Prefiltering in Iterative Feedback Tuning: Optimization of the Prefilter for Accuracy

R. Hildebrand, A. Lecchini, G. Solari, and M. Gevers

Abstract—**Iterative feedback tuning (IFT) is a data-based method for the tuning of restricted complexity controllers. At each iteration, an update for the controller parameters is estimated from data obtained partly from the normal operation of the closed loop system and partly from a special experiment, in which the output signal obtained under normal operation is fed back at the reference input. The choice of a prefilter for the input data to the special experiment is a degree of freedom of the method. In this note, the prefilter is designed in order to enhance the accuracy of the IFT update. The optimal prefilter produces a covariance of the new controller parameter vector that is strictly smaller than the covariance obtained with the standard constant prefilter.**

Index Terms—**Iterative feedback tuning (IFT), stochastic optimization.**

I. INTRODUCTION

Iterative feedback tuning (IFT) is a data-based method to tune the parameters of a controller with a given structure [2]–[4], [6], [7], [9]. The objective of IFT is to minimize a quadratic performance criterion defined on the space of controller parameters. IFT consists of an iterative stochastic gradient descent scheme. The gradient of the performance criterion is estimated from a set of data obtained partly from normal operation of the closed-loop system with the present controller, and partly from a special experiment on the plant, in which the output signal obtained under normal operation is fed back to the reference input. This gradient estimate is used to perform the next descent step in the iterations of the vector of controller parameters. Under suitable assumptions the algorithm converges to a local minimum of the performance criterion [3], [5].

In the IFT procedure, the user is given the possibility of prefiltering the input data to the special experiment. The choice of a particular prefilter is a degree of freedom of the procedure. In the original formulation of IFT [7] this degree of freedom was not used. In practice, this corresponds to using a trivial constant prefilter. In [2] and [4], it has been shown that the prefilter influences the covariance of the gradient estimate. A prefilter was derived which minimizes the weighted trace of the asymptotic covariance for a given weighting matrix. The main motivation for this result was the fact that near the optimal point the asymptotic convergence rate of IFT, measured as the rate of approach of the expected cost to the minimal value, can be optimized by minimizing the weighted trace of the asymptotic covariance of the gradient estimate for a specific and computable selection of the weight. By choosing that optimal weight, an optimal prefilter was derived that maximally improves the asymptotic convergence rate of the algorithm. The choice of the correct weight was based on an estimate of the Hessian of the design cost function near the optimal point.

R. Hildebrand is with the Laboratoire de Modélisation et Calcul (LMC), Université Joseph Fourier (UJF), 38041 Grenoble Cedex 9, France (e-mail: roland.hildebrand@imag.fr).

A. Lecchini is with the Department of Engineering, University of Cambridge, Cambridge CB2 1TN, U.K. (e-mail: al394@cam.ac.uk).

G. Solari and M. Gevers are with the Center for Systems Engineering and Applied Mechanics (CESAME), Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (e-mail: solari@csam.ucl.ac.be; gevers@csam.ucl.ac.be).

As far as we know, there are no clear results that allow one to formalize a design criterion for the convergence rate in the case where the current controller is far from the optimal one. As a matter of fact, the convergence depends on the global shape of the performance criterion. This shape is unknown and can hardly be estimated. Therefore, in this situation, the objective is mainly the improvement of the accuracy of every single IFT step. By reducing the deviation of the actual descent direction from the optimal direction at each iteration, one enhances the robustness of the whole iterative procedure.

Since we do not know the effect of the covariance of the gradient estimate on the expected cost of the IFT criterion for the reason explained above, we cannot compute the desired scalar measure of the covariance matrix of the gradient estimate, whose minimization would entail an optimal cost reduction. However, we show in this note that it is possible, by prefiltering, to obtain a covariance matrix that is smaller than the one obtained with the standard IFT step, i.e., with a trivial constant prefilter. By smaller we mean that the difference between the covariance matrix obtained with standard IFT and the new one is positive semidefinite. We propose a design criterion for the prefilter that is consistent with this goal and show how a prefilter can be computed that optimizes this design criterion. It turns out, as it was the case in [2] and [4], that the proposed prefilter can be estimated from data collected under normal operating conditions. Thus, the computation of the prefilter does not require any special experiment on the process and does not impose any additional cost.

Our optimal filter, computed and applied at every iteration of the IFT scheme, will have the effect of decreasing the spread (as measured by the covariance) of the gradient estimate that determines the descent direction in the iterative scheme. While this should, on average, improve the convergence speed of the search procedure and reduce the occurrences of converging to wrong local minima, it does in no way guarantee that the algorithm will not converge to an unsatisfactory local minimum as a result of poor initial conditions.

