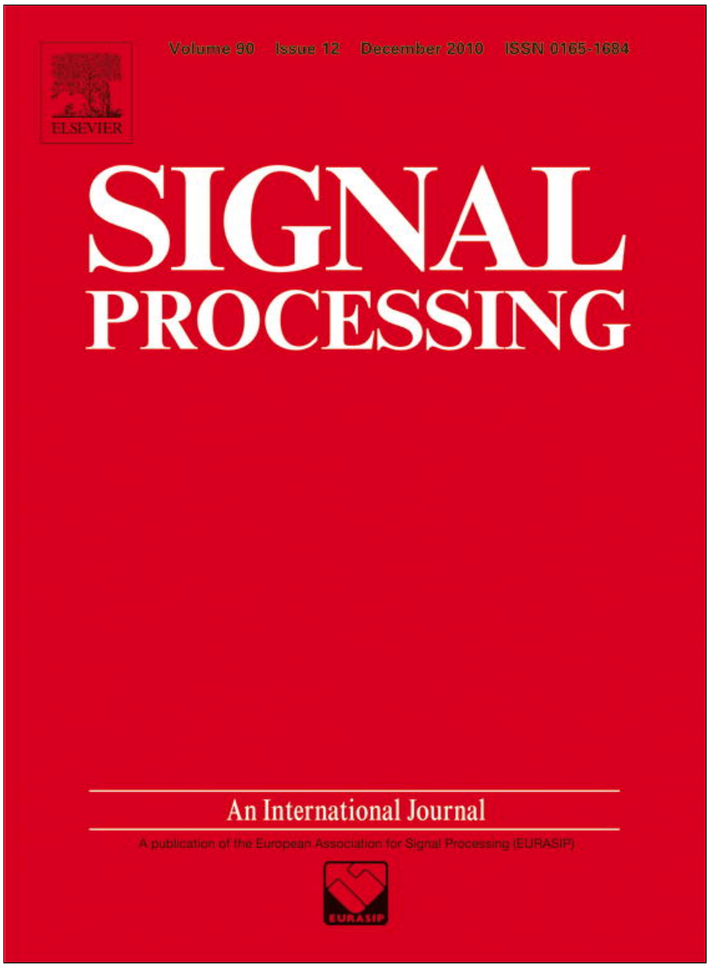
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Signal Processing 90 (2010) 3110–3121

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Robust deconvolution for ARMAX models with Gaussian uncertainties

R.H. Milocco [a,*], J.A. De Doná [b]

[a] Grupo Control Automático y Sistemas (GCAyS), Depto. Electrotecnia, Facultad de Ingeniería, Universidad Nacional del Comahue, Buenos Aires 1400, 8300 Neuquén, Argentina

[b] Centre for Complex Dynamic Systems and Control, School of Electrical Engineering and Computer Science, The University of Newcastle, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 22 October 2009
Received in revised form
29 March 2010
Accepted 12 May 2010
Available online 19 May 2010

Keywords:
Robust filtering
Robust deconvolution
Stochastic uncertainties
Quadratic programme
Truncated Gaussian distribution

ABSTRACT

In this paper we propose a robust deconvolution filter design that optimises a functional motivated by the *a posteriori* probability of the signals to be estimated. The problem is formulated in the framework of uncertain linear systems represented by discrete-time input–output ARMAX models, where the uncertainty is modelled as the realisation of a stochastic process with known statistics. The design is based on the use of a horizon of measurements in such a way that, for FIR systems, the functional to be optimised coincides with the one that maximises the *a posteriori* probability (MAP); and for ARMAX systems, the functional converges to the MAP functional as the length of the horizon is increased. The goal is to estimate signals with Gaussian or truncated Gaussian probability density functions based on measurements correlated with them. The robust design shows a very significant improvement, in a probabilistic sense for different systems, of the relative standard deviation of the estimation error when compared with the nominal model filter design.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The problem of robust filtering considered in this paper consists of estimating desired signals associated to an uncertain system, using measurements correlated with those signals. The measurements are the output of a discrete-time linear system corrupted by noise and the filtering problem consists in estimating the input signal to the system using a partial or imperfect model knowledge which constitute, in fact, a robust deconvolution problem. Robust filter designs can be based on a *deterministic approach*, where the uncertainty is modelled as an unknown element belonging to a family of admissible uncertainties, or they can be based on a *probabilistic approach*, where the uncertainty is assumed to be a stochastic process with known statistics. In the first approach, the objective consists in designing a stable and causal filter, which guarantees that the worst-case gain from the input signal to the filtering error remains bounded by a prescribed value for all admissible uncertainties. The filter can be obtained by minimising the H_∞ norm as in [1], and the references therein for linear systems represented in state space, or in [2] for input–output models. However, the H_∞ approach is known to be conservative. Mixed H_2/H_∞ designs allow a trade off between the best averaged performance of the minimum variance design and the best guaranteed worst case performance of the H_∞ estimator as in [3], and the references therein, where the design considers linear parameter varying systems in the state space model formulation.

The filter can also be obtained by minimising the error variance upper bound for all admissible noises and systems uncertainties, as in the so-called cost guaranteed filters. This design is achieved by minimising the H_2 norm

* Corresponding author. Tel./fax: +54 299 4488305.
E-mail addresses: milocco@uncoma.edu.ar, rmilocco@yahoo.com.ar (R.H. Milocco), Jose.Dedona@newcastle.edu.au (J.A. De Doná).

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.05.014

Author's personal copy

for the worst case as in [4,5], and the references therein for state-space models, or in [6–8], for input–output models.

A different approach, on the other hand, is the one based on a probabilistic description of uncertainties. In the context of state space approach, the finite-horizon filtering problem for uncertain discrete time-varying systems subject to both randomly varying measurements delays and norm bounded parameter uncertainties, using the Kalman filter, is given in [9]. In [10], a design is presented which minimises the expected value of the MSE cost with respect to the model uncertainties, when the latter are modelled as stochastic processes. In [11], a continuous-time solution is given to solve the H_∞ and the mixed H_2/H_∞ design for a stochastic description of the system uncertainties. Models with random uncertainties are common in the communication field where the problem of detecting finite alphabets transmitted through random channels is treated as in, e.g. [12–14].

In this paper, a robust deconvolution filter design is proposed which optimises a functional motivated by the *a posteriori* probability of the signals to be estimated. A horizon (or time window) of measurements is utilised, where the measurements are related to the signal to be estimated through an uncertain ARMAX model, including non-minimum phase systems. The uncertainty is modelled as the realisation of a stochastic process with known Gaussian distribution. For the particular cases of finite impulse response (FIR) models, the functional considered coincides, independently of the measurements horizon, with the one that maximises the *a posteriori* probability (MAP) of the estimations. For the more general case of ARMAX models, it is shown in the nominal case that, as the measurements horizon is increased, the estimates converge to the ones obtained with a maximum *a posteriori* optimisation criterion. In the case of the robust design, significant improvements are obtained in terms of reduction of the data horizons required for good performance, as shown in the simulation examples presented below. Thus, the robust design not only improves the quality of the estimates but significantly reduces the complexity of the on line estimations, extending the domains of application.

In addition to the scenario described above, a problem relevant to several engineering applications is the design of optimal filters subject to constraints. In particular, the problem of designing optimal filters for signals with *truncated Gaussian* distributions has attracted attention in recent times. In [15], this problem is treated in the context of state-space models, and its duality with a corresponding optimal control problem is established. Truncated Gaussian signals are also considered in applications of *data assimilation* for ocean models, see e.g. [16]. The design proposed in the current paper is optimal for signals with truncated Gaussian distributions. Thus, the work exposed in this paper covers a gap in the literature, where, to the best of the authors knowledge, very few or no treatments exist dealing with the problem of constrained robust filtering with *input–output* models. The improvement, in a probabilistic sense, of the standard deviation of the error that can be obtained, relative to the perfect model knowledge filter design, is illustrated through the average performance calculated for a family of stable linear systems.

In this paper, a common assumption about the variables that are stochastic in nature (both, external signals, noises, and uncertain system parameters) is that they have Gaussian distributions. The main reason for this assumption is the *mathematical tractability* of the resulting techniques. Of course, this represents an approximation to reality since, in any real life application, an exact characterisation of the probability distribution is hardly ever realistic. However, it is well known that many signals present in nature have distributions that are very close to Gaussian, a fact that is theoretically well-supported by the theorem of large numbers. In addition, as a slight generalisation to this assumption we also include *truncated* Gaussian distributions as part of the analysis. This generalisation allows to include constraints in the analysis while still keeping mathematical tractability.

The remainder of the paper is organised as follows: in Section 2 the problem is formally stated and the ARMAX model, using a finite horizon, is expressed in a linear matrix equation. In Section 3, the criterion to obtain the estimations using perfect model is derived for both, Gaussian and Gaussian truncated input signals. In Section 4 a strategy to improve the solution for the case of imperfect model is presented. In Section 5 simulation results of several examples are shown in order to illustrate the procedure and performance. Finally, in Section 6 we present the conclusions.

2. Problem formulation

Let us consider the following ARMAX model with scalar inputs and outputs:

$$y_k = a_1 y_{k-1} + \cdots + a_{na} y_{k-na} + b_0 u_k + \cdots + b_{nb} u_{k-nb} + e_k + c_1 e_{k-1} + \cdots + c_{nc} e_{k-nc} \quad (1)$$

where the input u_k, at each time instant k, is to be estimated using a horizon of present and past measurements y_k. The input e_k represents the measurement noise and the parameters a_i, b_i, and c_i are scalar constants. The problem can be easily extended to the multivariate case if the coefficients are matrices and the variables are vectors of appropriate dimensions. For simplicity, and without loss of generality, we will consider the scalar case. The input signals u_k and e_k are assumed to be i.i.d. stochastic processes with normal distributions $u_k \backsim \mathcal{N}(0,\sigma_u^2(0))$ and $e_k \backsim \mathcal{N}(0,\sigma_e^2(0))$, and independent of each other.

