METHODS FOR BLIND EQUALIZATION VIA JOINT DIAGNOSTIC

Xiang-yang Zhuang and A. Lee Swindlehurst

Department of Electrical and Computer Engineering
Brigham Young University, Provo, UT 84602
e-mail:{jeffz, swindle}@ee.byu.edu

Abstract

This paper addresses the problem of blind equalization in a multi-user multipath environment. Based on a generalized model in which different time-shifted versions of the signal are regarded as separate sources, the blind source separation methods JADE and ACM (which both employ a joint diagonalization operation) can be directly applied. An MMSE criterion is introduced to the eigendecomposition pre-whitening step of JADE to reduce noise enhancement. Because of the model, all significant replicas of the signals are resolved and the “best” delayed version is selected for each user. Unlike many algorithms which use the structural properties of the channel or data matrix, our unstructured blind methods are insensitive to channel order estimation and robust to ill-conditioned channels. The performance of JADE and ACM is compared using simulations and their respective computational complexities are also analyzed. Simulations exhibit promising results compared with those of the subspace method and the training based MMSE Wiener filter.

1. INTRODUCTION

Blind equalization in a wireless environment has recently been of considerable research interest. Adaptive and block methods have been proposed using diversity obtained by either oversampling or use of an antenna array (or both). In the familiar model $X = HS + N$, $H$ and $S$ are structured matrices, an important property that is exploited by most block methods. In particular, $H$ is a Sylvester matrix and $S$ is (block) Toeplitz where each row is a time-shifted version of a user sequence. In the general approach presented in this paper, we treat these rows as sequences from “different” users and we assume $H$ is unstructured. Under these assumptions, the model for $X$ is equivalent to that used in the standard blind source separation problem, where $H$ is treated as a generic “mixing” matrix.

An unstructured blind approach can be advantageous in situations where the implied structure is not precisely known. For example, in practice the channel length can often not be accurately determined. Errors in specifying the correct model order can cause significant degradation in the performance of methods that exploit the matrix structure of $H$ and $S$. Furthermore, structured methods are especially sensitive to an ill-conditioned channel matrix, which may be encountered when the channel order is overestimated. The methods proposed here are much more robust to these situations. In equalization, the delay to which the equalizer converges plays a very important role in symbol estimation performance, even for the training-based Wiener filter. Since the received signal consists of delayed components with different powers, when an equalizer tries to recover a weak replica with a certain delay, it has to suppress stronger paths and may also amplify the noise. The importance of delay has been observed in constant modulus algorithms (CMA), but there is no built-in control over which delay CMA will converge to, though some re-initialization and channel-surfing techniques have been proposed [6]. Our unstructured methods are able to easily choose the “best delay” since they resolve all significant signal replicas, although at the price of increased complexity.

In this paper, JADE (Joint Approximate Diagonalization of Eigenmatrices) [5] and ACM (Analytical Constant Modulus) [1] are two particular source separation algorithms that are modified and applied to the equalization problem. Source separation algorithms often assume a priori knowledge of the number of sources. In our case, the number depends on the channel order estimate. So some rank detection algorithm (e.g. MDL) is needed first. Then in JADE, we introduce an MMSE prewhitening step to reduce noise amplification. As MSE is contributed by both residual ISI and noise enhancement, we will analyze residual ISI and noise enhancement for different delays. Though higher order statistics are involved in JADE, it is still data efficient (about 100 samples seem to be enough in our simulations). When the ACM algorithm is applied, we may save some computation. The selection of the “best delay” for these methods is also discussed.

2. DATA MODEL

Each element in an $M$-sensor array collects data with an oversampling of factor $P$ to obtain $MP$ sub-channels. All sub-channel output samples at one instant are collected into an $MP$-tuple vector $x_{MP}(n) = [x_{MP}(n), x_{MP}(n+1), \ldots, x_{MP}(n+L-1)]^T$ such that

$$x_{MP}(n) = [h_0 \cdots h_{L-1}]_{MP \times L} \begin{bmatrix} s(n) \\ \vdots \\ s(n-L+1) \end{bmatrix} + n_{MP}(n) \tag{1}$$

If a temporal equalizer with $E$ taps is implemented for each of the $MP$ sub-channels, the model will be written as

$$x(n) = [x_{MP}(n) \cdots x_{MP}(n-E+1)]^T$$
JADE is well-known. The convergence proof of CMA requires an i.i.d. assumption as joined diagonalization minimizes a particular subset of cross terms. The real reason cross-cumulants should be zero. Criterion (4) actually minimizes outputs are truly independent sources, all their cross-cumulants should be zero. Criterion (4) actually minimizes a particular subset of cross terms. The real reason is that “it allows for an efficient optimization by means of joint diagonalization”.

