Blind Fractionally-Spaced Equalization Based on Cyclostationarity

Dirk T.M. Slock and Constantinos B. Papadias

Eurecom Institute, 2229 route des Crêtes

B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE

Abstract— **Equalization for digital communications constitutes a very particular blind deconvolution problem in that the received signal is cyclostationary. Oversampling (OS) (w.r.t. the symbol rate) of the cyclostationary received signal leads to a stationary vector-valued signal (polyphase representation (PR)). OS also leads to a fractionally-spaced channel model and equalizer. In the PR, channel and equalizer can be considered as an analysis and synthesis filter bank. Zero-forcing (ZF) equalization corresponds to a perfect-reconstruction filter bank. We show that in the OS case FIR ZF equalizers exist for a FIR channel. In the PR, the noise-free multichannel power spectral density matrix has rank one and the channel can be found as the (minimum-phase) spectral factor. The multichannel linear prediction of the noiseless received signal becomes singular eventually, reminiscent of the single-channel prediction of a sum of sinusoids. As a result, a ZF equalizer can be determined from the received signal second-order statistics by linear prediction in the noise-free case, and by using a Pisarenko-style modification when there is additive noise. In the given data case, Music (subspace) or ML techniques can be applied. We also present some Cramer-Rao bounds and compare them to the case of channel identification using a training sequence.**

I. Previous Work

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

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

where the a_k are the transmitted symbols, T is the symbol period, $h(t)$ 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. If the channel would be known, then one could pass the received signal through a matched filter and sample the output at the symbol rate. These samples would provide sufficient statistics for the detection of the transmitted symbols. 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 [1] 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

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

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. The results presented here generalize the results in [2] where an oversampling factor $m = 2$ was considered.

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

We assume the channel to be FIR with duration of approximately NT. With an oversampling factor m, the sampling instants for the received signal in (1) are $t_0{+}T(k + \frac{j}{m})$ for integer k and $j = 0, 1, \ldots, m{-}1$. We introduce the polyphase description of the received signal: $y_j(k) = y(t_0{+}T(k + \frac{j}{m}))$ for $j = 0, 1, \ldots, m{-}1$ are the m phases of received signal, and similarly for the channel impulse response and the additive noise. In principle, it suffices to introduce a restricted $t_0 \in [0,T)$ to be fully general. However, we shall take $t_0 = t_0' + dT$ where $t_0' \in [0,T)$ and d is chosen as the smallest integer such that

$$\left[h(t_0'+dT)\cdots h(t_0'+(d+\frac{m-1}{m})T)\right] \neq 0 \ . \qquad (3)$$

The channel being causal implies that d will be nonnegative; d represents an inherent delay. The oversampled received signal can now be represented in vector form at the symbol rate as

$$\mathbf{y}(k) \;=\; \sum_{i=0}^{N-1}\mathbf{h}(i)a_{k-i}+\mathbf{v}(k) \;=\; \mathbf{H}_N A_N(k)+\mathbf{v}(k)\ ,$$
$$\mathbf{y}(k)=\begin{bmatrix} y_1(k)\\ \vdots \\ y_m(k)\end{bmatrix},\mathbf{v}(k)=\begin{bmatrix} v_1(k)\\ \vdots \\ v_m(k)\end{bmatrix},\mathbf{h}(k)=\begin{bmatrix} h_1(k)\\ \vdots \\ h_m(k)\end{bmatrix}$$
$$\mathbf{H}_N=[\mathbf{h}(0)\cdots\mathbf{h}(N{-}1)]\,,A_N(k)=\left[a_k^H\cdots a_{k-N+1}^H\right]^H \qquad (4)$$

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

(AFIR) : $\mathbf{h}(0)\neq 0$, $\mathbf{h}(N{-}1)\neq 0$ and $\mathbf{h}(i)=0$ for $i<0$ or $i\geq N$.

