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

Research Report N° 93-002

Blind Fractionally-Spaced Equalization and Channel Identification from Second-Order Statistics or Data

Dirk T.M. Slock

August 17, 1993

Telephone:	+33 93 00 26 26	E-mail:
Dirk T.M. Slock:	+33 93 00 26 06	slock@eurecom.fr
Fax:	+33 93 00 26 27	

Abstract

Tong, Xu and Kailath have proposed a method for blindly identifying a linear channel from the second-order statistics of the cyclostationary oversampled received signal. However, their approach leads to a significant overparameterization of the channel. We propose to work with the polyphase description of the received signal. This leads to a multichannel formulation and a minimal parameterization of the channel for a given oversampling ratio. As a result of the oversampling, the covariance matrix of the noise-free stationary multichannel received signal for a FIR channel becomes singular when its dimension gets large enough. This implies that the received signal becomes perfectly predictible, reminiscent of the case of a signal consisting of a sum of sinusoids. As a result, the channel can be identified from the received signal second-order statistics by linear prediction in the noise-free case, and by using the Pisarenko, Music or other subspace fitting methods when there is additive noise. In the latter case, the familiar signal and noise subspaces emerge. The channel identification is not affected by the transmitted symbol sequence being correlated or not.

We show that for a FIR channel, there exist zero-forcing equalizers (ZFE) which are FIR also. Indeed, one can interpret channel and ZFE as the analysis and synthesis bank of a perfect-reconstruction filter bank. The prediction and equalization considerations lead to a convenient parameterization of the noise subspace, which parallels the one of the signal subspace. These dual parameterizations help to conveniently solve the deterministic Maximum Likelihood problem, the iterative solution of which can be initialized with an estimate obtained from a subspace fitting method based on data.

We elaborate in detail on the case of an oversampling factor equal to 2, and briefly discuss the implications of higher or lower oversampling ratios.

Contents

1. Previous Work

Consider linear digital modulation over a linear ISI channel with additive Gaussian noise so that the received signal can be written as

$$y(t) \;=\; \sum_k a_k h(t-kT) \;+\; v(t) \tag{1}$$

where the a_k are the transmitted symbols, T is the symbol interval, $h(t)$ is the combined impulse response of channel and transmitter and receiver filters, but is often called the channel response for simplicity. Assuming the $\{a_k\}$ and $\{v(t)\}$ to be (wide-sense) stationary, the process $\{y(t)\}$ is (wide-sense) cyclostationary with period T [1]. Gardner [1] has shown how the channel can be identified from the second-order statistics of the cyclostationary $\{y(t)\}$.

In order to detect the transmitted symbols, one samples the received signal. If $\{y(t)\}$ is sampled with period T, the sampled process is (wide-sense) stationary and its second-order statistics contain no information about the phase of the channel. Tong, Xu and Kailath [2],[3],[4] have proposed to oversample the received signal with a period $\Delta = T/m$, $m > 1$. In what follows, we assume $h(t)$ to have a finite duration. Tong *et al.* have shown that the channel can be identified from the second-order statistics of the oversampled received signal. They introduce an observation vector $\mathbf{y}(k)$ of received samples over a certain time window and consider a matrix linear model of the form

$$\mathbf{y}(k) \;=\; \mathbf{H}\,\mathbf{a}(k) + \mathbf{v}(k)\;. \tag{2}$$

The drawback of their approach is that they need the sampled channel matrix $\mathbf{H}$ to have full column rank. This leads to an unnecessary overparameterization of the channel as will become clear below (the matrix $\mathbf{H}$ could be parameterized in terms of the samples of the channel response, but this parameterization is not exploited by Tong *et al.*). Tong *et al.* found that the condition for identifiability of the (oversampled) channel from the second-order statistics of the received signal is that the z-transform of the oversampled channel should not have m equispaced zeros on a circle centered in the origin. One should also remark that the identification of the channel from the received signal second-order statistics can only be done up to a multiplicative constant (with magnitude one in certain cases), a not unusual phenomenon in blind equalization. This constant can be identified by other means. If the channel contains a delay, then this delay can also not be identified blindly.

2. Fractionally-Spaced Channels and Equalizers, and Filter Banks

We consider here an oversampling factor $m = 2$. We assume the channel to be FIR with duration NT. We consider the polyphase description of the received signal. With $m = 2$, let $y_1(k)$ and $y_2(k)$ denote the even and odd samples of $y(t)$ ($y_1(k) = y(t_0{+}kT)$, $y_2(k) = y(t_0{+}(k{-}\frac{1}{2})T)$), and similarly for the noise samples and channel response. Then the oversampled received signal can be represented in vector form at the symbol rate as

$$\mathbf{y}(k) \;=\; \sum_{i=0}^{N-1} \mathbf{h}(i) a_{k-i} + \mathbf{v}(k) \;=\; \mathbf{H}_N A_N(k) + \mathbf{v}(k)\;,\;\; \mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix},\;\; \mathbf{v}(k) = \begin{bmatrix} v_1(k) \\ v_2(k) \end{bmatrix},$$
$$\mathbf{h}(k) = \begin{bmatrix} h_1(k) \\ h_2(k) \end{bmatrix},\;\; \mathbf{H}_N = [\mathbf{h}(0) \cdots \mathbf{h}(N{-}1)]\,,\;\; A_N(k) = \left[a_k^H \cdots a_{k-N+1}^H\right]^H\;, \tag{3}$$

where superscript ^{H} denotes Hermitian transpose. We formalize the finite duration NT (approximately) assumption of the channel as follows

(AFIR): $\mathbf{h}(0) \neq 0$, $\mathbf{h}(N-1) \neq 0$, $\mathbf{h}(i) = 0$ for $i \geq N$, and any delay in the channel should be absorbed in a relabeling of the transmitted symbols.

The z-transform of the channel response at the sampling rate is $H(z) = H_1(z^2)+z^{-1}H_2(z^2)$. Similarly, consider a fractionally-spaced $(\frac{T}{2})$ equalizer of which the z-transform can also be decomposed into its polyphase components: $F(z) = F_1(z^2) + z^{-1}F_2(z^2)$, see Figure 1. As will become clear below, a unique ZF and properly scaled equalizer can be found (under certain conditions on the channel) when F_1 and F_2 are FIR filters of length $N-1$. However, in light of the prediction and noise subspace parameterization considerations to be discussed further, we take F_1, F_2 to be FIR of length N: $F_i(z) = \sum_{k=0}^{N-1} f_i(k)z^{-k}$, $i = 1, 2$. The condition for the equalizer to be zero-forcing is $F_1(z)H_1(z) + F_2(z)H_2(z) = z^{-q}$, $q \in \{0, 1, \ldots, 2N-2\}$. The equalization could be up to some delay q, but we shall consider $q = 0$ in what follows. If we introduce $\mathbf{f}(k) = [f_1(k)\ f_2(k)]$, $\mathbf{F}_N = [\mathbf{f}(0) \cdots \mathbf{f}(N-1)]$, then the zero-forcing condition can be written as

