

Blind Source Separation with a Time-Varying Mixing Matrix

Marcus R. DeYoung and Brian L. Evans
 Department of Electrical and Computer Engineering
 The University of Texas at Austin
 1 University Station, Austin, Texas, 78712
 {deyoung, bevans}@ece.utexas.edu

Abstract—Blind source separation (BSS) is the process of using multiple sensors to separate multiple random signals from a received linear combination. In this paper, we apply blind source separation techniques to mixtures of digital communications signals. We predict the achievable symbol error rate when the signals are received on the same carrier frequency and blindly separated. In a wireless environment where the sources are mobile or the environment is changing, the mixing matrix will vary with time. Our primary contribution is that we identify the major source of error in separation. We extend an adaptive step size algorithm to the complex-valued case to mitigate the errors in a dynamic environment.

I. INTRODUCTION

Blind source separation (BSS) is the process of using multiple sensors to separate multiple random signals from a received linear combination. Applications may be found in many areas of signal processing, such as separating speech, biomedical signals, and co-channel communication signals. The classic example is the “cocktail party” problem, where one uses a microphone array to separate the voices of speakers in a room. Specifically, this paper explores demixing co-channel communication signals. While a number of significant problems arise from such a setup, this paper will focus on the case where the mixing matrix, which relates the unknown input sources to the sensed observations, varies with time.

The goal of BSS is to separate m sources from $n \geq m$ sensors. The system can be modeled as a linear system of equations

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}_t, t = 1, 2, \dots \quad (1)$$

where \mathbf{s}_t is a vector of m unknown signals, \mathbf{x}_t is a vector of n sensor readings, and \mathbf{A} is an $n \times m$ mixing matrix. An additive noise vector \mathbf{n}_t may be included. Assuming the noise to be negligible and that an inverse or pseudo-inverse exists, blind source separation attempts to find a matrix \mathbf{W} that can recover the sources:

$$\mathbf{y}_t = \mathbf{W}\mathbf{x}_t = \mathbf{W}\mathbf{A}\mathbf{s}_t \quad (2)$$

The primary contributions of this paper are the following:

- Predict received symbol error rate (SER) of blindly separated communications signals
- Identify the major source of separation error for time-varying mixing matrices to be the condition number of the mixing matrix
- Extend an iterative, adaptive step size algorithm to decrease inter-signal interference for a time-varying but unknown mixing matrix to the complex valued case

II. BACKGROUND

We use the Equivariant Adaptive Source Identification (EASI) algorithm as a basis [1]. We assume the additive noise term to be negligible, ignoring it in the problem formulation as is done throughout the BSS literature. Through simulation we see that at reasonable signal-to-noise (SNR) ratios, the separating algorithms still perform well. A separating matrix \mathbf{W} is a time-varying matrix \mathbf{W}_t that is updated every time instance using an $n \times n$ matrix function $H(\cdot)$ according to:

$$\mathbf{W}_{t+1} = \mathbf{W}_t - \lambda_t H(\mathbf{y}_t) \mathbf{W}_t = (\mathbf{I} - \lambda_t H(\mathbf{y}_t)) \mathbf{W}_t \quad (3)$$

This is a serial update where \mathbf{W}_t is updated by left multiplication every time instance. The algorithm is a stochastic gradient approach, similar to the popular least mean-square (LMS) algorithm [2]. The multiplier λ_t is a step size, and $H(\cdot)$ is chosen such that $E\{H(\mathbf{y}_t)\} = 0$ when \mathbf{y} has independent components. The choice of λ_t and $H(\cdot)$ affect the convergence rate of the algorithm. By using a phase-preserving nonlinearity, we keep the constellations from rotating [3].

