SPARSE TARGET CANCELLATION FILTERS WITH APPLICATION TO SEMI-BLIND NOISE EXTRACTION

Jiří Málek and Zbyněk Koldovský

Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. E-mail: jiri.malek@tul.cz

ABSTRACT

Impulse responses of filters that perform spatial null in a target direction, so-called target-cancellation filters (CFs), are usually long and dense due to the reverberant acoustic environment. It is therefore hard to blindly estimate them from noisy recordings of the target. In this paper, we show that efficient sparse CFs having many coefficients equal to zero can be designed such that their cancellation performance is tolerably lower than the performance of dense CFs. We show that an efficient sparse CF can be blindly estimated from noisy data, provided that its support is known. The resulting filter is better than a dense CF which has been blindly estimated without any prior knowledge.

Index Terms— Target Cancellation Filters, Sparse Filters, Noise Extraction, Semi-Blind Audio Source Separation

1. INTRODUCTION

Target-cancellation filters (CFs) are important tools in audio signal processing tasks such as signal separation, noise suppression and speech enhancement. A CF is a multichannel filter that cancels the target signal but lets other signals (interferers and ambient noise) pass through. Its output thus provides a noise-reference signal that is useful for parallel processing of the original noisy recording of the target. For example, CFs are used within the blocking matrix part of generalized sidelobe cancellers [1], where the outputs are used for adaptive interference cancellation and post-filtering; see, e.g., [1, 2, 3]. This paper focuses on the problem of finding efficient CFs.

Consider a two-channel[1] noisy recording of a target whose position is fixed; this recording is described through

$$x_{\mathrm{L}}(n) = \{h_{\mathrm{L}} * s\}(n) + y_{\mathrm{L}}(n), \quad (1)$$

$$x_{\mathrm{R}}(n) = \{h_{\mathrm{R}} * s\}(n) + y_{\mathrm{R}}(n) \quad (2)$$

where n is the time index; $*$ denotes the convolution; x_{L} and x_{R} are, respectively, the signals from the left and right microphones; s is the target signal; and y_{L} and y_{R} are the remaining signals commonly referred to as noise. h_{L} and h_{R} denote the microphone-target impulse responses that depend on the target's position and on the acoustical environment.

A CF consists of two filters, g_{L} and g_{R}. Its output is $g_{\mathrm{L}} * x_{\mathrm{L}} - g_{\mathrm{R}} * x_{\mathrm{R}}$ and should not contain any contribution of s. Typically, the selection of g_{L} and g_{R} is such that $g_{\mathrm{L}} \approx h_{\mathrm{R}}$ and $g_{\mathrm{R}} \approx h_{\mathrm{L}}$. A more popular option is $g_{\mathrm{L}} \approx h_{\mathrm{R}} * h_{\mathrm{L}}^{-1} * \delta(n-d)$ and $g_{\mathrm{R}}(n) = \delta(n-d)$ (the delayed unit impulse); see, e.g., [1, 4, 5] or other alternatives in [6, 7]. Here, h_{L}^{-1} denotes the inverse filter of h_{L}, and $g_{\mathrm{rel}} = h_{\mathrm{R}} * h_{\mathrm{L}}^{-1}$ corresponds to the relative impulse response between the microphones. The latter need not be causal, so it is practical to select $d > 0$; typically, $d = 20$.

Since g_{L} and g_{R} depend on h_{L} and h_{R}, they must be estimated from data recorded on-site. Using a noise-free recording of the target ($y_{\mathrm{L}} = y_{\mathrm{R}} = 0$), the filters can be estimated using least squares [8]; frequency-domain estimates from [1, 4] allow for the presence of stationary noise. The challenge is to compute or update the filters when directional and non-stationary noise is present. Blind methods can be used [9, 10], however, since h_{L} and h_{R} are typically long and dense, the statistical error due to blind estimation may radically deteriorate the efficiency of the estimated CF. Recent efforts have therefore been made to lower the dimensionality of that problem using prior knowledge [11, 20, 25].