The ARMAX model (1) can be written in a compact form as

$$y_k = \mathbf{a}^T \mathbf{y}_{k-1} + \mathbf{b}^T \mathbf{u}_k + \mathbf{c}^T \mathbf{e}_k \quad (2)$$

where the different vectors are given by $\mathbf{a} = [a_{na}, \ldots, a_1]^T$, $\mathbf{b} = [b_{nb}, \ldots, b_0]^T$, and $\mathbf{c} = [c_{nc}, \ldots, 1]^T$, and

$$\mathbf{y}_{k-1} = \begin{pmatrix} y_{k-na} \\ \vdots \\ y_{k-1} \end{pmatrix}, \quad \mathbf{u}_k = \begin{pmatrix} u_{k-nb} \\ \vdots \\ u_k \end{pmatrix}, \quad \mathbf{e}_k = \begin{pmatrix} e_{k-nc} \\ \vdots \\ e_k \end{pmatrix} \quad (3)$$

then, from (2) we have that the following equation is

Author's personal copy

satisfied:

$$\mathbf{y}_k = A\mathbf{y}_{k-1} + B\mathbf{u}_k + C\mathbf{e}_k \tag{4}$$

where matrices A, B, and C of dimensions $na \times na$, $na \times (nb+1)$, and $na \times (nc+1)$, respectively, are given by

$$A = \begin{pmatrix} \mathbf{0} & | & I \\ - & - & - \\ & \mathbf{a}^T & \end{pmatrix}, \quad B = \begin{pmatrix} O \\ -- \\ \mathbf{b}^T \end{pmatrix}, \quad C = \begin{pmatrix} O \\ -- \\ \mathbf{c}^T \end{pmatrix} \tag{5}$$

where I is the identity, O is a matrix of zeros, and $\mathbf{0}$ is a vector of zeros. The measured value at time instant $k+N$, y_{k+N}, for $N \geq 0$, can be expressed as a function of the initial conditions $\mathbf{y}_{k-1}$ by using (2) and (4), as follows:

$$y_{k+N} = \mathbf{a}^T\mathbf{y}_{k+N-1} + \mathbf{b}^T\mathbf{u}_{k+N} + \mathbf{c}^T\mathbf{e}_{k+N}$$

$$y_{k+N} = \mathbf{a}^T[A\mathbf{y}_{k+N-2} + B\mathbf{u}_{k+N-1} + C\mathbf{e}_{k+N-1}] + \mathbf{b}^T\mathbf{u}_{k+N} + \mathbf{c}^T\mathbf{e}_{k+N}$$

$$\vdots = \vdots$$

$$y_{k+N} = \mathbf{a}^T A^N \mathbf{y}_{k-1} + \mathbf{a}^T A^{N-1}[B\mathbf{u}_k + C\mathbf{e}_k] + \cdots + \mathbf{a}^T[B\mathbf{u}_{k+N-1} + C\mathbf{e}_{k+N-1}] + [\mathbf{b}^T\mathbf{u}_{k+N} + \mathbf{c}^T\mathbf{e}_{k+N}] \tag{6}$$

Using the previous definitions, it is possible to write the sequence y_k in the interval $[-L_1, L_2]$ (where L_1 and L_2 are arbitrary integer numbers that satisfy $L_1 \geq 0$, $L_2 \geq 1$, defining a backward and a forward horizon, respectively, with respect to the time instant k as simple matrix operations of the following form:

$$\tilde{\mathbf{y}} = \mathbb{A}\mathbf{y}_0 + \mathbb{B}\mathbf{u} + \mathbb{C}\tilde{\mathbf{e}} \tag{7}$$

where

$$\tilde{\mathbf{y}} = \begin{pmatrix} y_{k-L_1} \\ \vdots \\ y_{k+L_2-1} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_{k-L_1-nb} \\ \vdots \\ u_{k+L_2-1} \end{pmatrix}, \quad \tilde{\mathbf{e}} = \begin{pmatrix} e_{k-L_1-nc} \\ \vdots \\ e_{k+L_2-1} \end{pmatrix} \tag{8}$$

$\mathbf{y}_0$ is a short notation for $\mathbf{y}_{k-L_1-1}$, and

$$\mathbb{A} = \begin{pmatrix} \mathbf{a}^T \\ \mathbf{a}^T A \\ \vdots \\ \mathbf{a}^T A^{L_1+L_2-1} \end{pmatrix}, \quad \mathbb{B} = \begin{pmatrix} \mathbf{b}^T & | & O \\ \mathbf{a}^T B & | & \vdots \\ \vdots & | & \vdots \\ \mathbf{a}^T A^{L_1+L_2-2} B & | & O \end{pmatrix} + \begin{pmatrix} 0 & 0 & | & O \\ \vdots & \mathbf{b}^T & | & \vdots \\ \vdots & \vdots & | & \vdots \\ 0 & \mathbf{a}^T A^{L_1+L_2-3} B & | & O \end{pmatrix} + \cdots + \begin{pmatrix} O & | & 0 \\ \vdots & | & \vdots \\ \vdots & | & 0 \\ O & | & \mathbf{b}^T \end{pmatrix} \tag{9}$$

and $\mathbb{C}$ defined similarly to $\mathbb{B}$ by exchanging B by C and $\mathbf{b}$ by $\mathbf{c}$. Notice that the vector $\mathbf{y}_0$ contains the minimum possible number of variables that, jointly with the vector of inputs $\mathbf{u}$ and $\tilde{\mathbf{e}}$, determine y_j for all $j \geq k-L_1$. In the case of ARMAX systems the vector $\mathbf{y}_0$ represents the *initial conditions*. Notice that, in the case of FIR systems, this vector is not required since $\mathbf{a}=0$. Finally, taking into account, from (4), that $\mathbf{y}_0=\mathbf{w}+\mathbf{v}_0$ where $\mathbf{w} = A\mathbf{y}_{k-L_1-2} + B\mathbf{u}_{k-L_1-1}$, and $\mathbf{v}_0=C\mathbf{e}_0$ with $\mathbf{e}_0$ a short notation for $\mathbf{e}_{k-L_1-1}$, the relationship between the input and output can be expressed in a simple matrix form by

$$\mathbf{y} = H\mathbf{x} + \mathbf{v} \tag{10}$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_0 \\ \tilde{\mathbf{y}} \end{pmatrix}, \quad H = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbb{A} & \mathbb{B} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} \tag{11}$$

$$\mathbf{v} = \begin{pmatrix} C & \mathbf{0} \\ \mathbb{A}C & \mathbb{C} \end{pmatrix} \begin{pmatrix} \mathbf{e}_0 \\ \tilde{\mathbf{e}} \end{pmatrix} \tag{12}$$

with covariance matrices

$$R_{\mathbf{u}} = \sigma_u^2 I, \quad R_{\mathbf{v}} = \sigma_e^2 \begin{pmatrix} CC^T & CC^T\mathbb{A}^T \\ \mathbb{A}CC^T & \mathbb{A}CC^T\mathbb{A}^T + \mathbb{C}\mathbb{C}^T \end{pmatrix} \tag{13}$$

for vectors $\mathbf{u}$ and $\mathbf{v}$. The goal is now to obtain an estimate $\hat{\mathbf{x}}^{opt}$ of the vector $\mathbf{x}$ from the knowledge of its statistics and the vector $\mathbf{y}$ given by the measurements $\mathbf{y}^d$. The processing to obtain such estimates will be done in 'blocks of data'.

3. Filter design based on perfect model

To motivate the estimation criterion proposed in this work, we will start by analysing the estimation criterion that maximises the *a posteriori* probability (MAP), given by

$$\hat{\mathbf{x}}^{MAP} = \underset{\hat{\mathbf{x}}}{\operatorname{argmax}}\{p_{\mathbf{x}|\mathbf{y}}(\hat{\mathbf{x}}|\mathbf{y}^d)\} \tag{14}$$

Using Bayes' rule we have

$$p_{\mathbf{x}|\mathbf{y}}(\hat{\mathbf{x}}|\mathbf{y}^d) = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^d|\hat{\mathbf{x}})p_{\mathbf{x}}(\hat{\mathbf{x}})}{p_{\mathbf{y}}(\mathbf{y}^d)} \tag{15}$$

Taking into account the following transformation:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{v} \end{pmatrix} \tag{16}$$

and the fact that $\mathbf{x}$ and $\mathbf{v}$ are independent of each other, we have that $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^d|\hat{\mathbf{x}}) = p_{\mathbf{v}}(\mathbf{y}^d - H\hat{\mathbf{x}})$.