The z-transform of the channel response at the sampling rate $\frac{m}{T}$ is $H(z) = \sum_{j=1}^{m} z^{-(j-1)}H_j(z^m)$. Similarly, consider a fractionally-spaced ($\frac{T}{m}$) equalizer of which the z-transform can also be decomposed into its polyphase components: $F(z) = \sum_{j=1}^{m} z^{(j-1)}F_j(z^m)$, see Fig. 1. Although this equalizer is slightly noncausal, this does not cause a problem because the discrete-time filter is not a sampled version of an underlying continuous-time function. In fact, a particular equalizer phase $z^{(j-1)}F_j(z^m)$ follows in cascade the corresponding channel phase $z^{-(j-1)}H_j(z^m)$ so that the cascade $F_j(z^m)H_j(z^m)$ is causal. We assume the equalizer phases to be causal and FIR of length L: $F_j(z)=\sum_{k=0}^{L-1}f_j(k)z^{-k}$, $j=1,\ldots,m$.

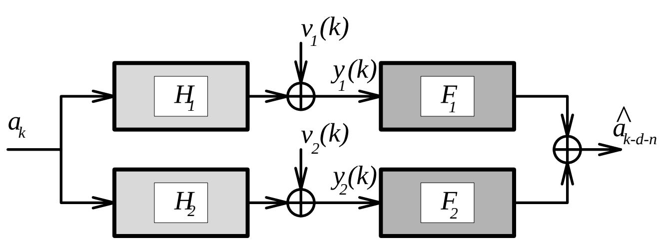

Fig. 1. Polyphase representation of the T/m fractionally-spaced channel and equalizer for $m=2$.

III. FIR Zero-Forcing (ZF) Equalization

We introduce $\mathbf{f}(k) = [f_1(k)\cdots f_m(k)]$, $\mathbf{F}_L = [\mathbf{f}(0)\cdots\mathbf{f}(L{-}1)]$, $\mathbf{H}(z) = \sum_{k=0}^{N-1}\mathbf{h}(k)z^{-k}$ and $\mathbf{F}(z) = \sum_{k=0}^{N-1}\mathbf{f}(k)z^{-k}$. The condition for the equalizer to be ZF is $\mathbf{F}(z)\mathbf{H}(z) = z^{-n}$ where $n=0,1,\ldots,N{+}L{-}2$. The ZF condition can be written in the time-domain as

$$\mathbf{F}_L\ \mathcal{T}_L(\mathbf{H}_N) \;=\; [0\cdots0\ \ 1\ \ 0\cdots0] \qquad (5)$$

where the 1 is in the $n{+}1st$ position and $\mathcal{T}_M(\mathbf{x})$ 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 $\mathbf{x}$). (5) is a system of $L{+}N{-}1$ equations in Lm unknowns. To be able to equalize, we need to choose the equalizer length L such that the system of equations (5) is exactly or underdetermined. Hence

$$L\ \geq \underline{L}\ =\ \left\lceil\frac{N-1}{m-1}\right\rceil\ . \qquad (6)$$

The matrix $\mathcal{T}_L(\mathbf{H}_N)$ is a generalized Sylvester matrix. It can be shown that for $L\geq\underline{L}$ it has full column rank if $\mathbf{H}(z)\neq 0$, $\forall z$ or in other words if the $H_j(z)$ have no zeros in common. This condition coincides with the identifiability condition of Tong *et al.* on $H(z)$ mentioned earlier. Assuming $\mathcal{T}_L(\mathbf{H}_N)$ to have full column rank, the nullspace of $\mathcal{T}_L^H(\mathbf{H}_N)$ has dimension $L(m{-}1){-}N{+}1$. 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 delay n), it suffices to take an equalizer length equal to $\underline{L}$. We can arbitrarily fix $\underline{L}(m{-}1){-}N{+}1$ equalizer coefficients (e.g. take $\underline{L}(m{-}1){-}N{+}1$ equalizer phases of length $\underline{L}{-}1$ only). The remaining $\underline{L}{+}N{-}1$ coefficients can be found from (5) if $\mathbf{H}(z)\neq 0$, $\forall z$. This shows that in the oversampled case, a FIR equalizer suffices for ZF equalization! With an oversampling factor $m=N$, the minimal required total number of equalizer coefficients N is found ($\underline{L}=1$).

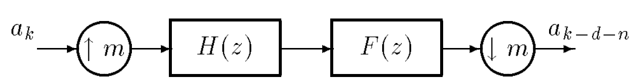

Fig. 2. Fractionally-spaced channel and equalizer.

