ROBUST FILTERING AND FEEDFORWARD CONTROL BASED ON PROBABILISTIC DESCRIPTIONS OF MODEL ERRORS

Mikael Sternad and Anders Ahlén *

October 1991, revised June 1992

Abstract

A new approach to robust estimation of signals, prediction of time–series and robust feedforward control is considered. Signal and system parameter deviations are represented as random variables, with known covariances. A robust design is obtained by minimizing the squared estimation error, averaged both with respect to model errors and the noise. A polynomial equations approach, based on averaged spectral factorizations and averaged Diophantine equations, is derived. Mild solvability conditions guarantee the existence of stable optimal filters and feedforward regulators. The robust design turns out to be no more complicated than the design of an ordinary Wiener filter or LQG regulator.

The proposed approach avoids two drawbacks of robust minimax design. First, probabilistic descriptions of model uncertainties may have *soft bounds*. These are more readily obtainable in a noisy environment than the hard bounds required for minimax design. Furthermore, not only the range of uncertainties, but also their likelihood is taken into account; common model deviations will have a greater impact on an estimator design than do very rare "worst cases". The conservativeness is thus reduced.

Subtitle: If model errors are represented by stochastic variables, performance robustness can be optimized by using a polynomial equations approach. Simple closed–form solutions are presented. They minimize quadratic criteria, averaged with respect to model error distributions.

Keywords: Robustness; Optimal estimation; Wiener filtering; Deconvolution; Prediction; Feedforward control.

---

*Systems and Control Group, Uppsala University, P. O. Box 27 S-751 03 Uppsala, Sweden. A shorter version of this paper was presented at the 2nd IFAC Workshop on System Structure and Control, Prague, September 3–5 1992. The work has been partially supported by the Swedish Board for Technical Developement, under contract 8701573 and by the Swedish Institute. It was carried out while Anders Ahlén was on leave at The Department of Electrical Engineering and Computer Science, The University of Newcastle, Australia.

1 INTRODUCTION

In *robust filter* synthesis, the ever present model uncertainty, and the whole range of expected system behaviour, is taken into account. We here propose a novel approach to robust design for open–loop problems such as signal estimation, state estimation and feedforward control. It is based on a stochastic description of model errors, shown to be related to the stochastic embedding concept of Goodwin and Salgado (1989). Compared to alternatives in the literature, the proposed method is simple and straightforward, yet flexible. It constitutes a generalization of the polynomial equations approach, which was pioneered by Kučera (1979, 1981, 1991).

In robust design problems, the true system is only partially known. Robust estimators are based on *nominal models*, $\mathcal{F}_0$, of transfer functions. *Error models* are also required. They specify modelling errors $\Delta\mathcal{F}$ as belonging to a certain class of systems. Nominal and error models will, taken together, be called *extended design models*. Error models are by necessity imprecise; exact modelling of the unmodelled dynamics would be a contradiction in terms. Hard bounds in the frequency domain are one example of specifications, commonly characterized as

$$|\Delta\mathcal{F}(\omega)| \leq L(\omega) \quad . \tag{1.1}$$

For other models of spectral uncertainty, see Kassam and Poor (1985), p 438, or Goodwin and Salgado (1989). Error models may be obtained from off–line experiments or via on–line adaptation. Ways to utilize estimates of model uncertainty are of value in adaptive as well as in fixed filter design.

We will consider mainly estimation problems, with scalar stationary measurements in discrete time. During the last two decades, several schemes for robust estimation have been suggested. The most obvious ad hoc. approach is perhaps to de–tune a filter, by increasing the measurement noise variance used in the design. Optimization of H_∞–criteria has been applied to estimation problems. So far, however, the H_∞–criterion appears to be of limited utility for obtaining robust filters [1]. Instead, almost all systematic schemes have considered *minimax optimization*, usually of the mean square estimation error (MSE). This approach goes back to work by D'Appolito and Hutchinson (1972) and Leondes and Pearson (1972) who dealt with large uncertainties in the noise covariances as well as in the plant dynamics.

Minimax design implies the search for a saddle–point solution. One may search for a least favourable pair of signal and noise spectra, in prespecified uncertainty classes. Under certain conditions, the optimal estimator is a filter designed for that pair. See Kassam and Poor (1985) and also Kassam and Lim (1977), Poor (1980), Moustakides and Kassam (1983,1985) and Vastola and Poor (1984). Uncertainties can be described in a state space framework. See e.g. Martin and Minz (1983),

[1]For methods, see eg. Nagpal and Khargonekar (1991). The main known connection of H_∞–filtering to robustness is the case when both signal and noise models have multiplicative errors, for which only power amplification bounds are known (Grimble and ElSayed 1989). Then, a H_∞–optimal filter can be shown to minimize the worst case mean square error. However, such situations are of rather limited practical relevance. A less conservative design could possibly be obtained from a H_2/H_∞ approach. See Limebeer et.al. (1991).

Wang et. al. (1987) or Haddad and Bernstein (1988). The computational effort involved is considerable. Closed–form solutions do mostly not exist. Furthermore, minimax implies a *worst case design*, with the following drawbacks:

- Even when extremely unlikely, a worst case still determines the estimator completely. There is a considerable risk that this may lead to a very conservative design, with poor performance in the normal range of model errors. (The conservativeness can be reduced by using fictitious tighter model error bounds. But then the whole point of using a worst case approach becomes unclear. Why design for a worst case which is not really worst?)

- An error model with *hard bounds* is required. Otherwise, a worst case is mostly impossible to find. In noisy environments, it may be possible to obtain statistical soft model error bounds, but much more difficult to to obtain reliable hard bounds, such as (1.1) [2]. Parameter errors obtained by system identification have, asymptotically, a Gaussian distribution under very general conditions. See e.g. Ljung (1987). With a distribution with infinite tails, hard bounds cannot really be justified.

There do exist applications where worst case design is well motivated. However, the *mean* performance of estimators and regulators is normally of greater interest. We will consider situations where *one* time–invariant linear filter is to be applied on a number of systems, with different dynamics. A probabilistic description of model errors is then appropriate. We utilize the criterion

$$\bar{E}(E|z(t)|^2) \tag{1.2}$$

where $z(t)$ is the estimation error, E denotes expectation over noise and $\bar{E}$ is the expectation over a (continuous or discrete) distribution of systems. If, for example, there is a single uncertain parameter, attaining three possible values h_1, h_2, h_3 with probabilities p_1, p_2, p_3, respectively, (1.2) becomes

$$p_1 Ez(t)^2\Big|_{h=h_1} + p_2 Ez(t)^2\Big|_{h=h_2} + p_3 Ez(t)^2\Big|_{h=h_3} \quad .$$

The criterion (1.2) takes not only the *effect*, but also the *likelihood* of different modelling errors into account. It has been proposed, and used, by Chung and Bélanger (1976) and by Grimble (1984). Related is also the work by Speyer and Gustafson (1975) who used a conditional expectation. For discrete distributions, Nahi and Knobbe (1976) derived robust filters of very high order. For continuous distributions, errors in time constants had to be assumed small, because these solutions were based on series expansion. However, as will be exemplified in Section 4, robust design is then of limited interest. One might often just as well use a nominal filter for small model errors. A quest for simpler and more useful design

[2] If the noise is bounded, hard bounds may be obtained. See Fogel and Huang (1982). These bounds are extremely wide, compared to statistical standard deviation estimates. Furthermore, if noise bounds are over–estimated, parameter error bounds do not converge to zero with an increasing number of data. If noise bounds are under–estimated, parameter set identification results in an empty set; algorithms such as those of Fogel and Huang will fail.

methods has motivated our present work. Its main contributions are listed below.

• A special type of *stochastic error model* for transfer functions is introduced in Section 3. We will argue that it is flexible enough to describe a wide range of uncertainties, structured and unstructured, large and small, with soft or hard bounds. Besides series expansion, it may be obtained from consideration of time–domain responses, identification by functional series expansion or from a stochastic frequency–domain description. Under mild conditions, only second order moments of stochastic model coefficients need to be known.

• Based on this error model and on the criterion (1.2), a *polynomial equations approach* to H_2 (or Wiener) design for open–loop problems is developed in Section 4 and the Appendix. The design equations are of the same type as for the nominal solution: spectral factorizations and Diophantine equations. Only trivial additional computations are required for obtaining a robust design.

• In Section 2–4, the method is developed in some detail for filtering, prediction and smoothing estimators of scalar signals, which we call *cautious Wiener filters.* The filters balance the importance of uncertainties due to noise and model errors. The peformance difference between robust and nominal designs is also discussed. It can be extremely large when the nominal design is sensitive. It increases for large model errors and decreases with an increased noise level.

• In Section 5, a polynomial approach to *robust state estimation* is outlined.

• Section 6 presents design equations and a design example for robust disturbance measurement *feedforward regulators* and *reference feedforward* filters. The way in which uncertainties in different transfer functions affect the solution is explained.

Remarks on the notation. For any complex polynomial in the backward shift operator q^{-1}, of degree np, $P(q^{-1}) = p_o + p_1 q^{-1} + \ldots + p_{np} q^{-np}$, the *conjugate polynomial* is defined as $P_*(q) \triangleq p_o^* + p_1^* q + \ldots + p_{np}^* q^{np}$. For matrices H, H^* means conjugate transpose. For polynomial matrices $\mathbf{P}(q^{-1})$, $\mathbf{P}_* \triangleq \mathbf{P}^*(q)$. In the frequency domain, z or $e^{i\omega}$ will be substituted for q. For convenience, polynomial arguments are often omitted. We call $P(q^{-1})$ *stable* if all zeros of $P(z^{-1})$ are in $|z| < 1$. Note that whenever P is stable, all zeros of P_* are in $|z| > 1$.

2 A SIGNAL ESTIMATION PROBLEM

2.1 The problem set–up

In the Sections 2–4, a generalized deconvolution problem will be considered, to illustrate the design principles. It includes e.g. ordinary output filtering and prediction of ARMA–processes as special cases. It also includes the design of linear recursive equalizers for digital communications. Measurements are described as

$$y(t) = \mathcal{G}(q^{-1})u(t-k) + w(t) \quad . \tag{2.1}$$

The stable, linear, causal, and possibly uncertain transfer function $\mathcal{G}(q^{-1})$ may e.g. represent a transducer or a transmission channel. The delay k denotes the minimum value of a possibly incompletely known time delay. The input $u(t)$ and the measurement noise $w(t)$ are described by possibly uncertain ARMA models

$$u(t) = \mathcal{F}(q^{-1})e(t) \quad ; \quad w(t) = \mathcal{H}(q^{-1})v(t) \tag{2.2}$$
$$E|e(t)|^2 = 1 \quad ; \quad E|v(t)|^2 \triangleq \rho \ .$$

The white time–series $e(t)$ and $v(t)$ are assumed mutually uncorrelated. They are stationary, with zero mean. All transfer functions are assumed time–invariant. In the next section, they will be partitioned into exactly known nominal models and additive errors $\mathcal{F}(q^{-1}) = \mathcal{F}_0(q^{-1}) + \Delta\mathcal{F}(q^{-1})$ etc, as in Fig. 1. For now, we represent the extended design models as fractions of (partly known) polynomials,

$$\mathcal{F}(q^{-1}) \triangleq \frac{C(q^{-1})}{D(q^{-1})} \quad ; \quad \mathcal{G}(q^{-1}) \triangleq \frac{B(q^{-1})}{A(q^{-1})} \quad ; \quad \mathcal{H}(q^{-1}) \triangleq \frac{M(q^{-1})}{N(q^{-1})} \ . \tag{2.3}$$

The polynomial degrees nc, nd etc, are known to (or determined by) the designer [3]. The polynomials D, A and N in (2.3) are *all assumed stable.* Model uncertainty more or less forces us to restrict attention to sets of stable systems [4]. Signals and polynomial coefficients may be complex–valued; this is the case e.g. in digital communications applications. A stable, linear and time–invariant estimator

$$\hat{f}(t|t+m) = \frac{Q(q^{-1})}{R(q^{-1})}y(t+m) \tag{2.4}$$

of the possibly filtered input

$$f(t) = \frac{S(q^{-1})}{T(q^{-1})}u(t) \tag{2.5}$$

is sought. See Fig. 1. The stable filter S/T is assumed to be specified by the user, and thus to be exactly known [5]. Depending on m, the estimator is a predictor ($m < 0$), a filter ($m = 0$) or a fixed lag smoother ($m > 0$). We minimize (1.2), or

$$\bar{E}(E|z(t)|^2) = \bar{E}E|f(t) - \hat{f}(t|t+m)|^2 \ . \tag{2.6}$$

[3]Note that we are talking about extended design *models*. In practice, they might only be approximations of classes of possibly infinite dimensional and time–varying true systems.

