Derivation and Design of Wiener Filters using Polynomial Equations

Anders Ahlén
Mikael Sternad

Systems and Control Group, Uppsala University
P. O. Box 27, S–751 03 Uppsala, Sweden

I INTRODUCTION

In this chapter, a polynomial approach to filter design is presented. Our goal is to demonstrate its utility in the area of signal processing and communications[1]. By studying specific model structures, it does, in particular, become evident that considerable engineering insight can be gained.

Minimization of mean–square error criteria by linear filters will be considered. We focus on the optimization of realizable discrete–time IIR-filters, to be used for prediction, filtering or smoothing of signals. Stochastic models of possibly complex-valued signals are assumed known.

Historically, such problems have been dealt with by applying the classical Wiener–Hopf approach. See e.g. [1], [2], [3], [4]. Even after the Kalman–filter breakthrough [5], many researchers still prefer a frequency domain approach, despite its inferior numerical properties for high order problems. One reason is the relative ease with which an obtained filter can be examined. A quick inspection of the poles and zeros roughly tells us what filter properties that could be expected.

While the classical Wiener solution is conceptually elegant it has, until recently, been rather intractable to perform the causal bracket operation $\{\cdot\}_+$ central to the design of realizable filters. In particular, Wiener–smoothing has not been straightforward. With the polynomial approach, pioneered by Kučera [6], *Diophantine equations* now offer an efficient way of automatizing the causul bracket operation. A (polynomial) Diophantine equation

[1]The chapter includes work partially supported by the Swedish National Board for Technical Development, NUTEK, under contract 87-01573. Parts of the text has been published previously as Chapter 5 of the book *Polynomial Methods in Optimal Control and Filtering*, ed. K. Hunt, Control Engineering Series, Peter Peregrinus, London, 1993.

CONTROL AND DYNAMIC SYSTEMS, VOL. 64

is a linear equation in two polynomials or polynomial matrices.

Polynomial equations were first applied to estimation problems by control engineers. An early result is due to Åström in 1970 [7]. To obtain minimum variance control laws, he derived a Diophantine equation for calculating the d–step prediction of an ARMA process

$$y(k) = \frac{C(q^{-1})}{D(q^{-1})}e(k) = \frac{(1 + c_1q^{-1} + \ldots + c_nq^{-n})}{(1 + d_1q^{-1} + \ldots + d_nq^{-n})}e(k) \quad . \tag{1}$$

Above, $y(k)$ is the measured discrete–time signal and q^{-1} denotes the backward shift operator ($q^{-1}x(k) = x(k-1)$). The polynomial $C(z^{-1})$ has zeros in $|z| < 1$ and $e(k)$ is white and zero mean. The linear d–step predictor, which minimizes the mean square estimation error, is given by

$$\hat{y}(k|k-d) = \frac{G(q^{-1})}{C(q^{-1})}y(k-d) \quad . \tag{2}$$

Here, the polynomial $G(q^{-1})$ of degree $n-1$, together with a polynomial $F(q^{-1})$ of of degree $d-1$, is the solution to the Diophantine equation

$$C(q^{-1}) = q^{-d}G(q^{-1}) + D(q^{-1})F(q^{-1}) \quad . \tag{3}$$

See Theorem 3.1 of [7]. This result was later generalized to multivariable systems by Borison [8]. Compared to earlier Wiener methods, which where based on the manipulation of auto- and cross covariance functions, the polynomial approach offered a considerable simplification. Another contribution (which did not explicitely use Diophantine equations), was the selftuning smoother of Hagander and Wittenmark [9]. See also [10]. Other early contributions were a filter for a signal vector in white measurement noise [11], and a polynomial method for computing the gain matrix of a Kalman filter [12], both by Kučera.

Fairly recently, the polynomial systems framework has been more systematically utilized to solve signal processing and communications problems. See e.g. [13]–[23]. There has been some work related to time–varying filter design [24]. Effort has also been spent on relating Kalman, Wiener, and polynomial methods [17], [25], [27], [28].

When a polynomial approach is used, estimators are calculated from three types of equations: Diophantine equations, polynomial spectral factorizations and (in some problems) coprime factorizations of polynomial matrices.

Two dominating approaches have been used for deriving these sets of equations: The *"completing the squares approach"* used e.g. in [6], [12], [17], [22] and [28] and a *"variational approach"* developed in [14], see also [31], [29], [30]. Lately, the *inner–outer factorization approach*, see e.g. [32], has been utilized in [33]. In [34], a correct use of that method is discussed as well as its relation to the polynomial approach.

The basis for our discussion will be a general linear filtering problem, outlined in Section II. In Section III, we shall discuss how the classical Wiener and inner–outer factorization approaches relate to the polynomial methods, based on variational arguments and completing the squares. The purpose of this discussion is not only to compare advantages and drawbacks, but also to emphasize similarities, and to link and increase understanding of the different viewpoints. To understand how they relate to one another, design equations for a simple scalar filtering problem will be derived using each approach.

The polynomial approach, based on variational arguments, is then used to study a collection of signal processing and communications problems in Sections IV–VI. We discuss deconvolution (Section IV), numerical differentiation and state estimation (Section V) and decision feedback equalization (Section VI). The selected special problems have features of general interest: multisignal estimation (Section IV), discrete time design based on a continuous time problem formulation (Section V), and the approximation of a problem involving a static nonlinearity by a linear–quadratic problem (Section VI). A summarizing discussion of characteristics and suitability of the polynomial approach is found in Section VII.

An guide describing connections between different sections of this chapter is provided by Figure 1. A reader who wish to obtain a good overview, without studying all details, is recommended the following path: Sections I,A; II; III,A, B, C, E, H, I, followed by IV,A and VII.

A main novel result is the multisignal estimator of Section IV. It constitutes a solution to the general filtering problem presented in Section II. The solution is simplified considerably, compared to previous solutions to similar problems derived by ourselves and by Grimble [14], [17], [19]. A key feature is that the signal–generating transfer functions are parametrized by joint use of common denominator forms and matrix fraction descriptions with diagonal denominators. This makes it possible to calculate estimators by solving only a polynomial matrix spectral factorization and one simple "unilateral" Diophantine equation. (Previous solutions required several coprime factorizations and were based on one, or several, "bilateral" Diophantine equations. Bilateral Diophantine equations require more elaborate numerical algorithms than "unilateral" equations.)

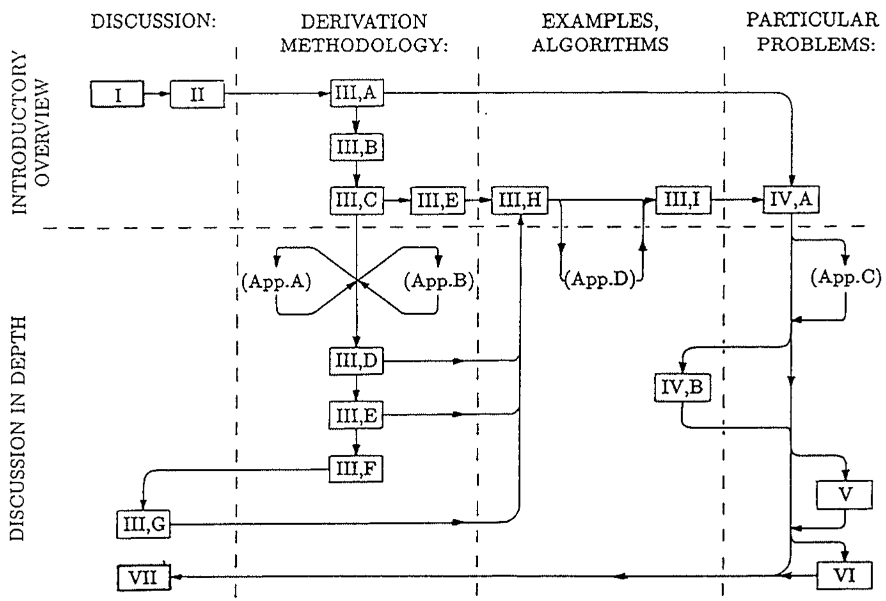

Figure 1: Reader's guide to the chapter.

A REMARKS ON THE NOTATION

Let p_j^* denote the complex conjugate of polynomial coefficient p_j. 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} ($q^{-1}y(k) = y(k-1)$), define

$$P_* \triangleq p_0^* + p_1^* q + \ldots + p_{np}^* q^{np}$$

$$\bar{P} \triangleq q^{-np} P_* = p_{np}^* + p_{np-1}^* q^{-1} + \ldots + p_0^* q^{-np} .$$

Whenever a polynomial in positive powers of q is introduced, it will be denoted with a star, P_*. Rational matrices, or transfer function matrices, are denoted by boldface script symbols, for example, as $\mathcal{R}(q^{-1})$. Polynomial matrices are denoted by boldface symbols like $\boldsymbol{P}(q^{-1})$ while constant matrices, for example covariance matrices, are denoted as $\mathbf{P}$.

For polynomial matrices, $\boldsymbol{P}_*$ means complex conjugate transpose. We denote the trace of $\boldsymbol{P}$ by tr$\boldsymbol{P}$. When appropriate, the complex variable z is substituted for the forward shift operator q. The *degree* of a polynomial matrix is the highest degree of any of its polynomial elements. Polynomial matrices $\boldsymbol{P}(q^{-1})$ are called *stable* if all zeros of det $\boldsymbol{P}(z^{-1})$ are located in $|z| < 1$. For *marginally stable* polynomial matrices, some zeros of det $\boldsymbol{P}(z^{-1})$ are located on $|z| = 1$. Arguments of polynomials and rational matrices are often omitted, when there is no risk for misunderstanding.

A rational matrix $\mathcal{G}(z^{-1})$ can be represented by polynomial matrices as a *matrix fraction description* (MFD), either left or right: $\mathcal{G} = \boldsymbol{A}_1^{-1}\boldsymbol{B}_1 = \boldsymbol{B}_2\boldsymbol{A}_2^{-1}$. See e.g. [35]. It can also be converted to *common denominator form* $\mathcal{G}(q^{-1}) = \boldsymbol{B}(q^{-1})/A(q^{-1})$, where $\boldsymbol{B}(q^{-1})$ is a polynomial matrix and the scalar monic polynomial $A(q^{-1})$ is the (least) common denominator of all rational transfer function elements in $\mathcal{G}(q^{-1})$.

A filter $\mathcal{V}(q^{-1})$ which whitens a signal $y(k)$,

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

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

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

is called an *innovations model* of the signal $y(k)$.

II A SET OF FILTERING PROBLEMS

A general linear filtering problem can be formulated in the following way. Based on measurements $z(k)$, up to time $k+m$, a real or complex-valued vector $f(k) = (f_1(k) \ldots f_\ell(k))^T$ of desired signals is sought. The signals are described by the linear discrete-time stochastic system

$$\begin{pmatrix} z(k) \\ f(k) \end{pmatrix} = \begin{pmatrix} \mathcal{G}_g(q^{-1}) \\ \mathcal{D}_g(q^{-1}) \end{pmatrix} u_g(k) \qquad (4)$$

and the estimator is

$$\hat{f}(k|k+m) = \mathcal{R}_z(q^{-1})z(k+m) \; . \qquad (5)$$

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

We will consider minimization of the estimation error covariance matrix

$$\mathbf{P} \triangleq E\varepsilon(k)\varepsilon^*(k) \tag{6}$$

where

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

Above, $\mathcal{W}(q^{-1})$ is a stable and causal transfer function weighting matrix, of dimension $\ell|\ell$. It may be used to emphasize performance in certain frequency ranges. The covariance matrix (6) 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 positive semidefinite. It is minimized under the constraint of realizability (internal stability and causality) of the filter $\mathcal{R}_z(q^{-1})$.

Minimization of (6) implies the minimization of the sum of the elementwise mean square errors (MSE)'s:

$$J = \mathrm{tr}E(\varepsilon(k)\varepsilon^*(k)) = E(\varepsilon^*(k)\varepsilon(k)) = \sum_{i=1}^{\ell} E|\varepsilon_i(k)|^2 \ . \tag{7}$$

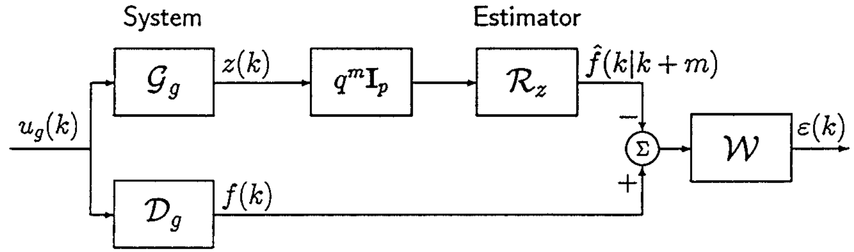

Figure 2: A general filtering problem formulation.

While the model (4) is general, it lacks sufficient degree of structure, to provide solutions which give useful engineering insight. For the purpose of of this chapter, we will therefore introduce a more detailed structure. It encompasses a number of special cases, to be separately studied in Sections III–VI. The vector $u_g(k)$ is split into two parts

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

where $w(k)$ represents additive measurement noise, uncorrelated to the desired signal $f(k)$. We also introduce explicit stochastic models

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

with $\mathcal{F}$ and $\mathcal{H}$ not necessarily stable. The noises $\{e(k)\}$ and $\{v(k)\}$ are mutually uncorrelated and stationary vector sequences. They have zero means[2] and covariance matrices $\phi \geq 0$ and $\psi \geq 0$.[3] Furthermore, define the measurement vector

$$z(k) \triangleq \begin{pmatrix} y(k) \\ a(k) \end{pmatrix} \tag{8}$$

where $y(k) = (y_1(k)\dots y_p(k))^T$ are noisy measurements and $a(k) = (a_1(k)\dots a_h(k))^T$ is an auxilliary measurement vector, uncorrupted by the noise $w(k)$. (One example could be directly measurable delayed inputs to the system.) The model structure (4) is thus specialized to

$$\begin{pmatrix} y(k) \\ a(k) \\ f(k) \end{pmatrix} = \begin{pmatrix} \mathcal{G}(q^{-1}) & \mathbf{I} \\ \mathcal{G}_a(q^{-1}) & 0 \\ \mathcal{D}(q^{-1}) & 0 \end{pmatrix} \begin{pmatrix} u(k) \\ w(k) \end{pmatrix}$$

$$\begin{pmatrix} u(k) \\ w(k) \end{pmatrix} = \begin{pmatrix} \mathcal{F}(q^{-1}) & 0 \\ 0 & \mathcal{H}(q^{-1}) \end{pmatrix} \begin{pmatrix} e(k) \\ v(k) \end{pmatrix} \tag{9}$$

Above, $\mathcal{G}$, $\mathcal{G}_a$, $\mathcal{F}$, $\mathcal{H}$, and $\mathcal{D}$ are transfer function matrices of appropriate dimensions. See Figure 3. From the measurements $z(k)$, up to time $k+m$, our aim is to optimize the linear estimator (5)

$$\hat{f}(k|k+m) = \mathcal{R}_z(q^{-1})z(k+m) = (\mathcal{R}(q^{-1}) \ \mathcal{R}_a(q^{-1})) \begin{pmatrix} y(k+m) \\ a(k+m) \end{pmatrix} \tag{10}$$

[2] How to handle nonzero means in on–line applications is explained in Appendix B.
[3] Sometimes it is convenient to normalize ϕ and ψ to unit matrices and include variance scalings in $\mathcal{F}$ and $\mathcal{H}$ respectively. The signals $e(k)$ and $v(k)$ may also be Bernoulli–Gaussian distributed. Such a distribution is useful for modelling shape–deterministic signals, for example, random step sequences.

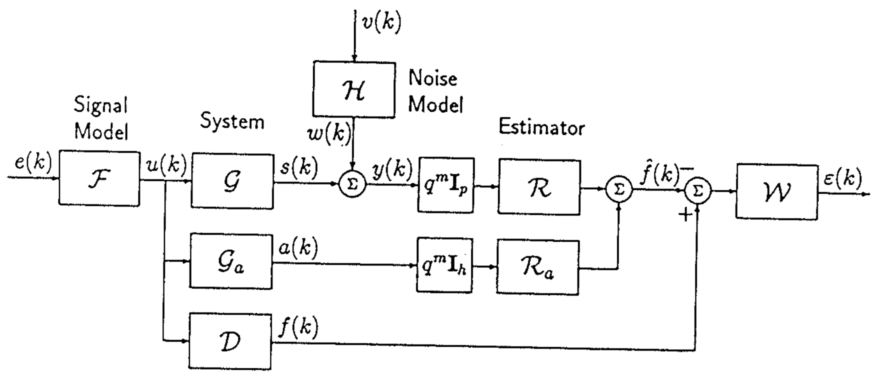

Figure 3: Unifying structure for a collection of filtering problems. The signal $f(k)$ is to be estimated from data up to time $k+m$.

with $\mathcal{R}$ and $\mathcal{R}_a$ being stable and causal transfer function matrices.

The structure depicted in Figure 3 covers a large set of different problems. We shall in this chapter discuss the following collection:

- Scalar prediction, filtering or smoothing: $\mathcal{G} = \mathcal{D} = \mathcal{W} = 1$, $\mathcal{G}_a = 0$. (Section III.)
- Multisignal deconvolution and linear equalization: $\mathcal{G}_a = 0$. (Section IV.)
- Numerical differentiation of scalar signals and state estimation: $\mathcal{W} = 1$, $\mathcal{G}_a = 0$, $u(k)$ state vector, $\mathcal{G}$ and $\mathcal{D}$ constant vectors. (Section V.)
- Decision feedback equalization of a scalar white symbol sequence: $\mathcal{W} = 1, \mathcal{F} = 1, \mathcal{D} = 1$, $\mathcal{G}_a = q^{-m-1}$. (Section VI.)

III DERIVATION METHODS

We have recently investigated and developed a *variational approach* for solving filtering and LQG control problems [14]. We open this section by presenting underlying general ideas, before going into details, and comparisons with other approaches. (Application to LQG, or $\mathcal{H}_2$–optimal, control is discussed in [29] and in [30].)

A USE OF VARIATIONAL ARGUMENTS

Consider the estimator (5) and the criterion (6), (7). Introduce an *alternative weighted estimate*

$$\hat{d}(k|k+m) = \mathcal{W}(q^{-1})\hat{f}(k|k+m)+\nu(k) = \mathcal{W}(q^{-1})\mathcal{R}_z(q^{-1})z(k+m)+\nu(k) \tag{11}$$

where a stationary signal $\nu(k)$ represents a modification of a (weighted) estimate (5). See Figure 4. The estimate $\hat{f}(k)$ is optimal if and only if no admissible variation can improve the criterion value.

