Realizable MIMO Decision Feedback Equalizers: Structure and Design

Claes Tidestav, Anders Ahlén
and Mikael Sternad

September 1998

UPPSALA UNIVERSITY
Signals and Systems

Abstract

We present and discuss optimum multivariable decision feedback equalizers (DFE:s). The equalizers are derived under the constraint of realizability, requiring causal and stable filters and finite smoothing lag. The design is based on a discrete-time channel model, where a digital signal passes through a dispersive multivariable channel with infinite impulse response. The additive noise is described by a multivariate ARMA model. Both minimum mean square error (MMSE) and zero-forcing (ZF) DFE:s are derived, under the assumption of correct past decisions.

For the MMSE DFE, the optimal structure is obtained, and it is noted that the conventional structure, with FIR filters in both the feedforward and the feedback links is optimal only under rather restrictive conditions. Simple design equations on closed form are also presented.

Conditions for the existence of a ZF DFE are presented, and we suggest that the existence of a ZF DFE guarantees *near-far resistance* of the corresponding MMSE DFE.

Simulations indicate that it may be advantageous to use a DFE with optimal structure as opposed to the conventional structure. However, in some cases, the conventional structure is close to optimal, and in these cases, the performance degradation is small for the conventional DFE. Also, the performance improvement of the optimum DFE is reduced when error propagation is taken into account.

Contents

Chapter 1

Introduction

During the last three decades, *decision feedback equalizers (DFE:s)* have been used in digital communication to suppress *intersymbol interference (ISI)*, i.e. to remove the effects of a frequency selective communication channel. The DFE constitutes a good compromise between performance and complexity: It provides much better performance than a linear equalizer, and it has a much lower complexity than the optimum detector, the maximum likelihood sequence estimator (MLSE).

A DFE consists of two filters and a decisions non-linearity. The ISI corrupted measurements are input to the *feedforward filter*. From the output of the feedforward filter, the output of the feedback filter is subtracted to remove the effect of residual ISI caused by the already detected symbols. A hard decision is then made to decide what symbol was transmitted. This decision is fed into the feedback filter to remove its effect on future symbol estimates. The coefficients of the feedforward and feedback filters are adjusted according to a criterion, the two most common being the zero-forcing (ZF) criterion and the minimum mean square error (MMSE) criterion. With a zero-forcing equalizer, all intersymbol interference is removed, whereas with an MMSE equalizer, the mean square difference between the transmitted signal and a soft signal estimate is minimized.

In the literature [1], two types of MMSE DFE designs have been proposed:

1. Optimal model based design, which results in a continuous-time non-causal feedforward filter. The structure is optimized based on the transfer function of the communication channel.

2. Fixed structure design, where the structure of the DFE is fixed prior to the design. The resulting DFE often has FIR filters of predetermined degrees both in the feedforward and feedback paths. The coefficients of these filters are then determined by solving the Wiener-Hopf equations.

The performance of the non-causal DFE is always better than that of a realizable DFE. The fixed structure DFE on the other hand may have a suboptimal structure, resulting in suboptimum performance. The optimum performance of a realizable DFE can thus only be bounded using these results.

This dilemma was resolved in [2], where an *optimum realizable* DFE was derived. The derivation was based on discrete time IIR models of the channel and the noise, and the resulting DFE had optimal structure with optimal filter degrees. Design equations on closed form were also presented.

During the last few years, channels with several inputs and/or outputs have gained increased interest. Such channels occur in many areas, e.g. in cellular communication systems where antenna arrays are used to improve the detection. Oversampled channel models can also be formulated as a channel with several outputs. With a detector based on a model with multiple inputs and/or outputs, it is possible to suppress not only intersymbol interference, but also co-channel interference, i.e. interference from other signals.

A *multiple input-multiple output (MIMO)* DFE is a DFE where both the feedforward and the feedback filter have multiple inputs and multiple outputs. The DFE is an attractive compromise between complexity and performance also in the MIMO case. As in the scalar case, studies of MIMO DFE:s are based on one of two principles: either a DFE with a non-causal feedforward filter [3] or a DFE whose structure is fixed prior to the design [4, 5, 6].

In this paper, we present a generalized DFE with several inputs and outputs, which minimizes the mean square error under the constraint of realizability. The resulting DFE utilizes multivariable IIR filters with optimal filter degrees, and its parameters can be obtained from closed form design equations. In the limit, when the smoothing lag tends to infinity, we also obtain the non-realizable MMSE DFE. Furthermore, we introduce the existence of a zero forcing MIMO DFE as a criterion for near-far resistance of the corresponding MMSE DFE. Our derivations are based on a discrete time system model, where the multivariable channel may have an infinite impulse response, and where the noise is described by a multivariate ARMA model.

The paper is organized as follows: In Chapter 2, we present scenarios which lead to multivariable channel models. A concise description of the equalization problem and the parameterization of the channel and noise models is then given in Chapter 3. In Chapter 4, the structure and the design equations for the optimum realizable minimum mean square error DFE are presented. We also discuss the conditions for the existence of a zero-forcing MIMO DFE and its implications on the design of MMSE DFE:s. Finally, we use our framework to derive the optimum *non-realizable* DFE. In Chapter 5, a numerical example illustrates advantages and drawbacks of using a DFE with optimal structure as opposed to a conventional DFE with FIR filters in both the feedforward and the feedback link. Finally, in Chapter 6, conclusions are drawn and topics for future studies are indicated.

Chapter 2

Multivariable channel models

2.1 Receiver frontend

The front-end of a receiver in the considered radio communication system is depicted in Figure 2.1.

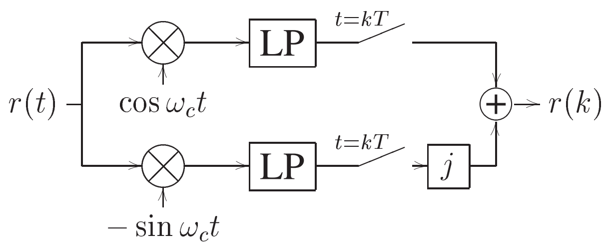

Figure 2.1: The front-end of a receiver in the considered communication system.

The in-phase and quadrature components of the received passband signal $r(t)$ are down-converted to the baseband. The baseband signal is passed through a fixed anti-aliasing filter and sampled. Depending on the application, the sampling rate either equals, or is a multiple of, the symbol rate. When designing the detectors, the transmitter filter, the multipath channel and the receiver filter are lumped together, and the resulting discrete-time channel from the transmitted symbols to the received sampled signal is used as a basis for detector design.

Note that all signals are represented by their complex envelopes and that all coefficients in the discrete-time channel models in general will be complex-valued. This is due to the fact that we are considering a communication system using radio frequency carriers.

Remark 1. As is apparent from Figure 2.1, the receiver front-end incorporates no filter matched to the received signal, as is common in optimal, model-based detector design [7]. This means that there is no guarantee that the sampled signal constitutes a sufficient statistic of the original continuous-time signal $r(t)$. Also, the noise suppression of the anti-aliasing filter is worse than that of an ideal matched filter. Still, we will not use a matched filter as an *a priori* component in our detector design. The reason for this is threefold:

1. As explained in Section 4.4, matched filtering may not be optimal for a detector with finite smoothing lag.

2. For channels having infinite impulse responses (IIR channels), the matched filter would not be realizable.
3. In a practical communication system, a fixed analog filter must be used.

□

2.2 Notations

Throughout the paper, channels and filters are assumed to be linear and time-invariant. A scalar discrete-time channel or filter will be represented as a rational function in the unit delay operator q^{-1}, i.e. as a ratio of polynomials in q^{-1} as exemplified below:

$$\begin{aligned} v(k) &= \mathcal{H}(q^{-1})u(k) = \frac{B(q^{-1})}{A(q^{-1})}u(k) \\ &= -A_1 v(k-1) - \cdots - A_{na} v(k-na) \\ &+ B_0 u(k) + B_1 u(k-1) + \cdots + B_{nb} u(k-nb) \ . \end{aligned}$$

When appropriate, the complex variable z will be substituted for the forward shift operator q. For convenience, polynomial arguments will often be omitted when there is no risk of misunderstanding.

A channel having multiple inputs and multiple outputs will be called a *multivariable* channel or a *multiple input-multiple output (MIMO)* channel. Multivariable channels will be described by matrices, whose elements are rational functions. Such matrices are called *rational matrices*. We can expand any stable and causal rational matrix $\boldsymbol{\mathcal{T}}(q^{-1})$ in the series

$$\boldsymbol{\mathcal{T}}(q^{-1}) = \sum_{n=0}^{\infty} \mathbf{T}_n q^{-n} \ . \tag{2.1a}$$

where $\mathbf{T}_n$ are the Markov parameters of $\boldsymbol{\mathcal{T}}(q^{-1})$. For any such rational matrix, we define

$$\boldsymbol{\mathcal{T}}_*(q) \stackrel{\triangle}{=} \sum_{n=0}^{\infty} \mathbf{T}_n^H q^n \tag{2.1b}$$

where $(\cdot)^H$ denotes complex conjugate transpose. We will also consider rational matrices which are non-causal. Such a matrix can be expanded in the series

$$\boldsymbol{\mathcal{T}}(q, q^{-1}) = \sum_{n=-\infty}^{\infty} \mathbf{T}_n q^{-n} \ . \tag{2.2a}$$

In this case, the definition (2.1b) can be generalized to

$$\boldsymbol{\mathcal{T}}_*(q, q^{-1}) \stackrel{\triangle}{=} \sum_{n=-\infty}^{\infty} \mathbf{T}_n^H q^n \ . \tag{2.2b}$$

In some cases, the denominators of all the matrix elements will be constants, rather than polynomials. In these cases, the multivariable channel can be described by a *polynomial matrix*:

$$\boldsymbol{P}(q^{-1}) \stackrel{\triangle}{=} \mathbf{P}_0 + \mathbf{P}_1 q^{-1} + \cdots + \mathbf{P}_{\delta P} q^{-\delta P} \ . \tag{2.3a}$$

The degree of a polynomial matrix $\boldsymbol{P}(q^{-1})$ equals the highest degree of any of its elements, and is denoted δP. For any polynomial matrix (2.3a), we also define

$$\boldsymbol{P}_*(q) \triangleq \mathbf{P}_0^H + \mathbf{P}_1^H q + \cdots + \mathbf{P}_{\delta P}^H q^{\delta P} \tag{2.3b}$$
$$\bar{\boldsymbol{P}}(q^{-1}) \triangleq q^{-\delta P}\boldsymbol{P}_*(q) = \mathbf{P}_{\delta P}^H + \mathbf{P}_{\delta P-1}^H q^{-1} + \cdots + \mathbf{P}_0^H q^{-\delta P}\ . \tag{2.3c}$$

2.3 Examples of multivariable channel models

With the receiver front-end depicted in Figure 2.1, the resulting channel is scalar, baseband and discrete-time. We will now describe scenarios where several such channels are combined to form a multivariable channel.

2.3.1 Detection using antenna arrays

In cellular communication systems, multi-element antennas, also known as antenna arrays, are frequently used to reject interference and to reduce the effect of fading and noise [8, 9].

The introduction of antennas with several elements at the receiver results in a channel with multiple outputs. Assume that the received signal is sampled at the symbol rate. The signal received at antenna i is denoted $y_i(k)$, and the additive noise received at the same antenna is denoted $v_i(k)$. Furthermore, assume that the scalar channel from the transmitter to receiver antenna i is described by the transfer operator $\mathcal{H}_i(q^{-1})$. To obtain a collective representation of the antenna signals, we form the following vectors:

$$y(k) \triangleq \begin{pmatrix} y_1(k) & y_2(k) & \dots & y_{n_y}(k) \end{pmatrix}^T \tag{2.4a}$$
$$\boldsymbol{\mathcal{H}}(q^{-1}) \triangleq \begin{pmatrix} \mathcal{H}_1(q^{-1}) & \mathcal{H}_2(q^{-1}) & \dots & \mathcal{H}_{n_y}(q^{-1}) \end{pmatrix}^T \tag{2.4b}$$
$$v(k) \triangleq \begin{pmatrix} v_1(k) & v_2(k) & \dots & v_{n_y}(k) \end{pmatrix}^T\ , \tag{2.4c}$$

where n_y is the number of antenna elements in the array. The vector of received samples can then be expressed as

$$y(k) = \boldsymbol{\mathcal{H}}(q^{-1})d(k) + v(k) \tag{2.5}$$

where $d(k)$ is the transmitted signal. The model (2.5) is a *single input-multiple output (SIMO)* model, and is depicted in Figure 2.2.

As is apparent from (2.4b), a SIMO filter is described by a column vector with rational elements. For the antenna array application considered here, the scalar channels $\mathcal{H}_i(q^{-1})$ are in fact accurately described by polynomials, and the SIMO channel can thus be modeled by a polynomial column vector.

The model (2.5) can be used as a basis for design of a detector performing *interference rejection*. This approach has been thoroughly investigated, in e.g. [8], [10] and [11]. In all these investigations, interference rejection improves the performance significantly.

In a scenario where several users transmit simultaneously, we can modify (2.5) to explicitly incorporate multiple users. For this purpose, we assume that the signal $d_j(k)$ is transmitted from user j, and define

$$d(k) \triangleq \begin{pmatrix} d_1(k) & d_2(k) & \dots & d_{n_d}(k) \end{pmatrix}^T\ . \tag{2.6a}$$

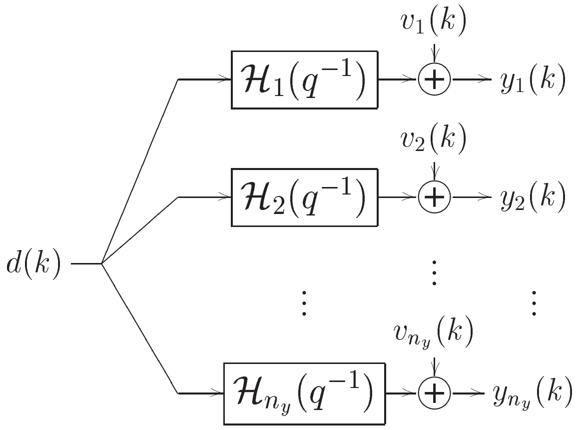

Figure 2.2: The presence of an antenna array at the receiver results in a single input-multiple output (SIMO) model.