[4]With uncertain unstable poles, the design problem becomes unsolvable, in the open–loop context considered here. Thus, only a single Diophantine equations will be required. (If unstable poles were exactly known, a finite estimation error could, theoretically, be obtained.)

[5]It can be used e.g. to reduce the estimator gain outside a restricted interesting frequency range, see Ahlén and Sternad (1989). For another application, see Carlsson et.al. (1991,1992). Alternatively, one could use a filter in the criterion, $J = \bar{E}E|(S/T)(u(t) - \hat{u}(t|t+m))|^2$. The design equations to be discussed are easily modified to apply to that problem.

Figure 1. A general scalar filtering problem. A filtered version $f(t)$ of the signal $u(t)$ is to be estimated, from noisy measurements $y(t)$. The uncertainties in transfer functions are represented by additive model errors $\Delta\mathcal{F}$, $\Delta\mathcal{G}$ and $\Delta\mathcal{H}$, described by stochastic error models.

2.2 The nominal solution

Assume a stable nominal model $\{C_0/D_0, B_0/A_0, M_0/N_0\}$ of (2.1)–(2.3) to be known. Assume C_0B_0 and ρM_0 to have no common zeros on $|z| = 1$. (When $\rho = 0$, C_0 and B_0 are not allowed to have zeros on $|z| = 1$.) A stable estimator may then be designed to minimize the mean square error (MSE) $E|f(t) - \hat{f}(t|t+m)|^2$, without taking the model uncertainty into account. This estimator is given by

$$\hat{f}_0(t|t+m) = \frac{Q_{10}N_0A_0}{T\beta_0}y(t+m) \quad . \tag{2.7}$$

Here, $\beta_0(q^{-1})$ is the numerator of the nominal innovations model of $y(t)$. It is the stable and monic solution of the polynomial spectral factorization equation

$$r_0\beta_0\beta_{0*} = C_0C_{0*}B_0B_{0*}N_0N_{0*} + \rho M_0M_{0*}A_0A_{0*}D_0D_{0*} \tag{2.8}$$

with r_0 being a scalar. The polynomial $Q_{10}(q^{-1})$, together with a polynomial $L_{0*}(q)$, is the unique solution to the Diophantine equation

$$q^{-m+k}SC_0C_{0*}B_{0*}N_{0*} = r_0\beta_{0*}Q_{10} + qD_0TL_{0*} \quad . \tag{2.9}$$

If the true system equals the nominal model, the minimal MSE is

$$E|z(t)|_0^2 = \frac{1}{2\pi i}\oint \frac{L_0L_{0*}}{r_0\beta_0\beta_{0*}} + \rho\frac{SS_*}{TT_*}\frac{C_0C_{0*}M_0M_{0*}A_0A_{0*}}{r_0\beta_0\beta_{0*}}\frac{dz}{z} \quad . \tag{2.10}$$

This is a Wiener filter, optimized by using polynomial equations. Its derivation is a special case of that in Appendix B. See Ahlén and Sternad (1989) for details.

3 PROBABILISTIC ERROR MODELS

3.1 Properties of the error models

Our goal is now to obtain a simple closed–form solution to the robust estimation problem (2.1)–(2.6). The perhaps most obvious way of parametrizing errors in polynomial coefficients,

$$\frac{C}{D} = \frac{C_0 + \Delta C}{D_0 + \Delta D}$$

is unsuitable in this respect. The reason is that ΔD would not, as opposed to ΔC, appear linearly when expectations with respect to parameters are evaluated in the criterion. This would cause severe difficulties in the derivation of design equations. A crucial simplification is to instead proceed from error models which are linear in all parameters treated as stochastic variables. Additive transfer functions $\Delta\mathcal{F}, \Delta\mathcal{G}, \Delta\mathcal{H}$, with partly unknown stochastic numerators and pre–specified denominators, will be considered. They provide flexibility, without sacrificing the linearity mentioned above. The extended design model (2.1)–(2.3) is specified as

$$\begin{aligned}
\frac{C}{D} &= \frac{C_0}{D_0} + \frac{C_1 \Delta C}{D_1} &= \frac{C_0 D_1 + D_0 C_1 \Delta C}{D_0 D_1} \\
\frac{B}{A} &= \frac{B_0}{A_0} + \frac{B_1 \Delta B}{A_1} &= \frac{B_0 A_1 + A_0 B_1 \Delta B}{A_0 A_1} \\
\frac{M}{N} &= \frac{M_0}{N_0} + \frac{M_1 \Delta M}{N_1} &= \frac{M_0 N_1 + N_0 M_1 \Delta M}{N_0 N_1}
\end{aligned} \tag{3.1}$$

Above, the nominal models C_0/D_0 etc. are assumed known and stable, with degrees nc_0, nd_0 etc. Stable "error denominators" D_1, A_1 and N_1, of degrees nd_1, na_1 and nn_1, as well as numerator factors C_1, B_1, M_1, of degrees nc_1, nb_1 and nm_1, may be specified by the designer or obtained from data. Coefficients of numerator polynomials

$$\Delta P(q^{-1}) = \Delta p_0 + \Delta p_1 q^{-1} + \ldots + \Delta p_{\delta p} q^{-\delta p} \tag{3.2}$$

are stochastic variables. They have zero means and parameter covariances $\bar{E}\Delta p_i \Delta p_j^*$, collected in the covariance matrices $\mathbf{P}_{\Delta P}$. (The zero–mean property defines the nominal model.) These coefficients are constant in time, so they are independent of the time series $e(t)$ and $v(t)$. Except for first and second order moments, their distributions need not be known, since they will not affect the design. Decomposition of numerators $P_1 \Delta P$ into known factors P_1 and stochastic factors ΔP simplifies the uncertainty modelling. In the sequel, we utilize two mild assumptions:

A1. The coefficients of ΔC are independent of those of ΔB.

A2. The matrices $\mathbf{P}_{\Delta C}, \mathbf{P}_{\Delta B}$ and $\mathbf{P}_{\Delta M}$, containing covariances of the coefficients of $\Delta C, \Delta B$ and ΔM, are Hermitian and positive semidefinite.

It is necessary to assure A2 when the covariance matrices are used pragmatically, as "robustness tuning knobs". Design equations may readily be derived for situations with correlations between ΔC and ΔB. Assumption A1 does, however, simplify the solution, and seems reasonable.

Model error covariances may be obtained from identification experiments, or from frequency domain data on system variability. If a fixed filter is to be designed for a large number of systems, the statistics may be obtained from a representative sample of systems. Probabilistic error models remain useful also when statistics is hard to obtain. Those who prefer a Bayesian view could then interprete error distributions as subjective probabilities. Others may just use them pragmatically, as robustness "tuning knobs". The covariances are then altered until satisfactory spectral properties of the filter are obtained. Let us now illustrate the versatility of the structure (3.1), and outline principles for tuning it.

If only the signal to noise ratio is uncertain, we set $\Delta C = \Delta B = \Delta M = 0$, and use a *higher equivalent noise variance.* A model (3.1) with uncertain noise variance, but well–defined noise spectrum is given by $M/N = M_0/N_0 + M_0\Delta m_0/N_0 = (M_0/N_0)(1+\Delta m_0)$, with a scalar stochastic Δm_0. It corresponds to regarding the noise as having variance $\rho(1+\bar{E}|\Delta m_0|^2)$.

Another special case is the use of FIR- or MA–filters (i.e. no denominators):

$$u(t) = (C_0 + \Delta C)e(t) \quad ; \quad y(t) = q^{-k}(B_0 + \Delta B)u(t) + (M_0 + \Delta M)v(t) \quad . \tag{3.3}$$

In (3.3), degrees of stochastic polynomials may be set higher than those of the nominal polynomials, $\delta c > nc_0$, etc. This can be used to guard against under–parametrization. However, for systems with long or infinite impulse responses, error models with denominators are more appropriate than FIR–models.

The structure (3.1) covers multiplicative as well as additive descriptions of model errors. A multiplicative error is obtained with e.g. $B_1 = B_0B_m, A_1 = A_0A_m$, with B_m, A_m arbitrary. It can be useful to extend (3.1) with several additive model error terms. For example, the total error of an identified model can be described by a variance part and a bias part. All results in this paper are readily modified to the case of several additive modelling errors. We will however, for clarity of presentation, confine the discussion to models with a single error term.

3.2 Time–domain determination of error models

Example 1: An uncertain time constant.

Assume that a signal $u(t)$ is known to be generated by a low order system, with a white input $e(t)$. Its impulse response is known to start at 1.0, but then becomes uncertain, because the dominant pole location is uncertain. With 95% probability, the pole is believed to be in the range [0.75...0.95], with values around 0.9 being most likely. From a first order model of this signal

$$u(t) = \mathcal{F}e(t) = \frac{1}{1 - pq^{-1}}e(t) \tag{3.4}$$

we might choose the nominal model $u(t) = \mathcal{F}_0 e(t)$, with

$$\mathcal{F}_0 = \frac{C_0}{D_0} = \frac{1}{1 - 0.9q^{-1}} \quad . \tag{3.5}$$

An additive error model could be derived by a first order series expansion of $\mathcal{F}$ in the uncertain coefficient p

$$\Delta\mathcal{F} = \left(\frac{\partial \mathcal{F}}{\partial p}\bigg|_{p=0.9} \right) \Delta p \ = \ \frac{q^{-1}\Delta p}{(1 - 0.9q^{-1})^2} \quad . \tag{3.6}$$

Here, Δp is seen as a zero mean stochastic variable. However, the fit of the impulse responses of $\mathcal{F} = \mathcal{F}_0 + \Delta\mathcal{F}$ to the expected range of impulse responses of the system (3.4) is not good, see Fig. 2. They tend to spread widely in the beginning and are too narrow for large t. A better approach is to keep the structure of (3.6), but with free user choices of denominator coefficients. Thus, we consider the error model

$$\Delta\mathcal{F} = \frac{C_1 \Delta C}{D_1} = \frac{q^{-1}\Delta c}{1 + d_1^1 q^{-1} + d_2^1 q^{-2}}$$

with Δc being a scalar stochastic variable. Its variance, and the coefficients of D_1, should be tailored to the actual amount of uncertainty. Better centering of the nominal model will also improve the fit. A solution which exactly fits the $\pm 95\%$–bounds can be found in this example. It is given by

$$\mathcal{F} = \frac{1 - 0.85q^{-1}}{(1 - 0.95q^{-1})(1 - 0.75q^{-1})} + \frac{q^{-1}\Delta c}{(1 - 0.95q^{-1})(1 - 0.75q^{-1})} \tag{3.7}$$

with $\bar{E}(\Delta c)^2 = 0.0025$, if Δc is assumed Gaussian. See Fig. 2. The model (3.7) reduces to $\mathcal{F} = 1/(1 - 0.75q^{-1})$ when $\Delta c = 0.10$ (2σ–limit) and to $\mathcal{F} = 1/(1 - 0.95q^{-1})$ when $\Delta c = -0.10$. A systematic algorithm for fitting error models to time–domain data is a topic of current research ■

For a system with measurable inputs, another way of obtaining extended design models of the type (3.1) is from identification experiments based on functional series expansions $\sum_{i=1}^{M} p_i \mathcal{B}_i(q^{-1})$. Here, $\mathcal{B}_i, i = 1 \ldots M$ represent a set of predetermined rational basis functions, such as e.g. discrete Laguerre functions. This model structure is linear in the parameters $\{p_i\}$. It has received increasing interest as a tool for system identification, see Wahlberg (1991) or Goodwin, Ninnes and Poor (1991). If an identification experiment provides parameter estimates $\{\bar{p}_{0i}\}$ and a covariance matrix for zero mean errors $\{\Delta\bar{p}_i\}$, we obtain the extended design model

$$\mathcal{P} \triangleq \sum_{i=1}^{M} \bar{p}_{0i}\mathcal{B}_i + \sum_{i=1}^{M} \Delta\bar{p}_i \mathcal{B}_i = \bar{\mathcal{P}}_0 + \Delta\bar{\mathcal{P}} \quad .$$

If bias–errors are small, the covariance matrix provides acceptable estimates of the modelling errors. Writing $\Delta\bar{\mathcal{P}}$ in common denominator form and use of the covariance matrix for $\{\Delta\bar{p}_i\}$ gives $\bar{E}(\Delta\bar{\mathcal{P}}\Delta\bar{\mathcal{P}}_*)$, needed in the robust design.

