MULTICHANNEL ADAPTIVE FILTERING IN COMPRESSIVE DOMAINS

Karim Helwani and Herbert Buchner◇*

* Huawei European Research Center,
◇ University of Cambridge, Information Engineering Division,
Email: karim.helwani@huawei.com, hb444@cam.ac.uk

ABSTRACT

In this paper, we give a study on reducing the coefficients to be estimated in an adaptive sparse multichannel system identification problem. We present an approach to perform the adaptation in a compressed representation of the sparse system without requiring prior knowledge about the dimensions in which the system has significant components. The presented technique exploits the ability of sparse systems to be compressed offering a reduction of the adaptive filter coefficients in addition to high convergence rates.

1. INTRODUCTION

Linear adaptive filters have found applications in diverse fields including communications, control, robotics, sonar, radar, seismics and biomedical engineering to name but a few [1]. The main classes of adaptive filtering problems (forward problems, inverse modelling, and prediction) share a structure in which a cost function is minimized iteratively. Hence, insights gained by studying one class of adaptive filtering problems can be used for other classes. In this study, we will concentrate on the system identification problem as a forward problem of adaptive filtering. System identification finds application in such diverse fields as layered earth modelling and acoustic echo cancellation (AEC) [1].
As can be expected intuitively , a systematic incorporation of prior knowledge about the system or the signals leads to a significant improvement of the adaptive filter performance. Typical examples of given prior knowledge about the systems could be a given distribution of its coefficients or a sparse structure of the system (i.e., only a small percentage of their components has significant magnitude while the rest is close to zero). In particular the decay of a typical single channel acoustic impulse responses or the structure of a system modelling the earth layers where only few peaks are different from zero (or being close to zero) motivates the assumption of sparseness of such impulse responses in the time domain [2, 3, 4]. For the multichannel case, an approach was presented in [5] to involve the prior knowledge about the spatio-temporal structure of a multichannel systems in the adaptive filter. It has been shown that by choosing suitable regularization strategies, high convergence rates can be achieved even for ill-conditioned problems, especially by exploring the sparsity of a system.
It has been shown in [6, 7, 8] that a spatially sparse representation can be enforced in multichannel acoustic systems by means of transformations into suitable subspaces where the multichannel system can be considered to be decoupled or diagonalized . A decoupled system acts on the independent modes at its input separately such that an input mode has influence on one single output mode and all other cross channels can be disregarded. Hence, these techniques to achieve a sparse structure of a multichannel systems give a hint on the position of the dimensions in which the transformed multichannel system has significant components. The identification reduces in theses techniques to estimate the system components in particular dimensions that are obtained from prior knowledge about either the system or the excitation signal.
In a more general case, the system could have some sparse representation however the knowledge about the relevant dimensions might be inaccessible a priori. Compressed sensing is a technique to compress and optimally reconstruct sparse signals regardless of the particular relevant dimensions of the signal. In the following, we will exploit insights from compressed sensing research on sparse signals to the multichannel system identification problem of sparse systems. For illustration we choose the setup of multichannel AEC as a prominent example of the system identification problem. Full-duplex communication in a hands-free communication scenario with multichannel setup (P loudspeakers) requires AEC. This aims at canceling the acoustic echoes from the microphone signals. Fig. 1 shows a block diagram of system identification scenario as it is e.g., in multichannel AEC with P reproduction channels and single microphone channel in the receiving room ('near-end', denoted more generally as 'plant' in Fig. 1). The signals of the P reproduction channels originate from speech- or audio sources in a transmission room ('far-end'). To cancel the echoes arising due to the reflections in the near

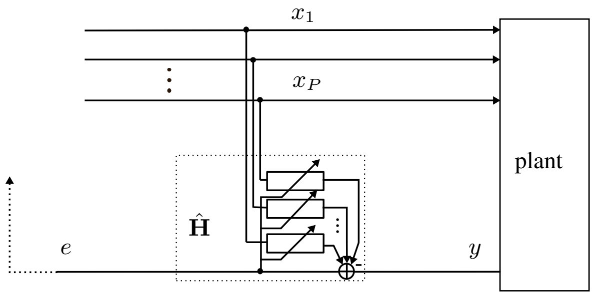

Fig. 1. Block diagram of multichannel adaptive system identification.

