A Short Introduction to Adaptive Equalization

Claes Tidestav
Signals and Systems
Uppsala University

April 1998

Contents

Preface

The present text is intended to give a short introduction to adaptive equalization, appearing in cellular digital communications. This text provides the background material for one of the miniprojects in the course "Adaptive signal processing" taught to undergraduate students at the Signals and Systems Group, Uppsala University.

In this presentation, we consider discrete-time channel models and equalizers, having finite impulse response (FIR structure). We begin with a short discussion of digital communications and the mobile telephony environment. We then discuss the equalization problem. Two different types of equalizers will be presented: the transversal linear equalizer and the transversal decision feedback equalizer. Direct and indirect methods to adaptive equalization will be described. First presented is the traditional, direct way of adapting the coefficients of the filters in an equalizer directly from data. Then, it is discussed how these coefficients can be adapted indirectly by first estimating channel parameters and then based on those estimates calculate the equalizer.

1 Introduction

A radio transmission system basically consists of three parts. At one end is the transmitter. The transmitter accepts information from a source, transforms it into a form that can be transmitted and sends it over a *Radio Frequency* (RF) channel. The channel possibly distorts the transmitted signal before it reaches the receiver. It is then the receivers job to figure out what signal was transmitted, and to turn it into understandable information. If everything goes well, the information the receiver delivers should coincide with the information fed into the transmitter.

Figure 1: *Basic components of a transmission system.*

Digital communication differs from its analogue counterpart in that it can only transmit a finite number of waveforms. The information to be transmitted is typically a stream of binary digits.

The information reaching the transmitter is typically *coded*, i.e. redundant bits have been added to the message to provide protection against transmission errors. In the same way, the information that leaves the receiver must be decoded before it can be used. Neither encoding nor decoding will be discussed further, and the treatment of digital communication systems in general is very brief in these notes. For a more complete treatment of such systems, see [9].

1.1 The Air Interface of a Digital Communication System

The data that is to be transmitted in digital communication systems are typically *bits*. In the transmitter, these binary digits are divided into groups and each group of bits is mapped into a so-called *symbol*. A symbol is a real or complex number.

Every discrete symbol is transformed into a continuous-time signal by a generalized form of DA-conversion known as *pulse shaping*, which is depicted in Figure 2. In its simplest form, pulse shaping is accomplished by using a zero-order hold DA converter, which will result in a rectangular pulse shape. However, more intelligent design of the pulse shaping filter improves the spectral efficiency of the continuous-time signal.

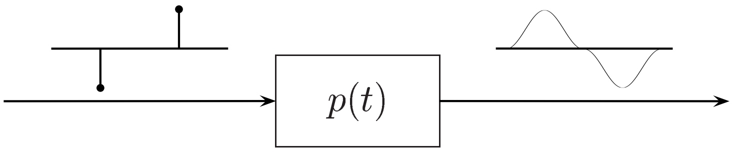


Figure 2: *Pulse shaping with the pulse shaping filter $p(t)$. In this example, the resulting pulse shape is a so-called* raised cosine *pulse.*

The continuous-time signal is called a *baseband signal*. Its spectrum is depicted in Figure 3 on the following page.

The baseband signal has a symmetric spectrum around the center frequency $f = 0$ if and only if the symbols are real. This is a well known fact from complex analysis: only real signals have spectra that are symmetric with respect to $f = 0$.

The baseband signal is then multiplied by a *carrier*, which is a high-frequency sine or cosine. After this multiplication, the signal is called a *passband signal*. This process is called

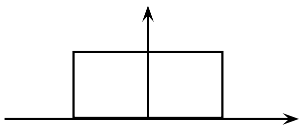
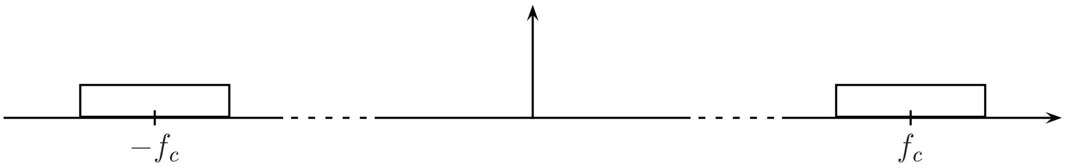


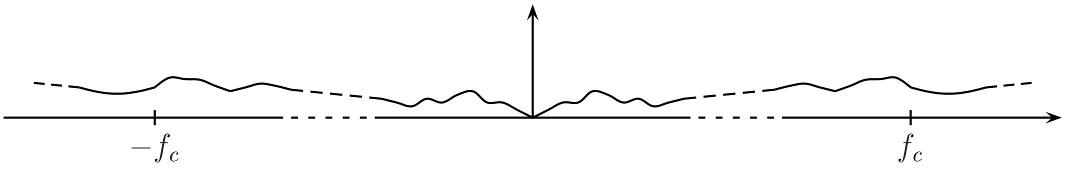
Figure 3: *The spectrum of the baseband signal. The spectrum is symmetric around $f = 0$ if and only if the symbols are real.*

modulation and has the effect that the baseband signal is shifted up in frequency, up to the passband. The center frequency of the passband signal is the frequency of the carrier. The spectrum of the transmitted passband signal is shown in Figure 4(a)

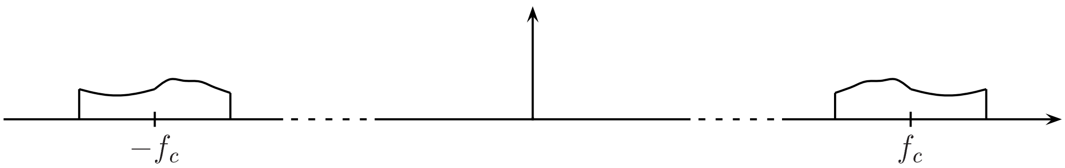
The passband signal passes through the channel. In general, the channel distorts the transmitted signal. The spectrum of the channel is shown in Figure 4(b).



(a) *Transmitted passband signal*



(b) *Channel*



(c) *Received passband signal*

At the receiver, the distorted passband signal has the spectrum shown in Figure 4(c). The receiver *demodulates* the passband signal, thereby shifting it back to the baseband with center frequency $f = 0$. The resulting baseband signal is however still distorted, as shown in Figure 4: Due to the distortion in the passband, the demodulated baseband signal has a spectrum that is

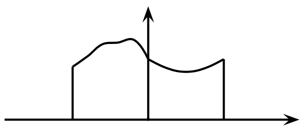


Figure 4: *Received baseband signal*

not symmetric with respect to $f = 0$. This means that the received baseband signal is complex, although the transmitted baseband signal is real. But then the transfer function from transmitted baseband signal to received baseband signal must be complex-valued! The real part of this transfer function is called the *in-phase* channel or *I-channel*, whereas the imaginary part is called the *quadrature* channel or *Q-channel*.

The fact that the spectrum is not flat means that the received signal is not white. In the time-domain, this means that the transmitted pulses will be smeared out so that the pulses corresponding to different symbols will not be separable. This phenomenon is known as *intersymbol interference* or *ISI*. The channel is said to be *(time-)dispersive*. To restore the transmitted sequence, some sort of signal processing must be performed at the receiver. This is known as *channel equalization*.

