COMBINED SUBSPACE TRACKING, PREWHITENING, AND CONTRAST OPTIMIZATION FOR NOISY BLIND SIGNAL SEPARATION

Scott C. Douglas

Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA

ABSTRACT

In many practical blind signal separation (BSS) applications, the measured mixtures contain additive noise that limits the performances of most existing BSS algorithms. In this paper, we present three new methods for blindly extracting independent sources from noisy linear mixtures. All of the methods combine approximate least-squares subspace tracking with contrast-based BSS in an elegant way. One of the BSS algorithms is designed to perform minimum mean-square-error (MMSE) or Wiener estimation of the unknown sources, and novel least-squares prewhitening and orthogonal contrast optimization techniques are introduced. Simulations verify the robust and accurate behaviors of the proposed methods for extracting unknown sources from noisy mixtures.

1. INTRODUCTION

This paper describes algorithms for noisy blind signal separation (BSS) of instantaneously-mixed sources, in which a measured vector sequence $\mathbf{x}(k) = [x_1(k) \; \cdots \; x_n(k)]^T$ is produced from a source vector sequence $\mathbf{s}(k) = [s_1(k) \; \cdots \; s_m(k)]^T$, $m \leq n$, as

$$\mathbf{x}(k) \quad = \quad \mathbf{A}\mathbf{s}(k) + \nu(k), \tag{1}$$

where $\mathbf{A}$ is an $(n \times m)$ unknown mixing matrix and $\nu(k) = [\nu_1(k) \; \cdots \; \nu_n(k)]^T$ is a zero mean jointly Gaussian noise vector sequence with covariance $\mathbf{R}_{\nu\nu} = \sigma_\nu^2\mathbf{I}$. The goal is to determine a matrix $\mathbf{B}(k)$ such that

$$\mathbf{y}(k) \quad = \quad \mathbf{B}(k-1)\mathbf{x}(k), \tag{2}$$

the output vector sequence, contains accurate estimates of the m source signals in $\mathbf{s}(k)$. When the noise is absent ($\nu(k) = \mathbf{0}$), the standard noiseless BSS task is obtained.

While many algorithms for noiseless BSS have been proposed recently, fewer algorithms for the noisy BSS task have been developed [1]–[7]. Many of these algorithms produce the zero-forcing solution; that is, the output signals have zero crosstalk at algorithm convergence [3, 6]. Such a condition generally does not correspond to the minimum mean-squared error (MMSE) or Wiener estimator commonly employed in traditional trained adaptive filtering tasks [8]. It is generally acknowledged (c.f. [2]) that the noisy BSS task is harder to solve than is the noiseless BSS task. Even so, it is important to consider this case as noise is likely to be present in all practical separation tasks.

In this paper, we develop three simple algorithms for blindly extracting source signals from noisy mixtures. All of these algorithms combine signal subspace estimation and signal separation in a convenient way with little to no additional computational effort. As discussed in [2] and shown in this paper, signal subspace estimation is a key component in constructing an MMSE separation solution. Moreover, while each of our methods can be viewed as extensions of the Equivariant Adaptive Separation via Independence (EASI) algorithm [9], they offer significant improvements when applied to noisy BSS, as verified by analysis and simulations:

1. The first of our new approaches maintains an exact inverse square root factor of the sample input signal autocorrelation matrix. This algorithm avoids the "small step size" approximations used in the EASI algorithm and has a level of numerical robustness that is unmatched by EASI or any other standard/natural gradient BSS scheme [10], especially with regard to initial convergence properties.

2. The second approach is a simplification of the first and combines EASI with an approximate version of the Projection Approximation Subspace Tracking (PAST) algorithm [11]. The proposed EASI Extended to Subspace Tracking (EASIEST) algorithm outperforms EASI in low signal-to-noise ratio environments; moreover, it requires fewer multiply/accumulates (MACs) than does EASI when the number of sources m is less than one-half the sensor number n.

3. The third approach modifies the EASIEST algorithm so that it produces true MMSE-based estimates of the sources. The EASIEST-MSE algorithm employs a bias removal technique described in [3] and requires $3mn + 2m^3 + \mathcal{O}(m^2)$ MACs to implement.

2. MMSE-BASED BLIND SIGNAL SEPARATION

Before presenting the new algorithms, we first review the form of the MMSE signal separator, also known as the Wiener beamformer in array processing [1], and we relate it to the standard or zero-forcing solution commonly used in noiseless BSS [12]. Assume without loss of generality that each of the source signals $s_i(k)$ are of unit variance such that $\mathbf{R}_{\mathbf{ss}} = E\{\mathbf{s}(k)\mathbf{s}^T(k)\} = \mathbf{I}$ and $\mathbf{R}_{\mathbf{xx}} = \mathbf{A}\mathbf{A}^T + \sigma_\nu^2\mathbf{I}$,

This material is based in part upon work supported by the Texas Advanced Technology Program under Grant No. 003613-0031-1999.

where $E\{\cdot\}$ denotes the expectation operator. Then, the MMSE separator matrix is

$$\mathbf{B}_{MSE} = E\{\mathbf{s}(k)\mathbf{x}^T(k)\}\mathbf{R}_{\mathbf{xx}}^{-1} = \mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T + \sigma_\nu^2\mathbf{I}\right)^{-1} \quad (3)$$

In practice, we allow for arbitrary row permutations and signal inversions of the MMSE solution, as the order and individual signs of the extracted sources are immaterial to the separation task. Thus, a blind signal separation system optimized for MMSE source estimates should converge to

$$\mathbf{B}_{ss} = \Phi\mathbf{J}\mathbf{B}_{MSE} = \Phi\mathbf{J}\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T + \sigma_\nu^2\mathbf{I}\right)^{-1}, \quad (4)$$

where Φ is any $(m \times m)$ permutation matrix and $\mathbf{J}$ is a diagonal matrix of ± 1's.

