Chapter 3

Separate Temporal and Spatial Parametric Channel Estimation

3.1 Introduction

Due to the increasing demand of capacity in wireless telecommunications, it is of great interest to find ways to increase the spectral efficiency, i.e. the number of mobiles per MHz and square kilometer. Since the total available bandwidth is limited, the focus is set on the spatial dimension. Spatial division multiple access schemes (SDMA) can be combined with more traditional access schemes such as F/TDMA (the mobiles are assigned a timeslot and a carrier frequency). In the F/TDMA-case either the reuse factor can be increased or more than one mobile can be allocated to the same frequency and timeslot, giving a great increase in the spectral efficiency. The spatial dimension can be exploited by means of an array antenna at the base station, in combination with a spatio-temporal equalizer. An indirect equalizer requires an estimate of the spatio-temporal channel from the mobile to the base station.

In [8], a maximum likelihood method for estimation of direction of arrivals (DOAs) and relative gains in a path model, for several possibly coherent signals, is derived. In [44] this method is utilized in order to exploit the spatial structure when identifying channels to an array of antennas. In this chapter we propose an improved method. First, an initial estimate of the channels to the individual antenna elements is formed, exploiting the *a priori* known temporal filtering in the transmitter and the receiver, as described in Chapter 2, but for the single user case. The spatial structure of the channel is then exploited by projecting the channel onto a subset of channels parameterized in DOAs and relative gains. For the

projection we use a spectrum norm, weighted with the inverse of the uncertainty of the spectrum of the initial estimate.

In Section 3.4, a simulation study is presented where the method proposed in this work is compared to other methods by means of the relative channel error and the BER (bit-error-rate).

3.2 Channel Identification Utilizing Pulse Shaping Information

Consider the uplink transmission, i.e. the transmission from the mobile to the array antenna at the base station. The array has M elements and an arbitrary structure. The symbols are pulse shaped at the transmitter, upconverted to the air interface frequency and transmitted through the air interface. Filtering and downconversion is then performed at the receiver. The physical channel from the transmitter to the receiver can be dispersive both in time and in space. The resulting system can be modeled by a continuous time filter with impulse response $p(t_c)$ (t_c denotes continuous time) containing all *a priori* known filtering at the transmitter and the receiver, a physical channel with impulse response $h_i(t_c)$ from the mobile to the i:th antenna element, and additive noise $n_i(t_c)$. A crucial assumption we make is that the channel is time invariant or at least slowly varying during the training sequence. Otherwise, the impulse response of the pulse shaping filter of the transmitter and the impulse response of the pulse shaping filer of the receiver cannot be accurately, convolved. We also assume the pulse shaping filters of all receivers to be identical. Figure 3.1 shows the continuous time model, where the derivative operator p is used as in Chapter 2.

In this model the impulse response $h_i(t_c)$ is a sum of a finite number of delayed dirac pulses each multiplied by a complex coefficient, corresponding to the reflections of the transmitted wave. The input symbols $d(t_c)$ are *modeled* by a train of dirac pulses, each pulse multiplied by a (possibly complex) discrete time symbol from the symbol alphabet. The additive noise term $n_i(t_c)$ contains noise sources such as thermal noise, but also co-channel interference that is not modeled.

Consider the transmission of a single symbol from the mobile to the base station. In the continuos time model of Figure 3.1, the discrete time symbol is modeled as a continuos time dirac pulse multiplied by the symbol. Convolution of the dirac pulse and the impulse response $p(t_c)$ of the pulse

shaping filter, gives a pulse scaled by the symbol. The pulse propagates through the physical channel resulting in a *superposition of delayed scaled pulses*. This is the impulse response of the complete system from the mobile to the base station. By sampling the continuous time impulse response, an FIR-model of the total channel is obtained.

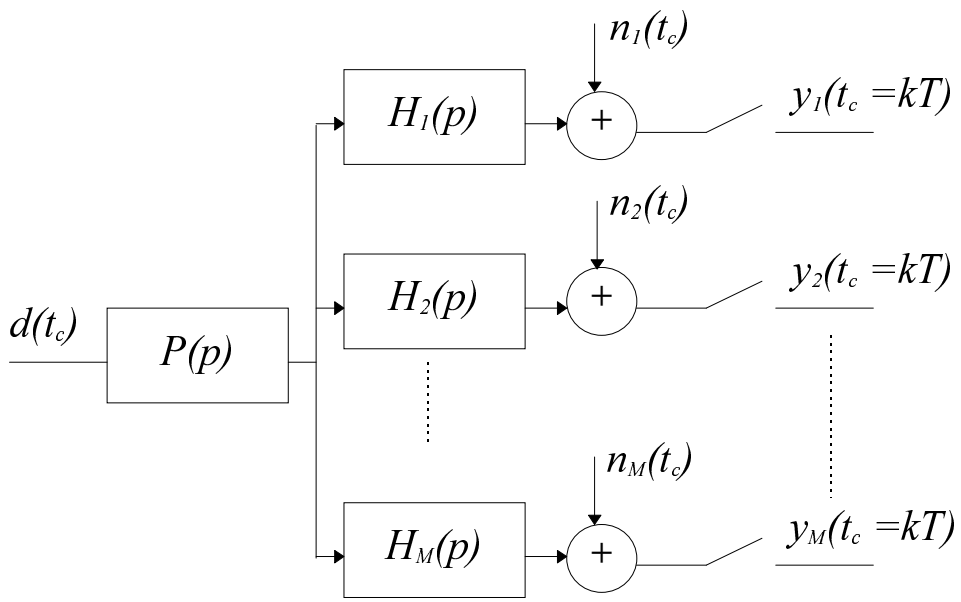


Fig. 3.1. Continuous time transmission model.

Setting the sampling period T to 1, we get the output sequence $y_i(t)$ for the i:th antenna, where t represents discrete time. Also, the input symbols are in the sequel considered to be discrete in time, and will be denoted by $d(t)$. The transfer function from the symbols $d(t)$ to the discrete time output $y_i(t)$ will therefore be modeled by the FIR-filter $B_i(q^{-1})$, where q^{-1} as before represents the backward shift operator, i.e. $q^{-1}x(t)=x(t-1)$. The continuous time noise $n_i(t_c)$ consists of co-channel interference and white thermal noise which is filtered by the receiver filter. Thus, the continuous time stochastic process $n_i(t_c)$ will be heavily colored. To obtain an accurate statistical representation of the noise in discrete time (conservation of the second order moments in the sampling points) stochastic sampling is required. See e.g. [45]. Here, we shall however refrain from deriving an exact statistical representation: It will simply be to complicated to derive. It will here be sufficient to observe that the discrete time noise will be colored with a certain distribution. The model is depicted in Figure 3.2.