1.2 The mobile telephony environment

The mobile environment presents the system designer with some problems not previously encountered in the construction of radio systems. One important aspect is that the two parties involved in the radio transmission, the *mobile* and the *base station*, seldom have a clear line of sight between them. This has an impact both in the transmission from the base station to the mobile (the *downlink* or *forward link*) and in the transmission from the mobile to the base station (the *uplink* or *reverse link*). However, we will concentrate on the downlink problem in this presentation.

In mobile communications, the mobile receives the signal through a great number of reflexes. A good model of the scenario is shown in Figure 5, where the mobile is surrounded by many secondary transmitters. These secondary transmitters are rather small objects like buildings, trees or cars, which are located quite close to the receiver. Larger objects like mountains or large buildings not in the immediate vicinity of the receiver give rise to the different incoming rays. These incoming rays reach the cluster of secondary transmitters separated by more than a symbol period. These rays will cause the intersymbol interference and each ray is said to give rise to a *tap* in the channel. Each of these taps will have the properties described below, which are caused by the cluster of secondary transmitters.

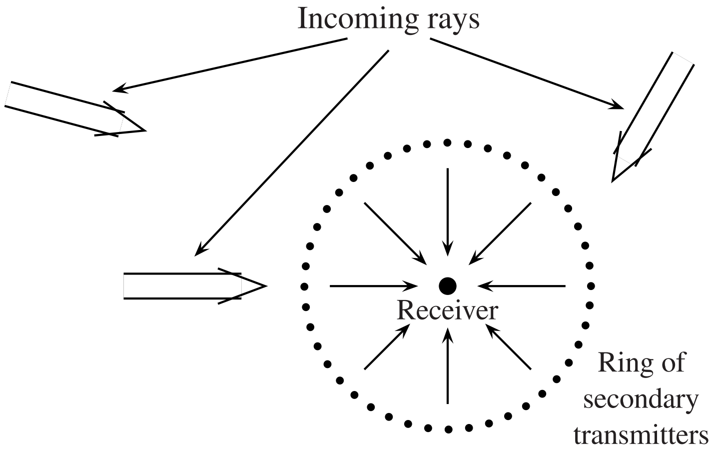


Figure 5: *The mobile in a receiving situation. The mobile receives very many reflected replicas of the transmitted signal.*

During a symbol interval, the mobile is reached by the reflexes from one incoming ray, impinging on the cluster of secondary transmitters. Due to the relative vicinity of the secondary transmitters, all the reflexes reach the mobile in a short period of time, so short that the received replicas are identical except for a difference in the phase of the carrier.

At the mobile, the signals from all the secondary transmitters are superimposed and give rise to an interference pattern. The exact nature of this interference pattern is determined by the phases of the signals coming from each of the secondary transmitters. These phase shifts in turn depend on the exact position of each scatterer. Since these positions cannot be determined, an adaptive equalizer is required to compensate for the attenuation and phase shift induced by this interference pattern. Moreover, as the mobile moves, it travels through this stationary interference pattern. This means that the attenuation and phase shift will vary stochastically in time. The rate of change depends on the speed of the mobile and the wavelength of the carrier. Typically, the tap will change completely as the mobile moves a distance corrresponding to half a carrier wavelength. With a carrier frequency of 900 MHz, which is common in today's systems, half a wavelength is 15 cm. If the mobile travels at 90 km/h, it only takes 6 ms to move 15 cm!

2 The transmission model

After the demodulation, the I- and Q-component of the continuous-time signal are sampled at the symbol rate and combined into a complex-valued measurement. Consider the received sampled data sequence $\{y(t)\}_{t=1}^{N}$[1], generated by transmission of one data *burst*, $\{d(t)\}_{t=1}^{N}$, over a time-variant and time-dispersive channel. By regarding the entire channel from the symbol sequence $d(t)$ to the sampled measurements $y(t)$ as a discrete-time system, it is possible to model the channel as

$$y(t) = H_t(q^{-1})d(t) + v(t) = h_0^t d(t) + \sum_{k=1}^{m} h_k^t d(t-k) + v(t) \quad t = 1, \ldots, N \tag{1}$$

$\{h_k^t\}_{k=0}^{m}$ is the complex-valued, time-varying impulse response of the equivalent discrete-time channel. $v(t)$ is measurement noise of zero mean, which is assumed to be independent of the symbol stream. We also make the assumption that the input stream is white, i.e. $E[d(t)d(\tau)] = 0$ for $t \neq \tau$.

Now, if the input $d(t)$ is to be recovered from $y(t)$, it is evident from the right hand side of (1) that the middle term, the intersymbol interference, and the noise $v(t)$ distorts the detection of $d(t)$ from $y(t)$. To extract the symbols, we use an equalizer. The design of the equalizer obviously depends on the impulse response of the channel. Since this is not only unknown, but also time-variant, an adaptive equalizer has to be used.

The problem of designing an equalizer is complicated by the fact that the possible regressors $\{d(t)\}$ are unknown as well. We cannot use adaptive modelling directly, since we do not know the input to the adaptive filter, and we cannot directly use inverse adaptive modelling since we do not know the desired signal. To cope with this problem, the data is transmitted in bursts of length N. A part of each transmitted burst is known at the receiver. This part, the so-called *training sequence*, can be located anywhere in the burst. For the rest of this article however, the training sequence is supposed to constitute the first N_{tr} symbols in the transmitted sequence. See Figure 6.

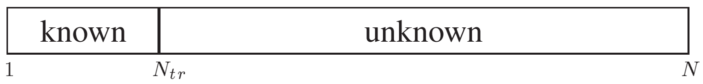


Figure 6: *Organization of the bursts*

By using the training sequence as regressors, it is thus possible to design an equalizer, which makes is possible to recover the first unknown symbols with reasonable accuracy. However, due

[1]In the following, we set $T = 1$.

to the time variation of the channel, the performance of the equalizer will deteriorate and the estimates become useless.

The solution is to use the initial, accurately estimated symbols as regressors and update the equalizer, thus enabling reliable communication during the entire burst.

It is thus natural to partition the adaptation procedure into two modes:

1. Learning-directed mode, $\{d(t), y(t)\}_{t=1}^{N_{tr}}$: The training sequence $\{d(t)\}_{t=1}^{N_{tr}}$ is utilized to initialize equalizer parameters.

2. Decision-directed mode, $\{\bar{d}(t), y(t)\}_{t=N_{tr}+1}^{N}$: Detected symbols, $\{\bar{d}(t)\}_{N_{tr}+1}^{N}$, are used as substitutes for $\{d(t)\}_{t=N_{tr}+1}^{N}$. The adaptation of the filter parameters is then based on decisioned data.

The adaptive equalization problem is depicted in Figure 7. An inherent difficulty is that adaptation has to be based on decisioned data, $\{\bar{d}(t)\}$. If incorrect decisions are made and subsequently used in the adaptation, there is a potential risk of losing tracking ability. If the tracking is lost, too many errors will occur, which in turn will make the equalizer useless. However, since a communication system is supposed to operate under such conditions that almost all decisions are correct, this is not a serious problem.

