Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas

Dirk T. M. Slock

Mobile Communications Department
Institut EURECOM, 2229 route des Crêtes, B.P. 193
06904, Sophia Antipolis Cedex, FRANCE

Abstract

We consider multiple (p) users that operate on the same carrier frequency and use the same linear digital modulation format. We consider $m > p$ antennas receiving mixtures of these signals through multipath propagation (equivalently, oversampling of the received signals of a smaller number of antenna signals could be used). We consider conditions on the matrix channel response for the existence of a Zero-Forcing Equalizer (ZFE) (which cancels inter-symbol and inter-user interference). In the noisefree case, we show how a ZFE can be obtained from linear prediction and the channel matrix itself can also be determined as a byproduct. The problem is one of signal and noise subspaces and we show a convenient way of solving the deterministic maximum likelihood problem using a minimal linear parameterization of the noise subspace. This parameterization is found as a byproduct in the linear prediction problem.

1 Matrix Channels

Consider linear digital modulation over a linear channel with additive Gaussian noise. Assume that we have p transmitters at a certain carrier frequency and m antennas receiving mixtures of the signals. We shall assume throughout that $m > p$. The received signals can be written in the baseband as

$$y_i(t) = \sum_{j=1}^{p}\sum_{k} a_j(k)h_{ij}(t-kT) + v_i(t) \quad (1)$$

where the $a_j(k)$ are the transmitted symbols from source j, T is the common symbol period, $h_{ij}(t)$ is the (overall) channel impulse response from transmitter j to receiver antenna i. Assuming the $\{a_j(k)\}$ and $\{v_i(t)\}$ to be jointly (wide-sense) stationary, the processes $\{y_i(t)\}$ are (wide-sense) cyclostationary with period T. If $\{y_i(t)\}$ is sampled with period T, the sampled process is (wide-sense) stationary. Sampling in this way leads to an equivalent discrete-time representation. We could also obtain multiple channels in the discrete time domain by oversampling the continuous-time received signals, see [1],[2],[3].

We assume the channels to be FIR. In particular, after sampling we assume the (vector) impulse response from source j to be of length N_j. Without loss of generality, we assume the first non-zero vector impulse response sample to occur at discrete time zero, and we can assume the sources to be ordered so that $N_1 \geq N_2 \geq \cdots \geq N_p$. Let $N = \sum_{j=1}^{p} N_j$. The discrete-time received signal can be represented in vector form as

$$\begin{aligned}\mathbf{y}(k) &= \sum_{j=1}^{p}\sum_{i=0}^{N_j-1} \mathbf{h}_j(i)a_j(k-i) + \mathbf{v}(k)\\ &= \sum_{i=0}^{N_1-1} \mathbf{h}(i)\mathbf{a}(k-i) + \mathbf{v}(k)\\ &= \sum_{j=1}^{p} \mathbf{H}_{j,N_j}A_{j,N_j}(k) + \mathbf{v}(k) = \mathbf{H}_N\mathbf{A}_N(k) + \mathbf{v}(k)\,,\end{aligned}$$

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k)\\ \vdots \\ y_m(k)\end{bmatrix}, \mathbf{v}(k) = \begin{bmatrix} v_1(k)\\ \vdots \\ v_m(k)\end{bmatrix}, \mathbf{h}_j(k) = \begin{bmatrix} h_{1j}(k)\\ \vdots \\ h_{mj}(k)\end{bmatrix}$$

$$\begin{aligned}&\mathbf{H}_{j,N_j} = [\mathbf{h}_j(N_j-1)\cdots\mathbf{h}_j(0)]\,, \mathbf{H}_N = [\mathbf{H}_{1,N_1}\cdots\mathbf{H}_{p,N_p}]\\ &\mathbf{h}(k) = [\mathbf{h}_1(k)\cdots\mathbf{h}_p(k)]\,, \mathbf{a}(k) = \left[a_1^H(k)\cdots a_p^H(k)\right]^H\\ &A_{j,N_j}(k) = \left[a_j^H(k-N_j+1)\cdots a_j^H(k)\right]^H\\ &\mathbf{A}_N(k) = \left[A_{1,N_1}^H(k)\cdots A_{p,N_p}^H(k)\right]^H\end{aligned} \quad (2)$$

where superscript ^{H} denotes Hermitian transpose.

