

# STATISTIQUE ET ANALYSE DES DONNÉES

LUC PRONZATO

ERIC WALTER

## **Détermination bayésienne des conditions opératoires maximisant un critère de qualité**

*Statistique et analyse des données*, tome 16, n° 2 (1991), p. 107-125

[http://www.numdam.org/item?id=SAD\\_1991\\_\\_16\\_2\\_107\\_0](http://www.numdam.org/item?id=SAD_1991__16_2_107_0)

© Association pour la statistique et ses utilisations, 1991, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## DETERMINATION BAYESIENNE DES CONDITIONS OPERATOIRES MAXIMISANT UN CRITERE DE QUALITE

Luc Pronzato et Eric Walter,

Laboratoire des Signaux et Systèmes,  
CNRS / Ecole Supérieure d'Electricité  
Plateau de Moulon,  
91192 Gif-sur-Yvette Cedex, France

### RESUME

*La détermination des conditions opératoires d'un système qui maximisent un critère de qualité est considérée. Quand la description de la relation supposée exister entre ce critère et ces conditions opératoires se fait à l'aide d'un modèle dépendant de paramètres inconnus, il est nécessaire d'estimer ces paramètres à partir de données expérimentales. Se pose alors un double problème de planification d'expérience et d'estimation. Ce problème est résolu ici dans un cadre Bayésien. Le critère de planification obtenu correspond à la notion de  $L_B$ -optimalité ; son optimisation peut donc s'effectuer à l'aide d'algorithmes connus. Le problème de la robustesse de l'approche proposée à des hypothèses erronées est abordé. Un exemple illustratif est présenté qui permet de comparer les résultats de cette approche à ceux d'une approche non Bayésienne plus classique.*

### ABSTRACT

*The determination of the operating conditions of a system that maximize a quality index is considered. When a model depending on unknown parameters is used to describe the relationship supposed to exist between this quality index and these operating conditions, these parameters must be estimated from experimental data. Hence a joint problem of parameter estimation and experimental design. This problem is solved here in a Bayesian framework. The design criterion obtained corresponds to the notion of  $L_B$ -optimality; its optimization can thus be performed using known algorithms. The issue of the robustness of the approach proposed to misspecifications of the priors is addressed. An illustrative example is treated, which allows a comparison to be drawn between the results of this approach and those of a more classical non-Bayesian approach.*

**Mots-clés :** Planification d'expériences, estimation Bayésienne, optimisation de qualité

## 1. INTRODUCTION

Dans de nombreux problèmes industriels, on souhaite maximiser une variable  $y$  (qui quantifie la qualité de la réponse du système considéré) par rapport à  $m$  variables d'entrée indépendantes (ou conditions opératoires). Le vecteur  $\underline{x}$  de ces conditions opératoires doit appartenir au domaine admissible  $\mathcal{X}$ . Ce problème a reçu une attention considérable dans la littérature et présente des implications évidentes dans le domaine de l'optimisation de la qualité. Sa solution est loin d'être triviale, car la dépendance de  $y$  en  $\underline{x}$  est en général inconnue et, de plus,  $y$  est sujet à une variabilité aléatoire. Nos connaissances concernant le système sont déduites de mesures de  $y$  sous diverses conditions opératoires  $\underline{x}^i$ . Ces connaissances sont utilisées pour prédire la partie déterministe de  $y$ , notée dans ce qui suit  $\eta(\underline{x})$ , puis en déduire le point optimal  $\underline{x}^*$  de la surface de réponse  $\{\eta(\underline{x}); \underline{x} \in \mathcal{X}\}$ . Dans ce contexte, une méthode performante devrait fournir une bonne estimée de  $\underline{x}^*$  à partir d'un nombre limité  $N$  de mesures  $y(\underline{x}^i)$ ,  $i = 1, \dots, N$ .

La technique la plus directe pour maximiser  $\eta(\underline{x})$  consiste à conduire les expériences séquentiellement en suivant une technique d'optimisation ne faisant pas appel à des dérivées, telle que la méthode du simplexe [1]. A chaque pas, compte tenu des observations passées, de nouvelles conditions expérimentales sont suggérées. Cette approche ne nécessite pas la spécification d'une structure de modèle pour  $\eta(\underline{x})$  mais ne prend pas en compte la variabilité de  $y$  due par exemple aux erreurs de mesure, de sorte que plus on se rapproche d'un maximum local de  $\eta$ , moins les modifications suggérées pour les conditions expérimentales sont significatives. Cet inconvénient peut être évité en utilisant une méthode de recherche fondée sur la procédure d'approximation stochastique de Kiefer et Wolfowitz [2]. Cependant, cette méthode nécessite à chaque pas l'évaluation d'une approximation de la dérivée de  $\eta$  par rapport à  $\underline{x}$ , ce qui implique au moins  $(m+1)$  mesures  $y(\underline{x}^i)$ . Comme, de plus, sa convergence est lente, le nombre total des mesures à effectuer semble prohibitif dans de nombreux cas. C'est là le prix à payer pour la généralité des approches qui n'utilisent pas de modèle mathématique pour décrire la dépendance de  $\eta$  en  $\underline{x}$ .

L'hypothèse d'une partie déterministe de  $y$  décrite par un modèle de structure connue permet de limiter le nombre de mesures nécessaires à la détermination de  $\underline{x}^*$ . Nous supposons que

$$y(\underline{x}) = \eta(\overline{\underline{\theta}}, \underline{x}) + \varepsilon(\underline{x}), \quad (1)$$

où les  $\epsilon(\underline{x})$  sont des variables aléatoires indépendantes identiquement distribuées suivant une loi normale  $\mathcal{N}(0, \sigma^2)$ . La fonction  $\eta(\underline{\theta}, \underline{x}) = E \{y(\underline{x})\}$  est la réponse d'un modèle déterministe (par exemple de type polynomial en  $\underline{x}$ ),  $\underline{\theta}$  est le vecteur des  $p$  paramètres de ce modèle (appartenant à un domaine admissible  $\Theta$ ), et  $\underline{\theta}^*$  est une valeur (inconnue mais supposée exister) de  $\underline{\theta}$  telle que (1) soit satisfaite. Par hypothèse,  $\frac{\partial^2 \eta(\underline{\theta}, \underline{x})}{\partial \underline{\theta} \partial \underline{\theta}^T}$  existe en tout point intérieur de  $\Theta$ . L'estimation de  $\sigma^2$  ne sera pas considérée ici. Le problème de la robustesse de l'approche proposée vis-à-vis de la connaissance de *a priori* de  $\sigma^2$  est abordé au paragraphe 2 (optimalité au sens minmax). Notons que l'estimation bayésienne conjointe de  $\sigma^2$  et des paramètres  $\underline{\theta}$  pourrait être envisagée en considérant une distribution *a priori* conjuguée inverse-gamma / normale. Une telle approche conduirait à remplacer  $\sigma^2$  par sa moyenne *a priori* (voir [6]).

