A SPARSE NONUNIFORMLY PARTITIONED MULTIDELAY FILTER FOR ACOUSTIC ECHO CANCELLATION

Daniele Giacobello and Joshua Atkins

Beats by Dr. Dre
1601 Cloverfield Blvd. Suite 5000N
Santa Monica, CA 90404
{giacobello, joshatkins}@ieee.org

ABSTRACT

In this paper, we propose a formulation of the multidelay adaptive filter for acoustic echo cancellation by modeling the echo path using sparse nonuniform partitions. The nonuniform partitioning allows for a low algorithmic delay without sacrificing the high order of the adaptive filter. It also further improves upon the computational efficiency of the uniformly partitioned multidelay filter by leveraging larger FFT sizes for certain partitions. The sparsity constraint allows for the definition of active and inactive regions of the adaptive filter, providing a better estimate of the order of the filter. Simulation results are provided showing increased convergence speed with the same steady-state misalignment compared to traditional multidelay filtering with both uniform and nonuniform partitioning.

***Index Terms*—** echo cancellation, multidelay filtering (MDF), sparse system identification, thresholding.

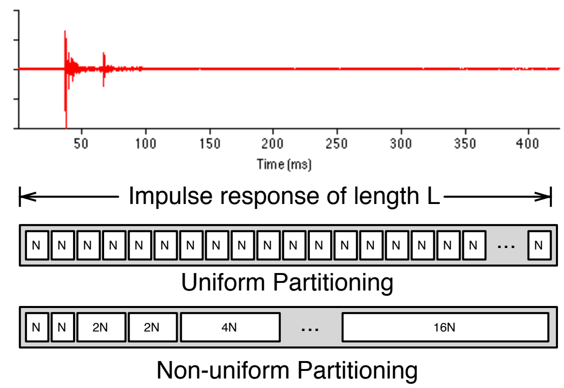

Figure 1: Example of uniform and nonuniform partitioning for MFD.

1. INTRODUCTION

Acoustic echo cancellation (AEC) is a fundamental feature in many devices that support full-duplex speech communication [1]. The fundamental premise is to model the echo path with an L tap adaptive filter and use it to replicate the unwanted signal at the near-end microphone. In currently deployed systems L can easily be on the order of thousands of taps, especially with the ongoing adoption of wideband speech standards.

An efficient way to compute an L tap adaptive filter is to define the problem in the frequency domain where the computational complexity can be greatly reduced by exploiting the efficient Fast Fourier Transform (FFT) implementation [2]. Moving the problem to the frequency domain also allows for decorrelation of the input signal, which is attractive from a convergence perspective [3].

The first approaches of adaptive filtering in the frequency domain were based on FFTs of the same size as the time-domain filter (see, e.g., [2] for an overview). This brought a few practical implementation problems, large quantization error in the FFT, large block delay, and difficulty with tracking time varying statistics due to the nonstationarity of the speech segment analyzed. A more flexible structure was then proposed in [4], where a filter of length L is segmented into K shorter subfilters, significantly reducing the algorithmic delay to $N = L/K$. However, K could be quite large for long impulse responses since delay must be kept reasonably low in real-time communication.

To keep the same processing delay while increasing the efficiency over conventional multidelay filtering (MDF), the partitions can be divided in a non-uniform manner, as shown in Fig. 1. The approach is generally to use shorter partitions near the beginning of the impulse response to achieve low latency and longer partitions towards the end to take advantage of increased computational efficiency (see, e.g., [5–7]).

In echo cancellation systems, the impulse response length L can vary considerably depending on the environment. The solution is then to employ an L tap filter, fixed at some compromise value determined by observation of typical scenarios [8]. It is well known that the minimum mean squared error of the filter output is a monotonic non-increasing function of the tap length [8]. However, in real scenarios a very long filter cannot track changes in echo response well and often suffers from an increase in adaptation noise. Several methods have been proposed to address this issue [9]. However, they all add a significant computational overhead.

