

Inversion d'un canal de communication non-linéaire par krigeage

Jean-Pierre COSTA, Luc PRONZATO, Eric THIERRY

Laboratoire I3S, CNRS-UNSA

Les Algorithmes / Bâtiment Euclide B, 2000 route des Lucioles BP 121, 06903 Sophia-Antipolis Cedex
costa@i3s.unice.fr,pronzato@i3s.unice.fr,et@i3s.unice.fr

Résumé – On considère le problème d'inversion de système non linéaire, où la structure du modèle sous-jacent est inconnue. Les modèles non linéaires classiques (Volterra et NARMAX) ont un grand nombre de paramètres, lesquels sont difficiles à estimer à partir d'une courte séquence d'apprentissage. Il en est de même pour les méthodes utilisant des réseaux de neurones. Nous proposons dans cet article, l'utilisation d'une approche semi-paramétrique appelée krigeage (kriging en anglais). Nous montrons sur un exemple que de bonnes performances sont obtenues pour de courtes séquences d'apprentissage.

Abstract – We consider a problem of nonlinear system-inversion, with unknown underlying model structure. Classical parametric behavioural models (Volterra and NARMAX models) involve a lot of parameters, which are difficult to estimate from short training sequences. A similar difficulty is encountered when methods based on neural networks are used. We suggest in this paper to use a semi-parametric approach called kriging. We show on an example that good performances are obtained for short training sequences.

1 Introduction

Nous considérons un système constitué de deux émetteurs x_1 et x_2 et de deux capteurs c_1 et c_2 . Sur le capteur c_2 on observe la somme de x_1 et x_2 après filtrage et passage dans une non linéarité instantanée (S). On suppose que l'on est capable d'estimer x_1 et on désire retrouver le signal x_2 . L'application envisagée correspond à l'écoute d'un signal x_2 de faible niveau en présence d'un signal de brouillage x_1 de fort niveau.

Afin de présenter le krigeage, considérons d'une manière plus générale un système non-linéaire F où les séquences d'entrée $\{x_k\}$ et de sortie $\{y_k\}$ sont observées sur un horizon n et où la réponse du système doit être prédite pour de nouveaux échantillons d'entrée.

Traditionnellement, on construit un modèle entrée/sortie pour F et on prédit à partir de la réponse du modèle pour de nouvelles valeurs d'entrée. Les modèles de Volterra et de Wiener [7], ou NARMAX (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) [6] sont des représentations paramétriques classiques pour des systèmes non-linéaires inconnus. Cependant, le choix des caractéristiques de la structure (degré du polynôme D , longueur de la mémoire m_x et m_y pour les séquences d'entrée et de sortie) est difficile. De plus, le nombre de paramètres du modèle, déterminé à partir de D , m_x et m_y , est généralement très grand, ce qui rend leur estimation difficile. Les réseaux neurones sont une autre approche possible, mais dans ce cas l'apprentissage est lent, c'est à dire nécessite beaucoup de données.

On se propose ici d'utiliser un modèle semi-paramétrique construit par krigeage. La prédiction par krigeage est un

outil statistique qui permet de modéliser des observations spatiales, avec ou sans erreurs d'observation, sans imposer la structure d'un modèle particulier [8]. La prédiction par krigeage a été introduite en géostatistique [5, 3] mais n'a pas été utilisée, à notre connaissance, en traitement du signal. Dans ce qui suit, nous présentons la prédiction par krigeage puis nous la comparons en simulation à celle obtenue avec un réseau RBF (Radial Basis Function).

2 Modélisation semi-paramétrique par krigeage

Le modèle construit peut être appelé semi-paramétrique car il contient une partie paramétrique et une partie non-paramétrique considérée comme la réalisation d'un processus aléatoire. Dans le cas où l'on suppose que le processus est gaussien, on peut estimer les paramètres de sa matrice de covariance par maximum de vraisemblance.

Par la suite, on notera S le système inconnu, y la sortie et \mathbf{x} le vecteur d'entrée $\in \mathbb{R}^q$. A partir des séquences d'apprentissage observées $\{\mathbf{x}_k\}$ et $\{y_k\}$, $k = 1, \dots, n$, on prédit les valeurs de y pour de nouvelles valeurs de \mathbf{x} ($\mathbf{x}_{n+1}, \mathbf{x}_{n+2}, \dots$) par interpolation des données précédentes.