2 FIR Zero-Forcing (ZF) Equalization

We consider a structure of equalizers as in Fig. 1 to not only cancel the intersymbol interference for every source separately, but also cancel the interference between different sources. We assume the equalizer filters to be FIR of length L: $F_{ji}(z) = \sum_{k=0}^{L-1} f_{ji}(k)z^{-k}$, $j = 1,\ldots,p, i = 1,\ldots,m$. We introduce $\mathbf{f}_j(k) = [f_{j1}(k)\cdots f_{jm}(k)]$, $\mathbf{f}(k) = \left[\mathbf{f}_1^H(k)\cdots\mathbf{f}_p^H(k)\right]^H$, $\mathbf{F}_{j,L} = [\mathbf{f}_j(L-1)\cdots\mathbf{f}_j(0)]$, $\mathbf{F}_L = [\mathbf{F}_{1,L}^H\cdots\mathbf{F}_{p,L}^H]^H$, $\mathbf{H}(z) = \sum_{k=0}^{N_1-1}\mathbf{h}(k)z^{-k}$ and

1058-6393/95 $4.00 © 1995 IEEE

$\mathbf{F}(z) = \sum_{k=0}^{L-1} \mathbf{f}(k) z^{-k}$. The condition for the equalizer to be ZF is $\mathbf{F}(z)\mathbf{H}(z) = \text{diag}\{z^{-n_1} \cdots z^{-n_p}\}$ where $n_j \in \{0, 1, \ldots, N_j+L-2\}$. The ZF condition can be written in the time-domain as

$$\mathbf{F}_L \; \mathcal{T}_{L,p}(\mathbf{H}_N) = \begin{bmatrix} 0\cdots0 \; 1 \; 0\cdots0 & \cdots & 0\cdots0 \\ \vdots & \ddots & \vdots \\ 0\cdots0 & \cdots & 0\cdots0 \; 1 \; 0\cdots0 \end{bmatrix} \tag{3}$$

where $\mathcal{T}_{M,p}(\mathbf{H}_N) = [\mathcal{T}_M(\mathbf{H}_{1,N_1}) \cdots \mathcal{T}_M(\mathbf{H}_{p,N_p})]$ and $\mathcal{T}_M(\mathbf{x})$ is a banded block Toeplitz matrix with M block rows and $[\mathbf{x} \; 0_{n\times(M-1)}]$ as first block row (n is the number of rows in $\mathbf{x}$). (3) is a system of $p(N+p(L-1))$ equations in Lmp unknowns. To be able to equalize, we need to choose the equalizer length L such that the system of equations (3) is exactly or underdetermined. Hence

$$L \geq \underline{L} = \left\lceil \frac{N-p}{m-p} \right\rceil \tag{4}$$

We assume that $\mathbf{H}_N$ has full rank if $N \geq m$. If not, it is still possible to go through the developments we consider below. But lots of singularities will appear and the non-singular part will behave in the same way as if we had a reduced number of channels, equal to the row rank of $\mathbf{H}_N$. Reduced rank in $\mathbf{H}_N$ can be detected by inspecting the rank of $\text{E}\mathbf{y}(k)\mathbf{y}^H(k)$. If a reduced rank in $\mathbf{H}_N$ is detected, the best way to proceed (also when quantities are estimated from data) is to preprocess the data $\mathbf{y}(k)$ by transforming them into new data of dimension equal to the row rank of $\mathbf{H}_N$.

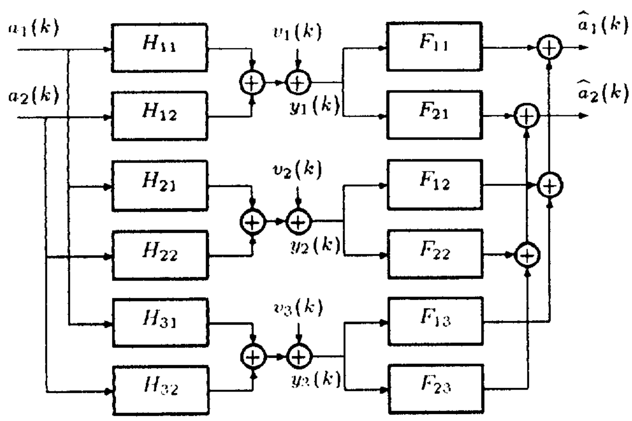

Figure 1: Channel and linear equalizer for $m = 3$ channels and $p = 2$ sources.

