EQUIVARIANT ALGORITHMS FOR SELECTIVE TRANSMISSION

S.C. Douglas
Department of Electrical Engineering
University of Utah
Salt Lake City, UT 84112 USA

ABSTRACT

In this paper, we consider the problem of *selective transmission*–the dual of the blind source separation task–in which a set of independent source signals are adaptively premixed prior to a non-dispersive physical mixing process so that each source can be independently monitored in the far field. We derive a stochastic gradient algorithm for iteratively-estimating the premixing matrix in the selective transmission problem, and through a simple modification, we obtain a second algorithm whose performance is equivariant with respect to the channel's mixing characteristics. We also describe an approximate version of the equivariant algorithm and other implementation issues. Simulations indicate the useful behavior of the premixing algorithms for selective transmission.

1. INTRODUCTION

In blind source separation of instantaneous mixtures, a set of m independent source signals are linearly-mixed by an unknown channel before being sensed by n sensors, and one desires to find an $(m \times n)$-dimensional separating matrix to linearly recombine these sensor signals and recover the individual sources. Fig. 1(a) shows the structure of the source separation task, in which the vectors $\mathrm{s}(k) = [s_1(k) \cdots s_m(k)]^T$, $\mathbf{x}(k) = [x_1(k) \cdots x_n(k)]^T$, and $\mathbf{y}(k) = [y_1(k) \cdots y_m(k)]^T$ contain samples of the source, sensor, and separated signals, respectively, and $\overline{\mathbf{H}}$ and $\overline{\mathbf{W}}(k)$ are the $(n \times m)$-dimensional mixing and $(m \times n)$-dimensional separating matrices, respectively. Much research effort has gone into finding simple, iterative algorithms for estimating the separating matrix $\mathbf{W}(k)$ adaptively [1]–[4]. Of particular note are the algorithms described in [3, 4] whose adaptation characteristics are independent of the mixing matrix $\mathbf{H}$. Recently, such algorithms have been extended to the multichannel deconvolution problem with some success [5].

In this paper, we consider the dual of blind source separation, a problem known as *selective transmission* [6, 7]. Fig. 1(b) shows the structure of this task. In this case, the m independent source signals in $\mathrm{s}(k)$ are *premixed* by an $(n \times m)$-dimensional matrix $\mathbf{W}(k)$ to produce a set of n transmitted signals in the vector

$$\mathbf{t}(k) = \mathbf{W}(k)\mathrm{s}(k), \tag{1}$$

[0]This material is based upon work supported in part by the National Science Foundation under Grant No. MIP-9501680.

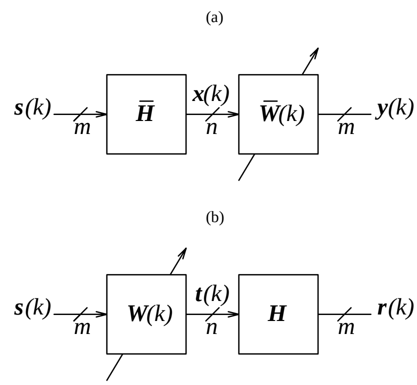

Fig. 1: Block diagrams for the (a) source separation and (b) selective transmission tasks.

and these signals are sent to a physical mixing system represented by the $(m \times n)$ matrix $\mathbf{H}$. The received signals are expressed by the vector $\mathbf{r}(k) = [r_1(k) \cdots r_m(k)]^T$ as

$$\mathbf{r}(k) = \mathbf{H}\mathbf{t}(k) = \mathbf{H}\mathbf{W}(k)\mathrm{s}(k). \tag{2}$$

The task is to iteratively adjust $\mathbf{W}(k)$ such that

$$\lim_{k \to \infty} \mathbf{H}\mathbf{W}(k) = \mathbf{P}\mathbf{D}, \tag{3}$$

where $\mathbf{P}$ is an $(m \times m)$-dimensional permutation matrix and $\mathbf{D}$ is a diagonal matrix of non-zero scaling factors d_{ii}, $1 \leq i \leq m$. The solution in (3) is similar to that for the source separation task. Unlike the source separation task, however, it is possible to determine both the ordering and scaling of the sources in $\mathbf{r}(k)$ because $\mathrm{s}(k)$ is known, such that (3) is more general than necessary. Even so, we consider (3) due to the useful algorithms that are obtained. Moreover, it is a relatively-simple matter to reorder and scale the transmitted source signals to obtain the desired source with the desired amplitude at the desired receiver.