Pour pouvoir utiliser (1) afin d'optimiser  $y$ , il faut estimer  $\underline{\theta}$  à partir de la mesure de  $y$  pour diverses valeurs de  $\underline{x}$ . Or le nombre des mesures envisageables est souvent limité, pour des raisons économiques, techniques voire éthiques. Il faut donc convenablement choisir les expériences à réaliser. La méthodologie classique de la surface de réponse [3-5] détermine les entrées optimales  $\underline{x}^*$  (au sens de  $y$ ) de façon séquentielle. A l'étape  $k$ , une petite région  $\mathcal{X}^{(k)}$  de  $\mathcal{X}$  est considérée. Tant que  $\mathcal{X}^{(k)}$  est loin de  $\underline{x}^*$ ,  $\eta$  peut être approximée par une fonction affine de  $\underline{x}$ . Plus près de  $\underline{x}^*$  une fonction quadratique est généralement plus adaptée. Le plus souvent, on applique à chaque étape des conditions expérimentales préséparées (par exemple des plans composites centrés) pour obtenir les estimées de  $\underline{\theta}$  qui seront utilisées pour calculer  $\underline{x}^*$ .

Une telle approche présente cinq inconvénients principaux :

- (i) la planification ne tient pas compte du but poursuivi (or les paramètres influent de façon différente sur le critère de qualité  $y$  et ne devraient donc pas tous se voir accorder la même importance) ;
- (ii) l'information *a priori* pouvant être éventuellement déduite d'expériences préalables sur le même type de processus n'est pas exploitée ;
- (iii) l'approche est intrinsèquement locale, et les contraintes éventuelles définissant  $\mathcal{X}$  ne sont pas prises en compte ;

- (iv) la structure du modèle est supposée linéaire ou quadratique en  $\underline{x}$ , ce qui peut ne pas être valide sur tout  $\mathcal{X}$ ;
- (v) l'estimation non Bayésienne de  $\underline{\theta}$  implique à chaque étape un nombre de mesures au moins égal à la dimension  $p$  de  $\underline{\theta}$ .

Nous proposons une nouvelle approche, reposant sur un critère d'optimalité Bayésien, qui prend en compte de façon explicite le but poursuivi par la modélisation (i), les informations *a priori* sur les paramètres du modèle (ii) et des contraintes éventuelles sur les conditions opératoires  $\underline{x}$  (iii). Elle permet en outre de considérer des modèles de structure plus générale que l'approche traditionnelle (iv) et peut se contenter de moins de  $p$  mesures à chaque étape (v). La réponse du modèle est supposée linéaire en  $\underline{\theta}$ , c'est à dire qu'elle peut s'écrire

$$\eta(\underline{\theta}, \underline{x}) = \underline{z}^T(\underline{x}) \underline{\theta}; \quad (2)$$

mais le modèle n'est pas nécessairement polynomial en  $\underline{x}$ .

Le paragraphe 2 définit le coût conditionnel et le risque Bayésien utilisés pour l'estimation et la planification. Sous des hypothèses techniques, nous montrons que le critère de planification obtenu par développement limité au second ordre correspond à la notion de  $L_B$ -optimalité. Ceci permet d'utiliser des procédures algorithmiques de la littérature. Un exemple numérique est traité au paragraphe 3 où l'approche Bayésienne proposée est comparée à une approche D-optimale classique combinée à une estimation par maximum de vraisemblance.

## 2. ESTIMATION ET PLANIFICATION BAYESIENNES

Notons  $\underline{x}^*(\underline{\theta})$  la valeur optimale de  $\underline{x}$  au sens de l'espérance de  $y$

$$\underline{x}^*(\underline{\theta}) = \arg \max_{\underline{x} \in \mathcal{X}} \eta(\underline{\theta}, \underline{x}), \quad (3)$$

et  $\mathcal{L}(\hat{\underline{\theta}} | \bar{\underline{\theta}})$  le coût de croire que les paramètres ont pour valeur  $\hat{\underline{\theta}}$  alors qu'ils valent  $\bar{\underline{\theta}}$ . Nous choisissons la fonction coût définie par

$$\mathcal{L}(\hat{\theta} | \bar{\theta}) = E_{\underline{y} | \bar{\theta}} \{y(\underline{x}^*(\bar{\theta})) - y(\underline{x}^*(\hat{\theta}))\} = \eta(\bar{\theta}, \underline{x}^*(\bar{\theta})) - \eta(\bar{\theta}, \underline{x}^*(\hat{\theta})). \quad (4)$$

Soit  $\underline{y}_x$  le vecteur des  $N$  observations ( $N$  fixé *a priori*) effectuées sur le processus pour estimer  $\theta$  avec le protocole entier  $X = \{\underline{x}^1, \underline{x}^2, \dots, \underline{x}^N\}$ , comportant éventuellement des répétitions ( $\underline{x}^i = \underline{x}^j, i \neq j$ ). On peut alors écrire le risque Bayésien

$$r(X) = E_{\theta, \underline{y}_x} \{ \mathcal{L}(\hat{\theta}(\underline{y}_x) | \theta) \} = \int p(\underline{y}_x) \left( \int \mathcal{L}(\hat{\theta}(\underline{y}_x) | \theta) p(\theta | \underline{y}_x) d\theta \right) d\underline{y}_x, \quad (5)$$

et prendre comme estimateur Bayésien l'estimateur du risque minimum

$$\hat{\theta}_B(\underline{y}_x) \in \arg \min_{\hat{\theta} \in \Theta} \int \mathcal{L}(\hat{\theta} | \theta) p(\theta | \underline{y}_x) d\theta = \arg \min_{\hat{\theta} \in \Theta} E_{\theta | \underline{y}_x} \{ \mathcal{L}(\hat{\theta} | \theta) \}, \quad (6)$$

ou de façon équivalente, d'après (4) et la linéarité (2) de  $\eta(\theta, \underline{x})$  vis-à-vis de  $\theta$ ,

$$\hat{\theta}_B(\underline{y}_x) \in \arg \min_{\hat{\theta} \in \Theta} \mathcal{L}(\hat{\theta} | E_{\theta | \underline{y}_x} \{ \theta \}), \quad (7)$$