The matrix $\mathcal{T}_{L,p}(\mathbf{H}_N)$ is a block Toeplitz block matrix. It can be shown that for $L \geq \underline{L}$ it has full column rank if the following assumptions are satisfied

(A1) $\text{rank}(\mathbf{H}(z)) = p, \forall z$ and $\text{rank}(\mathbf{h}(0)) = p$. In this case, $\mathbf{H}(z)$ is called irreducible in systems theory,

(A2) $\text{rank}([\mathbf{h}_1(N_1-1) \cdots \mathbf{h}_p(N_p-1)]) = p$, in which case $\mathbf{H}(z)$ is called column reduced, see [4].

Assuming $\mathcal{T}_{L,p}(\mathbf{H}_N)$ to have full column rank, the nullspace of $\mathcal{T}_{L,p}^H(\mathbf{H}_N)$ has dimension $L(m-p)-N+p$. If we take the entries of any vector in this nullspace as equalizer coefficients, then the equalizer output is zero, regardless of the transmitted symbols.

To find a ZF equalizer (corresponding to some delays n_j), it suffices to take an equalizer length equal to $\underline{L}$. We can arbitrarily fix $\underline{m} = \underline{L}(m-p)-N+p$ equalizer coefficients (e.g. take $\underline{m}$ equalizer filters of length $\underline{L}-1$ only). The remaining $p(p(\underline{L}-1)+N)$ coefficients can be found from (3). This shows that for $m > p$, a FIR equalizer suffices for ZF equalization (and interference cancellation)!

3 Channel Identification from Second-order Statistics: Frequency Domain Approach

Consider the noise-free case and let the sources be temporally white but possibly correlated among themselves with $p \times p$ covariance matrix $R_{\mathbf{a}}$. Then the power spectral density matrix of the stationary vector process $\mathbf{y}(k) = \mathbf{H}(z)\mathbf{a}(k)$ is

$$S_{\mathbf{yy}}(z) = \mathbf{H}(z) R_{\mathbf{a}} \mathbf{H}^H(z^{-*}) \,. \tag{5}$$

The following spectral factorization result has been brought to our attention by Loubaton [5]. Let $\mathbf{K}(z)$ be a $m \times p$ rational transfer function that is causal and stable. Then $\mathbf{K}(z)$ is called minimum-phase if $\mathbf{K}(z) \neq 0$, $|z| > 1$. Let $S_{\mathbf{yy}}(z)$ be a rational $m \times m$ spectral density matrix of rank p. Then there exists a rational $m \times p$ transfer matrix $\mathbf{K}(z)$ that is causal, stable , minimum-phase, unique up to a unitary $p \times p$ constant matrix, of (minimal) McMillan degree $\deg(\mathbf{K}) = \frac{1}{2}\deg(S_{\mathbf{yy}})$ such that

$$S_{\mathbf{yy}}(z) = \mathbf{K}(z) \mathbf{K}^H(z^{-*}) \,. \tag{6}$$

In our case, $S_{\mathbf{yy}}$ is polynomial (FIR channel) and $\mathbf{H}(z)$ is minimum-phase since we assume $\text{rank}(\mathbf{H}(z)) = p, \forall z$. Hence, the spectral factor $\mathbf{K}(z)$ identifies the channel

$$\mathbf{K}(z) = \mathbf{H}(z) R_{\mathbf{a}}^{1/2} \Phi \tag{7}$$

where $R_{\mathbf{a}}^{1/2}$ is any particular (e.g. triangular) matrix square-root of $R_{\mathbf{a}}$ and Φ is a $p \times p$ unitary matrix. So the channel identification from second-order statistics is simply a multivariate MA spectral factorization problem. The remaining factors $R_{\mathbf{a}}^{1/2}$ and Φ can be identified by exploiting higher-order moments (see [6] and references therein) or the discrete distribution nature of the sources [7].

4 Gram-Schmidt Orthogonalization, Triangular Factorization and Linear Prediction

UDL Factorization of the Inverse Covariance Matrix

Consider a vector of zero mean random variables $Y = [y_1^H\ y_2^H \cdots y_M^H]^H$. We shall introduce the notation $y_{1:M} = Y$. Consider Gram-Schmidt orthogonalization of the components of Y. We can determine the linear least-squares (lls) estimate $\widehat{y}_i$ of y_i given $y_{1:i-1}$ and the associated estimation error $\widetilde{y}_i$ as

$$\begin{aligned} \widehat{y}_i &= \widehat{y}_i|_{y_{1:i-1}} = R_{y_i y_{1:i-1}} R^{-1}_{y_{1:i-1} y_{1:i-1}} y_{1:i-1} , \\ \widetilde{y}_i &= \widetilde{y}_i|_{y_{1:i-1}} = y_i - \widehat{y}_i \end{aligned} \tag{8}$$