The note is structured as follows. In the next section we recall some results on the statistical properties of the gradient estimate in IFT. This enables us in Section III to establish a design criterion that has to be minimized with respect to the prefilter in order to reduce the covariance of the gradient estimate. We also show how to compute a prefilter that is optimal with respect to this design criterion. In Section IV, we demonstrate, by a simulation example, the gain in accuracy between the use of the optimal prefilter and the use of the trivial constant prefilter. Finally, we draw some conclusions in the last section.

II. IFT PARAMETER UPDATE

We assume that the plant to be controlled is a single-input–single-output (SISO) linear time-invariant system; its transfer function is denoted by $G(q)$. The output of the plant is affected by an additive stochastic disturbance $v(t) = H(q)e(t)$, where $H(q)$ is a monic, stable, and inversely stable transfer function and $e(t)$ is zero mean white noise with variance σ^2. The transfer functions $G(q)$ and $H(q)$ are unknown. We consider the closed loop system depicted in Fig. 1, where $C(q, \rho)$ belongs to a parameterized set of controllers with parameter vector $\rho \in \mathbf{R}^n$. The transfer function from $v(t)$ to $y(t, \rho)$ is denoted by $S(q, \rho)$. In this note, we consider the situation where the control objective is to perform disturbance rejection only. Thus, normal operating conditions are defined as those for which the reference signal $r(t)$ is set at zero. The goal is to find a minimizer for the cost function

$$J(\rho) = \frac{1}{2}\mathbf{E}\left[y(t, \rho)^2 + \lambda u(t, \rho)^2\right] \tag{1}$$

where $\lambda \geq 0$ is a penalty on the control effort chosen by the user.
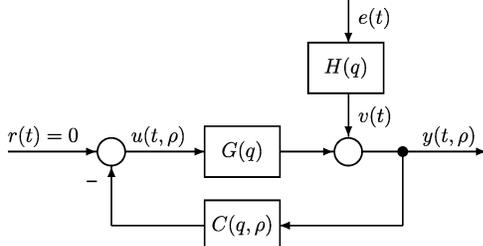
Fig. 1.   Control system under normal operating conditions.

The IFT method is an iterative procedure that gives a solution to this problem. It is based on the construction, at each iteration step, of an unbiased estimate of the gradient of $J(\rho)$ from data collected on the plant with some present controller $C(\rho)$ in operation. The cost function $J(\rho)$ is minimized with an iterative stochastic gradient descent scheme of Robbins–Monro type [1]. After each iteration, the controller parameter vector is updated ($\rho_i \rightarrow \rho_{i+1}$), and new data are collected on the closed-loop system with this updated controller in order to estimate the next gradient. Under some suitable assumptions [3], [7], the sequence of controllers converges to a local minimum of $J(\rho)$.

Since every iteration proceeds in the same fashion, and since this note analyzes (and optimizes in a sense to be defined) the covariance of the gradient estimate in a single iteration, we will in the sequel analyze the step where we move from the current controller parameter, denoted by ρ_0, to the updated parameter ρ_1. The IFT parameter update rule is given by

$$\rho_1 = \rho_0 - \gamma R^{-1} \text{est}_N \left[\frac{\partial J}{\partial \rho}(\rho_0) \right] \tag{2}$$

where γ is a positive step size and R is a positive definite matrix. The reader is referred to [7] for details of the algorithm to construct the gradient estimate.

Here, it suffices to recall that, in order to construct $\text{est}_N[(\partial J/\partial \rho)(\rho_0)]$, first a batch $\{u^1(t, \rho_0), y^1(t, \rho_0)\}_{t=1,\ldots,N}$ of N data is collected with the controller $C(q, \rho_0)$ in the loop under normal operating conditions, i.e., with $r(t) = 0$. Then, this batch of data is used to construct the signal sequence $\{r(t) = -K(q)y^1(t, \rho_0)\}_{t=1,\ldots,N}$ which is applied to the reference input of the system (see Fig. 1); this is the "special experiment," which deviates from normal operating conditions. It produces a second batch of input and output data $\{u^2(t, \rho_0), y^2(t, \rho_0)\}_{t=1,\ldots,N}$, which are used to construct the following estimates of the gradients of $u^1(t, \rho_0)$ and $y^1(t, \rho_0)$:

$$\text{est} \left[\frac{\partial u^1}{\partial \rho}(t, \rho_0) \right] = \frac{1}{K(q)} \frac{\partial C}{\partial \rho}(q, \rho_0) u^2(t, \rho_0)$$

$$\text{est} \left[\frac{\partial y^1}{\partial \rho}(t, \rho_0) \right] = \frac{1}{K(q)} \frac{\partial C}{\partial \rho}(q, \rho_0) y^2(t, \rho_0).$$

