Jacob Benesty
Yiteng (Arden) Huang

Adaptive Signal Processing: Application to Real-World Problems

SPIN Springer's internal project number, if known

Springer

Berlin Heidelberg New York
Barcelona Budapest Hong Kong
London Milan Paris
Santa Clara Singapore Tokyo

Contents

1 Algorithms for Adaptive Equalization in Wireless Applications

Markus Rupp[1] and Andreas Burg[2]

[1] TU Wien
Institute for Communication and RF Engineering
Gusshausstr. 25/389
A-1040 Vienna, Austria
E-mail: mrupp@nt.tuwien.ac.at
[2] ETH Zurich
Integrated Systems Laboratory
Gloriastr. 35
CH-8092 Zurich, Switzerland
E-mail: apburg@iis.ee.ethz.ch

Abstract. Since the introduction of adaptive equalizers in digital communication systems by Lucky [1], much progress has been made. Due to their particular constraints many new and different concepts in the wireless domain have been proposed. The wireless channel is typically time and frequency dispersive, making it difficult to use standard equalizer techniques. Also, due to its time varying nature, long transmission bursts may get corrupted and require a continuous tracking operation. Thus, transmission is often performed in short bursts, allowing only a limited amount of training data. Furthermore, quite recently, advantages of the multiple-input multiple-output character of wireless channels have been recognized. This chapter presents an overview of equalization techniques in use and emphasizes the particularities of wireless applications.

1.1 Introduction

Consider the following simplified problem: a linear channel, characterized by a time-discrete FIR filter function with L_C coefficients

$$C(q^{-1}) = \sum_{l=0}^{L_C-1} c_l q^{-l}, \tag{1.1}$$

where q^{-1} denotes the delay operator[1] and $c_l \in \mathbb{C}$ the coefficients. It is fed by transmitted symbols $s(k) \in \mathcal{A}$ from a finite alphabet[2] $\mathcal{A} \in \mathbb{C}$. The received

[1] Note that the operator style is utilized with shift operator q^{-1} throughout the chapter since it does not require the existence of a z-transform and thus can also be applied to noise sequences. Note also that the description of signals and systems is purely discrete-time assuming that equivalent discrete-time counterparts to continuous-time signals and systems exist.

[2] The transmitted signal $s(k)$ is assumed to be normalized such that $\mathrm{E}[|s(k)|^2] = 1$.

sequence $r(k) \in \mathbb{C}$ is given by the convolution of the transmitted symbols $s(k)$ with the channel filter and additive noise $v(k)$:

$$r(k) = C(q^{-1})[s(k)] + v(k) = \sum_{l=0}^{L_C-1} c_l s(k-l) + v(k). \tag{1.2}$$

Clearly, the coefficients of the channel do not only provide a means for transmitting the symbols, but also cause a linear distortion due to the time-dispersive nature of the wireless channel, called *inter-symbol-interference* (ISI). For the moment, its frequency-dispersive character is neglected and will be considered later. If a situation can be established in which only one coefficient exists, the decoding process can be simplified. Thus, the question arises whether there exists a linear filter that can guarantee the re-establishment of

$$F_D(q^{-1})C(q^{-1}) = g_D q^{-D} \tag{1.3}$$

with a finite delay D and $g_D \neq 0$. Following such an approach, the additive noise is neglected and a filter $F_D(q^{-1})$ of length L_F is desired. Such a criterion is called *zero-forcing* (ZF) since it forces all coefficients but one of the resulting filter to zero. The remaining coefficient g_D can be set to one due to the linear nature of the equalization filter $F_D(q^{-1})$ and will no longer be used explicitly in the context of ZF. So far, the noise component has been neglected. However, it impacts the decision by the filtered value $F_D(q^{-1})[v(k)]$, which can result in a noise enhancement. Therefore, it appears a better approach to take the noise into account when searching for the optimal filter $F_D(q^{-1})$, which can be obtained by minimizing the so called *mean square error* (MSE), i.e.,

$$\text{MSE} \stackrel{\Delta}{=} E\left[\left|q^{-D}s(k) - F_D(q^{-1})r(k)\right|^2\right]. \tag{1.4}$$

Minimizing (1.4) leads to the *minimum mean square error* (MMSE). Neither solution, (1.3) or (1.4) is straightforward to obtain. Depending on the equalizer structure and its length, different solutions may occur. The following Sect. 1.2 discusses criteria for optimal equalizer performance. For a time-invariant channel it is sufficient to compute the equalizer solution just once at the initialization phase of a transmission. Algorithms for this case will be addressed in Sect. 1.3 including fractionally spaced equalizers, *multiple-input multiple-output* (MIMO) channel equalizers and decision feedback structures. Adaptive algorithms based either on estimating the channel and noise variance or without explicitly knowing its impulse response can achieve the initial training of such structures. The most famous *least-mean-square* LMS algorithm is introduced for ZF and MMSE solutions and its implications are discussed in Sect. 1.4. With some modification such adaptive filters also perform well for time-variant channels as long as the rate of change is relatively slow and training is performed periodically.

Other techniques are presented in Sect. 1.6 that try to find the *most likely* transmitted sequence by a so-called *maximum likelihood* (ML) technique. This requires a-priori knowledge of the channel impulse response and thus Sect. 1.5 gives an overview of estimation techniques for time-variant channels and in MIMO settings. Finally, Sect. 1.7 presents a short overview of blind techniques as they are being applied today in wireless communications.

1.2 Criteria for Equalization

In data communications systems and in particular in wireless transmission the best criterion to compare the equalizer performance is either the resulting *bit-error-rate* (BER) or *symbol-error-rate* (SER). Unfortunately, such measures are usually not available and other performance metrics need to be utilized instead.

Next to the *signal to interference plus noise ratio* (SINR) at the equalizer output, the MMSE measure already mentioned in the previous section is the most common measure and will be considered first. Substituting (1.2) into the definition of the MSE (1.4) yields for the equalizer output

$$\text{MMSE} = \min_{F_D(q^{-1})} E\underbrace{\left[\left|\left\{q^{-D} - F_D(q^{-1})C(q^{-1})\right\} s(k)\right|^2\right]}_{\text{ISI part}} + E\underbrace{\left[\left|F_D(q^{-1})v(k)\right|^2\right]}_{\text{noise part}}. \tag{1.5}$$

The first part presents the remaining ISI while the second is the additive noise. Without equalization the SINR at the detector is given by

$$\text{SINR} = \frac{|c_D|^2}{\sum_{i=0,i\neq D}^{L_C-1} |c_i|^2 + \sigma_v^2}, \tag{1.6}$$

where the index $0 \leq D < L_C$ indicates the channel coefficient on which the signal is detected (typically the strongest one). Clearly, there is signal energy in the term $\sum_{i=0,i\neq D}^{L_C-1} |c_i|^2$ but without additional means it cannot be used as such and appears as a disturbance to the signal. Define the convolution of the channel and a finite length filter function by a new polynomial $G(q^{-1}) = F_D(q^{-1})C(q^{-1})$ with the coefficients $g_i;\ i = 0, ..., L_C + L_F - 2 = 0, ..., L_G - 1$. The impact of ISI after equalization can be described by either of the two following measures (note that $0 \leq D < L_G$ and in general $g_D \neq 1$)

$$\text{ISI} = \frac{\sum_{i=0}^{L_G-1} |g_i|^2}{|g_D|^2} - 1, \tag{1.7}$$

$$\text{PD} = \frac{\sum_{i=0}^{L_G-1} |g_i|}{|g_D|} - 1. \tag{1.8}$$

The second metric is called *peak distortion* (PD) measure. The convolution of channel and equalizer filter results in a new $\mathrm{SINR}_{\mathrm{eq}}$

$$\mathrm{SINR}_{\mathrm{eq}} = \frac{|g_D|^2}{\sum_{i=0,i\neq D}^{L_G-1}|g_i|^2 + \sigma_v^2\sum_{i=0}^{L_F-1}|f_i|^2} = \frac{1}{\mathrm{ISI} + \frac{\sigma_v^2}{|g_D|^2}\sum_{i=0}^{L_F-1}|f_i|^2}. \quad (1.9)$$

The problem is to find equalizer filter values $\{f_i\}$ such that (1.9) is maximized. Both criteria, MSE and SINR are related by

$$\mathrm{MSE} = \frac{|g_D|^2}{\mathrm{SINR}_{\mathrm{eq}}} + |1-g_D|^2, \quad (1.10)$$

which simply becomes $\mathrm{MSE}_{\mathrm{ZF}} = 1/\mathrm{SNR}_{\mathrm{eq}}$ in the ZF case. As mentioned before, the BER measure is the best criterion, however, it is very difficult to obtain. In the simple case of binary signaling (BPSK) or quadrature phase shift keying (QPSK) with Gray coding over an AWGN channel with constant gain c_D and delay D, the BER can be determined by evaluating the expression

$$\mathrm{BER}_{\mathrm{AWGN,B/QPSK}} = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{|c_D|^2}{2\sigma_v^2}}\right). \quad (1.11)$$

For other modulation schemes at least an upper bound of the form

$$\mathrm{BER}_{\mathrm{AWGN}} \leq K\mathrm{erfc}\left(\sqrt{\delta\mathrm{SNR}}\right) \quad (1.12)$$

exists [2], where K and δ are constants depending on the modulation scheme. SNR stands for *signal to noise ratio*, i.e., the SINR without interference. Once ISI is present, the BER measure can be modified to

$$\mathrm{BER}_{\mathrm{ISI}} = \frac{1}{2}\sum_{i=1}^{I} p_i(\boldsymbol{s}_i)\mathrm{erfc}\left(\sqrt{\frac{|c_D + c_{\mathrm{ISI}}(\boldsymbol{s}_i)|^2}{2\sigma_v^2}}\right), \quad (1.13)$$

where for all $I = P^{L_C-1}$ possibilities with probability $p_i(\boldsymbol{s}_i)$ signal corruption is caused by $c_{\mathrm{ISI}}(\boldsymbol{s}_i)$. The vectors $\boldsymbol{s}_i$ contain all possible combinations of I transmitted symbols. The formula is for QAM transmission with equidistant symbols and is only correct as long as the ISI is small enough ($\max|c_{\mathrm{ISI}}(\boldsymbol{s}_i)| < |c_D|$). For example, with BPSK the values $c_D + c_{\mathrm{ISI}}(\boldsymbol{s}_i)$ must remain positive and for QPSK they must remain in the first quarter of the complex plane. The value of P is 2 for BPSK and 4 for QPSK and can be very large, depending on the size of the symbol alphabet. Clearly, with a large number of coefficients the complexity of such an expression becomes very high. Applying a linear equalizer, the BER reads

$$\mathrm{BER}_{\mathrm{ISI,eq}} = \frac{1}{2}\sum_{i=1}^{I} p_i(\boldsymbol{s}_i)\mathrm{erfc}\left(\sqrt{\frac{|g_D + g_{\mathrm{ISI}}(\boldsymbol{s}_i)|^2}{2\sigma_v^2\sum_{i=0}^{L_F-1}|f_i|^2}}\right). \quad (1.14)$$

Optimizing such an expression is quite difficult. Approximations can be obtained for small ISI, i.e., $\max |g_{\mathrm{ISI}}(\boldsymbol{s}_i)| \ll |g_D|$. Then the corruption resulting from ISI can be regarded as a Gaussian process and can be added to the noise term:

$$\mathrm{BER}_{\mathrm{ISI,eq}} \approx \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{1}{2\left[\frac{\sigma_v^2}{|g_D|^2}\sum_{i=0}^{L_F-1}|f_i|^2 + \mathrm{ISI}\right]}}\right) = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\mathrm{SINR}_{\mathrm{eq}}}{2}}\right). \tag{1.15}$$

In the case of small PD < 1 another option for approximation is to derive an upper bound by the worst case

$$\mathrm{BER}_{\mathrm{ISI}} \leq \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{\Big|g_D - |g_D|\,\mathrm{PD}\Big|^2}{2\sigma_v^2\sum_{i=0}^{L_F-1}|f_i|^2}}\right), \tag{1.16}$$

with the above defined peak distortion measure (1.8). Note that (perfect) ZF solutions always result in simpler expressions of the form (1.11) with $g_D = 1$ and $g_{\mathrm{ISI}} = 0$, while MMSE will result in the much more complicated expression (1.13) with the need to use approximate results. The $\mathrm{erfc}(\cdot)$ links the BER to the $\mathrm{SNR}_{\mathrm{eq}}$ obtained from the ideal ZF solution and the BER to the $\mathrm{SINR}_{\mathrm{eq}}$ from the MMSE solution. However, in general it remains open which criterion leads to smaller BER[3]. Note also that there exist unbiased MMSE solutions as well. In this case $g_D = 1$, simply obtained by dividing the MMSE solution by g_D. In [4] it is argued that unbiased MMSE solutions give lower BER than standard MMSE.