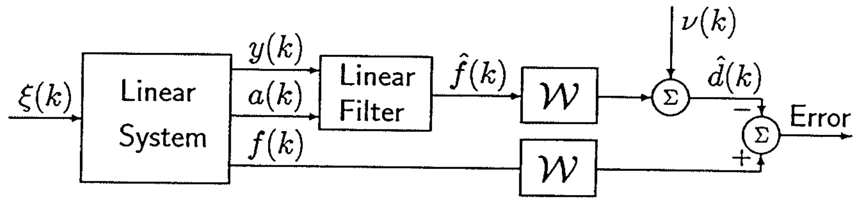

Figure 4: A variational approach to the general estimation problem (with driving noise $\xi(k) = (e(k)^T\ v(k)^T)^T$), discussed in Section II. The weighted estimate is perturbed by a variation $\nu(k)$.

All admissible variations can be represented by $\nu(k) = \mathcal{T}(q^{-1})z(k+m)$, where $\mathcal{T}(q^{-1})$ is some stable and causal rational matrix. Any nonstationary modes of $z(k)$ must be cancelled by zeros of $\mathcal{T}(q^{-1})$. Except for these requirements, $\mathcal{T}(q^{-1})$ is arbitrary. The use of the modified estimator (11) results in the covariance matrix

$$\begin{aligned}\bar{\mathbf{P}} &= E\{\mathcal{W}(q^{-1})f(k) - \hat{d}(k|k+m)\}\{\mathcal{W}(q^{-1})f(k) - \hat{d}(k|k+m)\}^* \\ &= E\varepsilon(k)\varepsilon(k)^* - E\varepsilon(k)\nu(k)^* - E\nu(k)\varepsilon(k)^* + E\nu(k)\nu(k)^* \end{aligned} \tag{12}$$

If the *cross–terms* in (12) are zero, $\nu(k) \equiv 0$ evidently minimizes $\bar{\mathbf{P}}$, since $E\nu(k)\nu(k)^*$ is positive semidefinite if any component of $\nu(k)$ has nonzero variance. Then, the estimator (5) is optimal. By taking the trace of (12) it is evident that (7) is also minimized.

Of the two cross terms, it is sufficient to consider only $E\varepsilon(k)\nu(k)^*$, for symmetry reasons. Now, assume $\varepsilon(k)$ in (6) to be *stationary.* (This is evidently true if $z(k)$ and $f(k)$ are stationary, since $\mathcal{W}(q^{-1})$ and $\mathcal{R}_z(q^{-1})$ are required to be stable. However, if $z(k)$ or $f(k)$ are generated by unstable models, stationarity will have to be verified separately, *after* the derivation. This will be exemplified in Appendix B.)

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

$$E\varepsilon(k)\nu(k)^* = \frac{1}{2\pi j}\oint_{|z|=1} \boldsymbol{\phi}_{\varepsilon\nu^*}\frac{dz}{z} = 0 \quad . \tag{13}$$

The rational $\ell|\ell$–matrix $\boldsymbol{\phi}_{\varepsilon\nu^*}$ is the cross spectral density. The expression (13) corresponds to the *elementwise orthogonality conditions*[4]

$$E\varepsilon_v(k)\nu_n(k)^* = \frac{1}{2\pi j}\oint_{|z|=1} \boldsymbol{\phi}^{vn}_{\varepsilon\nu^*}\frac{dz}{z} = 0 \qquad v = 1\ldots\ell\ ,\ n = 1\ldots\ell . \tag{14}$$

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

Using the common denominator form or the *left* polynomial matrix fraction description, the relations (14) 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.*

The variational approach can be summarized as a step by step procedure.

1. Parametrize the system by rational transfer functions, represented by polynomial fractions, common denominator forms or left MFD's. Define a spectral factorization from the spectral density of $y(k)$.
2. Define the estimation error $\varepsilon(k)$ and introduce an admissible variation $\nu(k)$ of the possibly weighted estimate. Express $E\varepsilon(k)\nu(k)^*$ in the frequency domain using Parseval's formula and simplify, by inserting the spectral factorization.

[4] When $\ell > 0$, these conditions imply (but are stronger than) orthogonality between the estimation error and any admissible linear function of the measurements, $\mathrm{tr}E[\varepsilon(k)\nu(k)^*] = \mathrm{tr}E\nu(k)^*\varepsilon(k) = 0$.

3. Fulfill the requirement $E\varepsilon(k)\nu(k)^* = 0$ by cancelling all poles in $|z| < 1$, in every element of the integrand, by zeros. This leads to a linear polynomial (matrix) equation, which determines the estimator. For *stable* systems, the derivation ends here.
4. For *marginally stable* and *unstable* signal–generating systems, verify stationarity of $\varepsilon(k)$, as outlined in Appendix B.

This method will now be exemplified, and compared to other methods, by solving a simple filtering problem.

B A SCALAR FILTERING PROBLEM

Consider the expressions (9)–(10) and set $\mathcal{G} = 1$, $\mathcal{G}_a = 0$ ($\Rightarrow \mathcal{R}_a = 0$), $\mathcal{F} = C/D$, $\mathcal{H} = M/N$, $\mathcal{D} = 1$, $\mathcal{W} = 1$, $\mathcal{R}_z = [\mathcal{R}\ \ 0]$. Let $\phi = \lambda_e$, $\psi = \lambda_v$. Thus, the signal

$$f(k) = s(k) = \frac{C(q^{-1})}{D(q^{-1})}e(k)$$

is to be estimated from noisy measurements

$$y(k) = s(k) + \frac{M(q^{-1})}{N(q^{-1})}v(k) \tag{15}$$

up to time $k+m$, using an estimator $\hat{s}(k|k+m) = \mathcal{R}(q^{-1})y(k+m)$, with $\mathcal{R}$ stable and causal. See Figure 5. All model polynomials, with degree nc, nd, etc, are monic. The signal and noise models are here assumed stable. (For a discussion of unstable models, refer to Appendix B.) The following assumption then guarantees problem solvability.

Assumption A. The signal and noise ARMA–models $s(k) = (C/D)e(k)$ and $w(k) = (M/N)v(k)$ are stable and causal, and are assumed to have no common zeros on the unit circle.

The measurements $\{y(k)\}$ can also be described by means of the innovations model

$$y(k) = \frac{\beta(q^{-1})}{D(q^{-1})N(q^{-1})}(\sqrt{\lambda_\epsilon}\epsilon(k)) \tag{16}$$

where the innovations sequence $\sqrt{\lambda_\epsilon}\epsilon(k)$ is white and has variance λ_ϵ. The monic polynomial $\beta(q^{-1}) = 1 + \beta_1 q^{-1} + \ldots + \beta_{n\beta}q^{-n\beta}$ is stable, under Assumption A. It is called the *polynomial spectral factor.* We shall now see how the solution to the filtering problem formulated here can be derived analytically in four different ways. A numerical illustration for a related problem is given in Section III,I.

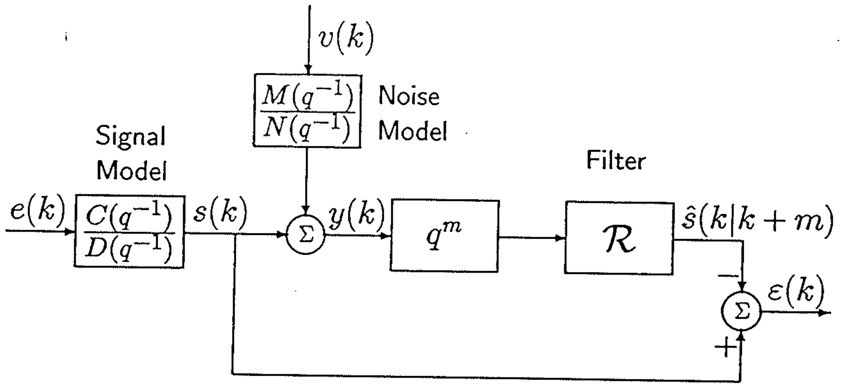

Figure 5: A scalar filtering, prediction or smoothing problem. The signal $\{s(k)\}$ is to be estimated from $\{y(k+m)\}$. In the noise-free case, this problem includes prediction of an ARMA-process $s(k) = (C/D)e(k)$.

C THE VARIATIONAL APPROACH

Let us follow the procedure for the variational approach in Section III,A.

1. Obtain spectral densities of $y(k)$, $\Phi_y(e^{j\omega})$, from both (15) and (16), and set them equal. This gives the spectral factorization equation

$$r\beta\beta_* = CC_*NN_* + \rho MM_*DD_* \tag{17}$$

where $r = \lambda_\epsilon/\lambda_e$, $\rho \triangleq \lambda_v/\lambda_e$ and $\beta(z^{-1})$ is stable and monic.

2. Use the error expression

$$\varepsilon(k) = (1 - q^m\mathcal{R})\frac{C}{D}e(k) - q^m\mathcal{R}\frac{M}{N}v(k) \tag{18}$$

and the estimator variation $\nu(k) = \mathcal{T}(q^{-1})y(k+m)$, where $\mathcal{T}(q^{-1})$ is any stable and causal transfer function. The first cross term in (12) becomes

$$E\varepsilon(k)\nu(k)^* =$$

$$E(1 - q^m\mathcal{R})\frac{C}{D}e(k)\left(\mathcal{T}q^m\frac{C}{D}e(k)\right)^* - Eq^m\mathcal{R}\frac{M}{N}v(k)\left(\mathcal{T}q^m\frac{M}{N}v(k)\right)^*$$

$$= \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\frac{(1 - z^m\mathcal{R})z^{-m}CC_*NN_* - \rho\mathcal{R}MM_*DD_*}{DD_*NN_*}\mathcal{T}_*\frac{dz}{z}$$

$$= \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\frac{(z^{-m}CC_*NN_* - \mathcal{R}r\beta\beta_*)}{DD_*NN_*}\mathcal{T}_*\frac{dz}{z} \quad . \tag{19}$$

To obtain (19) Parseval's formula and (17) were utilized.

3. In the integrand of (19), the stable polynomials D and N contribute poles in $|z| < 1$, while the poles of $\mathcal{T}_*$ and the zeros of D_*N_* are in $|z| > 1$. Furthermore, $\beta(z^{-1}) = \bar{\beta}(z)/z^{n\beta}$ may contribute poles at the origin. We see that N and β can be cancelled by $\mathcal{R}$, if $\mathcal{R} = \mathcal{R}_1N/\beta$ in (19). Thus, while β is cancelled directly, N may be factored out of the numerator, to cancel N in the denominator. The remaining poles of the integrand inside $|z| = 1$, are eliminated if (and only if)

$$\frac{(z^{-m}CC_*N_* - \mathcal{R}_1r\beta_*)}{D}\frac{1}{z} = L_*$$

for some polynomial $L_*(z)$. It suffices for $\mathcal{R}_1$ to be a polynomial, in order to obtain a (polynomial) Diophantine equation. With $\mathcal{R}_1 = Q_1$ and q exchanged for z, we obtain polynomials $Q_1(q^{-1})$, and $L_*(q)$, as the solution to the linear Diophantine equation

$$q^{-m}CC_*N_* = r\beta_*Q_1 + qDL_* \quad . \tag{20}$$

The solution is unique, see Appendix A. Thus, the optimal estimator

$$\hat{s}(k|k+m) = \frac{Q_1(q^{-1})N(q^{-1})}{\beta(q^{-1})}y(k+m) \tag{21}$$

is obtained by solving (17) for β (and r) and (20) for Q_1 (and L_*). It can be noted that the d-step predictor (2) for ARMA processes (1) is a special case of the solution above [5]. Also, note that the transfer function $\mathcal{T}_*$ in the variational term does not influence the solution at all.

[5] With a stable C and no measurement noise ($\rho = 0, M = N = 1$), we have $\beta = C, r = 1$. Then $\beta_* = C_*$ is a factor of two terms in (20), so it must also be a factor of qDL_*. Set $L_* = C_*L_{1*}$ in (20), cancel C_* and multiply by q^m, to obtain $C = q^mQ_1 + D(q^{m+1}L_{1*})$. Now, with $m = -d$, $G(q^{-1}) = Q_1(q^{-1})$ and $F(q^{-1}) = \bar{L}_1(q^{-1}) \triangleq q^{-d+1}L_{1*}(q)$, we obtain (3).

D COMPLETING THE SQUARES

While the variational method is based on manipulation of the orthogonality relation $E\varepsilon\nu^* = 0$, the "completing the squares" approach is a way of deriving the filter by manipulating $\mathrm{tr}E(\varepsilon\varepsilon^*)$ itself. The goal is to express the criterion as a sum of several terms, of which some can be minimized in a straightforward way. The other terms are either zero or unaffected by the filter.

The completing the squares approach has been used in the time domain by Kučera in e.g. [6], [28]. The frequency domain variant discussed below has been used, for example, by Roberts and Newmann in [22] and by Grimble in [17]. In the example of Section III,B, the criterion (7) can be expressed by

$$\begin{aligned} J &= E|s(k) - \mathcal{R}y(k+m)|^2 \\ &= E\left|(1-q^m\mathcal{R})\frac{C}{D}e(k)\right|^2 + E\left|q^m\mathcal{R}\frac{M}{N}v(k)\right|^2 \\ &= \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\left((1-z^m\mathcal{R})(1-z^{-m}\mathcal{R}_*)\frac{CC_*}{DD_*} + \rho\mathcal{R}\mathcal{R}_*\frac{MM_*}{NN_*}\right)\frac{dz}{z} \end{aligned}$$

$$= \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\left(\frac{CC_*}{DD_*} - z^m\mathcal{R}\frac{CC_*}{DD_*} - \frac{CC_*}{DD_*}z^{-m}\mathcal{R}_* + \mathcal{R}\mathcal{R}_*\frac{r\beta\beta_*}{DD_*NN_*}\right)\frac{dz}{z} \ .$$

In the third equality, we used Parseval's formula and in the last, the spectral factorization (17) was inserted. Completing the square gives

$$\begin{aligned} J &= \frac{\lambda_e}{2\pi j}\oint_{|z|=1} r\left(\frac{\beta}{DN}\mathcal{R} - \frac{z^{-m}CC_*N_*}{r\beta_*D}\right)\left(\frac{\beta_*}{D_*N_*}\mathcal{R}_* - \frac{z^mC_*CN}{r\beta D_*}\right)\frac{dz}{z} \\ &+ \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\left(\frac{CC_*}{DD_*} - \frac{CC_*CC_*NN_*}{r\beta\beta_*DD_*}\right)\frac{dz}{z} \triangleq J_1 + J_2 \ . \end{aligned} \tag{22}$$

The first term in (22), J_1, depends on $\mathcal{R}$ while the second term, J_2, does not. If $\mathcal{R}$ were not restricted to be realizable (internally stable and causal), the problem could have been solved by choosing $\mathcal{R}$ such that $J_1 = 0$. This would constitute the so-called non-realizable Wiener filter.

A realizable $\mathcal{R}$ can only eliminate the *causal* parts of the integrand of J_1. Since $(\beta/DN)\mathcal{R}$ is causal, it remains to partition $(z^{-m}CC_*N_*)/(r\beta_*D)$.

This expression can be partitioned as a sum of causal and noncausal terms, a partial fraction expansion, by introducing polynomials $Q_1(z^{-1})$ and $L_*(z)$ such that

$$\frac{z^{-m}CC_*N_*}{r\beta_*D} = \frac{Q_1}{D} + \frac{zL_*}{r\beta_*} \ . \tag{23}$$

The term $Q_1(z^{-1})/D(z^{-1})$ represents the causal part and $(zL_*(z)/(r\beta_*(z))$ the (strictly) noncausal part. By setting the right hand side of (23) on common denominator form, we obtain

$$z^{-m}CC_*N_* = r\beta_*Q_1 + zDL_* \tag{24}$$

which is (20), with z exchanged for q. Using (23) to express J_1 gives

$$J_1 = \frac{\lambda_e}{2\pi j}\oint_{|z|=1} r\left(\frac{\beta}{DN}\mathcal{R} - \frac{Q_1}{D} - \frac{zL_*}{r\beta_*}\right)\left(\frac{\beta}{DN}\mathcal{R} - \frac{Q_1}{D} - \frac{zL_*}{r\beta_*}\right)_*\frac{dz}{z} \ .$$

By expanding the integrand, J_1 may be written as a sum of four terms:

$$\begin{aligned} V_1 &= \frac{\lambda_e}{2\pi j}\oint_{|z|=1} r\left(\frac{\beta}{DN}\mathcal{R} - \frac{Q_1}{D}\right)\left(\frac{\beta_*}{D_*N_*}\mathcal{R}_* - \frac{Q_{1*}}{D_*}\right)\frac{dz}{z} \\ V_2 &= -\frac{\lambda_e}{2\pi j}\oint_{|z|=1}\left(\frac{\beta}{DN}\mathcal{R} - \frac{Q_1}{D}\right)\frac{z^{-1}L}{\beta}\frac{dz}{z} \\ V_3 &= -\frac{\lambda_e}{2\pi j}\oint_{|z|=1} z\frac{L_*}{\beta_*}\left(\frac{\beta_*}{D_*N_*}\mathcal{R}_* - \frac{Q_{1*}}{D_*}\right)\frac{dz}{z} \\ V_4 &= \frac{\lambda_e}{2\pi j}\oint_{|z|=1}\frac{LL_*}{r\beta\beta_*}\frac{dz}{z} \ . \end{aligned}$$

For any causal and stable choice of the rational filter $\mathcal{R}$, all poles of the integrand of V_3 will be located outside the unit circle, since β, D and N are all stable. Hence, $V_3 = 0$. (Note that it is crucial that zL_*/β_* is strictly noncausal, starting with a free z. This z cancels the pole at the origin of V_3.) For symmetry reasons[6], $V_2 = 0$. The term V_4 does not depend on $\mathcal{R}$. Thus, the criterion J_1 is minimized by minimizing V_1. We readily obtain $V_1 = 0$ by choosing

$$\frac{\beta}{DN}\mathcal{R} - \frac{Q_1}{D} = 0 \ .$$

This gives

$$\mathcal{R} = \frac{Q_1N}{\beta}$$

[6]Perform the variable substitution $w = z^{-1}$.

where Q_1, together with L_*, is the solution to (24) and β is the stable polynomial spectral factor, obtained from (17). The minimal criterion value is $J_{\min} = J_2 + V_4$. This derivation should be compared to steps 2. and 3. of the derivation in Section III,C.

E THE CLASSICAL WIENER SOLUTION

Wiener filters are traditionally designed by first whitening the measurements and then multiplying them by the cross spectral density, $\phi_{f\epsilon}$, between desired signal and whitened mesurement. See, for example, the classical paper by Bode and Shannon [2]. For the filtering problem of Section III,B, with $f(k) = s(k)$, the causal Wiener filter is

$$\hat{s}(k|k+m) = \{\phi_{s\epsilon}\}_+ \epsilon(k+m)$$

where $\epsilon(k+m) = \mathcal{V}(q^{-1})y(k+m)$ is the whitened measurement, cf (16), and the notation $\{\cdot\}_+$ represents the use of only the causal part of the expression $\{\cdot\}$. Thus,

$$\mathcal{R} = \{\phi_{s\epsilon}\}_+ \mathcal{V} = \{\phi_{sy}\mathcal{V}_*\}_+ \mathcal{V} \ . \tag{25}$$

Above, ϕ_{sy} is the cross–spectral density of desired signal and measurement. Apart from a scale factor, $1/\sqrt{\lambda_\epsilon}$, the whitening filter is the inverse of the innovations model (16):

$$\mathcal{V} \ = \ \frac{DN}{\sqrt{\lambda_\epsilon}\beta} \ . \tag{26}$$

The expression (25) is elegant. However, $\{\phi_{sy}\mathcal{V}_*\}_+$ is not explicit, in terms of polynomial coefficients of rational transfer functions of the signal and noise models. The polynomial systems framework is of help here. It can be used to evaluate the causal expression $\{\cdot\}_+$ in an efficient way.