Figure 2. Impulse responses of (3.4) in Example 1, for $p = 0.75, 0.9$ and 0.95 (solid). The $\pm 95\%$ probability limits of the extended design model $\mathcal{F} = \mathcal{F}_0 + \Delta\mathcal{F}$ are indicated for the model (3.7) (dashed) as well as for a series–expansion based Gaussian model (3.5), (3.6), with $\bar{E}(\Delta p)^2 = 0.010$ (dotted).

3.3 Frequency domain specifications

Probabilistic error models can be specified in the frequency domain, as parametrized distributions. If the input is measurable, the parameters of such models can be estimated from data, using a maximum likelihood method. See Goodwin et.al. (1991). Let us briefly recapitulate the stochastic embedding concept, proposed by Goodwin and Salgado (1989). An additive transfer function error $\Delta\mathcal{G}(e^{i\omega})$ is there viewed as a realization of a stochastic process in the frequency domain, with zero mean and

$$\bar{E}\{\Delta\mathcal{G}(e^{i\omega_1})\Delta\mathcal{G}_*(e^{i\omega_2})\} \triangleq \Gamma(e^{i\omega_1}, e^{i\omega_2}) \geq 0 \quad .$$

For stationary processes, $\Gamma(e^{i\omega_1}, e^{i\omega_2}) = \Gamma_s(e^{i(\omega_1 - \omega_2)})$. The shape of Γ_s is a measure of the assumed frequency domain smoothness of particular realizations of the model error. The variance $(\omega_1 - \omega_2 = 0)$ is a scale factor for the uncertainty.

The frequency domain stochastic process $\Delta\mathcal{G}(e^{i\omega})$ corresponds to a time–domain filter with stochastic, zero mean, impulse response coefficients

$$\Delta\mathcal{G}(q^{-1}) = \sum_{j=0}^{\infty} g_j q^{-j} \quad ; \quad \bar{E}(g_j, g_\ell) = \gamma(j, \ell) \quad . \tag{3.8}$$

Here, $\gamma(j, \ell)$ can be calculated from the inverse 2D discrete Fourier transform of $\Gamma(e^{i\omega_1}, e^{i\omega_2})$. For *stationary* stochastic processes in the frequency domain, the corresponding time–domain stochastic process will be *nonstationary and white*, with

$$\bar{E}(g_j, g_\ell) = \gamma_j \delta_{j,\ell} \quad . \tag{3.9}$$

For example, consider a frequency domain stochastic process $\mathcal{H}(e^{i\omega})$, with a zero mean Gaussian distribution and covariance function

$$\bar{E}\{\mathcal{H}(e^{i\omega_1})\mathcal{H}_*(e^{i\omega_2})\} = \frac{\alpha e^{i(\omega_1-\omega_2)}}{e^{i(\omega_1-\omega_2)} - \lambda} \quad . \tag{3.10}$$

It corresponds to the nonstationary time domain model (3.8), (3.9), with Gaussian distribution of independent parameters, with variances $\gamma_j = \alpha\lambda^j$. See Goodwin et.al. (1991). By truncating at some $j = M$ for which λ^M is small, we obtain $\mathcal{H}(q^{-1}) \approx h_0 + \ldots + h_M q^{-M}$, with $\bar{E}(h_j)^2 = \alpha\lambda^j$.

A priori information may be available about the frequency domain distribution of the unmodelled dynamics. It can be incorporated by using

$$\Delta\mathcal{G}(q^{-1}) = \mathcal{M}(q^{-1})\mathcal{N}'_{\Delta}(q^{-1}) \quad . \tag{3.11}$$

Here, $\mathcal{M}(q^{-1})$ is a known frequency shaping filter, and $\mathcal{N}'_{\Delta}(q^{-1})$ is a stationary process in the frequency domain, with covariance function $\Gamma'_s(e^{i(\omega_1-\omega_2)})$.

Example 2. A frequency–shaped error model.

Suppose that the *mean magnitude* of the model error in the frequency domain can be described by a magnitude response of a first order filter. Then, we may use (3.11), with

$$\mathcal{M}(q^{-1}) = \frac{1+\eta q^{-1}}{1+a_{11}q^{-1}} \quad . \tag{3.12}$$

Also, assume that the parameter λ in (3.10) can be tuned to give a reasonable description of the *degree of smoothness* (in the frequency domain) of the most probable model errors. The process in (3.10) can then be used to represent the stationary part $\Gamma'_s(e^{i(\omega_1-\omega_2)})$ of the frequency–domain process. Truncation of its corresponding time–domain impulse response gives a model (3.11), with the structure (3.1)

$$\Delta\mathcal{G}(q^{-1}) = \frac{1+\eta q^{-1}}{1+a_{11}q^{-1}}(h_0 + h_1 q^{-1} + \ldots + h_M q^{-M}) = \frac{B_1(q^{-1})\Delta B(q^{-1})}{A_1(q^{-1})} \quad . \tag{3.13}$$

The covariance matrix of $\{h_j\}$ is diag$(\alpha\lambda^j)$. Note that this model is fully determined by only five parameters, $\alpha, \lambda, a_{11}, \eta$ and the truncation length M ■

4 DESIGN OF ROBUST FILTERS

We proceed from the model (2.1), (2.2), (2.3) and (3.1). The coefficients of ΔC, ΔB and ΔM are random variables. They cause the polynomials C, B and M to be random variables as well. The expectation operator $\bar{E}$ in (2.6) will provide an ensemble average, yielding one estimator to be used on all systems.

4.1 The averaged spectral factorization

An averaged spectral factor $\beta(q^{-1})$ is defined as the numerator polynomial of an averaged innovations model of the utilized measurement. In the present case, it is given by the stable and monic solution to

$$r\beta\beta_* \triangleq \bar{E}\{CC_*BB_*NN_* + \rho MM_*AA_*DD_*\} \tag{4.1}$$

with scalar r. Define double–sided polynomials

$$\tilde{C}\tilde{C}_* \triangleq \bar{E}(CC_*) \quad ; \quad \tilde{B}\tilde{B}_* \triangleq \bar{E}(BB_*) \quad ; \quad \tilde{M}\tilde{M}_* \triangleq \bar{E}(MM_*) \quad .$$

Then, use of (3.1) gives

$$\begin{aligned}
\tilde{C}\tilde{C}_* &= C_0C_{0*}D_1D_{1*} + D_0D_{0*}C_1C_{1*}\bar{E}(\Delta C\Delta C_*) \\
\tilde{B}\tilde{B}_* &= B_0B_{0*}A_1A_{1*} + A_0A_{0*}B_1B_{1*}\bar{E}(\Delta B\Delta B_*) \\
\tilde{M}\tilde{M}_* &= M_0M_{0*}N_1N_{1*} + N_0N_{0*}M_1M_{1*}\bar{E}(\Delta M\Delta M_*) \quad .
\end{aligned} \tag{4.2}$$

We can now simplify (4.1).

Lemma 1.

Let assumption A1 hold. Then, (4.1) can be expressed as

$$r\beta\beta_* = \tilde{C}\tilde{C}_*\tilde{B}\tilde{B}_*NN_* + \rho\tilde{M}\tilde{M}_*AA_*DD_* \tag{4.3}$$

■

Proof: The coefficients of a polynomial ΔP are zero mean stochastic variables. Therefore, the coefficients of $\Delta P\Delta P_*$ will also be stochastic variables, having expected values given by (4.5) below. The coefficients of (ΔB, ΔC), are independent, and so are the coefficients of $\Delta B\Delta B_*$, $\Delta C\Delta C_*$. Using independence for complex parameters (see eg Loève (1963) Chapter V), the right–hand side of (4.1) becomes

$$\bar{E}(CC_*)\bar{E}(BB_*)NN_* + \rho\bar{E}(MM_*)AA_*DD_*$$

■

The averaged factors in (4.2) can be evaluated as follows. For a stochastic error model numerator $\Delta P(q^{-1})$, as in (3.2), let the Hermitian parameter covariance matrix be

$$\mathbf{P}_{\Delta P} = \begin{bmatrix} \bar{E}|\Delta p_0|^2 & \dots & \bar{E}(\Delta p_0\Delta p^*_{\delta p}) \\ \vdots & \ddots & \vdots \\ \bar{E}(\Delta p_{\delta p}\Delta p^*_0) & \dots & \bar{E}|\Delta p_{\delta p}|^2 \end{bmatrix} \tag{4.4}$$

Denote the sum of the diagonal elements g_0, the sum of the elements in the i'th super–diagonal g_i, the sum of elements in the i'th subdiagonal g_{-i}. Note that $g_{-i} = g_i^*$. Then it becomes evident, by direct multiplication of $\Delta P(q^{-1})\Delta P_*(q)$, and taking expectations, that

$$\bar{E}(\Delta P \Delta P_*) = g^*_{dp} q^{-dp} + \ldots + g^*_1 q^{-1} + g_0 + g_1 q + \ldots + g_{dp} q^{dp} \quad . \tag{4.5}$$

Thus, the averaged factors in (4.2) are readily obtained. Above, $dp \leq \delta p$, with $dp = 0$ if the coefficients are uncorrelated. For example, the coefficients h_i are uncorrelated in the model (3.13). The resulting polynomial $\bar{E}(\Delta B \Delta B_*)$ would then simply be a scalar, regardless of the degree of ΔB.

In (4.2), $\tilde{C}\tilde{C}_*$ will contain powers up to $q^{\pm n\tilde{c}}$, where $n\tilde{c} = \max\{nc_0 + nd_1, nd_0 + nc_1 + dc\}$, with analogous expressions for $n\tilde{b}, n\tilde{m}$. Since $N = N_0 N_1$ etc, the averaged spectral factor in (4.3) has degree

$$n\beta = \max\{n\tilde{c} + n\tilde{b} + nn_0 + nn_1, n\tilde{m} + na_0 + na_1 + nd_0 + nd_1\} \quad .$$

The factorization (4.3) is solvable with respect to a unique stable $\beta(z^{-1})$ if and only if its right–hand side is positive on $|z| = 1$. Introduce the assumptions

A3. C_0, $C_1\bar{E}(\Delta C \Delta C_*)$, ρM_0 and $\rho M_1 \bar{E}(\Delta M \Delta M_*)$ have no common zeros on $|z| = 1$

A4. B_0, $B_1\bar{E}(\Delta B \Delta B_*)$, ρM_0 and $\rho M_1 \bar{E}(\Delta M \Delta M_*)$ have no common zeros on $|z| = 1$.

Lemma 2.

Let D, A and N be stable and A2 hold. Then, a unique stable spectral factor β, satisfying (4.3), exists, if and only if both of A3 and A4 are true ■

Proof: See Appendix A.

The conditions A3 and A4 are mild. They will almost always be fulfilled, even if C_0, B_0 and M_0 have zeros on the unit circle. In fact, the conditions are more relaxed than for the nominal case, due to the presence of averaged factors $\bar{E}(\cdot)$.

4.2 The cautious Wiener filter

Theorem 1.