From (4), the row space of the optimal MMSE solution is the same as that of the source signal space and is simply the principal subspace of $\mathbf{x}(k)$. Let $\mathbf{P}(k) = \mathbf{P}$ denote any $(m \times n)$ principal subspace estimate of $\mathbf{x}(k)$, and define the m-element principal subspace signal vector $\mathbf{v}(k)$ as

$$\mathbf{v}(k) = \mathbf{P}(k-1)\mathbf{x}(k). \quad (5)$$

Then, $\mathbf{B}_{ss}$ in (4) can be equivalently written as

$$\mathbf{B}_{ss} = \Phi\mathbf{J}\mathbf{A}^T\mathbf{P}^T\left(\mathbf{P}\mathbf{A}\mathbf{A}^T\mathbf{P}^T + \sigma_\nu^2\mathbf{I}\right)^{-1}\mathbf{P}. \quad (6)$$

Furthermore, let

$$\mathbf{B}(k) = \mathbf{W}(k)\mathbf{P}(k), \quad (7)$$

where $\mathbf{W}(k)$ is an $(m \times m)$ separator matrix. Then, the optimal MMSE separation system is $\mathbf{B}_{ss} = \mathbf{W}_{MSE}\mathbf{P}$, where

$$\mathbf{W}_{MSE} = \Phi\mathbf{J}\mathbf{H}^T\mathbf{R}_{\mathbf{vv}}^{-1} \quad (8)$$

$$\mathbf{R}_{\mathbf{vv}} = \mathbf{H}\mathbf{H}^T + \sigma_\nu^2\mathbf{I}, \qquad \mathbf{H} = \mathbf{P}\mathbf{A}. \quad (9)$$

Thus, we can subdivide the MMSE BSS task into

- calculating a principal subspace estimate $\mathbf{P}(k)$, and
- determining a BSS system that calculates $\mathbf{W}_{MSE}$.

In this paper, it will be important to understand the relationship between the MMSE solution in (8) and the standard noiseless or zero-forcing BSS solution given by

$$\mathbf{W}_{ZF} = \Phi\mathbf{J}(\mathbf{P}\mathbf{A})^{-1} = \Phi\mathbf{J}\mathbf{H}^{-1}. \quad (10)$$

The following theorem gives the relationship between $\mathbf{W}_{ZF}$ and $\mathbf{W}_{MSE}$, the proof of which can be obtained via direct substitution of the singular-value-decomposition of $\mathbf{H}$.

Theorem 1: *The relationship between* $\mathbf{W}_{ZF}$ *and* $\mathbf{W}_{MSE}$ *is*

$$\mathbf{W}_{MSE} = \mathbf{W}_{ZF} - \sigma_\nu^2\mathbf{W}_{ZF}\mathbf{R}_{\mathbf{vv}}^{-1}. \quad (11)$$

3. SUBSPACE TRACKING METHODS

In this section, we briefly review an approximate recursive least-squares (RLS) algorithm for tracking principal subspaces as well as a modified version that possesses improved numerical properties. These methods will be combined with BSS techniques in later sections for noisy BSS.

The PAST algorithm computes a subspace matrix estimate $\mathbf{P}(k)$ that exactly minimizes the least-squares criterion

$$\mathcal{J}_{PAST}(\mathbf{P}(k)) = \sum_{l=1}^{k}\lambda^{k-l}||\mathbf{x}(l) - \mathbf{P}^T(k)\mathbf{v}(l)||^2 \quad (12)$$

at each time instant, where $||\cdot||$ denotes the Euclidean norm [11]. The resulting recursive updates for $\mathbf{P}(k)$ are

$$\mathbf{P}(k) = \mathbf{P}(k-1) + \mathbf{k}(k)\mathbf{e}^T(k) \quad (13)$$

$$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{P}^T(k-1)\mathbf{v}(k) \quad (14)$$

$$\mathbf{k}(k) = \frac{\mathbf{R}_{\mathbf{vv}}^{-1}(k-1)\mathbf{v}(k)}{\lambda + \mathbf{v}^T(k)\mathbf{R}_{\mathbf{vv}}^{-1}(k-1)\mathbf{v}(k)} \quad (15)$$

$$\mathbf{R}_{\mathbf{vv}}^{-1}(k) = \mathbf{R}_{\mathbf{vv}}^{-1}(k-1) - \mathbf{k}(k)\mathbf{v}^T(k)\mathbf{R}_{\mathbf{vv}}^{-1}(k-1). \quad (16)$$

Besides being of $\mathcal{O}(mn)$ complexity, this algorithm has a number of useful properties, among them fast convergence and the approximate orthonormality of the subspace matrix at its stationary points; see [11] for more details.

The PAST algorithm does not maintain the orthonormality of the subspace matrix estimate $\mathbf{P}(k)$ during adaptation. Although not critical to the system's tracking performance, imposing the orthonormality constraint

$$\mathbf{P}(k)\mathbf{P}^T(k) = \mathbf{I} \quad (17)$$

for all k is highly-desirable during the algorithm's initial convergence phase. In [13], a modification to the PAST algorithm is introduced that enforces the constraint in (17) at each iteration while approximately minimizing (12). This algorithm employs m identical Householder transformations to update $\mathbf{P}(k)$ at each time instant. The algorithm is

$$\mathbf{z}(k) = \mathbf{e}(k) - \frac{||\mathbf{e}(k)||^2}{2}\mathbf{P}^T(k-1)\mathbf{k}(k) \quad (18)$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) + \frac{\mathbf{k}(k)\mathbf{z}^T(k)}{1 + \frac{1}{4}||\mathbf{e}(k)||^2||\mathbf{k}(k)||^2}, \quad (19)$$

where $\mathbf{v}(k)$, $\mathbf{k}(k)$, and $\mathbf{R}_{\mathbf{vv}}^{-1}(k)$ are computed as in the original PAST algorithm. This algorithm is also of $\mathcal{O}(mn)$ complexity and offers improved numerical robustness over the PAST algorithm, particularly for small values of λ [13].