where $R_{ab} = \mathrm{E}ab^H$ for two random column vectors a and b. The Gram-Schmidt orthogonalization process consists of generating consecutively $\widetilde{Y} = [\widetilde{y}_1^H\ \widetilde{y}_2^H \cdots \widetilde{y}_M^H]^H$ starting with $\widetilde{y}_1 = y_1$. We can write the relation

$$L\,Y = \widetilde{Y} \tag{9}$$

where L is a unit-diagonal lower triangular matrix. The first $i-1$ elements in row i of L are $-R_{y_i y_{1:i-1}} R^{-1}_{y_{1:i-1} y_{1:i-1}}$. From (9), we obtain

$$\mathrm{E}(LY)(LY)^H = \mathrm{E}\widetilde{Y}\widetilde{Y}^H \Rightarrow L R_{YY} L^H = D = R_{\widetilde{Y}\widetilde{Y}} . \tag{10}$$

D is indeed a diagonal matrix since the $\widetilde{y}_i$ are decorrelated. Equation (10) can be rewritten as the UDL triangular factorization of R^{-1}_{YY}

$$R^{-1}_{YY} = L^H D^{-1} L . \tag{11}$$

If Y is filled up with consecutive samples of a random process, $Y = [y^H(k)\ y^H(k-1) \cdots y^H(k-M+1)]^H$, then the $\widetilde{y}_i$ become backward prediction errors of order $i-1$, the corresponding rows in L are backward prediction filters and the corresponding diagonal elements in D are backward prediction error variances. If the process is stationary, then R_{YY} is Toeplitz and the backward prediction errors filters and variances (and hence the UDL factorization of R^{-1}_{YY}) can be determined using a fast algorithm, the Levinson algorithm. If Y is filled up in a different order, *i.e.* $Y = [y^H(k)\, y^H(k+1) \cdots y^H(k+M-1)]^H$, then the backward prediction quantities become forward prediction quantities, which for the the prediction error filters and variances are the same as the backward quantities if the process $y(.)$ is scalar valued.

If the process $\mathbf{y}(.)$ is vector valued, we shall still carry out the Gram-Schmidt orthogonalization scalar component by scalar component. In the time-series case, this is multichannel linear prediction with sequential processing of the channels. If the matrix $R_{\mathbf{YY}}$ is singular, then there exist linear relationships between certain components of $\mathbf{Y}$. As a result, certain components y_i will be perfectly predictible from the previous components and their resulting orthogonalized version $\widetilde{y}_i$ will be zero. The corresponding diagonal entry in D will hence be zero also. For the orthogonalization of the following components, we don't need this y_i. As a result, the entries under the diagonal in the coresponding column of L can be taken to be zero (minimum-norm choice for the prediction filters in those rows). The (linearly independent) row vectors in L that correspond to zeros in D are vectors that span the null space of $R_{\mathbf{YY}}$. The number of non-zero elements in D equals the rank of $R_{\mathbf{YY}}$.

LDU Factorization of a Covariance Matrix

Assume at first that $R_{\mathbf{YY}}$ is nonsingular. Since the $\widetilde{y}_i$ form just an orthogonal basis in the space spanned by the y_i, $\mathbf{Y}$ can be perfectly estimated from $\widetilde{\mathbf{Y}}$. Expressing that the covariance matrix of the error in estimating $\mathbf{Y}$ from $\widetilde{\mathbf{Y}}$ is zero leads to

$$\begin{aligned} 0 &= R_{\mathbf{YY}} - R_{\mathbf{Y}\widetilde{\mathbf{Y}}} R^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}} R_{\widetilde{\mathbf{Y}}\mathbf{Y}} \Rightarrow \\ R_{\mathbf{YY}} &= R_{\mathbf{Y}\widetilde{\mathbf{Y}}} R^{-1}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}} R_{\widetilde{\mathbf{Y}}\mathbf{Y}} = U^H D^{-1} U \end{aligned} \tag{12}$$

where D is the same diagonal matrix as in (10) and $U = L^{-H}$ is a unit-diagonal upper triangular matrix. (12) is the LDU triangular factorization of $R_{\mathbf{YY}}$. In the stationary multichannel time-series case, $R_{\mathbf{YY}}$ is block Toeplitz and the rows of U and the diagonal elements of D can be computed in a fast way using a sequential processing version of the multichannel Schur algorithm.