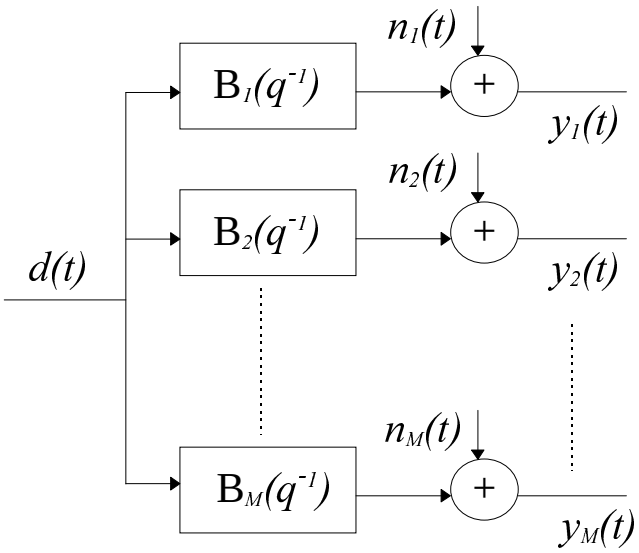


Fig. 3.2. Simple discrete time model.

The number of filter taps nb+1 in each channel model $B_i(q^{-1})$ has to be selected before identification. The channel coefficients can then be calculated by formulating a least squares problem utilizing the measured outputs $y_i(t)$ and a known sequence of the input $d(t)$, e.g. the training sequence normally present in each burst of F/TDMA systems.

Since the filtering performed at both the receiver and transmitter can be considered known, it is wiser to use this information and only estimate the unknown part of the channel. An identification method utilizing this information was presented in Chapter 2. For convenience, we recapitulate the basic principles here.

The discrete time version of the model is obtained by sampling a truncated impulse response $p(t_c)$ of the pulse shaping filter. The samples are used as taps of an FIR pulse shaping filter $P(q^{-1})$. The continuous time filter $h_i(t_c)$ is replaced by the FIR-filter $H_i(q^{-1})$ for each antenna. The model can be seen in Figure 3.3. By feeding the training sequence into the filter $P(q^{-1})$ the signal $x(t)$ is obtained. This signal, which is not white, can be used together with $y_i(t)$ to form a least-squares estimate of $H_i(q^{-1})$. The number of parameters that have to be estimated is therefore reduced compared to the estimation of $B_i(q^{-1})$ in Figure 3.2, especially if $P(q^{-1})$ has a long duration. Since the number of parameters are reduced we will obtain a variance reduction in the remaining parameter estimates. This is especially useful if the training sequence is short. One have to bear in mind that problems of identifiability may occur due to the filtering of the symbols in $P(q^{-1})$. If $P(q^{-1})$ is a lowpass filter and $H_i(q^{-1})$ a highpass filter, the accuracy in the parameter estimate will be low compared to a situation where $H_i(q^{-1})$ is a lowpass filter.

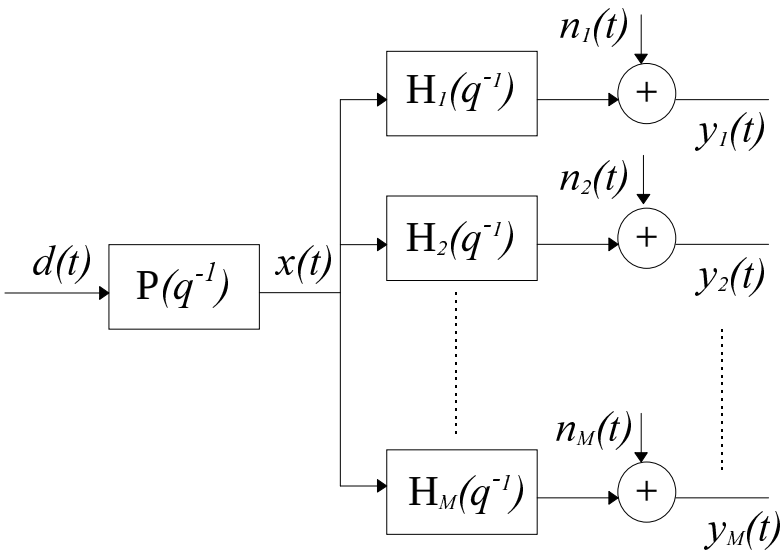


Fig. 3.3. Channel model with pulse shaping filter.

Since all sampled pulses cannot be well approximated by $P(q^{-1})$, two (or more) versions of the pulse shaping filter, sampled at different time instants, can be used. The training symbols are passed through these filters, and the channel can again be estimated by the use of the filtered signals and the output.

The method can also be extended to the case of fractional sampling. This doubles the amount of data whereas the number of parameters to estimate remain constant (see Chapter 2). However, one has to bear in mind that as the sampling rate increases, the noise samples tend to be increasingly correlated, due to the band-limited filter in the receiver: To some extent we only get more of the same data.

A model for a fractionally spaced system using two samples per symbol and two pulse shaping branches is depicted in Figure 3.4. The filter $P_{0.0}(q^{-1})$ corresponds to the pulse form sampled without delay relative to the ”center” of the pulse, the filter $P_{0.5}(q^{-1})$ to the pulse form sampled 0.5*T* *before* the center of the pulse and $P_{-0.5}(q^{-1})$ to the pulse form 0.5*T* *after* the center of the pulse.