We also denote the scalar channel from transmitter j to receiver antenna i as $\mathcal{H}_{ij}(q^{-1})$ and define the rational matrix

$$\boldsymbol{\mathcal{H}}(q^{-1}) \triangleq \begin{pmatrix} \mathcal{H}_{11}(q^{-1}) & \mathcal{H}_{12}(q^{-1}) & \dots & \mathcal{H}_{1n_d}(q^{-1}) \\ \mathcal{H}_{21}(q^{-1}) & \mathcal{H}_{22}(q^{-1}) & & \mathcal{H}_{2n_d}(q^{-1}) \\ \vdots & & \ddots & \vdots \\ \mathcal{H}_{n_y1}(q^{-1}) & \mathcal{H}_{n_y2}(q^{-1}) & \dots & \mathcal{H}_{n_yn_d}(q^{-1}) \end{pmatrix} . \tag{2.6b}$$

Using (2.4a), (2.4c), (2.6a) and (2.6b), we can now express the signal received at the antenna array by the multiple input-multiple output model

$$y(k) = \boldsymbol{\mathcal{H}}(q^{-1})d(k) + v(k) . \tag{2.7}$$

Of course, the model (2.5) is a special case ($n_d = 1$) of (2.7).

The channel model (2.7) can be used as a basis for design of a *multiuser detector*, which simultaneously detects the symbols transmitted from all users. This approach was first studied by Winters in [12] and more recently by Tidestav *et al.* in [6], where in addition, multiuser detection and interference rejection were compared.

2.3.2 Multiuser detection in DS-CDMA

In direct sequence code division multiple access (DS-CDMA) systems, the transmitted symbols are *spread* before transmission. Spreading implies that the signal is multiplied by a user-specific *spreading sequence*, a signal having (much) larger bandwidth than the information-bearing signal. At the receiver, the signals from the different users are *de-spread* by means of cross-correlation with the corresponding spreading sequence. If the spreading sequences are chosen almost orthogonal, the signals from different users will only interfere mildly with one another.

However, since the spreading sequences cannot be chosen completely orthogonal[1], the despread signal will contain contributions from all signals. The magnitude of this *multiple access interference (MAI)* is determined by the cross-correlation between the spreading sequences of the different users. The presence of MAI implies that we can formulate a

[1]In fact, this may not even be desirable in a real system.

multivariable model to describe this scenario. This model will have the symbols from all the users as input and the outputs of the cross-correlators, sampled at the symbol rate, as output. For this purpose, we collect the sampled outputs of the cross-correlators in a vector:

$$y'(k) \stackrel{\triangle}{=} \begin{pmatrix} y_1(k) & y_2(k) & \dots & y_{n_y}(k) \end{pmatrix}^T . \tag{2.8}$$

For a channel without delay spread, this vector is related to the vector of transmitted symbols $d(k)$ and a vector of noise samples $v'(k)$ through [13]

$$y'(k) = \mathbf{R}(1)\mathbf{W}d(k+1) + \mathbf{R}(0)\mathbf{W}d(k) + \mathbf{R}(-1)\mathbf{W}d(k-1) + v'(k) . \tag{2.9}$$

In (2.9), the matrices $\mathbf{R}(m)$ contain partial cross-correlations between the spreading sequences $s_i(t)$:

$$(\mathbf{R}(m))_{i,j} \stackrel{\triangle}{=} \int_{\tau_i}^{T_s+\tau_i} s_i(t-\tau_i)s_j^*(t+mT_s-\tau_j)dt ,$$

where T_s is the symbol period, $\tau_i \in [0, T_s[$ is the propagation delay of user i and $(\cdot)^*$ denotes complex conjugate. It is here assumed that $s_i(t) = 0$ outside $t \in [0, T_s[$. The elements in the diagonal matrix $\mathbf{W}$ represent the amplitudes and phase shifts of the respective signals.

The model (2.9) is non-causal. To obtain a causal model, we define $y(k) \stackrel{\triangle}{=} y'(k-1)$ and $v(k) \stackrel{\triangle}{=} v'(k-1)$, which can be inserted into (2.9) to yield:

$$y(k) = \boldsymbol{\mathcal{H}}(q^{-1})d(k) + v(k) ,$$

with a MIMO FIR channel of second order:

$$\boldsymbol{\mathcal{H}}(q^{-1}) = \mathbf{R}(1)\mathbf{W} + \mathbf{R}(0)\mathbf{W}q^{-1} + \mathbf{R}(-1)\mathbf{W}q^{-2} .$$

2.3.3 Fractionally spaced sampling.

We shall now turn our attention to *fractionally spaced sampling*. Fractionally spaced sampling implies that the received signal is sampled several, say p, times during a symbol period T_s, as depicted in Fig. 2.3.

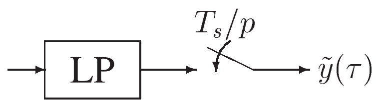

Figure 2.3: Fractionally spaced sampling

The sampling rate of the discrete-time signal $\tilde{y}(\cdot)$ equals p/T_s. Due to the oversampling, $\tilde{y}(\cdot)$ is not stationary, but cyclo-stationary: its moments vary periodically with a period equal to p. Since a new symbol is transmitted only once every p samples, we can express the received signal $\tilde{y}(\cdot)$ as

$$\tilde{y}(kp+r) = \tilde{H}(q^{-1})\tilde{d}(kp+r) + \tilde{v}(kp+r), \tag{2.10}$$

for

$$k = 0, 1, \ldots$$
$$r = 0, 1, \ldots, p-1 \, .$$

In (2.10), we have introduced a scalar sample-spaced, discrete-time input $\tilde{d}(\cdot)$, defined by

$$\tilde{d}(kp+r) = \begin{cases} d(k) & r = 0 \\ 0 & \text{otherwise} \end{cases}$$

and a scalar, sample-spaced channel

$$\tilde{H}(q^{-1}) = \tilde{H}_0 + \tilde{H}_1 q^{-1} + \tilde{H}_2 q^{-2} + \ldots \, .$$

To transform $\tilde{y}(kp+r)$ to a stationary signal, we collect p consecutive samples of $\tilde{y}(\cdot)$ in the vector

$$y(k) \stackrel{\triangle}{=} \begin{pmatrix} \tilde{y}(kp) & \tilde{y}(kp+1) & \ldots & \tilde{y}(kp+p-1) \end{pmatrix}^T \, . \tag{2.11}$$

The vector-valued stochastic process $y(k)$ is stationary and has sampling rate equal to the symbol rate $1/T_s$. To obtain the desired multivariable model, we introduce the single input-multiple output (SIMO) transfer operator

$$\boldsymbol{\mathcal{H}}(q^{-1}) \stackrel{\triangle}{=} \mathbf{H}_0 + \mathbf{H}_1 q^{-1} + \mathbf{H}_2 q^{-2} + \ldots \tag{2.12}$$

where

$$\mathbf{H}_m \stackrel{\triangle}{=} \begin{pmatrix} \tilde{H}_{mp} & \tilde{H}_{mp+1} & \ldots & \tilde{H}_{mp+p-1} \end{pmatrix}^T \, .$$

Analogous to (2.11), we stack the noise samples $\tilde{v}(kp+r)$ to form the vector $v(k)$:

$$v(k) \stackrel{\triangle}{=} \begin{pmatrix} \tilde{v}(kp) & \tilde{v}(kp+1) & \ldots & \tilde{v}(kp+p-1) \end{pmatrix}^T \, . \tag{2.13}$$

Using (2.12) and (2.13), we can express $y(k)$ as

$$y(k) = \boldsymbol{\mathcal{H}}(q^{-1}) d(k) + v(k) \, , \tag{2.14}$$

which is a SIMO model with symbol-spaced and stationary inputs and outputs.

Oversampling can be used to reduce the sensitivity of the receiver to synchronization errors, or simply to improve detector performance.

When several signals are transmitted over a common channel, an oversampled version of the received signal can be used to detect all of them. This can be implemented in CDMA systems, as described in [14, 15], and in xDSL systems, as described in [16]. The resulting model will have multiple inputs, as well as multiple outputs. The generalization of (2.14) to such a MIMO model is analogous to the generalization of (2.5) to (2.7) and will not be described any further.

2.4 Acquiring MIMO models

Channel models are seldom known *a priori*. Instead, they have to be estimated. However, since multivariable models can be parametrized in many different ways, ordinary prediction error methods [17] may be inappropriate, due to the large number of design variables involved. In this respect, *subspace identification* [18] offers a solution. With subspace identification, only one design variable is used: the order of the realization. The estimated model is a minimal state space realization, which can then be converted into any suitable form [19].

Fortunately, multivariate FIR models are still relatively easy to estimate, using least-squares methods. One remaining problem is that the number of parameters to estimate may be unnecessarily large. In particular, this is the case for the multivariable channel models described in section 2.3.1. These channels can in fact often be described by a *reduced rank* model [20]. This low-rank property can be taken into account when the channel is estimated [21], leading to more accurate models.

In the following, we will assume that all models are known without error. In practice, models are of course uncertain and better performance may be obtained with robust methods. With such methods, detectors which explicitly take the model uncertainty into account can be designed. See [22] and [23] for examples of such designs.

Chapter 3

Problem statement

Consider the received sequence of measurement vectors $y(k)$. Assume that each vector can be described as a sum of the output from a dispersive, multivariable channel and a multivariate noise term as depicted in Fig. 3.1. Both the channel and the noise model are

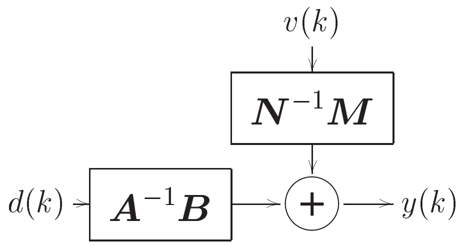

Figure 3.1: The multivariable system model. Both the IIR channel and the ARMA noise model are parametrized as left matrix fraction descriptions (MFD:s).

parameterized by *left matrix fraction descriptions (MFD:s)* [19]:

$$y(k) = \boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})d(k) + \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})v(k)\ . \tag{3.1}$$

The polynomial matrix $\boldsymbol{B}(q^{-1})$ has n_y rows and n_d columns, whereas $\boldsymbol{A}(q^{-1})$, $\boldsymbol{M}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are square polynomial matrices of dimension n_y. These three matrices are assumed to be *stably invertible*, i.e. the roots of

$$\begin{aligned} \det \boldsymbol{A}(z^{-1}) &= 0 \\ \det \boldsymbol{M}(z^{-1}) &= 0 \\ \det \boldsymbol{N}(z^{-1}) &= 0 \end{aligned} \tag{3.2}$$

all lie inside the unit circle $|z| = 1$.

We assume that the leading matrix coefficient in $\boldsymbol{M}(q^{-1})$ is non-singular, i.e. $\det \mathbf{P}_0 \neq 0$ in (2.3a). The denominator matrices $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are assumed to be monic (the leading matrix coefficient is equal to the identity matrix). To simplify the presentation, $\boldsymbol{A}(q^{-1})$ and $\boldsymbol{N}(q^{-1})$ are also assumed to be *diagonal*. This will result in less complex design equations, but might lead to unnecessarily high polynomial degrees in the matrix elements.[1] The polynomial elements in the matrices may have complex coefficients, and are assumed to be correctly estimated.

[1] Hence, neither $\boldsymbol{A}^{-1}\boldsymbol{B}$ nor $\boldsymbol{N}^{-1}\boldsymbol{M}$ constitute irreducible MFD:s.

Each element in the vector $d(k)$ is taken from a finite set of values, the so-called *alphabet*, i.e.

$$d_i(k) \in \mathcal{A}_i \ .$$

For instance, when binary phase shift keying (BPSK) is employed, $\mathcal{A}_i = \{+1, -1\}$. We will assume that each $d_i(k)$ is a stochastic variable with zero mean, which is uncorrelated with the disturbance vector $v(k)$. Finally, we assume that $d_i(k)$ is white with covariance matrix

$$Ed(k)d^H(k) = \lambda_d \mathbf{I} \ . \tag{3.3}$$

In a communication system employing interleaving, this assumption is in general valid.

The noise vector $v(k)$ in (3.1) has n_y elements. It is a possibly complex-valued, white stochastic process with zero mean and covariance matrix

$$E[v(k)v^H(k)] = \lambda_v \mathbf{I} \ . \tag{3.4}$$

For future reference, we also define

$$\rho \triangleq \frac{\lambda_v}{\lambda_d} \ . \tag{3.5}$$

Our primary goal is to reconstruct the sequence of symbol vectors $d(k)$ from the measurements of $y(k)$. For this purpose, we introduce the multiple input-multiple output general IIR decision feedback equalizer (GDFE):

$$\begin{aligned} \hat{d}(k-\ell|k) &= \boldsymbol{\mathcal{R}}(q^{-1})y(k) - \boldsymbol{\mathcal{F}}(q^{-1})\tilde{d}(k-\ell-1) \\ \tilde{d}(k-\ell) &= \mathrm{f}(\hat{d}(k-\ell|k)) \ . \end{aligned} \tag{3.6}$$

The feedforward filter $\boldsymbol{\mathcal{R}}(q^{-1})$ and the feedback filter $\boldsymbol{\mathcal{F}}(q^{-1})$ are stable and causal rational matrices. The design variable ℓ is known as the decision delay or the *smoothing lag*, i.e. the number of future measurements used to estimate the current symbol. The function $\mathrm{f}(\cdot)$ constitutes the decision non-linearity: For each element $\hat{d}_i(k-\ell|k)$ of the vector $\hat{d}(k-\ell|k)$, the decision device selects

$$\tilde{d}_i(k-\ell) = \arg \min_{d_i \in \mathcal{A}_i} |\hat{d}_i(k-\ell|k) - d_i|^2 \ .$$

The vector $\tilde{d}(k-\ell)$ thus constitutes the decision made on the estimate $\hat{d}(k-\ell|k)$. The GDFE is depicted in Fig. 3.2.

