SET THEORETIC ADAPTIVE FILTERING: USING PERIODOGRAM AND PROJECTIONS ONTO CONVEX SETS

L. Rey Vega, S. Tressens and H. Rey

Facultad de Ingeniería, Universidad de Buenos Aires
Paseo Colon 850 (1063), Buenos Aires, Argentina
{lrey@fi.uba.ar}{stres@fi.uba.ar}{hrey@fi.uba.ar}

ABSTRACT

In this paper we propose a novel adaptive filtering algorithm. Using the Set Theoretic Estimation framework, the algorithm exploits the information given by the power spectral density of the noise extracted from the periodogram of filtering error. With this information new appropriate sets are built and projections onto them are computed. The simulations results show that the algorithm has excellent convergence properties.

1. INTRODUCTION

The problem of adaptive filtering can be interpreted as one in which an unknown system has to be estimated. Adaptive filtering has a great number of applications such as channel equalization, noise cancellation, echo cancellation, etc. [11].

Set Theoretic Estimation has received considerable attention for the last 20 years [2]. It has been applied to a considerable number of problems like image processing [3], signal restoration [10], etc. The idea behind this approach is to use certain *a priori* information about the object to be estimated. The solution is required to be consistent with this information. This is the only requirement to be fulfilled.

The *a priori* information is used to build sets (*property sets*), in such a way that they contain the true object with a high degree of confidence. A solution to the problem can be stated in the following manner: find one element in the intersection of the sets. This task could be very difficult to implement in practice [2].

The application of this framework to adaptive filtering has been reported too. In [4], [6], [7], [9] it was proposed to bound the *feasibility set* (the intersection set built with the sets representing the pieces of *a priori* information) with hyperellipsoids at each time instant. In [12] a method based on parallel subgradient projection (PSP) techniques onto convex sets is utilized for recursive estimation of the true system. In [13] an interesting modification to the PSP algorithm is made, which improves its performance. In those previous works, information about additive noise is used for the construction of the *property sets*. The algorithms derived in those works show excellent convergence properties for highly-colored inputs and reduced number of updates.

This paper proposes a novel adaptive algorithm following the ideas given in [12] and using different *property* sets. These sets use information about the power spectral density of the noise. The periodogram of the filtering error plays a fundamental role in the algorithm for testing the consistency of the successive estimations with the information about the power spectral density of the noise.

This work has been supported in part by *Universidad de Buenos Aires* (I005 UBACyT).

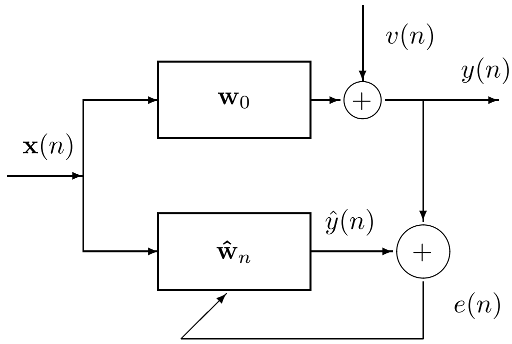

Figure 1: An adaptive filtering problem

Throughout the paper, the following notations are used: $\mathbb{R}^N$ and $\mathbb{C}^N$ are real and complex Hilbert spaces with inner products $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$ and $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H\mathbf{y}$ respectively, where the superscripts T and H denote transposition and complex conjugate transposition. For any nonempty closed convex set $\mathcal{C}$ in a Hilbert space $\mathcal{H}$, the *projection operator* $P_{\mathcal{C}} : \mathcal{H} \to \mathcal{C}$ is defined by $\|\mathbf{x} - P_{\mathcal{C}}(\mathbf{x})\| = \min_{\mathbf{y}\in\mathcal{C}} \|\mathbf{x} - \mathbf{y}\|\ \forall \mathbf{x} \in \mathcal{H}$.

2. PRELIMINARIES

Let $\mathbf{w}_0 = \left[w_0^0\ w_0^1 \dots\ w_0^{N-1}\right]^T \in \mathbb{R}^N$ be an unknown linear FIR system. The associated adaptive filtering problem is shown in Fig. 1. The input signal at time n, $\mathbf{x}(n) = [x(n)\ x(n-1) \dots\ x(n-N+1)]^T \in \mathbb{R}^N$ pass through the system giving an output $\mathbf{w}_0^T\mathbf{x}(n) \in \mathbb{R}$. This output is observed but in this process it usually appears a noise $v(n) \in \mathbb{R}$ which will be considered additive. Thus, each successive input $\mathbf{x}(n)$ gives an output $y(n) = \mathbf{w}_0^T\mathbf{x}(n) + v(n)$. The idea is to find $\hat{\mathbf{w}}_{n+1}$ to estimate $\mathbf{w}_0$. This filter receives the same input $\mathbf{x}(n)$, leading to an output estimation error $e(n) = y(n) - \hat{\mathbf{w}}_n^T\mathbf{x}(n)$.

