ON THE EQUIVALENCE OF BLIND EQUALIZERS BASED ON MRE AND SUBSPACE INTERSECTIONS

Alle-Jan van der Veen[1], David Gesbert[2], and Arogyaswami Paulraj[2]

[1] Delft Univ. of Technology, Dept. Elec. Eng./DIMES, 2628 CD Delft, The Netherlands
[2] Information Systems Laboratory, Stanford University, Stanford, CA 94305

Two classes of algorithms proposed for the blind equalization of multiple channels driven by a single source are the Mutually Referenced Equalizers (MRE) method by Gesbert *et al.*, and the Subspace Intersection (SSI) method by Van der Veen *et al*. Although these methods seem at first sight unrelated, we show here that a variant of the SSI method and a particular member in the class of MRE methods provide mathematically identical solutions.

1. INTRODUCTION

Blind equalization has been an active research area during the last two decades, but especially during the last few years. Two major factors appear to drive the wide interest in this topic. Firstly, there is an increasing number of interesting and promising applications in the area of digital communications, wireless or not. Secondly, the fact was recognized that channel oversampling, either temporally (fractionally spaced equalizers) or in space (antenna arrays), offers several new leverages for solving the blind equalization problem, and thus enhances its applicability.

By oversampling an analog channel carrying a single discrete source signal, this signal is effectively seen through several parallel linear channels. These may be the polyphase representations of a single physical channel or correspond to different physical links in the case of multiple antennas, or combinations of the two. In any case, we obtain a single-input, multiple-output (SIMO) channel model.

From an algebraic perspective, oversampling leads to a low-rank model for the output vector signal. This has been extensively exploited in the so-called second-order statistics and algebraic methods for the SIMO identification problem [1]. At least three classes can be identified. The first addresses the problem of estimating the channels, viz. e.g. [2–4], the second considers the estimation of channel inverses (equalizers) [5–7], and the third attempts to recover the transmitted symbols directly from a (typically small) batch of output samples without resorting to channel/equalizer estimates [8,9].

Categories 2 and 3 have the advantage of by-passing the channel estimation and inversion step, which in cases results in increased robustness. The direct symbol-estimation methods [8,9] have sometimes been called row-span methods as they exploit the row-span information of the data matrix to find the vector of unknown symbols. Following a seemingly different strategy, MRE techniques [5] estimate a collection of channel equalizers by forcing them to produce the same (unknown) output sequence, up to fixed equalization lags. As we show here, these two methods are in fact more or less identical, with small differences depending on the implementation.

We first review the row-span method of [8]. The symbol estimates produced by this technique can be regarded as the outputs of linear equalizers, averaged across all equalization lags. We show that these equalizers optimize a *maximal coherence (MC)* criterion. Finally, we demonstrate the equivalence between the MC criterion and a particular member in the class of MRE criteria.

2. DATA MODEL

2.1. Data matrices

Suppose that a digital symbol sequence $[s_i]$ is transmitted through a medium and received by an array of M sensors. The received signals are sampled at a rate which is P times faster than the symbol rate, here normalized to $T = 1$. Hence, during each symbol period, a total of MP scalar measurements are available, which can be stacked into MP-dimensional vectors $\mathbf{x}_i$ as

$$\mathbf{x}_i = \begin{bmatrix} x_i^1 \\ \vdots \\ x_i^{MP} \end{bmatrix}.$$

Assuming an FIR channel, we can model $\mathbf{x}_i$ as the output of an MP-dimensional vector channel with impulse response $[\mathbf{h}_0, \mathbf{h}_1, \cdots, \mathbf{h}_{L-1}]$, where L denotes the channel length in symbols. In the noise free case, a sample vector $\mathbf{x}_i$ is then given by the convolution

$$\mathbf{x}_i = \sum_{k=0}^{L-1} \mathbf{h}_k s_{i-k}. \tag{1}$$

Consider a finite block of data and define the $mMP \times N$ block-Toeplitz data matrix

$$\mathcal{X}^{(i)} = \begin{bmatrix} \mathbf{x}_i & \mathbf{x}_{i+1} & \ddots & \mathbf{x}_{i+N-1} \\ \mathbf{x}_{i-1} & \mathbf{x}_i & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots \\ \mathbf{x}_{i-m+1} & \ddots & \ddots & \ddots \end{bmatrix}.$$