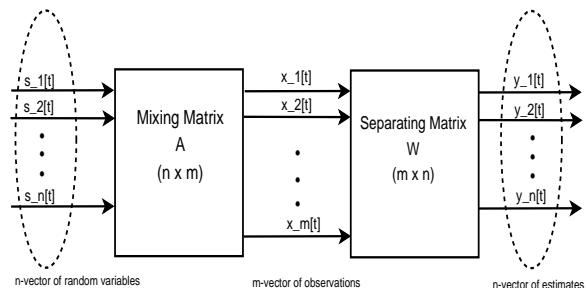


Fig. 1. Diagram illustrating the fundamental blind source separation system

M. DeYoung is supported by Zeta Associates, Inc., Fairfax, Virginia. This research was supported by an equipment grant from Intel.

Let $\langle \cdot | \cdot \rangle$ be the Euclidean inner product of two matrices:

$$\langle \mathbf{A} | \mathbf{B} \rangle = \text{Trace}(\mathbf{A}^T \mathbf{B}) \quad (4)$$

Consider an objective function $\phi(\mathbf{W})$, where \mathbf{W} is an $n \times m$ matrix. The standard gradient at \mathbf{W} is denoted $\frac{\partial \phi}{\partial \mathbf{W}}(\mathbf{W})$, where the first order Taylor approximation of $\phi(\mathbf{W})$ is

$$\phi(\mathbf{W} + \mathcal{E}) = \phi(\mathbf{W}) + \langle \frac{\partial \phi}{\partial \mathbf{W}} | \mathcal{E} \rangle + o(\mathcal{E}) \quad (5)$$

In this paper, to remain consistent with the literature, the relative gradient is denoted as $\nabla \phi(\mathbf{W})$ is defined such that

$$\phi(\mathbf{W} + \mathcal{E}\mathbf{W}) = \phi(\mathbf{W}) + \langle \nabla \phi(\mathbf{W}) | \mathcal{E} \rangle + o(\mathcal{E}) \quad (6)$$

From Equations 5 and 6, we see that $\nabla \phi(\mathbf{W}) = \frac{\partial \phi}{\partial \mathbf{W}}(\mathbf{W})\mathbf{W}^T$. This is useful for a number of reasons. First, the relative gradient is consistent with the serial update notion, where the separating matrix is updated at each iteration as $\mathbf{W}_{t+1} = \mathbf{W}_t + \mathcal{E}_t \mathbf{W}_t$. Secondly, it can be shown that when the objective function is in the form $\phi(\mathbf{W}) = E\{f(\mathbf{y})\} = E\{f(\mathbf{A}\mathbf{x})\}$, the relative gradient becomes $\nabla \phi(\mathbf{W}) = E\{f'(\mathbf{y})\mathbf{y}^T\}$ [1]. Thus, the relative gradient at any specific \mathbf{W} will only be a function of the received distribution $\mathbf{y} = \mathbf{W}\mathbf{x}$. A standard gradient would require us to use \mathbf{x} , which we assume is not available in this blind setting.

III. PERFORMANCE EFFECTS

In order to evaluate the execution of the separation algorithm, we first follow previous work defining the Inter-Signal Interference (ISI) as a performance measure to be [1]

$$ISI_{ij} = E\{ |(\mathbf{C}_t[i][j] - \mathbf{I})|^2 \} \quad (7)$$

where $\mathbf{C}_t = \mathbf{W}_t \mathbf{A}_t$ is ideally close to identity. Therefore ISI as defined above provides a measure of the distance from the desired identity matrix and is entirely source independent. Note that since \mathbf{C}_t may be a permuted scaled version of \mathbf{I} , we may have to permute the rows and scale back before making this measurement. The i -th received signal $\mathbf{y}_t[i]$ can be written as the desired signal plus interfering terms:

$$\mathbf{y}_t[i] = \mathbf{C}_t[i][i]\mathbf{y}[i]_t + \mathbf{n}_t + \sum_{j \neq i} \mathbf{C}_t[i][j]\mathbf{x}_j \quad (8)$$