While a typical room impulse response is hardly sparse [10], the effect of over-estimating the model order creates a problem where many of the coefficients at the end of the impulse response should be zero. Thus, another way to cope with this issue is to introduce sparsity criteria within the adaptive error minimization [11–15]. This allows for faster convergence because only the coefficients that contribute significantly to the energy in the error are updated.

In this paper, we propose an acoustic echo cancellation algorithm that exploits the computational efficiency and flexibility of the nonuniform MDF. A non-quadratic constraint that promotes sparsity is added in the partitions of the adaptive filter. The sparse constraint at each iteration results in a Landweber iteration with thresholding (or nonlinear shrinkage), for which a rich theory exists to prove its convergence in norm [15, 16].

2. MULTI-DELAY FILTER

2.1. Uniformly partitioned MDF

Let N be the MDF block size, K be the number of blocks and $\mathbf{F}_{2N}$ denote the $2N \times 2N$ Fourier transform matrix, we denote the frequency-domain signals for frame m as

$$\begin{aligned} \mathbf{e}(m) = &\quad \mathbf{F}_{2N}[\mathbf{0}_{1\times N}, e(mN), \ldots, e(mN+N-1)]^T, \\ \mathbf{X}_k(m) = &\quad \mathrm{diag}\{\mathbf{F}_{2N}[x((m-k-1)N-1), \ldots \\ &\quad \ldots, x((m-k+1)N-1)]^T\}, \\ \mathbf{d}(m) = &\quad \mathbf{F}_{2N}[\mathbf{0}_{1\times N}, d(mN), \ldots, d(mN+N-1)]^T. \end{aligned} \tag{1}$$

where $\mathbf{d}(m)$ is signal that we want to model, $\mathbf{e}(m)$ is the modeling error, and $\mathbf{X}_k(m)$ is the observed signal. The MDF algorithm then becomes

$$\mathbf{e}(m) = \mathbf{d}(m) - \hat{\mathbf{y}}(m), \tag{2}$$

$$\hat{\mathbf{y}}(m) = \sum_{k=0}^{K-1} \mathbf{G}_1 \mathbf{X}_k(m) \hat{\mathbf{h}}_k(m-1), \tag{3}$$

with model update

$$\forall k: \quad \hat{\mathbf{h}}_k(m) = \hat{\mathbf{h}}_k(m-1) + \mathbf{G}_2 \boldsymbol{\mu}(m) \boldsymbol{\nabla} \hat{\mathbf{h}}_k(m), \tag{4}$$

$$\boldsymbol{\nabla} \hat{\mathbf{h}}_k(m) = \mathbf{P}^{-1}_{\mathbf{X}_k\mathbf{X}_k}(m) \mathbf{X}_k^H(m) \mathbf{e}(m), \tag{5}$$

where $\mathbf{G}_1$ and $\mathbf{G}_2$ are matrices which select certain time-domain parts of the signal in the frequency domain,

$$\mathbf{G}_1 = \mathbf{F}_{2N} \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix} \mathbf{F}^{-1}_{2N},$$

$$\mathbf{G}_2 = \mathbf{F}_{2N} \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \end{bmatrix} \mathbf{F}^{-1}_{2N}.$$

The matrix $\mathbf{P}_{\mathbf{X}_k\mathbf{X}_k}(m) = \mathbf{X}_k^H(m)\mathbf{X}_k(m)$ is a diagonal approximation of the input power spectral density matrix [4]. To reduce the variance of the power spectrum estimate, the instantaneous power estimate is usually substituted by its smoothed version,

$$\mathbf{P}_{\mathbf{X}_k\mathbf{X}_k}(m) = \beta \mathbf{P}_{\mathbf{X}_k\mathbf{X}_k}(m-1) + (1-\beta)\mathbf{X}_k^H(m)\mathbf{X}_k(m), \tag{6}$$

where β is the smoothing term. We will also assume a fixed step-size for each partition $\boldsymbol{\mu}(m) = \mu_0 \mathbf{I}$.