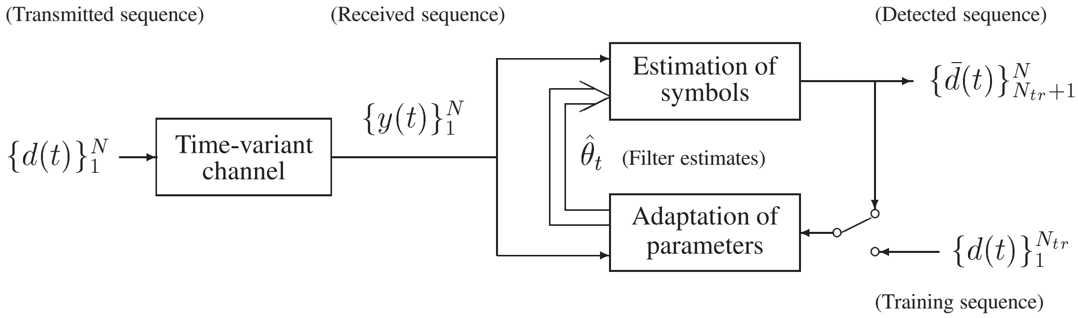


Figure 7: *A transmitted sequence of data $\{d(t)\}$, propagating through a time-variant channel, yields a received sampled sequence $\{y(t)\}$. The training sequence and corresponding data in the received sequence are used to initialize the adaptation. In decision-directed mode, the transmitted (unknown) symbols, $d(t)$, are replaced by decisions $\bar{d}(t)$ and adaptation of filter parameters works in tandem with a symbol estimator.*

Remark. In some mobile telephony systems (notably GSM[2]) the channel almost does not vary over the burst. This is because the duration of the entire burst is only 0.577 ms. During this time, the channel does not change significantly. The initial equalizer, obtained from identification using the training sequence, is then accurate enough to enable secure operation over the entire data sequence. In this case, it is not necessary for the equalizer to operate in decision-directed mode. On the other hand, in the digital telephony system presently used in North America, known as D-AMPS[3] or ADC[4], the duration of the burst is 6.8 ms. This is sufficient for the channel to change substantially during the burst.

3 Equalizer design

An equalizer is an input estimator. Since we are interested in making correct decisions, it is natural to choose an input estimator that minimizes the probability of making an error, i.e.

[2]GSM: originally Groupe Spécial Mobile, but today Global System for Mobile communication
[3]D-AMPS: Digital-Advanced Mobile Phone System
[4]ADC: American Digital Cellular

$\bar{d}(t) \neq d(t)$. Such an estimator is optimal under the so-called MAP-criterion (MAP=*Maximum A posteriori Probability*). If all values in the symbol alphabet are equally probable, this criterion is equivalent to the more familiar ML-criterion (ML=*Maximum Likelihood*). If the noise is white and gaussian, this optimal estimator computes the quantity

$$J = \sum_{t=N_{tr}+1}^{N} \left| y(t) - H_t(q^{-1})\bar{d}(t) \right|^2$$

for *all* possible input sequences $\{\bar{d}(t)\}_{t=N_{tr}+1}^{N}$ and chooses the sequence which results in the smallest J. Since the estimator makes decisions concerning a sequence of symbols rather than a single symbol, this detection scheme is called *MLSE, Maximum Likelihood Sequence Estimation*.

The MLSE can be relatively efficiently implemented using the so-called *Viterbi algorithm*[5], which also enables symbol-by-symbol detection. Still, the complexity of the detection algorithm increases exponentially with the length of the channel impulse response.

Due to the complexity of the optimal algorithm, suboptimal schemes based on linear filters can be used. An outline of such an approach is depicted in Figure 8. The filters produce an estimate $\hat{d}(t)$ of the transmitted symbol $d(t)$. This estimate is fed into a decision device to obtain a decisioned symbol $\bar{d}(t)$. The decision device selects the symbol which is closest, in Euclidean distance, to the estimate $\hat{d}(t)$. For example, if $d(t)$ takes the possible values ± 1, the decision device would simply be $\bar{d}(t) = \text{sgn}[\hat{d}(t)]$, where $\text{sgn}[\cdot]$ is the sign function. Our goal is still to minimize the probability of making an erronous decision, i.e. $\bar{d}(t) \neq d(t)$. To achieve this goal, we choose the filters to minimize the mean square error of the estimate $\hat{d}(t)$, i.e. to minimize $E[|d(t) - \hat{d}(t)|^2]$.

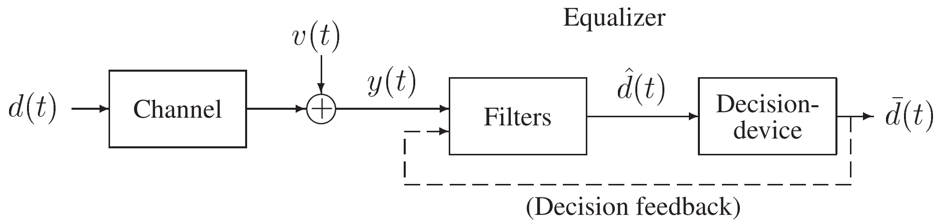


Figure 8: *The structure of an equalizer. The transmitted symbols $\{d(t)\}$ propagate through a time-variant and time-dispersive channel, having impulse response $\{h_k(t)\}_{k=0}^{m}$. The received signal $y(t)$ is corrupted by noise $v(t)$. The filters give an estimate $\hat{d}(t)$ and a decision device selects the symbol $\bar{d}(t)$.*

In the following, we shall present two types of equalizers which are often used in practice: the *transversal linear equalizer* and the *transversal decision feedback equalizer*. In contrast to the linear equalizer, the decision feedback equalizer utilizes the fact that the input attains only a finite number of values, e.g. ± 1. The linear equalizer is described also in Chapter 10 in [1]. The treatment in these notes is somewhat different to make the comparison with the decision feedback equalizer clearer and to stress the impact of the channel impulse response.

3.1 The linear transversal equalizer

The aim of a linear equalizer is to estimate $d(t)$ from (delayed) noisy measurements

$$\hat{d}(t) = C_t(q^{-1})y(t) \tag{2}$$

[5]This equalization scheme is in fact used in GSM.

As stated earlier, we wish to make the estimation error as small as possible. Let a time-variant channel model of (1) be expressed in transfer function form

$$y(t) = H_t(q^{-1})d(t) + v(t) \tag{3}$$
$$H_t(q^{-1}) = h_0^t + h_1^t q^{-1} + \cdots + h_m^t q^{-m}.$$

Substitution of $y(t)$ from (3) into (2) yields

$$\hat{d}(t) = C_t(q^{-1})H_t(q^{-1})d(t) + C_t(q^{-1})v(t).$$

The estimation error is now given by

$$\varepsilon(t) = d(t) - \hat{d}(t) = \underbrace{[1 - C_t(q^{-1})H_t(q^{-1})]d(t)}_{1} - \underbrace{C_t(q^{-1})v(t)}_{2}.$$

We observe that the estimation error consists of two terms, one term originating from the inter-symbol interference and the other from the noise. Ideally, if $H_t(q^{-1})$ is stably invertible, and there is no noise present, $C_t(q^{-1}) = 1/H_t(q^{-1})$ would make the error zero. There is however no guarantee whatsoever that $H_t(q^{-1})$ has all its zeroes inside the unit circle.

