1

FROM SINUSOIDS IN NOISE TO BLIND DECONVOLUTION IN COMMUNICATIONS

Dirk Slock

Eurecom Institute, Mobile Communications Dept.
B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE

Dedicated to Prof. Kailath at the occasion of his 60th birthday,
with fond memories of an inspiring teacher.

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. The multichannel formulation also arises in mobile communications, when multiple receiving antennas are used. In the multichannel case, 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 multichannel case FIR ZF equalizers exist for a FIR channel. 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. Due to the singularity and the FIR assumption, the spectral factorization reduces to the triangular factorization of a finite covariance matrix. In the given data case, Music (subspace) or ML techniques can be applied. We present these developments by drawing the parallel with existing techniques for the sinusoids in noise subspace problem.

1 INTRODUCTION

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

$$y(t) = \sum_k a(k)\mathrm{h}(t - kT) + \mathrm{v}(t) \tag{1.1}$$

where the $a(k)$ are the transmitted symbols, T is the symbol period, $\mathrm{h}(t)$ 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 $\{\mathrm{v}(t)\}$ to be (wide-sense) stationary, the process $\{\mathrm{y}(t)\}$ 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 $\{\mathrm{y}(t)\}$ 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 $\mathrm{h}(t)$ 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.

We review a number of developments in this context that have been inspired by their work. We shall concentrate here on the blind identification of a multichannel,i.e. without the use of training sequences for the transmitted symbols. In terms of second-order statistics, we shall only consider the statistics of the stationary vector received signal corresponding to the multichannel, and not cyclic statistics of possibly oversampled signals. We shall also not consider the exploitation of other side information such as the distribution of the symbols. The case of multiple sources was treated in e.g. [2].

1.1 Linear Multichannels

The multiple discrete-time channels we consider here come about by representing a diversity channel at the symbol rate. The diversity considered here can be spectral diversity, in which case oversampling w.r.t. the symbol rate exploits excess bandwidth (see Fig. 1), or spatial diversity obtained through multiple receiving antennas (and/or multiple polarizations) in wireless communications, see Fig. 2. A third case of diversity arises if the symbol constellation is one-dimensional (e.g. PAM or BPSK) and the transmitted signal is modulated. In

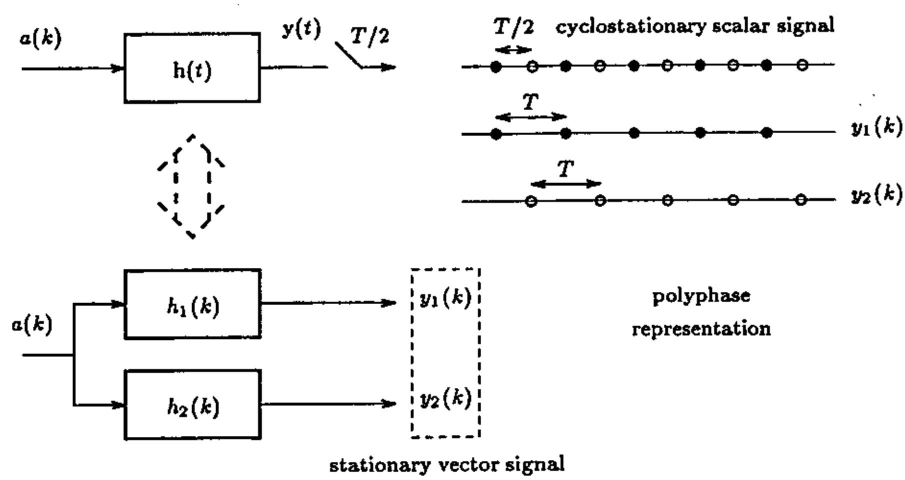

Figure 1 Polyphase and resulting vector channel representation of a $(T/2)$ oversampled received signal and channel.

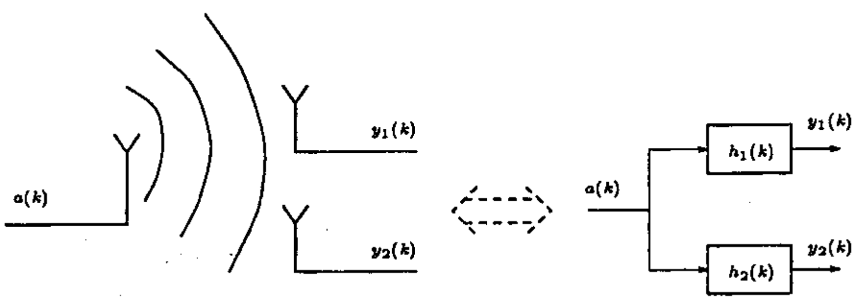

Figure 2 Vector channel representation of a signal received through multiple (2) antennas and sampled at the symbol rate.

that case the baseband channel impulse response has a real and an imaginary component, whereas the input is purely real. Hence, working with real signals only, we get a one-input two-output system. We assume in all cases the channel to be FIR with duration approximately NT. To further develop the case of oversampling, consider the sampling rate $\frac{m}{T}$. The sampling instants for the received signal in (1.1) 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 the case of Fig. 1, the two polyphase components are the even and odd samples. 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 $\boldsymbol{h}(0) = \left[\mathrm{h}(t'_0{+}dT)\cdots\mathrm{h}(t'_0{+}(d{+}\frac{m-1}{m})T)\right]^T \neq 0$ (superscript ^{T} denotes transpose). The channel being causal implies that d is nonnegative; d represents an inherent delay. The z-transform of the channel response at the sampling rate $\frac{m}{T}$ is $\mathrm{H}(z) = \sum_{j=1}^{m} z^{-(j-1)}\mathrm{H}_j(z^m)$.

The oversampled received signal can be represented in vector form at the symbol rate as

$$
\begin{aligned}
\boldsymbol{y}(k) &= \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k) = \boldsymbol{H}A_N(k) + \boldsymbol{v}(k)\,, \\
\boldsymbol{y}(k) &= \begin{bmatrix} y_1(k) \\ \vdots \\ y_m(k) \end{bmatrix}, \boldsymbol{v}(k) = \begin{bmatrix} v_1(k) \\ \vdots \\ v_m(k) \end{bmatrix}, \boldsymbol{h}(k) = \begin{bmatrix} h_1(k) \\ \vdots \\ h_m(k) \end{bmatrix} \\
\boldsymbol{H} &= [\boldsymbol{h}(N{-}1)\cdots\boldsymbol{h}(0)]\,, A_N(k) = \left[a(k{-}N{+}1)^H\cdots a(k)^H\right]^H
\end{aligned}
\tag{1.2}
$$

where superscript ^{H} denotes Hermitian transpose. We formalize the finite duration NT assumption of the channel as follows (AFIR): $\boldsymbol{h}(0) \neq 0$, $\boldsymbol{h}(N{-}1) \neq 0$ and $\boldsymbol{h}(i) = 0$ for $i < 0$ or $i \geq N$.

A sequence of received signal samples can be represented as

$$
\boldsymbol{Y}_L(k) = \mathcal{T}_L(\boldsymbol{H})\,A_{L+N-1}(k) + \boldsymbol{V}_L(k) \tag{1.3}
$$

where $\boldsymbol{Y}_L(k) = [\boldsymbol{y}^H(k{-}L{+}1)\cdots\boldsymbol{y}^H(k)]^H$ and similarly for $\boldsymbol{V}_L(k)$, and $\mathcal{T}_M(\boldsymbol{x})$ is a (block) Toeplitz matrix with M (block) rows and $\left[\boldsymbol{x}\ 0_{r\times(M-1)s}\right]$ as first (block) row ($r \times s$ is the size of the blocks in $\boldsymbol{x}$). The generic identifiability of the channel (and the symbols) in the multichannel case becomes apparent by considering equation (1.3) in the noiseless case: $\boldsymbol{v}(k) \equiv 0$. In this case, (1.3) represents a set of mL equations in mN (in $\boldsymbol{H}$) plus $L{+}N{-}1$ (in $A_{L+N-1}(k)$) unknowns. This set of equations can in principle be solved for the unknowns if $mL \geq mN{+}L{+}N{-}1$, hence if $L \geq \frac{(m+1)N-1}{m-1}$. For any finite channel length N a data record length L can be found to satisfy this constraint if the number of channels $m \geq 2$. The system of equations is always underdetermined however for a single channel ($m = 1$).