The ZF condition (with delay $n=0$) can be interpreted in the frequency domain as follows (see Fig. 2). Consider the cascade $G(z) \;=\; \sum_k g_k z^{-k} \;=\; H(z)\,F(z)$ of fractionally-spaced channel and equalizer. Then the ZF condition $\mathbf{F}(z)\mathbf{H}(z)=1$ becomes

$$\left\{\begin{array}{l} g_{mk} \;=\; \delta_{k0} \\ \displaystyle\sum_{j=0}^{m-1} G(\frac{f+j}{m}) \;=\; m \end{array}\right. \qquad (7)$$

This is similar to the Nyquist condition in the continuous-time case. If the channel is bandlimited with bandwidth $B \in (\frac{1}{T}, \frac{m}{T})$, this poses no particular problem for the determination of a ZF equalizer (assuming infinite length). If $B < \frac{1}{T}$ however, then the $H_j(f)$, $j = 1, \ldots, m$ are zero simultaneously for some f rendering ZF equalization impossible. This is the infinite length equivalent of the condition of no zeros in common in the FIR case.

IV. Channel Identification from Second-order Statistics: Frequency Domain Approach

Consider the noise-free case and let the transmitted symbols be uncorrelated with variance σ_a^2. Then the power spectral density matrix of the stationary vector process $\mathbf{y}(k)$ is

$$S_{\mathbf{yy}}(z) = \sigma_a^2 \mathbf{H}(z)\,\mathbf{H}^H(z^{-*}) . \tag{8}$$

The following spectral factorization result has been brought to our attention by Loubaton [3]. Let $\mathbf{K}(z)$ be a $m \times 1$ rational transfer function that is causal and stable. Then $\mathbf{K}(z)$ is called minimum-phase if $\mathbf{K}(z) \neq 0$, $|z| > 1$. $S_{\mathbf{yy}}(z)$ is a rational $m \times m$ spectral density matrix of rank 1. Then there exists a rational $m \times 1$ transfer matrix $\mathbf{K}(z)$ that is causal, stable , minimum-phase, unique up to a unitary constant, of (minimal) McMillan degree $\deg(\mathbf{K}) = \frac{1}{2}\deg(S_{\mathbf{yy}})$ such that

$$S_{\mathbf{yy}}(z) = \mathbf{K}(z)\,\mathbf{K}^H(z^{-*}) . \tag{9}$$

In our case, $S_{\mathbf{yy}}$ is polynomial (FIR channel) and $\mathbf{H}(z)$ is minimum-phase since we assume $\mathbf{H}(z) \neq 0$, $\forall z$. Hence, the spectral factor $\mathbf{K}(z)$ identifies the channel

$$\mathbf{K}(z) = \sigma_a\, e^{j\phi}\, \mathbf{H}(z) \tag{10}$$

up to a constant $\sigma_a e^{j\phi}$. So the channel identification from second-order statistics is simply a multivariate MA spectral factorization problem.

V. ZF Equalizer Determination from Second-order Statistics by Multichannel Linear Prediction

We consider again the noiseless case: $v(t) \equiv 0$. The input-output relation of the channel is

$$\mathbf{Y}_L(k) = \mathcal{T}_L(\mathbf{H}_N)\, A_{L+N-1}(k) \tag{11}$$

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

$$\mathrm{R}_L^{\mathbf{y}} = \mathrm{E}\mathbf{Y}_L(k)\mathbf{Y}_L^H(k) = \mathcal{T}_L(\mathbf{H}_N)\,\mathrm{R}^a_{L+N-1}\mathcal{T}_L^H(\mathbf{H}_N) \tag{12}$$

where $\mathrm{R}_L^a = \mathrm{E}A_L(k)A_L^H(k)$. When $mL > L{+}N{-}1$, $\mathrm{R}_L^{\mathbf{y}}$ is singular. If then L increases further by 1, the rank of $\mathrm{R}_L^{\mathbf{y}}$ increases by 1 and the dimension of its nullspace increases by $m{-}1$. Consider now the problem of predicting $\mathbf{y}(k)$ from $\mathbf{Y}_L(k-1)$ The prediction error can be written as

$$\begin{aligned} \widetilde{\mathbf{y}}(k)|_{\mathbf{Y}_L(k-1)} &= \mathbf{y}(k) - \widehat{\mathbf{y}}(k)|_{\mathbf{Y}_L(k-1)} \\ &= [I_m \;\; -\mathbf{P}_L]\,\mathbf{Y}_{L+1}(k). \end{aligned} \tag{13}$$

