OPTIMAL PREFILTERING IN ITERATIVE FEEDBACK TUNING [1]

R. Hildebrand, [*] **A. Lecchini,** [**] **G. Solari,** [**] **M. Gevers** [**]

[*] *Center for Operations Research and Econometrics (CORE)*
Université Catholique de Louvain
Voie du Roman Pays, 34
B-1348 Louvain-la-Neuve, Belgium
{hildebrand}@core.ucl.ac.be
[**] *Center for Systems Engineering and Applied Mechanics (CESAME)*
Université Catholique de Louvain
Bâtiment Euler, 4 Av. Georges Lemaître,
B-1348 Louvain-la-Neuve, Belgium
{lecchini, solari, gevers}@auto.ucl.ac.be

Abstract: Iterative Feedback Tuning (IFT) is a widely used procedure for controller tuning. It is a sequence of iteratively performed special experiments on the plant interlaced with periods of data collection under normal operating conditions. In this paper we derive the asymptotic convergence rate of IFT for disturbance rejection, which is one of the main fields of application. Further we present a method to improve the convergence of IFT by prefiltering the input data for the special experiment. At each iteration step the optimal prefilter is computed from data collected under normal operating conditions of the plant.

Keywords: Iterative Feedback Tuning, filter design

1. INTRODUCTION

Iterative Feedback Tuning (IFT) is a data based method for the tuning of restricted complexity controllers. It has proved to be very effective in practice and is now widely used in process control, often for disturbance rejection. Following the original formulation of the method in (Hjalmarsson *et al.*, 1998) many improvements and modifications of IFT have been suggested. The reader is referred to (Hjalmarsson, 2002) for a recent overview.
The objective of IFT is to minimize a quadratic performance criterion. IFT is a stochastic gradient descent scheme in a finitely parameterized controller space. The gradient of the cost function at each step is estimated from data. These data are collected with the actual controller in the loop. Under suitable assumptions the algorithm converges to a local minimum of the performance criterion. One of the advantages of IFT is that most data are collected while the process runs under normal operating conditions. These data are then used to design a special experiment, which yields a noisy, but unbiased, estimate of the cost function gradient. This gradient estimate is used to perform the next descent step in controller space. For more details of the procedure see (Hjalmarsson *et al.*, 1998).
In this and in the companion paper (Hildebrand *et al.*, 2002) we focus on IFT for disturbance rejection. We provide an analytic expression for the asymptotic convergence rate of the algorithm, as the number of data collected in each experiment tends to infinity.

[1] Paper supported by the the Belgian Programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture and the European Research Network on System Identification (ERNSI) funded by the European Union. The scientific responsibility rests with its authors.

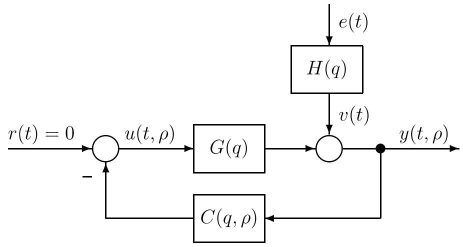

Fig. 1. The control system under normal operating conditions.

The convergence rate depends on the covariance of the gradient estimates. Therefore, the calculation of this covariance is a part of our analysis.

In (Hjalmarsson *et al.*, 1998) it was proposed to reduce the error in the gradient estimate by prefiltering the reference input data for the special experiment. Our second contribution is to optimize the corresponding prefilter with respect to both the convergence speed of the procedure and the accuracy of the finally obtained value. An analytical expression for the optimal prefilter is given. It depends on certain characteristics of the unknown process. However, in the spirit of IFT, these characteristics can be estimated from data collected under normal operating conditions. Thus the computation of the optimal prefilter does not necessitate any special experiment on the process and hence does not impose any additional cost.