The selective transmission problem appears in communications, in which a base station wishes to selectively transmit specific information to m different stationary receivers with fixed locations [6, 7]. In adaptive control, premixing would enable a controller to decouple its effort at m separate spatial points, thus simplifying the control law [8]. Although it might appear that blind source separation algorithms could be directly applied to the selective transmission problem, the order reversal of $\overline{\mathbf{W}}(k)\overline{\mathbf{H}}$ and $\mathbf{H}\mathbf{W}(k)$ in Fig. 1(a) and (b) leads to subtle differences in the algorithms for solving each problem. Moreover, while it is possible to estimate $\mathbf{H}$ explicitly and form its pseudo-inverse directly, such a solution may not be viable due to numerical and practical difficulties.

In this paper, we derive two stochastic gradient algorithms for iteratively-estimating the premixing matrix $\mathbf{W}(k)$ for selective transmission. Similar to blind source separation techniques [3, 4], we provide a suitable algorithm modification that yields an equivariant algorithm whose convergence behavior does not explicitly depend on the form of $\mathbf{H}$. Because implementing the equivariant algorithm requires knowledge of $\mathbf{H}$, we provide an approximate on-line version of the algorithm, and we describe methods for reordering and scaling the amplitudes of the source signals within the received signals near convergence. Simulations are provided to show the fast convergence behavior of the equivariant algorithm.

2. STOCHASTIC GRADIENT ALGORITHMS FOR SELECTIVE TRANSMISSION

To determine an algorithm for adjusting $\mathbf{W}(k)$ for selective transmission, we consider cost functions that yield (3) at their minima. In this regard, we consider one such cost function that has proven to be useful in blind source separation: the Kullback-Leibler divergence between the actual and a parametric model of the source signal distributions [9]. For selective transmission, the appropriate information-theoretic cost function is

$$\mathcal{J}(\mathbf{W}) \quad = \quad -\log p_{\mathbf{r}}(\mathbf{r}), \tag{4}$$

where the joint p.d.f. $p_{\mathbf{r}}(\mathbf{r})$ of $\mathbf{r}(k)$ is related to the joint p.d.f. $p_{\mathbf{s}}(\mathbf{s})$ of $\mathbf{s}(k)$ by a linear transformation of variables. An appropriate parametric model for $p_{\mathbf{r}}(\mathbf{r})$ is given by

$$p_{\mathbf{r}}(\mathbf{r}) \quad = \quad |\det(\mathbf{HW})| \prod_{i=1}^{m} p_i(r_i), \tag{5}$$

where $p_i(r)$ is the p.d.f. of the source signal at the ith receiver. Combining (4) and (5) in the case where $m = n$, we obtain

$$\mathcal{J}(\mathbf{W}) = -\log|\det\mathbf{H}| - \log|\det\mathbf{W}| - \sum_{i=1}^{m}\log p_i(r_i). \tag{6}$$

To minimize $\mathcal{J}(\mathbf{W})$ in (6), we first consider a simple stochastic gradient algorithm of the form

$$\mathbf{W}(k+1) \quad = \quad \mathbf{W}(k) - \mu(k)\frac{\partial\mathcal{J}(\mathbf{W}(k))}{\partial\mathbf{W}}, \tag{7}$$

where $\mu(k)$ is a step size sequence. It can be shown that $\partial\log|\det\mathbf{W}|/\partial\mathbf{W} = \mathbf{W}^{-T}$ [3]. Differentiating the third term on the RHS of (6), we obtain

$$-\frac{\partial\sum_{i=1}^{m}\log p_i(r_i)}{\partial w_{ij}} \quad = \quad \sum_{l=1}^{m} h_{li} f_l(r_l) s_j, \tag{8}$$

where we have defined $f_i(r) = -\partial\log p_i(r)/\partial r$. Thus, collecting the update terms yields the algorithm as

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left[\mathbf{W}^{-T}(k) - \mathbf{H}^T\mathbf{f}(\mathbf{r}(k))\mathbf{s}^T(k)\right], \tag{9}$$

where $\mathbf{f}(\mathbf{r}(k)) = [f_1(r_1(k)) \cdots f_m(r_m(k))]^T$.