2.2. Nonuniformly partitioned MDF

Let us now consider a generalization of the original uniformly partitioned MDF formulation with J partitions. In order to simplify the notation, we will assume the length of all the nonuniform partitions is a multiple integer of the first partition of length N. A j^{th} partition will have length $B_j N$, where

$$\mathbf{B} = [B_0, B_1, B_2, \ldots, B_{J-1}] \tag{7}$$

determines the length of the partitions normalized by N and $B_0 = 1$. Generalizing (3), the echo modeling becomes

$$\hat{\mathbf{y}}(m) = \sum_{j=0}^{J-1} \mathbf{G}_{1j} \mathbf{X}_j(m) \hat{\mathbf{h}}_j(m), \tag{8}$$

where $\mathbf{X}_j(m)$ is defined as

$$\begin{aligned} \mathbf{X}_j(m) = \mathrm{diag}\{\mathbf{F}_{2B_jN}[x((m - \sum_{i=1}^{j} B_i - B_j)N), \ldots \\ \ldots, x((m - \sum_{i=1}^{j} B_i + B_j)N - 1)]^T\}. \end{aligned} \tag{9}$$

$\mathbf{G}_{1j}$ is a $2N \times 2B_jN$ matrix that selects the last N samples in the vector and appends zeros on top after performing the inverse Fourier transform. It is a parallel-to parallel operation:

$$\mathbf{G}_{1j} = \mathbf{F}_{2N} \begin{bmatrix} \mathbf{0}_{N\times N} & \ldots & \ldots & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \ldots & \ldots & \mathbf{I}_{N\times N} \end{bmatrix} \mathbf{F}^{-1}_{2B_jN}.$$

The lower-right corner of the block matrix will contain B_j $N \times N$ identity matrices. The rest of the notation in Section 2.1 can then also be easily generalized for the nonuniform case. For clarity, $\mathbf{G}_2$ in (2.1), for the j^{th} component becomes

$$\mathbf{G}_{2j} = \mathbf{F}_{2B_jN} \begin{bmatrix} \mathbf{I}_{B_jN\times B_jN} & \mathbf{0}_{B_jN\times B_jN} \\ \mathbf{0}_{B_jN\times B_jN} & \mathbf{0}_{B_jN\times B_jN} \end{bmatrix} \mathbf{F}^{-1}_{2B_jN}.$$

3. ADAPTATION ALGORITHM WITH A SPARSE PARTITION CRITERION

We now define an adaptive criterion in order to minimize the model mismatch with respect to the filter coefficients. In particular, we are looking for a criterion to retrieve a sparse echo estimate that allows us to turn-off inactive partitions of the filter. This allows us to use a high order multidelay filter where only the partitions that correspond to the actual model in the echo cancellation are active. A parallelism can be seen with model order selection rather than sparse retrieval, as in general, the length of the retrieved echo response will be equal to the order of the filter [9, 12].

Considering the optimization problem at frame m, for each j^{th} component

$$\min_{\mathbf{h}_j(m)\in\mathbb{R}^{B_jN}} \varphi(\mathbf{e}_j(m)). \tag{10}$$

The common optimization problem proposes the minimization of the mean square error, represented by the 2-norm, $\varphi(\cdot) = \|\cdot\|_2^2$. Considering a second-order Taylor approximation of the cost function around the neighborhood of $\hat{\mathbf{h}}_j(m-1)$ [17], the next update point on the error curve is given by (4) and (5).