Since a legitimate unmixing matrix is \( B = H^T \), the JADE algorithm is actually targeted at estimating the three components of the SVD of \( H = UV \), among which \( U \) and \( \Lambda \) are estimated in the prewhitening step realized by an eigenstructure decomposition of the sample covariance matrix. Denoting the whitening matrix (also known as Mahalanobis transform) by \( W_w \), we have the whitened data as \( (\hat{z}^T \hat{x})(\hat{z}^T \hat{x})^H \) where \( \hat{z}^T \hat{x} = W_w \hat{x} \) (5)

The remaining \( V \) component is obtained as a unitary matrix which approximately jointly diagonalizes a matrix set \( \bar{N}' = \{ \bar{N}_r \} \) for \( r \leq s \). A candidate set is

\[ \bar{N}' = \{ \text{unvec}(C_z(:,r)) | 1 \leq r \leq r_s \} \] (6)

where \( \text{unvec} \) denotes the inverse matrix vectorization operation and \( C_z \) is an \( r_s \times r_s \) matrix with cross-cumulants as its entries, i.e.,

\[ C_z = \text{Cum} \left( [\hat{z}^T(n) \otimes \hat{z}(n)](\hat{z}^T(n) \otimes \hat{z}(n)) \right)' \] (7)

where \( \otimes \) denotes the matrix Kronecker product. We refer to this set as the whole cumulants set. Since theoretically \( \text{rank}(C_z) = r_s \), we can instead use a smaller set (eigenmatrices set)

\[ \bar{N}_s = \{ \text{unvec}(\xi_j, p_s, 1 \leq r \leq r_s) \} \] (8)

where \( \xi, p_r \) are the \( r_s \) most significant eigenpairs of \( C_z \). Joint diagonalization (JD) is implemented efficiently by iteratively applying successive Givens rotations to minimize off-diagonal entries. The coefficients of the Givens rotation can be solved for in closed form.

Once calculated, the rows of the unmixing matrix \( B \) are used to equalize the channel and recover individual elements of \( y(n) \). In the single user case, the recovered signals are delayed and arbitrarily (complex) scaled versions of each other, and an additional step is required to choose which of the signals should be used for making symbol decisions. If the signals are constant modulus, one way of doing this would be to choose the recovered signal with the smallest modulus variation. Alternatively, a decision-directed approach could be used in which the signal with the smallest post-detection residual is chosen. In the multi-user case, the process of assigning recovered signals to individual users and choosing the best signal for each user is somewhat more involved, but simple ad hoc methods are easily devised.

In the discussion below, we discuss a simple modification to a prewhitening step in JADE that reduces the effect of noise enhancement, especially in situations where the mixing matrix is ill-conditioned. This is of particular interest to the equalization problem, where the channel matrix may have a large condition number due to modeling of the channel length or impulse response tails with values near zero.

### 3.1. MMSE Prewhitening

With noise, the equalizer output is

\[ y = \hat{B}x = \hat{B}h(n) + \hat{B}n \] (9)
The closeness of $\tilde{I} = \hat{B}H$ to a permuted identity matrix reflects the goodness of the ISI and CCI cancellation. The second term above defines the noise amplification. The position of the largest entry in each row of $\tilde{I}$ corresponds to the delay. In JADE, the prewhitening matrix is chosen as

$$W_w = \hat{\Lambda}^{-1} \hat{U} = \text{diag} \left( \frac{1}{\hat{\lambda}_1}, \ldots, \frac{1}{\hat{\lambda}_{r_h}} \right) \hat{U}^H$$

