RELEVANT DECONVOLUTION FOR ACOUSTIC SOURCE ESTIMATION

Yuanqing Lin, Daniel D. Lee

GRASP Laboratory, Department of Electrical and Systems Engineering,
University of Pennsylvania, Philadelphia, PA 19104

ABSTRACT

We describe a robust deconvolution algorithm for simultaneously estimating an acoustic source signal and convolutive filters associated with the acoustic room impulse responses from a pair of microphone signals. In contrast to conventional blind deconvolution techniques which rely upon a knowledge of the statistics of the source signal, our algorithm exploits the nonnegativity and sparsity structure of room impulse responses. The algorithm is formulated as a quadratic optimization problem with respect to both the source signal and filter coefficients, and proceeds by iteratively solving the optimization in two alternating steps. In the *H-step*, the nonnegative filter coefficients are optimally estimated within a Bayesian framework using a relevant set of regularization parameters. In the *S-step*, the source signal is estimated without any prior assumption on its statistical distribution. The resulting estimates converge to a *relevant* solution exhibiting appropriate sparseness in the filters. Simulation results indicate that the algorithm is able to precisely recover both the source signal and filter coefficients, even in the presence of large ambient noise.

1. INTRODUCTION

The original motivation for this work was to accurately estimate the time difference of arrival between reverberant acoustic signals. This scenario is depicted in Fig.1 where the signals are measured by a pair of microphones. A single acoustic source signal $s(t)$ impinges on the two microphones, and the observed signals $x_m(t), m = 1, 2$ are given by the convolution of the source $s(t)$ with the corresponding acoustic room impulse responses $h_m(t), m = 1, 2$:

$$x_m(t) = \int dt' h_m(t') s(t - t') + n_m(t), \quad m = 1, 2 \qquad (1)$$

where $n_m(t)$ is random additive noise in the microphones. Theoretical models of the acoustic reflections indicate that the acoustic room impulse responses $h_m(t)$ should be nonnegative and display a sparse structure [1]. In recent work [2, 3], we used nonnegative deconvolution to estimate the filter coefficients when the source signal was known. In this submission, we describe a new algorithm based upon Bayesian regularization and nonnegativity constraints to estimate both an *unknown* source signal as well as the appropriate sparse filter coefficients.

The problem of simultaneously estimating unknown source signals and unknown filters from their convolved measurements has been extensively studied in the last decade. Most current techniques for *blind* deconvolution exploit some knowledge of the statistics of the source signal [4, 5, 6, 7]. These algorithms typically rely upon quantities such as higher order correlations in the estimated source signal to guide the blind deconvolution process. But in order to accurately calculate these statistics, large amounts of data need to be collected. In rapidly changing acoustic environments such as with a moving source, these algorithms may not be appropriate. Moreover, most of these blind deconvolution algorithms are also not very robust to the presence of noise.

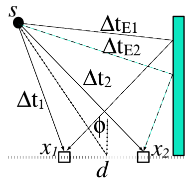

Fig. 1. An acoustic source signal $s(t)$ is measured by two microphones in a reverberant environment. The observed signals $x_1(t)$ and $x_2(t)$ consist of time delayed direct path signals, as well as echoes and ambient noise.

In the following work, we propose a relevant deconvolution framework for accurately resolving a single acoustic source signal $s(t)$ as well as the room impulse responses $h_m(t)$ from two convoluted measurements $x_m(t)$. Our algorithm does not assume anything about the nature of the source signal $s(t)$, and instead relies upon the sparse, nonnegative structure of the filters $h_m(t)$. Mathematically, our algorithm optimizes the following likelihood cost function with respect to both the source $s(t)$ and nonnegative filter $h_m(t)$:

$$\min_{h_m(t) \geq 0, s(t)} \sum_{m=1}^{2} \int dt \, |x_m(t) - h_m(t) * s(t)|^2 + \hat{\lambda}_m(t) h_m(t), \qquad (2)$$

where $*$ denotes convolution, and $\hat{\lambda}_m(t)$ are L_1-norm regularization parameters. The deconvolution algorithm proceeds by alternatively optimizing the estimated filter parameters (*H-step*) and the estimated source signal (*S-step*). The *H-step* consists of solving the non-negative least squares optimization for the filter coefficients while estimating the relevant regularization parameters within a Bayesian framework. In the *S-step*, the current filter estimates are used to recalculate the estimated source signal. Because the algorithm does not rely upon calculating source statistics and explicitly takes noise into account in the Bayesian regularization, it is quite computationally efficient and robust.

