, 1960), Fedorov, (1969, 1972), Wynn (1970), sans oublier l'article pionnier et particulièrement pédagogique de Box et Lucas (1959). Comme on peut formuler le modèle d'analyse de variance comme un modèle de régression (voir §3.3), on se basera sur celui-ci pour exposer cette théorie.

### 3.1 Le modèle de régression linéaire

Pour exprimer la relation qui lie une réponse continue  $y$  dépendant de  $h$  facteurs explicatifs continus, contrôlables,  $x_j$ , et leurs transformées continues, on considère simultanément le modèle de régression linéaire :

$$y_{ik} = \mathbf{f}(\mathbf{x}_i)^T \beta + \varepsilon_{ik} \quad ; \quad i = 1, \dots, N_S \quad ; \quad k = 1, \dots, r_i \quad (3)$$

et le plan d'expériences discret associé :

$$\xi_{N, N_S, \{r\}} = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \dots & \mathbf{x}_i & \dots & \mathbf{x}_{N_S} \\ r_1 & \dots & r_i & \dots & r_{N_S} \\ r_i \geq 1, \forall i & ; & \sum_{i=1}^{N_S} r_i = N & & \end{array} \right\} \quad (4)$$

où (les matrices et les vecteurs sont notés en gras, et  $T$  désigne l'opérateur de transposition usuel) :

- $y_{ik}$  est la  $k^{\text{ième}}$  observation de la réponse collectée au  $i^{\text{ième}}$  point de support  $\mathbf{x}_i$  défini ci-dessous; on notera  $\mathbf{y}$  le vecteur  $(N \times 1)$  des  $y_{ik}$ ,  $N$  est le nombre total d'observations  $y_{ik}$  collectées sur les  $N_S$  points de support, selon le schéma de répétition  $\{r\} = \{r_1, \dots, r_{N_S}\}$ .
- $\mathbf{x}_i$  est un vecteur  $(h \times 1)$  dont les composantes sont les valeurs des  $h$  facteurs explicatifs  $x_j$ ,  $j = 1, \dots, h$ , c'est-à-dire  $\mathbf{x}_i = (x_{i1}, \dots, x_{ij}, \dots, x_{ih})^T$  pour l'expérience  $i$  (appelée aussi une condition expérimentale ou un traitement); quand l'indice de l'expérience ne sera pas précisé on notera simplement  $\mathbf{x}$ ; on a  $\mathbf{x}_i \in \Xi \subset \mathbb{R}^h$ ,  $\Xi$  étant le domaine expérimental, supposé sous-ensemble compact de  $\mathbb{R}^h$ ; dans les cas les plus courants  $\Xi$  est un hyper-rectangle, ou un polyèdre convexe, ou un ellipsoïde (d'excentricité 0 ou plus).
- $\mathbf{f}(\mathbf{x}_i)$  est un vecteur  $(p \times 1)$ ,  $p \geq h$ , dont les composantes  $f_j(\mathbf{x}_i)$  sont les valeurs des fonctions de régression  $f_j(x)$ , au point  $\mathbf{x}_i$ , avec  $j = 0, \dots, p - 1$ ; par exemple, pour le modèle :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{23} x_2 x_3 + \beta_4 \text{Log}(x_3) + \varepsilon \quad (5)$$

on a  $h = 3$  et  $p = 7$ , et les fonctions de régression sont :  $f_0(x) = 1$ ,  $f_1(x) = x_1$ ,  $f_2(x) = x_2$ ,  $f_3(x) = x_3$ ,  $f_4(x) = x_1^2$ ,  $f_5(x) = x_2 x_3$ ,  $f_6(x) = \text{Log}(x_3)$ .

La matrice  $\mathbf{X}_N$ , de dimensions  $(N \times p)$ , est appelée matrice du modèle; ses lignes sont formées par les vecteurs lignes  $\mathbf{f}(\mathbf{x}_i)^T$ ; les  $h$  colonnes de  $\mathbf{X}_N$  relatives aux  $h$  facteurs constituent une matrice  $(N \times h)$ , appelée matrice du plan (4); les  $p - h$  colonnes restantes sont formées par les transformées des  $h$  facteurs explicatifs ( $f_0(x)$ ,  $f_4(x)$ ,  $f_5(x)$ ,  $f_6(x)$  dans (5)).

- $\beta$  est un vecteur  $(p \times 1)$  des  $p$  paramètres (ou coefficients), inconnus, certains, du modèle.
- $\xi_{N, N_S, \{r\}}$  est un plan discret à  $N$  expériences appartenant à  $\Psi$  l'ensemble discret des plans d'expériences engendré par toutes les valeurs possibles de  $N_S$  tel que  $p \leq N_S \leq N_{\max}$ , et pour tous les schémas possibles de répétition  $\{r\}$ . En général  $N_{\max}$  est fixé a priori pour un problème donné (dépend du budget alloué).
- $\varepsilon_{ik}$  est l'erreur (scalaire) aléatoire attachée à l'observation  $y_{ik}$ ; c'est une erreur expérimentale; le vecteur des  $N$  erreurs  $\varepsilon_{ik}$  sera noté  $\varepsilon$ , d'espérance  $E(\varepsilon) = 0$ , et de matrice de variance inconnue (en général)  $\Sigma = \text{diag} \{\sigma_i^2, i = 1, \dots, N\}$ ; si on suppose que la variance est homogène on posera  $\sigma_i^2 = \sigma^2, \forall i$ ; on notera  $\mathbf{W} = \Sigma^{-1}$  la matrice de pondération.

**Remarque :** dans la pratique, en situation de plan d'expériences (où l'on ne dispose pas encore des observations  $y_{ik}$ ) il faut disposer d'estimations préalables et sans biais de  $\sigma^2$  ou des  $\sigma_i^2$ , calculées à partir de données antérieures.

Matriciellement on pourra écrire :

$$\mathbf{y} = \mathbf{X}_N \beta + \varepsilon$$

## 3.2 Estimation

Si la variance est homogène un estimateur usuellement utilisé dans ce contexte est l'estimateur des moindres carrés  $\hat{\beta}_{MC}$  obtenu en minimisant par rapport à  $\beta$  le critère

”somme des carrés des erreurs”

$$\begin{aligned} S_N(\beta) &= \sum_{i=1}^N \varepsilon_i^2 \\ &= (\mathbf{y} - \mathbf{X}_N \beta)^T (\mathbf{y} - \mathbf{X}_N \beta) \end{aligned}$$

On obtient facilement, sous réserve d’inversibilité de  $\mathbf{X}^T \mathbf{X}$  :

$$\hat{\beta}_{MC} = (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \mathbf{X}_N^T \mathbf{y}$$

et sa variance (minimale) :

$$\mathbf{V}(\hat{\beta}_{MC}) = \sigma^2 (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \quad (6)$$

Si la variance est hétérogène (les  $\sigma_i$  sont différents) on minimise la ”somme des carrés des erreurs pondérées” :

$$S_N(\beta) = (\mathbf{y} - \mathbf{X}_N \beta)^T \mathbf{W} (\mathbf{y} - \mathbf{X}_N \beta)$$

et la solution est l’estimateur des moindres carrés pondérés  $\hat{\beta}_{MCP}$  :

$$\hat{\beta}_{MCP} = (\mathbf{X}_N^T \mathbf{W} \mathbf{X}_N)^{-1} \mathbf{X}_N^T \mathbf{W} \mathbf{y}$$

de variance

$$\mathbf{V}(\hat{\beta}_{MC}) = (\mathbf{X}_N^T \mathbf{W} \mathbf{X}_N)^{-1}$$

Pour alléger l’écriture par la suite, mais sans perte de généralité, on supposera la variance homogène égale à  $\sigma^2$ ; on notera de façon simplifiée  $\hat{\beta}_{MC}$  par  $\hat{\beta}$ .

On appelle  $\mathbf{M}_N = \mathbf{X}_N^T \mathbf{X}_N$  la matrice d’information du plan (4), et  $\mathbf{M}_N^N = (1/N) \mathbf{M}_N$  la matrice des moments ou matrice d’information moyenne par unité pour un plan discret à  $N$  expériences; elle est normalisée par définition. Cette normalisation est essentielle puisqu’elle va autoriser la comparaison des matrices d’information de plans à nombres d’expériences différents. Par ailleurs, les observations étant supposées indépendantes, la propriété d’additivité nous autorise à écrire cette matrice d’information comme une somme pondérée de  $N_S$  matrices  $\mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T$  de rang 1 :

$$\mathbf{M}_N^N = \sum_{i=1}^{N_S} \frac{r_i}{N} \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T \quad (7)$$

### 3.3 Nature et transformation des facteurs explicatifs

Par définition, dans un modèle de régression les facteurs explicatifs doivent être de nature au moins quantitative, et sont soit discrets (par exemple un facteur température ne pouvant prendre que les niveaux discrets 0, 10, 20, ..., 100, sur une échelle de 0 à 100°C), soit continus. Néanmoins, on peut vouloir étudier l’effet de facteurs de nature intrinsèquement qualitative (donc discrète), par exemple l’influence de trois types de matériels. Par ailleurs, les effets des facteurs explicatifs (quantitatifs), et de leurs transformées (telles les interactions et les effets quadratiques) se traduisent par les valeurs des composantes  $\hat{\beta}_j$  de  $\hat{\beta}$ .