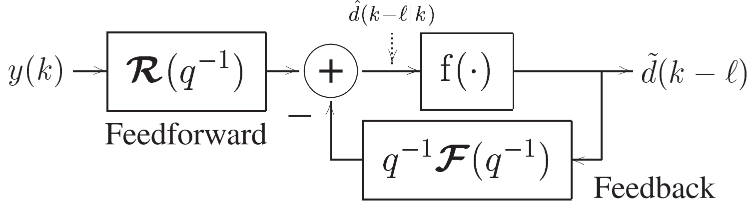

Figure 3.2: The general IIR decision feedback equalizer (GDFE).

It is important to note that the GDFE (3.6) must be *realizable*. This constraint implies that

- the smoothing lag ℓ must be finite, and
- the filters must be causal.

In Section 4.4, we will illustrate what happens when the constraint of realizability is relaxed.

Given the received sequence of symbol vectors $y(k)$ and the model (3.1), we want to find the stable and causal linear time-invariant MIMO filters $\{\boldsymbol{\mathcal{R}}(q^{-1}), \boldsymbol{\mathcal{F}}(q^{-1})\}$ which minimize the estimation error covariance matrix [2]

$$\mathbf{P} \stackrel{\triangle}{=} E\varepsilon(k-\ell)\varepsilon^H(k-\ell) \tag{3.7}$$

where the estimation error $\varepsilon(k-\ell)$ is defined as

$$\varepsilon(k-\ell) \stackrel{\triangle}{=} d(k-\ell) - \hat{d}(k-\ell|k) \; . \tag{3.8}$$

The GDFE which minimizes (3.7) will be called the *MMSE GDFE*.

We are also interested in finding the conditions under which a *zero-forcing* solution to the equalization problem exists. A scalar zero-forcing equalizer removes all intersymbol interference from the symbol estimate. A natural extension to the multivariable case [10] is to require that both the intersymbol interference and the co-channel interference, which is explicitely included in the channel model should be removed. A multivariable zero forcing (ZF) equalizer can then be defined accordingly:

Definition 1 *Consider the channel model* (3.1) *and a multivariable equalizer which forms the estimate* $\hat{d}(k-\ell|k)$ *of a transmitted symbol vector* $d(k-\ell)$*. If*

$$\hat{d}(k-\ell|k) = d(k-\ell) - \varepsilon(k-\ell) \tag{3.9}$$

where $\varepsilon(k-\ell)$ *is uncorrelated with all transmitted symbol vectors* $d(m)$ $\forall$ m*, then the equalizer is said to be* zero-forcing.

Because of the presence of the non-linear decision device in (3.6), closed form expressions for the parameters of the minimum mean square error or the zero-forcing GDFE cannot be found. To make derivation of optimum GDFE coefficients possible, we adopt the usual assumption that all past decisions affecting the current estimate are correct.[3] With this assumption, minimizing (3.7) becomes a quadratic optimization problem, and a zero-forcing equalizer can be found by solving a system of linear equations.

[2]The covariance matrix is minimized in the sense that any other admissible choice of $\{\boldsymbol{\mathcal{R}}(q^{-1}), \boldsymbol{\mathcal{F}}(q^{-1})\}$ will result in a estimation error covariance matrix $\tilde{\mathbf{P}}$ such that $\tilde{\mathbf{P}} - \mathbf{P}$ is positive definite.

[3]This assumption will almost surely be violated. Still, decision feedback equalizers generally work well, and more elaborate schemes give only small improvements [22].

Chapter 4

Optimum general decision feedback equalizers

We will now describe how to adjust the coefficients of the multivariable GDFE (3.6), so that the estimation error covariance matrix (3.7) is minimized. We also show how the GDFE can be tuned so that the zero-forcing condition (3.9) is satisfied.

4.1 The optimum MMSE GDFE

We introduce the following polynomial matrices:

$$\boldsymbol{\Gamma}(q^{-1}) \stackrel{\triangle}{=} \boldsymbol{A}(q^{-1})\boldsymbol{M}(q^{-1}) \tag{4.1a}$$

$$\boldsymbol{\tau}(q^{-1}) \stackrel{\triangle}{=} \boldsymbol{N}(q^{-1})\boldsymbol{B}(q^{-1}) \; . \tag{4.1b}$$

We also define the polynomial matrices $\tilde{\boldsymbol{\Gamma}}(q^{-1})$ and $\tilde{\boldsymbol{\tau}}(q^{-1})$ by the *coprime factorization*

$$\tilde{\boldsymbol{\tau}}(q^{-1})\tilde{\boldsymbol{\Gamma}}^{-1}(q^{-1}) = \boldsymbol{\Gamma}^{-1}(q^{-1})\boldsymbol{\tau}(q^{-1}) \; . \tag{4.2}$$

Without restriction, we assume that $\tilde{\boldsymbol{\tau}}(q^{-1})\tilde{\boldsymbol{\Gamma}}^{-1}(q^{-1})$ constitutes an irreducible MFD.[1]

We are now ready to formulate our main result.

Theorem 1 *Assume that a multivariable channel is described by* (3.1)*, and that the transmitted data is described by* (3.3)*, whereas the noise is described by* (3.4) *with* $\rho > 0$*. Assuming correct past decisions, the general multivariable DFE* (3.6) *minimizes the estimation error covariance matrix* (3.7) *if and only if*

$$\boldsymbol{\mathcal{R}}(q^{-1}) = \boldsymbol{S}(q^{-1})\boldsymbol{M}^{-1}(q^{-1})\boldsymbol{N}(q^{-1}) \tag{4.3a}$$

$$\boldsymbol{\mathcal{F}}(q^{-1}) = \boldsymbol{Q}(q^{-1})\tilde{\boldsymbol{\Gamma}}^{-1}(q^{-1}) \; . \tag{4.3b}$$

Above, $\boldsymbol{S}$ *and* $\boldsymbol{Q}$*, together with the polynomial matrices* $\boldsymbol{L}_1$ *and* $\boldsymbol{L}_2$ *can be calculated as the unique solution to the two coupled polynomial matrix equations*

$$\tilde{\boldsymbol{\Gamma}} - q^{\ell}\boldsymbol{S}\tilde{\boldsymbol{\tau}} + q^{-1}\boldsymbol{Q} = \boldsymbol{L}_{1*}\tilde{\boldsymbol{\Gamma}} \tag{4.4a}$$

$$q^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{\tau}_* - \rho\boldsymbol{S}\boldsymbol{\Gamma}_* = q\boldsymbol{L}_{2*} \tag{4.4b}$$

[1] A coprime factorization may be numerically sensitive. For a robust implementation, see [24].

where the degrees of the unknown polynomial matrices satisfy

$$\delta S = \ell \tag{4.5a}$$

$$\delta Q = \max(\delta\tilde{\Gamma}, \delta\tilde{\tau}) - 1 \tag{4.5b}$$

$$\delta L_1 = \ell \tag{4.5c}$$

$$\delta L_2 = \max(\delta\tau, \delta\Gamma) - 1 \; . \tag{4.5d}$$

Proof: See Appendix A. ■

Remark 1. The degrees listed in (4.5a)–(4.5d) are sufficiently high. When τ_0 (and hence $\tilde{\tau}_0$) has full rank, these degree conditions are also necessary.

Remark 2. The two coupled *Diophantine equations* (4.4a) and (4.4b) are *unilateral*, since all the unknown polynomial matrices appear on the same (in this case left) side of their respective coefficient polynomial matrices. Solving these unilateral Diophantine equations corresponds to solving a block-Toeplitz system of linear equations, as demonstrated in Appendix A, equations (A.18)–(A.21).

Remark 3. When the channel is equalized using the optimum MMSE DFE, the estimate of the transmitted vector will be given by

$$\begin{aligned}\hat{d}(k-\ell|k) &= \boldsymbol{\mathcal{R}}(q^{-1})y(k) - \boldsymbol{\mathcal{F}}(q^{-1})d(k-\ell-1)\\ &= (q^{-\ell}\mathbf{I} - \bar{\boldsymbol{L}}_1(q^{-1}))d(k) + \boldsymbol{S}(q^{-1})v(k)\end{aligned} \tag{4.6}$$

when all previous decisions are assumed correct. The *equivalent equalized channel* is thus given by

$$\boldsymbol{C}_{eq}(q^{-1}) = q^{-\ell}\mathbf{I} - \bar{\boldsymbol{L}}_1(q^{-1}) \; .$$

Also, from (4.6) we can calculate the resulting estimation error covariance matrix (3.7). Using the assumptions (3.3) and (3.4) we obtain

$$\mathbf{P} = \lambda_d \sum_{n=0}^{\ell} \mathbf{L}_{1n}^H \mathbf{L}_{1n} + \lambda_v \sum_{n=0}^{\ell} \mathbf{S}_n \mathbf{S}_n^H \; . \tag{4.7}$$

The first term in (4.7) is caused by residual intersymbol and co-channel interference from the first ℓ taps in the equalized channel. The deviation of the reference tap from the identity matrix also contributes to the term. The last term in (4.7) is caused by the noise.

Remark 4. The smoothing lag ℓ is a design variable and should be chosen as a trade-off between complexity and performance. In general, the smoothing lag should be chosen so that "enough" signal power can be collected by the feedforward filter before a decision is made. □

The presented MMSE solution provides an *optimal DFE structure*. It is evident that the conventional DFE structure, where both the feedforward and the feedback filters have finite impulse responses, is optimal *only* when $\boldsymbol{A}(q^{-1}) = \boldsymbol{M}(q^{-1}) = \mathbf{I}$. In other words, the channel must be described by a finite impulse response model, whereas the additive noise must be an autoregressive (or white) process.

In addition to providing an optimal DFE structure and optimal filter degrees, Theorem 1 gives guidelines on how to choose the filter degrees in a conventional structure when the use of the optimal structure is deemed inappropriate. For instance, when the noise can be described by a moving average model, Theorem 1 states that both feedforward and feedback filters should have infinite impulse responses. In that case, the transversal feedforward filter in the conventional DFE structure should have a long impulse response, particularly if the zeros of $\boldsymbol{M}(z^{-1})$ are located close to the unit circle.

4.2 The ZF GDFE

We now turn our attention to the general multivariable zero-forcing DFE.

Theorem 2 *Consider the multivariable channel model* (3.1) *and the general DFE* (3.6). *There exists a multivariable DFE satisfying the zero-forcing condition* (3.9) *if and only if there exist stable and causal rational matrices* $\boldsymbol{\mathcal{R}}(q^{-1})$ *and* $\boldsymbol{\mathcal{F}}(q^{-1})$ *such that*

$$q^{-\ell}\mathbf{I} = \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} - q^{-\ell-1}\boldsymbol{\mathcal{F}} \,. \tag{4.8}$$

Proof: See Appendix B. ■

Equation (4.8) may have several solutions. Hence, for a given channel, there may be many multivariable zero-forcing decision feedback equalizers. However, in some cases no solution to (4.8) will exist. The precise condition for this is stated in Lemma 1.

Lemma 1 *There exists a solution to* (4.8) *if and only if every common right factor of* $\boldsymbol{A}^{-1}\boldsymbol{B}$ *and* $q^{-\ell-1}\mathbf{I}$ *is also a right factor of* $q^{-\ell}\mathbf{I}$.[2]

Proof: The results follows from the general theory of Diophantine equations, see [26]. ■

Lemma 1 can be used to determine if a zero-forcing solution exists for any given channel. However, we can also use Lemma 1 to find cases where trivial channel characteristics preclude the existence of a zero-forcing equalizer. The following two corollaries exemplify two such cases.

Corollary 1 *If* $n_d > n_y$, *then* $\boldsymbol{A}^{-1}\boldsymbol{B}$ *and* $q^{-\ell-1}\mathbf{I}$ *will always have a common right factor which is not a right factor of* $q^{-\ell}\mathbf{I}$.

Proof: See Appendix C. ■

For the second corollary, we define

$$\Delta_i \stackrel{\triangle}{=} \text{the minimal delay of user } i \text{ in any channel,} \tag{4.9}$$

leading to the following formulation:

Corollary 2 *If* $\Delta_i > \ell$ *for some* i, *then* $\boldsymbol{A}^{-1}\boldsymbol{B}$ *and* $q^{-\ell-1}\mathbf{I}$ *will always have a common right factor which is not a right factor of* $q^{-\ell}\mathbf{I}$.

Proof: See Appendix C. ■

Remark 1. From the perspective of a detector designer, Corollary 2 is enlightening. When designing a multi-user detector, it is vital to choose a smoothing lag which is guaranteed to exceed the bulk delay of *all* users, even if we are interested only in the signal from one of them. □

[2] The right factors should be members of the ring of stable and causal rational matrices [26].

4.3 The ZF DFE solution and near-far resistance

The primary focus in the literature regarding equalizer design has been on MMSE equalizers. The reason for this is twofold:

1. When noise is present, an MMSE equalizer in general provides better performance than the corresponding ZF equalizer.

2. An MMSE equalizer is better suited for adaptive implementation, since it easy to derive an adaptive algorithm which recursively minimizes the MSE.

