Regression on Parametric Manifolds: Estimation of Spatial Fields, Functional Outputs, and Parameters from Noisy Data

Anthony T. Patera [a], Einar M. Rønquist [b]

[a] *Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
[b] *Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway*

Received *****; accepted after revision +++++

Presented by

Abstract

In this Note we extend the Empirical Interpolation Method (EIM) to a regression context which accommodates noisy (experimental) data on an underlying parametric manifold. The EIM basis functions are computed Offline from the noise-free manifold; the EIM coefficients for any function on the manifold are computed Online from experimental observations through a least-squares formulation. Noise-induced errors in the EIM coefficients and in linear-functional outputs are assessed through standard confidence intervals and without knowledge of the parameter value or the noise level. We also propose an associated procedure for parameter estimation from noisy data. *To cite this article: A.T. Patera, E.M. Rønquist, C. R. Acad. Sci. Paris, Ser. I XXX (2012).*

Résumé

Régression sur des Variétés Paramétriques : Estimation de Champs Spatiaux, Sorties Fonctionnelles, et Paramètres à Partir de Données Bruitées Nous étendons la méthode d'interpolation empirique, EIM en abrégé (pour Empirical Interpolation Method), au contexte de la régression en présence de données bruitées sur une variété paramétrique. Les fonctions de bases sont calculées hors-ligne sur la base de la variété sans bruit ; les coefficients EIM d'une fonction quelconque sur la variété sont calculés en-ligne sur la base des observations expérimentales à travers une formulation moindres carrés. Les erreurs induites par les données bruitées dans les coefficients EIM aussi bien que les sorties fonctionelle–linéaire associées sont quantifiées en intervalles de confiance et sans connaissance ni de la valeur du paramètre ni de la variance du bruit. Nous proposons aussi, dans le même esprit, une procédure d'estimation de paramètre . *Pour citer cet article : A.T. Patera, E.M. Rønquist, C. R. Acad. Sci. Paris, Ser. I XXX (2012).*

Email addresses: patera@mit.edu (Anthony T. Patera), ronquist@math.ntnu.no (Einar M. Rønquist).

Version française abrégée

Nous considérons une variété paramétrique $\mathcal{M} = \{u(\mu) \,|\, \mu \in \mathcal{D}\}$ où $\mu \in \mathcal{D} \to u(\cdot;\mu) \in C^0(\Omega)$ pour $\Omega \subset \mathbb{R}^d$. Nous introduisons ensuite une sortie $y(\mu) = \ell(u(\mu))$ pour ℓ une fonctionelle linéaire bornée.

La méthode d'interpolation empirique [1,2], EIM en abrégé (pour Empirical Interpolation Method), nous fournit un espace d'approximation vectoriel W de dimension n, des fonctions de base associées $q_j(x), 1 \le j \le n$, des noeuds de collocation $\tilde{x}_k, 1 \le k \le n$, et une matrice d'interpolation triangulaire inférieure $B_{kj} = q_j(\tilde{x}_k), 1 \le k, j \le n$. Nos approximations du champ $u(x;\mu)$ (respectivement, la sortie $y(\mu)$) $\tilde{u}(x;\mu)$ est donc (1) (respectivement, $\tilde{y}(\mu) = \ell(\tilde{u}(\cdot;\mu))$) où $\tilde{\beta}(\mu)$ est solution du système d'équations $(n \times n)$ $B\tilde{\beta}(\mu) = U(\mu)$ pour $U_k(\mu) = u(\tilde{x}_k;\mu), 1 \le k \le n$.

Nous considérons ensuite des données bruitées de la forme (2) dans laquelle μ^* est inconnu et ϵ suit une loi normale, de moyenne zéro, décorrélée en x, et homoscédastique d'écart–type σ ; notons que m' est le nombre d'observations (expérimentales) à chaque point de collocation $\tilde{x}_k, 1 \le k \le n$. On peut déduire de ces données bruitées une approximation de champ (4) où $\hat{\beta}(\mu^*)$ est solution des équations normales $S\hat{\beta}(\mu^*) = B^T V$ pour $S = B^T B$ et V donné en (3) ; la sortie associée est calculée ensuite comme $\hat{y}(\mu) = \ell(\hat{u}(\cdot;\mu))$. Nous définissons aussi l'écart-type d'échantillonnage, $\hat{\sigma}(\mu^*)$, voir (5). Pour la suite nous remarquons que S induit une norme $\|\cdot\|_S \equiv \|B\cdot\|$ où $\|\cdot\|$ répresente la norme euclidienne.