end the reproduction signals x_p are filtered with the adaptively estimated coefficients $\hat{\mathbf{h}}$, i.e., a replica of the actual acoustic multiple-input single-output (MISO) system, and the resulting signals are subtracted from the near-end microphone signals $y = \mathbf{h}^{\mathrm{T}}\mathbf{x}$, with $\mathbf{x}(n) = [\mathbf{x}_1^{\mathrm{T}}(n), \mathbf{x}_2^{\mathrm{T}}(n), \cdots, \mathbf{x}_P^{\mathrm{T}}(n)]^{\mathrm{T}}$, $\mathbf{x}_p(n) = [x_p(n), x_p(n-1), \cdots, x_p(n-L+1)]^{\mathrm{T}}$, $\widehat{\mathbf{h}}(n)$ denotes the PL MISO coefficient vector composed by P subfilters, $\hat{\mathbf{h}}_p = [\hat{h}_{p,0}, \hat{h}_{p,1}, \cdots, \hat{h}_{p,L-1}]^{\mathrm{T}}$ and n the time instant. If the estimated filter coefficients $\hat{\mathbf{h}}$ are equal to the true transfer paths $\mathbf{h}$, all disturbing echoes will be removed from the microphone signals. Most of the popular adaptive filtering algorithms are based on least-squares error minimization and aim at the so-called Wiener solution of the filter coefficients [1]. Typically, the cost function reads

$$J\left(\widehat{\mathbf{h}}(n)\right) := \hat{\mathcal{E}}\{e^2(n)\} = \hat{\mathcal{E}}\left\{\left(y(n) - \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{x}(n)\right)^2\right\}, \quad (1)$$

where $\hat{\mathcal{E}}\{\cdot\}$ denotes an estimate of the expectation.

2. SUBSPACE-BASED APPROCHES AS SPECIAL CLASS OF COMPRESSIVE DOMAINS

2.1. General considerations

We call an M-dimensional subspace of an L-dimensional space ($M \leq L$) a compressed domain, if the subspace is dense, i.e., any vector in the original space is concentrated in a lower dimensional subspace or if it can be fairly approximated in a lower dimensional subspace basis. From information theoretical point of view the condition for an ideal compression domain is to have maximal mutual information between a vector $\widehat{\mathbf{h}}$ and its compressed representation $\widehat{\underline{\mathbf{h}}}$ in the considered subspace [9]. The mutual information between the signal and its compressed version is given by

$$I(\underline{\mathbf{h}}, \mathbf{h}) = H(\underline{\mathbf{h}}) - H(\underline{\mathbf{h}}|\mathbf{h}). \quad (2)$$

The conditional entropy, $H(\underline{\mathbf{h}}|\mathbf{h})$, can be assumed to be zero when $\underline{\mathbf{h}}$ is completely determined by the value of $\mathbf{h}$. Hence, the mutual information is maximal if and only if the entropy of $\underline{\mathbf{h}}$ is maximal [10].

2.2. Illustration using second order statistics

To illustrate how a concrete compressive domain could look like, we give an example based on second-order statistics (SOS) with given correlation matrices $\mathbf{R}_{\mathbf{hh}} := \mathcal{E}\{\mathbf{h}\mathbf{h}^{\mathrm{T}}\}$. It is well known that the principle component analysis maximizes the mutual information between a high dimensional vector and a vector with lower dimensionality under a gaussian signal model [10]. Hence, the entropy of the compressed signal $\widehat{\underline{\mathbf{h}}} := \mathbf{\Phi}\widehat{\mathbf{h}}$ given by

$$H(\mathbf{h}) = -\,\mathcal{E}\{\log p_{\underline{\mathbf{h}}}(\underline{\mathbf{h}})\}, \quad (3)$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation. For ease of representation, we assume here the systems to be zero-mean. Hence, we have

$$p_{\underline{h}}(\underline{\mathbf{h}}) = \frac{1}{\sqrt{2\pi \det(\mathbf{R}_{\underline{\mathbf{hh}}})}} e^{-\frac{1}{2}\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{R}_{\underline{\mathbf{hh}}}^{-1}\underline{\mathbf{h}}}, \quad (4)$$

and the entropy (3) is maximal if the rectangular compression matrix $\mathbf{\Phi}$ of the size $M \times L$, $M \leq L$, contains the eigenvectors corresponding to the M largest eigenvalues of $\mathbf{R}_{\mathbf{hh}}$.
In the following we show that the least-squares minimization of the AEC problem in such compressive domains leads to an approximation of the solution of the normal equation in the uncompressed domain. Therefore, we define

$$\underline{\hat{y}}(n) := \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{x}(n), \quad (5)$$