The remainder of the paper is structured as follows. In the next section we summarize the details of the IFT algorithm for disturbance rejection. In Section 3 we derive an expression for the asymptotic convergence rate dependent on the covariance of the gradient estimates. In Section 4 the asymptotic expression of this covariance is calculated. This enables us in Section 5 to establish a design criterion for the optimization of the algorithm. We compute the optimal prefilter which minimizes the design criterion. In Section 6 we demonstrate the gains in convergence speed and accuracy in a simulation example. Finally, we draw some conclusions in the last section.

2. IFT FOR DISTURBANCE REJECTION

In this section we review the IFT method for the disturbance rejection problem with a classical LQ criterion. For a more general and detailed presentation of IFT the reader is referred to (Hjalmarsson *et al.*, 1998).

Consider a SISO discrete time system described by

$$y(t) = G(q)u(t) + v(t)\,, \qquad (1)$$

where $y(t)$ is the output, $u(t)$ is the input, $G(q)$ is a linear time-invariant transfer function, with q being the shift operator, and $v(t) = H(q)e(t)$ is the process disturbance. Here $H(q)$ 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.

Consider the feedback loop around $G(q)$ depicted in Figure 1, where $C(q,\rho)$ is a one-degree-of-freedom controller belonging to a parameterized set of controllers with parameter $\rho \in \mathbf{R}^n$. The transfer function from $v(t)$ to $y(t,\rho)$ is named sensitivity function and is denoted by $S(q,\rho)$. We assume that in the control system of Figure 1 the reference signal $r(t)$ is set at zero under normal operating conditions. Our goal is to tune the controller $C(q,\rho)$ so that the variance of the noise-driven closed loop output is as small as possible subject to a penalty on the control effort. Thus we want to find a minimizer for the cost function

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

where $\lambda \geq 0$ is chosen by the user.

The IFT method yields an approximate solution to the above problem. IFT is based on the possibility of obtaining an unbiased estimate of the gradient $\frac{\partial J}{\partial \rho}(\rho)$ of the cost function at $\rho = \rho_n$ from data collected from the closed-loop system with the controller $C(\rho_n)$ operating on the loop. The cost function $J(\rho)$ can then be minimized with an iterative stochastic gradient descent scheme of Robbins-Monro type (Blum, 1954). In the scheme a sequence of controllers $C(q,\rho_n)$ is computed and applied to the plant. In the n-th iteration step, data obtained from the system with the controller $C(\rho_n)$ operating on the loop are used to construct the next parameter vector ρ_{n+1}.

The data based iterative procedure is as follows.

IFT procedure

1. Collect a sequence $\{u^1(t,\rho_n),\, y^1(t,\rho_n)\}$ with $t = 1,\ldots,N$ of input-output data under normal operating conditions, i.e. without reference signal.

2. Collect a sequence $\{u^2(t,\rho_n),\, y^2(t,\rho_n)$ with $t = 1,\ldots,N$ of input-output data by performing a special experiment with reference signal

$$r_n^2(t) = -K_n(q)y^1(t,\rho_n)$$

where $K_n(q)$ is any stable minimum-phase prefilter.

3. Construct the estimates of the gradients of $u^1(t,\rho_n)$ and $y^1(t,\rho_n)$ as

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

4. Form the estimate of the gradient of $J(\rho)$ at ρ_n as

$$est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right] = \frac{1}{N}\sum_{t=1}^{N}\Bigg[y^1(t,\rho_n)$$
$$\times est\left[\frac{\partial y^1}{\partial \rho}(t,\rho_n)\right] + \lambda u^1(t,\rho_n) est\left[\frac{\partial u^1}{\partial \rho}(t,\rho_n)\right]\Bigg]$$

5. Calculate the new parameter vector ρ_{n+1} according to

$$\rho_{n+1} = \rho_n - \gamma_n R_n^{-1}\, est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right]$$

where γ_n is a positive step size and R_n is a positive definite matrix.