à condition que  $E_{\theta | \underline{y}_x} \{ \theta \} \in \Theta$ , ce que nous supposons. On pourra se reporter par exemple à [18, 19] pour une présentation générale de l'inférence Bayésienne. D'après (4), (7) s'écrit aussi

$$\hat{\theta}_B(\underline{y}_x) \in \{ \hat{\theta} \in \Theta; \underline{x}^*(\hat{\theta}) = \arg \max_{\underline{x} \in \mathcal{X}} \eta( E_{\theta | \underline{y}_x} \{ \theta \}, \underline{x}) \}, \quad (8)$$

ou encore

$$\hat{\theta}_B(\underline{y}_x) \in \{ \hat{\theta} \in \Theta; \underline{x}^*(\hat{\theta}) = \underline{x}^*( E_{\theta | \underline{y}_x} \{ \theta \}) \}. \quad (9)$$

Un estimateur de  $\theta$  qui minimise le risque est donc sa moyenne *a posteriori*, et nous poserons

$$\hat{\theta}_B(\underline{y}_x) = E_{\theta | \underline{y}_x} \{ \theta \}. \quad (10)$$

En remplaçant  $\hat{\theta}$  par  $\hat{\theta}_B$  dans l'expression du risque Bayésien  $r(X)$ , et compte tenu de (4),  $r(X)$  s'écrit

$$r(X) = \int \eta(\theta, \underline{x}^*(\theta)) p(\theta) d\theta - \int p(\underline{y}_X) \left( \int \eta(\theta, \underline{x}^*(\hat{\theta}_B(\underline{y}_X))) p(\theta|\underline{y}_X) d\theta \right) d\underline{y}_X. \quad (11)$$

ou encore, puisque  $\eta(\theta, \underline{x})$  est linéaire en  $\theta$ ,

$$r(X) = \int \eta(\theta, \underline{x}^*(\theta)) p(\theta) d\theta - \int \eta(\hat{\theta}_B(\underline{y}_X), \underline{x}^*(\hat{\theta}_B(\underline{y}_X))) p(\underline{y}_X) d\underline{y}_X. \quad (12)$$

Connaissant la distribution *a priori*  $p(\theta)$ , le risque  $r(X)$  peut donc, au moins en principe, être calculé pour tout  $X$  et servir de base à une planification d'expérience optimale. Cependant le calcul de  $r(X)$  est trop complexe pour que son optimisation soit aisée, et nous travaillerons sur une version approchée de ce critère.

Nous supposons que  $\theta$  est distribué suivant une loi normale  $\mathcal{N}(\bar{\theta}, \Omega)$ . La loi marginale  $p(\underline{y}_X)$  est alors normale  $\mathcal{N}(Z_X \bar{\theta}, \sigma^2 I_N + Z_X \Omega Z_X^T)$ , où  $I_N$  désigne la matrice identité de dimension  $N$  et  $Z_X$  la matrice  $N \times p$  dont la  $i^{\text{ème}}$  ligne est  $\underline{z}^T(\underline{x}^i)$ . La moyenne *a posteriori* est alors

$$\hat{\theta}_B(\underline{y}_X) = \left( \frac{1}{\sigma^2} Z_X^T Z_X + \Omega^{-1} \right)^{-1} \left( \frac{1}{\sigma^2} Z_X^T \underline{y}_X + \Omega^{-1} \bar{\theta} \right), \quad (13)$$

et des calculs élémentaires montrent que sa distribution prédictive correspond à une loi normale  $\mathcal{N}(\bar{\theta}, \Omega - \left( \frac{1}{\sigma^2} Z_X^T Z_X + \Omega^{-1} \right)^{-1})$ . L'estimée Bayésienne  $\hat{\theta}_B(\underline{y}_X)$  (13) sera utilisée pour calculer les conditions opératoires optimales  $\underline{x}^*(\hat{\theta}_B(\underline{y}_X))$  grâce à (3).

Nous utiliserons de plus un développement limité de  $\eta(\theta, \underline{x}^*(\theta))$  autour de  $\bar{\theta}$ ,

$$\begin{aligned} \eta(\theta, \underline{x}^*(\theta)) &\approx \eta(\bar{\theta}, \underline{x}^*(\bar{\theta})) + (\theta - \bar{\theta})^T \frac{\partial \eta(\theta, \underline{x}^*(\theta))}{\partial \theta} \Big|_{\bar{\theta}} \\ &+ \frac{1}{2} (\theta - \bar{\theta})^T \frac{\partial^2 \eta(\theta, \underline{x}^*(\theta))}{\partial \theta \partial \theta^T} \Big|_{\bar{\theta}} (\theta - \bar{\theta}). \end{aligned} \quad (14)$$

Sous l'hypothèse que  $\tilde{\theta} \in \text{int } \Theta$ , que  $\underline{x}^*(\tilde{\theta}) \in \text{int } X$  est un point stationnaire de  $\eta(\tilde{\theta}, \underline{x})$  et que  $\frac{\partial \underline{x}^{*T}(\theta)}{\partial \theta} \Big|_{\tilde{\theta}}$  existe, calculons le gradient et le Hessian de  $\eta$  intervenant dans (14).

Nous avons

$$\frac{\partial \eta(\tilde{\theta}, \underline{x})}{\partial \underline{x}} \Big|_{\underline{x}^*(\tilde{\theta})} = \underline{0}, \tag{15}$$

ce qui entraîne

$$\frac{\partial \eta(\tilde{\theta}, \underline{x}^*(\theta))}{\partial \theta} \Big|_{\tilde{\theta}} = \frac{\partial \underline{x}^{*T}(\theta)}{\partial \theta} \Big|_{\tilde{\theta}} \frac{\partial \eta(\tilde{\theta}, \underline{x})}{\partial \underline{x}} \Big|_{\underline{x}^*(\tilde{\theta})} = \underline{0}, \tag{16}$$

d'où

$$\frac{\partial \eta(\theta, \underline{x}^*(\theta))}{\partial \theta} \Big|_{\tilde{\theta}} = \frac{\partial \eta(\theta, \underline{x}^*(\tilde{\theta}))}{\partial \theta} \Big|_{\tilde{\theta}} \tag{17}$$

D'après (2), le gradient de  $\eta$  vérifie donc

$$\frac{\partial \eta(\theta, \underline{x}^*(\theta))}{\partial \theta} \Big|_{\tilde{\theta}} = \underline{z}(\underline{x}^*(\tilde{\theta})). \tag{18}$$

