1 Channel Estimation

GEERT LEUS and ALLE-JAN VAN DER VEEN†

1.1 INTRODUCTION

As demonstrated in other chapters in this book, the deployment of multiple antennas at the transmit and receive side (multiple-input multiple-output (MIMO)) can result in a significant capacity increase. This is due to two effects: (i) diversity, i.e., robustness against fading of the channel between a transmit and a receive antenna, and (ii) space-time coding, i.e., the parallel transmission of information via multiple transmit antennas. However, this capacity increase was based on an important assumption: all channels between the transmit antennas and the receive antennas are accurately known. In practice, these channels will have to be estimated, which is the focus of this chapter.

The wireless channel is highly complex. In general it is both frequency- and time-selective, and with multiple antennas, also the space-selectivity plays a role. Physical models such as Jakes' model [1] usually simplify this to a multipath propagation model where each path is parametrized by an angle at the receiver array, perhaps an angle at the transmitter array, and further a propagation delay and a complex amplitude. This can be refined by making statistical assumptions on the distribution of these parameters. For channel *modeling*, one tries to use a general model that allows to describe a large class of observed channels. For channel *estimation*, however, there is a trade-off: a sophisticated model with more parameters may turn out to be less accurate when the parameters have to be estimated with a finite set of observations.

It is clear that channel estimation is an extensive topic. To limit ourselves, we will cover only a small subset of channel models and possible estimation techniques:

- *Channel model: FIR-MIMO*. For broadband communications, the time-dispersion or frequency-selectivity due to multipath is important. For the sake of conciseness, we will restrict ourselves to single-carrier MIMO systems in a frequency-selective fading channel. The channels are modeled by simple finite impulse response (FIR) filters with a common order L, assumed to be known.

- *Estimation techniques: training-based and semi-blind.* In practical systems, channels are invariably estimated using periodic bursts of known training sym-

†Delft University of Technology, Dept. EEMCS, Mekelweg 4, 2628 CD Delft, The Netherlands

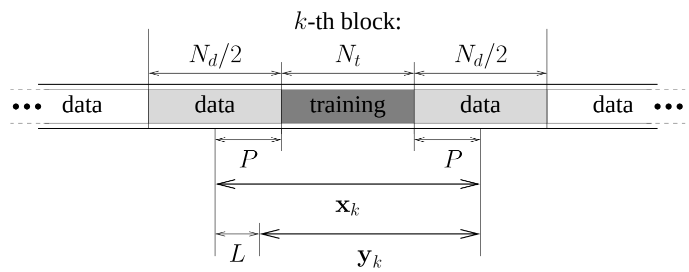


Fig. 1.1 Partitioning of the transmitted symbol vectors into blocks, each consisting of N_t training and N_d data symbol vectors.

> bols, therefore we focus mostly on these techniques. Conventional training-based methods only exploit the presence of the known training symbols. The results can be enhanced by also incorporating the convolutional properties of the surrounding unknown data symbols, which lead to so-called enhanced training-based methods. Also discussed are semi-blind methods that combine a training-based criterion with a purely blind criterion. Blind techniques do not exploit the knowledge of training symbols, and focus on deterministic or stochastic properties of the system. Note that all channel estimation methods considered in this chapter are *transparent to space-time coding*, i.e., any structure introduced by these codes is not exploited.

Suggestions for further reading are found at the end of the chapter.

Notation

Matrices and column vectors are written in boldface uppercase and lowercase, respectively. For a matrix or column vector, superscript ^{T} is the transpose, ^{H} the complex conjugate transpose, and † the pseudo-inverse (Moore-Penrose inverse). $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\mathbf{0}_{M\times N}$ ($\mathbf{0}_N$) is the $M \times N$ ($N \times N$) matrix for which all entries are equal to zero. $\mathrm{vec}(\mathbf{A})$ is a stacking of the columns of a matrix $\mathbf{A}$ into a column vector. $\|\cdot\|$ represents the Frobenius norm. $\otimes$ is the Kronecker product. A notable property is (for matrices of compatible sizes): $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$. Finally, $\mathrm{E}(\cdot)$ denotes the stochastic expectation operator.

1.2 DATA MODEL

Let us consider a convolutive MIMO system with A_t transmit antennas and A_r receive antennas. Suppose $\mathbf{x}(n)$ represents the $A_t \times 1$ symbol vector sequence transmitted at the A_t transmit antennas. Assuming symbol rate sampling at each receive antenna, the $A_r \times 1$ sample vector sequence received at the A_r receive antennas is then given by

$$\mathbf{y}(n) = \sum_{l=0}^{L} \mathbf{H}(l)\mathbf{x}(n-l) + \mathbf{e}(n), \tag{1.1}$$

where $\mathbf{e}(n)$ is the $A_r \times 1$ additive noise vector sequence on the A_r receive antennas, which we assume to be zero-mean white (spatially and temporally) Gaussian with variance σ_e^2, and $\mathbf{H}(l)$ is the $A_r \times A_t$ MIMO channel of order L (or length $L+1$). We will often make use of the vectorized form of $\mathbf{H}(l)$, which is obtained by stacking its columns: $\mathbf{h}(l) = \text{vec}[\mathbf{H}(l)]$.

In this chapter, we focus on estimating $\mathbf{H}(l)$ (or $\mathbf{h}(l)$) without assuming any structure on it. Hence, no calibration of the different transmit/receive antennas is required. We assume a burst of N symbol vectors is transmitted, in the form of K symbol blocks, where each symbol block consists of N_t training symbol vectors, surrounded at each side by $N_d/2$ unknown data symbol vectors, i.e., $N = K(N_t + N_d)$ (see figure 1.1). We will focus on training-based as well as semi-blind channel estimation algorithms, where the latter rely on the combination of a training-based with a purely blind criterion.

To describe the different training-based methods that will be discussed in this chapter, consider the following subvector of the k-th symbol block, as illustrated in figure 1.1:

$$\mathbf{x}_k = [\mathbf{x}^T(n_k - P), \dots, \mathbf{x}^T(n_k + N_t + P - 1)]^T, \tag{1.2}$$

where $n_k = k(N_t + N_d) + N_d/2$ indicates the start of the training symbol vectors in the k-th symbol block, and $0 \le P \le N_d/2$. This vector contains all the N_t training symbol vectors transmitted during the k-th symbol block, plus P unkwown data symbol vectors at each side of it. Due to the convolutive channel, the first $L-1$ received sample vectors corresponding to $\mathbf{x}_k$ are contaminated by preceding data symbol vectors. Therefore, the received samples that depend only on $\mathbf{x}_k$ are given by (see figure 1.1)

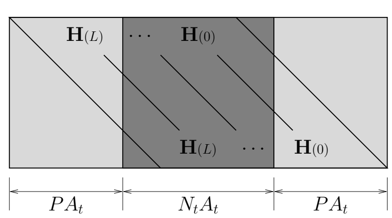
$$\mathbf{y}_k = [\mathbf{y}^T(n_k - P + L), \dots, \mathbf{y}^T(n_k + N_t + P - 1)]^T.$$

This vector can be expressed as

$$\mathbf{y}_k = \boldsymbol{\mathcal{H}}\mathbf{x}_k + \mathbf{e}_k, \tag{1.3}$$

where $\mathbf{e}_k$ is similarly defined as $\mathbf{y}_k$ and $\boldsymbol{\mathcal{H}}$ is the $A_r(N_t + 2P - L) \times A_t(N_t + 2P)$ block Toeplitz matrix representing the convolution by the channel:

$\boldsymbol{\mathcal{H}} =$



(1.4)

Define the training part of $\mathbf{x}_k$ as

$$\mathbf{x}_k^{(t)} = \left[\mathbf{x}^T(n_k), \ldots, \mathbf{x}^T(n_k + N_t - 1)\right]^T, \tag{1.5}$$

and the unknown data part as

$$\mathbf{x}_k^{(d)} = \left[\mathbf{x}^T(n_k - P), \ldots, \mathbf{x}^T(n_k - 1), \mathbf{x}^T(n_k + N_t), \ldots, \mathbf{x}^T(n_k + N_t + P - 1)\right]^T.$$

Then we can split (1.3) into

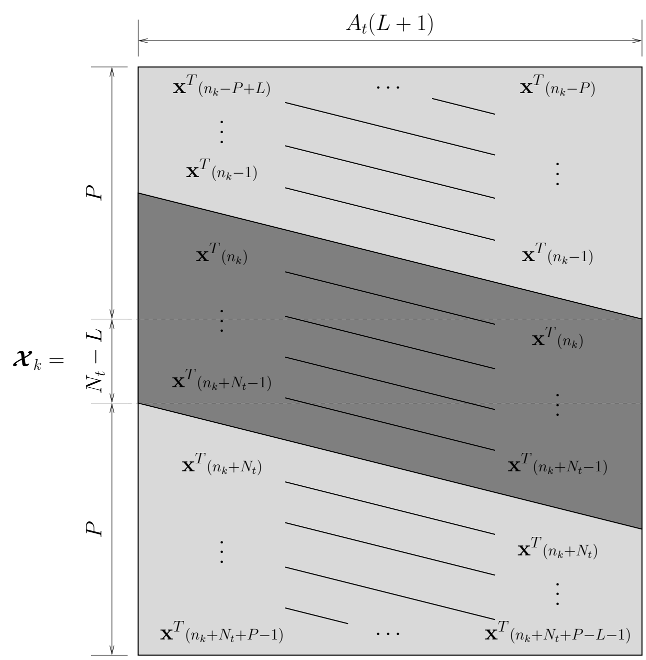
$$\mathbf{y}_k = \boldsymbol{\mathcal{H}}^{(t)}\mathbf{x}_k^{(t)} + \boldsymbol{\mathcal{H}}^{(d)}\mathbf{x}_k^{(d)} + \mathbf{e}_k, \tag{1.6}$$

where $\boldsymbol{\mathcal{H}}^{(t)}$ is the $A_r(N_t + 2P - L) \times A_t N_t$ matrix obtained by collecting the N_t middle block columns of $\boldsymbol{\mathcal{H}}$ (the dark shaded area in (1.4)), and $\boldsymbol{\mathcal{H}}^{(d)}$ is the $A_r(N_t + 2P - L) \times 2A_t P$ matrix obtained by collecting the P left and P right block columns (the light shaded area in (1.4)).