Assume an extended design model (2.1), (2.2), (2.3) and (3.1) to be given, with known covariances of the stochastic polynomial coefficients. Assume A1–A4 to hold. An estimator of $f(t)$ then minimizes (2.6), among all linear time–invariant estimators based on $y(t+m)$, if and only if it has the same coprime factors as

$$\hat{f}(t|t+m) = \frac{Q_1 N_0 N_1 A_0 A_1}{T\beta} y(t+m) \quad . \tag{4.6}$$

Here, $\beta(q^{-1})$ is obtained from (4.3), while $Q_1(q^{-1})$, together with $L_*(q)$, is the unique solution to

$$q^{-m+k} S\tilde{C}\tilde{C}_* B_{0*} A_{1*} N_{0*} N_{1*} = r\beta_* Q_1 + q D_0 D_1 T L_* \tag{4.7}$$

with generic [6] polynomial degrees

$$\begin{aligned} nQ_1 &= \max(ns + n\tilde{c} - k + m \ , \ nd_0 + nd_1 + nt - 1) \\ nL &= \max(n\tilde{c} + nb_0 + na_1 + nn_0 + nn_1 + k - m \ , \ n\beta) - 1 \ . \end{aligned} \tag{4.8}$$

For the ensemble of systems, the minimal criterion value becomes

$$\bar{E}E|z(t)|^2_{\min} = \frac{1}{2\pi i}\oint \frac{LL_*}{r\beta\beta_*} + \rho\frac{SS_*\tilde{C}\tilde{C}_*\tilde{M}\tilde{M}_*AA_*}{TT_*r\beta\beta_*} + \frac{SS_*\tilde{C}\tilde{C}_*\tilde{C}\tilde{C}_*\bar{E}(\Delta\mathcal{G}\Delta\mathcal{G}_*)AA_*NN_*}{TT_*DD_*r\beta\beta_*}\frac{dz}{z} \tag{4.9}$$

■

Proof: See Appendix B.

Remarks. For minimizing (2.6), the design equations are (4.3), (4.5) and (4.7). The only new type of computation, as compared to a nominal solution, is trivial: summation of covariance matrix elements, diagonalwise. There will, of course, exist a filter giving superior MSE for any particular system, namely the (unknown) Wiener filter for that system. There may also exist superior nonlinear filters.

Note that N_1 and A_1 affect the filter (4.6) directly. If $1/N_1$ or $1/A_1$ in the error models have resonance peaks, indicating large uncertainty, the filter (4.6) will have low gain at those frequencies. Stable common factors might exist in (4.6). (Thus, the formulation about coprime factors.) Equation (4.7) will have a unique solution, with degrees (4.8). Note that $\beta_*(z)$ (unstable) and $D_0(z^{-1})D_1(z^{-1})T(z^{-1})$ (stable) have no common factors. *Almost* common factors close to $|z| = 1$ tend to make Diophantine equations numerically sensitive. The risk for this is less in (4.7) than in (2.9), because zeros of β are in general more distant from unit circle than zeros of β_0. The filter (4.7) tends to have lower resonance peaks than (2.7).

It is easy to show that if the block S/T would contain an uncertainty, $S/T = S_0/T_0 + S_1\Delta S/T_1$, this would not affect the resulting filter. In (4.6) and (4.7), the nominal polynomials S_0 and T_0 would be substituted for S and T.

If the variance of broad–band measurement noise is increased, the gains of both the nominal and the robust filters decrease. If the noise level is high, performance differences between nominal and robust solutions tend to be small.

It is shown in Appendix B that if no uncertainty is assumed, (4.3),(4.6)–(4.9) reduce to the nominal solution. Also, with uncertainty, the structure of the design equations remains very similar to the nominal case. In fact, if the transducer B/A is known, i.e. $\Delta\mathcal{G} = 0$, we can substitute $\tilde{C}\tilde{C}_*$, $\tilde{M}\tilde{M}_*$, D, N for C_0C_{0*}, M_0M_{0*}, D_0, N_0 in (2.7)–(2.10), which then become identical to (4.3),(4.6)–(4.9). Robust and nominal filter design are then merged together into *one pair* of design equations, namely

[6] For special polynomials, the solution may have lower degrees.

(2.8) and (2.9). (The reasoning in Grimble (1984) is based on similar substitutions.) This analogy cannot be generalized to uncertain transducers B/A, because of the presence of $\tilde{B}\tilde{B}_*$ in (4.3) and of $B_{0*}A_{1*}$ in (4.7). The criterion expressions (2.10) and (4.9) will also differ for $\Delta\mathcal{G} \neq 0$, due to the additional third term in (4.9).

The integrand of (4.9) consists of three terms. Term 1 represents the effect of finite smoothing lag m. As in Carlsson et.al. (1991), it can be shown that $L_*(q) \to 0$ when $m \to \infty$. The second term mainly represents the effect of noise. It vanishes for $\rho = 0$. Finally, the third term represents degradation caused by errors $\Delta\mathcal{G}$ in the transducer model. It vanishes only when $\Delta\mathcal{G} = 0$. Note that if both ρ and $\Delta\mathcal{G}$ are zero, the impact of $\Delta\mathcal{F}$ diminishes as $m \to \infty$, because $LL_* \to 0$.

In situations with little noise and sufficiently large smoothing lag m, term 3 in (4.9) will dominate the error. This is not surprising; a deconvolution smoother then essentially inverts $\mathcal{G}$. This operation is sensitive to model errors there. Let $S = T = 1$, $\rho = 0$ and $m \to \infty$. Term 3 of (4.9) will then be $\tilde{C}\tilde{C}\bar{E}(\Delta\mathcal{G}\Delta\mathcal{G}_*)/\bar{E}(\mathcal{G}\mathcal{G}_*)DD_*$. It is proportional to the relative error in the spectrum of $\mathcal{G}$.

4.3 Analytical expressions for performance evaluation

Theorem 2.

Let a nominal estimator $Q_0/R_0 = Q_{10}N_0A_0/T\beta_0$ be designed by (2.7)-(2.9). Applying it, instead of (4.6), on an ensemble of systems results in an increase, compared to (4.9), of the mean MSE $\bar{E}E|z(t)|^2$. The increase is given by

$$\bar{E}E|z(t)|_0^2 - \bar{E}E|z(t)|_{\min}^2 = \frac{r}{2\pi i}\oint \left|\frac{\beta}{DAN}\right|^2 \left|\frac{Q_0}{R_0} - \frac{Q}{R}\right|^2 \frac{dz}{z} = \frac{r}{2\pi i}\oint \left|\frac{Q_{10}\beta - Q_1N_1A_1\beta_0}{D_0D_1A_1N_1T\beta_0}\right|^2 \frac{dz}{z} \tag{4.10}$$

where r, β is defined by (4.1)-(4.3) and Q/R is the optimal robust filter (4.6).

Proof: To obtain (4.10), the nominal filter Q_0/R_0 is expressed as $Q/R + (Q_0/R_0 - Q/R)$. The optimality of Q/R implies that any modification of it gives an orhtogonal contribution to the criterion. This, and the use of (4.1), gives the first equality of (4.10). Mixed terms vanish, due to the orthogonality. The second equality follows from (2.7) and (4.6) ■

Theorem 3.

Let a robust estimator Q/R be designed by (4.3)–(4.7). When applying it on the nominal system, the increased MSE, compared to the minimum (2.10), is

$$E|z(t)|^2 - E|z(t)|_0^2 = \frac{r_0}{2\pi i}\oint \left|\frac{\beta_0}{D_0A_0N_0}\right|^2 \left|\frac{Q}{R} - \frac{Q_0}{R_0}\right|^2 \frac{dz}{z} = \frac{r_0}{2\pi i}\oint \left|\frac{Q_1\beta_0A_1N_1 - Q_{10}\beta}{D_0T\beta}\right|^2 \frac{dz}{z} \tag{4.11}$$

Proof: Analogous to Theorem 2, by expressing Q/R as $Q_0/R_0 + (Q/R - Q_0/R_0)$ ■

Remarks. The middle expression in (4.10) can be used for arbitrary linear estimators Q_0/R_0, for example minimax–designs. We thus do not have to evaluate the mean performance of alternative designs by Monte–Carlo simulation. The averaged innovations model β/DAN in (4.10), and the nominal innovations model $\beta_0/D_0A_0N_0$ in (4.11) can be seen as weighting functions. In frequency regions where their magnitude is large, differences between the two estimators will have a large impact on the performance.

4.4 A numerical example

Example 3: Calculation and comparison of robust and nominal designs.

Consider an extended design model (3.3), with a FIR nominal model given by

$$C_0(q^{-1}) = 1 - 0.95q^{-1} \quad ; \quad B_0(q^{-1}) = 0.5 - 0.4q^{-1} \quad ; \quad M_0(q^{-1}) = 1 - 0.8q^{-1}$$
$$k = 1 \quad , \quad \rho = 0.001 \quad . \tag{4.12}$$

The covariance matrices of $\Delta C, \Delta B$ and ΔM are

$$\mathbf{P}_{\Delta C} = \begin{bmatrix} 0 & 0 \\ 0 & 0.040 \end{bmatrix} \; ; \; \mathbf{P}_{\Delta B} = \begin{bmatrix} 0.0025 & 0 \\ 0 & 0.0225 \end{bmatrix} \; ; \; \mathbf{P}_{\Delta M} = \begin{bmatrix} 0 & 0 \\ 0 & 0.010 \end{bmatrix}$$

The assumed standard deviations are thus 0.20 for c_1, 0.05 for b_0, 0.15 for b_1 and 0.10 for m_1. We would like to obtain a robust filter $\hat{u}(t|t) = (Q/R)y(t)$. This estimator should, essentially, perform a one step prediction to obtain $u(t)$, since the transducer $q^{-k}B(q^{-1})$ has a one step delay. Using (4.5), we obtain

$$\bar{E}(\Delta C \Delta C_*) = 0.040 \quad , \quad \bar{E}(\Delta B \Delta B_*) = 0.0250 \quad , \quad \bar{E}(\Delta M \Delta M_*) = 0.010 \quad . \tag{4.13}$$

Inserting (4.13) into (4.3) gives

$$r\beta\beta_* = (1.9425 - 0.95(q + q^{-1}))(0.435 - 0.2(q + q^{-1})) \; + \; 0.001(1.65 - 0.8(q + q^{-1})).$$

By solving for the stable monic $\beta(q^{-1})$ and r, we obtain

$$\beta(q^{-1}) = 1 - 1.4659q^{-1} + 0.5315q^{-2} \quad , \quad r = 0.3575 \quad .$$

Proceed to calculate the filter polynomial $Q_1(q^{-1})$ from (4.7), in which $S = A_1 = N_0 = N_1 = D_0 = D_1 = T = 1$. With degrees $nQ_1 = 0, nL = 2$, we obtain, with $Q_1(q^{-1}) = c$,

$$q(1.9425 - 0.95(q + q^{-1}))(0.5 - 0.4q) \; =$$

$$0.3575(1 - 1.4659q + 0.5315q^2)c + q(\ell_0 + \ell_1 q + \ell_2 q^2 \;) \quad .$$

Equating for different powers of q gives

$$Q_1(q^{-1}) = -1.3288 \quad ; \quad L_*(q) = 0.6549 - 0.9995q + 0.380q^2 \ .$$

The robust estimator (4.6) is $\hat{u}(t|t) = (Q_1/\beta)y(t)$, or

$$\hat{u}(t|t) = \frac{-1.3288}{1 - 1.4659q^{-1} + 0.5315q^{-2}} y(t) \ . \tag{4.14}$$

It has poles in $z = 0.8086$ and $z = 0.6573$. The nominal filter polynomials are

$$Q_{10}(q^{-1}) = -1.8413 \quad ; \quad \beta_0(q^{-1}) = 1 - 1.7206q^{-1} + 0.7365q^{-2} \tag{4.15}$$

with estimator poles in $z = 0.9206$ and $z = 0.80$. The robust estimator has decreased the gain and moved the poles inward. It has become more cautious.

The performance of nominal and robust estimators is exemplified in Fig. 3. The MSE is much lower for the robust filter, for most parameter values. For a wide range of parameter variations, the performance of the robust estimator is close to that which could be obtained if the true parameters were known.

In Fig. 4, the mean performance (2.6) was calculated, for nominal and robust estimators, as a function of the standard deviations of some of the parameters. As expected, the mean performance is much better for the robust filter for large parameter deviations. In situations with small model errors, we might just as well use the nominal estimator. This is also true when the noise level is high; when $Ev(t)^2 \geq 0.1$, the difference between robust and nominal filters is very small.