Therefore, in an actual implementation, an MMSE equalizer is in general preferable.

However, zero-forcing equalizers can provide information about the performance of their MMSE counterparts. In the multiuser case, the existence of a zero-forcing equalizer implies that all intersymbol and co-channel interference can be removed. In this case, the estimation error at the input of the decision device of the corresponding MMSE equalizer will vanish when the noise variance λ_v in (3.4) tends to zero. When no ZF equalizer can be found, there will be some residual interference at the input to the decision device of the corresponding MMSE equalizer, irrespective of the noise level. As the power of the interfering signals increases, so will the residual interference. In the limit as the power of the interfering signals goes to infinity, the MMSE detector will become useless. This phenomenon is called the *near-far* problem in the CDMA literature, and detectors that are capable of handling a situation with very disparate transmitter powers are said to be *near-far resistant*.

The discussion above suggests that the existence of a zero-forcing DFE can be used as an indicator of near-far resistance of the corresponding MMSE DFE. In fact, the existence of a zero-forcing DFE implies that the equalization problem is in some sense well-posed. For instance, Corollary 1 states that no ZF equalizer exists when $n_d > n_y$, an equalization problem which is badly posed. Even in this situation, an MMSE solution exists but will give a high MSE. In [6], the performance of different MMSE DFE:s is investigated in situations when no ZF solution exists.

4.4 The structure of decision feedback equalizers with asymptotically large smoothing lags

As mentioned in Section 2.1, it is common in theoretical investigations to let the received signal pass through a filter matched to the channel, and then design the DFE to operate on this signal. In other words, the matched filter is used as an *a priori* constraint on the DFE solution. In the MIMO case, a matched filter will have to be included for every scalar channel [3], resulting in a bank of $n_y \times n_d$ matched filters. In this section, we will show that such a bank of matched filters will indeed be present in the MSE optimal GDFE, but only when the smoothing lag ℓ is allowed to go to infinity.

To prove this suggestion, we rewrite (3.6) as

$$\begin{aligned}\hat{d}(k|k+\ell) &= q^{\ell}\boldsymbol{\mathcal{R}}(q^{-1})y(k) - \boldsymbol{\mathcal{F}}(q^{-1})\tilde{d}(k-1) \\ \tilde{d}(k) &= \mathrm{f}(\hat{d}(k|k+\ell)) \ .\end{aligned}$$

We now let ℓ tend to infinity to obtain the optimum non-realizable MIMO DFE:

$$\begin{aligned} \hat{d}^{\infty}(k) &= \boldsymbol{\mathcal{R}}^{\infty}(q, q^{-1})y(k) - \boldsymbol{\mathcal{F}}^{\infty}(q^{-1})\tilde{d}(k-1) \\ \tilde{d}(k) &= \mathrm{f}(\hat{d}^{\infty}(k)) \ , \end{aligned} \tag{4.10}$$

where we have defined

$$\begin{aligned} \hat{d}^{\infty}(k) &\triangleq \lim_{\ell\to\infty} \hat{d}(k|k+\ell) \\ \boldsymbol{\mathcal{R}}^{\infty}(q, q^{-1}) &\triangleq \lim_{\ell\to\infty} q^{\ell}\boldsymbol{\mathcal{R}}(q^{-1}) \\ \boldsymbol{\mathcal{F}}^{\infty}(q^{-1}) &\triangleq \lim_{\ell\to\infty} \boldsymbol{\mathcal{F}}(q^{-1}) \ . \end{aligned}$$

The coefficients of the non-realizable DFE can be obtained using the following theorem:

Theorem 3 *The non-causal feedforward filter* $\boldsymbol{\mathcal{R}}^{\infty}(q, q^{-1})$ *and causal feedback filter* $\boldsymbol{\mathcal{F}}^{\infty}(q^{-1})$ *of the optimum non-realizable MIMO MMSE GDFE* (4.10) *will be given by*

$$\boldsymbol{\mathcal{R}}^{\infty}(q, q^{-1}) = \frac{1}{\rho}\tilde{\mathbf{\Gamma}}_0\mathbf{W}^{-1}\boldsymbol{\beta}_*^{-1}\tilde{\mathbf{\Gamma}}_*\boldsymbol{\tau}_*\mathbf{\Gamma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{N} \tag{4.11a}$$

$$= \frac{1}{\rho}\tilde{\mathbf{\Gamma}}_0\mathbf{W}^{-1}\boldsymbol{\beta}_*^{-1}\tilde{\boldsymbol{\tau}}_*\boldsymbol{M}^{-1}\boldsymbol{N} \tag{4.11b}$$

$$\boldsymbol{\mathcal{F}}^{\infty}(q^{-1}) = q(\tilde{\mathbf{\Gamma}}_0\boldsymbol{\beta}\tilde{\mathbf{\Gamma}}^{-1} - \mathbf{I}) \ , \tag{4.11c}$$

where $\tilde{\mathbf{\Gamma}}_0$ *is the leading coefficient of* $\tilde{\mathbf{\Gamma}}$ *and* $\boldsymbol{\beta}$ *is the (monic) solution to the matrix spectral factorization*

$$\boldsymbol{\beta}_*\mathbf{W}\boldsymbol{\beta} = \tilde{\mathbf{\Gamma}}_*\tilde{\mathbf{\Gamma}} + \frac{1}{\rho}\tilde{\boldsymbol{\tau}}_*\tilde{\boldsymbol{\tau}} \ . \tag{4.12}$$

In (4.12)*,* $\mathbf{W}$ *is a constant, positive definite matrix which has been introduced to make* $\boldsymbol{\beta}$ *monic.*

Proof: See Appendix D. ■

Remark 1. Note that since $\tilde{\mathbf{\Gamma}}_0$ is non-singular, equation (4.12) has a unique solution. □

The filter $\boldsymbol{\tau}_*\mathbf{\Gamma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{N}$, which appears as a right factor of (4.11a), constitutes a bank of *whitening matched filters*. Hence, when the smoothing lag tends to infinity, there is no performance penalty associated with the introduction of such a filter prior to the optimization of the DFE. However, this is *not* true for the realizable DFE; no whitened matched filter is present in (4.3a).

4.5 A comparison between the DFE and MLSE

A maximum likelihood sequence estimator (MLSE) computes the transmitted signal which maximizes the conditional probability of the received signal, i.e. it selects

$$\{\tilde{d}(k)\}_{k=1}^{N} = \arg\max_{d(k)\in\mathcal{A}} p(\{y(k)\}_{k=1}^{N}|\{d(k)\}_{k=1}^{N}) \tag{4.13}$$

where $d(k) \in \mathcal{A}$ implies that element i in $d(k)$ should be taken from the alphabet $\mathcal{A}_i$. When the channel and noise statistics are known, the MLSE is the optimum sequence detector.

When the channel memory is finite, the optimization in (4.13) can be efficiently implemented using the Viterbi algorithm [27], which constitutes forward dynamic programming. Still, the complexity is very high. Assuming that the channel length is L and the size of the alphabet $\mathcal{A}_i$ is M_i, determination of the sequence $\{\tilde{d}(k)\}_{k=1}^{N}$ in (4.13) using the Viterbi algorithm requires on the order of

$$n_y^2 N \prod_{i=1}^{n_d} M_i^L$$

operations, which can be an enormously large number. On the other hand, calculating the coefficients of a corresponding multivariable MMSE DFE requires on the order of

$$n_y^3(\ell+1)^3 + n_y^2(\ell+1)^2 n_d$$

operations, while equalization of the N symbols requires approximately

$$N(n_y n_d(\ell+1) + n_d^2 L)$$

operations. For large n_d, alphabets and channel lengths, there is a large difference in complexity between the MLSE and the DFE. On the other hand, the difference in performance is generally considered small. [3]

[3]This is true when the SISO MLSE and SISO DFE are compared. To our knowledge, there are no results regarding the corresponding difference in performance between the MIMO MLSE and the MIMO DFE.

Chapter 5

A numerical example

To illustrate the potential performance improvements, which can be obtained by using the GDFE, a Monte Carlo simulation has been conducted. Consider the two input-two output FIR channel (cf. (3.1))

$$\boldsymbol{A} = \mathbf{I}, \quad \boldsymbol{B} = \begin{pmatrix} 0.979 + 0.204q^{-1} & 0.826 + 0.563q^{-1} \\ -0.843 - 0.538q^{-1} & 0.403 + 0.915q^{-1} \end{pmatrix} .$$

Over this channel, we transmit two BPSK modulated signals, i.e. $d_j(k) = \{+1, -1\},\ j = 1, 2$. At the receiver, noise is added. The noise is Gaussian and will be described by the first order moving average model

$$\boldsymbol{N} = \mathbf{I}, \quad \boldsymbol{M} = \begin{pmatrix} \text{-0.409} - 0.179q^{-1} & \text{-0.535} + 0.717q^{-1} \\ \text{-0.507} + 0.361q^{-1} & 0.761 - 0.181q^{-1} \end{pmatrix} .$$

This noise model has zeros in $z_{1,2} = 0.424 \pm 0.457i = 0.623e^{\pm j0.823}$. We compare the performance of two DFE:s with smoothing lag $\ell = 1$:

- The GDFE, with degrees and parameters given by Theorem 1.
- The conventional FIR DFE described in [28]. This DFE has FIR filters of degrees 2 and 1 in the feedforward and feedback links, respectively .[1]

In Fig. 5.1, the bit error rate (BER) is displayed as a function of the signal-to-noise ratio (SNR) of a single user. The SNR of user j is defined as

$$\text{SNR}_j \triangleq \frac{E||s_j(k)||^2}{E||w(k)||^2} \tag{5.1}$$

where

$$s_j(k) \triangleq \boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}_j(q^{-1})d_j(k) \ ; \quad w(k) \triangleq \boldsymbol{N}^{-1}(q^{-1})\boldsymbol{M}(q^{-1})v(k)$$

and $\boldsymbol{B}_j(q^{-1})$ is column j in $\boldsymbol{B}(q^{-1})$. We assume that the SNR:s of the two users are identical, i.e. that $\text{SNR}_1 = \text{SNR}_2$. The simulations are performed using both correct decisions and decisions from the decision device.

From Fig. 5.1, we see that in some cases, it is advantageous to take the noise model into account. The GDFE does this in an optimum way, whereas the conventional DFE

[1] These degrees have been chosen so that both DFE:s are described using the same number of parameters.

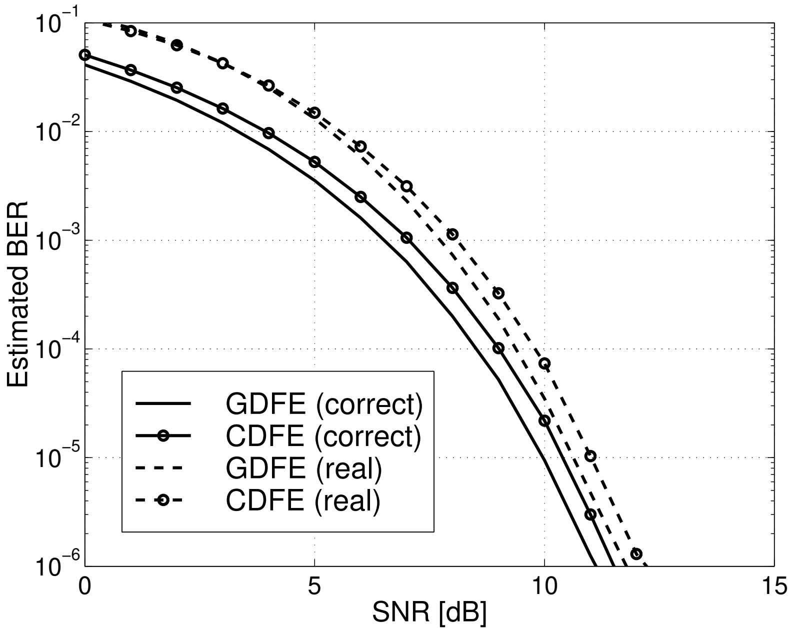

Figure 5.1: The BER of the GDFE compared to the BER of the conventional DFE (CDFE) for correct decisions (solid line) and real decisions (dashed line).

does not. With correct decisions, the GDFE is about 0.6 dB better than the conventional DFE over the range of investigated SNR:s. With real decisions, the performance of the two DFE:s is identical for low SNR:s, while it differs by 0.6 dB for high SNR:s.

By using higher filter degrees, the performance of the conventional DFE would be improved. In the limit, when the filter lengths go to infinity, it attains the performance of the GDFE.

From Fig. 5.1, we also see that with real decisions, the performance of both DFE:s worsen, and that the difference between the two DFE:s is smaller. This indicates that the DFE with optimal structure is more sensitive to incorrect past decisions.

For the considered FIR channel, the conventional DFE structure is optimal when the additive noise is temporally white. White noise corresponds to noise described by a moving average process, whose zeros are located in the origin. Therefore, the difference between the optimum DFE and the conventional DFE should be smaller, the closer to the origin the noise zeros lie. Correspondingly, when the noise zeros lie close to the unit circle, the difference should be larger.

To investigate this assumption, the locations of the zeros of the noise model are varied according to

$$z_{1,2}(r) = re^{\pm j0.823},\ r = 0.01, 0.1, 0.2, \ldots, 0.9, 0.95, 0.98, 0.99,$$

while the SNR, as defined in (5.1), is kept constant at 5 dB. Thus, the noise model zeros are moved along a radius, from the origin towards the unit circle. All other conditions for the simulation scenario are as in Fig. 5.1. The result is depicted in Fig. 5.2.