1.3 Channel Equalization

Channel equalization tries to restore the transmission signal $s(k)$ by means of linear or non-linear filtering. Such an approach seems straightforward and abundant literature is available, see for example [3], [4], [5], [6] to name a few. An overview of such techniques will be given in the following section where a trade-off has been made between a detailed description and sufficient information for wireless applications. Channel equalization as described in this section is not specific to wireless systems and can also be (and has successfully been) applied to other fields where time-invariant channels are common.

[3] For a non ISI channel, as it appears, for example, in OFDM (also called DMT) and PAM transmission, it can be shown that minimizing BER is equivalent to the ZF solution, i.e., $f_D = 1/c_D$.

1.3.1 Infinite Filter Length Solutions for Single Channels

It is quite educational to assume the equalizer filter solution to be of infinite length. Minimizing only the ISI part in (1.6), the ZF criterion is obtained, leading to the following expression:

$$F_{\mathrm{ZF},\infty}(q^{-1}) = \frac{C^*(q^{-1})}{|C(q^{-1})|^2} q^{-D} = \frac{q^{-D}}{C(q^{-1})}. \tag{1.17}$$

This solution is typically of infinite length as can be shown by a simple example. Assume the channel impulse response to be of the form

$$C(q^{-1}) = c_0 + c_1 q^{-1}. \tag{1.18}$$

Then the ZF solution (for $D = 0$) requires inversion of the channel, i.e.,

$$F_{\mathrm{ZF}}(q^{-1}) = \frac{1}{c_0} \frac{1}{1 + \frac{c_1}{c_0} q^{-1}}, \tag{1.19}$$

which is the structure of a first order recursive filter. If $|c_1/c_0| < 1$, a causal, stable filter solution with infinite length exists. On the other hand, if $|c_1/c_0| > 1$, a stable, anti-causal filter exists with impulse response spanning from $-\infty$ to 0. In practice such anti-causal solution can be handled by allowing additional delays $D > 0$ so that the part of the impulse response carrying most of its energy can be represented as a causal solution and the remaining anti-causal tail is neglected. In the following part of this section the difference of causal and anti-causal filters will not be considered and instead a general equalizer filter of double infinite length will be assumed. Substituting the general solution (1.17) into the MMSE expression above clearly zeroes the ISI part. The remaining term is controlled by the noise obtaining

$$\mathrm{MSE}_{\mathrm{ZF},\infty} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sigma_v^2}{|C(e^{-j\Omega})|^2} d\Omega. \tag{1.20}$$

In contrast to the ZF solution, the MMSE solution minimizing ISI and noise simultaneously, reads

$$F_{\mathrm{MMSE},\infty}(q^{-1}) = \frac{C^*(q^{-1})}{|C(q^{-1})|^2 + \sigma_v^2} q^{-D}. \tag{1.21}$$

The MMSE for this solution results in

$$\mathrm{MMSE}_{\mathrm{LIN},\infty} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sigma_v^2}{|C(e^{-j\Omega})|^2 + \sigma_v^2} d\Omega, \tag{1.22}$$

which is for $\sigma_v^2 > 0$ always smaller than the MSE of the ZF solution. Note that this advantage requires exact knowledge of the noise power as indicated in (1.21). Especially in wireless applications, due to time and frequency dispersive channels, such knowledge is usually not available and it can be very difficult and of high complexity to obtain a reliable estimate.

Example. Consider the above example of a two tap FIR filter for which $|c_0|^2 + |c_1|^2 = 1$ and $|c_1/c_0| < 1$. In case of the ZF receiver, the MSE is given by

$$\mathrm{MSE}_{\mathrm{ZF},\infty} = \frac{\sigma_v^2}{\sqrt{(|c_o|^2 - |c_1|^2)^2}} \tag{1.23}$$

while for the MMSE receiver

$$\mathrm{MMSE}_{\mathrm{LIN},\infty} = \frac{\sigma_v^2}{\sqrt{(|c_o|^2 - |c_1|^2)^2 + \sigma_v^4 + 2\sigma_v^2}} \leq \mathrm{MSE}_{\mathrm{ZF},\infty} \tag{1.24}$$

the MMSE is obtained.

1.3.2 Finite and Infinite Filter Length Solutions for Multiple Channels

Until here, only single channels were considered. In the near future, wireless systems are expected to use multiple antennas at the transmitter and/or receiver, resulting in so called *multiple-input multiple-output* (MIMO) systems. In this case, the received signal at antenna $n;\ n = 1, ..., N$ is a superposition of all transmitted symbols:

$$r_n(k) = \sum_{m=1}^{M} C_{nm}(q^{-1}) s_m(k) + v_n(k);\ n = 1, ..., N. \tag{1.25}$$

Two cases are of interest: 1) the symbols at every transmit antenna are identical (single-input multiple-output system (SIMO)) and 2) they are different.

SIMO Systems: If the transmitted symbol is identical at all transmit antennas (but different at different times), the M channels $C_{nm}(q^{-1})$ are simply combined to one new channel $C_n(q^{-1}) = \sum_m C_{nm}(q^{-1})$ and the received symbols can be written in vector form:

$$\boldsymbol{r}(k) = \boldsymbol{c}(q^{-1})[s(k)] + \boldsymbol{v}(k), \tag{1.26}$$

where $\boldsymbol{c}(q^{-1})$ is a column vector with N entries $C_n(q^{-1});\ n = 1, ..., N$. The linear equalizer solution is given by a vector $\boldsymbol{f}_D(q^{-1})$ with entries $F_n^*(q^{-1})$ chosen to re-establish the transmitted symbol $s(k)$ by computing $\boldsymbol{f}_D^H(q^{-1})\boldsymbol{r}(k) = \boldsymbol{f}_D^H(q^{-1})\boldsymbol{c}(q^{-1})[s(k)] + \boldsymbol{f}_D^H(q^{-1})\boldsymbol{v}(k)$. The corresponding ZF condition thus reads

$$\boldsymbol{f}_D^H(q^{-1})\boldsymbol{c}(q^{-1}) = \sum_{n=1}^{N} F_n(q^{-1}) C_n(q^{-1}) = q^{-D}. \tag{1.27}$$

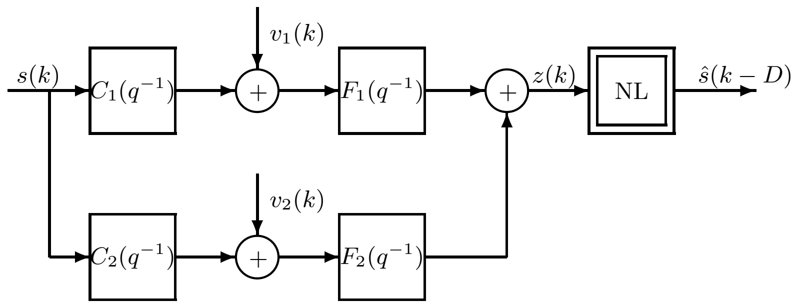

Fig. 1.1. Common structure for SIMO or fractionally spaced equalizers.

Figure 1.1 depicts a scenario for $N = 2$. If the antennas are spaced sufficiently far apart ($> \lambda/4$), the noise components can be assumed to be uncorrelated. A nonlinear device denoted by NL is typically used to reconstruct the symbols from a finite alphabet $\mathcal{A}$. In the simple case of BPSK and QPSK it is called a *slicer* since it slices the two halves or four quadrants of the complex plane into two or four equal-sized pieces.

A typical example of this principle is called receive-diversity where multiple receive antennas pick up different realizations of the same $s(k)$ caused by different transmission channels $C_n(q^{-1})$ from the one transmit antenna to each of the receive antennas. In a flat fading situation (no time-dispersion) for example, the channel functions are only given by a single value $C_n(q^{-1}) = c_{n0}$. The linear combination by the vector $\boldsymbol{f}$ is performed best for $f_{n0} = c_{n0}^*$ and is called *maximal ratio combining* (MRC). A second example of such a principle is the so-called RAKE receiver typically used in CDMA systems. Here, the multiple channels may not be caused through several receive antennas but by multiple delayed transmission paths which can be separated in the receiver. After this separation the signal appears as if several receive antennas (also called virtual antennas) were present.

A very similar situation to the multi-antenna case is obtained when so-called *fractionally spaced equalizers* (FSE) are utilized. In this case, the received signal is sampled more often than the Nyquist rate requires. Note that such oversampling is performed while the signal bandwidth remains constant.

In fractionally-spaced equalizers, the received signals are sampled with, say, N times higher rate. The corresponding N phases of the received signal experience different channels C_n. The received signals thus appear as if they were transmitted through N different channels very much as if N receive antennas were used. One difference though, is the correlation of the received noise. While in the multiple antenna case the noise component on each receive antenna can be uncorrelated, this is not the case for the fractionally spaced equalizer. The pulse shaping filter appearing at transmitter and receiver (in form of a matched filter) causes the noise components to be correlated for fractionally spaced equalizers. Thus, Fig. 1.1 also describes a scenario with $T/2$ spaced oversampling, where $C_1(q^{-1})$ describes the T-spaced part starting at time zero and $C_2(q^{-1})$ describes the part starting at time $T/2$. Note that the noise components, unlike in the multiple antenna case, are now correlated.

Example. Assume two channels of two coefficients each fed by the same symbols $s(k)$. The symbols are collected in a vector $\boldsymbol{s}(k) = [s(k), s(k-1)]^T$. The received vector $\boldsymbol{r}(k) = [r_1(k), r_2(k)]^T$ then reads in compact matrix notation

$$\boldsymbol{r}(k) = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \boldsymbol{s}(k) + \boldsymbol{v}(k). \tag{1.28}$$

The desired equalizer shall be a set of two FIR filters of length $L_F = 2$ each,

$$\boldsymbol{f}^H(q^{-1}) = [f_{11} + f_{12}q^{-1}, f_{21} + f_{22}q^{-1}]. \tag{1.29}$$

Thus, a ZF solution is obtained (stacked Toeplitz form) for:

$$\begin{bmatrix} c_{11} & 0 & c_{21} & 0 \\ c_{12} & c_{11} & c_{22} & c_{21} \\ 0 & c_{12} & 0 & c_{22} \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \tag{1.30}$$

Obviously, only three equations exist to determine four variables, i.e., one variable can be selected at will. For an MMSE solution, this additional freedom allows to achieve smaller MSE values while not violating the ZF solution too much. More about such finite MMSE length solutions will be said in Sect. 1.3.3. Length requirements for typical wireless channels are discussed in [5].

In general, if the reception of a symbol $s(k)$ can be written in terms of multiple polynomials caused by multiple sub-channel filters $C_n(q^{-1})$,

$$r_n(k) = C_n(q^{-1})s(k); \; n = 1, ..., N, \tag{1.31}$$

then a finite length ZF solution can be obtained when the sub-channels $C_n(q^{-1})$ are co-prime, i.e., they have no common zeros. They are co-prime if and only if there exists a set of polynomials $\{F_n(q^{-1})\}; \; n = 1, ..., N$ that satisfy the **Bezout identity** [7], [8], [9]:

$$\sum_{n=1}^{N} F_n(q^{-1})C_n(q^{-1}) = 1. \tag{1.32}$$

Thus, although an MMSE solution is obtained by

$$F_n(q^{-1}) = \frac{C_n^*(q^{-1})}{\sum |C_n(q^{-1})|^2 + \sigma_v^2} \tag{1.33}$$

in the case of independent and identical noise components, such an infinite length filter solution is not required and a finite length filter solution exists instead.