The traditional way of dealing with this problem is to use an FIR, or *transversal*, equalizer:

$$C_t(q, q^{-1}) = c_{-L}^t q^{-L} + \cdots + c_0^t + \cdots + c_L^t q^L$$

Notice that this is a *non-causal* filter, i.e. the output at time t depends not only on present and past inputs, but also on *future* ones. The non-causality is not a serious problem, since we in practise can introduce a sufficiently long delay in the estimator to make it causal. This means of course that the estimator instead can be written

$$\hat{d}(t - L) = C_t(q, q^{-1})y(t - L) \tag{4}$$

which is causal. The estimator above introduces an unwanted delay of L samples into the detection process. To estimate $d(t)$ without this delay would however be hazardous, since only one of the available measurements ($y(t)$) contains any information about $d(t)$. If h_0^t is small, this measurement could easily be destroyed by noise. The probability of correctly detecting $d(t)$ then diminishes. If more measurements ($y(t + j)$ $j = 1, \ldots, m$) are used in the detection process, the noise has to distort many measurements to cause an error. In general, the probability of error is reduced as the decision delay (or *smoothing lag*) L increases. Increasing L however means increased complexity, so the choice of filter size is a trade-off between complexity and performance.

The coefficients of $C_t(q, q^{-1})$ are determined to minimize the mean square error (MSE) of the estimate:

$$C_t(q, q^{-1}) = \arg\min E|\varepsilon(t)|^2 .$$

The optimum linear filter is derived in Appendix A.1 for white noise and a white symbol sequence with zero mean. The minimum MSE solution corresponds to solving a set of $2L + 1$ linear equations, the so-called *Wiener-Hopf equations*:

$$[\rho I + \mathcal{F}_t^T \mathcal{F}_t]\theta_c^t = h_t. \tag{5}$$

where $\theta_c^t \triangleq (c_{-L}^t \dots c_0^t \dots c_L^t)^T$, I is the identity matrix, $\rho \triangleq Ev^2(t)/Ed^2(t)$ and

$$\mathcal{F}_t \triangleq \begin{pmatrix} h_0^{t+L} & & & \mathbf{0} \\ \vdots & h_0^{t+L-1} & & \\ h_m^{t+L} & \vdots & \ddots & \\ & h_m^{t+L-1} & & h_0^{t-L} \\ & & \ddots & \vdots \\ & \mathbf{0} & & h_m^{t-L} \end{pmatrix} \qquad h_t \triangleq \begin{pmatrix} h_L^{t+L} \\ h_{L-1}^{t+L-1} \\ \vdots \\ h_0^t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The matrix $\mathcal{F}_t$ and the vector h_t are of dimension $m+2L+1|2L+1$ and $2L+1|1$, respectively.

The length of the linear transversal equalizer must often be chosen rather large if good performance is to be obtained.

3.2 The transversal Decision Feedback Equalizer (DFE)

A better equalizer is the decision feedback equalizer. It uses *past decisions* $\bar{d}(t)$ in the estimation, to remove interference from symbols which have already been detected. A transversal DFE with a finite *smoothing lag* (or decision delay) m_f has the structure

$$\hat{d}(t) = q^{m_f} S_t(q^{-1}) y(t) - R_t(q^{-1}) \bar{d}(t-1), \tag{6}$$

where

$$\begin{aligned} S_t(q^{-1}) &= s_0^t + s_1^t q^{-1} + \dots + s_{m_f}^t q^{-m_f} \\ R_t(q^{-1}) &= r_1^t + r_2^t q^{-1} + \dots + r_m^t q^{-(m-1)}. \end{aligned}$$

The smoothing lag is a design parameter, i.e. it is *chosen* by the system designer. As for the linear equalizer, the choice of smoothing lag is a trade-off between complexity and performance. Usually, $S_t(q^{-1})$ and $R_t(q^{-1})$ are referred to as the *forward filter* and the *feedback filter*, respectively. The structure of the DFE is depicted in Figure 9.

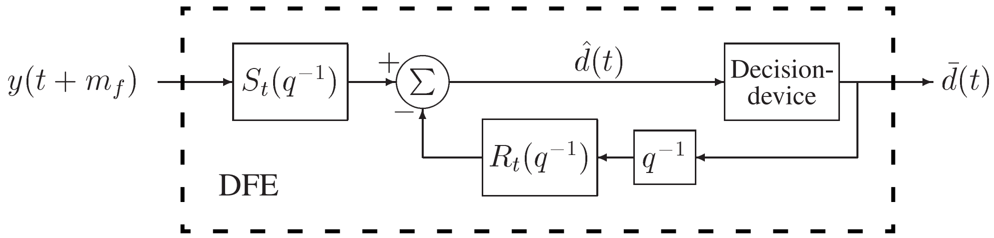


Figure 9: *The structure of a decision feedback equalizer. It consists of the filter $S_t(q^{-1})$ in the forward path, the filter $R_t(q^{-1})$ in the feedback path from $\bar{d}(t-1)$, and a decision device.*

We note from above that the order of the feedback filter is determined by the length of the channel impulse response. The order of the forward filter equals the smoothing lag. Thus, the transversal DFE is described by the following coefficient vector

$$\theta_{eq}^t = \begin{pmatrix} s_0^t & \dots & s_{m_f}^t & r_1^t & \dots & r_m^t \end{pmatrix}^T \triangleq \begin{pmatrix} \theta_s^t \\ \theta_r^t \end{pmatrix}.$$

Let us illustrate the basic idea behind the DFE. The most obvious way to utilize past decisions would be to rearrange (1), substitute decisioned data $\bar{d}(t)$ for $d(t)$ and then divide by h_0^t to obtain an estimate

$$\hat{d}(t) = \frac{1}{h_0^t}[y(t) - \sum_{k=1}^{m} h_k^t \bar{d}(t-k)]. \tag{7}$$

This estimator corresponds to setting $m_f = 0$ and choosing $S_t(q^{-1}) = 1/h_0^t$ and $R_t(q^{-1}) = 1/h_0^t\ [h_1^t + h_2^t q^{-1} + \cdots + h_m^t q^{-(m-1)}]$ in (6). Now, substitution of $y(t)$ from (1) into (7) yields

$$\hat{d}(t) = d(t) + \frac{1}{h_0^t}\sum_{k=1}^{m} h_k^t[d(t-k) - \bar{d}(t-k)] + \frac{1}{h_0^t}v(t).$$

If the past decisions $\{\bar{d}(t-k)\}_{k=1}^m$ were correct, the estimate becomes

$$\hat{d}(t) = d(t) + \frac{1}{h_0^t}v(t).$$

Clearly, this estimator removes all the intersymbol interference caused by the previously transmitted symbols $d(t-1), \ldots, d(t-m)$. In the noise-free case, we thus have $\hat{d}(t) = d(t)$. However, if h_0^t is small, the noise will be severely amplified. Again, reduction of the intersymbol interference must be balanced against noise amplification.