When $R_{\mathbf{YY}}$ is singular, then D will contain a number of zeros, equal to the dimension of the nullspace of $R_{\mathbf{YY}}$. Let J be a selection matrix (the rows of J are rows of the identity matrix) that selects the nonzero elements of D so that JDJ^H is a diagonal matrix that contains the consecutive non-zero diagonal elements of D. Then we can write

$$R_{\mathbf{YY}} = (JU)^H (JD^{-1}J^H)(JU) \tag{13}$$

which is a modified LDU triangular factorization of the singular $R_{\mathbf{YY}}$. $(JU)^H$ is a modified lower triangular matrix, its columns being a subset of the columns of the lower triangular matrix U^H. A modified version of the Schur algorithm to compute the generalized LDU factorization of a singular block Toeplitz matrix $R_{\mathbf{YY}}$ has been recently proposed in [8].

5 Signal and Noise Subspaces

Consider now the measured data with additive independent white noise $\mathbf{v}(k)$ with zero mean and assume $\mathrm{E}\mathbf{v}(k)\mathbf{v}^H(k) = \sigma_v^2 I_m$ with unknown variance σ_v^2 (in the complex case, real and imaginary parts are assumed to be uncorrelated, colored noise with known correlation structure but unknown variance could equally well be handled). A vector of L measured data can be expressed as

$$\mathbf{Y}_L(k) = \mathcal{T}_{L,p}(\mathbf{H}_N)\, A_{N+p(L-1)}(k+L-1) + \mathbf{V}_L(k) \tag{14}$$

where $\mathbf{Y}_L(k) = \left[\mathbf{y}^H(k)\cdots\mathbf{y}^H(k+L-1)\right]^H$ and $\mathbf{V}_L(k)$ is defined similarly. Therefore, the structure of the covariance matrix of the received signal $\mathbf{y}(k)$ is

$$\mathrm{R}_L^{\mathbf{y}} = \mathcal{T}_{L,p}(\mathbf{H}_N)\,\mathrm{R}^a_{N+p(L-1)}\mathcal{T}^H_{L,p}(\mathbf{H}_N) + \sigma_v^2 I_{mL} \quad (15)$$

where $\mathrm{R}^a_{N+p(L-1)} = \mathrm{E}A_{N+p(L-1)}(k)A^H_{N+p(L-1)}(k)$. We assume R^a_M to be nonsingular for any M. For $L \geq \underline{L}$, and assuming (A1), (A2), $\mathcal{T}_{L,p}(\mathbf{H}_N)$ has full column rank and σ_v^2 can be identified as the smallest eigenvalue of $\mathrm{R}_L^{\mathbf{y}}$. Replacing $\mathrm{R}_L^{\mathbf{y}}$ by $\mathrm{R}_L^{\mathbf{y}} - \sigma_v^2 I_{mL}$ gives us the covariance matrix for noise-free data. Given the structure of $\mathrm{R}_L^{\mathbf{y}}$ in (15), the column space of $\mathcal{T}_{L,p}(\mathbf{H}_N)$ is called the signal subspace and its orthogonal complement the noise subspace.

Consider the eigendecomposition of $\mathrm{R}_L^{\mathbf{y}}$ of which the real positive eigenvalues are ordered in descending order:

$$\begin{aligned}\mathrm{R}_L^{\mathbf{y}} &= \sum_{i=1}^{N+p(L-1)} \lambda_i V_i V_i^H + \sum_{i=N+p(L-1)+1}^{mL} \lambda_i V_i V_i^H \\ &= V_{\mathcal{S}}\Lambda_{\mathcal{S}}V_{\mathcal{S}}^H + V_{\mathcal{N}}\Lambda_{\mathcal{N}}V_{\mathcal{N}}^H\end{aligned} \quad (16)$$

where $\Lambda_{\mathcal{N}} = \sigma_v^2 I_{(m-p)L-N+p}$ (see (15)). The sets of eigenvectors $V_{\mathcal{S}}$ and $V_{\mathcal{N}}$ are orthogonal: $V_{\mathcal{S}}^H V_{\mathcal{N}} = 0$, and $\lambda_i > \sigma_v^2$, $i = 1,\ldots,N+p(L-1)$. We then have the following equivalent descriptions of the signal and noise subspaces

$$Range\{V_{\mathcal{S}}\} = Range\{\mathcal{T}_{L,p}(\mathbf{H}_N)\}\,,\; V_{\mathcal{N}}^H\mathcal{T}_{L,p}(\mathbf{H}_N) = 0. \quad (17)$$