In our case, we assume that the whole partition vector $\mathbf{h}$ is sparse. We can then add a penalization term to our original problem (10) in order to take this into account:

$$\min_{\mathbf{h}_j(m)\in\mathbb{R}^{B_jN}} \varphi(\mathbf{e}_j(m)) + \psi(\mathbf{h}_j(m)). \tag{11}$$

Measuring sparsity is often associated with the cardinality or 0-norm $\|\cdot\|_0$, a NP-hard problem of combinatorial nature. We solve this somewhat subtle problem by using the 1-norm $\|\cdot\|_1$, which is known throughout the sparse recovery literature to perform well as a relaxation of the 0-norm (see, e.g., [18]). The problem in (11) then becomes

$$\min_{\mathbf{h}_j(m)\in\mathbb{R}^{B_jN}} \|\mathbf{e}_j(m)\|_2^2 + \gamma_j \|\mathbf{h}_j(m)\|_1, \tag{12}$$

where γ_j controls *how sparse* the j^{th} filter should be. One of the most popular methods for solving problem (12) is in the class of iterative shrinkage-thresholding algorithms (ISTA) [16,18], where each iteration involves solving (4) followed by a shrinkage/thresholding step [15].

The common problem found throughout convex optimization uses a first-order gradient method. In our case we use a second-order approximation of the Taylor series [19], where the Hessian is approximated by $\boldsymbol{P}^{-1}_{\mathbf{X}_k\mathbf{X}_k}(m)$. The filter update step of our algorithm, (4), then becomes

$$\forall j: \quad \hat{\mathbf{h}}_j(m) = T_\varepsilon\left(\hat{\mathbf{h}}_j(m-1) + \mathbf{G}_{2_j}\mu_0\boldsymbol{\nabla}\hat{\mathbf{h}}_j(m)\right). \quad (13)$$

In the literature, the operator T generally functions on a element-by-element basis. In our case, we modify it to operate on a partition-by-partition basis. Furthermore, given the structure of our problem we employ a hard-thresholding operator:

$$T_\varepsilon(\mathbf{h}_j) = \begin{cases} 0, & \|\mathbf{h}_j\|_1 \le \varepsilon_j; \\ \mathbf{h}_j, & \|\mathbf{h}_j\|_1 > \varepsilon_j. \end{cases} \quad (14)$$

The hard thresholding in (14) can be seen as criterion to distinguish between active and inactive partitions. Furthermore, much of the theory of optimality of soft-thresholding for sparse approximation carries over to hard-thresholding [18]. The whole algorithm is summarized in Algorithm 1. Since $\varepsilon_j \propto \gamma_j$, the choice of the threshold is not trivial [18]. However, a reasonable choice for ε_j is to be in the order of the estimated noise level normalized for the block length. In the optimization literature, the thresholding theoretical basis can be traced back to the proximal forward-backward iterative scheme introduced in [20] where convergence results are also provided. Furthermore, as shown in [21], the thresholding operation can also be interpreted as the probability Maximization step of an Expectation-Maximization algorithm, where a sparse prior is employed.

Algorithm 1 SNUP-MDF Echo Canceller

Input: uplink signal segment $\mathbf{d}(m)$,
  downlink signal segment $\mathbf{x}(m)$
Output: residual echo $\hat{\mathbf{e}}(m)$
$m \leftarrow 0$
while halting criterion **do**
  for $j \in J$ **do**
    $B_T = \sum_{i=1}^{j} B_i$
    $\mathbf{X}_j = \text{diag}\{\mathbf{F}_{2B_jN}[x((m-B_T-B_j)N),$
      $\ldots, x((m-B_T+B_j)N-1)]^T\}$
  end for
  $\hat{\mathbf{y}}(m) = \sum_{j=0}^{J-1}\mathbf{G}_{1j}\,\mathbf{X}_j(m)\hat{\mathbf{h}}_j(m)$
  $\mathbf{d}_0(m) = \mathbf{F}_{2N}[\mathbf{0}_{1\times N}, d(mN), \ldots d(mN+N-1]\,)^T$
  $\mathbf{e}_0(m) = \mathbf{d}_0(m) - \hat{\mathbf{y}}(m)$
  $\mathbf{e}(t,m) =$ last N terms of $\mathbf{F}^{-1}_{2N}\mathbf{e}_0(m)$
  for $j \in J$ **do**
    $B_T = \sum_{i=1}^{j} B_i$
    $\mathbf{e}_j(m) = \mathbf{F}_{2B_TN}\left[\mathbf{0}_{1\times B_TN}, \mathbf{e}(m-B_T-B_j), \ldots, \mathbf{e}(m)\right]^T$
    $\mathbf{P}_{\mathbf{x}_j\mathbf{x}_j}(m) = \beta\mathbf{P}_{\mathbf{x}_j\mathbf{x}_j}(m) + (1-\beta)\mathbf{X}_j^H(m)\mathbf{X}_j(m)$
    $\boldsymbol{\nabla}\hat{\mathbf{h}}_j(m) = \mathbf{P}^{-1}_{\mathbf{X}_j\mathbf{X}_j}(m)\mathbf{X}_j^H(m)\mathbf{e}_j(m)$
    $\hat{\mathbf{h}}_j(m+1) = \hat{\mathbf{h}}_j(m) + \mathbf{G}_{2j}\mu_0\boldsymbol{\nabla}\hat{\mathbf{h}}_j(m)$
    $\hat{\mathbf{h}}_j(m+1) = T(\hat{\mathbf{h}}_j(m+1))$
  end for
  $m \leftarrow m+1$
end while