Alternatively, we can derive an algorithm for adjusting $\mathbf{W}(k)$ that minimizes the mean-squared error (MSE) between the source and received signals, since $\mathbf{s}(k)$ is known at the transmitter. The associated instantaneous cost function is

$$\check{\mathcal{J}}(\mathbf{W}) = ||\mathbf{e}||^2, \qquad \mathbf{e} = \mathbf{s} - \mathbf{r}, \tag{10}$$

and $\mathbf{e}$ is an m-dimensional vector of error signals. Following a similar derivation, the resulting update is

$$\mathbf{W}(k+1) \quad = \quad \mathbf{W}(k) + \mu(k)\mathbf{H}^T\mathbf{e}(k)\mathbf{s}^T(k). \tag{11}$$

The algorithms in (9) and (11) have several practical disadvantages. Firstly, they explicitly depend on the channel matrix $\mathbf{H}$, which is unknown. Secondly, as is shown via simulations, their convergence performances are highly-sensitive to the form of $\mathbf{H}$ as well as to the scaling of the sources in $\mathbf{s}(k)$. Finally, (9) requires the inverse of the premixing matrix, which is challenging to compute. In the following two sections, we provide modifications to the algorithm in (9) that overcome these limitations.

3. EQUIVARIANT ALGORITHMS FOR SELECTIVE TRANSMISSION

Recent advances in blind source separation have uncovered algorithms whose behaviors are independent of the mixing process. This remarkable property is called *equivariance* [4]. An equivariant blind source separation algorithm can be written in terms of the combined system $\overline{\mathbf{C}}(k) = \overline{\mathbf{W}}(k)\overline{\mathbf{H}}$ and the source signals in $\mathbf{s}(k)$, such that $\overline{\mathbf{H}}$ does not enter into the evolution of $\overline{\mathbf{C}}(k)$. Such algorithms can be obtained using either the *relative gradient* [4] or the *natural gradient* [3, 10, 11] concepts. While these two concepts differ in general [12], they are identical in the case of blind source separation.

In this section, we determine an equivariant algorithm for selective transmission. For this derivation, we use the methodology outlined in [5, 10] to determine the coefficient updates. Consider the differential of the cost function $\mathcal{J}(\mathbf{W})$ in (6), as given by

$$d\mathcal{J}(\mathbf{W}) \quad = \quad \mathcal{J}(\mathbf{W} + d\mathbf{W}) - \mathcal{J}(\mathbf{W}). \tag{12}$$

It can be shown that

$$d\mathcal{J}(\mathbf{W}) \quad = \quad -\mathrm{tr}[\mathbf{W}^{-1}d\mathbf{W}] + \mathbf{f}^T(\mathbf{r})d\mathbf{r}, \tag{13}$$

where $\mathbf{f}(\mathbf{r})$ is as defined previously and $d\mathbf{r} = \mathbf{H}d\mathbf{W}\mathbf{s}$. Define the modified coefficient differential $d\mathbf{X}$ as

$$d\mathbf{X} \quad = \quad \mathbf{W}^{-1}d\mathbf{W}. \tag{14}$$

We can evaluate (13) in terms of $d\mathbf{X}$ as

$$d\mathcal{J}(\mathbf{W}) \quad = \quad -\mathrm{tr}[d\mathbf{X}] + \mathbf{f}^T(\mathbf{r})\mathbf{HW}d\mathbf{X}\mathbf{s}. \tag{15}$$

