A PROBABILISTIC APPROACH TO MULTIVARIABLE ROBUST FILTERING AND OPEN-LOOP CONTROL

Kenth Öhrn, Anders Ahlén and Mikael Sternad*
Systems and Control Group, Uppsala University
P.O. Box 27, S–751 03 Uppsala, Sweden

April 24, 1996

Abstract

A new approach to robust filtering, prediction and smoothing of discrete-time signal vectors is presented. Linear time-invariant filters are designed to be insensitive to spectral uncertainty in signal models. The goal is to obtain a simple design method, leading to filters which are not overly conservative. Modelling errors are described by sets of models, parameterized by random variables with known covariances. These covariances could either be estimated from data, or be used as robustness "tuning knobs". A robust design is obtained by minimizing the $\mathcal{H}_2$-norm or, equivalently, the mean square estimation error, averaged with respect to the assumed model errors. A polynomial solution, based on an averaged spectral factorization and a unilateral Diophantine equation, is derived. The robust estimator is referred to as a *cautious* Wiener filter. It turns out to be only slightly more complicated to design than an ordinary Wiener filter. The methodology can be applied to any open loop filtering or control problem. In particular, we illustrate this for the design of robust multivariable feedforward regulators, decoupling and model matching filters.

*The work has been supported by the Swedish Research Council for Engineering Sciences (TFR), under grant 92-775.

1 INTRODUCTION

For any model-based filter, modelling errors are a potential source of performance degradation. Here, we will propose a *cautious Wiener filter* for the prediction, filtering or smoothing of discrete-time signal vectors. As in the scalar case, discussed in [35], it constitutes a generalization of the polynomial equations methodology pioneered by Kučera [20]. The design is based on a stochastic description of model errors, related to the stochastic embedding concept of Goodwin and co-workers [11], [12]. To be more specific, our approach is based on the following choices:

- A set of (true) dynamic systems is assumed to be well described by a set of discrete-time, stable, linear and time-invariant transfer function matrices

$$\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{F}}_o + \Delta\boldsymbol{\mathcal{F}} \quad . \tag{1.1}$$

  We call such a set an *extended design model.* Above, $\boldsymbol{\mathcal{F}}_o$ represents a stable nominal model, while an *error model* $\Delta\boldsymbol{\mathcal{F}}$ describes a set of stable transfer functions, parameterized by stochastic variables. The random variables enter linearly into $\Delta\boldsymbol{\mathcal{F}}$.

- A single robust linear filter is to be designed for the whole class of possible systems. Robust performance is obtained by minimizing the averaged mean square estimation error criterion

$$J = \text{trace}\bar{E}E(\varepsilon(k)\varepsilon(k)^*) \quad . \tag{1.2}$$

  Here, $\varepsilon(k)$ is the estimation error vector, E denotes expectation over noise and $\bar{E}$ is an expectation over the stochastic variables parameterizing the error model $\Delta\boldsymbol{\mathcal{F}}$.

The averaged mean square error has been used previously in the literature by e.g. Chung and Bélanger [9], Speyer and Gustafson [31] and by Grimble [13]. These works were based on assumptions of small parametric uncertainties and on series expansions of uncertain parameters. We suggest the use of the criterion (1.2), together with a particular description of the sets (1.1): transfer function elements in $\Delta\boldsymbol{\mathcal{F}}$ have stochastic numerators and fixed denominators. Such models can describe non-parametric uncertainty and under-modelling as well as parametric uncertainty. A discussion of the utility, and versatility, of linearly parameterized stochastic error models can be found in [35].

Most previous suggestions for obtaining robust filters have been based on some type of minimax approach [10], [23]. A paper [25] by Martin and Mintz takes both spectral uncertainty and uncertainty in the noise distribution into account. The resulting filter will, however, be of very high order. Minimax design of a filter $\boldsymbol{\mathcal{R}}$ becomes very complex, unless there exists either a saddle point or a boundary point solution. A crucial condition here is that $\min_{\mathcal{R}} \max_{\mathcal{F}}$ equals $\max_{\mathcal{F}} \min_{\mathcal{R}}$. If so, instead of finding the worst case with respect to a set of models, one can search for models whose optimal filter gives the worst (nominal) performance, and use the corresponding filter. This is a much simpler task, but can still be computationally demanding. See [18], [27], [30], [36] and the survey paper by Kassam and

Poor [19]. The condition $\min_{\mathcal{R}} \max_{\mathcal{F}} = \max_{\mathcal{F}} \min_{\mathcal{R}}$ is *not* fulfilled in numerous problems, which makes them very difficult to solve. See e.g. Example 5 in [35], and the example in Section 4 below.

Kalman filter-like estimators have recently been developed for systems with structured and possibly time-varying parametric uncertainty of the type

$$x(k+1) = (\mathbf{A} + \mathbf{D}\boldsymbol{\Delta}(k)\mathbf{E})\, x(k) + w(k)$$

where the matrix $\boldsymbol{\Delta}(k)$ contains norm-bounded uncertain parameters. See [29], [7] and [37] for continuous-time results and [38] for the discrete-time one-step-ahead predictor. See also [15] for a related method. For systems which are stable for all $\boldsymbol{\Delta}(k)$, an upper bound on the estimation error covariance matrix can be minimized by solving two coupled Riccati equations, combined with a one-dimensional numerical search. This represents a computational simplification, as compared to previous minimax designs. Still, the resulting estimators are quite conservative, partly because they rest on worst case design. This conservativism is illustrated and discussed in [28].

The method suggested in the present paper is computationally simpler than any of the minimax schemes referred to above. It also avoids two drawbacks of worst case designs. First, the stochastic variables in $\Delta\boldsymbol{\mathcal{F}}$ need not have compact support. Thus, the 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. Secondly, not only the range of the uncertainties, but also their likelihood is taken into account by using the expectation $\bar{E}(\cdot)$ of the MSE. Highly probable model errors will affect the estimator design more than do very rare "worst cases". Therefore, the performance loss in the nominal case, the price paid for robustness, becomes smaller than for a minimax design. In other words, conservativeness is reduced. There do, of course, exist applications where a worst case design is mandatory, e.g. for safety reasons. However, we believe that the average performance of estimators is often a more appropriate measure of performance robustness.

In the present paper, one of our goals will be to present transparent design equations, and to hold their number to a minimum, without sacrificing numerical accuracy. We use matrix fraction descriptions with diagonal denominators and common denominator forms. This leads to a solution which is, in fact, significantly simpler, and more numerically well-behaved, than the corresponding nominal $\mathcal{H}_2$-designs (without uncertainty) presented in [1] or [14]. Somewhat surprisingly, taking model uncertainty into account does not require any new types of design equations. We end up with just two equations for robust estimator design: a polynomial matrix spectral factorization and a unilateral Diophantine equation. The solution provides structural insight; important properties of a robust estimator are evident by direct inspection of the filter expression.

This paper is organized as follows. The filtering problem, the model structure (1.1) and the criterion (1.2) are discussed in more detail in Section 2 below. Section 3 presents the design equations and some tools for performance evaluation. The design procedure is illustrated by an elaborate numerical example in Section 4. The resulting estimator reduces the impact of model uncertainty and limited signal energy by using multiple sensors in an efficient way. In Section 5 the design of robust feedforward regulators, servos and model matching filters is discussed.

Remarks on the notation

Signals and polynomial coefficients may, in the following, be complex-valued. (This is required in e.g. communications applications.) Let p_j^* denote the complex conjugate (and transpose for matrices) of a polynomial coefficient p_j. For any polynomial

$$P(q^{-1}) = p_o + p_1 q^{-1} + \ldots + p_{np} q^{-np}$$

in the backward shift operator q^{-1}, define the conjugate polynomial

$$P_*(q) \triangleq p_o^* + p_1^* q + \ldots + p_{np}^* q^{np}$$

where q is the forward shift operator. A polynomial $P(q, q^{-1})$ having coefficients of both q and q^{-1} will be called *double sided*. Rational matrices, or transfer functions, are denoted by boldface calligraphic symbols, e.g. $\boldsymbol{\mathcal{R}}(q^{-1})$. Polynomial matrices are denoted by boldface symbols, such as $\boldsymbol{P}(q^{-1})$, while constant matrices are denoted as $\mathbf{P}$. For example, the identity matrix of dimension n is denoted $\mathbf{I}_n$. We denote the trace of $\boldsymbol{P}$ by tr$\boldsymbol{P}$. For polynomial or rational matrices, $\boldsymbol{P}_*(q)$ and $\boldsymbol{\mathcal{R}}_*(q)$ means complex conjugate, transpose and substitution of q for q^{-1}. When appropriate, the complex variable z or $e^{i\omega}$ is substituted for the forward shift operator q. Arguments of polynomials and matrices are often omitted, when there is no risk of misunderstanding. The *degree* of a polynomial matrix is the highest degree of any of its polynomial elements. Square polynomial matrices $\boldsymbol{P}(q^{-1})$ are called *stable* if all zeros of det $\boldsymbol{P}(z^{-1})$ are located in $|z| < 1$. A rational matrix is defined as stable if all its elements are stable. Causality is defined in the same way.

A rational matrix $\boldsymbol{\mathcal{G}}(q^{-1})$ may be represented by polynomial matrices as a matrix fraction description (MFD), either left $\boldsymbol{\mathcal{G}} = \boldsymbol{A}_1^{-1}\boldsymbol{B}_1$ or right $\boldsymbol{\mathcal{G}} = \boldsymbol{B}_2\boldsymbol{A}_2^{-1}$. It may also be represented in a common denominator form $\boldsymbol{\mathcal{G}} = \boldsymbol{B}/A$, where $\boldsymbol{B}$ is a polynomial matrix. The scalar and monic polynomial A is then the least common denominator of all elements in $\boldsymbol{\mathcal{G}}$. Denominator matrices in MFDs are assumed to have identity matrices as leading coefficients of their matrix polynomial representations, thus $\boldsymbol{A}_i(0) = \mathbf{I}$ above.

2 THE ROBUST ESTIMATION PROBLEM

Consider the following extended design model

$$\begin{aligned} y(k) &= \boldsymbol{\mathcal{G}}(q^{-1})u(k) + \boldsymbol{\mathcal{H}}(q^{-1})v(k) \\ u(k) &= \boldsymbol{\mathcal{F}}(q^{-1})e(k) \\ f(k) &= \boldsymbol{\mathcal{D}}(q^{-1})u(k) \end{aligned} \tag{2.1}$$

where $\mathcal{G}$, $\mathcal{H}$, $\mathcal{F}$ and $\mathcal{D}$ are stable and causal, but possibly uncertain, transfer functions of dimension $p|s$, $p|r$, $s|n$ and $\ell|s$, respectively. The noise sequences $\{e(k)\}$ and $\{v(k)\}$ are mutually uncorrelated and zero mean stochastic sequences. To obtain a simple notation they are assumed to have unit covariance matrices, so scaling and uncertainty of the covariances are included in $\mathcal{F}$ and $\mathcal{H}$, respectively. The signal $y(k)$ is assumed measurable, while $f(k)$ is the signal to be estimated.

2.1 Multisignal estimation

From data $y(k)$ up to time $k+m$, an estimator

$$\hat{f}(k|k+m) = \mathcal{R}(q^{-1})y(k+m) \tag{2.2}$$

of $f(k)$ is sought. See Figure 1. The estimator may be a predictor ($m < 0$), a filter ($m = 0$), or a fixed lag smoother ($m > 0$). Here $\mathcal{R}$, of dimension $\ell|p$, is required to be stable and causal. The transfer function $\mathcal{R}$ is designed to minimize the averaged mean square error (MSE) criterion (1.2)

$$J = \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*) = \bar{E}E(\varepsilon(k)^*\varepsilon(k)) = \sum_{i=1}^{\ell} \bar{E}E|\varepsilon_i(k)|^2 \tag{2.3}$$

where

$$\varepsilon(k) = (\varepsilon_1(k)\dots\varepsilon_\ell(k))^T \triangleq \mathcal{W}(q^{-1})(f(k) - \hat{f}(k|k+m)) \quad .$$

Above, $\mathcal{W}$ is a stable and causal $\ell|\ell$ rational weighting matrix, with a stable and causal inverse. It may be used by the designer to emphasize filtering performance in particular frequency bands. In filtering problems, $\mathcal{W}$ is not assumed uncertain.