In this paper, we show that it is possible to derive CFs that have many coefficients equal to zero (sparse CFs) while their cancellation performance is only slightly worse compared to dense CFs of the same length. The sparse CF could be seen as a vector of a low-dimensional subspace, where the dimension is equal to the number of nonzero coefficients of the CF. We assume that the support (indices of nonzero coefficients) of the sparse CF is given as prior knowledge, and, based on this, we propose a semi-blind approach to estimate the CF from noisy data. We show by experiments that the filter is significantly better compared to a dense CF when both are estimated blindly. Simultaneously, the semi-blind estimation is computationally simpler than the blind estimation.

In the following section, we discuss several approaches to compute sparse CFs from a noise-free recording using least squares penalized or constrained by sparsity-inducing norms. Two greedy approaches are proposed as well. The filters are

This work was supported by the Czech Science Foundation through Project No. 14-11898S.

[1]For simplicity, we consider only two channels, but the ideas of this paper can be generalized to a higher number of channels.

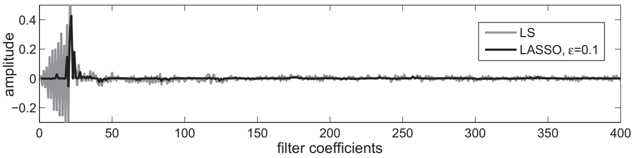

Fig. 1. An example of $g_{\rm rel}$ estimated using LS and LASSO (see Sections 2.1 and 2.2). The data were recorded in a room with $T_{60} =$ 800 ms, distance between microphones was 16 cm, and distance of target from microphones was 2 m.

assessed on real-world recordings in Section 3. In Section 4, the semi-blind approach is described and experimentally verified.

2. SPARSE TARGET-CANCELLATION FILTERS

Assume now that $x_{\rm L}(n)$ and $x_{\rm R}(n)$, $n = 1, \ldots, N$, are noise-free recordings of the target signal. We focus on the design of CFs through an approximation (estimation) of $g_{\rm rel}$. We assume that $g_{\rm rel}$ is nearly sparse. Although this is not guaranteed in general, such an assumption is often met even in highly reverberated rooms; see Fig. 1. When the microphones are close to each other, $g_{\rm rel}$ has a shape that is similar to that of a delayed unit impulse. The most significant coefficients of $g_{\rm rel}$ tend to be concentrated around the beginning of the filter due to direct path and early reflections of the target signal.

2.1. Least Squares Estimator

The least squares estimate of $g_{\rm rel}$ is given by elements of the vector

$$\hat{\mathbf{g}}_{\rm LS} = \arg\min_{\mathbf{g}} \|\mathbf{X}_{\rm L}\mathbf{g} - \mathbf{x}_{\rm R}\|_2^2 \tag{3}$$

where $\mathbf{X}_{\rm L}$ is the $N \times L$ Toeplitz matrix whose first column and first row are, respectively, $[x_{\rm L}(1), \ldots, x_{\rm L}(N)]^T$ and $[x_{\rm L}(1), 0, \ldots, 0]$, $\mathbf{x}_{\rm R} = [x_{\rm R}(1), \ldots, x_{\rm R}(N)]^T$, and L denotes the length of $\mathbf{g}$. The solution is $\hat{\mathbf{g}}_{\rm LS} = \mathbf{R}^{-1}\mathbf{p}$ where $\mathbf{R} = \mathbf{X}_{\rm L}^T\mathbf{X}_{\rm L}/N$ is a square $L \times L$ symmetric Toeplitz matrix and $\mathbf{p} = \mathbf{X}_{\rm L}^T\mathbf{x}_{\rm R}/N$. Hence, $\hat{\mathbf{g}}_{\rm LS}$ also satisfies $\|\mathbf{R}\hat{\mathbf{g}}_{\rm LS} - \mathbf{p}\|_2^2 = 0$, so we can recast (3) as

$$\hat{\mathbf{g}}_{\rm LS} = \arg\min_{\mathbf{g}} \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2. \tag{4}$$

The latter formulation is useful for defining the sparse estimates of $g_{\rm rel}$. These will be obtained through altering (4).