In addition, $p_{\mathbf{x}}(\hat{\mathbf{x}})$ in (15) can be written as

$$p_{\mathbf{x}}(\hat{\mathbf{x}}) = p_{\mathbf{w},\mathbf{u}}(\hat{\mathbf{w}},\hat{\mathbf{u}}) = p_{\mathbf{w}|\mathbf{u}}(\hat{\mathbf{w}}|\hat{\mathbf{u}})p_{\mathbf{u}}(\hat{\mathbf{u}}) \tag{17}$$

We assume that the pdf of $\mathbf{u}$ is known and that it belongs to a Gaussian, or truncated Gaussian, distribution. In the case of a truncated Gaussian distribution, we assume that the support (that is, the domain where the pdf is different from zero) is given by linear inequalities. In the case of FIR systems, since $\mathbf{w}$ should not be taken into account, we have $p_{\mathbf{x}}(\hat{\mathbf{x}}) = p_{\mathbf{u}}(\hat{\mathbf{u}})$ instead of (17) which gives in (15) the well known MAP estimation for FIR models. In the case of ARMAX systems, since we assume the probability distribution of $\mathbf{u}$ to be known, it is in principle possible to know the conditional probability $p_{\mathbf{w}|\mathbf{u}}(\hat{\mathbf{w}}|\hat{\mathbf{u}})$ in (17) by using the knowledge of H. However, if the input vector $\mathbf{u}$ has a truncated Gaussian pdf, the probability $p_{\mathbf{w}|\mathbf{u}}(\hat{\mathbf{w}}|\hat{\mathbf{u}})$ does not result, in general, in a Gaussian—nor truncated Gaussian—distribution, and its inclusion in the MAP functional (14) complicates significantly the computation of the estimations. Moreover, considering the case where H may have uncertainties as discussed in the next section, the calculation of conditional probability computed by

Author's personal copy

using the uncertain model can result in unacceptable inaccuracies. In its place, we propose to replace this *prior* knowledge of the statistics of $\mathbf{w}$ by its likelihood function, given by $p_{\mathbf{y}_0|\mathbf{w}}(\mathbf{y}_0^d|\hat{\mathbf{w}})$, which gives us an estimation of the statistics of $\mathbf{w}$ supposing that we have only access to the measurement $\mathbf{y}_0$. Taking into account that $p_{\mathbf{y}_0|\mathbf{w}}(\mathbf{y}_0^d|\hat{\mathbf{w}}) = p_{\mathbf{v}_0}(\mathbf{y}_0^d - \hat{\mathbf{w}})$, the functional is no more the *a posteriori* probability of the signal to be estimated, but, on the other hand, it allows us to work with Gaussian and truncated Gaussian density functions, and it enormously simplifies the solution of the estimation problem. In this way, the design also becomes independent of possible uncertainties of H. We will see later that, for stable systems, as the horizon L_1 is increased, the maximum of the proposed functional converges to the MAP functional. Thus, the probability $p_{\mathbf{x}}(\hat{\mathbf{x}})$ in (15) is substituted by $p_{\mathbf{v}_0}(\mathbf{y}_0^d - \hat{\mathbf{w}})\, p_{\mathbf{u}}(\hat{\mathbf{u}}) = p_{\mathbf{v}_0,\mathbf{u}}(\mathbf{y}_0^d - \hat{\mathbf{w}}, \hat{\mathbf{u}})$.

Finally, considering that $p_{\mathbf{y}}(\mathbf{y}^d)$ does not depend on the estimated values, we have that the estimation criterion can be posed as the following maximisation problem:

$$\hat{\mathbf{x}}^{opt} = \arg\max_{\hat{\mathbf{x}}}\{p_{\mathbf{v}}(\mathbf{y}^d - H\hat{\mathbf{x}})p_{\mathbf{v}_0,\mathbf{u}}(\mathbf{y}_0^d - \hat{\mathbf{w}}, \hat{\mathbf{u}})\} \tag{18}$$

In the case of strictly Gaussian distributions, we can pose the problem in *compact* form as

$$\hat{\mathbf{x}}^{opt} = \arg\min_{\hat{\mathbf{x}}}\{\|\mathbf{y}^d - H\hat{\mathbf{x}}\|^2_{R_{\mathbf{v}}^{-1}} + \|E\hat{\mathbf{x}} + F\mathbf{y}^d\|^2_{P^{-1}}\} \tag{19}$$

where the norm of a vector $\mathbf{z}$ weighted by a matrix R is $\|\mathbf{z}\|^2_R = \mathbf{z}^T R \mathbf{z}$,

$$P = \begin{bmatrix} R_{\mathbf{v}_0} & 0 \\ 0 & R_{\mathbf{u}} \end{bmatrix}, \quad E = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}, \quad F = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

$R_{\mathbf{v}_0} = \sigma_e^2 CC^T$ is the covariance matrix of the vector $\mathbf{v}_0$. Problem (19) can be formulated as a minimisation problem with quadratic cost (*quadratic programme*, or qp) without constraints, of the following form:

$$\hat{\mathbf{x}}^{opt} = \arg\min_{\hat{\mathbf{x}}}\left\{\frac{1}{2}\hat{\mathbf{x}}^T Q\hat{\mathbf{x}} - f^T\hat{\mathbf{x}}\right\} \tag{20}$$

The minimiser is $\hat{\mathbf{x}} = Q^{-1}f$, where

$$Q = H^T R_{\mathbf{v}}^{-1} H + P^{-1} \tag{21}$$

$$f = (-P^{-1}EF + H^T R_{\mathbf{v}}^{-1})\mathbf{y}^d \tag{22}$$

and where we have used the fact that $E^T P^{-1} E = P^{-1}$ and $E^T P^{-1} F = P^{-1} EF$.

In the case of truncated Gaussian density functions for the variable $\mathbf{u}$, the problem consists of minimising the same objective function, but now subject to linear constraints on the vector $\hat{\mathbf{u}} = [0 \;\; I]\hat{\mathbf{x}}$, as follows:

$$\hat{\mathbf{x}}^{opt} = \arg\min_{\hat{\mathbf{x}}}\left\{\frac{1}{2}\hat{\mathbf{x}}^T Q\hat{\mathbf{x}} - f^T\hat{\mathbf{x}}\right\}$$

$$\text{s.t.} \quad [0 \;\; S]\hat{\mathbf{x}} \le s \tag{23}$$

where the matrix S and vector s characterise[1] the constraint set.

[1] Inequalities involving vectors in this paper are interpreted elementwise.

By considering a horizon $L_2=1$, we have a filtering problem, whereas a horizon $L_2 > 1$ makes it a smoothing problem. The estimation is by blocks of data of length L_1+L_2+nb for $\hat{\mathbf{u}}$ and of length na for $\hat{\mathbf{w}}$. The processing strategy can be carried out by blocks of L_1+L_2+na samples updating the vector $\mathbf{y}$ with each new output measurement and keeping the previous ones. For each block estimation we keep the estimation at $\hat{\mathbf{u}}_{k+L_2-1}$. It is important to stress that in the case of FIR systems, the length of the horizon is not critical, since the estimates coincide with the MAP estimates for all horizons. In the case of ARMAX systems it is convenient to use horizons sufficiently large so that the estimates converge to the MAP estimates as it is proved in the following theorem.

Theorem 1. *Given the block structure of the ARMAX model* (10) *with inputs* $\mathbf{u}$ *Gaussian or truncated Gaussian and* $\mathbf{v}$ *strictly Gaussian, the proposed estimator converges asymptotically to the optimal MAP estimator when the horizon* L_1 *increases.*

For the proof see Appendix A.

In the next section the problem of robust filter design for uncertain models will be addressed.

4. Robust filter

The design of optimal filters requires the perfect knowledge of the system. This, however, is not generally possible since the system is identified experimentally using a known input. Even when using a persistent excitation, the best that can be obtained is a partial knowledge of the system. The use of parameters that only approximate the real ones from the system can considerably deteriorate the estimation of a given signal.

The system, in these cases, can be modelled as the sum of two components. One is a known component, given by a matrix of gains $\overline{H}$, sometimes called the nominal model of the system and that is the one obtained experimentally, and the other is a matrix Δ, which is unknown but assumed to belong to a family with known statistical properties. This kind of probabilistic models can be found in several robust approaches, see [10] for matrix polynomial representation of system dynamics, [9] in sensing, [14,12] in communication. The general formulation is as follows:

$$H = \overline{H} + \Delta, \quad \Delta_{i,j} \sim \mathcal{N}(0, \sigma^2_{\Delta ij}), \;\; \forall i,j \tag{24}$$

where $\Delta_{i,j}$ are the components of Δ. In our case, $\overline{H}$ is a matrix of known constants which is in fact the mean of the multivariable Gaussian distributed variable H. In this section, we develop a design for the estimator that is robust to uncertainties based on the knowledge of the system characterised in the form (24).

We assume that the covariances between the column vectors d_i of Δ are known in such a way that calling $L=L_1+L_2+na+nb$ the length of the vector $\mathbf{x}$, we have

$$\Delta = [d_1 \;\; \cdots \;\; d_L], \quad \mathcal{E}[d_i d_j^T] = R_{\Delta ij} \tag{25}$$

It is important to notice that uncertainties in the knowledge of the covariance matrices $R_{\mathbf{v}}$ and $R_{\mathbf{u}}$ in (19)

Author's personal copy

affect the cost in a much smaller extent than uncertainties in the system H. In Appendix B, it is shown that the *nominal* design minimises the variation of the cost with respect to uncertainties in the covariance matrix. On the other hand, with the nominal design there is no control over the variations of the cost with respect to uncertainties in the system. For this reason, in this work we deal with the problem of robust design against variations in the system, and we assume that the covariance matrices $R_{\mathbf{u}}$ and $R_{\mathbf{v}}$ are known, since any discrepancy in their values is effectively counteracted by the nominal design.