Deux problèmes se posent donc : le premier est de rendre tous les facteurs quantitatifs, et le second est de les exprimer dans des échelles telles que les effets  $\hat{\beta}_j$  soient tous

comparables entre eux. Pour résoudre le premier deux cas se présentent selon le nombre de modalités prises par le facteur qualitatif : si celui-ci présente deux modalités seulement, il suffira de les coder symétriquement autour de 0, par exemple -1 et +1; s'il existe  $q$  modalités ( $q > 2$ ) ce cas est plus délicat mais on peut utiliser les solutions proposées par Hardin et Sloane (1993). Pour le second problème, en supposant maintenant que tous nos facteurs sont quantitatifs, on doit rendre leurs échelles comparables. Ceci se fait facilement par une transformation au moyen des polynômes orthonormés (Kobilinsky, 1987). Par exemple, pour un facteur à quatre niveaux -3, -1, +1, +3 on obtiendra les niveaux respectifs  $-3/\sqrt{5}$ ,  $-1/\sqrt{5}$ ,  $+1/\sqrt{5}$ ,  $+3/\sqrt{5}$  pour coder sa transformée linéaire, puis les niveaux +1, -1, -1, +1 pour coder sa transformée quadratique, et enfin les niveaux  $-1/\sqrt{5}$ ,  $+3/\sqrt{5}$ ,  $-3/\sqrt{5}$ ,  $+1/\sqrt{5}$  pour coder sa transformée cubique. Les trois vecteurs correspondants sont tous de même norme (= 2) et orthogonaux entre eux.

### 3.4 Critère de $D$ -optimalité

Il existe aujourd'hui de nombreux critères d'optimalité. Il n'est pas question d'en dresser une liste exhaustive ici, mais uniquement de présenter les plus fréquemment utilisés. Pour être complet, il faudrait faire cette présentation dans le cadre de mesures continues (voir par exemple Silvey, 1980), mais dans le cadre de cet exposé on se contentera de les présenter dans le cadre de plans discrets comme définis en (4).

#### 3.4.1 Définition

Un plan d'expériences  $D$ -optimal ( $D$  comme déterminant) à  $N$  expériences minimise le déterminant de la matrice de variance de  $\hat{\beta}$ , ou de façon équivalente (ce qui est numériquement plus léger) maximise le déterminant de la matrice d'information  $\mathbf{M}_N$  :

$$\xi_{N, N_S^*, \{r^*\}}^D = Arg \left\{ \max_{\xi_N \in \Psi} \det(\mathbf{M}_N(\xi_{N, N_S, \{r\}})) \right\} \quad (8)$$

On remarque dans cette définition que pour un nombre total fixe d'expériences  $N$ , le problème de maximisation prend en compte aussi la recherche des valeurs optimales  $N_S^*, \{r^*\}$ . C'est un problème d'optimisation multiple. On peut aussi se contenter de rechercher le maximum pour  $N_S$  et  $\{r\}$  fixés, ou  $N_S$  fixé seulement, selon le type de problème expérimental à résoudre.

#### 3.4.2 Exemple

Pour le modèle de droite de régression  $y = \beta_0 + \beta_1 x + \varepsilon$ , pour lequel on suppose  $\sigma^2$  connue, on souhaite estimer au mieux (variances petites et covariance nulle) les paramètres  $\beta_0$  et  $\beta_1$ , à partir d'observations  $y_i$  recueillies en  $x \in [-1, +1]$ . Dans ce but on envisage quatre plans discrets que l'on donne dans le tableau I ainsi que leurs déterminants bruts et normés. On remarque que  $\det \mathbf{M}_N^N$  est bien un déterminant normé car :

$$\det \mathbf{M}_N^N = \det\left(\frac{1}{N} \mathbf{M}_N\right) = \det\left(\frac{1}{N} \mathbf{X}_N^T \mathbf{X}_N\right) = N^{-p} \det(\mathbf{X}_N^T \mathbf{X}_N) = N^{-p} \det(\mathbf{M}_N)$$

$\xi_{3,3,\{r\}}^1 = \begin{Bmatrix} -1 & 0 & +1 \\ 1 & 1 & 1 \end{Bmatrix}$	$\longrightarrow \det \mathbf{M}_N^{(1)} = 6 ; \det \mathbf{M}_N^{N(1)} = 0.67$
$\xi_{3,3,\{r\}}^2 = \begin{Bmatrix} -1 & -1 & +1 \\ 1 & 1 & 1 \end{Bmatrix}$	$\longrightarrow \det \mathbf{M}_N^{(2)} = 8 ; \det \mathbf{M}_N^{N(2)} = 0.89$
$\xi_{2,2,\{r\}}^3 = \begin{Bmatrix} -1 & +1 \\ 1 & 1 \end{Bmatrix}$	$\longrightarrow \det \mathbf{M}_N^{(3)} = 4 ; \det \mathbf{M}_N^{N(3)} = 1$
$\xi_{4,2,\{r\}}^4 = \begin{Bmatrix} -1 & +1 \\ 2 & 2 \end{Bmatrix}$	$\longrightarrow \det \mathbf{M}_N^{(4)} = 16 ; \det \mathbf{M}_N^{N(4)} = 1$

Tableau I : Quatre plans discrets pour estimer une droite de régression, et les déterminants bruts et normés correspondants.

A l'examen du tableau I le plan  $\xi_{4,2}^4$  apparaît très nettement comme le  $D$ -meilleur plan. Si on compare les déterminants normés, il est équivalent au plan  $\xi_{2,2}^3$ , c'est-à-dire que l'information (au sens de la  $D$ -optimalité) moyenne par expérience est équivalente, mais bien sûr le plan  $\xi_{4,2}^4$  apportera globalement plus d'information que le plan  $\xi_{2,2}^3$ . En outre, on remarque que les plans  $\xi_{3,3}^1$  et  $\xi_{3,3}^2$  apportent plus d'information (globalement) que le plan  $\xi_{2,2}^3$  mais l'efficacité de chacune des expériences de ces deux plans est moins bonne compte tenu des valeurs prises par leurs déterminants normés. Typiquement, l'expérience "centrale" du plan  $\xi_{3,3}^1$  est très peu performante vis-à-vis de notre objectif. Enfin, on rappelle qu'un plan orthogonal est tel que  $\mathbf{X}_N^T \mathbf{X}_N = N \mathbf{I}_N$  où  $\mathbf{I}_N$  est la matrice identité  $N \times N$ . On vérifie ainsi facilement que les plans  $\xi_{4,2}^3$  et  $\xi_{2,2}^4$  sont orthogonaux. En résumé, le plan  $\xi_{4,2}^4$  est très supérieur aux autres au sens de la  $D$ -optimalité : la formule (6) nous indique que les estimateurs de la pente et de l'ordonnée à l'origine auront une covariance nulle ; en outre, pour  $N = 4$ , ils seront de variance minimale.

### 3.4.3 $D$ -efficacité

On définit la  $D$ -efficacité d'un plan discret comme le pourcentage :

$$D_{eff} = 100 \times \left[ \frac{\det(\mathbf{M}_N^N)}{\det(\mathbf{M}(\pi_D))} \right]^{\frac{1}{p}} \quad (9)$$

où  $\mathbf{M}(\pi_D)$  est la matrice d'information d'un plan  $D$ -optimal à mesure continue qu'il nous faut définir succinctement maintenant, cette efficacité présentant un grand intérêt pratique dès lors que l'on sait calculer un plan  $D$ -optimal à mesure continue.

### 3.4.4 Plan $D$ -optimal à mesure continue

Cette notion est une idée majeure et originale due à Kiefer. Il considère le plan d'expériences comme une *mesure de probabilité continue* sur un domaine expérimental compact, on parle de plan à mesure continue ("design measure"). C'est donc un objet mathématique plutôt qu'un outil opérationnel pour faire des expériences réelles. Mais cet outil va s'avérer être très fructueux quand on va le compléter avec la *propriété de convexité* de l'ensemble des matrices d'information continues correspondantes. En effet, un des résultats importants sera que le déterminant ou son logarithme (fonction concave) d'une matrice d'information continue atteint un maximum global sur l'ensemble de définition, convexe, constitué des matrices d'information continues. La mesure correspondante à ce maximum est appelée une mesure  $D$ -optimale, notée ici  $\pi_D$ , et  $\mathbf{M}(\pi_D)$  est



la matrice d'information de cette mesure  $D$ -optimale. On comprend maintenant la pertinence et l'utilité de la  $D$ -efficacité définie en (9). Des algorithmes spécifiques (Fedorov, 1969, 1972, Wynn, 1970) ont été proposés pour calculer ce type de mesure et en déduire des approximations discrètes. En outre, il a été démontré que si on souhaite discrétiser la mesure  $\pi_D$  alors le maximum de points de support de la mesure discrète, donc du support d'un plan  $D$ -optimal discret, est  $N_{S_{\max}} = p(p + 1)/2$ .