The preceding equations have expressed $\mathbf{y}_k$ as a linear combination of the transmitted symbols $\mathbf{x}_k$. Alternatively, we can write the convolution operation (1.3) as a linear operation on the channel coefficient vector $\mathbf{h} = [\mathbf{h}^T(0), \ldots, \mathbf{h}^T(L)]^T$, which gives

$$\mathbf{y}_k = (\boldsymbol{\mathcal{X}}_k \otimes \mathbf{I}_{A_r})\mathbf{h} + \mathbf{e}_k, \tag{1.7}$$

where $\boldsymbol{\mathcal{X}}_k$ is the $(N_t + 2P - L) \times A_t(L+1)$ block Toeplitz symbol matrix given by



(1.8)

Similar as in (1.6), we can split (1.7) into a training and an unknown data part as

$$\mathbf{y}_k = (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I}_{A_r})\mathbf{h} + (\boldsymbol{\mathcal{X}}_k^{(d)} \otimes \mathbf{I}_{A_r})\mathbf{h} + \mathbf{e}_k, \tag{1.9}$$

where $\boldsymbol{\mathcal{X}}_k^{(t)}$ is obtained by setting the unknown data symbol vectors of $\mathbf{x}_k^{(d)}$ (the light shaded area in (1.8)) to zero in $\boldsymbol{\mathcal{X}}_k$, whereas $\boldsymbol{\mathcal{X}}_k^{(d)}$ is obtained by setting the training symbol vectors of $\mathbf{x}_k^{(t)}$ (the dark shaded area in (1.8)) to zero.

Although we will generally express the obtained results as a function of k, it will sometimes be convenient to stack all vectors $\mathbf{y}_k$ and all matrices $\boldsymbol{\mathcal{X}}_k^{(t)}$ and $\boldsymbol{\mathcal{X}}_k^{(d)}$ for $k = 0, \ldots, K-1$, leading to

$$\mathbf{y} = [\mathbf{y}_0^T, \ldots, \mathbf{y}_{K-1}^T]^T, \tag{1.10}$$

$$\boldsymbol{\mathcal{X}}^{(t)} = [\boldsymbol{\mathcal{X}}_0^{(t)T}, \ldots, \boldsymbol{\mathcal{X}}_{K-1}^{(t)T}]^T, \qquad \boldsymbol{\mathcal{X}}^{(d)} = [\boldsymbol{\mathcal{X}}_0^{(d)T}, \ldots, \boldsymbol{\mathcal{X}}_{K-1}^{(d)T}]^T. \tag{1.11}$$

We will now discuss training-based channel estimation using (1.6) and (1.9). We will make a distinction between conventional training-based channel estimation, which only takes received samples into account that solely depend on training symbols ($P = 0$), and enhanced training-based channel estimation, which next to these received samples also takes some surrounding received samples into account, which might depend on both training symbols and unknown data symbols or solely on unknown data symbols ($0 < P \leq N_d/2$). Although the latter techniques can be classified as semi-blind methods, we will in a subsequent section focus on semi-blind channel estimation methods that combine one of the previous training-based methods (conventional or enhanced) with a purely blind method.

1.3 CONVENTIONAL TRAINING-BASED METHODS

1.3.1 Channel Estimation

Conventional training solutions use only those received samples that solely depend on the training symbols. In other words, we consider $P = 0$, which allows us to simplify (1.9) as

$$\mathbf{y}_k = (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} + \mathbf{e}_k. \tag{1.12}$$

where $\mathbf{I} = \mathbf{I}_{A_r}$. Although many different channel estimation procedures can be applied to (1.12), we restrict ourselves to maximum likelihood (ML) channel estimation, which neither requires knowledge of the noise variance nor any statistical information about the channel [2].

The ML channel estimate related to (1.12) is obtained by solving the following optimization problem:

$$\begin{aligned} \mathbf{h}_{ML} &= \arg\min_{\mathbf{h}} \sum_{k=0}^{K-1} \| \mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} \|^2 \\ &= \arg\min_{\mathbf{h}} \ \| \mathbf{y} - (\boldsymbol{\mathcal{X}}^{(t)} \otimes \mathbf{I})\mathbf{h} \| \,. \end{aligned} \tag{1.13}$$

where the received sample vector $\mathbf{y}$ and training matrix $\boldsymbol{\mathcal{X}}^{(t)}$ are defined in (1.10) and (1.11). This is a standard least-squares (LS) problem, whose solution is given in terms of a pseudo-inverse, $(\boldsymbol{\mathcal{X}}^{(t)} \otimes \mathbf{I})^{\dagger}$, which is equal to $\boldsymbol{\mathcal{X}}^{(t)\dagger} \otimes \mathbf{I}$. Assuming $\boldsymbol{\mathcal{X}}^{(t)}$ has full column rank, which requires $K \geq A_t(L+1)/(N_t - L)$, we obtain

$$\begin{aligned} \mathbf{h}_{ML} &= (\boldsymbol{\mathcal{X}}^{(t)\dagger} \otimes \mathbf{I})\mathbf{y} \\ &= [(\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)})^{-1} \otimes \mathbf{I}](\boldsymbol{\mathcal{X}}^{(t)H} \otimes \mathbf{I})\mathbf{y} \\ &= \Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H}\boldsymbol{\mathcal{X}}_k^{(t)}\Big)^{-1} \otimes \mathbf{I}\Big] \sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\mathbf{y}_k \end{aligned} \tag{1.14}$$

If we insert the data model (1.12), it follows that

$$\mathbf{h}_{ML} = \mathbf{h} \ + \ \Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H}\boldsymbol{\mathcal{X}}_k^{(t)}\Big)^{-1} \otimes \mathbf{I}\Big] \sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\mathbf{e}_k \,, \tag{1.15}$$

which shows that the ML estimate is unbiased. Furthermore, since the noise term $\mathbf{e}_k$ has covariance $\mathrm{E}(\mathbf{e}_k\mathbf{e}_k^H) = \sigma_e^2\mathbf{I}$, the covariance of the channel estimation error is given by

$$\mathrm{E}[(\mathbf{h}_{ML} - \mathbf{h})(\mathbf{h}_{ML} - \mathbf{h})^H] = \sigma_e^2\Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H}\boldsymbol{\mathcal{X}}_k^{(t)}\Big)^{-1} \otimes \mathbf{I}\Big],$$

which is equal to the Cramer-Rao bound (CRB), and the mean square error (MSE) of the channel estimate can be expressed as

$$\begin{aligned} J_{ML} = \mathrm{E}\{\|\mathbf{h}_{ML} - \mathbf{h}\|^2\} &= \sigma_e^2 A_r \ \mathrm{tr}\Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H}\boldsymbol{\mathcal{X}}_k^{(t)}\Big)^{-1}\Big] \\ &= \sigma_e^2 A_r \ \mathrm{tr}[(\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)})^{-1}] \,. \end{aligned} \tag{1.16}$$

Note that this channel estimation problem can actually be decoupled into the different receive antennas, and is often presented as such. However, for the enhanced training-based methods discussed in the next section, the correlation between the different receive antennas will come into the picture, and the problem cannot be decoupled anymore. This does not mean that we cannot use the enhanced training-

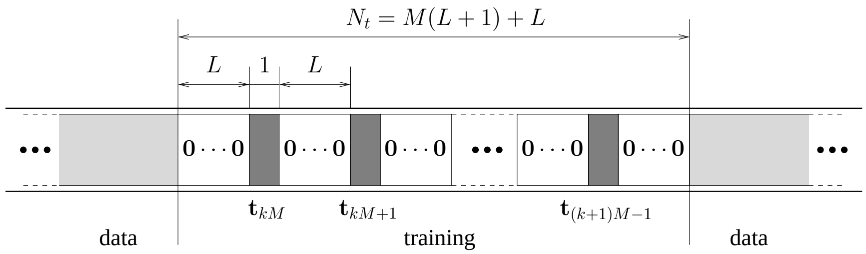


Fig. 1.2 Structure of the solution for optimal training.

based methods on smaller subgroups of receive antennas, it simply means that the performance of such an approach will be different (see also Section 1.6).

1.3.2 Optimal Training Design

It is possible to design the training symbol vectors such that J_{ML} is minimized under a total training power constraint. In other words, we solve

$$\min_{\{\mathbf{x}_k^{(t)}\}} \operatorname{tr}[(\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)})^{-1}] \quad \text{s.t.} \quad \sum_{k=0}^{K-1} \|\mathbf{x}_k^{(t)}\|^2 = E\,, \tag{1.17}$$

where E is a specified constant. To solve this problem, observe that

$$\operatorname{tr}[(\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)})^{-1}] \;\geq\; \sum_{i=1}^{A_t(L+1)} \frac{1}{\|\boldsymbol{\mathcal{X}}^{(t)}(:,i)\|^2}\,, \tag{1.18}$$

where equality is obtained if $\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)}$ is diagonal (the notation $\mathbf{A}(:,i)$ represents the i-th column of the matrix $\mathbf{A}$). From (1.8) (with $P = 0$), it is clear that each block column of $\boldsymbol{\mathcal{X}}_k^{(t)}$ only contains $N_t - L$ training symbol vectors of the total amount of N_t training symbol vectors collected in $\mathbf{x}_k^{(t)}$. Hence, there is no immediate connection between $\sum_{k=0}^{K-1} \|\mathbf{x}_k^{(t)}\|^2$ and $\|\boldsymbol{\mathcal{X}}^{(t)}\|^2$, which complicates matters. We proceed in the following way. We first find the minimum of the right hand side of (1.18) under the constraint, which is obtained when all terms are equal and as small as possible under this constraint, and we subsequently try to realize this minimum by a training design for which $\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)}$ is diagonal, in order to obtain equality in (1.18).

We will consider the following two cases: the number of training symbols $N_t \geq 2L+1$, and $N_t = L+1$. For the remaining case where $L+1 < N_t < 2L+1$, the optimization problem is hard to solve in analytical form.