The location of the noise zeros clearly affects the relative performances of the two algorithms. When the zeros are close to the origin, the performance of the two DFE:s are identical, but the further out towards the unit circle the zeros are moved, the larger the difference. One interesting discovery is the performance of the GDFE with real decisions when the noise zeros are located very close to the unit circle: In this scenario, error

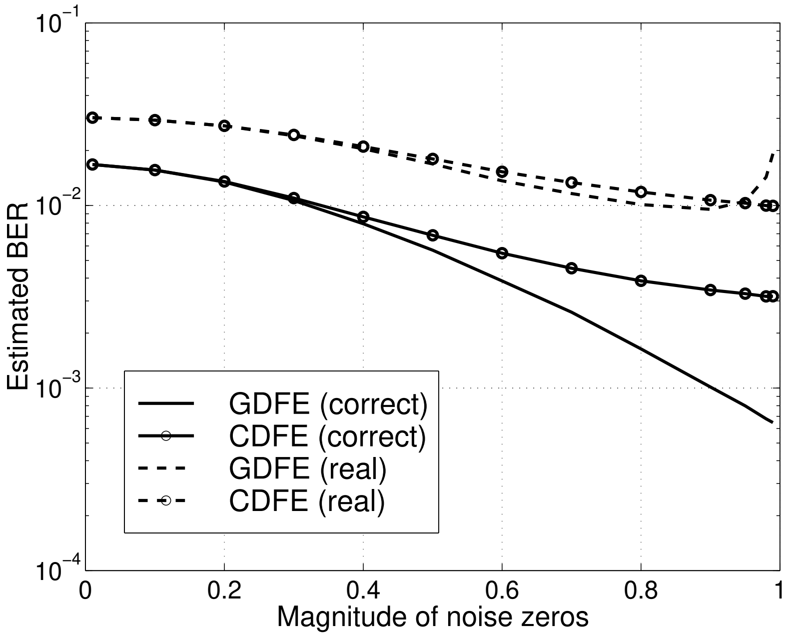

Figure 5.2: The BER of the GDFE as compared to the conventional DFE (CDFE) as a function of the location of the zeros of the noise model.

propagation causes very bad performance for the GDFE. The reason is that when the noise zeros are close to the unit circle, so are the poles of the feedback filter. The impulse response of the feedback filter then becomes very long, leading to a higher probability of error bursts.

Chapter 6

Conclusions

From a practical point of view, a decision feedback equalizer must be realizable. Also, optimum performance can be achieved only if the structure of the DFE is appropriate for the considered scenario. We have presented a generalized DFE with optimal structure, derived under the constraint of realizability. The IIR filters in this DFE are obtained from closed form design equations, which involve the channel and noise description. By allowing the smoothing lag to go to infinity, we have also derived the optimum non-realizable DFE.

New findings regarding the *near-far resistance* of the MIMO MMSE decision feedback equalizer have also been presented. By investigating the possible existence of a zero-forcing MIMO DFE, important conclusions can be drawn: If a ZF DFE does not exist, the corresponding MMSE DFE will not be near-far resistant.

The performance of the general DFE is demonstrated in a numerical example. This example indicates that for heavily colored noise, the GDFE outperforms the conventional FIR DFE. However, the conventional DFE seems to be less sensitive to the presence of incorrect past decisions: When incorrect decisions occur, the difference in performance is reduced.

Throughout the paper, we have assumed that channel and noise models are accurately known. In practice however, estimation of these models are prone to error. This is especially true for the noise model. The performance of the GDFE as compared to the conventional DFE when identified models are employed is a topic for future research. Another issue is to make the MIMO DFE more robust with respect to model errors and incorrect decisions, by taking the model and signal uncertainty into account already in the design.

Bibliography

[1] John G. Proakis, *Digital Communications*, McGraw–Hill, New York, NY, second edition, 1989.

[2] Mikael Sternad and Anders Ahlén, "The structure and design of realizable decision feedback equalizers for IIR channels with colored noise," *IEEE Transactions on Information Theory*, vol. 36, no. 4, pp. 848–858, July 1990.

[3] Alexandra Duel–Hallen, "Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 630–639, Apr. 1992.

[4] David D. Falconer, Majeed Abdulrahman, Norm W. K. Lo, Brent R. Petersen, and Asrar U. H. Sheikh, "Advances in equalization and diversity for portable wireless systems," *Digital Signal Processing*, vol. 3, no. 3, pp. 148–162, Mar. 1993.

[5] Claes Tidestav, Anders Ahlén, and Mikael Sternad, "Narrowband and broadband multiuser detection using a multivariable DFE," in *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, Sept. 1995, vol. 2, pp. 732–736.

[6] Claes Tidestav, Mikael Sternad, and Anders Ahlén, "Reuse within a cell—multiuser detection or interference rejection?," submitted.

[7] H. Vincent Poor, *An Introdution to signal detection and estimation*, Springer-Verlag, UK, second edition, 1995.

[8] Peter Monsen, "MMSE equalization of interference on fading diversity channels," *IEEE Transactions on Communications*, vol. 32, no. 1, pp. 5–12, Jan. 1984.

[9] Sören Andersson, Mille Millnert, Mats Viberg, and Bo Wahlberg, "An adaptive array for mobile communication systems," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 230–236, Feb. 1991.

[10] Philip Balaban and Jack 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, no. 5, pp. 885–894, 1992.

[11] Gregory E. Bottomley and Karim Jamal, "Adaptive arrays and MLSE equalization," in *Proceedings of the 45th IEEE Vehicular Technology Conference*, Chicago, July 1995, vol. 1, pp. 50–54.

[12] Jack Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Transactions on Communications*, vol. 35, no. 11, pp. 1222–1230, Nov. 1987.

[13] Ruxandra Lupas and Sergio Verdú, "Near-far resistance of multi-user detectors in asynchronous channels," *IEEE Transactions on Communications*, vol. 38, no. 4, pp. 496–508, Apr. 1990.

[14] Majeed Abdulrahman, Asrar U. H. Sheikh, and David D. Falconer, "Decision feedback equalization for CDMA in indoor wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 4, pp. 698–706, May 1994.

[15] Claes Tidestav, "Designing equalizers based on explicit channel models of DS-CDMA systems," in *Proceedings of the 5th IEEE International Conference on Universal Personal Communications*, Cambridge, MA, Oct. 1996, pp. 131–135.

[16] Brent R. Petersen and David D. Falconer, "Minimum mean square equalization in cyclostationary and stationary interference-analysis and subscriber line calculations," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 6, pp. 931–940, Aug. 1991.

[17] Lennart Ljung, *System Identification*, Prentice-Hall, 1987.

[18] Peter van Overschee and Bart L.R. De Moor, *Subspace Identification for linear systems: Theory—Implementation—Applications*, Kluwer Academic Publishers, 1996.

[19] Thomas Kailath, *Linear Systems*, Prentice Hall, 1980.

[20] Erik Lindskog and Claes Tidestav, "Reduced rank space-time equalization," in *Proceedings of the 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Boston, MA, Sept. 1998.

[21] Petre Stoica and Mats Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3069–3078, Dec. 1996.

[22] Mikael Sternad, Anders Ahlén, and Erik Lindskog, "Robust decision feedback equalizers," in *Proceedings of ICASSP*, Minneapolis, MN, Apr. 1993, vol. 3, pp. 555–558.

[23] Kenth Öhrn, Anders Ahlén, and Mikael Sternad, "A probalistic approach to multivariable robust filtering and open-loop control," *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 405–418, Mar. 1995.

[24] Didier Henrion, Martin Hromcik, Huibert Kwakernaak, Sonja Pejchová, Michael Šebek, and Rens C.W. Strijbos, "Polynomial toolbox version 1.6," 1998, See http://www.math.utwente.nl/polbox.

[25] Claes Tidestav, Anders Ahlén, and Mikael Sternad, "Realizable decision feedback equalizers: structure and design," Tech. Rep., Signals and Systems, Uppsala University, Uppsala, Sweden, Oct. 1998, See also http://www.signal.uu.se/Publications/abstracts/r982.html.

[26] Vladimír Kučera, *Analysis and Design of Discrete Linear Control Systems*, Prentice Hall, 1991.

[27] G. David Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[28] Claes Tidestav, Mikael Sternad, and Anders Ahlén, "Reuse within a cell — multiuser detection or interference rejection?," Tech. Rep., Signals and Systems, Uppsala University, Uppsala, Sweden, May 1998, See also http://www.signal.uu.se/Publications/abstracts/r981.html.

[29] Alan V. Oppenheim and Ronald W. Schafer, *Digital Signal Processing*, Prentice-Hall International, 1975.

[30] Anders Ahlén and Mikael Sternad, "Wiener filter design using polynomial equations," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2387–2399, Nov. 1991.

Appendix A

Proof of Theorem 1

Consider the channel described by (3.1), and the general multivariable decision feedback equalizer (3.6). Insert the expression for the symbol estimate $\hat{d}(k-\ell|k)$ into the expression (3.8) for the estimation error:

$$\varepsilon(k-\ell) = d(k-\ell) - \hat{d}(k-\ell|k) = d(k-\ell) - \boldsymbol{\mathcal{R}}y(k) + \boldsymbol{\mathcal{F}}\tilde{d}(k-\ell-1)\,. \tag{A.1}$$

Assume correct past decisions, i.e.

$$\tilde{d}(m) = d(m) \quad m \leq k-\ell-1\,. \tag{A.2}$$

Insert $\tilde{d}(k-\ell-1)$ from (A.2) and $y(k)$ from (3.1) into (A.1) and rearrange:

$$\begin{aligned}\varepsilon(k-\ell) &= d(k-\ell) - \boldsymbol{\mathcal{R}}(\boldsymbol{A}^{-1}\boldsymbol{B}d(k) + \boldsymbol{N}^{-1}\boldsymbol{M}v(k)) + \boldsymbol{\mathcal{F}}d(k-\ell-1) \\ &= (q^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + q^{-\ell-1}\boldsymbol{\mathcal{F}})d(k) - \boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k)\,.\end{aligned} \tag{A.3}$$

Introduce the alternative estimate $\hat{d}_a(k-\ell|k) \stackrel{\triangle}{=} \hat{d}(k-\ell|k) + n(k)$, where the variation $n(k)$ is a linear function of all signals which the estimate $\hat{d}(k-\ell|k)$ may be based upon. Thus,

$$n(k) \stackrel{\triangle}{=} n_1(k) + n_2(k)$$

where

$$n_1(k) \stackrel{\triangle}{=} \boldsymbol{\mathcal{G}}_1 y(k) \tag{A.4a}$$
$$n_2(k) \stackrel{\triangle}{=} \boldsymbol{\mathcal{G}}_2 d(k-\ell-1)\,. \tag{A.4b}$$

Above, $\boldsymbol{\mathcal{G}}_1(q^{-1})$ and $\boldsymbol{\mathcal{G}}_2(q^{-1})$ are arbitrary stable and causal rational matrices. If the estimation error obtained with (3.6) is orthogonal to any admissible variation (A.4a), (A.4b), i.e. if

$$E\varepsilon(k-\ell)n_1^H(k) = 0 \tag{A.5a}$$
$$E\varepsilon(k-\ell)n_2^H(k) = 0 \tag{A.5b}$$

then $\hat{d}_a(k-\ell|k) = \hat{d}(k-\ell|k)$ or equivalently $n(k) \equiv 0$ minimizes the estimation error covariance matrix (3.7): For any estimation error covariance matrix $\tilde{\mathbf{P}}$ obtained with $n(k) \neq 0$, $\tilde{\mathbf{P}} - \mathbf{P}$ will be positive definite. We must thus assure that (A.5a) and (A.5b) are fulfilled.

To compute the cross-correlations (A.5a) and (A.5b), we will use Parseval's relation for complex signals [29] and evaluate the expressions in the frequency domain. We thus insert (A.3) and (A.4a) into (A.5a) and use Parseval's relation to rewrite the result

$$E\varepsilon(k-\ell)n_1^H(k) = \frac{\lambda_d}{2\pi j}\oint\big\{(z^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-\ell-1}\boldsymbol{\mathcal{F}}) \times \\ \boldsymbol{B}_*\boldsymbol{A}_*^{-1} - \rho\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{N}_*^{-1}\big\}\,\boldsymbol{\mathcal{G}}_{1*}\frac{dz}{z}\,. \quad \text{(A.6)}$$

For an explanation of how Parseval's relation is used to obtain this expression, see [30]. Since $\boldsymbol{A}$ and $\boldsymbol{N}$ are both diagonal, we can express (A.6) as

$$E\varepsilon(k-\ell)n_1^H(k) = \frac{\lambda_d}{2\pi j}\oint\big\{(z^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-\ell-1}\boldsymbol{\mathcal{F}}) \times \\ \boldsymbol{B}_*\boldsymbol{N}_* - \rho\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_*\big\}\,\boldsymbol{A}_*^{-1}\boldsymbol{N}_*^{-1}\boldsymbol{\mathcal{G}}_{1*}\frac{dz}{z}\,. \quad \text{(A.7)}$$