An estimate of the gradient of $J(\rho)$ at ρ_0 is then obtained as

$$\text{est}_N \left[\frac{\partial J}{\partial \rho}(\rho_0) \right] = \frac{1}{N} \sum_{t=1}^{N} \left[y^1(t, \rho_0) \text{est} \left[\frac{\partial y^1}{\partial \rho}(t, \rho_0) \right] \right. $$
$$\left. + \lambda u^1(t, \rho_0) \text{est} \left[\frac{\partial u^1}{\partial \rho}(t, \rho_0) \right] \right]. \tag{3}$$

The choice of the prefilter $K(q)$ is a degree of freedom of the algorithm and is basically left to the user. In the original formulation of IFT the prefilter was not used, which corresponds to setting $K(q) = const.$

The prefilter $K(q)$ influences the statistical properties of the IFT update as has been shown in [7]. More specifically, we have the following proposition.

Proposition 2.1: Let $\bar{P} = \lim_{N \to \infty} N \mathbf{Cov}[\rho_1]$. Then, $\bar{P}$ can be decomposed as $\bar{P} = \bar{E} + \bar{S}$ where $\bar{E}$ is given by

$$\bar{E} = \gamma^2 R^{-1} \left[\frac{\sigma^4}{2\pi} \int_{-\pi}^{\pi} \frac{\left| S(e^{j\omega}, \rho_0) H(e^{j\omega}) \right|^4}{\left| K(e^{j\omega}) \right|^2} \right. $$
$$\left. \times \left[1 + \lambda \left| C(e^{j\omega}, \rho_0) \right|^2 \right]^2 \frac{\partial C}{\partial \rho}(e^{j\omega}, \rho_0) \frac{\partial C^*}{\partial \rho}(e^{j\omega}, \rho_0) d\omega \right] R^{-T}$$

and $\bar{S}$ is a constant matrix which does not depend on the choice of $K(q)$.

Proof: See [3].   □

This result shows how the covariance matrix of the parameter update depends on the prefilter $K(q)$. It is the sum of a constant term and a term that is a frequency weighted integral of the inverse of $K(q)$. In Section III, we will show how to choose $K(q)$ in order to make this covariance matrix smaller than that obtained with a constant prefilter.

III. DESIGN OF AN OPTIMAL PREFILTER

By Proposition 2.1, we can influence the covariance term $\bar{E}$ by choice of the prefilter $K(q)$. Our goal is to make $\bar{E}$ as small as possible by choosing $K(q)$ appropriately. However, here we deal with a matrix-valued object, minimization of which has no well-defined meaning. Nevertheless, there exists a partial ordering on the space of symmetric matrices. Namely, if the difference $\bar{P}_2 - \bar{P}_1$ of two symmetric matrices $\bar{P}_1$, $\bar{P}_2$ is positive definite, we can say that $\bar{P}_1$ is strictly smaller than $\bar{P}_2$. Clearly, if we have a choice between two different prefilters $K(q)$ yielding two different covariance matrices, which are comparable in this sense, then it is preferable to use that prefilter which leads to the smaller covariance matrix.

Specifically, if we can find a prefilter which leads to a covariance matrix $\bar{P}$ which is strictly smaller than the covariance matrix obtained by using no prefilter at all, then it is preferable to use this prefilter. In this section, we point out a subset of prefilters which lead to such covariance matrices, and we propose an algorithm to construct the prefilter in that set which leads to the smallest covariance matrix.

We will proceed as follows. First, we clarify the structure of the set of all covariance matrices $\bar{P}$ which can be obtained at all by using all possible prefilters $K(q)$. This will give us clues for the construction of a prefilter which leads to a covariance matrix that is smaller than the one obtained with some given prefilter. We will then construct a prefilter that yields a covariance matrix which is smaller than the one obtained by using no prefilter at all (i.e., by using the trivial constant prefilter).

A. Set of Achievable Covariance Matrices

By Proposition 2.1, we can write

$$\bar{E} = \int_{0}^{\pi} \frac{1}{\left| K(e^{j\omega}) \right|^2} M(\omega) d\omega \tag{4}$$

i.e., $\bar{E}$ is a weighted integral over a frequency-dependent real positive–semidefinite matrix $M(\omega)$, which itself does not depend on the prefilter $K(q)$. The prefilter is assigned the role of a positive weighting function. Thus, by choice of a suitable prefilter, we can assign to the expression $\bar{E}$ a value that is arbitrarily close to any given matrix in the convex conic hull of the matrix-valued curve $M(\omega)$. By Proposition 2.1, the closure of the set of all asymptotic covariance matrices $\bar{P}$ that
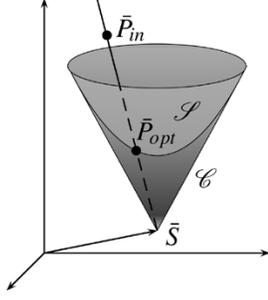
Fig. 2.   Representation of the optimization problem.

can be achieved by choosing a prefilter is an affine closed convex cone with offset $\bar{S}$. Let us denote this cone by $\mathcal{C}$.