### 3.4.5 Répétition des expériences

Dans le cas de plans discrets, pour  $N$  fixé, la valeur de  $\det(\mathbf{M}_N)$  dépend conjointement des quatre éléments suivants : la position et le nombre des  $N_S$  points de support, le nombre de répétitions et leur répartition sur ces  $N_S$  points de support. Illustrons ce phénomène avec l'exemple très simple proposé par Atkinson et Hunter (1968). Soit le modèle linéaire :

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (10)$$

avec  $x_1$  et  $x_2$  continus, définis sur  $[0, 1]$ , variance des erreurs constante. On peut vérifier facilement que tous les plans discrets de la forme :

$$\xi_{2,2,\{r\}} = \left\{ \begin{array}{cc} (x_1 = 0; x_2 = 1) & (x_1 = 1; x_2 = \alpha) \\ 1 & 1 \end{array} \right\} \quad (11)$$

avec  $0 \leq \alpha \leq 1$  sont globalement  $D$ -optimaux avec un déterminant  $\det(\mathbf{M}_2) = 1$ . Pour  $N = 6$ , la triple répétition de l'un quelconque de ces plans (11) conduit à  $\det[\mathbf{M}_6(\xi_{2,2,\{3,3,3\}})] = 9$  alors que le plan suivant constitué de la répétition d'un plan à trois points de support

$$\xi_{6,3,\{r\}} = \left\{ \begin{array}{ccc} (x_1 = 0; x_2 = 1) & (x_1 = 1; x_2 = 0) & (x_1 = 1; x_2 = 1) \\ 2 & 2 & 2 \end{array} \right\} \quad (12)$$

conduit à  $\det[\mathbf{M}_6(\xi_{6,3,\{2,2,2\}})] = 12$ , ce qui est nettement  $D$ -meilleur.

Il apparaît donc, comme annoncé plus haut, que la valeur du déterminant est liée de façon complexe à la structure du plan. Néanmoins, dans certaines conditions l'optimalité consiste à faire une répartition équitable, c'est-à-dire une répartition telle que l'écart entre les nombres de répétition ne diffèrent pas plus d'une unité, sur un support à nombre de points égal au nombre de paramètres du modèle ( $N_S = p$ ). Notamment, il faut tout d'abord déterminer le meilleur plan (globalement)  $D$ -optimal à  $N = N_S = p$  points (le plan  $D_p$ -optimal), et ensuite vérifier si la condition nécessaire et suffisante de Vila (1991) est satisfaite. Alors, si on note  $D_N$  le plan  $N$  expériences, tout plan discret globalement  $D_N$ -optimal, avec  $N$  multiple de  $p$  (nombre de paramètres du modèle), consiste en  $N/p$  répétitions du plan globalement  $D_p$ -optimal.

### 3.4.6 Construction algorithmique des plans $D$ -optimaux

Comme les plans  $D$ -optimaux sont largement répandus et sont des outils efficaces, on montre maintenant quelques algorithmes très utilisés pour les construire. On a vu plus haut la distinction de nature entre plans discrets et continus, les algorithmes respectifs sont différents aussi.

**Construction des plans  $D$ -optimaux discrets** Il existe maintenant de nombreux algorithmes pour construire des plans optimaux, mais certains des algorithmes les plus récents (par exemple basés sur le recuit simulé, ou les algorithmes génétiques) ne montrent pas une vitesse significativement plus grande (et souvent même inférieure en ce qui concerne le recuit simulé) que les algorithmes pionniers, dits d'échange, de Fedorov (1969, 1972) et de Mitchell (1974). Aussi, on exposera ici l'algorithme à échange double de Fedorov qui a le mérite d'être facile à programmer et à implémenter dans tout langage usuel. L'objectif est de sélectionner  $N$  points parmi un ensemble de  $N_C$  points candidats, ensemble défini a priori (typiquement par les noeuds d'une grille à maillage serré sur  $\Xi$ ), tels que le  $N$ -plan ait le déterminant normé de sa matrice d'information  $\mathbf{M}_N$  le plus grand possible parmi les déterminants de tous les  $N$ -plans possibles. On atteint cet objectif avec l'algorithme itératif à échange double de Fedorov, qui ne présente pas de preuve formelle de convergence, mais qui dans la pratique converge très souvent vers le maximum global, maximum que l'on sait calculer par ailleurs (voir plus loin la construction des plans continus).

*Algorithme d'échange double de Fedorov*

- *étape 1* : On choisit *aléatoirement* (ou selon un choix circonstancié)  $N$  expériences parmi les  $N_C$  pour constituer un  $N$ -plan d'expériences initial non dégénéré ( $\det(\mathbf{M}_N^N) \neq 0$ ) et tel que  $N = N_S \geq p$ ,
- *étape 2* : On souhaite échanger une expérience de ce plan initial par une expérience de l'ensemble candidat dans le but d'une augmentation maximale du déterminant de  $\mathbf{M}_N^N$ . La particularité de cet algorithme est que cet échange se fait *en un seul coup*. Si on désigne par  $i$  une expérience du plan initial et  $j$  une expérience de l'ensemble candidat on détermine le couple  $(i, j)$  le plus performant dans cette augmentation en calculant pour les  $N \times (N_C - 1)$  couples possibles la valeur de  $\Delta_F = \det(\mathbf{M}_N^N)$ . Si plusieurs couples distincts provoquent des augmentations comparables de  $\Delta_F$  (à un écart  $\varepsilon$  près, fixé au préalable) on choisit l'un de ces couples aléatoirement. Notons  $(i_{\max}, j_{\max})$  le meilleur couple.
- *étape 3* : On réalise effectivement l'échange de  $i_{\max}$  par  $j_{\max}$  correspondant à  $\Delta_{F \max}$  et on retourne à l'étape 2 si le critère d'arrêt n'est pas satisfait.

Pour calculer le déterminant de l'itération courante Fedorov s'appuie sur le théorème suivant.

**Théorème :**

*Après l'échange de l'expérience  $i$  par l'expérience  $j$  la nouvelle matrice d'information (à l'itération  $t + 1$ ) s'exprime en fonction de la matrice d'information de l'itération  $t$  comme suit :*

$$(\mathbf{X}_N^T \mathbf{X}_N)_{[t+1]} = (\mathbf{X}_N^T \mathbf{X}_N)_{[t]} - \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T + \mathbf{f}(\mathbf{x}_j) \mathbf{f}(\mathbf{x}_j)^T$$

*Et le nouveau déterminant est lié au précédent par :*

$$\det(\mathbf{X}_N^T \mathbf{X}_N)_{[t+1]} = \det(\mathbf{X}_N^T \mathbf{X}_N)_{[t]} \times [1 + \Delta(i, j)]$$

*avec :*

$$\Delta(i, j) = d(\mathbf{x}_j) - [d(\mathbf{x}_i) d(\mathbf{x}_j) - d^2(\mathbf{x}_i, \mathbf{x}_j)] - d(\mathbf{x}_i)$$

où :

$$\begin{aligned} d(\mathbf{x}_i) &= \mathbf{f}(\mathbf{x}_i)^T (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \mathbf{f}(\mathbf{x}_i) \\ d(\mathbf{x}_j) &= \mathbf{f}(\mathbf{x}_j)^T (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \mathbf{f}(\mathbf{x}_j) \\ d(\mathbf{x}_i, \mathbf{x}_j) &= \mathbf{f}(\mathbf{x}_i)^T (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \mathbf{f}(\mathbf{x}_j) \quad \text{fonction de covariance} \\ &= \mathbf{f}(\mathbf{x}_j)^T (\mathbf{X}_N^T \mathbf{X}_N)^{-1} \mathbf{f}(\mathbf{x}_i) \end{aligned}$$

Plusieurs tests d'arrêt sont possibles, Mathieu (1981) propose de stopper les itérations quand  $\Delta(i, j)$  devient inférieur à  $\varepsilon$ , valeur faible choisie a priori, et surtout montre comment les calculs peuvent être réduits de 25% à 75% suivant les cas. Enfin, on peut imposer à l'algorithme la possibilité de répéter ou non des expériences déjà sélectionnées. Comme on ne connaît pas a priori la meilleure valeur de  $N$ , on relance l'algorithme pour plusieurs valeurs de  $N$  comprises entre  $N = p$  et  $N_{\max} < N_C$ . En traçant ensuite le graphique de l'évolution du déterminant normé de  $\mathbf{M}_N$ , en fonction de ce nombre  $N$  variable, on peut repérer sur ce graphique des pics correspondants à des plans  $D$ -optimaux performants (voir l'illustration plus loin).

**Construction des plans  $D$ -optimaux continus** On se contentera d'exposer le principe de l'algorithme de Torsney (1988), algorithme rapide, encore accéléré grâce à la proposition de Pronzato (2004). Comme l'algorithme précédent on construit d'abord un ensemble de points candidats (points de support des expériences candidates) formé par les noeuds d'une grille à maillage très serré sur  $\Xi$ . Cet ensemble est le support discret d'une mesure uniforme de masse  $\pi^{[t]} = 1/N^{[t]}$  en chacun des points  $\mathbf{x}_i$ , avec  $\pi^{[t=0]} = 1/N_C$ . Soit :

$$d(\mathbf{x}_i, \pi^{[t]}) = \mathbf{f}(\mathbf{x}_i)^T \frac{1}{N^{[t]}} \mathbf{M}_{N^{[t]}}^{-1} \mathbf{f}(\mathbf{x}_i)$$

On calcule  $d(\mathbf{x}_i, \pi^{[t]})$  pour chaque  $\mathbf{x}_i$ , et on met à jour la nouvelle mesure par :

$$\pi_i^{[t+1]} = \pi_i^{[t]} \frac{d(\mathbf{x}_i, \pi_i^{[t]})}{p}, \quad i = 1, \dots, N^{[t]}$$

Si une mesure devient négligeable en un point  $\mathbf{x}_k$  à l'itération  $t = \tau$  alors on pose  $\pi_k^{[t]} = 0$ ,  $\forall t > \tau$ , et  $N^{[t+1]} = N^{[t]} - 1$ . Si pour tout point  $\mathbf{x}_u$  de mesure non négligeable on a  $d(\mathbf{x}_u, \pi_u^{[t]}) - p < \varepsilon$ ,  $\varepsilon$  étant un petit nombre réel positif satisfaisant à la précision souhaitée, alors on stoppe l'algorithme. La convergence (monotone) vers une mesure  $D$ -optimale  $\pi_D$  est démontrée. Si le maillage est très serré le déterminant correspondant de  $\mathbf{M}(\pi_D)$  est une bonne approximation du maximum global qui peut être atteint par ce déterminant.