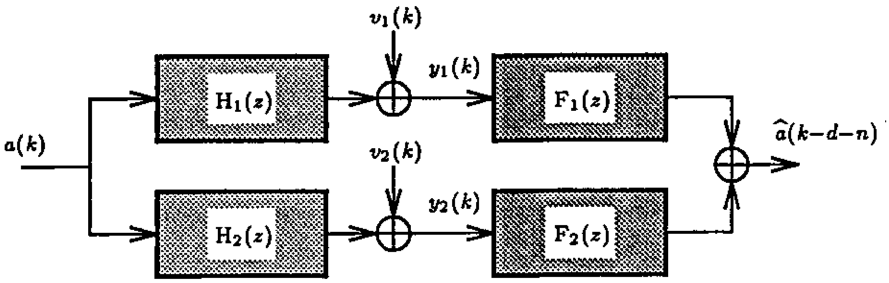

Figure 3 Polyphase representation of the T/m fractionally-spaced channel and equalizer for $m = 2$.

1.2 FIR Zero-Forcing (ZF) Equalization

In the multichannel representation, the Single Input Multiple Output (SIMO) vector channel transfer function from the single input $a(k)$ to the multiple outputs is $\mathrm{H}(z) = \sum_{i=0}^{N-1} h(i)z^{-1}$ so that we can write for the vector received signal: $y(k) = \mathrm{H}(q)\,a(k) + v(k)$, where $q^{-1}a(k) = a(k-1)$. Consider now (in the oversampling context) a fractionally-spaced ($\frac{T}{m}$) equalizer of which the z-transform can also be decomposed into its polyphase components: $\mathrm{F}(z) = \sum_{j=1}^{m} z^{-(m-j)}\mathrm{F}_j(z^m)$. We assume the equalizer phases to be causal and FIR of length L: $\mathrm{F}_j(z) = \sum_{k=0}^{L-1} f_j(k)z^{-k}$, $j = 1,\ldots,m$. The polyphase representation of the fractionally-spaced equalizer leads to a Multi Input Single Output (MISO) system representation $\mathrm{F}(z) = \sum_{k=0}^{N-1} f(k)z^{-k}$ with $f(k) = [f_1(k)\cdots f_m(k)]$ and $F = [f(L-1)\cdots f(0)]$, see Fig. 3 for $m = 2$. This MISO equalizer obvisouly applies to all multichannel formulations. The condition for the equalizer to be ZF is $\mathrm{F}(z)\mathrm{H}(z) = \sum_{j=1}^{m} \mathrm{F}_j(z)\mathrm{H}_j(z) = z^{-n}$ where $n \in \{0, 1, \ldots, N+L-2\}$. By equating equal powers of z^{-1}, we can write this in matrix form as

$$F\ \mathcal{T}_L(H) = [0\cdots 0\ 1\ 0\cdots 0] \tag{1.4}$$

where the 1 is in the n+1st position from the end. (1.4) is a system of $L+N-1$ equations in Lm unknowns. To enable zero-forcing (ZF) equalization, we need to choose the equalizer length L such that the system of equations (1.4) is exactly or underdetermined. Hence

$$L \geq \underline{L} = \left\lceil \frac{N-1}{m-1} \right\rceil . \tag{1.5}$$

The block Toeplitz matrix $\mathcal{T}_L(H)$ is also a generalized Sylvester matrix. It can be shown that for $L \geq \underline{L}$ it has full column rank if $\mathrm{H}(z) \neq 0$, $\forall z$ or in other words if the $H_j(z)$ have no zeros in common. Assuming $\mathcal{T}_L(\boldsymbol{H})$ to have full column rank, the nullspace of $\mathcal{T}_L^H(\boldsymbol{H})$ 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. We shall call such an equalizer a blocking equalizer.

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, or fix the excess degrees of freedom by requiring minimal noise enhancement). The remaining $\underline{L}+N-1$ coefficients can be found from (1.4) if $\mathbf{H}(z) \neq 0, \forall z$. This shows that in the multichannel case, a FIR equalizer suffices for ZF equalization! With $m = N$ channels, the minimal required total number of equalizer coefficients N is found ($\underline{L} = 1$). In the multichannel case, FIR ZF equalization is an issue of finding an FIR synthesis bank given an FIR analysis bank so that the overall filterbank has the perfect reconstruction property. In [3], another interpretation of the ZF condition is given in the oversampled case that is similar to the Nyquist condition in the continuous-time case.

2 SINUSOIDS IN NOISE: A REVIEW

As can be seen from (1.3), the blind equalization problem in the multichannel case is a signal subspace estimation problem. That means that all the techniques that have been developed for subspace problems are applicable here also. Therefore we shall review these techniques in the context of a basic subspace problem: the sinusoids in noise problem. Afterwards we shall show that all the methods applicable to the sinusoids in noise problem are also applicable to the blind multichannel equalization problem.

2.1 Second-Order Statistics

Consider a signal x_k consisting of a sum of sinusoids

$$x_k = \sum_{i=1}^{M} A_i \cos(\omega_i k + \phi_i) \; y_k = x_k + v_k \; . \tag{1.6}$$

We shall assume that the amplitudes A_i and the frequencies $f_i = \omega_i/2\pi$ are deterministic unknowns, while the phases ϕ_i are unknown and random, independent of each other and uniformly distributed over $[0, 2\pi)$. Hence x_k is a stationary process with zero mean. The actual measured signal y_k is a noisy version of x_k with v_k being independent white noise with variance σ_v^2. But we first concentrate on the noise-free signal x_k. The support of the power spectral density function $S_{xx}(f) = \sum_{i=1}^{M} \frac{A_i^2}{2}(\delta(f-f_i)+\delta(f+f_i))$ has measure zero. The covariance matrix R_{XX} has rank $2M$ whenever its dimension exceeds $2M$. Hence, the rank remains finite, even if the dimension goes to infinity. The frequencies and phases can easily be determined from $S_{xx}(f)$.