Since $\bar{E}$ is inversely proportional to the squared magnitude of the prefilter $K(q)$, we could make it as small as we wish by choosing a prefilter with a sufficiently large magnitude. However, this would be at the cost of a higher input energy for the reference signal $r(t) = -K(q)y^1(t, \rho_0)$ of the special experiment and this would represent a bigger perturbation of the process, which is not desirable. Thus, we have to restrict the set of allowed prefilters $K(q)$ by imposing some bound α on this input energy $\mathbf{E}[r(t)^2]$. This bound represents the level of acceptable perturbation of the normal operating conditions during the special experiment. Since $r(t) = -K(q)y^1(t, \rho_0)$ and $y^1(t, \rho_0) = S(q, \rho_0)H(q)e(t)$, we impose the following restriction on the magnitude of $K(q)$: $(\sigma^2/2\pi) \int_{-\pi}^{\pi} |K(e^{j\omega})|^2 |S(e^{j\omega}, \rho_0)H(e^{j\omega})|^2 d\omega \leq \alpha$. This can be written as

$$\int_0^\pi \left| K(e^{j\omega}) \right|^2 w(\omega) d\omega \leq 1 \qquad (5)$$

where $w(\omega) = (\sigma^2/\alpha\pi)|S(e^{j\omega}, \rho_0)H(e^{j\omega})|^2$ is a positive frequency-dependent scalar function. The set of covariance matrices $\bar{P}$ that can be obtained by prefiltering under restriction (5) is naturally smaller than the entire cone $\mathcal{C}$ defined previously. We will now investigate this set.

Let $\bar{P}_{\text{in}} = \bar{S} + \bar{E}_{\text{in}} \in \mathcal{C} \setminus \{\bar{S}\}$ be an arbitrary covariance matrix which can be achieved by prefiltering with some prefilter satisfying constraint (5). Then, the ray $\{P(\kappa) = \bar{S} + \kappa \bar{E}_{\text{in}} | \kappa \geq 0\}$ is contained in the cone $\mathcal{C}$. It is easily seen that there exists a unique $\kappa_{\text{opt}} > 0$ such that any matrix $P(\kappa) = \bar{S} + \kappa \bar{E}_{\text{in}}$ with $\kappa \geq \kappa_{\text{opt}}$ can be approximated arbitrarily well by choosing prefilters satisfying constraint (5), but matrices $P(\kappa)$ with $\kappa < \kappa_{\text{opt}}$ cannot. The union of the matrices $\bar{P}_{\text{opt}} = \bar{S} + \kappa_{\text{opt}} \bar{E}_{\text{in}}$ for all rays in $\mathcal{C}$ forms a section $\mathcal{S}$ of the cone $\mathcal{C}$ (see Fig. 2).

Proposition 3.1: The section $\mathcal{S}$ is a convex hypersurface.

Proof: It is sufficient to show that the closure of the set of all covariance matrices that can be achieved by prefiltering under restriction (5) is convex. Let $K_1(q)$, $K_2(q)$ be two prefilters satisfying (5) and yielding covariance matrices $\bar{P}_1, \bar{P}_2 \in \mathcal{C}$, respectively. Let $K_\tau(q)$ be such that $1/|K_\tau(q)|^2 = \tau(1/|K_1(q)|^2) + (1 - \tau)(1/|K_2(q)|^2)$, $\tau \in [0, 1]$. Then, by convexity of the function $f(x) = 1/x$ every filter $K_\tau(q)$ satisfies restriction (5). Thus, the matrices on the line segment between the covariance matrices $\bar{P}_1, \bar{P}_2$ can be approximated arbitrarily well along with the endpoints of this segment. $\square$

Prefilters $K(q)$ satisfying (5) and leading to covariance matrices on the surface $\mathcal{S}$ are optimal in the following sense. Suppose we are given some prefilter $K_{\text{in}}(q)$ satisfying (5) and leading to a covariance matrix $\bar{P}_{\text{in}} = \bar{S} + \bar{E}_{\text{in}}$ in the convex hull of $\mathcal{S}$. Then, the covariance matrix $\bar{P}_{\text{opt}} = \bar{S} + \kappa_{\text{opt}} \bar{E}_{\text{in}} \in \mathcal{S}$ is the *smallest* covariance matrix on the ray $\{\bar{S} + \kappa \bar{E}_{\text{in}} | \kappa \geq 0\}$ which can be approximated by using prefilters satisfying (5). In particular, we have $\kappa_{\text{opt}} \leq 1$ and $\bar{P}_{\text{opt}}$ is smaller than

$\bar{P}_{\text{in}}$. In this sense a prefilter $K_{\text{opt}}(q)$ which satisfies (5) and produces the covariance matrix $\bar{P}_{\text{opt}}$ is optimal along the given ray. In the next section, we will construct this optimal prefilter.