2.2. Sparse Approximations

LASSO (Least Absolute Shrinkage and Selector Operator, [13]) is the optimization program given by

$$\hat{\mathbf{g}}_{\rm LASSO} = \arg\min_{\mathbf{g}} \|\mathbf{g}\|_1 \quad \text{w.r.t.} \quad \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 \leq \epsilon, \tag{5}$$

where $\epsilon \geq 0$. This procedure is easy to interpret: The constraint $\|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 \leq \epsilon$ relaxes the property of $\mathbf{g}_{\rm LS}$, that is $\|\mathbf{R}\hat{\mathbf{g}}_{\rm LS} - \mathbf{p}\|_2^2 = 0$, while the sparsity-inducing ℓ_1-norm is minimized. For $\epsilon = 0$, $\hat{\mathbf{g}}_{\rm LASSO}$ coincides with $\hat{\mathbf{g}}_{\rm LS}$. On the other hand, there exists a sufficiently large ϵ such that $\hat{\mathbf{g}}_{\rm LASSO} = \mathbf{0}$; see the example in Fig. 1.

A program equivalent to (5) in the sense that the sets of solutions are the same for all possible choices of parameters is called Basis Pursuit Denoising (BPDN) and is given by [14]

$$\hat{\mathbf{g}}_{\rm BPDN} = \arg\min_{\mathbf{g}} \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 + \tau\|\mathbf{g}\|_1, \tag{6}$$

where $\tau \geq 0$. However, the correspondence between the parameters τ and ϵ is not trivial and is possibly discontinuous [15].

Weighted LASSO (WLASSO) [16] is a modified variant of (5) given by

$$\hat{\mathbf{g}}_{\rm WLASSO} = \arg\min_{\mathbf{g}} \|\mathbf{W}\mathbf{g}\|_1 \quad \text{w.r.t.} \quad \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 \leq \epsilon, \tag{7}$$

where $\mathbf{W}$ is a diagonal matrix with positive weights $w_1, \ldots, w_L$ on its diagonal. The weights allow to the placement of different emphasis on each filter coefficient. In our case, we select

$$w_i = |(\hat{\mathbf{g}}_{\rm LS})_i + \varsigma|^{-1}, \qquad i = 1, \ldots, L, \tag{8}$$

where $(\cdot)_i$ denotes the ith element of the argument, and $\varsigma > 0$ is a small positive constant to avoid division by zero. This choice means that "small" coefficients of $\hat{\mathbf{g}}_{\rm LS}$ are strongly forced to be zero, and vice versa. We will show later by experiments that this choice helps to increase the number of zero coefficients in $\hat{\mathbf{g}}_{\rm WLASSO}$ while the cancellation performance of the filter is preserved.

It is also possible to distribute the filter coefficients into groups and replace the ℓ_1-norm by the group $\ell_{1,2}$-norm [17] defined as $\|\mathbf{g}\|_{1,2} = \sum_{k=1}^{K} \|\mathbf{f}_k\|_2$, where the elements of $\mathbf{f}_k$ belong to the kth group and $\mathbf{g} = [\mathbf{f}_1^T, \ldots, \mathbf{f}_K^T]^T$. The penalized solution is then given by

$$\hat{\mathbf{g}}_{\rm GROUP} = \arg\min_{\mathbf{g}} \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 + \tau\|\mathbf{g}\|_{1,2}, \tag{9}$$

which is equivalent to (6) iff $K = L$, that is, when each filter coefficient is assigned to its individual group (singleton). With regard to the expected shape of $g_{\rm rel}$ (see Fig. 1), our selection of groups is such that first D elements of $\mathbf{g}$ are assigned to the first group $\mathbf{f}_1$ while the other elements are singletons.

2.3. Greedy Methods

We propose two greedy approaches to find sparse approximations of $g_{\rm rel}$ satisfying the constraint $\|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 \leq \epsilon$. The problem to solve is

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} \|\mathbf{g}\|_0 \quad \text{w.r.t.} \quad \|\mathbf{R}\mathbf{g} - \mathbf{p}\|_2^2 \leq \epsilon \tag{10}$$

Algorithm 1: The NAIVE greedy algorithm.

Input: $\mathbf{R}, \mathbf{p}, \mathbf{g}_{\text{LS}}, \epsilon$
Output: $\mathbf{g}^{I-1}$
$\alpha = 0, \mathbf{g}^0 = \mathbf{g}_{\text{LS}}, I = 0, \Gamma = \emptyset$;
while $\alpha < \epsilon$ **do**
  $\Gamma = \Gamma \cup \arg\min_i |(\mathbf{g}^I)_i|$;
  $\mathbf{g}^{(I+1)} = \arg\min_{\mathbf{g}} \left\{ \|\mathbf{Rg} - \mathbf{p}\|_2^2 \middle| (\mathbf{g})_i = 0 \text{ for } i \in \Gamma \right\}$;
  $\alpha = \|\mathbf{Rg}^{(I+1)} - \mathbf{p}\|_2^2$;
  $I = I + 1$;
end

Algorithm 2: The MINERR greedy algorithm.