Using a system description comprising an uncertain model, Eq. (10) can be rewritten as

$$\mathbf{y}_k = (\overline{H}+\Delta)\mathbf{x}_k + \mathbf{v}_k \tag{26}$$

considering the following transformation of variables:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} I & 0 \\ \overline{H} & I \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \Delta\mathbf{x}+\mathbf{v} \end{pmatrix} \tag{27}$$

we have that $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^d|\hat{\mathbf{x}})$ in (15) is given by

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^d|\hat{\mathbf{x}}) = \frac{p_{\mathbf{y},\mathbf{x}}(\mathbf{y}^d,\hat{\mathbf{x}})}{p_{\mathbf{x}}(\hat{\mathbf{x}})} \tag{28}$$

$$= \frac{p_{\Delta\mathbf{x}+\mathbf{v},\mathbf{x}}(\mathbf{y}^d-\overline{H}\hat{\mathbf{x}},\hat{\mathbf{x}})}{p_{\mathbf{x}}(\hat{\mathbf{x}})} \tag{29}$$

$$= p_{\Delta\mathbf{x}+\mathbf{v}|\mathbf{x}}(\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}|\hat{\mathbf{x}}) \tag{30}$$

$$= p_{\Delta\hat{\mathbf{x}}+\mathbf{v}}(\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}) \tag{31}$$

then, in this context, the criterion (18) can be rewritten as

$$\hat{\mathbf{x}}^{opt} = \arg\max_{\hat{\mathbf{x}}}\{p_{\Delta\hat{\mathbf{x}}+\mathbf{v}}(\mathbf{y}^d-\overline{H}\hat{\mathbf{x}})p_{\mathbf{v}_0,\mathbf{u}}(\mathbf{y}_0^d-\hat{\mathbf{w}},\hat{\mathbf{u}})\} \tag{32}$$

Given the fact that Δ has a Gaussian distribution, as well as $\mathbf{v}$, the probability density function $p_{\Delta\hat{\mathbf{x}}+\mathbf{v}}$ of (31), for any $\hat{\mathbf{x}}$, is also a Gaussian function of the form:

$$p_{\Delta\hat{\mathbf{x}}+\mathbf{v}}(\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}) = const. \times \frac{e^{-(1/2)\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}\|^2_{R^{-1}_{\mathbf{v}\Delta}}}}{|R_{\mathbf{v}\Delta}|^{1/2}} \tag{33}$$

where, taking into account that Δ is independent of $\mathbf{v}$, $R_{\mathbf{v}\Delta} = \mathcal{E}[(\Delta\hat{\mathbf{x}}+\mathbf{v})(\Delta\hat{\mathbf{x}}+\mathbf{v})^T]$ has the following form:

$$R_{\mathbf{v}\Delta} = \mathcal{E}(\Delta\hat{\mathbf{x}}\hat{\mathbf{x}}^T\Delta^T)+R_{\mathbf{v}} \tag{34}$$

$$= \sum_{i=1}^{L}\sum_{j=1}^{L}\hat{x}_i\hat{x}_j R_{\Delta ij}+R_{\mathbf{v}} \tag{35}$$

Then, the criterion in (32) is equivalent to the following minimisation problem:

$$\hat{\mathbf{x}}^{opt} = \arg\min_{\hat{\mathbf{x}}}\{\log(|R_{\mathbf{v}\Delta}|)+\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}\|^2_{R^{-1}_{\mathbf{v}\Delta}}+\|E\hat{\mathbf{x}}+F\mathbf{y}^d\|^2_{P^{-1}}\} \tag{36}$$

Notice that the covariance matrix $R_{\mathbf{v}\Delta}$ depends on the estimates $\hat{\mathbf{x}}$, thus problem (36) is not a quadratic programme as it was in the nominal case. However, it is possible to derive an algorithm to improve the nominal solution obtained as a minimiser of cost (19). To this end, we state the following problem: given the covariance matrix R_i, computed according to (35) with any initial estimation $\hat{\mathbf{x}}_i^{opt}$, we are interested in obtaining a new estimation $\hat{\mathbf{x}}_{i+1}^{opt}$, and their corresponding covariance matrix R_{i+1} computed with (35), which fulfils $J_i \geq J_{i+1}$ where

$$J_i = \log(|R_i|)+\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_i^{opt}\|^2_{R_i^{-1}}+\|E\hat{\mathbf{x}}_i^{opt}+F\mathbf{y}^d\|^2_{P^{-1}} \tag{37}$$

is the type of robust cost given in (36). Formally, we are interested in a new estimated vector $\hat{\mathbf{x}}_{i+1}$ such that the following holds:

$$\log(|R_i|)+\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_i^{opt}\|^2_{R_i^{-1}}+\|E\hat{\mathbf{x}}_i^{opt}+F\mathbf{y}^d\|^2_{P^{-1}}$$
$$\geq \log(|R_{i+1}|)+\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_{i+1}\|^2_{R^{-1}_{i+1}}+\|E\hat{\mathbf{x}}_{i+1}+F\mathbf{y}^d\|^2_{P^{-1}} \tag{38}$$

By subtracting $\|\mathbf{y}^d-H\hat{\mathbf{x}}_{i+1}\|^2_{R_i^{-1}}$ from both sides of (38) and reordering, the inequality can be written as

$$(\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_i^{opt}\|^2_{R_i^{-1}}+\|E\hat{\mathbf{x}}_i^{opt}+F\mathbf{y}^d\|^2_{P^{-1}})-(\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_{i+1}\|^2_{R_i^{-1}}+\|E\hat{\mathbf{x}}_{i+1}$$
$$+F\mathbf{y}^d\|^2_{P^{-1}}) \geq (\log(|R_{i+1}R_i^{-1}|)+\|\mathbf{y}^d-\overline{H}\hat{\mathbf{x}}_{i+1}\|^2_{R^{-1}_{i+1}-R^{-1}_i}) \tag{39}$$

In order to force the above inequality as much as possible, an optimal $\hat{\mathbf{x}}_{i+1}^{opt}$ can be obtained that minimises the second term on the left-hand side of (39). Note that, as a result, the cost to be minimised is similar to that of the nominal case and it is given by

$$\hat{\mathbf{x}}_{i+1}^{opt} = \arg\min_{\hat{\mathbf{x}}_{i+1}}\{\|\mathbf{y}^d-H\hat{\mathbf{x}}_{i+1}\|^2_{R_i^{-1}}+\|E\hat{\mathbf{x}}_{i+1}+F\mathbf{y}^d\|^2_{P^{-1}}\} \tag{40}$$

Thus, it is possible to derive a strategy to improve the cost obtained with the nominal solution. It consists in computing in a first iteration, $i=0$, the nominal solution, $\hat{\mathbf{x}}_1^{opt}$, with $R_0=R_{\mathbf{v}}$. The covariance matrix R_1 is obtained using (35) and $\hat{\mathbf{x}}_1^{opt}$. In the second iteration, $i=1$, we use R_1 to obtain $\hat{\mathbf{x}}_2^{opt}$ and the corresponding covariance R_2 using (35) and $\hat{\mathbf{x}}_2^{opt}$, also the cost J_2, using (37). Comparing both costs we decide if the new estimation is accepted $J_2<J_1$ or rejected $J_2 \geq J_1$. In the case that the cost is improved it is possible to continue with a third iteration, $i=2$, and so on until the cost ceases to improve. Otherwise, we keep the estimation of the previous iteration as final result. Extensive simulation results confirm that, with this algorithm, the robust design improves considerably the nominal estimation in the first iteration (see the simulation examples in Section 5).

Finally, we recall that the minimum of the cost (40) can be obtained in a similar way than in the nominal case as the minimiser of a quadratic programme of the following form:

$$\hat{\mathbf{x}}_{i+1}^{opt} = \arg\min_{\hat{\mathbf{x}}_{i+1}}\left\{\frac{1}{2}\hat{\mathbf{x}}_{i+1}^T Q\hat{\mathbf{x}}_{i+1}-f^T\hat{\mathbf{x}}_{i+1}\right\} \tag{41}$$

with minimiser $\hat{\mathbf{x}}_{i+1}^{opt} = Q^{-1}f$, and

$$Q = \overline{H}^T R_i^{-1}\overline{H}+P^{-1} \tag{42}$$

$$f = (-P^{-1}EF+\overline{H}^T R_i^{-1})\mathbf{y}^d \tag{43}$$

In the case in which $\mathbf{u}$ has truncated Gaussian distribution, the minimum is found with a similar quadratic programme but now subject to constraints on

Author's personal copy

the vector $\hat{\mathbf{u}}$ as follows:

$$\hat{\mathbf{x}}_{i+1}^{opt} = \underset{\hat{\mathbf{x}}_{i+1}}{\arg\min}\left\{\frac{1}{2}\hat{\mathbf{x}}_{i+1}^{T}Q\hat{\mathbf{x}}_{i+1} - f^{T}\hat{\mathbf{x}}_{i+1}\right\} \quad \text{s.t. } [0\ \ S]\hat{\mathbf{x}}_{i+1} \le s \tag{44}$$

where S and s are, respectively, a matrix and a vector of constants.

In the next section we present simulation results in order to illustrate the procedure and the performance of the estimation scheme.

5. Simulation results

We will see first in a simple example how the cost function is improved with the proposed strategy.

Example 1. We consider a first order FIR model given by $y_k = hu_k + v_k$ where u_k and v_k are independent white noise sequences with variance $\sigma_u^2 = 1$ and σ_v^2, respectively. The horizon is $L_2 = 0$ and $L_1 = 1$. The gain h is a Gaussian process $\backsim \mathcal{N}(\overline{h}, \sigma_d^2)$ with $\overline{h} = 1$. Thus, the problem can be accommodated to our notation by defining

$$\overline{H} = \begin{pmatrix} 1 & 0 \\ 0 & \overline{h} \end{pmatrix}, \quad P = \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}, \quad R_i = \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \hat{u}_i^2\sigma_d^2 + \sigma_v^2 \end{pmatrix} \tag{45}$$

$$\mathbf{y}^d = \begin{pmatrix} 0 \\ y^d \end{pmatrix}, \quad \hat{\mathbf{x}}_i = \begin{pmatrix} 0 \\ \hat{u}_i \end{pmatrix} \tag{46}$$

where i is the ith iteration. Replacing in (42) and (43) we obtain the optimal estimator for ith iteration which is given by

$$\hat{u}_{i+1} = \frac{\sigma_u^2\overline{h}}{\sigma_u^2\overline{h}^2 + \hat{u}_i^2\sigma_d^2 + \sigma_v^2} y_k^d \tag{47}$$

Note that for $i = 0$, in the first iteration, the solution for the nominal system coincides with the Wiener filter design.