In the above procedure we assume negligible measurement noise and independence between the disturbance realizations $v_n^1(t)$, in the first experiment, and $v_n^2(t)$, in the second experiment. Under these assumptions the estimate of the gradient turns out to be unbiased (Hjalmarsson *et al.*, 1998). The sequences γ_n and R_n are basically left to the choice of the user. The matrix R_n should be an approximation of the Hessian of the cost function in ρ_n. A biased estimate of the Hessian, obtained from data, has been proposed in (Hjalmarsson *et al.*, 1998).

3. ANALYSIS OF CONVERGENCE RATE IN IFT

In this section we quantify the effect of the variability of the gradient estimate on the asymptotic convergence rate of the algorithm.

The proposition below describes the asymptotic behavior of the sequence ρ_n. It follows from a general proposition on the convergence rate of for Robbins-Monro processes as can be found in (Nevelson and Khasminskii, 1976).

Proposition 1. Assume that the sequence ρ_n converges to a local isolated minimum $\bar{\rho}$ of $J(\rho)$ (the reader is referred to (Hildebrand *et al.*, 2002) for the conditions of convergence). Let H be the Hessian of $J(\rho)$ at $\rho = \bar{\rho}$. Suppose further that the following conditions hold.

1. The sequence γ_n of step sizes is given by $\gamma_n = \frac{a}{n}$, where a is a positive constant. There exists an index $\bar{n}$ and a matrix R such that $R_n = R$ for all $n > \bar{n}$.
2. The matrix $A = \frac{1}{2}I - aR^{-1}H$ is stable, i.e. the real parts of its eigenvalues are negative.
3. The covariance matrix $\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial \rho}(\rho)\right]\right]$ at $\rho = \bar{\rho}$ is positive definite.

Then the sequence of random variables $s_n = \sqrt{n}(\rho_n - \bar{\rho})$ converges in distribution to a normally distributed zero mean random variable with covariance matrix

$$\Sigma = \int_0^\infty e^{At} R^{-1}\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial \rho}(\bar{\rho})\right]\right] \times R^{-1} e^{A^T t} dt,$$

i.e. $\sqrt{n}(\rho_n - \bar{\rho}) \xrightarrow{D} \mathcal{N}(0, \Sigma)$. □

Proposition 1 shows that the asymptotic accuracy of the estimate crucially depends on the distribution of the error on the gradient. This distribution in turn can be influenced by the prefilters $K_n(q)$. Before turning to the question of designing the filters $K_n(q)$ for optimal accuracy, we analyze in detail how the covariance of the gradient estimate depends on $K_n(q)$. This will be done in the next section.

4. THE COVARIANCE OF THE GRADIENT ESTIMATE

This section is devoted to finding an explicit expression for the covariance of $est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right]$. We will show that this covariance can be written as the sum of two terms. These two contributions originate in the variability of the noise realizations in the first and second experiment of iteration n, respectively. Consequently, the first term is independent of the prefilter $K_n(q)$, because the filter is applied only to the reference signal for the second experiment. However, the second term can be influenced by the choice of this prefilter.

It can be shown that the estimates of the gradients of $u^1(t,\rho_n)$ and $y^1(t,\rho_n)$ obtained in Step 3 of the IFT procedure are corrupted by the realization $v_n^2(t)$ of the noise in the second experiment as follows

$$\begin{aligned}
est\left[\frac{\partial u^1}{\partial \rho}(t,\rho_n)\right] &= \frac{\partial u^1}{\partial \rho}(t,\rho_n) - \frac{S(q,\rho_n)}{K_n(q)} C(q,\rho_n) \\
&\quad \frac{\partial C}{\partial \rho}(q,\rho_n)\, v_n^2(t)\,, \\
est\left[\frac{\partial y^1}{\partial \rho}(t,\rho_n)\right] &= \frac{\partial y^1}{\partial \rho}(t,\rho_n) + \frac{S(q,\rho_n)}{K_n(q)} \\
&\quad \frac{\partial C}{\partial \rho}(q,\rho_n)\, v_n^2(t)\,.
\end{aligned}$$