Le Hessian de  $\eta$  est alors donné par

$$H(\tilde{\theta}) = \frac{\partial^2 \eta(\theta, \underline{x}^*(\theta))}{\partial \theta \partial \theta^T} \Big|_{\tilde{\theta}} = \frac{\partial \underline{z}(\underline{x})}{\partial \underline{x}^T} \Big|_{\underline{x}^*(\tilde{\theta})} \frac{\partial \underline{x}^*(\theta)}{\partial \theta^T} \Big|_{\tilde{\theta}}. \tag{19}$$

En remplaçant  $\eta(\cdot, \underline{x}^*(\cdot))$  par son développement limité (14) dans (12), nous obtenons

$$r(X) \approx \frac{1}{2N} \Psi_{LB}(X), \tag{20}$$

avec

$$\Psi_{LB}(X) = \text{trace} (H(\tilde{\theta}) M_B^{-1}(X)), \tag{21}$$

où  $M_B(X)$  est la matrice d'information Bayesienne (par échantillon) [6], donnée par

$$M_B(X) = \frac{1}{N} \left( \frac{1}{\sigma^2} Z_x^T Z_x + \Omega^{-1} \right), \tag{22}$$

et où  $H(\underline{\theta})$  est donnée par (19). Pour montrer que  $\Psi_{LB}(X)$  est un critère de  $L_B$ -optimalité [6] (à minimiser), il reste à prouver que  $H(\underline{\theta})$  est positif semi-défini. Soit  $f(\underline{\theta})$  la fonction  $\eta(\underline{\theta}, \underline{x}^*(\underline{\theta}))$ , avec  $\underline{x}^*(\underline{\theta})$  donné par (3). Supposons qu'il existe un voisinage convexe  $\mathcal{U}$  de  $\underline{\theta}$ ,  $\Theta \supset \mathcal{U}$ ,  $\forall \underline{\theta}^1, \underline{\theta}^2 \in \mathcal{U}$ ,  $\forall \alpha \in [0, 1]$ , nous avons

$$\begin{aligned} f((1-\alpha)\underline{\theta}^1 + \alpha\underline{\theta}^2) &= \eta((1-\alpha)\underline{\theta}^1 + \alpha\underline{\theta}^2, \underline{x}^*((1-\alpha)\underline{\theta}^1 + \alpha\underline{\theta}^2)), \\ &= (1-\alpha) \eta(\underline{\theta}^1, \underline{x}^*(\underline{\theta}^1)) + \alpha \eta(\underline{\theta}^2, \underline{x}^*(\underline{\theta}^2)), \\ &\leq (1-\alpha) \eta(\underline{\theta}^1, \underline{x}^*(\underline{\theta}^1)) + \alpha \eta(\underline{\theta}^2, \underline{x}^*(\underline{\theta}^2)) = (1-\alpha) f(\underline{\theta}^1) + \alpha f(\underline{\theta}^2). \end{aligned} \quad (23)$$

La fonction  $f$  est donc convexe, de sorte que  $H(\underline{\theta})$  est positif semi-défini.  $\Psi_{LB}$  est ainsi un critère de  $L_B$ -optimalité, et il est possible d'exploiter des résultats de la littérature pour calculer le protocole optimal au sens de (21).

Un théorème explicitant diverses conditions nécessaires et suffisantes de  $L_B$ -optimalité d'un protocole caractérisé par une distribution continue est démontré dans [6, 7]. Il permet de définir un algorithme dont la convergence vers un protocole continu  $L_B$ -optimal est garantie [6]. D'après [7], il existe un tel protocole possédant au plus  $r(2p-r+1)/2$  points de support, avec  $r = \text{rang } H(\underline{\theta})$ .

Puisque nous nous intéressons à des protocoles entiers de taille  $N$  fixée et non à des protocoles continus, nous suggérons d'utiliser l'algorithme d'échange proposé par Pilz [6] :

*Pas 1.* Choisir  $X^{(1)}$  (de taille  $N$ ) et poser  $k = 1$ .

*Pas 2.* Ajouter le point de support  $\underline{x}^{(k+)}$  au protocole  $X^{(k)}$ , avec

$$\underline{x}^{(k+)} = \arg \max_{\underline{x} \in \mathcal{X}} \frac{\underline{z}^T(\underline{x}) M_B^{-1}(X^{(k)}) H(\underline{\theta}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x})}{N + \frac{1}{\sigma^2} \underline{z}^T(\underline{x}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x})}.$$

Mettre à jour  $M_B^{-1}(X^{(k)})$  suivant

$$M_B^{-1}(X^{(k)}) := \frac{N+1}{N} \left( M_B^{-1}(X^{(k)}) - \frac{M_B^{-1}(X^{(k)}) \underline{z}(\underline{x}^{(k+)}) \underline{z}^T(\underline{x}^{(k+)}) M_B^{-1}(X^{(k)})}{N\sigma^2 + \underline{z}^T(\underline{x}^{(k+)}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x}^{(k+)})} \right)$$

*Pas 3.* Déterminer  $\underline{x}^{(k-)}$  donné par

$$\underline{x}^{(k-)} = \arg \min_{\underline{x} \in X^{(k)}} \frac{\underline{z}^T(\underline{x}) M_B^{-1}(X^{(k)}) H(\underline{\theta}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x})}{N + 1 - \frac{1}{\sigma^2} \underline{z}^T(\underline{x}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x})}$$

*Pas 4.* Si  $\underline{x}^{(k-)} = \underline{x}^{(k+)}$ , stop,

sinon supprimer  $\underline{x}^{(k-)}$  de  $X^{(k)}$ , mettre à jour  $M_B^{-1}(X^{(k)})$  suivant

$$M_B^{-1}(X^{(k)}) := \frac{N}{N+1} \left( M_B^{-1}(X^{(k)}) + \frac{M_B^{-1}(X^{(k)}) \underline{z}(\underline{x}^{(k-)}) \underline{z}^T(\underline{x}^{(k-)}) M_B^{-1}(X^{(k)})}{(N+1)\sigma^2 - \underline{z}^T(\underline{x}^{(k-)}) M_B^{-1}(X^{(k)}) \underline{z}(\underline{x}^{(k-)})} \right)$$

incrémenter k de 1 et retourner au pas 2.

Le protocole initial  $X^{(1)}$  utilisé au pas 1 peut être choisi de la façon suivante :

*Pas 1.1.* Poser  $n = 1$ ,  $M_B^{-1}(X^{(1)}) = \Omega$ ,  $X^{(1)} = \emptyset$ .

*Pas 1.2.* Appliquer le pas 2, avec N remplacé par n.

*Pas 1.3.* Si  $n = N$ , stop,

sinon incrémenter n de 1 et retourner au pas 1.2.