N is the block length, while m can be interpreted as the memory of an equalizer acting on the rows of $\mathcal{X}^{(i)}$. Let $n = L+m-1$. From (1), $\mathcal{X}^{(i)}$ has a factorization as $\mathcal{X}^{(i)} = \mathcal{H}\mathcal{S}^{(i)}$, where $\mathcal{H}$ is the $mMP \times n$ channel matrix,

$$\mathcal{H} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \\ \mathbf{0} & & \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} \end{bmatrix} \qquad (2)$$

and

$$\mathcal{S}^{(i)} = \begin{bmatrix} s_i & s_{i+1} & \ddots & s_{i+N-1} \\ \ddots & \ddots & \ddots & \ddots \\ s_{i-n+1} & \ddots & \ddots & \ddots \end{bmatrix}.$$

We will assume that $\mathcal{H}$ is full rank and that m and N are sufficiently large, so that this is a low rank factorization ($mMP \geq L+m-1 \leq N$). In that case, we can recover any row of $\mathcal{S}^{(i)}$ by taking linear combinations of the rows of $\mathcal{X}^{(i)}$.

2.2. Equalizers

An equalizer with delay k acting on $\mathcal{X}^{(i)}$ tries to reconstruct the $k+1$-st row of $\mathcal{S}^{(i)}$:

$$\mathbf{w}_k^* \mathcal{X}^{(i)} = [s_{i-k} \quad s_{i-k+1} \quad \cdots].$$

See figure 1(a). Since $\mathcal{S}^{(i)}$ has n rows, there is a total of n possible delays, and hence there are n different equalizers $\mathbf{w}_k$ ($k = 0, \cdots, n-1$). Note in particular that

$$\mathbf{w}_i^* \mathcal{X}^{(i)} = [s_0 \quad s_1 \quad \cdots]$$

so that

$$\mathbf{w}_i^* \mathcal{X}^{(i)} = \mathbf{w}_k^* \mathcal{X}^{(k)}, \qquad i, k = 0, \cdots, n-1.$$

Hence we can make the n equalizers all produce the same output sequence by properly delaying the data sequences.

2.3. Normalization

If m is sufficiently large, then $\mathcal{X}^{(i)}$ is rank deficient. This is a source of non-uniqueness for the equalizers $\{\mathbf{w}_k\}$, since any vector from the left null space of $\mathcal{X}^{(i)}$ may be added. Optimization algorithms avoid this by asking for equalizers of

minimal norm (this also helps in reducing noise amplification). Alternatively, we can achieve the same effect by defining equalizers to act on a minimal basis for the row span of $\mathcal{X}^{(i)}$, rather than $\mathcal{X}^{(i)}$ itself. Thus introduce the SVDs:

$$\mathcal{X}^{(i)} = U_i \Sigma_i V^{(i)}, \qquad i = 0, \cdots, n-1.$$

If $\mathcal{X}^{(i)}$ has rank n, then U_i has n orthonormal columns, $V^{(i)}$ has n orthonormal rows, and Σ_i is a diagonal matrix containing the n nonzero singular values. The rows of $V^{(i)}$ form an orthonormal basis for the row span of $\mathcal{X}^{(i)}$. A "normalized" equalizer acting on $V^{(i)}$ is called $\mathbf{t}_i$, which is related to $\mathbf{w}_i$ via

$$\mathbf{t}_i = \Sigma_i U_i^* \mathbf{w}_i$$

hence also satisfies

$$\begin{aligned} \mathbf{t}_i^* V^{(i)} &= [s_0 \quad s_1 \quad \cdots], & (3) \\ \mathbf{t}_i^* V^{(i)} &= \mathbf{t}_k^* V^{(k)}, & i, k = 0, \cdots, n-1. \end{aligned}$$

2.4. Super-equalizers

Define

$$X_T = \begin{bmatrix} \mathcal{X}^{(0)} \\ \vdots \\ \mathcal{X}^{(n-1)} \end{bmatrix}, \qquad V_T = \begin{bmatrix} V^{(0)} \\ \vdots \\ V^{(n-1)} \end{bmatrix}. \qquad (4)$$

"Super-equalizers" are long vectors that collect several equalizers with different delays, each reconstructing the same sequence $[s_0 \quad s_1 \quad \cdots]$. They act on the data X_T or on the normalized data V_T, respectively:

$$\mathbf{w}^* = [\mathbf{w}_0^* \cdots \mathbf{w}_{n-1}^*], \qquad \mathbf{t}^* = [\mathbf{t}_0^* \cdots \mathbf{t}_{n-1}^*].$$

Hence the super-equalizer combines the outputs of the regular equalizers, forming an average over all admissible delays. (By itself, it can also be interpreted as an ordinary equalizer of length $n+m-1$ at delay $n-1$.) See figure 1(b). There is an issue of how to weight the outputs of each equalizer and combine them in an optimal fashion.