(10)

where the $\hat{\lambda}_i$'s are the singular values of $\hat{H}$ estimated from an eigendecomposition of $R_x$. The power of the output noise for the $i^{th}$ row of $\tilde{I}$ is then

$$\sigma_i^2 = \frac{\hat{\sigma}_i^2}{\lambda_i^2} v_{i1}^2 + \cdots + \frac{\hat{\sigma}_i^2}{\lambda_{r_h}^2} v_{ir_h}^2, \quad (i = 1 \cdots r_h)$$

(11)

where $v_{ij}$ is the $ij$th entry of the unitary matrix $\hat{V}$ obtained via JD. We observe first that $\sigma_i/\lambda_i$, rather than only $\lambda_i$ or $\sigma_i$, contributes to the noise power. The noise enhancement could be large if $\hat{H}$ is nearly singular. Furthermore, we can see that different delays give different noise amplification as different rows of $\hat{V}$ (note $v_{i1}^2 + \cdots + v_{ir_h}^2 = 1$ because of unitary) have different weightings on $\hat{\sigma}_i/\lambda_i$. Some rows will result in better noise enhancement than other rows.

In the MMSE Wiener filter, the optimal filter is

$$b_{opt} = (R^ HT R_x)^{-1} V^ H \text{diag} \left( \frac{\lambda_1}{\lambda_1^2 + \sigma_n^2}, \ldots, \frac{\lambda_{r_h}}{\lambda_{r_h}^2 + \sigma_n^2} \right) U^H$$

(12)

which inspires us to modify the prewhitening matrix as

$$W_w = \text{diag} \left( \frac{\lambda_1}{\lambda_1^2 + \sigma_n^2}, \ldots, \frac{\lambda_{r_h}}{\lambda_{r_h}^2 + \sigma_n^2} \right) U^H$$

(13)

Then the equalizer output noise power for the $i^{th}$ delayed version $(i = 1 \cdots r_h)$ is now

$$\sigma_i^2 = \frac{1}{(\sum_{\lambda_i}^2 + \sigma_n^2)} v_{i1}^2 + \cdots + \frac{1}{(\sum_{\lambda_{r_h}}^2 + \sigma_n^2)} v_{ir_h}^2.$$ 

(14)

Different delays still have different noise amplification, but small $\lambda_i$ will not cause huge noise amplification as in (11). Actually, $\sigma_i^2$ is now bounded above by $1/4$. The price paid is the slight increase in residual ISI and CCI due to the diagonal entries of $(\hat{\lambda}_i/\lambda_i)/(\lambda_i^2 + \sigma_n^2)$ in the following output equation:

$$y = U^H \text{diag} \left( \frac{\lambda_1}{\lambda_1^2 + \sigma_n^2}, \ldots, \frac{\lambda_{r_h}}{\lambda_{r_h}^2 + \sigma_n^2} \right) V s + B n$$

(15)

The smaller $(\hat{\sigma}_i/\lambda_i)$ is, the better $(\hat{\lambda}_i/\lambda_i)/(\lambda_i^2 + \sigma_n^2)$ approaches 1 and consequently the smaller the increase in residual ISI.

The original prewhitening matrix is proposed with the aim of minimizing ISI only, as in zero-forcing equalization. Our MMSE prewhitening has same goal as MMSE equalization which performs much better, especially when the channel matrix is nearly singular or the SNR is low.

In source separation, whenever prewhitening is used, there exists a lower bound for the pairwise source separation rate as derived in [4]:

$$\lim_{\sigma_n \to 0} \frac{r_{PE} + r_{PE}'}{2} > \frac{1}{4}$$

(16)

where the asymptotic separation rate $r_{PE}$ is defined by $N E[\{I_{1n}^2\}] (N \to \infty)$. In our equalization case, the total residual ISI and CCI is easily found to be lower bounded by $r_h/(4N)$.

### 4. BLIND EQUALIZATION VIA ACM

The application of ACM to our model is straightforward. The $r_h$ elements of the vector $s(n)$ in (3) are assumed to have constant modulus. The ACM algorithm exploits the CM property analytically to directly find the unmixing matrix $B$ for all $r_h$ signals. Unlike JADE, it is not necessary that the elements of $s(n)$ be i.i.d., rather it is only required that they be linearly independent. However, at least $N > r_h$ samples must be collected in order to implement the algorithm. After a prewhitening step similar to that used in JADE, an $N-1 \times r_h$ matrix is formed from the data. Under certain conditions, this matrix has, in the noiseless case, an $r_h$-dimensional (right) nullspace. The unmixing matrix for ACM is calculated by jointly diagonalizing a set of linearly independent vectors (appropriately reshaped into square matrices) from this nullspace. As with JADE, in the single user case, each row of the unmixing matrix recovers a delayed and scaled version of the transmitted signal, and some technique is required to find the “best” delay.