Input: $\mathbf{R}, \mathbf{p}, \mathbf{g}_{\text{LS}}, \epsilon$
Output: $\mathbf{g}^{I-1}$
$\alpha = 0, \mathbf{g}^0 = \mathbf{g}_{\text{LS}}, I = 0, \Gamma = \emptyset$;
while $\alpha < \epsilon$ **do**
  $k = \arg\min_i \left\{ \min_{\mathbf{g}} \|\mathbf{Rg} - \mathbf{p}\|_2^2 \middle| (\mathbf{g})_j = 0, j \in \Gamma \cup \{i\} \right\}$;
  $\Gamma = \Gamma \cup \{k\}$;
  $\mathbf{g}^{(I+1)} = \arg\min_{\mathbf{g}} \left\{ \|\mathbf{Rg} - \mathbf{p}\|_2^2 \middle| (\mathbf{g})_i = 0 \text{ for } i \in \Gamma \right\}$;
  $\alpha = \|\mathbf{Rg}^{(I+1)} - \mathbf{p}\|_2^2$;
  $I = I + 1$;
end

where $\|\mathbf{g}\|_0$ is equal to the number of nonzero elements in $\mathbf{g}$ (the ℓ_0 pseudonorm). Although greedy algorithms are not guaranteed to solve (10) in general, they may successfully find suitable suboptimal results [18].

The first approach, denoted as NAIVE, starts from the least squares solution $\hat{\mathbf{g}}_{\text{LS}}$. It selects the smallest (in absolute value) nonzero element of $\hat{\mathbf{g}}_{\text{LS}}$ to constrain it to be zero and updates $\hat{\mathbf{g}}$ as the least squares solution under the constraint. This step is repeated until $\|\mathbf{R}\hat{\mathbf{g}} - \mathbf{p}\|_2^2 \leq \epsilon$. The algorithm is summarized in Algorithm 1.

A more sophisticated greedy algorithm, denoted as MINERR, also starts from $\hat{\mathbf{g}}_{\text{LS}}$. Compared to NAIVE, in each step, the algorithm selects the coefficient whose zeroing causes the smallest increase of $\|\mathbf{R}\hat{\mathbf{g}} - \mathbf{p}\|_2^2$. It is summarized in Algorithm 2.

In experiments, we use freely available online implementations for some of the considered minimization problems. We optimize the LASSO criterion via the ℓ_1 magic toolbox [21]. The BPDN criterion is minimized using the SLEP package [22]. The criteria WLASSO and GROUP are optimized using the SPArse Modeling Software (SPAMS) [23].

3. EXPERIMENTAL COMPARISON

The goal of this section is to find CFs with a minimum number of nonzero coefficients whose cancellation performance is close to that of $\hat{\mathbf{g}}_{\text{LS}}$. We compare the methods for computing sparse CFs from noise-free recordings introduced in the previous section.

The cancellation performance of each CF is evaluated in terms of Noise-to-Signal Ratio (NSR) measured in the output of the CF when the original noise-free recording of a target speaker (used to compute the filter) is mixed with a speech signal of another speaker. In this experiment, the signals are mixed so that the input NSR is 0 dB.

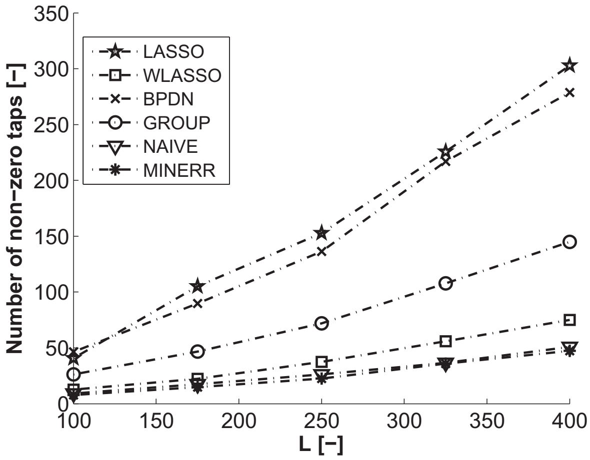

Fig. 2. Number of nonzero coefficients with respect to L averaged over all recordings.