4. COMBINING SUBSPACE TRACKING AND BLIND SIGNAL SEPARATION (ST-BSS)

In this section, we develop an algorithm that combines either PAST algorithm for subspace tracking with a new algorithm for contrast-based BSS. This new BSS method can be viewed as a modified EASI algorithm [9]; however, it possesses two important properties that the relative and natural gradient approaches in [9] and [10] lack:

- Our novel technique computes a separation matrix $\mathbf{W}(k)$ that is an *exact* square-root factor of $\mathbf{R}_{\mathbf{vv}}^{-1}(k)$ in (16). Hence, the second-order properties of our new method are more akin to least-squares than (equivariant) gradient-based approaches. The result is improved initial convergence performance and robustness with respect to the size of the prewhitening learning parameter.

• Our algorithm adjusts the column space of the separating matrix $\mathbf{W}(k)$ via a novel multiplicative orthogonal rotation technique that employs contrast optimization for separation, leaving the row space of $\mathbf{W}(k)$ unaltered. Hence, even though the prewhitening and separation tasks are combined into one separation matrix, the fidelity of $\mathbf{W}(k)$ as a least-squares prewhitening matrix is maintained.

Because of the above features, our new algorithm is a perfect complement to either PAST algorithm in the last section, because we can use $\mathbf{W}(k)$ to compute the Kalman gain vector $\mathbf{k}(k)$ for the subspace tracking updates. In addition, the resulting methods' complexities scale nearly linearly with the number of adaptive parameters within the systems.

4.1. Least-Squares Prewhitening

In contrast-based approaches to BSS, the separation matrix $\mathbf{W}(k-1)$ prewhitens the signal sequence $\mathbf{v}(k)$, such that

$$E\{\mathbf{y}(k)\mathbf{y}^T(k)\} = \mathbf{I}, \tag{20}$$

where the estimated source vector sequence is

$$\mathbf{y}(k) = \mathbf{W}(k-1)\mathbf{v}(k). \tag{21}$$

Consider a deterministic version of (20) using exponential averaging, for which the constraints are

$$\mathbf{W}(k)\mathbf{R}_{\mathbf{vv}}(k)\mathbf{W}^T(k) = \mathbf{I} \tag{22}$$

$$\mathbf{R}_{\mathbf{vv}}(k) = \lambda^k\delta\mathbf{I} + \sum_{l=1}^{k}\lambda^{k-l}\mathbf{v}(l)\mathbf{v}^T(l), \tag{23}$$

and δ is a small constant. From (22), we have

$$\mathbf{W}^T(k)\mathbf{W}(k) = \mathbf{R}_{\mathbf{vv}}^{-1}(k). \tag{24}$$

Thus, we seek to calculate in $\mathbf{W}(k)$ an exact square-root factor of $\mathbf{R}_{\mathbf{vv}}^{-1}(k)$ at each time instant.

Recently, a least-squares prewhitening method has been developed [14]. This algorithm is a modified version of the least-squares adaptive filter in [15] and is the least-squares equivalent of the natural gradient prewhitening procedure in [16]. The coefficient updates for this algorithm are

$$\mathbf{u}(k) = \mathbf{W}^T(k-1)\mathbf{y}(k) \tag{25}$$

$$\zeta(k) = \frac{1}{\lambda + ||\mathbf{y}(k)||^2 + \sqrt{\lambda}\sqrt{\lambda + ||\mathbf{y}(k)||^2}} \tag{26}$$

$$\mathbf{W}(k) = \frac{1}{\sqrt{\lambda}}\left[\mathbf{W}(k-1) - \zeta(k)\mathbf{y}(k)\mathbf{u}^T(k)\right], \tag{27}$$

where $\mathbf{W}(0) = \delta^{-1/2}\mathbf{I}$. It can be shown that this update satisfies (22) at each time instant. This update also possesses a number of other nice features, among them (i) fast convergence, (ii) robust behavior for any forgetting factor in the range $0 < \lambda \leq 1$, and (iii) good numerical properties. Equally importantly, this algorithm has similar complexity to its natural gradient counterpart given by

$$\mathbf{W}(k) = \mathbf{W}(k-1) + \frac{1-\lambda}{2}\left[\mathbf{I} - \mathbf{y}(k)\mathbf{y}^T(k)\right]\mathbf{W}(k-1). \tag{28}$$

In fact, the new prewhitening method requires only one MAC, one divide, and one square root more than the $4mn + m$ MACs of (28). See [14] for more details on these and other features of the method.