The model (2.1) offers considerable flexibility. For example, when estimating a signal $u(k)$ in coloured noise, we set $\mathcal{G} = \mathcal{D} = \mathbf{I}_s$, giving $f(k) = u(k)$. In deconvolution, or input estimation problems, $\mathcal{G}$ is a dynamic system and $\mathcal{D} = \mathbf{I}_s$. In a state estimation problem, $u(k)$ is the state vector, $\mathcal{G}$ and $\mathcal{D}$ are constant matrices while $\mathcal{H}\,v(k)$ represents (coloured) measurement noise. Other special cases are discussed in [2], [3] and [8].

Example. An application where uncertain dynamics in $\mathcal{G}$ is of interest is equalizer design for digital mobile radio communications [22]. A signal $u(k)$ then propagates along multiple paths, with different time delays, represented by delays in $\mathcal{G}$. The receiving antenna may have $p > 1$ elements (diversity design). See e.g. [4]. Thus, an appropriate model of $\mathcal{G}$ is a column vector of FIR channels, i.e. a vector of polynomials. The polynomials coefficients are estimated from short and noisy training sequences, with a known input $\{u(k)\}$. Estimation errors are inevitable. The task of a (robust) equalizer is to estimate $u(k)$, based on noisy measurements $y(k+m)$, a nominal model $\mathcal{G}_o$, and an estimate of the amount of model uncertainty $\square$

Figure 2.1: A general linear filtering problem formulation. Based on noisy measurements $y(k+m)$, the signal $f(k)$ is to be estimated. Model errors in transfer functions are described by stochastic error models.

2.2 Parameterization of the model

We choose to parameterize $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{H}}$ as left MFD's having *diagonal* denominators[1], while $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{W}}$ are parameterized in common denominator form;

$$\boldsymbol{\mathcal{G}} = \boldsymbol{A}^{-1}\boldsymbol{B} \quad ; \quad \boldsymbol{\mathcal{H}} = \boldsymbol{N}^{-1}\boldsymbol{M} \tag{2.4}$$

$$\boldsymbol{\mathcal{F}} = \frac{1}{D}\boldsymbol{C} \quad ; \quad \boldsymbol{\mathcal{D}} = \frac{1}{T}\boldsymbol{S} \quad ; \quad \boldsymbol{\mathcal{W}} = \frac{1}{U}\boldsymbol{V} \quad .$$

We have made these choices to obtain tidy and transparent design equations, and to avoid coprime factorizations, which are known to be numerically sensitive. In (2.4), it is assumed that $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{F}}$ may be uncertain. Introduction of uncertainty in the weighting matrix $\boldsymbol{\mathcal{W}}$ is not motivated in filtering problems. Its role in open-loop control, however, will be discussed in Section 3.2 below. It is shown in Appendix C that uncertainty in $\boldsymbol{\mathcal{D}}$ does not affect the optimal filter design, provided it is uncorrelated to uncertainties in other blocks. Therefore, uncertainty in $\boldsymbol{\mathcal{D}}$ is not introduced.

The extended design models, cf. (1.1) and (2.1),

$$\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}}_o + \Delta\boldsymbol{\mathcal{G}} \quad , \quad \boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_o + \Delta\boldsymbol{\mathcal{H}} \quad , \quad \boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{F}}_o + \Delta\boldsymbol{\mathcal{F}}$$

are now expressed in polynomial matrix form. Using $\hat{\boldsymbol{B}}_o = \boldsymbol{A}_1\boldsymbol{B}_o$, $\hat{\boldsymbol{B}}_1 = \boldsymbol{A}_o\boldsymbol{B}_1$ etc. we introduce

$$\begin{aligned}
\boldsymbol{\mathcal{G}} &= \boldsymbol{A}_o^{-1}\boldsymbol{B}_o + \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\Delta\boldsymbol{B} &&= \boldsymbol{A}_o^{-1}\boldsymbol{A}_1^{-1}(\hat{\boldsymbol{B}}_o + \hat{\boldsymbol{B}}_1\Delta\boldsymbol{B}) &&\triangleq \boldsymbol{A}^{-1}\boldsymbol{B} \\
\boldsymbol{\mathcal{H}} &= \boldsymbol{N}_o^{-1}\boldsymbol{M}_o + \boldsymbol{N}_1^{-1}\boldsymbol{M}_1\Delta\boldsymbol{M} &&= \boldsymbol{N}_o^{-1}\boldsymbol{N}_1^{-1}(\hat{\boldsymbol{M}}_o + \hat{\boldsymbol{M}}_1\Delta\boldsymbol{M}) &&\triangleq \boldsymbol{N}^{-1}\boldsymbol{M} \\
\boldsymbol{\mathcal{F}} &= \frac{1}{D_o}\boldsymbol{C}_o + \frac{1}{D_1}\boldsymbol{C}_1\Delta\boldsymbol{C} &&= \frac{1}{D_oD_1}(\hat{\boldsymbol{C}}_o + \hat{\boldsymbol{C}}_1\Delta\boldsymbol{C}) &&\triangleq \frac{1}{D}\boldsymbol{C} \quad .
\end{aligned} \tag{2.5}$$

[1]Note that this is a natural choice, if transfer functions are obtained by means of identification.

Above, $\boldsymbol{\mathcal{G}}_o = \boldsymbol{A}_o^{-1}\boldsymbol{B}_o$ represents the nominal model and $\Delta\boldsymbol{\mathcal{G}} = \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\Delta\boldsymbol{B}$ the error model. The same holds for $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{F}}$. The diagonal polynomial matrices $\boldsymbol{A} = \boldsymbol{A}_o\boldsymbol{A}_1$, $\boldsymbol{N} = \boldsymbol{N}_o\boldsymbol{N}_1$ and the polynomials $D = D_oD_1, T$ and U are all assumed to be stable, with causal inverses. Denominator polynomials are assumed monic. In the error models, the polynomial D_1, the diagonal matrices $\boldsymbol{A}_1$ and $\boldsymbol{N}_1$ and the matrices $\boldsymbol{C}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{M}_1$ are fixed. They can be used to tailor the error models for specific needs. For example, if multiplicative error models are deemed appropriate, we use $\boldsymbol{A}_1 = \boldsymbol{A}_o$, $\boldsymbol{B}_1 = \boldsymbol{B}_o\boldsymbol{B}_m$ etc., with $\boldsymbol{B}_m$ to be specified.

The matrices $\Delta\boldsymbol{B}$, $\Delta\boldsymbol{C}$ and $\Delta\boldsymbol{M}$ contain polynomials, with jointly distributed random variables as coefficients. These coefficients parameterize the class of assumed true systems. One particular modelling error is represented by one particular realization of the random coefficients. Element ij of a stochastic polynomial matrix $\Delta\boldsymbol{P}$ is denoted

$$\Delta P^{ij} \triangleq [\Delta\boldsymbol{P}]_{ij} = \Delta p_o^{ij} + \Delta p_1^{ij}q^{-1} + \ldots + \Delta p_{\delta p}^{ij}q^{-\delta p} \tag{2.6}$$

where δp is the degree of $\Delta\boldsymbol{P}$, i.e. the highest degree appearing in any polynomial ΔP^{ij}. All coefficients have zero means, so the nominal model is the average model in the set. Only the second order moments of the random coefficients need to be specified, since the type of distribution, and higher order moments, will not affect the filter design. The parameter covariances are denoted $\bar{E}(\Delta p_r^{ij})(\Delta p_s^{\ell k})^*$ and are collected in covariance matrices $\mathbf{P}_{\Delta\mathbf{P}}^{(ij,\ell k)}$, see Section 2.3.

We now introduce the assumption

A1. The coefficients of all polynomial elements of $\Delta\boldsymbol{C}$ are independent of those of $\Delta\boldsymbol{B}$.

It is possible to exclude Assumption **A1**, but it does simplify the solution and it is also reasonable in most practical cases.

Error models can be obtained from ordinary identification experiments, provided the model structures match. For SISO systems, error models can be estimated in presence of under-modelling, using a maximum likelihood approach [11]. Even if the statistics is hard to obtain, one could still use the elements of covariance matrices pragmatically, as robustness "tuning knobs". They are then used similarly as when weighting matrices are adjusted in LQG controller design. An objective could be to obtain reasonable performance for the uncertainty set, for a prespecified acceptable degradation of performance in the nominal case. The error models may also be used to account for a slowly time-varying dynamics [24].

One way of obtaining the models (2.5)–(2.6) is by series expansion of state space models with parametric uncertainty. Parameter deviations are represented by stochastic variables. For small uncertainties, a first order expansion can be used, which will directly lead to models of type (2.5). For larger uncertainties, a second order Taylor expansion is usually sufficient, see [28]. For non-parametric uncertainties, the error models can be adjusted directly to frequency domain data. In that context, a very useful concept is provided by the stochastic frequency domain theory of Goodwin and Salgado, see [12].

2.3 Covariance matrices for the stochastic coefficients

In order to represent the uncertainties of the system in a natural way, covariance matrices will be organized as follows. The ij-th element of a stochastic polynomial matrix $\Delta\boldsymbol{P}$ can be expressed as

$$\Delta P^{ij}(q^{-1}) = \varphi^T(q^{-1})\bar{p}_{ij} \tag{2.7}$$

where

$$\varphi^T(q^{-1}) = (1 \quad q^{-1}\dots q^{-\delta p}) \quad ; \quad \bar{p}_{ij} = (\Delta p_o^{ij} \quad \Delta p_1^{ij}\dots\Delta p_{\delta p}^{ij})^T \quad . \tag{2.8}$$

The cross covariance matrix $\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)}$, of dimension $\delta p+1|\delta p+1$, between elements of $\Delta P^{ij}(q^{-1})$ and $\Delta P^{\ell k}(q^{-1})$, is given by

$$\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)} = \bar{E}\,\overline{p}_{ij}\overline{p}_{\ell k}^* = \begin{bmatrix} \bar{E}(\Delta p_o^{ij})(\Delta p_o^{\ell k})^* & \dots & \bar{E}(\Delta p_o^{ij})(\Delta p_{\delta p}^{\ell k})^* \\ \vdots & \ddots & \vdots \\ \bar{E}(\Delta p_{\delta p}^{ij})(\Delta p_o^{\ell k})^* & \dots & \bar{E}(\Delta p_{\delta p}^{ij})(\Delta p_{\delta p}^{\ell k})^* \end{bmatrix} \tag{2.9}$$

where $\mathbf{P}_{\mathbf{\Delta P}}^{(ij,ij)}$ is Hermitian and positive semidefinite, while $\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)} = (\mathbf{P}_{\mathbf{\Delta P}}^{(\ell k,ij)})^*$. Thus,

$$\bar{E}(\Delta P^{ij}\Delta P_*^{\ell k}) = \bar{E}(\varphi^T(q^{-1})\bar{p}_{ij}\bar{p}_{\ell k}^*\varphi_*^T(q)) = \varphi^T\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)}\varphi_*^T \quad . \tag{2.10}$$

With autocovariances, $(ij) = (\ell k)$, we model the uncertainty within each input–output pair. Cross-dependencies between different transfer functions may also be known. For example, uncertainty in one single physical parameter may very well enter into several transfer functions between inputs and outputs. Such effects are captured by cross covariances, $(ij) \neq (\ell k)$.