4. EXPERIMENTAL RESULTS

For the experimental analysis, we have chosen to analyze the convergence of the algorithm and its behavior in tracking changes of the impulse response. In particular, we focused on changes in length, and thus sparsity, of the echo impulse response. We compared the behavior of the traditional uniformly partitioned MDF (MDF), nonuniformly partitioned MDF (NUP-MDF), and its sparse extension (SNUP-MDF). We considered the measure of sparsity for the model to identify as [22]:

$$\xi(\mathbf{h}) = \frac{L}{L-\sqrt{L}}\left(1 - \frac{\|\mathbf{h}\|_1}{\sqrt{L}\|\mathbf{h}\|_2}\right). \quad (15)$$

This measure is more accurate than $\|\cdot\|_0$ to measure the number of relevant coefficients, as an impulse response with identically zero components is not realistic due to measurement noise.

We considered two setups which are relevant in mobile scenarios. In the first setup, the true impulse response is initially very sparse, then changes into a significantly denser response ($\xi(\mathbf{h}_1) \approx 0.12 \rightarrow \xi(\mathbf{h}_2) \approx 0.67$). In the second setup, we used two similarly sparse responses, a denser one first and a sparser one later ($\xi(\mathbf{h}_1) \approx 0.19 \rightarrow \xi(\mathbf{h}_2) \approx 0.09$). The room impulse responses were calculated in real environments using the Audio Precision APx525 log-swept chirp signal through the *Beats by Dr. Dre Pill Portable Speaker* and truncated to the desired length (f_s=48kHz, resampled at 8 kHz).

We allocated a $L = 1408$ filter and defined $N = 32$, which determined $K = 44$ partitions for the MDF. For the traditional nonuniform case (NUP-MDF), we used the following vector to define the partitions, as shown in (7):

$$\mathbf{B} = [1, 1, 1, 1, 2, 2, 2, 2, 4, 4, 4, 4, 8, 8],$$

where $\sum_j NB_j = 1408$. We defined the downlink signal as an autoregressive signal $x(k) = 0.8x(k-1) + n(k)$, where $n(k)$ is a white Gaussian noise with zero mean and unitary variance. The signal is then convolved with the true impulse response and white noise is added with SNR=25 dB. The noise level determines ε_j used for the thresholding in (14). We also defined $\beta = 0.85$ and $\mu = 0.2$ in the adaptation algorithm.

We then compare the algorithm by measuring the misalignment between the true impulse response, $\mathbf{h}$,and estimated impulse response , $\hat{\mathbf{h}}$ [1]. The results are shown in Figure 2 and Figure 3.

