MULTICHANNEL ADAPTIVE FILTERING WITH SPARSENESS CONSTRAINTS

Karim Helwani, Herbert Buchner◇, and Sascha Spors**

* Quality and Usability Lab, Telekom Innovation Laboratories, ◇ Machine Learning Group,
Technische Universität Berlin, 10587 Berlin, Germany
Email:{karim.helwani, sascha.spors}@telekom.de, hb@buchner-net.com

ABSTRACT

The performance of adaptive filtering can be enhanced by incorporating prior system knowledge. In this paper, we systematically consider regularization strategies exploiting sparseness for the identification of acoustic room impulse responses specifically for multichannel systems. Due to the additional dimensions in the multichannel case, a structured regularization appears to be a natural choice. Based on this concept, we present a generic regularized Newton-type algorithm. This generic formulation allows us to discuss various properties specific to the multichannel case and forms a valuable basis for the future development of efficient algorithms.

Index Terms— multichannel adaptive filtering, structured regularization

1. INTRODUCTION

Full-duplex communication in a hands-free communication scenario with multichannel setup (M loudspeakers) requires acoustic echo cancellation (AEC). AEC aims at canceling the acoustic echoes from the microphone signals. Figure 1 shows a block diagram of multichannel AEC with M reproduction channels and a single microphone channel in the receiving room ('near-end'). The signals of the M reproduction channels originate from speech- or audio sources at the far-end. To cancel the echoes arising due to the acoustic path in the near-end the reproduction signals x_m are filtered with the adaptively estimated coefficients $\hat{\mathbf{h}}$, i.e., a replica of the actual acoustic multiple-input single-output (MISO) system. The resulting signal $\hat{y}(n)$ is subtracted from the near-end microphone signal $y = \mathbf{h}^{\mathrm{T}}\mathbf{x}$, with $\mathbf{x}(n) = [\mathbf{x}_1^{\mathrm{T}}(n), \mathbf{x}_2^{\mathrm{T}}(n), \cdots, \mathbf{x}_M^{\mathrm{T}}(n)]^{\mathrm{T}}$, $\mathbf{x}_m(n) = [x_m(n), x_m(n-1), \cdots, x_m(n-L+1)]^{\mathrm{T}}$, n denotes the time instant, and L the filter length. $\widehat{\mathbf{h}}(n)$ denotes the estimated MISO coefficient vector composed from M subfilters,

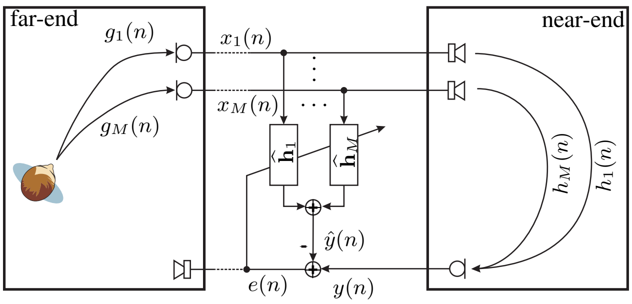

Fig. 1. Block diagram of multichannel acoustic echo cancelation.

$\hat{\mathbf{h}}_m = [\hat{h}_{m,0}, \hat{h}_{m,1}, \cdots, \hat{h}_{m,L-1}]^{\mathrm{T}}$. If the estimated filter coefficients $\hat{\mathbf{h}}$ are equal to the true transfer paths $\mathbf{h}$, all disturbing echoes will be canceled from the microphone signal. Note, that the multiple-input/output (MIMO) case can be considered as a series of independent MISO systems for each microphone channel [1]. Hence, the consideration of a MISO system in the near-end room is sufficient in the context of this work. Most of the popular adaptive filtering algorithms are based on least-squares error minimization [2]

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

and aim at the so-called Wiener solution. It is known that the recursive least-squares (RLS) algorithm is the optimum choice in terms of convergence speed for optimization problems in adaptive filtering based on the least-squares criterion [2]. For multichannel adaptive filtering, the important feature of RLS-type algorithms is that they explicitly take all autocorrelations and also all crosscorrelations between the filter input signals $x_1, \ldots, x_M$ into account for the adaptation process [3, 4]. However, one major problem of the RLS algorithm is the potential numerical instability caused by ill-conditioning due to correlated input signals.