$$\mathbf{F}_N \ \mathcal{T}_N(\mathbf{H}_N) \ = \ [1 \ \ 0 \cdots 0] \tag{4}$$

where $\mathcal{T}_M(\mathbf{x})$ is a (block) Toeplitz matrix with M (block) rows and $\left[\mathbf{x}\ 0_{p\times(M-1)}\right]$ as first (block) row (p is the number of rows in $\mathbf{x}$). (4) is a system of $2N-1$ equations in $2N$ unknowns. The equalizer has one degree of freedom more than necessary to be zero-forcing. Let us arbitrarily constrain $f_2(0) = 0$. Furthermore, consider equalization with removal of ISI, but up to a constant only and take $f_1(0) = 1$. Then the remaining equalizer coefficients can be found from

$$[\mathbf{f}(1) \cdots \mathbf{f}(N-1)]\, \mathcal{T}_{N-1}(\mathbf{H}_N) \ = \ -[h_1(1) \cdots h_1(N-1)\ 0 \cdots 0] \tag{5}$$

where $\mathcal{T}_{N-1}(\mathbf{H}_N)$ is now a square matrix of size $2N-2$. $\mathcal{T}_{N-1}(\mathbf{H}_N)$ is a Sylvester matrix (up to a permutation) which is known to be nonsingular if $H_1(z)$ and $H_2(z)$ have no zeros in common. This condition coincides with the identifiability condition of Tong *et al.* on $H(z)$. Let us denote the resulting equalizer coefficients as $\mathbf{F}_N^{p1}$. A set of equalizer coefficients that satisfies (4) is $\mathbf{F}_N = \mathbf{F}_N^{p1}/h_1(0)$ (assuming $h_1(0) \neq 0$). The ZF equalizer of length $N-1$ is

$$\mathbf{F}_{N-1} = [1\ 0 \cdots 0]\, \mathcal{T}_{N-1}^{-1}(\mathbf{H}_N) \ \ . \tag{6}$$

Note that there exists a set of blocking equalizer coefficients $\mathbf{F}_N^b$ for which no transmitted symbol has an influence on the equalizer output:

$$\mathbf{F}_N^b \mathcal{T}_N(\mathbf{H}_N) \ = \ 0 \tag{7}$$

(the nullspace of $\mathcal{T}_N^H(\mathbf{H}_N)$ has dimension one).

3. Channel Identification from Second-order Statistics by Multichannel Linear Prediction

In this section, we consider the noiseless case: $v(t) \equiv 0$. Similarly to $\mathbf{F}^{p1}$, we can introduce $\mathbf{F}^{p2} = [0\ 1\ * \cdots *]$ so that $\mathbf{F}^{p1}$, $\mathbf{F}^{p2}$ satisfy

$$\begin{bmatrix} \mathbf{F}_N^{p1} \\ \mathbf{F}_N^{p2} \end{bmatrix} \mathcal{T}_N(\mathbf{H}_N) \ = \ \begin{bmatrix} h_1(0) \\ h_2(0) \end{bmatrix} [1\ 0 \cdots 0] \ . \tag{8}$$

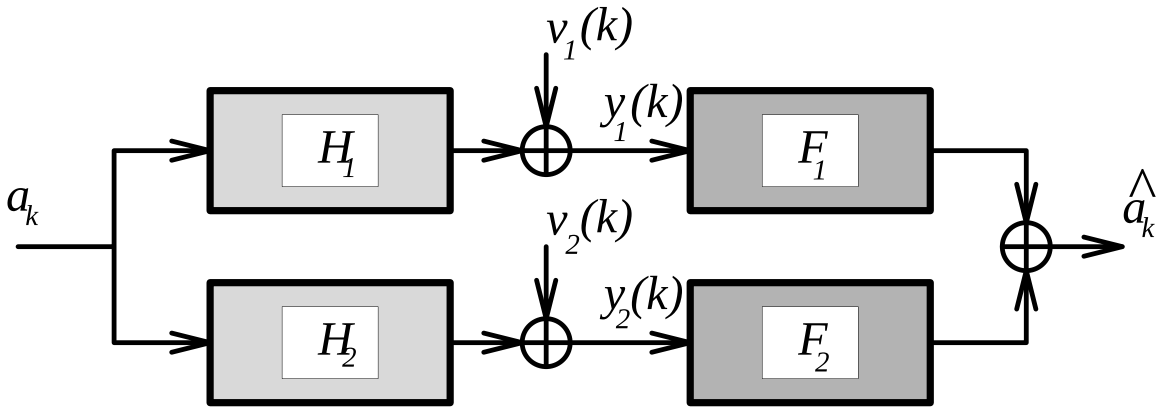

Figure 1: Polyphase representation of the T/2 fractionally-spaced channel and equalizer.

Consider now the problem of predicting $\mathbf{y}(k)$ from $\mathbf{Y}_{N-1}(k-1) = \left[\mathbf{y}^H(k-1) \cdots \mathbf{y}^H(k-N+1)\right]^H$. The prediction and prediction error can be written as

$$\begin{aligned} \widehat{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)} &= \mathbf{P}_{N-1}\mathbf{Y}_{N-1}(k-1) \, , \\ \widetilde{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)} &= \mathbf{y}(k) - \widehat{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)} = [I_2 \quad -\mathbf{P}_{N-1}]\, \mathbf{Y}_N(k). \end{aligned} \tag{9}$$

Minimizing the prediction error variance leads to the following optimization problem

$$\min_{\mathbf{P}_{N-1}} \; [I_2 \quad -\mathbf{P}_{N-1}]\, \mathrm{R}^{\mathbf{Y}}_N \, [I_2 \quad -\mathbf{P}_{N-1}]^H \tag{10}$$

where

$$\mathrm{R}^{\mathbf{Y}}_N = \mathrm{E}\; \mathbf{Y}_N(k)\mathbf{Y}^H_N(k) = \mathcal{T}_N\left(\mathbf{H}_N\right) \mathrm{R}^a_{2N-1} \mathcal{T}^H_N\left(\mathbf{H}_N\right) \;,\;\; \mathrm{R}^a_M = \mathrm{E}\; A_M(k) A^H_M(k) \, . \tag{11}$$

Hence $\mathbf{P}_{N-1}$ satisfies

$$[I_2 \quad -\mathbf{P}_{N-1}]\, \mathrm{R}^{\mathbf{Y}}_N = \left[\sigma^2_{\tilde{\mathbf{y}},N-1}\; 0 \cdots 0\right] \, . \tag{12}$$

Exploiting (11) in (12), and assuming that $H_1(z)$ and $H_2(z)$ have no common zeros, one can show that $\mathbf{P}_{N-1}$ satisfies

$$[I_2 \quad -\mathbf{P}_{N-1}]\, \mathcal{T}_N\left(\mathbf{H}_N\right) \begin{bmatrix} \mathbf{Q}_{2N-2} \\ I_{2N-2} \end{bmatrix} = 0 \, , \tag{13}$$

where $\mathbf{Q}_{2N-2}$ are the least-squares predictor coefficients for

$$\hat{a}(k)|_{A_{2N-2}(k-1)} = \mathbf{Q}_{2N-2}\, A_{2N-2}(k-1) \, , \tag{14}$$

$$\left[1 \quad -\mathbf{Q}_{2N-2}\right] \mathrm{R}^a_{2N-1} = \left[\sigma^2_{\tilde{a},2N-2}\; 0 \cdots 0\right] \, . \tag{15}$$