As they are block methods, both JADE and ACM are comparatively computationally expensive. Their applications are limited to low-order channels (e.g., such as those assumed for GSM) and small $d$. Denoting $m = MPE$, the total number of real flops for ACM is at least $36mN^2 + 9r_h^4N + O(r_h^6\delta)$, when we resolve $\delta$ outputs. The multiplicative coefficient of $O(\cdot)$ is very large, so the savings obtained by taking $\delta \ll r_h$ may be significant. The total number of real flops for JADE is roughly $(36m^3 + O(m^2N)) + O(r_h^4N) + 36r_h^4 + O(r_h^6)$. If the whole cumulants set is used, the SVD $(3O(r_h^2)$ flops of $C_x$ is saved at the expense of diagonalizing more matrices. The complexity is then $(36m^3 + O(m^2N)) + O(r_h^4N) + O(r_h^6)$ flops.

### 5. SIMULATIONS

We apply MDL for the rank detection in JADE and ACM. MSE is used as a measure after the rotation factor is compensated for. The Wiener filter performance is computed after obtaining the filter (of optimal delay) by training with all $N$ symbols. Results demonstrating the dramatic performance difference for different delays are not included here. The “best” delay is picked by measuring the Averaged Modulus Error (AME) or modulus variance of the output. A decision-directed approach [2] can be cheaply implemented as $R_{req}$ is already computed in our eigendecomposition step.

The single user case is simulated in Fig 1 under a three-ray channel whose impulse response is truncated by a window of width $4T$. The simulation parameters were $M = 2$.
sensors, an oversampling factor of $P = 2$, and temporal equalizer tap length of $E = 2$. Thus $\mathcal{H} \in \mathbb{C}^{8 \times 5}$ with $\text{cond}(\mathcal{H}^T) \approx 11.5$. The performance is averaged over 1000 Monte Carlo trials and SNR is defined as the channel output averaged over all subchannels. By comparison, the method of [3] was given the rank information $r_A = 5$. JADE and ACM clearly have better performance even though they estimate $r_A$. At low SNRs ($< 12dB$ in our simulation), the statistical method JADE has an MSE performance that outperforms ACM and closely approaches the Wiener MMSE bound. At higher SNR, the deterministic method ACM approaches the Wiener filter MSE while the curve for JADE levels off due to the residual ISI induced by the prewhitening bound. The performance of JADE and ACM under an empirical microwave channel ("chan4.mat" from SPIB) is shown in Figure 2. The sub-channels have a pair of almost overlapping zeros. The parameters are $P = 2$, $E = 4$ and no array is used here. Thus $\mathcal{H} \in \mathbb{C}^{8 \times 7}$ with $\text{cond}(\mathcal{H}^T) \approx 300$. A close look at the matrix $\mathcal{H}$ suggests that the small singular values are introduced by numerically small columns. Deleting them has a very small effect on other significant singular values. A case with two users (with different channel lengths) is presented in Figure 3. A 5-element ULA is used here and other parameters include $P = 2$ and $E = 1$. Thus $\mathcal{H} \in \mathbb{C}^{10 \times 8}$ and $\text{cond}(\mathcal{H}) \approx 21.7$. For multi-user cases, we can simply detect the sequence with the best AME first. Then we detect the second best sequence. If it is from the same user (detected for example using correlation or embedded user ID bits), we can discard it and proceed until all desired users are recovered. The slight fluctuation and leveling trend of the SER curve at high SNRs for JADE is again due to the prewhitening bound. However it is still good enough for the decision-directed post-processing to reduce the SER further to an undetectable level. Also worth noting is that the weaker sources will not be overwhelmed by stronger signals.

![Figure 1: Comparison (single user, M=1, P=2,E=2,N=80)](image1.png)

![Figure 2: Empirical channel (single user, P=2,E=4,M=1)](image2.png)

![Figure 3: Two users (M=5, P=2,E=1)](image3.png)

6. REFERENCES