Since $e(k)$ and the noise $v(k)$ are mutually uncorrelated and the measurement is $y(k+m) = s(k+m) + (M/N)v(k+m)$, we readily obtain

$$\phi_{sy} = \phi_{s(k)y(k+m)} = \phi_{s(k)s(k+m)} = \frac{C}{D} z^{-m} \frac{C_*}{D_*}\lambda_e \Big|_{z=e^{j\omega}} \ .$$

Thus, (25) becomes, with $r \triangleq \lambda_\epsilon / \lambda_e$,

$$\mathcal{R}(q^{-1}) = \left\{ \frac{C}{D} q^{-m} \frac{C_*}{D_*} \lambda_e \frac{D_* N_*}{\sqrt{\lambda_\epsilon}\beta_*} \right\}_+ \frac{DN}{\sqrt{\lambda_\epsilon}\beta} = \left\{ q^{-m} \frac{C C_* N_*}{D r \beta_*} \right\}_+ \frac{DN}{\beta} \ . \tag{27}$$

Extraction of the causal part $\{\cdot\}_+$ of the double–sided function, corresponds to performing a partial fraction expansion. Let

$$q^{-m} \frac{C(q^{-1}) C_*(q) N_*(q)}{D(q^{-1}) r \beta_*(q)} = \frac{Q_1(q^{-1})}{D(q^{-1})} + \frac{\tilde{L}_*(q)}{r\beta_*(q)} \tag{28}$$

for some polynomials $Q_1(q^{-1})$ and $\tilde{L}_*(q)$. Terms without delay should appear exclusively in the causal part, so the noncausal part starts with a free q–term. Thus, let $\tilde{L}_*(q) \triangleq qL_*(q)$. (This avoids the occurrence of an error pointed out by Chen [36].) Multiplying both sides of (28) by $Dr\beta_*$ then gives

$$q^{-m} C C_* N_* = r\beta_* Q_1 + qDL_* \ .$$

Once again, this is precisely the linear Diophantine equation (20). Thus, the causal Wiener filter, predictor or smoother is

$$\mathcal{R}(q^{-1}) = \left\{ \frac{Q_1}{D} + \frac{qL_*}{r\beta_*} \right\}_+ \frac{DN}{\beta} = \frac{Q_1 DN}{D\beta} \tag{29}$$

which, of course, coincides with the expression (21), if the stable factor D is cancelled. (Unstable systems are not allowed in the classical Wiener formulation.) The link between partial fraction expansion and Diophantine equations was noted by Grimble [17], and has also been independently noted by us and by others. This link also plays a key role in the "completing the squares"–reasoning, cf Eq. (23).

F INNER–OUTER FACTORIZATION

Vidyasagar [32] has discussed a factorization approach to optimal filtering. This subsection is based on that approach. To explain it, we need a brief recapitulation of inner and outer matrices and their properties. Consider rational matrices with n rows and m columns, having stable discrete–time transfer functions as elements. Let such matrices be denoted $\mathcal{P}^{n|m}(z^{-1})$, or just $\mathcal{P}$, and their conjugate transpose $\mathcal{P}_*^{m|n}(z)$ (or $\mathcal{P}_*$). We need the following definitions (see [32] and [37]).

- A stable rational matrix $\mathcal{P}^{n|m}(z^{-1})$, $n \geq m$, is *inner* if $\mathcal{P}_*\mathcal{P} = \mathbf{I}_m$ for almost all $|z| = 1$. It is *co–inner* if $n \leq m$ and $\mathcal{P}\mathcal{P}_* = \mathbf{I}_n$ for almost all $|z| = 1$.
- A stable rational matrix $\mathcal{P}^{n|m}(z^{-1})$, $n \leq m$, is *outer* if and only if it has full row rank n, $\forall\ |z| \geq 1$. In other words, it has no zeros in $|z| \geq 1$. It is *co–outer* when $n \geq m$ if and only if it has full column rank m, $\forall\ |z| \geq 1$.
- A stable rational matrix $\mathcal{P}^{n|m}(z^{-1})$, with full rank $p \triangleq \min\{m,n\}$ for all $z = e^{j\omega}$ (no zeros on the unit circle), has an *inner–outer factorization*

$$\mathcal{P}^{n|m} = \mathcal{P}_i^{n|p}\mathcal{P}_o^{p|m} \tag{30}$$

with the outer factor $\mathcal{P}_o$ having a stable right inverse. If $n \geq m$ $\mathcal{P}_o$ is square and its inverse is unique. It also has a *co–inner–outer factorization*

$$\mathcal{P}^{n|m} = \mathcal{P}_{co}^{n|p}\mathcal{P}_{ci}^{p|m} \tag{31}$$

with the co–outer factor $\mathcal{P}_{co}$ having a stable left inverse. If $n \leq m$, the co–outer matrix is square, and its inverse is unique.

Inner and co–inner matrices are generalizations of scalar all–pass links. Multiplication by a (co)inner matrix does not affect the spectral density or power of a signal vector. The important property of outer and co–outer matrices is that they are *stably* invertible. Additionally, the inverses are *causal* if the instantaneous gain matrices $\mathcal{P}_o(0)$ and $\mathcal{P}_{co}(0)$ have full rank$p$.

Now, minimizing (7) is, for the filtering problem of Section III,B, equivalent to minimizing

$$J = \left\| \left[\frac{C}{D}\lambda_e^{1/2} \ \ 0\right] - \mathcal{R}\left[z^m\frac{C}{D}\lambda_e^{1/2} \ \ z^m\frac{M}{N}\lambda_v^{1/2}\right]\right\|_2^2 \tag{32}$$

where

$$\| \mathcal{P}(z^{-1}) \|_2^2 = \frac{1}{2\pi j}\mathrm{tr}\oint_{|z|=1} \mathcal{P}\mathcal{P}_*\frac{dz}{z} \ .$$

The idea is now to factor the second term of (32) as

$$\boldsymbol{U} \triangleq \left[z^m\frac{C}{D}\lambda_e^{1/2} \ \ z^m\frac{M}{N}\lambda_v^{1/2}\right] = U_{co}\boldsymbol{U}_{ci} \tag{33}$$

where U_{co} is co–outer of dimension $1|1$ and $\boldsymbol{U}_{ci}$ is co–inner of dimension $1|2$. The scalar co–outer will have a stable inverse, if the left hand side of

(33) has full rank 1 for all $|z| = 1$. [7] The inverse $U_{co}^{-1}(z^{-1})$ is causal if and only if $U_{co}(0) \neq 0$.

By invoking (33), the criterion (32) can be written as

$$J = \left\| \left[\frac{C}{D}\lambda_e^{1/2} \ \ 0\right] - \mathcal{R}U_{co}\boldsymbol{U}_{ci}\right\|_2^2 \ . \tag{34}$$

Now, multiplying the interior of the norm in (34) from the right by $\boldsymbol{U}_{ci*}$, which is normpreserving, and using the co–inner property, $\boldsymbol{U}_{ci}\boldsymbol{U}_{ci*} = 1$ on $|z| = 1$, gives

$$J = \left\| \left[\frac{C}{D}\lambda_e^{1/2} \ \ 0\right]\boldsymbol{U}_{ci*} - \mathcal{R}U_{co}\right\|_2^2 \ .$$

By decomposing into *causal* and *noncausal* parts, the causal and stable filter $\mathcal{R}$, which minimizes J, is readily found from the requirement that $\mathcal{R}$ should eliminate the whole causal part. Thus,

$$\mathcal{R}U_{co} = \left\{\left[\frac{C}{D}\lambda_e^{1/2} \ \ 0\right]\boldsymbol{U}_{ci*}\right\}_+ \tag{35}$$

where $\{\cdot\}_+$, as before, represents the causal part. The optimal filter thus becomes

$$\mathcal{R} = \left\{\left[\frac{C}{D}\lambda_e^{1/2} \ \ 0\right]\boldsymbol{U}_{ci*}\right\}_+ U_{co}^{-1} \tag{36}$$

where the (left) inverse U_{co}^{-1} is stable by definition.

The factorization–based solution thus consists of first performing a co–inner–outer factorization (33) and then the causal–noncausal factorization required in (35). We will now emphasize the correspondence of these two steps to the previous solutions.

If the spectral factorization (17) has been solved, the co–inner and co–outer factors can be obtained as

$$U_{co} = \frac{\lambda_\epsilon^{1/2}\beta}{DN} \qquad \boldsymbol{U}_{ci} = \left[\frac{\lambda_e^{1/2}z^mCN}{\lambda_\epsilon^{1/2}\beta} \ \ \frac{\lambda_v^{1/2}z^mMD}{\lambda_\epsilon^{1/2}\beta}\right] \ . \tag{37}$$

It is easily verified that $\boldsymbol{U} = U_{co}\boldsymbol{U}_{ci}$ and that, with $r = \lambda_\epsilon/\lambda_e$, $\rho = \lambda_v/\lambda_e$, cf. (17), we obtain

$$\boldsymbol{U}_{ci}\boldsymbol{U}_{ci*} = \frac{\lambda_eCN(CN)_* + \lambda_vMD(MD)_*}{\lambda_\epsilon\beta\beta_*} = 1 \ . \tag{38}$$

[7] In other words, with $U = \frac{z^m}{DN}[CN\lambda_e^{1/2} \ \ MD\lambda_v^{1/2}]$, $CN\lambda_e^{1/2}$ and $MD\lambda_v^{1/2}$ should have no common factors with zeros on $|z| = 1$. This corresponds to the condition for existence of a stable spectral factor in (17), included in Assumption A.

Furthermore, U_{co} given by (37) has no zero in $|z| \geq 1$, and is therefore stably invertible, whenever a stable spectral factor β exists. The construction above is an application of the standard way of performing inner–outer factorizations: by means of spectral factorization, see e.g. [37].

Using (37), the optimal filter (36) can be expressed as

$$\mathcal{R} = \left\{ \left[\lambda_e^{1/2} \frac{C}{D} \quad 0 \right] U_{ci*} \right\}_+ U_{co}^{-1} = \left\{ \frac{\lambda_e z^{-m} C C_* N_*}{\lambda_\epsilon D \beta_*} \right\}_+ \frac{DN}{\beta} \tag{39}$$

where the scalar $\lambda_\epsilon^{-1/2}$ from U_{co}^{-1} has been absorbed into the $\{\cdot\}_+$–factor.

The causal bracket operation is the same as in the classical Wiener–solution. Thus, exchange q for z and introduce polynomials $Q_1(q^{-1})$ and $L_*(q)$, such that the rational function inside the brackets of (39) can be expressed as the sum of a causal and a noncausal term

$$\frac{\lambda_e q^{-m} C(q^{-1}) C_*(q) N_*(q)}{\lambda_\epsilon D(q^{-1}) \beta_*(q)} = \frac{Q_1(q^{-1})}{D(q^{-1})} + \frac{q L_*(q)}{\lambda_\epsilon \beta_*(q)} \quad . \tag{40}$$

Thus, the $\{\cdot\}_+$–factor equals Q_1/D. By setting the expression (40) on a common denominator, utilizing that $r = \lambda_\varepsilon/\lambda_e$, we obtain the Diophantine equation (20). The estimator (39) equals (21):

$$\mathcal{R} = \frac{Q_1}{D}\frac{DN}{\beta} = \frac{Q_1 N}{\beta} \quad . \tag{41}$$

Observe that the inverse of the co–outer, U_{co}, is nothing but the well–known whitening filter $\mathcal{V}$ in (26), from the classical Wiener solution. As in that case, unstable D–polynomials are not allowed.

G A COMPARATIVE DISCUSSION

Above, we have presented four different routes to the MSE–optimal solution. Which route to be preferred is more or less a matter of taste and background. There are, however, some aspects a problem solver should be aware of. We shall briefly summarize them below.

Evidently, the four approaches arrive, one way or another, at a polynomial spectral factorization and a Diophantine equation. In the variational approach, spectral factorization arises as an obvious simplification of the cross spectral density $\phi_{\varepsilon\mathcal{V}^*}$. See Eq. (19). The same is true for the "completing the squares"-approach, but there it simplifies the criterion expression.

In the classical Wiener solution, the spectral factorization determines the whitening filter while in the inner–outer factorization approach, it is part of the inner outer–factorization. In particular, it is defined by the inner property Eq. (38). The inverse of the outer matrix is the whitening filter in the classical solution, while the bracket term in Eq. (39) is just another way of writing $\{\phi_{sy}\mathcal{V}_*\}_+$, cf. Eqs. (25) and (27) .

Spectral factorization can be avoided in noise–free situations, with stably invertible models, such as the prediction problem (1)–(3). In problems with noise, it can be avoided only in very special cases, such as the optimization of decision feedback equalizers. That problem is discussed in Section VI.

It is interesting to note how the Diophantine equation arises. In all formulations except in the variational approach, it originates from a causal–noncausal partitioning, a partial fraction expansion, where the causal factor $\{\cdot\}_+$ is sought. In the variational approach, it arises from the requirement that the variational term should be orthogonal to the error.

One can also note that in multivariable problems ($\ell > 1$ in (14)), the variational optimization operates on the whole covariance matrix (6); in (13), elementwise orthogonality is required. The other three discussed methods depart from the scalar variance criterion (7). The solutions are the same. This means that all approaches do, in fact, result in a minimal covariance matrix, not only in a minimal scalar trace.

Problems with unstable models can be handled by the variational approach (see Appendix B) and by the "completing the squares" method. They cannot be handled by the classical Wiener approach or by inner–outer factorization.

A disadvantage with the variational approach is that some extra calculations are required to obtain the criterion value. In the "completing the squares" approach, the minimum of (7) comes as a bonus.

A disadvantage with the "completing the squares" approach is that it will, in difficult problems, be hard to complete the square: the solution has to be known (or suspected) in order to succeed. On the other hand it requires, in essence, the simplest mathematics: just quadratic forms are needed. (This is more apparent in a time domain formulation.) In the classical solution, it might be difficult to find the right track from the expression (25) to an explicit solution. In particular, this is not straightforward in Section VI. The same is true for the inner–outer factorization approach.

One advantage with the variational approach is that the free q–factor in e.g. (20) emerges automatically from the cancellation of the free z in the denominator of (19). In the other methods, one has to include direct terms only in the causal part of the causal–noncausal partitioning, to avoid a sub-optimal solution [36]. Variational minimization leads to the solution along a constructive and systematic route. This is of considerable importance in more difficult problem formulations. See e.g. Section VI.

H SOLUTION OF SPECTRAL FACTORIZATIONS

There exist efficient numerical algorithms for performing polynomial (FIR) spectral factorizations, such as (17):

$$\begin{aligned} rz^{n\beta}\beta(z^{-1})\beta_*(z) &= z^{n\beta}[C(z^{-1})C_*(z)N(z^{-1})N_*(z) \\ &+ \rho M(z^{-1})M_*(z)D(z^{-1})D_*(z)] \ . \end{aligned} \qquad (42)$$

The most obvious way is, perhaps, to calculate the $2n\beta$ roots of the right–hand side of (42). They are symmetrical with respect to the unit circle. Then, create the polynomial $z^{n\beta}\beta(z^{-1}) = z^{n\beta} + z^{n\beta-1}\beta_1 + \ldots + \beta_{n\beta}$ from the $n\beta$ factors with roots inside $|z| = 1$.

There exist several iterative, mostly Newton–based, algorithms for spectral factorization, see e.g. [6] or [38]. MATLAB–code for one of them is given in Appendix D. Convergence of iterative algorithms is in general fast, unless there are multiple roots close to $|z| = 1$.

Closed–form expressions exist for second–order spectral factors with real–valued coefficients [39]. If the right–hand side of (42) is given by

$$q^2[g_o + g_1(q+q^{-1}) + g_2(q^2+q^{-2})]$$

where we have substituted q for z, the scalar r and the polynomial $\beta(q^{-1}) = 1 + \beta_1 q^{-1} + \beta_2 q^{-2}$ are given by

$$\gamma \triangleq \frac{g_o}{2} - g_2 + \sqrt{\left(\frac{g_o}{2} + g_2\right)^2 - g_1^2}$$

$$r = \left(\gamma + \sqrt{\gamma^2 - 4g_2^2}\right)/2 \ ; \ \beta_1 = \frac{g_1}{r+g_2} \ ; \ \beta_2 = \frac{g_2}{r} \ . \qquad (43)$$

I A SCALAR DESIGN EXAMPLE

The purpose of this section is to illustrate what sort of numerical computations that are required, when working with the polynomial approach in a typical design case.

In Section III,B a scalar filtering problem was formulated. The signal, to be estimated there, could be regarded as a signal measured by an ideal transducer, having unit transfer function, and disturbed by noise.