In both cases, the algorithm shows much faster convergence rates than the implementation without the sparsity constraint. It is also easy to see that initially the algorithm very quickly "disables" the high order parts of the filter, thus speeding considerably the convergence. In Figure 2, the algorithm turns rapidly on inactive parts, thus adapting relatively quickly to the longer response. In the case of slight change in response length, as shown in Figure 3, all three algorithms perform roughly the same, however SNUP-MDF uses a filter significantly shorter than the other two methods, this allowing for computational savings. As a general remark, the NUP-MDF performs slightly worse than the traditional MDF due to the fact that we use the same step-size for all partitions [23].

In the current implementation, the complexity of the SNUP-MDF algorithm is in the same order as the complexity of the NUP-MDF. The savings are only related to avoiding the multiplication, FFT, and IFFT of the unused partitions to calculate the output in (8), and, in turn, the error driving the adaptation. In order to react fast to possible changes in the impulse response length, we still calculate

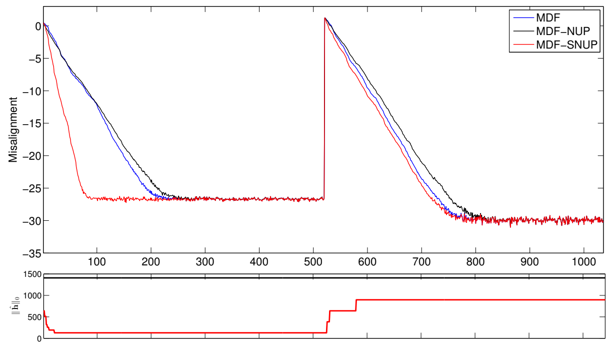

Figure 2: Misalignment comparison for the changing impulse response sparsity $\xi(\mathbf{h}_1) \approx 0.12$, $\xi(\mathbf{h}_2) \approx 0.67$, and estimated cardinality of $\hat{\mathbf{h}}$.

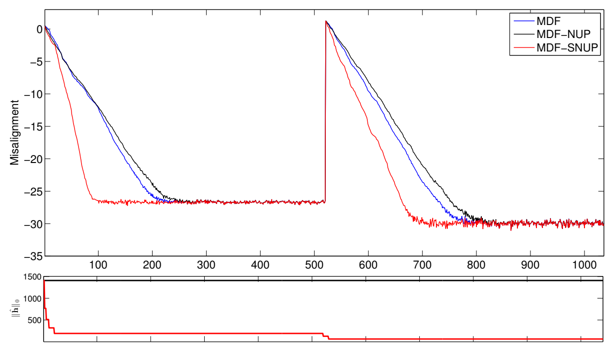

Figure 3: Misalignment comparison for the changing impulse response sparsity $\xi(\mathbf{h}_1) \approx 0.19$, $\xi(\mathbf{h}_2) \approx 0.09$, and estimated cardinality of $\hat{\mathbf{h}}$.

the gradient (5) and the filter update (4), before the thresholding. A more efficient approach would be "probing" the inactive parts only every n samples, using an approach similar to [24]. Furthermore, the long partitions do not need to be updated every B_0 samples, especially when $B_j >> B_0$ the statistics of the Hessian and gradient will vary very slowly. However, the uniformly partitioned MDF has the advantage of allowing a more straightforward implementation of the buffer (1) compared to (9), thus efficient implementation of the nonuniform convolution are still subject of research and definitely not trivial (see, e.g., [25] and references within).

5. CONCLUSIONS

Introducing a sparsity criterion in the nonuniform partition update has shown a substantial improvement in convergence speed and, potentially, a net improvement in computational complexity. However, given the complex architecture, an efficient implementation of the NUP-MDF and SNUP-MDF is not trivial, and it will be the subject of further studies.

6. REFERENCES

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.

[2] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Sig. Proc. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[3] H. Buchner, J. Benesty, T. Gansler, and W. Kellermann, "Robust Extended Multidelay Filter and Double-Talk Detector for Acoustic Echo Cancellation," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 14, no. 5, pp. 1633–1643, Aug. 2006.

[4] J.-S. Soo and K. K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Trans. on Acoustics, Speech and Sig. Proc.*, vol. 38, no. 2, pp. 373–376, Feb. 1990.