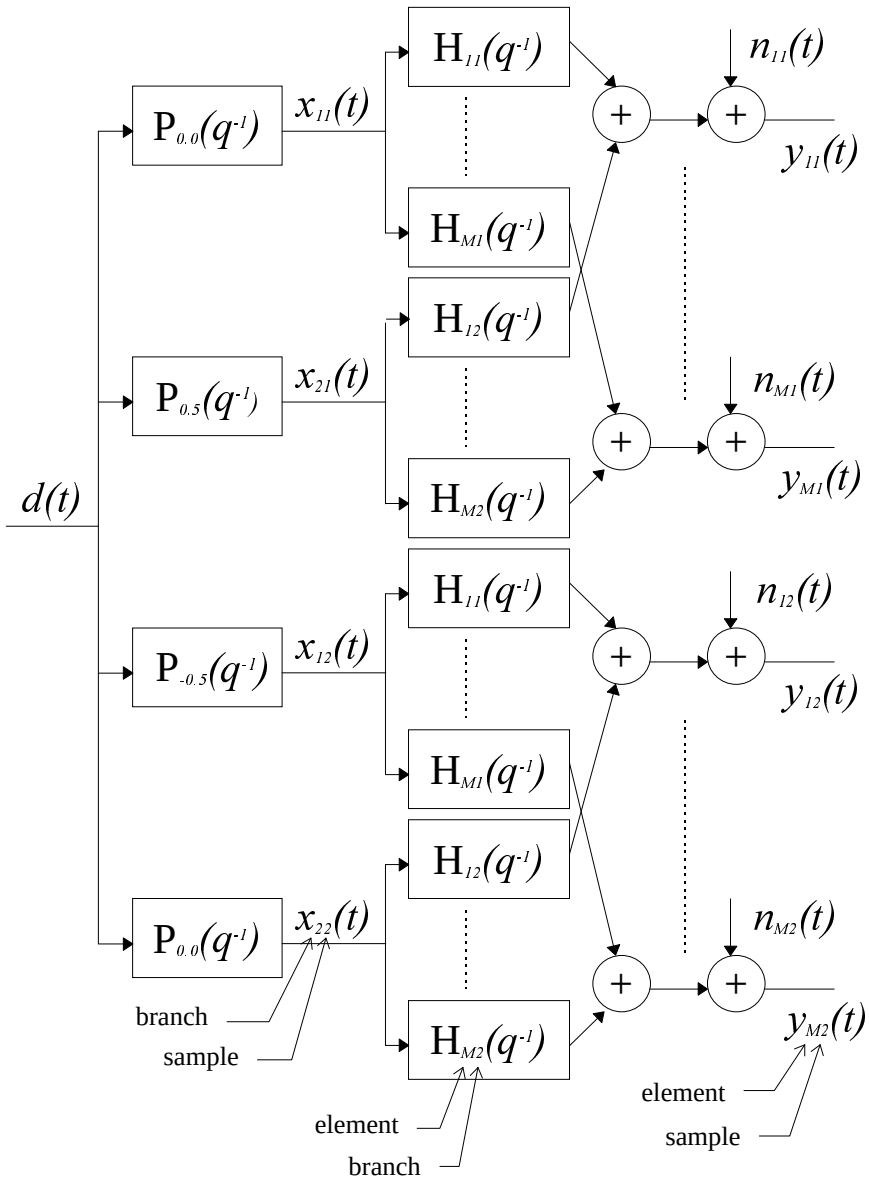


Fig. 3.4. Channel model utilizing multiple pulse shaping branches and oversampling.

The signal $x_{jk}(t)$ is the filtered input signal corresponding to the j:th branch and the k:th fractional sample, and $\mathrm{H}_{ij}(q^{-1})$ is the filter of the i:th antenna element and the j:th branch, defined as

$$\mathrm{H}_{ij}(q^{-1}) = h_{ij0} + h_{ij1}q^{-1} + \mathrm{K} + h_{ijnh}q^{-nh}. \tag{3.1}$$

The signals $y_{ik}(t)$ and $n_{ik}(t)$ are the output and noise, respectively, corresponding to the i:th antenna element and the k:th fractional sample. The indices i, j and k span over

i=1,2,…,M
j=1,2
k=1,2 .

The output samples and the noise samples at time t are collected in the Mx2 matrices $Y(t)$ and $N(t)$, given by

$$Y(t) = \begin{bmatrix} y_{11}(t) & y_{12}(t) \\ y_{21}(t) & y_{22}(t) \\ \mathrm{M} & \mathrm{M} \\ y_{M1}(t) & y_{M2}(t) \end{bmatrix} \tag{3.2}$$

and

$$N(t) = \begin{bmatrix} n_{11}(t) & n_{12}(t) \\ n_{21}(t) & n_{22}(t) \\ \mathrm{M} & \mathrm{M} \\ n_{M1}(t) & n_{M2}(t) \end{bmatrix}. \tag{3.3}$$

The filtered input symbols are generated according to

$$\begin{bmatrix} x_{11}(t) & x_{12}(t) \\ x_{21}(t) & x_{22}(t) \end{bmatrix} = \mathrm{P}(q^{-1})d(t) = \begin{bmatrix} \mathrm{P}_{0.0}(q^{-1}) & \mathrm{P}_{-0.5}(q^{-1}) \\ \mathrm{P}_{0.5}(q^{-1}) & \mathrm{P}_{0.0}(q^{-1}) \end{bmatrix} d(t) \tag{3.4}$$

with

$$\mathrm{P}_{0.0}(q^{-1}) = p_{0.0,0} + p_{0.0,1}q^{-1} + \mathrm{K} + p_{0.0,np}q^{-np} \tag{3.5}$$

The pulse shaping filters $\mathrm{P}_{0.5}(q^{-1})$ and $\mathrm{P}_{-0.5}(q^{-1})$ are defined similarly. The 2(nh+1)x2 matrix $X(t)$ is defined as

$$X(t) = \begin{bmatrix} x_{11}(t) & x_{11}(t-1) & \mathrm{K} & x_{11}(t-nh) & x_{21}(t) & x_{21}(t-1) & \mathrm{K} & x_{21}(t-nh) \\ x_{12}(t) & x_{12}(t-1) & \mathrm{K} & x_{12}(t-nh) & x_{22}(t) & x_{22}(t-1) & \mathrm{K} & x_{22}(t-nh) \end{bmatrix}^T \tag{3.6}$$