The single channel AEC problem must be regarded as ill-conditioned when the system to be identified is badly excited. This is the case if the input signals $\mathbf{x}_m$ are autocorrelated. The ill-conditioning becomes even worse in the multichannel case, e.g., with stereo reproduction systems. In this case the excitation is highly intra- and inter-channel correlated.

Strategies to cope with the mentioned ill-conditioning problem aim either at enhancing the conditioning by manipulating the input signals $\mathbf{x}_m$, as long as the manipulation can be perceptually tolerated [4, 5]. Or at regularizing the problem to determine an approximate solution that is stable under small changes in the initial data. Regularization incorporates supplementary prior solution knowledge into the ill-conditioned problem. A very popular regularization scheme is the energy-based regularization in the spirit of Tikhonov which can be understood as adding a constraint on the ℓ_2-norm of $\widehat{\mathbf{h}}(n)$. The resulting cost function reads [6]

$$J\left(\widehat{\mathbf{h}}(n)\right) := \hat{\mathcal{E}}\left\{\left(y(n) - \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{x}(n)\right)^2\right\} + \lambda\left\|\widehat{\mathbf{h}}(n)\right\|_2^2. \quad (2)$$

From a probabilistic point of view, regularization is strongly related to the maximum a posteriori criterion (MAP) which reads

$$\widehat{\mathbf{h}}_{\mathrm{opt}} = \arg\max_{\widehat{\mathbf{h}}} \quad p(\widehat{\mathbf{h}}|\mathbf{x}, y). \quad (3)$$

Note that we discarded the time dependency for clarity of presentation. $p(\widehat{\mathbf{h}}|\mathbf{x}, y)$ denotes the a posteriori probability distribution and is given by the Bayesian rule [6],

$$p(\widehat{\mathbf{h}}|\mathbf{x}, y) \propto p(y|\mathbf{x}, \widehat{\mathbf{h}}) \cdot p(\widehat{\mathbf{h}}). \tag{4}$$

The constraint in Eq. (2) corresponds to a prior multivariate normal distribution with zero mean and variance $\mathbf{\Sigma}_{\hat{h}} = \sigma^2_{\hat{h}}\mathbf{I}$

$$p(\widehat{\mathbf{h}}) = \frac{1}{\sqrt{(2\pi)^{ML}|\mathbf{\Sigma}_{\hat{h}}|}} e^{-\frac{1}{2}\widehat{\mathbf{h}}^{\mathrm{T}}\mathbf{\Sigma}_{\hat{h}}^{-1}\widehat{\mathbf{h}}}, \tag{5}$$

where $|\mathbf{\Sigma}_{\hat{h}}|$ denotes the determinant of $\mathbf{\Sigma}_{\hat{h}}$. It is easy to see that maximizing the a posteriori log-likelihood is equivalent to minimizing the cost function in Eq. (2).

Acoustic room impulse responses characterize the reverberant structure of a room. The presence of walls can be modeled by image sources which mirror the actual source and all other images with respect to the walls. Moreover, in practice the length of acoustic impulse responses is limited by the finite time over which the reflections need be considered [7]. This motivates the assumption of sparseness of typical room impulse responses in the time domain, i.e., only a small percentage of their components has significant magnitude while the rest is close to zero. Nonadaptive identification of sparse systems was the subject of several recent studies, e.g., [8, 9].

So far, many studies presented different techniques for sparse adaptive filtering in the single-channel case. E.g., proportionate normalized least mean squares (PNLMS) and exponentiated gradient [1, 10] are efficient gradient based algorithms that exploit the decaying structure of the acoustic impulse response in the time domain. A frequency-domain formulation of a sparse adaptive filtering approach has been developed in [11].

In the multichannel case interchannel correlations are present in addition to the intrachannel signal correlation which makes the ill-conditioning problem more challenging.