2.1 Prédiction sans erreurs sur les observations

On se place dans le cas où il n'y pas d'erreurs sur les observations, qui sont alors données par

$$y_k = y(\mathbf{x}_k) = F(\mathbf{x}_k), \quad (1)$$

avec $F(\cdot)$ une fonction non-linéaire inconnue. Ces observations sont modélisées par

$$y_k = \mathbf{f}^T(\mathbf{x}_k)\beta + Z(\mathbf{x}_k), \quad (2)$$

où le régresseur $\mathbf{f}(\mathbf{x}_k)$ est fonction de \mathbf{x}_k , $\beta \in \mathbb{R}^p$ est un vecteur inconnu de paramètres et $Z(\mathbf{x}_k)$ est la réalisation d'un processus stochastique. En pratique, choisir $\mathbf{f}(\mathbf{x}_k) = 1$ et β scalaire est généralement suffisant. Le processus $Z(\cdot)$ est supposé être de moyenne nulle et de covariance

$$E\{Z(\mathbf{x})Z(\mathbf{x}')\} = \sigma_Z^2 R(\mathbf{x} - \mathbf{x}')$$

avec

$$R(\mathbf{x} - \mathbf{x}') = \exp\left(\sum_{i=1}^q -\theta_i |x_i - x'_i|^{\gamma_i}\right). \quad (3)$$

On restreint les valeurs des paramètres γ_i à $\gamma_i \in [1, 2]$, $i = 1, \dots, q$. Le choix de la forme de la matrice de covariance est important. Celui donné en (3) offre suffisamment de flexibilité au travers des paramètres de corrélation θ_i et des paramètres de lissage γ_i (cf. [1],[2]).

On note \mathbf{y}_n le vecteur des observations de la séquence d'apprentissage,

$$\mathbf{y}_n = (y_1, \dots, y_n)^T,$$

et \mathbf{F}_n la matrice définie par $\mathbf{F}_n = (\mathbf{f}^T(\mathbf{x}_1) \dots \mathbf{f}^T(\mathbf{x}_n))^T$. La prédiction $y(\mathbf{x})$ pour une valeur \mathbf{x} est $\hat{y}(\mathbf{x}) = \mathbf{c}^T(\mathbf{x})\mathbf{y}_n$. La minimisation de l'erreur quadratique moyenne de ce prédicteur linéaire sous la condition de non biais $\mathbf{f}^T(\mathbf{x}) = \mathbf{c}^T(\mathbf{x})\mathbf{F}_n$ donne

$$\hat{y}(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\hat{\beta} + \mathbf{r}^T(\mathbf{x})\mathbf{V}_n^{-1}(\mathbf{y}_n - \mathbf{F}_n\hat{\beta}), \quad (4)$$

ce qui correspond au meilleur prédicteur linéaire non biaisé. Dans (4), $\mathbf{V}_n = \sigma_Z^2 \mathbf{R}_n$ est la matrice de covariance de $\mathbf{Z}_n = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))^T$, avec

$$[\mathbf{R}_n]_{ij} = R(\mathbf{x}_i - \mathbf{x}_j), \quad (5)$$

$\mathbf{r}(\mathbf{x}) = E\{Z(\mathbf{x})\mathbf{Z}_n\}$, c'est-à-dire $[\mathbf{r}(\mathbf{x})]_i = \sigma_Z^2 R(\mathbf{x} - \mathbf{x}_i)$, et

$$\hat{\beta} = (\mathbf{F}_n^T \mathbf{R}_n^{-1} \mathbf{F}_n)^{-1} \mathbf{F}_n^T \mathbf{R}_n^{-1} \mathbf{y}_n. \quad (6)$$

L'erreur quadratique moyenne de prédiction est alors

$$\sigma^2(\mathbf{x}) = \sigma_Z^2 - [\mathbf{f}^T(\mathbf{x}) \mathbf{r}^T(\mathbf{x})] \begin{bmatrix} \mathbf{O} & \mathbf{F}_n^T \\ \mathbf{F}_n & \mathbf{V}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}(\mathbf{x}) \\ \mathbf{r}(\mathbf{x}) \end{bmatrix}. \quad (7)$$