The filter taps of the filters $H_{ij}(q^{-1})$, i=1,2,…,M, j=1,2, are collected in the matrix H, according to

$$H = \begin{bmatrix} h_{110} & h_{111} & \mathrm{K} & h_{11nh} & h_{120} & h_{121} & \mathrm{K} & h_{12nh} \\ h_{210} & h_{211} & \mathrm{K} & h_{21nh} & h_{220} & h_{221} & \mathrm{K} & h_{22nh} \\ \mathrm{M} & \mathrm{M} & \mathrm{K} & \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{K} & \mathrm{M} \\ h_{M10} & h_{M11} & \mathrm{K} & h_{M1nh} & h_{M20} & h_{M21} & \mathrm{K} & h_{M2nh} \end{bmatrix}. \quad (3.7)$$

The model of Figure 3.4 can now be written on matrix from, according to

$$Y(t) = HX(t) + N(t). \quad (3.8)$$

The least-squares estimate of H is therefore given by

$$\hat{H} = \hat{R}_{YX} \hat{R}_{XX}^{-1} \quad (3.9)$$

where the estimate of the covariance matrices are given by

$$\hat{R}_{YX} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} \mathrm{Y}(t) X^{\mathrm{H}}(t) \quad (3.10)$$

$$\hat{R}_{XX} = \frac{1}{N - nb} \sum_{t=nb+1}^{N} X(t) X^{\mathrm{H}}(t) \quad (3.11)$$

with N being the length of the training sequence, $nb+1$ the length of the *total* channel and ^{H} complex conjugate transpose. The summation is taken from $nb+1$ to N since we want to make sure that the output only depend on the training symbols. The estimated ”physical channel” can now be written as a polynomial matrix, according to

$$\hat{\mathrm{H}}(q^{-1}) = \begin{bmatrix} \hat{\mathrm{H}}_{11}(q^{-1}) & \hat{\mathrm{H}}_{21}(q^{-1}) & \mathrm{K} & \hat{\mathrm{H}}_{M1}(q^{-1}) \\ \hat{\mathrm{H}}_{12}(q^{-1}) & \hat{\mathrm{H}}_{22}(q^{-1}) & \mathrm{K} & \hat{\mathrm{H}}_{M2}(q^{-1}) \end{bmatrix}^{T} \quad (3.12)$$

where $\hat{\mathrm{H}}(q^{-1})$ is formed by the elements of $\hat{H}$, i.e.

$$\hat{\mathrm{H}}_{ij}(q^{-1}) = \hat{h}_{ij0} + \hat{h}_{ij1} q^{-1} + \mathrm{K}\ \hat{h}_{ijnh} q^{-nh}. \quad (3.13)$$

The total estimated channel is then given by

$$\hat{B}(q^{-1}) = \hat{H}(q^{-1})P(q^{-1}) \tag{3.14}$$

where

$$\hat{B}(q^{-1}) = \begin{bmatrix} \hat{B}_{11}(q^{-1}) & \hat{B}_{21}(q^{-1}) & \cdots & \hat{B}_{M1}(q^{-1}) \\ \hat{B}_{12}(q^{-1}) & \hat{B}_{22}(q^{-1}) & \cdots & \hat{B}_{M2}(q^{-1}) \end{bmatrix}^T \tag{3.15}$$

and $\hat{B}_{ik}(q^{-1})$, given by

$$\hat{B}_{ik}(q^{-1}) = \hat{b}_{ik0} + \hat{b}_{ik1}q^{-1} + \cdots \hat{b}_{iknb}q^{-nb}, \tag{3.16}$$

is the total channel from the mobile to the *i*:th antenna element and the *k*:th fractional sample.

3.3 Projection onto a Spatially Parameterized Subset

In this section, the *temporal* channel estimate obtained according to the method described above, here denoted $\hat{B}_0(q^{-1})$, is projected onto a spatially parameterized subset $B_{\theta,\gamma}(q^{-1})$. This will in general improve the channel estimate since the *spatial structure* of the channel is exploited.

The *M*x2 polynomial matrix $\hat{B}_0(q^{-1})$ is used as an initial estimate of the channel from the mobile to the array antenna. Since the channel structure is assumed to arise from the multiple reflections of the transmitted signal, the true channel will belong to a subset, parameterized by angles and relative gains (signal strength) of the multipath components. The initial estimate $\hat{B}_0(q^{-1})$ will in general not belong to the subset $B_{\theta,\gamma}(q^{-1})$, and the idea is therefore to project $\hat{B}_0(q^{-1})$ onto $B_{\theta,\gamma}(q^{-1})$. The projection can be illustrated as in Figure 3.5.

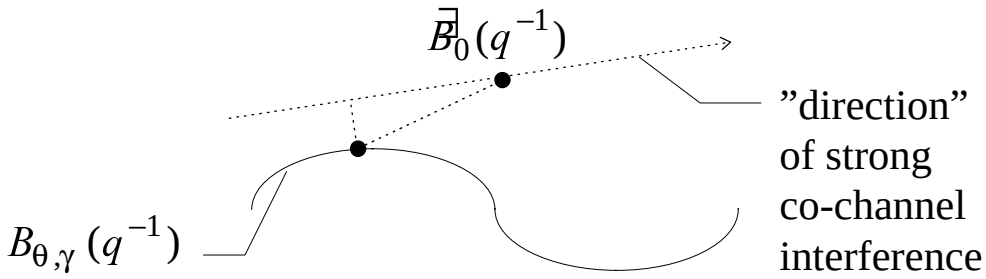


Fig 3.5. Illustration of the projection.

Errors in the initial estimate in ”directions” where we have strong co-channel interference are paid little attention. A more detailed description of the projection is given below, but first the involved parameters and variables are properly defined.

The parameterized subset $B_{\theta,\gamma}(q^{-1})$ is given by

$$B_{\theta,\gamma}(q^{-1}) = \sum_{l=0}^{nb} b_l(\theta_l,\gamma_l)q^{-l} = \sum_{l=0}^{nb} \left[b_{1l}(\theta_{1l},\gamma_{1l}) \quad b_{2l}(\theta_{2l},\gamma_{2l})\right]q^{-l}. \tag{3.17}$$