Figure 3. Variation of one parameter in Example 3, while the others are nominal. The MSE $Ez(t)^2$, as a function of the zero $-c_1$ of the signal model (left) and of $-b_1$ in the transducer (right). Performance of the nominal filter (2.7) (solid), of the robust filter (4.6) (dashed) and the lower bound (dotted), achievable with the knowledge of the true parameter value.

Figure 4. Mean performance $\bar{E}Ez(t)^2$ in Example 3, when standard deviations of c_1 (left) and either b_0 or b_1 (right) are varied, while other coefficients are nominal. Nominal filters (solid) are compared to robust filters, designed for the corresponding parameter variance (dashed).

Figure 5. Error variance $Ez(t)^2$ for nominal (solid) and robust (dashed) filters, as in Figure 3. We also illustrate a minimax–design assuming rectangular distributions (dashed–dotted), a minimax–design based on 5–point distributions (crosses) and also to a nominal filter, designed using $\rho = 0.1$ (dotted).

In Fig. 5, we compare with minimax–designs, assuming two distributions, both with variances as above. 1) rectangular (hard bound $\sqrt{3}*$standard deviations) and 2) 5–point distributions of each parameter (bound ± 0.5 for $c_1, b_1, m_1, \pm 0.2$ for b_0). Fig. 5 clearly reveals the weaknesses of minimax designs: for wide model error distributions, with unlikely remote values, the filter performance for the nominal case deteriorates. More reasonable filters are obtained when the most remote value is close to the standard deviation of the distribution. However, even the assumption of a rectangular distribution results in a more conservative design than our cautious Wiener filter. Additional details and comments are found in Appendix D.

The simplest way of robustifying an estimator is to just increase the noise variance ρ used in a nominal design. The performance of this technique is also illustrated in Fig. 5. The result around the nominal case is not as good as for our robust filter. The difference can be expected to be even larger in more high–order examples. The use of just one single robustification parameter mostly provides insufficient degrees of freedom. It can only vary β along a single root locus trajectory ■

4.5 Averaging over systems versus averaging over models

Conceptually, two types of "averaged MSE" could be defined:

Case 1. One filter is to be used on a class of *different systems*. The averaged performance $\bar{E}_{syst}(\cdot)$, with respect to the systems, is then to be minimized.

Case 2. A set of filters, based on *different design models*, are all to be used on one (unknown) true system. The average performance $\bar{E}_{mod}(\cdot)$, with respect to these models, is then of interest.

Our criterion (1.2) has been formulated with Case 1 in mind. Situations corresponding to Case 2 are, however, also encountered. An example is equalizer design for the pan–european digital mobile radio standard GSM. There, channel models of FIR type are estimated during short training sequences, in which the input $u(t)$ is known. Estimated models are then used to reconstruct unknown symbols during other times. If neither transmitter nor receiver moves, the channel will be time–invariant during many training events, but the estimated channels will differ due to noise. Robustness of equalizer performance according to the criterion of Case 2 is then a reasonable design objective.

An attempt to design filters directly by minimizing $\bar{E}_{mod}(E|z(t)|^2)$ does lead to intractable nonlinearities. We could, as a substitute, use our cautious Wiener filter (designed for Case 1) also in Case 2. For any nominal model, a filter would then be designed by minimizing $\bar{E}_{syst}(E|z(t)|^2)$ with respect to a (non-existing) set of systems around that particular model. Use of this procedure for all models would *mostly*, but *not always*, improve $\bar{E}_{mod}(E|z(t)|^2)$, compared to the use of nominal design. Exceptions exist, as illustrated by the following counter–example, suggested by Ari Kangas:

Example 3, cont'd.

The true system is assumed given by (4.12). 100 different design models were generated, by adding random numbers $\Delta c_1 \in \mathcal{N}(0, 0.040)$, $\Delta b_0 \in \mathcal{N}(0, 0.0025)$, $\Delta b_1 \in \mathcal{N}(0, 0.0225)$ and $\Delta m_1 \in \mathcal{N}(0, 0.010)$ to the corresponding system coefficients. For each of the 100 models, robust and nominal filters were designed, as in Example 3. On average, nominal designs outperformed robust filtering! We obtained $\bar{E}_{mod}(E|z|^2) = 1.22$ for the robust filters and 1.11 for the nominal ones ■

Fortunately, this kind of situation turns out to be a rather rare exception. The following Monte–Carlo simulation indicates that the cautious Wiener filter, on average, improves the performace similarly in Case 1 and Case 2.

Example 4. Performance for sets of randomly chosen models and systems.

We generated 500 different systems, with the structure

$$
\begin{aligned}
y(t) &= (b_0 + b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3} + b_4 q^{-4}) u(t-1) + (1 - 0.8 q^{-1}) v(t) \\
u(t) &= \frac{1}{1 - 1.4 q^{-1} + 0.65 q^{-2}} e(t) \;\; ; \;\; Ee(t)^2 = 1 \; ; \; Ev(t)^2 = 0.001
\end{aligned}
\tag{4.16}
$$

by using independent random numbers $b_i \in \mathcal{N}(0,1)$. The performance of filter estimators ($m = 0$) of $u(t)$ was investigated. In robust design, $\mathbf{P}_{\Delta B} = \text{diag}(0.04)$, $\mathbf{P}_{\Delta M} = 0$ and $\mathbf{P}_{\Delta C} = 0$ were used.

First, in an *Experiment 1* illustrating Case 1, the 500 systems (4.16) represented different nominal design models. Around each of these 500 models, ensembles of true systems, with B–polynomials $B_0 + \Delta B$; $\sqrt{\bar{E}(\Delta b_i)^2} = 0.2$, were assumed to exist. Using (4.9) and (4.10), the mean performances, over each ensemble of true systems, were calculated for robust and nominal filters. We also calculated the minimal MSE's, achievable for known systems. The minimal MSE's varied between ≈ 1.0 and ≈ 6.0, depending on how well a particular transmission channel $B(q^{-1})$ could be inverted. Their average, over 500 cases, was 3.33. The average mean MSE's for robust and nominal design were

$$
\frac{1}{500} \sum_{i=1}^{500} \bar{E}_{syst} E|z(t)|_i^2 \Big|_{rob} = 3.72 \;\; ; \;\; \frac{1}{500} \sum_{i=1}^{500} \bar{E}_{syst} E|z(t)|_i^2 \Big|_{nom} = 6.25
$$

The improvement resulting from a robust design was rather small in ≈ 400 of the 500 cases, but could be very large in the other ones. This is illustrated by the left-hand histogram of Fig. 6. Models for which a nominal design is sensitive have zeros of $B_0(z^{-1})$ close to the unit circle in the low–frequency range, where $u(t)$ has significant energy. See Fig. 7.

Next, an *Experiment 2* was performed to investigate Case 2. The 500 systems (4.16) here represented different true systems. For *each* of them, 100 design models were generated, by adding independent random coefficients $\Delta b_i \in \mathcal{N}(0, 0.04)$ to the B–coefficients of (4.16). Based on these models, robust and nominal filters were designed. For each system, estimates of their average performance over the model distribution, $\bar{E}_{mod}(E|z(t)|^2)$, were calculated [7]. The total averages, over the 500 different systems, were

$$
\frac{1}{500} \sum_{i=1}^{500} \bar{E}_{mod} E|z(t)|_i^2 \Big|_{rob} = 3.76 \;\; ; \;\; \frac{1}{500} \sum_{i=1}^{500} \bar{E}_{mod} E|z(t)|_i^2 \Big|_{nom} = 6.01
$$

[7] For each of the 100 design models, a robust and a nominal filter was obtained. Calculation of the MSE for each of those filters, and averaging over the results based on 100 models, gave estimates of $\bar{E}_{mod}E|z(t)|^2 \,|_{rob}$ and $\bar{E}_{mod}E|z(t)|^2 \,|_{nom}$, respectively.

$\bar{E}_{syst}(MSE)_{nom} - \bar{E}_{syst}(MSE)_{rob}$ $\bar{E}_{mod}(MSE)_{nom} - \bar{E}_{mod}(MSE)_{rob}$

Figure 6. Distribution of the reduction in mean MSE with robust, as compared to nominal, design in Example 4. In Experiment 1 (left), averages are taken over ensembles of systems, for each of 500 different nominal models of type (4.16). In Experiment 2 (right), we average over 100 models, for each of 500 true systems.

Figure 7. Magnitude spectra in one particular case in Example 4, experiment 1. Upper solid curve: input u. Lower solid curve: channel $B_0(e^{i\omega})$. Upper dashed: nominal filter. Lower dashed: robust filter. Dash-dotted: frequency distribution of the mean RMS performance loss (square root of the integrand in (4.10)) with nominal filtering, if the system is uncertain.

There were 18 cases out of 500 in which nominal designs, on average over 100 models, outperformed the robust ones. The performance difference was then small, in the range $0.01 \leq \bar{E}_{mod}(MSE)_{rob} - \bar{E}_{mod}(MSE)_{nom} \leq 0.22$. Typically, a robust design outperformed a nominal one. As in Experiment 1, the difference was mostly rather small, but was very large for a significant minority of systems. See Fig. 6.

A detailed investigation of the experiments revealed a very reliable behaviour of the robust filters, with mean MSE's seldomly more than 20% above MSE's obtainable with a known system. When a nominal design was sensitive, robust design eliminated the sensitivity, resulting in a greatly improved mean performance. No evidence of bad behaviour of robust filters was detected. No case of a large superiority of nominal design, as compared to robust design, were found in Experiment 2. To conclude, in situations described by either of Case 1 or Case 2, much could be gained, and little was lost, by using the cautious Wiener filter ■

5 ROBUST STATE ESTIMATION

Using the principles discussed in the previous sections, a polynomial equations approach to robust state estimation can be developed. Consider a stable stochastic model in state–space form. It has n states, a scalar output $s(t)$ and an ℓ–vector $f(t)$, to be estimated:

$$x(t+1) = Fx(t) + e_v(t) \tag{5.1}$$

$$s(t) = H_s x(t) \quad ; \quad f(t) = H_f x(t) \quad . \tag{5.2}$$

Here, $e_v(t)$ is white, with covariance matrix $R_v \geq 0$. If $\ell = n$ and $H_f = I_n$, full state estimation is desired. Measurements $y(t)$ of $s(t)$ are corrupted by coloured noise

$$y(t) = s(t) + \frac{M(q^{-1})}{N(q^{-1})} v(t) \tag{5.3}$$

with $E|v(t)|^2 \triangleq \rho$. Now, express (5.1) as

$$x(t) = \frac{1}{D(q^{-1})} \mathbf{C}(q^{-1}) e_v(t) \tag{5.4}$$

where $D(q^{-1}) = \det(I - q^{-1}F)$ is the characteristic polynomial and $\mathbf{C}(q^{-1}) = \mathrm{adj}(I - q^{-1}F)q^{-1}$ is a $n|n$ polynomial matrix. The $\ell|n$–matrix H_f is known. Other parts of the model (5.2)–(5.4) may be uncertain. We will consider an extended design model with the following structure.

$$H_s = H_0 + \Delta H \tag{5.5}$$

$$\frac{M(q^{-1})}{N(q^{-1})} = \frac{M_0(q^{-1})}{N_0(q^{-1})} + \frac{M_1(q^{-1}) \Delta M(q^{-1})}{N_1(q^{-1})} \tag{5.6}$$

$$\frac{1}{D(q^{-1})} \mathbf{C}(q^{-1}) = \frac{1}{D_0(q^{-1})} \mathbf{C_0}(q^{-1}) + \frac{1}{D_1(q^{-1})} \mathbf{C_1}(q^{-1}) \mathbf{\Delta C}(q^{-1}) \quad . \tag{5.7}$$

In (5.5), ΔH is a $1|n$–vector containing zero mean stochastic elements, with $\bar{E}(\Delta H^* \Delta H) \triangleq \mathbf{P}_H \geq 0$. The noise model (5.6), with covariance matrix $\mathbf{P}_{\Delta M} \geq 0$ and N_0, N_1 stable, is of the type introduced in Section 3. In (5.7), D_0 and D_1 are stable. The elements of $\mathbf{\Delta C}(q^{-1})$ are zero mean stochastic polynomials, defined as in (3.2). Different elements of $\mathbf{\Delta C}$ may be correlated, but $\mathbf{\Delta C}$ and ΔH are mutually independent. If elements of F in (5.1) are uncertain, series expansion gives a model of type (5.7). However, as was discussed in Example 1, a better description may be obtained for large uncertainties if $D_1(q^{-1}), \mathbf{C_1}(q^{-1})$ and $\mathbf{\Delta C}(q^{-1})$ are tuned to the uncertainty directly. Introduce the polynomial matrix

$$\tilde{\mathbf{P}} \triangleq \bar{E}(\mathbf{C}R_e\mathbf{C}_*) = D_1D_{1*}\mathbf{C_0}R_e\mathbf{C_{0*}} + D_0D_{0*}\mathbf{C_1}\bar{E}(\mathbf{\Delta C}R_e\mathbf{\Delta C}_*)\mathbf{C_{1*}} \ . \tag{5.8}$$

Elements of $\bar{E}(\mathbf{\Delta C}R_e\mathbf{\Delta C}_*)$ are linear combinations of polynomials of type (4.5).