In order to show that the nominal design is robust to uncertainties in the knowledge of noise covariance, contrary to uncertainties in the knowledge of the system coefficients, we will calculate the following functionals:

$$f_Q = \frac{J(Q,H,\mathbf{x}_0) - J(\tilde{Q},H,\tilde{\mathbf{x}}_Q)}{J(Q,H,\mathbf{x}_0)}, \quad f_H = \frac{J(Q,H,\mathbf{x}_0) - J(Q,\tilde{H},\tilde{\mathbf{x}}_H)}{J(Q,H,\mathbf{x}_0)} \tag{48}$$

where $J(Q,H,\mathbf{x})$ is a short notation of the cost of Eq. (19) where $\mathbf{x}_0$ is its minimiser, $\tilde{\mathbf{x}}_Q$ is the minimiser of $J(\tilde{Q},H,\mathbf{x})$, and $\tilde{\mathbf{x}}_H$ is the minimiser of $J(Q,\tilde{H},\mathbf{x})$. In Fig. 1 both functionals are shown versus the signal to noise ratio when a change of 10% in both σ_v^2, which produces a variation from Q to $\tilde{Q}$, and h, which produces a variation from H to $\tilde{H}$. Note that both functional are independent of y^d, due to cancelation in the numerator and denominator. It can be seen from the figure that the cost variation is much smaller in the case of differences in the knowledge of the noise variance than in the knowledge of the system gain. The differences increase as the signal-to-noise ratio increases.

In Fig. 2 the cost for three iterations, starting with the nominal solution, is shown for the case where $\sigma_d^2 = 0.4$ and $\sigma_v^2 = 10^{-2}$. Note that the improvement obtained in the first iteration is significant. In Fig. 3 the estimated values as a function of the measured output y are depicted. In Figs. 4 and 5 the same graphics are depicted but using $\sigma_d^2 = 0.1$ and $\sigma_v^2 = 10^{-2}$ with the same observations as in the previous case.

The same procedure can be used for higher order FIR filters. We recall that in the case of FIR systems, the length of the horizon is not critical, since the estimates coincide with the MAP estimates for all horizons.

Example 2. Systems with poles near the origin can be very well approximated by FIR filters. In contrast, systems

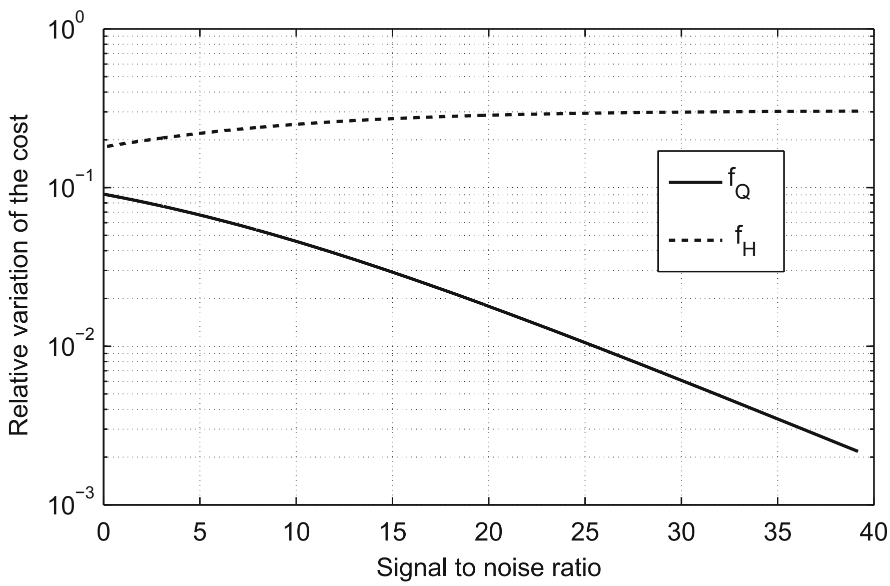


Fig. 1. Cost variation for uncertainties of 10% in σ_v^2 and h.

Author's personal copy

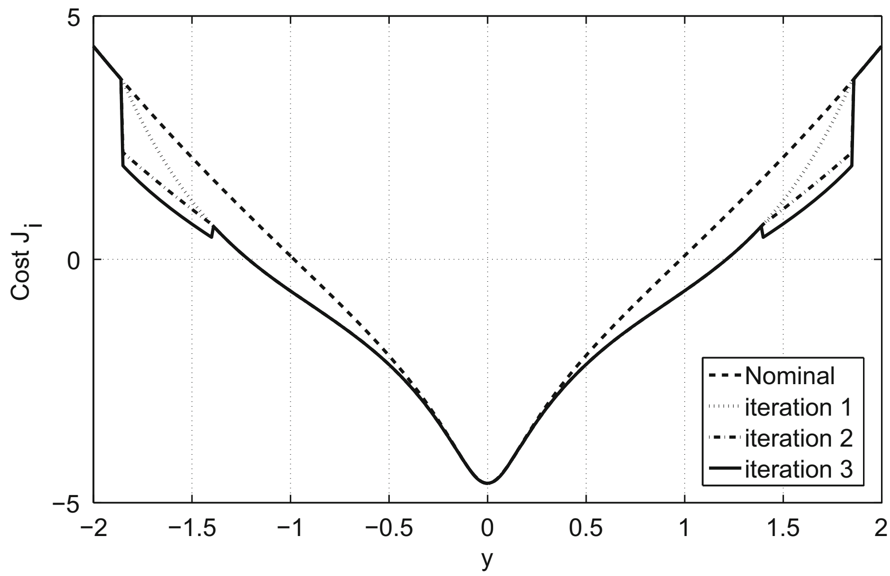


Fig. 2. Cost improvements for FIR y_k=u_k+v_k with $\sigma_d^2 = 0.4$, $\sigma_v^2 = 10^{-2}$.

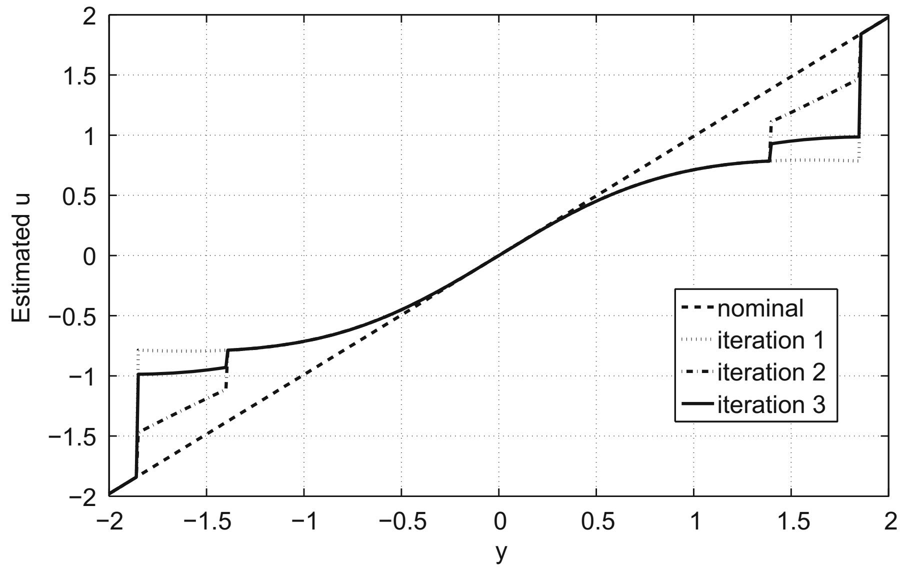


Fig. 3. Estimated values of u_k for FIR y_k=u_k+v_k with $\sigma_d^2 = 0.4$, $\sigma_v^2 = 10^{-2}$.

with poles near the unit circle can only be represented by IIR filters. We will see in this example the performance of the proposed heuristic design when it is used to estimate the input of a second order IIR system with poles and zeros near the unit circle given at $z_1 = -0.8 \pm j\ 0.51$ and poles at $\pm j\ 0.895$. First we show, for the case of perfect knowledge of the system, the convergence of the heuristic design versus the horizon length compared with the optimal Wiener filter design. The performance is depicted in Fig. 6 where it can be seen that, increasing the horizon, the heuristic design cost coincides with the one of the optimal Wiener filter.

We also are interested in knowing the performance of the nominal filter using Eqs. (20) and (21) in the case where the input is truncated Gaussian distributed versus the constrained design given by the quadratic programme solution in (23). For such purpose, a comparison between the relative standard deviation estimation error of both designs was performed by simulating a truncated Gaussian input signal distribution. The truncated Gaussian distribution of the input is obtained by taking samples from a Gaussian distribution $\backsim \mathcal{N}(0,1)$ and discarding the samples that fall outside of a window centred at zero and of width 2α. In the case of the nominal solution a decision element is incorporated such that $\hat{u} > \alpha \Rightarrow \hat{u} = \alpha$ and $\hat{u} < -\alpha \Rightarrow \hat{u} = -\alpha$. The result is shown in Fig. 7 for different noise amplitudes. It can be seen that, with qp (23), an improvement between 10% and 20% is obtained depending on the signal-to-noise ratio.