The remainder of the paper is arranged as follows. In Section 2, we describe the Bayesian regularization and nonnegative deconvolution procedure which forms the *H-step* in the relevant deconvolution algorithm. Then in Section 3, we introduce the update rule for iteratively estimating the source signal. The performance of our relevant deconvolution algorithm is shown in Section 4, and finally discussed in Section 5.

2. *H-STEP:* BAYESIAN REGULARIZATION AND NONNEGATIVE DECONVOLUTION (BRAND)

The *H-Step* of the deconvolution algorithm estimates the most relevant filter coefficients given the current source estimate. Within the context of a probabilistic Bayesian framework [8], the filter estimation is performed as a quadratic optimization with nonnegative constraints. The signals in Eq. 2 are first sampled in the discrete time domain, resulting in the matrix form:

$$\min_{\alpha^{(m)}\geq 0,\mathbf{s}} \sum_{m=1}^{2} \frac{1}{2}\|\mathbf{x}_m - \mathbf{S}^{(m)}\alpha^{(m)}\|^2 + (\hat{\lambda}^{(m)})^T\alpha^{(m)} \quad (3)$$

where $\mathbf{x}_m = [x_m(t_1)\ x_m(t_2)\ \ldots\ x_m(t_N)]^T$ is a $N \times 1$ vector containing the measured signal in the m-th microphone, and $\mathbf{S}^{(m)} = [\mathbf{s}(t - \Delta t_1^{(m)})\ \mathbf{s}(t - \Delta t_2^{(m)})\ \ldots\ \mathbf{s}(t - \Delta t_{M_m}^{(m)})]$ is a $N \times M_m$ matrix consisting of delayed patterns of the estimated source signal $\mathbf{s}(t) = [s(t_1)\ s(t_2)\ \ldots\ s(t_N)]^T$ as column vectors. The set of time delays is given by $\{\Delta t_i\}_{i=1}^{M_m}$, and $\alpha^{(m)}$ are the discrete samples of impulse responses $h_m(t)$ at those time delays. $\hat{\lambda}^{(m)}$ is a $M_m \times 1$ vector, where the i^{th} component corresponds to the Bayesian regularization parameter for $\alpha_i^{(m)}$.

Given the current estimate of the source $\mathbf{s}$, the best estimate of the filter coefficients is calculated by optimizing:

$$\min_{\alpha^{(m)}\geq 0} \frac{1}{2}\|\mathbf{x}_m - \mathbf{S}^{(m)}\alpha^{(m)}\|^2 + (\hat{\lambda}^{(m)})^T\alpha^{(m)} \quad m = 1, 2. \quad (4)$$

In order to properly define the regularization parameters, we show how this optimization arises from a probabilistic generative model. In the following, we omit the channel number $m = 1, 2$ from our notation since both channels are treated equivalently.

The probabilistic model assumes the measured signal $x(t)$ is contaminated by additive Gaussian white noise with zero-mean and covariance σ^2:

$$P(\mathbf{x}|\mathbf{S}, \alpha, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{S}\alpha\|^2\right). \quad (5)$$

Sparseness in the filter coefficients is achieved using independent exponential prior distributions. The priors only allow nonnegative values and their sharpness is controlled by the regularization parameters $\lambda = [\lambda_1\ \lambda_2\ \ldots\ \lambda_M]^T$:

$$P(\alpha|\lambda) = \prod_{i=1}^{M} \lambda_i \exp\{-\lambda_i\alpha_i\},\ \ \alpha \geq 0\,. \quad (6)$$

Rather than manually setting the regularization parameters σ^2 and λ, they are inferred from the data by maximizing the posterior distribution:

$$P(\lambda, \sigma^2|\mathbf{x}, \mathbf{S}) = \frac{P(\mathbf{x}|\lambda, \sigma^2, \mathbf{S})P(\lambda, \sigma^2)}{P(\mathbf{x}|\mathbf{S})}. \quad (7)$$