The result (13) can also be obtained directly, by expressing that

$$\widetilde{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)} = [I_2 \quad -\mathbf{P}_{N-1}]\, \mathcal{T}_N\left(\mathbf{H}_N\right)\, A_{2N-1}(k)$$

be orthogonal to $\mathbf{Y}_{N-1}(k-1) = \mathcal{T}_{N-1}\left(\mathbf{H}_N\right) A_{2N-2}(k-1)$. This is equivalent to expressing that $\widetilde{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)}$ be orthogonal to $A_{2N-2}(k-1)$ since $\mathcal{T}_{N-1}\left(\mathbf{H}_N\right)$ is invertible, which leads immediately to (13). Since the orthogonal complement of the rightmost matrix in (13) has dimension one, (13) leads to

$$[I_2 \quad -\mathbf{P}_{N-1}]\,\mathcal{T}_N\left(\mathbf{H}_N\right) \;=\; \mathbf{h}(0)\,\left[1 \quad -\mathbf{Q}_{2N-2}\right] \tag{16}$$

which, by postmultiplication with $A_{2N-1}(k)$, can be translated to

$$\tilde{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)} \;=\; \mathbf{h}(0)\;\;\tilde{a}(k)|_{A_{2N-2}(k-1)}\;. \tag{17}$$

Using (17), one can show that the variance of $\tilde{\mathbf{y}}(k)|_{\mathbf{Y}_{N-1}(k-1)}$ is

$$\min_{\mathbf{P}_{N-1}}\;[I_2 \quad -\mathbf{P}_{N-1}]\,\mathrm{R}^{\mathbf{Y}}_N\,[I_2 \quad -\mathbf{P}_{N-1}]^H \;=\; \sigma^2_{\tilde{a},2N-2}\,\mathbf{h}(0)\mathbf{h}^H(0) \;=\; \sigma^2_{\tilde{\mathbf{y}},N-1}\,. \tag{18}$$

By comparing (8) and (16), we see that when the transmitted data are uncorrelated ($\mathrm{R}^a_{2N-1} = \sigma^2_a I_{2N-1}$, $\mathbf{Q}_{2N-2} = 0$), we have

$$\begin{bmatrix} \mathbf{F}^{p1}_N \\ \mathbf{F}^{p2}_N \end{bmatrix} \;=\; [I_2 \quad -\mathbf{P}_{N-1}]\;. \tag{19}$$

If the transmitted symbols are correlated however, the prediction filter $\mathbf{P}_{N-1}$ is affected. Not completely however, since

$$[-h_2(0)\;\;h_1(0)]\,[I_2 \quad -\mathbf{P}_{N-1}] \;=\; [-h_2(0)\;\;h_1(0)]\begin{bmatrix} \mathbf{F}^{p1}_N \\ \mathbf{F}^{p2}_N \end{bmatrix} \tag{20}$$

which is proportional to $\mathbf{F}^b_N$ and only depends on $\mathbf{H}_N$. Note that $[-h_2(0)\;\;h_1(0)]$ can be determined up to a scalar multiple from the variance expression in (18) so that $\mathbf{F}^b_N$ can be determined from $\mathbf{P}_{N-1}$ and the variance expression, and hence from the prediction problem.

Let us introduce a block-componentwise transposition operator ^{t}, viz.

$$\begin{aligned} \mathbf{H}^t_N \;&=\; [\mathbf{h}(0)\cdots\mathbf{h}(N{-}1)]^t \;=\; \left[\mathbf{h}^T(0)\cdots\mathbf{h}^T(N{-}1)\right] \\ \mathbf{F}^t_N \;&=\; [\mathbf{f}(0)\cdots\mathbf{f}(N{-}1)]^t \;=\; \left[\mathbf{f}^T(0)\cdots\mathbf{f}^T(N{-}1)\right] \end{aligned} \tag{21}$$

where ^{T} is the usual transposition operator. Now the channel can be identified from the blocking equalizer. Indeed, (7) leads to

$$\mathbf{F}^b_N\,\mathcal{T}_N\left(\mathbf{H}_N\right) \;=\; 0 \quad\Longleftrightarrow\quad \sum_{i=1}^{2} F^b_i(z)\,H_i(z) \;=\; 0 \quad\Longleftrightarrow\quad \mathbf{H}^t_N\,\mathcal{T}_N\left(\mathbf{F}^{b\,t}_N\right) \;=\; 0\;. \tag{22}$$

This last equation allows one to determine the channel coefficients $\mathbf{H}_N$ up to a constant, if again $H_1(z)$ and $H_2(z)$ have no zeros in common (which is the same condition as $F^n_1(z)$ and $F^n_2(z)$ having no zeros in common since $F^b_1(z)/F^b_2(z) = -H_2(z)/H_1(z)$). A set of unique coefficients can be obtained by introducing one extra constraint. Traditionally, two types of constraints have been considered for this purpose:

(i) quadratic constraint: $\left\|\mathbf{H}^t_N\right\| = 1$. In this case, $\mathbf{H}^{tH}_N$ is the $2N^{\text{th}}$ left singular vector of $\mathcal{T}_N\left(\mathbf{F}^{b\,t}_N\right)$. This is a numerically very reliable way of determining $\mathbf{H}^t_N$, but is computationally somewhat costly.

(ii) linear constraint: $\mathbf{H}_N^t\, g = 1$. In this case $\mathbf{H}_N^{tH}$ is the last column vector of the Q matrix in the unnormalized QR factorization (Gram-Schmidt orthogonalization of the column vectors) of the matrix $\left[\mathcal{T}_N\left(\mathbf{F}_N^{b\,t}\right)\;\; g\right]$, divided by its length squared. In other words,

$$\mathbf{H}_N^t \;=\; \frac{1}{g^H\, P^{\perp}_{\mathcal{T}_N\left(\mathbf{F}_N^{b\,t}\right)}\, g}\; g^H\, P^{\perp}_{\mathcal{T}_N\left(\mathbf{F}_N^{b\,t}\right)}$$

where P_T is the projection matrix onto the column space of the matrix T and $P_T^{\perp} = I - P_T$. In conventional prediction theory, one takes $g = [1\;0\cdots0]^H$. However, with the channel not being necessarily minimum-phase, this may not be the most desirable choice. In particular, this choice doesn't work if $h_1(0) = 0$. So one should choose g as orthogonal as possible to the column space of $\mathcal{T}_N\left(\mathbf{F}_N^{b\,t}\right)$ in order for this procedure to be well-conditioned numerically.

If the symbol variance σ_a^2 is known, then from the prediction error variance expression in (18), we can identify $|h_1(0)|$ (or $|h_2(0)|$ if $h_1(0) = 0$). So we have identified the channel from the received signal second-order statistics, up to a factor $\frac{h_1(0)}{|h_1(0)|}$.