In Appendix A.2, the MSE optimal transversal DFE is derived. The derivation assumes white noise, a white symbol sequence and correct past decisions, $\bar{d}(t-j) = d(t-j)\ j = 1, \ldots, m$. The DFE is obtained by first solving a linear system of $m_f + 1$ equations, which gives the coefficients in the forward filter θ_s^t:

$$[\rho I + \mathcal{F}_t^T \mathcal{F}_t]\theta_s^t = h_t \tag{8}$$

where I is the identity matrix, $\rho = Ev^2(t)/Ed^2(t)$ and

$$\mathcal{F}_t \triangleq \begin{pmatrix} h_0^{t+m_f} & & & \mathbf{0} \\ h_1^{t+m_f} & h_0^{t+m_f-1} & & \\ \vdots & \vdots & \ddots & \\ h_{m_f}^{t+m_f} & h_{m_f-1}^{t+m_f-1} & \cdots & h_0^t \end{pmatrix} \quad h_t \triangleq \begin{pmatrix} h_{m_f}^{t+m_f} \\ \vdots \\ h_1^{t+1} \\ h_0^t \end{pmatrix}$$

The coefficients in the feedback filter θ_r^t are then obtained by pre-multiplying θ_s^t with a matrix $\mathcal{G}_t$:

$$\theta_r^t = \mathcal{G}_t \theta_s^t \tag{9}$$

where

$$\mathcal{G}_t \triangleq \begin{pmatrix} h_{m_f+1}^{t+m_f} & \cdots & h_2^{t+1} & h_1^t \\ \vdots & & \vdots & \vdots \\ & & h_m^{t+1} & h_{m-1}^t \\ 0 & \cdots & 0 & h_m^t \end{pmatrix}.$$

Above $h_k^t \equiv 0$ for $k > m$.

A drawback with the DFE is that single incorrect decisions may cause *error propagation*[6]. However, if the feedback filter $R_t(q^{-1})$ consists of few coefficients, that is the channel impulse response is short, the effect of single detection errors becomes small. In spite of this error propagation, the DFE normally gives much (orders of magnitude!) better performance than does the linear equalizer, when the signal-to-noise ratio is reasonably high.

4 Adaptive equalization structures

In this section, we describe two different approaches to adaptive equalization: the direct and the indirect equalization scheme. The direct scheme corresponds to the inverse modelling approach discussed in Chapter 10 in [1]. We thus use an adaptive algorithm to adjust the filter coefficients in order to minimize the difference between our desired response ($d(t)$ or $\bar{d}(t)$) and the output of the DFE ($\hat{d}(t)$).

In the indirect scheme we use adaptive modelling to estimate the impulse response of the channel. We thus adjust our adaptive filter so that its ouput matches the output of the channel when it is driven by $d(t$ or $\bar{d}(t)$. We then use (8) and (9) to design a DFE usgin the estimated channel impulse response.

The estimates produced by the equalizers discussed in the previous section can be obtained as an inner product of two vectors

$$\hat{d}(t) = \varphi_{eq}^T(t)\theta_{eq}^t,$$

where θ_{eq}^t is a vector of time-dependent equalizer parameters and $\varphi_{eq}(t)$ a vector of regressors. For the linear equalizer above we have that

$$\begin{aligned} \varphi_{eq}^T(t) &= \begin{pmatrix} y(t+L) & \dots & y(t) & \dots & y(t-L) \end{pmatrix} \\ \theta_{eq}^t &= \begin{pmatrix} c_{-L}^t & \dots & c_0^t & \dots & c_L^t \end{pmatrix}^T = \theta_c^t, \end{aligned}$$

and for the decision feedback equalizer we obtain, in decision-directed mode,

$$\begin{aligned} \varphi_{eq}^T(t) &= \begin{pmatrix} y(t+m_f) & \dots & y(t) & -\bar{d}(t-1) & \dots & -\bar{d}(t-m) \end{pmatrix} \\ \theta_{eq}^t &= \begin{pmatrix} s_0^t & \dots & s_{m_f}^t & r_1^t & \dots & r_m^t \end{pmatrix}^T = \begin{pmatrix} \theta_s^t \\ \theta_r^t \end{pmatrix}. \end{aligned}$$

To obtain decisioned symbols $\bar{d}(t)$, the estimates $\hat{d}(t)$ are fed into the decision device. Before we proceed to describe the two adaptive equalization approaches, let us make some notes about the time-varying estimation problem.

The goal in the estimation (tracking) of a time-varying system is to determine a sequence of parameter estimates, $\{\hat{\theta}^t\}_{t=1}^{t=N}$, such that the cumulative squared prediction errors are minimized,

$$\{\hat{\theta}^t\}_{t=1}^{t=N} = \arg\min_{\{\theta^t\}} \sum_{t=1}^{N} \varepsilon^2(t,\theta) \ .$$

In order to minimize this criterion, parameters of time-varying systems are often estimated by recursive algorithms having the structure

$$\hat{\theta}^t = \hat{\theta}^{t-1} + K(t)\varepsilon(t). \tag{10}$$

[6]By error propagation, we mean that a single decision error may cause subsequent erroneous decisions. If we are unlucky, it may take some time before we have recovered from such an error event.

The choice of gain vector sequence $\{K(t)\}$ differs in different algorithms. We note from (10) that $K(t)$ cannot tend to zero, and still retain tracking capability. For the particular case of a linear regression model we choose to express the gain as

$$K(t) = P(t)\varphi(t),$$

where $P(t)$ is a positive definite matrix. For more details see [3]. It should be mentioned that there exist other algorithm structures than (10), which are more suitable in some cases, see for example [2, 5, 6, 8].

4.1 Direct adaptation of the equalizer parameters

The traditional and the most straightforward way of adapting the coefficients in an equalizer, is to update the equalizer coefficient estimates, $\hat{\theta}^t_{eq}$, *directly* from data. The adjustment of the coefficients is based on recursive minimization of a function of the *input* prediction error. In training mode, this error is given by

$$\varepsilon_d(t) = d(t) - \hat{d}(t),$$

while in decision-directed mode, decisioned symbols are used as substitutes for the unknown transmitted symbols so that $\varepsilon_d(t) = \bar{d}(t) - \hat{d}(t)$. Thus, at each time instant, the equalizer parameters are updated by a recursive algorithm

$$\hat{\theta}^t_{eq} = \hat{\theta}^{t-1}_{eq} + K_{eq}(t)\varepsilon_d(t).$$

This is also known as inverse adaptive modelling, since we are trying to model the *inverse* of the channel.

4.2 Indirect adaptation of the equalizer parameters

Indirect adaptation is based on adaptive modelling of the *channel* coefficient vector

$$\theta^t_{ch} \triangleq \left(h^t_0 \ \ldots \ h^t_m\right)^T .$$

Using the updated channel parameters $\hat{\theta}^t_{ch}$, the equalizer parameters are calculated occasionally, possibly at every sample, by performing a mapping $\hat{\theta}^t_{eq} = f(\hat{\theta}^t_{ch})$. For the linear equalizer, this mapping is given by (5) and for the decision feedback equalizer, it is given by (8) and (9). The estimation is based on recursive minimization of a function of the *output* prediction error

$$\varepsilon_y(t) = y(t) - \hat{y}(t).$$

When the noise $v(t)$ in (1) is white, the one step prediction is simply

$$\begin{aligned} \hat{y}(t) &= \varphi^T_{ch}(t)\hat{\theta}^t_{ch} \\ \varphi^T_{ch}(t) &= \begin{pmatrix} d(t) & \ldots & d(t-m) \end{pmatrix} , \end{aligned}$$

where $\varphi_{ch}(t)$ is the channel regression vector. In decision-directed mode, we are forced to use decisioned symbols $\varphi^T_{ch}(t) = \left(\bar{d}(t) \ \ldots \ \bar{d}(t-m)\right)$ to replace the unknown transmitted symbols.