$$\underline{\mathbf{x}}(n) := \mathbf{\Phi}\mathbf{x}(n). \quad (6)$$

Minimization of the cost function in the compressive domain

$$\underline{J}\left(\widehat{\mathbf{h}}(n)\right) := \hat{\mathcal{E}}\left\{\left(y(n) - \underline{\hat{y}}(n)\right)^2\right\}, \quad (7)$$

leads to the least-squares solution, known as Wiener-Hopf equation [1]

$$\mathbf{R}_{\underline{\mathbf{xx}}}\widehat{\underline{\mathbf{h}}}_{\text{opt}} - \mathbf{r}_{\underline{\mathbf{x}}y} = 0. \quad (8)$$

with $\mathbf{r}_{\underline{\mathbf{x}}y}(n) := \hat{\mathcal{E}}\{\underline{\mathbf{x}}(n)y(n)\}$, $\mathbf{R}_{\underline{\mathbf{xx}}}(n) := \hat{\mathcal{E}}\{\underline{\mathbf{x}}(n)\underline{\mathbf{x}}^{\mathrm{T}}(n)\}$. Eq. (8) can be reformulated as

$$\mathbf{\Phi}\mathbf{R}_{\mathbf{xx}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\widehat{\mathbf{h}} - \mathbf{\Phi}\mathbf{r}_{\mathbf{x}y} = 0, \quad (9)$$

with $\mathbf{r}_{\mathbf{x}y}(n) := \hat{\mathcal{E}}\{\mathbf{x}(n)y(n)\}$ and $\mathbf{R}_{\mathbf{xx}}(n) := \hat{\mathcal{E}}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\}$. As mentioned, $\mathbf{\Phi}$ is computed by PCA. This technique aims at minimizing the least-squares error $e_{\text{PCA}} = \left\|(\widehat{\mathbf{h}} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\widehat{\mathbf{h}})\right\|_2^2$, hence we can write for $e_{\text{PCA}} \to 0$

$$\mathbf{\Phi}\mathbf{R}_{\mathbf{xx}}\widehat{\mathbf{h}} - \mathbf{\Phi}\mathbf{r}_{\mathbf{x}y} = 0, \quad (10)$$

$$\mathbf{\Phi}\left(\mathbf{R}_{\mathbf{xx}}\widehat{\mathbf{h}} - \mathbf{r}_{\mathbf{x}y}\right) = 0. \quad (11)$$

if $\left(\mathbf{R}_{\mathbf{xx}}\widehat{\mathbf{h}} - \mathbf{r}_{\mathbf{x}y}\right)$ is not in the null space of $\mathbf{\Phi}$, which is in general fulfilled by satisfied excitation, then we obtain again the Wiener-Hopf equation

$$\mathbf{R}_{\mathbf{xx}}\widehat{\mathbf{h}} = \mathbf{r}_{\mathbf{x}y}. \quad (12)$$

Therefore, we can conclude $\widehat{\mathbf{h}} = \widehat{\mathbf{h}}_{\text{opt}}$.

3. COMPRESSIVE DOMAINS FOR SPARSE SYSTEMS

Recent studies on compressive sensing state, that a sparse signal, e.g., $\widehat{\mathbf{h}}$, can be perfectly reconstructed from its undersampled version

$$\widehat{\underline{\mathbf{h}}} := \mathbf{\Phi}\widehat{\mathbf{h}} \quad (13)$$

with $\mathbf{\Phi}$ a random observation matrix [11]. The originality of the theory of compressed sensing bases on its implicit statement that a subspace spanned by $M = O(K\log(L/K))$ uncorrelated white vectors is dense in the space of K-sparse signals of length L [12]. The mentioned statement of compressive sensing motivates us to ask for the possibility of adaptive filtering for sparse systems in compressive domain analogously to the discussion in Sect. 2 but with random compression matrices and without an explicit knowledge about the relevant dimensions of the sparse systems.

