Chapter 5 by M Sternad and A Ahlén in

M Grimble and V Kučera ed.

A Polynomial Approach to H_2 and H_∞ Robust Control Design.

Springer-Verlag, London 1996.

Contents

Chapter 5

$\mathcal{H}_2$ DESIGN OF NOMINAL AND ROBUST DISCRETE TIME FILTERS

Polynomial methods were originally developed with control applications in mind [1, 2], but have turned out to be very useful also within digital signal processing and communications. The present chapter[1] will outline a polynomial equations framework for nominal and robust multivariable linear filtering and, at the same time, illustrate its utility for signal processing problems in digital communications.

5.1 Introduction

Wiener and Kalman techniques for model-based filter design have been used extensively by electrical engineers for decades. Today, these methods are well known as tools for the design of $\mathcal{H}_2$-optimal estimators. The use of Kalman filters has been the common choice within the control community. A primary reason is the development of the state feedback control theory [3], where Kalman observers constitute essential elements. Other reasons are flexibility and the ability to cope with time-varying systems, as well as the availability of Riccati equation solvers with good numerical behaviour.

[1]Parts of the chapter include material from [4], [5] and [6]. It is republished with permission by IEEE and by Academic Press. The chapter describes work supported by the Swedish National Board for Technical Development, NUTEK, under contracts 87-01583 and 9303294-2, as well as by the Swedish Research Council for Engineering Sciences (TFR), under grant 92-775.

Engineers working in the signal processing and communication fields have instead tended to prefer Wiener filters. One reason is their lower computational complexity; when only a few signals are to be estimated, the use of a time-varying estimator for the whole state vector is deemed unnecessary. The use of filters in input-output form will furthermore provide immediate engineering insight: A quick inspection of poles and zeros roughly indicate what filter properties can be expected.

There are additional differences between control and statistical signal processing. While rational transfer functions (IIR-models) are suitable as models of the often slow dynamics of industrial plants, complexity requirements and high speed applications have frequently forced designers in the signal processing and communication fields to restrict their attention to FIR-models and filters [7, 8, 9]. FIR-filters have the advantage of always being stable and they can readily be optimized, on line as well as off line [10, 11].

FIR filters can approximate any time-invariant impulse response, but this may require many filter coefficients. Long filters might be unacceptable from a complexity point of view. Furthermore, the use of many parameters may lead to overfitting [12]; the amount of information per parameter simply becomes too small. Superior results could therefore in many applications be achieved by the use of structures with fewer parameters, such as IIR-models and filters.

Realizable IIR Wiener filters, based on IIR-models, are conceptually easy to derive [13, 14, 15], but explicit solutions have been difficult to obtain. The polynomial systems framework has been of help here. The rather intractable causal bracket operation $\{\cdot\}_+$, which is central in the classical expression of realizable Wiener filters, can now be readily evaluated by means of a Diophantine equation [6, 17, 18]. The polynomial approach to the design of Wiener estimators for a wide class of filtering problems will be outlined in Sect. 5.2. The polynomial solution provides not only the optimal adjustment, but also the optimal structure and degree of the estimator. The resulting filters have a structure in which numerator and denominator polynomial matrices of the signal models appear directly. Such expressions provide immediate engineering insight into the properties of the solution.

Wiener filters are obtained by minimizing mean square error criteria. Although this is highly relevant in numerous applications, other types of criteria have been suggested as well. In particular, the use of robust filtering is of interest, since signal models are rarely exactly known.

In order to attain robustness, criteria of minimax type have frequently been used in the past. See, for example, [19, 20, 21]. A motivation comes from situations where, perhaps for safety reasons, the effect of a worst case scenario must be minimized, possibly in the presence of model errors. Robust filters obtained in this way tend to be rather conservative.

Another concept which has been considered in the context of robust design is $\mathcal{H}_\infty$-optimization. The design is then conducted with respect to signals of bounded power and unknown spectral density. See Chapters 2 and 4. In the control literature, a main motivation for the development of $\mathcal{H}_\infty$-optimization during the 1980's [22, 23] was to achieve robust stability for feedback systems. This motivation is absent in open-loop filtering problems. Applications of the $\mathcal{H}_\infty$ concept also to estimation problems have, however, been proposed. See, for instance, [24, 25] and [26]. The use of $\mathcal{H}_\infty$-optimization for filtering simply implies the minimization of the largest principal gain of the transfer function between unknown norm-bounded signals and the estimation error. The resulting filters will be inherently conservative, since they are designed to guard against extreme situations, where all disturbance energy is concentrated at the worst frequency. When such a design has to be robustified against modelling errors [25, 26], the conservatism will become even more pronounced. In view of this, it seems more promising to robustify a traditional $\mathcal{H}_2$ design, which focuses on the minimization of mean square estimation errors.

Robust $\mathcal{H}_2$-estimation can be formulated in minimax-$\mathcal{H}_2$ terms, see [19, 20, 21] and [27, 28, 29, 30], but we believe that in signal processing and communication applications, the *average* performance will be a more adequate measure of robust performance.

Robust filtering in an average $\mathcal{H}_2$ sense can be attained by parametrizing model uncertainties by sets of random variables. The average, with respect to these variables, of the mean square estimation error is then minimized. The result will be a single robust filter, designed with respect to the specified set of possible systems. The use of averaged $\mathcal{H}_2$ filtering criteria has been suggested previously in the literature in [31, 32, 33] and more recently in [34, 35] and [4]. The design of robust (cautious) Wiener filters, as presented in the last three references, will be summarized in Sect. 5.3. A corresponding design of robustified Kalman filters, based on a combined use of state-space and polynomial methods, was presented in [36]. This method is outlined in Sect.5.4. A comprehensive treatment can be found in the thesis [37] by Öhrn.

The robust filters derived in Sect. 5.3 and 5.4 are suitable for model-based design problems with moderate spectral uncertainties. They are also capable of accommodating slow time variations. An illustrative example can be found in [38]. If the uncertainties become large, or if the time variations are rapid, then the use of a single robust filter will no longer be appropriate. The use of filter banks or adaptive methods is then required.

Robust filtering can be used to obtain acceptable performance for a set of systems. It can also be used when uncertain models are obtained by system identification. The parameter covariance matrix will, for validated models, constitute a useful indication of the actual amount of model error [39].

When the amount of data available for model adjustment is limited, it is important to improve the model quality in frequency regions which matter most for the subsequent filtering or control. In the control community, this is known as *identification for control* [40]. The subject has received increased interest during recent years primarily for control problems [41], but the issue is equally relevant for estimation problems.

The development of improved methods for model identification can be seen as a complement to robust filter design. In digital communications, for example, equalization of channel dynamics is essential. Based on a channel model, the input signal to the channel is to be estimated. It would be desirable to use an identification algorithm which concentrates its accuracy in the frequency regions of most importance for the equalization. Improved results could then be obtained in the subsequent filtering step, since the range of dynamics over which a robust filter needs to operate will thus be smaller.[2] Unfortunately, little effort has been spent on methods of identification for filtering, based on small data sets. For a preliminary investigation of such problems, see the thesis [42] by Bigi.

The area of digital communications poses many other new challenges for the model-based design of robust, adaptive and multivariable filters. We shall briefly outline some aspects which have recently received attention.

5.1.1 Digital Communications: A Challenging Application Area

Digital mobile radio communications [43, 44] is one of the most rapidly expanding areas within the growing field of digital communications [45, 46]. A major reason is, of course, the introduction of cellular telephone systems, such as GSM[3] and D-AMPS[4]. These systems are now capable of providing both voice, fax and data services. However, various categories of users impose considerable pressure on the development of the systems, by continuously requiring higher capacity, improved quality and more advanced services, some of which involve Internet access.

To meet these demands, operators and manufacturers are already planning for the third generation of systems. Today, two leading technologies for third generation systems can be discerned, namely Time Division Multiple Access (TDMA), such as GSM, which is narrowband, and Code Division Multiple Access (CDMA), which is broadband. (For some details, refer to Sect. 5.2.4.) With either technology, third generation systems will be designed to operate at significantly higher carrier frequencies than the systems of today.

Currently, TDMA systems are most widely spread. Due to the often rather severe conditions and the limited time for calculations, algorithms are

[2] Connected to this problem is the question if filters should be based on estimated models (indirect tuning/adaptation) or if filter coefficients should be adjusted to the data directly. One aspect is that model estimation is performed by minimizing the *output* prediction error, while direct adjustment of an equalizer is performed by minimizing the *input* smoothing estimation error. The latter criterion is directly related to the purpose of an equalizer, namely input estimation. These questions are discussed further in Sect. 5.2.4.

[3] Global System for Mobile communications. The standard is used in Europe, as well as in many other parts of the world, including North American PCS networks.

[4] Digital Advanced Mobile Phone System. The standard is used in North America and a similar standard is used in Japan.

required to be highly efficient and of low complexity. Estimation problems occurring in TDMA and CDMA systems are challenging, for several reasons:

- *Multipath propagation.* In general, a signal travels to the receiver antenna along multiple paths with differing transmission delays. This is known as multipath propagation. Received symbols may thus be smeared out over several symbol intervals, causing intersymbol interference. To retrieve the transmitted symbol sequence, channel estimation and symbol estimation, equalization, will then be required. The design of equalizers is closely connected to the deconvolution problem, formulated in Sect. 5.2.3.

- *Short data records.* In TDMA systems, data are transmitted in bursts, where each burst is allocated to a specific user. A small fraction of the data, the training sequence, is known to the receiver. It is used for identification of the transmission channel. Since the training sequence is short, channel estimation errors are inevitable. The use of novel ways to perform identification for filtering, combined with a subsequent robust filter design, constitutes a promising path for improving the detection.

- *High disturbance levels.* The signal received from a particular mobile transmitter is contaminated by noise and interference, caused by other users in nearby cells and also on adjacent channels. The systems in use today often suffer from an inadequate transmission quality, with frequent interruptions of the radio connections. In some geographical areas, the capacity is also inadequate. A conceivable way to alleviate these problems, in CDMA as well as TDMA systems, is to use *many sensors* (antennas) on base stations and possibly also on mobile units. Multivariable methods for signal processing can then be utilized, to improve reception by nulling out interferers while increasing sensitivity in the direction of the transmitter. See, for example, [47, 48, 49] and [5]. In Sect. 5.2.4, we will discuss this problem further.

- *Rapidly time varying systems.* Mobiles which are travelling in urban areas will move through a standing wave pattern, due to radio waves reflected from surrounding nearby scatterers. The received signals will therefore have a time-varying amplitude, a phenomenon known as short-time *fading.* Depending on the speed of the mobile, the carrier frequency and the symbol rate, the fading will give rise to different degrees of time variability of the channel. In some systems, such as GSM, the time variations are slow or moderate. A detector which is robust to uncertain channel models may therefore be adequate. In other systems, such as D-AMPS, the time variations will be much faster, requiring detectors to be adaptive. As indicated in Sect. 5.5, this motivates the study of improved adaptation algorithms, as in [50, 51, 52, 53] and [54]. For a detailed study of the channel tracking problem, see the thesis

[54] by Lindblom, where also a novel systematic methodology for the design of adaptation algorithms, based on polynomial Wiener filtering concepts, is presented.

Summing up, estimation in mobile radio communications requires efficient algorithms, which not only provide insight, but also robustness and adaptivity as well as low complexity. It is our experience that the polynomial systems framework is an excellent tool for solving problems in this challenging area.

5.1.2 Remarks on the Notation

Signals, matrices and polynomial coefficients may, in the following, be complex-valued. This is, for example, required in communication applications. Let p^* denote the complex conjugate of a scalar p and $\mathbf{P}^*$ the complex conjugate transpose of a matrix $\mathbf{P}$. Let $\mathrm{tr}\mathbf{P}$ denote the trace of $\mathbf{P}$, while $\mathbf{P}^T$ is the transpose of $\mathbf{P}$.

For any complex-valued polynomial

$$P(q^{-1}) \ = \ p_0 + p_1 q^{-1} + \ldots + p_{np} q^{-np}$$

in the backward shift operator q^{-1}, where $q^{-1}y(k) = y(k-1)$, define the *conjugate polynomial*

$$P_*(q) \ \stackrel{\Delta}{=} \ p_0^* + p_1^* q + \ldots + p_{np}^* q^{np}$$

where q is the forward shift operator. A polynomial $P(q, q^{-1})$ in both positive and negative powers of q will be called *double-sided*. Rational matrices, or transfer function matrices, are denoted by boldface calligraphic symbols, for example as $\boldsymbol{\mathcal{R}}(q^{-1})$. Polynomial matrices are denoted by boldface symbols, for example $\boldsymbol{P}(q^{-1})$.

For polynomial matrices, $\boldsymbol{P}_*(q)$ denotes complex conjugate, transpose and substitution of q for q^{-1}. When appropriate, the complex variable z is substituted for the forward shift operator q. Arguments of polynomial and rational matrices are often omitted, when there is no risk for misunderstanding. The *degree* of a polynomial matrix $\boldsymbol{P}$, $\deg \boldsymbol{P}$ or np, is the highest degree of any of its polynomial elements. A polynomial (matrix) is called *monic* if it has a unit leading coefficient (matrix).

A rational matrix, $\boldsymbol{\mathcal{R}}(z^{-1})$, is defined as *stable* if all of its elements have poles within $|z| < 1$. A rational matrix is *causal* if all of its elements are causal transfer functions. Square *polynomial* matrices $\boldsymbol{P}(q^{-1})$ are called *stable* if all zeros of $\det \boldsymbol{P}(z^{-1})$ are located in $|z| < 1$. If $\boldsymbol{P}(z^{-1})$ is stable, then all poles of the elements of $\boldsymbol{P}^{-1}(z^{-1})$ will be located in $|z| < 1$, while all elements of $\boldsymbol{P}_*^{-1}(z)$ have poles in $|z| > 1$. For *marginally stable* square polynomial matrices, some zeros of $\det \boldsymbol{P}(z^{-1})$ are located on $|z| = 1$.

A rational matrix may be represented by polynomial matrices as a *matrix fraction description* (MFD), either left or right:

$$\boldsymbol{\mathcal{G}}(q^{-1}) = \boldsymbol{A}_1^{-1}(q^{-1})\boldsymbol{B}_1(q^{-1}) = \boldsymbol{B}_2(q^{-1})\boldsymbol{A}_2^{-1}(q^{-1}) \tag{5.1.1}$$

See [55]. It can also be converted to *common denominator form*

$$\boldsymbol{\mathcal{G}}(q^{-1}) = \frac{1}{A(q^{-1})}\boldsymbol{B}(q^{-1}) \tag{5.1.2}$$

where $\boldsymbol{B}(q^{-1})$ is a polynomial matrix. The scalar and monic polynomial $A(q^{-1})$ is then a common multiple of the denominators of all rational elements in $\boldsymbol{\mathcal{G}}(q^{-1})$.

A filter which whitens a stochastic process $y(k)$, in the sense that

$$\epsilon(k) = \boldsymbol{\mathcal{V}}(q^{-1})y(k) \quad , \quad E\epsilon(i)\epsilon(j)^T = \mathbf{0} \ , \ i \neq j$$

is called a *whitening filter*. The whitening filters considered in the discussion below are stably and causally invertible square rational matrices. The inverse of the above relation,

$$y(k) = \boldsymbol{\mathcal{V}}^{-1}(q^{-1})\epsilon(k) \tag{5.1.3}$$

represents an *innovations model* of the signal $y(k)$ [56].

5.2 Wiener Filter Design Based on Polynomial Equations

The use of polynomial methods for the (nominal) design of Wiener filters has been discussed during the last decade by several authors [17, 18, 57, 58, 59, 60, 61, 62]. We shall begin this section by presenting a fairly general problem formulation in Sect. 5.2.1 and Sect. 5.2.2, which includes many of the previously considered filtering problems as special cases. Then, as an example of the general setup, a deconvolution problem (Sect. 5.2.3) and a equalization problem (Sect. 5.2.4) will be discussed in some detail. The resulting estimators constitute multivariable model-based filters, predictors or fixed-lag smoothers for the nominal case, without model errors.

5.2.1 A General $\mathcal{H}_2$ Filtering Problem

Based on measurements $d(k)$ up to time $k+m$, a vector

$$z(k) = (z_1(k) \dots z_\ell(k))^T$$

of ℓ signals is to be estimated. The signals are modelled as the outputs of the linear time-invariant discrete-time stochastic system

$$\begin{pmatrix} d(k) \\ z(k) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mathcal{G}}_g(q^{-1}) \\ \boldsymbol{\mathcal{D}}_g(q^{-1}) \end{pmatrix} u_g(k) \tag{5.2.1}$$

and the estimator is represented as a transfer function matrix, operating on measurement data $d(k+m)$

$$\hat{z}(k|k+m) \;=\; \boldsymbol{\mathcal{R}}_d(q^{-1})d(k+m) \quad . \tag{5.2.2}$$

Here, $\boldsymbol{\mathcal{G}}_g$, $\boldsymbol{\mathcal{D}}_g$, and $\boldsymbol{\mathcal{R}}_d$ are rational matrices of appropriate dimensions and $\{u_g(k)\}$ is a stochastic process, not necessarily white. Depending on the *smoothing lag* m, the estimator would constitute a predictor ($m<0$), a filter ($m=0$) or a fixed lag smoother ($m>0$).