We collect all matrices of type (2.9) into one large covariance matrix, organized as

$$\mathbf{P}_{\mathbf{\Delta P}} = \begin{pmatrix} \begin{bmatrix} \mathbf{P}_{\mathbf{\Delta P}}^{(11,11)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(11,1m)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\mathbf{\Delta P}}^{(1m,11)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(1m,1m)} \end{bmatrix} & \dots & \begin{bmatrix} \mathbf{P}_{\mathbf{\Delta P}}^{(11,n1)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(11,nm)} \\ \vdots & & \vdots \\ \mathbf{P}_{\mathbf{\Delta P}}^{(1m,n1)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(1m,nm)} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} \mathbf{P}_{\mathbf{\Delta P}}^{(n1,11)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(n1,1m)} \\ \vdots & & \vdots \\ \mathbf{P}_{\mathbf{\Delta P}}^{(nm,11)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(nm,1m)} \end{bmatrix} & \dots & \begin{bmatrix} \mathbf{P}_{\mathbf{\Delta P}}^{(n1,n1)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(n1,nm)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\mathbf{\Delta P}}^{(nm,n1)} & \dots & \mathbf{P}_{\mathbf{\Delta P}}^{(nm,nm)} \end{bmatrix} \end{pmatrix} . \tag{2.11}$$

If $\Delta\boldsymbol{P}$ has dimension $n|m$, then $\mathbf{P}_{\mathbf{\Delta P}}$ has nm by nm blocks $\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)}$. The structure of (2.11) is useful from a design point of view. If, for example, a multivariable moving average model, or FIR model, is to be identified, then (2.11) is the natural way of representing the covariance matrix. If we instead prefer to use the blocks $\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)}$ of (2.11) as multivariable "tuning knobs", a given amount of uncertainty can be assign to a specific input-output pair.

3 DESIGN OF ROBUST FILTERS

3.1 An averaged spectral factorization

We define an averaged spectral factor $\boldsymbol{\beta}(q^{-1})$ as the numerator polynomial matrix of an averaged innovations model. It constitutes a key element of the robust filter. The average, over the set of models, of the spectral density matrix $\boldsymbol{\Phi}_y(e^{i\omega})$ of the measurement $y(k)$ is given by

$$\bar{E}\{\boldsymbol{\Phi}_y(e^{i\omega})\} = \frac{1}{DD_*}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1} \quad .$$

The square polynomial matrix $\boldsymbol{\beta}(z^{-1})$ is given by the stable solution to

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \bar{E}\{\boldsymbol{NBCC}_*\boldsymbol{B}_*\boldsymbol{N}_* + D\boldsymbol{AMM}_*\boldsymbol{A}_*D_*\} \quad . \tag{3.1}$$

Note that $\boldsymbol{N}^{-1}$ and $\boldsymbol{A}^{-1}$ are diagonal, and will thus commute. The averaged second order statistics of $y(k)$ is thus described by the same spectral density as for a vector-ARMA model

$$\bar{y}(k) = \frac{1}{D}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\,\epsilon(k) \tag{3.2}$$

where $\epsilon(k)$ is white with a unit covariance matrix. This model is denoted as the *averaged innovations model.* (Note that $\bar{y}(k) \neq y(k)$, but $\boldsymbol{\Phi}_{\bar{y}}(e^{i\omega}) = \bar{E}\{\boldsymbol{\Phi}_y(e^{i\omega})\}$). When constructing the right-hand side of (3.1), the following results are useful.

Lemma 1

Let $\boldsymbol{H}(q, q^{-1})$ be an $m|m$ polynomial matrix with double-sided polynomial elements having stochastic coefficients. Also, let $\boldsymbol{G}(q^{-1})$ be an $n|m$ polynomial matrix with polynomial elements having stochastic coefficients, independent of all those of $\boldsymbol{H}$. Then,

$$\bar{E}[\boldsymbol{GHG}_*] = \bar{E}[\boldsymbol{G}\bar{E}(\boldsymbol{H})\boldsymbol{G}_*] \tag{3.3}$$

$\square$

Proof. See Appendix A.

Now, introduce the double-sided polynomial matrices

$$\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* \triangleq \bar{E}(\boldsymbol{CC}_*) \quad ; \quad \tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} \triangleq \bar{E}(\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{B}_*) \quad ; \quad \tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_* \triangleq \bar{E}(\boldsymbol{MM}_*) \tag{3.4}$$

Invoking (2.5) and using the fact that the stochastic coefficients are assumed to be zero mean, gives

$$\begin{aligned}
\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* &= \hat{\boldsymbol{C}}_o\hat{\boldsymbol{C}}_{o*} + \hat{\boldsymbol{C}}_1\bar{E}(\Delta\boldsymbol{C}\Delta\boldsymbol{C}_*)\hat{\boldsymbol{C}}_{1*} \\
\tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} &= \hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*} + \hat{\boldsymbol{B}}_1\bar{E}(\Delta\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\Delta\boldsymbol{B}_*)\hat{\boldsymbol{B}}_{1*} \\
\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_* &= \hat{\boldsymbol{M}}_o\hat{\boldsymbol{M}}_{o*} + \hat{\boldsymbol{M}}_1\bar{E}(\Delta\boldsymbol{M}\Delta\boldsymbol{M}_*)\hat{\boldsymbol{M}}_{1*} \quad .
\end{aligned} \tag{3.5}$$

Factorizations to obtain $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{B}}_C$ etc. need not be performed. The double-sided polynomial matrices are expressed as $\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*$ etc. merely to simplify the notation.

Lemma 2

Let Assumption **A1** hold. By using (3.4), (3.5) and invoking Lemma 1, the averaged spectral factorization (3.1) can be expressed as

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{N}\tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*}\boldsymbol{N}_* + D\boldsymbol{A}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_*\boldsymbol{A}_*D_* \tag{3.6}$$

□

Proof. See Appendix A.

With a given right-hand side, equation (3.6) is just an ordinary polynomial matrix left spectral factorization. It is solvable under the following mild assumption

A2. The averaged spectral density matrix $\bar{E}\{\boldsymbol{\Phi}_y(e^{i\omega})\}$ is nonsingular for all ω.

This assumption is equivalent to the right-hand side of (3.6) being nonsingular on $|z| = 1$. Then, the solution to (3.6) is unique, up to a right unitary factor. (If $\boldsymbol{H}\boldsymbol{H}_* = \mathbf{I}$, then $\boldsymbol{\beta}\boldsymbol{\beta}_* = (\boldsymbol{\beta}\boldsymbol{H})(\boldsymbol{H}_*\boldsymbol{\beta}_*)$.) Under **A2**, a solution exists, with $\boldsymbol{\beta}$ having nonsingular leading coefficient matrix $\boldsymbol{\beta}(0)$. Its degree, $n\beta$, will be determined by the maximal degree of the two right-hand side terms in (3.6).[2]

To obtain the right-hand side of (3.6), averaged polynomial matrices $\bar{E}(\Delta\boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*)$ have to be computed, where $\boldsymbol{H}(q, q^{-1}) = \tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*$ or $\mathbf{I}$. It is shown in Appendix B that the ij-th element of $\bar{E}(\Delta\boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*)$ is given by

$$\bar{E}[\Delta\boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*]_{ij} = \mathrm{tr}\boldsymbol{H}\begin{bmatrix} \varphi^T & & 0 \\ & \ddots & \\ 0 & & \varphi^T \end{bmatrix}\begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(i1,j1)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(im,j1)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(i1,jm)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(im,jm)} \end{bmatrix}\begin{bmatrix} \varphi_*^T & & 0 \\ & \ddots & \\ 0 & & \varphi_*^T \end{bmatrix} \tag{3.7}$$

where φ^T is defined in (2.8). The block covariance matrix in (3.7) is obtained by taking the block-transpose of the ij-th block $[\,\cdot\,]$ of $\mathbf{P}_{\Delta\mathbf{P}}$ in (2.11). Average factors in (3.5) are readily obtained by substituting $\Delta\boldsymbol{C}$, $\Delta\boldsymbol{B}$ and $\Delta\boldsymbol{M}$ for $\Delta\boldsymbol{P}$ in (3.7).

3.2 A second spectral factorization

In the feedforward control problems discussed in Section 5, we shall allow for $\boldsymbol{\mathcal{W}}$ being uncertain. Using a common denominator form, $\boldsymbol{\mathcal{W}}$ is parameterized in a similar way as $\boldsymbol{\mathcal{F}}$ (cf. (2.5))

$$\boldsymbol{\mathcal{W}} \;=\; \boldsymbol{V}_o\frac{1}{U_o} + \Delta\boldsymbol{V}\boldsymbol{V}_1\frac{1}{U_1} \;=\; \frac{1}{U_oU_1}(\hat{\boldsymbol{V}}_o + \Delta\boldsymbol{V}\hat{\boldsymbol{V}}_1) \;\triangleq\; \frac{1}{U}\boldsymbol{V} \;. \tag{3.8}$$

[2]When solving (3.6), we have utilized an algorithm by Ježek and Kučera, presented in [17]. It provides a solution with an upper triangular full rank leading coefficient matrix. For an overview of spectral factorization algorithms, see [21].

A stable square matrix $\tilde{\boldsymbol{V}}$, with $\tilde{\boldsymbol{V}}(0)$ nonsingular, is introduced as a solution of the *right* spectral factorization

$$\tilde{\boldsymbol{V}}_*\tilde{\boldsymbol{V}} = \bar{E}(\boldsymbol{V}_*\boldsymbol{V}) = \hat{\boldsymbol{V}}_{o*}\hat{\boldsymbol{V}}_o + \hat{\boldsymbol{V}}_{1*}\bar{E}(\Delta\boldsymbol{V}_*\Delta\boldsymbol{V})\hat{\boldsymbol{V}}_1 \quad . \tag{3.9}$$

Also, introduce the following assumptions

A3. The coefficients of $\Delta\boldsymbol{V}$ are independent of all other stochastic coefficients.

A4. The right-hand side of (3.9) is nonsingular on the unit circle.

Whenever $\boldsymbol{\mathcal{W}}$ is known, ($\boldsymbol{\mathcal{W}} = \boldsymbol{V}_o/U_o = \boldsymbol{V}/U$), equation (3.9) need not be solved, and $\tilde{\boldsymbol{V}} = \boldsymbol{V}_o = \boldsymbol{V}$. This will be the case in filtering problems.

3.3 The cautious multivariable Wiener filter

Theorem 1

Assume an extended design model (2.1), (2.4), (2.5), (3.8) to be given, with known covariance matrices (2.11). Assume **A1**–**A4** to hold. A realizable estimator of $f(k)$ then minimizes the averaged MSE (2.3), among all linear time-invariant estimators based on $y(k+m)$, if and only if it has the same coprime factors as

$$\hat{f}(k|k+m) = \boldsymbol{\mathcal{R}}\, y(k+m) = \frac{1}{T}\tilde{\boldsymbol{V}}^{-1}\boldsymbol{Q}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A}\, y(k+m) \quad . \tag{3.10}$$

Here, $\boldsymbol{\beta}(q^{-1})$ is obtained from (3.6), $\tilde{\boldsymbol{V}}(q^{-1})$ from (3.9) while $\boldsymbol{Q}(q^{-1})$ together with $\boldsymbol{L}_*(q)$, both of dimensions $\ell|p$, is the unique solution to the unilateral Diophantine equation

$$q^{-m}\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_* = \boldsymbol{Q}\boldsymbol{\beta}_* + q\boldsymbol{L}_*UTD\,\mathbf{I}_p \tag{3.11}$$

with generic[3] degrees

$$\begin{aligned} nQ &= \max(n\tilde{v} + ns + n\tilde{c} + m, nu + nt + nd - 1) \\ nL_* &= \max(n\tilde{c} + n\hat{b}_o + nn - m, n\beta) - 1 \end{aligned} \tag{3.12}$$

where $ns = \deg \boldsymbol{S}$ etc. When applying the estimator (3.10) on an ensemble of systems, the minimal criterion value becomes

$$\mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*)_{\min} = \mathrm{tr}\frac{1}{2\pi j}\oint_{|z|=1}\Big\{\boldsymbol{L}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{L} +$$

$$+\frac{1}{UTDD_*T_*U_*}\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\left[\mathbf{I}_n - \tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\right]\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\tilde{\boldsymbol{V}}_*\Big\}\frac{dz}{z} \tag{3.13}$$

□

Proof. See Appendix C.