3.1. Adaptation in compressive domains for sparse systems

To answer the mentioned question about the adaptation in random compressive domains for sparse systems, we aim at formulating the cost function for the MISO system identification similarly to the formulation in (7) optimization problem in a compressive domain. Since the compression matrix is not given by an explicit eigenspace of the sparse system, we cannot assume a minimum reconstruction error by using the hermitian transpose of the compression matrix, as we assumed to conclude (11) and subsequently an optimal filtering in the compressive domain given by the system subspace. An optimal reconstruction by a transformation matrix can be obtained from the typical compressed sensing cost function which is based on exploiting the sparsity of the system

$$J'\left(\widehat{\mathbf{h}}(n)\right) = \lambda\left\|\widehat{\mathbf{h}}(n)\right\|_1 + \left\|\widehat{\underline{\mathbf{h}}}(n) - \mathbf{\Phi}\widehat{\mathbf{h}}(n)\right\|_2^2, \quad (14)$$

where λ denotes the Lagrange-multiplier. A minimum of the cost function can be found by setting its gradient w.r.t $\widehat{\mathbf{h}}$ to zero.
With

$$\left\|\widehat{\mathbf{h}}\right\|_1 = \operatorname{sgn}\left\{\widehat{\mathbf{h}}\right\}^{\mathrm{T}} \widehat{\mathbf{h}}, \tag{15}$$

hereby, sgn$\{\cdot\}$ stands for the sign function. Hence, the gradient reads

$$\nabla_{\widehat{\mathbf{h}}} J' = 2\lambda \operatorname{sgn}\left\{\widehat{\mathbf{h}}\right\} - 2\mathbf{\Phi}^{\mathrm{T}}(n)\left[\underline{\widehat{\mathbf{h}}}(n) - \mathbf{\Phi}(n)\widehat{\mathbf{h}}(n)\right], \tag{16}$$

the sign function can be approximated by

$$\operatorname{sgn}\left\{\widehat{\mathbf{h}}\right\} = \mathbf{E}^{-1}\widehat{\mathbf{h}}, \quad \text{with} \quad \mathbf{E} := \operatorname{diag}\left\{\left|\widehat{\mathbf{h}} + \epsilon\right|\right\}, \tag{17}$$

where ϵ is a parameter that prevents a division by zero. For an optimum point we have the condition

$$\begin{aligned} &\nabla_{\widehat{\mathbf{h}}_{\text{opt}}} J' = 0, \\ &\widehat{\mathbf{h}}_{\text{opt}}(n) = \left(\lambda \mathbf{E}^{-1}(n) + \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^{\mathrm{T}}\underline{\widehat{\mathbf{h}}}. \end{aligned} \tag{18}$$

Hence, we obtain a reconstruction matrix that we define as

$$\mathbf{\Phi}^{+}(n) := \left(\lambda \mathbf{E}^{-1}(n) + \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^{\mathrm{T}}. \tag{19}$$

Since $\widehat{\mathbf{h}}$ is a priori unknown, an iterative computation for the regularization matrix $\mathbf{E}^{-1}$ has to be performed

$$\mathbf{E}(n) = \operatorname{diag}\left\{\left|\widehat{\mathbf{h}}(n-1) + \epsilon\right|\right\}, \tag{20}$$

where $\mathbf{E}(0)$ is set to the unity matrix. Hence, the reconstruction matrix can be understood as an adaptive back transformation matrix from the the domain where the system has a compressed dense representation to the domain where it is sparsely represented. This back transformation matrix is adjusted to the sparse structure of the system. Now, we can rewrite the cost function

$$\underline{J}\left(\widehat{\mathbf{h}}(n)\right) := \hat{\mathcal{E}}\left\{\left(y(n) - \underline{\hat{y}}(n)\right)^2\right\}, \tag{21}$$

with the new definition

$$\underline{\hat{y}}(n) := \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}^{+\mathrm{T}}(n)\mathbf{x}(n), \tag{22}$$

here, $\mathbf{\Phi}^{+}(n)$ denotes the regularized inverse of $\mathbf{\Phi}$ which depends on $\widehat{\mathbf{h}}(n-1)$ as shown above, and

$$\underline{\mathbf{x}}(n) := \mathbf{\Phi}^{+\mathrm{T}}(n)\mathbf{x}(n). \tag{23}$$

We obtain analogously to the known least squares solution but in the compressed domain

$$\underline{\widehat{\mathbf{h}}}_{\text{opt}}(n) = \mathbf{R}_{\underline{\mathbf{x}}\underline{\mathbf{x}}}^{-1}(n)\mathbf{r}_{\underline{\mathbf{x}}y}(n). \tag{24}$$

Typically, the correlation matrix is estimated iteratively using the formula

$$\mathbf{R}_{\underline{\mathbf{x}}\underline{\mathbf{x}}}(n) = \alpha\, \mathbf{R}_{\underline{\mathbf{x}}\underline{\mathbf{x}}}(n-1) + \underline{\mathbf{x}}(n)\underline{\mathbf{x}}^{\mathrm{T}}(n), \tag{25}$$