Elle satisfait $\sigma^2(\mathbf{x}_k) = 0$, $k = 1, \dots, n$, c'est-à-dire que \hat{y}_k est un interpolateur parfait : $\hat{y}(\mathbf{x}_k) = y_k$, $k = 1, \dots, n_x$. Si l'on suppose que le processus $Z(\cdot)$ est gaussien, on peut construire un intervalle de confiance pour la prédiction. De plus, les paramètres θ_i et γ_i de la matrice de covariance peuvent alors être estimés par maximum de vraisemblance, ainsi que β et σ_Z^2 . On peut facilement montrer que

$$\{\hat{\theta}, \hat{\gamma}\} = \arg \min_{\{\theta \in \mathbb{R}^{+q}, \gamma \in [1,2]^q\}} [n \ln(\hat{\sigma}_Z^2) + \ln \det(\mathbf{R}_n)], \quad (8)$$

où $\hat{\sigma}_Z^2 = \frac{1}{n}(\mathbf{y}_n - \mathbf{F}_n\hat{\beta})^T \mathbf{R}_n^{-1}(\mathbf{y}_n - \mathbf{F}_n\hat{\beta})$.

2.2 Prédiction avec erreurs sur les observations

Le cas où il y a des erreurs sur les observations (séquence i.i.d.) peut être traité d'une manière similaire. Soit $y_k = F(\mathbf{x}_k) + \epsilon_k$, avec $\{\epsilon_k\}$ une séquence i.i.d. d'erreurs, de moyenne nulle et de variance σ_ϵ^2 . Les observations sont alors modélisées par

$$y_k = \mathbf{f}^T(\mathbf{x}_k)\beta + Z(\mathbf{x}_k) + \epsilon_k, \quad (9)$$

où $Z(\cdot)$ est un processus stochastique indépendant de $\{\epsilon_k\}$. On définit $\mathbf{V}_n = \sigma_\epsilon^2 \mathbf{I}_n + \sigma_Z^2 \mathbf{R}_n$, avec \mathbf{I}_n la matrice identité d'ordre n et \mathbf{R}_n donnée par (5). La prédiction en \mathbf{x} est donnée par (4), avec

$$\hat{\beta} = (\mathbf{F}_n^T \mathbf{V}_n^{-1} \mathbf{F}_n)^{-1} \mathbf{F}_n^T \mathbf{V}_n^{-1} \mathbf{y}_n, \quad (10)$$

qui coïncide avec (6) lorsque $\sigma_\epsilon^2 = 0$. Quand $\sigma_\epsilon^2 \neq 0$, le prédicteur n'est pas un interpolateur parfait : l'erreur quadratique moyenne est donnée par (7) et, en général, $\sigma^2(\mathbf{x}_k) \neq 0$ pour $k = 1, \dots, n$. Avec une distribution normale $\mathcal{N}(0, \sigma_\epsilon^2)$ pour ϵ_k et un processus gaussien $Z(\cdot)$, l'estimation du maximum de vraisemblance pour les paramètres de \mathbf{R} est encore valable. Si on définit α par $\alpha = \frac{\sigma_\epsilon^2}{\sigma_Z^2}$, alors $\mathbf{V}_n = \sigma_Z^2(\mathbf{R}_n + \alpha \mathbf{I}_n)$ et l'estimateur du maximum de vraisemblance de α et θ est

$$\{\hat{\alpha}, \hat{\theta}, \hat{\gamma}\} = \arg \min_{\{\alpha > 0, \theta \in \mathbb{R}^{+m_x}, \gamma \in [1,2]^{m_x}\}} [n \ln(\hat{\sigma}_Z^2) + \ln \det(\mathbf{R}_n + \alpha \mathbf{I}_n)], \quad (11)$$

où

$$\hat{\sigma}_Z^2 = \frac{1}{n}(\mathbf{y}_n - \mathbf{F}_n\hat{\beta})^T (\mathbf{R}_n + \alpha \mathbf{I}_n)^{-1} (\mathbf{y}_n - \mathbf{F}_n\hat{\beta}),$$

et $\hat{\beta}$ est donnée par (10) correspondent respectivement aux estimateurs du maximum de vraisemblance de σ_Z^2 et β .