When the model (5.2.1) is assumed exactly known, we will consider the minimization of the estimation error covariance matrix

$$\mathbf{P} \;\triangleq\; E\varepsilon(k)\varepsilon^*(k) \tag{5.2.3}$$

where

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

Above, $\boldsymbol{\mathcal{W}}(q^{-1})$ is a stable and causal transfer function weighting matrix, which may be used to emphasize filtering performance in important frequency ranges. The covariance matrix (5.2.3) is to be minimized, in the sense that any alternative estimator provides a covariance matrix $\bar{\mathbf{P}}$, for which $\bar{\mathbf{P}}-\mathbf{P}$, is nonnegative definite. The minimization is performed under the constraint of realizability (internal stability and causality) of the filter $\boldsymbol{\mathcal{R}}_d(q^{-1})$.

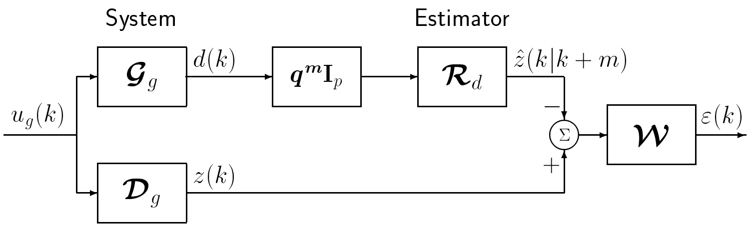

Figure 5.1: A general Wiener filtering problem.

Minimization of the covariance matrix (5.2.3) also implies the minimization of the sum of the elementwise mean square errors (MSE)'s:

$$J \;=\; \mathrm{tr}E(\varepsilon(k)\varepsilon^*(k)) \;=\; E(\varepsilon^*(k)\varepsilon(k)) \;=\; \sum_{i=1}^{\ell} E|\varepsilon_i(k)|^2 \ . \tag{5.2.4}$$

In Sect. 5.3 and 5.4, where the model (5.2.1) is assumed to be uncertain, we will instead consider the minimization of an *averaged* MSE criterion

$$\bar{J} \;=\; \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon^*(k)) \tag{5.2.5}$$

where $\bar{E}(\cdot)$ denotes an expectation over stochastic variables, which are used for parametrizing the set of admissible models.

5.2.2 A Structured Problem Formulation

While the model (5.2.1) is general, it is frequently of advantage to introduce additional structure, to obtain solutions which provide useful engineering insight. For the purpose of this chapter, we will therefore introduce a more detailed structure, which encompasses a number of special cases, some of which will be studied in more detail.

Let us partition the vector $u_g(k)$ in (5.2.1) into two parts

$$u_g(k) \quad = \quad \begin{pmatrix} u(k) \\ w(k) \end{pmatrix}$$

where $w(k)$ represents additive measurement noise, uncorrelated to the desired signal $z(k)$. The desired signal is assumed to be a filtered version of $u(k)$,

$$z(k) = \boldsymbol{\mathcal{D}}_g(q^{-1})u(k) \ .$$

We also introduce explicit stochastic models for the vector $u(k)$ and the noise

$$u(k) = \boldsymbol{\mathcal{F}}(q^{-1})e(k) \quad , \quad w(k) = \boldsymbol{\mathcal{H}}(q^{-1})v(k),$$

with $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$ being stable or marginally stable. The noise sequences $\{e(k)\}$ and $\{v(k)\}$ are assumed to be mutually uncorrelated, white and stationary. They have zero means and covariance matrices $\boldsymbol{\phi} \geq 0$ and $\boldsymbol{\psi} \ \geq 0$.[5]

If noise-free measurements are available, it can be of interest to handle them separately. We therefore partition the measurement vector as

$$d(k) \ \triangleq \ \begin{pmatrix} y(k) \\ a(k) \end{pmatrix} \tag{5.2.6}$$

where the noise $w(k)$ affects $y(k) = (y_1(k) \dots y_p(k))^T$ additively, while $a(k) = (a_1(k) \dots a_h(k))^T$ is uncorrupted by $w(k)$.

The model structure (5.2.1) is thus converted to the form

$$\begin{aligned} \begin{pmatrix} y(k) \\ a(k) \\ z(k) \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\mathcal{G}}(q^{-1}) & \mathbf{I} \\ \boldsymbol{\mathcal{G}}_a(q^{-1}) & \mathbf{0} \\ \boldsymbol{\mathcal{D}}(q^{-1}) & \mathbf{0} \end{pmatrix} \begin{pmatrix} u(k) \\ w(k) \end{pmatrix} \\ \begin{pmatrix} u(k) \\ w(k) \end{pmatrix} &= \begin{pmatrix} \boldsymbol{\mathcal{F}}(q^{-1}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{H}}(q^{-1}) \end{pmatrix} \begin{pmatrix} e(k) \\ v(k) \end{pmatrix} \end{aligned} \tag{5.2.7}$$

See Fig. 5.2. Above, $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\mathcal{G}}_a$, $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{H}}$, and $\boldsymbol{\mathcal{D}}$ are transfer function matrices of appropriate dimensions. The transfer functions will, in the following, be

[5]Frequently, it is convenient to normalize $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ to unit matrices and include variance scaling in $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$ respectively. This will be the case, for example, in Sect. 5.2.3 in the robust estimation problems discussed in Sect. 5.3

parametrized either by state-space models or by polynomial matrices in q^{-1} as MFD's (5.1.1) or common denominator forms (5.1.2).

All of the subsystems will in the present chapter be assumed stable. Corresponding problems with marginally stable blocks are discussed in [37]. Structured or unstructured model uncertainty may be present in any subsystem.

Based on the measurements $d(k)$, up to time $k+m$, our aim is thus to optimize the linear estimator (5.2.2)

$$\hat{z}(k|k+m) = \mathcal{R}_d(q^{-1})z(k+m) \;=\; (\mathcal{R}(q^{-1})\;\mathcal{R}_a(q^{-1})) \begin{pmatrix} y(k+m) \\ a(k+m) \end{pmatrix} \quad (5.2.8)$$

in which both $\mathcal{R}$ and $\mathcal{R}_a$ are required to be stable and causal transfer function matrices.

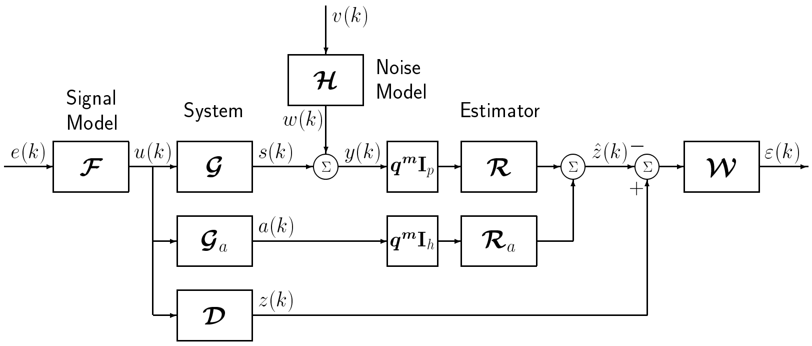

Figure 5.2: Unifying structure for a collection of $\mathcal{H}_2$ filtering problems. The signal $z(k)$ is to be estimated based on data $y(k)$ up to time $k+m$.

The structure depicted in Fig. 5.2 covers a large set of different problems. We shall in the present chapter discuss the following special cases:

- Multisignal deconvolution and linear equalization (Sect. 5.2.3);
- Decision feedback equalization of a white vector sequence of digital symbols: $\mathcal{W} = \mathbf{I}$, $\mathcal{F} = \mathbf{I}$, $\mathcal{D} = \mathbf{I}$, $\mathcal{G}_a = q^{-m-1}\,\mathbf{I}$ (Sect. 5.2.4);
- Estimation based on uncertain input-output models (Sect. 5.3);
- Estimation based on uncertain state-space models (Sect. 5.4).

5.2.3 Multisignal Deconvolution

The estimation of input signals to dynamic systems is known as *deconvolution*, or input estimation. In such problems, $\mathcal{G}$ in (5.2.7) constitutes a dynamic system. An application, discussed in more detail in Sect. 5.2.4, is the equalization of communication channels [45, 46, 63, 64]. Another recent interesting application of multivariable deconvolution is the reconstruction of stereophonic sound, as described by Nelson *et. al.* in [65]. Numerical differentiation may be formulated as the problem of estimating the input to a discrete-time approximation of an integrator [57, 58]. Applications to seismic signal processing are described in [66], and the references therein.

We will consider the problem of deconvolution with multiple inputs and multiple outputs, assuming the involved dynamic systems to be exactly known. All problems described by Fig. 5.1 and Fig. 5.2 are included.

The deconvolution problem could be set up and solved using general MFD's, see [16, 17] or [24]. Here, we will approach the problem by representing some transfer functions by MFD's having *diagonal denominator matrices*, while others are represented in *common denominator* form.

Parametrizing the problem in this way has several advantages. First, no coprime factorizations will be required, which results in a transparent solution. Thus, engineering insight is more easily obtained. Second, the solution involves a *unilateral* Diophantine equation instead of a bilateral one: The polynomial matrices to be determined appear on the same sides of different terms of the equation, instead of on opposite sides. This will make the solution attractive, both from a numerical and from a pedagogical point of view: Solving a unilateral Diophantine equation corresponds to solving a block-Toeplitz system of linear equations with multiple right-hand sides. For an example, see [4].

Let the measurement vector $y(k)$ and the input $u(k)$ be described by

$$\begin{aligned} y(k) &= \boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})u(k) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})v(k) \\ u(k) &= \frac{1}{D(q^{-1})}\boldsymbol{C}(q^{-1})e(k) \ . \end{aligned} \tag{5.2.9}$$

Here, $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{N}, \boldsymbol{M}, \boldsymbol{C}\}$ are polynomial matrices of dimensions $p|p$, $p|s$, $p|p$, $p|r$, and $s|n$, respectively, while D is a scalar polynomial. The matrices $\boldsymbol{A}$ and $\boldsymbol{N}$ are assumed *diagonal.* As indicated in the previous section, $\{e(k)\}$ and $\{v(k)\}$ are mutually uncorrelated zero mean stochastic processes. Here, they are normalized to have *unit* covariance matrices of dimensions $n|n$ and $r|r$, respectively. Since rows of $\boldsymbol{M}$ are allowed to be zero, noise-free measurements can be included. The polynomial matrix $\boldsymbol{B}$ need not be stably invertible. It may not even be square.

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

$$\hat{z}(k|k+m) = \boldsymbol{\mathcal{R}}(q^{-1})y(k+m) \tag{5.2.10}$$

of a filtered version $z(k)$ of the input $u(k)$

$$z(k) = \frac{1}{T(q^{-1})}\boldsymbol{S}(q^{-1})u(k)$$

is sought. The filter $\boldsymbol{S}/T$, with T scalar and $\boldsymbol{S}$ of dimension $\ell|s$, may represent additional dynamics in the problem description, cf. [58, 59], a frequency shaping weighting filter cf. [57], or the selection of particular states.

The covariance matrix (5.2.3), or the sum of MSE's (5.2.4), is to be minimized with dynamic weighting

$$\boldsymbol{\mathcal{W}}(q^{-1}) = \frac{1}{U(q^{-1})}\boldsymbol{V}(q^{-1}) \quad .$$

This problem formulation corresponds to the choice

$$\boldsymbol{\mathcal{G}} = \boldsymbol{A}^{-1}\boldsymbol{B} \ , \ \boldsymbol{\mathcal{G}}_a = \boldsymbol{0} \ , \ \boldsymbol{\mathcal{F}} = \boldsymbol{C}/D \ , \ \boldsymbol{\mathcal{H}} = \boldsymbol{N}^{-1}\boldsymbol{M} \ , \ \boldsymbol{\mathcal{D}} = \boldsymbol{S}/T \ , \ \boldsymbol{\mathcal{W}} = \boldsymbol{V}/U$$

and $\boldsymbol{\mathcal{R}}_z = [\boldsymbol{\mathcal{R}} \ \boldsymbol{0}]$ in (5.2.3),(5.2.7) and (5.2.8). See Fig. 5.3.

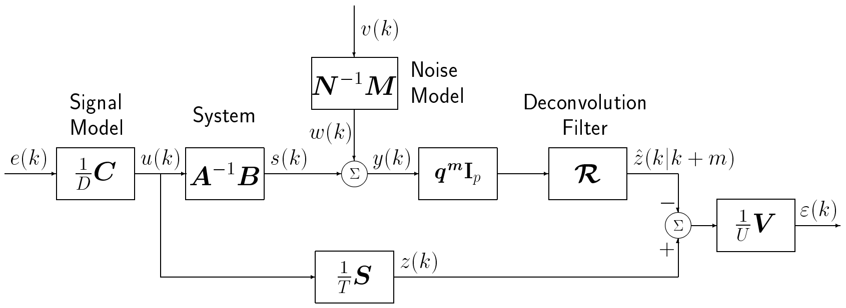

Figure 5.3: A generalized multi-signal deconvolution problem. The vector sequence $\{z(k)\}$ is to be estimated from the measurements $\{y(k)\}$, up to time $k+m$.

When $\{u(k)\}$ is a sequence of digital symbols in a communication network, the estimator (5.2.10) will constitute a multivariable *linear equalizer*. In order to retrieve the transmitted symbols, the estimate $\hat{z}(k) = \hat{u}(k)$ is then fed into a decision device.

Introduce the following assumptions.

Assumption 5.1. *The polynomials $D(q^{-1})$, $T(q^{-1})$, and $U(q^{-1})$ are all stable and monic, while the polynomial matrices $\boldsymbol{A}(q^{-1})$, $\boldsymbol{N}(q^{-1})$ and $\boldsymbol{V}(q^{-1})$ have stable determinants and unit leading coefficient matrices. (Thus, they have stable and causal inverses.)*

Assumption 5.2. *The spectral density of $y(k)$, $\boldsymbol{\Phi}_y(e^{j\omega})$, is nonsingular for all ω.*

From (5.2.9), we now obtain the spectral density matrix $\boldsymbol{\Phi}_y$ as

$$\boldsymbol{\Phi}_y = \frac{1}{DD_*}\boldsymbol{A}^{-1}\boldsymbol{BCC}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1} + \boldsymbol{N}^{-1}\boldsymbol{MM}_*\boldsymbol{N}_*^{-1} = \boldsymbol{\alpha}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{\alpha}_*^{-1} \tag{5.2.11}$$

where

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{NBCC}_*\boldsymbol{B}_*\boldsymbol{N}_* + DD_*\boldsymbol{AMM}_*\boldsymbol{A}_* \tag{5.2.12}$$

and

$$\boldsymbol{\alpha} \triangleq D\boldsymbol{NA} \ .$$

Note that $\boldsymbol{A}$ and $\boldsymbol{N}$ commute, since they are assumed diagonal.

Under Assumption 5.2, a stable $p|p$ spectral factor $\boldsymbol{\beta}$, with $\det \boldsymbol{\beta}(z^{-1}) \neq 0$ in $|z| \geq 1$ and with nonsingular leading matrix $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(0)$, can always be found. Thus, $\boldsymbol{\alpha}^{-1}\boldsymbol{\beta}$ constitutes an innovations model of the measurement vector, while $\boldsymbol{\beta}^{-1}\boldsymbol{\alpha}$ is a stable and causal whitening filter, cf. (5.1.3).

The optimal estimator can be derived in many ways, of which completing the squares, the variational approach, the inner-outer factorization approach [67] and the classical Wiener solution are the most well known. See [16] or [6] for a comparision. We shall here use the variational approach and a brief outline of this methodology is presented next.

Optimization by Variational Arguments [6, 16, 17]. Consider the estimator (5.2.2) and the criterion (5.2.3). Introduce an *alternative weighted estimate*

$$\hat{\delta}(k|k+m) = \boldsymbol{\mathcal{W}}(q^{-1})\hat{z}(k|k+m) + \nu(k) = \boldsymbol{\mathcal{W}}(q^{-1})\boldsymbol{\mathcal{R}}_d(q^{-1})d(k+m) + \nu(k) \tag{5.2.13}$$

where the column vector $\nu(k)$ of stationary signals represents a modification of the (weighted) estimate. See Fig. 5.4. The estimate $\hat{z}(k)$ is optimal if and only if no admissible variation can improve upon the criterion value.

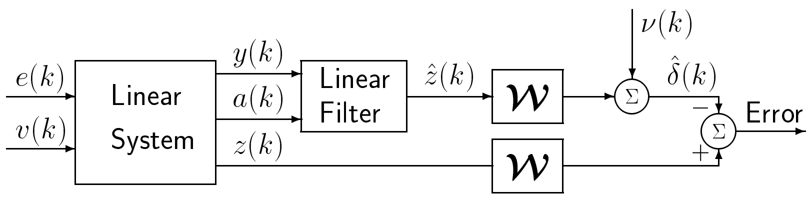

Figure 5.4: Setup for filter optimization via the variational approach in the general estimation problem defined by (5.2.7) and (5.2.8). The weighted estimate is perturbed by a variation $\nu(k)$.

All admissible variations can be represented by

$$\nu(k) = \boldsymbol{\mathcal{T}}(q^{-1})d(k+m)$$

where $\boldsymbol{\mathcal{T}}(q^{-1})$ is some stable and causal rational matrix. In problems with unstable models, any nonstationary mode of $d(k)$ must be cancelled by zeros of $\boldsymbol{\mathcal{T}}(q^{-1})$. Except for these requirements, $\boldsymbol{\mathcal{T}}(q^{-1})$ is an arbitrary.