**Illustration** Soit le modèle (1) et le domaine  $\Xi$  sous contraintes, polygonal et convexe dont les coordonnées des sommets et des milieux d'arêtes apparaissent au tableau II. On suppose que ces  $N_C = 17$  points forment un ensemble candidat de bonne qualité.

n° point de support	$X_1$	$X_2$
1	0.0	1.0
2	0.5	0.6
3	1.0	0.2
4	1.0	0.0
5	1.0	-0.2
6	0.9	-0.6
7	0.8	-1.0
8	0.2	-1.0
9	0.0	-1.0
10	-0.5	-0.9
11	-1.0	-0.8
12	-1.0	-0.2
13	-1.0	0.4
14	-0.9	0.7
15	-0.6	1.0
16	-0.3	1.0
17	0.0	0.0

Tableau II : Illustration : coordonnées des sommets et des milieux d'arêtes de  $\Xi$ . Il apparaît aussi le point  $(0, 0)$ ; ces  $N_C = 17$  points forment le réseau candidat.

Un algorithme d'échanges usuel (tel celui explicité plus haut) trouve en quelques secondes deux plans  $D$ -optimaux discrets correspondant à des pics du graphique obtenu comme expliqué plus haut :

$$\xi_{6,6,\{r\}}^D = \left\{ \begin{array}{l} n^\circ \text{ des points de support} \longrightarrow 1 \quad 3 \quad 7 \quad 11 \quad 14 \quad 17 \\ \{r\} \longrightarrow \quad \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \end{array} \right\}$$

$$\det(\mathbf{M}_6^6) = 0.001502$$

et :

$$\xi_{14,8,\{r\}}^D = \left\{ \begin{array}{l} n^\circ \text{ des points de support} \longrightarrow 1 \quad 3 \quad 7 \quad 9 \quad 11 \quad 13 \quad 15 \quad 17 \\ \{r\} \longrightarrow \quad \quad 2 \quad 2 \quad 2 \quad 1 \quad 2 \quad 2 \quad 1 \quad 2 \end{array} \right\}$$

$$\det(\mathbf{M}_{14}^{14}) = 0.001603$$

On remarque avec le second plan que l'on résout simultanément le problème d'optimisation multiple évoqué au §3.4.1, à savoir, pour  $N$  fixé, de trouver  $N_S^*$  et  $\{r^*\}$  en même temps que le maximum du déterminant. Puis l'algorithme de Torsney nous donne une approximation raisonnable ( $N_C$  ne vaut que 17, mais on pourrait montrer que le réseau candidat considéré est de bonne qualité relativement au modèle postulé) du maximum global correspondant à une mesure  $\pi_D$ , soit 0.001637. On peut maintenant qualifier ces deux plans par la  $D$ -efficacité calculée avec la formule (9). On trouve 98.6% pour le premier et 99.6% pour le second. Les deux plans obtenus sont de qualité suffisante pour être réellement mis en oeuvre au laboratoire. On préférera le second si le budget disponible l'autorise car les répétitions le rendent plus robuste à une erreur de manipulation. Toutefois, si le modèle postulé (1) n'est pas un bon modèle pour représenter le phénomène étudié ces plans  $D$ -optimaux ne sont pas robustes vis-à-vis de l'erreur de modèle. Pour se protéger de ce risque on conseille alors de s'orienter vers des plans construits à partir du critère  $J$  qui fait l'objet du paragraphe 3.10.

### 3.4.7 Extensions de la $D$ -optimalité

Parmi les nombreuses extensions de la  $D$ -optimalité citons les deux extensions suivantes assez largement utilisées :

**Critère de  $D_A$ -optimalité :** Sibson (1974) propose de généraliser le critère de  $D$ -optimalité si on s'intéresse à des combinaisons linéaires particulières des estimateurs. Si  $\mathbf{A}$  est la matrice ( $s \times p$ ), de rang  $s < p$ , des coefficients des combinaisons linéaires on a

$$\xi_{N,N_S^*,\{r^*\}}^{D_A} = Arg \left\{ \min_{\xi_N \in \Psi} \det(\mathbf{A}'\mathbf{M}_N^{-1}\mathbf{A}) \right\} \quad (13)$$

**Critère de  $D_S$ -optimalité :** Il arrive parfois que l'on attache plus d'importance à certains paramètres parmi l'ensemble des paramètres du modèle,  $s$  paramètres d'intérêt majeur parmi les  $p$ . Les  $r = p - s$  sont par exemple des paramètres de nuisance. Un critère d'optimalité peut se construire par restriction au sous-ensemble des  $s$  paramètres. La  $D_S$ -optimalité répond à ce but (Karlin et Studden, 1966, Silvey, 1980). On trouvera une application de ce critère dans Gauchi (2005).

### 3.5 Critère de $A$ -optimalité

Le critère  $A$  vise à minimiser la trace de  $\mathbf{M}_N^{-1}$  :

$$\xi_{N,N_S^*,\{r^*\}}^A = Arg \left\{ \min_{\xi_N \in \Psi} [Tr(\mathbf{M}_N^{-1})] \right\} \quad (14)$$

et donc minimiser ce critère équivaut à minimiser la somme des variances des estimateurs. Numériquement, le calcul d'un plan  $A$ -optimal est plus coûteux que celui d'un plan  $D$ -optimal puisqu'il faut inverser  $\mathbf{M}_N$ . Enfin, notons qu'un plan  $A$ -optimal dépend des unités des variables explicatives, ce qui est un handicap supplémentaire à l'utilisation de ce critère.

### 3.6 Critère de $E$ -optimalité

Soit  $\lambda_1, \dots, \lambda_p$  les valeurs propres de la matrice  $\mathbf{M}_N$  associée au plan  $\xi_N$ . Les valeurs propres de  $\mathbf{M}_N^{-1}$  sont  $1/\lambda_1, \dots, 1/\lambda_p$ . Un plan  $E$ -optimal à  $N$  expériences vise à minimiser la valeur propre maximale de  $\mathbf{M}_N^{-1}$ . C'est un critère de type min-max. On écrit :

$$\xi_N^E = Arg \left\{ \min_{\xi_N \in \Xi} [\max_i (1/\lambda_i)] \right\} \quad (15)$$

D'après certains auteurs la  $E$ -optimalité semble impliquer un étalement des expériences sur l'espace expérimental, contrairement à la  $D$ -optimalité. Remarquons enfin que la détermination d'un plan  $E$ -optimal est une tâche numériquement lourde, puisqu'elle nécessite le calcul répété des valeurs propres de  $\mathbf{M}_N$ ; ce peut être une raison de sa faible utilisation. En outre, il n'existe pas de résultats théoriques sur l'optimalité des répétitions d'expériences analogues à celui évoqué pour le critère  $D$ -optimalité.

### 3.7 Interprétation géométrique des critères $D$ , $A$ et $E$

Si  $\varepsilon \sim N(\mathbf{0}, \Sigma)$  on sait que le domaine de confiance de niveau  $\alpha$  pour  $\beta$  est un hyperellipsoïde centré sur  $\hat{\beta}$  dont la frontière est définie par l'inégalité :

$$(\beta - \hat{\beta})^T \mathbf{M}_N (\beta - \hat{\beta}) \leq \rho^2$$

où  $\rho^2$  est égal à :

- $ps^2 F_\alpha(p, N-p)$  si la variance expérimentale  $\sigma^2$  est inconnue et que  $s^2$  en est une estimation indépendante à  $N-p$  ddl ;  $F_\alpha(p, N-p)$  est le fractile à  $\alpha\%$  de la distribution  $\mathcal{F}$  de Fisher-Snedecor à  $p$  et  $N-p$  degrés de liberté,
- $\sigma^2 \chi_\alpha^2(p)$  sinon, avec  $\chi_\alpha^2(p)$  le fractile à  $\alpha\%$  de la distribution du Khi-2 à  $p$  degrés de liberté.

Comme  $\mathbf{M}_N$  intervient dans cette définition il apparaît ainsi que les critères d'optimalité définis plus haut dépendent de cet ellipsoïde relativement à ses caractéristiques de volume, de forme et d'orientation. Ces critères d'optimalité prennent alors une signification très naturelle.