To recapitulate, in the absence of additive noise, we have a singular prediction problem. From the multichannel prediction error variance and the prediction coefficients, one can identify the null space of the covariance matrix, the blocking equalizer $\mathbf{F}_N^b$. From $\mathbf{F}_N^b$, one can identify the channel up to multiplicative constant as indicated above. From (16), one can identify $\mathbf{Q}_{2N-2}$ and via (15), this leads to the identification of the (Toeplitz) symbol covariance matrix R_{2N-1}^a up to the multiplicative scalar σ_a^2 (which may be known). If the transmitted symbols are uncorrelated, then the prediction problem immediately provides ZF equalizers, see (19),(8). If the transmitted symbols are correlated, then a FIR ZF equalizer can still be found directly from the FIR channel. The ZF equalizer with shortest length is given in (6).

4. Signal and Noise Subspaces

Suppose now that we have additive white noise $v(t)$ with zero mean and unknown variance σ_v^2 (in the complex case, real and imaginary parts are assumed to be uncorrelated). Then since

$$\mathrm{R}_N^{\mathbf{y}} \;=\; \mathcal{T}_N\left(\mathbf{H}_N\right)\mathrm{R}_{2N-1}^a\mathcal{T}_N^H\left(\mathbf{H}_N\right) + \sigma_v^2 I_{2N}\;, \tag{23}$$

σ_v^2 can be identified as the smallest eigenvalue of $\mathrm{R}_N^{\mathbf{y}}$, and the corresponding eigenvector is $\mathbf{F}_N^b$. This is the Pisarenko method [5, page 500]. By replacing $\mathrm{R}_N^{\mathbf{y}}$ by $\mathrm{R}_N^{\mathbf{y}} - \sigma_v^2 I_{2N}$, all results of the prediction approach in the noiseless case still hold.

If the additive noise is correlated with known correlation sequence up to an unknown variance factor σ_v^2, $\mathrm{R}_{2N}^v = \sigma_v^2\,\overline{\mathrm{R}}_{2N}^v$, then we can easily generalize this approach. We can work with $\overline{\mathrm{R}}_{2N}^{-v/2}\mathrm{R}_N^{\mathbf{y}}\overline{\mathrm{R}}_{2N}^{-v\,H/2}$ which has a form similar to the one in (23), and of which the eigenvalues are the same and the eigenvectors are a transformation by $\overline{\mathrm{R}}_{2N}^{-v\,H/2}$ away from the generalized eigenvalues and eigenvectors of

$$\mathrm{R}_N^{\mathbf{y}}\, V_i \;=\; \lambda_i\, \overline{\mathrm{R}}_{2N}^v V_i\;. \tag{24}$$

In particular, the smallest generalized eigenvalue in (24) is again σ_v^2. We shall assume $\overline{\mathrm{R}}^v = I$ in what follows.

Consider now a covariance matrix of size $M \geq N$. Given $\mathrm{R}^{\mathbf{y}}_N$, we have been able to identify all the desired quantities in the case $M = N$. So given covariance information, there cannot be anything to be gained from considering $M > N$. However, this is not necessarily the case when the covariance sequence is estimated from data. So consider the block Toeplitz matrix $\mathcal{T}_M(\mathbf{H}_N)$ of dimension $2M \times (M+N-1)$. The following lemma is easy to show.

Lemma 1 *With assumption (AFIR) and assuming that $H_1(z)$ and $H_2(z)$ are coprime,*

$$rank\ (\mathcal{T}_M(\mathbf{H}_N)) \;=\; M+N-1\ , \quad M \geq N.$$

Hence, under the assumptions of the lemma, $\mathcal{T}_M(\mathbf{H}_N)$ has full column rank. The orthogonal complement of the space spanned by the columns of $\mathcal{T}_M(\mathbf{H}_N)$ therefore has dimension $M-N+1$. With the blocking equalizer $\mathbf{F}^b_N$ satisfying $\mathbf{F}^b_N\ \mathcal{T}_N(\mathbf{H}_N) = 0$, it is easy to see that

$$\mathcal{T}_{M-N+1}\left(\mathbf{F}^b_N\right)\ \mathcal{T}_M(\mathbf{H}_N) \;=\; 0\ , \quad M \geq N \tag{25}$$

where $\mathcal{T}_{M-N+1}\left(\mathbf{F}^b_N\right)$ is a $(M-N+1) \times 2M$ block Toeplitz matrix in which the blocks are 1×2. Under the conditions of the lemma above, $\mathcal{T}_{M-N+1}\left(\mathbf{F}^b_N\right)$ has full (row) rank. Hence, the columns of $\mathcal{T}^H_{M-N+1}\left(\mathbf{F}^b_N\right)$ span the orthogonal complement of the column space of $\mathcal{T}_M(\mathbf{H}_N)$. Given the structure of

$$\mathrm{R}^{\mathbf{y}}_M \;=\; \mathcal{T}_M(\mathbf{H}_N)\,\mathrm{R}^a_{M+N-1}\mathcal{T}^H_M(\mathbf{H}_N) + \sigma^2_v I_{2M}\ , \tag{26}$$

the column spaces of $\mathcal{T}_M(\mathbf{H}_N)$ and $\mathcal{T}^H_{M-N+1}\left(\mathbf{F}^b_N\right)$ are called the signal and noise subspaces respectively.

5. Covariance Matrix Characterization

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

$$\mathrm{R}^{\mathbf{y}}_M \;=\; \sum_{i=1}^{M+N-1}\lambda_i V_i V_i^H \;+\; \sum_{i=M+N}^{2M}\lambda_i V_i V_i^H \;=\; V_{\mathcal{S}}\Lambda_{\mathcal{S}}V_{\mathcal{S}}^H \;+\; V_{\mathcal{N}}\Lambda_{\mathcal{N}}V_{\mathcal{N}}^H \tag{27}$$

where $\Lambda_{\mathcal{N}} = \sigma^2_v I_{M-N+1}$ (see (26)). Assuming $\mathcal{T}_M(\mathbf{H}_N)$ and R^a_{M+N-1} to have full rank, the sets of eigenvectors $V_{\mathcal{S}}$ and $V_{\mathcal{N}}$ are orthogonal: $V_{\mathcal{S}}^H V_{\mathcal{N}} = 0$, and $\lambda_i > \sigma^2_v\ ,\ i = 1,\ldots,M+N-1$. We then have the following equivalent descriptions of the signal and noise subspaces

$$\begin{aligned} Range\ \{V_{\mathcal{S}}\} \;&=\; Range\ \{\mathcal{T}_M(\mathbf{H}_N)\} \\ Range\ \{V_{\mathcal{N}}\} \;&=\; Range\ \left\{\mathcal{T}^H_{M-N+1}\left(\mathbf{F}^b_N\right)\right\}\ . \end{aligned} \tag{28}$$

In particular,

$$V_{\mathcal{N}}^H\mathcal{T}_M(\mathbf{H}_N) = 0\ , \quad \mathcal{T}_{M-N+1}\left(\mathbf{F}^b_N\right)\ V_{\mathcal{S}} = 0\ . \tag{29}$$

6. Channel Estimation from an Estimated Covariance Sequence by Subspace Fitting

When the covariance matrix is estimated from data, it will no longer satisfy exactly the properties we have elaborated upon. A first (detection) problem then is to determine the dimension of the signal subspace. A number of techniques for doing this have been elaborated in the literature (typically based on an investigation of the eigenvalues) and we shall assume that the correct dimension $M+N-1$ (and hence the correct channel order N) has been detected. We shall again order the eigenvalues and eigenvectors as in (27). The signal subspace will now be defined as the space spanned by the eigenvectors corresponding to the $M+N-1$ largest eigenvalues, while the noise subspace corresponds to the $M-N+1$ remaining eigenvectors (as in (27), except that $\Lambda_{\mathcal{N}}$ is no longer a multiple of the identity matrix).