Remarks. The only new type of computation, as compared to the nominal case described in [1]–[3], is the calculation of averaged polynomials, using (3.7).

[3] In special cases, the degrees may be lower.

Since both $\tilde{\boldsymbol{V}}$ and $\boldsymbol{\beta}$ are stable, the estimator $\boldsymbol{\mathcal{R}}$ will be stable[4]. If Assumptions **A2** and **A4** hold, $\tilde{\boldsymbol{V}}(0)$ and $\boldsymbol{\beta}(0)$ are nonsingular, so $\boldsymbol{\mathcal{R}}$ will be causal.

Note that the diagonal matrix $\boldsymbol{NA} = \boldsymbol{N}_o\boldsymbol{N}_1\boldsymbol{A}_o\boldsymbol{A}_1$ appears explicitly in the filter (3.10). Important properties of the robust estimator are evident by direct inspection. For example, assume some diagonal elements of $\boldsymbol{N}_1^{-1}$ or $\boldsymbol{A}_1^{-1}$ in the error models to have resonance peaks, indicating large uncertainty at the corresponding frequencies. Then, the filter will have notches, so the filter gain from the uncertain components of $y(k+m)$ will be low at the relevant frequencies.

The nominal Wiener filter has as a component a whitening filter. The robust estimator has a similar structure. By multiplying $\boldsymbol{\mathcal{R}}$ by the stable common factor D/D, (3.10) will contain $\boldsymbol{\beta}^{-1}\boldsymbol{NA}D$ as right factor. This averaged counterpart of a whitening filter is the inverse of the averaged innovations model (3.2).

The model structure (2.4)–(2.5) was selected to obtain a few simple design equations. Other choices are possible, but lead to various complications. For example, if stochastic polynomials had been introduced in the denominators, no exact analytical solution could have been obtained. Also stability would have been a problem. The use of general left MFD representations, instead of forms with diagonal denominators or common denominators, would have led to a solution involving five coprime factorizations. Such a solution is presented in [28], but it provides less physical insight. It does also exhibit worse numerical behaviour, since algorithms for coprime factorization are numerically sensitive.

Furthermore, equation (3.11) is a *unilateral* Diophantine equation, since $\boldsymbol{Q}$ and $\boldsymbol{L}_*$ appear on the same side of the terms in which they are involved. (When the unknowns appear on opposite sides, the equation is bilateral.) This property is a consequence of our choice of U, T and D as scalar polynomials. Unilateral equations can easily be transformed into a system of linear equations, $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A}$ is a block-Toeplitz matrix. For an example, see Section 4 below.

Robust design also makes the solution less numerically sensitive. Almost common factors of $\det \boldsymbol{\beta}_*$ and UTD with zeros close to $|z| = 1$ would make the solution of (3.11) numerically sensitive. In the presence of model uncertainty, the risk for this is less than in the nominal case, due to the presence of averaged factors in (3.5). The averaged spectral factor $\boldsymbol{\beta}$ will, in general, have its zeros more distant from the unit circle than the nominal spectral factor, given by (3.20) below. This reduces the numerical difficulty of solving both (3.6) and (3.11).

The effect of uncertainties could be represented by equivalent coloured noises. In other words, (see [28]) for every cautious Wiener filter, there exists a system (without uncertainty) for which this estimator is the optimal Wiener filter. It would thus be possible to transform model uncertainties into an equivalent coloured noise,

[4]Stable common factors may exist in (3.10). They could be detected by calculating invariant polynomials of the involved matrices. If such factors have zeros close to the unit circle, it is advisable to cancel them before the filter is implemented. Otherwise, slowly decaying (initial) transients may deteriorate the filtering performance.

and then utilize a Wiener filter for that system. However, we do not recommend this approach for two reasons:

- It is far from trivial to obtain an equivalent noise representation of the uncertainties in the block $\boldsymbol{\mathcal{G}}$ in Figure 1. This is true in particular if the block $\boldsymbol{\mathcal{F}}$ is also uncertain, and if the problem is multivariable.
- It is of advantage from a design point of view to have separate tools which handle different aspects. Error models should represent the effect of modelling uncertainty; noise models should represent disturbances; criterion weighting functions should reflect the priorities of the user. A method which does not distinguish between these aspects will tend to confuse the designer.

The attainable performance improves monotonically with an increasing smoothing lag m. The following result gives the lower bound of the averaged estimation error. This bound can be approached pointwise in the frequency domain for $m < \infty$, by using a criterion filter $\boldsymbol{\mathcal{W}}$ with a high resonance peak.

Corollary 1

The limiting estimator for $m \to \infty$, the *non-realizable cautious Wiener filter*, can be expressed as

$$\lim_{m\to\infty} q^m \boldsymbol{\mathcal{R}} = \frac{1}{T}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A} \quad . \tag{3.14}$$

Its average performance is given by (3.13) with $\boldsymbol{L} = \mathbf{0}$. If $\boldsymbol{\mathcal{W}} = \mathbf{I}_\ell$, the spectral distribution of the lower bound of the estimation error $f(k) - \hat{f}(k|k+m)$ is

$$\lim_{m\to\infty} \mathrm{tr}\bar{E}E\,\boldsymbol{\Phi}_{f-\hat{f}}(e^{i\omega}) = \frac{1}{TDD_*T_*}\mathrm{tr}\{\boldsymbol{S}\tilde{\boldsymbol{C}}\left[\mathbf{I}_n - \tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\right]\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\}\,. \tag{3.15}$$

The bound can be attained at a frequency ω_1 by an estimator with finite smoothing lag, if it is designed using a weighted criterion where

$$U(e^{-i\omega_1}) \approx 0 \tag{3.16}$$

□

Proof: In a similar way as in Appendix A.3 of [8], it is straightforward to show that $\boldsymbol{L} \to \mathbf{0}$ as $m \to \infty$ in (3.11). Thus, (3.11) gives

$$\lim_{m\to\infty} q^m \boldsymbol{Q} = \tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1} \quad . \tag{3.17}$$

The substitution of this expression into (3.10) gives (3.14). The use of $\boldsymbol{L} = \mathbf{0}$, $\tilde{\boldsymbol{V}} = \mathbf{I}_\ell$ and $U = 1$ in the integrand of (3.13) gives (3.15). When $U(e^{-i\omega_1}) \approx 0$, we obtain the same effect on the Diophantine equation (3.11) at the frequency ω_1 as if $\boldsymbol{L} \to \mathbf{0}$: the rightmost term vanishes. Thus, at ω_1, the gain and the phase of the elements of the polynomial matrix $q^m\boldsymbol{Q}$ are approximately equal to those of (3.17) and the estimation error approaches the lower bound (3.15) ■

Remarks. Note that for realizable estimators (m finite) the lower bound (3.15) is only attainable at distinct frequencies ω_i by means of frequency weighting. For frequencies outside the bandwidth of $\boldsymbol{\mathcal{W}}$, the estimator may be severely degraded. The results of Corollary 1 are illustrated at the end of the example in Section 4.

3.4 Analytical expressions for performance evaluation

Theorem 2

Let a nominal estimator $\boldsymbol{\mathcal{R}}_n$ be designed based on a nominal model, with no uncertainties. Applying it, instead of (3.10), on an ensemble of systems results in an increase, as compared to (3.13), of the averaged MSE. The increase is given by

$$\mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*)_n - \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*)_{\min} = \bar{E}\| \boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{R}}_n - \boldsymbol{\mathcal{R}})z^m D^{-1}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\|_2^2 \tag{3.18}$$

where $\boldsymbol{\beta}$ is defined by (3.1)–(3.6) and $\boldsymbol{\mathcal{R}}$ is the robust estimator (3.10) □

Proof. In order to obtain (3.18), the nominal filter $\boldsymbol{\mathcal{R}}_n$ is expressed as $\boldsymbol{\mathcal{R}}+(\boldsymbol{\mathcal{R}}_n - \boldsymbol{\mathcal{R}})$. The optimality of $\boldsymbol{\mathcal{R}}$ implies that any modification gives an orthogonal contribution to the criterion. This, and the use of the averaged innovations model (3.2), gives (3.18). Mixed terms vanish, due to the orthogonality ■

Theorem 3

Let a robust estimator $\boldsymbol{\mathcal{R}}$ be designed by (3.6)–(3.11). When applying it on a system equal to the nominal model, the increased MSE, as compared to the minimum obtainable with a nominal estimator $\boldsymbol{\mathcal{R}}_n$, is

$$\mathrm{tr}E(\varepsilon(k)\varepsilon(k)^*) - \mathrm{tr}E(\varepsilon(k)\varepsilon(k)^*)_n = \|\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{R}} - \boldsymbol{\mathcal{R}}_n)z^m D_o^{-1}\boldsymbol{A}_o^{-1}\boldsymbol{N}_o^{-1}\boldsymbol{\beta}_o\|_2^2 \quad . \tag{3.19}$$

Here, $D_o^{-1}\boldsymbol{A}_o^{-1}\boldsymbol{N}_o^{-1}\boldsymbol{\beta}_o$ is the nominal innovations model and $\boldsymbol{\beta}_o$ is obtained from the nominal spectral factorization

$$\boldsymbol{\beta}_o\boldsymbol{\beta}_{o*} = \boldsymbol{N}_o\boldsymbol{B}_o\boldsymbol{C}_o\boldsymbol{C}_{o*}\boldsymbol{B}_{o*}\boldsymbol{N}_{o*} + D_o\boldsymbol{A}_o\boldsymbol{M}_o\boldsymbol{M}_{o*}\boldsymbol{A}_{o*}D_{o*} \tag{3.20}$$

□

Proof. Analogous to that of Theorem 2, by expressing $\boldsymbol{\mathcal{R}}$ as $\boldsymbol{\mathcal{R}}_n + (\boldsymbol{\mathcal{R}} - \boldsymbol{\mathcal{R}}_n)$ ■

Remarks. The expression (3.18) can be used for arbitrary linear estimators $\boldsymbol{\mathcal{R}}_n$, for example, minimax-designs. The expression (3.19) quantifies the price paid in nominal performance for obtaining a robust design. The averaged innovations model in (3.18), and the nominal innovations model in (3.19), together with the filter $\boldsymbol{\mathcal{W}}$, can be seen as weighting functions.

The largest effect of robust filtering is obtained at moderate and high signal to noise ratios. 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.