Minimizing the prediction error variance leads to the following optimization problem

$$\min_{\mathbf{P}_L} [I_m \;\; -\mathbf{P}_L]\,\mathrm{R}^{\mathbf{y}}_{L+1}\,[I_m \;\; -\mathbf{P}_L]^H = \sigma^2_{\widetilde{\mathbf{y}},L} \tag{14}$$

or hence

$$[I_m \;\; -\mathbf{P}_L]\,\mathrm{R}^{\mathbf{y}}_{L+1} = \left[\sigma^2_{\widetilde{\mathbf{y}},L} \;\; 0\cdots 0\right] . \tag{15}$$

When $mL > L{+}N{-}1$, $\mathcal{T}_L(\mathbf{H}_N)$ has full column rank. Hence, using (11),

$$\widetilde{\mathbf{y}}(k)|_{\mathbf{Y}_L(k-1)} = \widetilde{\mathbf{y}}(k)|_{A_{L+N-1}(k-1)} . \tag{16}$$

Now, $\widetilde{\mathbf{y}}(k)|_{A_{L+N-1}(k-1)} =$

$$[I_m \;\; -\mathbf{P}_L]\,\mathcal{T}_{L+1}(\mathbf{H}_N)\; A_{L+N}(k) \;\perp\; A_{L+N-1}(k{-}1) \tag{17}$$

which leads to

$$[I_m \;\; -\mathbf{P}_L]\,\mathcal{T}_{L+1}(\mathbf{H}_N)\,\mathrm{R}^a_{L+N}\begin{bmatrix} 0\cdots 0 \\ I_{L+N-1}\end{bmatrix} = 0 . \tag{18}$$

Now let us consider the prediction problem for the transmitted symbols. We get similarly

$$\hat{a}(k)|_{A_M(k-1)} = \mathbf{Q}_M\, A_M(k{-}1) , \tag{19}$$

$$[1 \;\; -\mathbf{Q}_M]\,\mathrm{R}^a_{M+1} = \left[\sigma^2_{\widetilde{a},M} \;\; 0\cdots 0\right] . \tag{20}$$

Comparing (18) and (20), we find

$$[I_m \;\; -\mathbf{P}_L]\,\mathcal{T}_{L+1}(\mathbf{H}_N) = \mathbf{h}(0)\,\left[1 \;\; -\mathbf{Q}_{L+N-1}\right] . \tag{21}$$

which, using (14), leads to

$$\sigma^2_{\widetilde{\mathbf{y}},L} = \sigma^2_{\widetilde{a},L+N-1}\mathbf{h}(0)\mathbf{h}^H(0) . \tag{22}$$

All this holds for $L \geq \underline{L}$. We can summarize:

$$\operatorname{rank}\left(\sigma^2_{\widetilde{\mathbf{y}},L}\right) \begin{cases} = 1 & ,\; L \geq \underline{L} \\ \in \{2,3,\ldots,m\} & ,\; L = \underline{L}{-}1 \\ = m & ,\; L < \underline{L}{-}1 \end{cases} \tag{23}$$

We continue, assuming $L \geq \underline{L}$. Then (22) allows us to find $\mathbf{h}(0)$ up to a scalar multiple. If the transmitted symbols are uncorrelated, then from (21) we see that $\frac{\mathbf{h}^H(0)}{\mathbf{h}^H(0)\mathbf{h}(0)}[I_m \;\; - \mathbf{P}_L]$ is a ZF equalizer (and using (5), we could also determine the channel $\mathbf{H}_N$ up to a scalar multiple)! In this case, the prediction problem allows us also (in theory) to check whether the H_j have zeros in common. Indeed, the common factor colors the transmitted symbols (MA process) and hence once $\sigma^2_{\tilde{\mathbf{y}},L}$ becomes of rank 1, its one nonzero eigenvalue $\sigma^2_{\tilde{a},L+N-1}\mathbf{h}^H(0)\mathbf{h}(0)$ continues to decrease as a function of L since for a MA process, $\sigma^2_{\tilde{a},L}$ is a decreasing function of L.