In the sequel we define the $M \times 1$ output data vector $\mathbf{y}(n) = [y(n-M+1)\ y(n-M+2) \dots\ y(n)]^T$ and the $N \times M$ input data matrix $\mathbf{X}(n) = [\mathbf{x}(n-M+1)\ \mathbf{x}(n-M+2) \dots\ \mathbf{x}(n)]$. It can be defined the $M \times 1$ error vector $\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{X}^T(n)\hat{\mathbf{w}}_n$.

3. THE SET THEORETIC FORMULATION

3.1 Constructing the *property sets*

In the Set Theoretic Estimation framework the solution has to be consistent with the available *a priori* information. In this paper it is assumed that there is some *a priori* information about the additive noise. In fact it is assumed that its power spectral density is known. If a perfect estimation of $\mathbf{w}_0$ is available, $\hat{\mathbf{w}}_n = \mathbf{w}_0\ \forall n$, then $v(n) = y(n) - \mathbf{x}^T(n)\hat{\mathbf{w}}_{n+1}\ \forall n$. It can be proved that $v(n)$ and $y(n) - \mathbf{x}^T(n)\hat{\mathbf{w}}_{n+1}$ have the same probability theoretic properties [1]. Defining the following set:

$$\mathcal{S}_k^n = \left\{ \hat{\mathbf{w}}_{n+1} \in \mathbb{R}^N : \frac{1}{M} \left| \left(\mathbf{y}(n) - \mathbf{X}^T(n)\hat{\mathbf{w}}_{n+1}\right)^T \mathbf{s}_k \right|^2 \leq \xi_k \right\}, \qquad (1)$$

where $\mathbf{s}_k = \left[1\ e^{-j\frac{2\pi k}{M}} \ldots\ e^{-j\frac{2\pi k(M-1)}{M}}\right]^T$, it could be known with a probability $0 < P_k < 1$ that the true system $\mathbf{w}_0$ is in $\mathcal{S}_k^n$. The probability P_k depends on the distribution of the noise and on the parameter ξ_k. It is easy to see that $\mathcal{S}_k^n$ is built by taking the periodogram at frequency bin k of the vector $\mathbf{y}(n) - \mathbf{X}^T(n)\hat{\mathbf{w}}_{n+1}$. It is known that the periodogram is a simple statistic for the spectral density power of a stationary stochastic process. The set $\mathcal{S}_k^n$ is known as a *property set*[2]. In the Set Theoretic framework it is reasonable to seek the solution in this set provided that P_k is close to 1.

3.2 Determining ξ_k

Strictly, to determine ξ_k to guarantee that P_k is closed to 1 the noise's probability distribution has to be known. This kind of knowledge can be difficult to have. But if the noise $v(n)$ is white and gaussian with variance σ^2, it can be shown that I_0/σ^2 and $I_{M/2}/\sigma^2$ have a χ_1^2 distribution, and $2I_1\sigma^2, \ldots\ 2I_{M/2-1}/\sigma^2$ have a χ_2^2 distribution, where:

$$I_k = \frac{1}{M}\left|\mathbf{v}^T(n)\mathbf{s}_k\right|^2 \quad k = 0, 1, \ldots, M-1, \qquad (2)$$

and $\mathbf{v}(n) = [v(n-M+1)\ v(n-M+2) \ldots\ v(n)]^T$. For $k = M/2+1, \ldots, M-1$, the results are the same due to the even symmetry of the periodogram of real signals. The determination of ξ_k for a required probability P_k can be accomplished using chi-squared tables. Moreover, if $v(n)$ is not gaussian or white but it is a strongly mixing process [1] with summable second- and fourth-order cummulant functions and spectral density $g(f)$ with $0 \leq f_k = k/M \leq 1/2 \quad k = 0, 1, \ldots, M/2$, it can be shown that $I_0/g(0)$ and $I_{M/2}/g(1/2)$ are asymptotically distributed as χ_1^2, and $2I_{f_1}/g(f_1), \ldots\ 2I_{f_{M/2-1}}/g(f_{M/2-1})$ are asymptotically distributed as χ_2^2. As a result, in the general case, the sets $\mathcal{S}_k^n$ can be built having knowledge of the spectral density provided that $v(n)$ satisfies the above mentioned hypothesis.