6 The Instantaneous Mixture Case

We shall consider the noiseless case and we can assume w.l.o.g. that the first p rows of $\mathbf{h}(0)$ are linearly independent (the ordering of the channels can always be permuted to achieve this). The covariance matrix of $\mathbf{y}(k) = \mathbf{h}(0)\mathbf{a}(k)$ is $\mathrm{R}_1^{\mathbf{y}} = \mathbf{h}(0)R_{\mathbf{a}}\mathbf{h}^H(0)$. By carrying out the Gram-Schmidt orthogonalization of the components of $\mathbf{y}(k)$, we obtain the triangular factorizations we discussed above. In particular

$$\begin{aligned}&L\mathrm{R}_1^{\mathbf{y}}L^H = D = blockdiag\left\{D_p, 0_{(m-p)\times(m-p)}\right\}\\ &\Rightarrow \mathrm{R}_1^{\mathbf{y}} = U_p^H D_p^{-1} U_p\end{aligned} \quad (18)$$

where U_p^H is a $m \times p$ matrix of the generalized lower triangular form we discussed above. Taking $R_{\mathbf{a}}^{1/2}$ to be triangular, we arrive at

$$\mathbf{h}(0) = U_p^H D_p^{-1/2}\Phi R_{\mathbf{a}}^{-1/2} \quad (19)$$

where Φ is a $p \times p$ unitary matrix. Φ and $R_{\mathbf{a}}^{1/2}$ represent $\frac{1}{2}p(p-1)$ and $\frac{1}{2}p(p+1)$ degrees of freedom respectively. If we don't know $R_{\mathbf{a}}$, we can determine $\mathbf{h}(0)$, using the LDU factorization of $\mathrm{R}_1^{\mathbf{y}}$, as $U_p^H D_p^{-1/2}$, up to p^2 degrees of freedom. If $R_{\mathbf{a}}$ is known, e.g. $R_{\mathbf{a}} = \sigma_a^2 I_p$, then $U_p^H D_p^{-1/2}$ determines $\mathbf{h}(0)$ up to only Φ, *i.e.* up to only $\frac{1}{2}p(p-1)$ degrees of freedom.

In general, if $\mathbf{h}(0)$ is determined using subspace techniques from U_p^H, then the only part of $\mathbf{h}(0)$ that can be determined uniquely from $\mathrm{R}^{\mathbf{y}}$ is $\mathbf{h}(0)T = \mathbf{h}'(0) = [I_p\ *]^H$ which is related to $\mathbf{h}(0)$ by a nonsingular $p \times p$ matrix T, representing p^2 degrees of freedom. Note also that $LU_p^H = [I_p\ 0]^H$. Hence

$$\widetilde{\mathbf{y}}(k) = L\mathbf{y}(k) = L\mathbf{h}(0)\mathbf{a}(k) = \begin{bmatrix} I_p \\ 0\end{bmatrix} D_p^{-1/2}\Phi R_{\mathbf{a}}^{-1/2}\mathbf{a}(k) \quad (20)$$

or $\widetilde{y}_{1:p}(k)$ is just a linear transformation of $\mathbf{a}(k)$.

7 Blind Equalization and Channel Identification from Second-order Statistics by Multichannel Linear Prediction

ZF Equalizer and Noise Subspace Determination

We consider again the noiseless covariance matrix or equivalently assume noisefree data: $v(t) \equiv 0$. We shall also assume the transmitted symbols to be uncorrelated, $\mathrm{R}_M^{\mathbf{a}} = R_{\mathbf{a}} \otimes I_M$, though the noise subspace parameterization we shall obtain also holds when the transmitted symbols are correlated.

Consider now the Gram-Schmidt orthogonalization of the consecutive (scalar) elements in the vector $\mathbf{Y}_L(k)$. We start building the UDL factorization of $\mathrm{R}_L^{\mathbf{y}}$ and obtain the consecutive prediction error filters and variances. No singularities are encountered until we arrive at block row $\underline{L}$ in which we treat the elements of $\mathbf{y}(k+\underline{L}-1)$. From the full column rank of $\mathcal{T}_{\underline{L},p}(\mathbf{H}_N)$, we infer that we will get $\underline{m} \in \{0,1,\ldots,m-p-1\}$ singularities. If $\underline{m} > 0$, then the following scalar components of $\mathbf{Y}$ become zero after orthogonalization: $\widetilde{y}_i(k+\underline{L}-1) = 0$, $i = m+1-\underline{m},\ldots,m$. So the corresponding elements in the diagonal factor D are also zero. We shall call the corresponding rows in the triangular factor L singular prediction filters.