Therefore we can separate $est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right]$ as

$$\begin{aligned}
& est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right] = S_N(\rho_n) + E_N(\rho_n)\,, \\
& S_N(\rho_n) = \frac{1}{N}\sum_{t=1}^{N}\Bigg[y^1(t,\rho_n)\frac{\partial y^1}{\partial \rho}(t,\rho_n) \\
& + \lambda u^1(t,\rho_n)\frac{\partial u^1}{\partial \rho}(t,\rho_n)\Bigg] \\
& E_N(\rho_n) = \frac{1}{N}\sum_{t=1}^{N}\Bigg[y^1(t,\rho_n)\Bigg[\frac{S(q,\rho_n)}{K_n(q)} \\
& \frac{\partial C}{\partial \rho}(q,\rho_n)\, v_n^2(t)\Bigg] + \lambda u^1(t,\rho_n) \\
& \times \Bigg[-\frac{C(q,\rho_n)S(q,\rho_n)}{K_n(q)}\frac{\partial C}{\partial \rho}(q,\rho_n)\, v_n^2(t)\Bigg]\Bigg]\,.
\end{aligned}$$

The term $S_N(\rho_n)$ corresponds to the sampled estimate of the gradient of $J(\rho)$. This term is entirely dependent on the realization $v_n^1(t)$ of the noise in the first experiment. The second term $E_N(\rho_n)$ is an error due to the corruption of the estimates of the gradients of $u^1(t,\rho_n)$ and $y^1(t,\rho_n)$ by $v_n^2(t)$. The estimate $est_N\left[\frac{\partial J}{\partial\rho}(\rho_n)\right]$ turns out to be unbiased under the assumption that the two experiments in the algorithm are sufficiently separated in time. In fact, under this assumption, the realization $v_n^2(t)$ can be considered as being independent of the signals coming form the first experiment and therefore the mean of $E_N(\rho_n)$ is zero. The dispersion of $est_N\left[\frac{\partial J}{\partial\rho}(\rho_n)\right]$ is described in the following proposition.

Proposition 2.
1. The following relation holds

$$\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial\rho}(\rho_n)\right]\right] = \mathbf{Cov}\left[S_N(\rho_n)\right] + \mathbf{Cov}\left[E_N(\rho_n)\right] .$$

2. The following asymptotic frequency-domain expression of $\mathbf{Cov}\left[E_N(\rho_n)\right]$ holds

$$\lim_{N\to\infty} N\mathbf{Cov}\left[E_N(\rho_n)\right] = \frac{\sigma^4}{2\pi}\int_{-\pi}^{\pi}\frac{1}{|K_n(e^{j\omega})|^2}|S(e^{j\omega},\rho_n)H(e^{j\omega})|^4\left[1 + \lambda|C(e^{j\omega},\rho_n)|^2\right]^2 \times \frac{\partial C}{\partial\rho}(e^{j\omega},\rho_n)\frac{\partial C^*}{\partial\rho}(e^{j\omega},\rho_n)\,d\omega\,.$$

3. Under the additional assumption that the 4th order cumulants of the noise v are zero (e.g the noise is normally distributed), the following asymptotic frequency-domain expression of $\mathbf{Cov}\left[S_N(\rho_n)\right]$ holds

$$\lim_{N\to\infty} N\mathbf{Cov}\left[S_N(\rho_n)\right] = 2\cdot\frac{\sigma^4}{2\pi}\int_{-\pi}^{\pi}|S(e^{j\omega},\rho_n)H(e^{j\omega})|^4 \times \mathcal{R}e\left\{\left[G(e^{j\omega}) - \lambda\bar{C}(e^{j\omega},\rho_n)\right]S(e^{j\omega},\rho_n)\frac{\partial C}{\partial\rho}(e^{j\omega},\rho_n)\right\} \times\mathcal{R}e\left\{\left[G(e^{j\omega}) - \lambda\bar{C}(e^{j\omega},\rho_n)\right]S(e^{j\omega},\rho_n) \times\frac{\partial C}{\partial\rho}(e^{j\omega},\rho_n)\right\}^T d\omega\,.$$