### 3.8 Critère de $G$ -optimalité

#### 3.8.1 Définition

Plutôt que s'intéresser directement à la précision de l'estimateur  $\hat{\beta}$  on peut souhaiter un modèle avec lequel les prédictions elles-mêmes seront les plus précises possible. Le critère de  $G$ -optimalité vise à minimiser le maximum de la variance de la réponse prédite. Si on appelle  $d(\mathbf{x}_i) = \mathbf{f}(\mathbf{x}_i)^T \mathbf{M}_N^{-1} \mathbf{f}(\mathbf{x}_i)$  la fonction de variance au point  $\mathbf{x}_i$ , on a  $var(\hat{y}(\mathbf{x}_i)) = \sigma^2 d(\mathbf{x}_i)$ . Un plan  $G$ -optimal discret à  $N$  expériences est défini par :

$$\xi_N^G = Arg \left\{ \min_{\xi_N \in \Psi} \left\{ \max_{\mathbf{x}_i \in \Xi} d(\mathbf{x}_i) \right\} \right\} \quad (16)$$

#### 3.8.2 Exemple

Pour le plan  $\xi_{3,3}^1$  du §3.4.2 on a (à  $\sigma^2$  près) :  $var(\hat{y}(\mathbf{x})) = \frac{1}{3} + \frac{1}{2}x^2$  et pour le plan  $\xi_{3,3}^2$  on a  $var(\hat{y}(\mathbf{x})) = \frac{1}{8}(3 + 2x + 3x^2)$ . Aucun de ces deux plans n'est  $G$ -meilleur que l'autre pour toutes les valeurs de  $x \in [-1, +1]$ . Mais, sur le même intervalle, pour le premier plan on a  $\max\{var(\hat{y}(\mathbf{x}))\} = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}$  tandis que pour le second on a  $\max\{var(\hat{y}(\mathbf{x}))\} = \frac{1}{8}(3 + 2 + 3) = 1$ . Donc le plan  $\xi_{3,3}^1$  est  $G$ -meilleur que le plan  $\xi_{3,3}^2$ . On voit ainsi qu'un plan discret  $D$ -optimal n'est pas forcément  $G$ -optimal sur le même domaine. En revanche, on verra ci-dessous qu'un plan continu  $D$ -optimal est aussi  $G$ -optimal.

### 3.9 Equivalence entre $G$ - et $D$ -optimalités continues

#### 3.9.1 Le TEG, un théorème fondamental

Cette équivalence, dans le cas continu uniquement, s'appuie sur le théorème fondamental d'équivalence générale (TEG) de Kiefer et Wolfowitz (1960), énoncé pour un modèle linéaire. Il a été étendu par la suite (Whittle, 1973) à tout critère différentiable et convexe. Il est explicité dans Gauchi (1997). Ce théorème montre, notamment, qu'un

plan  $D$ -optimal continu est aussi un plan  $G$ -optimal continu et inversement, d'une part, et d'autre part que  $[d(\mathbf{x}_i, \pi_D)] = p$ ,  $i = 1, \dots, N_S$ , c'est-à-dire que la fonction de variance vaut toujours  $p$  en tous les points de support  $\mathbf{x}_i$  où la mesure continue  $D$ -optimale  $\pi_D$  est discrétisée. Nous donnons maintenant deux exemples très simples d'utilisation du TEG.

*Exemple 1 : un plan  $G$ -optimal pour la régression simple*

Soit le modèle :

$$E(y_x) = \beta_0 + \beta_1 x \quad , \quad x \in [-1, +1]$$

On souhaite réaliser 6 expériences (c'est-à-dire recueillir 6 observations). Dans ce but on propose le plan discret  $\xi_{6,2,\{3,3\}}$  pour lequel les points de support sont  $-1$  et  $+1$ . Ce plan est  $D$ - et  $G$ -optimal car par la formule (7) :

$$\mathbf{M}_6^6 = \frac{3}{6} \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix} + \frac{3}{6} \begin{pmatrix} +1 & +1 \\ +1 & +1 \end{pmatrix} = \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix}$$

La fonction de variance au point-support  $-1$  s'écrit :

$$d(-1) = \begin{pmatrix} +1 & -1 \end{pmatrix} \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix} \begin{pmatrix} +1 \\ -1 \end{pmatrix} = 2$$

et au point-support  $+1$  :

$$d(+1) = \begin{pmatrix} +1 & +1 \end{pmatrix} \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix} \begin{pmatrix} +1 \\ +1 \end{pmatrix} = 2$$

Donc :

$$\max_{-1 \leq x \leq 1} \{d(x, \xi_{6,2,\{3,3\}})\} = \max_{-1 \leq x \leq 1} (1 + x^2) = 2 = p$$

En chaque point-support du plan la fonction de variance vaut  $p$ .

*Exemple 2 : un plan  $G$ -optimal pour la régression quadratique*

Soit le modèle :

$$E(y_x) = \beta_0 + \beta_1 x + \beta_{11} x^2 \quad , \quad x \in [-1, +1]$$

Le plan discret :

$$\xi_{9,3,\{r\}} = \left\{ \begin{array}{ccc} -1 & 0 & +1 \\ 3 & 3 & 3 \end{array} \right\}$$

est  $D$ - et  $G$ -optimal car :

$$\begin{aligned} \mathbf{M}_9^9 &= \frac{3}{9} \begin{pmatrix} +1 & -1 & +1 \\ -1 & +1 & -1 \\ +1 & -1 & +1 \end{pmatrix} + \frac{3}{9} \begin{pmatrix} +1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{3}{9} \begin{pmatrix} +1 & +1 & +1 \\ +1 & +1 & +1 \\ +1 & +1 & +1 \end{pmatrix} \\ &= \begin{pmatrix} +1 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/3 \end{pmatrix} \end{aligned}$$

et :

$$(\mathbf{M}_9^9)^{-1} = \frac{3}{4} \begin{pmatrix} +4 & 0 & -4 \\ 0 & +2 & 0 \\ -4 & 0 & +6 \end{pmatrix}$$

d'où :

$$\max_{-1 \leq x \leq 1} \{d(x, \xi_{9,3,\{3,3,3\}})\} = \max_{-1 \leq x \leq 1} \left\{ \frac{3}{4}(4 - 6x^2 + 6x^4) \right\} = 3 = p$$

et donc :

$$d(-1) = 3 ; d(0) = 3 ; d(+1) = 3$$

On peut donc vérifier facilement si un plan discret présente simultanément les propriétés de  $D$ - et  $G$ -optimalité en calculant sa fonction de variance  $d(\mathbf{x}_i)$  en chaque point de son support.

## 3.10 Critère de $J$ -optimalité

### 3.10.1 Définition

Le critère  $J$  – critère appelé IMSE en anglais pour "Integrated Mean Square Error" – est connu en français sous le nom de critère de "variance+biais". Il a été proposé par Box et Draper (1959). On le trouvera largement explicité dans Khuri et Cornell (1987). L'originalité de ce critère par rapport aux critères déjà vus plus haut est qu'il prend en compte à la fois la variance de la prédiction de la réponse et le biais induit sur les estimateurs du modèle postulé quand celui-ci n'est pas forcément le "bon" modèle. En général, le "bon" modèle est un sur-modèle, c'est-à-dire un modèle englobant le modèle postulé et qui contient des termes supplémentaires. On cherche à se "protéger" contre ce sur-modèle, c'est-à-dire à minimiser l'impact de l'omission des termes supplémentaires sur la qualité des estimateurs obtenus avec le "faux" modèle postulé. Pour présenter plus facilement ce critère, mais sans sacrifier à l'esprit pédagogique de l'article, on se placera dans un cadre continu et on utilisera donc une mesure de référence  $\pi \in \mathcal{D}$  ensemble des mesures possibles sur  $\Xi$  (de dimension  $h$ ).

Soit  $\mathbf{x}$  le vecteur dont les composantes sont les  $h$  facteurs explicatifs continus. Posons :

- $y(\mathbf{x}) = \mathbf{f}_1(\mathbf{x})^T \beta_1 + \varepsilon$ , le modèle postulé pour la réponse  $y$ , où  $\mathbf{f}_1(\mathbf{x})^T$  est le vecteur ligne ( $1 \times p_1$ ) des  $p_1$  fonctions de régression de  $\mathbf{x}$ , et  $\beta_1$  le vecteur des  $p_1$  paramètres à estimer,
- $y(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon$ , le "vrai" modèle (ou sur-modèle) avec  $\eta(\mathbf{x}) = \mathbf{f}_1(\mathbf{x})^T \beta_1 + \mathbf{f}_2(\mathbf{x})^T \beta_2$ , où  $\mathbf{f}_2(\mathbf{x})^T$  est le vecteur ligne ( $1 \times p_2$ ) des  $p_2$  fonctions de régression de  $\mathbf{x}$  "oubliées", et  $\beta_2$  le vecteur des  $p_2$  paramètres correspondants à estimer.

Le critère "variance +biais" consiste à minimiser par rapport à la mesure  $\pi$  la quantité suivante :

$$J = \frac{ND}{\sigma^2} \int_{\Xi} E [\hat{y}(\mathbf{x}) - \eta(\mathbf{x})]^2 d\pi(\mathbf{x})$$

avec :

- $N$  = nombre total d'expériences du plan
- $\mathcal{D}^{-1} = \int_{\Xi} d\pi(\mathbf{x})$ ,
- $\hat{y}(\mathbf{x})$  : la prédiction du modèle postulé,

Matriciellement posons :

- $\mathbf{y} = \mathbf{X}_{N,1}\beta_1 + \varepsilon$
- $\hat{\mathbf{y}} = \mathbf{X}_{N,1}\hat{\beta}_1$  la prédiction du modèle postulé où  $\hat{\beta}_1 = (\mathbf{X}_{N,1}^T \mathbf{X}_{N,1})^{-1} \mathbf{X}_{N,1}^T \mathbf{y}$ ,
- $\eta(\mathbf{x}) = \mathbf{X}_{N,1}\beta_1 + \mathbf{X}_{N,2}\beta_2$



L'estimation  $\check{\beta}_1$  de  $\beta_1$  par le "vrai" modèle est telle que :

$$E\left(\check{\beta}_1\right) = \beta_1 + (\mathbf{X}_{N,1}^T \mathbf{X}_{N,1})^{-1} \mathbf{X}_{N,1}^T \mathbf{X}_{N,2} \beta_2 = \beta_1 + \mathbf{A} \beta_2$$

avec  $\mathbf{A} = (\mathbf{X}_{N,1}^T \mathbf{X}_{N,1})^{-1} \mathbf{X}_{N,1}^T \mathbf{X}_{N,2}$  appelée matrice d'alias.