Assuming a flat prior for $P(\lambda, \sigma^2)$ [9], estimating σ^2 and λ is then equivalent to maximizing the likelihood:

$$\begin{aligned} P(\mathbf{x}|\lambda, \sigma^2, \mathbf{S}) &= \int_{\alpha\geq 0} d\alpha P(\mathbf{x}|\mathbf{S}, \alpha, \sigma^2)P(\alpha|\lambda) \quad (8) \\ &= \frac{\prod_i \lambda_i}{(2\pi\sigma^2)^{N/2}} \int_{\alpha\geq 0} d\alpha\ \exp[-F(\alpha)] \end{aligned}$$

where

$$F(\alpha) = \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{S}\alpha)^T(\mathbf{x} - \mathbf{S}\alpha) + \lambda^T\alpha. \quad (9)$$

Since the integral in Eq. 8 cannot be directly maximized, we derive the following iterative update rules for λ and σ^2 using Expectation-Maximization (EM):

$$\frac{1}{\lambda_i} \longleftarrow \int_{\alpha\geq 0} d\alpha\, \alpha_i Q(\alpha) \quad (10)$$

$$\sigma^2 \longleftarrow \frac{1}{N}\int_{\alpha\geq 0} d\alpha\, (\mathbf{x} - \mathbf{S}\alpha)^T(\mathbf{x} - \mathbf{S}\alpha)Q(\alpha) \quad (11)$$

where the expectations are taken over the distribution

$$Q(\alpha) = \frac{\exp[-F(\alpha)]}{\mathcal{Z}_\alpha}\ , \quad (12)$$

with normalization $\mathcal{Z}_\alpha = \int_{\alpha\geq 0} d\alpha\ \exp[-F(\alpha)]$. Since the integrals in Eq. 10 and Eq. 11 are still intractable, we make a factorized approximation for $Q(\alpha)$.

The maximum likelihood estimate for α^{ML} are determined by solving the nonnegative quadratic programming (NNQP) problem:

$$\min_{\alpha\geq 0} \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{S}\alpha)^T(\mathbf{x} - \mathbf{S}\alpha) + \lambda^T\alpha. \quad (13)$$

where the linear term is related to Eq. 4 by $\hat{\lambda} = \sigma^2\lambda$. This optimization can be solved using either a modified simplex method or multiplicative updates as we have shown previously [3]. Using this solution, we approximate the distribution $Q(\alpha)$ with the factorized form:

$$Q(\alpha) \approx Q_I(\alpha_I)Q_J(\alpha_J) \quad (14)$$

where the vector α is partitioned into two distinct subsets α_I and α_J, consisting of components $i \in I$ such that $(\alpha^{ML})_i = 0$, and components $j \in J$ such that $(\alpha^{ML})_j > 0$, respectively.

Since the non-zero components α_J are not greatly restricted by nonnegativity constraints, $Q_J(\alpha_J)$ is approximated by the unconstrained Gaussian with mean α_J^{ML} and inverse covariance given by the submatrix $\mathbf{A_{JJ}}$ of $\mathbf{A} = \frac{1}{\sigma^2}\mathbf{S}^T\mathbf{S}$.

The other components α_I are restricted by nonnegativity to only vary in the positive direction, so their marginal distribution is given by the following functional form:

$$Q_I(\alpha_I) \propto \exp[-(\mathbf{A}\alpha_{ML} + \mathbf{b})_I^T\alpha_I - \frac{1}{2}\alpha_I^T\mathbf{A}_{II}\alpha_I],\ \ \alpha_I \geq 0. \quad (15)$$

To calculate approximate expectations over this distribution, we use an independent exponential distribution:

$$\hat{Q}_I(\alpha_I) = \prod_{i\in I} \frac{1}{\mu_i} e^{-\alpha_i/\mu_i},\ \ \alpha_i \geq 0, \mu_i \geq 0 \quad (16)$$

By minimizing the KL-divergence between $\hat{Q}_I(\alpha_I)$ and $Q_I(\alpha_I)$, we obtain the mean-field parameters μ_i.

With the factorized approximation $Q(\alpha) = \hat{Q}_I(\alpha_I)Q_J(\alpha_J)$, the expectations in Eqs. 10–11 can be analytically calculated. The mean value of α under this distribution is given by:

$$\bar{\alpha}_i = \begin{cases} \alpha_i^{ML} & \text{if} \quad i \in J \\ \mu_i & \text{if} \quad i \in I \end{cases} \quad (17)$$

and its covariance $\mathbf{C}$ is:

$$C_{ij} = \begin{cases} (\mathbf{A_{JJ}}^{-1})_{ij} & \text{if } i,j \in J \\ \mu_i^2 \delta_{ij} & \text{otherwise} \end{cases}$$