In the case of imperfect knowledge of the system coefficient, the performance of the robust design at the

Author's personal copy

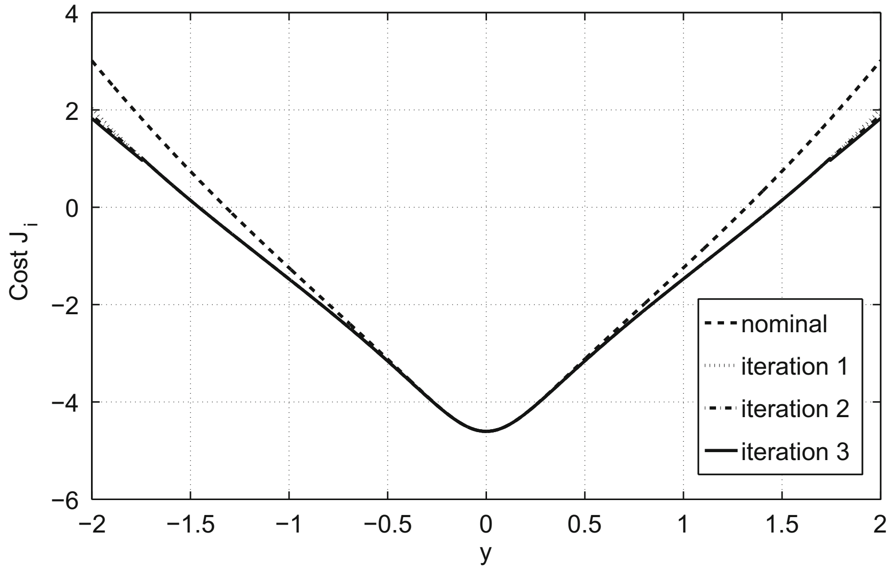


Fig. 4. Cost improvements for FIR y_k=u_k+v_k with $\sigma_d^2=0.1$, $\sigma_v^2=10^{-2}$.

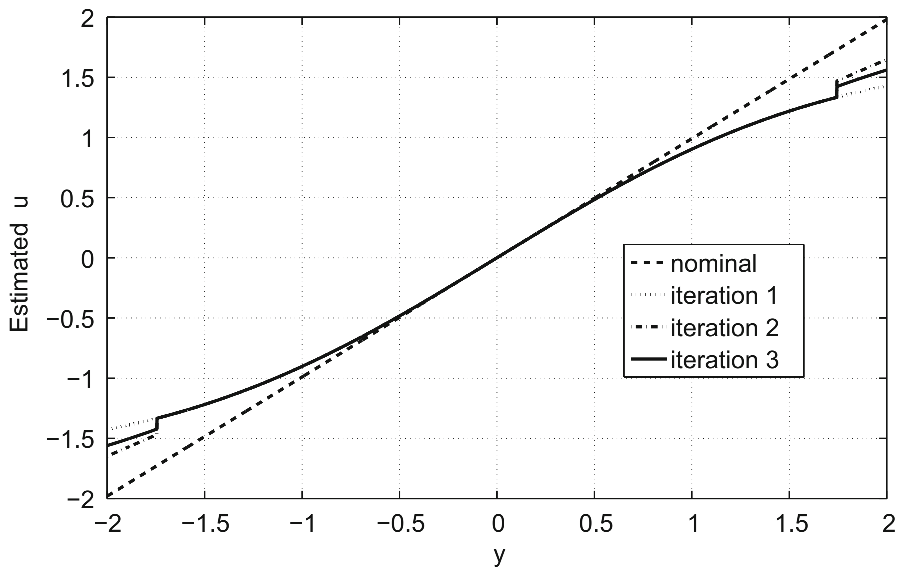


Fig. 5. Estimated values of u_k for FIR y_k=u_k+v_k with $\sigma_d^2=0.1$, $\sigma_v^2=10^{-2}$.

first iteration i=1 using (41), together with the nominal design, i=0, versus the amplitude γ of the uncertainty was obtained by simulation. The uncertainty is modelled by matrices Δ whose elements are realisations of a white Gaussian process with covariance matrix $\gamma^2 HH^T$, where γ is a scalar. For each realisation, comprising input sequences of 2000 samples, the relative standard deviation of the estimation error was calculated for the nominal design and for the robust design. Then, the average over 200 realisations was calculated. A signal to noise ratio of 20 dB was considered, and the horizons utilised were L_1=15 and 30 in all cases with L_2=1. The relative standard deviation of the error is shown in Fig. 8.

It is also important to note that the main improvement is achieved in the first iteration. Note that the performance of the nominal design improves when the horizon is increased, reaching a value without further improvements when the horizon is sufficiently large so as to capture all the dynamics of the optimal ARMAX filter. It can be seen that in the nominal case, with horizons between L_1=15 and 30 the filter converges as it was predicted in Fig. 6. It is important to highlight that the performance of the nominal design with horizon L_1=30 practically coincides with the Wiener solution. Also note that the improvements of the robust solution compared with the nominal design is significant. Note that, even though a

Author's personal copy

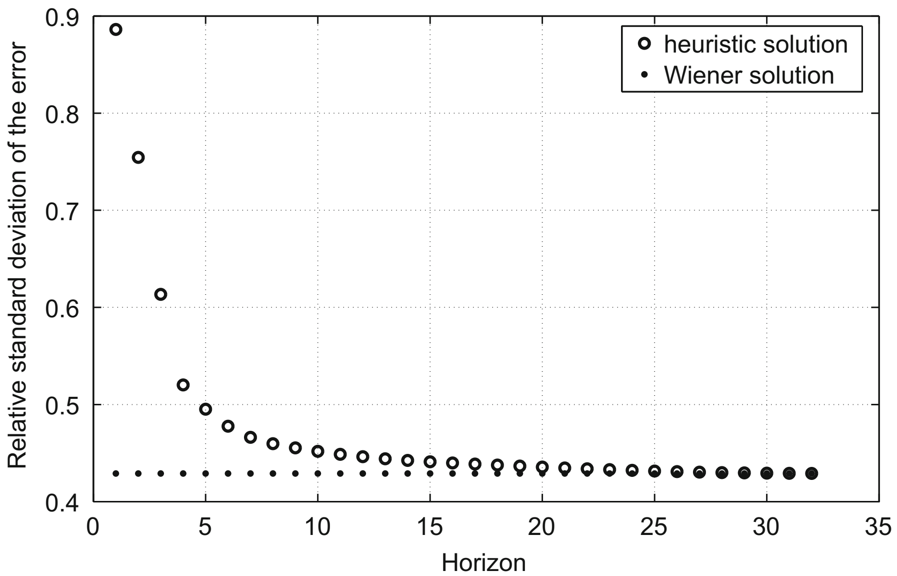


Fig. 6. Convergence of the heuristic cost to the Wiener solution as a function of the horizon length.

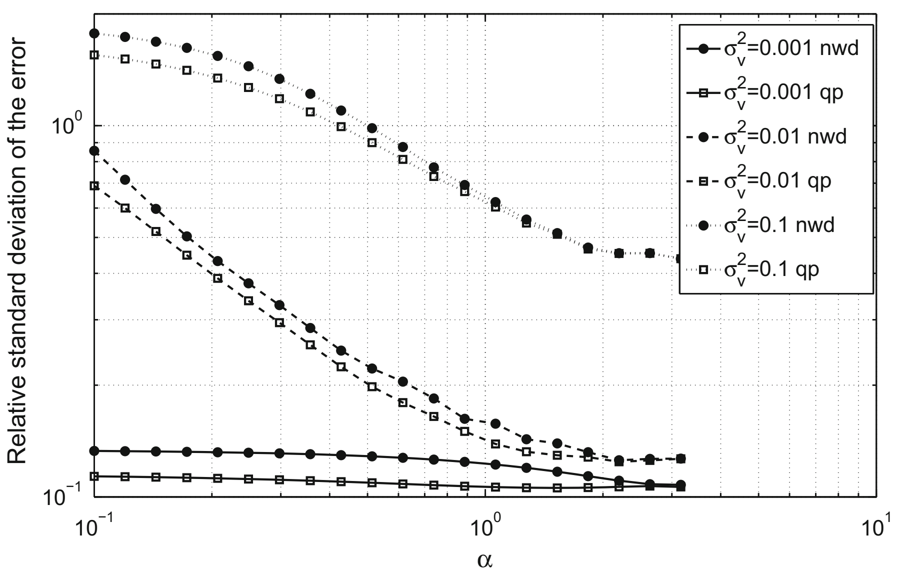


Fig. 7. Relative estimation error standard deviation for nominal with decision (nwd) and quadratic programme (qp) designs versus the half width α of the truncated Gaussian.

relative standard deviation of the error greater than one in the case of the nominal design has no significance, it is still shown to highlight how uncertainties deteriorate significantly the performance and the high improvement achieved with respect to the nominal design.

Example 3. In this example we consider the averaged performance over a family of stable systems. The systems considered are second order ARMAX (na=2;nb=2;nc=0) and consist of all possible combinations, without repetition, of zeros and poles located inside the unit disc, in positions corresponding to magnitudes of [0.9; 0.6; 0.3] and phases $[\pi/3; 2\pi/3; \pi]$. A signal to noise ratio, σ_u^2/σ_v^2 of 20 dB was considered, both Gaussian distributed.