Due to the form of (14), $d\mathbf{X}$ describes an infinitesimal direction that is not the gradient of any well-defined surface. Even so, it is useful to consider $d\mathbf{X}$ as a search direction for minimizing $\mathcal{J}(\mathbf{W})$ when translated into the space of premixing matrices $\mathbf{W}$. Using (14) and (15), the proposed algorithm for $\mathbf{W}(k)$ is

$$\mathbf{W}(k+1) \quad = \quad \mathbf{W}(k) - \mu(k)\mathbf{W}(k)\frac{\partial\mathcal{J}(\mathbf{W}(k))}{\partial\mathbf{X}} \tag{16}$$

$$\frac{\partial\mathcal{J}(\mathbf{W})}{\partial\mathbf{X}} \quad = \quad -\mathbf{I} + \mathbf{W}^T\mathbf{H}^T\mathbf{f}(\mathbf{r})\mathbf{s}^T. \tag{17}$$

The proposed update is thus

$$\mathbf{W}(k+1) = \mathbf{W}(k)+\mu(k)\mathbf{W}(k)\big[\mathbf{I}-\mathbf{C}^T(k)\mathbf{f}(\mathbf{r}(k))\mathbf{s}^T(k)\big] \quad (18)$$

where $\mathbf{C}(k)$, the combined premixing-channel matrix, is

$$\mathbf{C}(k) \quad = \quad \mathbf{HW}(k). \quad (19)$$

Comparing (9) and (18), we see that the algorithm in (18) could also be obtained using the rule

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \mu(k)\mathbf{W}(k)\mathbf{W}^T(k)\frac{\partial \mathcal{J}(\mathbf{W}(k))}{\partial \mathbf{W}(k)}. \quad (20)$$

Since $\mathbf{W}(k)\mathbf{W}^T(k)$ is always positive-definite when $\mathbf{W}(k)$ is non-singular, (18) has the same local stability properties about a selective transmission solution as those of the algorithm in (9) for small step size values.

The usefulness of (18) is seen if one considers the update in the combined system matrix $\mathbf{C}(k)$. Premultiplying both sides of (18) by $\mathbf{H}$, we obtain

$$\begin{aligned}&\mathbf{C}(k+1)\\ &\quad= \mathbf{C}(k) + \mu(k)\mathbf{C}(k)\big[\mathbf{I} - \mathbf{C}^T(k)\mathbf{f}(\mathbf{C}(k)\mathbf{s}(k))\mathbf{s}^T(k)\big] \end{aligned} \quad (21)$$

which does not explicitly depend on $\mathbf{H}$. This result proves that the algorithm in (18) is equivariant. The algorithm in (18) also has the added benefit that the matrix inverse $\mathbf{W}^{-T}(k)$ is no longer required. Moreover, (18) can be applied when $m \leq n$, in which case $\mathbf{W}(k)$ lies in the column space of $\mathbf{W}(0)$ for all k [13].

Although one could apply the modified update in (20) to the MSE cost function in (10), simulations of the resulting algorithm indicate that it can converge to spurious local minima in which $\mathbf{HW}(k)$ has one or more zero eigenvalues. By contrast, (18) never converged to such minima in any of our extensive tests, indicating that the algorithm shares the apparent robust behavior of similar blind source separation algorithms. While these behaviors are currently under study, they suggest that thorough testing of an algorithm under different signal and channel conditions is necessary when employing the modification in (20) to any particular cost function for selective transmission.

4. IMPLEMENTATION ISSUES

As in blind source separation, the local stability of the algorithm in (18) about any coefficient solution satisfying (3) depends on the nonlinearities in $\mathbf{f}(\mathbf{r})$ and on the the p.d.f.'s of the source signals $\{s_i(k)\}$. In [13], the stability conditions for (18) are derived and are shown to be similar to those of the dual algorithm for blind source separation in [3, 10]. Thus, one can use the same nonlinearity $f(r)$ for each entry of $\mathbf{f}(\mathbf{r})$ for certain classes of source distributions. For example, if all of the sources have a negative kurtosis, then $f(r) = r^3$ will yield a stable selective transmission solution for $\mathbf{W}(k)$.