Consider now the following subspace fitting problem

$$\min_{\mathbf{H}_N,T} \|\mathcal{T}_M(\mathbf{H}_N) - V_{\mathcal{S}}\,T\|_F \tag{30}$$

where the Frobenius norm of a matrix Z can be defined in terms of the trace operator: $\|Z\|_F^2 = \text{ tr }\left\{Z^H Z\right\}$. Note that this subspace fitting problem differs from the one considered in [6] (in a different context) by the fact that the roles of $\mathcal{T}_M(\mathbf{H}_N)$ and $V_{\mathcal{S}}$ are interchanged. The problem considered in (30) is quadratic in both $\mathbf{H}_N$ and T. If $V_{\mathcal{S}}$ contains the signal subspace eigenvectors of the actual covraiance matrix $\mathrm{R}_M^{\mathbf{y}}$, then the minimal value of the cost function in (30) is zero. Indeed, if the column spaces of two matrices with full column rank are identical (as in (28)), then one of the matrices can be transformed into the other one by postmultiplication with a unique nonsingular square matrix. If $\mathrm{R}_M^{\mathbf{y}}$ is estimated from a finite amount of data however, then its eigenvectors (and eigenvalues) are perturbed w.r.t. their theoretical values. Therefore, in general there will be no value for $\mathbf{H}_N$ for which the column space of $\mathcal{T}_M(\mathbf{H}_N)$ coincides with the signal subspace *Range* $\{V_{\mathcal{S}}\}$. But it is clearly meaningful to try to estimate $\mathbf{H}_N$ by taking that $\mathcal{T}_M(\mathbf{H}_N)$ into which $V_{\mathcal{S}}$ can be transformed with minimal cost. This leads to the subspace fitting problem in (30). The optimization problem in (30) is separable. With $\mathbf{H}_N$ fixed, the optimal matrix T can be found to be (assuming $V_{\mathcal{S}}^H V_{\mathcal{S}} = I$)

$$T \;=\; V_{\mathcal{S}}^H \mathcal{T}_M(\mathbf{H}_N)\ . \tag{31}$$

Modulo the result in (31), we get the following equivalences

$$\begin{aligned}
&\min_{\mathbf{H}_N,T} \|\mathcal{T}_M(\mathbf{H}_N) - V_{\mathcal{S}}\,T\|_F^2 \\
&= \min_{\mathbf{H}_N} \text{ tr }\left\{P_{V_{\mathcal{S}}}^{\perp}\mathcal{T}_M(\mathbf{H}_N)\,\mathcal{T}_M^H(\mathbf{H}_N)\right\} \;=\; \min_{\mathbf{H}_N} \text{ tr }\left\{P_{V_{\mathcal{N}}}\mathcal{T}_M(\mathbf{H}_N)\,\mathcal{T}_M^H(\mathbf{H}_N)\right\} \\
&\quad = \min_{\mathbf{H}_N} \text{ tr }\left\{\mathcal{T}_M^H(\mathbf{H}_N)\,V_{\mathcal{N}}\,V_{\mathcal{N}}^H\,\mathcal{T}_M(\mathbf{H}_N)\right\} \\
&\quad = \min_{\mathbf{H}_N} \left\|V_{\mathcal{N}}^H\,\mathcal{T}_M(\mathbf{H}_N)\right\|_F^2 \;=\; \min_{\mathbf{H}_N} \sum_{i=M+N}^{2M} \left\|V_i^H\,\mathcal{T}_M(\mathbf{H}_N)\right\|_2^2 \\
&= \min_{\mathbf{H}_N}\left[\text{ tr }\left\{\mathcal{T}_M(\mathbf{H}_N)\,\mathcal{T}_M^H(\mathbf{H}_N)\right\} - \text{ tr }\left\{P_{V_{\mathcal{S}}}\mathcal{T}_M(\mathbf{H}_N)\,\mathcal{T}_M^H(\mathbf{H}_N)\right\}\right] \\
&\quad = \min_{\mathbf{H}_N}\left[M\,\|\mathbf{H}_N\|_F^2 - \left\|V_{\mathcal{S}}^H\,\mathcal{T}_M(\mathbf{H}_N)\right\|_F^2\right]
\end{aligned} \tag{32}$$

Due to the commutativity of the convolution operation, we can again write

$$V_i^H \mathcal{T}_M(\mathbf{H}_N) \;=\; \mathbf{H}_N^t \mathcal{T}_M\left(V_i^{H\,t}\right) \tag{33}$$

where V_i^H (like $\mathbf{F}_N$) is considered a block vector with M blocks of size 1×2. Hence we can write

$$\begin{aligned}\min_{\mathbf{H}_N, T} \|\mathcal{T}_M(\mathbf{H}_N) - V_{\mathcal{S}}\, T\|_F^2 \;&=\; \min_{\mathbf{H}_N^t} \mathbf{H}_N^t \left(\sum_{i=M+N}^{2M} \mathcal{T}_M\left(V_i^{H\,t}\right) \mathcal{T}_M^H\left(V_i^{H\,t}\right)\right) \mathbf{H}_N^{t\,H} \\ =\; \min_{\mathbf{H}_N^t} &\left[M \left\|\mathbf{H}_N^t\right\|_2^2 \;-\; \mathbf{H}_N^t \left(\sum_{i=1}^{M+N-1} \mathcal{T}_M\left(V_i^{H\,t}\right) \mathcal{T}_M^H\left(V_i^{H\,t}\right)\right) \mathbf{H}_N^{t\,H}\right] .\end{aligned} \tag{34}$$

These optimization problems have to be augmented with a nontriviality constraint on $\mathbf{H}_N^t$ such as the quadratic or the linear constraints we have discussed in section 3.. In case we choose the quadratic constraint $\left\|\mathbf{H}_N^t\right\|_2 = 1$, then the last term in (34) leads equivalently to

$$\max_{\left\|\mathbf{H}_N^t\right\|_2 = 1} \mathbf{H}_N^t \left(\sum_{i=1}^{M+N-1} \mathcal{T}_M\left(V_i^{H\,t}\right) \mathcal{T}_M^H\left(V_i^{H\,t}\right)\right) \mathbf{H}_N^{t\,H} \tag{35}$$

the solution of which is the eigenvector corresponding to the maximum eigenvalue of the matrix appearing between the brackets. In the case of an oversampling factor equal to two, the noise subspace always has a lower dimension than the signal subspace. Hence it is computationally more interesting to estimate $\mathcal{T}_M(\mathbf{H}_N)$ by optimizing its orthogonality to the noise subspace, rather than by optimizing its fit to the signal subspace.