### 3.10.2 Décomposition du critère $J$

On peut le décomposer en deux termes :  $J = V + B$  où :

$$V = \frac{N\mathcal{D}}{\sigma^2} \int_{\Xi} \text{var}(\hat{y}(\mathbf{x})) d\pi(\mathbf{x})$$

est la variance de la prédiction intégrée et normalisée et :

$$B = \frac{N\mathcal{D}}{\sigma^2} \int_{\Xi} [E(\hat{y}(\mathbf{x})) - \eta(\mathbf{x})]^2 d\pi(\mathbf{x})$$

est le carré du biais intégré sur  $\Xi'$  et normalisé.

Si on désigne la matrice des moments du plan par  $\mathbf{M}_{N,X_1} = N^{-1} \mathbf{X}_{N,1}^T \mathbf{X}_{N,1}$  on peut exprimer  $V$  sous une forme plus concise en utilisant  $\text{var}(\hat{y}(\mathbf{x})) = \sigma^2 N^{-1} \mathbf{f}_1(\mathbf{x})^T \mathbf{M}_{N,X_1}^{-1} \mathbf{f}_1(\mathbf{x})$ . On aboutit ainsi à :

$$V = \text{Tr}[\mu_{11} \mathbf{M}_{N,X_1}^{-1}] = N \text{Tr}[\mu_{11} (\mathbf{X}_{N,1}^T \mathbf{X}_{N,1})^{-1}] \quad (17)$$

où la matrice  $(p_1 \times p_1)$  des moments  $\mu_{11}$  est définie par :

$$\mu_{11} = \mathcal{D} \int_{\Xi} \mathbf{f}_1(\mathbf{x}) \mathbf{f}_1(\mathbf{x})^T d\pi(\mathbf{x})$$

De même on peut exprimer  $B$  sous la forme :

$$B = \frac{N}{\sigma^2} \beta_2^T \Delta \beta_2 \quad (18)$$

où :

$$\begin{aligned} \Delta &= A^T \mu_{11} A - A^T \mu_{12} - \mu_{21} A + \mu_{22} \\ \mu_{12} &= \mathcal{D} \int_{\Xi} \mathbf{f}_1(\mathbf{x}) \mathbf{f}_2(\mathbf{x})^T d\pi(\mathbf{x}) \\ \mu_{21} &= \mu_{12}^T \\ \mu_{22} &= \mathcal{D} \int_{\Xi} \mathbf{f}_2(\mathbf{x}) \mathbf{f}_2(\mathbf{x})^T d\pi(\mathbf{x}) \end{aligned}$$

On remarque que (18) dépend du vecteur de paramètres inconnus, et donc il faudra apporter des informations (par une approche bayésienne par exemple) sur ce vecteur pour calculer (18).

Si on construit un plan en ignorant  $B$  cela revient à minimiser  $V$  ce qui revient à définir un critère de  $V$ -optimalité (Fedorov, 1969, 1972, le dénomme critère de  $Q$ -optimalité). Minimiser  $B$  en ignorant  $V$  est plus rare. En résumé le critère  $J$  revient à minimiser :

$$\gamma_1 V + \gamma_2 B \quad ; \quad 0 < \gamma_1, \gamma_2 \leq 1$$

même si le plus souvent on choisit  $\gamma_1 = \gamma_2 = 1$ . Illustrons ce critère avec les modèles élémentaires suivants.

### 3.10.3 Application aux modèles de droite et de parabole

Pour une illustration claire de ce critère on reprend ici l'exemple de Khuri et Cornell (1987). Soit à ajuster une droite à partir de  $N$  données  $(x_i, y_i)$  sur  $\Xi = [-1, +1]$ . Sans perte de généralité on supposera les  $x_i$  centrés de sorte que  $\sum_{i=1}^N x_i = 0$ .

On a :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \text{et} \quad y_{vrai} = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

a/ Calculons le terme  $V$

On a :

$$var(\hat{y}) = var(\hat{\beta}_0) + x^2 var(\hat{\beta}_1) = \frac{\sigma^2}{N} + x^2 \frac{\sigma^2}{\sum_{i=1}^N x_i^2}$$

et ainsi il vient en posant le moment  $[11] = \frac{\sum_{i=1}^N x_i^2}{N}$  :

$$V = \frac{1}{2} \int_{-1}^{+1} \left( 1 + \frac{x^2}{\sum_{i=1}^N x_i^2} \right) dx = 1 + \frac{1}{3 \sum_{i=1}^N x_i^2} = 1 + \frac{1}{3 [11]}$$

b/ Calculons le terme  $B$

On part de :

$$B = \frac{N}{2\sigma^2} \int_{-1}^{+1} [E(\hat{y}) - \beta_0 - \beta_1 x - \beta_2 x^2]^2 dx$$

En posant  $[111] = N^{-1} \sum_{i=1}^N x_i^3$  on obtient :

$$\begin{aligned} B &= \frac{N\beta_2^2}{\sigma^2} \int_{-1}^{+1} \left\{ \frac{([11]^2 + [111]x)}{[11]} - x^2 \right\}^2 dx \\ &= \frac{N\beta_2^2}{\sigma^2} \left\{ [11]^2 + \frac{[111]^2}{3[11]^2} - \frac{2}{3}[11] + \frac{1}{5} \right\} \\ &= \frac{N\beta_2^2}{\sigma^2} \left\{ [11]^2 - \frac{2}{3}[11] + \frac{1}{5} \right\} \end{aligned}$$

Si on pose  $[111] = 0$  correspondant au cas où les  $x_i$  sont symétriquement répartis autour de 0 on a :

$$J = 1 + \frac{1}{3[11]} + \left( \frac{\beta_2}{\sigma/\sqrt{N}} \right)^2 \left\{ \left( [11] - \frac{1}{3} \right)^2 + \frac{4}{45} \right\}$$

Il nous faut maintenant émettre des hypothèses sur  $\beta_2$  :

- Un premier cas extrême : quand la courbure est très petite par rapport à l'écart-type de l'échantillon de taille  $N$  c'est-à-dire quand  $\beta_2 \ll \sigma/\sqrt{N}$ , alors  $J$  est minimisé en rendant  $[11]$  grand, c'est-à-dire en minimisant  $V$ .
- Un second cas extrême : quand  $\beta_2 \gg \sigma/\sqrt{N}$  est grand, alors  $J$  est minimisé en posant  $[11] = \frac{1}{3}$ , c'est-à-dire en minimisant  $B$ .
- Pour les cas intermédiaires où  $V$  et  $B$  sont simultanément différents de zéro, Box et Draper (1987) proposent de considérer le rapport  $g = \frac{V}{B}$  et établissent le tableau suivant :

$g = \frac{V}{B}$	$\frac{\sqrt{N}\beta_2}{\sigma}$	$[11]_{\text{optimal}}$	$J_{\text{optimal}}$	$J_{\text{tout biais}} (V = 0, B > 0)$
1/2	6.540	0.363	5.755	5.799
1	4.499	0.388	3.718	3.798
2	2.994	0.433	2.656	2.797
4	1.822	0.519	2.052	2.296
10	0.501	1.000	1.467	2.022

Tableau III : Influence du rapport  $g$  sur le critère  $J$ .

On note dans ce tableau l'importante contribution du biais dans le critère. Appliquons les résultats du tableau III dans le cas de nos trois hypothèses. On sait déjà que les expériences doivent se répartir symétriquement autour de zéro, centre de l'intervalle  $[-1, +1]$ .

- a/ Un plan discret "tout-variance" ( $V > 0, B = 0$ ) conduit à :
  - placer  $N/2$  expériences respectivement en  $-1$  et  $+1$  si  $N$  est pair,
  - si  $N$  est impair une expérience sera placée au centre.
- b/ Un plan "tout biais" ( $V = 0, B > 0$ ) conduit à :
  - placer  $N/2$  expériences en  $-0.58$  et  $+0.58$ , si deux niveaux seulement sont autorisés,
  - si  $N$  est impair une expérience est placée au centre et  $(N - 1)/2$  expériences respectivement en  $x_i = \pm\sqrt{N/3(N - 1)}$ ,
  - si trois niveaux de  $x$  sont autorisés et si  $N$  est multiple de 3, alors  $N/3$  expériences se placent respectivement en  $-\sqrt{0.5}, 0, +\sqrt{0.5}$ .
- c/ Un plan équilibré  $V = B$  conduit à :
  - placer  $N/2$  expériences respectivement en  $\pm 0.623$  si  $N$  est pair,
  - sinon, placer une expérience au centre et  $(N - 1)/2$  respectivement en des sites dépendants de  $N$  (à chaque fois  $[11]$  vaut 0.388), soit :
    - si  $N = 3$  alors  $x_i = \pm 0.763$
    - si  $N = 5$  alors  $x_i = \pm 0.696$
    - si  $N = 7$  alors  $x_i = \pm 0.672$
  - si  $N$  est multiple de 3 et si on place  $N/3$  en chaque site, les sites sont toujours  $x_i = \pm 0.763$ .