B. Optimal Prefilter

The problem of finding the optimal prefilter $K_{\text{opt}}(q)$ for which the corresponding covariance matrix $\bar{P}_{\text{opt}}$ is smaller than some given initial covariance matrix $\bar{P}_{\text{in}}$ can be cast as the following optimization problem [compare (4) and (5)]. Given an initial prefilter $K_{\text{in}}(q)$ which realizes a covariance matrix $\bar{P}_{\text{in}} = \bar{S} + \bar{E}_{\text{in}}$, minimize κ, by choice of $K(e^{j\omega})$, under the constraints

$$\int_0^\pi \left| K(e^{j\omega}) \right|^2 w(\omega) d\omega \leq 1 \quad \int_0^\pi \frac{1}{|K(e^{j\omega})|^2} M(\omega) d\omega = \kappa \bar{E}_{\text{in}}.$$

The solution of this optimization problem yields K_{opt} and $\kappa_{\text{opt}} \leq 1$, with $\bar{P}_{\text{opt}} = \bar{S} + \kappa_{\text{opt}} \bar{E}_{\text{in}} \leq \bar{P}_{\text{in}}$. A natural choice of the initial prefilter $K_{\text{in}}(q)$ is the trivial constant prefilter with maximum gain satisfying the energy constraint.

We will now solve this problem. By associating a scalar Lagrange multiplier $\lambda_0 \geq 0$ and a matrix-valued Lagrange multiplier Λ with the two constraints, we obtain the Lagrange function

$$L = \kappa + \lambda_0 \left(\int_0^\pi \left| K(e^{j\omega}) \right|^2 w(\omega) d\omega - 1 \right)$$
$$+ \left\langle \Lambda, \int_0^\pi \frac{1}{|K(e^{j\omega})|^2} M(\omega) d\omega - \kappa \bar{E}_{\text{in}} \right\rangle$$

where $\langle A, B \rangle$ denotes $\text{trace}(A^T B)$. By setting the gradient of L with respect to the design variables κ and $K(\omega)$ to zero and inserting the resulting equations into the constraints, we get

$$|K_{\text{opt}}(\omega)|^2$$
$$= \frac{\sqrt{\langle \Lambda, M(\omega) \rangle}}{w(\omega) \int_0^\pi \sqrt{\langle \Lambda, M(\omega) \rangle w(\omega)}}$$
$$\frac{\partial}{\partial \Lambda} \left(-\ln \int_0^\pi \sqrt{\langle \Lambda, M(\omega) \rangle w(\omega)} d\omega + \frac{1}{2} \langle \Lambda, \bar{E}_{\text{in}} \rangle \right)$$
$$= \frac{\partial f(\Lambda)}{\partial \Lambda} = 0. \qquad (6)$$

The function $f(\Lambda)$ is convex with respect to Λ. In order to determine Λ, we thus have to solve the following convex optimization problem.

$$\text{minimize } f(\Lambda) \quad \text{s.t. } \langle \Lambda, M(\omega) \rangle \geq 0 \qquad \forall \omega. \qquad (7)$$

The matrix-valued function $M(\omega)$ is rational in $\cos(\omega)$. Therefore, the constraint $\langle \Lambda, M(\omega) \rangle > 0 \ \forall \omega$ is semidefinite representable and can be formulated as a linear matrix inequality (LMI). Problem (7) is thus a standard convex optimization problem for which efficient numerical solution algorithms are available. Note that the gradient of f tends to infinity as Λ approaches the boundary of the feasible set. Hence, at the optimum, if it exists, the inequality in (7) is strict. Having determined Λ, one can compute the magnitude of the prefilter according to (6). Then, there exist standard tools to approximate a given magnitude function by a minimum phase filter.