If the transmitted symbols are correlated, we proceed as follows (Pisarenko-style [4, page 500]). Linear prediction corresponds to the LDU factorization $\mathrm{L}\mathrm{R}^{\mathbf{y}}\mathrm{L}^H = \mathrm{D}$. The prediction filters are rows of L while the prediction variances are the diagonal elements of D. Let's take l prediction filters corresponding to singularities in D and assume the longest one has block length L. So we obtain $\mathbf{F}^b_L$ of size $l \times mL$. We introduce a block-componentwise transposition operator ^{t}, viz.

$$\begin{aligned} \mathbf{H}^t_N &= [\mathbf{h}(0)\cdots\mathbf{h}(N-1)]^t = \left[\mathbf{h}^T(0)\cdots\mathbf{h}^T(N-1)\right] \\ \mathbf{F}^t_N &= [\mathbf{f}(0)\cdots\mathbf{f}(N-1)]^t = \left[\mathbf{f}^T(0)\cdots\mathbf{f}^T(N-1)\right] \end{aligned} \tag{24}$$

where ^{T} is the usual transposition operator. Due to the singularities, we have

$$\mathbf{F}^b_L\, \mathcal{T}_L(\mathbf{H}_N) = 0 \iff \mathbf{H}^t_N\, \mathcal{T}_N\left(\mathbf{F}^{b\,t}_L\right) = 0\ . \tag{25}$$

Since $\mathbf{F}^b_L \mathbf{Y}_L(k) = 0$, we call $\mathbf{F}^b_L$ a *blocking* equalizer. We find: if $l(L+N-1) \geq mN-1$, then

$$\dim\left(Range^{\perp}\left\{\mathcal{T}_N\left(\mathbf{F}^{b\,t}_L\right)\right\}\right) = 1\ . \tag{26}$$

In that case, we can identify the channel $\mathbf{H}^{t\,H}_N$ (up to scalar multiple) as the last right singular vector of $\mathcal{T}_N\left(\mathbf{F}^{b\,t}_L\right)$. In particular, let $\mathbf{h}^{\perp}$ be $m \times (m-1)$ of rank $m-1$ such that $\mathbf{h}^{\perp\,H}\mathbf{h}(0) = 0$, then with $L = \underline{L}+1$ and $l = m-1$, we can take

$$\mathbf{F}^b_{\underline{L}+1} = \mathbf{h}^{\perp\,H}\left[I_m \;\; - \mathbf{P}_{\underline{L}}\right]\ . \tag{27}$$

¿From (21), one can furthermore identify $\mathbf{Q}_{L+N-1}$ and via (20), this leads to the identification of the (Toeplitz) symbol covariance matrix R^a_{L+N} up to the multiplicative scalar σ^2_a (which may be known).

VI. Signal and Noise Subspaces

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

$$\mathrm{R}^{\mathbf{y}}_L = \mathcal{T}_L(\mathbf{H}_N)\,\mathrm{R}^a_{L+N-1}\mathcal{T}^H_L(\mathbf{H}_N) + \sigma^2_v I_{mL}\ , \tag{28}$$

For $L \geq \underline{L}$, σ^2_v can be identified as the smallest eigenvalue of $\mathrm{R}^{\mathbf{y}}_L$. Replacing $\mathrm{R}^{\mathbf{y}}_L$ by $\mathrm{R}^{\mathbf{y}}_L - \sigma^2_v I_{mL}$, all results of the prediction approach in the noiseless case still hold. Given the structure of $\mathrm{R}^{\mathbf{y}}_L$ in (28), the column space of $\mathcal{T}_L(\mathbf{H}_N)$ is called the signal subspace and its orthogonal complement the noise subspace. In [2], a linear parameterization of the noise subspace is given in terms of a blocking equalizer for $m = 2$. For $m > 2$ however, a linear minimal parameterization of the noise subspace does not exist.