The use of the modified estimator (5.2.13) results in the covariance matrix

$$\begin{aligned}\bar{\mathbf{P}} &= E\{\boldsymbol{\mathcal{W}}(q^{-1})z(k) - \hat{\delta}(k|k+m)\}\{\boldsymbol{\mathcal{W}}(q^{-1})z(k) - \hat{\delta}(k|k+m)\}^* \\ &= E\varepsilon(k)\varepsilon(k)^* - E\varepsilon(k)\nu(k)^* - E\nu(k)\varepsilon(k)^* + E\nu(k)\nu(k)^* \ . \end{aligned} \qquad (5.2.14)$$

If the *cross-terms* in (5.2.14) are zero, then $\nu(k) \equiv \mathbf{0}$ will evidently minimize $\bar{\mathbf{P}}$, since $E\nu(k)\nu(k)^*$ is positive semidefinite if any component of $\nu(k)$ has nonzero variance. Then, the estimator (5.2.2) is optimal.[6]

Of the two cross terms, it is sufficient to consider only $E\varepsilon(k)\nu(k)^*$, for symmetry reasons. The estimation error $\varepsilon(k)$ in (5.2.3) is required to be *stationary.* This will be the case if $z(k)$ and $d(k)$ are stationary, since $\boldsymbol{\mathcal{W}}(q^{-1})$ and $\boldsymbol{\mathcal{R}}_d(q^{-1})$ are required to be stable[7].

With $\{\varepsilon(k)\}$ and $\{\nu(k)\}$ being stationary sequences, Parsevals formula can now be used to convert the requirement $E\varepsilon(k)\nu(k)^* = \mathbf{0}$ into the frequency-domain relation

$$E\varepsilon(k)\nu(k)^* = \frac{1}{2\pi j}\oint_{|z|=1} \boldsymbol{\phi}_{\varepsilon\nu^*}\frac{dz}{z} = \mathbf{0} \quad . \qquad (5.2.15)$$

The rational $\ell|\ell$-matrix $\boldsymbol{\phi}_{\varepsilon\nu^*}$ is the cross spectral density. The expression (5.2.15) corresponds to the elementwise orthogonality conditions[8]

$$E\varepsilon_v(k)\nu_n(k)^* = \frac{1}{2\pi j}\oint_{|z|=1} \boldsymbol{\phi}^{vn}_{\varepsilon\nu^*}\ \frac{dz}{z} = 0 \qquad v = 1\ldots\ell \ , \ n = 1\ldots\ell \ . \qquad (5.2.16)$$

These ℓ^2 conditions determine the estimator $\boldsymbol{\mathcal{R}}_d(q^{-1})$. They are fulfilled if the integrands are made analytic inside the integration path $|z| = 1$. All poles of the integrands inside the unit circle should thus be cancelled by zeros.

Using the common denominator form or the left polynomial matrix fraction description, the relations (5.2.16) can be evaluated collectively, rather than individually, when $\ell > 1$. They then reduce to a linear polynomial (matrix) equation, a bilateral or unilateral *Diophantine equation.*

In robust design problems with uncertain models, the operator $\bar{E}E(\cdot)$ is substituted for $E(\cdot)$ in the reasoning above [4].

[6] By taking the trace of (5.2.14), it is evident that the scalar MSE criterion (5.2.4) is also minimized by $\nu(k) = \mathbf{0}$.

[7] If $z(k)$ or $d(k)$ were generated by marginally stable models, stationarity of the estimation error would have to be verified separately, *after* the derivation. See [6, 16, 37].

[8] When $\ell > 0$, these conditions imply (but are stronger than) orthogonality between the estimation error and any admissible perturbation of the estimate, which corresponds to $\mathrm{tr}E[\varepsilon(k)\nu(k)^*] = \mathrm{tr}E\nu(k)^*\varepsilon(k) = E\nu(k)^*\varepsilon(k) = 0$.

Derivation of the Deconvolution Estimator. The methodology outlined above will now be exemplified on the problem specified by (5.2.9)-(5.2.12) and by the Assumptions 5.1-5.2. Let

$$\varepsilon(k) = \frac{1}{U}\boldsymbol{V}(z(k) - \hat{z}(k|k+m))$$

be the filtered error and $\nu(k) = \boldsymbol{\mathcal{T}}(q^{-1})y(k+m)$ the variation. Since the noises $e(k)$ and $v(k)$ are assumed uncorrelated, and since all the involved systems are assumed stable, we obtain the cross covariance matrix

$$\begin{aligned} E\varepsilon(k)\nu^*(k) = {} & E\frac{1}{U}\boldsymbol{V}\left[\left(\frac{1}{T}\boldsymbol{S} - q^m\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B}\right)\frac{1}{D}\boldsymbol{C}e(k) - q^m\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k)\right] \\ & \left[\boldsymbol{\mathcal{T}}q^m\left(\frac{1}{D}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}e(k) + \boldsymbol{N}^{-1}\boldsymbol{M}v(k)\right)\right]^* \\ = {} & \frac{1}{2\pi j}\oint_{|z|=1}\frac{1}{U}\boldsymbol{V}\left\{z^{-m}\frac{1}{TDD_*}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1}\right. \\ & \left. -\boldsymbol{\mathcal{R}}\left[\frac{1}{DD_*}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1} + \boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{N}_*^{-1}\right]\right\}\boldsymbol{\mathcal{T}}_*\frac{dz}{z} \quad . \end{aligned} \tag{5.2.17}$$

The use of the expression (5.2.11) in (5.2.17) gives, with $\boldsymbol{\alpha}_*^{-1} = D_*^{-1}\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}$,

$$E\varepsilon(k)\nu^*(k) = \frac{1}{2\pi j}\oint\frac{1}{U}\left\{\frac{z^{-m}}{TD}\boldsymbol{V}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* - \boldsymbol{V}\boldsymbol{\mathcal{R}}\boldsymbol{\alpha}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\right\}\boldsymbol{\alpha}_*^{-1}\boldsymbol{\mathcal{T}}_*\frac{dz}{z} \quad . \tag{5.2.18}$$

The integral vanishes if the estimator $\boldsymbol{\mathcal{R}}$ is adjusted so that no element of the integrand in (5.2.18) has poles inside the integration path $|z| = 1$. Since $\boldsymbol{A}$, $\boldsymbol{N}$ and D are stable, $\boldsymbol{\alpha}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{N}^{-1}D^{-1}$ has poles only in $|z| < 1$. Elements of $\boldsymbol{\beta}(z^{-1})$ may contribute poles at the origin.[9] These factors can be cancelled directly by $\boldsymbol{\mathcal{R}}$. Moreover, if $\boldsymbol{\mathcal{R}}$ contains $\boldsymbol{V}^{-1}/T$ as a left factor, the matrix $\boldsymbol{V}$ is cancelled and $1/TD$ can be factored out from the two terms of the integrand, to be cancelled later. Thus, we select

$$\boldsymbol{\mathcal{R}} = \frac{1}{T}\boldsymbol{V}^{-1}\boldsymbol{Q}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A} \tag{5.2.19}$$

where $\boldsymbol{Q}(q^{-1})$, of dimension $\ell|p$, is undetermined. With the filter (5.2.19) inserted, the cross covariance matrix (5.2.18) becomes

$$E\varepsilon(k)\nu^*(k) = \frac{1}{2\pi j}\oint_{|z|=1}\frac{1}{UTD}\{z^{-m}\boldsymbol{V}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* - \boldsymbol{Q}\boldsymbol{\beta}_*\}\boldsymbol{\alpha}_*^{-1}\boldsymbol{\mathcal{T}}_*\frac{dz}{z} \quad .$$

[9] A polynomial $\beta(z^{-1})$ can, alternatively, be represented as a rational function in z, by multiplying and dividing with $z^{n\beta}$. The elements of this rational function have poles at the origin $z = 0$. Thus, we have to eliminate all numerator polynomials having z^{-1} as argument.

All poles of every element of $\boldsymbol{\alpha}_*^{-1}\boldsymbol{\mathcal{T}}_*$ are located outside $|z|=1$, since $\boldsymbol{\alpha}$ is a stable polynomial matrix and the rational matrix $\boldsymbol{\mathcal{T}}$ is causal and stable. The remaining factor of the integrand may contribute poles in $|z|<1$. In order to attain orthogonality, we therefore require that

$$z^{-m}\boldsymbol{VSCC}_*\boldsymbol{B}_*\boldsymbol{N}_* \;=\; \boldsymbol{Q\beta}_* + z\boldsymbol{L}_*UTD\,\mathbf{I}_p \tag{5.2.20}$$

for some polynomial matrix $\boldsymbol{L}_*(z)$. We then obtain an integrand with only strictly unstable rational functions in z as elements, so the integral

$$E\varepsilon(k)\nu^*(k) = \frac{1}{2\pi j}\oint_{|z|=1}\boldsymbol{L}_*(z)\boldsymbol{\alpha}_*^{-1}(z)\boldsymbol{\mathcal{T}}_*(z)dz \tag{5.2.21}$$

will vanish. Equation (5.2.20) is a linear polynomial matrix equation, a *unilateral* Diophantine equation. Here, $\boldsymbol{Q}(q^{-1})$ and $\boldsymbol{L}_*(q)$ are polynomial matrices, of dimension $\ell|p$, with generic degrees[10]

$$nQ = \max(nc+ns+nv+m, nt+nd+nu-1)$$

$$nL = \max(nc+nb+nn-m, n\beta)-1 \tag{5.2.22}$$

where $nc=\deg\boldsymbol{C}, ns=\deg\boldsymbol{S}$ etc. Unique solvability of (5.2.20) with respect to $\boldsymbol{Q}$ and $\boldsymbol{L}_*$ can be demonstrated, see, for example, [4] or [6].

The design equations thus consist of the left spectral factorization (5.2.12), the Diophantine equation (5.2.20) and the filter (5.2.19).[11] For scalar systems, the solution reduces to the one presented in [57]. See also [59]. A numerical illustration can be found in [6].

A Wiener filter works by first whitening the measurement. By introducing D as a stable common factor, it is evident that the estimator (5.2.19) contains a whitening filter $\boldsymbol{\beta}^{-1}\boldsymbol{\alpha}=\boldsymbol{\beta}^{-1}\boldsymbol{NA}D$ as a right factor.

The spectral factor $\boldsymbol{\beta}$ is unique up to a right orthogonal matrix. (If $\boldsymbol{FF}_*=\mathbf{I}$ then $\boldsymbol{\beta\beta}_*=(\boldsymbol{\beta F})(\boldsymbol{F}_*\boldsymbol{\beta}_*)$.) There exist several efficient algorithms for polynomial matrix spectral factorization, some of which are based on state space methods. A survey of algorithms is presented in [68]. See also [69, 70].

The Attainable MSE. The minimal (scalar) MSE criterion value is obtained by inserting (5.2.19), (5.2.12), and (5.2.20), in this order, into (5.2.4). We thus obtain, with $\boldsymbol{H}\triangleq\boldsymbol{VSC}$ and $P\triangleq UTD$,

$$\mathrm{tr}E(\varepsilon(k)\varepsilon^*(k))_{\min} =$$

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

[11] For models with poles in $|z|>1$, a second Diophantine equation will be required. See Ch. 3 and [6]. Such models are, however, of limited practical interest in open-loop filtering.

$$\frac{1}{2\pi j}\oint \mathrm{tr}\{\boldsymbol{L}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{L} + \frac{1}{PP_*}\boldsymbol{H}(\mathbf{I}_n - \boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\ \boldsymbol{NBC})\boldsymbol{H}_*\}\frac{dz}{z}\ . \tag{5.2.23}$$

The minimal criterion value consists of two terms. The first term represents the error caused by incomplete inversion of the system $\boldsymbol{A}^{-1}\boldsymbol{B}$. Only the use of an infinite smoothing lag will cause this term to vanish, unless the system is minimum phase and there is no noise. One can show that $\boldsymbol{L} \rightarrow \mathbf{0}$ when $m \rightarrow \infty$, see [58]. Thus, the second part of (5.2.23) represents the limit of performance approached by a noncausal Wiener filter.

There exists a very special case in which perfect input estimation is possible with finite smoothing lags. It is the case of minimum-phase systems without noise ($\boldsymbol{M} = \mathbf{0}$, $\boldsymbol{N} = \mathbf{I}$), with $q^m\boldsymbol{B}$ square and stably and causally invertible. Consider this situation and let $\boldsymbol{S} = \mathbf{I}_\ell$ and $T = 1$. Then, the direct inversion of the transducer dynamics

$$\boldsymbol{\mathcal{R}} = \boldsymbol{B}^{-1}\boldsymbol{A}\ q^{-m}$$

will result in a vanishing integrand in (5.2.18) .

Adaptive and Blind Deconvolution. For scalar systems, the deconvolution problem has also been studied in an adaptive setting. An interesting feature here is that the spectral factor $\boldsymbol{\beta}$ need not be calculated from the equation (5.2.12). Instead, it can be obtained directly from data, by estimating an appropriately parametrized innovations model

$$y(k) \ = \ \boldsymbol{\alpha}^{-1}(q^{-1})\boldsymbol{\beta}(q^{-1})\epsilon(k)$$

of the measurement vector, cf. (5.2.11) and (5.1.3). See [73]. Based on this fact, multivariable adaptive deconvolution, for the special case of white input and noise, has been discussed in [71, 72] and [60]. Crucial for an adaptive algorithm to work in more general situations, with coloured input and noise, is that the model polynomials can be estimated from the output only. Algorithms which can be applied in the scalar case with signals and noises of unknown colour, but with a *known system*, have been presented in [73] and [74]. In [75], the identifiability properties of the scalar deconvolution problem are investigated and conditions for parameter identifiability are given, when $\boldsymbol{\mathcal{G}}$ is known, while $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$ have to be estimated. The conditions for identifiability in [75] are based on the use of second order moments only.

Another challenging problem is that of *blind deconvolution*, where both the input signal $u(k)$ and the transducer/channel $\boldsymbol{\mathcal{G}}$ have to be estimated from data. The unique estimation of a possibly non-minimum phase system $\boldsymbol{\mathcal{G}}$, based on only the second order statistics of its output is, in general, impossible. The second order statistics does not provide the appropriate phase information. If the input is non-Gaussian while the noise is Gaussian, then higher order statistics can be utilized [76]. Higher order statistics is used, directly or indirectly, in most proposed algorithms for blind deconvolution.

Algorithms based on higher order statistics require long convergence times, and the quality of the estimates may be sensitive to the assumption that the noise is Gaussian. A recent discovery is therefore of significant interest: Blind identification *is* possible, for cyclo-stationary inputs, if several output samples are available per input sample. The continuous-time baseband signals used in digital communications are cyclostationary; the result is therefore of interest for *blind equalization*. The number of received samples per symbol can here either be increased by oversampling (fractionally spaced equalization), or by the use of multiple antennas/sensors, so that $y(k)$ is a vector, while $u(k)$ is scalar [77, 78]. Radio systems with several receiver antennas are of increasing interest in particular for mobile applications, see Sect. 5.2.4.

A Duality to Feedforward Control. The set of problems for which the solution above is relevant can be enlarged further. The considered deconvolution problem turns out to be dual to the LQG (or $\mathcal{H}_2$)-feedforward control problem, with rational weights on control and output signals. See [79]. It is very simple to demonstrate this duality. Reverse all arrows, interchange summation points and node points and transpose all rational matrices in Fig. 5.3. Then, the block diagram for the LQG problem is obtained. The optimization problem remains the same for $\mathcal{H}_2$ problems, and indeed for the minimization for any transfer function norms that are invariant under transposition. Thus, the equations (5.2.12) and (5.2.20) can be used also to design disturbance measurement feedforward regulators, reference feedforward filters and feedforward decoupling filters.

5.2.4 Decision Feedback Equalizers

We now turn our interest to an important problem in digital communications and outline a polynomial solution, which was originally presented in [80] for the scalar case, and later in [6, 16]. The multivariable solution discussed here has been presented in [5, 81].

Digital Communications in the presence of intersymbol interference and co-channel interference. When digital data are transmitted over multiple cross-coupled communication channels, *intersymbol interference*, *co-channel interference* and noise will prevent perfect detection.

The transmitted sequence $\{u(k)\}$ in question is white, and it may be real- or complex-valued.[12] It is to be reconstructed from a sampled received signal $y(k)$. Whenever the channel has an impulse response of length ≥ 1, we are

[12] One example is the use of p-ary symmetric Pulse Amplitude Modulated (PAM) signals. Then, each component of $u(k)$ is a real, white, zero mean sequence which attains values $\{-p+1, \ldots, -1, +1, \ldots, p-1\}$ with some probability distribution. In other modulation schemes, such as Quadrature Amplitude Modulation (QAM), the signal $u(k)$ is complex-valued. For example, in 4-QAM, the symbols represented by the elements of $u(k)$ may attain the values $\{1+i, \;\; 1-i, \;\; -1+i \;\; -1-i\}$. See, for example, [45, 46].

said to encounter *intersymbol interference*: not only the symbol at time k, but also previous symbols contribute to the current received measurement $y(k)$. The channel is then said to be *dispersive*. Dispersive channels occur in digital mobile radio systems such as GSM. Other situations include modem connections over cable, transmission of data to and from hard discs in computers, as well as underwater acoustic communication. In radio communications and underwater acoustics, intersymbol interference is caused by *multipath propagation*: The transmitted signal travels along several paths with differing transmission delays.