4.2. A Novel Prewhitened Blind Signal Separation Algorithm Using Contrasts

Signal separation using orthogonal contrast functions is a well-studied problem in which numerous algorithms and methods have been developed [17]–[20]. The common goal of all of these approaches can be described as follows: Given a set of prewhitened measurements $\mathbf{y}(k)$, find an orthogonal rotation matrix $\Xi(k)$ such that

$$\mathcal{J}_C(\Xi(k)) = \sum_{i=1}^{m} E\{\phi_i(\overline{y}_i(k))\} \tag{29}$$

is maximized, where $\phi_i(y)$ is the ith contrast function and

$$\overline{\mathbf{y}}(k) = d(k)\Xi(k)\mathbf{y}(k). \tag{30}$$

Here, $d(k)$ is a constant that counteracts the scaling of $\mathbf{y}(k)$ due to the exponentially-weighted least-squares prewhitening procedure. A useful choice for $d(k)$ is

$$d(k) = \left(\frac{1-\lambda^k}{1-\lambda}\right)^{\frac{1}{2}}. \tag{31}$$

By suitable choice of $\phi_i(y)$ relative to the source signal probability density functions (p.d.f.'s), extraction of all of the source signals can be achieved through this optimization process. Moreover, in the case where $\mathbf{s}(k)$ contains at least $(m-1)$ nonzero-kurtosis source signals, choosing

$$\phi_i(y) = \mathrm{sgn}[\kappa_i]|y|^4 \tag{32}$$

with $\kappa_i = E\{|y_i(k)|^4\} - 3$ guarantees extraction of the m source signals in $\mathbf{s}(k)$ [17, 18]. To allow an iterative solution, we define the orthogonal update matrix $\Theta(k)$ satisfying

$$\Xi(k) = \Theta(k)\Xi(k-1). \tag{33}$$

Previously-proposed methods for constructing $\Theta(k)$ to iteratively maximize (29) have generally employed orthogonal (Givens) rotations that require costly cos / sin operations [18]. In what follows, we provide a new technique that uses the *Cayley transformation matrix*

$$\Theta(k) = [\mathbf{I} + 0.5\mathbf{S}(k)][\mathbf{I} - 0.5\mathbf{S}(k)]^{-1} \tag{34}$$

$$= 2[\mathbf{I} - 0.5\mathbf{S}(k)]^{-1} - \mathbf{I}, \tag{35}$$

where $\mathbf{S}(k) = -\mathbf{S}^T(k)$ is skew-symmetric. It can be proven that $\Theta(k)$ is orthonormal; that is, $\Theta(k)\Theta^T(k) = \mathbf{I}$ [21]. Moreover, through a Taylor series expansion, we see that

$$\Theta(k) = \mathbf{I} + \mathbf{S}(k) + \mathcal{O}(\mathbf{S}^2(k)), \tag{36}$$

such that (33) becomes for small $||\mathbf{S}(k)||_F$

$$\Xi(k) \approx \Xi(k-1) + \mathbf{S}(k)\Xi(k-1). \tag{37}$$

When $\mathbf{S}(k)\Xi(k-1)$ is the gradient of a cost function within the space of $(n \times n)$ orthonormal matrices, (37) is a discretized version of a gradient search over the *Stiefel manifold* [20, 22]. Unlike (37), however, the update in (33) preserves the exact orthogonality of $\Xi(k)$ at each time instant.

At this point, we must choose an anti-symmetric matrix $\mathbf{S}(k)$ so that the contrast function in (29) is maximized by

Table 1: Combined ST-BSS Algorithm

Equation	*MACs*
$\mathbf{v}(k) = \mathbf{P}(k-1)\mathbf{x}(k)$	mn
$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{P}^T(k-1)\mathbf{v}(k)$	mn
$\mathbf{y}(k) = \mathbf{W}(k-1)\mathbf{v}(k)$	m^2
$\mathbf{u}(k) = \mathbf{W}^T(k-1)\mathbf{y}(k)$	m^2
$\mathbf{k}(k) = \dfrac{\mathbf{u}(k)}{\lambda + \lvert\lvert\mathbf{y}(k)\rvert\rvert^2}$	$2m$ $(1\div)$
$\mathbf{z}(k) = \mathbf{e}(k) - \mathbf{P}^T(k-1)\left[0.5\lvert\lvert\mathbf{e}(k)\rvert\rvert^2\mathbf{k}(k)\right]$	$mn+n$ $+m+1$
$\mathbf{P}(k) = \mathbf{P}(k-1) + \dfrac{\mathbf{k}(k)\mathbf{z}^T(k)}{1+0.25\lvert\lvert\mathbf{e}(k)\rvert\rvert^2}$	$2mn+2m$ $+2$ $(1\div)$
$\zeta(k) = \left[\lambda + \lvert\lvert\mathbf{y}(k)\rvert\rvert^2 + \sqrt{\lambda}\sqrt{\lambda + \lvert\lvert\mathbf{y}(k)\rvert\rvert^2}\right]^{-1}$	1 $(1\div, 1\sqrt{\cdot})$
Compute $f_i(y_i(k))$, $1 \leq i \leq m$	$2m$
Compute $\alpha_i(k)$, $i \in \{1,2,3,4\}$	$2m+21$ $(1\div)$
$\mathbf{W}(k) = \dfrac{1}{\sqrt{\lambda}}\{\mathbf{W}(k-1) + [\alpha_1(k)\mathbf{f}(\mathbf{y}(k))$ $+(\alpha_2(k)-\zeta(k))\mathbf{y}(k)]\,\mathbf{u}^T(k)$ $+[\alpha_3(k)\mathbf{y}(k)$ $+\alpha_4(k)\mathbf{f}(\mathbf{y}(k))]\,\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k-1)\}$	$4m^2+4m$ $+1$
Total: $4mn + 6m^2 + n + 11m + 25$ $(4\div, 1\sqrt{\cdot})$	

the procedure in (33). This choice is straightforward. The standard gradient of $\mathcal{J}_C(\Xi)$ is