Let H_{fi} denote the i'th row of H_f. Introduce the following polynomials

$$\tilde{P}_{fi} \ \triangleq \ \bar{E}(H_{fi}\mathbf{C}R_e\mathbf{C}_*H_s^*) = H_{fi}\tilde{\mathbf{P}}H_0^* \tag{5.9}$$

$$\tilde{P}_s \ \triangleq \ \bar{E}(H_s\mathbf{C}R_e\mathbf{C}_*H_s^*) = \mathrm{tr}[\tilde{\mathbf{P}}\bar{E}(H_s^*H_s)] = \mathrm{tr}[\tilde{\mathbf{P}}(H_0^*H_0 + \mathbf{P}_H)] \ . \tag{5.10}$$

Introduce the stable and monic average spectral factor $\beta(q^{-1})$, and a scalar r, from

$$r\beta\beta_* \triangleq \bar{E}\{H_s\mathbf{C}R_e\mathbf{C}_*H_s^*NN_* + \rho DD_*MM_*\} = \tilde{P}_sNN_* + \rho DD_*\tilde{M}\tilde{M}_* \tag{5.11}$$

where $N = N_0N_1, D = D_0D_1$ and where $\tilde{M}\tilde{M}_*$ is defined as in (4.2). A stable β exists if and only if $\tilde{P}_s$ and $\rho\tilde{M}\tilde{M}_*$ have no common factors on $|z| = 1$.

Theorem 4. If a stable β from (5.11) exists, a stable causal estimator, which minimizes

$$\bar{E}E\mathrm{tr}[z(t)z(t)^*] = \bar{E}E\sum_{i=1}^{\ell}|f_i(t) - \hat{f}_i(t|t+m_i)|^2$$

is

$$\hat{f}(t) = (q^{m_1}Q_1 \dots q^{m_\ell}Q_\ell)^T\frac{N_0N_1}{\beta}y(t) \tag{5.12}$$

where the polynomials $Q_i(q^{-1})$, together with $L_{i*}(q)$, are unique solutions to

$$q^{-m_i}\tilde{P}_{fi}N_{0*}N_{1*} = r\beta_*Q_i + qD_0D_1L_{i*} \quad , \quad i = 1\dots\ell \tag{5.13}$$

with generic degrees

$$nQ_i = \max\{n\tilde{p}_{fi} + m_i \ , \ nd_0 + nd_1 - 1\} \ \ ; \ \ nL_i = \max\{nn_0 + nn_1 - m_i \ , \ n\beta\} - 1$$

■

Proof: See Appendix C.

Remarks. Note that the lags m_i may differ. After calculating (5.8)–(5.10), the design equations consist of the scalar spectral factorization (5.11) and of the ℓ uncoupled scalar Diophantine equations (5.13). (Since only the left–hand sides of (5.13) differ, they can be solved as a single system of linear equations.) For models without uncertainty, (5.12) is a transfer function formulation of a stationary Kalman estimator. For the nominal case with scalar $f(t)$, see Carlsson et.al. (1991).

6 ROBUST FEEDFORWARD CONTROL

It is easy to utilize the proposed technique in order to improve the performance robustness of the LQG feedforward controllers discussed, for example, by Sternad and Sderstrm (1988) or Hunt (1989). Consider a stable system with scalar output to be controlled $y(t)$, input $u(t)$ and a measurable signal $w(t)$. With a similar notation as in (2.1)–(2.3), it is described by

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t-k) + \frac{D(q^{-1})}{F(q^{-1})}w(t-d) \quad ; \quad w(t) = \frac{G(q^{-1})}{H(q^{-1})}v(t) \quad . \tag{6.1}$$

Here, $v(t)$ is white, with unit variance. As in Section 3, we use the structure

$$\frac{B}{A} = \frac{B_0}{A_0} + \frac{B_1 \Delta B}{A_1} \quad ; \quad \frac{D}{F} = \frac{D_0}{F_0} + \frac{D_1 \Delta D}{F_1} \quad ; \quad \frac{G}{H} = \frac{G_0}{H_0} + \frac{G_1 \Delta G}{H_1} \tag{6.2}$$

where ΔB, ΔD and ΔG are stochastic polynomials. They have known positive semidefinite autocovariance matrices $\mathbf{P}_{\Delta B}$ etc. All denominators are stable.

A stable and causal feedforward filter, operating on $w(t)$, is to be designed in order to minimize an infinite horizon quadratic criterion, with input penalty $\rho \geq 0$

$$J = \bar{E}E(|y(t)|^2 + \rho|u(t)|^2) \quad . \tag{6.3}$$

In a *disturbance measurement feedforward* problem, $w(t)$ represents the disturbance. In a *command feedforward* problem, $w(t)$ is a command signal and $(G/H)v(t)$ is a stochastic model, describing its second order properties. A servo filter is then to be designed, so that $(B/A)u(t-k)$ optimally follows a response model $-(D/F)w(t-d)$. The solution to both of these problems is presented below. First, calculate a stable and monic spectral factor $\tilde{G}(q^{-1})$ and a scalar s from

$$s\tilde{G}\tilde{G}_* = \bar{E}(GG_*) = G_0 G_{0*} H_1 H_{1*} + H_0 H_{0*} G_1 G_{1*} \bar{E}(\Delta G \Delta G_*) \quad . \tag{6.4}$$

It exists if and only if $\{G_0, \bar{E}(\Delta G \Delta G_*)\}$ have no common zeros on $|z| = 1$. Then, calculate another stable and monic spectral factor $\beta(q^{-1})$, and a scalar r, from

$$r\beta\beta_* = B_0 B_{0*} A_1 A_{1*} + \bar{E}(\Delta B \Delta B_*) B_1 B_{1*} A_0 A_{0*} + \rho A_0 A_{0*} A_1 A_{1*} \quad . \tag{6.5}$$

It exists if either $\rho > 0$ or if $\{B_0, B_1 \bar{E}(\Delta B \Delta B_*)\}$ lack common zeros on $|z| = 1$.

Theorem 5. A stable and causal feedforward filter minimizing (6.3) is given by

$$u(t) = -\frac{Q_1 A_0 A_1}{\beta F_0 \tilde{G}} w(t) \tag{6.6}$$

where $Q_1(q^{-1})$, together with $L_*(q)$, is the unique solution to

$$q^{-d+k} B_{0*} A_{1*} D_0 \tilde{G} = r\beta_* Q_1 + q F_0 H_0 H_1 L_* \tag{6.7}$$

with generic degrees

$$\begin{aligned} nQ_1 &= \max(nd_0 + n\tilde{g} + d - k \ , \ nf_0 + nh_0 + nh_1 - 1) \\ nL &= \max(nb_0 + na_1 - d + k \ , \ n\beta) - 1 \end{aligned}$$

■

A direct proof closely follows the reasoning in Appendix B and is omitted here. The duality discussed in Bernhardsson and Sternad (1992) can also be used. The various transfer function uncertainties affect the design equations differently:

- The uncertainty ΔG of the disturbance or reference enters via (6.4). It affects the design in a similar way as would a *measurement noise* on $w(t)$.
- The uncertainty ΔB of the system enters via (6.5), in a similar way as does the *input penalty* ρ.[8] Also, note the presence of A_1 in (6.6). If $|B_1/A_1|$ is large at frequency ω_1, indicating large uncertainty, the filter has low gain at ω_1.
- Perhaps somewhat surprisingly, uncertainty in D/F has no effect at all on the regulator design. By duality (see Bernhardsson and Sternad 1992), this corresponds to uncertainty in S/T having no effect in Section 4.

The intuitive notion that assumption of measurement noise on $w(t)$ and increase of the input penalty ρ would improve the performance robustness of a nominal design, is thus supported. With the above equations, the robustification can be tailored more exactly to the expected type and amount of uncertainty.

Example 5. Frequency domain adjustment of an error model and robustness improvement of a disturbance measurement feedforward regulator.

Assume the disturbance $w(t)$ in (6.1) to be white. The disturbance transfer function is $D/F = 1/(1-0.5q^{-1})$, with delay $d = 2$. No uncertainty in G/H or D/F is assumed, for simplicity. The transfer function

$$q^{-k}\frac{B}{A} = \frac{q^{-2}b_0(1+0.8q^{-1})}{(1-z_1q^{-1})(1-z_2q^{-1})}$$

has uncertain complex–conjugated poles

$$z_{1,2} = 1 - rp \pm 0.67pi \quad ; \quad r \in [0.4, 0.9] \quad , \quad p \in [0.2, 0.4] \quad .$$

The static gain is assumed to be exactly $= 1$, so we normalize to $b_0 = (r^2 + 0.67^2)p^2/1.8$. At higher frequencies, the uncertainty increases. Both the bandwith and the damping can vary significantly. See Fig. 8. Information about the exact distribution of r and p is not assumed to be available. (This is often the case in practice.) Still, a multiplicative error model

$$\frac{B}{A} = \frac{B_0}{A_0} + \frac{B_1\Delta B}{A_1} \quad ; \quad \frac{B_1}{A_1} = \frac{B_0}{A_0}\frac{B_m}{A_m}$$

can approximate the set of systems. Its properties are straightforward to adjust by inspection in the frequency domain, to roughly cover the expected range of

[8] In fact, the effects of ρ and of $\bar{E}(\Delta B \Delta B_*)$ are identical if $A_1 = 1$ and $B_1 = q^{-c}$, when $\mathbf{P}_{\Delta B}$ is diagonal. ($\bar{E}(\Delta B \Delta B_*)$ is then a scalar, cf (4.5).)

variations in the dynamics. As nominal model, we select the model for $r = 0.5, p = 0.3$, which is in the central part of the area of pole locations:

$$\frac{B_0}{A_0} = \frac{0.035(1 + 0.8q^{-1})}{1 - 1.70q^{-1} + 0.7617q^{-2}}$$

Since the relative uncertainty is low at low frequencies, we select a high–pass filter as fixed multiplicative error gain. A good fit is obtained with

$$\frac{B_m}{A_m} = 0.80\frac{(1 - 0.95q^{-1})^2}{(1 - 0.80q^{-1})^2} \quad .$$

The covariance matrix $\mathbf{P}_{\Delta B}$ of the coefficients of ΔB is selected as $\mathrm{diag}(\alpha\lambda^i)$, as in Example 2, Section 3.2. Use of $\alpha = 0.5$ and $\lambda = 0.5$ gives a reasonable fit to the expected range of variations in the dynamics. See Fig. 8. Based on this model, nominal and robust feedforward filters were designed for the input penalty $\rho = 0.0001$. The robust filter was calculated from (6.4)–(6.7), with result $s = 1, \tilde{G} = 1, r = 0.00334, Q_1(q^{-1}) = 12.78$ and

$$\beta(q^{-1}) = A_0(q^{-1})(1 - 1.2490q^{-1} + 0.0146q^{-2} + 0.2634q^{-3} + 0.0146q^{-4}) \quad .$$

As is evident from Fig. 9, the sensitivity of a robust design is reduced significantly, as compared to the nominal regulator ■

Figure 8. Pole location (left) and Bode magnitude plot of the system B/A in Example 5. Four corner points $(r, p) = (0.9, 0.4), (0.4, 0.4)$ (dashed), $(0.9, 0.2), (0.4, 0.2)$ (dash–dotted) are compared to 10 systems chosen randomly, using a rectangular distribution, from the set defined by the adjusted error model (solid).