Considering (21), it is impossible to scale each of the sources in $\mathbf{s}(k)$ arbitrarily and then absorb this scaling into the combined system matrix $\mathbf{C}(k)$. Therefore, unlike that of blind source separation algorithms, the performance of (18) depends on the absolute scales of the sources. To obtain better convergence performance, we propose the following signal normalization scheme:

$$s_i(k) \quad = \quad \widehat{s}_i(k)/\psi(\widehat{\rho}_i(k)) \quad (22)$$

$$\widehat{\rho}_i(k) \quad = \quad (1-\alpha)\widehat{\rho}_i(k) + \alpha\widehat{s}_i(k)f(\widehat{s}_i(k)), \quad (23)$$

where $\widehat{s}_i(k)$ are the original source signals before scaling, α is a small constant, and $\psi(s)$ is the inverse of the function $sf(s)$ for $s \geq 0$. With this signal normalization, the amplitudes of the signals in $\mathbf{r}(k)$ are the same on average as those in $\mathbf{s}(k)$ at convergence, *i.e.,* the absolute values of the scaling factors d_{jj} of $\mathbf{D}$ in (3) are unity-valued. With this choice, excellent convergence behavior for (18) is provided, independently of the choice of $\mathbf{W}(0)$.

In practice, individual amplitude control over each of the m received signals is desired. Such control can be provided without changing the equivariant performance of the system by replacing $\mathbf{f}(\mathbf{r}(k))$ in (18) with $\mathbf{A}^T\mathbf{f}(\mathbf{r}(k))$, where $\mathbf{A}$ is a diagonal matrix of amplitudes a_{ii}. It can be shown that such an algorithm depends only on the combined system matrix $\mathbf{AHW}(k)$, and $\mathbf{AHW}(k)$ approximately converges to $\mathbf{PJ}$ if (22)–(23) is used, where $\mathbf{J}$ is a diagonal matrix of ± 1 values. Therefore, the amplitude of the source at the ith receiver is scaled by a_{ii}^{-1} at convergence.

The algorithm in (18) employs the unknown mixing channel $\mathbf{H}$. While $\mathbf{H}$ could be estimated prior to the application of (18), such a strategy defeats the purpose of the selective transmission task. Since the combined premixing-channel matrix $\mathbf{C}(k)$ appears in (18), we can develop a simple technique for estimating this matrix for use within (18). If the signal scaling method in (22)–(23) is used, then a simple estimator for $\mathbf{C}(k)$ is

$$\widehat{\mathbf{C}}(k) \quad = \quad (1-\beta)\widehat{\mathbf{C}}(k-1) + \beta\mathbf{r}(k)\mathbf{f}(\mathbf{s}^T(k)), \quad (24)$$

where β is a small constant. In practice, $\widehat{\mathbf{C}}(k)$ is used in place of $\mathbf{C}(k)$ in (18). Since $\widehat{\mathbf{C}}(k)$ is not exactly equal to $\mathbf{C}(k)$, the performance of this approximate algorithm is somewhat different from that of the equivariant algorithm in (18), as is shown via simulations.

For the selective transmission task, one must broadcast a specific source to a particular receiver. The algorithm in (18) does not guarantee a specific ordering or the proper signs of the source signals in $\mathbf{r}(k)$. However, since $\widehat{\mathbf{C}}(k)$ in (24) converges to $\mathbf{PJ}$ if the signal scaling in (22)–(23) is used, we can use $\widehat{\mathbf{C}}(k)$ to reorder and change the signs of the source signals prior to the premixing process. Define $\widehat{\mathbf{P}}(k)$ as the matrix obtained by replacing each $\widehat{c}_{ij}(k)$ in $\widehat{\mathbf{C}}(k)$ by either 1, 0, or -1, whichever is closest in value. By using $\widehat{\mathbf{P}}^T(k)\mathbf{s}(k)$ in place of $\mathbf{s}(k)$ within the system, we obtain the desired source with the desired sign at the desired receiver. When the same nonlinearity is used for each entry of $\mathbf{f}(\mathbf{r}(k))$ and the scaling method in (22)–(23) is employed, no other change to the algorithm is required. In fact, the order and signs of the sources can be arbitrarily switched at will in this case without any adverse effects to the convergence behavior of the system.