A linear equalizer, for example the MSE-optimal design of Sect. 5.2.3, can be used to retrieve the transmitted symbols. A linear MSE-optimal equalizer performs an approximate inversion of the channel. This inversion may result in noise amplification at the filter output, which severely limits the attainable performance.

The lowest error rate would be attained by maximum likelihood estimation of the entire transmitted sequence, implemented through the Viterbi algorithm [82], which constitutes forward dynamic programming. Close to optimal performance, at a much lower computational load, can often be obtained with a Decision Feedback Equalizer (DFE). A DFE is a nonlinear filter, which involves a decision circuit and a feedback of decisioned data through a linear filter to improve the current estimate. See, for example, [83, 84, 85], and the references therein. The attainable bit error rate of a DFE is in many cases several orders of magnitude lower than for a linear equalizer. We shall here consider DFE design as an application where noise free auxiliary information $a(k)$ is explicitely taken into account in the general structure (5.2.7). In order to make the discussion general, we will consider transmission and reception of several signals, that is, we will consider a Multivariable DFE. The design or adaptation of the DFE is here assumed to be based on an indirect approach, i.e. on an explicit multivariable channel model.

Let us describe a channel model structure which is appropriate for the purpose of mobile radio communications.

A FIR Channel Model. Consider a sampled data vector sequence $y(k)$, which represents measurements from a receiver in a radio communication system. The received signal is down-converted to the baseband [45, 46]. Under such circumstances, multiple cross-coupled communication channels are adequately modelled as FIR systems represented by polynomial matrices, with complex elements. The channel model includes pulse shaping, receiver filters and possible transmission delays. It will be described by the multivariable linear stochastic discrete-time model

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{ny} \end{pmatrix} = \begin{pmatrix} B_{11}(q^{-1}) & \dots & B_{1n_u}(q^{-1}) \\ \vdots & \ddots & \vdots \\ B_{n_y1}(q^{-1}) & \dots & B_{n_yn_u}(q^{-1}) \end{pmatrix} \begin{pmatrix} u_1(k) \\ \vdots \\ u_{n_u}(k) \end{pmatrix} + \begin{pmatrix} v_1(k) \\ \vdots \\ v_{n_y}(k) \end{pmatrix}$$

or

$$\begin{aligned} y(k) &= \boldsymbol{B}(q^{-1})u(k) + v(k) \qquad (5.2.24) \\ &= \mathbf{B}_0 u(k) + \ldots + \mathbf{B}_{n_b} u(k - n_b) + v(k) \quad . \end{aligned}$$

Here, we have denoted by n_b the highest degree occurring in any matrix element $B_{ij}(q^{-1})$. The sequence $\{v(k)\}$ is assumed to be discrete-time white noise, which is zero mean, stationary and uncorrelated with $u(k)$.

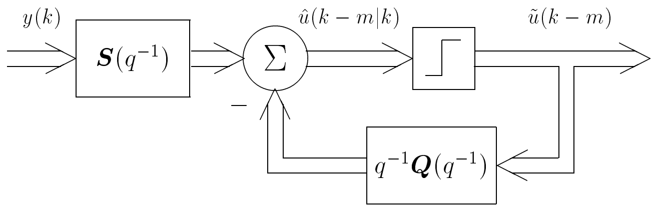

Figure 5.5: Structure of the multivariable DFE.

The Multivariable DFE. A detector designed to estimate only one message $u_i(k)$ would have to treat the co-channel interference from the remaining signals, $\sum_j B_{ij}u_j$, $i \neq j$, as noise. A multivariable detector, which estimates all components of $u(k)$ simultaneoulsy, can utilize the fact that the signals have discrete and known amplitude distributions. This knowledge is utilized in an efficient way by a multivariable DFE.

When the channel is adequately modelled by a polynomial matrix and the noise is white or autoregressive, the appropriate structure of a DFE, cf. [80], is described by

$$\hat{u}(k-m|k) = \boldsymbol{S}(q^{-1})y(k) - \boldsymbol{Q}(q^{-1})\tilde{u}(k-m-1) \quad . \qquad (5.2.25)$$

Here, $\tilde{u}(k-m-1)$ denotes previously detected symbols, while $\hat{u}(k-m|k)$ represents the amplitude-continuous (soft) estimate, obtained with smoothing lag m. Detected symbols $\tilde{u}$ are obtained by feeding $\hat{u}$ through a decision device, which, in its simple form, is given by $\text{sign}(\hat{u})$ for binary PAM signals. The polynomial matrices $\boldsymbol{S}(q^{-1})$ and $\boldsymbol{Q}(q^{-1})$ have dimension $n_u|n_y$ and $n_u|n_u$, respectively. See Fig. 5.5.

Our multivariable DFE is obtained by minimizing the MSE criterion[13]

$$J = E \parallel u(k-m) - \hat{u}(k-m|k) \parallel_2^2 \quad . \qquad (5.2.26)$$

[13] It could be argued that a more relevant criterion is minimum probability of decision errors (MPE), which leads to a nonlinear optimization problem. However, Monsen [84] has concluded that consideration of MPE and MSE lead to essentially the same error probability. A more recent discussion of this issue can be found in [86].

An obstacle, preventing a direct minimization of J, is the presence of the nonlinear decision element. It is impossible to obtain closed form solutions, but the problem can be simplified by assuming *correct past decisions*. This is a common simplification, which allows us to replace previous decisions $\tilde{u}(j)$ by the correct values $u(j)$ for $j = k - m - 1, k - m - 2 \ldots$ in (5.2.25).

If previous decisions are correct, they can be used to completely eliminate the contamination caused by past symbols, at the current time instant. In contrast to linear equalizers, this can be achieved without any noise amplification, since we need not invert the channel. Instead, a feedforward measurement of $a(k) = u(k - m - 1)$ is used. Under the assumption of correct previous decisions, the MIMO-DFE can be included in the general structure (5.2.7), cf Fig. 5.2, by setting

$$\boldsymbol{\mathcal{G}}_a = q^{-m-1}\mathbf{I}_{n_u} \quad , \quad \boldsymbol{\mathcal{F}} = \mathbf{I}_{n_u} \quad , \quad \boldsymbol{\mathcal{D}} = \mathbf{I}_{n_u} \quad , \quad \boldsymbol{\mathcal{R}}_d = [\boldsymbol{S}\ , \boldsymbol{Q}] \ .$$

Thus, we have obtained an ordinary LQ-optimization problem. The resulting MIMO DFE, to be presented next, was derived in [81] and presented in [5].

Theorem 5.1. *Consider the DFE described by (5.2.25) and the channel model described by (5.2.24), where* $E[v(k)v(l)^*] = \Psi\delta_{kl}$. *Assuming correct past decisions, the polynomial matrices* $\boldsymbol{S}(q^{-1})$ *and* $\boldsymbol{Q}(q^{-1})$ *of order* m *and* $n_b - 1$ *respectively, which minimize (5.2.26), are obtained as follows:*

- *The feedforward filter* $\boldsymbol{S}(q^{-1}) = \mathbf{S}_0 + \mathbf{S}_i q^{-1} + \ldots + \mathbf{S}_m q^{-m}$ *is obtained by solving the system of linear equations*

$$\begin{pmatrix} \mathbf{B}_0^* & \ldots & \mathbf{0} & \mathbf{I}_{n_u} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_m^* & \ldots & \mathbf{B}_0^* & \mathbf{0} & \ldots & \mathbf{I}_{n_u} \\ -\Psi & \ldots & \mathbf{0} & \mathbf{B}_0 & \ldots & \mathbf{B}_m \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & -\Psi & \mathbf{0} & \ldots & \mathbf{B}_0 \end{pmatrix} \begin{pmatrix} \mathbf{S}_0^* \\ \vdots \\ \mathbf{S}_m^* \\ \mathbf{L}_{1m} \\ \vdots \\ \mathbf{L}_{10} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{I}_{n_u} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \tag{5.2.27}$$

  with respect to the matrices $\mathbf{S}_i^*$ *and* $\mathbf{L}_{1j}$. *On the right-hand side, the zero matrices have dimension* $n_u|n_u$ *and* $\mathbf{I}_{n_u}$ *is positioned vertically in block number* $m + 1$.

- *The coefficients matrices of the feedback filter*

$$\boldsymbol{Q}(q^{-1}) = \mathbf{Q}_0 + \mathbf{Q}_1 q^{-1} + \ldots + \mathbf{Q}_{n_b-1} q^{-n_b+1}$$

  are then given by

$$\mathbf{Q}_i = \sum_{j=0}^{m} \mathbf{S}_{m-j}\mathbf{B}_{j+i+1} \tag{5.2.28}$$

□

Proof: See [81].

Remarks. The derivation, which follows the same principles as outlined in Sect. 5.2.3, leads to two coupled Diophantine equations. These equations can be transformed to the linear system of equations (5.2.27).

It can be shown that whenever the noise covariance matrix Ψ is nonsingular, the system of equations (5.2.27) will have a unique solution, regardless of the properties of the channel $\boldsymbol{B}(q^{-1})$.

The performance of the equalizer improves monotonically with an increased smoothing lag.

Adaptation and Robustness. The equalizer coefficients can be calculated from data *directly*. They can also be adjusted *indirectly*, via model estimation and filter computation.[14] The *input* MSE criterion (5.2.26) relevant for equalization is used also for filter adjustment by a direct algorithm. An estimator of the model (5.2.24) would instead typically minimize the *output* prediction error. For very long data windows, direct and indirect methods provide the same performance, if the channel is time-invariant. For *scalar* $y(k)$ and time-invariant channels, a nominal indirect design will tend to perform worse than a direct one if the data record is of medium lenght (40-300 data) [87]. A possible explanation is a higher sensitivity to model errors, due to the mismatch between the criteria used of identification and for filtering [42]. For *vector* measurements, the situation seems to be reversed; suitably parameterized indirect methods then outperform direct adaptation of DFE's [49], since a smaller number of parameters need to be adjusted.

For short data windows, indirect methods clearly outperform direct ones, for scalar as well as vector-valued received signals. For example, we advocate the indirect approach when tracking rapidly time-varying channels, a situation which is common in mobile radio communications due to the presence of fading. The reason for this is twofold. First, the time-variations of radio channel coefficients will tend to be smooth, while the corresponding optimal adjustments of the equalizer parameters will have strongly time-varying rates of change. This poses a difficult problem for the selection of the gains of a direct adaptive algorithm [54]. Second, the number of parameters in the equalizer will, in general, be larger than the number of channel coefficients, in particular if the smoothing lag is large. This makes it more difficult to directly adjust the filter to short data sets [49].

If an indirect multivariable adaptive DFE is applied to a time-varying channel, the filter coefficients will have to be recomputed periodically, using Theorem 5.1. Note that the solution steps presented in Theorem 5.1 require *no spectral factorization*. This reduces the computational complexity.

[14] When training data are available, they are used for identification. Otherwise, estimates $\hat{u}(k)$ from the DFE can be used for so-called decision-directed adaptation.

A major drawback with the DFE is that a single erroneous decision under unfortunate conditions may cause a whole sequence of errors, *an error burst.* This phenomenon is known as error propagation. It occurs, in particular, if the feedback filter $\boldsymbol{Q}(q^{-1})$ has a long impulse response.

In [38] and [88], *robust equalizers* are discussed for the scalar case. These algorithms are based on uncertain channel models and are optimized with respect to the averaged MSE criterion (5.2.5). They also provide means to control the error bursts and can basically trade shorter but more frequently occurring error bursts for a decreased frequency of long bursts. The system can then be designed with a coding scheme which gives rise to a smaller delay. The robust DFE is designed by assuming that the feedback signal is corrupted by white noise, with an adjustable variance. As this noise variance is increased, the gain of the feedback filter is reduced, until a (robust) linear equalizer, without decision feedback, is obtained as a limiting case. Channel uncertainty is taken into account as in Sect. 5.3 below.

Next, we will present some problems of digital mobile radio communications in which multivariable DFE:s are applicable.

Example 5.1. *Combined temporal and spatial equalization.*

Consider a digital radio communications scenario where there is only one transmitted message $u(k)$ and n_y receiver antennas. Let the channel from the transmitter to receiver antenna i be described by

$$y_i(k) = B_i(q^{-1})u(k) + v_i(k) \quad . \tag{5.2.29}$$

The scenario for $n_y = 2$ is depicted in Fig. 5.6. The use of two antennas will improve the attainable performance significantly if the channels B_1 and B_2 differ. It will improve the performance moderately even for identical channels, if the noises v_1 and v_2 are not identical.

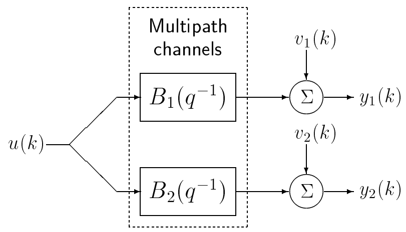

Figure 5.6: Multivariable channel model in a system with two receiver antennas.

Channel models of the type (5.2.29) occur not only in systems for digital mobile radio communications, but also in systems for acoustic underwater *communication. The structure is of increasing interest, since it enables spatial filtering: The antenna may be adapted to have spatial nulls in the direction of interferers while maintaining high gain in the directions of arrival of $u(k)$. See, for example, [5, 47, 48, 49, 81] and [86].*

The equalizer described in Theorem 5.1 can be applied directly to the above scenario. □

Example 5.2. *Narrowband multiuser detection in cellular digital mobile radio systems.*

The multivariable DFE can be used to detect multiple users on the same channel in the same cell simultaneously [5, 81, 89, 90]. See Fig. 5.7. The user u_2 could represent a second user in the same cell. Alternatively, u_2 could represent a co-channel interferer located in a nearby cell. The influence of this interferer on the detection of u_1 should then be minimized.

Consider a scenario with n_u transmitters and n_y receiver antennas. Let the channel from transmitter j to receiver antenna i be represented by $B_{ij}(q^{-1})$. The received signal at antenna i, $y_i(k)$, can in this case be expressed as

$$y_i(k) = \sum_{j=1}^{n_u} B_{ij}(q^{-1}) u_j(k) + v_i(k) \quad .$$

By collecting the antenna signals in vector form, we obtain the model (5.2.24). Application of the DFE at the receiver can be seen as a way to utilize spatial diversity. If the transmitters are at different locations, then the transfer functions B_{ij} will differ for different transmitters j. It is here interesting to note that multipath propagation which leads to intersymbol interference will actually be of advantage. Propagations through different paths will tend to make the channels unequal, which improves the attainable performance.

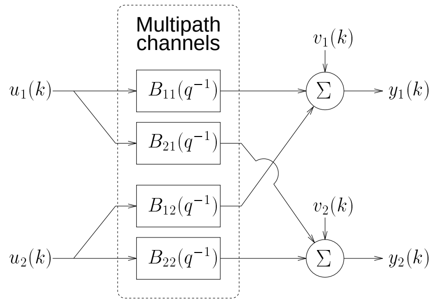

Figure 5.7: Multivariable channel model describing a situation where multiple users simultaneously share the same baseband channel.

□

Example 5.3. Multiuser detection in asynchronous DS-CDMA (Direct Sequence - Code Division Multiple Access).

Multiple users within a cell of a cellular radio systems can share resources by utilizing different frequencies (frequency division), different time slots (time division) or different codes (code division).[15]

In systems using DS-CDMA, all active users within a cell transmit on the same frequency band at the same time [91]. In order to distinguish one message from another, each message is convolved with a different code (or signature) sequence at the transmitter. The sampling rate of the code sequence, the chip rate, is much higher than the symbol rate. The convolution with the code will transform the narrowband message into a rather broad band signal. A DS-CDMA system will thus constitute a spread spectrum transmission system.

A DS-CDMA radio system with n_u simultaneous users can be represented within the model structure (5.2.24). Define $u(k) = (u_1(k)\ \ u_2(k) \dots u_{n_u}(k))^T$ as the vector of transmitted (scalar) symbols at time k. For simplicity, assume the channels to cause no intersymbol interference. The received scalar signal is assumed to be convolved with the appropriate code sequence and then sampled once per symbol. Following Verdú [92], the sampled outputs from a bank of n_u filters, which are matched to the signature sequences of the individual users, may then be collected in vector form, as

$$y'(k) = \mathbf{R}(1)\mathbf{A}u(k+1) + \mathbf{R}(0)\mathbf{A}u(k) + \mathbf{R}^T(1)\mathbf{A}u(k-1) + v'(k)\ . \quad (5.2.30)$$

Here, the matrices $\mathbf{R}(n)$, of dimension $n_u|n_u$, contain partial crosscorrelations between the signature sequences used to spread the data, whereas $v'(k)$ denotes noise. The square and diagonal matrix $\mathbf{A}$ contains channel coefficients associated with the different users. By redefining the measurement and the noise as $y(k) = y'(k-1)$ and $v(k) = v'(k-1)$ in (5.2.30), we obtain a causal multivariable channel model of the form (5.2.24), with $n_y = n_u$ outputs

$$y(k) = (\mathbf{R}(1)\mathbf{A} + \mathbf{R}(0)\mathbf{A}q^{-1} + \mathbf{R}^T(1)\mathbf{A}q^{-2})u(k) + v(k)\ . \quad (5.2.31)$$

If the channel causes intersymbol interference, the constant matrix $\mathbf{A}$ would be substituted by a diagonal polynomial matrix. If multiple antennas are utilized, the dimensions of all matrices are increased correspondingly. □

[15] These strategies can be combined. They can also be complemented and enhanced by utilizing spatial diversity with multiple antennas, as indicated by Examples 5.1 and 5.2.