Thus, in this approach we first estimate the impulse response and then calculate the equalizer parameters:

$$\hat{\theta}^t_{ch} = \hat{\theta}^{t-1}_{ch} + K_{ch}(t)\varepsilon_y(t)$$
$$\hat{\theta}^t_{eq} = f(\hat{\theta}^t_{ch}).$$

The calculation of "optimal" equalizer parameters was described in Section 3. Here we substitute estimates $\hat{\theta}^t_{ch}$ for the true channel coefficients. For the decision feedback equalizer, we have to solve (8) and (9) possibly at each time instant:

$$[\hat{\rho}I + \widehat{\mathcal{F}}^T_t \widehat{\mathcal{F}}_t]\hat{\theta}^t_s = \hat{h}^t_s \qquad \hat{\theta}^t_r = \widehat{\mathcal{G}}_t\hat{\theta}^t_s.$$

The indirect approach is illustrated in Figure 10. In contrast to the direct approach, this approach requires the noise power $Ev^2(t)$ to be estimated.

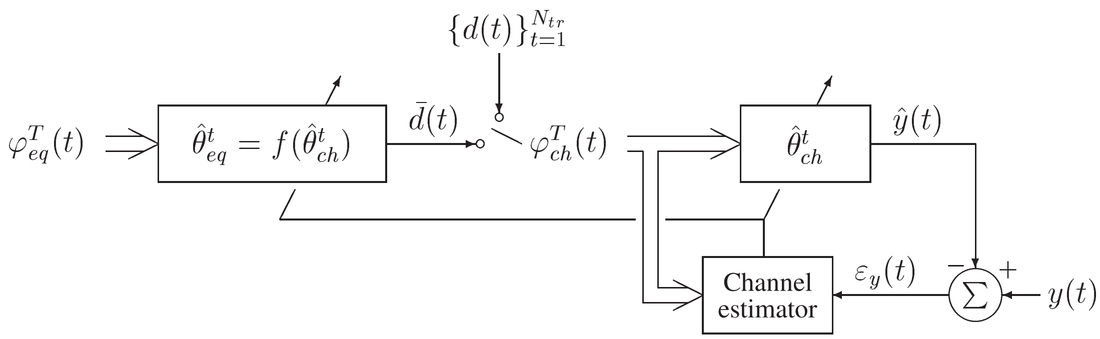


Figure 10: *The structure of an indirect adaptive equalizer. At each time instant, the channel model coefficients are updated via the estimate $\hat{\theta}^t_{ch}$, produced by the channel estimator. Inputs to the channel estimator are the regression vector $\varphi_{ch}(t)$ and the output prediction error $\varepsilon_y(t)$, respectively. Based on this channel estimate, the equalizer parameters are updated by making the mapping $\hat{\theta}^t_{eq} = f(\hat{\theta}^t_{ch})$.*

4.3 Comparison

From the discussion above, it follows that the most straightforward approach to adaptive equalization is to use the direct form. However, this approach might lead to a very difficult tracking problem. A closer examination of the (MSE optimal) equalizer parameters, derived and discussed in Section 3, indicates a non-linear relationship between the parameters and the channel coefficients. This follows from the fact that an equalizer attains channel inversion in some way. The typical relation is $\sim 1/h^t_0$, where h^t_0 is the direct term in the channel impulse response. A problem arises when h^t_0 passes zero. Then, large and rapid changes in the equalizer parameters occur. These parameter "jumps" are the reason why direct adaptation of equalizer parameters could result in a very difficult tracking problem. This problem is avoided if the equalizer parameters are updated via channel estimates. The channel coefficients might change rapidly, but normally smoothly, without large jumps. The following example illustrates a scenario of how the "optimal" equalizer parameters may change.

Example 1: Consider a noise-free channel model with two coefficients $y(t) = h_0^t d(t) + h_1^t d(t-1)$. The parameters in the considered DFE are given by

$$s_0^t = \frac{1}{h_0^t} \qquad r_1^t = \frac{h_1^t}{h_0^t},$$

and correspond to the estimator (7) for $m = 1$. Figure 11 shows the changes of the channel coefficients and the corresponding changes of the equalizer parameters. It is evident that accu-

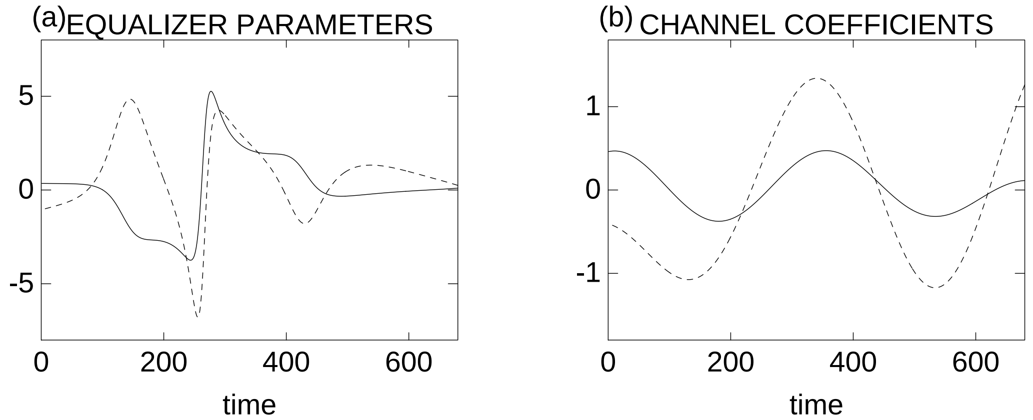


Figure 11: *A segment of the time-variant channel, h_0^t, h_1^t (solid,dashed), and the corresponding DFE parameters s_0^t, r_1^t (solid,dashed).*

rate tracking of equalizer parameters from noisy data would imply a more intricate tuning of a recursive algorithm compared to tracking of the channel coefficients. ■

Another drawback with the direct approach is that the size of the parameter vector θ_{eq}^t is often (much) larger than θ_{ch}^t. A large parameter vector will decrease the accuracy of the parameter estimates. This is true, in particular, in adaptation of rapidly time-varying parameters. Since a larger smoothing lag means a larger prefilter, for the direct appoach, performance is not necessarily improved by increasing the smoothing lag.

The direct method however also has a number of advantages. First, it is very simple to implement. Second, the equalizer coefficients are obtained directly from data, eliminating the need for a separate mapping from the channel coefficients to the equalizer coefficients. Third, since the algorithm does not assume any model of the channel, it is very unsensitive to undermodelling. This can easily happen due to the assumption of the noise $v(t)$ being white. In practise, the noise is often colored. Then the indirect algorithms derived in the appendices are not optimal anymore. For the direct method, on the other hand, the equalizer parameters converge to the optimal ones, taking the color of the noise into account.

References

[1] B Widrow and S Stearns, *Adaptive Signal Processing*. Prentice-Hall, Upper Saddle River, NJ, 1985

[2] A Benveniste, M Métivier and P Priouret, *Adaptive Algorithms and Stochastic Approximations*. Springer–Verlag, Berlin Heidelberg, 1990.

[3] S Haykin, *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ, 1991.

[4] W C Jakes, *Microwave Mobile Communications*. Wiley, New York, NY, 1974.

[5] G Kubin, "Adaptation in rapidly time–varying environments using coefficient filters," *Proceedings of ICASSP*, April 1991, Toronto, Ontario, Canada, vol. 3, pp. 2097–2100.

[6] L Lindbom, *A Wiener filtering approach to the design of tracking algorithms with applications in mobile radio communications*. PhD thesis, Department of technology, Uppsala University, Sweden, 1995.

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

[8] L Ljung and S Gunnarsson, "Adaptation and tracking in system identification – A survey," *Automatica*, vol. 26, pp. 7–21, 1990.