Proof See (Hildebrand *et al.*, 2002). $\square$

In Proposition 2 it has been shown that the covariance of the gradient estimate can be represented as the sum of the covariances of the separate contributions $S_N(\rho_n)$ and $E_N(\rho_n)$ (i.e. $S_N(\rho_n)$ and $E_N(\rho_n)$ are uncorrelated). Both $\mathbf{Cov}\left[S_N(\rho_n)\right]$ and $\mathbf{Cov}\left[E_N(\rho_n)\right]$ decay asymptotically proportionally to $1/N$ as the number of data tends to infinity. Their asymptotic frequency domain expressions as $N\to\infty$ have been given.

By Proposition 1, $\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial\rho}(\rho_n)\right]\right]$ enters linearly into the asymptotic covariance Σ of the estimated controller parameter value. Hence this covariance can as well be expressed as a sum of two terms, dependent on the covariances of $S_N(\rho_n)$ and $E_N(\rho_n)$, respectively. Since $S_N(\rho_n)$ does not depend on $K_n(q)$, the corresponding term in Σ can be regarded as constant for the purpose of designing the prefilter $K_n(q)$. This observation leads us to the design of an optimal prefilter.

5. DESIGN OF THE OPTIMAL PREFILTER

We are now ready to specify the criterion for the design of the prefilter $K_n(q)$. In this section we state this criterion and deliver the expressions of the corresponding optimal prefilter. We assume that the current controller is near the optimal one. Then the convergence rate of the procedure is measured by the accuracy of the estimate. Therefore, in order to construct a design criterion for the prefilter, one can employ the asymptotic results on the accuracy given in Section 3.

5.1 The design criterion

Let the sequence γ_n of step lengths in the IFT procedure be proportional to $1/n$, i.e. $\gamma_n = \frac{a}{n}$. Define $\Delta\bar{\rho}_n = \rho_n - \bar{\rho}$, where $\bar{\rho}$ is the optimal parameter. Let us take $\mathbf{E}[J(\rho_n)] - J(\bar{\rho})$ as a measure of quality of the controller $C(\rho_n)$. Expanding $J(\rho)$ into a Taylor series around $\bar{\rho}$ and retaining only terms up to the second order, we obtain $\mathbf{E}\left[\Delta\bar{\rho}_n^T\mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right]$ as an approximation of $\mathbf{E}[J(\rho_n)] - J(\bar{\rho})$. Here $\mathbf{H}(\rho)$ denotes the Hessian of $J(\rho)$.
Following Proposition 1, $\sqrt{n}\Delta\bar{\rho}_n$ is asymptotically normally distributed with zero mean and covariance

$$\Sigma = a^2\int_0^\infty e^{At}R^{-1}\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial\rho}(\bar{\rho})\right]\right] \times\left[R^{-1}\right]^T e^{A^Tt}dt,$$

where $R = \lim_{n\to\infty}R_n$ and $A = \frac{1}{2}I - aR^{-1}\mathbf{H}(\bar{\rho})$. Let us now assume $R = \mathbf{H}(\bar{\rho})$, i.e. we consider a Gauss-Newton scheme. Then we obtain $A = \left(\frac{1}{2} - a\right)I$ and

$$\Sigma = \frac{a^2}{2a-1}R^{-1}\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial\rho}(\bar{\rho})\right]\right]\left[R^{-1}\right]^T \quad (3)$$

This yields

$$\lim_{n\to\infty} n\mathbf{E}\left[\Delta\bar{\rho}_n^T\mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right] = \frac{a^2}{2a-1} \times\mathrm{Trace}\left[\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial\rho}(\bar{\rho})\right]\right]\left[R^{-1}\right]^T\right].$$

We shall take this expression as the criterion to be minimized for the design of the optimal prefilter.