Figure 9. Criterion value in Example 5 for well damped systems, $r = 0.9$ (left) and lightly damped systems, $r = 0.4$ (right). Use of the nominal controller (solid) is compared to use of the robust one (dashed), when the parameter p of the system is varied. Compare to the LQG criterion value achievable for a known system (lower dotted) and to the output variance without control (upper dotted).

7 CONCLUSIONS

Estimation and feedforward control, based on imperfectly known linear discrete–time models, has been considered. Model errors were represented as additive transfer functions, with random numerators. A robust design was obtained by minimizing the squared estimation error, averaged both with respect to model errors and noise. This allows large but unlikely model errors to be taken into account, without dominating the design. The resulting filter becomes cautious, but not conservative.

With the presented polynomial equations approach, the robust filter design becomes simple and straightforward: just sum elements along diagonals of covariance matrices. Then, solve somewhat modified, "averaged", spectral factorizations and Diophantine equations. Application to related problems such as decision feedback equalization, discussed in Sternad and Ahlén (1990), is straightforward. See Sternad and Ahlén (1992). Nonsingular continuous–time problems can be solved, with obvious modifications of the equations. Multivariable design is under investigation.

In the open–loop problems considered here, filter stability could easily be assured. The use of stochastic error models in robust *feedback* design is more problematic. An interesting conceptual shift is that the notion of guaranteed robust stability mostly has to be abandoned, for unbounded (e.g. Gaussian) error distributions. Only stability with a certain probability can then be ascertained. See e.g. Stengel and Ray (1991). This may seem unappealing from a theoretical standpoint. However, it is the price to be paid for allowing infinte tail model error distributions.

There is a need for further work on systematic ways to estimate error models, both in the time domain and in the frequency domain. The method of Goodwin et.al. (1991) requires known inputs. Parameter set identification for ARMA–models needs to be developed. A method for transforming an ordinary ARMA-model C_0/D_0, with covariance data on C_0 and D_0, into a well approximating extended design model of type $C_0/D_0 + C_1\Delta C/D_1$ is under development.

With very large system variations, the performance of even a robust linear filter will be unsatisfactory. An approach analogous to gain scheduling in feedback control can then be of use. A bank of filters is designed, with each filter attuned to a subset of the total system class. By using the output or auxiliary information, the most likely subset is selected, and the corresponding filter is used. See e.g. Lainiotis (1976) or Padilla and Haddad (1976). Robust design is a complement to this appoach. By utilizing robust filters, which give acceptable behaviour for large model subsets, the number of filters in the filter bank may be reduced.

Robust design could also complement adaptation. Adaptive robust filtering/control, based on on–line estimation of nominal models and also of error model parameters, is a challenging subject. It is a main goal motivating our present research.

Acknowledgements. The authors wish to thank Ari Kangas, Vladimir Kučera, Kjell Nordstrm and Torbjrn Wigren for valuable comments on the manuscript.

APPENDICES:

A Proof of Lemma 2.

Introduce a vector

$$f(\omega) \triangleq (\ 1\ e^{i\omega} \ldots\ e^{i\omega dp}\ )^T \quad .$$

Then,

$$f^*(\omega)\mathbf{P}_{\Delta P} f(\omega) = g_{dp}^* e^{-i\omega dp} + \ldots + g_1^* e^{-i\omega} + g_0 + g_1 e^{i\omega} + \ldots + g_{dp} e^{i\omega dp} \quad .$$

This real–valued scalar is the polynomial $\bar{E}(\Delta P \Delta P_*)$ from (4.5), evaluated on $z = e^{i\omega}$. Thus, since $\mathbf{P}_{\Delta C}, \mathbf{P}_{\Delta B}$ and $\mathbf{P}_{\Delta M}$ are all assumed positive semidefinite, the corresponding polynomials from (4.5) will be non–negative on the unit circle. This is a sufficient condition for the expressions (4.2) to be non–negative on the unit circle. If $M_1\bar{E}(\Delta M \Delta M_*)$ has no zeros on the unit circle, the same will then be true for $\tilde{M}\tilde{M}$: it can only have zeros on $|z| = 1$ which are common to $M_0 M_{0*}$ and $M_1 M_{1*}\bar{E}(\Delta M \Delta M_*)$. Such zeros are zeros also of $\tilde{C}\tilde{C}_*$ or $\tilde{B}\tilde{B}_*$ in (4.3) only if either of A3 or A4 is violated. This holds for the special case $\rho = 0$ as well.

B Proof of Theorem 1.

In Ahlén and Sternad (1991), a novel technique for simple constructive derivation of polynomial design equations for Wiener filters is presented. It is based on the evaluation of orthogonality in the frequency domain. This technique will be utilized here, and be shown to be applicable to the criterion (2.6). If there is no model uncertainty, this derivation reduces to the derivation of (2.7)–(2.10).

With (2.1)–(2.3), the estimation error is

$$z(t) = \left(\frac{S}{T} - q^{m-k}\frac{Q}{R}\frac{B}{A}\right)\frac{C}{D}e(t) - q^m \frac{QM}{RN}v(t) \quad . \tag{B.1}$$

All admissible alternatives to a proposed estimate $\hat{f}(t|t+m)$, given by (2.4), can be described by

$$\hat{d}(t) = (Q/R)y(t+m) + n(t) \quad ; \quad n(t) = \mathcal{M}y(t+m) \quad .$$

Here, $\mathcal{M}(q^{-1})$ is a causal, stable but otherwise arbitrary rational function. Optimality of (2.4) is obtained if no perturbation $n(t)$ can improve the mean estimator performance. This occurs if and only if the corresponding error $z(t)$ is orthogonal to any admissible estimator variation $n(t)$, i.e. $\bar{E}Ez(t)n^*(t) = \bar{E}Ez(t)^* n(t) = 0$. Then, the perturbed criterion value reduces to

$$\begin{aligned} \bar{E}E|f(t) - \hat{d}(t)|^2 &= \bar{E}E(\ |z(t)|^2 - z(t)n^*(t) - z^*(t)n(t) + |n(t)|^2\ ) \\ &= \bar{E}E(\ |z(t)|^2 + |n(t)|^2\ ) \end{aligned} \tag{B.2}$$

which is obviously minimized by $n(t) = 0$. Since all systems included in the extended design model (2.1)–(2.3) are assumed stable, both $z(t)$ and $n(t)$ will be stationary. Parsevals formula can then be used, to express $\bar{E}Ez(t)n^*(t)$ as

$$\bar{E}E\left\{\left(\frac{S}{T}-q^{m-k}\frac{QB}{RA}\right)\frac{C}{D}e(t)-q^m\frac{QM}{RN}v(t)\right\}\left\{q^{m-k}\mathcal{M}\frac{BC}{AD}e(t)+q^m\mathcal{M}\frac{M}{N}v(t)\right\}^*$$

$$=\bar{E}\frac{1}{2\pi i}\oint_{|z|=1}\left\{\left(z^{k-m}\frac{S}{T}-\frac{QB}{RA}\right)\frac{C}{D}\frac{B_*C_*}{A_*D_*}-\rho\frac{QMM_*}{RNN_*}\right\}\mathcal{M}_*\frac{dz}{z}$$

$$=\frac{1}{2\pi i}\oint\frac{\bar{E}\{z^{k-m}SCC_*B_*N_*NAR-TQ(CC_*BB_*NN_*+\rho MM_*AA_*DD_*)\}}{DTNARN_*A_*D_*}\mathcal{M}_*\frac{dz}{z}\ . \tag{B.3}$$

The denominator has no zeros on the unit circle. We are allowed to move the expectation $\bar{E}$ inside the integration, since, for any particular realization of $\Delta C, \Delta B$ and ΔM, the integrand is Riemann integrable on the unit circle. See e.g. Jazwinski (1970), Theorem 3.8. The expectation $\bar{E}$ operates on the numerator, since stochastic variables are present there only. Using the spectral factorization (4.1), and the fact that neither of S, N, A, R, T or Q contain stochastic variables, (B.3) can be expressed as

$$\frac{1}{2\pi i}\oint\frac{(z^{k-m}SNN_*AR\bar{E}(CC_*B_*)-TQr\beta\beta_*)}{DTNARN_*A_*D_*}\mathcal{M}_*\frac{dz}{z}\ . \tag{B.4}$$

Now, $\bar{E}Ez(t)n^*(t) = 0$ is fullfilled if all poles in $|z| < 1$ of the integrand are cancelled by zeros. Since $\mathcal{M}$ and NAD are stable, $(1/N_*A_*D_*)\mathcal{M}_*$ will have poles only in $|z| > 1$. All other poles are in $|z| < 1$. Thus, we require

$$z^{k-m}SNN_*AR\bar{E}(CC_*B_*)-TQr\beta\beta_* = zL_*DTNAR$$

for some polynomial $L_*(z)$ or, equivalently,

$$(z^{k-m}SN_*\bar{E}(CC_*B_*)-zL_*DT)NAR = Qr\beta_*\beta T\ . \tag{B.5}$$

The right–hand side of (B.5) must contain R as a factor. Since R must be stable, its factors cannot include factors of β_*. Thus, $\beta T = RH$ for some stable $H(q^{-1})$. Now, cancel R in (B.5). Observe that NA must be factor of QH, i.e. $QH = Q_1NA$. The filter $Q/R = (Q_1NA/H)/(\beta T/H)$ is (4.6). Cancel NA and exchange q for z in (B.5) to obtain

$$q^{k-m}SN_*\bar{E}(CC_*B_*) = r\beta_*Q_1+qDTL_*\ . \tag{B.6}$$

This is (4.7), since by using independence

$$\bar{E}(CC_*B_*) = \bar{E}(CC_*)\bar{E}(B_0A_1+A_0B_1\Delta B)_* = \tilde{C}\tilde{C}_*B_{0*}A_{1*}\ . \tag{B.7}$$

The "only if"–part of the result follows because choices of Q/R other than (4.6) correspond to $n(t) \neq 0$, which, according to (B.2), increase the criterion value.

Remark on the degrees (4.8). Diophantine equations in general have an infinite number of solutions. In (4.7), however, causality requires Q_1 to be a polynomial only in q^{-1}, while optimality requires L_* to be a polynomial in q. [9] Their generic degrees (4.8) are then uniquely determined by the requirement that the highest occuring powers of q^{-1} and q, respectively, must be covered by the variables in (4.7). This gives an equal number of equations and unknowns. The linear system of equations is nonsingular, since β_* and DT have no common factors [10].

The expression (4.9). The minimal mean MSE is derived by inserting (B.1), (4.2), (4.6) and (B.6), in this order, into (2.6). This gives, after some calculations, the expression $\bar{E}E|z(t)|^2_{\min} =$

$$= \frac{1}{2\pi i}\oint \frac{SS_*\bar{E}(CC_*)r\beta\beta_* + DD_*TT_*LL_* - SS_*NN_*\bar{E}(CC_*B_*)\bar{E}(CC_*B)}{r\beta\beta_*TT_*DD_*}\frac{dz}{z}$$

$$= \frac{1}{2\pi i}\oint \frac{LL_*}{r\beta\beta_*} + \frac{SS_*\tilde{C}\tilde{C}_*\{r\beta\beta_* - \tilde{C}\tilde{C}_*\tilde{B}\tilde{B}_*NN_* + \tilde{C}\tilde{C}_*(\tilde{B}\tilde{B}_* - B_0B_{0*}A_1A_{1*})NN_*\}}{r\beta\beta_*DD_*TT_*}\frac{dz}{z} \quad . \tag{B.8}$$

In the last equality of (B.8), we used $\bar{E}(CC_*B_*) = \tilde{C}\tilde{C}_*B_{0*}A_{1*}$ and exchanged $B_0B_{0*}A_1A_{1*}$ for $B_0B_{0*}A_1A_{1*} + \tilde{B}\tilde{B}_* - \tilde{B}\tilde{B}_*$. Now, utilizing (4.3) and, from (4.2), $\tilde{B}\tilde{B}_* - B_0B_{0*}A_1A_{1*} = A_0A_{0*}B_1B_{1*}\bar{E}(\Delta B\Delta B_*)$ we obtain (4.9) by using $B_1\Delta B = \Delta\mathcal{G}A_1$.