Here, we will make a slight modification of that problem set–up, by assuming the signal to be observed through a transducer having transfer function $B(q^{-1})$. Thus, the signal

$$f(k) = u(k) = \frac{C(q^{-1})}{D(q^{-1})}e(k) \qquad (44)$$

is to be estimated from noisy observations

$$y(k) = B(q^{-1})u(k) + \frac{M(q^{-1})}{N(q^{-1})}v(k) \qquad (45)$$

up to time k. Compared to Section III,C, where $B(q^{-1}) = 1$, we have to make some minor modifications. Since the spectral density $\Phi_y(e^{i\omega})$ is now affected also by $B(q^{-1})$, the spectral factorization (17) becomes

$$r\beta\beta_* = CC_*BB_*NN_* + \rho MM_*DD_* \ . \qquad (46)$$

Moreover, by substituting $\mathcal{R}B$ for $\mathcal{R}$ in the first of the right hand terms of (18), we obtain the Diophantine equation

$$q^m CC_*B_*N_* = r\beta_*Q_1 + qDL_* \ . \qquad (47)$$

Now, let the polynomials in (44) and (45) be given by

$$C(q^{-1}) = M(q^{-1}) = N(q^{-1}) = 1$$

$$D(q^{-1}) = 1 - 0.5q^{-1} \ ; \ B(q^{-1}) = 1 - 1.4q^{-1} + 0.92q^{-2} \ .$$

The noise variances are $\lambda_e = 1$, $\lambda_v = \rho = 0.01$ and the smoothing lag, m, is zero. With these choices we find that the spectral factor in (46) is of order two, and can thus be calculated by means of the analytical expressions of Section III,H. We obtain

$$g_o + g_1(q+q^{-1}) + g_2(q^2+q^{-2}) = 3.8189 - 2.6930(q+q^{-1}) + 0.92(q^2+q^{-2}) \qquad (48)$$

which gives $\gamma = 1.8575$. Inserting this into (43) leads to

$$\begin{aligned} r &= 1.0560 \\ \beta(q^{-1}) &= 1 - 1.3629q^{-1} + 0.8712q^{-2} \ . \end{aligned} \tag{49}$$

The Diophantine equation (47) shall be solved for Q_1 and L_*. Their orders are determined so that they match the highest power of q^{-1} and q on both sides of (47).[8] Here Q_1 and L_* will be of order $nd - 1 = 0$ and $n\beta - 1 = 1$, respectively. We thus have to solve

$$1 - 1.4q + 0.92q^2 = 1.056(1 - 1.3629q + 0.8712q^2)Q_{10} + (-0.5 + q)(\ell_o + \ell_1 q) \ . \tag{50}$$

By equating (50) for equal powers of q, we obtain a system of linear equations[9]

$$\begin{pmatrix} 1 \\ -1.4 \\ 0.92 \end{pmatrix} = \begin{pmatrix} 1.056 & -0.5 & 0 \\ -1.4392 & 1 & -0.5 \\ 0.92 & 0 & 1 \end{pmatrix} \begin{pmatrix} Q_{10} \\ \ell_o \\ \ell_1 \end{pmatrix} . \tag{51}$$

The solution is

$$\begin{aligned} Q_1 = Q_{10} &= 0.9357 \\ L_* &= -0.0238 + 0.0592q \ . \end{aligned}$$

Thus, the optimal estimator, cf (21), becomes

$$\hat{u}(k|k) = \frac{Q_1 N}{\beta} y(k) = \frac{0.9357}{1 - 1.3629q^{-1} + 0.8712q^{-2}} y(k) \tag{52}$$

and the minimal mean square error is readily calculated from[10]

$$J_{\min} = \frac{\lambda_e}{2\pi j} \oint_{|z|=1} \frac{LL_* + \rho CMC_*M_*}{r\beta\beta_*} \frac{dz}{z} = \frac{\lambda_e}{2\pi j} \oint_{|z|=1} \frac{LL_* + \rho}{r\beta\beta_*} \frac{dz}{z} \ . \tag{53}$$

In this example, with $\rho = 0.01$ and L_* from above, $J_{min} = 0.1006$.

In Figure 6 the Bode magnitude plots of the transducer $B(q^{-1})$ and the optimal filter $\hat{u}(k|k)$ are depicted. Note the location of the notch and peak of the transducer and optimal filter respectively. The zeros of the transducer are located at $0.7 \pm i0.6557$, while the poles of the filter are located at $0.68 \pm i0.6378$. If it were not for the additional noise, perfect inversion would have been obtained. Here, the filter poles are moved slightly inward.

[8] If the orders of Q_1 and L_* are chosen too high, superfluous coefficients will be zero. On the other hand, if too low orders are chosen, no solution can be found. For more details, see Appendix A.

[9] Diophantine equations can also be solved by means of the Euclidian algorithm, see for example, [26].

[10] See, for example, [13] Eq. (3.5) or add J_2 and V_4 of Section III,D utilizing (17).

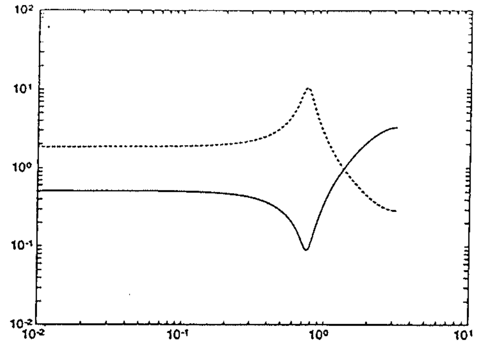

Figure 6: Bode magnitude plots for the transducer $B(q^{-1})$ (solid) and the optimal filter $\hat{u}(k|k)$ (dashed).

IV MULTISIGNAL DECONVOLUTION

In many areas, it is of interest to estimate the input to a linear system, cf the example above. A recent interesting application in the area is the reconstruction of stereophonic sound, described by Nelson et.al. [40]. Others are described in [13], [41], [42], [43], and the references therein.

We will consider the problem of deconvolution with multiple inputs and multiple outputs. The formulation used here includes all problems described by the general structure of Figure 2. The solution illustrates the application of the variational approach to multi–signal filtering problems.

In [14], the deconvolution problem was set up and solved using MFD's. Here we will approach the problem using a different parametrization: some transfer functions will be represented 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* are needed, 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 polynomials to be determined appear on the same side in different terms of the equation, instead of on opposite sides. This will make the solution attractive, both from a numerical and a pedagogical point of view: solving a unilateral Diophantine equation corresponds to solving a block–Toeplitz system of linear equations. How such a system of equations is formed and solved, is illustrated in Section IV,B.

A DERIVING THE SOLUTION

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

$$\begin{aligned} y(k) &= \boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})u(k) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})v(k) \\ u(k) &= \frac{1}{D(q^{-1})}\boldsymbol{C}(q^{-1})e(k) \ . \end{aligned} \quad (54)$$

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 *diagonal*. As in previous sections, $\{e(k)\}$ 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. The 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

$$\hat{f}(k|k+m) = \boldsymbol{\mathcal{R}}(q^{-1})y(k+m) \quad (55)$$

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

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

is sought. The covariance matrix (6), or the sum of MSE's (7), is to be minimized with dynamic weighting

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

This corresponds to the choice $\boldsymbol{\mathcal{G}} = \boldsymbol{A}^{-1}\boldsymbol{B}$, $\boldsymbol{\mathcal{G}}_a = \boldsymbol{0}$ ($\Rightarrow \boldsymbol{\mathcal{R}}_a = \boldsymbol{0}$), $\boldsymbol{\mathcal{F}} = \boldsymbol{C}/D$, $\boldsymbol{\mathcal{H}} = \boldsymbol{N}^{-1}\boldsymbol{M}$, $\boldsymbol{\mathcal{D}} = \boldsymbol{S}/T$, $\boldsymbol{\mathcal{W}} = \boldsymbol{V}/U$, and $\boldsymbol{\mathcal{R}}_z = [\boldsymbol{\mathcal{R}}\ \boldsymbol{0}]$ in (6),(9) and (10). See Figure 7. The filter $\boldsymbol{S}/T$, with T scalar and $\boldsymbol{S}$ of dimension $\ell|s$, may represent additional dynamics in the problem description (cf [15],[16]), a frequency shaping weighting filter (cf [13]), or a selection of certain states.

When $\{u(k)\}$ constitutes a sequence of digital symbols in a communication network, (55) is a multivariable version of a *linear equalizer*. In order to recover the transmitted symbols, the estimate $\hat{u}(k)$ is fed into a decision device. See e.g. [44] and [23], for a discussion of the scalar case.

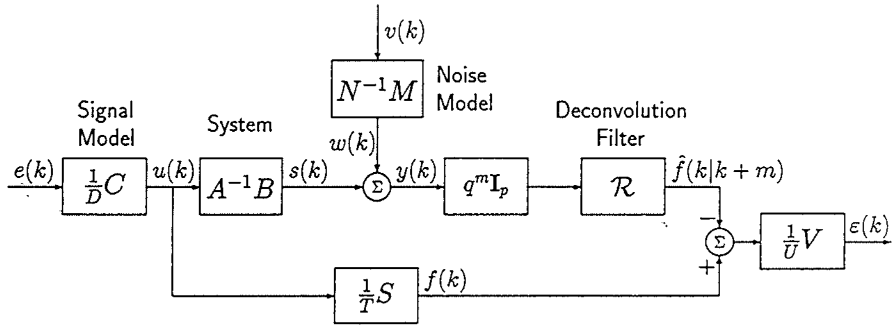

Figure 7: A generalized multi–signal deconvolution problem. The vector sequence $\{f(k)\}$ is to be estimated from the measurements $\{y(k)\}$, up to time $k+m$.

Introduce the following assumptions.

Assumption 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 2: The spectral density of $y(k)$, $\boldsymbol{\Phi}_y(e^{j\omega})$, is nonsingular for all ω.

From (54) we now obtain the spectral density matrix $\boldsymbol{\Phi}_y$ as

$$\boldsymbol{\Phi}_y = \frac{1}{DD_*}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1} + \boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{N}_*^{-1} = \boldsymbol{\alpha}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\boldsymbol{\alpha}_*^{-1} \quad (56)$$

where

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{N}\boldsymbol{B}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* + DD_*\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_* \quad (57)$$

and

$$\boldsymbol{\alpha} \triangleq D\boldsymbol{N}\boldsymbol{A} \ .$$

Under Assumption 2, a stable $p|p$ spectral factor $\boldsymbol{\beta}$, with $\det\boldsymbol{\beta}(z^{-1}) \neq 0$ in $|z| \geq 1$ and nonsingular leading matrix $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(0)$, can always be found. Now, the optimal estimator can be derived as outlined in Section III,A. Let $\varepsilon(k) = (\boldsymbol{V}/U)(f(k) - \hat{f}(k|k+m))$ be the filtered error and $\nu(k) = \boldsymbol{\mathcal{T}}(q^{-1})y(k+m)$ the variation. Since $e(k)$ and $v(k)$ are assumed uncorrelated, we obtain

$$E\varepsilon(k)\nu^*(k) = E\frac{1}{U}\boldsymbol{V}[(\frac{1}{T}\boldsymbol{S} - q^m\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B})\frac{1}{D}\boldsymbol{C}e(k) - q^m\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k)]$$

$$[\boldsymbol{\mathcal{T}}q^m(\frac{1}{D}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}e(k) + \boldsymbol{N}^{-1}\boldsymbol{M}v(k))]^*$$

$$= \frac{1}{2\pi j}\oint_{|z|=1}\frac{1}{U}\boldsymbol{V}\{z^{-m}\frac{1}{TDD_*}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1}$$

$$-\boldsymbol{\mathcal{R}}[\frac{1}{DD_*}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{A}_*^{-1} + \boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{N}_*^{-1}]\}\boldsymbol{\mathcal{T}}_*\frac{dz}{z} \quad . \tag{58}$$

The use of (56) in (58) gives, with $\boldsymbol{\alpha}_*^{-1} = D_*^{-1}\boldsymbol{N}_*^{-1}\boldsymbol{A}_*^{-1}$,

$$E\varepsilon(k)\nu^*(k) = \frac{1}{2\pi j}\oint\frac{1}{U}\{\frac{z^{-m}}{TD}\boldsymbol{V}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* - \boldsymbol{V}\boldsymbol{\mathcal{R}}\boldsymbol{\alpha}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\}\boldsymbol{\alpha}_*^{-1}\boldsymbol{\mathcal{T}}_*\frac{dz}{z} \quad . \tag{59}$$

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 $\boldsymbol{\beta}$ may contribute poles at the origin, since they are polynomials in z^{-1}. These factors can be cancelled directly by $\boldsymbol{\mathcal{R}}$. Moreover, if $\boldsymbol{\mathcal{R}}$ contains $\boldsymbol{V}^{-1}/T$ as a left factor, $\boldsymbol{V}$ is cancelled and $1/TD$ can be factored out from the two terms of the integrand, to be cancelled later. Thus,

$$\boldsymbol{\mathcal{R}} = \frac{1}{T}\boldsymbol{V}^{-1}\boldsymbol{Q}\boldsymbol{N}\boldsymbol{A}\boldsymbol{\beta}^{-1} \tag{60}$$

where $\boldsymbol{Q}(q^{-1})$, of dimension $\ell|p$, is undetermined. With $P \triangleq UTD$ and (60) inserted, (59) becomes

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

All poles, of every element, of $\boldsymbol{\alpha}_*^{-1}\boldsymbol{\mathcal{T}}_*$ are located outside $|z| = 1$, since $\boldsymbol{\alpha}$ is stable and $\boldsymbol{\mathcal{T}}$ is causal and stable. In order to fulfill (14), we require

$$z^{-m}\boldsymbol{V}\boldsymbol{S}\boldsymbol{C}\boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_* = \boldsymbol{Q}\boldsymbol{\beta}_* + z\boldsymbol{L}_*P\mathbf{I}_p \quad . \tag{61}$$

This is a linear polynomial matrix equation, a *unilateral* Diophantine equation. Here, $\boldsymbol{Q}(q^{-1})$ and $\boldsymbol{L}_*(q)$ are polynomial matrices, of dimension $\ell|p$, of degree

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

$$nL \leq \max(nc + nb + nn - m, n\beta) - 1 \tag{62}$$

where $nc = deg\boldsymbol{C}, ns = deg\boldsymbol{S}$ etc. The unique solvability of (61) is demonstrated in Appendix C.

The design equations thus consist of the left spectral factorization (57), the Diophantine equation (61) and the filter expression (60). The solution above constitutes a simplification of the one in [14], but extended with a frequency dependent weighting $\boldsymbol{\mathcal{W}} = \boldsymbol{V}/U$ of the error signal. For scalar systems, the solution reduces to the one presented in [13]. See also [16].

The spectral factor $\boldsymbol{\beta}$ is unique up to a right orthogonal matrix. (If $\boldsymbol{H}\boldsymbol{H}_* = \boldsymbol{I}$ then $\boldsymbol{\beta}\boldsymbol{\beta}_* = (\boldsymbol{\beta}\boldsymbol{H})(\boldsymbol{H}_*\boldsymbol{\beta}_*)$.) There exist severel efficient algorithms for polynomial matrix spectral factorization, some of which are based on state space methods. An survey of different algorithms is presented in [45]. See also [46].

The minimal (scalar) criterion value is obtained by inserting (60), (56), and (61), in this order, into J in (7). We thus obtain, with $\boldsymbol{H} \triangleq \boldsymbol{V}\boldsymbol{S}\boldsymbol{C}$,

$$\frac{1}{2\pi j}\oint \mathrm{tr}\{\boldsymbol{L}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{L} + \frac{1}{PP_*}\boldsymbol{H}(\mathbf{I}_n - \boldsymbol{C}_*\boldsymbol{B}_*\boldsymbol{N}_*\boldsymbol{\beta}_*^{-1}\boldsymbol{\beta}^{-1}\boldsymbol{N}\boldsymbol{B}\boldsymbol{C})\boldsymbol{H}_*\}\frac{dz}{z} \quad . \tag{63}$$

The minimal criterion value consists of two terms. The first term involves the sometimes so called "dummy"–polynomial matrix $\boldsymbol{L}_*$. In deconvolution problems, it can be given a nice interpretation: it represents the error caused by incomplete inversion of the system $\boldsymbol{A}^{-1}\boldsymbol{B}$. Only use of an infinite smoothing lag can make the first term vanish, unless the system is minimum phase and there is no noise. One can show that $\boldsymbol{L} \to 0$ when $m \to \infty$ [15].

The rule in the derivation technique used here is to cancel what can be cancelled directly, by means of $\mathcal{R}$. The rest of the terms contributing poles in $|z| < 1$ must be factored out, to be taken care of by $\boldsymbol{L}_*$ and $\boldsymbol{Q}$. It is instructive to note how the Diophantine equation interacts with the cross–term (59). It has to absorb contributing parts of the integrand which cannot be cancelled directly by $\mathcal{R}$. The polynomial matrix $\boldsymbol{L}_*$ represents the remainder.

There exists a very special case in which perfect input estimation is possible. It is the case of minimum–phase systems without noise, 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, $\mathcal{R} = \boldsymbol{B}^{-1}\boldsymbol{A}q^{-m}$ makes the integrand of (59) zero directly. Consequently, there is nothing left for $\boldsymbol{L}_*$ to take care of, so $\boldsymbol{L}_*$ must be zero.

It is well motivated to ask if the choosen parametrization, with scalar or diagonal denominators, implies any disadvantages. If transfer functions are identified from measured data it is natural to use diagonal denominator matrices. Such a parameterization can hardly be regarded as a restriction. Use of the common denominator form might, however, impose some limitations as compared to MFD's: elements of polynomial matrices may have unnecessarily high order. This might cause numerical problems in spectral factorization algorithms, if involved terms have common factors very close to the unit circle. In general this is, however, not the case so the common denominator form used here, will not be an essential restriction.

For scalar systems, the deconvolution problem has also been studied in an adaptive setting, see [47], [48]. Multivariable adaptive deconvolution, for the special case of white input and noise, has been discussed in [42] and [20]. Crucial for an adaptive algorithm to work, is that the model polynomials can be estimated from the output only. In [49], the identifiability properties of the scalar deconvolution problem are investigated and conditions for parameter identifiability are given. If similar conditions exist for the multivariable problem, is still an open question.

The considered deconvolution problem turns out to be dual to the *LQG (or $\mathcal{H}_2$)–feedforward* control problem (with rational weights). See [50]. It is very simple to demonstrate this duality. Reverse all arrows, interchange summation points and node points and transpose all rational matrices. Then, the block diagram for the other problem is obtained.

Note that the problem set–up discussed here contains the *general filtering problem* described by (4), (5) as the special case $v(k) = 0$. See Figure 2. The solution derived here thus solves *all* problems discussed in this chapter. (By duality, it also solves all $\mathcal{H}_2$–feedforward control problems.) However, it does not provide the same degree of expliciteness as do the solutions in Section III and Sections V–VI. One can simply not "see through" all the generality.[11] One of our convictions is that *structure gives insight.* Therefore, we have, in each specific problem, abandoned generality for structure, in order to gain insight, and also to simplify the solution.

B A MULTIVARIABLE DESIGN EXAMPLE

The purpose of this example is to illustrate the different types of numerical computations involved in a typical design problem and at the same time exemplify what can be achieved by multivariable filtering, as compared to scalar filtering. The numerical algorithms used in this example were implemented by Kent Öhrn. They are available upon request from the authors.

Asssume that a scalar signal $u(k)$ is to be estimated. It is described by a first order autoregressive process

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

Thus, $\boldsymbol{S}/T = 1, D = 1-0.5q^{-1}, \boldsymbol{C} = 1$, cf Figure 7. The signal is observed through two transducers (p=2). The measurements are contaminated by independent white noises, $w_1(k)$ and $w_2(k)$, having zero means and variances $\lambda = 0.01$. Thus,

$$y(k) = \boldsymbol{B}u(k) + w(k)$$

$$w(k) = \boldsymbol{M}v(k) = 0.1\,\mathbf{I}_2 v(k) \ . \tag{64}$$

The transducers are modelled by

$$\boldsymbol{B} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 0.100 + 0.080q^{-2} \\ 1 - 1.4q^{-1} + 0.92q^{-2} \end{pmatrix} . \tag{65}$$

They have zeros at $\pm i0.894$ and $0.7 \pm i0.656$ respectively. Note that the noise contaminating channel B_2 is the same as was used in the scalar design

[11] For example, the solution to the decision feedback equalization problem does not involve any (polynomial) spectral factorization. To see this from the general solution would be very hard.

example of Section III,I. However, in contrast to that example, we here have an additional measurement represented by channel B_1. A filter

$$\hat{u}(k) = \mathcal{R}y(k) = [Q^1 \;\; Q^2]/R$$

minimizing the unweighted ($\mathcal{W} = 1$) criterion (7) is sought. See Figure 8.