2.2 Noise-free Prediction Problem

A sum of sinusoids satisfies a homogeneous difference equation. In particular

$$\mathrm{P}(q)\, x_k = 0\,, \quad \mathrm{P}(z) = \prod_{i=1}^{M}(1-2\cos\omega_i\, z^{-1}+z^{-2})\,. \tag{1.7}$$

Hence, x_k is perfectly predictible from its previous $2M$ samples. $\mathrm{P}(z)$ and hence the ω_i can be found by linear prediction: this is *Prony*'s method. The normal equations governing the linear prediction problem are

$$P\, R_{XX} = [0 \cdots 0\; \sigma^2]\,, \quad \sigma^2 = 0 \tag{1.8}$$

where

$$\begin{aligned} R_{XX} &= \mathrm{E}\, X X^T\,,\; X = [x_0 \cdots x_{2M}]^T \\ P &= [P_{2M} \cdots P_1\; P_0] \\ P_0 &= 1\,,\; P_i = P_{2M-i}\,,\; i = 0,\ldots,M-1\,. \end{aligned} \tag{1.9}$$

2.3 Signal and Noise Subspaces

The signal structure can be revealed by considering $X_k = \mathcal{V}\, S$:

$$\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ \cos\omega_1 & \sin\omega_1 & \cdots & \cos\omega_M & \sin\omega_M \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos\omega_1 k & \sin\omega_1 k & \cdots & \cos\omega_M k & \sin\omega_M k \end{bmatrix} \begin{bmatrix} A_1\cos\phi_1 \\ -A_1\sin\phi_1 \\ \vdots \\ A_M\cos\phi_M \\ -A_M\sin\phi_M \end{bmatrix} \tag{1.10}$$

where $\mathcal{V} = \mathcal{V}(\Omega)$ with $\Omega = [\omega_1 \cdots \omega_M]^T$ is a block Vandermonde matrix. Indeed, two consecutive block elements of a block column of $\mathcal{V}$ are proportional:

$$[\cos\omega_i k \;\; \sin\omega_i k] = [\cos\omega_i(k-1) \;\; \sin\omega_i(k-1)] \begin{bmatrix} \cos\omega_i & \sin\omega_i \\ -\sin\omega_i & \cos\omega_i \end{bmatrix}\,. \tag{1.11}$$

One calls

$$\begin{array}{rcl} Range\{\mathcal{V}\} & = & \text{SSS } = \text{signal subspace} \\ (Range\{\mathcal{V}\})^{\perp} & = & \text{NSS } = \text{noise subspace} \end{array} \tag{1.12}$$

We get for the covariance structure of the noisy signal

$$Y_k = X_k + V_k = \mathcal{V}_k\, S + V_k \;\Rightarrow\; R_{YY} = \mathcal{V}\, R_{SS}\, \mathcal{V}^T + \sigma_v^2 I \tag{1.13}$$

where $R_{SS} = \frac{1}{2}\,\mathrm{diag}\left\{A_1^2, A_1^2, \ldots, A_M^2, A_M^2\right\}$. The signal component X_k of the measurement vector Y_k can only live in the signal subspace, whereas in the noise subspace only noise can be found.

Consider the eigendecomposition of R_{YY} $(\lambda_1 \geq \lambda_2 \geq \cdots)$:

$$R_{YY} = \sum_{i=1}^{2M} \lambda_i V_i V_i^H + \sum_{i=2M+1}^{k+1} \lambda_i V_i V_i^H = V_S \Lambda_S V_S^H + V_{\mathcal{N}} \Lambda_{\mathcal{N}} V_{\mathcal{N}}^H \tag{1.14}$$

where $V_S = [V_1 \cdots V_{2M}]$, $V_{\mathcal{N}} = [V_{2M+1} \cdots V_{k+1}]$, $\Lambda_S = \mathrm{diag}\{\lambda_1, \ldots, \lambda_{2M}\}$ and $\Lambda_{\mathcal{N}} = \sigma_v^2 I_{k+1-2M}$. Assuming $\mathcal{V}$ and R_{SS} to have full rank (all $\omega_i \in (0, \pi)$ are different, all $A_i > 0$), the sets of eigenvectors V_S and $V_{\mathcal{N}}$ are orthogonal: $V_S^H V_{\mathcal{N}} = 0$, and $\lambda_i > \sigma_v^2$, $i = 1, \ldots, 2M$. The following equivalent descriptions of the signal and noise subspaces result:

$$\begin{array}{rcl} \text{SSS} & = & Range\{V_S\} = Range\{\mathcal{V}(\Omega)\} \\ \text{NSS} & = & Range\{V_{\mathcal{N}}\} \;:\; V_{\mathcal{N}}^T \mathcal{V} = 0\,. \end{array} \tag{1.15}$$

So far we have a parametric description of the SSS and we can compute both subspaces from the eigendecomposition of R_{YY}. We can also find a parametric description of the NSS. Indeed, we have $\mathrm{P}(q)\, \cos\omega_i k = 0$ and $\mathrm{P}(q)\, \sin\omega_i k = 0$, hence $\mathcal{G}(P)^T \mathcal{V} = 0$ where

$$\mathcal{G}(P)^T = \mathcal{T}_{k+1-2M}(P) = \begin{bmatrix} P_{2M} & \cdots & P_1 & P_0 & 0 & \cdots & 0 \\ 0 & P_{2M} & \cdots & P_1 & P_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & P_{2M} & \cdots & P_1 & P_0 \end{bmatrix} \tag{1.16}$$

is a Toeplitz matrix of full rank, equal to the dimension of the NSS. Hence

$$\text{NSS} = Range\{\mathcal{G}(P)\}\,. \tag{1.17}$$

Note that both the SSS, $\mathcal{V}(\Omega)$, and the NSS, $\mathcal{G}(P)$, are parameterized by M independent parameters.

2.4 Frequency Estimation from Second-Order Statistics

Linear Prediction from Denoised Statistics

In the case of additive white noise, we can retrieve the noise-free prediction coefficients by replacing the monic constraint of linear prediction by a norm constraint in the minimization of the prediction error variance. We get the *Pisarenko* method

$$\begin{aligned} &\min_{\|P\|=1} P\,R_{YY}\,P^T = \min_{\|P\|=1} P\,R_{XX}\,P^T + \sigma_v^2 \\ &\Rightarrow \quad P\,R_{XX} = [0\cdots 0]\,,\ P^T = V_{2M+1} \end{aligned} \tag{1.18}$$

This is the case of $k = 2M$: NSS dimension = 1. Equivalently, we can identify $\sigma_v^2 = \lambda_{2M+1}$ and apply linear prediction to the denoised statistics $R_{XX} = R_{YY} - \lambda_{2M+1} I_{2M+1}$.

Signal Subspace Fitting

We have two theoretically equivalent signal subspace descriptions: $\mathcal{V}(\Omega)$ and $V_S = V_S(R_{YY})$: both matrices have the same column space, hence one matrix can be transformed into the other one. V_S can be computed from the covariance matrix. By fitting $\mathcal{V}(\Omega)$ to V_S, we can determine Ω. With an estimated covariance matrix, V_S is approximate, so consider the following subspace fitting criterion

$$\min_{\Omega,T} \|\mathcal{V}(\Omega) - V_S T\|_F \tag{1.19}$$

where T is a square transformation matrix and the Frobenius norm is defined as $\|A\|_F^2 = \mathrm{tr} A^H A$ with tr denoting trace. The minimal value of the criterion is zero if V_S is exact. This criterion differs from the original subspace fitting strategy proposed in [4], which would propose $\min_{\Omega,T} \|\mathcal{V}(\Omega)T - V_S\|_F$ as criterion. We propose (1.19) because it leads to a simpler optimization problem. Both approaches can be made to be equivalent by the introduction of column space weighting. The criterion in (1.19) is separable. In particular, it is quadratic in T. Minimization w.r.t. T leads to $T = V_S^T \mathcal{V}$ and $\mathcal{V} - V_S T = P_{V_S}^{\perp} \mathcal{V}$ where $P_{V_S}^{\perp} = I - P_{V_S}$ and P_{V_S} is the projection matrix on the signal subspace (V_S).