5.3 Design of Robust Filters in Input-Output Form Based on Averaged $\mathcal{H}_2$ Criteria

For any model-based filter, modelling errors will constitute a potential source of performance degradation. In this section, we propose a *cautious Wiener filter* for the prediction, filtering or smoothing of discrete-time signal vectors. The methodology has been presented in [4, 35] and [34]. A comprehensive exposition can be found in the thesis [37] by Öhrn.

The design of robust multivariable estimators, as it will be presented here, will be based on a stochastic description of model errors, related to the stochastic embedding concept of Goodwin and co-workers [12, 93].

To be more specific, the suggested approach is based on the following foundations:

- 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

$$\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{F}}_o + \Delta\boldsymbol{\mathcal{F}} \quad . \tag{5.3.1}$$

  We will 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 (5.2.5)

$$\bar{J} = \mathrm{tr}\bar{E}E(\varepsilon(k)\varepsilon(k)^*) \quad . \tag{5.3.2}$$

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

The averaged mean square error has been used previously in the literature by, for example, Chung and Bélanger [31], Speyer and Gustafson [32] and by Grimble [33, 94]. These works were based on assumptions of small parametric uncertainties and on series expansions of uncertain parameters. We suggest the use of the criterion (5.3.2), together with a particular description of the sets (5.3.1): Transfer function elements in $\Delta\boldsymbol{\mathcal{F}}$ are postulated to have stochastic numerators and fixed denominators. Such models can describe non-parametric uncertainty and under-modelling as well as parametric uncertainty.

5.3.1 Approaches to Robust $\mathcal{H}_2$ Estimation

The most obvious *ad hoc* approach for increasing the robustness against model errors is perhaps to detune a filter, by increasing the measurement noise variance used in the design. This will work, in principle, if the transducer $\boldsymbol{\mathcal{G}}$ and/or the noise model $\boldsymbol{\mathcal{H}}$ is uncertain. It might, however, be very difficult to find an appropriate noise colour or covariance structure without a more systematic technique. When the signal model $\boldsymbol{\mathcal{F}}$ is uncertain, a detuning approach will not work at all. The filter gain should in such situations instead be *increased*, in an appropriate way.

Most previous suggestions for obtaining robust filters in a systematic way have been based on some type of minimax approach [19, 95]. For example, a paper [96] by Martin and Mintz takes both spectral uncertainty and uncertainty in the noise distribution into account. The resulting filter will 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 $\min_{\mathcal{R}} \max_{\mathcal{M}}$ equals $\max_{\mathcal{M}} \min_{\mathcal{R}}$. If so, instead of finding the worst case with respect to a set of models $\boldsymbol{\mathcal{M}}$, 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 [20, 97, 98, 99] and the survey paper [21] by Kassam and Poor. $\min_{\mathcal{R}} \max_{\mathcal{M}} = \max_{\mathcal{M}} \min_{\mathcal{R}}$ is *not* fulfilled in numerous problems, which makes them very difficult to solve. See, for instance, Example 5 in [34], and the example in [4].

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 [27, 28, 30] and [100] for continuous-time results and [29] for the discrete-time one-step-ahead predictor. See also [101] 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 conservatism is illustrated and discussed in [36, 37].

The method suggested in the present section and in Sect. 5.4 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 the error model 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.

One of our goals will be to present transparent design equations, and to hold their number to a minimum, without sacrificing numerical accuracy. As in Sect. 5.2.3, we therefore 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 [17] or [24]. 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 has a strong formal similarity to the nominal design of Sect. 5.2.3 and it provides structural insight; important properties of a robust estimator are evident by direct inspection of the filter expression.

5.3.2 The Averaged $\mathcal{H}_2$ Estimation Problem

Consider the following extended design model

$$\begin{aligned} y(k) &= \boldsymbol{\mathcal{G}}(q^{-1})u(k) + \boldsymbol{\mathcal{H}}(q^{-1})v(k) \\ u(k) &= \boldsymbol{\mathcal{F}}(q^{-1})e(k) \\ z(k) &= \boldsymbol{\mathcal{D}}(q^{-1})u(k) \end{aligned} \tag{5.3.3}$$

where $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\mathcal{H}}$, $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{D}}$ 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 $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$, respectively.

As before, an estimator

$$\hat{z}(k|k+m) = \boldsymbol{\mathcal{R}}(q^{-1})y(k+m) \tag{5.3.4}$$

of $z(k)$ is sought. See Fig. 5.8. The transfer function $\boldsymbol{\mathcal{R}}$ is designed to minimize the averaged mean square error (MSE) criterion (5.2.5).

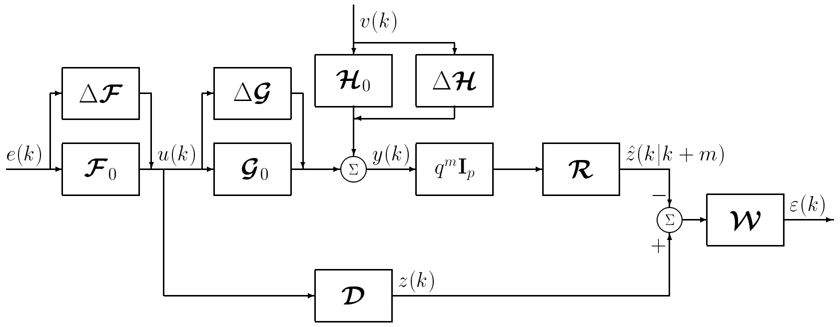

Figure 5.8: A general linear filtering problem formulation with uncertain linear models. Based on noisy measurements $y(k+m)$, the signal $z(k)$ is to be estimated, as in Fig. 5.3. Model errors in transfer functions are described by stochastic error models. Signals have the same dimension as in Sect.5.2.3.

Example 5.4. *Equalization based on an uncertain channel model.*

An application where uncertain dynamics in $\boldsymbol{\mathcal{G}}$ is of interest is equalizer design for digital mobile radio communications. See Sect. 5.2.4.

A signal $u(k)$ then propagates along multiple paths, with different time delays, represented by delays in $\boldsymbol{\mathcal{G}}$. In present systems, both the transmitted symbol sequence $u(k)$ and the received baseband signal $y(k)$ are scalar. The use of multiple antenna elements ($p > 1$), are however of increasing interest, see Example 5.1 in Sect. 5.2.4. An appropriate model of $\boldsymbol{\mathcal{G}}$ is then a column vector of FIR channels, (5.2.29), i.e. a vector of polynomials. The polynomial coefficients are normally estimated based on a short training sequence $\{u(k)\}$ which is known to the receiver. Estimation errors are inevitable. Furthermore, the channel coefficients will be slowly time-varying during the time intervals between training events in, for example, the GSM system [38].

The task of a (robust) equalizer is then to estimate $u(k)$, based on $y(k+m)$, a nominal model $\boldsymbol{\mathcal{G}}_o$, and an estimate of the model uncertainty $\Delta\boldsymbol{\mathcal{G}}$. □

5.3.3 Parametrization of the Extended Design Model

As in Sect. 5.2.3, we choose to parametrize $\boldsymbol{\mathcal{G}}$ and $\boldsymbol{\mathcal{H}}$ as left MFDs having *diagonal* denominators, while $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{W}}$ are parameterized in common denominator form

$$\boldsymbol{\mathcal{G}} = \boldsymbol{A}^{-1}\boldsymbol{B} \;\; ; \;\; \boldsymbol{\mathcal{H}} = \boldsymbol{N}^{-1}\boldsymbol{M} \tag{5.3.5}$$

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

It will be assumed that $\boldsymbol{\mathcal{G}}$, $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{F}}$ may be uncertain. The weighting matrix

$\mathcal{W}$ is assumed to be exactly known. It is shown in [4] that uncertainty in $\mathcal{D}$ does not affect the optimal filter design, provided it is uncorrelated to uncertainties in other blocks. Therefore, uncertainty in $\mathcal{D}$ is not introduced.

The extended design models, cf. (5.3.1), which consist of a nominal model and an uncertainty model according to

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

These models 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

$$\begin{aligned}
\mathcal{G} &= \boldsymbol{A}_o^{-1}\boldsymbol{B}_o + \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\Delta\boldsymbol{B} = \boldsymbol{A}_o^{-1}\boldsymbol{A}_1^{-1}(\hat{\boldsymbol{B}}_o + \hat{\boldsymbol{B}}_1\Delta\boldsymbol{B}) \triangleq \boldsymbol{A}^{-1}\boldsymbol{B} \\
\mathcal{H} &= \boldsymbol{N}_o^{-1}\boldsymbol{M}_o + \boldsymbol{N}_1^{-1}\boldsymbol{M}_1\Delta\boldsymbol{M} = \boldsymbol{N}_o^{-1}\boldsymbol{N}_1^{-1}(\hat{\boldsymbol{M}}_o + \hat{\boldsymbol{M}}_1\Delta\boldsymbol{M}) \triangleq \boldsymbol{N}^{-1}\boldsymbol{M} \\
\mathcal{F} &= \frac{1}{D_o}\boldsymbol{C}_o + \frac{1}{D_1}\boldsymbol{C}_1\Delta\boldsymbol{C} = \frac{1}{D_oD_1}(\hat{\boldsymbol{C}}_o + \hat{\boldsymbol{C}}_1\Delta\boldsymbol{C}) \triangleq \frac{1}{D}\boldsymbol{C} \quad .
\end{aligned} \tag{5.3.6}$$

Above, $\mathcal{G}_o = \boldsymbol{A}_o^{-1}\boldsymbol{B}_o$ represents the nominal model and $\Delta\mathcal{G} = \boldsymbol{A}_1^{-1}\boldsymbol{B}_1\Delta\boldsymbol{B}$ the error model. The same holds for $\mathcal{H}$ and $\mathcal{F}$. 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 polynomial matrices $\boldsymbol{A}_1$ and $\boldsymbol{N}_1$ and the polynomial 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 are used to fit the model class to the set of true systems. One particular modelling error is represented by one particular realization of the random coefficients.[16] An element ij of a stochastic polynomial matrix $\Delta\boldsymbol{P}$ is denoted

$$\Delta P^{ij} \triangleq [\Delta\boldsymbol{P}]_{ij} = \Delta p_o^{ij} + \Delta p_1^{ij}q^{-1} + \ldots + \Delta p_{\delta p}^{ij}q^{-\delta p} \tag{5.3.7}$$

where δp is the degree of $\Delta\boldsymbol{P}$. 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,

[16] For a given system realization, the random coefficients are assumed time-invariant and independent of the time-series $e(k)$ and $v(k)$. This is in contrast to the approach of Haddad and Bernstein in [102], who represent the effect of uncertainties by multiplicative noises. For a given uncertainty variance, a noise representation would under-estimate the true effect of (time-invariant) parameter deviations on the dynamics.

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 $\mathbf{P}_{\Delta\mathbf{P}}^{(ij,\ell k)}$, discussed further in Sect. 5.3.5.

We now introduce the following assumption.

Assumption 5.3. *The coefficients of all polynomial elements of $\Delta\boldsymbol{C}$ are independent of those of $\Delta\boldsymbol{B}$.*

It is possible to exclude Assumption 5.3, but it does simplify the solution by eliminating the need of taking fourth order moments with respect to elements of $\Delta\boldsymbol{C}$ and $\Delta\boldsymbol{B}$ into account. The assumption is reasonable in most practical cases.

5.3.4 Obtaining Error Models

Error models can be obtained from ordinary identification experiments, provided the model structures match. SISO transfer function models can also be obtained in the presence of undermodelling, using a maximum likelihood approach [12]. We shall next outline various ways to adjust error models to the variability of the dynamics within sets of possible SISO-systems.

Obtaining Extended Design Models from Identification. Model error estimates are obtained from many types of identification algorithms, for example prediction error methods. In a Bayesian setting, a model error estimate can be said to represent the characteristics of a possible set of true systems, which might have generated the data used for identification. It is conventional to decompose the estimation error into a variance error, caused by noise in finite data sets, and a bias error, which would remain even for infinite data sets. The bias error is caused by the selection of an inappropriate model structure. Model structure selection can sometimes be difficult, but is aided by systematic procedures for model validation [104].

If the model error caused by bias is small, and if the data series is not too short, then the estimated parameter covariance matrix of an identified model provides acceptable estimates of the modelling errors.[17]

For model structures with denominator polynomials, such as AR, ARX and ARMAX models [103, 104], the estimated model uncertainty of denominator coefficients will have to be transformed into an additive error model by series expansion. Methods for series expansion are discussed in detail in [37]. In general, first or modified second order expansions will provide an approximation with sufficient accuracy. The estimation of error models for

[17] For models of time-invariant systems, which pass standard validation tests, the variance error will, in general, dominate the bias error, see [39]. A reasonable, but perhaps somewhat conservative, estimate of the total model error is then obtained by doubling the parameter covariance matrix, which is a measure of the variance error.

ARMAX structures, based on short data records generated by high order systems, has been investigated recently by Bigi in [42].

For a system with measurable inputs, one way of directly obtaining extended design models of the type (5.3.6) is from identification experiments based on functional series expansions $\sum_{i=1}^{M} p_i\mathcal{B}_i(q^{-1})$. Here, $\mathcal{B}_i, i = 1\ldots M$ represents a set of predetermined rational basis functions, such as, e.g. discrete Laguerre functions. A functional series model is linear in the parameters $\{p_i\}$. The model structure has received increasing interest as a useful tool in system identification, see [105] or [106]. If an identification experiment provides parameter estimates $\{\bar{p}_{0i}\}$ and covariances for zero mean errors $\{\Delta\bar{p}_i\}$, we directly obtain the extended design model

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

Writing $\Delta\bar{\mathcal{P}}$ in common denominator form and using the covariance matrix for $\{\Delta\bar{p}_i\}$ gives the frequency domain variance $\bar{E}(\Delta\bar{\mathcal{P}}\Delta\bar{\mathcal{P}}_*)$, which will be needed in the robust design. See Chapter 5 of [107].

Adjustment to Sets of Spectra or Nyquist Plots. Error models representing nonparametric uncertainties 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 [93].

We will next very briefly recapitulate their stochastic embedding concept. An additive transfer function error $\Delta\mathcal{G}(e^{i\omega})$ is viewed as a realization of a stochastic process in the frequency domain, with zero mean and covariance function

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

For stationary processes, the covariance depends only on the difference in frequency, $\Gamma(e^{i\omega_1}, e^{i\omega_2}) = \Gamma_s(e^{i(\omega_1-\omega_2)})$. The shape of Γ_s is a measure of the assumed frequency domain smoothness of 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

$$\Delta\mathcal{G}(q^{-1}) = \sum_{j=0}^{\infty} g_j q^{-j} \ ; \qquad \bar{E}(g_j, g_\ell) = \gamma(j,\ell) \ . \tag{5.3.8}$$

Here, $\gamma(j,\ell)$ can be calculated from the inverse two-dimensional 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 process will be white, with

$$\bar{E}(g_j, g_\ell) = \gamma_j\delta_{j,\ell} \ . \tag{5.3.9}$$

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

$$\bar{E}\{\mathcal{H}(e^{i\omega_1})\mathcal{H}_*(e^{i\omega_2})\} = \frac{\alpha e^{i(\omega_1-\omega_2)}}{e^{i(\omega_1-\omega_2)} - \lambda} \ . \tag{5.3.10}$$

This process corresponds to the time domain model (5.3.8), (5.3.9), with Gaussian distributed independent parameters with variances $\gamma_j = \alpha\lambda^j$. See [12] by Goodwin *et al.* 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 $\bar{E}(h_j)^2 = \alpha\lambda^j$ and $\bar{E}(h_j) = 0$.

A priori information may be available about the frequency domain distribution of the unmodelled dynamics. It can be incorporated by using a fixed prefilter, to obtain the total model

$$\Delta\mathcal{G}(q^{-1}) = \mathcal{M}(q^{-1})\mathcal{N}'_{\Delta}(q^{-1}) \ . \tag{5.3.11}$$

Here, $\mathcal{M}(q^{-1})$ is a known shaping filter and $\mathcal{N}'_{\Delta}(q^{-1})$ is a stationary process in the frequency domain, with covariance function $\Gamma_s(e^{i(\omega_1-\omega_2)})$. Examples of the use of this modelling procedure can be found in [34, 37] and [93].

Example 5.5. *A frequency-shaped error model.*

Assume that the variance in the frequency domain of the model error in can be described by a squared magnitude response of a first order filter. Then, we may use a model of the type (5.3.11), with

$$\mathcal{M}(q^{-1}) = \frac{1+\eta q^{-1}}{1+a_{11}q^{-1}} \ . \tag{5.3.12}$$

Also, assume that the parameter λ in (5.3.10) 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 (5.3.10) 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 (5.3.11). The error model has the structure introduced in (5.3.6)

$$\begin{aligned}\Delta\mathcal{G}(q^{-1}) &= \frac{1+\eta q^{-1}}{1+a_{11}q^{-1}}(h_o + h_1 q^{-1} + \ldots + h_M q^{-M}) \\ &= \frac{B_1(q^{-1})\Delta B(q^{-1})}{A_1(q^{-1})} \ . \end{aligned} \tag{5.3.13}$$

The covariance matrix of $\{h_j\}$ is $\mathbf{P}_{\Delta B} = \mathrm{diag}(\alpha\lambda^j)$. Note that the model is characterized by only five parameters: α, λ, a_{11}, η and the truncation length M. □

Pragmatic Tuning of Covariances. 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. One objective could be to obtain as good a performance as possible, under the constraint of a prespecified level of degradation in the nominal case. Another objective could be to limit the maximal error within a specified range of model dynamics. The error models may also be used to account for a slowly time-varying dynamics, see [38].