It has been shown that multichannel acoustic impulse responses can be regarded as sparse in suited transform domains, such as the frequency or wave domain [12, 13]. In this paper we focus on a spatio-temporal regularization in the time domain and present a rigorous derivation of a Newton-based algorithm for adaptive filtering which takes explicitly the spatio-temporal probability distribution of the multichannel system into account. Furthermore, we discuss some special cases and present simulation results.

2. STRUCTURED REGULARIZATION

As mentioned before, the subfilters of a full-duplex multichannel acoustic communication system are typically sparse in the time domain. Therefore, it is desirable that a space-time regularization strategy exploits this time sparsity.

It has been shown that sparse systems are likely to be Laplace-distributed. Incorporating this prior information in the MAP estimatior leads to a constraint on the ℓ_p-norm, with $p \to 1$ of the subfilters. Prior knowledge about the spatial structure of the impulse responses can be understood as a prior distribution of the norms of every channel. It can be related to a ℓ_q-norm constraint on a vector with components being composed from the ℓ_p-norms of the individual channels. E.g., in many cases one could intuitively assume the ℓ_p-norms of the channels to be normal distributed.

Hence, the concept of structured regularization [14] is promising for MISO adaptive filtering. This aims at minimizing the $\ell_{p,q}$-norm which is defined as

$$\|\widehat{\mathbf{h}}\|_{p,q} := \left(\sum_m \|\widehat{\mathbf{h}}_m\|_p^q\right)^{\frac{1}{q}} = \left(\sum_m \left(\sum_l |\hat{h}_{m,l}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}. \tag{6}$$

Please note, that the traditional ℓ_p-norm can be regarded as the special case of taking the $\ell_{p,p}$-norm.

3. $\ell_{p,q}$-NORM CONSTRAINED ADAPTIVE FILTERING

The sparseness of room impulse responses offers us the possibility to transform the traditional minimization process in the AEC problem into a constrained optimization problem. Hence, the cost function from Eq. (1) can be modified using the Lagrange multipliers formulation into

$$J\left(\widehat{\mathbf{h}}(n)\right) = \hat{\mathcal{E}}\left\{\left(y - \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{x}(n)\right)^2\right\} + \lambda \left\|\widehat{\mathbf{h}}(n)\right\|_{p,q}^q, \tag{7}$$

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.

$$\nabla_{\widehat{\mathbf{h}}_{\mathrm{opt}}} J \stackrel{!}{=} \mathbf{0}. \tag{8}$$

The gradient of the cost function is

$$\begin{aligned}\nabla_{\widehat{\mathbf{h}}} J &= -2\hat{\mathcal{E}}\left\{\mathbf{x}(n)\left[y(n) - \widehat{\mathbf{h}}^{\mathrm{T}}(n)\mathbf{x}(n)\right]\right\} + \lambda\nabla_{\widehat{\mathbf{h}}}\|\widehat{\mathbf{h}}\|_{p,q}^q \\ &= \hat{\mathcal{E}}\left\{-2\mathbf{x}(n)\cdot e(n)\right\} + \lambda\nabla_{\widehat{\mathbf{h}}}\|\widehat{\mathbf{h}}\|_{p,q}^q. \end{aligned} \tag{9a}$$

It can be easily verified that the entries of the vector $\nabla_{\widehat{\mathbf{h}}}\|\widehat{\mathbf{h}}\|_{p,q}^q$ are given as

$$\frac{\partial\|\widehat{\mathbf{h}}\|_{p,q}^q}{\partial\hat{h}_{m,l}} = q\|\widehat{\mathbf{h}}_m\|_p^{q-p}\frac{|\hat{h}_{m,l}|^p}{\hat{h}_{m,l}}. \tag{9b}$$

Determining the zeros of $\nabla_{\widehat{\mathbf{h}}} J$ can be done iteratively with the Newton algorithm.

$$\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) - (\nabla_{\widehat{\mathbf{h}}}\nabla_{\widehat{\mathbf{h}}}^{\mathrm{T}} J(\widehat{\mathbf{h}}(n-1)))^{-1}\nabla_{\widehat{\mathbf{h}}} J(\widehat{\mathbf{h}}(n-1)). \tag{9c}$$