α denotes a forgetting factor.
The uncompressed estimated filter is then given by

$$\widehat{\mathbf{h}}(n) = \mathbf{\Phi}^{+}(n)\underline{\widehat{\mathbf{h}}}(n). \tag{26}$$

The zeros of $\nabla_{\widehat{\mathbf{h}}}\underline{J}$ can be determined iteratively with the Newton algorithm. The main advantage of Newton-type adaptation algorithms is its quadratic convergence rate compared to the linear convergence rate of the gradient-based algorithms [1].
Compared with algorithms that incorporate system prior knowledge in the uncompressed domain such as the approach presented in [5], the shown algorithm introduced here deals with reduced dimensions of the signal autocorrelation matrix $\mathbf{R}_{\mathbf{xx}}$. A block diagram to illustrate the structure of the presented algorithm is presented in Fig. 2.

3.2. Approximations to avoid explicit estimation of the reconstruction matrix for Newton based algorithms

As we have shown above, an adaptive computation of the reconstruction matrix is required for an optimal estimation of the sparse system in the compressive domain. The computational complexity of the back transformation matrix is relatively high compared with algorithms aiming at estimating only a sparse signal out of incomplete measurements, such as the message passing algorithms [13]. Hence, it is desirable to combine the high convergence rate of the Newton-type algorithms with the idea of a fast estimation of a sparse system from incomplete measurements. Let us consider the update rule of a Newton-type algorithm. This reads in the multichannel case [14]

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \mathbf{R}_{\mathbf{xx}}^{-1}(n) \cdot \mathbf{x}(n)e(n), \tag{27}$$

where the Kalman gain, $\mathbf{k}(n)$ is given as

$$\mathbf{k}(n) := \mathbf{R}_{\mathbf{xx}}^{-1}(n) \cdot \mathbf{x}(n). \tag{28}$$

By multiplying both equation sides of this equation by a measurement matrix $\mathbf{\Phi}$ we obtain with (13)

$$\underline{\widehat{\mathbf{h}}}(n) = \underline{\widehat{\mathbf{h}}}(n-1) + \underline{\mathbf{k}}(n)e(n), \tag{29}$$

where

$$\underline{\mathbf{k}}(n) := \mathbf{\Phi}\mathbf{k}(n). \tag{30}$$

Assuming the input signal is white, we make the approximation of the Kalman gain

$$\underline{\mathbf{k}}(n) \approx \left(\mathbf{\Phi}^{\dagger}\mathbf{R}_{\mathbf{xx}}(n)\mathbf{\Phi}^{\dagger^{\mathrm{T}}}\right)^{-1} \cdot \mathbf{\Phi}^{\dagger^{\mathrm{T}}}\mathbf{x}(n), \tag{31}$$

where $\mathbf{\Phi}^{\dagger}$ denotes an ℓ_2 regularized inverse of $\mathbf{\Phi}$. This is motivated by the assumption that the input is normally distributed.
Hence, we obtain an approximate Newton type algorithm as

$$\underline{\widehat{\mathbf{h}}}(n) = \underline{\widehat{\mathbf{h}}}(n-1) + \underline{\mathbf{k}}(n)e(n). \tag{32}$$

Note that the error $e(n)$ is given by the difference between $y(n)$ and the estimated output of the uncompressed system which can be efficiently estimated e.g., by the message passing algorithms.

4. EXPERIMENTS

As a proof of concept for the compressive domain adaptive filtering we have simulated in Fig. 3, a 2-channel MISO system identification scenario. Here the system environment was simulated as 10 randomly weighted and separated pulses, the filter length $L = 64$. Each channel was excited by uncorrelated white noise. To the captured signal, additive white noise was added to obtain an SNR of 60 dB. Five adaptation runs were simulated. As reference, the red curve shows the system distance by non compressive adaptation by

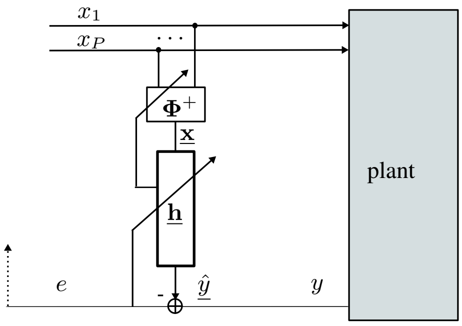

Fig. 2. Block diagram of multichannel adaptive system identification in compressive domain.