The l:th *vector* tap corresponding to the k:th fractional sample can be written as

$$b_{kl}(\theta_{kl},\gamma_{kl}) = [b_{1kl} \quad b_{2kl} \quad \mathrm{K} \quad b_{Mkl}]^T \tag{3.18}$$

or as

$$b_{kl}(\theta_{kl},\gamma_{kl}) = \sum_{m=1}^{K_{kl}} \gamma_{klm} a(\theta_{klm}) \tag{3.19}$$

where $a(\theta_{klm})$ is the Mx1 array response vector parameterized by the angle θ_{klm}. The number of impinging plane waves for the l:th vector tap and the k:th fractional sample is given by K_{kl}. The angles and relative gains corresponding to each of the plane waves are collected in the parameter *vectors* θ_{kl} and γ_{kl}, respectively, given by

$$\begin{aligned} \theta_{kl} &= \left[\theta_{kl1} \quad \mathrm{K} \quad \theta_{klK_{kl}}\right]^T \\ \gamma_{kl} &= \left[\gamma_{kl1} \quad \mathrm{K} \quad \gamma_{klK_{kl}}\right]^T. \end{aligned} \tag{3.20}$$

The parameter *matrices* θ and γ are defined as

$$\begin{aligned}\theta &= \begin{bmatrix}\theta_{10} & \theta_{20} & \theta_{11} & \theta_{21} & \mathrm{K} & \theta_{1nb} & \theta_{2nb}\end{bmatrix}\\ \gamma &= \begin{bmatrix}\gamma_{10} & \gamma_{20} & \gamma_{11} & \gamma_{21} & \mathrm{K} & \gamma_{1nb} & \gamma_{2nb}\end{bmatrix}\end{aligned} \tag{3.21}$$

In a realistic situation, one would have to consider the effects of over- and underestimation of the number of paths K_{kl} per sample and tap. This remains to be investigated. In this study, we also make the assumption that all angles are distinct, i.e. no angular spread.

The initial channel estimate $\hat{B}_0(q^{-1})$ is projected onto the parameterized subset $B_{\theta,\gamma}(q^{-1})$ in a *spectrum norm* sense, i.e.

$$\hat{\theta},\hat{\gamma} = \arg\min_{\theta,\gamma}\left\|\hat{B}_0(q^{-1}) - B_{\theta,\gamma}(q^{-1})\right\|^2_{R^{-1}_{\Delta\hat{B}_0,\Delta\hat{B}_0}(q,q^{-1})} \tag{3.22}$$

giving $B_{\hat{\theta},\hat{\gamma}}(q^{-1})$. The norm $\left\|A(q^{-1})\right\|^2_{W(q,q^{-1})}$ is defined as the *trace of the constant term* of the matrix polynomial

$$A^H(q)W(q,q^{-1})A(q^{-1}). \tag{3.23}$$

The *error* of the initial estimate $\hat{B}_0(q^{-1})$ is defined as

$$\Delta\hat{B}_0(q^{-1}) = \hat{B}_0(q^{-1}) - B_{true}(q^{-1}) \tag{3.24}$$

where $B_{true}(q^{-1})$ is the true channel (in the subset $B_{\hat{\theta},\hat{\gamma}}(q^{-1})$). The double-sided matrix polynomial $R_{\Delta\hat{B}_0,\Delta\hat{B}_0}(q,q^{-1})$ is the expectation of the spectrum of $\Delta\hat{B}_0(q^{-1})$, i.e.

$$R_{\Delta\hat{B}_0,\Delta\hat{B}_0}(q,q^{-1}) = E\left[\Delta\hat{B}_0(q^{-1})\Delta\hat{B}_0^{\,H}(q)\right]. \tag{3.25}$$

The weighting $R^{-1}_{\Delta\vec{B}_0,\Delta\vec{B}_0}(q,q^{-1})$ puts high weights in the areas of the spatio-temporal spectrum $\vec{B}_0(q^{-1})\vec{B}_0^H(q)$ that have low uncertainties, and low weights in the areas of $\vec{B}_0(q^{-1})\vec{B}_0^H(q)$ that have high uncertainties.

We now derive an expression for $R_{\Delta\vec{B}_0,\Delta\vec{B}_0}(q,q^{-1})$. The output matrix can be expressed as

$$Y(t) = \mathrm{H}(q^{-1})X(t) \tag{3.26}$$

where $X(t)$ here (as opposed to (3.6)) is defined as

$$X(t) = \begin{bmatrix} x_{11}(t) & x_{12}(t) \\ x_{21}(t) & x_{22}(t) \end{bmatrix} \tag{3.27}$$

and

$$\mathrm{H}(q^{-1}) = \begin{bmatrix} \mathrm{H}_{11}(q^{-1}) & \mathrm{H}_{21}(q^{-1}) & \mathrm{K} & \mathrm{H}_{M1}(q^{-1}) \\ \mathrm{H}_{12}(q^{-1}) & \mathrm{H}_{22}(q^{-1}) & \mathrm{K} & \mathrm{H}_{M2}(q^{-1}) \end{bmatrix}^T. \tag{3.28}$$

The polynomial matrix $\mathrm{H}(q^{-1})$ can also be expressed in spectral densities, see e.g. [42], i.e.

$$H(q^{-1}) = R_{YX}(q,q^{-1})R^{-1}_{XX}(q,q^{-1}) \tag{3.29}$$

where the spectral densities are given by

$$R_{YX}(q,q^{-1}) = \sum_{m=-np}^{np+nh} E\left[Y(t+m)X^H(t)\right]q^{-m} \tag{3.30}$$

$$R_{XX}(q,q^{-1}) = \sum_{m=-np}^{np} E\left[X(t+m)X^H(t)\right]q^{-m}. \tag{3.31}$$

The obvious approximation is therefore to use

$$\vec{H}(q^{-1}) = \vec{R}_{YX}(q,q^{-1})\vec{R}^{-1}_{XX}(q,q^{-1}) \tag{3.32}$$

where

$$\hat{R}_{YX}(q,q^{-1}) = \frac{1}{N_u}\sum_{t}\sum_{m=-np}^{np+nh} Y(t+m)X^H(t)q^{-m} \tag{3.33}$$

$$\hat{R}_{XX}(q,q^{-1}) = \frac{1}{N_u}\sum_{t}\sum_{m=-np}^{np} X(t+m)X^H(t)q^{-m} \tag{3.34}$$

and

$$N_u = \text{the number of terms used in the approximation.} \tag{3.35}$$

Above, np+1 and nh+1 are the number of taps in the pulse shaping filter and in the physical channel, respectively. The term $Y(t+m)X^H(t)$ will be small for $m \notin [-np, np+nh]$, since in this region, $Y(t+m)$ and $X^H(t)$ do not depend on any mutual symbols $d(t)$. The same holds for $X(t+m)X^H(t)$, $m \notin [-np, np]$. That is the motivation for summation indices in (3.30) and (3.31).