Hence

$$\begin{aligned}\min_T \|\mathcal{V}(\Omega) - V_S T\|_F^2 &= \|P_{V_S}^{\perp}\mathcal{V}\|_F^2 \\ &= \operatorname{tr}\mathcal{V}^T P_{V_S}^{\perp}\mathcal{V} = \operatorname{tr}\mathcal{V}^T P_{V_\mathcal{N}}\mathcal{V} = \|V_\mathcal{N}^T\mathcal{V}\|_F^2 \\ &= \sum_{i=2M+1}^{k} \|V_i^T\mathcal{V}\|^2 = \sum_{j=1}^{M}\sum_{i=2M+1}^{k} |V_i(\omega_j)|^2\end{aligned} \tag{1.20}$$

where $V_i(\omega) = [1\ e^{j\omega}\cdots e^{j\omega k}]\,V_i$ is the Fourier transform of the elements of V_i. The last expression in (1.20) needs to be minimized w.r.t. the ω_i. An approximate solution can be found as follows. Plot as a function of ω and find the ω_i as the abscissae of the M largest peaks of

$$\frac{1}{\sum_{i=2M+1}^{k} |V_i(\omega)|^2}\,. \tag{1.21}$$

This method is called MUSIC.

Noise Subspace Fitting

We have again two theoretically equivalent noise subspace descriptions: $\mathcal{G}(P)$ and $V_\mathcal{N} = V_\mathcal{N}(R_{YY})$. By fitting $\mathcal{G}(P)$ to $V_\mathcal{N}$, we can determine P. We introduce the following noise subspace fitting criterion

$$\min_{P,T} \|\mathcal{G}(P) - V_\mathcal{N}T\|_F\,. \tag{1.22}$$

Minimization w.r.t. T first leads again to $T = V_\mathcal{N}^T\mathcal{G}$ and $\mathcal{G} - V_\mathcal{N}T = P_{V_\mathcal{N}}^{\perp}\mathcal{G}$ and hence

$$\begin{aligned}\min_{P,T} \|\mathcal{G}(P) - V_\mathcal{N}T\|_F &= \|P_{V_\mathcal{N}}^{\perp}\mathcal{G}\|_F^2 = \operatorname{tr}\mathcal{G}^T P_{V_\mathcal{N}}^{\perp}\mathcal{G} = \operatorname{tr}\mathcal{G}^T P_{V_S}\mathcal{G} \\ &= \|V_S^T\mathcal{G}\|_F^2 = \sum_{i=1}^{2M} \|\mathcal{G}^T V_i\|^2\end{aligned} \tag{1.23}$$

Due to the commutativity of convolution, we can write $\mathcal{G}^T V_i = \mathcal{H}_i\,P^T$ where $\mathcal{H}_i = \mathcal{H}(V_i)$ is Hankel. The symmetry of P can be expressed as $P = P\,J$ where J is the reverse identity matrix (ones on the main antidiagonal). We can assure the symmetry of the solution for P by explicitly expressing the symmetry of P. In this way, minimization of the last criterion in (1.23) w.r.t. P leads to

$$\min_P P\left[(\sum_{i=1}^{2M}\mathcal{H}_i^T\mathcal{H}_i) + J\,(\sum_{i=1}^{2M}\mathcal{H}_i^T\mathcal{H}_i)\,J\right] P^T \tag{1.24}$$

subject to $P_0 = 1$ or $\|P\| = 1$. Roughly this approach has been proposed recently in [5].

2.5 Frequency Estimation from Data: ML

With white Gaussian additive noise v_k, the likelihood function becomes the following least-squares criterion

$$\min_{\Omega, S} \|Y - \mathcal{V}(\omega)\, S\|^2 \,. \tag{1.25}$$

The criterion is again separable and minimization w.r.t. S first leads to $S = (\mathcal{V}^T\mathcal{V})^{-1}\mathcal{V}^T Y$. Hence

$$\min_{S} \|Y - \mathcal{V}\, S\|^2 = Y^T P_{\mathcal{V}}^{\perp} Y = Y^T P_{\mathcal{G}(P)} Y = P\mathcal{Y}^T(\mathcal{G}(P)^T\mathcal{G}(P))^{-1}\mathcal{Y}P^T \tag{1.26}$$

where we again exploited the commutativity of convolution: $\mathcal{G}(P)^T Y = \mathcal{H}(Y)\, P^T = \mathcal{Y}\, P^T$. Note that in (1.26), we went from the SSS parameterization in terms of Ω to the equivalent NSS parameterization in terms of P. The reason is that now a straightforward iterative procedure suggests itself, known as the Iterative Quadratic Maximum Likelihood (IQML) procedure [6]. At iteration n, we get the following quadratic criterion:

$$\min_{P^{(n)}} P^{(n)}\mathcal{Y}^T(\mathcal{G}(P^{(n-1)})^T\mathcal{G}(P^{(n-1)}))^{-1}\mathcal{Y}\, P^{(n)T} \tag{1.27}$$

subject to $P_0 = 1$ or $\|P\| = 1$. We could again incorporate the symmetry of P. The IQML iterations are not guaranteed to converge. However, with a consistent initialization (such as obtained from the second-order statistics based methods), only one iteration is required to get an Asymptotically Best Consistent (ABC) estimate for P and hence Ω.

2.6 Adaptive Notch Filtering

The model for the sum of sinusoids $\mathrm{P}(q)x_k = 0$ naturally leads to the following constrained ARMA representation for the measured signal y_k:

$$\mathrm{P}(q)y_k = \mathrm{P}(q)v_k \;\Rightarrow\; v_k = \frac{\mathrm{P}(q)}{\mathrm{P}(q/\rho)}\, y_k = \mathrm{N}(q)\, y_k \quad \text{as } \rho \lesssim 1 \tag{1.28}$$

where as indicated we can in principle recover the additive noise v_k from the measurements y_k using an infinitely sharp notch filter. This infinitely sharp notch filter is in practice approximated by the IIR notch filter $N(z) = P(z)/P(z/\rho)$ with zeros $e^{\pm j\omega_i}$ and poles $\rho e^{\pm j\omega_i}$. When the coefficients of P in $N(z)$ (the notches) are not properly chosen, the notch filter output will not equal (approximate) v_k but can be written in general as $\epsilon_k = N(q)\, y_k = N(q)x_k + N(q)v_k$. The notch filter output variance is ($N(f) = N(e^{j2\pi f})$)

$$\begin{aligned} \mathrm{E}\, \epsilon_k^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |N(f)|^2 S_{xx}(f)df + \int_{-\frac{1}{2}}^{\frac{1}{2}} |N(f)|^2 S_{vv}(f)df \\ &= \sum_{i=1}^{M} \frac{A_i^2}{2} |N(f_i)|^2 + \sigma_v^2 \end{aligned} \tag{1.29}$$

where the expression for the second term is valid for infinitely sharp notches. We can find P by minimizing the notch filter output variance w.r.t. it since indeed

$$\min_P \ \mathrm{E}\, \epsilon_k^2 = \sigma_v^2 \,. \tag{1.30}$$

Due to the long transients of the notch filter, this approach lends itself to adaptive filtering. An adaptive notch filter can be obtained by applying any adaptive filtering strategy to the MMSE criterion. Remark in particular that the ML criterion is in fact the sum of squares of the additive noise $v_k = y_k - x_k$.

3 BLIND CHANNEL ESTIMATION

Now we shall draw the parallel of the previous approaches for the blind equalization problem.

3.1 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 (psd) matrix of the stationary vector process $\boldsymbol{y}(k)$ is

$$S_{\boldsymbol{yy}}(z) = \sigma_a^2 \mathrm{H}(z)\, \mathrm{H}^{\dagger}(z) \tag{1.31}$$

where $\mathrm{H}^{\dagger}(z) = \mathrm{H}^H(z^{-*})$. The following spectral factorization result can be found in [7]. Let $\mathrm{K}(z)$ be a $m \times 1$ rational transfer function that is causal and stable. Then $\mathrm{K}(z)$ is called minimum-phase if $\mathrm{K}(z) \neq 0$, $|z| > 1$. Let $S_{\boldsymbol{yy}}(z)$ be a rational $m \times m$ spectral density matrix of rank 1. Then there exists a rational

$m \times 1$ transfer matrix $K(z)$ that is causal, stable , minimum-phase, unique up to a unitary constant, of (minimal) McMillan degree $\deg(K) = \frac{1}{2}\deg(S_{yy})$ such that

$$S_{yy}(z) = K(z)\, K^{\dagger}(z)\,. \tag{1.32}$$

In our case, S_{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 $K(z)$ identifies the channel

$$K(z) = \sigma_a\, e^{j\phi}\, \mathbf{H}(z) \tag{1.33}$$

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.