Practical aspects: The solution of (7) and, hence, the computation of the optimal prefilter, requires knowledge of $M(\omega)$ and $w(\omega)$: see (4), (5), and Proposition 2.1. The only unknown in $M(\omega)$ and $w(\omega)$ is the spectral density $|S(e^{j\omega}, \rho_0)H(e^{j\omega})|^2$ of the signal $y(t, \rho_0)$ which is the output of the plant under normal operating conditions, i.e., with
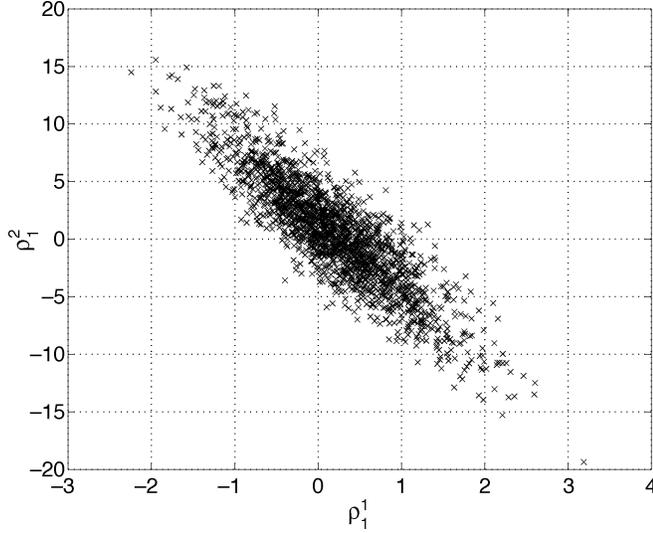
Fig. 3.   2048 parameters ρ_1 obtained using the constant prefilter $K_{\mathrm{in}}(q)$.

Fig. 4.   2048 parameters ρ_1 obtained using the optimal prefilter $K_{\mathrm{opt}}(q)$.

$r(t) = 0$. Such estimate can be obtained with standard techniques in the time or in the frequency domain [8], [10] from the measured signal $y(t, \rho_0)$. By assuming periods of normal operating conditions to be much longer than the length of the special experiment from which the gradient is estimated, the contribution of the variability in the estimate of $|S(e^{j\omega}, \rho_0)H(e^{j\omega})|^2$ to the variability of the gradient estimate can be considered as being negligible.

Before closing this section, let us remind the reader that the optimal prefilter is guaranteed to yield a smaller covariance matrix of the updated parameter vector than the one which would be obtained with the standard IFT procedure. It is optimal in the sense described in the previous subsection.

IV. SIMULATION EXAMPLE

Consider the system described by $G(q) = (q^{-1} - 0.5q^{-2})/(1 - 0.3q^{-1} - 0.28q^{-2})$ and $H(q) = 1/(1 + 0.9q^{-1})$ with $\sigma^2 = 1$. Let the class of controllers be $C(q, \rho) = \rho^1/(1 + \rho^2 q^{-1})$ and set $\lambda = 0.6$ in (1). Let the current stabilizing controller be given by: $\rho_0 = [-1.25\ 0.83]$. We assume that the constraint on the reference signal $r(t)$ of the special experiment is that this signal must have the same energy as the output of the plant in normal operating conditions. In the following, we quantify the accuracy improvement on the parameter vector update ρ_1, when the optimal prefilter $K_{\mathrm{opt}}(q)$, given by (6), is used instead of the trivial constant prefilter $K_{\mathrm{in}}(q)$, both of which having a maximum gain satisfying the energy constraint on $\mathbf{E}[r(t)^2]$.

We consider the IFT update (2) with step size $\gamma = 1$ and $R = I$. By using Proposition 2.1 and results from [3] we can find the numerical value of $\bar{P} = \lim_{N \to \infty} N\mathbf{Cov}[\rho_1]$.

For the constant prefilter $K_{\mathrm{in}}(q)$, we obtain

$$\bar{P}_{\mathrm{in}} = 10^4 \times \begin{bmatrix} 0.0524 & -0.3019 \\ -0.3019 & 2.2765 \end{bmatrix} = \bar{S} + \bar{E}_{\mathrm{in}}$$

in which the $K(q)$-dependent component is given by

$$\bar{E}_{\mathrm{in}} = 10^4 \times \begin{bmatrix} 0.0472 & -0.3053 \\ -0.3053 & 2.2735 \end{bmatrix}.$$