From (A.7) we see that $E\varepsilon(k-\ell)n_1^H(k)=0$ if and only if the integrand is analytic inside the unit circle. According to the assumption (3.2), $\boldsymbol{A}^{-1}$ and $\boldsymbol{N}^{-1}$ are stable. This implies that $\boldsymbol{A}_*^{-1}$ and $\boldsymbol{N}_*^{-1}$ are analytic inside the unit circle. The same applies for $\boldsymbol{\mathcal{G}}_{1*}$, since $\boldsymbol{\mathcal{G}}_1$ is required to be stable. If

$$(z^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-\ell-1}\boldsymbol{\mathcal{F}})\boldsymbol{B}_*\boldsymbol{N}_* - \rho\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_* = z\boldsymbol{\mathcal{L}}_{2*} \quad \text{(A.8)}$$

for some rational matrix $\boldsymbol{\mathcal{L}}_{2*}(z)$ with all its poles *outside* the unit circle, the integrand will thus be analytic inside the unit circle. Proceeding in the same way with (A.5b) results in a second condition

$$z^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-\ell-1}\boldsymbol{\mathcal{F}} = z^{-\ell}\boldsymbol{\mathcal{L}}_{1*} \quad \text{(A.9)}$$

for some rational matrix $\boldsymbol{\mathcal{L}}_{1*}(z)$ with all its poles *outside* the unit circle. However, none of the terms on the left hand side of (A.9) can have poles outside the unit circle. Therefore, we conclude that $\boldsymbol{\mathcal{L}}_{1*}(z)$ must be a polynomial matrix. Thus,

$$z^{-\ell}\mathbf{I} - \boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-\ell-1}\boldsymbol{\mathcal{F}} = z^{-\ell}\boldsymbol{L}_{1*} \quad \text{(A.10)}$$

for some polynomial matrix $\boldsymbol{L}_{1*}(z)$. We can now insert (A.10) into (A.8) to obtain

$$z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{B}_*\boldsymbol{N}_* - \rho\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_* = z\boldsymbol{\mathcal{L}}_{2*}\,. \quad \text{(A.11)}$$

We realize that neither of the terms on the left hand side of (A.11) can have any poles outside the unit circle. Therefore, $\boldsymbol{\mathcal{L}}_{2*}(z)$ cannot have any such poles either, and we conclude that $\boldsymbol{\mathcal{L}}_{2*}(z)$ must be a polynomial, rather than a rational, matrix. Hence we can express equation (A.11) as

$$z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{B}_*\boldsymbol{N}_* - \rho\boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}\boldsymbol{M}_*\boldsymbol{A}_* = z\boldsymbol{L}_{2*}\,. \quad \text{(A.12)}$$

for some polynomial matrix $\boldsymbol{L}_{2*}(z)$. In the integrand (A.12), the poles contributed by $\boldsymbol{N}^{-1}(z^{-1})$ must be canceled by a corresponding factor in $\boldsymbol{\mathcal{R}}(z^{-1})$. Also, the polynomial matrix $\boldsymbol{M}(z^{-1})$ contributes poles in the origin, which must be canceled by a corresponding factor in $\boldsymbol{\mathcal{R}}(z^{-1})$. We therefore insert

$$\boldsymbol{\mathcal{R}} = \boldsymbol{S}\boldsymbol{M}^{-1}\boldsymbol{N}$$

where $\boldsymbol{S}$ is an arbitrary polynomial matrix, into (A.10) and (A.12) and rearrange:

$$\begin{aligned} \mathbf{I} - z^{\ell}\boldsymbol{S}\boldsymbol{M}^{-1}\boldsymbol{N}\boldsymbol{A}^{-1}\boldsymbol{B} + z^{-1}\boldsymbol{\mathcal{F}} &= \boldsymbol{L}_{1*} \\ z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{B}_*\boldsymbol{N}_* - \rho\boldsymbol{S}\boldsymbol{M}_*\boldsymbol{A}_* &= z\boldsymbol{L}_{2*}\,. \end{aligned} \tag{A.13}$$

We now use that $\boldsymbol{A}^{-1}$ and $\boldsymbol{N}$ are diagonal and hence commute. We also insert (4.1a) and (4.1b) into (A.13):

$$\begin{aligned} \mathbf{I} - z^{\ell}\boldsymbol{S}\boldsymbol{\Gamma}^{-1}\boldsymbol{\tau} + z^{-1}\boldsymbol{\mathcal{F}} &= \boldsymbol{L}_{1*} \\ z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{\tau}_* - \rho\boldsymbol{S}\boldsymbol{\Gamma}_* &= z\boldsymbol{L}_{2*}\,. \end{aligned} \tag{A.14}$$

Equation (A.14) can be further simplified by using the coprime factorization (4.2) and multiplying with $\tilde{\boldsymbol{\Gamma}}(q^{-1})$ from the right:

$$\tilde{\boldsymbol{\Gamma}} - z^{\ell}\boldsymbol{S}\tilde{\boldsymbol{\tau}} + z^{-1}\boldsymbol{\mathcal{F}}\tilde{\boldsymbol{\Gamma}} = \boldsymbol{L}_{1*}\tilde{\boldsymbol{\Gamma}} \tag{A.15a}$$

$$z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{\tau}_* - \rho\boldsymbol{S}\boldsymbol{\Gamma}_* = z\boldsymbol{L}_{2*} \tag{A.15b}$$

Since $\boldsymbol{\mathcal{F}}$ is the only remaining rational matrix in (A.15a), its poles must be canceled by a corresponding factor in $\tilde{\boldsymbol{\Gamma}}$. We thus conclude that

$$\boldsymbol{\mathcal{F}} = \boldsymbol{Q}\tilde{\boldsymbol{\Gamma}}^{-1}$$

where $\boldsymbol{Q}$ is an undetermined polynomial matrix. Note that $\boldsymbol{Q}$ and $\tilde{\boldsymbol{\Gamma}}$ may have common factors. We can now insert this expression into (A.15a) to yield

$$\tilde{\boldsymbol{\Gamma}} - z^{\ell}\boldsymbol{S}\tilde{\boldsymbol{\tau}} + z^{-1}\boldsymbol{Q} = \boldsymbol{L}_{1*}\tilde{\boldsymbol{\Gamma}} \tag{A.16a}$$

$$z^{-\ell}\boldsymbol{L}_{1*}\boldsymbol{\tau}_* - \rho\boldsymbol{S}\boldsymbol{\Gamma}_* = z\boldsymbol{L}_{2*} \tag{A.16b}$$

By exchanging the unknown z for q, equation (A.16a) coincides with (4.4a), whereas (A.16b) coincides with (4.4b).

The Diophantine equations (A.16a) and (A.16b) are *double sided*, i.e. they contain powers of both z^{-1} and z. Thus, both the powers of z^{-1} and z on the left hand side must match the corresponding powers on the right hand side. For this purpose, we list the degrees in z^{-1} and z for each term in (A.16a) and (A.16b):
Equation (A.16a):

$$z^{-1}: \quad \delta\tilde{\Gamma},\ \delta S + \delta\tilde{\tau} - \ell,\ \delta Q + 1,\ \delta\tilde{\Gamma} \tag{A.17a}$$

$$z: \quad 0,\ \ell,\ 0,\ \delta L_1 \tag{A.17b}$$

Equation (A.16b):

$$z^{-1}: \quad \ell,\ \delta S,\ 0 \tag{A.17c}$$

$$z: \quad \delta L_1 + \delta\tau - \ell,\ \delta\Gamma,\ \delta L_2 + 1 \tag{A.17d}$$

From (A.17b) and (A.17c) we immediately obtain the conditions for the degrees of $\boldsymbol{L}_1$ and $\boldsymbol{S}$ respectively:

$$\delta L_1 = \ell; \quad \delta S = \ell\,.$$

If we insert $\delta S = \ell$ into (A.17a), we obtain (4.5b). Finally, by inserting $\delta L_1 = \ell$ into (A.17d), we obtain (4.5d).

It remains to show that equations (A.16a) and (A.16b) have a solution with the degrees specified by (4.5a)–(4.5d). For this purpose, we rewrite (A.16a) and (A.16b) as two systems of linear equations. Two matrix polynomials are identical if and only if all the corresponding coefficient matrices are identical. We must thus adjust the coefficients of $\boldsymbol{S}$, $\boldsymbol{Q}$, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ so that the expressions for the matrix coefficients for each power of z and z^{-1} are equal on the left and right hand side of (A.16a) and (A.16b). We thus evaluate the expressions for the matrix coefficients, conjugate, transpose and equate the left and right hand sides. For (A.16a) we obtain

$$\begin{pmatrix} \tilde{\boldsymbol{\tau}}_0^H & & 0 & \tilde{\mathbf{\Gamma}}_0^H & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ \tilde{\boldsymbol{\tau}}_{\delta\alpha}^H & & \tilde{\boldsymbol{\tau}}_0^H & \tilde{\mathbf{\Gamma}}_{\delta\alpha}^H & & \tilde{\mathbf{\Gamma}}_0^H \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \tilde{\boldsymbol{\tau}}_{\delta\alpha}^H & 0 & & \tilde{\mathbf{\Gamma}}_{\delta\alpha}^H \end{pmatrix} \begin{pmatrix} \mathbf{S}_0^H \\ \vdots \\ \mathbf{S}_\ell^H \\ \mathbf{L}_{1\ell} \\ \vdots \\ \mathbf{L}_{10} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \tilde{\mathbf{\Gamma}}_0^H \\ \boldsymbol{\alpha}_1^H \\ \vdots \\ \boldsymbol{\alpha}_{\delta\alpha}^H \end{pmatrix} \tag{A.18}$$

where we have defined

$$\boldsymbol{\alpha}(z^{-1}) = \tilde{\mathbf{\Gamma}}(z^{-1}) + z^{-1}\boldsymbol{Q}(z^{-1})\,. \tag{A.19}$$

Note that $\delta\alpha = \max(\delta\tilde{\tau}, \delta\tilde{\Gamma})$. In (A.18) $\tilde{\boldsymbol{\tau}}_m = 0$ if $m > \delta\tilde{\tau}$ and $\tilde{\mathbf{\Gamma}}_m = 0$ if $m > \delta\tilde{\Gamma}$. Proceeding in the same manner with (A.16b) results in

$$\begin{pmatrix} -\rho\mathbf{\Gamma}_{\delta L_2+1} & & 0 & \boldsymbol{\tau}_{\delta L_2+1} & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ -\rho\mathbf{\Gamma}_0 & & -\rho\mathbf{\Gamma}_{\delta L_2+1} & \boldsymbol{\tau}_0 & & \boldsymbol{\tau}_{\delta L_2+1} \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & -\rho\mathbf{\Gamma}_0 & 0 & \dots & \boldsymbol{\tau}_0 \end{pmatrix} \begin{pmatrix} \mathbf{S}_0^H \\ \vdots \\ \mathbf{S}_\ell^H \\ \mathbf{L}_{1\ell} \\ \vdots \\ \mathbf{L}_{10} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{2\delta L_2} \\ \vdots \\ \mathbf{L}_{20} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{A.20}$$

where $\mathbf{\Gamma}_m = 0$ if $m > \delta\Gamma$ and $\boldsymbol{\tau}_m = 0$ if $m > \delta\tau$.

Neither (A.18) nor (A.20) can be solved directly, since unknown coefficient matrices appear on their right hand sides. However, we can combine the first $(\ell+1)n_d$ equations from (A.18) with the last $(\ell+1)n_y$ equations from (A.20) to obtain a system with equal number of equations and unknowns

$$\begin{pmatrix} \tilde{\boldsymbol{\tau}}_0^H & & 0 & \tilde{\mathbf{\Gamma}}_0^H & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ \tilde{\boldsymbol{\tau}}_\ell^H & \dots & \tilde{\boldsymbol{\tau}}_0^H & \tilde{\mathbf{\Gamma}}_\ell^H & \dots & \tilde{\mathbf{\Gamma}}_0^H \\ -\rho\mathbf{\Gamma}_0 & \dots & -\rho\mathbf{\Gamma}_\ell & \boldsymbol{\tau}_0 & & \boldsymbol{\tau}_\ell \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & \dots & -\rho\mathbf{\Gamma}_0 & 0 & \dots & \boldsymbol{\tau}_0 \end{pmatrix} \begin{pmatrix} \mathbf{S}_0^H \\ \vdots \\ \mathbf{S}_\ell^H \\ \mathbf{L}_{1\ell} \\ \vdots \\ \mathbf{L}_{10} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \tilde{\mathbf{\Gamma}}_0^H \\ 0 \\ \vdots \\ 0 \end{pmatrix}\,, \tag{A.21}$$

where only known coefficient matrices appear on the right hand side. We will now show that this system of linear equations has a unique solution whenever $\rho > 0$. Define

$$\mathbf{T} \triangleq \begin{pmatrix} \boldsymbol{\tau}_0 & \dots & \boldsymbol{\tau}_\ell \\ & \ddots & \vdots \\ 0 & \dots & \boldsymbol{\tau}_0 \end{pmatrix};\ \tilde{\mathbf{T}} \triangleq \begin{pmatrix} \tilde{\boldsymbol{\tau}}_0 & \dots & \tilde{\boldsymbol{\tau}}_\ell \\ & \ddots & \vdots \\ 0 & \dots & \tilde{\boldsymbol{\tau}}_0 \end{pmatrix};\ \mathbf{G} \triangleq \begin{pmatrix} \mathbf{\Gamma}_0 & \dots & \mathbf{\Gamma}_\ell \\ & \ddots & \vdots \\ 0 & \dots & \mathbf{\Gamma}_0 \end{pmatrix};\ \tilde{\mathbf{G}} \triangleq \begin{pmatrix} \tilde{\mathbf{\Gamma}}_0 & \dots & \tilde{\mathbf{\Gamma}}_\ell \\ & \ddots & \vdots \\ 0 & \dots & \tilde{\mathbf{\Gamma}}_0 \end{pmatrix}$$

$$\theta \triangleq \begin{pmatrix} \mathbf{S}_0 & \dots & \mathbf{S}_\ell & \mathbf{L}_{1\ell}^H & \dots & \mathbf{L}_{10}^H \end{pmatrix}^H ; \ \tilde{\mathbf{g}} \triangleq \begin{pmatrix} 0 & \dots & \tilde{\mathbf{\Gamma}}_0 & 0 & \dots & 0 \end{pmatrix}^H .$$

Equation (A.21) can then be written as

$$\underbrace{\begin{pmatrix} \tilde{\mathbf{T}}^H & \tilde{\mathbf{G}}^H \\ -\rho\mathbf{G} & \mathbf{T} \end{pmatrix}}_{\mathbf{W}} \theta = \tilde{\mathbf{g}} .$$

The matrix $\mathbf{W}$ is non-singular since

$$\begin{aligned} \det \mathbf{W} &= (-1)^r \det \tilde{\mathbf{G}}^H \det \left(\rho\mathbf{G} + \mathbf{T}\tilde{\mathbf{G}}^{-H}\tilde{\mathbf{T}}^H\right) \\ &= (-1)^r \det \tilde{\mathbf{G}}^H \det \left(\rho\mathbf{G} + \mathbf{T}(\mathbf{G}^{-1}\mathbf{T})^H\right) \\ &= (-1)^r \det \tilde{\mathbf{G}}^H \det \left(\rho\mathbf{G}\mathbf{G}^H + \mathbf{T}\mathbf{T}^H\right) \det \mathbf{G}^{-H} \end{aligned}$$

where $(\cdot)^{-H} = ((\cdot)^{-1})^H$. Above, the integer r compensates for the sign shifts caused by the elementary operations performed on the determinant. In the second equality, we have used that $\tilde{\mathbf{T}}\tilde{\mathbf{G}}^{-1} = \mathbf{G}^{-1}\mathbf{T}$, which is a consequence of the coprime factorization (4.2). Now, since both $\mathbf{G}$ and $\tilde{\mathbf{G}}$ are non-singular, so are $\mathbf{G}^{-H}$ and $\tilde{\mathbf{G}}^H$. Also, $\mathbf{G}\mathbf{G}^H$ is positive definite. Therefore, the matrix $\mathbf{W}$ is non-singular whenever $\rho > 0$.