5.3.5 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

$$\Delta P^{ij}(q^{-1}) = \varphi^T(q^{-1})\bar{p}_{ij} \tag{5.3.14}$$

where

$$\varphi^T(q^{-1}) = (1 \quad q^{-1} \dots q^{-\delta p}) \quad ; \quad \bar{p}_{ij} = (\Delta p_o^{ij} \quad \Delta p_1^{ij} \dots \Delta p_{\delta p}^{ij})^T \quad . \tag{5.3.15}$$

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

$$\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)} = \bar{E}\,\bar{p}_{ij}\bar{p}_{\ell k}^{\,*} = \begin{bmatrix} \bar{E}(\Delta p_o^{ij})(\Delta p_o^{\ell k})^* & \dots & \bar{E}(\Delta p_o^{ij})(\Delta p_{\delta p}^{\ell k})^* \\ \vdots & \ddots & \vdots \\ \bar{E}(\Delta p_{\delta p}^{ij})(\Delta p_o^{\ell k})^* & \dots & \bar{E}(\Delta p_{\delta p}^{ij})(\Delta p_{\delta p}^{\ell k})^* \end{bmatrix} \tag{5.3.16}$$

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,

$$\bar{E}(\Delta P^{ij}\Delta P_*^{\ell k}) = \bar{E}(\varphi^T(q^{-1})\bar{p}_{ij}\bar{p}_{\ell k}^*\varphi_*^T(q)) = \varphi^T\mathbf{P}_{\mathbf{\Delta P}}^{(ij,\ell k)}\varphi_*^T \quad . \tag{5.3.17}$$

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 (5.3.16) into one large covariance matrix, organized as $\mathbf{P}_{\Delta\mathbf{P}} =$

$$\begin{pmatrix} \begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(11,11)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(11,1m)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(1m,11)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(1m,1m)} \end{bmatrix} & \dots & \begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(11,n1)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(11,nm)} \\ \vdots & & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(1m,n1)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(1m,nm)} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(n1,11)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(n1,1m)} \\ \vdots & & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(nm,11)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(nm,1m)} \end{bmatrix} & \dots & \begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(n1,n1)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(n1,nm)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(nm,n1)} & \dots & \mathbf{P}_{\Delta\mathbf{P}}^{(nm,nm)} \end{bmatrix} \end{pmatrix} . \tag{5.3.18}$$

If $\Delta\boldsymbol{P}$ has dimension $n|m$, then $\mathbf{P}_{\Delta\mathbf{P}}$ has nm by nm blocks $\mathbf{P}_{\Delta\mathbf{P}}^{(ij,\ell k)}$. The structure of (5.3.18) is useful from a design point of view. If, for example, a multivariable moving average model, or FIR model, is to be identified, then (5.3.18) is the natural way of representing the covariance matrix. If we instead prefer to use the blocks $\mathbf{P}_{\Delta\mathbf{P}}^{(ij,\ell k)}$ of (5.3.18) as multivariable "tuning knobs", a given amount of uncertainty can be assigned to a specific input-output pair.

5.3.6 Design of the Cautious Wiener Filter

Nominal Wiener filter design involves two key elements, namely a spectral factorization and a Diophantine equation. These two elements will remain crucial also when models are uncertain. However, they will now be based on the average behaviour with respect to an underlying set of systems.

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 $y(k)$ in (5.3.3) 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

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \bar{E}\{\boldsymbol{NBCC}_*\boldsymbol{B}_*\boldsymbol{N}_* + D\boldsymbol{AMM}_*\boldsymbol{A}_*D_*\} \quad . \tag{5.3.19}$$

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

$$\bar{y}(k) = \frac{1}{D}\boldsymbol{A}^{-1}\boldsymbol{N}^{-1}\boldsymbol{\beta}\,\epsilon(k) \tag{5.3.20}$$

where $\epsilon(k)$ is white with a unit covariance matrix. This model is denoted 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 (5.3.19), the following results are useful.

Lemma 5.1. *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,*

$$\bar{E}[\boldsymbol{G}\boldsymbol{H}\boldsymbol{G}_*] = \bar{E}[\boldsymbol{G}\bar{E}(\boldsymbol{H})\boldsymbol{G}_*] \tag{5.3.21}$$

□

Proof. See [4].

Now, introduce the double-sided polynomial matrices

$$\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* \triangleq \bar{E}(\boldsymbol{C}\boldsymbol{C}_*) \;\; ; \;\; \tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} \triangleq \bar{E}(\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\boldsymbol{B}_*) \;\; ; \;\; \tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_* \triangleq \bar{E}(\boldsymbol{M}\boldsymbol{M}_*) \; . \tag{5.3.22}$$

Invoking (5.3.6) and using the fact that the stochastic coefficients are assumed to be zero mean, gives

$$\begin{aligned}
\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_* &= \hat{\boldsymbol{C}}_o\hat{\boldsymbol{C}}_{o*} + \hat{\boldsymbol{C}}_1\bar{E}(\Delta\boldsymbol{C}\Delta\boldsymbol{C}_*)\hat{\boldsymbol{C}}_{1*} \\
\tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*} &= \hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*} + \hat{\boldsymbol{B}}_1\bar{E}(\Delta\boldsymbol{B}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\Delta\boldsymbol{B}_*)\hat{\boldsymbol{B}}_{1*} \\
\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_* &= \hat{\boldsymbol{M}}_o\hat{\boldsymbol{M}}_{o*} + \hat{\boldsymbol{M}}_1\bar{E}(\Delta\boldsymbol{M}\Delta\boldsymbol{M}_*)\hat{\boldsymbol{M}}_{1*} \; .
\end{aligned} \tag{5.3.23}$$

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 5.2. *Let Assumption 5.3 hold. By using (5.3.22), (5.3.23) and invoking Lemma 5.1, the averaged spectral factorization (5.3.19) can be expressed as*

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{N}\tilde{\boldsymbol{B}}_C\tilde{\boldsymbol{B}}_{C*}\boldsymbol{N}_* + D\boldsymbol{A}\tilde{\boldsymbol{M}}\tilde{\boldsymbol{M}}_*\boldsymbol{A}_*D_* \tag{5.3.24}$$

□

Proof. See [4].

With a given right-hand side, equation (5.3.24) is just an ordinary polynomial matrix left spectral factorization, of the type encountered in (5.2.12).

It is solvable under the following mild assumption.

Assumption 5.4. *The averaged spectral density matrix $\bar{E}\{\mathbf{\Phi}_y(e^{i\omega})\}$ is nonsingular for all ω.*

This assumption is equivalent to the right-hand side of (5.3.24) being nonsingular on $|z| = 1$. Then, the solution to (5.3.24) is unique, up to a right orthogonal factor. Under Assumption 5.4, 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 (5.3.24).

To obtain the right-hand side of (5.3.24), averaged polynomial matrices like $\bar{E}(\Delta \boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*)$ have to be computed, where $\boldsymbol{H}(q,q^{-1}) = \tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*$ or $\mathbf{I}$. It is shown in [4] that the ij-th element of $\bar{E}(\Delta \boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*)$ is given by

$$\bar{E}[\Delta \boldsymbol{P}\boldsymbol{H}\Delta\boldsymbol{P}_*]_{ij} \quad =$$

$$\mathrm{tr}\boldsymbol{H}\begin{bmatrix} \varphi^T & & 0 \\ & \ddots & \\ 0 & & \varphi^T \end{bmatrix}\begin{bmatrix} \mathbf{P}_{\Delta\mathbf{P}}^{(i1,j1)} & \cdots & \mathbf{P}_{\Delta\mathbf{P}}^{(im,j1)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\Delta\mathbf{P}}^{(i1,jm)} & \cdots & \mathbf{P}_{\Delta\mathbf{P}}^{(im,jm)} \end{bmatrix}\begin{bmatrix} \varphi_*^T & & 0 \\ & \ddots & \\ 0 & & \varphi_*^T \end{bmatrix} \tag{5.3.25}$$

where φ^T is defined in (5.3.15). The block covariance matrix in (5.3.25) is obtained by taking the block-transpose of the ij-th block $[\ \cdot\ ]$ of $\mathbf{P}_{\Delta\mathbf{P}}$ in (5.3.18). Average factors in (5.3.23) are readily obtained by substituting $\Delta\boldsymbol{C}$, $\Delta\boldsymbol{B}$ and $\Delta\boldsymbol{M}$ for $\Delta\boldsymbol{P}$ in (5.3.25).

We are now ready to present the solution to the robust $\mathcal{H}_2$ filter design problem.

The Cautious Multivariable Wiener Filter.

Theorem 5.2. *Assume an extended design model (5.3.3), (5.3.5), (5.3.6), to be given, with known covariance matrices (5.3.18). Under Assumptions 5.3 and 5.4, a realizable estimator of $z(k)$ then minimizes the averaged MSE (5.2.5), among all linear time-invariant estimators based on $y(k+m)$, if and only if it has the same coprime factors as*

$$\hat{z}(k|k+m) = \boldsymbol{\mathcal{R}}\, y(k+m) = \frac{1}{T}\boldsymbol{V}^{-1}\boldsymbol{Q}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A}\, y(k+m) \ . \tag{5.3.26}$$

Here, $\boldsymbol{\beta}(q^{-1})$ is obtained from (5.3.24), while $\boldsymbol{Q}(q^{-1})$ together with $\boldsymbol{L}_*(q)$, both of dimensions $\ell|p$, is the unique solution to the unilateral Diophantine equation

$$q^{-m}\boldsymbol{V}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_* = \boldsymbol{Q}\boldsymbol{\beta}_* + q\boldsymbol{L}_*UTD\,\mathbf{I}_p \tag{5.3.27}$$

with generic[18] *degrees*

$$\begin{aligned} nQ &= \max(n\tilde{v} + ns + n\tilde{c} + m, nu + nt + nd - 1) \\ nL_* &= \max(n\tilde{c} + n\hat{b}_o + nn - m, n\beta) - 1 \end{aligned} \tag{5.3.28}$$

where $ns = \deg \boldsymbol{S}$ *etc. When applying the estimator (5.3.26) on an ensemble of systems, the minimal criterion value becomes*

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

$$+ \frac{1}{UTDD_*T_*U_*}\boldsymbol{V}\boldsymbol{S}\tilde{\boldsymbol{C}}\left[\mathbf{I}_n - \tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\boldsymbol{C}\right]\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\boldsymbol{V}_*\Big\}\frac{dz}{z} \tag{5.3.29}$$

□

Proof. See [4], where the variational approach outlined in Sect. 5.2.3 is utilized, with $\bar{E}E(\cdot)$ substituted for $E(\cdot)$. In the case considered here, with a known filter $\boldsymbol{\mathcal{W}}$, it is straightforward to show that the cautious Wiener filter minimizes not only the scalar averaged MSE, but also the average covariance matrix $\bar{E}E\varepsilon(k)\varepsilon(k)^*$.

Remarks. Note the very close formal similarity of (5.3.26), (5.3.27) and (5.3.29) to their nominal counterparts (5.2.19), (5.2.20) and (5.2.23), respectively. The only new required type of computation, as compared to the nominal case described in Sect. 5.2.3, is the calculation of averaged polynomials in (5.3.23) performed by using (5.3.25). Design examples can be found in [4] and in [37].

Since both $\boldsymbol{V}$ and $\boldsymbol{\beta}$ are stable, the estimator $\boldsymbol{\mathcal{R}}$ will be stable[19]. Since $\boldsymbol{V}(0)$ and $\boldsymbol{\beta}(0)$ are nonsingular, $\boldsymbol{\mathcal{R}}$ 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 (5.3.26). 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 (5.2.19) has a whitening filter as a right factor. The robust estimator has a similar structure. After multiplying $\boldsymbol{\mathcal{R}}$ by the stable common factor D/D, the cautious filter (5.3.26) will contain $\boldsymbol{\beta}^{-1}\boldsymbol{NA}D$

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

[19] Stable common factors may exist in (5.3.26). 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.

as right factor. This averaged counterpart of a whitening filter is the inverse of the averaged innovations model (5.3.20).

The model structure (5.3.5)-(5.3.6) 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 [37], but it provides less physical insight. It does also exhibit worse numerical behaviour, since algorithms for coprime factorization are numerically sensitive.

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 (5.3.27) numerically sensitive. 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 (5.2.12). This reduces the numerical difficulty of solving both (5.3.24) and (5.3.27).

The Equivalent-Noise Interpretation. For every cautious Wiener filter, there exists a system (without uncertainty) for which this estimator is the optimal Wiener filter, see [37]. For example, if $\boldsymbol{\mathcal{G}} = \mathbf{I}_p$, then we can utilize the modified signal and noise spectral density matrices

$$\boldsymbol{\mathcal{F}}_o\boldsymbol{\mathcal{F}}_{o*} + \bar{E}(\Delta\boldsymbol{\mathcal{F}}\Delta\boldsymbol{\mathcal{F}}_*) \quad ; \quad \boldsymbol{\mathcal{H}}_o\boldsymbol{\mathcal{H}}_{o*} + \bar{E}(\Delta\boldsymbol{\mathcal{H}}\Delta\boldsymbol{\mathcal{H}}_*) \tag{5.3.30}$$

to obtain the averaged innovations model. The spectral densities (5.3.30) may be obtained directly from frequency domain data if such are available, by simply averaging the spectral densities of the model sets $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{H}}$.

It is also possible to represent model uncertainties by coloured noises, and then to design a Wiener filter for the corresponding system. This correspondence provides a way of understanding the structure of the above design equations. However, we do not recommend the use of an equivalent noise-approach in the actual design, for two reasons:

- It is far from trivial to obtain an equivalent noise representation of the uncertainties in the block $\boldsymbol{\mathcal{G}}$, with appropriate colour and covariance structure. This is true in particular if the block $\boldsymbol{\mathcal{F}}$ is also uncertain, and if the problem is multivariable.

- It is an advantage from a design point of view to have separate tools which handle different aspects of the design: Error models to represent the effect of modelling uncertainty; noise models to represent disturbances; criterion weighting functions to reflect the priorities of the user. A method which does not distinguish between these aspects will tend to confuse the designer.

The Attainable Limit of Estimator Performance. 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 5.1. *The limiting estimator for $m \to \infty$, the non-realizable cautious Wiener filter, can be expressed as*

$$\lim_{m\to\infty} q^m \boldsymbol{\mathcal{R}} = \frac{1}{T}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{A} \quad . \tag{5.3.31}$$

Its average performance is given by (5.3.29) with $\boldsymbol{L} = \mathbf{0}$. If $\boldsymbol{\mathcal{W}} = \mathbf{I}_\ell$, the trace of the spectral density of the lower bound of the estimation error $z(k)-\hat{z}(k|k+m)$ is

$$\lim_{m\to\infty} \mathrm{tr}\bar{E}E\ \boldsymbol{\Phi}_{z-\hat{z}}(e^{i\omega}) =$$

$$= \frac{1}{TDD_*T_*}\mathrm{tr}\left\{\boldsymbol{S}\tilde{\boldsymbol{C}}\left[\mathbf{I}_n - \tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\hat{\boldsymbol{B}}_o\tilde{\boldsymbol{C}}\right]\tilde{\boldsymbol{C}}_*\boldsymbol{S}_*\right\} \quad . \tag{5.3.32}$$

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

$$U(e^{-i\omega_1}) \approx 0 \tag{5.3.33}$$

□

Proof. In a similar way as in Appendix A.3 of [59], it is straightforward to show that $\boldsymbol{L} \to \mathbf{0}$ as $m \to \infty$ in (5.3.27). Thus, (5.3.27) gives

$$\lim_{m\to\infty} q^m \boldsymbol{Q} = \boldsymbol{V}\boldsymbol{S}\tilde{\boldsymbol{C}}\tilde{\boldsymbol{C}}_*\hat{\boldsymbol{B}}_{o*}\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1} \quad . \tag{5.3.34}$$

The substitution of this expression into (5.3.26) gives (5.3.31). The use of $\boldsymbol{L} = \mathbf{0}$, $\boldsymbol{V} = \mathbf{I}_\ell$ and $U = 1$ in the integrand of (5.3.29) gives (5.3.32). When $U(e^{-i\omega_1}) \approx 0$, we obtain the same effect on the Diophantine equation (5.3.27) 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 (5.3.34) and the estimation error approaches the lower bound (5.3.32). ■

Remark. Note that for realizable estimators (m finite) the lower bound (5.3.32) is only attainable at distinct frequencies ω_i by means of frequency weighting. For frequencies outside the bandwidth of $\boldsymbol{\mathcal{W}}$, the estimate may be severely degraded.

5.4 Robust $\mathcal{H}_2$ Filter Design Based on State-Space Models with Parametric Uncertainty

Let us now consider state space models with parametric uncertainty obtained, for example, by physical modelling. The section illustrates the utility of combining state space and polynomial representations. It summarizes results from [36] and [37]. Extended design models of the type (5.3.1) will be obtained by a series expansion methodology outlined in Sect. 5.4.1 below, and discussed in more detail in [37]. The method is an improvement upon a similar suggestion by Speyer and Gustafsson [32], in that ℓ'th order expansions will lead to modified state space models of order $n(\ell+1)$, rather than $n(\ell+1)^2$. The subsequent design of an averaged robust $\mathcal{H}_2$ estimator can be performed either by using the cautious Wiener estimator of Theorem 5.2 or by designing a robustified Kalman estimator, as outlined in Sect. 5.4.2.