4 A DESIGN EXAMPLE

Assume that a scalar signal $u(k)$ is to be estimated. It is described by a first order AR-process without uncertainty

$$u(k) = \frac{1}{1-0.5q^{-1}}e(k) \quad ; \quad Ee(k)^2 = 1 \quad .$$

Thus, $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{S}}/T = 1$, $D_1 = 1$, $D = D_o = 1 - 0.5q^{-1}$, $\hat{\boldsymbol{C}}_o = 1$ and $\hat{\boldsymbol{C}}_1 = 0$. This signal is measured by two transducers ($p = 2$), with nominal models being second order FIR filters. The transducers are modelled by

$$y(k) = (\boldsymbol{B}_o + \boldsymbol{A}_1^{-1}\Delta\boldsymbol{B})u(k) + w(k)$$

with

$$\boldsymbol{B}_o = \begin{pmatrix} B_o^{11} \\ B_o^{21} \end{pmatrix} = \begin{pmatrix} 0.100 + 0.080q^{-2} \\ 1 - 1.4q^{-1} + 0.92q^{-2} \end{pmatrix} \tag{4.1}$$

$$\boldsymbol{A}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 - 0.6q^{-1} \end{pmatrix} \quad ; \quad \Delta\boldsymbol{B} = \begin{pmatrix} \Delta B^{11} \\ \Delta B^{21} \end{pmatrix} = \begin{pmatrix} \Delta b_o^{11} + \Delta b_2^{11} q^{-2} \\ \Delta b_o^{21} + \Delta b_1^{21} q^{-1} + \Delta b_2^{21} q^{-2} \end{pmatrix}$$

Thus, $\boldsymbol{B}_1 = \boldsymbol{A}_o = \mathbf{I}_2$ and $\boldsymbol{A} = \boldsymbol{A}_1$ are used in (2.5). See Figure 2.

Figure 4.1: Model and filter structure in the design example.

In the first transducer B^{11}, there is only a single uncertain parameter. It affects the coefficients Δb_o^{11} and Δb_2^{11} with opposite signs, so they have zero mean, variance r_1^2 and cross-covariance $-r_1^2$. In the second transducer, the stochastic coefficients are assumed mutually uncorrelated, with zero means and equal variance r_2^2. Thus, the auto-covariance matrices are

$$\mathbf{P}_{\Delta\mathbf{B}}^{(11,11)} = r_1^2 \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad ; \quad \mathbf{P}_{\Delta\mathbf{B}}^{(21,21)} = r_2^2\, \mathbf{I}_3 \ . \tag{4.2}$$

The scale factors (standard deviations) of the uncertainties are set to

$$r_1 = 0.02 \ \ ; \ \ r_2 = 0.10 \ . \tag{4.3}$$

Coefficients of ΔB^{11} and ΔB^{21} are assumed mutually uncorrelated. The complete covariance matrix (2.11) then becomes

$$\mathbf{P}_{\Delta\mathbf{B}} = \begin{pmatrix} \mathbf{P}_{\Delta\mathbf{B}}^{(11,11)} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{\Delta\mathbf{B}}^{(21,21)} \end{pmatrix} . \tag{4.4}$$

The measurement noises $w_i(k)$ have variance 0.01. They are white and mutually uncorrelated. Thus, $w(k) = \boldsymbol{M}v(k)$, with

$$\boldsymbol{M} = \boldsymbol{M}_o = 0.1\,\mathbf{I}_2 \quad . \tag{4.5}$$

The goal is now to design a filter ($m = 0$), which estimates $u(k)$ based on the two measurements $y_1(k)$ and $y_2(k)$. Frequency weighting is not used used here ($\boldsymbol{\mathcal{W}} = 1$), but its influence will be illustrated at the end of the example. In (3.4)–(3.5), we obtain

$$\begin{aligned} \tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* &= \quad \bar{E}(\boldsymbol{C}\boldsymbol{C}_*) \quad = 1 \\ \tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_* &= \quad \bar{E}(\boldsymbol{M}\boldsymbol{M}_*) \quad = 0.01\,\mathbf{I}_2 \\ \tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} &= \bar{E}(\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{B}_*) = \hat{\boldsymbol{B}}_o\hat{\boldsymbol{B}}_{o*} + \bar{E}(\Delta\boldsymbol{B}\Delta\boldsymbol{B}_*) \quad . \end{aligned} \tag{4.6}$$

The expression (3.7) or (2.10) gives

$$\begin{aligned} \bar{E}(\Delta B^{11}\Delta B^{11}_*) = \varphi^T\mathbf{P}^{(11,11)}_{\Delta\mathbf{B}}\varphi^T_* &= (\ 1\ q^{-1}\ q^{-2})r_1^2 \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ q \\ q^2 \end{pmatrix} \\ &= \ r_1^2(\ -q^2\ \ +2\ \ -q^{-2}\ ) \quad . \end{aligned} \tag{4.7}$$

Note that $\bar{E}(\Delta B^{11}\Delta B^{11}_*)$ has zeros at $z = 1$ and at $z = -1$. Thus, the static gain and the high-frequency gain are assumed to be exactly known. Furthermore,

$$\bar{E}(\Delta B^{21}\Delta B^{21}_*) = \varphi^T\mathbf{P}^{(21,21)}_{\Delta\mathbf{B}}\varphi^T_* = 3r_2^2 \quad . \tag{4.8}$$

The spectral factorization (3.6) has dimension $p|p = 2|2$. Using (4.6), it reduces to

$$\begin{aligned} \boldsymbol{\beta}\boldsymbol{\beta}_* &= \tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} + D\boldsymbol{A}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_*\boldsymbol{A}_*D_* \\ &= \boldsymbol{A}_1\boldsymbol{B}_o\boldsymbol{B}_{o*}\boldsymbol{A}_{1*} + \bar{E}(\Delta\boldsymbol{B}\Delta\boldsymbol{B}_*) + 0.01D\boldsymbol{A}_1\boldsymbol{A}_{1*}D_* \end{aligned}$$

where

$$\bar{E}(\Delta\boldsymbol{B}\Delta\boldsymbol{B}_*) = \begin{pmatrix} r_1^2(-q^2+2-q^{-2}) & 0 \\ 0 & 3r_2^2 \end{pmatrix} .$$

By using the Newton-based algorithm described in [17], a stable averaged spectral factor, with $\boldsymbol{\beta}(0)$ nonsingular, was found to be

$$\boldsymbol{\beta} =$$

$$\begin{pmatrix} 0.1339 - 0.01867q^{-1} + 0.01622q^{-2} & 0.07862 - 0.01488q^{-1} + 0.06905q^{-2} \\ -0.1474q^{-1} + 0.2908q^{-2} - 0.1325q^{-3} & 1.1585 - 2.0327q^{-1} + 1.6219q^{-2} - 0.4765q^{-3} \end{pmatrix} . \tag{4.9}$$

In the Diophantine equation (3.11), we use $m = 0$, $\tilde{\boldsymbol{V}} = \boldsymbol{V} = 1$, $\boldsymbol{S} = 1$, $\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* = 1$, $\boldsymbol{N}_* = \mathbf{I}_2$, $U = 1$ and $T = 1$. Equation (3.11) thus reduces to

$$\boldsymbol{B}_{o*}\boldsymbol{A}_{1*} = \boldsymbol{Q}\boldsymbol{\beta}_* + \boldsymbol{L}_*qD\,\mathbf{I}_2 \quad . \tag{4.10}$$

The degrees (3.12) are $nQ = 0, nL_* = 2$. By expressing the polynomial matrices as matrix polynomials, equation (4.10) becomes

$$
\begin{aligned}
(\mathbf{B}_0^* + \mathbf{B}_1^* q + \mathbf{B}_2^* q^2)(\mathbf{I}_2 + \mathbf{A}_1^* q) &= \\
= \mathbf{Q}_0(\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1^* q + \boldsymbol{\beta}_2^* q^2 + \boldsymbol{\beta}_3^* q^3) + (\mathbf{L}_0^* + \mathbf{L}_1^* q + \mathbf{L}_2^* q^2)(-0.5 + q)\mathbf{I}_2 &\quad .
\end{aligned}
$$

Transpose this equation, and note that since the coefficient matrices are real-valued, $\mathbf{P}_i^{*T} = \mathbf{P}_i$. By equating the two sides for each power of q separately, a linear system of eight equations, in block-Toepliz form, is obtained:

$$
\begin{pmatrix} \mathbf{B}_0 \\ \mathbf{B}_1 + \mathbf{A}_1\mathbf{B}_0 \\ \mathbf{B}_2 + \mathbf{A}_1\mathbf{B}_1 \\ \mathbf{A}_1\mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_0 & -0.5\mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\beta}_1 & \mathbf{I}_2 & -0.5\mathbf{I}_2 & \mathbf{0} \\ \boldsymbol{\beta}_2 & \mathbf{0} & \mathbf{I}_2 & -0.5\mathbf{I}_2 \\ \boldsymbol{\beta}_3 & \mathbf{0} & \mathbf{0} & \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{Q}_0^T \\ \mathbf{L}_0 \\ \mathbf{L}_1 \\ \mathbf{L}_2 \end{pmatrix} .
$$

With numerical values from (4.1) and (4.9), we obtain

$$
\begin{pmatrix} .1 \\ 1 \\ 0 \\ -2 \\ .08 \\ 1.76 \\ 0 \\ -.552 \end{pmatrix} = \begin{pmatrix} .1339 & .07862 & -.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.1585 & 0 & -.5 & 0 & 0 & 0 & 0 \\ -.01867 & -.01488 & 1 & 0 & -.5 & 0 & 0 & 0 \\ -.1474 & -2.0327 & 0 & 1 & 0 & -.5 & 0 & 0 \\ .01622 & .06905 & 0 & 0 & 1 & 0 & -.5 & 0 \\ .2908 & 1.6219 & 0 & 0 & 0 & 1 & 0 & -.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -.1325 & -.4765 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Q_o^{11} \\ Q_o^{12} \\ \ell_o^{11} \\ \ell_o^{12} \\ \ell_1^{11} \\ \ell_1^{12} \\ \ell_2^{11} \\ \ell_2^{12} \end{pmatrix}
$$

The solution is

$$
\begin{aligned}
\boldsymbol{Q} &= (\ 0.4005 \quad 0.7746\ ) \\
\boldsymbol{L}_* &= (\ 0.0290 + 0.0200q \quad -0.2053 + 0.3224q - 0.1299q^2\ ) \quad .
\end{aligned}
\tag{4.11}
$$

Finally, the robust estimator (3.10) becomes

$$
\boldsymbol{\mathcal{R}} = \boldsymbol{Q}\boldsymbol{\beta}^{-1}\boldsymbol{A}_1 = \frac{1}{R_r}(K_r^{11} \ K_r^{12}) \tag{4.12}
$$

where the monic denominator is $R_r(q^{-1}) = \det \boldsymbol{\beta}(q^{-1}) / \det \boldsymbol{\beta}_0$. We obtain

$$
\begin{aligned}
K_r^{11} &= 2.9922 - 4.5138q^{-1} + 2.7365q^{-2} - 0.5687q^{-3} \\
K_r^{12} &= 0.4655 - 0.3341q^{-1} - 0.06445q^{-2} + 0.05841q^{-3} \\
R_r &= 1 - 1.8193q^{-1} + 1.6043q^{-2} - 0.6584q^{-3} + 0.08479q^{-4} + 0.009182q^{-5} \quad .
\end{aligned}
$$

A corresponding nominal estimator (with no uncertainty assumed) is given by $(K_n^{11} \ K_n^{12})/R_n$, with

$$
\begin{aligned}
K_n^{11} &= 0.7419 - 1.0943q^{-1} + 0.3617q^{-2} \\
K_n^{12} &= 0.8792 - 0.3767q^{-1} - 0.03145q^{-2} \\
R_n &= 1 - 1.7786q^{-1} + 1.4269q^{-2} - 0.3938q^{-3} \quad .
\end{aligned}
$$

Figure 4.2: Bode magnitude plots for the nominal models of the two transducers $B^{11}(q^{-1})$ and $B^{21}(q^{-1})$ (solid). The dotted curves show fifteen realizations of possible true systems. Magnitude plots for the gains from $y_1(k)$ (upper) and $y_2(k)$ (lower), are shown for the robust estimator (dash-dotted) and for the nominal Wiener filter design (dashed).

Figure 3 shows the Bode magnitude plots for the robust and nominal estimators. Also shown is the nominal transducer model, and fifteen randomly chosen systems. These were generated by using $\boldsymbol{B} = \boldsymbol{B}_o + \boldsymbol{A}_1^{-1}\Delta\boldsymbol{B}$, with covariance matrix (4.4) and Gaussian distributions. The channel B^{11} has its uncertainty concentrated around the notch, while B^{21} is uncertain mainly at low frequencies.