2.3 Optimisation

Une optimisation numérique est nécessaire pour la détermination de $\hat{\theta}$ et $\hat{\gamma}$ (Eq. 8). Bien que le problème soit difficile (cf [10]), les solutions montrent qu'une estimation exacte de $\hat{\theta}$ et $\hat{\gamma}$ n'est pas nécessaire pour obtenir une prédiction fine; notre expérience [1] [2] montre que les minimum locaux sont généralement acceptables. En pratique, on prendra $\gamma = 2$ ce qui correspond à une covariance gaussienne. Le processus $Z(\cdot)$ est alors continu et indéfiniment dérivable en moyenne quadratique, ce qui suppose que la non linéarité à estimer est suffisamment lisse. Il est recommandé d'imposer une contrainte sur θ , telle que $\theta_i \geq \delta > 0$, afin que \mathbf{R}_n demeure définie-positive lors de l'optimisation.

3 Inversion de système

3.1 Définition du problème

Le problème est décrit par la Figure 1, où les observations satisfont la relation d'entrée/sortie suivante

$$\begin{aligned} y_k &= \varphi(z_k) \\ z_k &= \sum_{i=0}^{n_a} a_i x_{1k-i} + \sum_{i=0}^{n_b} b_i x_{2k-i} \end{aligned}$$

avec $\varphi(\cdot)$ une non-linéarité sans mémoire,

$$\varphi(z) = 2/[1 + \exp(-0.5z)] - 1, \quad (12)$$

voir Figure 2, et $\{x_{1k}\}, \{x_{2k}\}$ les séquences d'entrées.

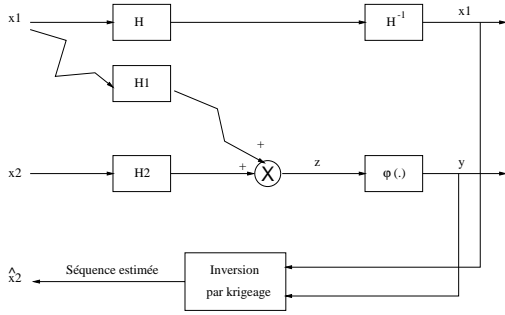


FIG. 1: Inversion du système-Reconstruction

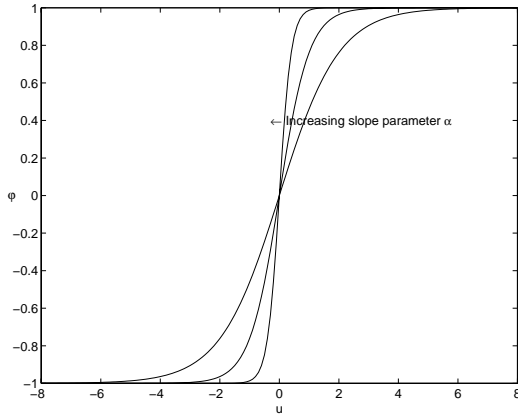


FIG. 2: Fonction sigmoïde

On suppose que les séquences d'apprentissage $\{x_{1k}, x_{2k}, y_k\}, k = 1, \dots, n$, sont disponibles et on souhaite inverser le système afin de reconstruire la séquence d'entrée $\{x_{2k}\}$ comme une fonction de

$$\underline{\mathbf{x}}_k = (y_k, y_{k-1}, \dots, y_{k-m_y+1}, x_{1k}, x_{1k-1}, \dots, x_{1k-m_x+1}),$$

pour $k > n$. La séquence d'entrée $\{x_{1k}\}$ est également observée au travers d'un canal linéaire H et reconstruite, par exemple, par une méthode de déconvolution aveugle telle que la maximisation du kurtosis [9]. La présence de la non-linéarité est due à ce que le niveau du signal $\{x_{2k}\}$ peut être plus faible que celui du signal de brouillage $\{x_{1k}\}$. Le gain de l'amplificateur est alors à sa valeur maximale ce qui entraîne la présence de saturations. On note que lorsque $n_b > 0$, m_y et m_x doivent être choisis grands car

x_{2k} , fonction de z_k et x_{1k} , contient une partie autoregressive. Les performances sont évaluées en termes d'erreur quadratique moyenne normalisée E_r :

$$E_r = 10 \log \frac{(\mathbf{x}_{2n+1}^N - \hat{\mathbf{x}}_{2n+1}^N)^T (\mathbf{x}_{2n+1}^N - \hat{\mathbf{x}}_{2n+1}^N)}{(\mathbf{x}_{2n+1}^N)^T \mathbf{x}_{2n+1}^N}, \quad (13)$$

où \mathbf{x}_{2n+1}^N représente le vecteur des observations $(x_{2n+1}, \dots, x_{2N})$ et $\hat{\mathbf{x}}_{2n+1}^N$ représente le vecteur des prédictions $(\hat{x}_{2n+1}, \dots, \hat{x}_{2N})$.