3. BLIND EQUALIZATION

3.1. Subspace intersection method

From an algebraic perspective, the problem of blind equalization is, for given a data matrix $\mathcal{X}$, to find a factorization $\mathcal{X} = \mathcal{H}\mathcal{S}$ where $\mathcal{S}$ meets the required Toeplitz structure. Since a Toeplitz matrix is generated by a single vector in a linear way, this translates to finding

$$\mathbf{s} = [s_0 \quad s_1 \quad \cdots \quad s_{N-1}]$$

such that

$$\begin{aligned} \mathbf{s} &\in \text{row}(\mathcal{X}^{(0)}) \\ \mathbf{s} &\in \text{row}(\mathcal{X}^{(1)}) \\ &\vdots \\ \mathbf{s} &\in \text{row}(\mathcal{X}^{(n-1)}) \end{aligned}$$
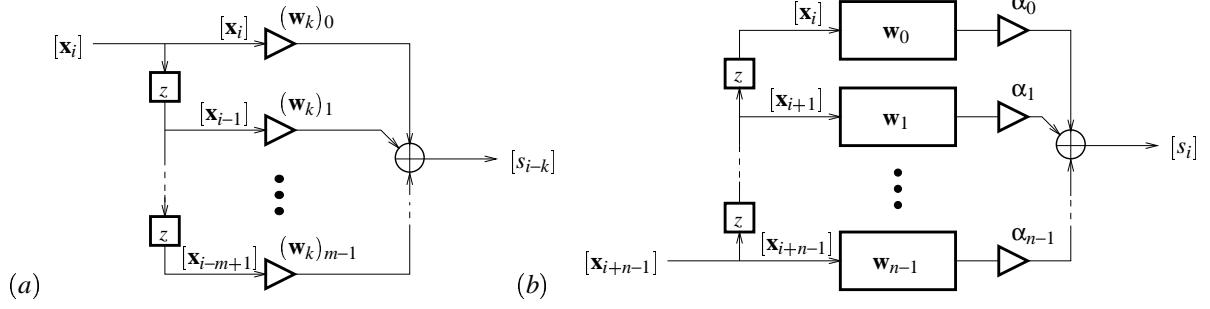
Fig. 1. (*a*) Equalizer with delay k; (*b*) Super-equalizer, combining the outputs of several equalizers at different delays

where 'row($\cdot$)' stands for the row span. Thus, we have to find a single (hopefully unique) vector **s** which is in the intersection of all n subspaces.

Numerically, there are several ways to compute the intersection. The usual way is to construct the union of the complement of all subspaces, and take the complement again. The problem with this is that the complementary spaces can be highly dimensional. It was proven in [8] that, since the rows of $V^{(i)}$ form a minimal and *orthonormal* basis for row($\mathcal{X}^{(i)}$), the same result can be obtained by constructing the matrix V_T in (4) and looking for the right singular vector corresponding to the largest singular values of V_T. In the noise-free case it is equal to the vector in the intersection; with noise perturbations, we find a sequence that "best" fits all subspaces. The corresponding left singular vector can be interpreted as the equalizer that returns this sequence.

In particular, it is proven in [8] that, if $\mathbf{t}_{ssi}$ is the principal left singular vector of V_T and $n = L + m - 1$, then

$$\mathbf{t}_{ssi}^* V_T = \alpha[s_0 \quad s_1 \; \cdots \; s_{N-1}]$$

where α is some nonzero scalar that makes the output sequence have norm 1. The reason, essentially, is that because of the normalization, the largest singular value of V_T is bounded by $\sqrt{n}$. This bound is attained when $\mathbf{t}_{ssi}^* = [\mathbf{t}_0^* \cdots \mathbf{t}_{n-1}^*]$ where each component by itself is an equalizer on the normalized signals (viz. (3)), returning a multiple α_i of $[s_0 \quad s_1 \; \cdots]$. In fact, one can show that all scalings α_i will be the same.