Nous fournissons en Proposition 2.1 un intervalle de confiance en norme $\|\cdot\|_S$ pour les coefficients d'EIM $\tilde{\beta}(\mu^*)$ en fonction des coefficients de régression $\hat{\beta}(\mu^*)$, du nombre d'observations expérimentales à chaque point de collocation, m', et de la quantité $\rho(\mu^*)$, comprenant l'écart–type d'échantillonnage $\hat{\sigma}(\mu^*)$ et le quantile de la distribution F au niveau de confiance γ. Ensuite nous proposons, voir Corollary 2.1, un intervalle de confiance pour la sortie $\tilde{y}(\mu^*)$ en fonction de m' et $\rho(\mu^*)$ mais en plus du vecteur de sortie $L_j = \ell(q_j), 1 \le j \le n$. (Nous supposons dans cette Note que l'erreur d'approximation EIM, le deuxième terme dans la borne de Corollary 2.1, est négligeable.) Des résultats pour une fonction gaussienne à deux paramètres perturbée par un bruit synthétique confirment le comportement prévu pour les intervalles de confiance des coefficients EIM aussi bien que la sortie.

Pour conclure, la Proposition 3.1 fournit la borne, au niveau de confiance γ, $|\tilde{\beta}(\overline{\mu}) - \hat{\beta}(\mu^*)|_S \le \xi(\mu^*; r) \equiv \rho(\mu^*)/\sqrt{m'} + r$, pour toute valeur de paramètre $\overline{\mu}$ dans un ensemble de valeurs candidates $\Upsilon \subset \mathcal{D}$ tel que μ^* se trouve dans un voisinage $\mathcal{N}(\overline{\mu}; r) \equiv \{\mu' \in \mathcal{D} \;|\; \|U(\mu') - U(\overline{\mu})\| \le r\}$. Cette borne de la Proposition 3.1 peut servir comme critère pour identifier, dans l'ensemble des valeurs candidates Υ, un ensemble de valeurs cohérentes $\Upsilon_{\rm con}$ — valeurs de paramètres compatibles avec les observations expérimentales — donné en (6). Nous présentons, voir Figure 1, des résultats numériques pour notre example de gaussienne à deux paramètres (avec le paramètre $\mu^* = (0.55, 0.55)$) et un ensemble de valeurs candidates Υ uniforme de cardinalité $40,000$: les cercles ouverts indiquent les valeurs de paramètres en $\Upsilon_{\rm con}$.

1. Introduction

Recent advances in model order reduction — in this paper we focus on the Empirical Interpolation Method (EIM) [1,2] — exploit an underlying parametric manifold for purposes of field or state approximation, functional output approximation, and also parameter estimation. In the EIM we first construct a low–dimensional approximation space to represent the manifold and identify an associated set of *ad hoc* collocation points; we then approximate any particular function (field) on the parametric manifold by interpolation. In [3,4] the EIM is extended to an experimental context in which the space and collocation points are generated by an appropriate model for the manifold — for example, solutions of a partial differential equation — but the interpolation data is then provided by measurements. In this note we extend the "experimental version" of the EIM to a regression context [5] which accommodates noisy data and furthermore provides an assessment of noise–induced error through standard confidence intervals.

We assume that we are given a parametric manifold of functions, $\mathcal{M} = \{u(\cdot,\mu) \,|\, \mu \in \mathcal{D}\}$, where for any given μ in the compact parameter domain $\mathcal{D} \in \mathbb{R}^P$ the field $u(\cdot;\mu)$ is a function in $C^0(\Omega)$ for some prescribed d–dimensional spatial domain Ω. We further introduce an output $y(\mu) = \ell(u(\mu))$, where ℓ is a bounded linear functional. We presume that $u(x;\mu)$ is piecewise linear over some fine simplex discretization of Ω.