3.3 Solving the problem

It is required to find a point in $\mathcal{S}_k^n$ because this is the consistency condition that any valid solution has to fulfill. Actually, we need to find a point in:

$$\mathcal{S}^n = \bigcap_{k=0}^{M-1} \mathcal{S}_k^n, \qquad (3)$$

to be consistent with all spectrum information. $\{\mathcal{S}_k^n\}_{k=0}^{M-1}$ are closed and convex sets in a Hilbert space. It can be proved easily that $\mathcal{S}^n$ is also a closed and convex set. Then, the concept of a projection in Hilbert space can be applied to find a point in $\mathcal{S}^n$ given an arbitrarily point in the total space[8]. However, the computation of the projection over $\mathcal{S}^n$ can be a formidable task, while the projections over each $\mathcal{S}_k^n$ can be more easily obtained. The *POCS* (*Projections onto Convex Sets*) method can be utilized to find a point in the intersection of a family of closed and convex sets using the individual projections [2]. However, its application in a real time problem which is the nature of adaptive filtering problem can be difficult or even impossible.

In [12] a general algorithm of potential application to a real time problem using the individual projections is derived. It was used with other *property sets*, but it can be used with the sets $\{\mathcal{S}_k^n\}_{k=0}^{M-1}$ defined in (1). Using these sets, the algorithm can be expressed in the following manner:

$$\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n + L_n \left(\sum_{k=0}^{M-1} \lambda_k^n P_{\mathcal{S}_k^n}(\hat{\mathbf{w}}_n) - \hat{\mathbf{w}}_n \right), \qquad (4)$$

where $P_{\mathcal{S}_k^n}$ is the projector onto $\mathcal{S}_k^n$, $\lambda_k^n > 0\ \forall n, k$ and $\sum_{k=0}^{M-1} \lambda_k^n = 1\ \forall n$. The parameter $L_n \in (0, 2M_n)$ is a relaxation parameter[3] and M_n is:

$$M_n = \begin{cases} \frac{\sum_{k=0}^{M-1} \lambda_k^n \| P_{\mathcal{S}_k^n}(\hat{\mathbf{w}}_n) - \hat{\mathbf{w}}_n \|^2}{\| \sum_{k=0}^{M-1} \lambda_k^n P_{\mathcal{S}_k^n}(\hat{\mathbf{w}}_n) - \hat{\mathbf{w}}_n \|^2} & \text{if } \hat{\mathbf{w}}_n \notin \bigcap \mathcal{S}_k^n \\ 1 & \text{otherwise} \end{cases}. \qquad (5)$$

It can be proved that $M_n \geq 1$. In [12] it is shown that the algorithm has the *Fejér-monotonicity* property: for every $\mathbf{w}^* \in \bigcap_{k=0}^{M-1} \mathcal{S}_k^n$:

$$\|\mathbf{w}^* - \hat{\mathbf{w}}_{n+1}\| \leq \|\mathbf{w}^* - \hat{\mathbf{w}}_n\|. \qquad (6)$$

If we assume that $\mathbf{w}_0 \in \bigcap_{k=0}^{M-1} \mathcal{S}_k^n\ \forall n$, the property is true for $\mathbf{w}_0$. These results are still valid taking the projections onto closed and convex sets $\mathcal{C}_k^n$ that satisfy:

$$\mathcal{S}_k^n \subset \mathcal{C}_k^n \text{ and } \hat{\mathbf{w}}_n \notin \mathcal{S}_k^n \Rightarrow \hat{\mathbf{w}}_n \notin \mathcal{C}_k^n. \qquad (7)$$

This last result allows the use of computable projections, if the ones onto the *property sets* are difficult to obtain. In view of this last result in [12], the projections are computed using subgradients of convex functions.