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

$$\begin{aligned} \mathrm{R}^{\mathbf{y}}_L &= \sum_{i=1}^{L+N-1}\lambda_i V_i\, V_i^H + \sum_{i=(m-1)L-N+1}^{mL}\lambda_i V_i\, V_i^H \\ &= V_{\mathcal{S}}\Lambda_{\mathcal{S}}V^H_{\mathcal{S}} + V_{\mathcal{N}}\Lambda_{\mathcal{N}}V^H_{\mathcal{N}} \end{aligned} \tag{29}$$

where $\Lambda_{\mathcal{N}} = \sigma^2_v I_{(m-1)L-N+1}$ (see (28)). Assuming $\mathcal{T}_L(\mathbf{H}_N)$ and R^a_{L+N-1} to have full rank, the sets of eigenvectors $V_{\mathcal{S}}$ and $V_{\mathcal{N}}$ are orthogonal: $V^H_{\mathcal{S}}V_{\mathcal{N}} = 0$, and $\lambda_i > \sigma^2_v$, $i = 1,\ldots,L+N-1$. We then have the following equivalent descriptions of the signal and noise subspaces

$$Range\,\{V_{\mathcal{S}}\} = Range\,\{\mathcal{T}_L(\mathbf{H}_N)\}\ ,\ V^H_{\mathcal{N}}\mathcal{T}_L(\mathbf{H}_N) = 0\ . \tag{30}$$

VII. 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. We assume that the detection problem of the signal subspace dimension $L+N-1$ has been solved correctly. The signal subspace will now be defined as the space spanned by the eigenvectors corresponding to the $L+N-1$ largest eigenvalues, while the noise subspace is its orthogonal complement. Consider now the following subspace fitting problem

$$\min_{\mathbf{H}_N,T}\|\mathcal{T}_L(\mathbf{H}_N) - V_{\mathcal{S}}\,T\|_F \tag{31}$$

where the Frobenius norm of a matrix Z can be defined in terms of the trace operator: $\|Z\|^2_F = \operatorname{tr}\left\{Z^H Z\right\}$. The problem considered in (31) is quadratic in both $\mathbf{H}_N$ and T. If $V_{\mathcal{S}}$ contains the signal subspace eigenvectors of the actual covariance matrix $\mathrm{R}^{\mathbf{y}}_L$, then the minimal value of the cost function in (31) is zero. If $\mathrm{R}^{\mathbf{y}}_L$ 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 $\mathcal{T}_L(\mathbf{H}_N)$ coincides with the signal subspace $Range\{V_{\mathcal{S}}\}$. But it is clearly meaningful to try to estimate $\mathbf{H}_N$ by taking that $\mathcal{T}_L(\mathbf{H}_N)$ into which $V_{\mathcal{S}}$ can be transformed with minimal cost. This leads to the subspace fitting problem in (31). The optimization problem in (31) is separable. With $\mathbf{H}_N$ fixed, the optimal matrix T can be found to be (assuming $V_{\mathcal{S}}^H V_{\mathcal{S}} = I$)

$$T = V_{\mathcal{S}}^H \mathcal{T}_L(\mathbf{H}_N) . \tag{32}$$

Using (32) and the commutativity of the convolution operator as in (25), one can show that (31) is equivalent to

$$\begin{aligned}&\min_{\mathbf{H}_N^t} \mathbf{H}_N^t \left(\sum_{i=(m-1)L-N+1}^{mL} \mathcal{T}_L\left(V_i^{H\,t}\right) \mathcal{T}_L^H\left(V_i^{H\,t}\right) \right) \mathbf{H}_N^{t\,H} \\ &= \min_{\mathbf{H}_N^t} \left[L \left\| \mathbf{H}_N^t \right\|_2^2 - \mathbf{H}_N^t \left(\sum_{i=1}^{L+N-1} \mathcal{T}_L\left(V_i^{H\,t}\right) \mathcal{T}_L^H\left(V_i^{H\,t}\right) \right) \mathbf{H}_N^{t\,H} \right]\end{aligned} \tag{33}$$

where V_i^H (like $\mathbf{F}_L$) is considered a block vector with L blocks of size $1 \times m$. These optimization problems have to be augmented with a nontriviality constraint on $\mathbf{H}_N^t$. In case we choose the quadratic constraint $\left\| \mathbf{H}_N^t \right\|_2 = 1$, then the last term in (33) leads equivalently to

$$\max_{\left\| \mathbf{H}_N^t \right\|_2 = 1} \mathbf{H}_N^t \left(\sum_{i=1}^{L+N-1} \mathcal{T}_L\left(V_i^{H\,t}\right) \mathcal{T}_L^H\left(V_i^{H\,t}\right) \right) \mathbf{H}_N^{t\,H} \tag{34}$$

the solution of which is the eigenvector corresponding to the maximum eigenvalue of the matrix appearing between the brackets.