This measure is reasonable for any source separation problem. In digital communication channels, the fundamental performance measure is usually the probability of symbol error. We can determine the expected probability of error based on the expected ISI if we treat the extra terms as additional Gaussian interferers. It is a rough estimate, but works well as the number of interferers gets large and is often used to characterize the effects of intersymbol interference. Fig. III illustrates the accuracy of predicting the symbol error rate this way with a stationary matrix. Additionally, the ISI can be predicted for a stationary matrix for a given λ [1]. For a time-varying mixing matrix, we look at the instantaneous ISI

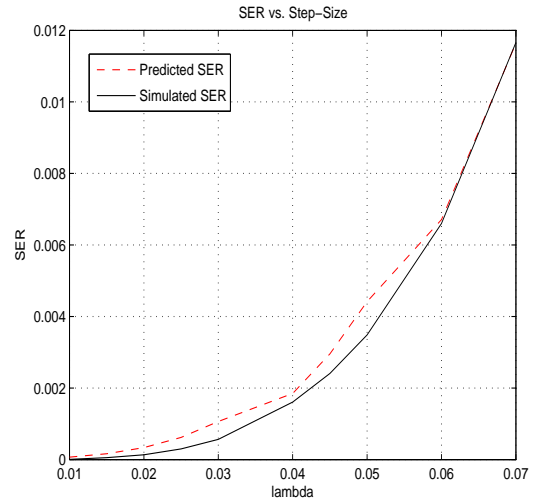


Fig. 2. Simulations show the effect of step size on symbol error rate vs. predicted for $n = m = 8$ sources and sensors. SNR was set to 15 dB; \mathbf{A} is complex Gaussian and constant for all t

instead of the expected value and average each of the non-diagonal ISI terms. Consequently, the average ISI used in Fig. 4 is defined as:

$$ISI_{avg} = \frac{1}{m^2 - m} \sum_{i,j,i \neq j} |(\mathbf{C}_t[i][j] - \mathbf{I})|^2 \quad (9)$$

A larger step size converges faster but has more error in the steady state while a smaller step size converges slower and has less error after convergence.

The example separation problem examined in this work considers mixtures of narrowband signals in the presence of diffuse multipath where the delay spread is shorter than the symbol time. This assumption allows us to maintain that \mathbf{A}_t is still instantaneous (otherwise, in a frequency selective channel, we would have a linear combination of multipath taps for each symbol period leading to the related problem of blind deconvolution). We follow Clarke's model for wide sense stationary uniform scattering [4]. The major source of error in this environment is the ill-conditioning of the matrix. The condition number of \mathbf{A} is denoted $\kappa(\mathbf{A}) = \frac{\sigma_{max}}{\sigma_{min}}$, the ratio of the maximum and minimum singular values. If \mathbf{A} is square, then the expression is equivalently $\kappa(\mathbf{A}) = \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\|$, where we use the L_2 -norm. When the condition number is large, small perturbations in \mathbf{A} result in large changes in the inverse (or pseudo-inverse) of \mathbf{A} , and an iterative algorithm may have trouble tracking the changes. Fig. 3 illustrates the direct correlation between the condition number and the average ISI using the EASI algorithm with a fixed step-size.

Besides the fact that higher ISI implies more symbol errors, there is a more fundamental problem with this ill-conditioning. The objective function is not convex, in fact there are many local minimums. These local minimums correspond to separating matrices, which are permuted forms of the identity matrix,