The main advantage of Newton-type adaptation algorithms is their quadratic convergence rate compared to the linear convergence rate of the gradient-based algorithms. Newton type algorithms require the computation of the Hessian matrix [15]

$$\nabla_{\widehat{\mathbf{h}}}\nabla_{\widehat{\mathbf{h}}}^{\mathrm{T}} J(\widehat{\mathbf{h}}(n-1)) = \mathbf{R}_{\mathbf{xx}} + \lambda \cdot \underbrace{\nabla_{\widehat{\mathbf{h}}}\nabla_{\widehat{\mathbf{h}}}^{\mathrm{T}}\|\widehat{\mathbf{h}}\|_{p,q}^q}_{:=\mathbf{G}}. \tag{9d}$$

Usually, the correlation matrix is estimated iteratively using

$$\mathbf{R}_{\mathbf{xx}}(n) = \alpha\,\mathbf{R}_{\mathbf{xx}}(n-1) + \mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n), \tag{9e}$$

where α denotes a forgetting factor.

Hence, in our special case the Hessian reduces to an estimate of the regularized correlation matrix. The entries of the $\mathbf{G}$ are given by differentiation of Eq. (9b) and we derive after several straightforward calculation steps

$$\begin{aligned}\frac{\partial^2\|\widehat{\mathbf{h}}\|_{p,q}^q}{\partial\hat{h}_{m,l}\partial\hat{h}_{m',l'}} =& \delta_{mm'}\,q(q-p)\|\widehat{\mathbf{h}}_m\|_p^{(q-2p)}\frac{|\hat{h}_{m,l}|^p}{\hat{h}_{m,l}}\frac{|\hat{h}_{m,l'}|^p}{\hat{h}_{m,l'}} \\ &+\delta_{mm'}\delta_{ll'}\,q(p-1)\|\widehat{\mathbf{h}}_m\|_p^{(q-p)}\frac{|\hat{h}_{m,l}|^p}{\hat{h}_{m,l}^2},\end{aligned} \tag{9f}$$

hereby, $\delta_{mm'}$ denotes the Kronecker delta.

Hence, the regularization matrix can be decomposed into the sum of two matrices, one *block-diagonal* matrix $\mathbf{G}_{\mathrm{bdiag}}$ with entries given by the first summand of the right hand side of Eq. (9f), and one *diagonal* matrix $\mathbf{G}_{\mathrm{diag}}$ reflected by the second summand.

4. DISCUSSION OF SPECIAL CASES

From Eq. (9f) it can be deduced that for cases where the norm $\ell_{p,q}$ with $p = q$ is considered, the matrix $\mathbf{G}_{\text{bdiag}}$ becomes a zero-matrix and the regularization can be described by adding a diagonal matrix to the correlation matrix. Moreover, for the choice $p = q = 2$ we get for $\mathbf{G}$ the unity matrix multiplied by a scalar which is consistent with the known Tikhonov regularization.

4.1. Multichannel sparse adaptive filtering

In the following we discuss the special case of setting $p = q$. Studying this case offers insights into the properties of the sparseness based regularization in the context of multichannel adaptive filtering and as we will see this choice of the norm parameter leads to an efficient implementation strategy since the regularization matrix $\mathbf{G}$ becomes diagonal as discussed above.
By this p, q configuration the gradient (9b) simplifies to

$$\frac{\partial \|\widehat{\mathbf{h}}_m\|_p^p}{\partial \hat{h}_{m,l}} = p|\hat{h}_{m,l}|^{(p-1)}\text{sgn}(\hat{h}_{m,l}), \tag{10}$$

hereby, $\text{sgn}(\cdot) = \frac{|\cdot|}{\cdot}$ stands for the sign function.
The entries on the main diagonal of $\mathbf{G}$ according to the second term of (9f) are then given as

$$\frac{\partial^2 \|\widehat{\mathbf{h}}_m\|_p^p}{\partial \hat{h}_{m,l}^2} = p(p-1)|\hat{h}_{m,l}|^{(p-2)}. \tag{11}$$