5. SIMULATIONS

We now explore the behaviors of the various algorithms via simulation. For the examples in this section, we generate three random i.i.d. source signals $\{s_i(k)\}$ with specific negative-kurtosis distributions, and $f_i(r) = r^3$ in these examples. The behaviors of the algorithms in (9), (11), (18)–(19), and $\{$(18), (24)$\}$ are simulated for different choices of $\mathbf{H}$. In each case, we calculate the performance factor

$$\gamma(k) \quad = \quad \min_{\{i_1,i_2,i_3\}\in\{1,2,3\}} \frac{1}{3}\sum_{l=1}^{3}\sum_{j=1,j\neq i_l}^{3}\frac{c_{i_l j}^2(k)}{c_{i_l i_l}^2(k)}, \quad (25)$$

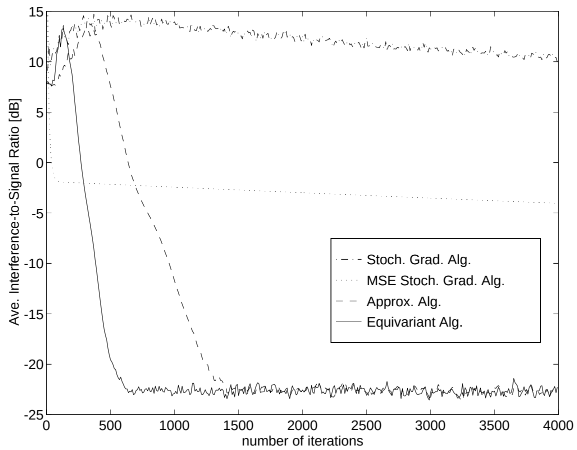

Fig. 2: Convergence of the average value of $\gamma(k)$ for the four algorithms in the first example.

where $i_1 \neq i_2 \neq i_3 \neq i_1$, as averaged over 100 different simulation runs. We select $\mathbf{W}(0)$ as a random orthogonal matrix with eigenvalues equal to δ for each simulation run.

In the first example, each of the sources is binary-$\{\pm 1\}$-distributed, so that source scaling is not required. The channel matrix $\mathbf{H}$ for these simulations was chosen as

$$\mathbf{H} = \begin{bmatrix} 0.3 & 0.4 & 0.4 \\ -0.1 & -0.2 & 0.2 \\ -0.3 & -0.4 & -0.5 \end{bmatrix}. \quad (26)$$

The condition number of $\mathbf{H}\mathbf{H}^T$ in this case is 18924, indicating a severely ill-conditioned channel. For (9), (18)–(19), and $\{(18), (24)\}$, step sizes of 0.05, 0.02 and 0.0085 are chosen such that $\gamma(k) \approx 0.0055$ in steady-state in each case, and for (11), $\mu(k) = 0.7$ is chosen to provide the fastest convergence without instability. Fig. 2 shows the performance of each algorithm in this case, where $\beta = 0.025$ and $\delta = 0.5$. It is seen that the equivariant algorithm has the best performance, followed by the approximate equivariant and the two stochastic gradient algorithms, respectively. Moreover, using $\widehat{\mathbf{C}}(k)$ in place of $\mathbf{C}(k)$ in (18) does not significantly degrade this algorithm's performance.