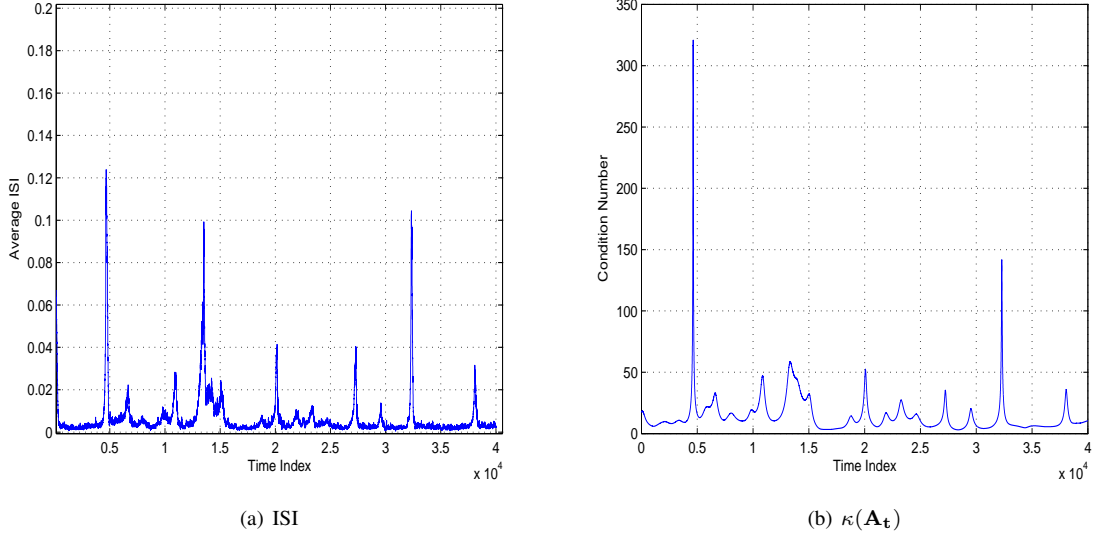


Fig. 3. Illustration of the effects of ill-conditioned mixing matrix on inter-signal interference. Notice the direct correlation between a higher κ value (condition number) and higher ISI.

with perhaps some phase shift. For example the following matrices could both be considered separating matrices for a 4×4 system:

$$\mathbf{WA}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{WA}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

We can see that \mathbf{WA}_1 will permute the sources such that $\mathbf{WA}_1[s_1, s_2, s_3, s_4]^T = [s_2, s_4, s_1, s_3]^T$, and \mathbf{WA}_2 will permute the sources in a new order as $\mathbf{WA}_2[s_1, s_2, s_3, s_4]^T = [s_1, s_3, s_2, s_4]^T$. This creates a problem when trying to sort out and demodulate the sources. If the separating matrix changes, the output signals must be re-synchronized somehow to determine which stream is which. The proposed algorithm reduces the number of times that this occurs, as is illustrated in the simulation results in Section V.

IV. VARIABLE STEP-SIZE ALGORITHM

When the matrix becomes ill-conditioned, we would like the step size to increase to accommodate the rapid changes in the separating matrix. In other stochastic gradient descent type algorithms, for example the least mean square (LMS) algorithm, research has shown improved performance by allowing the step size to vary [5], [6]. In many of these algorithms, the step size is updated according to it's own stochastic gradient descent in the form

$$\lambda_{t+1} = \lambda_t - \rho_t \frac{\partial \mathcal{J}(\cdot)}{\partial \lambda_t} \quad (11)$$

where $\mathcal{J}(\cdot)$ is the cost function and ρ is a small constant. In the LMS case, the cost function would be the squared error between the filtered and desired signal. In blind source

separation, it will be the function measuring independence. Note that now we are looking at the gradient of this cost function with respect to the step size.

A few papers have looked into using adaptive step size techniques for source separation. Douglas and Cichocki adapted Amari's natural gradient approach [7]. An approach following for the EASI algorithm with real-valued data was presented in [8].

An adaptive step size approach following for the EASI algorithm with real-valued data was presented in [8]. We build upon this work to find an adaptive step size algorithm for the complex-valued EASI algorithm and analyze it in a wireless environment. Following from Section II, we find the relative gradient taken with respect to the step size to be

$$\nabla_{\lambda} \phi|_{\lambda=\lambda_t} = \langle \Gamma^H(t+1), \Gamma(t) \rangle \quad (12)$$

$$\Gamma(t) = [\mathbf{y}_t \mathbf{y}_t^H - \mathbf{I} + g(\mathbf{y}_t) \mathbf{y}_t^H - \mathbf{y}_t g(\mathbf{y}_t)^H] \mathbf{W}_t \quad (13)$$