For the limiting case $p = 1$ we derive the sign function for the first derivative, and hence, the following update equation

$$\begin{aligned}\widehat{\mathbf{h}}(n) = \widehat{\mathbf{h}}(n-1) + \Big(\lambda \cdot \text{diag}\Big\{\boldsymbol{\delta}(\widehat{\mathbf{h}}(n-1))\Big\} + \mathbf{R}_{\mathbf{xx}}(n)\Big)^{-1} \\ \cdot \Big(-\lambda \cdot \mu \cdot \text{sgn}(\widehat{\mathbf{h}}(n-1)) + \mathbf{x}(n)e(n)\Big),\end{aligned} \tag{12}$$

where μ is a weighting factor for the gradient of the norm that takes into account the different estimation approaches in practical implementations for the Hessian and the gradient. The Hessian is usually estimated in a recursive way in contrast to the estimation of the gradient which is mostly done by taking the instantaneous value of the vector $\hat{\mathcal{E}}\{\mathbf{x}e\}$. $\boldsymbol{\delta}(\cdot)$ denotes a component-wise Dirac impulse. Hence, once some of the filter coefficients converged to zero, the algorithm can change their values only slowly. This results in relatively bad tracking properties of the adaptive filter. This statement clarifies why most well known single channel sparse adaptive filtering approaches are strongly related to minimization of the ℓ_p-norm for $p \in]1, 2[$. For instance, the IPNLMS algorithm [16, 17].

4.2. Efficient computation of the regularized inverse

As stated in Eq. (8) the optimization process requires the inversion of an $ML \times ML$ matrix in each iteration. This results in very high complexity. The following observation leads to a reduction of the computational complexity.
Since the correlation matrix is estimated iteratively using Eq. (9e). Let assume $\mathbf{R}_{\mathbf{xx}}^{-1}(n-1)$ to be known and

$$\Big(\mathbf{R}_{\mathbf{xx}}(n-1) + \mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n) + \mathbf{G}(n)\Big)^{-1}$$

is required. Since $\mathbf{G} := \nabla_{\widehat{\mathbf{h}}}\nabla_{\widehat{\mathbf{h}}}^{\mathrm{T}}\|\widehat{\mathbf{h}}\|_p^p$ is diagonal, a unitary matrix $\mathbf{U}$ and $\boldsymbol{\Lambda}$ can be efficiently computed representing the eigensystem of a diagonal plus rank-1 matrix [18, 19]

$$\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n) + \mathbf{G}(n) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}. \tag{13}$$

Preliminary experiments of the authors have shown that the rank k, i.e., the number of nonzero eigenvalues is much smaller than ML. Applying the inversion lemma leads to

$$\begin{aligned}(\mathbf{R}_{\mathbf{xx}}(n) + \mathbf{G}(n))^{-1} = \mathbf{R}_{\mathbf{xx}}^{-1}(n-1) - \mathbf{R}_{\mathbf{xx}}^{-1}(n-1) \\ \cdot\mathbf{U}\Big(\boldsymbol{\Lambda}^{-1} + \mathbf{U}^{\mathrm{T}}\mathbf{R}_{\mathbf{xx}}^{-1}(n-1)\mathbf{U}\Big)^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{R}_{\mathbf{xx}}^{-1}(n-1).\end{aligned} \tag{14}$$

Hence, only the inversion of the much smaller matrix

$$\mathbf{A} := \Big(\boldsymbol{\Lambda}^{-1} + \mathbf{U}^{\mathrm{T}}\mathbf{R}_{\mathbf{xx}}^{-1}(n-1)\mathbf{U}\Big)^{-1}, \tag{15}$$

with the size $k \times k$ is needed.

5. ILL-CONDITIONING IN MULTICHANNEL ADAPTIVE FILTERING AND SPARSENESS CONSTRAINT

An advantage of the regularization due to a ℓ_2-constraint is that the ℓ_2 regularization aims at adding the same value to all eigenvalues of an ill-conditioned system. This has the positive effect that all eigenvalues are prevented from becoming zero, hence, they can be inverted and an inversion of the resulting regularized system is ensured. But the resulting system could still have eigenvalues with high multiplicity. Hence, the inversion of the resulting matrix is not unique. In contrast, $\ell_{p\to 1}$ regularization aims at adding large values to the diagonal of the ill-conditioned system at the positions corresponding to the unknown parameters which are likely to become zero. Adding large regularization ($r \to \infty$) for $p = 1$ to the i-th element of the diagonal of $\mathbf{R}_{\mathbf{xx}}$ results in zeroing out the i-th column and i-th row of $\mathbf{R}_{\mathbf{xx}}^{-1}$. Hence, to measure the resulting misalignment [12] we should adapt its definition to