1.3.2.1 Case $N_t \geq 2L+1$. In this case, the terms on the right hand side of (1.18) are equal and as small as possible under the constraint $\sum_{k=0}^{K-1} \|\mathbf{x}_k^{(t)}\|^2 = E$, if

$$\begin{cases} \mathbf{x}(n_k+l) & = & \mathbf{0}_{A_t \times 1} \\ \mathbf{x}(n_k+N_t-1-l) & = & \mathbf{0}_{A_t \times 1} \end{cases}, \quad l = 0, 1, \ldots, L-1, \; k = 0, 1, \ldots, K-1, \tag{1.19}$$

and

$$\|\boldsymbol{\mathcal{X}}^{(t)}(:, i)\|^2 = E/A_t, \quad i = 1, \ldots, A_t(L+1). \tag{1.20}$$

If we also choose $\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)}$ diagonal, in order to obtain equality in (1.18), then the latter condition can be written as

$$\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)} = E/A_t \mathbf{I}_{A_t(L+1)}, \tag{1.21}$$

which requires $K \geq A_t(L+1)/(N_t - L)$.

As an example, consider an integer $M \geq 1$ and set $N_t = M(L+1) + L$. An optimal solution is then given by using dispersed training symbol vectors separated by L zero vectors:

$$\mathbf{x}_k^{(t)} = [\mathbf{0}_{A_t L \times 1}^T, \mathbf{t}_{kM}^T, \mathbf{0}_{A_t L \times 1}^T, \mathbf{t}_{kM+1}^T, \ldots, \mathbf{0}_{A_t L \times 1}^T, \mathbf{t}_{(k+1)M-1}^T, \mathbf{0}_{A_t L \times 1}^T]^T \tag{1.22}$$

where $\mathbf{T} = [\mathbf{t}_0, \ldots, \mathbf{t}_{KM-1}]$ is an $A_t \times KM$ matrix that satisfies $\mathbf{T}\mathbf{T}^H = E/A_t \, \mathbf{I}_{A_t}$, which requires $K \geq A_t/M$. The structure of the solution is shown in figure 1.2.

Often, only $M = A_t$ (hence, $N_t = A_t(L+1) + L$) and a single message block ($K = 1$) is considered, since this option minimizes the total training overhead for a fixed total burst length N. For instance, the optimal training approach of [3], which maximizes a lower bound on the ergodic capacity assuming linear minimum mean square error (LMMSE) channel estimation, falls within this special class. However, N_t and K cannot always be chosen freely. As a result, the other options might also be useful in practice, as well as the case that is considered next.

1.3.2.2 Case $N_t = L+1$. First of all, note from (1.8) (with $P = 0$) that in this case $\boldsymbol{\mathcal{X}}_k^{(t)} = \mathbf{x}_k^{(t)T}$, and consequently $\boldsymbol{\mathcal{X}}^{(t)} = [\mathbf{x}_0^{(t)}, \ldots, \mathbf{x}_{K-1}^{(t)}]^T$. Therefore, the terms on the right hand side of (1.18) are equal and as small as possible under the constraint $\sum_{k=0}^{K-1} \|\mathbf{x}_k^{(t)}\|^2 = E$, if

$$\|\boldsymbol{\mathcal{X}}^{(t)}(:, i)\|^2 = E/(A_t(L+1)), \quad i = 1, \ldots, A_t(L+1). \tag{1.23}$$

Again, if we also choose $\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)}$ diagonal, this condition becomes

$$\boldsymbol{\mathcal{X}}^{(t)H}\boldsymbol{\mathcal{X}}^{(t)} = E/(A_t(L+1)) \, \mathbf{I}_{A_t(L+1)}, \tag{1.24}$$

which requires $K \geq A_t(L+1)$. Since $\boldsymbol{\mathcal{X}}^{(t)}$ has no specific block Toeplitz structure, condition (1.24) is easy to satisfy.

1.4 ENHANCED TRAINING-BASED METHODS

In the previous section, we only considered those received samples that solely were depending on training symbols. However, an enhanced channel estimate can be obtained if we also take some surrounding received samples into account. Referring to Figure 1.1, we will consider $0 < P \leq N_d/2$, which, this time, does not allow us to simplify (1.9).

We will again apply ML channel estimation. However, since also unknown data symbols are involved, we can adopt different options now. We will here focus on deterministic ML (DML) and Gaussian ML (GML). In DML, we assume the data symbols are unknown deterministic parameters, whereas in GML we assume that they are unknown random variables with a Gaussian distribution. Both methods do not take the finite alphabet property of the data symbols into account, because this often leads to more complex algorithms. For this extension, we refer the interested reader to [4], where the data symbols are viewed as discrete deterministic parameters, and to [5], where they are regarded as random variables with a discrete distribution (only flat-fading MIMO channels are considered in [4] and [5]).

Since we restrict P to $0 < P \leq N_d/2$, two successive vectors $\mathbf{x}_k^{(d)}$ and $\mathbf{x}_{k+1}^{(d)}$ never overlap, which allows us to process them independently. See [6] for an overview of similar enhanced training-based ML methods for the single-input multiple-output (SIMO) case with $K = 1$.

1.4.1 Deterministic ML

Viewing the data symbols as unknown deterministic parameters, the ML channel estimate related to (1.9) is obtained by solving the following optimization problem:

$$(\mathbf{h}_{ML}, \{\mathbf{x}_{k,ML}^{(t)}\}) = \arg \min_{(\mathbf{h}, \{\mathbf{x}_k^{(t)}\})} \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} - (\boldsymbol{\mathcal{X}}_k^{(d)} \otimes \mathbf{I})\mathbf{h}\|^2 \tag{1.25}$$

$$= \arg \min_{(\mathbf{h}, \{\mathbf{x}_k^{(t)}\})} \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} - \boldsymbol{\mathcal{H}}^{(d)}\mathbf{x}_k^{(d)}\|^2. \tag{1.26}$$

This can for instance be solved by alternating minimizations between $\mathbf{h}$ and $\{\mathbf{x}_k^{(d)}\}$ (initialized by $\mathbf{x}_k^{(d)} = \mathbf{0}_{2A_tP\times 1}$). In this context, note that the solution for $\mathbf{h}$ of (1.25) for a given estimate $\hat{\mathbf{x}}_k^{(d)}$ is, similar as in (1.14),

$$\begin{aligned}\mathbf{h}_{ML}(\{\hat{\mathbf{x}}_k^{(d)}\}) &= \arg\min_{\mathbf{h}} \sum_{k=0}^{K-1} \| \mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} - (\hat{\boldsymbol{\mathcal{X}}}_k^{(d)} \otimes \mathbf{I})\mathbf{h} \|^2 \\ &= \Big\{ \Big[\sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)} + \hat{\boldsymbol{\mathcal{X}}}_k^{(d)})^H (\boldsymbol{\mathcal{X}}_k^{(t)} + \hat{\boldsymbol{\mathcal{X}}}_k^{(d)}) \Big]^{-1} \otimes \mathbf{I} \Big\} \\ &\quad \cdot \sum_{k=0}^{K-1} [(\boldsymbol{\mathcal{X}}_k^{(t)} + \hat{\boldsymbol{\mathcal{X}}}_k^{(d)})^H \otimes \mathbf{I}]\mathbf{y}_k, \end{aligned} \tag{1.27}$$

whereas the solution for $\mathbf{x}_{k,ML}^{(d)}$ of (1.26) for a given estimate $\hat{\mathbf{h}}$ is

$$\begin{aligned}\mathbf{x}_{k,ML}^{(d)}(\hat{\mathbf{h}}) &= \arg\min_{\mathbf{x}_k^{(d)}} \| \mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\hat{\mathbf{h}} - \hat{\boldsymbol{\mathcal{H}}}^{(d)} \mathbf{x}_k^{(d)} \|^2 \\ &= \hat{\boldsymbol{\mathcal{H}}}^{(d)\dagger} [\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\hat{\mathbf{h}}]\,. \end{aligned} \tag{1.28}$$

Note that we assume here that $\boldsymbol{\mathcal{X}}^{(t)} + \hat{\boldsymbol{\mathcal{X}}}^{(d)}$ always has full column rank $A_t(L+1)$, which requires $K \geq A_t(L+1)/(N_t + 2P - L)$.

We can also plug (1.28) in (1.26) to obtain

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \sum_{k=0}^{K-1} \| \mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h} \|^2_{\mathbf{P}_{\text{col}}^{\perp}(\boldsymbol{\mathcal{H}}^{(d)})}\,, \tag{1.29}$$

where $\mathbf{P}_{\text{col}}^{\perp}(\boldsymbol{\mathcal{H}}^{(d)}) = \mathbf{I}_{A_r(N_t+2P-L)} - \boldsymbol{\mathcal{H}}^{(d)}\boldsymbol{\mathcal{H}}^{(d)\dagger}$ is the projection matrix onto the orthogonal complement of the column space of $\boldsymbol{\mathcal{H}}^{(d)}$. This problem can be solved using gradient techniques which, unfortunately, are rather complex and probably not worth the additional effort. Note that the simplified quadratic DML techniques proposed for SIMO systems [6] cannot be applied for MIMO systems, because it is impossible to find a linear parametrization of the null space of $\boldsymbol{\mathcal{H}}^{(d)}$ as a function of $\mathbf{h}$.