Example. Assume two channel filters with one common zero $C_1(q^{-1}) = (1+c_1q^{-1})\bar{C}_1(q^{-1})$ and $C_2(q^{-1}) = (1+c_1q^{-1})\bar{C}_2(q^{-1})$, where $\bar{C}_1(q^{-1})$ and $\bar{C}_2(q^{-1})$ are co-prime. In this case,

$$
\begin{aligned}
1 &= C_1(q^{-1})F_1(q^{-1}) + C_2(q^{-1})F_2(q^{-1}) \\
&= (1+c_1q^{-1})\left[\bar{C}_1(q^{-1})F_1(q^{-1}) + \bar{C}_2(q^{-1})F_2(q^{-1})\right]
\end{aligned}
$$

cannot be satisfied.

MIMO Systems: If the transmitted symbols are all different[4], the received symbol vector can be written as

$$
\boldsymbol{r}(k) = \boldsymbol{C}\,\boldsymbol{s}(k) + \boldsymbol{v}(k), \tag{1.34}
$$

where $\boldsymbol{C}$ is a matrix with $N \times M$ entries $C_{nm}(q^{-1})$; $m = 1, ..., M$, $n = 1, ..., N$. In this case, a matrix $\boldsymbol{F}$ is desired to re-establish the transmitted vector $\boldsymbol{s}(k) = [s_1(k), ..., s_M(k)]^T$ by computing $\boldsymbol{F}^H\boldsymbol{r}(k) = \boldsymbol{F}^H\boldsymbol{C}\,\boldsymbol{s}(k) + \boldsymbol{F}^H\boldsymbol{v}(k)$. Consider a simple case in which two symbols are transmitted (one on each transmit antenna) and three receive antennas are utilized:

$$
\boldsymbol{r}(k) = \begin{bmatrix} C_{11}(q^{-1}) & C_{12}(q^{-1}) \\ C_{21}(q^{-1}) & C_{22}(q^{-1}) \\ C_{31}(q^{-1}) & C_{32}(q^{-1}) \end{bmatrix} \boldsymbol{s}(k) + \boldsymbol{v}(k). \tag{1.35}
$$

According to the generalized Bezout Identity [9], [10], a ZF solution exists for $N > M$ if $\boldsymbol{C}$ is of full rank for all q^{-1}. Since each filter function C_{mn} can be expressed as a vector $\boldsymbol{c}_{mn}$, stacking the past $L_C - 1$ symbols of $\boldsymbol{s}(k)$ in an even larger vector, allows for reformulating the matrix of dimension $N \times M$ into a new matrix $\bar{\boldsymbol{C}}$ of higher dimension with constant entries. At the receiver a minimum norm ZF solution can be obtained (see also next section) by computing $\boldsymbol{F}_{\text{ZF}} = \bar{\boldsymbol{C}}^H[\bar{\boldsymbol{C}}^H\bar{\boldsymbol{C}}]^{-1}$ or alternatively, an MMSE solution by computing $\boldsymbol{F}_{\text{MMSE}} = [\bar{\boldsymbol{C}}^H\bar{\boldsymbol{C}} + \sigma_v^2\boldsymbol{I}]^{-1}\bar{\boldsymbol{C}}^H$. The matrix inversion of the so-obtained $(ML_C) \times (ML_C)$ matrix can be numerically challenging. Better methods are obtained by so-called iterative BLAST receivers [11] also in combination with CDMA [12] or OFDM transmission [13]. An overview of ZF, MMSE, and DFE techniques for MIMO is given in [14].

1.3.3 Finite Filter Length Solutions for Single Channels

As mentioned before, the ideal filter length of linear equalizers for single channels is doubly infinite. Practically, the filter length is fixed by the affordable complexity. Therefore, in this section, suboptimal finite length solutions

[4] This is also called *Vertical-Bell-labs-LAyered-Space-Time* (V-BLAST) transmission.

will be considered. To achieve this goal, the channel impulse response will be assumed to be of finite length L_C and will be written in vector notation, i.e., $\tilde{\boldsymbol{C}} \in \mathbb{C}^{(L+L_C-1)\times L}$ denotes a channel transmission matrix describing the following model (only shown for $L_C = 3$)

$$\begin{bmatrix} r(L+2) \\ r(L+1) \\ r(L) \\ r(L-1) \\ \vdots \\ r(2) \\ r(1) \end{bmatrix} = \begin{bmatrix} c_2 & 0 & 0 & 0 & \dots & 0 \\ c_1 & c_2 & 0 & 0 & \dots & 0 \\ c_0 & c_1 & c_2 & 0 & \dots & 0 \\ 0 & c_0 & c_1 & c_2 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & c_0 & c_1 & c_2 \\ 0 & \dots & 0 & 0 & c_0 & c_1 \\ 0 & \dots & 0 & 0 & 0 & c_0 \end{bmatrix} \begin{bmatrix} s(L) \\ s(L-1) \\ \vdots \\ s(2) \\ s(1) \end{bmatrix} + \begin{bmatrix} v(L+2) \\ v(L+1) \\ v(L) \\ v(L-1) \\ \vdots \\ v(2) \\ v(1) \end{bmatrix}, \tag{1.36}$$

allowing us to express the transmission of L symbols $s(k);\ k = 1, ..., L$:

$$\boldsymbol{r} = \tilde{\boldsymbol{C}}\,\boldsymbol{s} + \boldsymbol{v}. \tag{1.37}$$

Note that such notation includes the beginning and end of a transmission and grows with L. For zero-forcing, a filter vector $\boldsymbol{f}_D \in \mathbb{C}^{(L+L_C-1)\times 1}$ can be found such that

$$\tilde{\boldsymbol{f}}_D^H \tilde{\boldsymbol{C}} = \boldsymbol{e}_D^T, \tag{1.38}$$

(where $\boldsymbol{e}_D = [0...010...0]^T$ is a unit vector of appropriate dimension whose D-th component is 1) i.e., the filter selects a single value $\boldsymbol{e}_D^T \boldsymbol{s}(k) = s(k-D)$. This description allows for computing the ZF and MMSE solutions:

$$\tilde{\boldsymbol{f}}_{\mathrm{ZF},D} = \tilde{\boldsymbol{C}}[\tilde{\boldsymbol{C}}^H \tilde{\boldsymbol{C}}]^{-1} \boldsymbol{e}_D \tag{1.39}$$

$$\tilde{\boldsymbol{f}}_{\mathrm{MMSE},D} = \tilde{\boldsymbol{C}}[\tilde{\boldsymbol{C}}^H \tilde{\boldsymbol{C}} + \sigma_v^2 \boldsymbol{I}]^{-1} \boldsymbol{e}_D. \tag{1.40}$$

Example. Consider again the example of (1.23) and (1.24) for a two tap FIR channel filter. The channel matrix reads

$$\tilde{\boldsymbol{C}} = \begin{bmatrix} c_1 & 0 \\ c_0 & c_1 \\ 0 & c_0 \end{bmatrix}, \tag{1.41}$$

The two ZF solutions are

$$\tilde{\boldsymbol{f}}_1^T = \frac{\left[c_1\left(|c_0|^2 + |c_1|^2\right), c_0|c_0|^2, -c_1^* c_0^2\right]}{\left(|c_0|^2 + |c_1|^2\right)^2 - |c_0 c_1|^2}, \tag{1.42}$$

$$\tilde{\boldsymbol{f}}_2^T = \frac{\left[-c_0^* c_1^2, c_1|c_1|^2, c_0\left(|c_0|^2 + |c_1|^2\right)\right]}{\left(|c_0|^2 + |c_1|^2\right)^2 - |c_0 c_1|^2}, \tag{1.43}$$

while the two MMSE solutions read

$$\tilde{\boldsymbol{f}}_1^T = \frac{\left[c_1\left(|c_0|^2+|c_1|^2+\sigma_v^2\right), c_0(|c_0|^2+\sigma_v^2), -c_1^* c_0^2\right]}{\left(|c_0|^2+|c_1|^2+\sigma_v^2\right)^2-|c_0c_1|^2}, \tag{1.44}$$

$$\tilde{\boldsymbol{f}}_2^T = \frac{\left[-c_0^* c_1^2, c_1(|c_1|^2+\sigma_v^2), c_0\left(|c_0|^2+|c_1|^2+\sigma_v^2\right)\right]}{\left(|c_0|^2+|c_1|^2+\sigma_v^2\right)^2-|c_0c_1|^2}. \tag{1.45}$$

While these results seem convincing, they are misleading. Clearly, the description (1.36) requires us to observe the transmission from beginning to end; in particular, the values in which not all symbols are present are of importance. If, however, for a continuous transmission, a snapshot over $L_1 < L$ symbols is taken, say $\boldsymbol{s}^T(k) = [s(k), s(k+1), ..., s(k+L_1-1)]$, the following matrix of Toeplitz structure is obtained

$$\boldsymbol{C} = \begin{bmatrix} c_0 & c_1 & ... & c_{L_C-1} & 0 & ... & 0 \\ 0 & c_0 & c_1 & ... & c_{L_C-1} & 0 & ... \\ 0 & 0 & c_0 & c_1 & ... & c_{L_C-1} & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & ... & 0 & c_0 & c_1 & ... & c_{L_C-1} \end{bmatrix}, \tag{1.46}$$

instead. Such a matrix $\boldsymbol{C} \in \mathbb{C}^{L_1 \times (L_C+L_1-1)}$ is not of sufficient rank to compute a linear solution by the pseudo-inverse [of dimension $(L_C+L_1-1) \times (L_C+L_1-1)$] as was done in (1.39). For ZF, a different pseudo-inverse leading to a minimum norm solution exists while for MMSE the standard pseudo-inverse exists and with the above notation, and vectors $\boldsymbol{e}_D$ of appropriate size for each solution,

$$\boldsymbol{f}_{\mathrm{ZF},D} = [\boldsymbol{C}\boldsymbol{C}^H]^{-1}\boldsymbol{C}\boldsymbol{e}_D \tag{1.47}$$

$$\boldsymbol{f}_{\mathrm{MMSE},D} = \boldsymbol{C}[\boldsymbol{C}^H\boldsymbol{C} + \sigma_v^2\boldsymbol{I}]^{-1}\boldsymbol{e}_D. \tag{1.48}$$

Extending such solutions to the MIMO case (including fractionally spaced equalizers) is straightforward[5].

Example. For an FSE with $T/2$ spacing including the matched filters the MMSE solution is considered. Note that in order to describe the oversampled system, the introduction of interpolation filters is required. Typically a matched filter pair $\boldsymbol{R}_p$ is used at the transmitter and receiver. So far, such filters were not shown explicitly and were incorporated in the channel impulse response. For oversampled systems, however, they play an important role and need to be considered explicitly. The received signal for each phase can be written in the form

$$\boldsymbol{r}_p(k) = \boldsymbol{R}_p\boldsymbol{C}_p\boldsymbol{R}_p\boldsymbol{s}(k) + \boldsymbol{R}_p\boldsymbol{v}(k);\; p = 1, ..., P. \tag{1.49}$$

[5] Note that in the SIMO or FSE case the elements in $\boldsymbol{C}$ become vectors and the condition that the Toeplitz matrix $\boldsymbol{C}$ (1.46) is of full column rank is equivalent to Bezout's Identity.

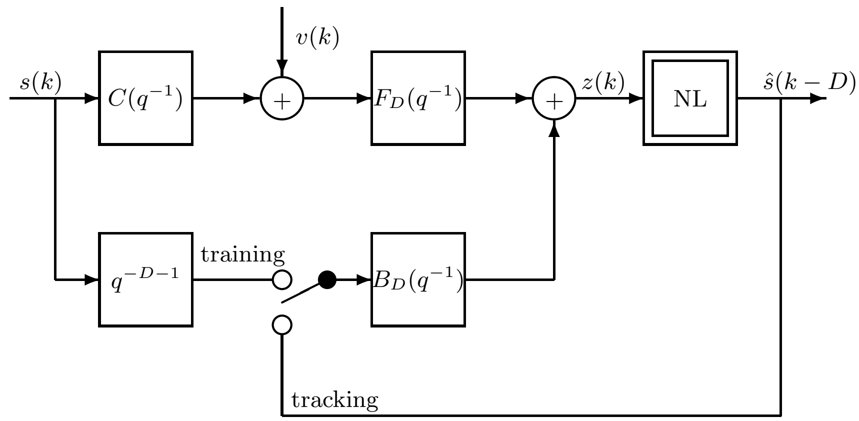

Fig. 1.2. Structure of a DFE receiver.