To give a quantitative illustration of the averaged improvements that can be obtained when using the robust design in a problem of deconvolution with an uncertain linear operator H, extensive simulations were performed on the same system set up described above. Each of those systems belongs to a family of uncertainties given by matrices Δ whose elements are realisations of a white Gaussian process with covariance matrix $\gamma^2 HH^T$, where γ is a scalar. For each realisation, comprising input sequences of 2000 samples, the relative standard deviation of the estimation error was calculated for the nominal design and for the robust design. Then, for each system considered, the average over 200 realisations was calculated. A signal to noise ratio of 20 dB was considered, and the horizons utilised were L_1=25, 50, 100 and 500, in all cases with L_2=1. The result of averaging all the relative standard deviations of the error is shown in Fig. 9. It is

Author's personal copy

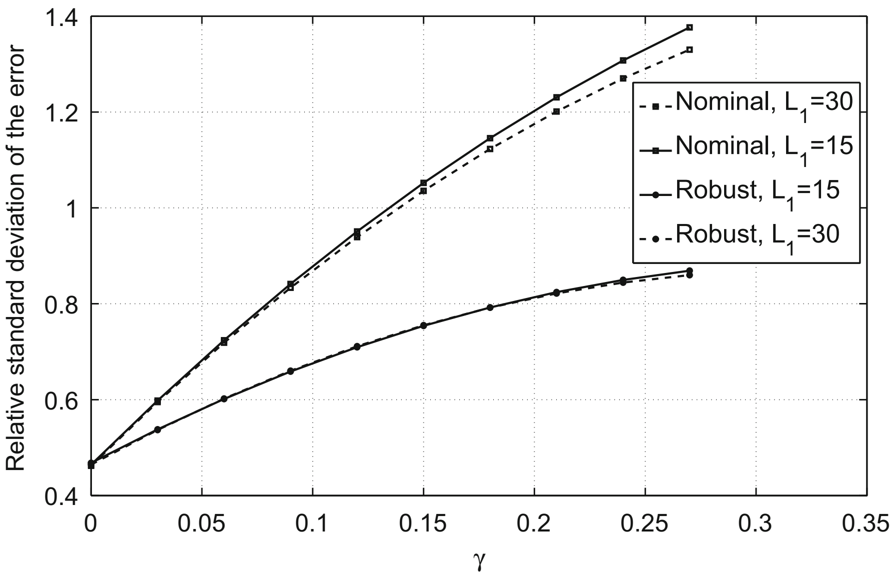


Fig. 8. Relative estimation error standard deviation for nominal and robust design versus amplitude of uncertainty.

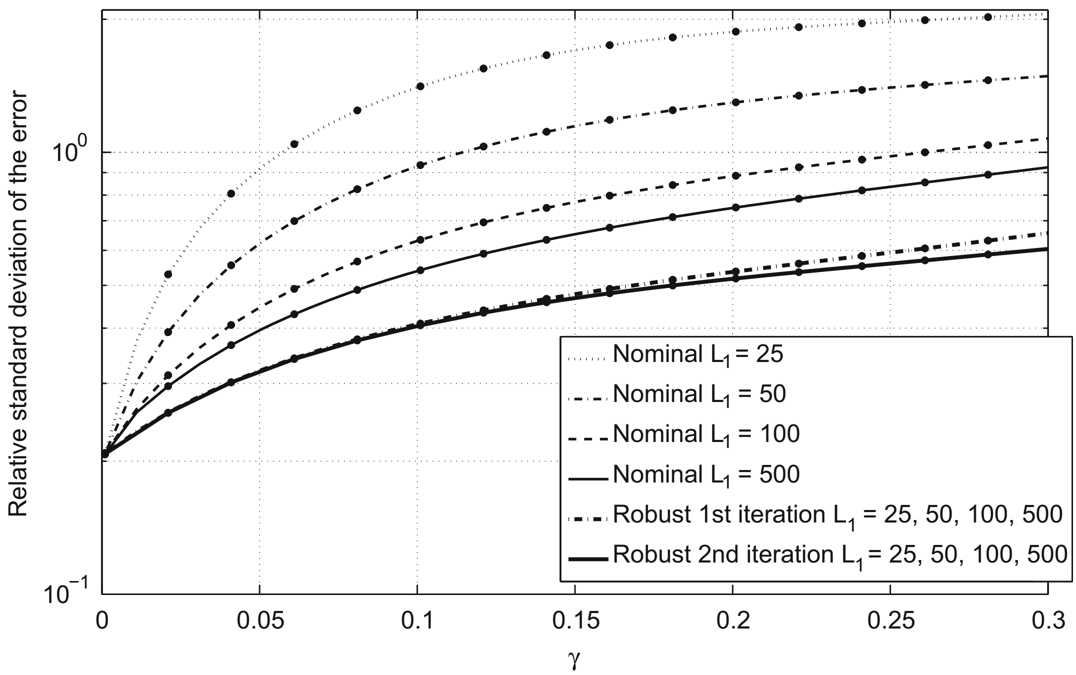


Fig. 9. Average value of the relative estimation error standard deviation for nominal and robust design versus uncertainty amplitude.

also important to note that the main improvement is achieved in the first iteration.

In the simulations, no significant variations were observed beyond L_1=500. In the case of the robust design, the changes in performance due to different horizons are significantly smaller and only for horizons smaller than L_1=10 the differences start to become noticeable. This confirms that the filter is robust against variations in the horizon of measurements. Thus, as it was stated in the Introduction, the robust design not only improves the quality of the estimates but significantly reduces the length of the horizons. Then the complexity of the on line estimations become simpler, extending the domains of application.

In Fig. 10, the performance obtained with the robust filter for truncated Gaussian inputs u_k is shown. The same quantities as before were averaged in this case, but now with a distribution for the input u_k given by a Gaussian distribution with zero mean, and with truncated negative values. The problem is solved as a quadratic programming problem subject to constraints. The simulations shown are for the uncertainties interval $\gamma \in [0.01,0.3]$.

Author's personal copy

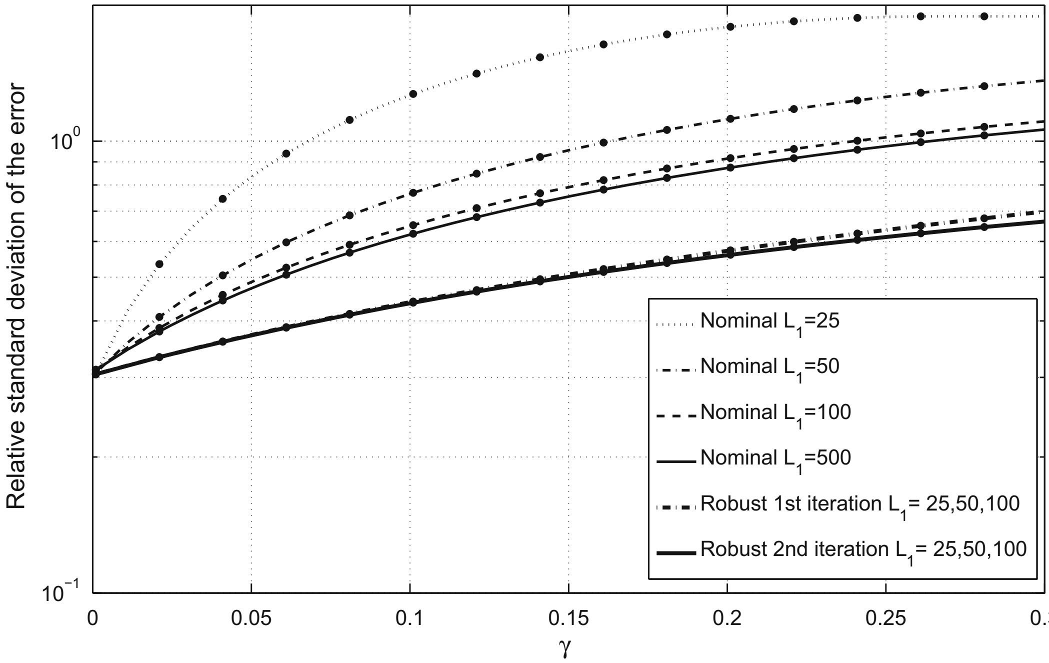


Fig. 10. Average value of the relative estimation error standard deviation for nominal and robust constrained design versus uncertainty amplitude.

In the case of perfect model, the average improvements between the constrained and unconstrained designs can be appreciated by comparing $\gamma = 0$.

6. Conclusion

A robust design for the problem of filtering with finite measurement horizons has been presented. The design maximises a functional motivated by the *a posteriori* probability and consists in iteratively improving the solution obtained for the nominal model. First, an estimation using the nominal model is carried out, and then the result is used to obtain, iteratively, an improved robust estimation. It is important to highlight that the greatest improvement, in all examples considered, is reached in the first iteration. Because of its formulation, the design is also feasible for the case of truncated Gaussian inputs.

For FIR systems the design is independent of the horizon. In contrast, for IIR systems the solution improves as the horizon is increased. For the latter case it was shown that for perfect model with Gaussian input the cost converges to the optimal solution given by the Wiener filter.

Acknowledgments

This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Agencia Nacional de Promoción Científica y Técnica, and Universidad Nacional del Comahue; all of República Argentina. The second author acknowledges support from the Centre for Complex Dynamic Systems and Control and the University of Newcastle, Australia.