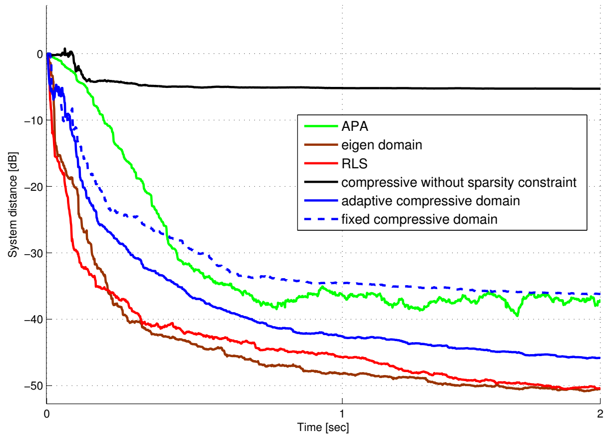

Fig. 3. Identification performance in compressive domains

simulating the recursive least squares algorithm (RLS). The brown curve shows the reached system distance by adaptation in a compressive domain obtained by the first 40 principal components of a training dataset. The training dataset consisted of 15 randomly generated MISO systems, each channel has 10 randomly weighted and separated pulses in the first 16 taps. Finally, the impulse response to be identified was not part of the training dataset.
The blue curve was produced according to the results in Sect. 3.1, it corresponds to a compression length of 40. The approximation given in in Sect. 3.2 is reflected in dotted blue curve. The green curve is the system distance can be reached by affine projection algorithm (APA) with a projection order of 40. Note that the performance of APA and the approximated compressive domain algorithm with a fixed back transformation are similar. However the compressive domain algorithm converges faster. The black curve shows the system distance when the estimation of the sparse system is done in a compressed domain which is not necessarily the eigenspace of the system and the back transformation is done by an unconstrained least squares. The simulations have proven that the concept of compressive domain adaptive filtering is technically realizable and it offers acceptable convergence rates.

5. CONCLUSION

In this paper we gave a study on the multichannel adaptive system identification in compressive domains. On the one hand, estimating a sparse system in a compressed domain can be understood as a special subspace adaptive filtering algorithm where the back transformation matrix is given by means of a constrained cost function. On the other hand, subspace adaptive filtering itself can be understood as a special case of sparse adaptive filters where the relevant dimensions in which the compressed system has significant components are known.

6. REFERENCES

[1] S. Haykin, *Adaptive filter theory*. Prentice Hall, Inc., 1991.

[2] J. Benesty, C. Paleologu, and S. Ciochină, “Proportionate adaptive filters from a basis pursuit perspective,” *IEEE Signal Processing Letters*, vol. 17, pp. 985–988, Dec. 2010.

[3] P. Naylor, J. Cui, and M. Brookes, “Adaptive algorithms for sparse echo cancellation,” *Applied Speech and Audio Processing*, vol. 86, pp. 1182–1192, June 2006.

[4] P. Loganathan, X. Lin, A. Khong, and P. Naylor, “Frequency-domain adaptive multidelay algorithm with sparseness control for acoustic echo cancellation,” in *in Proc. EUSIPCO*, 2009.

[5] K. Helwani, H. Buchner, and S. Spors, “Multichannel adaptive filtering with sparseness constraints,” *in Proc. IWAENC 2012*, pp. 1–4, Sept. 2012.

[6] H. Buchner, S. Spors, and W. Kellermann, “Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field synthesis,” in *in Proc. IEEE ICASSP*, vol. 4, 2004.

[7] S. Spors, H. Buchner, and R. Rabenstein, “Eigenspace adaptive filtering for efficient pre-equalization of acoustic MIMO systems,” in *in Proc. EUSIPCO, 2006*, pp. 1–5.

[8] K. Helwani, H. Buchner, and S. Spors, “Source-domain adaptive filtering for MIMO systems with application to acoustic echo cancellation,” in *in Proc. IEEE ICASSP*, pp. 321–324, Mar. 2010.

[9] D. Barber and F. Agakov, “The IM algorithm: a variational approach to information maximization,” *NIPS MIT Press*, 2003.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

[11] E. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *IEEE Trans. on nformation Theory,*, vol. 52, no. 12, pp. 5406–5425, 2006.

[12] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[13] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18914–18919, 2009.

[14] H. Buchner, J. Benesty, and W. Kellermann, “Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication,” *Signal Processing*, vol. 85, no. 3, pp. 549–570, 2005.