The gains of the nominal estimator (dashed curves) are determined exclusively by the nominal signal to noise ratios. The gains of the robust estimator (dash-dotted) are determined by the balance between noise levels and model uncertainties in the two channels. For example, the robust filter "knows" that channel 1 is well known, as compared to channel 2. Consequently, a higher gain is used from $y_1(k)$ as compared to the nominal case, and a lower gain from $y_2(k)$. The difference, as compared to nominal design, is largest at low frequencies. There, the dynamics of channel 1 is almost perfectly known, while channel 2 is very uncertain. The nominal filter gain in channel 2 is an approximate inverse of the nominal transducer, In contrast to the nominal filter, the robust filter has hardly any peak at the (uncertain) notch around $\omega = 0.7$. It utilizes channel 1 more at this frequency.

Figure 4 shows Bode magnitude plots of transfer functions from $u(k)$ to $\hat{u}(k|k)$. Since the noise levels are rather low, the nominal estimator performs an almost complete inversion of the nominal transducers. The robust estimator is somewhat more cautious, but it also accomplishes a rather good inversion. However, it utilizes the two measurement signals differently than the nominal estimator.

Figure 5 shows the mean square estimation error, when *one* of the uncertain pa-

Figure 4.3: Bode magnitude plots for the transfer function from $u(k)$ to $\hat{u}(k|k)$ for the nominal system, with robust and nominal estimators.

Figure 4.4: MSE for robust (dashed) and nominal filter (solid). One of the four uncertain parameters is varied, while the others are held at nominal values. Also shown is the variance of $u(k)$ (upper dotted). The lower dotted curve is the lower bound, achievable with knowledge of the true parameter values. Rings (o) indicate the two standard deviation limits of each parameter.

rameters

$$\Delta b_o^{11} = -\Delta b_2^{11} \triangleq r \quad ; \quad \Delta b_o^{21} \quad ; \quad \Delta b_1^{21} \quad ; \quad \Delta b_2^{21}$$

is varied, while the others are zero. The four parameters above span the set of assumed true systems, the extended design model. On average, over the four uncorrelated stochastic coefficients, the MSE is 0.32 for the robust filter and 0.90 for the nominal design. Note, however, that when r is varied, the robust design (dashed) is actually slightly *more* sensitive than the nominal design. This is a price paid for reducing the sensitivity in the other dimensions[5].

[5]It is natural that the robust filter has a somewhat increased sensitivity to model errors in

The robust estimator does, of course, not perform as well as the nominal one in the nominal case. This is mainly due to its somewhat lower gain from $u(k)$ to $\hat{u}(k|k)$, see Figure 4. It is evident from Figure 5 that this performance loss is very small, as compared to the improvement in non-ideal situations[6].

As an alternative, we tried to investigate minimax designs, i.e. worst case $\mathcal{H}_2$-designs, assuming rectangular parameter distributions. This turned out to be prohibitively difficult, since no point where $\min_{\mathcal{R}} \max_{\Delta B} \varepsilon(k)^2 = \max_{\Delta B} \min_{\mathcal{R}} \varepsilon(k)^2$ could be found. These difficulties were in marked contrast to the ease of designing a cautious Wiener filter based on the averaged $\mathcal{H}_2$-criterion.

Let us finally illustrate the effect of using a frequency-dependent weighting function in the criterion (2.3). Assume that the performance at frequencies close to $\omega = 0.9$ is of a particular importance. The choice

$$\boldsymbol{\mathcal{W}} = \frac{V}{U} = \frac{1}{1 - 1.2184q^{-1} + 0.9604q^{-2}} \tag{4.13}$$

with a high resonance peak ($|z| = 0.98$) at $\omega = 0.9$ should, according to Corollary 1, result in a performance, at that frequency, close to the lower bound. Figure 4.5 confirms this. However, performance is substantially degraded at higher frequencies, where estimation accuracy is not emphasized.

Figure 4.5: The average spectral density of the estimation error $u(k) - \hat{u}(k|k)$. Shown in the plot are the lower bound, according to (3.15), (dash-dotted) and the spectral densities obtained with (dashed) and without (solid) frequency weighting.

channel 1; it has higher gain in that channel. This result is due to the much larger uncertainty in channel 2 at most frequencies. A designer worried about this effect could simply increase the value of the standard deviation r_1 used in the design.

[6] It can also be noted that it is of advantage to use both channels. The minimal MSE, for channels equal to the nominal models, is 0.07 if both channels are used. It is 0.59 if only channel 1 is used and 0.11 if only channel 2 is used. The *average* MSE of the robust filter (0.32) is in fact lower than the *nominal* MSE for an estimator which uses only channel 1 (0.59).

5 ROBUST FEEDFORWARD CONTROL

A class of feedforward control problems turns out to be dual to the filtering problems discussed in Section 2. We include a brief separate discussion of them, since it offers important engineering insights.
Feedforward compensation does not affect the classical sensitivity function. However, the *effect* of an $x\%$ model deviation at a particular frequency, on e.g. the step response, will very much depend on the (nominal) magnitude of the transfer function at that frequency. As the gain at a particular frequency is increased by a feedforward link, model errors at that frequency become more and more noticeable. Therefore, it is of value to take model uncertainty into account explicitly in the feedforward design.

To stress the duality to filtering problems, the output of an uncertain but stable model will be described by

$$y(k) = (\boldsymbol{\mathcal{G}}_o^T + \Delta\boldsymbol{\mathcal{G}}^T)u(k) + \boldsymbol{\mathcal{D}}^T w(k) = (\boldsymbol{B}_o^T \boldsymbol{A}_o^{-T} + \Delta\boldsymbol{B}^T \boldsymbol{B}_1^T \boldsymbol{A}_1^{-T})u(k) + \frac{1}{T}\boldsymbol{S}^T w(k) \tag{5.1}$$

where $\boldsymbol{A}^{-T}$ denotes inverse and transpose. The rational and polynomial matrices above have properties as outlined in Section 2. The matrices $\boldsymbol{\mathcal{G}}^T$ and $\boldsymbol{\mathcal{D}}^T$ may contain delays. Based on possibly delayed or advanced measurements of

$$w(k) = (\boldsymbol{\mathcal{W}}_o^T + \Delta\boldsymbol{\mathcal{W}}^T)v(k) = \left(\frac{1}{U_o}\boldsymbol{V}_o^T + \frac{1}{U_1}\Delta\boldsymbol{V}^T\boldsymbol{V}_1^T\right)v(k) \quad ; \quad E(v(k)^T v(k)) = \mathbf{I}_\ell \tag{5.2}$$

a stable controller

$$u(k) = -\boldsymbol{\mathcal{R}}^T w(k+m) \tag{5.3}$$

is to be designed to minimize the averaged $\mathcal{H}_2$-norm of the transfer function from v to $(z_1^T \; z_2^T)^T = ((\boldsymbol{\mathcal{F}}^T y)^T \; (\boldsymbol{\mathcal{H}}^T u)^T)^T$:

$$J' = \bar{E}\left\| \begin{array}{c} \boldsymbol{\mathcal{F}}^T(\boldsymbol{\mathcal{D}}^T - \boldsymbol{\mathcal{G}}^T\boldsymbol{\mathcal{R}}^T z^m)\boldsymbol{\mathcal{W}}^T \\ -\boldsymbol{\mathcal{H}}^T\boldsymbol{\mathcal{R}}^T z^m \boldsymbol{\mathcal{W}}^T \end{array} \right\|_2^2 \quad . \tag{5.4}$$

The control weighting $\boldsymbol{\mathcal{H}}^T = \boldsymbol{M}^T\boldsymbol{N}^{-T}$ and the output weighting $\boldsymbol{\mathcal{F}}^T = \boldsymbol{C}^T/D$ are normally specified by the designer and exactly known, i.e. $\Delta\boldsymbol{\mathcal{H}}^T = 0$, $\Delta\boldsymbol{\mathcal{F}}^T = 0$.

The problem formulation (5.1)–(5.4) may represent a disturbance measurement feedforward design. When $m > 0$, the disturbance $w(k)$ can be measured before it affects the system via $\boldsymbol{\mathcal{D}}^T$. (Such a situation could, equivalently, be described by a delay q^{-m} in $\boldsymbol{\mathcal{D}}^T$.) The formulation also covers reference feedforward problems, decoupling and model matching. Then, $w(k)$ is a command signal and $\boldsymbol{\mathcal{W}}^T v(k)$ is a (possibly uncertain) stochastic model, describing its second order properties. A servo filter $\boldsymbol{\mathcal{R}}^T$ is to be designed, so that the output $-\boldsymbol{\mathcal{G}}^T u(k)$ optimally follows the response model $\boldsymbol{\mathcal{D}}^T w(k)$[7]. In decoupling problems, $\boldsymbol{\mathcal{D}}^T$ is diagonal.

[7] The corresponding nominal result (without uncertainty) was discussed in [34] and in [5]. See also [26], where common denominator forms were used. The robust controller for SISO systems was presented in [35]. In [6], the applicability of the averaged $\mathcal{H}_2$-criterion is demonstrated for somewhat more general class of open-loop type problems.

Figure 5.1: A block diagram which is dual to the one in Figure 1. Arrows are reversed, summation points and node points are interchanged and transfer functions are transposed. The corresponding robust control problem is to design $\boldsymbol{\mathcal{R}}^T$ to minimize the average, over the class of models, of the $\mathcal{H}_2$-norm of the transfer function from v to $z = (z_1^T \; z_2^T)^T$.

The duality of feedforward control to the previously discussed filtering problem, has been described in [5] for the nominal case. The corresponding result in the presence of model errors is given below.

Theorem 4. A feedforward filter, which solves the robust control problem (5.1)–(5.4) under assumptions **A1**–**A4** is given by the transpose of (3.10), or

$$u(k) = -\boldsymbol{A}^T \boldsymbol{N}^T \boldsymbol{\beta}^{-T} \boldsymbol{Q}^T \tilde{\boldsymbol{V}}^{-T} \frac{1}{T} w(k+m) \tag{5.5}$$

where $\boldsymbol{\beta}$, $\boldsymbol{Q}$ and $\tilde{\boldsymbol{V}}$ are given by (3.6), (3.11) and (3.9), respectively □

Proof: The averaged $\mathcal{H}_2$-norm is invariant under transposition. Thus, it fulfills the basic requirement of [5]. By extending the discussion in [5] to uncertain models the result is obtained. ■

The uncertainty $\Delta\boldsymbol{\mathcal{W}}^T$ of the disturbance or reference model (5.2) enters via the spectral factorization (3.9). Transposition of (3.9) gives

$$\tilde{\boldsymbol{V}}^T \tilde{\boldsymbol{V}}_*^T = U_1 \boldsymbol{V}_o^T \boldsymbol{V}_{o*}^T U_{1*} + U_o \boldsymbol{V}_1^T \bar{E}(\Delta \boldsymbol{V}^T \Delta \boldsymbol{V}_*^T) \boldsymbol{V}_{1*}^T U_{o*} \quad . \tag{5.6}$$

This is the kind of left spectral factorization encountered when two noise sources are described by one innovations model. In fact, the uncertainty $\Delta\boldsymbol{\mathcal{W}}^T$ has exactly the same effect on the controller design as would a measurement noise on $w(k)$, with spectral density $\boldsymbol{V}_1^T \bar{E}(\Delta \boldsymbol{V}^T \Delta \boldsymbol{V}_*^T) \boldsymbol{V}_{1*}^T / U_1 U_{1*}$. We do not need to solve a right spectral factorization (3.9) in this problem. The left spectral factorization (5.6) can be solved instead.