The noise is thus also filtered by such a filter. Note that the noise is present at each path, originating from identical noise sources but differently filtered: $\boldsymbol{v}_p = \boldsymbol{R}_p\boldsymbol{v}$. The MMSE equations read

$$\begin{bmatrix} \boldsymbol{R}_1\boldsymbol{C}_1\boldsymbol{R}_1\boldsymbol{R}_1^H\boldsymbol{C}_1^H\boldsymbol{R}_1^H + \sigma_v^2\boldsymbol{R}_1\boldsymbol{R}_1^H & \boldsymbol{R}_1\boldsymbol{C}_1\boldsymbol{R}_1\boldsymbol{R}_2^H\boldsymbol{C}_2^H\boldsymbol{R}_2^H \\ \boldsymbol{R}_2\boldsymbol{C}_2\boldsymbol{R}_2\boldsymbol{R}_1^H\boldsymbol{C}_1^H\boldsymbol{R}_1^H & \boldsymbol{R}_2\boldsymbol{C}_2\boldsymbol{R}_2\boldsymbol{R}_2^H\boldsymbol{C}_2^H\boldsymbol{R}_2^H + \sigma_v^2\boldsymbol{R}_2\boldsymbol{R}_2^H \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{R}_1\boldsymbol{C}_1\boldsymbol{R}_1\boldsymbol{e}_D \\ \boldsymbol{R}_2\boldsymbol{C}_2\boldsymbol{R}_2\boldsymbol{e}_D \end{bmatrix}. \tag{1.50}$$

1.3.4 Decision Feedback Equalizers

As mentioned above and substantiated by the Bezout theorem, once several transmission paths (sub-channels) are available for the same sequence, the equalization problem can be solved by a finite length filter, whereas this is not the case for a single transmission path. There are several methods available to obtain independent paths and they do not necessarily require more than one antenna at the transmit or receiver; FSE was already mentioned in this respect. A similar method is known under the name of *decision feedback equalization* (DFE). Unlike the so far considered *linear equalizers*, the DFE also uses past estimated symbols $s(k-D-1), ..., s(k-D-L_B)$ to detect the current symbol $s(k-D)$.

Figure 1.2 displays the DFE structure. The switch allows the operation mode to be changed from training to tracking, thus from feeding correct symbols to using estimated ones. The estimated symbols result from a nonlinear device (NL) that maps the soft symbols $z(k)$ onto the nearest possible symbol in the allowed $\mathcal{A}$. During training, the feedback path $B_D(q^{-1})$ of length L_B

$$B_D(q^{-1}) = \sum_{i=0}^{L_B-1} b_i q^{-i};\ b_0 \neq 0, \tag{1.51}$$

can be considered a second transmission path and thus Bezout's theorem can be satisfied. In the training mode, the ZF condition reads

$$F_D(q^{-1})C(q^{-1}) + q^{-D-1}B_D(q^{-1}) = q^{-D}, \tag{1.52}$$

while the MMSE condition is given by:

$$\begin{aligned}\text{MMSE} = \min_{F_D,B_D} &\underbrace{E\left[\left|\left\{q^{-D} - F_D(q^{-1})C(q^{-1}) - q^{-D-1}B_D(q^{-1})\right\} s(k)\right|^2\right]}_{\text{ISI part}} \\ &+E\underbrace{\left[\left|F_D(q^{-1})v(k)\right|^2\right]}_{\text{noise part}}. \end{aligned}\tag{1.53}$$

As compared to (1.6), the equation above allows one more degree of freedom. By modifying $B_D(q^{-1})$ it is now possible to cancel ISI components while the noise term remains unchanged. In the tracking mode, the estimated symbols are fed back. As long as the SER is relatively small, it can still be considered as a second transmission path.

Infinite Length Solution. Classically the DFE solution was given by J. Salz [15]. The infinite MMSE solution requires a term $|C(q^{-1})|^2 + \sigma_v^2$ in the denominator. Assume a monic, causal, and minimum-phase polynomial[6] $M(q^{-1})$ selected so that $M_o M(q^{-1})M^*(q) = |C(q^{-1})|^2 + \sigma_v^2$, where M_o is a scaling constant. For unit transmitted signal energy, the DFE solution is given by

$$F(q^{-1}) = \frac{1}{M_o}\frac{C^*(q)}{M^*(q)}, \tag{1.54}$$

$$B(q^{-1}) = 1 - M(q^{-1}). \tag{1.55}$$

Obviously, the feedforward path requires an anti-causal solution while the feedback path is strictly causal. It can be shown [15], [3] that the corresponding MSE is given by

$$\text{MMSE}_{\text{DFE},\infty} = \frac{\sigma_v^2}{M_o} = \sigma_v^2 \exp\left(-\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left[|C(e^{j\Omega})|^2 + \sigma_v^2\right] d\Omega\right) \tag{1.56}$$

[6] Monic polynomials are of the form $m_0 = 1$, causal ones have only terms for non-negative delays and minimum-phase is equivalent to having zeroes only inside the unit circle.

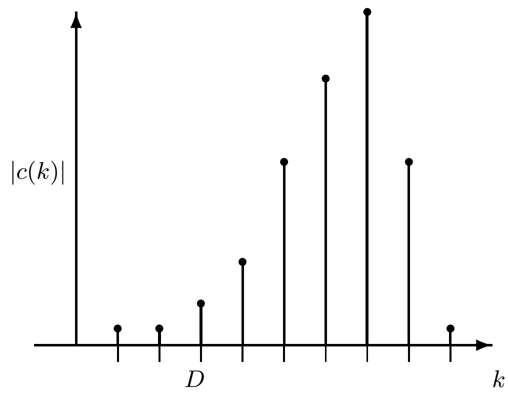

Fig. 1.3. Possible channel impulse response magnitude and selected cursor position D.

and that the so obtained MMSE never exceeds the MMSE for a linear equalizer solution[7]. A comprehensive overview for DFE solutions is given in [3].

Finite Length Solution. Figure 1.3 depicts a possible channel impulse response. Assume $D = 0$, i.e., it is set to the first value of the impulse response. In this case by proper choice of $B_D(q^{-1})$ all following values can be cancelled. For the BER of a BPSK (QPSK) system,

$$\text{BER} = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{|c_0 f_0|^2}{2\sum |f_i|^2 \sigma_v^2}}\right) \tag{1.57}$$

is obtained. Since the noise term is minimal for $f_i = 0,\ i \neq 0$, it can be recognized that the influence of f_0 is cancelled and thus the quality of the BER is only dependent on the first channel term c_0. For wireless channels this will be in general a small term and thus only poor BER performance can be achieved. If D is increased much better values can be obtained. This does not need to come at the expense of additional ISI resulting form the pre-cursor position. In [16], it is shown that a ZF solution is obtained if

$$L_F + L_B \geq \max\{L_F + L_C - 1, D + L_B - 1\}. \tag{1.58}$$

However, it remains difficult to find the optimal selection of D since it is dependent on the actual channel $C(q^{-1})$.

Since minimizing BER remains problematic, a better starting point for optimization is again the ZF and the MMSE criterion. In the finite filter

[7] Also a second solution exists with a stable and causal feedforward filter but with anti-causal feedback filter.

length case, the impact of the channel and the feedback part can be combined into one matrix

$$\tilde{\boldsymbol{C}} = \begin{bmatrix} c_0 \; c_1 \; ... \; c_{L_C-1} & 0 & ... & 0 \\ 0 \; c_0 \; c_1 & ... \; c_{L_C-1} & 0 & ... \\ 0 \;\; 0 \; c_0 & c_1 & ... \; c_{L_C-1} & 0 \\ \vdots \; \vdots \; \vdots & \vdots & \ddots \quad \ddots & \vdots \\ 0 \; ... \;\; 0 & c_0 & c_1 \quad ... & c_{L_C-1} \\ 0 \; ... \;\; 0 & 1 & 0 \quad ... & 0 \\ 0 \; ... \;\; 0 & 0 & 1 \quad 0 & ... \\ 0 \; ... \;\; 0 & 0 & 0 \quad 1 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{I}_D \end{bmatrix}, \tag{1.59}$$

where a new matrix

$$\boldsymbol{I}_D \triangleq \begin{bmatrix} 0\,0\,0\,1 & 0 & ... & 0\,0 \\ 0\,0\,0\,0 & 1 & 0 & ... \; 0 \\ 0\,0\,0\,0 & 0 & 1 & ... \; 0 \\ \vdots \; \vdots & & \ddots & 0 \\ 0\,0\,0\,0 & ... & 0 & 1 \end{bmatrix} \tag{1.60}$$

has been introduced which is a shifted identity matrix with a delay of D steps (here $D = 3$ has been shown). The task is to find two vectors $\boldsymbol{f}_D$ and $\boldsymbol{b}_D$ in order to minimize either the ZF or MMSE condition

$$\mathrm{MSE_{ZF}} = \min_{\boldsymbol{f}_D, \boldsymbol{b}_D} \left\| \left[\boldsymbol{e}_D - \boldsymbol{f}_D^H \boldsymbol{C} - \boldsymbol{b}_D^H \boldsymbol{I}_D \right] \boldsymbol{s} \right\|^2, \tag{1.61}$$

$$\mathrm{MMSE} = \min_{\boldsymbol{f}_D, \boldsymbol{b}_D} \left\| \left[\boldsymbol{e}_D - \boldsymbol{f}_D^H \boldsymbol{C} - \boldsymbol{b}_D^H \boldsymbol{I}_D \right] \boldsymbol{s} \right\|^2 + \|\boldsymbol{f}_D\|^2 \sigma_v^2. \tag{1.62}$$

For MMSE, after differentiation with respect to $\boldsymbol{f}_D$ and $\boldsymbol{b}_D$,

$$\begin{bmatrix} \boldsymbol{C}\boldsymbol{C}^H + \sigma_v^2 \boldsymbol{I} & \boldsymbol{C}\boldsymbol{I}_D^T \\ \boldsymbol{I}_D \boldsymbol{C}^H & \boldsymbol{I}_D \boldsymbol{I}_D^T \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_D \\ \boldsymbol{b}_D \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}\, \boldsymbol{e}_D \\ \boldsymbol{I}_D\, \boldsymbol{e}_D \end{bmatrix} \tag{1.63}$$

is obtained. The matrix is not of full rank when the noise term is zero. Thus, the ZF solution can again be obtained via the minimum norm solution.

Error Propagation. So far it was assumed that the correct symbols are fed back. In general, this will only be the case when utilizing a known training sequence. Once the equalizer is used in tracking (or decision directed) mode, the detected symbols are fed back. In this case an erroneously detected symbol can cause further errors. In simulations [18] DFE structures show about 2-3 dB loss in BER due to error propagation[8]. Possibilities to avoid such undesired behavior are either the application of a particular precoding (Tomlinson-Harashima [18]) or in general the utilization of error correcting codes such as trellis codes. See also Sect. 1.6.1 for more details in combination with the Viterbi algorithm.

[8] A computation of BER is also possible. See, for example, [17].

1.4 Adaptive Algorithms for Channel Equalization

The equations for finite filter solutions already provide a possible solution for an adaptive filter. Assuming the channel remains constant over a certain time period, the channel (and the noise variance) can be estimated and by applying equations (1.39) and (1.40), the ZF or MMSE solution can be computed [19]. Such a method not only requires an additional channel estimation step, but also the computation of a matrix inverse. Although complexity-wise often acceptable, the numerical challenge of such inversion can be quite high. For rapidly changing channels, the complexity aspect becomes burdensome as well. Thus, simpler methods with low complexity and high numerical robustness are desired. Since only the equalizer solution is wanted, it is not required to estimate the channel first. Applying gradient-type, adaptive algorithms is a straightforward solution, resulting in much less complexity for this application. In the following, adaptive algorithms to minimize ZF as well as MMSE based on reference models are considered.