Appendix A

Proof of Theorem 1. We will compare the MAP cost, given by the following equation:

$$\|\mathbf{y}^d - H\hat{\mathbf{x}}\|^2_{R_{\mathbf{v}}^{-1}} + \|\hat{\mathbf{x}}\|^2_{R_{\mathbf{x}}^{-1}} \tag{49}$$

where $R_{\mathbf{x}} = \mathcal{E}(\mathbf{x}\mathbf{x}^T)$ with the heuristic cost of Eq. (19) for any vector $\hat{\mathbf{x}}$, different from zero. We will show that the relative error between both criteria with respect to the MAP cost converge asymptotically to zero as L_1 increases, that is,

$$\lim_{L_1 \to \infty} \frac{\|\hat{\mathbf{x}}\|^2_{R_{\mathbf{x}}^{-1}} - \|E\hat{\mathbf{x}} + F\mathbf{y}^d\|^2_{P^{-1}}}{\|\mathbf{y}^d - H\hat{\mathbf{x}}\|^2_{R_{\mathbf{v}}^{-1}} + \|\hat{\mathbf{x}}\|^2_{R_{\mathbf{x}}^{-1}}} = 0 \tag{50}$$

To this end, consider an appropriate uncorrelated partition of the vector $\mathbf{x}$ as follows:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \mathbf{x}_a = \begin{pmatrix} \mathbf{w} \\ u_{k-L_1-nb} \\ \vdots \\ u_{k-L_1-1} \end{pmatrix}, \quad \mathbf{x}_b = \begin{pmatrix} u_{k-L_1} \\ \vdots \\ u_{k+L_2-1} \end{pmatrix} \tag{51}$$

Then, the covariance matrix $R_{\mathbf{x}}$ can be written as

$$R_{\mathbf{x}} = \begin{pmatrix} \mathcal{E}\mathbf{x}_a\mathbf{x}_a^T & \mathcal{E}\mathbf{x}_a\mathbf{x}_b^T \\ \mathcal{E}\mathbf{x}_b^2\mathbf{x}_a^T & \mathcal{E}\mathbf{x}_b\mathbf{x}_b^T \end{pmatrix} \tag{52}$$

By using the definition of $\mathbf{w}$ we can see that $\mathcal{E}(\mathbf{x}_a\mathbf{x}_b^T) = \mathcal{E}(\mathbf{x}_b\mathbf{x}_a^T) = 0$. Hence, we have

$$R_{\mathbf{x}}^{-1} = \begin{pmatrix} R_{\mathbf{x}_a}^{-1} & 0 \\ 0 & R_{\mathbf{x}_b}^{-1} \end{pmatrix} \tag{53}$$

where $R_{\mathbf{x}_a} = \mathcal{E}(\mathbf{x}_a, \mathbf{x}_a^T)$ and the same for $R_{\mathbf{x}_b}$. Then, for the first term of the numerator of (50) we have $\|\hat{\mathbf{x}}\|^2_{R_{\mathbf{x}}^{-1}} = \|\hat{\mathbf{x}}_a\|^2_{R_{\mathbf{x}_a}^{-1}} + \|\hat{\mathbf{x}}_b\|^2_{R_{\mathbf{x}_b}^{-1}}$.

Author's personal copy

For the second term we have $\|E\hat{\mathbf{x}}+F\mathbf{y}^d\|^2_{P^{-1}} = \|[\hat{\mathbf{v}}_0^T,\hat{\mathbf{u}}^T]^T\|^2_{P^{-1}}$. Using an equivalent partition to the one used previously, of the form $\|[\hat{\mathbf{v}}_0^T,\hat{\mathbf{u}}^T]^T\|^2_{P^{-1}} = \|[\hat{\mathbf{z}}^T,\hat{\mathbf{x}}_b^T]^T\|^2_{P^{-1}}$ where $\hat{\mathbf{z}}^T = [\hat{\mathbf{v}}_0^T,\hat{u}_{k-L_1-nb},\ldots,\hat{u}_{k-L_1-1}]$, we have that $\|E\hat{\mathbf{x}}+F\mathbf{y}^d\|^2_{P^{-1}} = \|\hat{\mathbf{z}}\|^2_{R_{\mathbf{z}}^{-1}} + \|\hat{\mathbf{x}}_b\|^2_{R_{\mathbf{x}_b}^{-1}}$.

Finally, substituting in (50) the quotient is given by

$$\lim_{L_1\to\infty} \frac{\|\hat{\mathbf{x}}_a\|^2_{R_{\mathbf{x}_a}^{-1}} - \|\hat{\mathbf{z}}\|^2_{R_{\mathbf{z}}^{-1}}}{\|\mathbf{y}^d - H\hat{\mathbf{x}}\|^2_{R_{\mathbf{v}}^{-1}} + \|\hat{\mathbf{x}}_a\|^2_{R_{\mathbf{x}_a}^{-1}} + \|\hat{\mathbf{x}}_b\|^2_{R_{\mathbf{x}_b}^{-1}}} = 0 \tag{54}$$

When L_1 grows, both terms of the numerator are bounded, whereas the term in the denominator, that depends on $\mathbf{x}_b$, grows unbounded. Thus, it is clear that for ARMAX systems (recall that for FIR systems, the heuristic cost coincides with the MAP cost), by increasing the horizon of the proposed estimator both criteria, the heuristic and the MAP criterion, converge asymptotically for any possible estimation. Thus, it follows that with long enough horizon, the minimiser of the MAP cost is the same minimiser of the heuristic cost. □

Appendix B

The effect on the variation of the cost that the uncertainties of the covariance matrix have can be quantified analysing expression (19). To this end, we write the cost in compact form as follows:

$$J(Q,H,\mathbf{x}) = \left\| \begin{pmatrix} I & -H \\ F & E \end{pmatrix} \begin{pmatrix} \mathbf{y}^d \\ \mathbf{x} \end{pmatrix} \right\|^2_Q \tag{55}$$

where Q=diag$(R_{\mathbf{v}}^{-1},P^{-1})$. The effect on the cost, caused by a lack of knowledge of the exact value of the matrix Q, can be quantified by computing the variation when, instead of Q, a positive definite matrix $\tilde{Q}$ is used, where $Q \le \gamma\tilde{Q}$, with γ a given positive scalar. By calling $\mathbf{x}_0$ the minimiser of $J(Q,H,\mathbf{x})$ and $\tilde{\mathbf{x}}$ the minimiser of $J(\tilde{Q},H,\mathbf{x})$, the change in the value of the functional is bounded by

$$J(Q,H,\mathbf{x}_0) \le J(Q,H,\tilde{\mathbf{x}}) \le \gamma J(\tilde{Q},H,\tilde{\mathbf{x}}) \tag{56}$$

It is important to note that the bound is minimised by $\tilde{\mathbf{x}}$. Hence, the design is indeed robust to variations in the matrix Q.

A similar analysis can be performed to calculate the variation of the cost for uncertainties in the knowledge of the system. Let us consider a percentage variation in the system when we use $\overline{H}$ instead of H, where $H = \overline{H}(1+\gamma^2)$. Then, the variation of the cost is

$$\left\| \begin{pmatrix} I & -H \\ F & E \end{pmatrix} \begin{pmatrix} \mathbf{y}^d \\ \hat{\mathbf{x}} \end{pmatrix} \right\|^2_Q = \left(\left\| \begin{pmatrix} I & -\overline{H} \\ F & E \end{pmatrix} \begin{pmatrix} \mathbf{y}^d \\ \hat{\mathbf{x}} \end{pmatrix} \right\|_Q + \gamma^2 \|\overline{H}\hat{\mathbf{x}}\|_{P^{-1}} \right)^2 \tag{57}$$

We conclude that the first term of the right-hand side of cost (57) is minimised by the design, but the second term depends on the values of the system multiplied by the estimation, which might be arbitrarily large.

References

[1] B. Zhang, J. Lam, S. Xu, Deconvolution filtering for stochastic systems via homogeneous polynomial Lyapunov functions, Signal Processing 89 (2009) 605–614.

[2] A.T. Erdogan, B. Hassibi, T. Kailath, On linear H_∞ equalization of communication channels, IEEE Trans. Signal Process. 48 (11) (2000) 3227–3231.

[3] J. Mohammadpour Velni, K.M. Grigoriadis, Rate-dependent mixed H_2/H_∞ filter design for parameter-dependent state delayed LPV systems, IEEE Trans. Circuit Systems I 55 (7) (2007) 2097–2105.

[4] C. de Souza, U. Shaked, Robust H_2 filtering for uncertain systems with measurable inputs, IEEE Trans. Signal Process. 47 (8) (1999) 2286–2290.

[5] J.C. Geromel, R.C.D. Martins, A new method to H_2 robust filter design, Linear Algebra Appl. 430 (2009) 145–154.

[6] R. Milocco, C. Muravchik, Cost guaranteed robust sampled-data parallel model design using polynomial approach, Signal Processing 85 (2005) 751–765.

[7] H. Zhang, D. Zhang, L. Xie, J. Lin, Robust filtering under stochastic parametric uncertainties, Automatica 40 (2004) 1583–1589.

[8] Y. Guo, B.C. Levy, Robust MSE equalizer design for MIMO communication systems in the presence of model uncertainties, IEEE Trans. Signal Process. 54 (5) (2006) 1840–1852.

[9] F. Yang, Z. Wang, G. Feng, Xi. Liu, Robust filtering with randomly varying sensor delay: the finite-horizon case, IEEE Trans. Circuit Systems I 56 (3) (2009) 664–672.

[10] K. Öhrn, A. Ahlén, M. Sternad, A probabilistic approach to multivariable robust filtering and open-loop control, IEEE Trans. Autom. Control 10 (1995) 405–418.

[11] E. Gershon, D.J.N. Limebeer, U. Shaked, I. Yaesh, Robust H filtering of stationary continuous-time linear systems with stochastic uncertainties, IEEE Trans. Autom. Control 46 (2001) 1788–1793.

[12] E.G. Larsson, P. Stoica, J. Li, On maximum-likelihood detection and decoding for space–time coding systems, IEEE Trans. Signal Process. 50 (4) (2002) 937–944.

[13] L.M. San José-Revuelta, J. Cid-Sueiro, Robust joint channel and noise estimation in Bayesian blind equalizers, Signal Processing 84 (2004) 535–548.

[14] X. Zhang, D.P. Palomar, B. Ottersten, Statistically robust design of linear MIMO transceivers, IEEE Trans. Signal Process. 56 (8) (2008) 3678–3689.

[15] G. Goodwin, M. Seron, J. De Doná, Constrained Control and Estimation. An Optimisation Approach, Springer-Verlag, London, 2005.

[16] C. Lauvernet, J. Brankart, F. Castruccio, G. Broquet, P. Brasseur, J. Verron, A truncated Gaussian filter for data assimilation with inequality constraints: application to the hydrostatic stability condition in ocean models, Ocean Modelling 27 (2009) 1–17.