The CRB for the DML channel estimate can be derived as in [7] and is given by

$$\text{CRB}_{DML} = \sigma_e^2 \Big\{ \sum_{k=0}^{K-1} [(\boldsymbol{\mathcal{X}}_k^{(t)} + \boldsymbol{\mathcal{X}}_k^{(d)})^H \otimes \mathbf{I}]\mathbf{P}_{\text{col}}^{\perp}(\boldsymbol{\mathcal{H}}^{(d)})[(\boldsymbol{\mathcal{X}}_k^{(t)} + \boldsymbol{\mathcal{X}}_k^{(d)}) \otimes \mathbf{I}] \Big\}^{-1}. \tag{1.30}$$

Remark 1: If $N_t > L$, $A_r \leq A_t$, and $\boldsymbol{\mathcal{H}}^{(d)}$ has a non-empty left null space of dimension $A_r(N_t - L)$ (true with probability one if $N_t > L$ and $A_r \leq A_t$), then the matrix $\mathbf{P}_{\text{col}}^{\perp}(\boldsymbol{\mathcal{H}}^{(d)}) = \mathbf{S}$, where $\mathbf{S}$ is the $A_r(N_t + 2P - L) \times A_r(N_t + 2P - L)$ selection matrix that selects the $A_r(N_t - L)$ middle rows and removes the A_rP top and bottom rows. As a result, the iterative approach is not required, since we can solve (1.29) in closed form. Its solution is then given by

$$\mathbf{h}_{ML} = \Big[\sum_{k=0}^{K-1}(\boldsymbol{\mathcal{X}}_k^{(t)H}\otimes\mathbf{I})\mathbf{S}(\boldsymbol{\mathcal{X}}_k^{(t)}\otimes\mathbf{I})\Big]^{-1}\sum_{k=0}^{K-1}(\boldsymbol{\mathcal{X}}_k^{(t)H}\otimes\mathbf{I})\mathbf{S}\mathbf{y}_k. \quad (1.31)$$

The CRB (1.30) can then be expressed as

$$\begin{aligned}\mathrm{CRB}_{DML} &= \sigma_e^2\Big\{\sum_{k=0}^{K-1}[(\boldsymbol{\mathcal{X}}_k^{(t)}+\boldsymbol{\mathcal{X}}_k^{(d)})^H\otimes\mathbf{I}]\mathbf{S}[(\boldsymbol{\mathcal{X}}_k^{(t)}+\boldsymbol{\mathcal{X}}_k^{(d)})\otimes\mathbf{I}]\Big\}^{-1}\\ &= \sigma_e^2\Big[\sum_{k=0}^{K-1}(\boldsymbol{\mathcal{X}}_k^{(t)H}\otimes\mathbf{I})\mathbf{S}(\boldsymbol{\mathcal{X}}_k^{(t)}\otimes\mathbf{I})\Big]^{-1},\end{aligned}$$

where we have used that the $A_r(N_t-L)$ middle rows of $\boldsymbol{\mathcal{X}}_k^{(d)}$ contain zero entries. It is clear that these results are exactly the same as the ones for the conventional training-based method (applying $\mathbf{S}$ is the same as taking $P=0$).

Remark 2: Another special case arises when $N_t \le L$, $A_r \le A_t$, and the matrix $\boldsymbol{\mathcal{H}}^{(d)}$ has an empty left null space (true with probability one if $N_t \le L$ and $A_r \le A_t$). In that case, $\mathbf{P}_{\mathrm{col}}^{\perp}(\boldsymbol{\mathcal{H}}^{(d)}) = \mathbf{0}_{A_r(N_t+2P-L)}$, which actually means that the DML problem is underdetermined ((1.29) is underdetermined and the CRB (1.30) is infinity). However, the iterative approach can still be applied in order to find a reasonable channel estimate. Actually, the iterative approach will converge in one step to the solution that is obtained after the first step of the GML method (see Section 1.4.2). Hence, it can never outperform the GML method under these circumstances.

1.4.2 Gaussian ML

Recall equation (1.9), viz.,

$$\mathbf{y}_k = (\boldsymbol{\mathcal{X}}_k^{(t)}\otimes\mathbf{I})\mathbf{h} \quad + \quad [(\boldsymbol{\mathcal{X}}_k^{(d)}\otimes\mathbf{I})\mathbf{h}+\mathbf{e}_k]\,. \quad (1.32)$$

Viewing the data symbols as unknown random variables with a Gaussian distribution, the term in brackets is a Gaussian noise term with covariance

$$\mathbf{Q} = \boldsymbol{\mathcal{H}}^{(d)}\mathbf{R}_{x^{(d)}}\boldsymbol{\mathcal{H}}^{(d)H} + \sigma_e^2\mathbf{I}_{A_r(N_t+2P-L)}, \quad (1.33)$$

where $\mathbf{R}_{x^{(d)}} = \mathrm{E}\{\mathbf{x}_k^{(d)}\mathbf{x}_k^{(d)H}\}$ is the covariance of $\mathbf{x}_k^{(d)}$, which we assume to be known. Following standard techniques, the ML channel estimate related to (1.32) is then obtained by solving the optimization problem

$$(\mathbf{h}_{ML},\sigma_{e,ML}^2) = \arg\min_{(\mathbf{h},\sigma_e^2)} K\ln|\mathbf{Q}| + \sum_{k=0}^{K-1}\|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)}\otimes\mathbf{I})\mathbf{h}\|_{\mathbf{Q}^{-1}}^2\,. \quad (1.34)$$

This problem can be solved using gradient techniques. However, since these are rather complex, we will simplify (approximate) the problem by assuming that $\mathbf{Q}$ is an arbitrary matrix that is independent from $\mathbf{h}$ and σ_e^2. This approach has been proposed in [8] for the single-input single-output (SISO) case, and it has been shown there that the effect on the CRB is negligible. We then obtain

$$(\mathbf{h}_{ML}, \mathbf{Q}_{ML}) = \arg \min_{(\mathbf{h},\mathbf{Q})} K \ln |\mathbf{Q}| + \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h}\|^2_{\mathbf{Q}^{-1}}, \tag{1.35}$$

which can be solved using alternating minimizations between $\mathbf{h}$ and $\mathbf{Q}$ (initialized by $\mathbf{Q} = \mathbf{I}_{A_r(N_t+2P-L)}$). Indeed, using similar derivations as in [8], we can show that for a given estimate $\hat{\mathbf{Q}}$, the optimal estimate for $\mathbf{h}$ is

$$\begin{aligned}\mathbf{h}_{ML}(\hat{\mathbf{Q}}) &= \arg \min_{\mathbf{h}} K \ln |\hat{\mathbf{Q}}| + \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h}\|^2_{\hat{\mathbf{Q}}^{-1}} \\ &= \Big[\sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\hat{\mathbf{Q}}^{-1}(\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\Big]^{-1} \sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\hat{\mathbf{Q}}^{-1}\mathbf{y}_k,\end{aligned} \tag{1.36}$$

whereas for a given estimate $\hat{\mathbf{h}}$, the optimal estimate for $\mathbf{Q}$ is

$$\begin{aligned}\mathbf{Q}_{ML}(\hat{\mathbf{h}}) &= \arg \min_{\mathbf{Q}} K \ln |\mathbf{Q}| + \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\hat{\mathbf{h}}\|^2_{\mathbf{Q}^{-1}} \\ &= K^{-1} \sum_{k=0}^{K-1} [\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\hat{\mathbf{h}}][\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\hat{\mathbf{h}}]^H.\end{aligned} \tag{1.37}$$

Note that we assume here that $\boldsymbol{\mathcal{X}}^{(t)}$ has full column rank $A_t(L+1)$, which requires $K \geq A_t(L+1)/(N_t+2P-L)$, and that $\hat{\mathbf{Q}}$ is always invertible, which requires $K \geq A_r(N_t+2P-L)$.

As already mentioned, the CRB does not change much by simplifying (approximating) the GML problem formulation. Hence, we will only show the CRB of the simplified (approximated) GML channel estimate. This can be derived as in [8] and is given by

$$\text{CRB}_{GML} = \Big[\sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\mathbf{Q}^{-1}(\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\Big]^{-1}. \tag{1.38}$$

Remark 3: If we set the first and last L training symbol vectors to zero in each symbol block, as we did for the optimal training strategy of Section 1.3.2.1 (see (1.19)), then it is easy to show that

$$\mathbf{Q}^{-1}(\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I}) = \sigma_e^{-2}(\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I}).$$

As a result, the CRB (1.38) can be expressed as

$$\text{CRB}_{GML} = \sigma_e^2 \Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H} \boldsymbol{\mathcal{X}}_k^{(t)} \Big)^{-1} \otimes \mathbf{I} \Big]. \tag{1.39}$$

Because of (1.19), this CRB does not change with P, and is hence equivalent to the CRB for the conventional training-based method (which takes $P = 0$). Therefore, the GML method has no advantage over the conventional training-based method in this case.

Remark 4: A closed-form GML channel estimate can be obtained when $N_t = 1$ and $P = L$. First, observe that when $N_t = 1$ and $P = L$, $\boldsymbol{\mathcal{X}}_k^{(t)}$ can be written as

$$\boldsymbol{\mathcal{X}}_k^{(t)} = \mathbf{I}_{L+1} \otimes \mathbf{x}_k^{(t)T} = \mathbf{I}_{L+1} \otimes \mathbf{x}^T(n_k).$$

Due to this special structure, it can be shown that (1.36) becomes independent of $\hat{\mathbf{Q}}$ as long as it is invertible. As a result, the iterative approach converges in one step and the closed-form GML channel estimate is given by

$$\mathbf{h}_{ML} = \Big[\mathbf{I}_{L+1} \otimes \Big(\sum_{k=0}^{K-1} \mathbf{x}^*(n_k) \mathbf{x}^T(n_k) \Big)^{-1} \otimes \mathbf{I} \Big] \sum_{k=0}^{K-1} (\mathbf{I}_{L+1} \otimes \mathbf{x}^*(n_k) \otimes \mathbf{I}) \mathbf{y}_k. \tag{1.40}$$

The CRB (1.38) becomes

$$\text{CRB}_{GML} = \Big[\sum_{k=0}^{K-1} (\mathbf{I}_{L+1} \otimes \mathbf{x}^*(n_k) \otimes \mathbf{I}) \mathbf{Q}^{-1} (\mathbf{I}_{L+1} \otimes \mathbf{x}^T(n_k) \otimes \mathbf{I}) \Big]^{-1}.$$

Note that in this case, we can again decouple the problem into the different receive antennas.

1.5 SEMI-BLIND CHANNEL ESTIMATION

1.5.1 A Combined Cost Function

The enhanced training-based methods discussed in the previous section can be viewed as semi-blind channel estimation methods, albeit in a limited form. We discuss in this section the combination of training-based methods (conventional or enhanced) with purely blind methods. We can arrive at such a method by combining the cost functions in Sections 1.3 or 1.4 with a blind criterion. We will limit ourselves to quadratic blind criteria which usually result from deterministic blind methods or stochastic blind methods based on second-order statistics (SOS). It would also be possible to exploit higher-order statistics (HOS) [9, 10] or constant modulus and finite alphabet properties.