$$\begin{aligned}\mu_{\min} := 10\log\left(\frac{\sigma_\nu^2}{\sigma_x^2}\right)\frac{(1-\alpha)^2}{\|\mathbf{h}\|_2^2}\kappa\{\mathbf{R}_{\mathbf{xx}}^{1/2}\}, \\ \kappa\{\mathbf{R}_{\mathbf{xx}}^{1/2}\} = \sum_{\{\eta:t_\eta<T\}}\sum_{\{\eta':t_{\eta'}^{-1}>\frac{1}{T}\}} t_\eta(t_{\eta'})^{-1}.\end{aligned} \tag{16}$$

where $t_{\eta,\eta'}$ denotes eigenvalues of $\mathbf{R}_{\mathbf{xx}}$ and κ the condition number. Since the condition number considers a smaller matrix it is always smaller or equal to the condition number of the original matrix. Therefore, smaller misalignment could be expected.
It should be noted that the non uniqueness [4] is still not solved. The correlation of the loudspeakers signals leads in general to violation of the convexity assumption on the search space. However, simulations have shown that the sparseness constraint enhances the tracking ability of the algorithm, see Sect. 6, and the adaptive system manages the identification with significantly less preprocessing effort such that improved perceptual quality could be expected.
Moreover, using structured regularization $\ell_{p\to 1,2}$ for estimating multichannel systems with sparse subvectors leads only then to reasonable estimations when the $\ell_{p\to 1}$-norm of the subvectors are in assimilable dimensions otherwise, minimizing the $\ell_{p\to 1,2}$-norm would converge to a solution that equalizes the $\ell_{p\to 1}$-norms of the subvectors in the ℓ_2-sense.

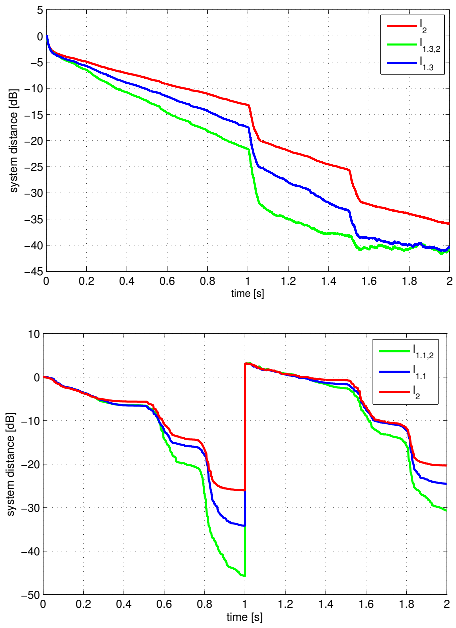

Fig. 2. (a) Achieved enhancement by considering constraints on the ℓ_2, $\ell_{1.3}$, and $\ell_{1.3,2}$-norms, $L = 256$, $M = 2$. (b) Tracking properties of the presented algorithm. Constraints on the ℓ_2, $\ell_{1.1}$, and $\ell_{1.1,2}$-norms for sparse system of length $L = 64$, $M = 2$.