Note that the psd matrix of the noise-free signal y_k is of rank one and hence is singular in the multichannel case $m > 1$. We recall that the input-output relation of the channel is

$$\boldsymbol{Y}_L(k) = \mathcal{T}_L(\boldsymbol{H})\, A_{L+N-1}(k). \tag{1.34}$$

Therefore, the structure of the covariance matrix of the received signal $y(k)$ is

$$R_L^y = E\, \boldsymbol{Y}_L(k) \boldsymbol{Y}_L^H(k) = \mathcal{T}_L(\boldsymbol{H})\, R_{L+N-1}^a \mathcal{T}_L^H(\boldsymbol{H}) \tag{1.35}$$

where $R_L^a = E\, A_L(k) A_L^H(k) > 0$. When $mL > L+N-1$, R_L^y is singular. If then L increases further by 1, the rank of R_L^y increases by 1 and the dimension of its nullspace increases by $m-1$. In fact, $\frac{\text{rank}}{\text{dimension}} = \frac{L+N-1}{mL} \xrightarrow{L\to\infty} \frac{1}{m}$.

So the channel can in principle be identified by spectral factorization, an iterative procedure that represents an infinite number of computations. We shall see however that due to the singularity and the FIR assumption the channel can be identified from the triangular factorization of a finite covariance matrix.

3.2 Noise-free Prediction Problem

Multichannel Linear Prediction

Consider now the problem of predicting $y(k)$ from $\boldsymbol{Y}_L(k-1)$ The prediction error can be written as

$$\widetilde{y}(k)|_{\boldsymbol{Y}_L(k-1)} = y(k) - \widehat{y}(k)|_{\boldsymbol{Y}_L(k-1)} = \boldsymbol{P}_L \boldsymbol{Y}_{L+1}(k) \tag{1.36}$$

with $P_L = [P_{L,L} \cdots P_{L,1}\ P_{L,0}]$, $P_{L,0} = I_m$. Minimizing the prediction error variance leads to the following optimization problem

$$\min_{P_L : P_{L,0} = I_m} P_L \mathrm{R}^{y}_{L+1} P_L^H \;=\; \sigma^2_{\widetilde{y},L} \tag{1.37}$$

or hence

$$P_L \mathrm{R}^{y}_{L+1} \;=\; [0 \cdots 0]\;\; \sigma^2_{\widetilde{y},L}\ . \tag{1.38}$$

When $mL > L{+}N{-}1$, $\mathcal{T}_L(H)$ has full column rank. Hence, using (1.34) and (1.36),

$$\begin{aligned} \widetilde{y}(k)|_{Y_{L}(k-1)} &= \widetilde{y}(k)|_{A_{L+N-1}(k-1)} = y(k) - \widehat{y}(k)|_{A_{L+N-1}(k-1)} \\ &= \sum_{i=0}^{N-1} h(i)a(k{-}i) - \sum_{i=0}^{N-1} h(i)\ \widehat{a}(k{-}i)|_{A_{L+N-1}(k-1)} \\ &= \sum_{i=0}^{N-1} h(i)a(k{-}i) - \sum_{i=1}^{N-1} h(i)a(k{-}i) - h(0)\ \widehat{a}(k)|_{A_{L+N-1}(k-1)} \\ &= h(0)\ \widetilde{a}(k)|_{A_{L+N-1}(k-1)} \end{aligned} \tag{1.39}$$

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

$$\widetilde{a}(k)|_{A_M(k-1)} = a(k) - \widehat{a}(k)|_{A_M(k-1)} = Q_M\, A_M(k{-}1),\, Q_M \mathrm{R}^{a}_{M+1} = \left[0 \cdots 0\; \sigma^2_{\widetilde{a},M}\right] \tag{1.40}$$

where $Q_M = [Q_{M,M} \cdots Q_{M,1}\ 1]$. We find from (1.39),(1.40)

$$\widetilde{y}(k)|_{Y_L(k-1)} = P_L \mathcal{T}_{L+1}(H)\, A_{L+N}(k) = \mathbf{h}(0)\, Q_{L+N-1} A_{L+N}(k) \tag{1.41}$$

for all $A_{L+N}(k)$ and hence

$$P_L \mathcal{T}_{L+1}(H) \;=\; \mathbf{h}(0)\, Q_{L+N-1}\ . \tag{1.42}$$

From (1.39), we also get

$$\sigma^2_{\widetilde{y},L} \;=\; h(0)\sigma^2_{\widetilde{a},L+N-1} h^H(0)\ . \tag{1.43}$$

All this holds for $L \geq \underline{L}$. As a function of L, the rank profile of $\sigma^2_{\widetilde{y},L}$ behaves like

$$\operatorname{rank}\left(\sigma^2_{\widetilde{y},L}\right) \begin{cases} = 1 & ,\ L \geq \underline{L} \\ = m{-}\underline{m} \in \{2,3,\ldots,m\} & ,\ L = \underline{L}{-}1 \\ = m & ,\ L < \underline{L}{-}1 \end{cases} \tag{1.44}$$

where $\underline{m} = m\underline{L} - (\underline{L}+N-1) \in \{0, 1, \ldots, m-2\}$ represents the degree of singularity of $R^{y}_{\underline{L}}$. Note that multichannel linear prediction corresponds to block triangular factorization of (some generalized) inverse of R^{y}. Indeed,

$$\boldsymbol{L}_L \, R^{y}_{L} \boldsymbol{L}_L^H = \boldsymbol{D}_L \; , \; (\boldsymbol{L}_L)_{i,j} = \boldsymbol{P}_{i-1,i-j} \; , \; (\boldsymbol{D}_L)_{i,i} = \sigma^2_{\widetilde{y},i-1} \tag{1.45}$$

where $\boldsymbol{L}_L$ is block lower triangular and $\boldsymbol{D}_L$ is block diagonal. (A slight generalization to the singular case of) the multichannel Levinson algorithm can be used to compute the prediction quantities and hence the triangular factorization above in a fast way. In the case that $R^{y}_{\underline{L}}$ is singular, some precaution is necessary in the determination of the last block coefficient $P_{\underline{L},\underline{L}}$ (see [8]). Similar singularities will then arise at higher orders.

Uncorrelated Symbols

We shall now concentrate on the case in which the symbols $a(k)$ are uncorrelated. In this case the noise-free received signal is a singular multivariate MA process. Observe that for $L = \underline{L}$ we have

$$\boldsymbol{y}(k) + \sum_{i=1}^{\underline{L}} P_{\underline{L},i}\boldsymbol{y}(k-i) = \widetilde{\boldsymbol{y}}_{\underline{L}}(k) \; = \; \boldsymbol{h}(0)\, \widetilde{a}_{\underline{L}+N-1}(k) \; = \; \boldsymbol{h}(0)\, a(k) \tag{1.46}$$

so that the prediction error is a singular white noise. This means that the noise-free received signal $\boldsymbol{y}(k)$ is also a singular multivariate AR process. Hence

$$\boldsymbol{P}_L = [\cdots 0 \;\; \boldsymbol{P}_{\underline{L}}] \; , \; \sigma^2_{\widetilde{y},L} = \sigma^2_{\widetilde{y},\underline{L}} \; , \; L > \underline{L} \, . \tag{1.47}$$

Hence the factors $\boldsymbol{L}_L$ and $\boldsymbol{D}_L$ in the factorization (1.45) become block Toeplitz after $\underline{L}$ lines.