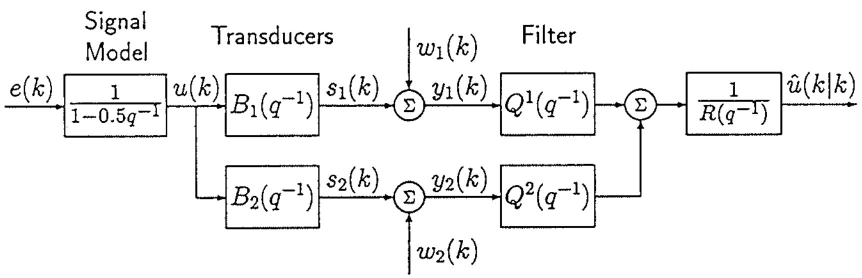

Figure 8: Model and filter structure. The scalar signal $u(k)$ is to be estimated using two measurement signals $y_1(k)$ and $y_2(k)$.

The spectral factorization (57) has dimension $p|p = 2|2$. In this example it becomes

$$\beta\beta_* = BB_* + DD_*MM_* = \begin{pmatrix} 0.008q^{-2} + 0.0164 + 0.008q^2 & 0.080q^{-2} - 0.112q^{-1} + 0.1736 - 0.14q + 0.092q^2 \\ 0.092q^{-2} - 0.14q^{-1} + 0.1736 - 0.112q + 0.08q^2 & 0.92q^{-2} - 2.688q^{-1} + 3.8064 - 2.688q + 0.92q^2 \end{pmatrix}$$

$$+ 0.01(-0.5q^{-1} + 1.25 - 0.5q)\,\mathbf{I}_2 \ . \tag{66}$$

By using the Newton–based algorithm described in [46] a stable spectral factor, (here rounded so that the elements contain at least three significant digits) with $\beta(0)$ nonsingular, was found to be

$$\beta = \begin{pmatrix} 0.1064 - 0.0423q^{-1} + 0.00286q^{-2} & 0.0982 - 0.00214q^{-1} + 0.0784q^{-2} \\ 0.00847q^{-1} + 0.0329q^{-2} & 1.0206 - 1.4013q^{-1} + 0.9014q^{-2} \end{pmatrix} . \tag{67}$$

The Diophantine equation (61) becomes

$$\boldsymbol{B}_* = \boldsymbol{Q}\boldsymbol{\beta}_* + q\boldsymbol{L}_*D\,\mathbf{I}_2 \ . \tag{68}$$

with degrees $n\boldsymbol{Q} = 0$ and $n\boldsymbol{L}_* = 1$. By expressing the polynomial matrices as matrix polynomials, equation (68) can be written as

$$(\mathbf{B}_0^* + \mathbf{B}_1^* q + \mathbf{B}_2^* q^2) =$$

$$= \mathbf{Q}_0(\beta_0^* + \beta_1^* q + \beta_2^* q^2) + (\mathbf{L}_0^* + \mathbf{L}_1^* q)(-0.5 + q)\,\mathbf{I}_2 \ .$$

Transpose this equation, note that $\mathbf{P}_i^{*T} = \mathbf{P}_i$, since $\mathbf{P}_i$ are real–valued, and equate both sides, for equal powers of q. We then obtain a linear system of six equations in block–Toeplitz form

$$\begin{pmatrix} \mathbf{B}_0 \\ \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \beta_0 & -0.5\mathbf{I}_2 & \mathbf{0} \\ \beta_1 & \mathbf{I}_2 & -0.5\mathbf{I}_2 \\ \beta_2 & \mathbf{0} & \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{Q}_0^T \\ \mathbf{L}_0 \\ \mathbf{L}_1 \end{pmatrix} .$$

By inserting numerical values for β_i, we obtain

$$\begin{pmatrix} .1 \\ 1 \\ 0 \\ -1.4 \\ .08 \\ .92 \end{pmatrix} = \begin{pmatrix} .10636 & .09818 & -.5 & 0 & 0 & 0 \\ 0 & 1.0206 & 0 & -.5 & 0 & 0 \\ -.04232 & -.00214 & 1 & 0 & -.5 & 0 \\ .00847 & -1.4013 & 0 & 1 & 0 & -.5 \\ .00286 & .07839 & 0 & 0 & 1 & 0 \\ .03285 & .90144 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Q_o^1 \\ Q_o^2 \\ \ell_o^1 \\ \ell_o^2 \\ \ell_1^1 \\ \ell_1^2 \end{pmatrix} .$$

Solving the linear system of equations for ${\boldsymbol{Q}_o}^T$ and $\boldsymbol{L}_*$ gives[12]

$$\boldsymbol{Q} = (\, 0.07890 \quad 0.97011 \,)$$
$$\boldsymbol{L}_* = (\, 0.00728 + 0.00373q \quad -0.01983 + 0.04291q \,) \ . \tag{69}$$

Finally, the estimator (60) becomes

$$\mathcal{R} = \boldsymbol{Q}\boldsymbol{\beta}^{-1} = \frac{1}{R}(Q^1 \; Q^2) \tag{70}$$

[12]Note that the elements of the displayed matrix are rounded to contain at least three significant digits.

where the monic denominator is $R(q^{-1}) = \det \beta(q^{-1}) / \det \beta_0$. We obtain

$$\begin{aligned} Q^1 &= 0.74186 - 1.0943q^{-1} + 0.36168q^{-2} \\ Q^2 &= 0.87917 - 0.37668q^{-1} - 0.03145q^{-2} \\ R &= 1 - 1.7786q^{-1} + 1.4269q^{-2} - 0.39381q^{-3} \end{aligned}$$

In Figure 9 the Bode magnitude plots for the estimator and the transfer function from $u(k)$ to $\hat{u}(k|k)$ are depicted.

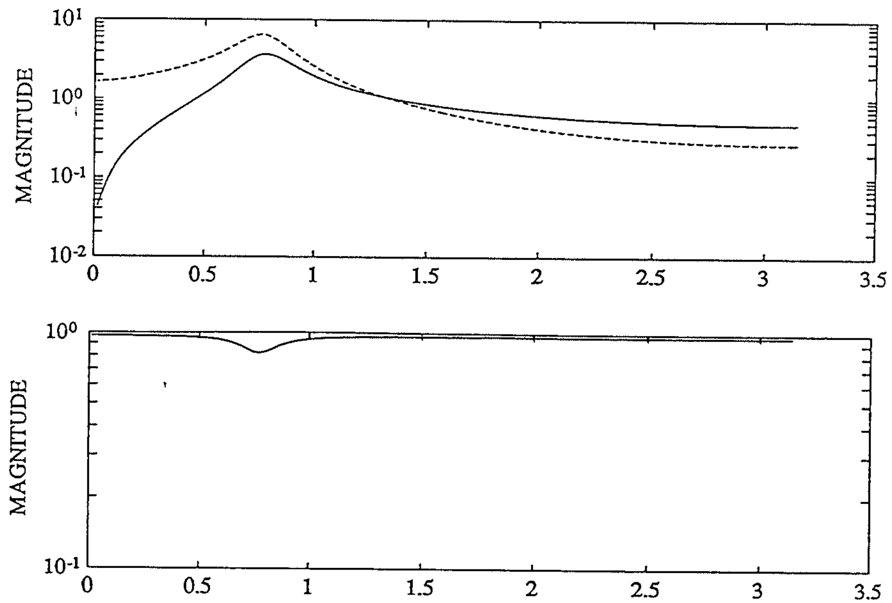

Figure 9: Bode magnitude plots for a: Multivariable estimator Q^1/R solid, Q^2/R dashed b: Transfer function from $u(k)$ to $\hat{u}(k|k)$.

From Figure 9a, it is evident that it is efficient to increase the gain for channel two at low frequencies whereas the gain should be low for channel one over the same frequency range. This is natural, since the signal $u(k)$ has much energy at low frequencies and channel one has very low amplification for that frequency range. By utilizing channel two, which has a larger local signal to noise ratio, unnecessary noise amplification is avoided.

Because of a low noise level and minimum phase transducers, the transfer function from $u(k)$ to $u(k|k)$ should be close to unity. Except for frequencies around $\omega = 0.76$, where channel two has a notch, this is also the case

as shown by Figure 9b. If the gain were increased more around the notch, the noise would be amplified too much, compared to the optimal solution.

From this example, we can learn that it is rewarding to use more than one measurement signal, if possible. The filter will use different signals in different frequency ranges, according to which one has the best local signal to noise ratio. This effect is further emphasized by comparing the minimum mean square errors obtained by a multivariable filter and a scalar filter, respectively. In this example the minimal mean square error is 0.0702 if both channels are used, whereas it is 0.5961 and 0.1006 if only channel one or channel two is used solely. The last number corresponds to the scalar design example of Section III,I.

V DIFFERENTIATION AND STATE ESTIMATION

The problem of estimating derivatives from measured data can be treated as an application of state estimation. It is an important engineering problem, which has been extensively studied over the years, see for example [51]–[55], and the references therein. In radar applications, velocity estimation from position data is of interest [56]. Other applications are, for example, estimation of heating rates from temperature data or net flow rates in tanks from level data. The estimation of derivatives is challenging because of its sensitivity to measurement noise. We will here describe a design method developed together with our colleague Dr Bengt Carlsson. For more details, see [52], [15] and [16].

Since the derivative is a continuous–time concept, it is appropriate to base the discrete–time filter design on a continuous–time problem formulation. Let a continuous-time scalar signal $s(t)$ be characterized as a linear stochastic process

$$s(t) = G(p)e_c(t) \tag{71}$$

where $e_c(t)$ is zero mean white noise, with spectral density $\lambda_c/2\pi$. The argument t denotes continuous time. and $G(p)$ is a rational function of the derivative operator $p \triangleq d/dt$.

$$G(p) = \frac{b_o p^{\delta-n-1} + b_1 p^{\delta-n-2} + \ldots + b_{\delta-n-1}}{p^\delta + a_1 p^{\delta-1} + \ldots + a_\delta} \ . \tag{72}$$

The transfer function has order $\delta \geq n+1$ and pole excess (relative degree) $\geq n+1$. Here, we think of the expression (71) as a model describing the spectral properties of the signal. We assume λ_c and $G(p)$ to be time–invariant. The signal $s(t)$ is sampled with sampling period h.

The objective is to estimate the n'th order derivative of the signal $s(t)$

$$\begin{aligned} f(t) &\triangleq \frac{d^n s(t)}{dt^n} = p^n G(p) e_c(t) \\ &= \frac{b_o p^{\delta-1} + b_1 p^{\delta-2} + \ldots + b_{\delta-n-1} p^n}{p^\delta + a_1 p^{\delta-1} + \ldots + a_\delta} e_c(t) \end{aligned} \quad (73)$$

at the time instants $t = kh$; $t = 0, 1, \ldots$.

Let us outline one solution, which is derived and discussed in more detail in [15]. The stochastic model (71)–(73) can be represented in state space form, denoting the state vector $u(t)$, as

$$\begin{aligned} du(t) &= \mathbf{A}u(t)dt + \mathbf{B}dW(t) \\ s(t) &= \mathbf{H}_1 u(t) \\ f(t) &= \mathbf{H}_2 u(t) \ . \end{aligned} \quad (74)$$

Here, $dW(t) \doteq e_c(t)dt$ represents Wiener increments. A convenient representation of (74) is

$$\mathbf{A} = \begin{pmatrix} -a_1 & \ldots & \ldots & -a_\delta \\ 1 & & & \\ & \ddots & & \\ 0 & & 1 & 0 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{aligned} \mathbf{H}_1 &= (\underbrace{0 \ldots 0}_{\leftarrow n \rightarrow} \quad b_o \ldots b_{\delta-n-1}) \\ \mathbf{H}_2 &= (b_o \ldots b_{\delta-n-1} \quad \underbrace{0 \ldots 0}_{\leftarrow n \rightarrow}) \end{aligned} \quad (75)$$

Stochastic sampling of (74), results in the discrete-time representation

$$\begin{aligned} u(k+1) &= \mathbf{F}u(k) + e_v(k) \\ s(k) &= \mathbf{H}_1 u(k) \\ f(k) &\triangleq \left. \frac{d^n s(t)}{dt^n} \right|_{t=kh} = \mathbf{H}_2 u(k) \end{aligned} \quad (76)$$

where $\mathbf{F} = e^{\mathbf{A}h}$, see e.g. [7]. Note that $f(k)$ is *exactly* the derivative at the sampling instants $t = kh$. We assume the system to have poles in $|z| \leq 1$, and the pair $(\mathbf{F}, \mathbf{H}_1)$ to be detectable. (Possible unobservable modes must be stable.) The column vector $e_v(k)$ consists of discrete-time stationary white noise elements with zero mean. The covariance matrix equals

$$Ee_v(k)e_v(k)^T = \lambda_c \int_0^h \exp(\mathbf{A}\tau)\mathbf{B}\mathbf{B}^T \exp(\mathbf{A}^T\tau)d\tau \triangleq \lambda_c \mathbf{R}_e \ . \quad (77)$$

Note that while the continuous-time noise process $e_c(t)$ is scalar, $e_v(k)$ will be a column vector of dimension $s = \dim \mathbf{A}$. In general, $\mathbf{R}_e$ has full rank. The effect of all components of $e_v(k)$ on $f(k) = \mathbf{H}_2 u(k)$ can *not*, in general, be calculated exactly from their effect on $s(k) = \mathbf{H}_1 u(k)$, *unless* the covariance matrix $\mathbf{R}_e$ has rank 1. When the sampling frequency increases, $\mathbf{R}_e$ approaches a rank 1-matrix.

Measurements of the signal $s(k)$ are assumed to be corrupted by a discrete-time noise $w(k)$, described below by a discrete–time ARMA model

$$y(k) = s(k) + w(k) \ . \quad (78)$$

In order to fit this problem into the parametrization (9), we will convert the state space model (76) into a transfer-function based model. For this reason, introduce the characteristic polynomial $D(q^{-1})$, of degree nd equal to the number of states $s = \dim\mathbf{A}$, and the polynomial matrix $C(q^{-1})$ as

$$D(q^{-1}) \triangleq \det(I - q^{-1}\mathbf{F}) \ ; \ C(q^{-1}) \triangleq \mathrm{adj}(I - q^{-1}\mathbf{F})q^{-1} \ . \quad (79)$$

Note the distinction between $C(q^{-1})$, a polynomial matrix, and the constant matrix $\mathbf{F}$. Hence, the sampled system can be expressed as

$$\begin{aligned} u(k) &= \frac{1}{D(q^{-1})} C(q^{-1}) e_v(k) & Ee_v(k)e_v(k)^T &= \lambda_c \mathbf{R}_e \\ w(k) &= \frac{M(q^{-1})}{N(q^{-1})} v(k) & Ev(k)^2 &= \lambda_v \end{aligned} \quad (80)$$

$$y(k) = \mathbf{H}_1 u(k) + w(k) \ ; \ f(k) = \mathbf{H}_2 u(k) \ .$$

Assume the parameters of the continuous-time model (71)–(73), and those of the noise description, to be known a priori or correctly estimated in some way. The discrete-time model (80) is then obtained by stochastic sampling. We seek the stable time-invariant linear estimator of the n'th order derivative

$$\hat{f}(k|k+m) = \frac{Q^c(q^{-1})}{R^c(q^{-1})} y(k+m) \qquad (81)$$

which minimizes the mean square estimation error $\mathrm{E}|\varepsilon(k)|^2 = E|f(k) - \hat{f}(k|k+m)|^2$.

This corresponds to the choices $\mathcal{G} = \mathbf{H}_1$, $\mathcal{G}_a = 0$ ($\Rightarrow \mathcal{R}_a = 0$), $\mathcal{F} = C/D$, $\mathcal{H} = M/N$, $\mathcal{D} = \mathbf{H}_2$, $\mathcal{W} = 1$ and $\mathcal{R}_z = [Q^c/R^c \;\; 0]$ in (6), (9) and (10). See Figure 10.

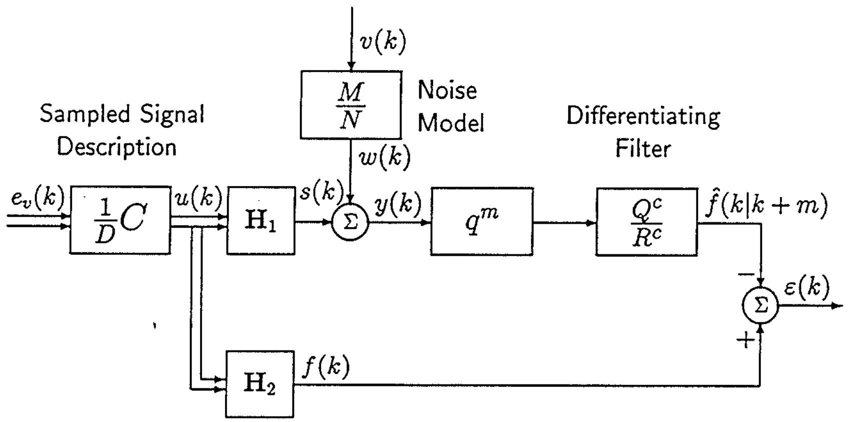

Figure 10: A state estimation problem, which here represents a differentiation problem based on a continuous time model. The state variable $f(k) = u(k)$ is the derivative of $s(k)$. It is to be estimated from measurements $y(k+m)$.

Based on the model (80), introduce the following double–sided polynomials, $P_{ij}(q, q^{-1})$,

$$P_{ij} = p_{nc}^{ij} q^{nc} + \dots p_o^{ij} + \dots p_{-nc}^{ij} q^{-nc} \triangleq \mathbf{H}_i C(q^{-1}) \mathbf{R}_e C_*(q) \mathbf{H}_j^T \;, \; i,j = 1,2 \;. \qquad (82)$$

Also, with $\eta \triangleq \lambda_v/\lambda_c$, introduce the polynomial spectral factorization

$$\tau\beta\beta_* = P_{11} N N_* + \eta D D_* M M_* \qquad (83)$$

defining the stable and monic spectral factor $\beta(q^{-1}) = 1 + \beta_1 q^{-1} + \dots + \beta_{n\beta} q^{-n\beta}$ of degree $n\beta = \max\{nc + nn, nd + nm\}$ and a scalar τ. As mentioned before, a stable spectral factor β exists, if and only if the two terms on the right hand side of (83) have no common factors with zeros on the unit circle. (If $\eta = 0$, the first term should have no zeros on the unit circle.)

The polynomials P_{ij} and β have specific interpretations. Note, from (76), (80) and (82), that for stationary signals (stable D and N), the spectral densities of $\{s(k)\}$ and $\{f(k)\}$ are given by

$$\Phi_s(\omega) = \frac{\lambda_c}{2\pi}\frac{P_{11}}{DD_*} \qquad \Phi_f(\omega) = \frac{\lambda_c}{2\pi}\frac{P_{22}}{DD_*} \qquad \Phi_{fs}(\omega) = \frac{\lambda_c}{2\pi}\frac{P_{21}}{DD_*} \qquad (84)$$

where $e^{-i\omega h}$ and $e^{i\omega h}$ have been substituted for q^{-1} and q in all polynomials.