Alternatively, we may consider the following subspace fitting problem

$$\min_{\mathbf{F}_N^b, T} \left\|\mathcal{T}_{M-N+1}^H\left(\mathbf{F}_N^b\right) - V_{\mathcal{N}}\, T\right\|_F \tag{36}$$

which leads again to either a minimization problem optimizing the orthogonality to the signal subspace, or a maximization problem optimizing the fit to the noise subspace. In this case, the latter will be computationally more interesting. The channel $\mathbf{H}_N$ can then be identified from $\mathbf{F}_N^b$ as we discussed before.

When $M = N$, the subspace fitting problem in (30) leads to the Pisarenko method discussed before. When $M > N$, the Pisarenko method generalizes to the Music method [5, page502] (corresponding to (34)). When the exact covariance matrix is given, any value of $M \geq N$ will lead to the same value for $\mathbf{H}_N$, namely the true channel (up to a multiplicative scalar). When the covariance matrix is estimated from data, the estimated covariance lags can be considered as a noisy version of the true ones and hence a better estimate should be obtained as more data are incorporated, as M increases. However, as M increases, the quality of the covariance matrix estimate from a fixed finite amount of data goes down. So there should be some optimal value for M, compromising for these two opposite effects.

7. Channel and Transmitted Symbols Estimation from Data using Deterministic Maximum Likelihood

In the case of given data (samples of $y(.)$), the subspace fitting approach of the previous section involves the data through the sample covariance matrix. Though this leads to computationally tractable optimization problems, this may not lead to very efficient estimates from an estimation theoretic point of view. Therefore we consider here a deterministic or conditional maximum likelihood (DML) method. The likelihood is conditional on the transmitted symbols and the channel parameters, which are hence treated as deterministic unknowns. The stochastic part only comes from the additive noise, which we shall assume Gaussian and white with zero mean and unknown variance σ_v^2 ($\overline{\mathrm{R}}^v_{2M} = I_{2M}$, though the generalization to any known $\overline{\mathrm{R}}^v_{2M}$ is straightforward). We assume the data $\mathbf{Y}_M(k)$ to be available. The maximization of the likelihood function boils down to the following least-squares problem

$$\min_{\mathbf{H}_N, A_{M+N-1}(k)} \|\mathbf{Y}_M(k) - \mathcal{T}_M(\mathbf{H}_N) A_{M+N-1}(k)\|_2^2 \ . \tag{37}$$

This criterion consists of a sum of $2M$ terms and involves $2N$ unknowns in $\mathbf{H}_N$ and $M+N-1$ unknowns in $A_{M+N-1}(k)$. Hence we need $M \geq 3N-1$ for the criterion to make sense. The optimization problem in (37) is again separable. For fixed $\mathbf{H}_N$, the optimal transmitted symbol estimates are

$$A_{M+N-1}(k) \ = \ \left(\mathcal{T}_M^H(\mathbf{H}_N)\,\mathcal{T}_M(\mathbf{H}_N)\right)^{-1} \mathcal{T}_M^H(\mathbf{H}_N)\,\mathbf{Y}_M(k) \ . \tag{38}$$

Eliminating $A_{M+N-1}(k)$ in terms of $\mathbf{H}_N$ via (38) from (37), we get

$$\min_{\mathbf{H}_N} \left\| P^{\perp}_{\mathcal{T}_M(\mathbf{H}_N)} \ \mathbf{Y}_M(k) \right\|_2^2 \ . \tag{39}$$

Now we can use the equivalent parameterization through $\mathbf{F}_N^b$ of the orthogonal complement of *Range* $\{\mathcal{T}_M(\mathbf{H}_N)\}$ to obtain

$$\begin{aligned} &\min_{\mathbf{H}_N} \left\| P^{\perp}_{\mathcal{T}_M(\mathbf{H}_N)} \ \mathbf{Y}_M(k) \right\|_2^2 \ = \ \min_{\mathbf{F}_N^b} \left\| P_{\mathcal{T}^H_{M-N+1}\left(\mathbf{F}_N^b\right)} \ \mathbf{Y}_M(k) \right\|_2^2 \\ &= \ \min_{\mathbf{F}_N^b} \ \mathbf{Y}_M^H(k)\mathcal{T}^H_{M-N+1}\left(\mathbf{F}_N^b\right) \left[\mathcal{T}_{M-N+1}\left(\mathbf{F}_N^b\right)\mathcal{T}^H_{M-N+1}\left(\mathbf{F}_N^b\right)\right]^{-1} \mathcal{T}_{M-N+1}\left(\mathbf{F}_N^b\right) \ \mathbf{Y}_M(k) \ . \end{aligned} \tag{40}$$

Because of the commutativity of convolution, we can again rewrite

$$\mathcal{T}_{M-N+1}\left(\mathbf{F}_N^b\right) \ \mathbf{Y}_M(k) \ = \ \mathcal{H}_{M-N+1}\left(\mathbf{Y}_M^t(k)\right) \ \mathbf{F}_N^{bT} \tag{41}$$

where $\mathcal{H}_L(\mathbf{x})$ is a block Hankel matrix with L block rows, obtained by taking the block entries from the block vector $\mathbf{x}$ and filling up a Hankel matrix starting from the top left corner. (41) allows us to rewrite the criterion (40) as

$$\min_{\mathbf{F}_N^b} \ \mathbf{F}_N^{b*}\,\mathcal{H}^H_{M-N+1}\left(\mathbf{Y}_M^t(k)\right) \ \left(\mathcal{T}_{M-N+1}\left(\mathbf{F}_N^b\right)\mathcal{T}^H_{M-N+1}\left(\mathbf{F}_N^b\right)\right)^{-1} \mathcal{H}_{M-N+1}\left(\mathbf{Y}_M^t(k)\right) \ \mathbf{F}_N^{bT} \tag{42}$$

where $\mathbf{F}_N^{b*} = \mathbf{F}_N^{bTH}$. Again, this criterion has to be augmented with a nontriviality constraint. The optimization problem in (42) is nonlinear. It can easily be solved iteratively in such a way that in each iteration, a quadratic problem appears [7]. Namely, at a given iteration, plug in the $\mathbf{F}_N^b$ estimate from the previous iteration (or initialization) in the matrix being inverted in the middle of (42), and minimize the criterion w.r.t. the two outer factors $\mathbf{F}_N^b$ appearing in (42) to obtain the next estimate. With a non-convex cost function, it is important to provide the iterative solution with a good initial guess. Such an initial estimate may be obtained from the subspace fitting approach discussed above. Given the fact that this initial estimate is consistent, it can be shown (along the lines of [8]) that one iteration of the iterative solution process outlined above suffices to obtain an asymptotically best consistent (ABC) estimate!