The EIM then provides an n–dimensional approximation space W; an associated set of basis functions $q_j(x), 1 \le j \le n$, such that $W = \mathrm{span}\{q_j, 1 \le j \le n\}$; and a set of collocation points $\tilde{x}_k, 1 \le k \le n$. The functions $q_j(x)$ are normalized such that $\max_{x\in\Omega} |q_j(x)| = 1$. We next construct an $n \times n$ interpolation matrix B of the form $B_{kj} = q_j(\tilde{x}_k), 1 \le k, j \le n$, which is lower triangular with unity main diagonal. We now fix n and approximate $u(x;\mu)$ for any given $\mu \in \mathcal{D}$: we define the $n \times 1$ vector $U(\mu)$ with elements $U_k(\mu) = u(\tilde{x}_k;\mu), 1 \le k \le n$; find coefficients $\tilde{\beta}_j(\mu), 1 \le j \le n$, solution of the $n\times n$ system $B\tilde{\beta}(\mu) = U(\mu)$; construct the EIM interpolant as

$$\tilde{u}(x;\mu) = \sum_{j=1}^{n} \tilde{\beta}_j(\mu) q_j(x); \tag{1}$$

and evaluate our output approximation from $\tilde{y}(\mu) = \ell(\tilde{u}(\cdot;\mu))$. The EIM is an interpolation scheme: $\tilde{u}(\tilde{x}_k;\mu) = u(\tilde{x}_k;\mu), 1 \le k \le n$. In this paper we generate the space W by a Greedy procedure which provides the error bound $\sup_{\mu\in\Xi} \sup_{x\in\Omega} |u(x;\mu) - \tilde{u}(x;\mu)| \le \tau$ where Ξ is a "training" set of points in $\mathcal{D}$.

2. Regression Framework

In this paper we shall presume that we are provided with experimental data of the form

$$u^{\exp}(\tilde{x}_k;\omega_{k;i}) = u(\tilde{x}_k;\mu^*) + \epsilon(\tilde{x}_k;\omega_{k;i}), \quad 1 \le k \le n, \quad 1 \le i \le m', \tag{2}$$

where ϵ is *assumed* normal, zero-mean, uncorrelated in space, and homoscedastic with standard deviation σ. Note that $\omega_{k;i}, 1 \le k \le n, 1 \le i \le m'$, corresponds to m' realizations — repeated measurements — at collocation point $\tilde{x}_k$ such that, in total, $m = m'n$ measurements are available. The conceit is that neither μ^* nor σ is known and that we wish to determine $u(x;\mu^*)$ (state estimation), $y(\mu^*)$ (output estimation), and perhaps also μ^* (parameter estimation).

We pursue a least–squares approximation in linear regression fashion [5]. We form the $n \times n$ matrix $S \equiv B^T B$ (superscript T refers to transpose) and the $n \times 1$ vector V

$$V_k = \frac{1}{m'} \sum_{i=1}^{m'} u^{\exp}(\tilde{x}_k;\omega_{k;i}), \quad 1 \le k \le n. \tag{3}$$

We then find $\hat{\beta}(\mu^*)$ from the normal equations $S\hat{\beta}(\mu^*) = B^T V$ to obtain our state approximation

$$\hat{u}(x;\mu^*) = \sum_{j=1}^{n} \hat{\beta}(\mu^*) q_j(x), \tag{4}$$

and subsequently our output approximation $\hat{y}(\mu^*) = \ell(\hat{u}(\cdot;\mu^*))$. [1] We also define the sample standard deviation as

1. Note that the normal equations take a particularly simple form with respect to B given our assumption of m' replicated measurements for each point $\tilde{x}_k, 1 \le k \le n$.

$$\hat{\sigma}(\mu^*) = \sqrt{\frac{1}{m-n}\sum_{k=1}^{n}\sum_{i=1}^{m'}(u^{\exp}(\tilde{x}_k;\omega_{k;i}) - \hat{u}(\tilde{x}_k;\mu^*))^2}. \quad (5)$$

We make two notational remarks: we suppress ω but in fact $\hat{\beta}(\mu^*)$ and $\hat{\sigma}(\mu^*)$ are random; we let the context determine $\hat{\beta}(\mu^*)$ and $\hat{\sigma}(\mu^*)$ as either random variables or as realizations of random variables.