For $L = \underline{L}+1$, $\mathcal{T}_{\underline{L}+1,p}(\mathbf{H}_N)$ has m more rows than $\mathcal{T}_{\underline{L},p}(\mathbf{H}_N)$ but only p more columns. Hence the (column) rank increases by p. As a result $\widetilde{y}_i(k+\underline{L})$, $i = 1,\ldots,p$ are not zero in general while $\widetilde{y}_i(k+\underline{L}) = 0$, $i = p+1,\ldots,m$. Furthermore, since $\mathcal{T}_{\underline{L},p}(\mathbf{H}_N)$ has full column rank, the orthogonalization of $y_{1:p}(k+\underline{L})$ w.r.t. $\mathbf{Y}_{\underline{L},p}(k)$ is the same as the orthogonalization of $y_{1:p}(k+\underline{L})$ w.r.t. $A_{N+p(\underline{L}-1)}(k+\underline{L}-1)$. Hence, since the $\mathbf{a}(k)$ are assumed to be uncorrelated, only the components of $y_{1:p}(k+\underline{L})$ along $\mathbf{a}(k+\underline{L})$ remain: $\widetilde{\mathbf{y}}(k+\underline{L})|_{\mathbf{Y}_{\underline{L},p}(k)} = \mathbf{h}(0)\mathbf{a}(k+\underline{L})$ and for the rest of the details of the orthogonalization of the components of $\mathbf{y}(k+\underline{L})$, we can refer to section 6. In particular, $\widetilde{y}_{1:p}(k+\underline{L})$ are just a linear transformation of $\mathbf{a}(k+\underline{L})$.

This means that the corresponding (p outputs) prediction filter is (proportional to) a ZF equalizer! Since the prediction error is white, a further increase in the length of the prediction span will not improve the prediction. Hence $\widetilde{\mathbf{y}}(k+L) = \mathbf{h}(0)\mathbf{a}(k+L)$, $L \geq \underline{L}$ and the (block of m) prediction filters in the corresponding block row $L+1$ will be appropriately shifted versions of the (block) prediction filter in (block) row $\underline{L}+1$. In particular also for the prediction errors that are zero, a further increase of the length of the prediction span cannot possibly improve the prediction. Hence $\widetilde{y}_i(k+L) = 0$, $i = p+1, \ldots, m$, $L \geq \underline{L}$. The singular prediction filters further down in the triangular factor L are appropriately shifted versions of the first $m-p$ singular prediction filters. Furthermore, the entries in these first $m-p$ singular prediction filters that appear under the 1's ("diagonal" elements) are zero for reasons we explained before in the general orthogonalization context. So we get a (rank p) white prediction error with a finite prediction order. Hence the channel ouput process $\mathbf{y}(k)$ is *autoregressive*. Due to the structure of the remaining rows in L being shifted versions of the first ZF equalizer and the first $m-p$ singular prediction filters, after a finite "transient", L becomes a banded lower triangular block Toeplitz matrix.

Consider now $L > \underline{L}$ and let us collect all consecutive singular prediction filters in the triangular factor L into a $((m-p)(L-\underline{L})+\underline{m}) \times (mL)$ matrix $\mathcal{G}_L$. The row space of $\mathcal{G}_L$ is the (transpose of) the noise subspace. Indeed, every singular prediction filter belongs to the noise subspace since $\mathcal{G}_L \mathcal{T}_{L,p}(\mathbf{H}_N) = 0$, all rows in $\mathcal{G}_L$ are linearly independent since they are a subset of the rows of a unit-diagonal triangular matrix, and the number of rows in $\mathcal{G}_L$ equals the noise subspace dimension. $\mathcal{G}_L$ is a banded block Toeplitz matrix of which the first $m-p-\underline{m}$ rows have been omitted. $\mathcal{G}_L$ is in fact parameterized by the first $m-p$ singular prediction filters. Let us collect the nontrivial entries in these $m-1$ singular prediction filters into a column vector G_N. So we can write $\mathcal{G}_L(G_N)$. The length of G_N can be calculated to be $Nm - p^2$ which equals the number of degrees of freedom in $\mathbf{H}_N$ for identification with a subspace technique (in which case we can only identify $\mathbf{h}(k)T = \mathbf{h}'(k)$ where T is such that $\mathbf{h}'(0) = [I_p \;\; *]^H$). So $\mathcal{G}_L(G_N)$ represents a minimal linear parameterization of the noise subspace.