For the sake of simplicity, we restrict our attention to the conventional training-based criterion (1.13). Combining this criterion with a quadratic blind criterion, we obtain a semi-blind problem that often can be formulated as

$$\mathbf{h}_o = \arg\min_{\mathbf{h}} \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h}\|^2 + \alpha\|\mathbf{V}\mathbf{H}\|^2, \tag{1.41}$$

where $\mathbf{H} = [\mathbf{H}^T(0), \ldots, \mathbf{H}^T(L)]^T$, and $\mathbf{V} = [\mathbf{V}(0), \ldots, \mathbf{V}(L)]$ is a matrix which depends on the received data and can be constructed in many different ways depending on the blind criterion that will be adopted (see later on).

To rewrite this as a single condition on $\mathbf{h}$, let $\mathbf{W} = [\mathbf{W}(0), \ldots, \mathbf{W}(L)]$ be a matrix such that $\mathrm{vec}(\mathbf{VH}) = \mathbf{Wh}$. In particular, since

$$\mathrm{vec}(\mathbf{V}(l)\mathbf{H}(l)) = (\mathbf{I}_{A_t} \otimes \mathbf{V}(l))\mathrm{vec}(\mathbf{H}(l)) = \mathbf{W}(l)\mathbf{h}(l)\,, \tag{1.42}$$

$\mathbf{W}(l)$ is given by $\mathbf{W}(l) = \mathbf{I}_{A_t} \otimes \mathbf{V}(l)$. In terms of $\mathbf{W}$, the problem (1.41) can be rewritten as

$$\mathbf{h}_o = \arg\min_{\mathbf{h}} \sum_{k=0}^{K-1} \|\mathbf{y}_k - (\boldsymbol{\mathcal{X}}_k^{(t)} \otimes \mathbf{I})\mathbf{h}\|^2 + \alpha\|\mathbf{W}\mathbf{h}\|^2\,, \tag{1.43}$$

the solution of which is given by

$$\mathbf{h}_o = \Big[\Big(\sum_{k=0}^{K-1} \boldsymbol{\mathcal{X}}_k^{(t)H}\boldsymbol{\mathcal{X}}_k^{(t)}\Big) \otimes \mathbf{I} + \alpha\mathbf{W}^H\mathbf{W}\Big]^{-1} \sum_{k=0}^{K-1} (\boldsymbol{\mathcal{X}}_k^{(t)H} \otimes \mathbf{I})\mathbf{y}_k. \tag{1.44}$$

Note that the choice of a good weighting factor α will be crucial. We will come back to this in Section 1.5.4.

The matrix $\mathbf{V}$ in (1.41) can be obtained from many recently proposed blind MIMO channel estimation algorithms that determine the channel $\mathbf{H}$ up to an invertible (sometimes unitary) matrix. Usually, HOS are used to resolve this ambiguity. Here, the training-based part will take care of that (we assume that the channel is identifiable using only the training part). Generally, only the received sample vectors that depend completely on unknown data symbol vectors are taken into account in the blind criterion. However, in the considered setup, this would mean that a lot of blind structural information is lost, because the training symbol vectors break up the total burst in many different pieces. In this work, we therefore construct the blind criterion based on all received sample vectors in order not to lose any blind information.

As examples on the construction of $\mathbf{V}$, we will present in Sections 1.5.2 and 1.5.3 two deterministic blind algorithms, capable of perfectly estimating $\mathbf{H}$ (up to an invertible matrix) using a finite number of samples in the absence of noise. The first is the subspace (SS) approach [11], [12], [13] whereas the second is the least-squares smoothing (LSS) approach [14], [15].

1.5.2 Subspace Approach

The first example of a blind MIMO channel estimation algorithm is the so-called subspace approach, defined in [12], [13] (introduced in [11] for the SIMO case). Let $\boldsymbol{\mathcal{Y}}$ be an extended data matrix constructed from the received sample vectors as

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \mathbf{y}(L) & \mathbf{y}(L+1) & \cdots & \mathbf{y}(N-Q-1) \\ \vdots & & & \vdots \\ \mathbf{y}(L+Q) & \mathbf{y}(L+Q+1) & \cdots & \mathbf{y}(N-1) \end{bmatrix}. \tag{1.45}$$

$\boldsymbol{\mathcal{Y}}$ is an $A_r(Q+1)\times(N-L-Q)$ block Hankel matrix, where Q is a design parameter that can be interpreted as the filter order of an equalizer acting on the sequence $\mathbf{y}(n)$, and is usually called the smoothing factor. This matrix can be written as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{H}}_Q \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}}, \tag{1.46}$$

where the additive noise matrix $\boldsymbol{\mathcal{E}}$ is similarly defined as $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{H}}_Q$ is the $A_r(Q+1) \times A_t(L+Q+1)$ block Toeplitz channel matrix that is similarly defined as $\boldsymbol{\mathcal{H}}$ in (1.4) but with different dimensions, and $\boldsymbol{\mathcal{X}}$ is the $A_t(L+Q+1) \times (N-L-Q)$ block Hankel data symbol matrix given by

$$\boldsymbol{\mathcal{X}} = \begin{bmatrix} \mathbf{x}(0) & \cdots & \mathbf{x}(N-L-Q-1) \\ \vdots & & \vdots \\ \mathbf{x}(L+Q) & \cdots & \mathbf{x}(N-1) \end{bmatrix}. \tag{1.47}$$

Let us assume that there is no noise. If $\boldsymbol{\mathcal{H}}_Q$ is tall and of full column rank, and $\boldsymbol{\mathcal{X}}$ is wide and of full row rank, then $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{H}}_Q \boldsymbol{\mathcal{X}}$ is a low-rank factorization, and therefore $\boldsymbol{\mathcal{Y}}$ has a non-empty left null space of dimension $d_{SS} = A_r(Q+1) - A_t(L+Q+1)$. Denoting the $A_r(Q+1) \times d_{SS}$ matrix $\mathbf{U}_{SS}$ as the left singular vectors of $\boldsymbol{\mathcal{Y}}$ corresponding to this null space, we have

$$\boldsymbol{\mathcal{H}}_Q^H \mathbf{U}_{SS} = \mathbf{0}_{A_t(L+Q+1)\times d_{SS}}. \tag{1.48}$$

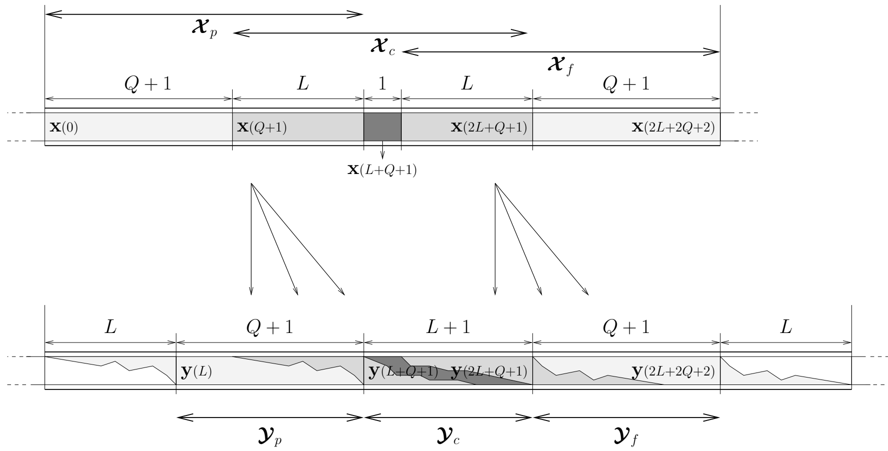


Fig. 1.3 Least-squares smoothing data model.

Each block row of (1.48) gives a linear relation on the block entries $\mathbf{H}(l)$ inside the structured matrix $\boldsymbol{\mathcal{H}}_Q$. Subsequently, we use this structure to rewrite the equation:

$$\boldsymbol{\mathcal{H}}_Q^H \mathbf{U}_{SS} = \begin{bmatrix} \mathbf{H}^H(L) & & \\ \vdots & \ddots & \\ \mathbf{H}^H(0) & & \mathbf{H}^H(L) \\ & \ddots & \vdots \\ & & \mathbf{H}^H(0) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{SS}(0) \\ \vdots \\ \mathbf{U}_{SS}(Q) \end{bmatrix} = \mathbf{0}_{A_t(L+Q+1)\times d_{SS}}$$

$$\Updownarrow$$

$$\begin{bmatrix} \mathbf{H}^H(0) \cdots \mathbf{H}^H(L) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{SS}(Q) \cdots \mathbf{U}_{SS}(0) & & \\ & \ddots \qquad \ddots & \\ & & \mathbf{U}_{SS}(Q) \cdots \mathbf{U}_{SS}(0) \end{bmatrix} = \mathbf{H}^H \boldsymbol{\mathcal{U}}_{SS} = \mathbf{0}_{A_t \times d_{SS}} .$$

As a result, we can choose $\mathbf{V} = \mathbf{V}_{SS} = \boldsymbol{\mathcal{U}}_{SS}^H$ in equation (1.41). In the noiseless case, the solution is unique (up to an invertible matrix) if $\boldsymbol{\mathcal{H}}_Q$ has full column rank and $\boldsymbol{\mathcal{X}}$ has full row rank.