8. Channel and Transmitted Symbols Estimation from Data using Discrete Stochastic Maximum Likelihood

The channel estimate obtained (via $\mathbf{F}_N^b$) with DML may be of acceptable quality. However, the transmitted symbol estimates (ultimately the quantities of interest) obtained from DML via (38) and proceesed further by passing them trough a detector, may not yet be sufficiently accurate. The ABC estimate mentioned above is the asymptotically best estimate if all we known about the transmission problem is captured in the DML problem formulation. However, we often know more about the problem. In the discrete stochastic ML (DSML) approach, the transmitted symbols are no longer considered to be deterministic unknowns, but a stochastic process. Hence we need a description for the probability distribution of the stochastic process. The description that the DSML method uses is very incomplete however: it only uses the fact that the symbols come from a known finite alphabet. In other words, the marginal distributions of the symbols are modeled as discrete distributions in which the discrete values are known. However, nothing is said about the respective probabilities of those discrete values in the marginal distributions, nor about how the marginal distributions interact to form the joint distribution.

The problem formulation of the DSML method leads again to (37) with now the additional constraint that the symbols a_k belong to a given alphabet. It can again be solved in an iterative fashion [9]. We will again need to exploit the commutativity of convolution, viz.

$$\begin{aligned} &\mathcal{T}_M\left(\mathbf{H}_N\right)\, A_{M+N-1}(k) \;=\; \mathcal{A}_{2M,2N}(k)\mathbf{H}_N^{tT}\,, \\ &\mathcal{A}_{2M,2N}(k) \triangleq \left[\mathcal{H}_M\left(\left[A_{M+N-1}(k)\;\; 0_{(M+N-1)\times 1}\right]^T\right)\; \mathcal{H}_M\left(\left[0_{(M+N-1)\times 1}\;\; A_{M+N-1}(k)\right]^T\right)\right]\,. \end{aligned} \tag{43}$$

Each iteration now comprises the following two steps

(i) Given an estimate of the channel $\mathbf{H}_N$, estimate the symbols using ML sequence estimation (Viterbi algorithm).

(ii) Given an estimate of the symbols $A_{M+N-1}(k)$, find the channel estimate from (37) as

$$\mathbf{H}_N^{tT} \;=\; \left(\mathcal{A}_{2M,2N}^H(k)\mathcal{A}_{2M,2N}(k)\right)^{-1} \mathcal{A}_{2M,2N}^H(k)\mathbf{Y}_M(k)\,. \tag{44}$$

Again it is important to provide the iterative scheme with a good initial guess of the channel $\mathbf{H}_N$, which may be obtained with the DML method. The quality of this initialization should be such that only very few iterations (in the limit only the first half of the first iteration) should be needed. All this leads to a blind equalization procedure in three steps:

(a) identify the channel using subspace fitting,

(b) improve upon this estimate by running a few (possibly only one) iterations of the DML algorithm,

(c) use the resulting channel estimate in a Viterbi algorithm to decode the symbols (half an iteration of the DSML algorithm).

9. Oversampling Factor OF $= m > 2$

We shall briefly discuss the case of an oversampling factor higher than two. A lot of the previous discussion goes through immediately by simply replacing 2 by m. In particular, the identifiability condition remains unchanged. With an estimated covariance matrix, the channel can still be estimated by expressing an optimal fit to the signal subspace or orthogonality to the noise subspace. One implication for ZF FIR equalizers is that their minimal order decreases as m increases. In the limit, if $m = N$, then a simple constant (filter length equal to 1) can be taken in each phase of the polyphase description. This leads to an equalizer with a total of only N coefficients.

The most interesting part is perhaps the parameterization of the orthogonal complement of *Range* $\{\mathcal{T}_M(\mathbf{H}_N)\}$ (needed for DML), which is a bit more complicated now, but again the noise-free prediction problem provides a solution. When $\mathcal{T}_M(\mathbf{H}_N)$ has reached full column rank, then as M increases to $M{+}1$, m rows get added, but only one column gets added, which makes the rank increase by one. As a result, the $m \times m$ prediction error variance $\sigma^2_{\tilde{\mathbf{y}},N-1} = \sigma_a^2 \mathbf{h}(0)\mathbf{h}^H(0)$ is of rank one. We can take for the blocking equalizer

$$\mathbf{F}^b_N \;=\; B\,[I_n \quad -\mathbf{P}_{N-1}] \tag{45}$$

where B is any full rank $(m{-}1) \times m$ matrix such that $B\,\mathbf{h}(0) = 0$ $(B\,\sigma^2_{\tilde{\mathbf{y}},N-1} = 0)$. The parameterization of the orthogonal complement of *Range* $\{\mathcal{T}_M(\mathbf{H}_N)\}$ follows again from $\mathcal{T}\left(\mathbf{F}^b_N\right)$ except that the first block row gets modified to include in general only a subspace of the rows of $\mathbf{F}^b_N$. Indeed, at the smallest value of M for which the noise-free $\mathrm{R}^{\mathbf{y}}_M$ hits singularity, the singularity of $\mathrm{R}^{\mathbf{y}}_M$ and of the corresponding $\sigma^2_{\tilde{\mathbf{y}}}$ can be anywhere from 1 to $m{-}1$.

10. Oversampling Factor OF $= \frac{m}{n} \in (1,2)$

If on the other hand we want to oversample by less than a factor of 2, then we need to oversample by a rational number $\frac{m}{n}$ (a rational number is needed in order to have a polyphase description and hence stationarity of the multichannel problem). The case $\frac{m}{n} = \frac{3}{2}$ is depicted in Figure 2. On the one hand we have the expected filter bank aspect, in which an analysis filter bank splits the signal in $m = 3$ components and a synthesis filter bank recombines the m components into one signal. On the other hand we also have a transmultiplexing aspect in which a synthesis filter bank combines $n = 2$ signals into one, which is then unraveled into n components by an analysis filter bank.

For the equalizer in Figure 2 to be ZF, we need (with sampling period $2T$)

$$\begin{bmatrix} F_{11}(z) & F_{21}(z) & F_{31}(z) \\ F_{12}(z) & F_{22}(z) & F_{32}(z) \end{bmatrix} \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \\ H_{31}(z) & H_{32}(z) \end{bmatrix} \;=\; I_2\,. \tag{46}$$

For the general case of OF $= \frac{m}{n}$ (in which m and n are of course taken to be coprime), and assuming that all mn FIR equalizer filters $F_{ij}(z)$ have L coefficients, we have mP parameters to satisfy $n\left(L{+}\frac{N}{n}{-}1\right)$ ZF constraints. Hence we need $L \geq \frac{N-n}{m-n}$. In particular, we find $L \geq 1$ for the case OF $= N$ as mentioned before. But we also find $L \geq 1$ for the case OF $= 1{+}\frac{1}{N-1}$.