1.4.1 Adaptively Minimizing ZF

As previously mentioned, a finite equalization filter $\boldsymbol{f}_o$ cannot be expected to deliver a perfect ZF-solution. Assuming a linear channel with finite impulse response $\boldsymbol{c}^T = [c_0, c_1, ..., c_{L_C-1}]$, $\boldsymbol{f}_o$ will result in a sub-optimal ZF solution. Composing a matrix of size $L_1 \times (L_C + L_1 - 1)$ as in (1.46) allows us to write the obtained ZF solution as

$$\boldsymbol{C}^H \boldsymbol{f}_o = \boldsymbol{e}_D + \delta_D, \tag{1.64}$$

where $\boldsymbol{e}_D$ indicates the unit vector with one entry "one" at the $D-$th position and δ_D is a residual error vector. Following this approach, the optimal ZF solution is the one given by

$$\boldsymbol{f}_o = \arg\min_{\boldsymbol{f}} \|\boldsymbol{C}^H \boldsymbol{f} - \boldsymbol{e}_D\|^2. \tag{1.65}$$

Note that the solution of (1.65) is not straightforward due to the rank deficiency of $\boldsymbol{C}$. A minimum norm solution exists (1.47), causing a gain at time instant D not necessarily being one. Or, equivalently, the contribution of δ_D at time D may not be zero. With such a reference system $\boldsymbol{f}_o$, an iterative algorithm known as the **adaptive zero-forcing algorithm** is given by

$$\boldsymbol{f}^*(k) = \boldsymbol{f}^*(k-1) + \mu(k)\boldsymbol{s}_k^*[s(k-D) - \hat{s}(k-D)], \tag{1.66}$$

where the vector $\boldsymbol{f}^*(k)$ of dimension L_F estimates the equalizer solution $\boldsymbol{f}_o$ and the regression vector $\boldsymbol{s}^T(k) = [s(k), s(k-1), ..., s(k-D), ..., s(k-L_F+1)]$. In the literature (see for example [18]) the behavior of this algorithm is usually referred to the *least-mean-square* (LMS) algorithm. It is demonstrated next that this is a simplification. By introducing a signal vector $\boldsymbol{t}(k)$

$$\boldsymbol{t}^T(k) = [s(k), s(k-1), ..., s(k - L_F - L_C + 1)]$$

of appropriate length, the received signal is given by

$$\boldsymbol{r}(k) = \boldsymbol{C}\boldsymbol{t}(k) + \boldsymbol{v}(k)$$

where a noise vector $\boldsymbol{v}(k)$ of dimension L_F is added as well. Applying the optimal equalizer (1.64),

$$\boldsymbol{f}_o^H \boldsymbol{r}(k) = \boldsymbol{f}_o^H \boldsymbol{C}\boldsymbol{t}(k) + \boldsymbol{f}_o^H \boldsymbol{v}(k) = \underbrace{\boldsymbol{e}_D^T \boldsymbol{t}(k)}_{s(k-D)} + \underbrace{\delta_D^H \boldsymbol{t}(k) + \boldsymbol{f}_o^H \boldsymbol{v}(k)}_{\bar{v}_o(k)} \tag{1.67}$$

is obtained. This needs to be compared to the estimated equalizer output

$$\begin{aligned} \boldsymbol{f}^H(k-1)\boldsymbol{r}(k) &= \boldsymbol{f}^H(k-1)\boldsymbol{C}\boldsymbol{t}(k) + \boldsymbol{f}^H(k-1)\boldsymbol{v}(k) \\ &= \hat{s}(k-D) + \underbrace{\hat{\delta}_D^H \boldsymbol{t}(k) + \boldsymbol{f}^H(k-1)\boldsymbol{v}(k)}_{\bar{v}(k)}. \end{aligned} \tag{1.68}$$

Thus, the error signal $s(k-D) - \hat{s}(k-D)$ can be reformulated as

$$\begin{aligned} s(k-D) - \hat{s}(k-D) =& \ \boldsymbol{r}^T(k)\boldsymbol{f}_o^* - \bar{v}_o(k) - \boldsymbol{r}^T(k)\boldsymbol{f}^*(k-1) + \bar{v}(k) \\ &= [\boldsymbol{r}^T(k) - \boldsymbol{v}^T(k)][\boldsymbol{f}_o^* - \boldsymbol{f}^*(k-1)] \\ &\quad + \boldsymbol{t}^T(k)[\delta_D^* - \hat{\delta}_D^*]. \end{aligned} \tag{1.69}$$

This formulation utilizes the noise-free received value $\boldsymbol{y}(k) = \boldsymbol{r}(k) - \boldsymbol{v}(k) = \boldsymbol{C}\boldsymbol{t}(k)$. Reformulating the update equation (1.66) in terms of the parameter error vector $\tilde{\boldsymbol{f}}(k) = \boldsymbol{f}_o - \boldsymbol{f}(k)$ results in

$$\begin{aligned} \tilde{\boldsymbol{f}}^*(k) &= \tilde{\boldsymbol{f}}^*(k-1) - \mu(k)\boldsymbol{s}^*(k)[\boldsymbol{y}^T(k)\tilde{\boldsymbol{f}}^*(k-1) - \boldsymbol{t}^T(k)[\delta_D^* - \hat{\delta}_D^*]] \quad (1.70) \\ &= \left[\boldsymbol{I} - \mu(k)\boldsymbol{s}^*(k)\boldsymbol{y}^T(k)\right] \tilde{\boldsymbol{f}}^*(k-1) - \mu(k)\boldsymbol{s}^*(k)\boldsymbol{t}^T(k)[\delta_D^* - \hat{\delta}_D^*]. \end{aligned}$$

The additional term $\mu(k)\boldsymbol{s}(k)^*\boldsymbol{t}^T(k)[\delta_D^* - \hat{\delta}_D^*]$ causes a noise floor very similar to the additive noise floor when the LMS algorithm is applied for system identification [21]. However, when approaching the optimal solution, δ_D also tends to $\hat{\delta}_D$ and the difference vanishes. Obviously, the quality of the ZF solution has an impact on initial behavior of the iterative solution. If the ZF solution results in a large value for δ_D, the iterative solution has to cope with a large additional noise. By selecting a small step-size, this effect can be decreased at the expense of a slower convergence rate. The convergence of such algorithm is thus dependent on the properties of the matrix

$$\left[\boldsymbol{I} - \mu(k)\boldsymbol{s}^*(k)\boldsymbol{y}^T(k)\right] = \left[\boldsymbol{I} - \mu(k)\boldsymbol{s}^*(k)\boldsymbol{t}^T(k)\boldsymbol{C}^T\right].$$

If convergence in the mean is considered, the correlation $\mathrm{E}[\boldsymbol{s}^*(k)\boldsymbol{t}^T(k)]=\boldsymbol{G}$ simplifies the condition, so that the eigenvalues of $\boldsymbol{I} - \mu(k)\boldsymbol{G}\boldsymbol{C}^T$ need to be smaller than 1. However, this is not guaranteed for general channels $\boldsymbol{C}$. A more practical solution for this problem can be obtained by a particular step-size rule. Consider the eigenvalues of $\boldsymbol{I} - \mu(k)\boldsymbol{s}^*(k)\boldsymbol{y}^T(k)$. For this matrix

$L_F - 1$ eigenvalues are 1 and one eigenvalue equals $1 - \mu(k)\boldsymbol{y}^T(k)\boldsymbol{s}^*(k)$. A good choice for the step-size $\mu(k)$ is thus

$$\mu(k) = \alpha \frac{\boldsymbol{y}^T(k)\boldsymbol{s}^*(k)}{\|\boldsymbol{y}(k)\|^2\|\boldsymbol{s}(k)\|^2 + \epsilon}, \tag{1.71}$$

where a small positive value ϵ is added to ensure that the denominator is strictly positive. More details on such step-size selection can be found in [20].

1.4.2 Adaptively Minimizing MMSE

The corresponding gradient-type algorithm for minimizing MMSE is much easier to analyze. The update equation reads

$$\boldsymbol{f}^*(k) = \boldsymbol{f}^*(k-1) + \mu(k)\boldsymbol{r}^*(k)[s(k-D) - \hat{s}(k-D)]. \tag{1.72}$$

Utilizing the same terms as in the ZF-case,

$$\tilde{\boldsymbol{f}}^*(k) = \left[\boldsymbol{I} - \mu(k)\boldsymbol{r}^*(k)\boldsymbol{y}^T(k)\right]\tilde{\boldsymbol{f}}^*(k-1) - \mu(k)\boldsymbol{r}^*(k)\boldsymbol{t}^T(k)[\delta_D - \hat{\delta}_D] \tag{1.73}$$

is obtained. The essential difference now is the appearance of the vector $\boldsymbol{r}(k)$ in place of the former $\boldsymbol{s}(k)$. Although not perfectly symmetrical [note that the noisy term $\boldsymbol{r}(k)$ and the noise-free term $\boldsymbol{y}(k)$ appear together], the algorithm can be analyzed with conventional methods including the independence assumption. The algorithm behaves very much like an LMS algorithm for system identification, however, with slightly different learning dynamic due to the additional noise in the driving signal and a different noise behavior.

1.4.3 Training and Tracking

The previously considered algorithms find ZF and MMSE solutions assuming the correct symbols $s(k-D)$ are available. Once the training period is over, the system is fed with estimated symbols $\hat{s}(k-D)$ instead. This mode of operation is called the decision directed, or tracking, mode. Again, a reference model delivering an optimal solution is assumed. This time, a nonlinear device is included to map the linearly estimated symbols into symbols of the transmitted alphabet $\mathcal{A}$. Note that such reference structure cannot guarantee error-free symbols $s(k-D)$. Due to the additive Gaussian noise, a small error probability remains. In the following, the length L_F of the reference filter $\boldsymbol{f}_o$ is assumed to be sufficiently long so that such an error occurrence is negligible. Figure 1.4 depicts the structure, exhibiting four different error signals under consideration:

$$e_{\mathrm{LIN}}(k) = z(k) - \hat{z}(k), \tag{1.74}$$

$$e_{\mathrm{TK}}(k) = \hat{s}(k-D) - \hat{z}(k), \tag{1.75}$$

$$e_{\mathrm{TR}}(k) = s(k-D) - \hat{z}(k), \tag{1.76}$$

$$e_{\mathrm{NL}}(k) = s(k-D) - z(k). \tag{1.77}$$

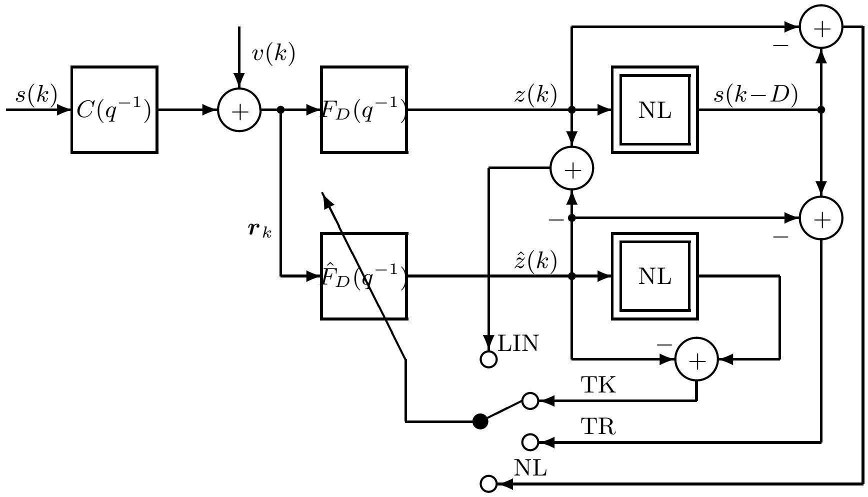

Fig. 1.4. Reference model for adaptive filtering showing various adaptation errors.