For each L, the sparse CFs are unified through the parameter τ or ϵ so that each filter loses exactly 1.5 dB of the output NSR compared to the optimal $\hat{\mathbf{g}}_{\text{LS}}$. In other words, the cancellation performance loss due to CF sparsity is 1.5 dB in terms of the NSR. The number of nonzero coefficients is then the criterion of our interest.

We utilize two distinct datasets of multichannel recordings of speech performed in real-world situations. The locations of sound sources (target and interferer) are fixed during the recordings. Each source is recorded separately, which allows us to create an arbitrary mixture and evaluate the NSR.

The first dataset originates from the task "Robust blind linear/non-linear separation of short two-sources-two-microphones recordings" of SiSEC 2010[2]. We take 24 speech recordings with various mutual positions of the speakers where the distance of the speakers to microphones ranges from 0.9 m to 1.75 m. The recordings are approximately 1.5-3 s long, sampled at 16 kHz. The second dataset is described in [20]. It contains recordings from an office room with $T_{60} \approx 490$ ms. There are 32 recordings, 16 for a male target voice interfered with male speech and 16 for a female target voice and interfered with another female speaker. Four channels are available; here we use channels 1 and 4. In total, we use 56 recordings to perform the experiment.

The average number of nonzero coefficients of the computed CFs depending on the filter length L is shown in Fig. 2. The sparsest CFs were obtained using the greedy algorithms NAIVE and MINERR. WLASSO yields similar results. The

[2] http://sisec2010.wiki.irisa.fr

filters by LASSO and BPDN have significantly higher numbers of nonzero coefficients. The results of GROUP ($D = 30$) are between those of BPDN and WLASSO.

The results indicate that the inclusion of some prior information about the desired form of the CF helps us improve its sparsity [19]. To this end, WLASSO, NAIVE and MINERR use the properties of the least squares estimate $\hat{\mathbf{g}}_{\mathrm{LS}}$.

4. SEMI-BLIND NOISE EXTRACTION

This section presents a method for blindly estimating an efficient sparse CF from noisy data. This method is based on an assumption that the support of the filter is known. Such prior information can be obtained, for example, during target-only intervals; cf. the previous section. This approach is compared with a similar one that is completely blind. The experiments are evaluated on the data described in the preceding section.

The method derived here is based on the time-domain blind audio separation approach where an observation space is defined; it is decomposed into independent components using Independent Component Analysis (ICA) [24, 25]. The observation space is spanned by rows of a data matrix

$$\mathbf{B} = \begin{bmatrix} x_{\mathrm{L}}(1-D_1) & x_{\mathrm{L}}(2-D_1) & \dots & x_{\mathrm{L}}(N-D_1) \\ x_{\mathrm{L}}(1-D_2) & x_{\mathrm{L}}(2-D_2) & \dots & x_{\mathrm{L}}(N-D_2) \\ \vdots & \vdots & \vdots & \\ x_{\mathrm{L}}(1-D_P) & x_{\mathrm{L}}(2-D_P) & \dots & x_{\mathrm{L}}(N-D_P) \\ x_{\mathrm{R}}(1-d) & x_{\mathrm{R}}(2-d) & \dots & x_{\mathrm{R}}(N-d) \end{bmatrix}, \quad (11)$$

where $D_1, \dots, D_P$ are integer delays. To find the independent components, the BGSEP algorithm from [26] is used. It yields a $(P + 1) \times (P + 1)$ de-mixing matrix $\mathbf{W}$, and the independent components are $\mathbf{C} = \mathbf{W}\mathbf{B}$.

Each row of $\mathbf{W}$ corresponds to a two-input-single-output CF in which the nonzero coefficients of g_{L} have indices $D_1, \dots, D_P$ and $g_{\mathrm{R}} = \delta(n - d)$. The corresponding row of $\mathbf{C}$ is the output of this filter.