As in the dual filtering case, uncertainty in the direct feedthrough, or response model $\boldsymbol{\mathcal{D}}^T = \boldsymbol{S}^T / T$ does not affect the optimal solution, if it is independent of the uncertainties in $\boldsymbol{\mathcal{G}}^T$.

6 CONCLUSIONS

A method for designing robust filters and feedforward controllers, based on imperfectly known linear models, has been presented. Modelling errors were described by sets of models, parameterized by random variables with known covariances. A robust design was obtained by minimizing the $\mathcal{H}_2$-norm, averaged with respect to the assumed model errors. The estimator minimizes this criterion by balancing model uncertainties against noise properties, at different frequencies and in different measurement channels. When using robust filtering, the greatest sensitivity reduction is obtained at moderate and high signal to noise ratios. Dually, the largest impact of robust control is obtained for designs with low input penalties.

One variant of the discussed filtering problems is to explicitly define a part of the measurement vector as being a noise-free signal. This signal could e.g. represent known inputs to the system. Such a formulation is also of use in the optimization of decision feedback equalizers for digital communications [32], [33].

There exist efficient numerical algorithms based on a polynomial equations approach. (We have implemented them as MATLAB .m-files, and the code is available upon request.) For multivariable problems of high order and high signal vector dimension p, Riccati-based algorithms do, however, perform better numerically. For dimensions of the measurement vector up to, say, four, algorithms based directly on polynomial manipulation can in general be used safely. For higher dimensions, we recommend analytical solutions to be obtained by an input-output approach, to gain engineering insight, but algorithms e.g. for spectral factorization to be based on state space formulations, cf. [21]

Appendix

A Proofs of Lemmas

Proof of Lemma 1

Let $\boldsymbol{G} = [\boldsymbol{G}_1^T \dots \boldsymbol{G}_n^T]^T$, where $\boldsymbol{G}_1 \dots \boldsymbol{G}_n$ represent the n polynomial row vectors of the $n|m$ matrix. Then, with $\boldsymbol{H}$ of dimension $m|m$, the ij-th element of $\bar{E}(\boldsymbol{GHG_*})$ can be expressed as

$$[\bar{E}(\boldsymbol{GHG}_*)]_{ij} = \bar{E}(\boldsymbol{G}_i\boldsymbol{H}\boldsymbol{G}_{j*}) = \bar{E}(\mathrm{tr}\boldsymbol{G}_{j*}\boldsymbol{G}_i\boldsymbol{H}) = \mathrm{tr}\bar{E}(\boldsymbol{G}_{j*}\boldsymbol{G}_i)\bar{E}(\boldsymbol{H}) \quad .$$

In the last equality, we used the fact that all elements of $\boldsymbol{G}_{j*}\boldsymbol{G}_i$ are independent of all those of $\boldsymbol{H}$. We also have that

$$\mathrm{tr}\bar{E}(\boldsymbol{G}_{j*}\boldsymbol{G}_i)\bar{E}(\boldsymbol{H}) = \bar{E}(\boldsymbol{G}_i\bar{E}(\boldsymbol{H})\boldsymbol{G}_{j*}) = [\bar{E}(\boldsymbol{G}\bar{E}(\boldsymbol{H})\boldsymbol{G}_*)]_{ij}$$

which proves (3.3), since $\bar{E}(\cdot)$ operates on all elements of $\boldsymbol{GHG_*}$ ∎

Proof of Lemma 2

If the coefficients of the elements of a polynomial matrix $\Delta\boldsymbol{P}(q^{-1})$ are stochastic variables, then so are the coefficients of the elements in $\Delta\boldsymbol{P}\Delta\boldsymbol{P}_*$. The coefficients of the polynomial elements in $\Delta\boldsymbol{C}$ and $\Delta\boldsymbol{B}$ are independent and so are the coefficients of the elements in $\Delta\boldsymbol{C}\Delta\boldsymbol{C}_*$ and $\Delta\boldsymbol{B}\Delta\boldsymbol{B}_*$. By defining $\boldsymbol{CC}_*$ as $\boldsymbol{H}$ and $\Delta\boldsymbol{B}$ as $\boldsymbol{G}$ in Lemma 1 and using (3.4), the right hand side of (3.1) becomes

$$\begin{aligned}
&\boldsymbol{N}\bar{E}[\boldsymbol{BCC}_*\boldsymbol{B}_*]\boldsymbol{N}_* + D\boldsymbol{A}\bar{E}(\boldsymbol{MM}_*)\boldsymbol{A}_*D_* \; = \\
&= \; \boldsymbol{N}\bar{E}[\boldsymbol{B}\bar{E}(\boldsymbol{CC}_*)\boldsymbol{B}_*]\boldsymbol{N}_* + D\boldsymbol{A}\bar{E}(\boldsymbol{MM}_*)\boldsymbol{A}_*D_* \\
&= \; \boldsymbol{N}\bar{E}[\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{B}_*]\boldsymbol{N}_* + D\boldsymbol{A}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_*\boldsymbol{A}_*D_* \quad .
\end{aligned}$$

By once more utilizing (3.4), we obtain (3.6) ∎

B Calculation of averaged polynomial matrices

Consider the matrices discussed in Lemma 1. Let the polynomial matrices $\boldsymbol{G} = \Delta\boldsymbol{P}$ and $\boldsymbol{H}$ be of dimensions $n|m$ and $m|m$, respectively. Denote the i-th row of $\Delta\boldsymbol{P}$ by

$$\Delta\boldsymbol{P}_i = [\Delta P^{i1} \dots \Delta P^{im}]$$

where ΔP^{ij} are polynomials with stochastic coefficients. If $\boldsymbol{H}$ is assumed deterministic, the ij-th element of $\bar{E}(\Delta\boldsymbol{PH}\Delta\boldsymbol{P}_*)$ can be written as

$$\begin{aligned}
\bar{E}[\Delta\boldsymbol{PH}\Delta\boldsymbol{P}_*]_{ij} \; &= \; \mathrm{tr}\bar{E}(\Delta\boldsymbol{P}_{j*}\Delta\boldsymbol{P}_i)\boldsymbol{H} \; = \\
&= \; \mathrm{tr}\boldsymbol{H}\bar{E}[\Delta P_*^{j1} \dots \Delta P_*^{jm}]^T[\Delta P^{i1} \dots \Delta P^{im}] \\
&= \; \mathrm{tr}\boldsymbol{H}\bar{E}\begin{bmatrix} \Delta P^{i1}\Delta P_*^{j1} & \dots & \Delta P^{im}\Delta P_*^{j1} \\ \vdots & \ddots & \vdots \\ \Delta P^{i1}\Delta P_*^{jm} & \dots & \Delta P^{im}\Delta P_*^{jm} \end{bmatrix}
\end{aligned}$$

By using (2.10), we readily obtain (3.7) ∎

C Proof of Theorem 1

A technique for constructive derivation of polynomial design equations for Wiener filters was presented in [1]. This method is utilized here to minimize (2.3). The estimation error is given by

$$\varepsilon(k) = \boldsymbol{\mathcal{W}}\,(f(k) - \hat{f}(k|k+m)) \quad . \tag{C.1}$$

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

$$\hat{d}(k) = \boldsymbol{\mathcal{R}}\,y(k+m) + \tilde{\nu}(k) \quad ; \quad \tilde{\nu}(k) = \boldsymbol{\mathcal{M}}\,y(k+m) \quad . \tag{C.2}$$

Here, $\boldsymbol{\mathcal{M}}$ is a rational, stable and causal, but otherwise arbitrary transfer function. Define the weighted variation

$$\nu(k) \triangleq \boldsymbol{\mathcal{W}}\,\tilde{\nu}(k) = \frac{1}{U}\boldsymbol{V}\boldsymbol{\mathcal{M}}\,q^m \left(\boldsymbol{A}^{-1}\boldsymbol{B}\frac{1}{D}\boldsymbol{C}e(k) + \boldsymbol{N}^{-1}\boldsymbol{M}v(k)\right) \quad .$$

Optimality of (2.2) is obtained if no perturbation $\nu(k)$ will improve the average estimator performance. This occurs if and only if the error $\varepsilon(k)$ is orthogonal to any admissible weighted estimator variation $\nu(k)$. In other words,

$$\mathrm{tr}\bar{E}E(\varepsilon(k)\nu(k)^*) = \mathrm{tr}\bar{E}E(\varepsilon(k)^*\nu(k)) = 0 \quad .$$

Then, the perturbed criterion value becomes

$$\begin{aligned}
\bar{J} &= \mathrm{tr}\bar{E}E[\boldsymbol{\mathcal{W}}\,(f(k) - \hat{d}(k))][\boldsymbol{\mathcal{W}}\,(f(k) - \hat{d}(k))]^* \\
&= \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^* - \varepsilon(k)\nu(k)^* - \varepsilon(k)^*\nu(k) + \nu(k)\nu(k)^*) \\
&= \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^* + \nu(k)\nu(k)^*) \quad .
\end{aligned} \tag{C.3}$$

This expression is evidently minimized by $\nu(k) = 0$.

Since all transfer functions in (2.4) are assumed stable, both $\varepsilon(k)$ and $\nu(k)$ are stationary. Parseval's formula may then be used in order to express $\mathrm{tr}\bar{E}E\varepsilon(k)\nu(k)^*$ as

$$\begin{aligned}
&\mathrm{tr}\bar{E}E\frac{1}{U}\boldsymbol{V}\left\{\left(\frac{1}{T}\boldsymbol{S} - q^m\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\frac{1}{D}\boldsymbol{C}e(k) - q^m\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k)\right\} \\
&\qquad \left\{\frac{1}{U}\boldsymbol{V}\boldsymbol{\mathcal{M}}q^m\left(\boldsymbol{A}^{-1}\boldsymbol{B}\frac{1}{D}\boldsymbol{C}e(k) + \boldsymbol{N}^{-1}\boldsymbol{M}v(k)\right)\right\}^* \\
= \;&\mathrm{tr}\bar{E}\frac{1}{2\pi j}\oint_{|z|=1}\frac{1}{UU_*DD_*}\boldsymbol{V}\left\{z^{-m}\frac{1}{T}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1}\right. \\
&\left.-\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\left(\boldsymbol{N}\boldsymbol{B}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* + D\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_*D_*\right)\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}\right\}\boldsymbol{\mathcal{M}}_*\boldsymbol{V}_*\frac{dz}{z}
\end{aligned} \tag{C.4}$$

Note that $\boldsymbol{A}$ and $\boldsymbol{N}$ commute, since they are diagonal. We are allowed to move the expectation $\bar{E}$ inside the integration, since, for any particular realization of the elements of $\Delta\boldsymbol{C}$, $\Delta\boldsymbol{B}$, $\Delta\boldsymbol{M}$ and $\Delta\boldsymbol{V}$, all elements of the integrand are Riemann integrable on the unit circle, see e.g. [16], Theorem 3.8. The use of the trace rotation, $\mathrm{tr}\boldsymbol{V}\{\cdots\}\boldsymbol{\mathcal{M}}_*\boldsymbol{V}_* = \mathrm{tr}\boldsymbol{V}_*\boldsymbol{V}\{\cdots\}\boldsymbol{\mathcal{M}}_*$, the spectral factorizations (3.1) and (3.9) and the Assumption **A3**, leads to

$$\mathrm{tr}\bar{E}E\varepsilon(k)\nu(k)^* \;=$$

$$\mathrm{tr}\frac{1}{2\pi j}\oint\frac{1}{UU_*DD_*}\tilde{\boldsymbol{V}}_*\tilde{\boldsymbol{V}}\left\{z^{-m}\bar{E}(\frac{1}{T}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*)\boldsymbol{A}_*^{-1} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}\right\}\boldsymbol{\mathcal{M}}_*\frac{dz}{z}\,. \tag{C.5}$$

From (C.5) it is now easy to see that uncertainties in

$$\boldsymbol{\mathcal{D}} \;=\; \frac{1}{T_o}\boldsymbol{S}_o + \frac{1}{T_1}\boldsymbol{S}_1\Delta\boldsymbol{S} \;=\; \frac{1}{T_oT_1}\left(\hat{\boldsymbol{S}}_o + \hat{\boldsymbol{S}}_1\Delta\boldsymbol{S}\right) \;\triangleq\; \frac{1}{T}\boldsymbol{S} \tag{C.6}$$

independent of $\Delta\boldsymbol{C}$, $\Delta\boldsymbol{B}$ and $\Delta\boldsymbol{V}$, will not affect the filter design since, using **A1**

$$\bar{E}(\frac{1}{T}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*) = \bar{E}(\frac{1}{T}\boldsymbol{S})\bar{E}(\boldsymbol{C}\boldsymbol{C}_*)\bar{E}(\boldsymbol{B}_*) = \frac{1}{T}_o\boldsymbol{S}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*} \tag{C.7}$$

(The minimal criterion value will, however, be affected, which is evident form (C.12), below.) In the sequel we thus use $T_o = T$ and $\boldsymbol{S}_o = \boldsymbol{S}$.