The adaptation error $e(k) = s(k-D) - \hat{s}(k-D)$ of the two adaptive algorithms in the previous sections is not shown explicitly. Note that in practice, the signal $e_{\mathrm{TR}}(k)$ is used for training as long as training symbols are available, while in the tracking mode $e_{\mathrm{TK}}(k)$ is used. The adaptive algorithm thus reads

$$\hat{\boldsymbol{f}}^*(k) = \hat{\boldsymbol{f}}^*(k-1) + \mu(k)\boldsymbol{r}^*(k) \begin{cases} e_{\mathrm{TR}}(k) & ;\text{training mode} \\ e_{\mathrm{TK}}(k) & ;\text{tracking mode} \end{cases} . \tag{1.78}$$

The relation of the two errors to the error signal $e_{\mathrm{LIN}}(k)$ is given by

$$e_{\mathrm{LIN}}(k) = e_{\mathrm{TR}}(k) - e_{\mathrm{NL}}(k) \tag{1.79}$$

$$= e_{\mathrm{TK}}(k) - e_{\mathrm{NL}}(k) + e(k). \tag{1.80}$$

Thus, adaptation error in the training mode can be regarded as the combination of the MMSE obtained by a linear system identification and an additional corruption term $e_{\mathrm{NL}}(k)$, controlled only by the optimal MMSE estimate $z(k)$. Using an adaptive filter algorithm, the following can be concluded:

- The adaptive filter works in a system identification setting.
- In the training mode, an additive noise term is present, defined in terms of the nonlinear error $e_{\mathrm{NL}}(k)$.
- In the tracking mode, the error signal consists of an additional error term $e(k)$.
- The excitation for the system identification problem is given by a composite signal, consisting of the transmitted symbols filtered by the linear channel and a white noise source.

In particular, the last point causes problems when applying low-complexity adaptive filters like the LMS algorithm. Its learning behavior is very slow for highly correlated signals [21] as is in general the case with filtered symbols. In [22] it has been found that the learning of oversampled DFE equalizers is hampered even further. Several possible solutions to overcome this problem are available:

1. A solution to this problem could be the recursive least-squares (RLS) algorithm. However, cheap fixed-point solutions on general purpose DSPs do not seem to be available [23]. Only if floating point hardware is available, the RLS algorithm in some formulation [21] can be considered for implementation.
2. A solution guaranteeing fast convergence rate is the subband approach with polyphase filters [24], [25]. Here, the entire frequency band is split into smaller bands in which the signal appears to be more or less white. Faster convergence is a result. Moreover, because of down-sampling, complexity can be saved. However, the price for these advantages is an additional delay due to the inherent block processing.
3. The convergence of the LMS algorithm can be sped up by using particular step-size matrices [26]. However, modifying the entries of such a matrix during the adaptation process can lead to unstable behavior [27].
4. Another possibility is to modify the reference model and to include the nonlinear device [28]. In this case, the conditions for the adaptive filter change considerably and standard theory for system identification cannot be applied.
5. Similar to the previous point is the idea of using particular nonlinear error signals constructed out of the estimated values $\hat{z}(k)$ only. Such algorithms are called blind when they exclude a training signal for reference entirely. More on this will be considered in Sect. 1.7.
6. Due to fading in wireless channels, a fixed MMSE solution may not exist over the entire observation horizon of a data burst. In such case, adaptive algorithms will be used to track the channel alterations. However, in a fading environment it can happen that the instantaneous channel gain becomes so weak that it disappears in the noise for a few symbols. Adaptive algorithms easily lose performance in such situations. If only one fade exists in a data burst, the algorithm can run forward beginning with a training preamble and backwards with the preamble of the consecutive burst, or like in GSM where a training midamble exists in the forward and backward direction. Forward-backward DFE structures typically offer advantages compared to just unidirectional adaptation [29], [30].

1.5 Channel Estimation

One of the major problems with the previously considered equalizers is the typically high correlation of the received sequence, hampering the learning rate. When using transmitted symbols, the learning rate can be expected to be much higher since the transmitted symbols are typically uncorrelated (white) and can be considered to be a statistically independent sequence. Consider the received signal vector

$$\boldsymbol{r}(k) = \boldsymbol{C}\boldsymbol{s}(k) + \boldsymbol{v}(k) = \boldsymbol{S}(k)\boldsymbol{c} + \boldsymbol{v}(k), \tag{1.81}$$

where the first form describes the channel in matrix form and the transmitted symbols are organized in a vector, while in the second form, this description has been swapped. Once, the training sequence $\boldsymbol{S}(k)$ is known, it can be used to estimate the channel impulse response $\boldsymbol{c}$. The *least-squares* (LS) estimator is given by

$$\hat{\boldsymbol{c}}_{\mathrm{LS}} = [\boldsymbol{S}^H(k)\boldsymbol{S}(k)]^{-1}\boldsymbol{S}^H(k)\boldsymbol{r}(k). \tag{1.82}$$

Under the assumption of additive white noise $\boldsymbol{v}(k)$, the LS estimate is known to be the *best linear unbiased estimator* (BLUE), i.e., the estimator with lowest error variance[9]. More importantly, under Gaussian noise, the LS estimator achieves the Cramer-Rao (CR) bound [21], thus delivering an efficient estimate that cannot be improved by any other unbiased estimator [31]. Its variance is given by trace$([\boldsymbol{S}^H(k)\boldsymbol{S}(k)]^{-1})$ and thus, the CR bound is smallest if $\boldsymbol{S}^H(k)\boldsymbol{S}(k) = \boldsymbol{I}$. Hence, orthogonal sequences are of highest interest for fastest training [31], [32].

1.5.1 Channel Estimation in MIMO Systems

In MIMO systems with M transmit and N receive antennas it is of interest to estimate all MN paths simultaneously. Thus, M transmit sequences have to be found that do not interfere with each other. In [33], it is shown that for optimal MIMO training trace$([\boldsymbol{S}^H(k)\boldsymbol{S}(k)]^{-1})$ must be minimized, where $\boldsymbol{S}(k)$ is a matrix of the form

$$\boldsymbol{S}(k) \triangleq \begin{bmatrix} s_1(L) & \dots & s_1(L-L_C+1) & \dots & s_M(L) & \dots & s_M(L-L_C+1) \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ s_1(L_C) & \dots & s_1(1) & \dots & s_M(L_C) & \dots & s_M(1) \end{bmatrix},$$

in which all M training sequences of length L for estimating channels of length $L_C < L$ are combined. It turns out that the minimum LS error is obtained if and only if $\boldsymbol{S}^H(k)\boldsymbol{S}(k) = (L - L_C + 1)\boldsymbol{I}$ (a necessary condition being $L - L_C + 1 \geq ML_C$) and that such minimum error is given by

$$\min_{\boldsymbol{S}(k)} E\left[\|\boldsymbol{c} - \hat{\boldsymbol{c}}_{\mathrm{LS}}\|^2\right] = \frac{ML_C}{L - L_C + 1}\sigma_v^2, \tag{1.83}$$

allowing the interesting interpretation that the error increases proportional to the growing number of transmit antennas, while a growing training length L can compensate for this effect.

[9] Without any further a-priori knowledge on given statistics.

1.5.2 Estimation of Wireless Channels

Depending on the wireless environment, channel impulse responses can be as short as 25 ns (in small rooms [34]) and in mountainous areas as long as 60 μs. Thus, given the data rate and modulation scheme, a certain symbol length is defined and if it is much smaller than the channel impulse response, the channel estimation vector can be expected to be of large dimension. However, typical wireless channels display a specific structure. Most of the energy is contained in only a few taps of the impulse response. If concentrating only on such points, most of the channel energy with much less complexity can be captured. Typically, four positions (called fingers) are sufficient; up to ten may be used. A receiver architecture exploiting such channel structure is called the RAKE receiver [18]. Again, ZF or MMSE techniques can be applied to reconstruct the transmitted symbol with a channel equalizer, or maximum-likelihood and Viterbi techniques can be used based on the channel information. Note, however that in addition to the low complexity RAKE structure, further algorithms are required to find the optimal finger locations and possibly track such locations when the channel is time-varying.

Once the channels are not static but time varying, as is expected in mobile communications, the tracking behavior of adaptive algorithms becomes important. For small movement ($f_D/f_C < 10^{-7}$, with f_D the Doppler and f_C the carrier frequency) typical algorithms like LMS and RLS track channel quite well [18], [21]. However, as the mobile movement becomes larger, standard algorithms can no longer track well. Of particular interest in recent years is the estimation of rapidly changing channels. Since mobiles are expected to move with speeds up to 350 kmph (fast TGV Train in France, for example), channel estimation becomes challenging. Some methods to improve estimation quality in this environment will be discussed in the following.

1.5.3 Channel Estimation by Basis Functions

An approach for achieving better results when applying channel estimation techniques to rapidly changing channels is the utilization of basis functions [35], [36]. If the channel is considered only for a limited time period, each coefficient $c_i(k); i = 0, ..., L_C - 1$ is assumed to vary in time, and thus describes a function in time in its real and imaginary parts. Such a function can be approximated by orthogonal basis functions, the simplest of them being the exponential function. A model for one coefficient over a certain time period can thus be written

$$c_i(k) = \sum_{l=0}^{L_g-1} a_{il} \exp(j\Omega_l k) = \boldsymbol{a}_i^T \boldsymbol{g}(k), \tag{1.84}$$

where the coefficients a_{il} are gathered in a vector $\boldsymbol{a}_i$ and the exponential terms are based on the frequencies Ω_l in a second time-varying vector $\boldsymbol{g}(k)$. The received signal is thus given by

$$\begin{aligned} r(k) &= \sum_{i=0}^{L_C-1} c_i(k)s(k-i) + v(k) \\ &= \sum_{i=0}^{L_C-1} \boldsymbol{a}_i^T \boldsymbol{g}(k)s(k-i) + v(k) \\ &= \boldsymbol{a}^T \tilde{\boldsymbol{s}}(k) + v(k), \end{aligned} \tag{1.85}$$

in which the vector $\boldsymbol{a}$ consists of all time-constant parameters a_{il} and $\tilde{\boldsymbol{s}}(k)$ of all time varying components $\exp(j\Omega_l k)s(k-i)$. If $\tilde{\boldsymbol{s}}(k)$ is known *a-priori*, the unknown vector $\boldsymbol{a}$ can be estimated from the observations $r(k)$ over a time period larger than the dimension of $\boldsymbol{a}$ [36]. Another approach allows us to reformulate the received vector in matrix form:

$$\boldsymbol{r}(k) = \boldsymbol{G}(k)\boldsymbol{A}\boldsymbol{s}(k) + \boldsymbol{v}(k), \tag{1.86}$$

with new matrices $\boldsymbol{G}(k)$ and $\boldsymbol{A}$ containing the basis functions and the column vectors $\boldsymbol{a}_i$, respectively. The matrix $\boldsymbol{G}(k)$ exhibits the particular structure of an FFT matrix when $\Omega_l = \Omega_o l$, in which case the various basis functions can be interpreted as Doppler spectral components of the time-varying channel. Thus, its inverse is simply given by $\boldsymbol{G}^H(k)$ and can be applied to the received sequence $\boldsymbol{r}(k)$, making the values independent of the time variations. The matrix elements of $\boldsymbol{A}$ can then be estimated by conventional estimation techniques like LS. By reformulating $\boldsymbol{A}\boldsymbol{s}(k)$ into $\boldsymbol{S}(k)\boldsymbol{a}$, the LS estimate is $\hat{\boldsymbol{a}}_{LS} = [\boldsymbol{S}^H(k)\boldsymbol{S}(k)]^{-1}\boldsymbol{S}^H(k)\boldsymbol{G}^H(k)\boldsymbol{r}(k)$. Note that this is a particularly low-complexity solution since $\boldsymbol{G}^H(k)\boldsymbol{r}(k)$ can be computed by an FFT and $[\boldsymbol{S}^H(k)\boldsymbol{S}(k)]^{-1}\boldsymbol{S}^H(k)$ can be pre-computed for the training mode.