There are different methods to satisfy the condition $R = \mathbf{H}(\bar{\rho})$. A classical method to obtain a sequence of estimated matrices R_n which converges to the Hessian is to fit a regression model using the gradient estimates obtained in the previous iterations. The reader is referred to (Wei, 1985; Yin, 1988).

In practice, in order to use (4) as a criterion for the design of the optimal filter, at the iteration n one has to replace the optimal parameter $\bar{\rho}$ on the right-hand side by the current parameter ρ_n, since (as it will be shown in Subsection 5.2) the prefilter is estimated from data obtained under the current operating conditions. In the same way, R has to be replaced by the current estimate of the Hessian R_n. This estimate could be the (biased) data-based estimate of the Hessian proposed in (Hjalmarsson *et al.*, 1998) which is constructed, at each step, with the data of the first and second experiment. However, in order to not violate a certain condition for the convergence of ρ_n (see (Hildebrand *et al.*, 2002)), the estimated R_n has to be uncorrelated with the noise realizations $v_n^1(t)$ and $v_n^2(t)$. Therefore it has to be calculated by using the data of iteration $n-1$. These approximations are reasonable, because we assume that the current controller is near the optimal controller.

5.2 *The optimal prefilter*

The quantity that has to be optimized in the design of the prefilter is thus (4) with $\bar{\rho}$ replaced by ρ_n. The optimal prefilter minimizes the weighted trace of the covariance of the gradient estimate. In order to obtain a bounded solution we have to restrict the gain of the prefilter. A straightforward constraint is a bound on the energy of the reference signal $r_n^2(t)$, i.e. on the input of the second experiment. This bound represents the level of acceptable perturbation to the normal operating conditions during the second experiment at each step. We thus arrive at the following optimization problem:

$$K_n^{opt} = \arg\min_K \text{Trace}\left[R_n^{-1}\mathbf{Cov}\left[est_N\left[\frac{\partial J}{\partial \rho}(\rho_n)\right]\right]\right]$$
$$\text{subject to} \quad \mathbf{Var}\left[r_n^2(t)\right] \leq \alpha,$$

where α is selected by the user. By Proposition 2, and recalling that $\mathbf{Cov}\left[S_N(\rho_n)\right]$ does not depend on the prefilter, we can rewrite the problem as follows:

$$K_n^{opt} = \arg\min_K \text{Trace}\left[R_n^{-1}\mathbf{Cov}\left[E_N(\rho_n)\right]\right] \quad (4)$$
$$\text{subject to} \quad \mathbf{Var}\left[r_n^2(t)\right] \leq \alpha.$$

The explicit solution of this problem is characterized by the following proposition.

Proposition 3. The optimal prefilter solving (4) satisfies the following relation:

$$|K_n^{opt}(e^{j\omega})|^4 = const \cdot |S(e^{j\omega}, \rho_n)H(e^{j\omega})|^2 [1 + \lambda |C(e^{j\omega}, \rho_n)|^2]^2 \,\text{Trace}\left\{R_n^{-1}\frac{\partial C}{\partial \rho}(e^{j\omega}, \rho_n) \times \frac{\partial C^*}{\partial \rho}(e^{j\omega}, \rho_n)\right\}, \quad (5)$$

where the constant is determined by the design restriction.

Proof. See (Hildebrand *et al.*, 2002). □

In order to compute the optimal prefilter in practice, one needs an estimate of the unknown spectral density $|S(e^{j\omega}, \rho_n)H(e^{j\omega})|^2$ of the signal $y(t, \rho_n)$ which is the output of the plant under normal operating conditions, i.e. with zero reference signal. The estimate can be obtained with standard techniques in the time or in the frequency domain (Ljung, 1999; Pintelon and Schoukens, 2001). Note that since the data needed to estimate this quantity do not stem from a special experiment they are available in large amounts. In fact periods of normal operating conditions can be interlaced with the IFT special experiments. By assuming these periods 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 $|S(e^{j\omega}, \rho_n)H(e^{j\omega})|^2$ to the variability of the gradient estimate can be considered as being negligible.