Ce critère  $J$  est un critère très performant quand le modèle postulé est suspecté d'oublier des termes que l'on peut spécifier, car de nombreuses simulations ont montré que les plans obtenus sont plus robustes que ceux obtenus avec le critère  $D$ . La procédure GOSSET de Hardin et Sloane (1993) permet de les construire.

## 4 Plans optimaux pour modèles de régression non linéaire

### 4.1 Le cadre non linéaire

Ce paragraphe mériterait à lui seul tout un article didactique. Pour cette raison nous serons volontairement très succincts. Néanmoins, nous indiquerons précisément le cadre du modèle non linéaire concerné, et citerons les principales approches rencontrées dans la littérature. Ces approches sont détaillées dans Gauchi (1997).

La non linéarité des modèles de régression paramétrique considérés peut s'exercer de différentes manières, éventuellement simultanément :

- Dans la forme analytique du modèle elle-même, relativement à ses paramètres, on utilise alors la dénomination de modèle intrinsèquement non linéaire. On peut en donner une définition plus formelle :

**Définition**

*Un modèle de régression est dit intrinsèquement non linéaire si au moins une dérivée première de la fonction du modèle par rapport aux paramètres dépend d'au moins un de ces paramètres. Si certaines dérivées ne dépendent pas des paramètres, le modèle sera qualifié de partiellement non linéaire.*

Par exemple, le modèle de Michaelis-Menten :  $\eta(\theta, x) = \theta_1 x / (\theta_2 + x)$  est partiellement intrinsèquement non linéaire en les paramètres.

- Dans l'expression de la variance des erreurs, représentée alors par un modèle paramétrique :  $var(\varepsilon_i) = \sigma^2(\mathbf{x}, \theta, \beta)$  où  $\beta$  est le paramètre spécifique de la variance. Par exemple, on peut imaginer  $\sigma^2(\mathbf{x}, \beta) = \exp(\beta x)$  avec  $\beta > 0$ .

Dans ce cadre non linéaire le point crucial par rapport à la construction d'un plan d'expériences est que la matrice d'information dépend maintenant des paramètres inconnus  $\theta^*$  (et éventuellement  $\beta$  aussi), contrairement au cas linéaire du paragraphe précédent. Quels que soient les critères d'optimalité imaginés  $\theta^*$  sera présent (sous une forme ponctuelle ou de distribution) dans la définition de ces critères. Indiquons maintenant les notations relatives au modèle de régression non linéaire.

## 4.2 Le modèle de régression non linéaire

On considère simultanément le modèle de régression non linéaire paramétrique suivant

$$y_{ik} = \eta_\xi(\mathbf{x}_i, \theta^*) + \varepsilon_{ik} \quad ; \quad i = 1, \dots, N_S \quad ; \quad k = 1, \dots, r_i \tag{19}$$

et le plan d'expériences associé (4), où, en plus des notations du §3.1,  $\eta_\xi(\cdot)$  est une fonction scalaire continue définie sur  $\Xi \times \Theta$ , évaluée dans  $\mathbb{R}$  pour le plan (4), doublement dérivable relativement aux paramètres et aux facteurs explicatifs ; on notera  $\eta_\xi(\cdot)$  le vecteur  $N \times 1$  correspondant. On supposera les erreurs gaussiennes. Pour l'estimation de  $\theta^*$  on utilise l'estimateur des moindres carrés non linéaire défini par :

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{Argmin}} \|\mathbf{y} - \eta_\xi(\mathbf{x}, \theta^*)\|^2 \tag{20}$$

qui jouit de bonnes propriétés asymptotiques dans le cas non linéaire (Jenrich 1969) : il est asymptotiquement distribué comme  $\mathcal{N}(\theta^*, \mathbf{M}^F(\xi, \theta^*)^{-1})$ , où  $\mathbf{M}^F(\xi, \theta^*)$  est la matrice d'information de Fisher (voir, par exemple, Atkinson and Doneev (1992) pour des détails sur celle-ci) définie ici pour un plan à  $N$  expériences par :

$$\mathbf{M}_N^F(\xi, \theta^*) = \frac{1}{\sigma^2} \mathbf{J}_N(\mathbf{x}, \theta^*)^T \mathbf{J}_N(\mathbf{x}, \theta^*) \tag{21}$$

où  $\mathbf{J}_N(\mathbf{x}, \theta^*)$  est la matrice jacobienne  $N \times p$ , de terme général  $\partial \eta(\mathbf{x}, \theta^*) / \partial \theta_j^*$ ,  $j = 1, \dots, p$ , et où le symbole  $\xi$  remplace  $\xi_{N, N_S, \{r\}}$  pour simplifier l'écriture.

## 4.3 Le plan $D$ -optimal local

Le principe de l'optimalité locale est dû à Chernoff (1953) : la matrice jacobienne  $\mathbf{J}_N(\mathbf{x}, \theta^*)$  joue le rôle de matrice du modèle, et  $\theta^*$  inconnu sera remplacé par une valeur

a priori  $\theta_0$ . L'emploi de la matrice jacobienne provient de la linéarisation du modèle  $\eta(\cdot)$  autour de la solution aux moindres carrés  $\hat{\theta}$ . Les définitions et théorèmes du paragraphe précédent restent alors valables "localement". Le critère de  $D$ -optimalité devient le critère de  $D$ -optimalité locale, qu'on note parfois  $D(\theta_0)$ -optimalité, et se définit comme en (8), avec  $\mathbf{M}_N(\xi)$  remplacée par  $\mathbf{M}_N^F(\xi, \theta_0)$ . Ainsi, même le TEG se généralise au cas non linéaire (White, 1973). On utilise les mêmes algorithmes que pour le cas linéaire pour construire le plan  $D(\theta_0)$ -optimal.

On peut faire le commentaire suivant quant au choix  $\theta_0$ . Celui-ci est basé sur une connaissance préalable (théorique ou expérimentale) qui implique par exemple que l'on sache que les dérivées du modèle varient lentement dans le domaine étudié et que le domaine paramétrique  $\Theta$  soit bien choisi (qu'il contienne effectivement  $\theta^*$ ).

### Exemple

Considérons le modèle (2) de la situation expérimentale du §2.5. Après avoir calculé les dérivées  $\partial\eta(t, \theta^*)/\partial\theta_1^*$  et  $\partial\eta(t, \theta^*)/\partial\theta_2^*$ , on peut exprimer la matrice  $\mathbf{J}_2(\theta_0, \xi(t_1, t_2))$  en fonction de deux variables  $t_1$  et  $t_2$ , avec  $\theta_0 = (0.7 \ 0.2)^T$  fourni par Box et Lucas (1959). Le déterminant de la matrice d'information locale s'écrit alors :

$$\Delta_2 = \det(\mathbf{J}_2^T(\theta_0, \xi(t_1, t_2))\mathbf{J}_2(\theta_0, \xi(t_1, t_2))) = [\det \mathbf{J}_2(\theta_0, \xi(t_1, t_2))]^2$$

On cherche le couple  $(t_1^*, t_2^*)$  qui le maximise. Comme il n'y a pas de solution analytique ici, Box et Lucas (1959) utilisent une méthode numérique classique qui conduit au plan optimal cherché (à deux points de support) :

$$t_1^* = 1.229 \quad t_2^* = 6.858 \quad \text{avec } \Delta_2^{\max} = 0.657$$

Par ailleurs, on rappelle (Vila, 1985) que pour  $N > 2$  les plans  $D(\theta_0)$ -optimaux se concentrent tous sur le même support (1.229 ; 6.858). Si l'information sur  $\theta^*$ , au travers de  $\theta_0$ , est très douteuse, on aura un plan localement  $D(\theta_0)$ -optimal de mauvaise qualité. Une alternative est alors le plan suivant.

## 4.4 Le plan $D$ -optimal bayésien

Intuitivement, il est raisonnable de penser que, si on dispose d'une connaissance a priori assez fiable concernant la loi de probabilité du paramètre  $\theta^*$ , le plan sera plus efficace (robuste) en injectant cette information lors de sa construction, plutôt que de se contenter d'une valeur isolée  $\theta_0$ . Le principe de cette approche est donc basé sur le théorème de Bayes qui permet d'exprimer la densité de probabilité a posteriori du paramètre inconnu  $\theta^*$  ayant observé le vecteur  $\mathbf{y}$  et en disposant d'une loi a priori pour  $\theta^*$  :

$$p(\theta^* | \mathbf{y}(\xi)) = \frac{\mathcal{L}(\mathbf{y}(\xi) | \theta^*) \mathcal{P}_0(\theta_0, \mathbf{H})}{\int_{\Theta} \mathcal{L}(\mathbf{y}(\xi) | \theta^*) \mathcal{P}_0(\theta_0, \mathbf{H}) d\theta^*}$$

avec :

- $\mathbf{y}(\xi)$  vecteur des observations recueillies au plan  $\xi$ ,
- $\mathcal{L}(\mathbf{y}(\xi) | \theta^*)$  la vraisemblance de ce vecteur,
- $\mathcal{P}_0(\theta_0, \mathbf{H})$  la densité de probabilité a priori de  $\theta^*$ , d'espérance  $\theta_0$  et de matrice de variance  $\mathbf{H}$ .
- $p(\theta^* | \mathbf{y}(\xi))$  la densité de probabilité a posteriori de  $\theta^*$  ayant observé le vecteur  $\mathbf{y}(\xi)$ .