$$\mathrm{grad}(\mathcal{J}(\Xi)) = \frac{\partial \mathcal{J}_C(\Xi)}{\partial \Xi} = \mathbf{f}(\overline{\mathbf{y}}(k))d(k)\mathbf{y}^T(k), \quad (38)$$

where $\mathbf{f}(\mathbf{y}) = [f_1(y_1) \; \cdots \; f_m(y_m)]^T$ and $f_i(y) = \partial\phi_i(y)/\partial y$. Then, the tangent gradient in the Stiefel manifold is [22]

$$\mathrm{grad}_S(\mathcal{J}(\Xi)) = \frac{\partial \mathcal{J}_C(\Xi)}{\partial \Xi} - \Xi\left[\frac{\partial \mathcal{J}_C(\Xi)}{\partial \Xi}\right]^T \Xi \quad (39)$$

$$= \mathbf{f}(\overline{\mathbf{y}}(k))d(k)\mathbf{y}^T(k) - \overline{\mathbf{y}}(k)\mathbf{f}^T(\overline{\mathbf{y}}(k))\Xi. \quad (40)$$

Finally, we modify (40) using the fact that $\Xi^T\Xi = \mathbf{I}$ to get

$$\mathrm{grad}_S(\mathcal{J}(\Xi)) = \mathbf{f}(\overline{\mathbf{y}}(k))d(k)\mathbf{y}^T(k)\Xi^T\Xi - \overline{\mathbf{y}}(k)\mathbf{f}^T(\overline{\mathbf{y}}(k))\Xi \quad (41)$$

$$= \left[\mathbf{f}(\overline{\mathbf{y}}(k))\overline{\mathbf{y}}^T(k) - \overline{\mathbf{y}}(k)\mathbf{f}^T(\overline{\mathbf{y}}(k))\right]\Xi. \quad (42)$$

Thus, we choose

$$\mathbf{S}(k) = \beta\left[\mathbf{f}(\overline{\mathbf{y}}(k))\overline{\mathbf{y}}^T(k) - \overline{\mathbf{y}}(k)\mathbf{f}^T(\overline{\mathbf{y}}(k))\right], \quad (43)$$

where β is the step size of the contrast-based update. When $|\beta|$ is small, (33) as this choice produces a valid geodesic trajectory in the Stiefel manifold to maximize $\mathcal{J}_C(\Xi)$.

The updating scheme for $\Xi(k)$ in (33) can be easily combined with the prewhitening update for $\mathbf{W}(k)$ in (27), as described next. Eqn. (27) can be written as

$$\mathbf{W}(k) = \frac{1}{\sqrt{\lambda}}\mathbf{W}(k-1)\left[\mathbf{I} - \zeta(k)\mathbf{x}(k)\mathbf{u}^T(k)\right], \quad (44)$$

where $\zeta(k)$ depends on the length of $\mathbf{y}(k)$. Since the last term inside large brackets on the RHS of (44) does not depend on the relative orientations of the columns of $\mathbf{W}(k-1)$, we can apply an arbitrary rotation $\Theta(k)$ to the columns

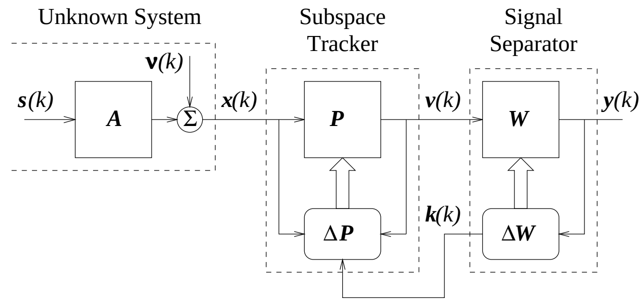

Fig. 1: Combined ST-BSS algorithm.

of the RHS of (44) and the resulting update still maintains (24). This family of updates is

$$\mathbf{W}(k) = \Theta(k)\frac{1}{\sqrt{\lambda}}\left[\mathbf{W}(k-1) - \zeta(k)\mathbf{y}(k)\mathbf{u}^T(k)\right]. \quad (45)$$

Substituting $\Theta(k)$ in (35) into (45) gives us the combined prewhitening/contrast optimization procedure as

$$\mathbf{W}(k) = \frac{1}{\sqrt{\lambda}}[\mathbf{I} + 0.5\mathbf{S}(k)][\mathbf{I} - 0.5\mathbf{S}(k)]^{-1} \times \left[\mathbf{W}(k-1) - \zeta(k)\mathbf{y}(k)\mathbf{u}^T(k)\right] \quad (46)$$

$$\mathbf{S}(k) = \beta d^4(k)\left[\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) - \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))\right] \quad (47)$$

$$f_i(y) = \mathrm{sgn}[\kappa_i]|y|^2 y. \quad (48)$$

Note that this update effectively redefines $\overline{\mathbf{y}}(k)$ as $\mathbf{y}(k)$, whereas $\mathbf{u}(k)$ is unchanged.

Although mathematically-correct, the update in (46) involves a matrix inverse as written. Fortunately, because $\mathbf{S}(k)$ is a rank-two matrix, we may apply the matrix inversion lemma twice to simplify the form of $[\mathbf{I} - 0.5\mathbf{S}(k)]^{-1}$. Then, multiplying out the remaining terms, we find that the resulting update for $\mathbf{W}(k)$ has the form of the last equation within Table 1, where $\alpha_i(k)$, $i \in \{1,2,3,4\}$ depend on $\mathbf{y}(k)$, $\mathbf{f}(\mathbf{y}(k))$, λ, and β. We have omitted the exact expressions of $\alpha_i(k)$ for brevity as they can be derived from direct linear algebraic manipulations of the corresponding matrices.