After having solved (A.21) for $\mathbf{S}_m^H$ and $\mathbf{L}_{1m}$, we use (A.18) to calculate the coefficients of $\boldsymbol{\alpha}$. We then compute the feedback filter $\boldsymbol{Q}$ with the aid of (A.19):

$$\boldsymbol{Q}(z^{-1}) = z(\boldsymbol{\alpha}(z^{-1}) - \tilde{\mathbf{\Gamma}}(z^{-1})) . \tag{A.22}$$

Since the solution to (A.21) is unique, and $\boldsymbol{\alpha}(z^{-1})$ and $\boldsymbol{Q}(z^{-1})$ are determined explicitly from (A.18) and (A.22), we conclude that there exists a unique solution to (A.16a) and (A.16b).

Appendix B

Proof of Theorem 2

We start with the expression (3.6), and use the assumption on correct past symbols:

$$\hat{d}(k-\ell|k) = \boldsymbol{\mathcal{R}}(q^{-1})y(k) - \boldsymbol{\mathcal{F}}(q^{-1})d(k-\ell-1) \; .$$

We then insert $y(k)$ from (3.1) and rearrange:

$$\begin{aligned} \hat{d}(k-\ell|k) &= (\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} - q^{-\ell-1}\boldsymbol{\mathcal{F}})d(k) + \boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k) \\ &= d(k-\ell) + (\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} - q^{-\ell-1}\boldsymbol{\mathcal{F}} - q^{-\ell}\mathbf{I})d(k) + \boldsymbol{\mathcal{R}}\boldsymbol{N}^{-1}\boldsymbol{M}v(k) \end{aligned} \quad \text{(B.1)}$$

Since $d(k)$ and $v(m)$ are uncorrelated for all k and m, the zero-forcing condition (3.9) can be satisfied if and only if

$$(\boldsymbol{\mathcal{R}}\boldsymbol{A}^{-1}\boldsymbol{B} - q^{-\ell-1}\boldsymbol{\mathcal{F}} - q^{-\ell}\mathbf{I}) = 0 \; ,$$

which coincides with (4.8).

Appendix C

Proof of Corollaries 1 and 2

C.1 Proof of Corollary 1

Define

$$\begin{aligned} \boldsymbol{B}_\nu^{(1)}(q^{-1}) &\triangleq \text{The first } \nu - 1 \text{ columns of } \boldsymbol{B}(q^{-1}) & \text{(C.1a)} \\ \boldsymbol{b}_\nu(q^{-1}) &\triangleq \text{Column } \nu \text{ in } \boldsymbol{B}(q^{-1}) & \text{(C.1b)} \\ \boldsymbol{B}_\nu^{(2)}(q^{-1}) &\triangleq \text{The last } n_d - \nu \text{ columns of } \boldsymbol{B}(q^{-1})\ . & \text{(C.1c)} \end{aligned}$$

Introduce the polynomial matrices $\boldsymbol{r}_\nu(q^{-1})$, $\boldsymbol{t}_\nu(q^{-1})$ and $\tilde{\boldsymbol{b}}_\nu(q^{-1})$ which satisfy the Diophantine equation

$$\boldsymbol{b}_\nu = \boldsymbol{B}_\nu^{(1)}\boldsymbol{r}_\nu + \boldsymbol{B}_\nu^{(2)}\boldsymbol{t}_\nu + q^{-\ell-1}\tilde{\boldsymbol{b}}_\nu\ . \tag{C.2}$$

In [28], it is demonstrated that when $n_d > n_y$, equation (C.2) always has a solution for some ν.

We will now demonstrate, that when $n_d > n_y$

$$\boldsymbol{R}(q^{-1}) \triangleq \begin{pmatrix} \mathbf{I}_{\nu-1} & \boldsymbol{r}_\nu(q^{-1}) & 0 \\ 0 & q^{-\ell-1} & 0 \\ 0 & \boldsymbol{t}_\nu(q^{-1}) & \mathbf{I}_{n_d-\nu} \end{pmatrix} \tag{C.3}$$

will be a right factor of $\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})$ and $q^{-\ell-1}\mathbf{I}$ but not of $q^{-\ell}\mathbf{I}$, for some $1 \leq \nu \leq n_d$.

We must thus verify three properties of $\boldsymbol{R}(q^{-1})$:

1. $\boldsymbol{R}(q^{-1})$ is a right factor of $\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})$.
2. $\boldsymbol{R}(q^{-1})$ is a right factor of $q^{-\ell-1}\mathbf{I}$.
3. $\boldsymbol{R}(q^{-1})$ is not a right factor of $q^{-\ell}\mathbf{I}$.

Property 1:

If $\boldsymbol{R}(q^{-1})$ is to be a right divisor of $\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})$, we must have

$$\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1}) = \boldsymbol{\mathcal{T}}(q^{-1})\boldsymbol{R}(q^{-1}) \tag{C.4}$$

for some stable and causal rational matrix $\boldsymbol{\mathcal{T}}(q^{-1})$. We now demonstrate that

$$\boldsymbol{\mathcal{T}}(q^{-1}) = \boldsymbol{A}^{-1}(q^{-1})\begin{pmatrix}\boldsymbol{B}_\nu^{(1)}(q^{-1}) & \tilde{\boldsymbol{b}}_\nu(q^{-1}) & \boldsymbol{B}_\nu^{(2)}(q^{-1})\end{pmatrix}$$

satisfies (C.4). Above, $\boldsymbol{B}_\nu^{(1)}(q^{-1})$ and $\boldsymbol{B}_\nu^{(2)}(q^{-1})$ are defined in (C.1a) and (C.1c) respectively whereas $\tilde{\boldsymbol{b}}_\nu(q^{-1})$ satisfies (C.2). We thus multiply $\boldsymbol{\mathcal{T}}(q^{-1})$ with $\boldsymbol{R}(q^{-1})$ given by (C.3), from the right:

$$\boldsymbol{\mathcal{T}}\boldsymbol{R} = \boldsymbol{A}^{-1}\begin{pmatrix}\boldsymbol{B}_\nu^{(1)} & \boldsymbol{B}_\nu^{(1)}\boldsymbol{r}_\nu + \boldsymbol{B}_\nu^{(2)}\boldsymbol{t}_\nu + q^{-\ell-1}\tilde{\boldsymbol{b}}_\nu & \boldsymbol{B}_\nu^{(2)}\end{pmatrix}$$

But from (C.2), $\boldsymbol{B}_\nu^{(1)}\boldsymbol{r}_\nu + \boldsymbol{B}_\nu^{(2)}\boldsymbol{t}_\nu + q^{-\ell-1}\tilde{\boldsymbol{b}}_\nu = \boldsymbol{b}_\nu$ which implies that

$$\boldsymbol{\mathcal{T}}\boldsymbol{R} = \boldsymbol{A}^{-1}\begin{pmatrix}\boldsymbol{B}_\nu^{(1)} & \boldsymbol{b}_\nu & \boldsymbol{B}_\nu^{(2)}\end{pmatrix} = \boldsymbol{A}^{-1}\boldsymbol{B}\,.$$

where the last equality follows from the definitions (C.1a)–(C.1c). Thus, $\boldsymbol{R}(q^{-1})$ is a right divisor of $\boldsymbol{A}^{-1}(q^{-1})\boldsymbol{B}(q^{-1})$.

Property 2:

One immediately sees that $\boldsymbol{R}(q^{-1})$ is a right divisor of $q^{-\ell-1}\mathbf{I}$, since

$$\begin{pmatrix} q^{-\ell-1}\mathbf{I}_{\nu-1} & -\boldsymbol{r}_\nu(q^{-1}) & 0 \\ 0 & 1 & 0 \\ 0 & -\boldsymbol{t}_\nu(q^{-1}) & q^{-\ell-1}\mathbf{I}_{n_d-\nu}\end{pmatrix}\begin{pmatrix}\mathbf{I}_{\nu-1} & \boldsymbol{r}_\nu(q^{-1}) & 0 \\ 0 & q^{-\ell-1} & 0 \\ 0 & \boldsymbol{t}_\nu(q^{-1}) & \mathbf{I}_{n_d-\nu}\end{pmatrix} = q^{-\ell-1}\mathbf{I}\,. \tag{C.5}$$

Property 3:

We will now show that it is impossible to find a stable and causal rational matrix $\boldsymbol{\mathcal{L}}(q^{-1})$ such that

$$q^{-\ell}\mathbf{I} = \boldsymbol{\mathcal{L}}(q^{-1})\boldsymbol{R}(q^{-1})\,. \tag{C.6}$$

Note that causality is required since we are only looking for solutions to (B.1) which are causal. Partition $\boldsymbol{\mathcal{L}}$ as

$$\boldsymbol{\mathcal{L}} \triangleq \begin{pmatrix}\boldsymbol{\mathcal{L}}_{11} & \boldsymbol{\mathcal{L}}_{12} & \boldsymbol{\mathcal{L}}_{13} \\ \boldsymbol{\mathcal{L}}_{21} & \mathcal{L}_{22} & \boldsymbol{\mathcal{L}}_{23} \\ \boldsymbol{\mathcal{L}}_{31} & \boldsymbol{\mathcal{L}}_{32} & \boldsymbol{\mathcal{L}}_{33}\end{pmatrix}$$

with the block widths equal to the block heights of $\boldsymbol{R}(q^{-1})$ in (C.3). If we perform the matrix multiplication $\boldsymbol{\mathcal{L}}(q^{-1})\boldsymbol{R}(q^{-1})$ using (C.3) and the partitioning of $\boldsymbol{\mathcal{L}}$, condition (C.6) gives

$$\begin{aligned}\boldsymbol{\mathcal{L}}_{11} &= q^{-\ell}\mathbf{I}_{\nu-1} \\ \boldsymbol{\mathcal{L}}_{13} &= \boldsymbol{\mathcal{L}}_{21} = \boldsymbol{\mathcal{L}}_{23} = \boldsymbol{\mathcal{L}}_{31} = 0 \\ \boldsymbol{\mathcal{L}}_{33} &= q^{-\ell}\mathbf{I}_{n_d-\nu}\,.\end{aligned}$$

If we use these relations, we also obtain

$$q^{-\ell-1}\boldsymbol{\mathcal{L}}_{12} = -q^{-\ell}\boldsymbol{r}_\nu \tag{C.7a}$$

$$q^{-\ell-1}\mathcal{L}_{22} = q^{-\ell} \tag{C.7b}$$

$$q^{-\ell-1}\boldsymbol{\mathcal{L}}_{32} = -q^{-\ell}\boldsymbol{t}_\nu\,. \tag{C.7c}$$

But (C.7b) has no causal solution $\mathcal{L}_{22}(q^{-1})$. Therefore we conclude that $\boldsymbol{R}(q^{-1})$ cannot be a right divisor of $q^{-\ell}\mathbf{I}$.

Since all three properties have been established, we conclude that (4.8) cannot have a solution when $n_d > n_y$.

C.2 Proof of Corollary 2

Assume that user ν has a bulk delay which exceeds ℓ. We will now verify that

$$\boldsymbol{R}(q^{-1}) = \operatorname{diag}\big(\underbrace{1 \;\; \ldots \;\; 1}_{\nu-1} \;\; q^{-\ell-1} \;\; \underbrace{1 \;\; \ldots \;\; 1}_{n_d-\nu}\big) \tag{C.8}$$

is a right divisor of $\boldsymbol{A}^{-1}\boldsymbol{B}$ and $q^{-\ell-1}\mathbf{I}$, but not of $q^{-\ell}\mathbf{I}$.

With the definition (4.9), we can write

$$\boldsymbol{B}(q^{-1}) = \tilde{\boldsymbol{B}}(q^{-1})\boldsymbol{D}(q^{-1}) \tag{C.9}$$

where

$$\boldsymbol{D}(q^{-1}) = \operatorname{diag}\big(q^{-\Delta_1} \quad q^{-\Delta_2} \quad \ldots \quad q^{-\Delta_{n_d}}\big)\ . \tag{C.10}$$

If we compare (C.8) and (C.10), we see that

$$\boldsymbol{D}(q^{-1}) = \tilde{\boldsymbol{D}}(q^{-1})\boldsymbol{R}(q^{-1}) \tag{C.11}$$

with

$$\tilde{\boldsymbol{D}}(q^{-1}) = \operatorname{diag}\left(q^{-\tilde{\Delta}_1} \quad q^{-\tilde{\Delta}_2} \quad \ldots \quad q^{-\tilde{\Delta}_{n_d}}\right)$$

and

$$\tilde{\Delta}_i = \begin{cases} \Delta_i & i \neq \nu \\ \Delta_\nu - \ell - 1 & i = \nu \end{cases} .$$

If we now insert $\boldsymbol{B}$ from (C.9) and $\boldsymbol{D}$ from (C.11) into $\boldsymbol{A}^{-1}\boldsymbol{B}$, we obtain

$$\boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{A}^{-1}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{D}}\boldsymbol{R}\,.$$

Since $\Delta_\nu \geq \ell + 1$, $\tilde{\boldsymbol{D}}$ is a causal rational matrix, and $\boldsymbol{R}$ is a right divisor of $\boldsymbol{A}^{-1}\boldsymbol{B}$.