Comme c'est généralement le cas pour les algorithmes de détermination d'un protocole entier optimal, la convergence vers un protocole  $L_B$ -optimal n'est pas prouvée. C'est pourquoi nous effectuons plusieurs optimisations à partir de différents protocoles initiaux (éventuellement déduits de la détermination d'un protocole optimal continu). Quand l'algorithme d'échange n'est plus capable de progresser, nous commutons sur une technique d'optimisation locale par programmation non linéaire classique. Notons que l'algorithme d'échange précédent pourrait sans doute être amélioré en exploitant les idées mises en œuvre dans des algorithmes récents pour la détermination de protocoles entiers D-optimaux (voir par exemple [8, 9] et leurs références).

L'expérience optimale au sens de (21) dépend des valeurs de  $N$ ,  $\underline{\theta}$ ,  $\Omega$  et  $\sigma^2$ . La valeur de  $N$  peut être fixée *a priori*, et il est toujours possible de déterminer l'expérience optimale pour  $N' > N$  afin d'évaluer le gain qui résulterait de  $N' - N$  mesures supplémentaires. Remarquons par ailleurs que l'utilisation de l'information *a priori* permet de choisir des protocoles comportant moins de  $p$  mesures. D'après (22), le critère

(21) dépend seulement de  $\Omega$  et  $\sigma^2$  par leur rapport. Si l'on suppose que  $\Omega$  est connue à un facteur d'échelle près ( $\Omega = k \Omega_0$ , avec  $k$  inconnu,  $k \in \mathcal{R}^+$  et  $\Omega_0$  une matrice définie positive connue), et que  $\sigma^2$  est inconnu, on peut suivre une politique minimax, et minimiser  $\max_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} \Psi_{LB}(X, \alpha)$  avec  $\alpha = k/\sigma^2$ . Puisque

$$\frac{d}{d\alpha^{-1}} \Psi_{LB}(X, \alpha) = -\frac{1}{N\sigma^2} \text{trace} [M_B^{-1}(X, \alpha) H(\hat{\theta}) M_B^{-1}(X, \alpha) \Omega_0^{-1}] < 0 \quad (24)$$

le protocole optimal au sens minimax est obtenu pour  $\alpha = \alpha_{\max}$ .

**Remarque :**

La première utilisation du risque Bayésien  $r(X)$  (5) pour la planification en vue de l'optimisation d'un critère de qualité semble due à Chatterjee et Mandal [10]. Dans le cas d'un modèle quadratique en  $\underline{x}$ , et sous des approximations analogues à celles utilisées ici, ils obtiennent une expression analytique de  $r(X)$  lorsque l'estimateur est celui des moindres carrés. Si de plus le domaine admissible  $X$  est ellipsoïdal, le protocole optimal peut alors être explicité. Le cas de domaines  $X$  correspondant à des pavés de  $\mathcal{R}^m$  est également considéré dans [11]. L'approche proposée ici présente sur celle de Chatterjee et Mandal l'avantage de minimiser le risque non seulement vis-à-vis du protocole, mais aussi vis-à-vis de l'estimateur utilisé.  $\diamond$

L'exemple suivant permettra d'illustrer la robustesse de l'approche Bayésienne proposée vis-à-vis d'erreurs dans la spécification de  $\underline{\theta}$ , et de comparer les résultats obtenus avec ceux d'une approche plus classique.

### 3. EXEMPLE

Pour permettre une présentation graphique des résultats, nous considérons un critère de qualité  $y$  ne dépendant que de deux conditions opératoires  $x_1$  et  $x_2$ . Ces conditions doivent satisfaire les contraintes suivantes qui définissent  $X$ ,

$$\sum_{i=1}^2 x_i \leq 1, x_i \geq 0, i = 1, 2. \quad (25)$$

Nous supposons  $\eta(\underline{\theta}, \underline{x})$  donné par le modèle quadratique

$$\eta(\underline{\theta}, \underline{x}) = q_1 x_1 + q_2 x_2 + \frac{1}{2} (Q_{11} x_1^2 + 2 Q_{12} x_1 x_2 + Q_{22} x_2^2), \quad (26)$$

avec

$$\underline{\theta} = (q_1, q_2, Q_{11}, Q_{22}, Q_{12})^T, \quad \underline{x} = (x_1, x_2)^T. \quad (27)$$

Les résultats obtenus suivant l'approche Bayésienne vont être comparés avec ceux déduits de l'utilisation combinée d'une estimation au sens du maximum de vraisemblance et d'une planification reposant sur la matrice d'information de Fisher (voir par exemple [12-15]). Nous ne considérerons ici que la notion de D-optimalité, qui est la plus largement utilisée en planification. Un protocole D-optimal est défini par le critère suivant

$$\Psi_D(X) = \det (Z_X^T Z_X), \quad (28)$$

qu'il s'agit de maximiser. Notons que le but final de l'expérience (c'est à dire la maximisation du critère de qualité  $y$ ) n'est pas pris en compte. Le protocole D-optimal de taille 5 calculé pour l'exemple spécifié par (25-27) est

$$X_D = \{(0, 1), (1, 0), (0, 0.5), (0.5, 0), (0.5, 0.5)\}. \quad (29)$$

L'estimateur au sens du maximum de vraisemblance correspond ici à l'estimateur des moindres carrés,

$$\hat{\underline{\theta}}_{MV}(\underline{y}_X) = (Z_X^T Z_X)^{-1} Z_X^T \underline{y}_X. \quad (30)$$

Sa valeur est utilisée pour déterminer  $\underline{x}^*(\hat{\underline{\theta}}_{MV}(\underline{y}_X))$  grâce à (3).

Pour l'approche Bayésienne, nous supposons que  $\Omega = k \Omega_0$ , avec  $\Omega_0 = I_5$ , et  $1 \leq k/\sigma^2 \leq 10$ . D'après le paragraphe 2, l'expérience optimale au sens minimax est obtenue pour  $k/\sigma^2 = 10$ . Avec  $\underline{\theta}$  donné par

$$\underline{\theta} = (5, 10, -10, -10, -5)^T, \quad (31)$$

l'algorithme d'échange décrit précédemment combiné à une optimisation locale conduit au protocole  $L_B$ -optimal

$$X_{LB}(\underline{\theta}) = \{(0, 1), (1, 0), (0.481, 0.519), (0.481, 0.519), (0.481, 0.519)\}. \quad (32)$$

Notons que ce protocole présente des répétitions et possède moins de points de support qu'il n'y a de paramètres dans le modèle.