The update rules for λ and σ^2 are then given by:

$$\lambda_i \quad \longleftarrow \quad \frac{1}{\bar{\alpha}_i} \tag{18}$$

$$\sigma^2 \quad \longleftarrow \quad \frac{1}{N}[(\mathbf{x} - \mathbf{S}\bar{\alpha})^T(\mathbf{x} - \mathbf{S}\bar{\alpha}) + \mathrm{Tr}(\mathbf{S}^T\mathbf{S}\mathbf{C})] \tag{19}$$

To initialize the regularization parameters in λ, we start by assuming that they are all uniform instead of being independent. This improves the global convergence of the algorithm since there are fewer optimization parameters. With a uniform prior on the Bayesian regularization, namely:

$$P(\alpha|\lambda') = (\lambda')^M \exp\{-\lambda' \sum_i \alpha_i\}, \ \ \alpha \geq 0\ , \tag{20}$$

the Bayesian update rules are similar to the independent case except that Eq. 9, Eq. 10 and Eq. 18 become

$$F(\alpha) = \frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{S}\alpha)^T(\mathbf{x} - \mathbf{S}\alpha) + \lambda' \mathbf{e}^T \alpha. \tag{21}$$

$$\frac{1}{\lambda'} \quad \longleftarrow \quad \frac{1}{M}\int_{\alpha \geq 0} d\alpha\, \mathbf{e}^T \alpha Q(\alpha) \tag{22}$$

$$\lambda' \quad \longleftarrow \quad \frac{M}{\sum_i \bar{\alpha}_i} \tag{23}$$

respectively, where $\mathbf{e} = [1\ 1\ 1\ \ldots\ 1]^T$. Our algorithm proceeds by initially beginning with a uniform Bayesian regularization for the first few iterations, and then the independent regularization is used to further refine the solution.

3. *S-STEP:* SOURCE UPDATE RULE

The alternating *S-step* of the deconvolution algorithm optimizes the most probable source signal $\mathbf{s}$ with respect to the current estimate of the filter parameters $\alpha^{(m)}$ $(m = 1, 2)$ from Eq. 3. The optimal source is derived from the optimization:

$$\min_{\mathbf{s}} \sum_{m=1}^{2} \frac{1}{2}\|\mathbf{x}_m - \mathbf{A}_m \mathbf{s}\|^2, \tag{24}$$

where $\mathbf{A}^m$ is a Toeplitz matrix containing the nonnegative filter coefficients of the m-th room impulse response. This quadratic optimization can be solved analytically, giving the estimate: $\mathbf{s} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}$. However, since the dimensionality of $\mathbf{s}$ can be very large, a direct pseudo-inverse computation can be very costly. We employ an alternative algorithm for computing $\mathbf{s}$ by splitting the variables $\mathbf{s} = \mathbf{s}^+ - \mathbf{s}^-$ where both $\mathbf{s}^+$ and $\mathbf{s}^-$ are nonnegative, and by solving the resulting nonnegative quadratic programming problem using a multiplicative update rule. These updates do not require the adjustment of any rate parameters, and can also easily incorporate the addition of source priors in the optimization.

As a standard nonnegative quadratic programming problem, Eq. 24 becomes:

$$\min_{\hat{\mathbf{s}} \geq 0} \frac{1}{2}\hat{\mathbf{s}}^T \mathbf{H} \hat{\mathbf{s}} + \mathbf{b}^T \hat{\mathbf{s}} \tag{25}$$

where $\hat{\mathbf{s}} = [\mathbf{s}^+; \mathbf{s}^-]$, $\mathbf{b}^T = [-\sum_{m=1}^{2} \mathbf{x}_m^T \mathbf{A}_m \ \ \sum_{m=1}^{2} \mathbf{x}_m^T \mathbf{A}_m]$, and

$$\mathbf{H} = \sum_{m=1}^{2} \begin{bmatrix} \mathbf{A}_m^T \mathbf{A}_m & -\mathbf{A}_m^T \mathbf{A}_m \\ -\mathbf{A}_m^T \mathbf{A}_m & \mathbf{A}_m^T \mathbf{A}_m \end{bmatrix}. \tag{26}$$

The multiplicative updates for solving $\hat{\mathbf{s}}$ are

$$\hat{s}_i \leftarrow \hat{s}_i \left[\frac{-b_i + \sqrt{b_i^2 + 4(\mathbf{H}^+\hat{\mathbf{s}})_i(\mathbf{H}^-\hat{\mathbf{s}})_i}}{2(\mathbf{H}^+\hat{\mathbf{s}})_i} \right]. \tag{27}$$

where $\mathbf{H} = \mathbf{H}^+ - \mathbf{H}^-$ is the decomposition of the matrix into its positive and negative components. Due to the Toeplitz structure of $\mathbf{A}_m$, the matrix-vector multiplications of $\mathbf{H}^+\hat{\mathbf{s}}$ and $\mathbf{H}^-\hat{\mathbf{s}}$ can be efficiently computed using fast Fourier transformations (FFTs).