For $L = \underline{L}$, (1.43) allows us to find $\boldsymbol{h}(0)$ up to a scalar multiple. We see from (1.42) that $\frac{\boldsymbol{h}^H(0)}{\boldsymbol{h}^H(0)\boldsymbol{h}(0)}\boldsymbol{P}_{\underline{L}}$ is a zero-delay ZF equalizer. Given $\boldsymbol{h}(0)$ from (1.43) and $\boldsymbol{P}_{\underline{L}}$, we can solve for the channel impulse response $\boldsymbol{H}$ from (1.42). The channel can alternatively be found from

$$\boldsymbol{P}_{\underline{L}} \, \mathrm{E}\boldsymbol{Y}_{\underline{L}+1}(k)\boldsymbol{Y}^H_N(k+N-1) \; = \; \sigma^2_a \, \boldsymbol{h}(0)[\boldsymbol{h}^H(0) \cdots \boldsymbol{h}^H(N-1)] \tag{1.48}$$

or from $\mathbf{P}_{\underline{L}}(z)\mathbf{H}(z) = \boldsymbol{h}(0) \;\Rightarrow\; \mathbf{H}(z) = \mathbf{P}^{-1}_{\underline{L}}(z)\,\boldsymbol{h}(0)$ using the lattice parameterization for $\mathbf{P}_{\underline{L}}(z)$ obtained with the Levinson algorithm.

In the uncorrelated symbols 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 $\sigma^2_{\tilde{y},L}$ becomes of rank 1, its one nonzero eigenvalue $\sigma^2_{\tilde{a},L+N-1} h^H(0)h(0)$ continues to decrease as a function of L since for a MA process, $\sigma^2_{\tilde{a},L}$ is a decreasing function of L.

Correlated Symbols

Now consider the case in which the symbols $a(k)$ are correlated. Still, for $L = \underline{L}$, (1.43) allows us to find $h(0)$ up to a scalar multiple. Let $h^\perp$ be $m \times (m-1)$ of rank $m-1$ such that $h^{\perp H} h(0) = 0$, then

$$F^b_{\underline{L}+1} \;=\; h^{\perp H}\, P_{\underline{L}} \tag{1.49}$$

is a set of $m-1$ blocking equalizers since indeed $F^b Y_L(k) = 0$. We introduce a block-componentwise transposition operator ^{t}, viz.

$$\begin{aligned} H^t \;&=\; [h(N-1)\cdots h(0)]^t \;=\; \left[h^T(N-1)\cdots h^T(0)\right] \\ F^t_L \;&=\; [\mathrm{f}(L-1)\cdots \mathrm{f}(0)]^t \;=\; \left[\mathrm{f}^T(L-1)\cdots \mathrm{f}^T(0)\right] \end{aligned} \tag{1.50}$$

where ^{T} is the usual transposition operator. Due to the commutativity of convolution, we find

$$F^b_L\, \mathcal{T}_L\,(H_N) \;=\; 0 \quad\Longleftrightarrow\quad \mathrm{H}^t_N\, \mathcal{T}_N\left(F^{b\,t}_L\right) \;=\; 0\,. \tag{1.51}$$

Now

$$\dim\left(Range^\perp\left\{\mathcal{T}_N\left(F^{b\,t}_{\underline{L}+1}\right)\right\}\right) = 1 \tag{1.52}$$

so that we can identify the channel $\mathrm{H}^{t\,H}_N$ (up to scalar multiple) as the last right singular vector of $\mathcal{T}_N\left(F^{b\,t}_{\underline{L}+1}\right)$ (a QR factorization would require less computations but might be less reliable numerically). From (1.42), one can furthermore identify $Q_{\underline{L}+N-1}$ and via (1.40), this leads to the identification of the (Toeplitz) symbol covariance matrix $\mathrm{R}^a_{\underline{L}+N}$ up to a multiplicative scalar.

Modular Multichannel Linear Prediction

We noted previously that the consecutive multichannel linear prediction problems correspond to a block triangular factorization. They also correspond to Gram-Schmidt orthogonalization of the block components of the vector Y. We can alternatively introduce sequential processing in the orthogonalization process and orthogonalize scalar component by scalar component the elements of the vector Y. This leads to cyclic prediction filters and a true (non-block) triangular factorization. To make the orthogonalization process unique in the singular case, we need to introduce a convention. We shall assume that components of Y of which the orthogonalized versions are zero are not used in the orthogonalization of the further components (this corresponds to some minimum-norm choice). The consequence of this convention is that zeros will appear in the triangular factor (cyclic prediction filters). We get

$$L_L^{'} \, \mathrm{R}_L^{\mathcal{Y}} L_L^{'H} = D_L^{'} \tag{1.53}$$

where $L_L^{'}$ is a unit-diagonal lower triangular factor and $D_L^{'}$ is a diagonal matrix. After $m\underline{L}$ rows, both matrices become block Toeplitz again in the case of uncorrelated symbols. The steady-state diagonal elements of $D_L^{'}$ become in that case $\sigma_a^2|h_1(0)|^2$ followed by $m{-}1$ zeros (since $h(0) \neq 0$, we can w.l.o.g. assume that $h_1(0) \neq 0$). If we introduce a permutation matrix $\mathcal{P}$

$$\mathcal{P} L_L^{'} \, \mathrm{R}_L^{\mathcal{Y}} L_L^{'H} \mathcal{P}^H = \mathcal{P} D_L^{'} \mathcal{P}^H \Rightarrow \begin{bmatrix} L^{''} \\ \mathcal{G} \end{bmatrix} \mathrm{R}_L^{\mathcal{Y}} \begin{bmatrix} L^{''} \\ \mathcal{G} \end{bmatrix}^H = \begin{bmatrix} D^{''} & 0 \\ 0 & 0 \end{bmatrix} \tag{1.54}$$

so that the non-singular and singular parts get separated ($D^{''}$ is non-singular). $L^{''}$ eventually becomes block Toeplitz with $1 \times m$ blocks. Its repeated row then corresponds to a zero-delay ZF equalizer (up to a scalar multiple). $\mathcal{G} = \mathcal{G}(G)$ is block Toeplitz with $(m{-}1) \times m$ blocks and contains $m{-}1$ blocking equalizers parameterized by G. The number of elements in G is $mN{-}1$ (the number of degrees of freedom in H that can be determined blindly). Apart from the elements of G, $\mathcal{G}$ also contains 1's.