La comparaison des deux approches est effectuée par simulation numérique. Pour chacun des protocoles  $X_D$  et  $X_{LB}(\underline{\theta})$ , 1000 vecteurs d'observations  $\mathbf{y}_x^{(i)}$  sont générés suivant

$$\mathbf{y}_x^{(i)} = Z_x \overline{\underline{\theta}} + \underline{\boldsymbol{\varepsilon}}^{(i)}, \quad i = 1, \dots, 1000, \quad (33)$$

avec  $\underline{\boldsymbol{\varepsilon}}^{(i)}$  distribué suivant une loi normale  $\mathcal{N}(\mathbf{0}, 0.1 \times I_5)$ . Le protocole  $L_B$ -optimal dépend de la valeur des paramètres, contrairement au protocole  $D$ -optimal. Afin de tester la robustesse de l'approche Bayésienne vis-à-vis d'erreurs sur la moyenne *a priori*  $\underline{\theta}$ , nous choisissons donc  $\overline{\underline{\theta}}$  différent de  $\underline{\theta}$ ,

$$\overline{\underline{\theta}} = (3.75, 7.5, -7.5, -13.33, -5)^T. \quad (34)$$

Les conditions opératoires optimales estimées  $\hat{\mathbf{x}}^{*(i)}$  sont calculées pour chacun des ensembles de données  $\mathbf{y}_x^{(i)}$ ,  $i = 1, \dots, 1000$ , et leur position est indiquée sur la figure 1a pour  $X_D$  et la figure 1b pour  $X_{LB}$ .

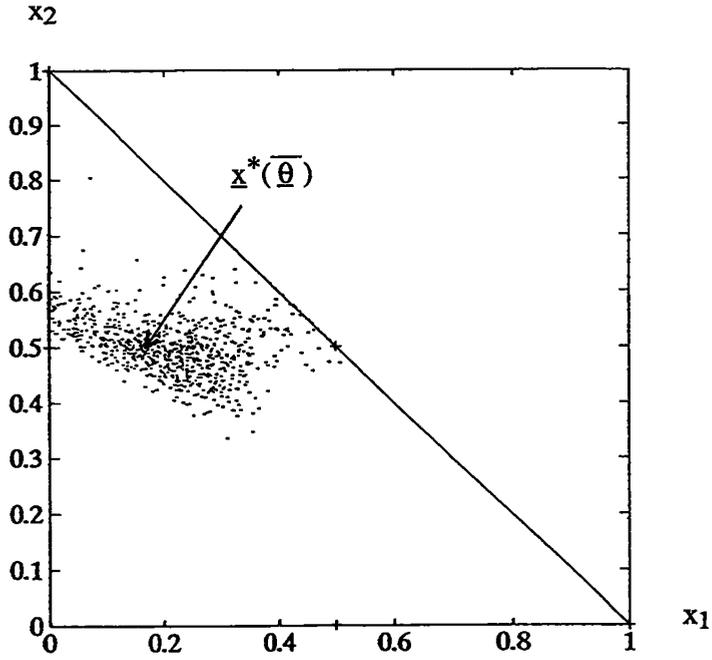


Figure 1a. — Protocole D-optimal et estimateur du maximum de vraisemblance.

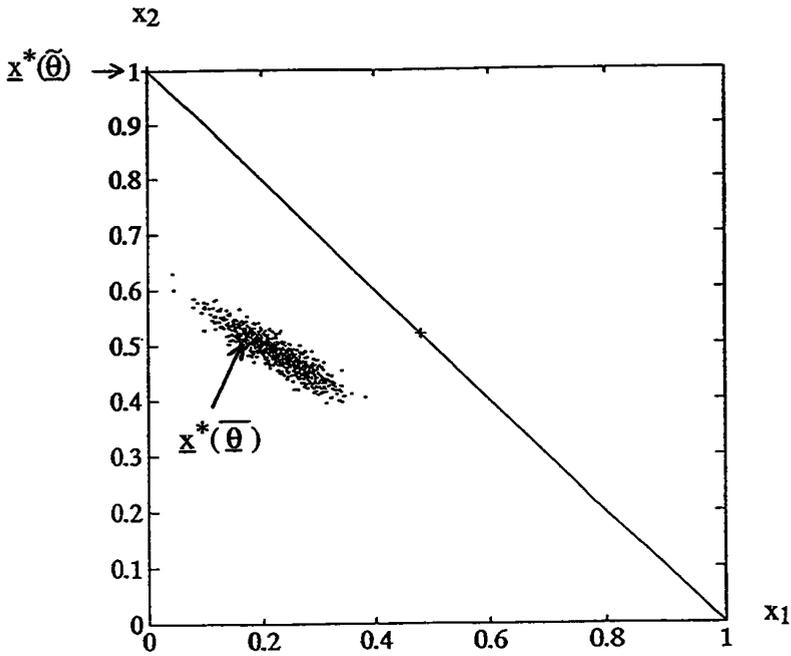


Figure 1b. — Protocole L<sub>B</sub>-optimal et estimateur Bayésien.

L'approche Bayésienne conduit à une plus grande concentration des  $\hat{\underline{x}}^{*(i)}$ ,  $i = 1, \dots, 1000$ , autour des conditions opératoires réellement optimales  $\underline{x}^*(\bar{\underline{\theta}})$ . Bien que l'information *a priori* sur la position de  $\underline{x}^*(\bar{\underline{\theta}})$  soit tout à fait inexacte, le biais dans la localisation de  $\underline{x}^*$  demeure raisonnable.

Supposons à présent la moyenne *a priori* des paramètres donnée par

$$\bar{\underline{\theta}} = (3, 3, -10, -10, 0)^T, \quad (35)$$

l'algorithme d'échange décrit au paragraphe 2 combiné à une optimisation locale conduit à

$$X_{LB}(\bar{\underline{\theta}}) = \{(0, 1), (1, 0), (0, 0.897), (0.533, 0.467), (0.533, 0.467)\}. \quad (36)$$

Pour chacun des protocoles  $X_D$  et  $X_{LB}(\bar{\underline{\theta}})$  nous générons 1000 vecteurs d'observations  $\underline{y}_x^{(i)}$  suivant

$$\underline{y}_x^{(i)} = Z_x \bar{\underline{\theta}}^{(i)} + \underline{\varepsilon}^{(i)}, \quad i = 1, \dots, 1000, \quad (37)$$

avec  $\underline{\varepsilon}^{(i)}$  et  $\bar{\underline{\theta}}^{(i)}$  indépendamment distribués, respectivement suivant des lois  $\mathcal{N}(\underline{0}, 0.1 \times I_5)$  et  $\mathcal{N}(\bar{\underline{\theta}}^0, I_5)$ . Supposons que

$$\bar{\underline{\theta}}^0(\beta) = (3+2\beta, 3+2\beta, -10, -10, 0)^T. \quad (38)$$

Quand  $\beta = 0$ ,  $\bar{\underline{\theta}}^0(\beta) = \bar{\underline{\theta}}$ , de telle sorte que  $\beta$  caractérise l'erreur sur la moyenne *a priori*. La figure 2 présente le lieu des conditions opératoires optimales  $\underline{x}^*(\bar{\underline{\theta}}^0(\beta))$  pour  $\beta$  variant de -1 à 1, ce qui correspond à des erreurs considérables sur la position *a priori* de ces conditions opératoires comme on peut le constater sur la figure.