Le paramètre vectoriel  $\theta^*$  est alors considéré comme un paramètre aléatoire. Sur ce principe Draper et Hunter (1967) proposent une approche qui revient à définir un plan  $D$ -optimal bayésien. Ensuite, Pronzato (1986) propose plusieurs critères bayésiens, plus robustes. Puis Chaloner et Verdinelli (1995) proposent dans ce contexte bayésien un critère basé sur la maximisation du logarithme du déterminant de la matrice d'information de Fisher. On trouvera une application de ce critère dans Gauchi (2005).

## 4.5 Le plan $X$ -optimal

Dans cette approche on ne fonde plus le critère d'optimalité sur la linéarisation implicite du modèle non linéaire autour d'une valeur a priori  $\theta_0$  du paramètre vectoriel inconnu  $\theta^*$ , mais plutôt sur la minimisation de l'espérance du volume d'une région de confiance exacte pour le paramètre  $\theta^*$ . On rappelle qu'une telle région est dite *exacte* si sa probabilité  $C(\theta^*)$  de recouvrir la vraie valeur  $\theta^*$  est  $1 - \alpha$ , pour un niveau  $\alpha$  choisi et  $\forall N$ . Le critère se nomme le critère de  $X$ -optimalité, et il en existe aussi une version bayésienne appelée le critère de  $X^l$ -optimalité (Vila et Gauchi, 2005).

## 4.6 Les plans $T_{XP}$ -optimal et $D_{XP}$ -optimal

Cette approche, très rigoureuse mais qui conduit à des critères très techniques et très lourds numériquement, est essentiellement due à Pazman et Pronzato (1992) et Gauchi et Pazman (2003, 2005). On ne peut pas détailler ici cette approche, mais on indique seulement qu'elle est basée sur la fonction de densité exacte de probabilité (à distance finie) de l'estimateur des moindres carrés non linéaire, densité établie par Pazman (1984).

## 5 Conclusion

Bien sûr, d'autres critères d'optimalité existent, répondant à des objectifs précis. On renvoie le lecteur à Gauchi (1997) pour un panorama assez large sur le sujet. Par ailleurs, il faut souligner que, dans le cas non linéaire, il est nécessaire de faire appel à des méthodes puissantes d'optimisation, notamment stochastiques, pour calculer les plans optimaux, elles sont évoquées par exemple dans Vila et Gauchi (2005), et Gauchi et Pazman (2003).

Enfin, il faut indiquer qu'il existe plusieurs logiciels commerciaux pour calculer des plans optimaux, essentiellement dans le cas linéaire. Par exemple, sans être exhaustif et sans préjuger de la qualité des logiciels non nommés, on peut citer la procédure OPTEX du logiciel SAS/QC (SAS Institute), les logiciels STATGRAPHICS (Société Sigma Plus), NEMROD (Société LPRAI), NEUROPEX (Société NETRAL). Ce dernier calcule aussi des plans  $X$ -optimaux. Toutes les informations sur ces logiciels sont facilement accessibles par le moteur de recherche Google sur Internet.

## Références

- Atkinson, A.C., Donev, A.N. (1992). Optimum experimental designs. Oxford : Clarendon Press.
- Atkinson, A.C., Hunter, W.G. (1968). The design of experiments for parameter estimation. *Technometrics*, 10, 2, 271-289.

- Box, G.E.P., Draper, N.R. (1959). A basis for selection of a response surface design. *J. of the Amer. Stat. Association*, 54, 622-653.
- Box, G.E.P., Draper, N.R. (1987). *Empirical Model Building and Response Surfaces*. Wiley. New-York.
- Box, G.E.P., Lucas, H.L. (1959). Design of experiments in nonlinear situations. *Biometrika*, 46, 77-90.
- Chaloner, K., Verdinelli, I. (1995). Bayesian experimental design : a review. *Statistical Science*, 10, 3, 273-304.
- Chernoff, H., (1953). Locally optimum designs for estimating parameters. *Ann. of Math. Statist.* 24, 586-602.
- Draper, N.R., Hunter, W.G. (1967). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations. *Biometrika*, 54, 147-153.
- Fedorov, V.V. (1969, en russe, 1972, en anglais). *Theory of Optimal Experiments*. Academic Press, New York.
- Fedorov, V.V. (1980). Convex design theory. *Math. Operationsforsch. Statist., Ser. Statistics*, 11, 3, 403-411.
- Gauchi, J.-P. (1997). Plans d'expériences optimaux pour modèles linéaires, Chap. 7 et 8, In *Plans d'Expériences-Applications à l'Entreprise*, Editeur TECHNIP, Paris 1997, (Editeurs scientifiques : J.J. Dreesbeke, J. Fine et G. Saporta).
- Gauchi, J.-P. (2005). Optimal statistical designs for the accurate estimation of the parameters of a growth rate model for *Listeria monocytogenes*. Technical Report n°2005-2 of the MIA Unit, INRA, Research Center of Jouy-en-Josas, France. On line at : <http://www.inra.fr/miaj/public/nosdoc/inrajouymia2005T2.pdf>.
- Gauchi, J.-P., Pázman, A. (2003). Distribution of the least squares estimator, stochastic optimization and optimum designs in nonlinear models. Technical Report of the Biometrics Unit, n°2003-7, INRA, Jouy-en-Josas, France.
- Gauchi, J.-P., Pázman, A. (2005). Designs in nonlinear regression by stochastic minimization of functionals of the mean square error matrix. In Press in *J. of Stat. Planning and Inference*.
- Hardin, R.H., Sloane, N.J.A. (1993). A new approach to the construction of optimal design. *J. of Stat. Planning and Inference*, 37, 339-369.
- Jenrich, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. of Stat.*, 40, 2, 633-643.
- Karlin, S., Studden, W.J. (1966). Optimal experimental designs. *Ann. of Math. Stat.*, 37, 783-815.
- Khuri, A.I., Cornell, J.A. (1987). *Response surfaces : designs and analyses*. Marcel Dekker.
- Kiefer, J. (1959). Optimum experimental designs. *J.R. Statist. Soc. B.*, 21,272-319.
- Kiefer, J. (1961). Optimum designs in regression problems II. *Ann. of Math. Stat.*, 32, 298-325.
- Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. of Statist.*, 2, 849-879.

- Kiefer, J., Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. of Math. Statist.*, 30, 271-294.
- Kiefer, J., Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.*, 12, 363-366.
- Kobilinsky, A. (1988). *Tactiques en analyse de variance et en régression*. *Revue Modulad*, 1, 25-58.
- Mathieu, D. (1981). *Contribution de la méthodologie de la recherche expérimentale à l'étude des relations structure-activité*. Thèse. Université d'Aix-Marseille. France.
- Mitchell, T.J. (1974). An algorithm for the construction of  $D$ -optimal experimental designs. *Technometrics*, 16, 203-210.
- Pázman, A. (1984). Probability distribution of the multivariate nonlinear least squares estimator. *Kybernetika*, 20, 209-230.
- Pázman, A., Pronzato, L. (1992). Nonlinear experimental design based on the distribution of estimators. *J. of Stat. Planning and Inference*, 33, 385-402.
- Pronzato, L. (1986). *Synthèse d'expériences robustes pour modèles à paramètres incertains*. Thèse. Université d'Orsay.
- Pronzato, L. (2004). A minimax equivalence theorem for optimum bounded design measure. *Statistics & Probability Letters*, 68, 325-331.
- Sibson, R. (1974).  $DA$ -optimality and duality, in J. Gani, K.Sarkadi, I. Vincze (Eds), *Progress in Statistics. Proceedings 9th European Meeting of Statisticians*, Budapest 1972, 2, 667-692. Amsterdam, North Holland.
- Silvey, S.D. (1980). *Optimal designs : an introduction to the theory for parameter estimation*. London : Chapman and Hall.
- Torsney, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing, in *Optimal design and Analysis of experiments* (Dodge, Fedorov, Wynn Eds), 361-370.
- Vila, J.P. (1985). *Etude et comparaison de critères de plans d'expériences optimaux pour l'estimation des paramètres d'un modèle de régression non linéaire*. Thèse Université d'Orsay.
- Vila, J.P. (1991). Local optimality of replications from a minimal  $D$ -optimal design in regression : a sufficient and a quasi-necessary condition. *J. of Stat. Planning and Inference*, 29, 261-277.
- Vila, J.P., Gauchi, J.-P. (2005). Optimal designs based on exact confidence regions for parameter estimation of a nonlinear regression model. En cours de soumission à *J. of Stat. Planning and Inference*.
- White, L.V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, 60, 2, 345-348.
- Whittle, P. (1973). Some general points in the theory of optimum experimental design. *J. Roy. Statist. Soc, B*, 35, 123-130.
- Wynn, H.P. (1970). The sequential generation of  $D$ -optimum experimental designs. *Ann. of Math. Stat.*, 41, 1655-1664.