4. THE NEW ALGORITHM

It can be shown that the projections onto the sets $\{\mathcal{S}_k^n\}_{k=0}^{M-1}$ defined in (1) are very difficult to obtain. It can be possible to follow the same steps that in [12] using subgradients. However another approach is possible. In this paper the following sets $\{\mathcal{C}_k^n\}_{k=0}^{M-1}$ are considered:

$$\mathcal{C}_k^n = \left\{\hat{\mathbf{w}}_{n+1} \in \mathbb{C}^N : \frac{1}{M} \left|(\mathbf{y}(n) - \mathbf{X}^T(n)\hat{\mathbf{w}}_{n+1})^T \mathbf{s}_k\right|^2 \leq \xi_k\right\}. \quad (8)$$

These sets are built in $\mathbb{C}^N$ and have the property (7) assuming that $\mathbf{X}(n)$, $\mathbf{y}(n)$ and $\hat{\mathbf{w}}_n$ are real quantities. The projections onto the sets $\{\mathcal{C}_k^n\}_{k=0}^{M-1}$ for each k can be computed more easily using the Lagrange multipliers[8]:

$$P_{\mathcal{C}_k^n}(\hat{\mathbf{w}}_n) = \hat{\mathbf{w}}_n + \alpha_k^n \frac{\mathbf{X}(n)\mathbf{s}_k\mathbf{s}_k^H\mathbf{e}(n)}{\|\mathbf{X}(n)\mathbf{s}_k\|^2}, \quad (9)$$

where

$$\alpha_k^n = \begin{cases} 0 & \text{if } \hat{\mathbf{w}}_n \in \mathcal{C}_k^n \\ 1 - \frac{\sqrt{M\xi_k}}{|\mathbf{e}^T(n)\mathbf{s}_k|} & \text{otherwise} \end{cases} \quad (10)$$

Replacing these results in (4), the algorithm is obtained. For the calculation of α_k^n it is necessary to check if $\hat{\mathbf{w}}_n$ belongs to $\mathcal{C}_k^n$. It is not difficult to show that the following rule applies:

$$\text{If } \frac{1}{M}\left|\mathbf{e}^T(n)\mathbf{s}_k\right|^2 \leq \xi_k \Rightarrow \hat{\mathbf{w}}_n \in \mathcal{C}_k^n. \quad (11)$$

$$\text{If } \frac{1}{M}\left|\mathbf{e}^T(n)\mathbf{s}_k\right|^2 > \xi_k \Rightarrow \hat{\mathbf{w}}_n \notin \mathcal{C}_k^n. \quad (12)$$

The equations (10) and (12) show that the periodogram of the filtering error has to be evaluated for checking the membership of $\hat{\mathbf{w}}_n$ to $\mathcal{C}_k^n$ (and because of (7), to $\mathcal{S}_k^n$). Then the periodogram of the filtering error evaluates the degree of consistency of $\hat{\mathbf{w}}_n$ with the information about the power of the noise at frequency bin k. If this degree of consistency is high enough there is no need of update at this frequency bin.

The parameter α_k^n controls the update in each frequency k. If $\alpha_k^n = 0\ \forall k$ at a given n, it is not difficult to see that $\hat{\mathbf{w}}_{n+1} = \hat{\mathbf{w}}_n$. This possible absence of updates has been reported in the literature in others adaptive algorithms derived according to the Set Theoretic Estimation ideas [4], [6], [7], [9]. Significant saving of computations can be achieved due to this feature of this adaptive algorithm.

It can be shown that the result in (9) is a complex vector. This can be a problematic situation since the final vector $\hat{\mathbf{w}}_{n+1}$ must be real because the true system is assumed to be real. In order to handle with this situation we have proved the following proposition:

Proposition 1 *Given (4) and (5) where each projection is given by (9) and (10) and assuming that* $\mathbf{X}(n)$, $\mathbf{y}(n)$ *and* $\hat{\mathbf{w}}_n$ *are real quantities and* $\lambda_{k-M/2}^n = \lambda_{k+M/2}^n\ \forall k = 1, 2, \ldots, M/2 - 1$ *with* M *even, it can be proved that* $\hat{\mathbf{w}}_{n+1}$ *is a real vector.*