We immediately see that $\boldsymbol{R}(q^{-1})$ is a right divisor of $q^{-\ell-1}\mathbf{I}$, since

$$q^{-\ell-1}\mathbf{I} = \begin{pmatrix} q^{-\ell-1}\mathbf{I}_{\nu-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & q^{-\ell-1}\mathbf{I}_{n_d-\nu} \end{pmatrix} \boldsymbol{R}\,.$$

However, if we try to find a causal rational matrix $\boldsymbol{\mathcal{L}}(q^{-1})$ such that

$$q^{-\ell}\mathbf{I} = \boldsymbol{\mathcal{L}}\boldsymbol{R}$$

we must require

$$\boldsymbol{\mathcal{L}} = \operatorname{diag}\big(\underbrace{q^{-\ell} \quad \ldots \quad q^{-\ell}}_{\nu-1} \quad q \quad \underbrace{q^{-\ell} \quad \ldots \quad q^{-\ell}}_{n_d-\nu}\big)\ .$$

Since $\boldsymbol{\mathcal{L}}$ is not causal, we conclude that $\boldsymbol{R}$ is not a right factor of $q^{-\ell}\mathbf{I}$.

Since $\boldsymbol{R}$ is a right factor of $\boldsymbol{A}^{-1}\boldsymbol{B}$ and $q^{-\ell-1}\mathbf{I}$, but not of $q^{-\ell}\mathbf{I}$, equation (4.8) lacks a solution.

Appendix D

Proof of Theorem 3

We will now derive the design equations (4.11a), (4.11b) and (4.11c). We will also show that a bank of matched filters can constitute a part of the MSE optimum GDFE if we allow the smoothing lag ℓ to tend to infinity. For this purpose, we use the GDFE to estimate the symbol transmitted at time k, using measurements up to time $k+\ell$:

$$\hat{d}(k|k+\ell) = \boldsymbol{\mathcal{R}}(q^{-1})y(k+\ell) - \boldsymbol{\mathcal{F}}(q^{-1})\tilde{d}(k-1) \; . \tag{D.1}$$

Note that (D.1) is merely a shifted version of (3.6). We now let the smoothing lag tend to infinity. We can then equivalently express the symbol estimate of the asymptotic GDFE as

$$\hat{d}^{\infty}(k) = \boldsymbol{\mathcal{R}}^{\infty}(q,q^{-1})y(k) - \boldsymbol{\mathcal{F}}(q^{-1})\tilde{d}(k-1) \; . \tag{D.2}$$

where we have defined

$$\boldsymbol{\mathcal{R}}^{\infty}(q,q^{-1}) \stackrel{\triangle}{=} \lim_{\ell\to\infty} q^{\ell}\boldsymbol{\mathcal{R}}(q^{-1}) \; . \tag{D.3}$$

Note that $\boldsymbol{\mathcal{R}}^{\infty}(q,q^{-1})$ is stable but non-causal.

Assuming correct past decisions, i.e. $\tilde{d}(m) = d(m)$ $m \leq k-1$, and using the channel model (3.1) the estimation error can be expressed as

$$\begin{aligned}
\varepsilon(k) &= d(k) - \hat{d}^{\infty}(k) \\
&= (\mathbf{I} - \boldsymbol{\mathcal{R}}^{\infty}(q,q^{-1})\boldsymbol{A}^{-1}\boldsymbol{B} + q^{-1}\boldsymbol{\mathcal{F}})d(k) \\
&- \boldsymbol{\mathcal{R}}^{\infty}(q,q^{-1})\boldsymbol{N}^{-1}\boldsymbol{M}v(k)
\end{aligned}$$

Introduce the alternate estimate $\hat{d}_a^{\infty}(k)$

$$\hat{d}_a^{\infty}(k) \stackrel{\triangle}{=} \hat{d}^{\infty}(k) + n(k)$$

where the variation $n(k)$ is based on all signals the estimate $\hat{d}^{\infty}(k)$ may be based upon:

$$n(k) \stackrel{\triangle}{=} n_1(k) + n_2(k)$$

with

$$n_1(k) \stackrel{\triangle}{=} \boldsymbol{\mathcal{G}}_1^{\infty}(q,q^{-1})y(k) \tag{D.4a}$$
$$n_2(k) \stackrel{\triangle}{=} \boldsymbol{\mathcal{G}}_2(q^{-1})d(k-1) \tag{D.4b}$$

and where $\boldsymbol{\mathcal{G}}_1^{\infty}(q, q^{-1})$ and $\boldsymbol{\mathcal{G}}_2(q^{-1})$ are arbitrary stable rational matrices. In addition, $\boldsymbol{\mathcal{G}}_2(q^{-1})$ is causal. If the estimation error obtained with (D.2) is orthogonal to any admissible variation (D.4a), (D.4b), i.e. if

$$E\varepsilon(k) n_1^H(k) = 0 \tag{D.5a}$$

$$E\varepsilon(k) n_2^H(k) = 0 \tag{D.5b}$$

then $\hat{d}_a^{\infty}(k) = \hat{d}^{\infty}(k)$ or equivalently $n(k) \equiv 0$ minimizes the MSE (3.7). We must thus assure that (D.5a) and (D.5b) are fulfilled.

To derive the design equations of the MSE optimum GDFE with asymptotically large smoothing lag, we use Parseval's relation to evaluate the cross-correlation (D.5a) in the frequency domain. Proceeding as in Appendix A, we obtain

$$\begin{aligned} E\varepsilon(k) n_1^H(k) = \frac{\lambda_d}{2\pi j} \oint \big\{ (\mathbf{I} - \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{A}^{-1} \boldsymbol{B} + z^{-1} \boldsymbol{\mathcal{F}}) \times \\ \boldsymbol{B}_* \boldsymbol{N}_* - \rho \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{N}^{-1} \boldsymbol{M} \boldsymbol{M}_* \boldsymbol{A}_* \big\} \boldsymbol{A}_*^{-1} \boldsymbol{N}_*^{-1} \boldsymbol{\mathcal{G}}_{1*}^{\infty} \frac{dz}{z} \,. \end{aligned} \tag{D.6}$$

For (D.6) to equal zero, the integrand has to be made analytic inside the unit circle. However, since $\boldsymbol{\mathcal{G}}_1^{\infty}$ is non-causal, its Laurent expansion will contain powers of z, and hence the Laurent expansion of $\boldsymbol{\mathcal{G}}_{1*}^{\infty}$ will contain powers of z^{-1}. But this implies that $\boldsymbol{\mathcal{G}}_{1*}^{\infty}$ contributes poles in the origin. The only way to ensure that the integrand is analytic within $|z| = 1$ is thus to require that

$$(\mathbf{I} - \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{A}^{-1} \boldsymbol{B} + z^{-1} \boldsymbol{\mathcal{F}}) \boldsymbol{B}_* \boldsymbol{N}_* - \rho \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{N}^{-1} \boldsymbol{M} \boldsymbol{M}_* \boldsymbol{A}_* = 0 \,. \tag{D.7}$$

For (D.5b), the cross-correlation can be simplified to yield

$$\mathbf{I} - \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{A}^{-1} \boldsymbol{B} + z^{-1} \boldsymbol{\mathcal{F}} = \boldsymbol{L}_{1*} \,. \tag{D.8}$$

for some polynomial matrix $\boldsymbol{L}_{1*}$. Insert (D.8) into (D.7):

$$\boldsymbol{L}_{1*} \boldsymbol{B}_* \boldsymbol{N}_* - \rho \boldsymbol{\mathcal{R}}^{\infty} \boldsymbol{N}^{-1} \boldsymbol{M} \boldsymbol{M}_* \boldsymbol{A}_* = 0 \,.$$

We can now express the feedforward filter as

$$\boldsymbol{\mathcal{R}}^{\infty} = \frac{1}{\rho} \boldsymbol{L}_{1*} \boldsymbol{B}_* \boldsymbol{N}_* \boldsymbol{A}_*^{-1} \boldsymbol{M}_*^{-1} \boldsymbol{M}^{-1} \boldsymbol{N} \,.$$

If we use the definitions (4.1a) and (4.1b), we obtain

$$\boldsymbol{\mathcal{R}}^{\infty} = \frac{1}{\rho} \boldsymbol{L}_{1*} \boldsymbol{\tau}_* \boldsymbol{\Gamma}_*^{-1} \boldsymbol{M}^{-1} \boldsymbol{N} \,. \tag{D.9}$$

We now insert (D.9) into (D.8) and rearrange:

$$\begin{aligned} \mathbf{I} + z^{-1} \boldsymbol{\mathcal{F}} &= \boldsymbol{L}_{1*} \left(\mathbf{I} + \frac{1}{\rho} \boldsymbol{\tau}_* \boldsymbol{\Gamma}_*^{-1} \boldsymbol{\Gamma}^{-1} \boldsymbol{\tau} \right) = \boldsymbol{L}_{1*} \left(\mathbf{I} + \frac{1}{\rho} \tilde{\boldsymbol{\Gamma}}_*^{-1} \tilde{\boldsymbol{\tau}}_* \tilde{\boldsymbol{\tau}} \tilde{\boldsymbol{\Gamma}}^{-1} \right) \\ &= \boldsymbol{L}_{1*} \tilde{\boldsymbol{\Gamma}}_*^{-1} \left(\tilde{\boldsymbol{\Gamma}}_* \tilde{\boldsymbol{\Gamma}} + \frac{1}{\rho} \tilde{\boldsymbol{\tau}}_* \tilde{\boldsymbol{\tau}} \right) \tilde{\boldsymbol{\Gamma}}^{-1} = \boldsymbol{L}_{1*} \tilde{\boldsymbol{\Gamma}}_*^{-1} \boldsymbol{\beta}_* \mathbf{W} \boldsymbol{\beta} \tilde{\boldsymbol{\Gamma}}^{-1} \end{aligned}$$

where we in the second equality used the coprime factorization (4.2) and in the last equality the spectral factorization (4.12). We now collect polynomial matrices in z^{-1} on the left hand side and polynomial matrices in z on the right hand side:

$$(\mathbf{I} + z^{-1}\boldsymbol{\mathcal{F}})\tilde{\mathbf{\Gamma}}\boldsymbol{\beta}^{-1} = \boldsymbol{L}_{1*}\tilde{\mathbf{\Gamma}}_*^{-1}\boldsymbol{\beta}_*\mathbf{W} \; . \tag{D.10}$$

The left hand side contains only powers of z^{-1} and the right hand side contains only powers of z. The only way for this equality to hold is to require that both sides equal a constant matrix. Since $\boldsymbol{\beta}$ is monic, we see that the constant term of the left hand side equals $\tilde{\Gamma}_0$, the constant term of $\tilde{\mathbf{\Gamma}}$. We must thus require

$$\begin{aligned}(\mathbf{I} + z^{-1}\boldsymbol{\mathcal{F}})\tilde{\mathbf{\Gamma}}\boldsymbol{\beta}^{-1} &= \tilde{\mathbf{\Gamma}}_0 \\ \boldsymbol{L}_{1*}\tilde{\mathbf{\Gamma}}_*^{-1}\boldsymbol{\beta}_*\mathbf{W} &= \tilde{\mathbf{\Gamma}}_0 \; ,\end{aligned}$$

or, equivalently

$$\boldsymbol{L}_{1*} = \tilde{\mathbf{\Gamma}}_0\mathbf{W}^{-1}\boldsymbol{\beta}_*^{-1}\tilde{\mathbf{\Gamma}}_* \tag{D.11}$$

$$\boldsymbol{\mathcal{F}} = z(\tilde{\mathbf{\Gamma}}_0\boldsymbol{\beta}\tilde{\mathbf{\Gamma}}^{-1} - \mathbf{I}) \; . \tag{D.12}$$

We can now insert (D.11) into (D.9) to arrive at our final expression for the asymptotic DFE filters:

$$\boldsymbol{\mathcal{R}}^{\infty} = \frac{1}{\rho}\tilde{\mathbf{\Gamma}}_0\mathbf{W}^{-1}\boldsymbol{\beta}_*^{-1}\tilde{\mathbf{\Gamma}}_*\boldsymbol{\tau}_*\mathbf{\Gamma}_*^{-1}\boldsymbol{M}^{-1}\boldsymbol{N} \tag{D.13a}$$

$$= \frac{1}{\rho}\tilde{\mathbf{\Gamma}}_0\mathbf{W}^{-1}\boldsymbol{\beta}_*^{-1}\tilde{\boldsymbol{\tau}}_*\boldsymbol{M}^{-1}\boldsymbol{N} \tag{D.13b}$$

$$\boldsymbol{\mathcal{F}} = z(\tilde{\mathbf{\Gamma}}_0\boldsymbol{\beta}\tilde{\mathbf{\Gamma}}^{-1} - \mathbf{I}) \; . \tag{D.13c}$$

If we replace z and z^{-1} with q and q^{-1} respectively, (D.13a) coincides with (4.11a), (D.13b) coincides with (4.11b) and (D.13c) coincides with (4.11c).