1.5.4 Channel Estimation by Predictive Methods

Another class of adaptive algorithms includes knowledge of the time varying process of the channel filter coefficients [37]. Here, a model of this process is incorporated into the adaptive process, for example by Kalman filtering [38]. The channel coefficients can be described by

$$\boldsymbol{c}(k) = \boldsymbol{A}(k)\boldsymbol{c}(k-1) + \boldsymbol{u}(k), \tag{1.87}$$

where $\boldsymbol{A}(k)$ describes the transition of the channel states from time instant $k-1$ to k and $\boldsymbol{u}(k)$ is a driving process. If the channel is described by statistic methods, $\boldsymbol{A}(k) = \boldsymbol{A}$ can be as simple as a filter causing a particular autocorrelation function and spectrum of the elements $c_i(k)$ of $\boldsymbol{c}(k)$. If the time-variation of the channel is caused by two-dimensional motion of the transmitter and/or receiver, the so-called Jakes-spectrum is obtained [39].

The optimal adaptive filter in such situations [described by (1.87)] is the Kalman filter [40]. However, this requires not only knowledge of the filter structure of $\boldsymbol{A}(k)$ but also precise knowledge of its parameters. A similar but complexity-simpler approach is to apply the *Wiener least-mean-square* (WLMS) algorithm [41], [42]. Here, the dynamic behavior of the channel is described by a simpler one-dimensional autoregressive (AR) model which is included in the parameter estimation part.

If not even the model structure is known, simple predictive methods [43] can be utilized. The estimates from LMS filtering, for example, can be linearly combined by

$$\hat{\hat{\boldsymbol{c}}}(k) = \sum_{l=0}^{L_p} \gamma_l \hat{\boldsymbol{c}}(k-l-D), \tag{1.88}$$

with a fixed positive delay D defining the prediction order. Having statistical knowledge about the random process $\boldsymbol{c}(k)$, the optimal coefficients γ_l can be pre-computed. Good results were obtained with the simple approach

$$\hat{\hat{\boldsymbol{c}}}(k) = \hat{\boldsymbol{c}}(k-D) + \frac{p}{L_p}\left[\hat{\boldsymbol{c}}(k-D) - \hat{\boldsymbol{c}}(k-D-L_p)\right], \tag{1.89}$$

where p is the prediction step-size. Obviously, the parameters of the random process primarily depend on the selection of the step-size p.

Other approaches do not exhibit the prediction process in the algorithmic modification. The *proportionate LMS* (PLMS) [26], [27], for example, assigns individual step-sizes according to the energy of the various weights. This can also be interpreted as a predictive method.

1.6 Maximum Likelihood Equalization

Having available a good estimate of the channel coefficients simplifies the detection process considerably. In this case, an optimal detection method can be considered. The *maximum a-posteriori probability* (MAP) detector is known to provide the best performance [18]. Its decision is based on the Bayes' rule

$$P(\boldsymbol{s}_m|\boldsymbol{r}) = \frac{P(\boldsymbol{r}|\boldsymbol{s}_m)P(\boldsymbol{s}_m)}{P(\boldsymbol{r})}, \tag{1.90}$$

where $P(\boldsymbol{r}|\boldsymbol{s}_m)$ is the conditional PDF of the observed vector $\boldsymbol{r}$ given the transmitted signal $\boldsymbol{s}_m$. In the MAP approach $P(\boldsymbol{s}_m|\boldsymbol{r})$ has to be maximized in order to find the most likely sequence $\boldsymbol{s}_m$. Since $P(\boldsymbol{r}) = \sum_m P(\boldsymbol{r}|\boldsymbol{s}_m)P(\boldsymbol{s}_m)$, this term deserves no further attention. Assuming that all sequences are transmitted with the same probability $P(\boldsymbol{s}_m)$, the expression simplifies and the so-called *maximum likelihood* (ML) estimator is obtained. In case of additive noise, the conditional PDF $P(\boldsymbol{r}|\boldsymbol{s}_m)$ can be computed. Assuming a noise with Gaussian distribution, maximizing $P(\boldsymbol{s}_m|\boldsymbol{r})$ becomes equivalent to minimizing the Euclidian distance $\|\boldsymbol{r} - \boldsymbol{s}_m\|$. The principle can be applied to transmission with ISI. In this case, the minimization has to include the channel information, i.e.,

$$\hat{\boldsymbol{s}}_{\mathrm{ML}} \stackrel{\Delta}{=} \arg\min_{\boldsymbol{s}\in\mathcal{A}} \|\boldsymbol{r} - \boldsymbol{C}\boldsymbol{s}\|^2 . \tag{1.91}$$

Suppose a sequence of L symbols with a P-ary alphabet $\mathcal{A}$ has been transmitted. Then, the minimization has to be performed over P^L possible values, a complexity growing exponentially with the number of transmitted symbols. Such high complexity is the reason why brute-force ML equalization has not been used much in the past. However, recently, ML as a realizable receiver technique has drawn more attention since with MIMO antenna systems, typically a small number of transmit and receive antennas are considered and thus an ML technique can be used to search for the most likely instantaneous symbols [13], [12].

1.6.1 Viterbi Algorithm

A technique that allows performance close to the ML technique but with linear complexity is the *Viterbi algorithm* (VA) [18], [44], [45], [46]. The VA structures the ML search recursively based on a finite state machine (FSM) description for the signal generation. Such an FSM is given in the context of a channel of length L_C with the past symbol values $s(k-1), s(k-2), ..., s(k-L_C)$ as the states. If the next symbol $s(k)$ is to be detected, its impact on all possible L_C positions of the estimated channel filter has to be computed, i.e, roughly P^{L_C} operations. Among those possibilities, only the P best ones, i.e., those with smallest metric, are selected, called survivor paths, and based on those survivor metrics, the next symbol is investigated. Thus, for L symbols, the complexity is roughly LP^{L_C} and for $L_C \ll L$, the VA has much lower complexity than ML. Once all symbols have been taken into account, a backward algorithm is required to find the optimal symbol sequence in the set of smallest metrics.

An important issue is the delay until the detected symbol is available. It turns out that with high probability the survivor paths agree on a symbol $s(k)$ after D steps as long as $D \geq 5L_C$. If they do not agree, the decision can be based on the most probable path with smallest metric. The Viterbi algorithm can be used to reduce the error propagation in a DFE structure. Since the length L_B of a DFE filter can be much shorter than the actual channel length L_C, the complexity of the VA becomes smaller. Such *reduced-state sequence estimation* (RSSE) techniques have been exploited in [47], [18] where also trellis codes have been considered. In a recent development [48] a tap-selective DFE has been implemented in which only the most energy carrying coefficients are utilized in the search, reducing the states considerably. For time varying channels, however, the latency in the decision of the VA can be quite prohibitive. The minimum decision latency of L_B can further be reduced by *parallel decision feedback decoding* (PDFD) techniques [49], running several DFE structures in parallel and selecting at a later stage the most likely one. Another promising technique to deal with time-varying channels combines the concept of basis functions [see (1.5.3)] with the VA [50].

1.7 Blind Algorithms

When training signals are entirely absent, the transmission is called blind, and adaptive algorithms for estimating the transferred symbols and possibly estimating channel or equalizer information are called blind algorithms [6], [51]. Since training information is not available, a reliable reference is missing, leading to very slow learning behavior in such algorithms. Thus, blind methods are typically of interest when a large amount of data is available and quick detection not important. Their major application field is thus broadcasting for digital radio or TV. However, recently the concept of basis functions [see (1.5.3)] to describe time-variant channels has been incorporated, proving that blind techniques also have potential for time-varying channels, in particular, for MIMO transmissions [52].

Various principles allowing blind algorithms to successfully detect the symbols can be distinguished:

1. Algorithms utilizing the *constant modulus* (CM) property of the transmitted signal [53], [54], [55] are historically probably the first blind algorithms. The *constant modulus algorithm* (CMA)

$$\boldsymbol{f}^*(k) = \boldsymbol{f}^*(k-1) + \mu(k)\boldsymbol{x}^*(k)y(k)g[y(k)], \tag{1.92}$$

$$y(k) = \boldsymbol{f}^H(k-1)\boldsymbol{x}(k), \tag{1.93}$$

   is the most well-known procedure and depends on the nonlinear function[10] $g[\cdot]$ in many variations [56], [57] and applications [58]. While the convergence analysis of such algorithms is limited to very few cases [56], the analysis of its tracking behavior, i.e., its steady-state performance has made progress. In [57], the feedback approach from [56] has been extended and conditions were presented under which the steady state error can be computed. In particular, for CMA1-2 and CMA2-2, it was shown that the proper selection of $g[\cdot]$ can lead to improved performance.
2. Algorithms based on *higher order statistics* (HOS), also called higher order moments, have been introduced by Shalvi and Weinstein [59]. In particular, the kurtosis, i.e., the ratio of the fourth-order moment to the squared second-order moment is of interest (also in the context of multiuser detection [8]). This ratio is 3 for Gaussian sequences but typically

[10] The nonlinear function used is mostly of the form $g[y] = \gamma - |y|^n$.

smaller for transmission symbols. It has been recognized that most algorithms based on the CM property also satisfy this condition [53], [54], [55].

3. *Second Order Statistics* (SOS) [60] usually do not carry phase information and thus can not be used to identify linear systems. If more than one transmission channel is present, however, missing phase information can be delivered by SOS techniques. A simple example is a two channel scenario transmitting the same sequence $s(k)$ on both channels $C_1(q^{-1})$ and $C_2(q^{-1})$. In the absence of noise, the received signals are $r_1(k) = C_1(q^{-1})s(k)$ and $r_2(k) = C_2(q^{-1})s(k)$. Thus, $C_2(q^{-1})r_1(k) = C_1(q^{-1})r_2(k)$. The selection of

$$[\boldsymbol{R}_1, \boldsymbol{R}_2] \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix} = \mathbf{0}, \tag{1.94}$$

guarantees a unique solution[11] up to a constant if $C_1(q^{-1})$ and $C_2(q^{-1})$ do not share any common zeroes, where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are Toeplitz matrices of the received sequences $r_1(k)$ and $r_2(k)$. In the presence of noise, this condition can be modified to

$$\arg \min_{||\boldsymbol{c}_1||^2 + ||\boldsymbol{c}_2||^2 = 1} \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}^H [\boldsymbol{R}_1, \boldsymbol{R}_2]^H [\boldsymbol{R}_1, \boldsymbol{R}_2] \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix}. \tag{1.95}$$

The solution of such a problem is typically found by *singular-value decomposition* (SVD) methods. SVD methods (also called subspace techniques) are of high complexity and usually numerically challenging. Not only is an extension of (1.95) to multiple channels possible, but also various variants exist, using cross-correlation functions or spectra. A good overview of such techniques can be found in [61], [10]. An interesting modification for MIMO transmission is the *space-time matrix modulation* method (STMM) [62]. The received signal consists of

$$\boldsymbol{r}(k) = \boldsymbol{C} \sum_{l=1}^{L} \boldsymbol{m}_l(k) s_l(k) + \boldsymbol{v}(k), \tag{1.96}$$

where $\boldsymbol{m}_l(k)s_l(k)$ is a term of the mixture matrix combining L different data streams onto a certain number of transmit antennas. With a fixed mixture vector sequence $\boldsymbol{m}_l(k)$ known at the transmitter and receiver it can be shown that the channel $\boldsymbol{C}$ and the symbols $s_l(k)$ can be separated uniquely up to a constant. The advantage of STMM is that not only SVD methods can be applied, but also ML and even a much simpler projection algorithm leads to successful equalization [63]. Combinations with basis function approaches are possible as well [63], [64].

[11] In order to obtain a unique solution there is also a persistent excitation condition on the input signal $s(k)$. If the two channels do share at least one common zero, there are multiple solutions.