The output matrix $Y(t)$ can also be written as

$$Y(t) = Y_{true}(t) + N(t) \tag{3.36}$$

where Y_{true} is the ”noise-free” output given by

$$Y_{true}(t) = \mathrm{B}_{true}(q^{-1})d(t) \tag{3.37}$$

and $N(t)$ is the Mx2 noise matrix defined in (3.3). It is important to note that we have neglected errors in the channel estimate due to an incorrect model structure, i.e. only errors due to the noise is considered. Define the residues $\varepsilon(t)$ as

$$\varepsilon(t) = Y(t) - \hat{B}_0(q^{-1})d(t)\,. \tag{3.38}$$

The channel error is then given by

$$\begin{aligned}\Delta\hat{B}_0(q^{-1}) &= \hat{B}_0(q^{-1}) - B_{true}(q^{-1}) = (\hat{H}(q^{-1}) - H_{true}(q^{-1}))P(q^{-1}) \\ &= \Delta H(q^{-1})P(q^{-1}) = \hat{R}_{\varepsilon X}(q,q^{-1})\hat{R}_{XX}^{-1}(q,q^{-1})P(q^{-1})\end{aligned} \tag{3.39}$$

where $\vec{R}_{\varepsilon X}(q,q^{-1})$ is defined similarly to (3.33). For a large number of training symbols N, the following holds

$$\vec{R}_{XX}(q,q^{-1}) \approx P(q^{-1})P^H(q) \tag{3.40}$$

and

$$\begin{aligned}
&R_{\Delta\vec{B}_0\Delta\vec{B}_0}(q,q^{-1}) = E\left[\Delta\vec{B}_0(q^{-1})\Delta\vec{B}_0^{\,H}(q)\right] \approx \\
&E\{\sum_{t_1,m_1,t_2,m_2}\varepsilon(t_1+m_1)d^H(t_1)q^{m_2-m_1}P^H(q)\left[P(q^{-1})P^H(q)\right]^{-1}P(q^{-1})\times \\
&\times P^H(q)\left[P(q^{-1})P^H(q)\right]^{-H}P(q^{-1})d^H(t_2)\varepsilon(t_2+m_2)\} \approx \\
&E\{\varepsilon(t)P^H(q)\left(P(q^{-1})P^H(q)\right)^{-H}P(q^{-1})\varepsilon^H(t)\}.
\end{aligned} \tag{3.41}$$

In the approximations above we have assumed the symbols *d(t)* and the residues $\varepsilon(t)$ to be white. In practice, with co-channel interference, the noise will be spatially colored, but we still make the assumption above, since this leads to a radically simplified decoupled algorithm, as will be explained below.

For the special case of two branches and two samples per symbol, $P(q^{-1})$ will be a square matrix, and the expression above simplifies to

$$R_{\Delta\vec{B}_0\Delta\vec{B}_0}(q,q^{-1}) = E\left[\Delta\vec{B}_0(q^{-1})\Delta\vec{B}_0^{\,H}(q)\right] \approx E\left[\varepsilon(t)\varepsilon^H(t)\right] = \vec{R}_{\Delta\vec{B}_0\Delta\vec{B}_0}(q,q^{-1}) \tag{3.42}$$

i.e. a constant matrix (no q-dependence), which means that $\vec{R}_{\Delta\vec{B}_0,\Delta\vec{B}_0}(q,q^{-1}) \equiv \vec{R}_{\Delta\vec{B}_0,\Delta\vec{B}_0}$ is purely spatial. This means that the projection in this case is purely spatial, and therefore we call the projection here a *parametric spatial projection*. The minimization of the norm in (3.22) now decouples as shown below, i.e.

$$\begin{aligned}
&\left\|\vec{B}_0(q^{-1}) - B_{\theta,\gamma}(q^{-1})\right\|^2_{\vec{R}^{-1}_{\Delta B,\Delta B}} = \\
&= tr\left\{\Delta b_0^H \vec{R}^{-1}_{\Delta B_0,\Delta B_0}\Delta b_0 + \ldots + \Delta b_{nb}^H \vec{R}^{-1}_{\Delta B_0,\Delta B_0}\Delta b_{nb}\right\} \\
&= tr\left\{\Delta b_0^H \vec{R}^{-1}_{\Delta B_0,\Delta B_0}\Delta b_0\right\} + \ldots + tr\left\{\Delta b_{nb}^H \vec{R}^{-1}_{\Delta B_0,\Delta B_0}\Delta b_{nb}\right\}.
\end{aligned} \tag{3.43}$$

where

$$\Delta b_l = \hat{b}_l - b_l(\theta_l, \gamma_l). \tag{3.44}$$

The terms of (3.43) can therefore be minimized independently. The l:th term can be expressed as

$$\begin{aligned} tr\left\{\Delta b_l^H \hat{R}_{\Delta B_0,\Delta B_0}^{-1} \Delta b_l\right\} &= tr\left\{\begin{bmatrix} \Delta b_{1l}^H \\ \Delta b_{2l}^H \end{bmatrix} \hat{R}_{\Delta B_0,\Delta B_0}^{-1} \begin{bmatrix} \Delta b_{1l} & \Delta b_{2l} \end{bmatrix}\right\} \\ &= \Delta b_{l1}^H \hat{R}_{\Delta B_0,\Delta B_0}^{-1} \Delta b_{l1} + \Delta b_{l2}^H \hat{R}_{\Delta B_0,\Delta B_0}^{-1} \Delta b_{l2}. \end{aligned} \tag{3.45}$$

We thus get two independent minimizations per vector tap, and the minimizing angles and gains are therefore given by

$$\hat{\theta}_{kl}, \hat{\gamma}_{kl} = \arg\min_{\theta_i, \gamma_i} \left\{ \left\| \hat{b}_{kl} - b_{kl}(\theta_{kl}, \gamma_{kl}) \right\|_{\hat{R}_{\Delta B_0,\Delta B_0}^{-1}} \right\}. \tag{3.46}$$