1.5.3 Least-Squares Smoothing Approach

A second blind MIMO channel estimation technique is the so-called least-squares smoothing approach [15] (introduced in [14] for the SIMO case). Define the following block Hankel received sample matrices (viz. Figure 1.3),

$$\boldsymbol{\mathcal{Y}}_p = \begin{bmatrix} \mathbf{y}(L) & \cdots & \mathbf{y}(N-2Q-L-3) \\ \vdots & & \vdots \\ \mathbf{y}(Q+L) & \cdots & \mathbf{y}(N-Q-L-3) \end{bmatrix} \quad : A_r(Q+1)\times[N-2(Q+L+1)]\,,$$

$$\boldsymbol{\mathcal{Y}}_c = \begin{bmatrix} \mathbf{y}(Q+L+1) & \cdots & \mathbf{y}(N-Q-L-2) \\ \vdots & & \vdots \\ \mathbf{y}(Q+2L+1) & \cdots & \mathbf{y}(N-Q-2) \end{bmatrix} \quad : A_r(L+1)\times[N-2(Q+L+1)]\,,$$

$$\boldsymbol{\mathcal{Y}}_f = \begin{bmatrix} \mathbf{y}(Q+2L+2) & \cdots & \mathbf{y}(N-Q-1) \\ \vdots & & \vdots \\ \mathbf{y}(2Q+2L+2) & \cdots & \mathbf{y}(N-1) \end{bmatrix} \quad : A_r(Q+1)\times[N-2(Q+L+1)]\,,$$

corresponding to the “past”, “current” and “future”. L is the smoothing factor for $\boldsymbol{\mathcal{Y}}_c$, and Q is the smoothing factor for $\boldsymbol{\mathcal{Y}}_p$ and $\boldsymbol{\mathcal{Y}}_f$. These matrices have the following models:

$$\begin{aligned} \boldsymbol{\mathcal{Y}}_p &= \boldsymbol{\mathcal{H}}_Q\boldsymbol{\mathcal{X}}_p + \boldsymbol{\mathcal{E}}_p, \\ \boldsymbol{\mathcal{Y}}_c &= \boldsymbol{\mathcal{H}}_L\boldsymbol{\mathcal{X}}_c + \boldsymbol{\mathcal{E}}_c, \\ \boldsymbol{\mathcal{Y}}_f &= \boldsymbol{\mathcal{H}}_Q\boldsymbol{\mathcal{X}}_f + \boldsymbol{\mathcal{E}}_f, \end{aligned} \tag{1.49}$$

where $\boldsymbol{\mathcal{H}}_Q$ and $\boldsymbol{\mathcal{H}}_L$ are $A_r(Q+1) \times A_t(Q+L+1)$ and $A_r(L+1) \times A_t(2L+1)$ block Toeplitz channel matrices, and $\boldsymbol{\mathcal{X}}_p$, $\boldsymbol{\mathcal{X}}_c$, and $\boldsymbol{\mathcal{X}}_f$ are the past, current and future block Hankel data symbol matrices, given by

$$\boldsymbol{\mathcal{X}}_p = \begin{bmatrix} \mathbf{x}(0) & \cdots & \mathbf{x}(N-2Q-2L-3) \\ \vdots & & \vdots \\ \mathbf{x}(Q+L) & \cdots & \mathbf{x}(N-Q-L-3) \end{bmatrix} \quad : A_t(Q+L+1)\times[N-2(Q+L+1)]\,,$$

$$\boldsymbol{\mathcal{X}}_c = \begin{bmatrix} \mathbf{x}(Q+1) & \cdots & \mathbf{x}(N-Q-2L-2) \\ \vdots & & \vdots \\ \mathbf{x}(Q+2L+1) & \cdots & \mathbf{x}(N-Q-2) \end{bmatrix} \quad : A_t(2L+1)\times[N-2(Q+L+1)]\,,$$

$$\boldsymbol{\mathcal{X}}_f = \begin{bmatrix} \mathbf{x}(Q+L+2) & \cdots & \mathbf{x}(N-Q-L-1) \\ \vdots & & \vdots \\ \mathbf{x}(2Q+2L+2) & \cdots & \mathbf{x}(N-1) \end{bmatrix} \quad : A_t(Q+L+1)\times[N-2(Q+L+1)]\,.$$

Note that all the block rows of $\boldsymbol{\mathcal{X}}_c$, except for the block row

$$\mathbf{X} = \left[\mathbf{x}(Q+L+1), \ldots, \mathbf{x}(N-Q-L-2)\right], \tag{1.50}$$

are contained in $\boldsymbol{\mathcal{X}}_p$ and $\boldsymbol{\mathcal{X}}_f$. Hence, all different block rows can be collected in

$$\boldsymbol{\mathcal{X}}_{tot} = [\boldsymbol{\mathcal{X}}_p^T, \mathbf{X}^T, \boldsymbol{\mathcal{X}}_f^T]^T. \tag{1.51}$$

Let us assume that there is no noise. If $\boldsymbol{\mathcal{H}}_Q$ is tall and of full column rank and $\boldsymbol{\mathcal{X}}_{tot}$ is wide and of full row rank, then the orthogonal projection of $\boldsymbol{\mathcal{Y}}_c$ onto the orthogonal complement of the row space of $\boldsymbol{\mathcal{X}}_{p,f} = [\boldsymbol{\mathcal{X}}_p^T, \boldsymbol{\mathcal{X}}_f^T]^T$ is equal to the orthogonal projection of $\boldsymbol{\mathcal{Y}}_c$ onto the orthogonal complement of the row space of $\boldsymbol{\mathcal{Y}}_{p,f} = [\boldsymbol{\mathcal{Y}}_p^T, \boldsymbol{\mathcal{Y}}_f^T]^T$, and it is given by

$$\boldsymbol{\mathcal{Y}}_c \mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{Y}}_{p,f}) = \boldsymbol{\mathcal{Y}}_c \mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{X}}_{p,f}) = \mathbf{H}\mathbf{X}\mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{X}}_{p,f}), \tag{1.52}$$

where $\mathbf{P}_{\text{row}}^{\perp}(\mathbf{A}) = \mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A}$ is the projection matrix onto the orthogonal complement of the row space of $\mathbf{A}$. In addition, $\mathbf{H}$ has full column rank and $\mathbf{X}\mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{X}}_{p,f})$ has full row rank, which means that the column span of $\boldsymbol{\mathcal{Y}}_c \mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{Y}}_{p,f})$ coincides with the column span of $\mathbf{H}$. Let $\mathbf{U}_{LSS}$ be a matrix containing the left null space vectors of $\boldsymbol{\mathcal{Y}}_c \mathbf{P}_{\text{row}}^{\perp}(\boldsymbol{\mathcal{Y}}_{p,f})$. $\mathbf{U}_{LSS}$ has size $A_r(L+1) \times d_{LSS}$, where $d_{LSS} = A_r(L+1) - A_t$, and

$$\mathbf{H}^H \mathbf{U}_{LSS} = \mathbf{0}_{A_t \times d_{LSS}} . \tag{1.53}$$

In terms of the semi-blind criterion, we can take $\mathbf{V} = \mathbf{V}_{LSS} = \mathbf{U}_{LSS}^H$ in (1.41). In the noiseless case, the blind solution is unique (up to an invertible matrix) if $\boldsymbol{\mathcal{H}}_Q$ has full column rank and $\boldsymbol{\mathcal{X}}_{tot}$ has full row rank.

1.5.4 Weighting Factor

In equation (1.41), a weighting factor α scales the blind equation error relative to the training error. The choice of α is important: with an incorrect setting, the channel estimate can be worse than a training-only estimate! Ideally, one would want to choose α to minimize the channel MSE.

This is a well-known but essentially unsolved problem in semi-blind channel estimation; an extensive discussion can be found in [6]. One heuristic way of handling the problem is trying to avoid it by adapting the blind cost function in such a way that the MSE becomes less sensitive to α, e.g., by "denoising" [6]: a technique where the smallest eigenvalue of $\mathbf{W}^H\mathbf{W}$ is forced to zero by replacing $\mathbf{W}^H\mathbf{W}$ by $\mathbf{W}^H\mathbf{W} - \lambda_{\min}\mathbf{I}$, where $\lambda_{\min}$ is the minimal eigenvalue of $\mathbf{W}^H\mathbf{W}$. One could also try to find the optimal α in terms of the channel MSE, but this is usually very hard and represents a large computational cost.

1.5.5 Other Blind Channel Estimation Algorithms

The SS and LSS approaches basically require that all sources have the same channel order and that this channel order is known. Related to this latter requirement, it must be said that in contrast to the SS approach, the LSS approach allows for a simple joint order-channel estimation technique (illustrated in [14] for the SIMO case).

To avoid the sensitive requirement on knowledge of the channel length as present in the SS and LSS techniques, some interesting stochastic blind techniques based on SOS have been developed, such as the outer-product decomposition (OPD) approach [16], [17], [15], and the multi-step linear prediction (MSLP) approach [18],

[19], [15], which is a generalization of the earlier (one-step) linear prediction (LP) approach [20], [21], [22], [23], [15]. Note that for the latter, a corresponding semi-blind MIMO channel estimation procedure has been developed in [24]. The OPD and MLSP are closely related to each other and can be viewed as a stochastic version of the LSS approach. They do not require that all sources have the same channel order and that this channel order is known. However, they require the different sources to be zero-mean white (spatially and temporally), which is not always the case, e.g., when space-time coding is used. Moreover, since the training-based part that we include in the cost function can remove any ambiguity problems due to the above identifiability requirements, we have observed that the deterministic techniques are to be preferred in a semi-blind context if the total burst length N is short.

1.6 SIMULATION RESULTS

To finish the chapter, the proposed techniques are illustrated by means of simulation. We assume that the $A_t \times A_r$ MIMO channel $\mathbf{H}(l)$ of order L is Rayleigh fading, i.e., has zero-mean Gaussian distributed channel taps. We further assume that these channel taps are i.i.d. with variance σ_h^2. Although the proposed methods are able to handle sources with correlations that are spatially or temporally colored, e.g., due to space-time coding, we assume here a simple spatial multiplexing approach where the unknown data symbols are zero-mean i.i.d. (spatially and temporally) QPSK modulated with symbol energy 1. If possible, the training symbols are designed according to one of the optimal training strategies discussed in Section 1.3.2 in such a way that the average symbol energy is 1. If not possible, e.g., when $N_t = 1$ as illustrated below, they are designed in a similar way as the unknown data symbols, with unit energy. The noise variance is σ_e^2, and the received signal-to-noise ratio (SNR) per transmit antenna is defined as

$$\text{SNR} = \frac{(L+1)\sigma_h^2}{\sigma_e^2}. \tag{1.54}$$

To compare the different channel estimation methods, we will use the normalized MSE (NMSE) as performance measure, which can be defined as

$$\text{NMSE} = \frac{1}{R}\sum_{r=0}^{R-1} \frac{\|\hat{\mathbf{h}}^{(r)} - \mathbf{h}^{(r)}\|^2}{(L+1)A_t A_r \sigma_h^2}, \tag{1.55}$$

where the superscript r indicates the simulation run, and R is the total number of runs. Note that in each run, we will consider a new channel, data, and noise realization. First, we will compare the different training-based methods. Next, we study the performance of the presented semi-blind channel estimation procedures.