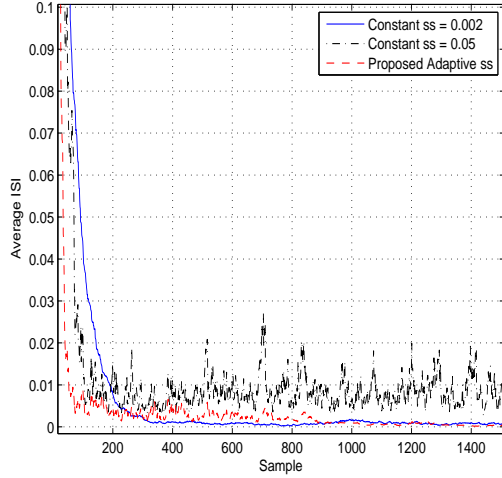
This is a complex-valued gradient, which provides information on the descent direction in both the real and imaginary planes. We propose a simple combination approach by choosing the average of the real and imaginary components:

$$\lambda_{t+1} = \lambda_t - \rho \frac{1}{2} [\Re\{\nabla_{\lambda} \phi|_{\lambda=\lambda_t}\} + \Im\{\nabla_{\lambda} \phi|_{\lambda=\lambda_t}\}] \quad (14)$$

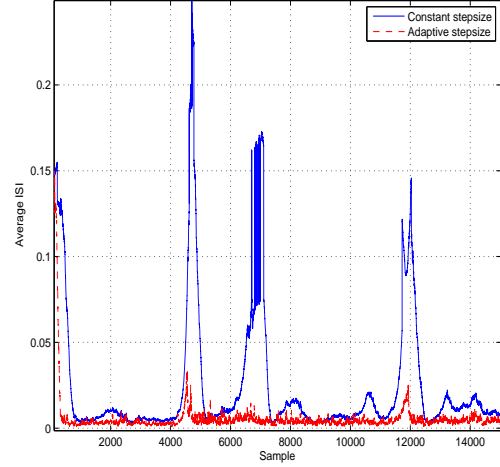
We now illustrate the advantages of using the complex-valued variable step size approach.

V. SIMULATION RESULTS

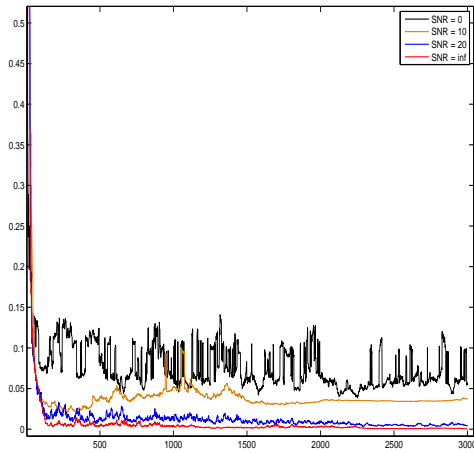
The advantages of the complex-valued adaptive step size are twofold. In addition to the main points discussed in Section IV with a time-varying mixing matrix, with a constant (or very slowly changing) mixing matrix, the proposed algorithm



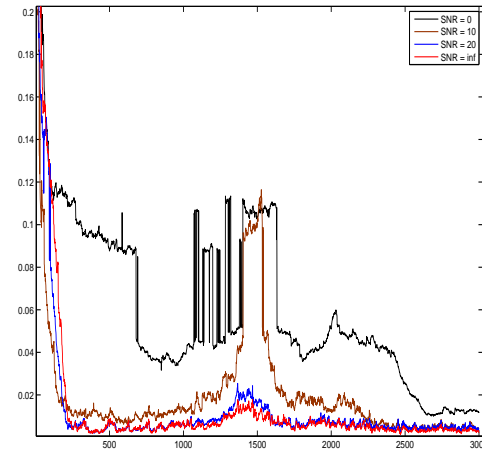
(a) Average ISI - Constant Mixing Matrix