The results of [36] are here generalized slightly to include time-varying measurement equations. The resulting estimators are then directly applicable as robust adaptive algorithms, which can be applied for tracking the parameters of linear regression models, as discussed briefly in Sect. 5.5.

5.4.1 Series Expansion

Assume a set of stable discrete-time models

$$\begin{aligned} x(k+1) &= (\mathbf{A}_0 + \Delta\mathbf{A}(\rho))x(k) + (\mathbf{B}_0 + \Delta\mathbf{B}(\rho))e(k) \\ y(k) &= \mathbf{C}(k)x(k) + (\mathbf{M}_0 + \Delta\mathbf{M}(\rho))v(k) \\ z(k) &= \mathbf{L}x(k) \end{aligned} \tag{5.4.1}$$

where $x(k) \in \mathbf{R}^n$ is the state vector and $e(k) \in \mathbf{R}^{n_e}$ is zero mean process noise with unit covariance matrix. The output $y(k) \in \mathbf{R}^p$ is the measurement signal, with $v(k) \in \mathbf{R}^p$ being white zero mean noise having unit covariance matrix. The vector $z(k) \in \mathbf{R}^l$ is to be estimated. The nominal model is

$$x_0(k+1) = \mathbf{A}_0 x_0(k) + \mathbf{B}_0 e(k) \tag{5.4.2}$$

$$y_0(k) = \mathbf{C}(k)x_0(k) + \mathbf{M}_0 v(k) \quad .$$

We assume the matrices $\Delta\mathbf{A}, \Delta\mathbf{B}$ and $\Delta\mathbf{M}$ to be known functions of the unknown parameter vector ρ. The vector ρ may, for example, contain uncertain physical parameters of a continuous-time model. The robust estimation of $z(k)$ will be founded on the following assumptions:

- The uncertain parameters ρ are treated as if they were stochastic variables. Their realizations represent particular models in the set.

- All models (5.4.1) are assumed internally and BIBO stable. In other words, the eigenvalues of $\mathbf{A}_0 + \Delta\mathbf{A}(\rho)$ are located in $|z| < 1$ for all admissible ρ, and the elements of $\mathbf{C}(k)$ are bounded.
- The effect of ρ on the set of models (5.4.1) is described by known covariances between elements of the matrices $\Delta\mathbf{A}, \Delta\mathbf{B}$ and $\Delta\mathbf{M}$. The nominal model (5.4.2) is selected as the average model of the set; $\Delta\mathbf{A}, \Delta\mathbf{B}$ and $\Delta\mathbf{M}$ have mean value zero.

The aim is to obtain an approximate modified Kalman estimator which minimizes the criterion (5.2.5), i.e. the average, over the set of models, of the mean square estimation error.

In order to apply the framework of Sect. 5.3, a model with uncertainties in the system matrix must be approximated by a new model, in which the uncertainties appear only in the input matrix. One way of doing this is to use series expansion, based on the denominator terms of a transfer function representation of (5.4.1). See Sect. 5 of [34] or Chapter 3 of [37]. Here we shall, instead, perform the expansion directly in the state space representation, by augmenting the nominal state vector $x_0(k+1)$ by additional vectors. These vectors correspond to sets of perturbations caused by the different powers of $\Delta\mathbf{A}$ occurring in a series expansion.

Introduce the set of possible state trajectory variations $\delta x(k+1)$, caused by $\Delta\mathbf{A}(\rho)$ and $\Delta\mathbf{B}(\rho)$, such that

$$x(k+1) = x_0(k+1) + \delta x(k+1) \tag{5.4.3}$$

where x_0 is the nominal state vector given by (5.4.2) and

$$\delta x(k+1) = \Delta\mathbf{A}x_0(k) + \mathbf{A}_0\delta x(k) + \Delta\mathbf{B}e(k) + \Delta\mathbf{A}\delta x(k) \ . \tag{5.4.4}$$

The equality (5.4.4) is an exact expression derived from (5.4.1). We now express $\delta x(k)$ in (5.4.4) as

$$\delta x(k) = x_1(k) + x_2(k) + \ldots + x_d(k)$$

for a given expansion order d. The terms $x_m(k), m < d$ are defined as being affected by powers of $\Delta\mathbf{A}$ up to m only. Specifying state equations for the additional state vectors, $x_m(k)$, is now a matter of pairing terms $x_m(k+1)$ on the left-hand side of (5.4.4) with appropriate terms on the right-hand side. The choice

$$\begin{aligned}
x_1(k+1) &= \Delta\mathbf{A}x_0(k) + \mathbf{A}_0 x_1(k) + \Delta\mathbf{B}e(k) \\
x_2(k+1) &= \mathbf{A}_0 x_2(k) + \Delta\mathbf{A}x_1(k) \\
&\vdots \\
x_d(k+1) &= \mathbf{A}_0 x_d(k) + \Delta\mathbf{A}x_{d-1}(k) + \Delta\mathbf{A}x_d(k)
\end{aligned}$$

yields the augmented state space model

$$\begin{bmatrix} x_0(k+1) \\ x_1(k+1) \\ \vdots \\ x_d(k+1) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{A}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \Delta\mathbf{A} & \mathbf{A}_0 & & \vdots \\ & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & & \Delta\mathbf{A} & \mathbf{A}_0 + \Delta\mathbf{A} \end{bmatrix}}_{\bar{\mathbf{A}}} \begin{bmatrix} x_0(k) \\ x_1(k) \\ \vdots \\ x_d(k) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_0 \\ \Delta\mathbf{B} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} e(k)$$

$$x(k) = x_0(k) + x_1(k) + \dots x_d(k) \; . \tag{5.4.5}$$

So far, no approximation has been made. The term $\Delta\mathbf{A}$ in the lower right corner of $\bar{\mathbf{A}}$ represents the effect of $(d+1)$'th and higher powers of $\Delta\mathbf{A}$ on $x(k)$. We *neglect this term* from now on, and thus discard terms of higher order than d. The characteristic polynomial is then given by

$$\det(z\mathbf{I}_{(d+1)n} - \bar{\mathbf{A}}) = \det(z\mathbf{I}_n - \mathbf{A}_0)^{d+1}$$

so perturbations will no longer affect any transfer function denominator.

To keep the notation simple, we shall in the sequel specialize to first order expansions. By transforming to an input-output model in forward shift operator form, we then obtain

$$x(k) = [\; \mathbf{I}_n \;\; \mathbf{I}_n \;] \begin{bmatrix} q\mathbf{I}_n - \mathbf{A}_0 & \mathbf{0} \\ -\Delta\mathbf{A} & q\mathbf{I}_n - \mathbf{A}_0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B}_0 \\ \Delta\mathbf{B} \end{bmatrix} e(k) \; . \tag{5.4.6}$$

Now, introduce the $n|n$ polynomial matrices $\widetilde{\boldsymbol{D}}(q)$ and $\widetilde{\Delta\boldsymbol{A}}(q)$ as a solution to the coprime factorization

$$\widetilde{\boldsymbol{D}}(q)\Delta\mathbf{A} = \widetilde{\Delta\boldsymbol{A}}(q)(q\mathbf{I}_n - \mathbf{A}_0) \tag{5.4.7}$$

where $\widetilde{\boldsymbol{D}}(q)$ should contain no stochastic coefficients[20] and $\deg\det\widetilde{\boldsymbol{D}} = n$. Then, (5.4.6) can be written as a left matrix fraction description

$$\begin{aligned} x(k) &= \\ &(q\mathbf{I}_n - \mathbf{A}_0)^{-1}(\mathbf{B}_0 + \Delta\mathbf{B} + \Delta\mathbf{A}(q\mathbf{I}_n - \mathbf{A}_0)^{-1}\mathbf{B}_0)e(k) \\ &= (q\mathbf{I}_n - \mathbf{A}_0)^{-1}\widetilde{\boldsymbol{D}}^{-1}(q)(\widetilde{\boldsymbol{D}}(q)(\mathbf{B}_0 + \Delta\mathbf{B}) + \widetilde{\Delta\boldsymbol{A}}(q)\mathbf{B}_0)e(k) \\ &\triangleq \boldsymbol{D}^{-1}(q)\boldsymbol{C}(q)e(k) \end{aligned} \tag{5.4.8}$$

[20]This step is superfluous if the original system is realized in diagonal form. As explained in [37], the factorization actually corresponds to a polynomial matrix spectral factorization. A d'th order expansion will require d factorizations of the type (5.4.7).

where $\deg\det \boldsymbol{D}(q) = 2n$. This representation is of the form (5.3.1), with

$$\begin{aligned}\boldsymbol{\mathcal{F}}_o &= (q\mathbf{I}_n - \mathbf{A}_0)^{-1}\mathbf{B}_0 \\ \Delta\boldsymbol{\mathcal{F}} &= (q\mathbf{I}_n - \mathbf{A}_0)^{-1}\widetilde{\boldsymbol{D}}^{-1}(q)(\widetilde{\boldsymbol{D}}(q)\Delta\mathbf{B} + \widetilde{\Delta\boldsymbol{A}}(q)\mathbf{B}_0) \ .\end{aligned}$$

It can easily be converted to backward shift operator form. The input-output representation could be complemented by stochastic additive error models which represent unmodelled higher-order dynamics.

5.4.2 The Robust Linear State Estimator

The model (5.4.8) can form the basis of a robust Wiener filter design, as described in Sect. 5.3, in which also uncertainty in a time-invariant matrix $\mathbf{C}$ of (5.4.1) can be handled. If we prefer to work with state space estimators, the set of models (5.4.8) can be realized on observable state space form [55], with $2n$ states:

$$\xi(k+1) = \mathbf{F}\xi(k) + (\mathbf{G}_0 + \Delta\mathbf{G})e(k) \ ; \ x(k) = \mathbf{H}\xi(k) \tag{5.4.9}$$

where $\Delta\mathbf{G}$ has zero mean. Note that since the denominator matrix $\boldsymbol{D}(q)$ of (5.4.8) contains no uncertain coefficients, neither will $\mathbf{F}$ in (5.4.9). The covariance matrix of the uncertain elements of $\Delta\mathbf{G}$ in (5.4.9) can be calculated straightforwardly from the covariances of the elements of $\Delta\mathbf{A}$ and $\Delta\mathbf{B}$ in (5.4.1).

Let us restrict attention to linear estimators[21]. The model (5.4.9) can now be utilized for designing robust Kalman predictors, filters and smoothers, using well-known techniques [56]. For example, it can be shown, see [37] Chapter 7, that if $\Delta\mathbf{M}$ is independent of $\Delta\mathbf{A}, \Delta\mathbf{B}$ and if $e(k)$ is uncorrelated to $v(s)$ for all k, s, then the one-step predictor minimizing (5.3.2) is given by

$$\begin{aligned}\hat{\xi}(k+1) &= \mathbf{F}\hat{\xi}(k) + \mathbf{K}(k)(y(k) - \mathbf{C}_1(k)\hat{\xi}(k)) \\ \hat{z}(k+1) &= \mathbf{L}\mathbf{H}\hat{\xi}(k+1)\end{aligned}$$

with $\mathbf{K}(k)$ calculated from

$$\begin{aligned}\mathbf{K}(k) &= \mathbf{F}\mathbf{P}(k)\mathbf{C}_1^T(k)(\mathbf{C}_1(k)\mathbf{P}(k)\mathbf{C}_1^T(k) + \mathbf{R}_2)^{-1} \\ \mathbf{P}(k+1) &= \mathbf{F}\mathbf{P}(k)\mathbf{F}^T + \mathbf{R}_1 \\ &\quad - \mathbf{F}\mathbf{P}(k)\mathbf{C}_1^T(k)(\mathbf{C}_1(k)\mathbf{P}(k)\mathbf{C}_1^T(k) + \mathbf{R}_2)^{-1}\mathbf{C}_1(k)\mathbf{P}(k)\mathbf{F}^T\end{aligned} \tag{5.4.10}$$

where $\mathbf{C}_1(k) \triangleq \mathbf{C}(k)\mathbf{H}$, with initial values

$$\hat{\xi}(0) = \bar{E}E\xi(0) \triangleq \xi^0$$

[21]The variable $\Delta Ge(k)$ will, in general, not be Gaussian. Robustified Kalman estimators are, however, the optimal *linear* estimators for arbitrary noise and uncertainty distributions.

$$\mathbf{P}(0) \; = \; \bar{E}E(\xi(0)-\xi^0)(\xi(0)-\xi^0)^T \quad .$$

The robustifying modified covariance matrices are given by

$$\mathbf{R}_1 = \mathbf{G}_0\mathbf{G}_0^T + \bar{\mathbf{E}}(\Delta\mathbf{G}\Delta\mathbf{G}^T)$$

$$\mathbf{R}_2 = \mathbf{M}_0\mathbf{M}_0^T + \bar{\mathbf{E}}(\Delta\mathbf{M}\Delta\mathbf{M}^T) \tag{5.4.11}$$

with $\Delta\mathbf{G}$ and $\Delta\mathbf{M}$, both with zero mean, introduced in (5.4.9) and (5.4.1), respectively.

Numerical illustrations can be found in [36] and in Chapter 7 of [37].

5.5 Parameter Tracking

In digital communication and, in particular, in digital mobile radio applications, the problem of adjusting filters to a rapidly time-varying dynamics is encountered. If models cannot be re-adjusted with sufficient speed and accuracy, sophisticated tools for model-based filtering will be of little utility.

Models of time-varying communication channels, can often be represented in a linear regression form

$$y(k) = \varphi^*(k)\hat{\theta}(k) + \varepsilon(k) \tag{5.5.1}$$

where $\varphi^*(k)$ is a regressor matrix containing known signals, $\hat{\theta}(k)$ is a column parameter vector and $\varepsilon(k)$ is a vector of residuals.[22] If the model structure is correct, it becomes meaningful to formulate the problem of model adjustment as a problem of following (tracking) a vector $\theta(k)$ which parametrizes a system

$$y(k) = \varphi^*(k)\theta(k) + w(k) \tag{5.5.2}$$

where the disturbance $w(k)$ is assumed independent of both $\varphi^*(k)$ and $\theta(k)$.

Parameter tracking is normally performed by utilizing the LMS algorithm or exponentially windowed RLS [7, 8, 10, 11, 103]. The initial convergence of the Newton-based RLS algorithm is much faster than that of the gradient-based LMS, but their performance when tracking continuously changing parameters is about equal [108]. In general, both LMS and windowed RLS algorithms turn out to be structurally mismatched to the tracking problem at hand [50, 54].

If the dynamics of the elements of $\theta(k)$ and the noise have known second order moments, then a Kalman-based adaptation algorithm represents the optimal scheme, which attains the best tracking performance. In many problems, the dynamic properties of time-varying parameters are approximately

[22] In an adaptive equalizer which runs in decision-directed mode, the elements of $\varphi^*(k)$ contain estimates of transmitted symbols. These estimates are obtained from an equalizer or Viterbi detector which, in its turn, is adjusted based on the the changing channel model.

known. They are, however, rarely known exactly. Methods for *robust* model-based parameter tracking would therefore be of interest. The robust Kalman scheme outlined in Sect. 5.4 can be used in such cases.

Let $z(k) = \mathbf{L}x(k)$ in (5.4.1) represent the parameter vector $\theta(k)$, to be tracked. The state update equation will then represent the set of possible parameter dynamics, whereas the measurement equation corresponds to the linear regression (5.5.2)

$$y(k) = \varphi^*(k)\theta(k) + w(k) = \varphi^*(k)\mathbf{L}x(k) + (\mathbf{M}_o + \Delta\mathbf{M})v(k) \ . \tag{5.5.3}$$

Thus, $\mathbf{C}(k) = \varphi^*(k)\mathbf{L}$. The cautious Kalman predictor outlined in Sect. 5.4.2 can now be used directly for tracking the parameters of linear regression models.

Kalman-based adaptation algorithms require a Riccati update step at each sample, and they will therefore have an unacceptable complexity in many high-speed applications. Algorithms of lower computational complexity, which can still take *a priori* knowledge about the statistical properties of the parameter variations into account, have therefore been sought [50, 109, 110].

By using the polynomial approach, a family of algorithms with low computational complexity, but still close to optimal performance, has been derived by Lindbom in [54]. The algorithms within this family have a *constant* adaptation gain, and will require no Riccati update. They are characterized by the general recursive structure

$$\varepsilon(k) = y(k) - \boldsymbol{\varphi}^*(k)\hat{\theta}(k|k-1) \tag{5.5.4}$$

$$\hat{\theta}(k+m|k) = \boldsymbol{F}(q^{-1})\hat{\theta}(k+m-1|k-1) + \boldsymbol{\mathcal{G}}_m(q^{-1})\varphi(k)\varepsilon(k)$$

which includes predictors, filters and fixed-lag smoothers. Above, the polynomial matrix $\boldsymbol{F}$ is obtained directly from the assumed stochastic model for the parameter variations, while the stable rational matrix $\boldsymbol{\mathcal{G}}_m$ is obtained via spectral factorization and Diophantine design equations, which need to be recomputed only if the model describing the parameter variations changes. In simple but useful special cases, no design equations at all need to be solved. The complexity of the algorithm (5.5.4) will increase only linearly with the dimension of $\hat{\theta}(k)$, if the regressor correlation matrix is known.