VIII. Channel Estimation from Data using Deterministic ML

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, the following deterministic Maximum Likelihood approach leads to more efficient estimates. The stochastic part is considered to come only from the additive noise, which we shall assume Gaussian and white with zero mean and unknown variance σ_v^2. 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

$$\min_{\mathbf{H}_N, A_{M+N-1}(k)} \left\| \mathbf{Y}_M(k) - \mathcal{T}_M(\mathbf{H}_N) A_{M+N-1}(k) \right\|_2^2 . \tag{35}$$

The optimization problem in (35) is again separable. Eliminating $A_{M+N-1}(k)$ in terms of $\mathbf{H}_N$, we get

$$\min_{\mathbf{H}_N} \left\| P_{\mathcal{T}_M(\mathbf{H}_N)}^{\perp} \mathbf{Y}_M(k) \right\|_2^2 \text{ or } \max_{\mathbf{H}_N} \left\| P_{\mathcal{T}_M(\mathbf{H}_N)} \mathbf{Y}_M(k) \right\|_2^2 . \tag{36}$$

Since $\mathcal{T}_M^H(\mathbf{H}_N) \mathbf{Y}_M(k) = \mathcal{T}_N^T\left(\mathbf{Y}_M^{t\,T}\right) \mathbf{H}_N^{t\,H}$, we can rewrite the second approach in (36) as

$$\begin{aligned}&\max_{\left\| \mathbf{H}_N^t \right\|_2 = 1} \mathbf{H}_N^t \, \mathcal{T}_N^*\left(\mathbf{Y}_M^{t\,T}(k)\right) \\ &\left(\mathcal{T}_M^H(\mathbf{H}_N) \mathcal{T}_M(\mathbf{H}_N)\right)^{-1} \mathcal{T}_N^T\left(\mathbf{Y}_M^{t\,T}(k)\right) \mathbf{H}_N^{t\,H}\end{aligned} \tag{37}$$

This optimization problem can now easily be solved iteratively in such a way that in each iteration, a quadratic problem appears [4]. An initial estimate may be obtained from the subspace fitting approach discussed above. To determine the CR bound, note that $\mathcal{T}_M(\mathbf{H}_N) A_{M+N-1}(k) = \mathcal{A}_{M,N}(k) \mathbf{H}_N^{t\,T}$ where $\mathcal{A}_{M,N}(k) = A_{M,N}(k) \otimes I_m$ and

$$A_{M,N}(k) = \begin{bmatrix} a(k) & \cdots & a(k-N+1) \\ \vdots & \cdot^{\cdot^{\cdot}} & \vdots \\ a(k-M+1) & \cdots & a(k-M-N+2) \end{bmatrix} \tag{38}$$

(Hankel matrix). This leads to a singularity in the joint information matrix for $A_{M+N-1}(k)$ and $\mathbf{H}_N^{t\,T}$, which translates into a singularity for the information matrix for $\mathbf{H}_N^{t\,T}$ separately (we can only determine $\mathbf{H}_N$ up to a scalar multiple). If we consider the estimation of the channel modulo the problem of determining the proper scale factor, then the $\frac{1}{\sigma_v^2} CRB_{\widehat{\mathbf{H}}_N^{t\,T}}$ can be shown to be

$$\left[\mathcal{A}_{M,N}^H(k) P_{\mathcal{T}_M(\mathbf{H}_N)}^{\perp} \mathcal{A}_{M,N}(k) \right]^+ \geq \left[\mathcal{A}_{M,N}^H(k) \mathcal{A}_{M,N}(k) \right]^{-1}$$

(pseudo-inverse). The last expression is the CR bound if the data $A_{M+N-1}(k)$ were known (training sequence). For small m (e.g. 2), we find that the quality of the channel estimate may be relatively bad if the channel impulse response tapers off near the ends (channel length detection problem!). For large m however, the CR bound appraches the value corresponding to known data (which is independent of the channel)!

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] Ph. Loubaton. "Egalisation autodidacte multi-capteurs et systèmes multivariables". GDR 134 (Signal Processing) working document, february 1994, France.

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