6. EXPERIMENTS

To illustrate the properties of the developed algorithms, an AEC application scenario is considered. The simulation aims at a proof of our concept. More efficient implementations for complex scenarios can be obtained by considering a block formulation for the presented algorithm in a similar manner to the approach in [15].
The near-end room is a small room with a reverberation time (T_{60}) of approximately 20ms containing two loudspeakers, spaced by 1m. In a distance of 1.5m an omni directional microphone is placed. The filter length is $L = 256$ at a sampling rate of $f_s = 8$kHz. Noise with a level of approximately -60dB with respect to the echo was added to the microphone signals, in order to simulate microphone and other noise sources at the near-end. The far-end is a stereo system rendering a virtual source of white noise randomly located between the two loudspeakers. The virtual sources were positioned during the simulation at three different points (position changes after 1s and 1.5s). The stereo signals were preprocessed, as suggested in [4] with a non-linearity rate of only 0.1. The forgetting factor α is set to 0.99, the Lagrange multiplier $\lambda = 0.15$, and μ was set to $3 \cdot 10^{-6}$. The update (9c) was implemented using the pseudoinverse. The red, blue, and green curves in Fig. 6(a) depict the achieved system distance of the estimated MISO system by using a constraint on the ℓ_2, $\ell_{1.3}$, and $\ell_{1.3,2}$-norms respectively. The simulations show the achieved enhancement of the convergence rate by using a sparseness constraint. To show the tracking performance of the presented algorithm systems we give a second example with filter length $L = 64$ and very similar scenario but here we simulated a system change after 1s, by changing the microphone position. Again we simulate a stereo system with one microphone but here each simulated acoustic impulse response is zero except at ten random points. The simulation scenario is suitably adopted for the chosen short filter length by taking a sampling frequency of 1kHz and the virtual source movements in the far-end is done by delaying one of the loudspeaker signals, we preprocessed the loudspeaker signals with a non-linearity rate of 0.05. The position changes in the far-end were now after 0.5s, 0.75s, 1.5s, and 1.75s. All other simulation parameters are the same as in the first experiment. The simulations demonstrate the relation between the sparseness degree of the system and the suitable norm constraint. In Fig. 6(b) the artificially generated system is sparser than the measured one in Fig. 6(a) hence, a constraint on the $\ell_{1.1,2}$ offered the best results.

7. CONCLUSION

In this paper we presented a rigorous derivation of a Newton-based algorithm for adaptive filtering which takes explicitly the spatio-temporal probability distribution of the multichannel system into account. Furthermore, we discussed some special cases. Future work should focus on a block formulation of the presented algorithm for a more efficient implementation in the frequency domain. Exploiting the well known link between the NLMS and the Newton-based algorithms allows the derivation of efficient algorithms with sparseness constraint as special cases of the presented algorithm.

8. REFERENCES

[1] Y. Huang, *Acoustic MIMO signal processing*, Springer, New York, 2006.

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

[3] T. Gänsler and J. Benesty, "Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview," *International Journal of adaptive control and signal processing*, Feb. 2000.

[4] J. Benesty, D.R. Morgan, and M.M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, 1998.

[5] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE ICASSP*, Apr. 2007, vol. 1, pp. I–17–I–20.

[6] Ch. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006.

[7] J. B Allen and D. A Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, 1979.

[8] D. L Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[10] P.A Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation," *Applied Speech and Audio Processing*, vol. 86, no. 6, pp. 1182–1192, June 2006.

[11] P. Loganathan, X.S. Lin, A.W.H. Khong, and P.A. Naylor, "Frequency-domain adaptive multidelay algorithm with sparseness control for acoustic echo cancellation," in *Proc. EUSIPCO*, 2009.

[12] 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.

[13] H. Buchner, S. Spors, and W. Kellermann, "Wave-domain adaptive filtering: Acoustic echo cancellation for full-duplex systems based on wave-field synthesis," in *Proc. IEEE ICASSP*, 2004, pp. iv–117–iv–120.

[14] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[15] H. Buchner, J. Benesty, T. Gansler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1633–1644, 2006.

[16] J. Benesty and S.L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, 2002, vol. 2, pp. 1881–1884.

[17] J. Benesty, C. Paleologu, and S. Ciochină, "Proportionate adaptive filters from a basis pursuit perspective," *Signal Processing Letters, IEEE*, vol. 17, no. 12, pp. 985–988, Dec. 2010.

[18] G.H. Golub and C.F. Van Loan, *Matrix computations*, Johns Hopkins University Press, 1996.

[19] K. Helwani, H. Buchner, and S. Spors, "On the robust and efficient computation of the kalman gain for multichannel adaptive filtering with application to acoustic echo cancellation," in *Proc. IEEE SSC (ASILOMAR)*, Nov. 2010, pp. 988–992.