Having an estimate of $|S(e^{j\omega}, \rho_n)H(e^{j\omega})|^2$ one can construct the magnitude of the optimal prefilter by calculating the 4-th root of the right-hand side of (5). Then, there exist standard tools to approximate a given magnitude function by a stable minimum phase filter.

6. SIMULATION EXAMPLE

Consider the system described by

$$G(q) = \frac{q^{-1} - 0.5q^{-2}}{1 - 0.3q^{-1} - 0.28q^{-2}},$$
$$H(q) = \frac{1}{1 + 0.9q^{-1}}$$

with $\sigma^2 = 1$. Let the class of controllers be $C(q, \rho) = \rho^1 + \rho^2 q^{-1}$ and set $\lambda = 0.6$ in (2). The (local) minimizer $\bar{\rho} = [-0.69058\ 0.33105]$ has been found numerically. Let us assume that the constraint on the reference signal $r_n^2(t)$ during the IFT procedure is that this signal has to have one half the energy of the output of the first experiment. In the following we will quantify the performance improvement between the trivial constant filter satisfying the energy constraint and the optimal filter given by (5) when the criterion $\mathbf{E}\left[\Delta\bar{\rho}_n^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right]$ is used. We run the IFT procedure with experiment length $N = 512$, step sizes $\gamma_n = a/n$ with $a = 1$ and $R_n = \mathbf{H}(\bar{\rho})$.

Using the asymptotic approximations of the covariance of the gradient estimate given in Section 3 we can

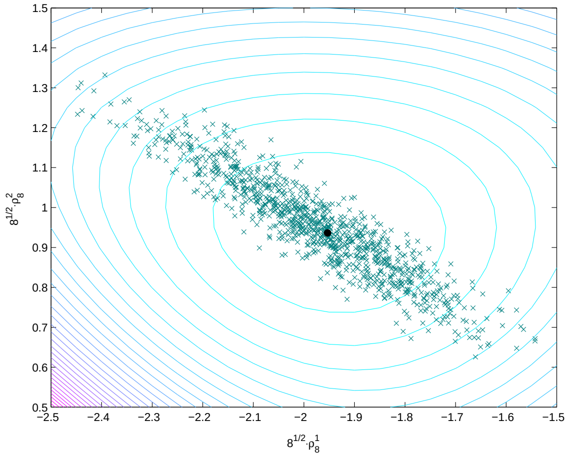

Fig. 2. The parameter $\bar{\rho}$ (•), the 1024 parameters ρ_8 obtained using the constant filter ($\times$) and the contour lines of $J(\rho)$.

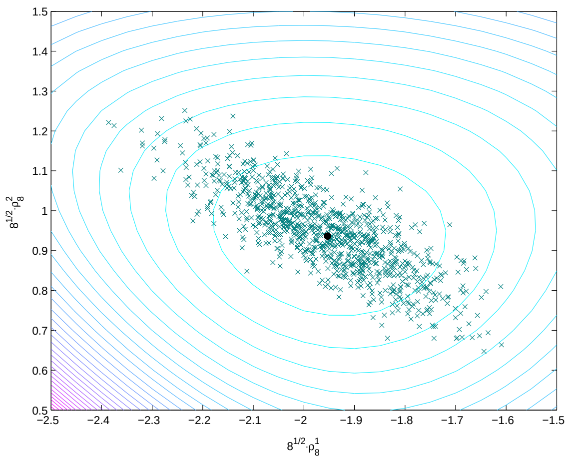

Fig. 3. The parameter $\bar{\rho}$ (•), the 1024 parameters ρ_8 obtained using the optimal filter ($\times$) and the contour lines of $J(\rho)$.

find the asymptotic covariance of $\sqrt{n}\Delta\bar{\rho}_n$ according to (3). Then, for the constant prefilter we obtain that the asymptotic value of $n\mathbf{E}\left[\Delta\bar{\rho}_n^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right]$ is $8.39 \cdot 10^{-2}$. For the optimal prefilter the asymptotic value of $n\mathbf{E}\left[\Delta\bar{\rho}_n^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right]$ is $5.79 \cdot 10^{-2}$. The improvement in the value of $\mathbf{E}\left[\Delta\bar{\rho}_n^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_n\right]$ between the two cases is 31%.