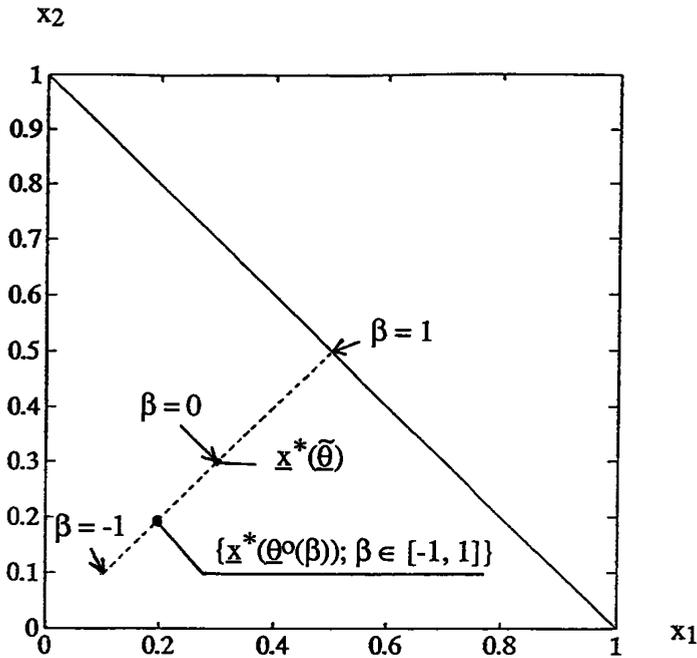


Figure 2. — Lieu des conditions opératoires optimales  $\underline{x}^*(\theta^o(\beta))$  ( $\theta^o(\beta)$  donné par (38),  $\beta \in [-1, 1]$ ) et des conditions opératoires optimales *a priori*  $\underline{x}^*(\tilde{\theta})$  ( $\tilde{\theta}$  donné par (35)).

Définissons le risque relatif

$$\rho(X, \beta) = \frac{\sum_{i=1}^{1000} \eta(\bar{\theta}^{(i)}, \underline{x}^*(\bar{\theta}^{(i)})) - \eta(\bar{\theta}^{(i)}, \hat{\underline{x}}^{*(i)})}{\sum_{i=1}^{1000} \eta(\bar{\theta}^{(i)}, \underline{x}^*(\bar{\theta}^{(i)}))}, \beta \in [-1, 1], X = X_D, X_{LB}(\tilde{\theta}), \quad (39)$$

où les conditions opératoires optimales estimées  $\hat{\underline{x}}^{*(i)}$ ,  $i = 1, \dots, 1000$ , sont déterminées à partir des observations simulées obtenues suivant (37). La figure 3 décrit l'évolution de  $\rho(X, \beta)$  en fonction de  $\beta$  pour le protocole D-optimal (29) combiné avec une estimation au sens du maximum de vraisemblance ainsi que pour l'approche Bayésienne. La seconde approche donne de bien meilleurs résultats quelque soit  $\beta$  compris entre -1 et 1.

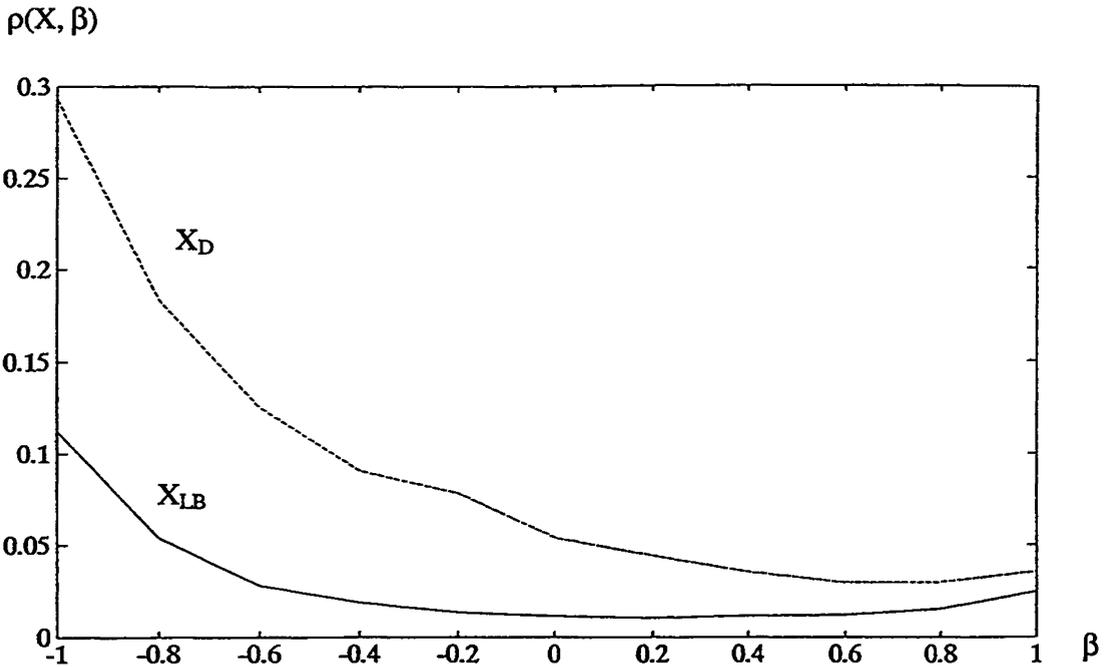


Figure 3. — Evolution du risque relatif  $\rho(X, \beta)$  (39) en fonction de  $\beta$  pour les protocoles  $X_D$  (29) ----- et  $X_{LB}$  (36) ———.