The case without uncertainty. With $\Delta C = \Delta B = \Delta M = 0$, the spectral factorization (4.3) gives $r = r_0$ and $\beta = \beta_0 D_1 A_1 N_1$. Equation (4.7) then becomes

$$q^{-m+k}SC_0C_{0*}D_1D_{1*}B_{0*}A_{1*}N_{0*}N_{1*} = r_0\beta_{0*}D_{1*}A_{1*}N_{1*}Q_1 + qD_0D_1TL_* \quad .$$

It is evident that D_1 must be a factor of Q_1 and $D_{1*}A_{1*}N_{1*}$ a factor of L_*. With $Q_1 = Q_{10}D_1$ and $L_* = L_{0*}D_{1*}A_{1*}N_{1*}$ and common factors cancelled, the above equation reduces to (2.9). The filter (4.6) equals (2.7).

C Proof of Theorem 4.

The derivation follows the reasoning in the proof of Theorem 1 closely. Consider estimation of a single component $f_i(t)$ of $f(t)$, with a scalar $\hat{f}_i(t) = (Q/R)y(t+m_i)$. Use the orthogonality relation

$$\bar{E}Ez(t)n(t)^* = \bar{E}E(f_i(t) - (Q/R)y(t+m_i))(\mathcal{M}y(t+m_i))^* = 0 \quad .$$

Use of (5.2)–(5.4) and of (5.8)–(5.11) gives $\bar{E}Ez(t)n(t)^* =$

[9]If L_* were allowed to have negative powers of z as arguments, poles at the origin would be introduced in the integrand of (B.4). The path integral would then not vanish.

[10]A more general discussion of these points can be found in Section IV of Ahlén and Sternad(1991) and in Kučera (1979,1991).

$$\bar{E}E\left\{\frac{[RH_{fi}\mathbf{C}-q^{m_i}QH_s\mathbf{C}]}{RD}e_v(t)-q^{m_i}\frac{Q}{R}\frac{M}{N}v(t)\right\}\left\{\mathcal{M}q^{m_i}\left(\frac{H_s\mathbf{C}}{D}e_v(t)+\frac{M}{N}v(t)\right)\right\}^*$$

$$=\bar{E}\frac{1}{2\pi i}\oint\left(\frac{RH_{fi}\mathbf{C}R_e\mathbf{C}_*H_s^*z^{-m_i}}{RDD_*}-\frac{QH_s\mathbf{C}R_e\mathbf{C}_*H_s^*}{RDD_*}-\frac{Q}{R}\frac{MM_*}{NN_*}\rho\right)\mathcal{M}_*\frac{dz}{z}$$

$$=\frac{1}{2\pi i}\oint\frac{(R\tilde{P}_{fi}z^{-m_i}NN_*-Qr\beta\beta_*)}{RDD_*NN_*}\mathcal{M}_*\frac{dz}{z}\quad . \tag{C.1}$$

All poles inside $|z|=1$ of the integrand of (C.1) are cancelled by zeros if and only if

$$R\tilde{P}_{fi}z^{-m_i}NN_*-Qr\beta\beta_*=zRDNL_*$$

for some polynomial $L_*(q)$ or, with $z\leftrightarrow q$,

$$R(\tilde{P}_{fi}q^{-m_i}N_*-qL_*D)N=Qr\beta\beta_*\quad . \tag{C.2}$$

We see that β must have R as a factor, i.e. $RH=\beta$ for some stable $H(q^{-1})$. Furthermore, N must be a factor of QH, i.e. $QH=Q_iN$. The scalar filter $q^{m_i}Q/R=(q^{m_i}Q_iN/H)/(\beta/H)$ can now be seen as row i of (5.12). Cancelling of NR in (C.2) reduces it to the i'th of the equations (5.13). Estimation of component i of $f(t)$ does not affect the estimate of component j. The total estimator of $f(t)$ can be obtained as ℓ parallel scalar estimators of $f_i(t)$, derived as above. Thus, we obtain (5.12).

D Additional comments on Example 3.

In Fig. 3, it is interesting to note that the robust filter attains its minimal estimation error for parameter values other than the nominal. The assymetry of the ideal performance surface in parameter space (the dotted curves in Fig. 3) seems to be a main cause of the displacement.

In the minimax design, filters were constructed for the worst case (giving highest MSE). When these filters were applied to other systems in the class, the MSE was never higher that in the design case. Thus, the minimax solution had been found. The worst case for the rectangular distribution was

$$C(q^{-1})=1-1.3q^{-1}\quad,\quad B(q^{-1})=0.41-0.66q^{-1}\quad,\quad M(q^{-1})=1-0.63q^{-1}\quad .$$

The worst case for the 5–point distribution was

$$C(q^{-1})=1-1.45q^{-1}\quad,\quad B(q^{-1})=0.3-0.9q^{-1}\quad,\quad M(q^{-1})=1-0.3q^{-1}\quad .$$

We feel that the comparision with minimax design is reasonably fair; both distributions, in particular the rectangular one, have considerable weights far from their mean values. One could easily have chosen other, more extreme, distributions, for which parameter values close to the worst case are much more unlikely.

REFERENCES

Ahlén, A. and M. Sternad (1989). Optimal deconvolution based on polynomial methods. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol 37, pp 217–226.

Ahlén, A. and M. Sternad, (1991). Wiener filter design using polynomial equations. *IEEE Transactions on Signal Processing*, vol 39, pp 2387–2399.

Bernhardsson, B. and M. Sternad (1992). Feedforward control is dual to deconvolution. To appear in the *International Journal of Control*.

Carlsson, B., A. Ahlén and M. Sternad (1991). Optimal differentiation based on stochastic signal models. *IEEE Transactions on Signal Processing*, vol 39, pp 341–353.

Carlsson, B., M. Sternad and A. Ahlén (1992). Digital differentiation of noisy data measured through a dynamic system. *IEEE Transactions on Signal Processing*, vol 40, pp 218–221.

Chung, R. C. and P. R. Bélanger (1976). Minimum–sensitivity filter for linear time–invariant stochastic systems with uncertain parameters. *IEEE Transactions on Automatic Control*, vol 21, pp 98–100.

D'Appolito, J. A. and C. E. Hutchinson (1972). A minimax approach to the design of low sensitivity state estimators. *Automatica*, vol 8, pp 599–608.

Fogel E. and Y. F. Huang (1982). On the value of information in system identification–bounded noise case. *Automatica*, vol 18, pp 229-238.

Goodwin, G. C., M. Gevers and B. Ninness (1991). Quantifying the error in estimated transfer functions with application to model order selection. *Submitted.*

Goodwin, G. C. and M. E. Salgado (1989). A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *International Journal of Adaptive Control and Signal Processing* vol–3, pp 333–356.

Goodwin, G. C., B. Ninness and H. V. Poor (1991). Choice of basis functions for continuous and discrete system modelling. *IFAC Symposium on Identification and System Parameter Estimation , Hungary.*

Grimble, M. J. (1984). Wiener and Kalman filters for systems with random parameters. *IEEE Transactions on Automatic Control*, vol 29, pp 552–554.

Grimble, M. J. and A. Elsayed (1989). H_∞ robust linear estimator. *IFAC Symposium on Adaptive Control and Signal Processing* ACASP–89, pp 159–164, Glasgow, UK.

Haddad, W. F. and D. S. Bernstein (1988). Robust, reduced–order, nonstrictly proper state estimation via the optimal projection equations with guaranteed cost bounds. *IEEE Transactions on Automatic Control*, 33, pp 591–595.

Hunt, K. J. (1989). *Stochastic Optimal Control Theory with Application to Self–tuning Control.* Springer–Verlag, Berlin.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Kassam, S. A. and T. L. Lim (1977). Robust Wiener filters. *Journal of The Franklin Institute*, pp 171–185.

Kassam, S. A. and H. V. Poor (1985). Robust techniques for signal processing: a survey. *Proceedings of the IEEE*, vol 73, pp 433– 481.

Kučera, V. (1979). *Discrete Linear Control. The Polynomial Equations Approach.* Wiley, New York.

Kučera, V. (1981). New results in state estimation and regulation. *Automatica*, vol 17, pp 745–748.

Kučera, V. (1991). *Analysis and Design of Discrete Linear Control Systems.* Academia, Prague and Prentice Hall International, London.

Lainiotis, D. G. (1976). Partitioning: a unifying framework for adaptive systems. Part I: estimation. *Proceedings of the IEEE*, vol 64, pp 1126–1143.

Leondes, C. T. and J. O. Pearson (1972). A minimax filter for systems with large plant uncertainties. *IEEE Transactions on Automatic Control*, pp 266–268.

Limebeer, D. J. N., B. D. O. Anderson and B. Hendel (1991). Mixed H_2/H_∞ filtering by the theory of Nash Games. Submitted for publication.

Ljung, L. (1987). *System Identification–Theory for the User.* Prentice Hall, Englewood Cliffs, NJ.

Loève, M. (1963). *Probability Theory.* D. Van Nostrand Company, New York.

Martin, C. J. and M. Mintz (1983). Robust filtering and prediction for linear systems with uncertain dynamics: a game–theoretic approach. *IEEE Transactions on Automatic Control*, vol 28, pp 888–896.

Moustakides, G. and S. A. Kassam (1983). Robust Wiener filters for random signals in correlated noise. *IEEE Transactions on Information Theory*, vol 29, pp 614–619.

Moustakides, G. and S. A. Kassam (1985). Minimax equalization for random signals. *IEEE Transactions on Communications*, vol 33, pp 820–825.

Nagpal, K. M. and P. P. Khargonekar (1991). Filtering and smoothing in an H_∞–setting. *IEEE Transactions on Automatic Control*, vol 36, pp 152–166.

Nahi, N. E. and E. J. Knobbe (1976). Optimal linear recursive estimation with uncertain system parameters. *IEEE Transactions on Automatic Control*, vol 21, pp 263–266.

Padilla, R. A. and A. H. Haddad (1976). On the estimation of uncertain signals using an estimation–detection scheme. *IEEE Transactions on Automatic Control*, vol 21, pp 509–512.

Poor, H. V. (1980). On robust Wiener filtering. *IEEE Transactions on Automatic Control*, vol 25, pp 531–536.

Speyer, J. L. and D. E. Gustafson (1975). An approximation method for estimation in linear systems with parameter uncertainty. *IEEE Transactions on Automatic Control*, vol 20, pp 354–359.

Stengel, R. F. and L. R. Ray (1991). Stochastic robustness of linear time–invariant control systems. *IEEE Transactions on Automatic Control*, vol 36, pp 82–87.

Sternad, M. and A. Ahlén (1990). The structure and design of realizable decision feedback equalizers for IIR channels with colored noise. *IEEE Transactions on Information Theory*, vol 36, pp 848–858.

Sternad, M. and A. Ahlén (1992). Robust decision feedback equalization. *Proc. NUTEK Workshop on Digital Communications*, Uppsala, Sweden, May 25–26 1992, pp118-121.

Sternad, M. and T. Söderström (1988). LQG–optimal feedforward regulators. *Automatica*, vol 24, pp 557-561.

Vastola, K. S. and H. V. Poor (1984). Robust Wiener–Kolmogorov theory. *IEEE Transactions on Information Theory*, vol 30, pp 316–327.

Wahlberg, B. (1991). System identification using Laguerre models. *IEEE Transactions on Automatic Control*, vol 36, pp 551–562.

Wang, S.D., T.S. Kuo and C. F. Hsu (1987). Optimal observer design for linear dynamical systems with uncertain parameters. *International Journal of Control*, vol 45, pp 701–711.