This can be solved as (see [8])

$$\begin{cases} \hat{\gamma}_{kl} = \left\{ \hat{R}_{\Delta B,\Delta B}^{-1/2} A(\hat{\theta}_{kl}) \right\}^+ \hat{R}_{\Delta B,\Delta B}^{-1/2} \hat{b}_{kl} \\ \hat{\theta}_{kl} = \arg\min_{\theta_{kl}} \left\{ \hat{b}_{kl}^H \left[\hat{R}_{\Delta B,\Delta B}^{-1} - \hat{R}_{\Delta B,\Delta B}^{-1} A(\theta_{kl}) \right.\right. \\ \left.\left. \times \left[A^H(\theta_{kl}) \hat{R}_{\Delta B,\Delta B}^{-1} A(\theta_{kl}) \right]^{-1} A^H(\theta_{kl}) \hat{R}_{\Delta B,\Delta B}^{-1} \right] \hat{b}_{kl} \right\} \end{cases} \tag{3.47}$$

where

$$A(\theta_{kl}) = \left[a(\theta_{kl1}) \quad a(\theta_{kl2}) \quad \Lambda \quad a(\theta_{klK_k}) \right] \tag{3.48}$$

and $^+$ denotes pseudo inverse. The above non-quadratic minimization with respect to θ_{kl} can be solved by some gradient method, and the resulting angles and relative gains of each path in each tap give the final channel estimate.

3.4 Examples

In this section the channel identification methods discussed in Sections 3.2 and 3.3 are applied to a multipath channel with inter-symbol-interference. Independent Rayleigh fading is assumed, corresponding to scatterers close to the mobile. The example physical channel have taps of equal mean power at delays 0, 0.33T, 0.67T and T. A raised cosine pulse with roll-off factor 0.35 is used for the pulse shaping. For simplicity a ULA (uniform linear array) is used. Distinct angles are assumed (no distributed transmitters).

Four identification methods have been investigated and compared; unparameterized least-squares identification (LS), parametric spatial projection initialized with the LS-estimate of the channel (LS-PSP), the method of Chapter 2 utilizing the pulse shaping information with two branches and two times oversampling (PS), and finally the here derived parametric spatial projection initialized by PS (PS-PSP). It should be noted that the gradient method for the non-linear minimization in the PSP-methods are initialized by the true angles. The true angles would of course not be known in a real situation, but if the angles vary slowly, estimated angles from the previous frame could be used to initiate the angle calculation of the current frame. We only need initial estimates good enough to avoid false local minima.

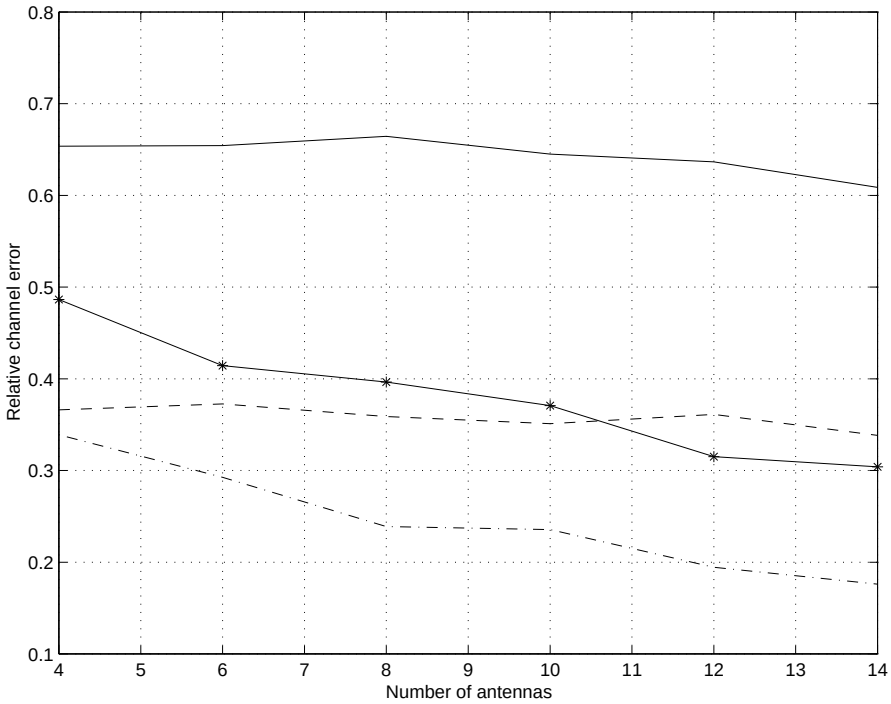
The methods can be compared by studying the relative channel error, defined as

$$\left\| \hat{B}_0 - B_{true} \right\| \Big/ \left\| B_{true} \right\| \tag{3.49}$$

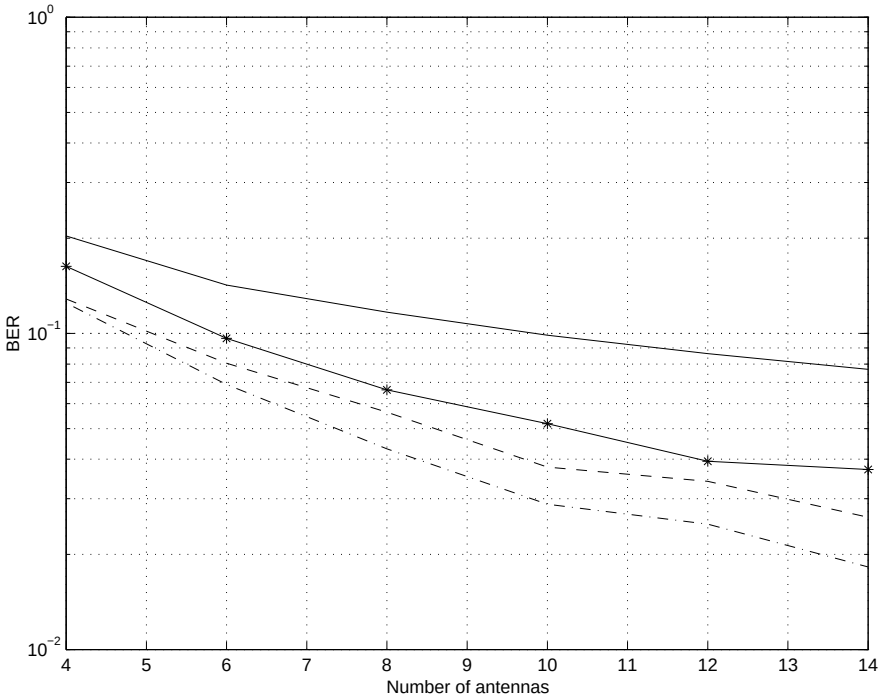
using the Frobenius norm, where $\hat{B}_0$ is a matrix with columns formed by the Mx2 matrix taps of $\hat{B}_0(q^{-1})$. A more useful performance criterion is the bit-error-rate (BER) achieved by using the channel estimates in an equalizer/symbol detector. In this chapter we use the multidimensional maximum likelihood sequence estimator (MLSE) discussed in [7].