#### 4. CONCLUSION

Une nouvelle approche Bayésienne de la détermination des conditions opératoires maximisant un critère de qualité a été proposée. Elle s'applique quand la relation entre ce critère et les conditions opératoires est décrite par un modèle dépendant de paramètres inconnus. La maximisation du critère passe alors par l'estimation de ces paramètres à partir de données expérimentales, et il convient de déterminer au mieux les conditions de leur recueil. Se pose donc un double problème de planification et d'estimation. Contrairement à l'approche classique reposant sur la méthodologie de la surface de réponse, l'approche Bayésienne proposée permet d'exploiter des informations *a priori* éventuellement disponibles, de prendre en compte le but poursuivi par la modélisation et les contraintes sur les conditions opératoires. Elle n'est pas limitée aux modèles

polynomiaux en les conditions opératoires, la seule restriction étant que le critère de qualité soit linéaire en les paramètres. Le critère de planification proposé est un critère de  $L_B$ -optimalité, de sorte que son optimisation peut être effectuée à l'aide d'un algorithme de la littérature et se révèle aussi simple que celle du critère classique de D-optimalité. Quand elle est fiable, l'utilisation de l'information *a priori* permet de se contenter d'un nombre de mesures inférieur au nombre de paramètres, ce qui peut être particulièrement intéressant quand il s'agit de maximiser la qualité d'une série de systèmes de même nature. Le problème de la robustesse de cette approche Bayésienne vis-à-vis d'hypothèses erronées a été abordé. Pour ce qui concerne l'incertitude sur la covariance *a priori* des paramètres et la variance du bruit de mesure, une approche minimax a été proposée qui n'augmente pas la complexité des calculs. La robustesse vis-à-vis de la moyenne *a priori* des paramètres a été illustrée par un exemple simulé. Le même exemple a mis en évidence la supériorité de l'approche Bayésienne sur une approche classique reposant sur l'utilisation combinée d'un protocole D-optimal et d'une estimation des paramètres par maximum de vraisemblance.

Nous avons supposé que le critère de qualité était convenablement décrit par la fonction  $\eta(\underline{\theta}, \underline{x})$ , et la robustesse de l'approche à une erreur sur la structure du modèle n'a pas été considérée. Les problèmes posés par ce type de robustesse en planification pour l'estimation de paramètres ont été largement abordés dans la littérature depuis les travaux de Box et Draper [16] (voir par exemple l'article de synthèse [17]). Les idées développées devraient aussi pouvoir être exploitées dans le contexte de la planification pour la maximisation d'un critère de qualité.

Enfin, le problème abordé ici est de nature non séquentielle. Il s'agit, partant d'information *a priori*, de choisir de façon optimale l'expérience (comportant un nombre prédéfini de mesures) à effectuer pour localiser  $\underline{x}^*$ . On peut se demander comment une telle approche pourrait être transposée au séquentiel. Une première approche (naïve) consisterait à effectuer uniquement les expériences successives associées aux valeurs prédites de  $\underline{x}^*$ . Cependant, dans ce cas il n'y a pas convergence de l'estimée des paramètres vers leur vraie valeur, à cause du caractère trop peu informatif de l'expérience associée à  $\underline{x}^*$  (ceci pose un problème analogue à celui de l'excitation persistante en commande adaptative : la commande optimale au sens de la régulation du système pouvant être trop peu informative pour l'estimation des paramètres [20]). Afin d'éviter ce problème, on peut penser à une utilisation séquentielle de l'approche proposée ici : à chaque étape les mesures effectuées permettent de mettre à jour l'information *a priori* devant être utilisée pour choisir les expériences suivantes.

## 5. BIBLIOGRAPHIE

- [1] Nelder J.A. et Mead R., 1965. A simplex method for function minimization, *Computer J.*, **7**, pp. 308-313.
- [2] Kiefer J. et Wolfowitz J., 1952. Stochastic estimation of the maximum of a regression function, *Annals of Math. Stat.*, **23**, pp. 462-466.
- [3] Box G.E.P. et Wilson K.B., 1951. On the experimental attainment of optimum conditions (with discussion), *J. Royal Stat. Soc.*, **B13**, n°1, pp. 1-45.
- [4] Hill W.J. et Hunter W.G., 1966. A review of response surface methodology: a literature survey, *Technometrics*, **8**, pp. 571-590.
- [5] Myers R.H., 1976. *Response Surface Methodology*, Allyn and Bacon, Inc., Boston.
- [6] Pitz J., 1983. *Bayesian Estimation and Experimental Design in Linear Regression Models*, Teubner Texte zur Mathematik, Leipzig ; John Wiley, New York, 1991.
- [7] Chaloner K., 1984. Optimal Bayesian experimental design for linear models, *The Annals of Stat.*, **12**(1), pp. 283-300.
- [8] Atkinson A.C. et Donev A.N., 1988. Algorithms, exact design and blocking in response surface and mixture designs, in *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov et H.P. Wynn (eds.), North-Holland, Amsterdam, pp. 61-69.
- [9] Yonchev H., 1988. New computer procedures for constructing D-optimal designs, in *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov et H.P. Wynn (eds.), North-Holland, Amsterdam, pp. 71-80.
- [10] Chatterjee S.K. et Mandal N.K., 1981. Response surface designs for estimating the optimal point, *Calcutta Statist. Assoc. Bull.*, **30**, pp. 145-169.

- [11] **Mandal N.K.**, 1989. D-optimal designs for estimating the optimum point in a quadratic response surface—rectangular region, *J. Statist. Planning and Inference*, **23**, pp. 243-252.
- [12] **Fedorov V.V.**, 1972. *Theory of Optimal Experiments*, Academic Press, New York.
- [13] **Silvey S.D.**, 1980. *Optimal Design*, Chapman and Hall, London.
- [14] **Pazman A.**, 1986. *Foundations of Optimum Experimental Design*, D. Reidel, Dordrecht et Veda, Bratislava.
- [15] **Walter E. et Pronzato L.**, 1990. Qualitative and quantitative experiment design for phenomenological models — a survey, *Automatica*, **26(2)**, pp. 195-213.
- [16] **Box G.E.P. et Draper N.R.**, 1959. A basis for the selection of a response surface design, *J. Amer. Statist. Assoc.*, **54**, pp. 622-654.
- [17] **Steinberg D.M. et Hunter W.G.**, 1984. Experimental design: review and comment (with discussion), *Technometrics*, **26(2)**, pp. 71-130.
- [18] **Box G.E.P. et Tiao G.C.**, 1973. *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading.
- [19] **Hartigan J.A.**, 1983. *Bayes Theory*, Springer-Verlag, New York.
- [20] **Goodwin G.C. et Payne R.L.**, 1977. *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York.