Thus, $\mathbf{t}_{ssi}$ is a super-equalizer in the sense of section 2.4. The corresponding equalizer on unnormalized data X_T is denoted by $\mathbf{w}_{ssi}$, related to $\mathbf{t}_{ssi}$ via

$$\mathbf{w}_{ssi} = [\mathbf{w}_0^* \cdots \mathbf{w}_{n-1}^*]^*, \qquad \mathbf{w}_i = U_i \Sigma_i^{-1} \mathbf{t}_i. \quad (5)$$

3.2. Maximal coherence criterion

The principal left singular vector $\mathbf{t}_{ssi}$ of V_T can also be expressed in terms of a criterion on the unnormalized received data. Indeed, $\mathbf{t}_{ssi}$ can be written as

$$\mathbf{t}_{ssi} = \arg \max_{\|\mathbf{u}\|^2=1} \mathbf{u}^* \mathcal{R}_V \mathbf{u}$$

where $\mathcal{R}_V = V_T V_T^*$. Define the (empirical) correlation matrices

$$R_{i,j} = \mathcal{X}^{(i)} \mathcal{X}^{(j)^*},$$

$$\mathcal{R}_X = X_T X_T^* = \begin{bmatrix} R_{0,0} & \cdots & R_{0,n-1} \\ \vdots & & \vdots \\ R_{n-1,0} & \cdots & R_{n-1,n-1} \end{bmatrix},$$

$$\mathcal{R}_0 = \begin{bmatrix} R_{0,0} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & R_{n-1,n-1} \end{bmatrix}.$$

Then $\mathcal{R}_X = \mathcal{R}_0^{1/2} \mathcal{R}_V \mathcal{R}_0^{1/2}$, where

$$\mathcal{R}_0^{1/2} = \begin{bmatrix} R_{0,0}^{1/2} & & 0 \\ & \ddots & \\ \mathbf{0} & & R_{n-1,n-1}^{1/2} \end{bmatrix}, \qquad R_{i,i}^{1/2} := U_i \Sigma_i.$$

It follows that $\mathbf{w}^* \mathcal{R}_X \mathbf{w} = \mathbf{u}^* \mathcal{R}_V \mathbf{u}$ for $\mathbf{w} = \mathcal{R}_0^{-1/2*} \mathbf{u}$. Now denote by $\mathbf{w}_{ssi}$ the corresponding super-equalizer provided by the SSI method (related to $\mathbf{t}_{ssi}$ as in equation (5). By substitution, $\mathbf{w}_{ssi}$ is found to optimize the following constrained criterion:

$$\mathbf{w}_{ssi} = \arg \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w}=1} \mathbf{w}^* \mathcal{R}_X \mathbf{w} = \arg \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w}=1} J_{ssi}, \quad (6)$$

where J_{ssi} is given by

$$J_{ssi} := \| \sum_{i=0}^{n-1} \mathbf{w}_i^* \mathcal{X}^{(i)} \|^2$$

and the constraint can be written as

$$\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = \sum_{i=0}^{n-1} \| \mathbf{w}_i^* \mathcal{X}^{(i)} \|^2 = 1. \quad (7)$$

Thus, the subspace intersection solution is also obtained by maximizing the power of the sum of all equalizers outputs,

subject to the constraint that the sum of the powers is kept to a constant. The SSI method tends to maximize the coherence of the equalizers outputs. Indeed, in the noise-free case, all equalizers return the *same* output sequence $[s_0 \; s_1 \; \cdots]$, up to a common scaling. Note that this result holds true only in the case of the constraint specified in (7).