4.3. Combining the Algorithms

We can easily combine our new prewhitened/contrast-based BSS algorithm with either of the PAST algorithms of Section 3. This combination is much more computationally-efficient than two separate implementations, as $\mathbf{u}(k)$ and $\mathbf{y}(k)$ can be used to compute $\mathbf{k}(k)$ in (15) as

$$\mathbf{k}(k) = \frac{\mathbf{u}(k)}{\lambda + ||\mathbf{y}(k)||^2}. \quad (49)$$

Table 1 contains the combined subspace tracking/blind signal separation (ST-BSS) algorithm, in which the Householder-based PAST algorithm has been employed. The overall system uses only $4mn + 6m^2 + \mathcal{O}(n)$ MACs at each time instant. Alternatively, the standard PAST update in (13) can be substituted for (19), producing an algorithm that requires $3mn + 6m^2 + \mathcal{O}(n)$ MACs per iteration. Fig. 1 shows the signal flows for these combined ST-BSS algorithms, in which the Kalman gain vector $\mathbf{k}(k)$ is computed within the signal separator update module before being passed along to the subspace tracker update module.

Table 2: EASIEST Algorithm

Equation	*MACs*
$\mathbf{v}(k) = \mathbf{P}(k-1)\mathbf{x}(k)$	mn
$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{P}^T(k-1)\mathbf{v}(k)$	mn
$\mathbf{y}(k) = \mathbf{W}(k-1)\mathbf{v}(k)$	m^2
$\mathbf{u}(k) = \mathbf{W}^T(k-1)\mathbf{y}(k)$	m^2
Compute $f_i(y_i(k))$, $1 \leq i \leq m$	$2m$
$\mathbf{P}(k) = \mathbf{P}(k-1) + 2\mu\mathbf{u}(k)\mathbf{e}^T(k)$	$mn + m$
$\mathbf{W}(k) = [1+\mu]\,\mathbf{W}(k-1)$ $+ [\beta\mathbf{f}(\mathbf{y}(k)) - \mu\mathbf{y}(k)]\,\mathbf{u}^T(k)$ $- \beta\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k-1)$	$4m^2 + 3m$
Total: $3mn + 6m^2 + 6m$	

5. EASIEST: A SIMPLIFIED ST-BSS ALGORITHM

The ST-BSS algorithm in Table 1 offers a number of useful features, such as fast prewhitening convergence and robust (non-divergent) behavior for any values of λ and β in the ranges $0 < \lambda \leq 1$ and $\beta > 0$, respectively. Even so, the algorithm's complexity relative to other BSS approaches is large if m and n are small. For this reason, we derive an approximate version of this ST-BSS algorithm that sacrifices algorithm robustness for computational simplicity. As the new algorithm combines the EASI BSS method with an approximate version of the PAST algorithm, we have termed it EASI Extended to Subspace Tracking (EASIEST).

The EASIEST algorithm assumes that $||\mathbf{k}(k)||^2$ is small relative to $\mathbf{v}(k)$, which is reasonable for forgetting factors close to unity. In such situations, we can approximate

$$\mathbf{k}(k) \approx 2\mu\mathbf{u}(k), \qquad [1 + ||\mathbf{k}(k)||^2||\mathbf{e}(k)||^2]^{-1} \approx 1, \quad (50)$$

$$\frac{1}{\sqrt{\lambda}} \approx 1 + \mu, \quad \text{and} \quad \zeta(k) \approx \frac{1}{2} + \mu, \quad (51)$$

where we have defined

$$\mu = \frac{1-\lambda}{2}. \quad (52)$$

Finally, ignoring all terms on the RHS that are of $\mathcal{O}(\mu^2)$, $\mathcal{O}(\beta^2)$, and $\mathcal{O}(\mu\beta)$ and higher, we obtain the EASIEST algorithm given in Table 2.

Although similar to the original EASI algorithm, the EASIEST algorithm possesses two important improvements:

- Unlike EASI, the EASIEST algorithm can alter the row space of the overall separation matrix $\mathbf{W}(k)\mathbf{P}(k)$ to ignore the noise subspace of $\mathbf{x}(k)$ when measurement noise is present. Hence, it typically provides better estimates of the unknown source signals for noisy BSS.
- In situations where

$$m < \frac{n}{2}, \quad (53)$$

the EASIEST algorithm actually requires fewer MACs than does the EASI algorithm. In other words, EASIEST is indeed "easier" than EASI when extracting a few source signals from a large-dimensional measurement space, a common scenario in many array processing applications.

Table 3: EASIEST-MSE Algorithm

Equation	*MACs*
$\mathbf{v}(k) = \mathbf{P}(k-1)\mathbf{x}(k)$	mn
$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{P}^T(k-1)\mathbf{v}(k)$	mn
$\mathbf{y}(k) = \mathbf{W}(k-1)\mathbf{v}(k)$	m^2
$\mathbf{u}(k) = \mathbf{W}^T(k-1)\mathbf{y}(k)$	m^2
$\mathbf{t}(k) = \mathbf{G}(k-1)\mathbf{v}(k)$	m^2
$\mathbf{y}_{MSE}(k) = \mathbf{y}(k) - \sigma_\nu^2\mathbf{W}(k-1)\mathbf{t}(k)$	$m^2 + m$
Compute $f_i(y_i(k))$, $1 \leq i \leq m$	$2m$
$\mathbf{P}(k) = \mathbf{P}(k-1) + 2\mu\mathbf{t}(k)\mathbf{e}^T(k)$	$mn + m$
$\mathbf{G}(k) = [1+2\mu]\,\mathbf{G}(k-1) - 2\mu\mathbf{t}(k)\mathbf{t}^T(k)$	$2m^2$
$\mathbf{W}(k) = [1+\mu]\mathbf{W}(k-1)$ $+ \mu\sigma_\nu^2\mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{W}(k-1)$ $+ [\beta\mathbf{f}(\mathbf{y}(k)) - \mu\mathbf{y}(k)]\,\mathbf{u}^T(k)$ $- \beta\mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))\mathbf{W}(k-1)$	$2m^3$ $+5m^2$ $+3m$
Total: $3mn + 2m^3 + 11m^2 + 5m$	