There is a uniform time delay and scaling factor that is invariant to the deconvolution optimization. We fix these factors by choosing the filter coefficient of the direct path propagation of one of the channels to have zero time delay and a fixed unity amplitude.

In summary, the complete algorithm for relevant deconvolution is:

1. Initialize σ_1^2, σ_2^2, $\lambda^{(1)}$, $\lambda^{(2)}$, $\mathbf{s}$, and the discrete time delays $\{\Delta t_i^{(1)}\}$ and $\{\Delta t_i^{(2)}\}$. Without loss of generality, $\{\Delta t_i^{(2)}\} \geq 0$ while $\{\Delta t_i^{(1)}\}$ may be either positive or negative.
2. Solve the nonnegative quadratic program problem in Eq. 4 for the signal $\mathbf{x}_2$ to estimate $\alpha^{(2)}$. The estimated signals are scaled appropriately so that $\alpha^{(2)}(\Delta t = 0) = 1$. Then σ_2^2 and $\lambda^{(2)}$ are re-estimated based upon the current estimates of $\mathbf{s}$ and $\alpha^{(2)}$.
3. Solve nonnegative quadratic program problem in Eq. 4 for the signal $\mathbf{x}_1$ to estimate $\alpha^{(1)}$. Then σ_1^2 and $\lambda^{(1)}$ are re-estimated based upon the current estimates of $\mathbf{s}$ and $\alpha^{(1)}$.
4. Repeat Steps 2-3 with a uniform regularization prior, and then with an independent regularization prior.
5. A new estimate for the source $\mathbf{s}$ is computed from Eq. 27 using the previous estimate as an initial value.
6. Go back to Step 2 until convergence.

4. SIMULATION RESULT

In this section, the performance of the relevant deconvolution algorithm is illustrated using a speech recording as a source signal. The speech was sampled at 16 kHz, and 2048 samples were used as shown in Fig. 2. The source signal was convolved with two filters $h_1(t)$ and $h_2(t)$ to generate two observation signals $x_1(t)$ and $x_2(t)$, respectively. The resulting $x_1(t)$ and $x_2(t)$ are then optionally corrupted with Gaussian white noise.

For the deconvolution algorithm, σ_1^2, σ_2^2, $\lambda^{(1)}$, $\lambda^{(2)}$ are initialized to be some small values, $\{\Delta t_i^{(1)}\}$ and $\{\Delta t_i^{(2)}\}$ to be $0, T_s, 2T_s, ..., +63T_s$ where T_s is the sample interval. The generalized cross-correlation is used to initially estimate the primary time delay between $x_1(t)$ and $x_2(t)$, and the traditional beamformed solution is used to initial the estimate of $s(t)$.

The mean squared error ($\|\hat{s}(t) - s(t)\|^2/\|s(t)\|^2$) of the estimated source $\hat{s}(t)$ at each iteration is shown in Fig 2. To illustrate the robustness of the algorithm, $x_1(t)$ and $x_2(t)$ were corrupted with various levels of Gaussian white noise. The deconvolution results indicates that the relevant deconvolution algorithm is able

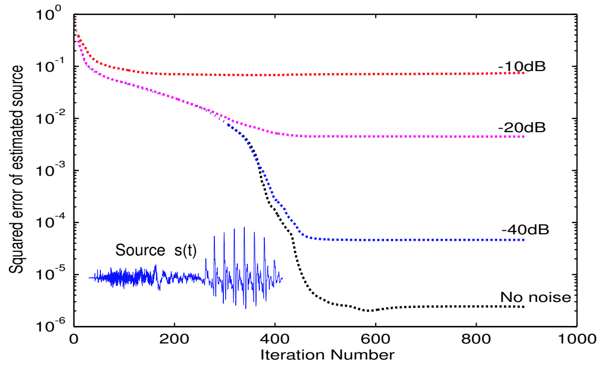

Fig. 2. Mean square error of the estimated source signal $\hat{s}(t)$: $\|\hat{s}(t) - s(t)\|^2/\|s(t)\|^2$. The measured observations $x_1(t)$ and $x_2(t)$ were contaminated with either zero, -40dB, -20dB, -10dB Gaussian white noise, respectively.

to precisely and robustly recover the source signal. The estimated source signal displays less errors than the added noise level, showing that deconvolution algorithm is not amplifying the input noise.