For the optimal prefilter $K_{\mathrm{opt}}(q)$, we have

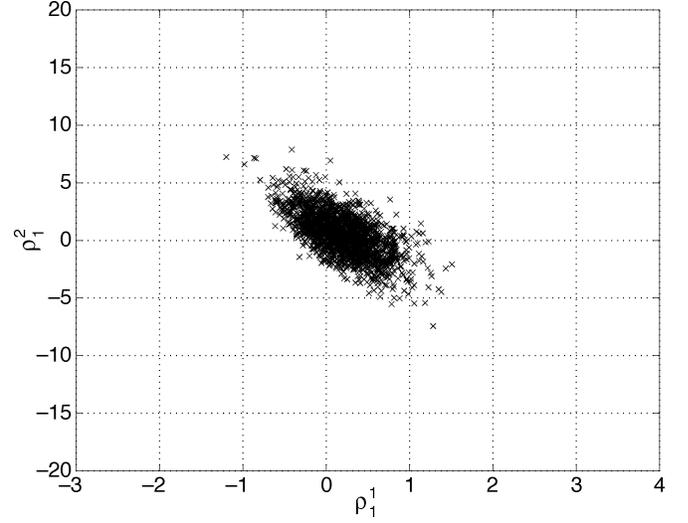$$\bar{P}_{\mathrm{opt}} = 10^3 \times \begin{bmatrix} 0.1245 & -0.4309 \\ -0.4309 & 3.4975 \end{bmatrix} = \bar{S} + \bar{E}_{\mathrm{opt}}$$

where $\bar{E}_{\mathrm{opt}} = \kappa_{\mathrm{opt}}\bar{E}_{\mathrm{in}}$ with $\kappa_{\mathrm{opt}} = 0.15$. The difference in the total covariance for the two cases is then

$$\bar{\Delta} = \bar{P}_{\mathrm{in}} - \bar{P}_{\mathrm{opt}} = 10^4 \times \begin{bmatrix} 0.0399 & -0.2588 \\ -0.2588 & 1.9267 \end{bmatrix} > 0.$$

Thus, whatever scalar measure of the covariance matrix one might use to evaluate the spread of ρ_1, the use of $K_{\mathrm{opt}}(q)$ leads to a strict improvement.

The aforementioned theoretical values can be verified by a Monte Carlo simulation. The parameter vector ρ_1 has been extracted 2048 times, by performing 2048 times the parameter vector update (2) with a different noise realization for the two batches of data, each batch being of length $N = 1024$. The 2048 parameter vectors obtained in this way are shown in Fig. 3 for the case of the constant prefilter. The corresponding sampled estimate of $N\mathbf{Cov}[\rho_1]$ is given by

$$\hat{P}_{\mathrm{in}} = 10^4 \times \begin{bmatrix} 0.0523 & -0.3003 \\ -0.3003 & 2.2338 \end{bmatrix}.$$

The 2048 parameter vectors obtained for the case of the optimal prefilter are shown in Fig. 4. Note that the optimal prefilter K_{opt} is identical for all 2048 runs, since it only depends on the statistics of the output process during normal operation, and not on individual signal realizations. In this case, the corresponding sampled estimate of $N\mathbf{Cov}[\rho_1]$ is given by

$$\hat{P}_{\mathrm{opt}} = 10^3 \times \begin{bmatrix} 0.1270 & -0.4344 \\ -0.4344 & 3.5599 \end{bmatrix}.$$

We observe that the results of the Monte Carlo simulation are remarkably close to the theoretical results.

V. CONCLUSION

In this note, we have exploited the possibility to prefilter the reference input in the special experiment of the IFT procedure in order to enhance the robustness of the algorithm. Based on the relationship between the prefilter and the covariance of the parameter update, we have formulated a design criterion for the prefilter which, in each IFT iteration, replaces the covariance of the parameter update obtained with standard IFT by a smaller covariance. The problem of optimizing the prefilter with respect to this criterion amounts to a convex optimization problem with LMI constraints.

REFERENCES

[1] J. R. Blum, "Multidimensional stochastic approximation methods," *Ann. Math. Statist.*, vol. 25, pp. 737–744, 1954.

[2] R. Hildebrand, A. Lecchini, G. Solari, and M. Gevers, "Optimal pre-filtering in iterative feedback tuning," in *Proc. 13th IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003, pp. 513–518.

[3] ——, "Asymptotic accuracy of iterative feedback tuning," *IEEE Trans. Automat. Contr.*, 2004, submitted for publication.

[4] ——, "Optimal prefiltering in iterative feedback tuning," *IEEE Trans. Automat. Contr.*, 2004, submitted for publication.

[5] H. Hjalmarsson, "Performance analysis of iterative feedback tuning," Dept. Signals, Sensors and Systems, Royal Inst. Technol., Stockholm, Sweden, Tech. Rep., 1998.

[6] ——, "Iterative feedback tuning—An overview," *Int. J. Adapt. Control Signal Processing*, vol. 16, no. 5, pp. 373–395, 2002.

[7] H. Hjalmarsson, M. Gevers, S. Gunnarson, and O. Lequin, "Iterative feedback tuning: theory and applications," *IEEE Control Syst. Mag.*, vol. 18, pp. 26–41, Aug. 1998.