6. MODIFYING THE EASIEST ALGORITHM FOR MMSE-BASED BSS

While the EASIEST algorithm achieves some noise immunity in its separation performance, it does not compute the MMSE separation solution, and a full analysis of the EASIEST algorithm is beyond the scope of this paper. Instead, we modify the EASIEST algorithm so that it extracts MMSE source estimates at convergence. The modifications are based on the following observation, verified by the analytical work in [3]: It is possible to introduce a bias correction term to the natural gradient signal separation method in [10] so that, at convergence, the algorithm achieves the zero-forcing solution $\mathbf{W}_{ZF}$ in (10). This bias-corrected version is

$$\begin{aligned}\mathbf{W}(k) &= \mathbf{W}(k-1) + \beta\left[\mathbf{W}(k-1) - \mathbf{f}(\mathbf{y}(k))\mathbf{u}^T(k)\right] \\ &\quad + \beta\sigma_\nu^2\mathbf{D}'_f(k)\mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{W}(k-1)\end{aligned} \quad (54)$$

where the diagonal matrix $\mathbf{D}'_f(k)$ contains estimates of the quantities $E\{f'_i(y_i(k))\}$ along its main diagonal.

We can modify the EASIEST algorithm so that it too achieves the zero-forcing solution. In this case, only the prewhitening portion of the algorithm needs to be considered, as $E\{\mathbf{S}(k)\} = \mathbf{0}$ at a zero-forcing solution when Gaussian noise is present. This modified algorithm is

$$\begin{aligned}\mathbf{W}(k) &= \mathbf{W}(k-1) + \mu\left[\mathbf{W}(k-1) - \mathbf{y}(k)\mathbf{u}^T(k)\right] \\ &\quad + \beta\left[\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) - \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k))\right]\mathbf{W}(k-1) \\ &\quad + \mu\sigma_\nu^2\mathbf{W}(k-1)\mathbf{W}^T(k-1)\mathbf{W}(k-1).\end{aligned} \quad (55)$$

It should be noted that the resulting algorithm is no longer a natural gradient approach, and the additional term within the coefficient updates could cause the algorithm to converge to non-separating solutions. Even so, the algorithm has been observed to have the desired steady-state estimation characteristics in simulation, a situation similar to that observed in [3] for the update in (54).

Finally, we can use the solution obtained by the above procedure to construct the MMSE signal estimate vector, denoted as $\mathbf{y}_{MSE}(k) = \mathbf{W}_{MSE}\mathbf{v}(k)$. Post-multiplying (11) by $\mathbf{v}(k)$ and letting $\mathbf{W}_{ZF} = \mathbf{W}(k)$, we obtain

$$\mathbf{y}_{MSE}(k) = \mathbf{y}(k) - \sigma_\nu^2\mathbf{W}(k)\mathbf{t}(k), \quad (56)$$

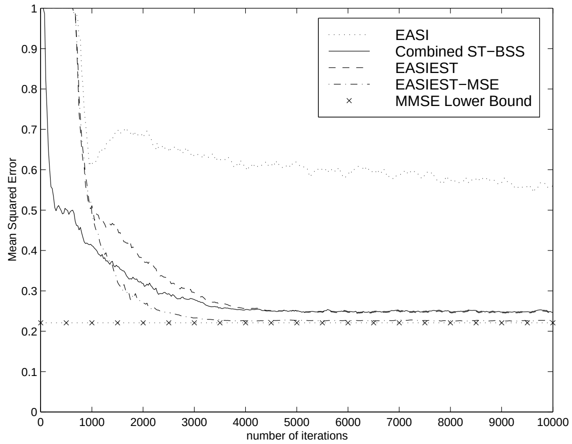

Fig. 1: Evolutions of $\mathcal{J}_{MSE}(k)$ for the four algorithms as well as the MMSE lower bound in the simulation example.

where $\mathbf{t}(k) = \mathbf{G}(k-1)\mathbf{v}(k)$ and $\mathbf{G}(k-1)$ is an estimate of $\mathbf{R}_{\mathbf{vv}}^{-1}$ in (9). While we could estimate $\mathbf{G}(k)$ using (15)–(16), an even simpler approach that matches (15)–(16) up to terms of $\mathcal{O}(\mu)$ can be used. Table 3 contains the resulting EASIEST-MSE algorithm. This algorithm requires $3mn + \mathcal{O}(m^3)$ operations at each time instant.