To select the CF (the row of $\mathbf{W}$) that efficiently cancels the target, we use the same idea as in [25]. The CF is selected according to the largest element (in absolute value) of the last column of $\mathbf{W}$. This element determines the contribution of x_{R} in the filter output, which must be significant: Signals from both microphones must be involved in the filter's output to achieve spatial null in the target direction.

When no prior knowledge is available, we cannot prefer any special choice of $D_1, \dots, D_P$. Therefore, blind methods such as the one in [24] select all integer delays from 0 through L where L is the length of the CF. Hence, we consider this blind approach where $D_i = i-1, i = 1, \dots, L$, and $L = 100$.

In the semi-blind approach, we select $D_1, \dots, D_P$ as indices of nonzero coefficients of the CF computed by MINERR from a target-only recording. The reason to choose MINERR is that it yields the minimum number of nonzero coefficients in the computed CF as shown in the preceding section. P depends on the size of the support and ranges between 10 and 20.

The results of the comparison, with respect to different values of the initial NSR, are shown in Fig. 3. The results obtained by the proposed semi-blind approach are, on average, better by 2 dB than those by the blind approach. We also include the results of the CFs computed from noise-free data (of the length $L = 100$), by LS and MINERR. The latter results (in terms of the NSR improvement) are independent of the initial NSR, and LS is by 1.5dB better than MINERR, which agrees with the selected tolerance. Note that the blind approach could theoretically achieve the performance of LS while the semi-blind approach is limited by that of MINERR. However, the statistical error due to ICA causes that the semi-blind approach is finally better that the blind one. Moreover, the computational savings are substantial, because the complexity of this time-domain approach rapidly grows with P.

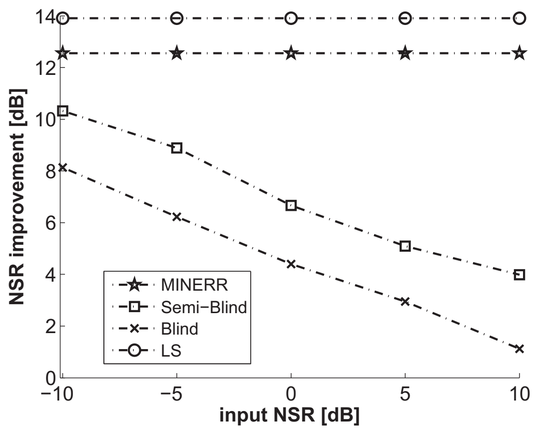

Fig. 3. Comparison of performance of sparse CFs estimated by blind and semi-blind approach. The output NSR is averaged over all test data.

5. CONCLUSIONS

We have presented sparse CFs which are able to achieve target suppression comparably to their dense counterparts and have up to 90 % of the taps in the impulse response equal to zero. We have also proposed a semi-blind method for estimation of sparse CFs from noisy signals which assumes prior knowledge of the filter support. The resulting sparse CFs achieve better target suppression (by more than 2 dB NSR) compared to dense CFs estimated by a fully blind approach.

We recommend another paper [27] of ours presented at this conference as material closely related to the topic of the present paper.

6. REFERENCES

[1] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[2] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann, “Multichannel Bin-Wise Robust Frequency-Domain Adaptive Filtering and Its Application to Adaptive Beamforming,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1340–1351, May 2007.

[3] S. Gannot and I. Cohen, “Speech Enhancement Based on the General Transfer Function GSC and Postfiltering,” *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 6, pp. 561–571, Nov. 2004.

[4] A. Krueger, E. Warsitz, and R. Haeb-Umbach, “Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation,” *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 19, no. 1, Jan. 2011.

[5] S. Doclo and M. Moonen, “GSVD-based optimal filtering for single and multimicrophone speech enhancement,” *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[6] Y. Lin, J. Chen, Y. Kim and D. Lee, “Blind channel identification for speech dereverberation using ℓ_1 norm sparse learning,” *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems*, MIT Press, Vancouver, British Columbia, Canada, December 3-6, 2007.

[7] Z. Koldovský, P. Tichavský, D. Botka, “Noise Reduction in Dual-Microphone Mobile Phones Using A Bank of Pre-Measured Target-Cancellation Filters,” *Proc. of ICASSP 2013*, pp. 679–683, Vancouver, Canada, May 2013.