The estimated filters corresponding to no noise, -40dB, -20dB, and -10dB ambient noise are plotted in Fig. 3. For noise levels of -20dB or less, the estimated filter coeffients match the true filters. Even with -10dB noise, the general structure of the filter coefficients is still properly computed.

5. DISCUSSION

We have described the relevant deconvolution algorithm for simultaneously estimating a single acoustic source and the associated room impulse responses from two convoluted observations. In contrast to conventional blind deconvolution algorithms, our algorithm assumes no knowledge of the statistics of the source. This approach has several distinct advantages. Relatively few measurement samples are needed since the algorithm does not rely upon calculating source statistics. Also, the signals do not need to be prewhitened, and the algorithm can estimate the sources with a variety of bandwidths. The algorithm is also quite robust to the ambient noise, as observed in Fig. 2. This shows that nonnegativity constraints and Bayesian regularization are powerful methods to help solve the deconvolution problem. Furthermore, this general framework can easily be extended to incorporate signals from possibly more sensors, and to estimate perhaps more than one source.

Although we have emphasized the role of nonnegativity and sparsity of the filter coefficients in this work, the algorithm does not preclude incorporating prior knowledge about the source. Preliminary work indicates that the convergence and robustness of the algorithm is even further improved by incorporating a simple Laplacian prior on the source signal. Further experimentation will illustrate the utility of the relevant deconvolution algorithm for real-time deconvolution problems.

We acknowledge discussions with Jihun Ham, Fei Sha, and Lawrence Saul. We also acknowledge support from the U.S. National Science Foundation and the Army Research Office.

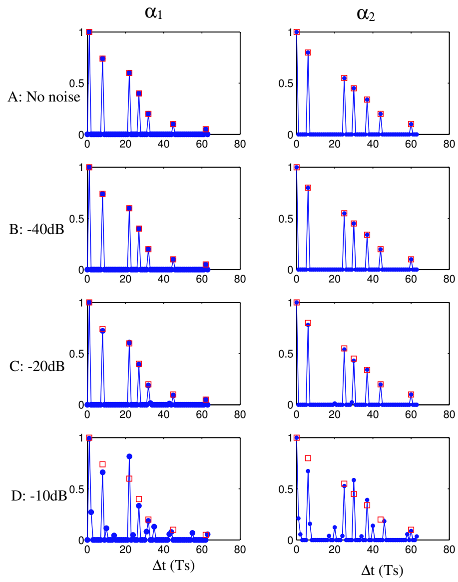

Fig. 3. Estimated filter parameters when the signals $x_1(t)$ and $x_2(t)$ were corrupted with various levels of Gaussian white noise: A) no noise, B) -40dB, C) -20dB, and D) -10dB noise. The left and right columns show the estimated filter parameters associated with $x_1(t)$ and $x_2(t)$, respectively. The hollow squares indicate the true filter coefficients.

6. REFERENCES

[1] Jont B. Allen and David A. Berkley, "Image method for efficient simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[2] Yuanqing Lin, Daniel D. Lee, and Lawrence K. Saul, "Nonnegative deconvolution for time of arrival estimation," in *ICASSP*, 2004.

[3] Yuanqing Lin and Daniel D. Lee, "Bayesian regularization and nonnegative deconvolution for time delay estimation," in *The Conference on Advances in Neural Information Processing Systems, 2004*.

[4] Sandro Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind deconvolution*, Simon Haykin, Ed. 1994, Prentice-Hall.

[5] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[6] H. Attias and C.E. Schreiner, "Blind source separation and deconvolution: the dynamic component analysis algorithm," *Neural Computation*, vol. 10, pp. 1373–1424, 1998.

[7] Hagai Attias, "New em algorithms for source separation and deconvolution," in *ICASSP, 2003*.

[8] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[9] D.J.C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, pp. 415–447, 1992.