If D and N are stable, and with $e_v(k)$ and $v(k)$ uncorrelated, the spectral density of the measurement sequence $\{y(k)\}$ is given by

$$\begin{aligned} \Phi_y(\omega) &= \Phi_s(\omega) + \Phi_w(\omega) = \frac{\lambda_c}{2\pi}\frac{P_{11}}{DD_*} + \frac{\lambda_v}{2\pi}\frac{MM_*}{NN_*} \\ &= \frac{\lambda_c}{2\pi}\frac{\tau\beta\beta_*}{DD_*NN_*} \;. \end{aligned} \qquad (85)$$

As usual, the spectral factor β thus represents the numerator of an innovations model $y(k) = (\beta/DN)\epsilon(k)$.

Theorem 1

Consider the sampled signal model described by (80), with all zeros of D and N in $|z| \le 1$. Assume that a stable spectral factor β, defined by (83), exists. A stable linear estimator (81) of the derivative then attains the minimum of the criterium $E|f(k) - \hat{f}(k|k+m)|^2$ if and only if it has the same coprime factors as

$$\frac{Q^c}{R^c} = \frac{Q_1^c N}{\beta} \;. \qquad (86)$$

Here $Q_1^c(q^{-1})$, together with a polynomial $L_*^c(q)$, is the unique solution to the linear polynomial equation

$$q^{-m} P_{21} N_* = \tau\beta_* Q_1^c + qDL_*^c \qquad (87)$$

with polynomial degrees

$$\begin{aligned} nQ_1^c &= \max\{nc+m, nd-1\} \\ nL^c &= \max\{nc+nn-m, n\beta\}-1 \ . \end{aligned} \tag{88}$$

The minimal variance of the estimation error is given by

$$E|\varepsilon(k)|^2_{\min} = \frac{\lambda_c}{2\pi i}\oint\{\underbrace{\frac{L^cL^c_*}{\tau\beta\beta_*}}_{I}+\eta\underbrace{\frac{MM_*P_{22}}{\tau\beta\beta_*}}_{II}+\underbrace{\frac{NN_*[P_{11}P_{22}-P_{12}P_{21}]}{\tau\beta\beta_*DD_*}}_{III}\}\frac{dz}{z} \tag{89}$$

□

Proof: See [15].

This solution considers estimation of one state variable only: the derivative of order n of the signal $s(k)$. It is straightforward to estimate several state variables, or even all of them, with different smoothing lags for each one. If $f(k)$ is a vector, the estimation of component i of $f(k)$ does not affect the estimation of component j. The total estimator of $f(k)$ can then be obtained as ℓ *parallel* scalar estimators. We thus obtain a set of independent Diophantine equations of type (87), one for each estimated state variable. The scalar spectral factorization remains unaltered. For details, see [57].

This way of expressing a state estimator can be seen as an alternative way of computing a stationary Kalman filter/predictor/smoother, for systems with scalar measurements $y(k)$. (For systems with multiple measurements, a matrix spectral factorization would be required.)

If the characteristic polynomial D is marginally stable, both $f(k)$ and the estimate $\hat{f}(k)$ will, in general, be nonstationary sequences. The estimation error $\varepsilon(k) = f(k) - \hat{f}(k)$ will, however, be a stationary zero mean sequence, with a finite minimal variance given by (89). (This implies that marginally stable factors of D, in the denominator of term III in (89), are cancelled by numerator factors.)

The three terms of the minimal cost (89) can be interpreted as follows.

Term I represents the effect of a *finite smoothing lag m*. As is shown in [15], $L^c \to 0$ when $m \to \infty$. Term 1 then vanishes.

Term II depends on the noise $w(k)$. It represents the unavoidable performance degradation due to noise, which cannot be eliminated, even with an arbitrarily large smoothing lag m. The term vanishes in the noise-free case ($\eta = 0$).

Term III remains even when $m \to \infty$ and $\eta = 0$. It represents aliasing effects. Asymptotically, when $h \to 0$ and the covariance matrix $\lambda_c R_e$ (defined by (77)) approaches a rank 1-matrix, the term vanishes. See [15].

The differentiation problem can also be posed in a discrete time setting, without assuming an explicit underlying continuous–time model. The problem then becomes a scalar variant of the general problem of Section IV, with $u(k)$ being the signal of interest and $\boldsymbol{S}/T$, cf Figure 7, representing a discrete–time approximation of the derivative operator $(i\omega)^n$. Based on such an approximation, which can be designed by well–known means [52], [54], one can construct an estimator which optimally takes noise and transducer dynamics into account. See [15], [16]. The use of a discrete–time approximation of the derivative operator mostly results in an additional perfomance loss, compared to the formulation outlined above. However, if the continuous time system is known, this loss can be eliminated by using an optimal derivative approximation, presented in [15].

VI DECISION FEEDBACK EQUALIZERS

We finally turn our interest to an important problem in digital communications, and present a polynomial solution derived in [23]. When digital data are transmitted over a communication channel, intersymbol interference and noise prevent a receiver from always detecting the symbols correctly. Consider a received sampled data sequence $y(k)$. It is described as a sum of channel output $s(k)$ and noise $w(k)$ by the following linear stochastic discrete time model

$$y(k) = s(k) + w(k) = q^{-d}\frac{B(q^{-1})}{A(q^{-1})}u(k) + \frac{M(q^{-1})}{N(q^{-1})}v(k) \ . \tag{90}$$

The first right–hand term of (90) represents a dispersive linear communication channel, with $\{u(k)\}$ being the transmitted data sequence. The channel model includes pulse shaping, receiver filter and a transmission delay of d samples. Baseband operation on a complex channel is assumed. The second term describes a coloured noise, where the colour may be caused, for example, by effects of receiver filters or leakage from other channels. The sequence $\{v(k)\}$ is a discrete-time white noise. It is zero mean, stationary and independent of $u(k)$.

The polynomials in (90), having degrees δa, δb etc, are assumed known a priori or correctly estimated. Except for the $B(q^{-1})$-polynomial, which has an arbitrary nonzero leading coefficient b_o, all polynomials are monic. It is realistic to assume $A(q^{-1})$ and $M(q^{-1})$ to be stable polynomials, while $B(q^{-1})$ can have zeros anywhere and $N(q^{-1})$ may have zeros in $|z| \leq 1$.

The sequence $\{u(k)\}$ is here assumed to be white. It may be real or complex. One example is the use of p-ary symmetric Pulse Amplitude Modulated (PAM) signals. Then, $u(k)$ 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 model coefficients and signals in (90) are complex-valued. Now, define

$$\lambda_u \triangleq E|u(k)|^2 \quad ; \quad \rho \triangleq E|v(k)|^2/E|u(k)|^2 \ . \tag{91}$$

The data sequence $u(k)$ is to be reconstructed from measurements of $y(k)$. As has been mentioned in Section IV, this can be accomplished by a linear equalizer. Superior performance is, however, achieved with a Decision Feedback Equalizer (DFE) for moderate and high signal to noise ratios. ADFE is a nonlinear filter, which involves a decision circuit. Decisioned data are fed back through a linear filter to improve the estimate. See e.g. [58]–[62], and the references therein. The bit error rate of a DFE is in many cases several orders of magnitude lower than for a linear equalizer.

Previously available design principles for the linear filters of a DFE have either provided optimal filters that are not realizable, or realizable filters with a suboptimal transversal (FIR) structure. Here, we will introduce a *general IIR decision feedback equalizer* (GDFE), see Figure 11,

$$\hat{u}(k|k+m) = \underbrace{\frac{S(q^{-1})}{R(q^{-1})}}_{\text{forward filter}} y(k+m) - \underbrace{\frac{Q(q^{-1})}{P(q^{-1})}}_{\text{feedback filter}} \tilde{u}(k-1) \ . \tag{92}$$

Above, m is the number of lags (smoothing lag) and $\tilde{u}(k-1)$ is decisioned data, for example sign$(\hat{u}(k-1))$, when PAM is used with $p=2$. The denominator polynomials $R(q^{-1})$ and $P(q^{-1})$ are assumed to be monic, and required to be stable. The sampling rate is assumed to equal the symbol rate. Note that the coefficients of the filters may be complex.

Given a received sequence $y(k)$, a model (90), (91) and a smoothing lag $m \geq d$, the problem is to find polynomials (S, R, Q, P) which minimize the

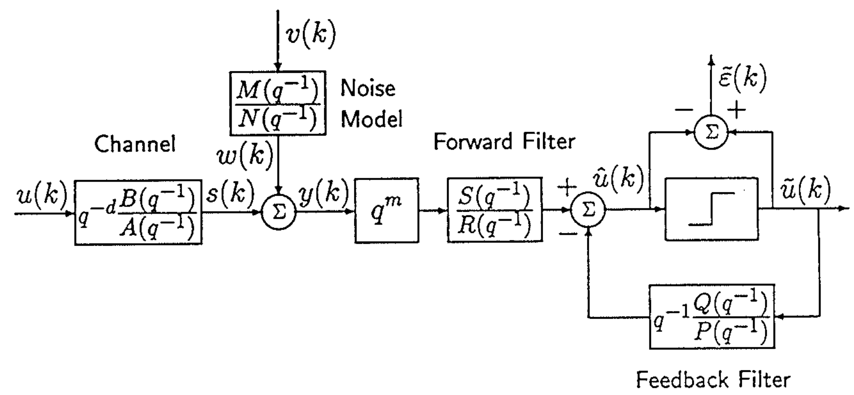

Figure 11: The decision feedback equalization problem. The estimate $\hat{u}(k|k+m)$ is obtained by subtracting old decisioned data $\tilde{u}(k-1)$, fed back through a filter. The estimate $\hat{u}(k|k+m)$ is fed into a decision device to recover the transmitted sequence.

MSE criterion $E|\varepsilon(k)|^2 = E|u(k) - \hat{u}(k|k+m)|^2$. [13] Because of the presence of a nonlinear decision circuit, it is impossible to obtain closed-form expressions for an optimal estimator. As in most previous treatments of the DFE–problem, we will simplify the problem by assuming *correct past decisions.*

If previous decisions are correct, they can be used to completely eliminate the interference, caused by past symbols, at the current received signal. In contrast to linear equalizers, this can be achieved without any noise amplification. This is more easily understood if Figure 11 is redrawn as in Figure 12. No inversion has to be done. Instead, a feedforward from $u(k-1)$ is used. The nonlinearity is now outside the signal path from $u(k)$ and $v(k)$ to $\varepsilon(k)$. Thus, by assuming correct past decisions we can transform the problem into an LQ– optimization problem[14].

[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 [60] 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 [63].

[14] For low signal to noise ratios, the assumption of correct past decisions is not appropriate. Because of the high noise level, incorrect decisions will occur. They may even start a burst of errors. This phenomenon is known as *"error propagation"*. If too many error bursts occur, they will deteriorate the performance considerably and could in fact make the equalizer useless. For a discussion, see [59] and [23].

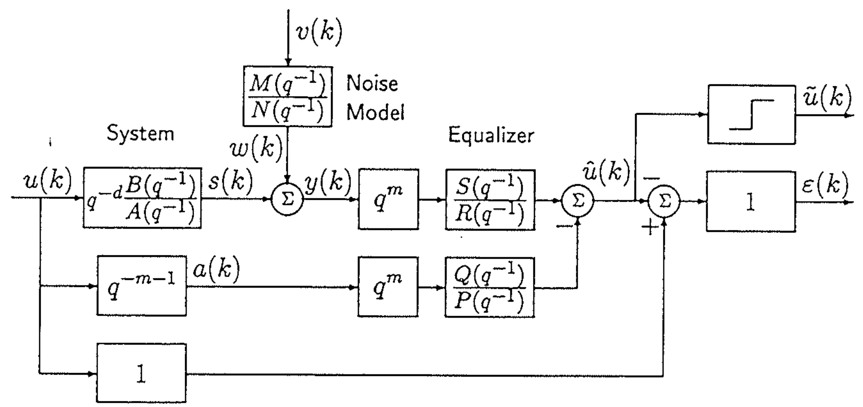

Figure 12: The decision feedback equalization problem, when correct past decisions are assumed. The signal $u(k)$ is to be estimated from measurements $z(k+m) = (y(k+m)\ u(k-1))^T$.

This problem formulation corresponds to the choices $\mathcal{G} = q^{-d}B/A$, $\mathcal{G}_a = q^{-m-1}$, $\mathcal{F} = \mathcal{D} = 1$, $\mathcal{H} = M/N$, $\mathcal{W} = 1$, and $\mathcal{R}_z = [S/R \quad -Q/P]$ in (6), (9) and (10).

Introduce the following polynomials:

$$\begin{aligned}
\tau(q^{-1}) &\triangleq BN = \tau_o + \tau_1 q^{-1} + \ldots + \tau_{\delta\tau} q^{-\delta\tau} \\
\gamma(q^{-1}) &\triangleq AM = 1 + \gamma_1 q^{-1} + \ldots + \gamma_{\delta\gamma} q^{-\delta\gamma} \\
\alpha(q^{-1}) &\triangleq \gamma + q^{-1}Q_1 = 1 + \alpha_1 q^{-1} + \ldots + \alpha_{\delta\alpha} q^{-\delta\alpha} \ .
\end{aligned} \tag{93}$$

We are now able to state the following result.

Theorem 2

Assume the received data to be accurately described by (90),(91). The general DFE (92) then attains the global minimum of $J = E|u(k) - \hat{u}(k|k+m)|^2$, if and only if the filters S/R and Q/P have the same coprime factors as

$$\frac{S}{R} = \frac{S_1 N}{M} \qquad \frac{Q}{P} = \frac{Q_1}{AM} \tag{94}$$

where $S_1(q^{-1})$ and $Q_1(q^{-1})$, together with polynomials $L_{1*}(q)$ and $L_{2*}(q)$, satisfy the two coupled polynomial equations

$$q^{m-d}\tau S_1 + \gamma L_{1*} = \alpha = \gamma + q^{-1}Q_1 \tag{95a}$$
$$-\rho\gamma_* S_1 + q^{-m+d}\tau_* L_{1*} = qL_{2*} \tag{95b}$$

with polynomial degrees

$$\begin{aligned}
\delta S_1 &= \delta L_1 = m - d \\
\delta Q_1 &= \delta L_2 = \max(\delta\gamma, \delta\tau) - 1 \ .
\end{aligned} \tag{96}$$

The minimal mean square estimation error is

$$E|\varepsilon(k)|^2_{\min} = \frac{\lambda_u}{2\pi j}\oint_{|z|=1} L_1 L_{1*} + \rho S_1 S_{1*} \frac{dz}{z} = \lambda_u \Big(\sum_{j=0}^{m-d} |\ell_j|^2 + \rho |s_j|^2\Big) \tag{97}$$

□

Proof: See [23], where optimality is verified using a non-constructive variant of the variational approach. A variation $\nu(k) = \mathcal{T}_1 y(k+m) + \mathcal{T}_2 u(k-1)$ with two terms is introduced. Orthogonality with respect to each of these terms is verified. The Diophantine equations arise from the two orthogonality requirements.

Remark: Note that (95a) and (95b) represent two coupled polynomial equations, containing *four* unknown polynomials ($S_1(q^{-1})$, $Q_1(q^{-1})$, $L_{1*}(q)$, $L_{2*}(q)$). Also, note that *no spectral factorization is required.* The solution obtained here is, of course, a special case of the one obtained in Section IV. It is, however, difficult to derive it from that solution. Instead, by formulating a scalar problem which utilizes both $y(k)$ and $a(k)$ of Figure 3, the solution is readily obtained.

An explicit solution to (95a) and (95b) is given by the following result.

Theorem 3

The polynomials S_1, $\overline{L}_1$ and Q_1, calculated in the following way, provide the unique solution to the polynomial equations (95a) and (95b).

1. Solve for the coefficients of the polynomials $S_1(q^{-1})$ and $\overline{L}_1(q^{-1})$ in

$$\left[\begin{array}{cccc|cccc} \tau_o & & & 0 & 1 & & & 0 \\ \vdots & \ddots & & & \gamma_1 & \ddots & & \\ \tau_{m-d} & \cdots & & \tau_o & \gamma_{m-d} & \cdots & \gamma_1 & 1 \\ \rho & \rho\gamma_1^* & \cdots & \rho\gamma_{m-d}^* & -\tau_0^* & & \cdots & -\tau_{m-d}^* \\ & \ddots & & \rho\gamma_1^* & & & \ddots & \vdots \\ 0 & & & \rho & 0 & & & -\tau_0^* \end{array}\right] \left[\begin{array}{c} s_o \\ \vdots \\ s_{m-d} \\ l_{m-d}^* \\ \vdots \\ l_0^* \end{array}\right] = \left[\begin{array}{c} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{array}\right] \quad (98)$$

2. With $\{s_i\}$ and $\{\ell_j^*\}$ obtained from step 1, perform the multiplication

$$\left[\begin{array}{ccc|ccc} \tau_o & & 0 & 1 & & 0 \\ \vdots & \ddots & & \gamma_1 & \ddots & \\ \vdots & & \tau_o & \vdots & & 1 \\ \tau_{\delta\tau} & & \vdots & \gamma_{\delta\gamma} & & \gamma_1 \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \tau_{\delta\tau} & 0 & & \gamma_{\delta\gamma} \end{array}\right] \left[\begin{array}{c} s_o \\ \vdots \\ s_{m-d} \\ l_{m-d}^* \\ \vdots \\ l_o^* \end{array}\right] = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ 1 \\ \alpha_1 \\ \vdots \\ \alpha_{\delta\alpha} \end{array}\right] \quad (99)$$

yielding the coefficients of the polynomial $\alpha(q^{-1})$.

3. Finally, calculate the polynomial $Q_1(q^{-1})$ from (93)

$$Q_1(q^{-1}) = q(\alpha(q^{-1}) - \gamma(q^{-1})) \ . \quad (100)$$

□

The equivalent equalized channel (from $u(k+m)$ to $\hat{u}(k|k+m)$) will be

$$C_{eq} = q^{-d}\frac{BNS_1}{AM} - q^{-m-1}\frac{Q_1}{AM} = q^{-m} - q^{-d}\overline{L}_1(q^{-1}) \quad (101)$$

Equation (98) is obtained in the following way. Write (95a) and (95b) in matrix form. Select all rows with *known* right hand sides and combine them into a new system of equations, in the coefficients of S_1 and L_{1*} only. For details, see Appendix B of [23]. By further substitutions, a linear system for determining S_1 of only half the size of (98) is derived in [23]. (Observe that the polynomial α is defined monic in (93). With the leading coefficient of α fixed and nonzero, we avoid the trivial solution $S_1 = L_{1*} = 0$.)

The matrix blocks in (98) are quadratic. If $\tau(q^{-1})$ or $\gamma(q^{-1})$ are of order $< m-d$, zeros are used to fill up the corners of the blocks. The second step of Theorem 3 represents calculation of α from equation (95a), with known S_1 and $\overline{L}_1$: $\tau S_1 + \gamma\overline{L}_1 = q^{-m+d}\alpha$.

An important question is if a unique solution to (98) can always be found without any restrictions on, for example, the coprimeness of $\tau(q^{-1})$ and $\gamma(q^{-1})$.