Using (C.7) in (C.5) and extracting T to the left and $\boldsymbol{A}_*^{-1}$ to the right now gives

$$\mathrm{tr}\bar{E}E\varepsilon(k)\nu(k)^* \;=$$

$$\mathrm{tr}\frac{1}{2\pi j}\oint\frac{1}{UU_*TDD_*}\tilde{\boldsymbol{V}}_*\tilde{\boldsymbol{V}}\left\{z^{-m}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*} - T\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{N}_*^{-1}\right\}\boldsymbol{A}_*^{-1}\boldsymbol{\mathcal{M}}_*\frac{dz}{z}\,. \tag{C.8}$$

To make (C.8) zero, all poles inside $|z| = 1$ are eliminated. This is achieved if, in every element of the integrand, all such poles are cancelled by zeros. We first cancel what can be cancelled by means of $\boldsymbol{\mathcal{R}}$ directly. Thus,

$$\boldsymbol{\mathcal{R}} = \frac{1}{T}\tilde{\boldsymbol{V}}^{-1}\boldsymbol{Q}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A} \tag{C.9}$$

where $\boldsymbol{Q}(z^{-1})$ is undetermined. Inserting (C.9) into (C.8) gives $\mathrm{tr}\bar{E}E\varepsilon(k)\nu(k)^*$ as

$$\mathrm{tr}\frac{1}{2\pi j}\oint\frac{1}{UU_*TDD_*}\tilde{\boldsymbol{V}}_*\left\{z^{-m}\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_* - \boldsymbol{Q}\boldsymbol{\beta}_*\right\}\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}\boldsymbol{\mathcal{M}}_*\frac{dz}{z} \quad . \tag{C.10}$$

Now, U, T, D, $\boldsymbol{A}$, $\boldsymbol{N}$ are all stable, so they have zeros only inside $|z| = 1$. The poles of $\boldsymbol{\mathcal{M}}$ are inside $|z| = 1$. This means that U_*^{-1}, D_*^{-1}, $\boldsymbol{A}_*^{-1}$, $\boldsymbol{N}_*^{-1}$ and $\boldsymbol{\mathcal{M}}_*$ have all their poles outside $|z| = 1$. No poles will thus exist inside $|z| = 1$ in (C.10), if and only if

$$z^{-m}\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_* - \boldsymbol{Q}\boldsymbol{\beta}_* = z\boldsymbol{L}_*UTD\,\mathbf{I}_p \tag{C.11}$$

for some polynomial matrix $\boldsymbol{L}_*(z)$. This is equation (3.11), if q is substituted for z. The filter (C.9) coincides with (3.10). Necessity follows because choices of $\boldsymbol{\mathcal{R}}$ other that (C.9) correspond to $\nu(k) \neq 0$ in (C.3).

Unique solvability of (3.11) is demonstrated as follows. The Diophantine equation will always have one or several solutions, since the invariant polynomials of $UTD\,\mathbf{I}_p$ are all stable, while those of $\boldsymbol{\beta}_*$ are all unstable. Thus, there exist no common invariant factors. Let $(\boldsymbol{Q}_0, \boldsymbol{L}_{0*})$ be one solution pair. Every solution to (3.11) can then be expressed as

$$(\boldsymbol{Q}, \boldsymbol{L}_*) = (\boldsymbol{Q}_0 - \boldsymbol{X} qUTD\, \mathbf{I}_p\ ,\ \boldsymbol{L}_{0*} + \boldsymbol{X}\boldsymbol{\beta}_*)$$

where the polynomial matrix $\boldsymbol{X}(q, q^{-1})$ is undetermined. Now, $\boldsymbol{Q}$ is required to be causal, so it can not have any positive powers of q as arguments, while $\boldsymbol{L}_*$ must contain no negative powers of q, to assure optimality. Thus, $\boldsymbol{X}(q, q^{-1}) = 0$ is the only choice. We conclude that the solution to (3.11) is unique.

The degrees (3.12) are determined by the requirement that the maximum powers of q^{-1} and q are covered on both sides of (3.11). For details, see [1] or [3].

The minimal average estimation error, $J_{min} = \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*)_{min}$, is obtained as follows. First insert (3.10) into the criterion (2.3), use Parseval's formula, take expectation and use (3.1), (3.9) and (3.5) in this order. Then we obtain

$$\begin{aligned} J_{min} \;=\; & \mathrm{tr}\frac{1}{2\pi j}\oint_{|z|=1} \frac{1}{UU_*TT_*DD_*}\tilde{\boldsymbol{V}}_*\tilde{\boldsymbol{V}}\Big\{\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{S}_* \\ & - \boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{A}_*^{-1}\boldsymbol{\mathcal{R}}_*T_*z^{-m} - z^m T\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{S}_* \\ & + T\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}\boldsymbol{\mathcal{R}}_*T_*\Big\}\frac{dz}{z}\,. \end{aligned} \tag{C.12}$$

Now, the use of (C.9), $\mathrm{tr}\tilde{\boldsymbol{V}}_*\tilde{\boldsymbol{V}}\{\cdots\} = \mathrm{tr}\tilde{\boldsymbol{V}}\{\cdots\}\tilde{\boldsymbol{V}}_*$ and completing the square gives

$$\begin{aligned} J_{min} \;=\; & \mathrm{tr}\frac{1}{2\pi j}\oint_{|z|=1} \frac{1}{UU_*TT_*DD_*}\Big\{\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\tilde{\boldsymbol{V}}_* \\ & + \left(z^{-m}\tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1} - \boldsymbol{Q}\right)\left(\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\tilde{\boldsymbol{V}}_*z^m - \boldsymbol{Q}_*\right) \\ & - \tilde{\boldsymbol{V}}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\tilde{\boldsymbol{V}}_*\Big\}\frac{dz}{z}\,. \end{aligned}$$

Finally, use the Diophantine equation (C.11) in the middle term and rearrange the terms to obtain equation (3.13). ■

References

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

[2] A. Ahlén and M. Sternad, "Optimal filtering problems," in K. Hunt, ed. *Polynomial methods in optimal control and filtering.* Control Engineering Series, Peter Peregrinus, London, 1993.

[3] A. Ahlén and M. Sternad, "Derivation and design of Wiener filters using polynomial equations," In C. T. Leondes ed., *Academic Press Theme volumes on Digital Signal Processing Techniques and Applications*, Academic Press, New York, 1994, to appear.

[4] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio – part I: theoretical considerations," *IEEE Transactions on Communications*, vol. 40, pp. 885–894, 1992.

[5] B. Bernhardsson and M. Sternad, "Feedforward control is dual to deconvolution," *International Journal of Control*, vol. 57, pp. 393–405, 1993.

[6] B. Bernhardsson, "Robust stochastic performance optimization", IFAC World Congress, Sydney, Australia, July 1993, Vol. 8, pp. 67–70. Also in *Topics in Digital and Robust Control of Linear Systems.* PhD Thesis TFRT–1039–SE, Dept. of Automatic Control, Lund Institute of Technology, Lund, Sweden, 1992.

[7] P. Bolzern, P. Colaneri and G De Nicolao, "Optimal robust filtering for linear systems subject to time-varying parameter perturbations," *Proc. 32nd Conf. on Decision and Control*, San Antonio, Texas, pp 1018–1023, Dec. 1993.

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

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

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

[11] G. C. Goodwin, M. Gevers and B. Ninness, "Quantifying the error in estimated transfer functions with application to model order selection," *IEEE Transactions on Automatic Control*, vol. 27, pp. 913–928, 1992.

[12] G. C. Goodwin and M. E. Salgado, "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, 1989.

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

[14] M. J. Grimble and A. ElSayed, "Solution to the $\mathcal{H}_\infty$ optimal linear filtering problem for discrete-time systems," *IEEE Transactions on Signal Processing*, vol. 38, pp. 1092–1104, 1990.

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

[16] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[17] J. Ježek and V. Kučera, "Efficient algorithm for matrix spectral factorization," *Automatica*, vol. 21, pp. 663–669 1985.

[18] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *Journal of The Franklin Institute*, pp. 171–185, 1977.

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

[20] V. Kučera, *Discrete Linear Control. The Polynomial Equations Approach.* Wiley, New York, 1979.

[21] V. Kučera, "Factorization of rational spectral matrices: a survey," Preprints, *IEE Control'91*, Edinburgh, pp. 1074–1078, March 1991.

[22] W. Lee, *Mobile Cellular Telecommunications Systems*, McGraw Hill, New York, NY, 1989.

[23] C. T. Leondes and J. O. Pearson, "A minimax filter for systems with large plant uncertainties," *IEEE Transactions on Automatic Control*, vol. 17, pp. 266–268, 1972.

[24] E. Lindskog, M. Sternad and A. Ahlén, "Designing decision feedback equalizers to be robust with respect to channel time variations," *Nordic Radio Society Symposium on disturbance resistant radio and radar*, Uppsala, Sweden, Nov 10–11 1993.

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

[26] R. H. Milocco, "Optimal design of feedforward regulators for multivariable systems," Automatic Control Conference, Chicago, IL, June 24–26, 1992, pp. 780–784.

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

[28] K. Öhrn, "Design of Multivariable Cautious Wiener filters," Licentiate thesis, Department of Technology, Uppsala University, Uppsala, Sweden, Jun. 1994.

[29] I. R. Petersen and D. C. McFarlane, "Robust state estimation for uncertain systems," *Proc. 30th Conf. on Decision and Control*, Brighton, England, pp 2630–2631, Dec. 1991.

[30] H. V. Poor, "On robust Wiener filtering," *IEEE Transactions on Automatic Control*, vol. 25, pp. 531–536, 1980.

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

[32] M. Sternad and A. Ahlén, "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, 1990.

[33] M. Sternad, A. Ahlén and E. Lindskog "Robust decision feedback equalization," *ICASSP'93, International Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, vol. III, pp. 555–558, April 26–30, 1993.

[34] M. Sternad and A. Ahlén, "A novel derivation methodology for polynomial-LQ controller design," *IEEE Transactions on Automatic Control*, vol. 38, pp. 116–121, 1993.

[35] M. Sternad and A. Ahlén, "Robust filtering and feedforward control based on probabilistic descriptions of model errors", *Automatica*, vol. 29, pp. 661–679, 1993.

[36] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Transactions on Information Theory*, vol. 30, pp. 316–327, 1984.

[37] L. Xie and Y. C. Soh, "Robust Kalman filtering for uncertain systems," Systems & Control Letters, vol 22, pp 123–129, 1994.

[38] L. Xie, Y. C. Soh and C. E. de Souza, "Robust L_2 filtering: the discrete-time case," *Proc. 32nd Conf. on Decision and Control*, San Antonio, Texas, pp 1250–1251, Dec. 1993.