[9] J G Proakis, *Digital Communications*. McGraw Hill, New York, NY, 1984.

[10] S U H Qureshi, "Adaptive equalization," *Proceedings of IEEE*, vol. 73, pp. 1349–1387, 1985.

[11] L Ahlin and J Zander *Digital radiokommunikation*. Studentlitteratur, Lund, Sweden, 1992.

Appendix A

A.1 Derivation of the MSE optimal transversal linear equalizer

We want to determine the coefficients of filter $C_t(q, q^{-1})$ so that the mean square error of the estimate (4) is minimized. The estimate can equivalently be written

$$\hat{d}(t) = \sum_{i=-L}^{L} c_i^t y(t-i) = (\theta_c^t)^T \mathrm{y}_t \ , \tag{A.1}$$

where

$$\begin{aligned} \mathrm{y}_t &= \begin{pmatrix} y(t+L) & y(t+L-1) & \dots & y(t-L) \end{pmatrix}^T \\ \theta_c^t &= \begin{pmatrix} c_{-L}^t & c_{-L+1}^t & \dots & c_L^t \end{pmatrix}^T \ . \end{aligned}$$

The MSE optimal estimator is the one which minimizes

$$E[|d(t) - \hat{d}(t)|^2] \ .$$

The MSE is minimized if it is uncorrelated with all signals the estimate may be based upon, i.e. $\{y(t+L), y(t+L-1), \dots y(t-L)\}$, or equivalently if

$$E[(d(t) - \hat{d}(t))\mathrm{y}_t] = 0 \ . \tag{A.2}$$

Inserting (A.1) into (A.2) gives the so-called *normal equations* of this problem

$$E[\mathrm{y}_t \mathrm{y}_t^T]\theta_c^t = E[\mathrm{y}_t d(t)] \ . \tag{A.3}$$

To arrive at an expression for the equalizer coefficients, we express the outputs y_t in the inputs $d(t)$, the channel coefficients h_k^t and the noise $v(t)$. From (1), $y(t+\ell), \ -L \leq \ell \leq L$ can be expressed as

$$\begin{aligned} y(t+\ell) &= \sum_{k=0}^{m} h_k^{t+\ell} d(t+\ell-k) + v(t+\ell) \\ &= \begin{pmatrix} h_0^{t+\ell} & h_1^{t+\ell} & \dots h_m^{t+\ell} \end{pmatrix} \begin{pmatrix} d(t+\ell) \\ d(t+\ell-1) \\ \vdots \\ d(t+\ell-m) \end{pmatrix} + v(t+\ell) \ . \end{aligned} \tag{A.4}$$

Using (A.4), y_t can be expressed in matrix notation as

$$\begin{pmatrix} y(t+L) \\ y(t+L-1) \\ \vdots \\ y(t-L) \end{pmatrix} =$$

$$= \underbrace{\begin{pmatrix} h_0^{t+L} & \dots & h_m^{t+L} & 0 & \dots & 0 \\ 0 & h_0^{t+L-1} & \dots & h_m^{t+L-1} & 0 \quad \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_0^{t-L} & \dots & h_m^{t-L} \end{pmatrix}}_{\mathcal{F}_t^T} \times$$

$$\begin{pmatrix} d(t+L) \\ \vdots \\ d(t) \\ d(t-1) \\ \vdots \\ d(t-L) \\ \vdots \\ d(t-L-m) \end{pmatrix} + \begin{pmatrix} v(t+L) \\ v(t+L-1) \\ \vdots \\ v(t-L) \end{pmatrix} ,$$

which can be compactly written as

$$\mathrm{y}_t = \mathcal{F}_t^T \mathrm{d}_t + \mathrm{v}_t \,, \tag{A.5}$$

where

$$\mathrm{d}_t = \begin{pmatrix} d(t+L) & \dots & d(t) & d(t-1) & \dots & d(t-L) & \dots & d(t-L-m) \end{pmatrix}^T$$

and

$$\mathrm{v}_t = \begin{pmatrix} v(t+L) & v(t+L-1) & \dots & v(t-L) \end{pmatrix}^T \,.$$

By inserting (A.5), the normal equations (A.3) become

$$E[(\mathcal{F}_t^T \mathrm{d}_t + \mathrm{v}_t)(\mathrm{d}_t^T \mathcal{F}_t + \mathrm{v}_t^T)]\theta_c^t = E[(\mathcal{F}_t^T \mathrm{d}_t + \mathrm{v}_t) d(t)] \,. \tag{A.6}$$

If this is expanded, four terms are obtained. However, only two of them are non-zero, since $d(t)$ and $v(\tau)$ are uncorrelated for all t and τ. Equation (A.6) is thus reduced to

$$\mathcal{F}_t^T E[\mathrm{d}_t \mathrm{d}_t^T] \mathcal{F}_t + E[\mathrm{v}_t \mathrm{v}_t^T] = \mathcal{F}_t^T E[\mathrm{d}_t d(t)]$$

Suppose

$$E[\mathrm{d}_t \mathrm{d}_t^T] = E[d^2(t)] \times \mathrm{I}$$
$$E[\mathrm{v}_t \mathrm{v}_t^T] = E[v^2(t)] \times \mathrm{I}$$

and define

$$\rho \triangleq \frac{E[v^2(t)]}{E[d^2(t)]} \,.$$

Then

$$\begin{aligned}(\mathcal{F}_t^T\mathcal{F}_t+\rho\mathrm{I})\theta_c^t &= \mathcal{F}_t^T\begin{pmatrix}0 & \dots & 0 & 1 & 0 & \dots & 0\end{pmatrix}^T \\ &= \begin{pmatrix}h_L^{t+L} & h_{L-1}^{t+L-1} & \dots & h_0^t & 0 & \dots & 0\end{pmatrix}^T \\ &= h_c^t\end{aligned}$$

which is (5).

A.2 Derivation of the MSE optimal transversal DFE

The estimator (6) can equivalently be written as

$$\hat{d}(t)=\sum_{i=0}^{m_f}s_i^t y(t+m_f-i)-\sum_{j=1}^{m}r_j^t\bar{d}(t-j)=(\theta_s^t)^T\mathrm{y}_t-(\theta_r^t)^T\bar{\mathrm{d}}_t\;, \tag{A.7}$$

where

$$\begin{aligned}\mathrm{y}_t&=\begin{pmatrix}y(t+m_f) & y(t+m_f-1) & \dots & y(t)\end{pmatrix}^T\\ \bar{\mathrm{d}}_t&=\begin{pmatrix}\bar{d}(t-1) & \bar{d}(t-2) & \dots & \bar{d}(t-m)\end{pmatrix}^T\\ \theta_s^t&=\begin{pmatrix}s_0^t & \dots & s_{m_f}^t\end{pmatrix}^T\\ \theta_r^t&=\begin{pmatrix}r_1^t & \dots & r_m^t\end{pmatrix}^T\;.\end{aligned}$$

The MSE optimal estimator is the one that minimizes

$$E[|d(t)-\hat{d}(t)|^2]\;. \tag{A.8}$$

Again, the estimation error $\varepsilon(t)=d(t)-\hat{d}(t)$ must be uncorrelated with all signals we are allowed to base our estimate upon, i.e. $y(t+m_f),\dots,y(t),-\bar{d}(t-1),\dots,-\bar{d}(t-m)$ or equivalently

$$E\left[(d(t)-\hat{d}(t))\begin{pmatrix}\mathrm{y}_t\\ \bar{\mathrm{d}}_t\end{pmatrix}\right]=0 \tag{A.9}$$