The thesis [54] also includes an analysis of tracking algorithms and a case study, describing adaptive equalization in the US D-AMPS standard (IS-54). For these two purposes, as well as for the design of constant-gain tracking algorithms, the use of a polynomial equations approach has turned out to be very fruitful.

Acknowledgement. We thank Claes Tidestav for many useful comments on the manuscript for the present chapter.

Bibliography

[1] Kučera V. Discrete Linear Control. The Polynomial Equation Approach. Wiley, Chichester, 1979

[2] Kučera V. Analysis and Design of Discrete Linear Control Systems. Academia, Prague and Prentice Hall International, London, 1991

[3] Kwakernaak H, Sivan R. Linear Optimal Control Systems. Wiley, New York, 1972.

[4] Öhrn K, Ahlén A, Sternad M. A probabilistic approach to multivariable robust filtering and open-loop control. IEEE Transactions on Automatic Control,1995; 40:405-418

[5] Tidestav C, Ahlén A, Sternad M. Narrowband and broadband multiuser detection using a multivariable DFE. IEEE PIMRC'95. Toronto, 1995, pp 732-736

[6] Ahlén A, Sternad M. Derivation and design of Wiener filters using polynomial equations. In: Leondes CT (ed) Control and Dynamic Systems, vol 64: Stochastic Techniques in Digital Signal Processing Systems. Academic Press, New York, 1994

[7] Honig ML, Messerschmidt DG. Adaptive Filters: Structures, Algorithms and Applications. Kluwer Academic, Boston, 1984

[8] Haykin S. Adaptive Filter Theory. Prentice-Hall, Englewood Cliffs, Third ed. 1996.

[9] Oppenheim AV, Shafer RW. Digital Signal Processing. Prentice-Hall, Englewood Cliffs, 1975

[10] Widrow B, Stearns SD. Adaptive Signal Processing. Prentice Hall, Englewood Cliffs, 1985

[11] Treichler JR, Johnson Jr CR, Larimore MG. Theory and Design of Adaptive Filters. Wiley, New York, 1987

[12] Goodwin GC, Gevers M, Ninness B. Quantifying the error in estimated transfer functions with application to model order selection. IEEE Transactions on Automatic Control. 1992; 27:913-928

[13] Bode HW, Shannon CE. A simplified derivation of linear least square smoothing and prediction theory. Proceedings of the I.R.E. 1950; 38:417-425

[14] Kailath T. Lectures on Wiener and Kalman Filtering. Springer, Wien, 1981

[15] Cadzow JA. Foundations of Digital Signal Processing and Data Analysis. Macmillan, New York, 1987

[16] Ahlén A, Sternad M. Optimal Filtering Problems. In: Hunt K (ed) Polynomial Methods in Optimal Control and Filtering, Chapter 5. Control Engineering Series, Peter Peregrinus, London, 1993

[17] Ahlén A, Sternad M. Wiener filter design using polynomial equations. IEEE Transactions on Signal Processing. 1991; 39:2387-2399

[18] Grimble MJ. Polynomial systems approach to optimal linear filtering and prediction. International Journal of Control. 1985; 41:1545-1564

[19] Leondes CT, Pearson JO. A minimax filter for systems with large plant uncertainties. IEEE Transactions on Automatic Control. 1972; 17:266-268

[20] Kassam SA, Lim TL. Robust Wiener filters. Journal of The Franklin Institute. 1977; 171-185

[21] Kassam SA, Poor HV. Robust techniques for signal processing: a survey. Proceedings of the IEEE. 1985; 73:433-481

[22] Francis B, Zames G. On $\mathcal{H}_\infty$ optimal sensitivity theory for SISO feedback systems. IEEE Transactions on Automatic Control, 1984;29:9-16

[23] Doyle J, Glover K, Khargonekar P, Francis B. State space solutions to standard $\mathcal{H}_\infty$ and $\mathcal{H}_\infty$ control problems. IEEE Transactions on Automatic Control, 1989;34:831-847

[24] Grimble MJ, ElSayed A. Solution to the $\mathcal{H}_\infty$ optimal linear filtering problem for discrete-time systems. IEEE Transactions on Signal Processing. 1990; 38:1092-1104

[25] Shaked U, Theodor Y. A frequency domain approach to the problems of $\mathcal{H}_\infty$-minimum error state estimation and deconvolution. IEEE Transactions on Signal Processing. 1992; 40:3001-3011.

[26] Theodor Y, Shaked U, de Souza C E. A game theory approach to robust discrete-time $\mathcal{H}_\infty$-estimation. IEEE Transactions on Signal Processing. 1994; 42:1486-1495.

[27] Bolzern P, Colaneri P, De Nicolao G. Optimal robust filtering for linear systems subject to time-varying parameter perturbations In: Proceedings of the IEEE 32nd Conference on Decision and Control. San Antonio, 1993, pp 1018-1023

[28] Xie L, Soh YC. Robust Kalman filtering for uncertain systems. Systems & Control Letters. 1994; 22:123-129

[29] Xie L, Soh YC, de Souza CE. Robust Kalman filtering for uncertain discrete-time systems. IEEE Transactions on Automatic Control. 1994; 39:1310-1314

[30] Shaked U, de Souza C. Robust minimum variance filtering. IEEE Transactions on Signal Processing. 1995; 43:2474-2483

[31] Chung RC, Bélanger PR. Minimum-sensitivity filter for linear time-invariant stochastic systems with uncertain parameters. IEEE Transactions on Automatic Control. 1976; 21:98-100

[32] Speyer JL, Gustafson DE. An approximation method for estimation in linear systems with parameter uncertainty. IEEE Transactions on Automatic Control. 1975; 20:354-359

[33] Grimble MJ. Wiener and Kalman filters for systems with random parameters. IEEE Transactions on Automatic Control. 1984; 29:552-554

[34] Sternad M, Ahlén A. Robust filtering and feedforward control based on probabilistic descriptions of model errors. Automatica. 1993; 29:661-679

[35] Öhrn K, Ahlén A, Sternad M. A probabilistic approach to multivariable robust filtering, prediction and smoothing. In: Proceedings of the 32nd IEEE Conference on Decision and Control. San Antonio, 1993, pp 1227-1232

[36] Sternad M, Öhrn K, Ahlén A. Robust $\mathcal{H}_2$ filtering for structured uncertainty: the performance of probabilistic and minimax schemes. European Control Conference. Rome, 1995, pp 87-92

[37] Öhrn K. Design of Multivariable Cautious Wiener Filters: A Probabilistic Approach. PhD thesis, Uppsala University, 1996

[38] Lindskog E, Sternad M, Ahlén A. Designing decision feedback equalizers to be robust with respect to channel time variations. Nordic Radio Society Symposium on interference resistant radio and radar. Uppsala, 1993

[39] Guo L, Ljung L. The role of model validation for assessing the size of the unmodelled dynamics. In: Proceedings of the 33rd IEEE Conference of Decision and Control. Lake Buena Vista, 1994, pp 3894-3899

[40] Gevers M. Towards a joint design of identification and control. In Trentelman HL and Willems JC ed. Essays on Control: Perspectives in the Theory and itn Applications. Birkhauser, 1993,111-151.

[41] Kosut R L, Goodwin G C and Polis, eds. Special issue on system identification for robust control design. IEEE Transactions on Automatic Control, 1992;37(7)

[42] Bigi S. On the use of system identification for design purposes and parameter estimation. Licentiate Thesis, Uppsala University, 1995

[43] Lee W. Mobile Communications Engineering. McGraw Hill, New York, 1982

[44] Lee W. Mobile Cellular Telecommunications Systems McGraw Hill, New York, 1989

[45] Proakis JG. Digital Communications. McGraw Hill, New York, 3rd ed. 1995

[46] Lee EA, Messerschmitt DG. Digital Communication. Kluwer Academic Publisher, Boston, Second ed. 1994

[47] Lindskog E. On Equalization and Beamforming for Mobile Radio Applications. Licentiate Thesis, Uppsala University, 1995

[48] Lindskog E, Ahlén A, Sternad M. Combined spatial and temporal equalization using an adaptive antenna array and a decision feedback equalization scheme. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Detroit, 1995

[49] Lindskog E, Ahlén A, Sternad M. Spatio-temporal equalization for multipath environments in mobile radio application. In: Proceedings of the 45th IEEE Vehicular Technology Conference, Chicago, 1995, pp 399-403

[50] Benveniste A. Design of adaptive algorithms for tracking of time-varying systems. International Journal of Adaptive Control and Signal Processing. 1987; 2:3-29

[51] Kubin G. Adaptation in rapidly time-varying environments using coefficient filters. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol 3. Toronto, 1991,pp 2097-2100

[52] Lindbom L. Simplified Kalman estimation of fading mobile radio channels: high performance at LMS computational load. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol 3. Minneapolis, 1993, pp 352-355

[53] Lindbom L. Adaptive equalizers for fading mobile radio channels. Licenciate Thesis, Uppsala University, 1992

[54] Lindbom L. A Wiener filtering approach to the design of tracking algorithms: with applications in mobile radio communications. PhD Thesis, Uppsala University, 1995

[55] Kailath T. Linear Systems. Prentice-Hall, Englewood Cliffs, New Jersey, 1980

[56] Anderson BDO, Moore JB. Optimal Filtering. Prentice-Hall, Englewood Cliffs, 1979

[57] Ahlén A, Sternad M. Optimal deconvolution based on polynomial methods. IEEE Transactions on Acoustics, Speech and Signal Processing. 1989; 37:217-226

[58] Carlsson B, Ahlén A, Sternad M, Optimal differentiation based on stochastic signal models. IEEE Transactions on Signal Processing. 1991; 39:341-353

[59] Carlsson B, Sternad M, Ahlén A. Digital differentiation of noisy data measured through a dynamic system. IEEE Transactions on Signal Processing. 1992; 40:218-221

[60] Moir TJ. A polynomial approach to optimal and adaptive filtering with application to speech enhancement. IEEE Transactions on Signal Processing. 1991; 39:1221-1224

[61] Roberts AP, Newmann MM. Polynomial approach to Wiener filtering. International Journal of Control. 1988; 47:681-696

[62] Grimble MJ. Time-varying polynomial systems approach to multichannel optimal linear filtering. Proceedings of the American Control Conference. WA6-10:45, Boston, 1985, pp 168-174

[63] Fitch SM, Kurz L. Recursive equalization in data transmission - A design procedure and performance evaluation. IEEE Transactions on Communication. 1975; 23:546-550

[64] Qureshi SUH. Adaptive equalization. Proceedings of the IEEE. 1985; 73:1349-1387

[65] Nelson PA, Hamada H and Elliot SJ. Adaptive inverse filters for stereophonic sound reproduction. IEEE Transactions on Signal Processing. 1992; 40:1621-1632

[66] Mendel JM. Optimal Seismic Deconvolution. An Estimation-Based Approach. Academic Press, New York, 1983

[67] Ahlén A, Sternad M. Filter design via inner-outer factorization: Comments on "Optimal deconvolution filter design based on orthogonal principle". Signal Processing, 1994; 35:51-58

[68] Kučera V. Factorization of rational spectral matrices: a survey. In: Preprints of IEE Control'91. Edinburgh, 1991, pp 1074-1078

[69] Ježek J, Kučera V. Efficient algorithm for matrix spectral factorization. Automatica. 1985; 21:663-669

[70] Demeure CJ, Mullis CT. A Newton-Raphson method for moving average spectral factorization using the Euclid algorithm. IEEE Transactions on Acoustics, Speech and Signal Processing. 1990; 38:1697-1709

[71] Deng ZL. White-noise filter and smoother with application to seismic data deconvolution. In: Preprints 7th IFAC/IFORS Symp. Identification Syst. Parameter Estimation. York, 1985, pp 621-624

[72] Deng Z-L, Zhang H-S, Liu S-J, Zhou L. Optimal and self-tuning white noise estimator with applications to deconvolution and filtering problems. Automatica. 1996; 32:199-216

[73] Ahlén A, Sternad M. Adaptive input estimation. IFAC Symposium ACASP-89, Adaptive Systems in Control and Signal Processing. Glasgow, 1989, pp 631-636

[74] Ahlén A, Sternad M. Adaptive deconvolution based on spectral decomposition. SPIE Annual Symposium. vol 1565. San Diego, 1991, pp 130-142

[75] Ahlén A. Identifiability of the deconvolution problem. Automatica. 1990; 26:177-181

[76] Nikias CL, Petropulu AP. Higher Order Spectral Analysis: A Nonlinear Signal Processing Framework. Prentice Hall, Englewood Cliffs, 1993.

[77] Tong L, Xu G, Kailath T. Fast blind equalization via antenna arrays. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Minneapolis, 1993, 4:272-275.

[78] Moulines E, Duhamel P, Cardoso J-F, Mayrargue S. Subspace methods for the blind identification of multichannel FIR filters. IEEE Transactions on Signal Processing. 1995; 43(2):516-525.

[79] Bernhardsson B, Sternad M. Feedforward control is dual to deconvolution. International Journal of Control. 1993; 47:393-405

[80] Sternad M, Ahlén A. The structure and design of realizable decision feedback equalizers for IIR-channels with coloured noise. IEEE Transactions on Information Theory. 1990; 36:848-858

[81] Tidestav C. Optimum diversity combining in multi-user digital mobile telephone systems. Master's Thesis, Uppsala University,1993

[82] Forney Jr GD. The Viterbi algorithm. In: Proceedings of IEEE. 1973; 61:268-278

[83] Belfiore CA, Park JH. Decision feedback equalization. Proceedings of the IEEE. 1979;67:1143-1156

[84] Monsen P. Feedback equalization for fading dispersive channels. IEEE Transactions on Information Theory. 1971;17:56-64

[85] Salz J. Optimum mean-square decision feedback equalization. Bell System Technical Journal. 1973; 52:1341-1374

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

[87] Lo NWK, Falconer DD, Sheikh AUH. Adaptive equalizer MSE performance in the presence of multipath fading, interference and noise. In: Proceesings of the IEEE Vehicular Technology Conference, Rosemont, 1995, 409-413

[88] Sternad M, Ahlén A, Lindskog E. Robust decision feedback equalization. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, 1993;3:555-558.

[89] Winters J. On the capacity of radio communication systems with diversity in a Rayleigh fading environment. IEEE Journal on Selected Areas in Communication. 1987; 5(5):871-878

[90] Falconer DD, Abdulrahman M, Lo NWK, Petersen BR, Sheikh AUH. Advances in equalization and diversity for portable wireless systems. Digital Signal Processing. 1993; 3(3):148-162

[91] Gilhousen KS, Jacobs IW, Padovani R, Viterbi A, Weaver LA, Wheatly CE. On the capacity of a cellular CDMA system. IEEE Transactions on Vehicular Technology. 1991; 40:303-312

[92] Verdú S. Optimum Multi-user Signal Detection. PhD thesis, University of Illinois, 1984.

[93] Goodwin GC. Salgado ME. A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. International Journal of Adaptive Control and Signal Processing. 1989; 3:333-356

[94] Grimble MJ. Generalized Wiener and Kalman filters for uncertain parameters. In: Proceedings of the 21th IEEE Conference on Decision and Control. Orlando, 1982, pp 221-227

[95] D'Appolito JA, Hutchinson CE. A minimax approach to the design of low sensitivity state estimators. Automatica. 1972; 8:599-608

[96] Martin CJ, Mintz M. Robust filtering and prediction for linear systems with uncertain dynamics: a game-theoretic approach. IEEE Transactions on Automatic Control. 1983; 28:888-896

[97] Moustakides G. Kassam SA. Robust Wiener filters for random signals in correlated noise. IEEE Transactions on Information Theory. 1983; 29:614-619

[98] Poor HV. On robust Wiener filtering. IEEE Transactions on Automatic Control. 1980; 25:531-536

[99] Vastola KS, Poor HV. Robust Wiener-Kolmogorov theory. IEEE Transactions on Information Theory. 1984; 30:316-327

[100] Petersen IR, McFarlane DC. Robust state estimation for uncertain systems. In: Proceedings of the 30th Conference on Decision and Control. Brighton, 1991, pp 2630-2631

[101] Haddad WF, Bernstein DS. Robust, reduced-order, nonstrictly proper state estimation via the optimal projection equations with guaranteed cost bounds. IEEE Transactions on Automatic Control. 1988; 33:591-595

[102] Haddad WF, Bernstein DS. The optimal projection equations for reduced-order discrete-time state estimation for linear systems with multiplicative noise. System & Control Letters. 1987; 8:381-388

[103] Ljung L, Söderström T. Theory and Practice of Recursive Identification. MIT Press, Cambridge, 1983

[104] Ljung L. System Identification: Theory for the User. Prentice Hall, Englewood Cliffs, 1987

[105] Wahlberg B. System identification using Laguerre models. IEEE Transactions on Automatic Control. 1991; 36:551-562

[106] P. Lindskog and B. Walhberg, Applications of Kautz models in system identification. In: Proceedings of the IFAC World Congress, Sydney, 1993(5):309-312

[107] Hakvoort R. System identification for robust process control. PhD Thesis, Delft University of Technology, The Netherlands, 1994.

[108] Widrow B, Walach E. On the statistical efficiency of the LMS algorithm with nonstationary inputs. IEEE Transactions on Information Theory. 1984; 30:211-221

[109] Benveniste A, Métivier M, Priouret P. Adaptive Algorithms and Stochastic Approximations. Springer, Berlin, 1990

[110] Clark AP. Adaptive Detectors for Digital Modems. Pentech Press, London, 1989