Theorem 4

If the noise variance ratio ρ and the leading coefficient of B, b_o, are not both zero, then (98) will always have a unique solution, $(S_1, \overline{L}_1)$. □

Proof: See [23].

Remark: When both $|b_o|$ ($= |\tau_o|$) and ρ are small, the system (98) may be badly conditioned.

Summing up, one can conclude that an equalizer can be calculated using (94) and (98)-(100) (Theorem 3). This procedure works under very general conditions (Theorem 4). The resulting equalizer is MSE-optimal (Theorem2). The minimal criterion value is given by (97), assuming correct past decisions. The properties of the optimal DFE are emphasized in some more detail below:

1. It is efficient to whiten the noise. The forward filter S/R contains the inverse noise description in cascade with a transversal filter S_1 of order $m-d$. After noise inversion, we have to equalize a channel $q^{-d}\tau/\gamma = q^{-d}BN/AM$. Therefore, the polynomials S_1, Q_1 and P are determined exclusively by the polynomials τ and γ, *not* by their separate factors A, B, M and N.

2. A conventional DFE-structure (transversal filters both in the forward and feedback parts of (92)) is optimal if and only if $M = 1$ and $A = 1$. In other words, the channel must be adequately described by a transversal filter, and the noise statistics by an autoregressive process.

3. Theorem 2 provides us with an optimal *filter structure* and optimal *polynomial degrees*. Hence, unnecessary overparametrization is avoided. It also gives guidelines about how to choose filter degrees in a conventional DFE. The number of smoothing lags "m" is a user

choice. It may often be chosen rather small (yet $m \geq d$). Usually, m should be chosen larger than the channel bulk delay, so that the major part of the received impulse energy, caused by $u(k)$, can be used by the filter at time $k + m$.

4. In the minimal criterion value (97), the second term $\rho S_1 S_{1*}$ represents noise transmission. The first term $L_1 L_{1*}$ is caused by residual intersymbol interference from the first $m - d$ taps of the equalized channel $(\lambda_u \sum_{j=1}^{m-d} |\ell_j|^2)$. It is also caused by the deviation of the reference tap (at time index $m - d$) from1 $(\lambda_u |\ell_o|^2)$. See (101). We thus get a nice interpretation of one of the extra "dummy"–polynomials. As in all DFE's, the equalized channel impulse response beyond time index $m - d$ is cancelled completely by the feedback filter. See (101). Thus, past symbols do not affect the present decision.

5. In the noise-free case ($\rho = 0$), $L_{1*} = L_{2*} = 0$, see (95b). For any $m \geq d$, this gives $|\varepsilon(k)|^2 = 0$, *even when B is unstable*, and the channel has no stable inverse. The reason for this remarkable property is that, instead of inverting the channel, the estimator uses feedforward from $u(k-1)$. See Figure 12.

6. The denominator polynomials R and P are stable by construction, since A and M are stable. In adaptive algorithms, stability of the estimates $\hat{A}$ and $\hat{M}$, or of $\hat{R}$ and $\hat{P}$, would be required.

The algorithm presented above has been used in an adaptive equalizer designed for the North–American digital mobile radio standard. See [64]. Combined with a novel and efficient channel estimator, it has achieved very good performance. Ageneralization of the above result, which takes model uncertainty and decision errors into account, is presented in [65].

VII CONCLUDING DISCUSSION

Why and when should a filter designer use the polynomial approach? What advantages does it offer from an engineering point of view, compared to e.g. a state space approach [66] or Wiener design of FIR filters [67]. Some answeres to these questions are given next.

• Many properties of the resulting filter can be disclosed by inspection only. See, for example, the remarks to the solutions obtained for the problems in Sections IV–VI. Such information is hard to obtain from a corresponding state space approach. The obtained filters can also be examined directly

using classical concepts, such as frequency response, poles and zeros.

• The solution is often explicit, in terms of the model polynomials. (Note, for example, the presence of the noise model denominator N as numerator factor of the filters (20), (60), and (86).) This will not only help a designer to gain engineering insight, but also to build in design requirements. An example is the suggestion in Appendix B to use integrating signal models to avoid bias for non–zero mean signals. The minimal criterion value can often be interpreted in terms of effects of different design constraints. For example, in the differentiation problem of Section V, performance of the estimator is limited by the effects of aliasing, noise and finite smoothing lag. A designer will not only be able to calculate the limits of performance, but also to understand them.

• If an incorrect filter structure, with insufficient degrees of freedom is assumed, a solution will not exist. In the polynomial derivation techniques, the warning signals for this are inconsistencies or degenerated polynomial degrees. A polynomial solution thus leads to the optimal structure and degrees of filters and their polynomials. In contrast to the Wiener design of FIR filters [67], unnecessary overparametrization is avoided. This is of considerable importance, if the solution is to be used in an indirect adaptive algorithm.

• In contrast to Kalman–based methods, fixed lag smoothing does not complicate the solution or decrease insight, nor do singular situations (where white noise is not present in all measurements).

• The simplicity of the polynomial approach is rather evident for scalar filtering problems. Low order filters can even be designed by hand calculation. See the design example presented in Section III,I. However, any conceptual advantage, compared to Kalman filtering, has been rather questionable in cases of multi–signal estimation. When multivariable filter design is based on general MFD models, polynomial design equations tend to become rather unattractive and complicated. One reason is the, in some problems, unavoidable introduction of coprime factorizations: (polynomial) matrices do, in general, not commute. This complication can be avoided if the the filter design is based on signal models in common denominator form, and MFD's with diagonal denominators. See Section IV. This solution gives as much structural insight as in the scalar case. It provides much more insight than a Kalman filter based solution, and has about the same algorithmic complexity.

Of course, the polynomial approach has limitations as well as strengths. Compared to Kalman filtering, polynomial methods seem less well suited to some off–line problems such as e.g. fixed interval smoothing [66]. For linear time–varying systems, which generate non–stationary signals, Kalman estimation is without rival. Apart from some work by Rissanen [69],[70],[71] and Grimble [24], very little work has been done on polynomial and input–output approaches to the design of time–varying estimators. It is not clear at present if they can offer a useful alternative viewpoint.

It is well known that the zeros of polynomials of high order are sensitive to variations in the coefficients. Therfore, solutions based on the polynomial approach will often have inferior numerical properties, as compared to a corresponding state space approach, in particular for high order problems. There exist algorithms for solutions of Riccati–equations that are very well–behaved numerically [45]. Therefore, we suggest that for high order problems, a designer uses the polynomial approach in order to derive optimal filters and to gain engineering insight, but uses a state space approach for performing spectral factorizations.

Performance robustness is another important issue related to the discussed approach, as well as to any other filter design method. How well does a designed filter perform under non–ideal conditions and in presence of modelling errors? The performance of the estimators designed in this chapter can be sensitive to model errors, in particular if the filters have poles or zeros close to the unit circle.

Such issues have led us to investigate the design of robust filters. In this area we have, again, found polynomial equations to be of great utility. Anew approach to robust estimation of signals and prediction of time–series is presented in [57], [65], [68] and [72]. Possible modelling errors are described by sets of models, parametrized by random variables, with known covariances. A robust design is obtained by minimizing the squared estimation error, averaged both with respect to model errors and noise. A polynomial solution, based on averaged spectral factorizations and averaged Diophantine equations, is derived. The robust estimator is called a *cautious Wiener filter*. It turns out to be no more complicated to design than an ordinary Wiener filter. Whenever an ordinary (nominal) design is sensitive, the robust design radically reduces the sensitivity. The methodology can be applied to any open loop filtering or control problem.

In [57], robust designs are obtained for the filter discussed in Section III, as well as for scalar deconvolution estimators, feedforward controllers, and the

state estimator discussed in Section V. In [72], the general multivariable estimator of Section IV has been robustified. (This turns out to require only trivial additional computations.) In [65], robustness with respect to model uncertainty and decision errors is built into the decision feedback equalizer design methodology outlined in Section VI. Combination with robustification with respect to noise "outliers" and sampling timing jitter is also of interest.

In the years to come, we believe that the polynomial approach will play a key role in adaptive robust filtering. In such schems, an adaptive algorithm estimates not only the system and noise models, but also a measure of their uncertainties. Filters are adjusted to take the time–varying properties *and* their uncertainty into account.

APPENDIX A: SCALAR DIOPHANTINE EQUATIONS

While Diophantine equations in general have an infinite number of solutions, equations arising from linear quadratic design problems mostly have a unique solution. This is a consequence of two requirements, which are easily seen for the scalar Diophantine equations obtained in this chapter, see e.g. (20):

1. Filter causality requires Q_1 to be a polynomial only in q^{-1}.
2. Optimality restricts L_* to be a polynomial only in q.

For equations with these properties, the following result can be established.

Theorem A1

Consider the scalar Diophantine equation

$$C(q,q^{-1}) = A(q,q^{-1})X(q^{-1}) + B(q,q^{-1})Y(q) \qquad (102)$$

where

$$\begin{aligned} C(q,q^{-1}) &\triangleq c_{nc1}q^{nc1} + \ldots + c_0 + \ldots + c_{-nc2}q^{-nc2} \\ A(q,q^{-1}) &\triangleq a_{na1}q^{na1} + \ldots + a_0 + \ldots + a_{-na2}q^{-na2} \neq 0 \\ B(q,q^{-1}) &\triangleq b_{nb1}q^{nb1} + \ldots + b_0 + \ldots + b_{-nb2}q^{-nb2} \neq 0 \ . \end{aligned}$$

Let d be the the number of linearly dependent equations in the corresponding system of linear equations. Then, (102) has a *unique* solution

$$X(q^{-1}) = x_o + x_1q^{-1} + \ldots + x_{nx}q^{-nx} \ ; \ Y(q) = y_o + y_1q + \ldots + y_{ny}q^{ny}$$

with degrees

$$nx = \max\{nc2, nb2\} - na2 \ ; \ ny = \max\{nc1, na1\} - nb1 \qquad (103)$$

if and only if common factors of A and B are also factors of C and

$$nb1 + na2 - d = 1 \ . \qquad (104)$$

□

Proof. Let g_1 and g_2 denote the highest powers of q and q^{-1} respectively, present anywhere in (102). The degree of $Y(q)$ must be selected such that the highest power of q in $B(q,q^{-1})Y(q)$ equals the highest power of q in any of the two other terms. (To increase ny above this value would be useless; with no matching terms, the superfluous coefficients would become zero.) Thus,

$$g_1 = nb1 + ny = \max\{na1, nc1\} \ . \qquad (105)$$

For similar reasons, the degree of $X(q^{-1})$, nx, must satisfy

$$g_2 = na2 + nx = \max\{nb2, nc2\} \ . \qquad (106)$$

Equation (102) corresponds to $g_1 + g_2 + 1$ linear simultaneous equations. A solution exists only if factors of A and B are also factors of C. (Write $A = TA_1$ and $B = TB_1$, where T is the greatest common factor of A and B in (102). Since T is a factor of the left–hand side of (102), $T(A_1X + B_1Y)$, it must also be a factor of C.) A unique solution then exists if and only if the number of *linearly independent* equations equals the number of unknowns (coefficients of X and Y):

$$g_1 + g_2 + 1 - d = nx + 1 + ny + 1 \ . \qquad (107)$$

The use of (105) and (106) in (107) gives (104). If the left–hand side of (104) is < 1, the solution is non–unique. If it is > 1, no solution exists. □

The equation (20) fulfills (104). There we have $nb1 = 1$ (because of the free q–factor) and $na2 = 0$. Since β_* (unstable) and D (stable) cannot have common factors, the corresponding system of equations has full rank. Consequently, $d = 0$ in (104). From (103), we obtain the degrees

$$nQ_1 = \max(nc + m, nd - 1), \ nL = \max(nc + nn - m, n\beta) - 1 \ .$$

APPENDIX B: UNSTABLE MODELS

Let us remove the assumption of stability of D and N in the problem described in Section III,B. The complete solution then turns out to include a second Diophantine equation. However, we will argue that the original equation is sufficient in filtering problems of practical interest. Assumption A in Section III,B is now exchanged for

Assumption B. The signal and noise models $s(k) = (C/D)e(k)$ and $w(k) = (M/N)v(k)$ are causal and have no unstable hidden modes. They have no common zeros on the unit circle and no common poles on or outside the unit circle.

The requirement of no common unstable modes corresponds to detectability of a state space model.

For unstable systems, an innovations model (16) can still be defined. It should, more properly, be called a generalized innovations model, with $\beta(q^{-1})$ being a generalized (polynomial) spectral factor [27]. Under Assumption B, the spectral factorization equation (17) will have a unique stable solution.

In the variational approach, the stationarity of the variation $\nu(k) = \mathcal{T}y(k+m)$ has to be guaranteed. (The modified estimation error is $\varepsilon(k) + \nu(k)$. Assuming $\varepsilon(k)$ to be zero mean stationary, we could never obtain a lower MSE by adding to it a nonstationary signal $\nu(k)$, with variance tending to infinity.) Stationarity of $\nu(k)$ is guaranteed by requiring $\mathcal{T}$ to contain all unstable poles as zeros. For example, set $\mathcal{T} = DN\mathcal{T}_1$ with $\mathcal{T}_1$ stable and causal. Then, the factor $(1/D_*N_*)\mathcal{T}_*$ in (19) has poles only outside $|z| = 1$. The rest of the reasoning remains unchanged. The optimal linear estimator still satisfies (20) and (21).

For unstable signal (and noise) models, two different situations are now possible.

Case 1: β_* and D have no common factors. Under Assumption B, the equation (20) remains *uniquely solvable*. Since β_* has zeros only in $|z| > 1$, this holds for *marginally stable* models, where D (or N) has zeros on $|z| = 1$. These are the unstable models of most interest in filtering problems. They are used for describing signals and noise with drifting or sinusoid behaviour. The use of signal models with poles at $z = 1$ is also a trick for avoiding bias, when estimating stationary signals with nonzero mean.

Stationarity of the error $\varepsilon(k)$ in Section III,C is verified in the following way. The use of (21) in (18) gives

$$\varepsilon(k) = \left(1 - q^m \frac{Q_1 N}{\beta}\right) \frac{C}{D} e(k) - q^m \frac{Q_1 N}{\beta} \frac{M}{N} v(k) \quad .$$

Cancellation of N in the last term is assumed to be exact. (If D is stable and N unstable, $\varepsilon(k)$ is therefore stationary.) Let us evaluate the first term at the zeros of D in $|z| \geq 1$, denoted $\{z_j\}$. When (17) and (20) are evaluated at $\{z_j\}$, their most right–hand terms (but no other terms) vanish. Use of this fact gives

$$1 - q^m \frac{Q_1 N}{\beta} \bigg|_{z=z_j} = 1 - q^m \left(q^{-m} \frac{CC_*N_*}{r\beta_*}\right) \frac{N}{\beta} \bigg|_{z=z_j} = 0 \quad .$$

Thus, the transfer function from $e(k)$ to the error $\varepsilon(k)$ remains finite for all $z : |z| \geq 1$, including $\{z_j\}$. Unstable poles are cancelled by zeros. In SISO problems, the reasoning above is straightforward. In multi–signal estimation problems, additional conditions will often have to be imposed, to avoid "impossible" problem formulations, for which no finite minimal criterion value exists.

Note that for signals with nonzero mean, the presence of a zero at $z = 1$ in the transfer function from $s(k)$ to $\varepsilon(k)$ precludes biased estimates. The steady–state value of $\varepsilon(k)$ will be zero, even if $E(s(k)) \neq 0$. The presence of such a zero is assured by including a pole at $z = 1$ in the signal model C/D.

Case 2: β_* and D have common factors. Under Assumption B, those factors must also be factors of the left–hand side of (20) [15]. Thus, the Diophantine equation remains solvable, but the solution becomes *non–unique.* We obtain a linear dependence in the equations, represented by $d > 0$ in (104). Only one of these solutions corresponds to a stationary error [16]. The correct solution is obtained by *requiring* that D is cancelled in the transfer function from $e(k)$ to $\varepsilon(k)$. Thus, we require that

$$\beta - q^m Q_1 N = XD \tag{108}$$

for some polynomial $X(q^{-1})$. This is a *second Diophantine equation.* An alternative variant is obtained by multiplying (108) by $r\beta_*$. This gives

$$r\beta\beta_* = q^m(r\beta_* Q_1)N + r\beta_* XD \quad .$$

The use of (17) and (20), and cancellation of D, gives the equation

$$\rho MM_*D_* = -q^{m+1}L_*N + r\beta_* X \quad . \tag{109}$$

Any one of the equations (108) or (109) can be solved in conjunction with (20). Then, the unique optimal $Q_1(q^{-1})$ is obtained, together with $L_*(q)$ and $X(q^{-1})$.

The need for a second Diophantine equation in certain situations has been emphasized by Kučera [73] for feedback control problems and by Grimble

[15] From the spectral factorization (17), it is evident that factors common to D and $r\beta\beta_*$ must also appear in CC_*NN_*. Since (D,N) and (D,C) are not allowed to have unstable common factors, these factors must be present in C_*N_*.

[16] The demonstration of a finite transfer function utilized in Case 1 cannot be used for the zeros common to β_* and D. Both the spectral factorization (17) and the Diophantine equation (20) vanish *completely* at those zeros.

[74] and Chisci and Mosca [75] for filtering problems.

The additional equation complicates the solution, but it is required only in the exceptional Case 2. In the open loop filtering problems considered here, that situation is furthermore of little practical interest. It corresponds to estimation of exponentially increasing, "exploding", time series.[17] There would be severe problems with variable overflow, except for short data series. Furthermore, the stationarity of the error depends on exact cancellation. Arbitrarily small modelling errors or roundoff errors would ruin the result completely in the long run. When signals are nonstationary, the problem of model errors is furthermore larger than for stationary signals. In a nonlinear world, linear time–invariant models are good (but not perfect) descriptions of time series only around *stationary* operating points. The sensitivity problem is still serious, but more acceptable, in the important case of poles on $|z| = 1$.

For these reasons, estimation problems for strictly unstable models, with a theoretical need for an additional Diophantine equation, is not considered in this chapter.

APPENDIX C: MULTIVARIABLE DIOPHANTINE EQUATIONS.

Unique solvability of (61) is demonstrated as follows. The Diophantine equation will always have solutions, since the invariant polynomials of $P\mathbf{I}_p$ are all stable, while those of $\boldsymbol{\beta}_*$ are all unstable. Thus, there exist no common invariant factors. Let $(\boldsymbol{Q}_0, \boldsymbol{L}_{0*})$ be one solution pair. Every solution to (61) can be expressed as

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

where the polynomial matrix $\boldsymbol{X}(q, q^{-1})$ is undetermined, cf [6]. Since $\boldsymbol{Q}$ is required to be causal, it can have no positive powers of q as arguments, while $\boldsymbol{L}_*$ must contain no negative powers of q, to assure optimality. Thus, $\boldsymbol{X}(q, q^{-1}) = 0$ is the only choice. We conclude that the solution to (61) is unique. The degrees (62) are determined by the requirement that the maximum powers of q^{-1} and q are covered on both sides of (61), cf Appendix A.