[8] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[9] "Special section on algorithms and applications of iterative feedback tuning," *Control Eng. Pract.*, vol. 11, no. 9, pp. 1021–1094, Sept. 2003.

[10] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. New York: IEEE Press, 2001.

Nonlinearity Recovering in Wiener System Driven With Correlated Signal

Włodzimierz Greblicki

Abstract—The characteristic of the nonlinear part of the Wiener system is estimated. The system is driven by a Gaussian random process which may not be white. Three algorithms are presented, two semirecursive and one of the offline type. Their pointwise convergence in probability is shown and results of numerical simulation are given.

Index Terms—Nonparametric identification, nonparametric regression, system identification, Wiener system.

I. INTRODUCTION

In the block-oriented approach, we identify systems consisting of linear dynamic and nonlinear memoryless subsystems. In particular, in the Wiener system, a linear dynamic part is followed by a nonlinear memoryless one. The objective is to recover descriptions of the subsystems from input–output observations of the whole system. In this note, we focus our attention on the nonlinear subsystem.

In the literature, the *a priori* information about both subsystems is usually parametric; see, e.g., [1], [3], [7], [8], and [10]. As a result, the nonlinear characteristic is a function known up to some coefficients [10], or just a polynomial [3], [7], [8]. The random input signal is Gaussian [3], [8], but not necessarily, [10], white, [3], or correlated, [8], [10]. Due to serious theoretical difficulty, properties of some algorithms are only suggested by simulation examples [1], [7].

The characteristic can be recovered also when the *a priori* information is smaller than parametric, i.e., nonparametric. Nonparametric
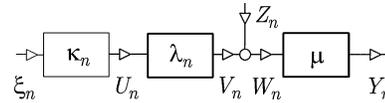
Fig. 1. Wiener system (thick line) and the system forming its input signal (thin line).

algorithms work properly even if we know nothing about the nonlinear characteristic which can be any Borel measurable function [5]. Needless to say that parametric algorithms are then completely useless. Thus, for a user, nonparametric methods have a great advantage, since they can be applied when the knowledge about the system is so poor that its description cannot be parameterized.

Offline [4], [5] and semirecursive, [6], nonparametric kernel algorithms have been examined. In those papers, the input signal is a Gaussian white random process. The novelty of this note is the fact that we admit also a correlated input signal. To estimate the nonlinear characteristic, we propose three kernel algorithms, of which two are semirecursive and one is offline. We show that all are pointwise consistent and present results of numerical simulation.

II. SYSTEM AND ALGORITHMS

The Wiener system with input U_n and output Y_n, shown in Fig. 1 with the thick line, consists of two subsystems. First is linear, dynamic while the other nonlinear, memoryless. The dynamic subsystem has an impulse response $\{\lambda_n; n = 0, 1, \dots\}$ which means that

$$V_n = \sum_{i=-\infty}^{n} \lambda_{n-i} U_i \tag{1}$$

where $\{U_n; n = \dots, -1, 0, 1, \dots\}$. The impulse response $\{\lambda_n\}$ is unknown but we assume that

$$|\lambda_n| \leq \delta_1 \varepsilon_1^n \tag{2}$$

with some unknown δ_1 and ε_1, where $0 \leq \varepsilon_1 < 1$. Observe that the restriction is satisfied by every stable ARMA system.

We assume that the input signal is correlated and is of the following form:

$$U_n = \sum_{i=-\infty}^{n} \kappa_{n-i} \xi_i \tag{3}$$

where $\{\xi_n; n = \dots, -1, 0, 1, \dots\}$ is a stationary white random Gaussian process with zero mean and unknown variance σ_ξ^2. The signal $\{\xi_n\}$ is not measured. The weighting function $\{\kappa_n\}$ is unknown but satisfies the following inequality:

$$|\kappa_n| \leq \delta_2 \varepsilon_2^n \tag{4}$$

with some unknown δ_2 and ε_2, where $0 \leq \varepsilon_2 < 1$. Thus, $\{U_n\}$ is output of some system with unknown impulse response $\{\kappa_n\}$ driven by $\{\xi_n\}$; see the thin line in Fig. 1.

Output of the subsystem is disturbed by additive stationary white Gaussian noise $\{Z_n; n = \dots, -1, 0, 1, \dots\}$ independent of both $\{U_n\}$ and $\{V_n\}$ and having zero mean and unknown finite variance. Owing to all that, both $\{V_n; n = \dots, -1, 0, 1, \dots\}$ as well $\{W_n; n = \dots, -1, 0, 1, \dots\}$ are also stationary and Gaussian.

The unknown characteristic of the nonlinear part, denoted by μ, has a continuous derivative such that, for all $w \in R$

$$0 < |\mu'(w)| \leq c_\mu \tag{5}$$