Pour les simulations, x_{1k} et x_{2k} sont uniformément distribués respectivement avec les valeurs discrètes $\{\pm 3, \pm 5\}$ et $\{\pm 1\}$, $n = 50$, $N = 250$, $n_a = 5$, $n_b = 3$,

$$a = [1, 0.04, -0.08, 0.03, -0.01],$$

$$b = [1, 0.09, -0.04].$$

$m_x = 5$ et $m_y = 7$. La Figure 3 représente une réalisation des séquences $\{z_k\}$ et $\{y_k\}$.

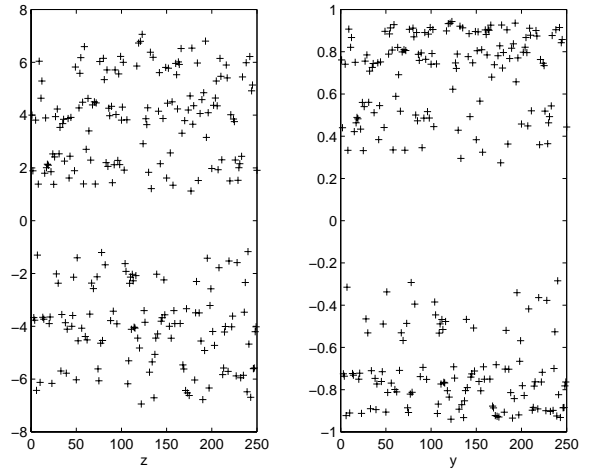


FIG. 3: Séquences $\{z_k\}$ et $\{y_k\}$

3.2 Prédiction par un modèle linéaire

Considérons un modèle linéaire où x_{2k} est modélisé par

$$x_{2k} = \sum_{i_1=0}^{m_y-1} \alpha_1 y_{k-i_1} + \sum_{i_2=0}^{m_x-1} \alpha_2 x_{1k-i_2},$$

avec les paramètres α_i estimés par moindres carrés. La moyenne sur 10 réalisations indépendantes de E_r est de -3.443 dB. Comme l'on pouvait s'y attendre, ce modèle linéaire ne donne pas de résultats satisfaisants. On note que les modèles de Volterra et NARMAX ne peuvent pas être utilisés ici car ils font intervenir trop de paramètres qui ne peuvent pas être estimés à partir de la séquence d'apprentissage ($n = 50$).

3.3 Prédiction par un réseau RBF

Le réseau RBF (Radial Basis Function) est constitué d'une couche d'entrée, une couche cachée correspondant à une unité de traitement non linéaire, et une couche de sortie linéaire par rapport aux paramètres. La relation

d'entrée-sortie du réseau RBF est donnée par la relation suivante

$$x_{2_k} = w_0 + \sum_{i=1}^n w_i \varphi_i(\underline{\mathbf{x}}_k; \mathbf{t}_i),$$

où le terme $\varphi_i(\underline{\mathbf{x}}_k; \mathbf{t}_i)$ correspond à la $i^{\text{ème}}$ fonction radiale de base qui calcule la distance entre le vecteur d'entrée $\underline{\mathbf{x}}_k$ et le centre \mathbf{t}_i . Traditionnellement on utilise comme fonction radiale de base la fonction gaussienne. Lorsque les centres \mathbf{t}_i coïncident avec les entrées de la base d'apprentissage $\underline{\mathbf{x}}_i$, $i = 1, \dots, n$, on obtient

$$x_{2_k} = w_0 + \sum_{i=1}^n w_i \exp\left(-\frac{1}{\sigma_i^2} \|\underline{\mathbf{x}}_k - \underline{\mathbf{x}}_i\|^2\right) \quad (14)$$

où σ_i est la largeur de la $i^{\text{ème}}$ fonction radiale de base [4], et est fixée par l'utilisateur. On prend pour notre exemple $\sigma_i = 1$. Les paramètres w_i peuvent être estimés par moindres carrés. La moyenne sur 10 réalisations indépendantes de E_r est de -0.023 dB . Le réseau RBF dans cette configuration n'est pas capable de prédire correctement la séquence de sortie $\{x_{2_k}\}$. Ceci est essentiellement dû à la faible longueur de la séquence d'apprentissage et à la valeur de σ_i fixée a priori.