The above theoretical values can be illustrated by a Monte-Carlo simulation. The parameter vector ρ_8 has been extracted 1024 times. The extractions have been performed by starting the IFT procedure at $\rho_0 = \bar{\rho}$, in order to eliminate the transient effect of the initial condition, and then running 8 steps up to the parameter vector ρ_8. The 1024 parameter vectors obtained this way are shown in Figure 2 for the case of the constant prefilter. The corresponding sampled estimate of $8 \cdot \mathbf{E}\left[\Delta\bar{\rho}_8^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_8\right]$ was $8.00 \cdot 10^{-2}$. The parameter vectors obtained for the case of the optimal prefilter are shown in Figure 3. In this case, the corresponding sampled estimate of $8 \cdot \mathbf{E}\left[\Delta\bar{\rho}_8^T \mathbf{H}(\bar{\rho})\Delta\bar{\rho}_8\right]$ was $5.48 \cdot 10^{-2}$. The estimated improvement between the two cases hence equals 31.5%, as predicted by the theoretical calculations made above.

7. CONCLUSIONS

In this contribution we have investigated the convergence properties of the IFT algorithm for disturbance rejection and shown how its performance can be improved by prefiltering the reference input for the special experiment.

The asymptotic convergence rate of the algorithm for a step size sequence proportional to $1/n$ was quantified as a function of the covariance of the gradient estimate at the optimum (Proposition 1). An expression for this covariance with an arbitrary controller in the loop was given (Proposition 2).

We investigated how to optimize the accuracy of the gradient estimate and the asymptotic convergence rate of the algorithm by a prefilter in the special experiment. An expression for the optimal prefilter was derived for constrained reference input energy (Proposition 3). It was shown how to construct the optimal prefilter from data collected during normal operating conditions of the process.

The effect of the prefilter amounts to decreasing the contribution of the process noise in the special experiment to the error in the gradient estimate at each step. Hence the prefilter will be the more effective the bigger the contribution of the process noise $v_n^2(t)$ is. This is the case if a low energy level of the reference signal for the special experiment is required.

REFERENCES

Blum, Julius R. (1954). Multidimensional stochastic approximation methods. *Annals of Mathematical Statistics* **25**, 737–744.

Hildebrand, R., A. Lecchini, G. Solari and M. Gevers (2002). Convergence analysis and optimal prefiltering in iterative feedback tuning. *Submitted*.

Hjalmarsson, H. (2002). Iterative Feedback Tuning - an overview. *International Journal of Adaptive Control and Signal Processing* **16**(5), 373–395.

Hjalmarsson, H., M. Gevers, S. Gunnarson and O. Lequin (1998). Iterative Feedback Tuning: theory and applications. *IEEE Control Systems* **18**(4), 26–41.

Ljung, L. (1999). *System Identification: Theory for the user*. Prentice Hall.

Nevelson, M. B. and R. Z. Khasminskii (1976). *Stochastic Approximation and Recursive Estimation*. Vol. 47 of *Translations of Mathematical Monographs*. third printing 2000 ed.. American Mathematical Society.

Pintelon, R. and J. Schoukens (2001). *System Identification: a frequency domain approach*. IEEE Press.

Wei, C.Z. (1985). Asymptotic properties of least-squares estimates in stochastic regression models. *The Annals of Statistics* **13**(4), 1498–1508.

Yin, G. (1988). A stopped stochastic approximation algorithm. *Systems and Control Letters* **11**, 107–115.