4. *Blind ML methods* try to estimate the channel or, alternatively, the equalizer, and the transmitted data sequence at the same time. Hereby, two criteria are common

$$\min_{\{\boldsymbol{c},\boldsymbol{s}\}} \|\boldsymbol{r} - \boldsymbol{C}\boldsymbol{s}\|^2 = \min_{\{\boldsymbol{c},\boldsymbol{s}\}} \|\boldsymbol{r} - \boldsymbol{S}\boldsymbol{c}\|^2 \tag{1.97}$$

$$\min_{\{\boldsymbol{f},\boldsymbol{s}\}} \|\boldsymbol{s} - \boldsymbol{F}\boldsymbol{r}\|^2 = \min_{\{\boldsymbol{f},\boldsymbol{s}\}} \|\boldsymbol{s} - \boldsymbol{R}\boldsymbol{f}\|^2 , \tag{1.98}$$

where $\boldsymbol{C}$ and $\boldsymbol{F}$ are Toeplitz matrices of the channel $\boldsymbol{c}$ and equalizer $\boldsymbol{f}$, respectively, and $\boldsymbol{S}$ and $\boldsymbol{R}$ those of the sequences $\boldsymbol{s}$ and $\boldsymbol{r}$, respectively. Typically, one minimizes the first term with a fixed channel/equalizer and then the second with a fixed data sequence, and runs such procedure several times until convergence is observed. Hereby, the Toeplitz structure of the matrices, as well as the property that $s(k)$ stems from a finite alphabet, is utilized to optimize the technique. In [6], [61] the first criterion is utilized while the second can be found in [65]. In order to succeed, a good starting value is required. This is usually achieved by a very short training sequence, hence such techniques are called semi-blind.

1.8 Conclusions

In this chapter, an overview of adaptive equalizer techniques was presented. Special emphasis was given to techniques applied in modern wireless systems where channels are frequency- and time-dispersive. Many basic concepts were explained and brought into the context of multiple-input multiple-output systems as they appear in the near future in wireless communication systems. A short overview of blind techniques was given demonstrating the potential of new signal processing techniques even better suited for the particular needs of wireless communications.

Acknowledgment

The authors would like to thank Harold Artés for his carefully reading of the manuscript and pointing out many inconsistencies.

References

1. R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, Apr. 1965.
2. G. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Information Theory*, vol. 18, no. 3, pp. 363–378, 1972.
3. J. M. Cioffi, G. Dudevoir, M. Eyuboglu, and G. D. Forney, Jr, "MMSE decision feedback equalization and coding–Part I," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

4. N. Al-Dhahir and J. M. Cioffi, "MMSE decision feedback equalizers: finite length results," *IEEE Trans. Information Theory*, vol. 41, no. 4, pp. 961–975, July 1995.
5. J. R. Treichler, I. Fijalkow, and C.R.Johnson, Jr, "Fractionally spaced equalizers," *IEEE Signal Processing Mag.*, pp. 65–81, May 1996.
6. J. K. Tugnait, L. Tong, and Z.Ding, "Single user channel estimation and equalization," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 17–28, May 2000.
7. P. A. Fuhrman, *A Polynomial Approach to Linear Algebra*. Springer, N.Y., 1996.
8. S. Haykin, *Unsupervised Adaptive Filtering*. Wiley-Interscience, N.Y., 2000.
9. T. Kailath, *Linear Systems*. Prentice Hall, Englewood Cliffs, N.J., 1980.
10. G. B. Giannnakis et al., *Signal Processing Advances in Wireless & Mobile Communications*. Prentice Hall, vol. 1, 2001.
11. P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for achieving very high data rates over rich-scattering wireless channels," in *Conf. Record of ISSSE*, Pisa, Italy 98.
12. A. Burg, E. Beck, M. Rupp, D. Perels, N. Felber, and W. Fichtner, "FPGA implementation of a MIMO receiver front-end for UMTS," in *Conf. Record of Int. Zuerich Seminar on Broadband Commun.*, 2002, pp. 8_1–8_6.
13. R. Van Nee, A. van Zelst, and G. Awater, "Maximum likelihood decoding in a space division multiplexing system," in *Conf. Record of VTC Japan*, 2000.
14. B. A. Bjerke and J. G. Proakis, "Equalization and decoding for multiple-input multiple-output wireless channels," *Eurasip Journal on Applied Signal Processing*, vol. 3, pp. 249–266, 2002.
15. J. Salz, "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, no. 8, Oct. 1973.
16. C. B. Papadias and A. J. Paulraj, "Unbiased decision feedback equalization," in *Conf. Record of the IEEE Intern. Symp. on IT*, 1998, pp. 448.
17. C. B. Papadias and M. Rupp, "Performance analysis of finite-length DFE receivers based on a polyphase representation," in *Conf. Record of the 32nd. Asilomar Conf. of Signals, Systems and Computers*, 1998, pp. 374–378.
18. J. G. Proakis, *Digital Communications*. McGraw Hill, Fourth edition, 2001.
19. I. Ghauri and D. T. M. Slock, "Linear receivers for the DS-CDMA downlink exploiting orthogonality of spreading sequences," in *Conf. Record of the 32nd. Asilomar Conf. of Signals, Systems and Computers*, 1998.
20. M. Rupp, "Normalization and convergence of gradient-based algorithms for adaptive IIR filters," *Signal Processing*, vol. 46, no. 1, pp. 15–30, Sept. 1995.
21. S. Haykin, *Adaptive Filter Theory*. Fourth Edition, Prentice Hall, 2001.
22. M. Rupp, "On the learning behavior of decision feedback equalizers," in *Conf. Record of the 33rd. Asilomar Conf. of Signals, Systems and Computers*, vol. 1, 1999, pp. 514–518.
23. A. P. Liavas and P. A. Regalia, "On the numerical stability and accuracy of the conventional recursive least squares algorithm," *IEEE Trans. Signal Processing*, pp. 88–96, Jan. 1999.
24. H. Mohamad, S. Weiss, M. Rupp, and L. Hanzo, "A fast converging fractionally spaced equalizer," in *Conf. Record of the 35th Asilomar Conf. of Signals, Systems and Computers*, 2001.
25. H. Mohamad, S. Weiss, M. Rupp, and L. Hanzo, "Fast adaptation of fractionally spaced equalizers," *Electronic Letters*, vol. 38, no. 2, pp. 96–98, Jan. 17, 2002.

26. S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Conf. Record of the 32nd. Asilomar Conf. of Signals, Systems and Computers*, 1998, pp. 394–398.
27. M. Rupp and J. Cezanne, "Robustness conditions of the LMS algorithm with time-variant matrix step-size," *Signal Processing*, vol. 80, no. 9, pp. 1787–1794, Sept. 2000.
28. M. Rupp and A. H. Sayed, "Robustness and convergence of adaptive schemes in blind equalization," in *Conf. Record of the 30th. Asilomar Conf. on Signals, Systems and Computers*, vol. 1, 1996, pp. 271–275.
29. A. Bahai and M. Rupp, "Adaptive DFE algorithms for IS-136 based TDMA cellular phones," in *Conf. Record of IEEE International Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, 1997, pp. 2489–2492.
30. J. Balakrishnan and C. R. Johnson, Jr.,"Time-reversal diversity in decision feedback equalization," in *Conf. Record of Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), 2000.
31. S. N. Crozier, D. D. Falconer, and S. A. Mahmoud, "Least sum of squared errors channel estimation," *IEE Proc. F*, vol. 138, no. 4, pp. 371–378, Aug. 1991.
32. M. Rupp, "Fast implementation of the LMS algorithm," in *Conf. Record of Eusipco*, Tampere, 2000.
33. J. Balakrishnan, M. Rupp, and H. Vishwanatan, "Optimal channel training for multiple antenna systems," in *Conf. Record of Multiaccess, Mobility and Teletraffic for Wireless Communications*, 2000.
34. T. S. Rappaport, *Wireless Communications*. Prentice Hall, 1996.
35. L. Greenstein and B. Czekaj, "Modeling multipath fading responses using multitone probing signals and polynomial approximation," *Bell Syst. Tech. J.*, vol. 60, pp. 193–214, 1981.
36. M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adaptive Control Signal Processing*, vol. 10, pp. 159–176, 1996.
37. A. Duel-Hallen, S. Hu, and H. Hallen, "Long range prediction of fading signals," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.
38. R. A. Iltis and A. W. Fuxjaeger, "A digital DS spread-spectrum receiver with joint channel and Doppler shift estimation," *IEEE Trans. Commun.*, vol. 39, no. 8, Aug. 1991,
39. W. C. Jakes, *Microwave Mobile Communication*. IEEE Press, 1974.
40. T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 1999.
41. L. Lindbom, M. Sternad, and A. Ahlen, "Tracking of time-varying mobile radio channels: part I. The Wiener LMS algorithm," *IEEE Trans. Commun.*, pp. 2207–2217, Dec. 2001.
42. L. Lindbom, A. Ahlen, M. Sternad, and M. Falkenstrom, "Tracking of time-varying mobile radio channels: part II. A case study," *IEEE Trans. Commun.*, pp. 156–167, Jan. 2002.
43. M. C. Chiu and C. Chao, "Analysis of LMS-adaptive MLSE equalization on multipath fading channels," *IEEE Trans. Commun.*, pp. 1684–1692, Dec. 1996.
44. A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Information Theory*, vol. IT-13, Apr. 1967.

45. E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Kluwer, 2nd edition, 1994.
46. H. L. Lou, "Implementing the Viterbi algorithm," *IEEE Signal Processing Mag.*, vol. 12, no. 5, pp. 42–52, Sept. 1995.
47. M. Eyuboglu and S. Qureshi, "Reduced-State sequence estimation for coded modulation on intersymbol interference channels," *IEEE Journal Sel. Areas Commun.*, vol. 7, pp. 989–995, Aug. 1989.
48. E. F. Haratsch, A. J. Blanksby, and K. Azadet, "Reduced state sequence estimation with tab-selective decision feedback," in *IEEE Int. Conf. on Commun.*, vol. 1, 2000, pp. 372–376.
49. H. L. Lou, M. Rupp, R. L. Urbanke, H. Viswanatan, and R. Krishnamoorthy, "Efficient implementation of parallel decision feedback decoders for broadband applications," in *Conf. Record of the 6th IEEE Int. Conf. on Electronics, Circuits and Systems*, vol. 3, pp. 1475–1478, 1999.
50. J. Bakkoury, D. Roviras, M. Ghogho, and F. Castanie, "Adaptive MLSE receiver over rapidly fading channels," *Signal Processing*, vol. 80, pp. 1347–1360, 2000.
51. S. Haykin, *Blind Deconvolution*. Prentice Hall, Englewood Cliffs, N.J., 1994.
52. H. Liu and G. B. Giannakis, "Deterministic approaches for blind equalization of time-varying channels with antenna arrays," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3003–3013, Nov. 1998.
53. Y. Sato, "A method of self-recovering equalization for multi level amplitude modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.
54. D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.
55. G. J. Foschini, "Equalization without altering or detecting data," *AT&T Tech. Journal*, vol. 64, pp. 1885–1911, 1985.
56. M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant modulus signals," *IEEE Trans. Comm.*, vol. 48, no. 5, pp. 795–803, May 2000.
57. J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally spaced blind equalizers," *IEEE Trans. Signal Processing*, vol. 48, no. 1, pp. 80–91, Jan. 2000.
58. J. Treichler and C. R. Johnson, Jr., "Blind fractionally spaced equalization of digital cable TV," in *Conf. Record of the 7th IEEE DSP Workshop*, 1996, pp. 122–130.
59. O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Information Theory*, vol. IT-39, pp. 292–297, Jan. 1990.
60. L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Conf. Record of the 25th. Asilomar Conf. of Signals, Systems and Computers*, 1991.
61. L. Tong and S. Perreau, "Multichannel blind channel estimation: from subspace to maximum likelihood methods," *Proc. of IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.
62. H. Artes and F. Hlawatsch, "Blind equalization of MIMO channels using deterministic precoding," in *Conf. Record of IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, vol. 4, 2001, pp. 21530–2156.

63. H. Artes, F. Hlawatsch, and G. Matz "Efficient POCS algorithms for deterministic blind equalization of time-varying channels," in *Conf. Record of IEEE Globecom*, 2000, pp. 1031–1035.
64. H. Artes and F. Hlawatsch, "Blind multiuser equalization equalization for time-varying channels," in *Conf. Record of third IEEE Signal Processing Workshop SPAWC*, 2001, pp. 102–105.
65. J. Laurila, R. Tschofen, and E. Bonek, "Semi-blind space-time estimation of co-channels signals using least squares projections," in *Conf. Record of the 50th IEEE Vehicular Technology*, vol. 3, 1999, pp. 1310–1315.

Index