The disadvantage of a rational oversampling factor is that the sampled received signal becomes cyclostationary with period nT in some sense. Indeed, for the vector of measurement

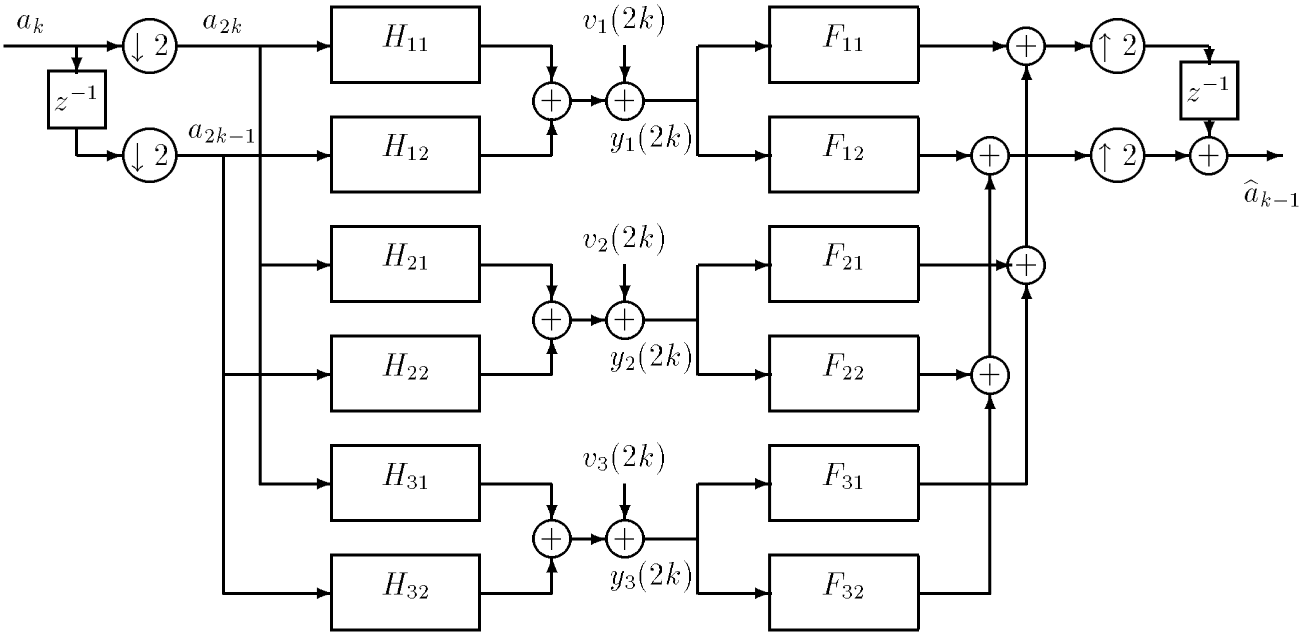

Figure 2: Polyphase representation of the 2T/3 fractionally-spaced channel and equalizer.

signals $\mathbf{y}(.)$, we get measurments with period nT. Hence, if we have a finite amount of data available, there is less room for time avaraging as n increases, leading to more noisy estimates. Time averaging is explicitly used in e.g. the subspace fitting method based on a sample covariance matrix, but these considerations also have repercussions for the accuracy of the ML methods.

11. Identifiability Issues

When OF$=m$, and the channel is no longer considered of finite length, then the identifiability condition (of the sampled channel) (from second-order statistics) still remains that the different channel filters (which are causal) in the different phases of the polyphase representation should not contain common causal factors. On the other hand, there is a result from [10] which states that the continuous-time impulse response of a band-limited channel cannot be identified from the output covariance function. Hence, if the oversampling factor is such that the Nyquist criterion is satisfied for the channel impulse response, then the sampled impulse response cannot be identified (in that case, identifying sampled or continuous-time channel amounts to the same thing).

Oversampling (by a large factor) is useful if the channel has a large bandwidth, since then a lot more information is obtained by oversampling. However, given a fixed channel bandwidth, even if the channel identifiability conditions do not get violated strictly speaking, as the oversampling factor increases, the channel estimation problem becomes ill-conditioned (e.g. the Sylvester matrix becomes ill-conditioned). This motivates the use of fractional oversampling factors when only a little extra bandwitdh (over $\frac{1}{T}$) is available. However, as the oversampling factor approaches 1, the estimation problem may be well conditioned, but the estimation accuracy degrades since less time averaging can be done. This may pose a problem if only a relatively small amount of data is available or if the channel varies rapidly. So there is a compromise to be made.

12. Concluding Remarks

The case of a rational oversampling factor $OF = \frac{m}{n}$ admits some generalizing interpretations. The factor m could be interpreted as the *diversity* factor. We obtain diversity of a factor m by oversampling with a factor m. However, diversity can be obtained in other ways. For instance, we obtain a diversity factor m in mobile radio communications if we receive the signal via m antennas [11]. However, in that case, it is for identifiability reasons not necessary to oversample the received signals (leading to combined spatial and temporal diversity), as is advocated in [11] (this is related to the requirement by Tong *et al.* for $\mathbf{H}$ to be full column rank as we discussed in our introduction).

The factor n on the other hand could be called the *multi-user* factor. It arises when the received signal from a single user gets subsampled. However, the subsampled phases a_{2k} and a_{2k-1} in Figure 2 could equivalently be interpreted as arriving from two users that are transmitting symbols at the rate $\frac{1}{2T}$. In fact, though we have obtained Figure 2 for the case of oversampling by a factor $\frac{3}{2}$ the signal received from a single user by a single antenna, Figure 2 could equivalently represent the case of no oversampling with two users and three receiver antennas.

References

[1] W.A. Gardner. "A New Method of Channel Identification". *IEEE Trans. Communications*, (6):813–817, June 1991.

[2] 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.

[3] L. Tong, G. Xu, and T. Kailath. "Blind Identification and Equalization of Multipath Channels". In *Proc. of Int'l Conf. on Communications (ICC)*, pages 1513–1517, Chicago,IL, June 14-18 1992.

[4] L. Tong, G. Xu, and T. Kailath. "Blind Identification and Equalization of Multipath Channels: A Time Domain Approach". submitted to IEEE-IT on 9/10/91.

[5] L.L. Scharf. *Statistical Signal Processing.* Addison-Wesley, Reading, MA, 1991.

[6] M. Viberg and B. Ottersten. "Sensor Array Procesing Based on Subspace Fitting". *IEEE Trans. Acoust., Speech and Sig. Proc.*, ASSP-39(5):1110–1121, May 1991.

[7] Y. Bresler and A. Macovski. "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise". *IEEE Trans. Acoust., Speech and Sig. Proc.*, ASSP-34:1081–1089, Oct. 1986.

[8] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung. "An Instrumental Variable Approach to Subspace Based System Identification". draft, May 10, 1993.

[9] M. Ghosh and C.L. Weber. "Maximum-likelihood blind equalization". *Optical Engineering*, 31:1224–1228, June 1992.

[10] Y. Chen and C.L. Nikias. "Blind Identifiability of a Band-Limited Nonminimum Phase System from its Output Autocorrelation". In *Proc. ICASSP Conf.*, volume IV, pages 444–447, Minneapolis,MN, April 1993.

[11] L. Tong, G. Xu, and T. Kailath. "Fast Blind Equalization Via Antenna Arrays". In *Proc. ICASSP*, volume IV, pages 272–275, Minneapolis, MN, April 27-30 1993.