3.4 Prédiction par krigeage

On utilise le modèle suivant

$$x_{2_k} = \beta + Z(\underline{\mathbf{x}}_k) + \varepsilon_k \quad (15)$$

et on fixe $m_{x_1} = 5$, $m_y = 7$, $f = 1$, $q = m_y + m_{x_1} = 12$ et $\gamma_i = 2$, $i = 1, \dots, 12$. La moyenne de E_r sur 10 réalisations est -17.221 dB . La Figure 4 présente la reconstruction de la séquence $\{\hat{x}_{2_k}\}$. Comme les 50 premiers échantillons sont utilisés comme séquence d'apprentissage, il y a une parfaite reconstruction. Si l'on considère le fait que l'on a fixé de faibles valeurs pour m_{x_1} et m_y , qu'il n'y a pas de partie paramétrique modélisant la non-linéarité, et que la longueur de la séquence d'apprentissage est faible ($n = 50$), les résultats sont très satisfaisants.

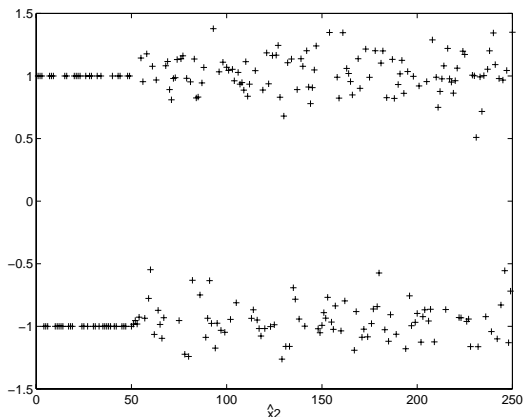


FIG. 4: Reconstruction de $\{x_{2_k}\}$

4 Conclusions

La prédiction par krigeage semble être une méthode intéressante pour le filtrage non-linéaire. La présence d'une partie non-paramétrique dans le modèle permet une grande flexibilité. De plus, le choix d'un simple terme constant pour la partie paramétrique donne généralement des résultats satisfaisants, même en présence de fortes non-linéarités et de séquences d'apprentissage très courtes.

L'étude de l'inversion d'un canal de communication non linéaire a été considérée, avec la sortie du système dépendant de deux entrées, l'une connue et l'autre devant être reconstruite. L'exemple utilisé a permis de valider cette approche par la comparaison avec des résultats obtenus par d'autres techniques, linéaires et non-linéaires.

Références

- [1] J.-P. Costa, L. Pronzato, and E. Thierry. Nonlinear predicting by kriging, with application to noise cancellation. *Internal report*, 1998.
- [2] J.-P. Costa, L. Pronzato, and E. Thierry. Nonlinear filtering by kriging, with application to system inversion. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 3, 1999.
- [3] N. Cressie. Kriging nonstationary data. *Journal of the American Statistical Association*, 81:625-634, 1986.
- [4] S. Haykin. *Adaptive Filter Theory*. Third Edition, Prentice Hall, Inc, A Simon & Schuster Company, New Jersey, 1996.
- [5] D. Krige. A statistical approach to some mine valuation and allied problems on the Witwatersrand. Master Thesis, University of Witwatersrand, 1951.
- [6] I. Leontaritis and S. Billings. Input-Output parametric models for nonlinear systems part 2: stochastic nonlinear systems. *International Journal of Control*, 41(2):329-344, 1985.
- [7] W. J. Rugh. *Nonlinear System Theory: The Volterra / Wiener Approach*. The Johns Hopkins University Press, Baltimore, 1981.
- [8] J. Sacks, W. Welch, T. Mitchell, and H. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409-435, 1989.
- [9] O. Shalvi and E. Weinstein. "New criteria for blind deconvolution of nonminimum phase systems (channels)". *IEEE Transactions on Information Theory*, 36(2):312-321, March 1990.
- [10] J. Warnes and B. Ripley. Problems with likelihood estimation of covariance functions of spatial gaussian processes. *Biometrika*, 74(3):640-642, 1987.