We now explore the behavior of $\{(18), (24)\}$ when the signal scaling in (22)–(23) is employed. In this case, we generate $\widehat{s}_1(k)$ and $\widehat{s}_2(k)$ as i.i.d. binary-$\{\pm 1\}$ and uniform-$[-1, 1]$ random variables, respectively, and the distribution of the i.i.d. signal $\widehat{s}_3(k)$ is $\widehat{p}_3(s) = 1 - |s|$ for $|s| < 1$ and is zero otherwise. The premixing matrix $\mathbf{H}$ employed here is taken from [14], where the condition number of $\mathbf{H}\mathbf{H}^T$ is 422.4. Fig. 3 shows the evolution of the average value of the performance factor for three versions of $\{(18), (24)\}$, where $\alpha = 0.01$, $\beta = 0.03$, $\delta = 1$, and $\mu(k) = 0.01$. In addition to the original algorithm, we show the behavior of the algorithm when $\mathbf{f}(\mathbf{r}(k))$ is replaced by $\mathbf{A}^T\mathbf{f}(\mathbf{r}(k))$ with $\mathbf{A} = \mathrm{diag}\{1, 0.1, 0.01\}$, and we show the behavior of the algorithm where we have randomly-selected the order of the normalized sources $s_i(k)$ within $\mathbf{s}(k)$ at *every* time step. In this case, the performance of the approximate algorithm is found to be nearly independent of the amplitudes of the nonlinearities in $\mathbf{f}(\mathbf{r}(k))$ as well as to the order of the source signals so long as the signal scaling in (22)–(23) is employed. Moreover, using $\mathbf{A}^T\mathbf{f}(\mathbf{r}(k))$ in the updates caused the source signals to have amplitudes of 1, 10, and 100, respectively, at the three receivers, as predicted.

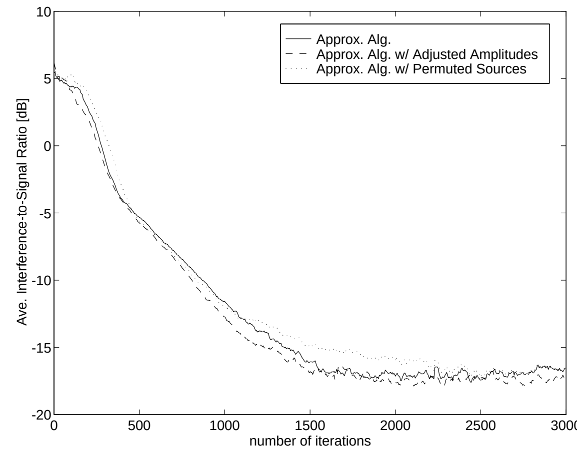

Fig. 3: Convergence of the average value of $\gamma(k)$ for the three algorithm versions in the second example.

6. CONCLUSIONS

This paper describes an equivariant algorithm for selective transmission, the dual of the blind source separation task. We have also described an approximate equivariant algorithm that does not require exact knowledge of the channel to operate, and we have addressed other practical implementation issues. Simulations have been provided to indicate the useful behavior of the algorithms. Additional details about the algorithms, as well as a formal stability analysis, is provided in [13].

REFERENCES

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

[2] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.,* vol. 7, no. 6, pp. 1129-1159, Nov. 1995.

[3] 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.

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

[5] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution using the natural gradient," *Proc. IEEE Workshop Signal Proc. Adv. Wireless Commun.,* Paris, France, pp. 101-104, Apr. 1997.

[6] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array system for spatially selective transmission," *IEEE Trans. Vehicular Tech.,* vol. 44, pp. 651-660, Aug. 1995.

[7] H. Liu and G. Xu, "Multiuser blind channel estimation and spatial channel pre-equalization," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Detroit, MI, vol. 3, pp. 1756-1759, May 1995.

[8] B. Widrow and E. Walach, *Adaptive Inverse Control* (Upper Saddle River, NJ: Prentice-Hall PTR, 1996).

[9] J.-F. Cardoso, "Infomax and maximum likelihood in source separation," *IEEE Signal Processing Lett.,* vol. 4, pp. 112-114, Apr. 1997.

[10] S. Amari, "Natural gradient works efficiently in learning," accepted for publication in *Neural Networks;* to appear.

[11] S.C. Douglas and S. Amari, "Natural gradient adaptation," *Proc. IEEE,* vol. 86, no. 6, June 1998 (to appear).

[12] J.-F. Cardoso, private communication, July 1997.

[13] S.C. Douglas, "Equivariant adaptive selective transmission," submitted to *IEEE Trans. Signal Processing.*

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

[0] The author acknowledges helpful discussions with S. Amari and A. Cichocki, RIKEN, Japan, regarding blind source separation topics.