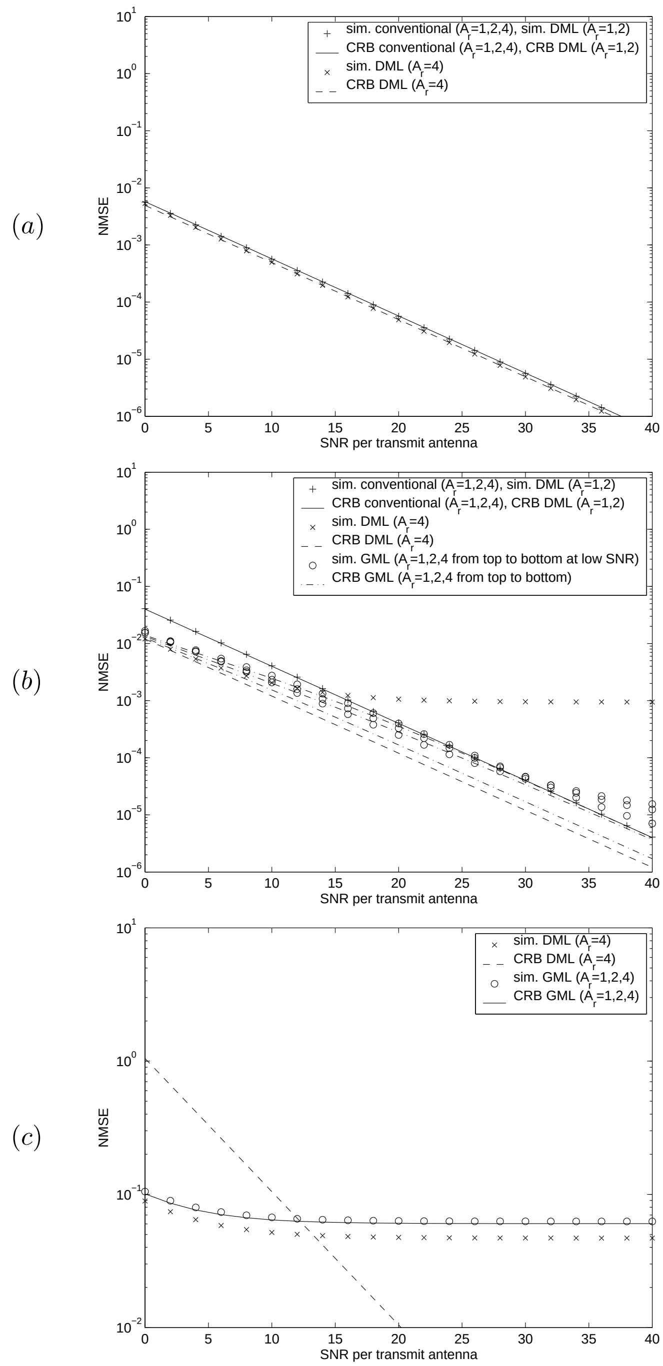


Fig. 1.4 Training-based algorithms: Comparison of the different methods (a) for $N_t = 2L + 1 = 7$, (b) for $N_t = L + 1 = 4$, (c) for $N_t = 1$.

1.6.1 Training-Based Results

We consider a MIMO system with $A_t = 2$ transmit antennas and $A_r = 1, 2, 4$ receive antennas. The channel order we simulate is $L = 3$. For the enhanced training-based methods, we take $P = L = 3$ and we carry out three iterations (unless of course the method converges in one step). The total number of symbol blocks that will be taken into account is given by $K = 100$. To make a fair comparison between the different number of receive antennas, we consider $R = 400, 200, 100$ simulation runs for $A_r = 1, 2, 4$ receive antennas, respectively. Note that decreasing A_r can either be viewed as reducing the number of receive antennas, or as treating the different receive antennas in smaller subgroups.

First, we take $N_t = 2L + 1 = 7$, such that we can implement the optimal training strategy of Section 1.3.2.1. From Remark 1, we know that the conventional ML method with $A_r = 1, 2, 4$ and the DML method with $A_r = 1, 2$ will produce the same result in this case. The DML method with $A_r = 4$ and the GML method with $A_r = 1, 2, 4$, on the other hand, will result in a different performance. However, since the GML method will not be able to outperform the conventional ML method in this case (see Remark 3), we will not consider it here. The simulation results and CRB's are shown in Figure 1.4(a). We observe that the simulation results match the CRB's well. We also notice that the DML method with $A_r = 4$ does a little bit better than the other methods.

Next, we take $N_t = L + 1 = 4$, such that we can implement the optimal training strategy of Section 1.3.2.2; the results are shown in Figure 1.4(b). From Remark 1, we can again deduce that the conventional ML method with $A_r = 1, 2, 4$ and the DML method with $A_r = 1, 2$ are the same. As before, the DML method with $A_r = 4$ and the GML method with $A_r = 1, 2, 4$ will be different, but this time there is no indication that the GML method can not beat the conventional ML method. Hence, all methods will be considered. From Figure 1.4(b), we observe that the conventional ML method with $A_r = 1, 2, 4$ and the DML method with $A_r = 1, 2$ closely approach their CRB, which is by the way the worst among all CRB's. The DML method with $A_r = 4$ has the best CRB, but the simulated performance saturates at high SNR. The saturation level can be lowered by increasing the number of iterations, but it goes slowly. The CRB of the GML method is somewhere in between and improves with A_r. As for the DML method with $A_r = 4$, its simulated performance moves away from the CRB, but not as much. In addition, only a few more iterations will shift the simulated performance close to its CRB.

We finally consider a situation where the conventional ML method can not be used due to the fact that $N_t < L + 1$. This happens for instance for the pilot symbol assisted modulation (PSAM) scheme [25], where training symbols (pilots) are periodically inserted in the total burst. In the considered MIMO setup, the PSAM approach corresponds to taking $N_t = 1$. From Remark 2, we know that the DML method with $A_r = 1, 2$ corresponds to the first step of the GML method with $A_r = 1, 2$, and thus should not be considered. Remark 4, on the other hand, tells us that the GML method converges in one step and performs the same for $A_r = 1, 2, 4$. Only the DML method with $A_r = 4$ does better, as can be observed from Figure 1.4(c).

Notice that whereas the GML method performs close to its CRB, the performance of the DML method with $A_r = 4$ is generally far from its CRB (in a positive sense at low SNR, but in a negative sense at high SNR). This gap reduces by increasing the number of iterations, but this goes very slowly.

1.6.2 Semi-Blind Results

In this section, we illustrate the performance of the semi-blind methods. We consider a MIMO system with $A_t = 2$ transmit antennas and $A_r = 4$ receive antennas. Note that $A_r > A_t$ is required for the semi-blind criterion to be useful (this is in contrast with the enhanced training-based methods). The channel order is again assumed to be $L = 3$. We consider $K = 10$ symbols blocks, $N_t = 7$ training symbol vectors per block, and $N_d = 80$ unknown data symbol vectors per block (as a result, we have a total of $N = K(N_t+N_d) = 870$ symbol vectors in one burst). Since $N_t = 2L+1 = 7$, we again use the optimal training strategy of Section 1.3.2.1. We consider $R = 100$ simulation runs and the received SNR per transmit antenna is set to 15 dB. Figure 1.5 compares the performance of the semi-blind method using the subspace and the least-squares smoothing criterion with the conventional training-based method as a function of α. Clearly, at the optimal α (different for the two aproaches), the semi-blind method outperforms the conventional training-based method. We also observe that the subspace approach does better than the least-squares smoothing approach. A similar behavior was observed for different settings.

1.7 CONCLUSIONS AND ADDITIONAL POINTERS

This chapter has covered only a very small part of the available literature. As a general reference to wireless channel estimation algorithms, we suggest the edited book [26], and the overview article [27].

Optimal training symbol placement and design has recently become a popular topic, e.g., in context of capacity maximization; some contributions are [28–32]. In particular relevant to the discussion in this chapter is [30], which gives optimal placement of training symbols in the semiblind context based on Cramer-Rao Bound (CRB) considerations (i.e., minimizing the MSE of the channel estimate and unknown data symbols).

In this chapter, we did not cover multi-carrier MIMO systems (e.g., based on orthogonal frequency division multiplexing (OFDM)). For these systems, the frequency-selective channel can be transformed into a set of parallel frequency-flat (or instantaneous) channels. Although each subband is a special case of a single-carrier MIMO system, more optimal methods can be adopted by exploiting the OFDM structure (the dependencies between subbands) . For more details in this direction, we refer to [31–35] and references therein.

A new trend in the field of channel estimation considers *linear time-varying (LTV) channels*. Here, the channel is assumed to vary over a data block. There are two approaches: (i) channel tracking, where an initial estimate is updated as time pro-

(a)

(b)

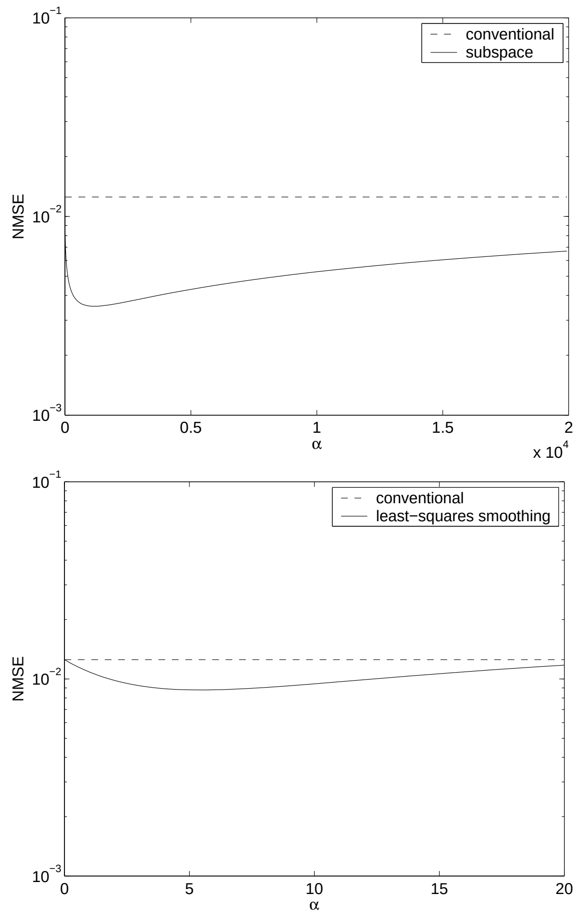


Fig. 1.5 Semi-blind algorithms: Comparison of the different methods (a) subspace method, (b) least-squares smoothing method.

gresses, and (ii) model-based block solutions, where the number of unknown channel coefficients is limited via a parametric model for the channel time variation, such as an exponential or polynomial basis expansion model. These parametric models are studied in, e.g., [36–39].