(b) Average ISI $f_d = 25$ Hz, $f_s = 100$ kHz



(c) Average ISI - Constant Mixing Matrix



(d) Average ISI $f_d = 25$ Hz, $f_s = 100$ kHz

Fig. 4. Sample runs comparing the proposed adaptive step size algorithm to constant step size. Note that Fig. 4(a) is with a constant matrix, so a smaller number of samples are shown to illustrate the fast convergence. Fig. 4(b) shows a longer sequence to illustrate the adaptation under ill-conditioning for a time-varying matrix. Fig. 4(c) shows the performance in different noise environments of the adaptive algorithm in a constant mixing matrix. Fig. 4(d) presents the same SNR comparison in a fading channel.

adjusts the step size to be large at first for fast convergence then shrinks the step size in the steady state. Notice the performance of the proposed algorithm when \mathbf{A}_t is constant for all t , as seen in Fig. 4(a). The adaptive algorithm is compared to constant step sizes of $\lambda = 0.002$ and $\lambda = 0.05$. The larger step size converges more quickly, and the smaller step size has less steady-state error after convergence. The adaptive algorithm uses a large step size in the beginning so as to converge faster than $\lambda = 0.05$, and then shrinks λ_t to achieve a steady state error consistent with the small step size. Thus the adaptive step size provides a clear advantage even in the constant mixing matrix case.

Fig. 4(b) illustrates the advantage to using the adaptive algorithm. The simulations were run with $n = m = 4$ and

the average SNR set to 20 dB. No assumptions are made about the sensor geometry except that the observations are independent. The fixed step size was $\lambda = 0.01$, and the adaptive step size was computed with $\rho = 5 \times 10^{-5}$ with $0.0001 \leq \lambda_t \leq 0.10$. The adaptive step size algorithm adapts well to the ill-conditioned matrix. Note by looking at the y-axis that the average ISI increases as $\frac{f_d}{f_s}$ increases, as expected. The adaptive step size algorithm helps smooth out the jumps by increasing the step size appropriately.

In the formulation of the EASI algorithm, noise is removed from the problem. Here we show that at reasonably high SNR the noise effects are rather small. Figures 4(c) and 4(d) present the average ISI for different noise powers in the constant and time-varying mixing matrix cases respectively with the same

variable step size parameters as in Fig. 4(b) described above. For each SNR value, the mixing matrix is the same at each time instant but with varying additive noise. In both cases, once the SNR reaches 20 dB, the algorithm performs almost as well as when noise is absent in the system.

VI. CONCLUSION

This paper has examined the problem of separating multiple co-channel digital communications signals in a wireless environment. We have illustrated the relationship between the ability to separate digital communications signals measured by inter-signal interference and the maximum likelihood detector symbol error rate. Under the assumption of diffuse multipath and spatially independent channels, we have shown that the major difficulty in separation is the ill-conditioning of the channel matrix. In response to this challenge, we have proposed an adaptive step size EASI-based algorithm for complex-valued data which performs well in simulation compared to the constant step size approach.

REFERENCES

- [1] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [2] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of the IEEE*, vol. 63, pp. 719–720, Apr. 1975.
- [3] L. He, T. Thaiupathump, and S. Kassam, "Blind separation of complex IQ independent sources with phase recovery," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 419–422, May 2005.
- [4] R. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. Journal*, vol. 47, pp. 957–1000, July 1968.
- [5] W.-P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size lms algorithms," *IEEE Transaction on Signal Processing*, vol. 49, no. 4, pp. 805–810, April 2001.
- [6] V. Mathews and Z. Xie, "Stochastic gradient adaptive filters with gradient adaptive step sizes," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3-6 April 1990, pp. 1385–1388vol.3.
- [7] S. Douglas and A. Cichocki, "Adaptive step size techniques for decorrelation and blind source separation," in *Asilomar Conference on Signals, Systems & Computers*, vol. 2, 1-4 Nov. 1998, pp. 1191–1195vol.2.
- [8] J. Chambers, M. Jafari, and S. McLaughlin, "Variable step-size EASI algorithm for sequential blind source separation," *IEE Electronics Letters*, vol. 40, no. 6, pp. 393–394, March 2004.