[5] W. G. Gardner, "Efficient Convolution without Input-Output Delay," *J. of the Audio Eng. Soc.*, vol. 43, no. 3, pp. 127–136, Feb. 1995.

[6] G. Garcia, "Optimal Filter Partition for Efficient Convolution with Short Input/Output Delay," in *113th AES Convention*, 2002.

[7] G. P. Egelmeers and P. C. Sommen, "A New Method For Efficient Convolution in Frequency Domain by Nonuniform Partitioning for Adaptive Filtering," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3123–3129, 1996.

[8] Y. Gong and C. F. N. Cowan, "A Novel Variable Tap-Length Algorithm for Linear Adaptive Filters," in *IEEE Int. Conf. on Acoustics Speech and Sig. Proc.*, vol. 3, no. 1, 2004, pp. 825–828.

[9] X. Wei, D. G. M. Cruickshank, B. Mulgrew, and F. Riera-Palou, "A Unified Approach to Dynamic Length Algorithms for Adaptive Linear Equalizers," *IEEE Trans. on Sig. Proc.*, vol. 55, no. 3, pp. 908–920, Mar. 2007.

[10] J. Benesty, Y. Huang, J. Chen, and P. Naylor, "Adaptive Algorithms for the Identification of Sparse Impulse Responses," in *Topics in Acoustic Echo and Noise Control*, Springer-Verlag, 2006.

[11] D. L. Duttweiler, "Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancelers," *IEEE Trans. on Speech and Audio Proc.*, vol. 8, no. 5, pp. 508–518, 2000.

[12] R. Martin, W. Sethares, R. Williamson, and J. Johnson, C.R., "Exploiting Sparsity in Adaptive Filters," *IEEE Trans. on Sig. Proc.*, vol. 50, no. 8, pp. 1883–1894, 2002.

[13] P. A. Naylor, J. Cui, and M. Brookes, "Adaptive Algorithms for Sparse Echo Cancellation," *Sig. Proc.*, vol. 86, no. 6, pp. 1182–1192, 2006.

[14] J. Benesty, Y. Huang, and D. R. Morgan, "On a Class of Exponentiated Adaptive Algorithms for the Identification of Sparse Impulse Responses," in *Adaptive Signal Processing*, Springer, 2003, pp. 1–22.

[15] Y. Murakami, M. Yamagishi, M. Yukawa, and I. Yamada, "A Sparse Adaptive Filtering Using Time-Varying Soft-Thresholding Techniques," in *IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, 2010, pp. 3734–3737.

[16] A. Beck and M. Teboulle, "A fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM J. on Imag. Sc.*, vol. 2, no. 1, pp. 183–202, 2009.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[18] D. L. Donoho, "De-noising by Soft-Thresholding," *IEEE Trans. on Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[19] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*, Springer, 2007.

[20] R. Bruck, "On the Weak Convergence of an Ergodic Iteration for the Solution of Variational Inequalities for Monotone Operators in Hilbert Space," *J. of Math. An. and App.*, vol. 61, no. 1, pp. 159–164, 1977.

[21] M. A. Figueiredo and R. D. Nowak, "An EM Algorithm for Wavelet-Based Image Restoration," *IEEE Trans. on Image Proc.*, vol. 12, no. 8, pp. 906–916, 2003.

[22] P. O. Hoyer, "Non-Negative Matrix Factorization with Sparseness Constraints," *The J. of Mach. Learn. Res.*, vol. 5, pp. 1457–1469, 2004.

[23] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, New York, 2003.

[24] Y. Sun, J. Chen, and K. K. Parhi, "Multi-Delay Block Frequency Domain Adaptive Filters with Sparse Partial Subblock Update," in *Rec. 43rd Asilomar Conf. on Sig., Sys. and Comp.*, pp. 206–209.

[25] E. Battenberg and R. Avizienis, "Implementing Real-Time Partitioned Convolution Algorithms on Conventional Operating Systems," in *14th Int. Conf. on Dig. Audio Effects DAFx*, no. 1, 2011.