[17] This claim does not hold for estimation within a stabilized closed loop. One example is a state estimator used in conjunction with a state feedback, which stabilizes the unstable mode.

APPENDIX D: MATLAB ALGORITHMS FOR FILTER DESIGN

Polynomial spectral factorization.

The following MATLAB code is based on an algorithm for polynomial spectral factorization given in [6]. It has been implemented by T. Söderström. Note that polynomial coefficients are assumed to be real valued.

```
function [Beta,l2]=spefac1(Right)
%
% function [Beta,l2]=spefac1(Right)
%
% Given the symmetric polynomial Right(z,z^-1),
% the sequence Beta(1) .. Beta(n) and the real l2
% is computed such that
%    l2*[z**n + Beta(1)*z**(n-1) + .. + Beta(n)]
%    [z**(-n) + Beta(1)*z**(-n+1) + .. + Beta(n)] = Right
% 'Right' assumed to be positive definite on the unit circle.
%
r = Right((length(Right)+1)/2:length(Right));
if r(1)<max(abs(r)),error('r not pos. definite'),return,end
[n,n2]=size(r);if n<n2,n=n2;r=r';end
a=r'/sqrt(r(1));da=1;k=0;
while da>1e-14,k=k+1;
 for i=1:n
   aa(i,:)=[a(i:n),zeros(1,i-1)]+[zeros(1,i-1),a(1:n+1-i)];
 end
 x=2*(aa\r);a1=(a+x')/2;da=norm(a1-a);a=a1;
 if k==300 ,error('No convergence'),return,end
end
l2=a(1)^2; Beta = a/a(1);
```

Scalar Diophantine equations.

The following program solves a polynomial Diophantine equation, which is a scalar variant of the equation (61) in Section IV. In the special case $S = T = B = A = 1, k = 0$, it constitutes eq. (20). The linear system of equations (see e.g. (51)) is also given.

```
function[Q,Lstar,As,Bs]=polysolve(S,CCstar,B,N,rBeta,DT,m,k)
%
% function[Q,Lstar,As,Bs]=polysolve(S,CCstar,B,N,rBeta,DT,m,k)
%
% Solves the polynomial equation
%      -m+k
%    z     S(CC_*)B_*N_* = rBeta_* Q + z(DT)L_*
%
%  with respect to Q and L_* of degree
%
%   nQ=max(ns+nc+m-k,nd+nt-1), nL=max(nc+nb+nn-m+k,nbta)-1
%
%  The polynomials may have complex coefficients.
%  In the input, CCstar = C(z^-1)*C_*(z)
%                rBeta  = r*Beta(z^-1)
%               m,k can have any sign. (k is transducer delay)
%  On exit, Q and L_* polynomials are delivered, as well as
%   the linear system As*x = Bs.
%  USES: sylv
%
ns=length(S)-1;       nb=length(B)-1;
nc=(length(CCstar)-1)/2;  nd=length(DT)-1;
nn=length(N)-1;       BTA =rBeta ;
nbta=length(BTA)-1;   BN = conv(B,N);

nal=nc+nb+nn-m+k;     x=[nal nbta];
nll=max(x);           y=[ns+nc+m-k,nd-1];
nQ=max(y);

if nQ==-1
  Q=0; return;
end

nrow=nQ+nll+1;              hl=zeros(1,nrow);
ni=ns+nc+nb+nn+nc;          CCBBN=zeros(1,nrow);
```



```
BBN=conj(BN(nb+nn+1:-1:1));    CCBBN=conv(BBN,conv(CCstar,S));

 if nal>=nbta
   BBTA=conj(BTA(nbta+1:-1:1));
   BBTA=[zeros(1,nal-nbta) BBTA];
   s=sylv(DT,nll,BBTA,nQ+1,nrow);
   hl(1:ni+1)=CCBBN;
   th=s\hl';
 else
   BBTA=conj(BTA(nbta+1:-1:1));
   CCBBN=[zeros(1,nbta-nal) CCBBN];
   s=sylv(DT,nll,BBTA,nQ+1,nrow);
   hl(1:ni+nbta-nal+1)=CCBBN;
   th=s\hl';
 end
Q=th(nll+1:nll+nQ+1,1)';
LL=th(1:nll,1)';          Lstar=LL(nll:-1:1);
Bs=hl(nrow:-1:1)';        As=s(nrow:-1:1,nrow:-1:1);


function s = sylv(x,nx,y,ny,nrow)
% Creates Sylvester matrix s for scalar equation.
lx=length(x);
ly=length(y);
s=zeros(nrow,nx+ny);
for j=1:nx , s(j:j+lx-1,j)=x' ; end
for j=1:ny , s(j:j+ly-1,j+nx)=y' ; end
```

Multisignal filter design:

MATLAB .m-files for performing spectral factorization and solution of unilateral and bilateral Diophantine equations can be obtained upon request from the authors. We have implemented a spectral factorization algorithm by Ježek and Kučera [46]. It is restricted to real–valued polynomial matrices. The solution of unilateral Diophantine equation is straightforward, and follows the principles outlined in Section IV,B. A solution to bilateral equation has also been implemented.

It is somewhat inconvenient that MATLAB only works with two–dimensional matrices. Polynomial matrices of degree np

$$P(q^{-1}) = \begin{pmatrix} P^{11}(q^{-1}) & \dots & P^{1\ell}(q^{-1}) \\ \vdots & & \vdots \\ P^{k1}(q^{-1}) & \dots & P^{k\ell}(q^{-1}) \end{pmatrix}$$

are assumed to be represented by (block) matrices

$$\begin{pmatrix} p_o^{11} & \dots & p_{np}^{11} & \dots & p_o^{1\ell} & \dots & p_{np}^{1\ell} \\ & \vdots & & & & \vdots & \\ p_o^{k1} & \dots & p_{np}^{k1} & \dots & p_o^{k\ell} & \dots & p_{np}^{k\ell} \end{pmatrix}$$

while matrices $M_*(q)$ have the orders of indexes reversed. Double–sided matrices $B_*(q,q^{-1})$ are represented as above, with coefficients of each polynomial ordered in decreasing power of q towards the right. As example, the header of an .m–file for spectral factorization is provided below.

```
function [A,na,why,nr]=pmspefac(B,p,niter,lim)
%                                    Kenth Ohrn 18-Jan-93
% Performs left spectral-factorization for a "twosided"
% discrete-time polynomial-matrix B, i.e. it solves
%   A(z^-1)*A(z)' =  B(z,z^-1),  with respect to A(z^-1),
%               where det A     not equal 0 for |z|>=1.
% (See Jezek and Kucera, Automatica, vol 21, 663-669, 1985.)
% B symmetric:      B'(z)=B(z^-1)
% B pos. definite:  B(exp(iw))>0, for real w and -pi<w<=pi.
% The z^0-coefficient-matrix of B has to be positive definite.
%
% Input:  B  : Two-sided-polynomial-matrix to be factorized.
%         p  : The maximal order, in z or z^-1, in any
%               of the elements in B(z,z^-1).
%      niter : Maximal number of iterations (default 10);
%        lim : coefficient error limit for convergece
%               max(|A'(z))A(z^-1)-B|) < lim (default 1e-6)
%
% Output: A  : The spectral-factor matrix.
%         na : The maximal order of any polynomial in A
%         why: The iteration-stop-cause.
%         nr : Number of iterations
```


References

[1] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* The Technology Press and Wiley, New York, 1950.

[2] H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," *Proceedings of the I.R.E.*, vol. 38, pp. 417–425 (1950).

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

[4] J. A. Cadzow, *Foundations of Digital Signal Processing and Data Analysis.* Macmillan, New York, 1987.

[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. AMSE, Journal of Basic Engineering, vol. 82D, pp. 35-45 (1960).

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

[7] K. J. Åström, *Introduction to Stochastic Control Theory.* Academic Press, New York, 1970.

[8] U. Borison, "Self–tuning regulators for a class of multivariable systems," *Automatica*, vol 15, pp. 209–215 (1979).

[9] P. Hagander and B. Wittenmark, "A self–tuning filter for fixed–lag smoothing," *IEEE Transactions on Information Theory*, vol. 23 pp. 377–384 (1977).

[10] T. M. Moir and M.J. Grimble, "Optimal self–tuning filtering, smoothing and prediction for discrete time multivariable processes," *IEEE Transactions on Automatic Control*, vol. 19, pp. 128–137 (1984).

[11] V. Kučera, "Transfer–function solution of the Kalman–Bucy filtering problem," *Kybernetika (Prague)*, vol 14, pp. 110–122 (1978).

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

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

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

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

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

[17] M. J. Grimble, "Polynomial systems approach to optimal linear filtering and prediction," *International Journal of Control*, vol. 41, pp. 1545–1564 (1985).

[18] M. J. Grimble and M. A. Johnsson, *Optimal Control and Stochastic Estimation.* Wiley, Chichester, 1988.

[19] M. J. Grimble and A. ElSayed, "Solution of the H_∞ optimal linear filtering problem for discrete–time systems," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 1092–1104 (1990).

[20] T. J. Moir, "Optimal deconvolution smoother," *IEE proceedings*, vol. 133, pp. 13–18 (1986).

[21] T. J. Moir, "A polynomial approach to optimal and adaptive filtering with application to speech enhancement," *IEEE Transactions on Signal Processing*, vol 39, pp. 1221–1224 (1991).

[22] A. P. Roberts and M. M. Newmann, "Polynomial approach to Wiener filtering," *International Journal of Control*, vol. 47, pp. 681–696 (1988).

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

[24] M. J. Grimble, "Time–varying polynomial systems approach to multichannel optimal linear filtering," *Proc. American Control Conference*, Boston, WA6-10:45, pp. 168-174 (1985).

[25] J. F. Barret and T. J. Moir, "A unified approach to multivariable discrete–time filtering based on the Wiener theory," *Kybernetika (Prague)*, vol. 23, pp 177–197 (1987).

[26] J. Ježek "New algorithm for minimal solution of linear polynomial equations," *Kybernetika (Prague)*, vol. 18, pp 505–516 (1982).

[27] U. Shaked, "A generalized transfer function approach to linear stationary filtering and steady–state optimal control problems," *International Journal of Control*, vol. 29, pp. 741–770 (1976).

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

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

[30] M. Sternad and A. Ahlén, "LQ controller design and self–tuning control," Chapter 3 of *Polynomial Methods in Optimal Contol and Filtering*, ed. K. Hunt, Control Engineering Series, Peter Peregrinus, London, 1993.

[31] A. Ahlén and M. Sternad "Optimal Filtering Problems," Chapter 5 of *Polynomial Methods in Optimal Contol and Filtering*, ed. K. Hunt, Control Engineering Series, Peter Peregrinus, London, 1993.

[32] M. Vidyasagar, *Control Systems Synthesis. A Factorization Approach.* MIT Press, Cambridge, MA, 1985.

[33] B–S. Chen and S–C. Peng, "Optimal deconvolution filter design based on orthogonal principle," *Signal Processing*, vol. 25, pp. 361–372 (1991).

[34] A. Ahlén and M. Sternad, "Filter design via inner–outer factorization: Comments on 'Optimal deconvolution filter design based on orthogonal principle'," to appear in Signal Processing, 1994.

[35] T. Kailath, *Linear Systems.* Prentice–Hall, Englewood Cliffs, New Jersey, 1980.

[36] C.–T. Chen, "On digital Wiener filters," *Proc IEEE*, vol. 64, pp. 1736–1737 (1976).

[37] B. A. Francis, *A Course in H_∞ Control Theory.* Springer–Verlag, Berlin, 1987.

[38] C. J. Demeure and C. T. Mullis, "A Newton–Raphson method for moving average spectral factorization using the Euclid algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol 38, pp 1697–1709 (1990).

[39] V. Peterka, "Predictor–based self–tuning control". *Automatica*, vol 20, pp 39–50 (1984).

[40] P. A. Nelson, H. Hamada and S. J. Elliot, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Transactions on Signal Processing*, vol 40, pp. 1621–1632 (1992).

[41] G. Demoment and R. Reynaud, "Fast minimum variance deconvolution," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 1324–1326 (1985).

[42] Z. L. Deng, "White–noise filter and smoother with application to seismic data deconvolution," in *Preprints 7th IFAC/IFORS Symp. Identification Syst. Parameter Estimation*, York, U.K., pp. 621–624 (1985).

[43] J. M. Mendel, *Optimal Seismic Deconvolution. An Estimation–Based Approach.* Academic Press, New York, 1983.

[44] S. M. Fitch and L. Kurz, "Recursive equalization in data transmission - A design procedure and performance evaluation," *IEEE Transactions on Communication*, vol. 23, pp. 546-550 (1975).

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

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

[47] A. Ahlén and M. Sternad, "Adaptive input estimation," *IFAC Symposium ACASP-89, Adaptive Systems in Control and Signal Processing*, pp. 631-636, Glasgow, UK (1989).

[48] A. Ahlén and M. Sternad, "Adaptive deconvolution based on spectral decomposition," *SPIE Annual Symposium*, San Diego, USA, vol 1565, pp 130–142 (1991).

[49] A. Ahlén, "Identifiability of the deconvolution problem," *Automatica*, vol. 26, pp. 177–181 (1990).

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

[51] R. S. Anderssen and P. Bloomfield, "Numerical differentiation procedures for non-exact data," *Numerische Mathematik*, vol. 22, pp. 157-182 (1974).

[52] B Carlsson, Digital differentiating filters and model–based fault detection. *Ph.D. thesis* (Uppsala Dissertations from the Faculty of Science, no 28), Uppsala University, Sweden, 1989.

[53] S. Fioretti, L. Jetto and T. Leo, A discrete-time delay system for efficient simultaneous derivative estimation. In *Signal Processing III: Theories and applications*, eds I T Young et al, Elsevier Science Publishers B V (North-Holland), pp. 89-92, 1986.

[54] L. R. Rabiner and K. Steiglitz, "The design of wide-band recursive and nonrecursive digital differentiators," *IEEE Transactions on Audio Electronics*, vol. 18, pp. 204-209 (1970).

[55] S. Usui and I. Amidror, "Digital low-pass differentiation for biological signal processing," *IEEE Transactions on Biomedical Engineering*, vol. 29, pp. 686-693 (1982).

[56] F. Farina and F. A. Studer, *Radar Data Processing. Volume I.* Research Student Press and Wiley, New York, 1985.

[57] M. Sternad and A. Ahlén, "Robust filtering and feedforward control based on probabilistic descriptions of model errors," To appear in *Automatica*, vol 29, no 3 (1993).

[58] C. A. Belfiore and J. H. Park, "Decision feedback equalization," *Proc IEEE*, vol. 67, pp. 1143-1156 (1979).

[59] D. L. Duttweiler, J. E. Mazo and D. G. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *IEEE Transactions on Information Theory*, vol. 20, pp. 490-497 (1974).

[60] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Transactions on Information Theory*, vol. 17, pp. 56-64 (1971).

[61] S. U. H. Qureshi, "Adaptive equalization," *Proc IEEE*, vol. 73, pp. 1349-1387 (1985).

[62] J. Salz, "Optimum mean-square decision feedback equalization," *Bell System Technical Journal*, vol. 52, pp. 1341-1374 (1973).

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

[64] L. Lindbom, *Adaptive equalizers for fading mobile radio channels.* Licenciate Thesis, Department of Technology, Uppsala University, Sweden, November, 1992.

[65] M. Sternad, A. Ahlén and E. Lindskog, "Robust decision feedback equalization," *ICASSP'93, International Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, April 26-30, (1993).

[66] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Prentice–Hall, Englewood Cliffs, NJ, 1979.

[67] S. Haykin, *Adaptive Filter Theory.* Prentice–Hall, Englewood Cliffs, NJ, 1986.

[68] M. Sternad and A. Ahlén, "Robust filtering based on probabilistic descriptions of model errors," *2nd IFAC Workshop on System Structure and Control*, Prague, Sept. 3–5 (1992).

[69] J. Rissanen and L. Barbosa, "Properties of infinite covariance matrices and stability of optimum predictors," *Information Science*, vol 1, pp 221–236 (1969).

[70] J. Rissanen and L. Barbosa, "A factorization problem and the problem of predicting nonstationary vector–valued stochastic processes," *Z. Wahrsch. Verw. Geb.* vol 2, pp 255-266 (1969).

[71] P. Caines, *Linear Stochastic Systems.* Wiley, New York, 1987.

[72] K. Öhrn, A. Ahlén and M. Sternad, "A probabilistic approach to multivariable robust filtering and open loop control," Submitted. Also available as Report UPTEC 93008R, Systems and Control Group, Uppsala University, Sweden.

[73] V. Kučera, "The LQG control problem: a study of common factors," *Problems of Control and Information Theory (Hungary)*, vol 13, pp. 239–251 (1984).

[74] M. J. Grimble, "Single versus double Diophantine equation debate: comments on 'A polynomial approach to Wiener filtering'," *International Journal of Control*, vol 48, pp. 2161–2165 (1988).

[75] L. Chisci and E. Mosca, "A general polynomial solution to the MMSE deconvolution problem," *IEEE Transactions on Signal Processing*, vol 39, pp. 962–965 (1991).

INDEX

A

ARMA, *see* Autoregressive moving-average processes
Array manifold, cyclostationary signals and, 4
Autoregressive moving-average processes
  minimum phase model, input and parameters, 187–188, 201
    bootstrap algorithm, 212–214
    estimation, 187–188
    Gaussian white noise input, 187–189, 191–196, 199–200
    simulation results, 180–182
  nonminimum phase model, input and parameters, 173–201, 203–214
    algorithms, 178–180, 207–212
    approximate inverse systems, 175–178
      least-squares approximation method, 205–206
      long-division method, 203–205
      simulation results, 182–187
    background, 173–174
    estimation, 188–200
    Gaussian white noise input, 187–189, 191–196, 199–200
    impulse sequence input, 188, 190, 197–198
    Kalman filtering, 178–180
    simulation results, 180–182

B

Backward smoothing, 312
Bandstop filtering, cyclostationary signals and, 38, 40–41
Bartlett window, 86, 88, 90–91
Beamforming
  finite data case and, 310–311
  smoothing and, 301–302, 335
Blackman window, 86, 88, 90–91
Broadband signals, smoothing and, 331–346
  background, 331–337
  beamforming, 335
  correlation method, 335–337
  spatial smoothing, 337–338
  weighted smoothing, 338–343

C

Canonical correlation analysis, 25–26; *see also* Programmable Canonical Correlation Analysis
Chebyshev approximation, finite impulse response filter design and, 262–264
  optimal, 249–253
  Rivlin and Shapiro, 251–252
Chebyshev window, 86, 89–91
Cholesky factors, stochastic time series modeling and, 157, 158
Coherence estimation, 129–130
Constrained conditional maximum likelihood, cyclostationary signals and, 26–30
Convex hull, finite impulse response filter design and, 251
Cycle frequency, cyclostationarity and, 9
Cyclostationarity, 8–9
Cyclostationary signals, 1–63
  cycle frequencies, 9