Figure 3.6 shows the relative channel error and the BER as a function of the number of antennas. Each data point is the mean of 40 noise and fading realizations. The signal-to-noise ratio (SNR) was chosen to -5 dB. A co-channel-interferer with the same delay profile as the signal of interest but with different DOA:s was simulated, giving a signal-to-interference ratio (SIR) of -3 dB. The length of the BPSK-modulated training sequence was

chosen to 26 symbols. Since LS and PS are purely temporal methods, it is not surprising that the channel errors remain constant as we increase the number of antennas. LS-PSP and PS-PSP utilizes the spatial structure of the channel, and therefore the channel errors decrease as the number of antennas increases. Many more antennas than 14 cannot be studied in this case since the estimate of the covariance matrix becomes increasingly inaccurate as the number of antennas approaches the length of the training sequence.



(a)

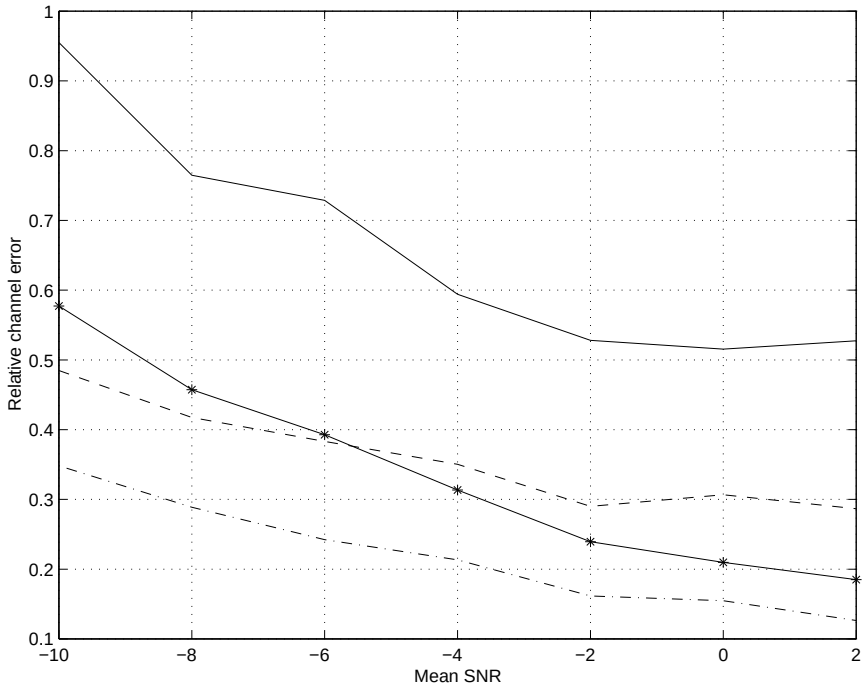


(b)

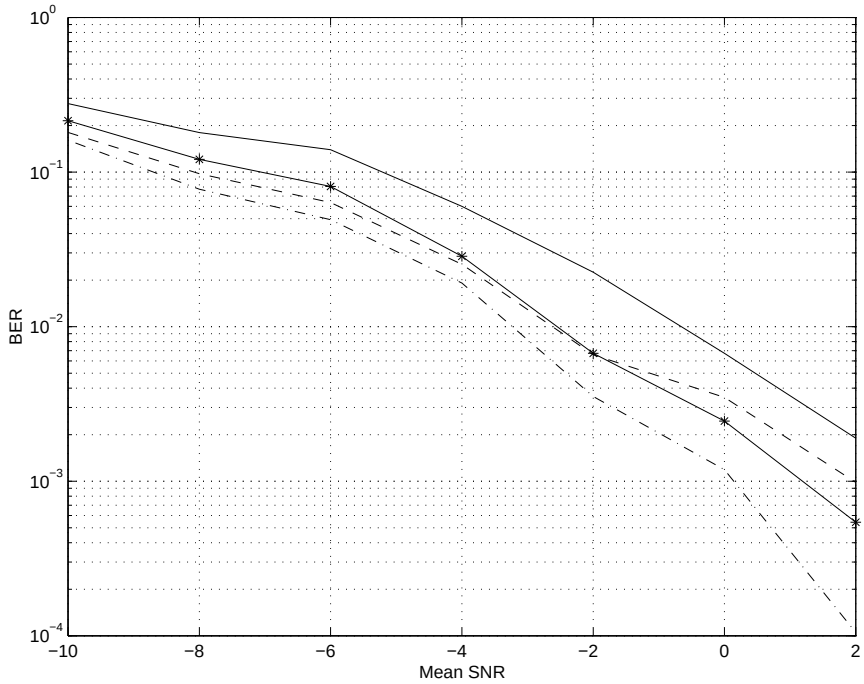
Fig. 3.6. Relative channel error (a) and BER (b) as a function of the number of antennas. Solid is LS, asterisk LS-PSP, dashed PS and dash-dot PS-PSP.

We can see that the BER is significantly improved for PS-PSP compared to LS-PSP. If 4 % is an acceptable BER, the number of antennas can be reduced from 12 to 9. Note that the BER for all methods decreases as the number of antennas increases. This is because the performance of the multidimensional MLSE improves with the number of antennas.

Figure 3.7 shows the relative channel error and the BER as a function of the average SNR per antenna. The number of antennas is ten, otherwise the situation is the same as above.



(a)



(b)

Fig. 3.7. Relative channel error (a) and BER (b) as a function of the mean SNR per antenna. Solid is LS, asterisk LS-PSP, dashed PS and dash-dot PS-PSP.

The relative channel error is improved by approximately a factor two independently of the SNR for PS-PSP compared to LS-PSP. This is also reflected in the BER.