Another recent trend is to consider *superimposed training*. The schemes in this chapter were based on time-division multiplexing. Superimposed training is to add to the data stream $x^{(d)}(t)$ a known training signal $x^{(t)}(t)$, so that $x(t) = x^{(d)}(t) + x^{(t)}(t)$. Channel estimation is possible e.g., if the data symbols are i.i.d. and zero-mean, whereas the pilot has certain periodicities (cyclostationary properties) [40].

For the FIR-SIMO case, semi-blind methods of the type we discussed have been presented in [6], [41]. Some of the earlier (FIR-SISO) semi-blind papers are [42–45]. In Section 1.5, we presented two deterministic techniques for blind channel estimation. In general, such techniques may exploit (i) the convolutive structure (Hankel/Toeplitz), via oversampling or multiple channels, (ii) instantaneous properties of the sources, such as their finite alphabet and constant modulus, or (iii) stochastic properties such as statistical independence and cyclostationarity. This area has seen tremendous activity in the 1990s. Overviews can be found in [26, 46–48]. There are also several examples of algorithms which combine training with blind source properties, but the topic has not systematically been researched. As mentioned in Section 1.5.4, the problem of "α"-scaling of the training vs. blind parts of the cost function remains essentially open, although several heuristics are known [6].

Acknowledgement

This work was supported by NWO-STW under the VICI programme (DTC.5893), and the Commission of the EC under contract no. 506790 (IST STREP "UBROAD").

1.8 REFERENCES

1. W. Jakes, ed., *Microwave Mobile Communications*. New York: John Wiley, 1974.

2. O. Simeone and U. Spagnolini, "Lower bounds on the channel estimation error for fast-varying frequency-selective Rayleigh MIMO channels," in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Hong Kong), pp. V–69–72, April 2003.

3. X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," in *Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 1107–1111, November 2002.

4. S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – part I: Algorithms," *IEEE Trans. Signal Proc.*, vol. 44, pp. 1184–1197, May 1996.

5. V. Barroso, J. Moura, and J. Xavier, "Blind array channel division multiple access (AChDMA) for mobile communications," *IEEE Trans. Signal Proc.*, vol. 46, pp. 737–752, March 1998.

6. E. de Carvalho and D. T. M. Slock, "Semi-blind methods for FIR multichannel estimation," in *Signal Processing Advances in Wireless & Mobile Communications: Trends in Channel Estimation and Equalization* (G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds.), vol. 1, ch. 7, pp. 211–254, Prentice Hall, 2000.

7. E. de Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE Workshop on Signal Processing Advances for Wireless communications (SPAWC)*, (Paris, France), pp. 129–132, April 1997.

8. O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "Gaussian maximum likelihood channel estimation with short training sequences," *IEEE Trans. Wireless Comm.*, 2004. to appear; See also Proc. Globecom'03.

9. J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Signal Proc.*, vol. 45, pp. 658–672, March 1997.

10. J. K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise," *IEEE Trans. Signal Proc.*, vol. 46, pp. 3117–3123, November 1998.

11. E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Proc.*, vol. 43, pp. 516–525, February 1995.

12. K. Abed-Meraim, P. Loubaton, and E. Moulines, “A subspace algorithm for certain blind identification problems,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, March 1997.

13. P. Loubaton, E. Moulines, and P. Regalia, “Subspace method for blind identification and deconvolution,” in *Signal Processing Advances in Wireless & Mobile Communications: Trends in Channel Estimation and Equalization* (G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds.), vol. 1, ch. 3, pp. 63–112, Prentice Hall, 2000.

14. L. Tong and Q. Zhao, “Joint order detection and blind channel estimation by least squares smoothing,” *IEEE Trans. Signal Proc.*, vol. 47, pp. 2345–2355, September 1999.

15. Z. Ding, “Linear predictive algorithms for multichannel identification,” in *Signal Processing Advances in Wireless & Mobile Communications: Trends in Channel Estimation and Equalization* (G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds.), vol. 1, ch. 6, pp. 179–210, Prentice Hall, 2000.

16. Z. Ding, “Matrix outer-product decomposition method for blind multiple channel identification,” *IEEE Trans. Signal Proc.*, vol. 45, pp. 3053–3061, December 1997.

17. Z. Ding and L. Qiu, “Blind MIMO channel identification from second order statistics using rank deficient channel convolution matrix,” *IEEE Trans. Signal Proc.*, vol. 51, pp. 535–544, February 2003.

18. D. Gesbert and P. Duhamel, “Robust blind channel identification and equalization based on multi-step predictors,” in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Munich, Germany), pp. 3621–3624, April 1997.

19. J. K. Tugnait and B. Huang, “Multistep linear predictors-based blind identifications and equalization of multi-input multi-output channels,” *IEEE Trans. Signal Proc.*, vol. 48, pp. 26–38, January 2000.

20. D. T. M. Slock, “Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction,” in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Adelaide, Australia), pp. IV/585–IV/588, April 1994.

21. K. Abed-Meraim, E. Moulines, and P. Loubaton, “Prediction error method for second-order blind identification,” *IEEE Trans. Signal Proc.*, vol. 45, pp. 694–705, March 1997.

22. D. T. M. Slock, “Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas,” in *Asilomar Conf. on Signals, Systems, and Computers*, (Pacific Grove, California), pp. 1154–1158, October/November 1994.

23. A. Gorokhov and P. Loubaton, “Blind identification of MIMO-FIR systems: a generalized linear prediction approach,” *Signal Processing*, vol. 73, pp. 105–124, 1999.

24. A. Medles, D. T. M. Slock, and E. de Carvalho, “Linear prediction based semi-blind estimation of MIMO FIR channels,” in *IEEE Workshop on Signal Proc. Adv. for Wireless Comm. (SPAWC)*, (Taoyuan, Taiwan), pp. 58–61, March 2001.

25. J. Cavers, “An analysis of pilot symbol assisted modulation for Rayleigh fading channels,” *IEEE Trans. Vehic. Technol.*, vol. 40, pp. 686–693, November 1991.

26. G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds., *Signal Processing Advances in Wireless & Mobile Communications: Trends in Channel Estimation and Equalization*, vol. 1. Prentice Hall, 2000.

27. J. K. Tugnait, L. Tong, and Z. Ding, “Single-user channel estimation and equalization,” *IEEE Signal Proc. Magazine*, vol. 17, pp. 17–28, May 2000.

28. S. Adireddy and L. Tong, “Detection with embedded known symbols: Optimal symbol placement and equalization,” in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Istanbul, Turkey), June 2000.

29. H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, “Optimal training for frequency-selective channels,” in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Salt Lake City, Utah), May 2001.

30. M. Dong and L. Tong, “Optimal design and placement of pilot symbols for channel estimation,” *IEEE Trans. Signal Proc.*, vol. 50, pp. 3055–3069, Dec. 2002.

31. B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?,” *IEEE Trans. Informat. Th.*, vol. 48, pp. 951–963, Apr. 2003.

32. S. Adireddy and L. Tong, “Optimal placement of training for frequency-selective block-fading channels,” *IEEE Trans. Informat. Th.*, vol. 48, pp. 2338–2353, Aug. 2003.

33. T.-L. Tung and K. Yao, “Channel estimation and optimal power allocation for a multiple-antenna OFDM system,” *Eurasip Journal on Applied Signal Proc.*, pp. 330–339, March 2002.

34. I. Barhumi, G. Leus, and M. Moonen, “Optimal training design for MIMO OFDM systems in mobile wireless channels,” *IEEE Trans. Signal Proc.*, vol. 51, pp. 1615–1624, June 2003.

35. H. Bölcskei, R. W. Heath, Jr., and A. J. Paulraj, “Blind channel identification and equalization in OFDM-based multiantenna systems,” *IEEE Trans. Signal Proc.*, vol. 50, pp. 96–109, January 2002.

36. M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. Journal Adaptive Control Signal Proc.*, vol. 10, pp. 159–176, March 1996.

37. G. B. Giannakis and C. Tepedelenlioğlu, "Basis expansion models and diversity techniques for blind equalization of time-varying channels," *Proceedings of the IEEE*, pp. 1969–1986, 1998.

38. X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Proc.*, vol. 51, pp. 1351–1366, May 2003.

39. G. Leus, I. Barhumi, and M. Moonen, "MMSE time-varying FIR equalization of doubly-selective channels," in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, (Hong Kong), April 2003.

40. J. Tugnait and W. Luo, "On channel estimation using superimposed training and first-order statistics," *IEEE Trans. Signal Proc.*, vol. 51, pp. 413–415, Dec. 2003.

41. V. Buchoux, O. Cappé, E. Moulines, and A. Gorokhov, "On the performance of semi-blind subspace-based channel estimation," *IEEE Trans. Signal Proc.*, vol. 48, pp. 1750–1759, June 2000.

42. E. de Carvalho and D. T. M. Slock, "Cramér-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *IEEE workshop on Signal Processing Advances in Wireless Communications*, (Paris), pp. 129–132, Apr. 1997.

43. E. de Carvalho and D. T. M. Slock, "Maximum-likelihood blind FIR multi-channel estimation with Gaussian prior for the symbols," in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pp. 3593–3596, 1997.

44. D. Pal, "Fractionally spaced equalization of multipath channels: A semi-blind approach," in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pp. 9–12 vol.3, 1993.

45. A. Gorokhov and P. Loubaton, "Semi-blind second order identification of convolutive channels," in *IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, pp. 3905–3908, 1997.

46. L. Tong and S. Perreau, "Multichannel blind identification: From subspace to Maximum Likelihood Methods," *Proceedings of the IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

47. A.-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proceedings of the IEEE*, vol. 86, pp. 1987–2008, Oct. 1998.

48. K. Abed-Meraim, W. Qui, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, pp. 1310–1322, Aug. 1997.