7. SIMULATIONS

We now explore the performances of the proposed BSS algorithms via simulations. We have generated measurements according to (1), where $m = 3$, $n = 6$, $\sigma_\nu^2 = 0.25$, and

$$\mathbf{A}^T = \begin{bmatrix} 0.9 & 0.3 & 0.8 & 0.1 & 0.9 & 0.2 \\ 0.9 & 0.5 & 0.6 & 0.2 & 0.1 & 0.7 \\ 0.3 & 0.7 & 0.2 & 0.9 & 0.3 & 0.5 \end{bmatrix}. \quad (57)$$

With these choices, the signal-to-noise ratio across the signal space varies from approximately 1.7dB to 13dB. We have chosen $s_1(k)$, $s_2(k)$, and $s_3(k)$ to be zero-mean binary, uniform, and uniform distributed i.i.d. sequences, such that $\kappa_i < 0$ in (48) for all i. The results of twenty simulations have been used to estimate the MMSE cost function

$$\mathcal{J}_{MSE}(k) = \frac{1}{m} E\{||\mathbf{s}(k) - \mathbf{J}\Phi^T\mathbf{W}(k)\mathbf{P}(k)\mathbf{x}(k)||^2\} \quad (58)$$

for each algorithm. In each simulation run, random orthogonal matrices were chosen satisfying $\mathbf{W}(0)\mathbf{W}^T(0) = 0.05\mathbf{I}$ and $\mathbf{P}(0)\mathbf{P}^T(0) = \mathbf{I}$ so as to not bias the results towards any fixed initial condition.

Fig. 1 shows the evolutions of $\mathcal{J}_{MSE}(k)$ for the combined ST-BSS, EASI, EASIEST, and EASIEST-MSE algorithms, where $\beta = 0.0007$ and $\lambda = 0.993$. As can be seen, the EASI algorithm does not estimate the sources very well, whereas the three proposed methods provide much better MSE performance. The least-squares-based algorithm gives the fastest initial convergence, and its steady-state behavior is quite similar to the EASIEST algorithm, as expected. The EASIEST-MSE algorithm produces the lowest steady-state MSE of all the approaches, and its MSE value of 0.226 is quite close to the minimum achievable MSE of 0.221 for these signal statistics. Clearly, the performance improvements of the new methods over EASI are substantial, and in the case of EASIEST, they can often be had at no additional computational cost.

8. CONCLUSIONS

In this paper, we have introduced three novel algorithms for noisy BSS tasks. All of the algorithms integrate subspace tracking, prewhitening, and contrast optimization in a convenient manner. New least-squares prewhitening and orthogonal contrast optimizations are given, and one of the approaches is designed to perform MMSE source estimation. Simulations verify the usefulness of the algorithms in blindly-extracting unknown signals under low signal-to-noise ratio conditions.

9. REFERENCES

[1] S. Bose and B. Friedlander, "On the direct estimation of the Wiener beamformer using second and fourth order cumulants," *Proc. 30th Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, vol. 1, pp. 370-374, Nov. 1996.

[2] J.-F. Cardoso, "Blind separation of noisy mixtures," *Proc. IEEE Workshop, Neural Networks Signal Processing,* Amelia Island, FL, p. 387, Sept. 1997.

[3] S.C. Douglas, A. Cichocki, and S. Amari, "Bias removal technique for blind source separation with noisy measurements," *Electron. Lett.,* vol. 34, no. 14, pp. 1379-1380, 9 July 1998.

[4] H. Attias, "EM algorithms for independent component analysis," *Proc. IEEE Workshop Neural Networks Signal Processing,* Cambridge, UK, pp. 132-141, Aug. 1998.

[5] T.-W. Lee, M.S. Lewicki, M. Girolami, and T.J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Lett.,* vol. 6, pp. 87-90, Apr. 1999.

[6] A. Hyvärinen, "Gaussian moments for noisy independent component analysis," *IEEE Signal Processing Lett.,* vol. 6, pp. 145-147, June 1999.

[7] L.-Q. Zhang, A. Cichocki, and S. Amari, "Natural gradient algorithm for blind separation of overdetermined mixture with additive noise," *IEEE Signal Processing Lett.,* vol. 6, pp. 293 - 295, Nov. 1999.

[8] S. Haykin, *Adaptive Filter Theory,* 3rd ed. (Upper Saddle River, NJ: Prentice-Hall, 1996).

[9] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing,* vol. 44, pp. 3017-3030, Dec. 1996.

[10] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Proc. Sys. 8* (Cambridge, MA: MIT Press, 1996), pp. 757-763.

[11] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing,* vol. 43, pp. 95-107, Jan. 1995.

[12] S.C. Douglas and S. Haykin, "Relationships between blind deconvolution and blind source separation," in *Unsupervised Adaptive Filtering, Vol. II: Blind Deconvolution,* S. Haykin, ed. (New York: Wiley, in press), Chap. 3.

[13] S.C. Douglas, "Numerically-robust adaptive subspace tracking using Householder transformations," *Proc. 1st IEEE Sensor Array Multichannel Workshop,* Boston, MA, Mar. 2000 (in press).

[14] S.C. Douglas, "Numerically-robust $\mathcal{O}(N^2)$ RLS algorithms using least-squares prewhitening," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Istanbul, Turkey, June 2000 (in press).

[15] A.A. Rontogiannis and S. Theodoridis, "Inverse factorization adaptive least-squares algorithms," *Signal Processing,* vol. 52, no. 1, pp. 35-47, July 1996.

[16] S.C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing,* vol. 45, pp. 2829-2842, Nov. 1997.

[17] P. Comon, "Independent component analysis: A new concept?" *Signal Processing,* vol. 36, no. 3, pp. 287-314, Apr. 1994.

[18] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal Processing,* vol. 45, no. 1, pp. 59-83, July 1995.

[19] S.-Y. Kung and C. Mejuto, "Extraction of independent components from hybrid mixture: KuicNet learning algorithm and applications," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* Seattle, WA, vol. II, pp. 1209-1212, May 1998.

[20] S.C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," accepted for publication in *IEEE Trans. Neural Networks;* to appear.

[21] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. I* (New York: Wiley, 1953).

[22] A. Edelman, T. Arias, and S.T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.,* vol. 20, no. 2, pp. 303-353, Apr. 1998.