Inserting (A.7) into (A.9) gives the normal equations for this problem

$$E\left[\begin{pmatrix}\mathrm{y}_t\\ -\bar{\mathrm{d}}_t\end{pmatrix}\begin{pmatrix}\mathrm{y}_t^T & -\bar{\mathrm{d}}_t^T\end{pmatrix}\right]\begin{pmatrix}\theta_s^t\\ \theta_r^t\end{pmatrix}=E\left[\begin{pmatrix}\mathrm{y}_t\\ -\bar{\mathrm{d}}_t\end{pmatrix}d(t)\right]\;. \tag{A.10}$$

Assume that *all previous decision are correct* (i.e. assume that $\bar{d}(t-j)=d(t-j),\ 1\le j\le m$) and define

$$\mathrm{d}_t=\begin{pmatrix}d(t-1) & d(t-2) & \dots & d(t-m)\end{pmatrix}^T\;.$$

The normal equations can then be rewritten

$$E\left[\begin{pmatrix}\mathrm{y}_t\\ -\mathrm{d}_t\end{pmatrix}\begin{pmatrix}\mathrm{y}_t^T & -\mathrm{d}_t^T\end{pmatrix}\right]\begin{pmatrix}\theta_s^t\\ \theta_r^t\end{pmatrix}=E\left[\begin{pmatrix}\mathrm{y}_t\\ -\mathrm{d}_t\end{pmatrix}d(t)\right]\;. \tag{A.11}$$

Use (A.4) with $0 \leq \ell \leq m_f$ to obtain a channel model for the vector output y_t:

$$\begin{pmatrix} y(t+m_f) \\ y(t+m_f-1) \\ \vdots \\ y(t) \end{pmatrix} =$$

$$= \underbrace{\begin{pmatrix} h_0^{t+m_f} & \dots & h_m^{t+m_f} \\ 0 & h_0^{t+m_f-1} & \dots \\ \vdots & \ddots & \ddots \\ 0 & \dots & h_0^t \end{pmatrix}}_{\mathcal{F}_t^T} \underbrace{\begin{pmatrix} 0 & \dots & & 0 \\ h_m^{t+m_f-1} & 0 & \dots & 0 \\ \ddots & \ddots & & \vdots \\ h_1^t & \dots & & h_m^t \end{pmatrix}}_{\mathcal{G}_t^T} \times \begin{pmatrix} d(t+m_f) \\ \vdots \\ d(t) \\ \vdots \\ d(t-m) \end{pmatrix} + \begin{pmatrix} v(t+m_f) \\ v(t+m_f-1) \\ \vdots \\ v(t) \end{pmatrix} .$$

Above, $\mathcal{F}_t^T$ is of dimension $m_f+1|m_f+1$, whereas $\mathcal{G}_t^T$ is of dimension $m_f+1|m$.
This implies that y_t can be written

$$\mathrm{y}_t = \mathcal{F}_t^T \mathrm{d}_t' + \mathcal{G}_t^T \mathrm{d}_t + \mathrm{v}_t \,, \tag{A.12}$$

where

$$\mathrm{d}_t' = \begin{pmatrix} d(t+m_f) & d(t+m_f-1) & \dots & d(t) \end{pmatrix}^T$$

and

$$\mathrm{v}_t = \begin{pmatrix} v(t+m_f) & v(t+m_f-1) & \dots & v(t) \end{pmatrix}^T .$$

Inserting (A.12) into the normal equations (A.11) results in

$$\begin{aligned} E\left[\begin{pmatrix} \mathcal{F}_t^T \mathrm{d}_t' + \mathcal{G}_t^T \mathrm{d}_t + \mathrm{v}_t \\ -\mathrm{d}_t \end{pmatrix} \begin{pmatrix} {d'}_t^T \mathcal{F}_t + \mathrm{d}_t^T \mathcal{G}_t + \mathrm{v}_t^T & -\mathrm{d}_t^T \end{pmatrix} \right] \begin{pmatrix} \theta_s^t \\ \theta_r^t \end{pmatrix} \\ = E\left[\begin{pmatrix} \mathcal{F}_t^T \mathrm{d}_t' + \mathcal{G}_t^T \mathrm{d}_t + \mathrm{v}_t \\ -\mathrm{d}_t \end{pmatrix} d(t) \right] . \end{aligned} \tag{A.13}$$

Both $d(t)$ and $v(t)$ are white sequences and mutually uncorrelated, i.e. $E[d(t)v(\tau)] = 0 \;\forall\; t, \tau$. Using this, the following identities are easily established

$$\begin{aligned} E[\mathrm{d}_t \mathrm{d}_t^T] &= E[d^2(t)] \times I \\ E[\mathrm{d}_t' {d'}_t^T] &= E[d^2(t)] \times I \\ E[\mathrm{d}_t' \mathrm{d}_t^T] &= 0 \\ E[\mathrm{d}_t d(t)] &= 0 \\ E[\mathrm{d}_t' d(t)] &= E[d^2(t)] \begin{pmatrix} 0 & \dots & 0 & 1 \end{pmatrix}^T \\ E[\mathrm{v}_t \mathrm{v}_t^T] &= E[v^2(t)] \times I. \end{aligned}$$

Inserting this into (A.13), the result is

$$\begin{aligned} \begin{pmatrix} E[d^2(t)]\mathcal{F}_t^T \mathcal{F}_t + E[d^2(t)]\mathcal{G}_t^T \mathcal{G}_t + E[v^2(t)]I & -E[d^2(t)]\mathcal{G}_t^T \\ -E[d^2(t)]\mathcal{G}_t & E[d^2(t)]I \end{pmatrix} \begin{pmatrix} \theta_s^t \\ \theta_r^t \end{pmatrix} \\ = E[d^2(t)]\mathcal{F}_t^T \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix}^T . \end{aligned} \tag{A.14}$$

Use the second row in (A.14) to get an expression for the feedback filter:

$$\theta_r^t = \mathcal{G}_t \theta_s^t \ . \tag{A.15}$$

The first equation in (A.14) gives

$$\begin{aligned}(E[d^2(t)]\mathcal{F}_t^T\mathcal{F}_t + E[v^2(t)]I)\theta_s^t + E[d^2(t)]\mathcal{G}_t^T(\mathcal{G}_t\theta_s^t - \theta_r^t))& \\ = E[d^2(t)]\mathcal{F}_t^T \begin{pmatrix}0 & \dots & 0 & 1\end{pmatrix}^T \ .&\end{aligned} \tag{A.16}$$

Due to the choice (A.15), the second term in (A.16) vanishes. Insert $\rho \stackrel{\triangle}{=} E[v^2(t)]/E[d^2(t)]$ into (A.16). This gives the design equations for the feedforward filter coefficients

$$(\mathcal{F}_t^T\mathcal{F}_t + \rho I)\theta_s^t = \mathcal{F}_t^T \begin{pmatrix}0 \\ \vdots \\ 0 \\ 1\end{pmatrix} \ .$$

By defining $h_t = \mathcal{F}_t^T \begin{pmatrix}0 & 0 & \dots & 0 & 1\end{pmatrix}^T$, (8) is obtained.