We now define $\|\cdot\|$ as the Euclidean norm and then $\|v\|_S = \sqrt{v^{\mathrm{T}} S v}$ ($= \|Bv\|$). We further define $\rho(\mu^*) = \hat{\sigma}(\mu^*)\sqrt{nF(n, m-n, \gamma)}$ where $F(k_1, k_2, \gamma)$ is the F–statistic quantile [5]. We may then state

Proposition 2.1 *With confidence level γ, $\|\tilde{\beta}(\mu^*) - \hat{\beta}(\mu^*)\|_S \leq \frac{1}{\sqrt{m'}}\rho(\mu^*)$.*

We now sketch the proof. We first note that $\mathbb{E}[u^{\exp}(\tilde{x}_k;\omega_{k;i})] = U_k(\mu^*) = (B\tilde{\beta})_k(\mu^*), 1 \leq i \leq m', 1 \leq k \leq n$, where $\mathbb{E}$ denotes expectation. Hence our confidence ellipse [5] is given by $(\tilde{\beta}(\mu^*) - \hat{\beta}(\mu^*)) m' B^T B (\tilde{\beta}(\mu^*) - \hat{\beta}(\mu^*)) \leq \rho^2(\mu^*)$. The result then directly follows from the definition of S and $\|\cdot\|_S$.

The crucial point is that the EIM model is unbiased due to first, the assumed form of the experimental data, (2), as a perturbation on our manifold, and second, the interpolation property of the EIM approximation. Note that we interpret confidence levels in the frequentist sense: the bound of Proposition 2.1 will obtain in a fraction γ of (sufficiently many) realizations.

We now define the $n \times 1$ output vector L as $L_j = \ell(q_j), 1 \leq j \leq n$, such that $\hat{y}(\mu^*) = L^{\mathrm{T}}\hat{\beta}(\mu^*)$. We may then further prove

Corollary 2.1 *With confidence level γ, $|y(\mu^*) - \hat{y}(\mu^*)| \leq \Delta^y(\mu^*) + |y(\mu^*) - \tilde{y}(\mu^*)|$, where $\Delta^y(\mu^*) \equiv \frac{\rho(\mu^*)}{\sqrt{m'}}\sqrt{L^{\mathrm{T}} S^{-1} L}$.*

We now sketch the proof. We first note from the definition of L and the triangle inequality that $|y(\mu^*) - \hat{y}(\mu^*)| \leq |y(\mu^*) - \tilde{y}(\mu^*)| + |L^T(\tilde{\beta}(\mu^*) - \hat{\beta}(\mu^*))|$. We next note that the maximum of $L^T\alpha$ (respectively, minimum of $L^T\alpha$) subject to the constraint $\alpha^T S \alpha \leq C^2$ is given by $C\sqrt{L^T S^{-1} L}$ (respectively, $-C\sqrt{L^T S^{-1} L}$). The result then directly follows from Proposition 2.1. Note we may apply Corollary 2.1 (jointly) over any number of different outputs, including (with appropriate regularity assumptions) point values of the field.

We describe a paradigm in which Corollary 2.1 might prove useful. (We consider the case, as in the examples below, in which the error in the EIM approximation is sufficiently small such that the second term in the bound of Corollary 2.1 may be neglected.) We presume that the EIM approximation and in particular S is formed in an Offline stage. Then, in the Online stage, we conduct an experiment to form V in m operations, find $\hat{\beta}(\mu^*)$ in n^2 operations, calculate $\hat{\sigma}(\mu^*)$ in m operations, evaluate $\hat{y}(\mu^*)$ in n operations, and finally compute $\Delta^y(\mu^*)$ in n^2 operations (for any L defined in the Online stage). Thus all computations over Ω are replaced by calculations over the very few points $\tilde{x}_k, 1 \leq k \leq n$. Note that μ^* is not known, nor is μ^* deduced as part of the Online calculations.