3.3. The MRE method

The idea behind the mutually referenced equalizer (MRE) method for blind equalization [5] is to find a vector of n equalizers $\mathbf{w} = [\mathbf{w}_0^* \; \cdots \; \mathbf{w}_{n-1}^*]^*$ that optimizes[1]

$$\min_{\mathbf{w}} \sum_{i=0}^{n-2} \|\mathbf{w}_i^* \mathcal{X}^{(i)} - \mathbf{w}_{i+1}^* \mathcal{X}^{(i+1)}\|^2 . \tag{8}$$

Each $\mathbf{w}_i$ plays the role of an equalizer with delay i and serves as a training for the next equalizer. The criterion provides a collection of exact channel inverses in the noise free case [5]. To avoid trivial solutions, $\mathbf{w}$ should be constrained, e.g. by fixing one of its entries or its norm. Another suitable constraint is one that keeps the sum of output powers to a constant, $\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1$. The motivation for this particular choice is that it avoids trivial null space solutions $\mathbf{w}_i^* \mathcal{X}^{(i)} = \mathbf{0} \; \forall i$, which is necessary at least in the noise free case.

The MRE method can be extended to look at *all* available cross-differences [5], i.e. to solve

$$\mathbf{w}_{mre} := \arg \min_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} \sum_{i=0}^{n-1} \sum_{k=0}^{n-1} \|\mathbf{w}_i^* \mathcal{X}^{(i)} - \mathbf{w}_k^* \mathcal{X}^{(k)}\|^2 . \tag{9}$$

Elaborating on this expression, we find

$$
\begin{aligned}
J_{mre} &:= \sum_{i=0}^{n-1} \sum_{k=0}^{n-1} \|\mathbf{w}_i^* \mathcal{X}^{(i)} - \mathbf{w}_k^* \mathcal{X}^{(k)}\|^2 \\
&= \sum_{i=0}^{n-1} \sum_{k=0}^{n-1} \mathbf{w}_i^* R_{ii} \mathbf{w}_i + \mathbf{w}_k R_{kk} \mathbf{w}_k - \mathbf{w}_i^* R_{ik} \mathbf{w}_k - \mathbf{w}_k^* R_{ki} \mathbf{w}_i \\
&= 2\mathbf{w}^* \begin{bmatrix} (n-1)R_{0,0} & -R_{0,1} & \cdots & -R_{0,n-1} \\ -R_{1,0} & (n-1)R_{1,1} & & \cdot \\ \vdots & & & \vdots \\ -R_{n-1,0} & \cdots & & (n-1)R_{n-1,n-1} \end{bmatrix} \mathbf{w}.
\end{aligned}
$$

It thus follows that

$$J_{mre} + 2J_{ssi} = 2n \mathbf{w}^* \mathcal{R}_0 \mathbf{w} .$$

Under the constraint $\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1$, we finally obtain

$$\min_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{mre} = 2n - \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{ssi} ,$$

so that

$$\mathbf{w}_{mre} = \mathbf{w}_{ssi} .$$

Hence we conclude that the SSI method and the extended MRE method under output power constraint provide identical solutions.

---

[1] In [5], the MRE criterion is presented in a stochastic setting. A finite sample implementation of the method leads to (8).

3.4. Remarks

The SSI method as described here is slightly different from that in [8]. There, the sequence to be recovered was extended with additional tail symbols, which changed the definition of V_T in such a way that only a single matrix $V^{(0)}$ was needed. The advantage of this is that only a single data matrix has to be normalized, leading to substantial computational savings. Otherwise, the two variants of the method lead to similar results that are asymptotically identical. The variant as presented here was chosen for expository reasons.

With noise, the SSI method on normalized data V_T and on original data X_T are slightly different. The reason is that, with noise, each $\mathcal{X}^{(i)}$ is always full rank, whereas $V^{(i)}$ is presumably obtained from a truncated SVD, resulting in an approximate n-dimensional basis for the row span of $\mathcal{X}^{(i)}$. If we omit the truncation, i.e. define $V^{(i)}$ to contain all mMP right singular vectors of $\mathcal{X}^{(i)}$, then the maximum coherence solution is exactly equal to the SSI method on V_T.

4. REFERENCES

[1] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Processing*, vol. 50, no. 1-2, pp. 83–99, Apr. 1996.

[2] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[3] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Information Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[4] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind equalization," *IEEE Trans. Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[5] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2307–2317, Sept. 1997.

[6] G. Giannakis and S. Halford, "Blind fractionally spaced equalization of noisy FIR channels: Direct and adaptive solutions," *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2277–2292, Sept. 1997.

[7] D.T.M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. ICASSP 94 Conference*, Adelaide, Australia, Apr. 1994, pp. IV.585–588.

[8] A.J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 173–190, Jan. 1997.

[9] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2714–2723, Nov. 1995.