Whereas a modular multichannel Levinson algorithm (slightly adapted to handle singularities) can be used to find this factorization and the prediction quantities involved fast, similarly, a corresponding modular multichannel Schur algorithm [9] can be used to find the LDU factorization of $\mathrm{R}_L^{\mathcal{Y}}$ itself:

$$\mathrm{R}_L^{\mathcal{Y}} = U_L^H \, D_L^{'} \, U_L \tag{1.55}$$

where we get the same diagonal factor and U_L is a unit-diagonal upper triangular matrix. Also U_L becomes block Toeplitz after $m\underline{L}$ rows. If we again introduce the permutation matrix $\mathcal{P}$, then

$$\begin{aligned} \mathrm{R}_L^{\mathcal{Y}} &= \left(U_L^H \mathcal{P}^H\right) \left(\mathcal{P} D_L^{'} \mathcal{P}^H\right) (\mathcal{P} U_L) \\ &= \left[U^{''} \; U^{'}\right] \begin{bmatrix} D^{''} & 0 \\ 0 & 0 \end{bmatrix} \left[U^{''} \; U^{'}\right]^H = U^{''} D^{''} U^{''H} \end{aligned} \tag{1.56}$$

After a finite number of columns, U'' becomes block Toeplitz with $m \times 1$ blocks and the column that gets repeated contains (a multiple of) the channel impulse response for reasons that are related to the fact that L'' contains a ZF equalizer and to (1.48). This shows our earlier claim that the spectral factorization of $S_{\mathcal{Y}\mathcal{Y}}(z)$ can be replaced by a triangular factorization of a finite covariance matrix $R^{\mathcal{Y}}$. More details can be found in [10]

3.3 Signal and Noise Subspaces

Consider now the measured data with additive independent white noise $v(k)$ with zero mean and assume $\mathrm{E}\, v(k)v^H(k) = \sigma_v^2 I_m$ 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

$$Y_L(k) = \mathcal{T}_L(H)\, A_{L+N-1}(k) + V_L(k)\,. \tag{1.57}$$

Therefore, the structure of the covariance matrix of the received signal $y(k)$ is

$$R_L^{\mathcal{Y}} = \mathrm{E}\, Y_L(k)Y_L^H(k) = \mathcal{T}_L(H)\, R_{L+N-1}^a \mathcal{T}_L^H(H) + \sigma_v^2 I_{mL}\,. \tag{1.58}$$

Clearly, the column space of $\mathcal{T}_L(H_N)$ is the signal subspace. Since $\mathcal{G}(G)\mathcal{T}_L(H) = 0$, the column space of $\mathcal{G}^H$ is the noise subspace and G provides a linear parameterization for it.

Consider the eigendecomposition of $R_L^{\mathcal{Y}}$ of which the real positive eigenvalues are ordered in descending order:

$$R_L^{\mathcal{Y}} = \sum_{i=1}^{L+N-1} \lambda_i V_i V_i^H + \sum_{i=L+N}^{mL} \lambda_i V_i V_i^H = V_{\mathcal{S}}\Lambda_{\mathcal{S}}V_{\mathcal{S}}^H + V_{\mathcal{N}}\Lambda_{\mathcal{N}}V_{\mathcal{N}}^H \tag{1.59}$$

where $\Lambda_{\mathcal{N}} = \sigma_v^2 I_{(m-1)L-N+1}$ (see (1.58)). The sets of eigenvectors $V_{\mathcal{S}}$ and $V_{\mathcal{N}}$ are orthonormal: $V_{\mathcal{S}}^H V_{\mathcal{N}} = 0$, and $\lambda_i > \sigma_v^2$, $i = 1, \ldots, L{+}N{-}1$. We then have the following equivalent descriptions of the signal and noise subspaces

$$\begin{aligned} \mathrm{SSS} &= \mathit{Range}\{V_{\mathcal{S}}\} = \mathit{Range}\{\mathcal{T}_L(H)\} \\ \mathrm{NSS} &= \mathit{Range}\{V_{\mathcal{N}}\} = \mathit{Range}\{\mathcal{G}(G)^H\}\,. \end{aligned} \tag{1.60}$$

The noise subspace parameterization that we consider here is prediction based. It can perhaps be more easily expressed in the frequency domain by noting that

$$\mathrm{F}^b(z)\mathrm{H}(z) = 0\,,\ \ \mathrm{F}^b(z) = h^{\perp H}\mathrm{P}(z)\,. \tag{1.61}$$

G above is not the same but is related to this choice of $\mathbf{F}^b(z)$. Another set of blocking equalizers and hence another linear parameterization of the noise subspace is channel based, e.g.

$$\mathbf{F}^b(z) = \begin{bmatrix} \mathrm{H}_2(z) & -\mathrm{H}_1(z) & & 0 \\ \vdots & & \ddots & \\ \mathrm{H}_m(z) & 0 & & -\mathrm{H}_1(z) \end{bmatrix} \qquad (1.62)$$

Many other choices of $\mathbf{F}^b(z)$ are possible involving other pairs of channels. It is also possible to consider more than $m-1$ blocking equalizers, possibly involving up to all $\frac{m(m-1)}{2}$ possible pairs of channels.

3.4 Channel Estimation from Second-Order Statistics

Linear Prediction from Denoised Statistics

σ_v^2 can again be identified as the smallest eigenvalue of R_L^y. Replacing R_L^y by $\mathrm{R}_L^y - \sigma_v^2 I_{mL}$ gives us the covariance matrix for noise-free data, to which the prediction techniques discussed previously can be applied.

Signal Subspace Fitting

Consider now the following subspace fitting problem

$$\min_{\boldsymbol{H},T} \|\mathcal{T}_L(\boldsymbol{H}) - V_S T\|_F \ . \qquad (1.63)$$

The optimal transformation matrix T can again be found to be

$$T = V_S^H \mathcal{T}_L(\boldsymbol{H}) \ . \qquad (1.64)$$

Using (1.64) and the commutativity of the convolution operator, one can show that (1.63) is equivalent to

$$\begin{aligned} &\min_{\boldsymbol{H}^t} \boldsymbol{H}^t \left(\sum_{i=L+N}^{mL} \mathcal{T}_N\left(V_i^{H\,t}\right) \mathcal{T}_N^H\left(V_i^{H\,t}\right) \right) \boldsymbol{H}^{t\,H} \\ &= \min_{\boldsymbol{H}^t} \left[L \left\|\boldsymbol{H}^t\right\|_2^2 - \boldsymbol{H}^t \left(\sum_{i=1}^{L+N-1} \mathcal{T}_N\left(V_i^{H\,t}\right) \mathcal{T}_N^H\left(V_i^{H\,t}\right) \right) \boldsymbol{H}^{t\,H} \right] \end{aligned} \qquad (1.65)$$

where V_i^H 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 $\boldsymbol{H}^t$. In case we choose the quadratic constraint $\|\boldsymbol{H}^t\|_2 = 1$, $\boldsymbol{H}^t$ is found as the minimum eigenvector of the first matrix in brackets in (1.65). This solution reflects orthogonalization of parameterized SSS and estimated NSS. Alternatively, the last term in (1.65) leads equivalently to

$$\max_{\|\boldsymbol{H}^t\|_2=1} \boldsymbol{H}^t \left(\sum_{i=1}^{L+N-1} \mathcal{T}_N \left(V_i^{H\,t}\right) \mathcal{T}_N^H \left(V_i^{H\,t}\right) \right) \boldsymbol{H}^{t\,H} \tag{1.66}$$

the solution of which is the eigenvector corresponding to the maximum eigenvalue of the matrix appearing between the brackets. This solution reflects really the attempt to fit parameterized and estimated SSS's.

Noise Subspace Fitting

Alternatively we may work with the parameterized noise subspace and consider the following subspace fitting approach

$$\min_{\boldsymbol{F}^b, T} \left\| \mathcal{T}\left(\boldsymbol{F}^b\right)^H - V_{\mathcal{N}}\, T \right\|_F . \tag{1.67}$$

One choice would be

$$\min_{G,T} \left\| \mathcal{G}\,(G)^H - V_{\mathcal{N}}\, T \right\|_F . \tag{1.68}$$

Again, two possible solutions can be obtained, depending on whether we attempt to orthogonalize the parameterized NSS to the estimated SSS or to fit it to the estimated NSS. The choice of $\boldsymbol{F}^b$ as in (1.62) corresponds to Xu's deterministic least-squares channel identification approach.