We now turn to numerical results. In particular, we introduce the parametrized function $u(x;\mu) = \exp(-((x_1 - \mu_1)^2 + (x_2 - \mu_2)^2)/0.02)$ for $x = (x_1, x_2) \in \Omega \equiv (0,1)^2$ and a parameter vector $\mu = (\mu_1, \mu_2) \in \mathcal{D} \equiv [0.4, 0.6]^2$. We construct an EIM approximation with $n = 33$ terms which yields error $\tau = 10^{-3}$ over a 200×200 uniform grid Ξ. Our output functional is $\ell(v) = v(0.4, 0.5) - v(0.6, 0.5)$. We now consider the particular choice $\mu^* = (0.55, 0.55)$ with associated output $y(\mu^*) = -0.4922$. We first verify Proposition 2.1 for the case $\gamma = 0.95$, $m' = 16$, and $\sigma = 0.01$: the inequality is satisfied in 95% of 10,000 realizations. We next consider Corollary 2.1 for $\gamma = 0.95$ and $m' = 16$ and present results for the sample standard deviation, $\hat{\sigma}(\mu^*)$, output approximation, $\hat{y}(\mu^*)$, and output error bound, $\Delta^y(\mu^*)$: (i) $\sigma = 0.0010$ gives $\hat{\sigma}(\mu^*) = 0.0011$, $\hat{y}(\mu^*) = -0.4922$ and $\Delta^y(\mu^*) = 0.0037$; (ii) $\sigma = 0.0100$ gives $\hat{\sigma}(\mu^*) = 0.0096$, $\hat{y}(\mu^*) = -0.4963$ and $\Delta^y(\mu^*) = 0.0340$; (iii) $\sigma = 0.0500$ gives $\hat{\sigma}(\mu^*) = 0.0503$, $\hat{y}(\mu^*) = -0.4768$ and

$\Delta^y(\mu^*) = 0.1824$. In conclusion, $\hat{\sigma}(\mu^*) \approx \sigma$, and $\Delta^y(\mu^*) \approx 3.5\,\hat{\sigma}(\mu^*)$. Finally, we again consider the case $\gamma = 0.95$ and $\sigma = 0.0100$ but now for $m' = 64$: we obtain $\hat{\sigma}(\mu^*) = 0.0099$, $\hat{y}(\mu^*) = -0.4893$ and $\Delta^y(\mu^*) = 0.018$; as expected, the output error bound is decreased twofold when m' is increased fourfold.

3. Parameter Estimation

We can also apply the framework to parameter estimation [6]. Towards that end, we introduce a candidate set $\Upsilon \subset \mathcal{D}$ of cardinality K. We also define $\mathcal{N}(\mu; r) \equiv \{\mu' \in \mathcal{D} \,|\, \|U(\mu') - U(\mu)\| \le r\}$, which represents a "ball" near μ, and $\xi(\mu^*; r) = \frac{1}{\sqrt{m'}}\rho(\mu^*) + r$. We furthermore assume that Υ is fine enough to give good estimates for the sensitivity of U with respect to μ based on neighboring points in parameter space. We may then claim

Proposition 3.1 *With confidence level greater than γ, $\|\tilde{\beta}(\overline{\mu}) - \hat{\beta}(\mu^*)\|_S \le \xi(\mu^*; r)$ for any $\overline{\mu} \in \Upsilon$ such that $\mu^* \in \mathcal{N}(\overline{\mu}; r)$.*

We now sketch the proof. We first note from the triangle inequality that $\|\tilde{\beta}(\overline{\mu}) - \hat{\beta}(\mu^*)\|_S \le \|\tilde{\beta}(\overline{\mu}) - \tilde{\beta}(\mu^*)\|_S + \|\tilde{\beta}(\mu^*) - \hat{\beta}(\mu^*)\|_S$. We next note from the EIM system $B\tilde{\beta}(\mu) = U(\mu)$ that $\|\tilde{\beta}(\overline{\mu}) - \tilde{\beta}(\mu^*)\|_S = \|B(\tilde{\beta}(\overline{\mu}) - \tilde{\beta}(\mu^*))\| = \|U(\overline{\mu}) - U(\mu^*)\|$. The result then follows from the definition of $\mathcal{N}(\mu; r)$ and Proposition 2.1.

We emphasize that the accuracy of the EIM approximation does not affect the validity of our claim. We can also develop from Proposition 3.1 a test for the hypothesis that the experimental data is indeed obtained from the postulated manifold. Note that the restriction to the manifold effectively regularizes the inverse problem.