Channel Identification

From the discussion above, it is now not difficult to see that in the LDU factorization of $\mathrm{R}^{\mathbf{y}}$, the lower triangular factor $(JU)^H$ is banded and becomes block Toeplitz after a finite transient. Indeed, for $L \geq \underline{L}$, the $L+1^{\text{st}}$ block column of $(JU)^H$ is $\mathrm{E}\,\mathbf{y}(k:\infty)\widetilde{y}^H_{1:p}(k+L) =$

$$\left[0^H_{mL\times p}\;\mathbf{h}^H(0)\cdots\mathbf{h}^H(N_1-1)\;0\cdots\right]^H \;\mathrm{E}\mathbf{a}(k+L)\widetilde{y}^H_{1:p}(k+L) \tag{21}$$

which hence contains the channel impulse response, apart from a multiplicative factor.

Channel Estimation from Data using Deterministic ML

See [3] for channel estimation from an estimated covariance sequence by subspace fitting for $p = 1$. That approach can straightforwardly be extended to the case of general p. The details for deterministic maximum likelihood have been worked out in [9] for $p = 1$. Basically, we use $P^{\perp}_{\mathcal{T}_{M,p}(\mathbf{H}_N)} = P_{\mathcal{G}^H_M(G_N)}$. The essential number of degrees of freedom in $\mathbf{H}_N$ and G_N is $mN-p^2$ for both. So $\mathbf{H}_N$ can be uniquely determined from G_N and vice versa. Due to the (almost) block Toeplitz character of $\mathcal{G}_M$, the product $\mathcal{G}_M\mathbf{Y}_M(k)$ represents a convolution. Due to the commutativity of convolution, we can write $\mathcal{G}_M(G_N)\mathbf{Y}_M(k) = \mathcal{Y}_N(\mathbf{Y}_M(k))[1 \;\; G_N^H]^H$ for some properly structured $\mathcal{Y}_N(\mathbf{Y}_M(k))$. This leads us to formulate the DML problem as

$$\min_{G_N} \begin{bmatrix} 1 \\ G_N \end{bmatrix}^H \mathcal{Y}_N^H(\mathbf{Y}_M(k)) \left(\mathcal{G}_M^H(G_N)\,\mathcal{G}_M(G_N)\right)^{-1} \mathcal{Y}_N(\mathbf{Y}_M(k)) \begin{bmatrix} 1 \\ G_N \end{bmatrix} \tag{22}$$

which can be solved iteratively in the IQML fashion.

References

[1] L. Tong, G. Xu, and T. Kailath. "A New Approach to Blind Identification and Equalization of Multipath Channels". In *Proc. of the 25th Asilomar Conference on Signals, Systems & Computers*, pages 856–860, Pacific Grove, CA, Nov. 1991.

[2] D.T.M. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction". In *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[3] D.T.M. Slock and C.B. Papadias. "Blind Fractionally-Spaced Equalization Based on Cyclostationarity". In *Proc. Vehicular Technology Conf.*, Stockholm, Sweden, June 1994.

[4] T. Kailath. *Linear Systems*. Prentice-Hall, NJ, 1980.

[5] Ph. Loubaton. "Egalisation autodidacte multicapteurs et systèmes multivariables". GDR 134 (Signal Processing) working document, february 1994, France.

[6] B. Laheld and J.-F. Cardoso. "Adaptive Source Separation with Uniform Performance". In *Proc. EUSIPCO-94, SIGNAL PROCESSING VII: Theories and Applications*, pages 183–186, Edinburgh, Scotland, Sept. 13-16 1994.

[7] S. Talwar, M. Viberg, and A. Paulraj. "Blind Estimation of Multiple Co-Channel Digital Signals Using an Antenna Array". *IEEE Signal Processing Letters*, SPL-1(2):29–31, Feb. 1994.

[8] P. Van Dooren. "Regularization of Low-Rank Toeplitz and Quasi-Toeplitz Matrices". In *Proc. 3rd International Workshop on SVD and Signal Processing*, Leuven, Belgium, Aug. 22-25 1994.

[9] D.T.M. Slock. "Subspace Techniques in Blind Mobile Radio Channel Identification and Equalization using Fractional Spacing and/or Multiple Antennas". In *Proc. 3rd International Workshop on SVD and Signal Processing*, Leuven, Belgium, Aug. 22-25 1994.