3.5 Channel Estimation from Data: ML

The transmitted symbols a_k are considered deterministic, the stochastic part is considered to come only from the additive Gaussian white noise. We assume the data $\boldsymbol{Y}_M(k)$ to be available. The maximization of the likelihood function boils down to the following least-squares problem

$$\min_{\boldsymbol{H}, A_{M+N-1}(k)} \|\boldsymbol{Y}_M(k) - \mathcal{T}_M(\boldsymbol{H})\, A_{M+N-1}(k)\|_2^2 . \tag{1.69}$$

The optimization problem in (1.69) is separable. Eliminating $A_{M+N-1}(k)$ in terms of $\boldsymbol{H}$, we get

$$\min_{\boldsymbol{H}} \left\| P^{\perp}_{\mathcal{T}_M(\boldsymbol{H})} \boldsymbol{Y}_M(k) \right\|_2^2 \tag{1.70}$$

subject to a nontriviality constraint on $\boldsymbol{H}$. In order to find an attractive iterative procedure for solving this optimization problem, we should work with a minimal parameterization of the noise subspace, which we have obtained before. Indeed,

$$P^{\perp}_{\mathcal{T}_M(\boldsymbol{H})} = P_{\mathcal{G}_M^H(G)} \ . \tag{1.71}$$

The number of degrees of freedom in H and G is both $mN-1$ (the proper scaling factor cannot be determined). So H can be uniquely determined from G and vice versa. Hence, we can reformulate the optimization problem in (1.70) as

$$\min_{G} \left\| P_{\mathcal{G}_M^H(G)} \boldsymbol{Y}_M(k) \right\|_2^2 \ . \tag{1.72}$$

Due to the (almost) block Toeplitz character of $\mathcal{G}_M$, the product $\mathcal{G}_M \boldsymbol{Y}_M(k)$ represents a convolution. Due to the commutativity of convolution, we can write $\mathcal{G}_M(G)\boldsymbol{Y}_M(k) = \mathcal{Y}_N(\boldsymbol{Y}_M(k))[1 \quad G^H]^H$ for some properly structured $\mathcal{Y}_N(\boldsymbol{Y}_M(k))$. This leads us to rewrite (1.72) as

$$\min_{G} \begin{bmatrix} 1 \\ G \end{bmatrix}^H \mathcal{Y}_N^H(\boldsymbol{Y}_M(k)) \left(\mathcal{G}_M(G)\mathcal{G}_M^H(G)\right)^{-1} \mathcal{Y}_N(\boldsymbol{Y}_M(k)) \begin{bmatrix} 1 \\ G \end{bmatrix} \tag{1.73}$$

This optimization problem can now easily be solved iteratively in the classical IQML style. An initial estimate may be obtained from the subspace fitting approach discussed above. Such an initial estimate is consistent and hence one iteration of (1.73) will be sufficient to generate an estimate that is asymptotically equivalent to the global optimizer of (1.73). Cramer-Rao bounds have been obtained and analyzed in [3]. The choice of the noise subspace parameterization in (1.62) using all pairs of channels leads to Yingbo Hua's ML method. More discussion on the ML method can be found in [11],[12].

3.6 Constrained IIR Filter DFE

Here we consider an equalizer structure with decision feedback. The approach is in fact a multichannel extension of the adaptive notch filter approach for sinusoids in noise. As a consequence, the method will continue to work well even if the additive noise and/or the transmitted symbols are colored.

Let $\mathrm{P}_{\widehat{y}_L}(z)$ and $\mathrm{P}_{\widetilde{y}_L}(z)$ be the z-transforms of the forward prediction and prediction error filters (of the noise-free case) so that $\mathrm{P}_{\widetilde{y}_L}(z) = I_m - z^{-1}\mathrm{P}_{\widehat{y}_L}(z)$. To alleviate the notation, $\mathbf{P}(z)$ will continue to represent $\mathrm{P}_{\widetilde{y}_L}(z)$ (as before). Since $\mathbf{P}(z)\mathbf{H}(z) = \boldsymbol{h}(0)$, the noise-free received vector signal $\boldsymbol{y}(k) = \mathbf{H}(q)a_k$, which is a multichannel MA process, is also a (singular) multichannel AR process: $\mathbf{P}(q)\boldsymbol{y}(k) = \boldsymbol{h}(0)a_k$. For the noisy received signal $\boldsymbol{y}(k) = \mathbf{H}(q)a_k + \boldsymbol{v}(k)$, we get

$$\mathbf{P}(q)\boldsymbol{y}(k) = \boldsymbol{h}(0)a_k + \mathbf{P}(q)\boldsymbol{v}(k) \tag{1.74}$$

which is a constrained multichannel ARMA process, apart from the term $\boldsymbol{h}(0)a_k$, which will require detection. In the scalar case, the prediction error filter is minimum-phase. For the multichannel case, the extension is that $\det[\mathbf{P}(z)]$ is minimum-phase, even in the singular case. So we can recover $\boldsymbol{v}(k)$ as follows:

$$\boldsymbol{v}(k) = \mathbf{P}^{-1}(q)[\mathbf{P}(q)\boldsymbol{y}(k) - \boldsymbol{h}(0)a_k]\,. \tag{1.75}$$

This can be more straightforwardly implemented by the following procedure

$$\left\{\begin{array}{l} s(k) = \mathrm{P}_{\widetilde{y}_L}(q)\boldsymbol{y}(k) - \mathrm{P}_{\widehat{y}_L}(q)\widehat{\boldsymbol{v}}(k-1) \\ \widehat{a}_k = \mathrm{dec}[\frac{\boldsymbol{h}^H(0)}{\boldsymbol{h}^H(0)\boldsymbol{h}(0)}s(k)] \\ \widehat{\boldsymbol{v}}(k) = s(k) - \boldsymbol{h}(0)\widehat{a}_k \end{array}\right. \tag{1.76}$$

where dec denotes the decision operation, whose argument is ideally $a_k + \frac{\boldsymbol{h}^H(0)}{\boldsymbol{h}^H(0)\boldsymbol{h}(0)}\boldsymbol{v}(k)$. Various algorithms are now possible to adapt the coefficients $\mathrm{P}_{\widehat{y}_L}$ such as the Recursive Prediction Error Method and its simplifications.

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 Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas". In *Proc. 28th Asilomar Conf. on Signals, Systems and Computers*, pages 1154–1158, Pacific Grove, CA, Oct. 31 - Nov. 2 1994.

[3] D.T.M. Slock and C.B. Papadias. "Blind Fractionally-Spaced Equalization Based on Cyclostationarity". In *Proc. Vehicular Technology Conf.*, pages 1286–1290, Stockholm, Sweden, June 1994.

[4] 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.

[5] D. Kundu and A. Mitra. "Estimating the parameters of exponentially damped/undamped sinusoids in noise: A non-iterative approach". *Signal Processing (EURASIP)*, 46:363–368, 1995.

[6] 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.

[7] Y.A. Rozanov. *Stationary Random Processes*. Holden-Day, San Fransisco, 1967. *Translated by* A. Feinstein.

[8] B. Porat. *Contributions to the Theory and Applications of Lattice Filters*. PhD thesis, Stanford University, Stanford, CA, Aug. 1982.

[9] K. Gallivan, S. Thirumalai, and P. Van Dooren. "A Block Toeplitz Look-Ahead Schur Algorithm". In *Proc. 3rd International Workshop on SVD and Signal Processing*, Leuven, Belgium, Aug. 22-25 1994.

[10] 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.

[11] D.T.M. Slock and C.B. Papadias. "Further Results on Blind Identification and Equalization of Multiple FIR Channels". In *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.

[12] D.T.M. Slock. "Spatio-Temporal Training-Sequence-Based Channel Equalization and Adaptive Interference Cancellation". In *Proc. ICASSP 96 Conf.*, Atlanta, Georgia, May 1996.