We briefly describe a paradigm associated with Proposition 3.1. In the Offline stage we compute for all $\overline{\mu}$ in Υ the EIM interpolant and associated $\tilde{\beta}(\overline{\mu})$: total storage Kn. We also choose r to be a maximum distance $\|U(\overline{\mu}_1) - U(\overline{\mu}_2)\|$ between neighboring points $\overline{\mu}_1, \overline{\mu}_2$ in Υ to ensure adequate coverage of $\mathcal{D}$. Then, in the Online stage [2], we find a consistent set Υ_{con} — a set of parameter values consistent with the experimental data — given by

$$\Upsilon_{\text{con}} = \{\overline{\mu} \in \Upsilon \;|\; \|\tilde{\beta}(\overline{\mu}) - \hat{\beta}(\mu^*)\|_S \le \xi(\mu^*; r)\}. \tag{6}$$

The construction of Υ_{con} requires Kn operations. From Proposition 3.1, in a fraction $\ge \gamma$ of all realizations, $\|U(\overline{\mu}) - U(\mu^*)\| > r$, $\forall \overline{\mu} \in \Upsilon \backslash \Upsilon_{\text{con}}$.

We now turn to numerical results for the problem and EIM approximation ($n = 33$) introduced in the previous section. We consider the case $\gamma = 0.95$, $m' = 16$, and $\sigma = 0.02$ for the particular candidate set $\Upsilon \equiv \Xi$ with $K = 40,000$. In the Offline stage we estimate $r = 0.0136$ based on simple nearest neighbor considerations in Υ. Then, in the Online stage, we identify Υ_{con} as shown in Figure 1.

In future work we will combine the results described here with model order reduction in order both to efficiently generate the EIM models (in the Offline stage) and also to efficiently assess the EIM contribution to the output error and parameter estimation (in the Online stage).

2. We presume that $\hat{\beta}(\mu^*)$ and $\rho(\mu^*)$ are already available as discussed in the context of Corollary 2.1; note to deduce these quantities we do not need to know μ^*.

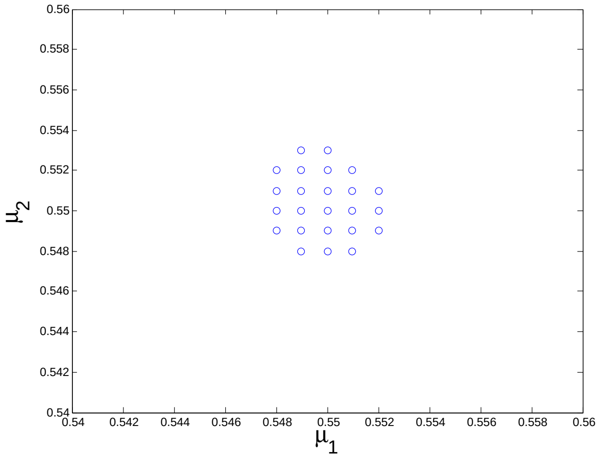

Figure 1. The plot depicts the consistent set $\Upsilon_{\rm con}$ in (6) for a particular realization. All the consistent parameter vectors are within the range $\mu^* \pm \Delta\mu^*$, with $\mu^* = (0.55, 0.55)$ and $\Delta\mu^* = (0.003, 0.003)$; note the restricted range of the axes relative to the full parameter domain $\mathcal{D} = [0.4, 0.6]^2$.

Acknowledgements

This work was funded by AFOSR/OSD Grant FA9550-09-1-0613, ONR Grant N00014-11-0713, and by the Norwegian University of Science and Technology while the second author was Visiting Professor at MIT during sabbatical leave. The authors thank Dr. Jens Lohne Eftang for EIM software and also James Penn for helpful comments.

References

[1] M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, C. R. Acad. Sci. Paris, Ser. I 339 (2004) 667-672.

[2] S. Chaturantabut, D.C. Sorensen, Discrete Empirical Interpolation for Nonlinear Model Reduction, TR09-05, CAAM, Rice University, 2009.

[3] N.C. Nguyen, A.T. Patera, J. Peraire, A 'best points' interpolation method for efficient approximation of parametrized functions, Int. J. Numer. Meth. in Engrg. 73 (2008) 521-543.

[4] N.C. Nguyen, J. Peraire, An interpolation method for the reconstruction and recognition of face images, in: A. Ranchordas, H. Araújo, J. Vitrià (Eds.), VISAPP 2007 Proceedings Second International Conference on Computer Vision Theory and Applications - Volume 2, March 8-11, 2007, Barcelona, Spain, pp. 91-96.

[5] N.R. Draper, H. Smith, Applied regression analysis, Wiley, 1998.

[6] H.T. Banks, K. Kunisch, Estimation techniques for distributed parameter systems, Birkhauser, 1989.