[8] L. Tong, G. Xu, and T. Kailath, “Blind identification and equalization based on second-order statistics: A time domain approach,” *IEEE Trans. Information Theory*, vol. 40, no. 2, pp. 340–349, 1994.

[9] L. Juan, F. Qiang, Y. Yonghong, “An approach of adaptive blocking matrix based on frequency domain independent component analysis in generalized sidelobe canceller,” *IEEE 10th International Conference on Signal Processing (ICSP)*, pp. 231–234, Oct. 2010.

[10] K. Reindl, S. Markovich-Golan, H. Barfuss, S. Gannot, W. Kellermann, “Geometrically Constrained TRINICON-based relative transfer function estimation in underdetermined scenarios,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 1–4, 2013.

[11] R. Talmon and S. Gannot, “Relative transfer function identification on manifolds for supervised GSC beamformers,” in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sep. 2013.

[12] J. Li, S. Sakamoto, S. Hongo, M. Akagi and Y. Suzuki, “Two-stage binaural speech enhancement with Wiener filter for high-quality speech communication,” *Speech Communication*, vol. 53, no. 5, pp. 677–689, June 2011.

[13] R. Tibshirani, ”Regression Shrinkage and Selection Via the Lasso,” *Journal of the Royal Statistical Society*, Series B, Vol. 58, pp. 267–288, 1994.

[14] S. S. Chen, D. L. Donoho, M. A. Saunders, ”Atomic Decomposition by Basis Pursuit”, *SIAM Journal on Scientific Computing*, Vol. 20, No. 1., pp. 33–61, 1999.

[15] D. L. Donoho, Y. Tsaig, ”Fast Solution of l1-Norm Minimization Problems When the Solution May Be Sparse”, *IEEE Transactions on Information Theory*, Vol. 54, No. 11., pp. 4789–4812, 2008.

[16] T. Shimamura, S. Imoto, R. Yamaguchi, and S. Miyano, ”Weighted lasso in graphical gaussian modeling for large gene network estimation based on microarray data”, *Genome Informatics*, vol. 19, no. 142, 2007.

[17] M. Yuan, Y. Lin, ”Model selection and estimation in regression with grouped variables”, *Journal of the Royal Statistical Society*, Series B, vol. 68, pp. 49–67, 2006.

[18] S. F. Cotter and B. D. Rao, “Sparse Channel Estimation via Matching Pursuit With Application to Equalization,” *IEEE Trans. Commun.*, vol. 50. no. 3, pp. 374–377, March 2002.

[19] E. J. Candès , M. B. Wakin , S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[20] J. Málek, Z. Koldovský, S. Gannot, and P. Tichavský, “Informed Generalized Sidelobe Canceler Utilizing Sparsity of Speech Signals,” *Proc. of IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2013)*, Southampton, UK, Sept. 2013.

[21] E. J. Candès and J. Romberg, ”Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions, ” *Foundations of Computational Mathematics*, Vol. 6, No. 2, pp 227–254, April 2006.

[22] J. Liu, S. Ji, and J. Ye., SLEP: Sparse Learning with Efficient Projections, Arizona State University, [online], 2009.

[23] J. Mairal, F. Bach, J. Ponce and G. Sapiro., “Online Learning for Matrix Factorization and Sparse Coding,” *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[24] Z. Koldovský and P. Tichavský, “Time-domain blind separation of audio sources on the basis of a complete ICA decomposition of an observation space”, *IEEE Trans. on Speech, Audio and Language Process.*, vol. 19, no. 2, pp. 406–416, Feb. 2011.

[25] Z. Koldovský, J. Málek, P. Tichavský, and F. Nesta, “Semi-blind Noise Extraction Using Partially Known Position of the Target Source,” *IEEE Trans. on Speech, Audio and Language Processing*, vol. 21, no. 10, pp. 2029-2041, Oct. 2013.

[26] P. Tichavský, and A. Yeredor, “Fast Approximate Joint Diagonalization Incorporating Weight Matrices,” *IEEE Trans. on Signal Processing*, vol. 57, pp. 878–891, 2009.

[27] Z. Koldovský and P. Tichavský, “A Homotopy Recursive-in-Model-Order Algorithm for Weighted Lasso,” *Proc. of ICASSP 2014 (this conference)*, Florence, Italy, May 2014.