5. NUMERICAL RESULTS

To verify the efficacy of the proposed algorithm, it is compared with the algorithm in [12] (PSP) and the APA algorithm, which is a well-established adaptive algorithm [5] when the input signal is highly-colored. The true system to be estimated is $\mathbf{w}_0 \in \mathbb{R}^{64}$. The input signal is generated by filtering a white, zero-mean, gaussian random sequence through a first-order system $G(z) = 1/1 - 0.95z^{-1}$. This input is highly-colored. The noise is white, zero-mean and gaussian with SNR= $10\log_{10}\left(E\left[\left|\mathbf{w}_0^T\mathbf{x}(n)\right|^2\right]/E\left[|v(n)|^2\right]\right)$=20 dB. The system mismatch, $10\log_{10}\left(\|\mathbf{w}_0 - \hat{\mathbf{w}}_n\|^2/\|\mathbf{w}_0\|^2\right)$ $[dB]$ $\forall n$, is evaluated. The PSP algorithm uses $q = 1$ and $\rho = (r + \sqrt{2r})\sigma^2$ for the parameter that define the corresponding *property sets* [12], where $r = 8$ and σ^2 is the variance of the noise. The order of the APA algorithm is $p = 8$ and $\mu = 1$. The regularization of the APA algorithm take the value of 20 times the power of the input signal, thus following [5]. The proposed algorithm uses $M = 8$. The parameters $\{\xi_k\}_{k=0}^{M-1}$ are computed with chi-squared tables to obtain $P_k = 0.99$ $k = 0, 1, \ldots, M-1$. The coefficients λ_k^n are set equal to $1/M$ $\forall n, k$ in the proposed algorithm and in the PSP algorithm. The technique developed in [13] could be applied to this algorithm to improve its convergence properties. The curves shown are the result of the ensemble of 50 independent trials.

In Fig. 2 the proposed algorithm is compared with the APA algorithm and the PSP algorithm. The proposed algorithm presents almost the same speed of convergence than the APA algorithm with a lower final error. The PSP algorithm, under this kind of input signal, shows a lower speed of convergence, but a lower final error than the APA algorithm. In Fig. 3 the proposed algorithm is tested under different values of M. The speed of convergence and the final error are improved as this parameter becomes larger. However, the performance of the algorithm with $M = 2$ is still good. The algorithm was tested in other conditions (other input signals, different filter lenght, etc.) but the results are not shown due to the lack of space.

Finally, we compared the computational cost of the algorithms. In Fig. 4 we computed the normalized average number of "effective" projectors ($\alpha_k^n \neq 0$) per iteration. This could be thought as an estimator of the probability of computing an "effective" projector at each iteration. In this simulation, both algorithms had nearly the same mismatch curve (not shown). As we can see, the proposed algorithm requires less computations. The total average number of "effective" projectors for the PSP algorithm was 32719.92, and for the proposed one, it was 3022.08.

6. CONCLUSIONS

A novel adaptive algorithm has been proposed in which information about power spectral density of the noise is used. The algorithm has a reduced number of updates and shows excellent convergence properties under highly-colored inputs. This fact make the algorithm suitable for treating problems like echo cancellation. The information about the power spectral density of the

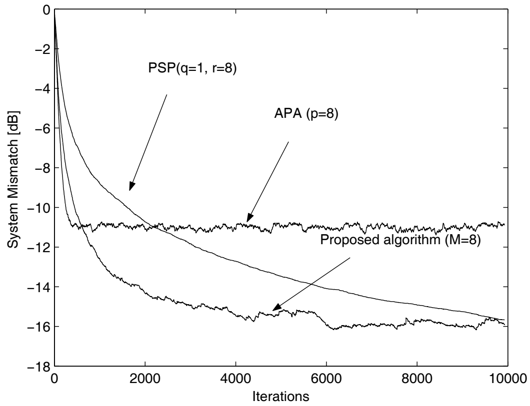

Figure 2: System mismatch for the proposed algorithm ($M = 8$), the APA ($p = 8$) and the PSP algorithm ($r = 8$, $q = 1$) under SNR=20 dB.

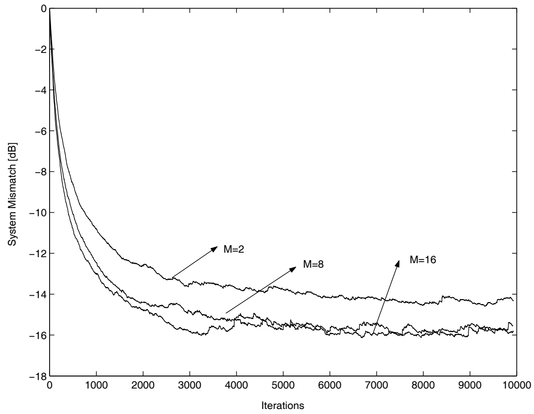

Figure 3: System mismatch for the proposed algorithm with M=2, M=8 and M=16.

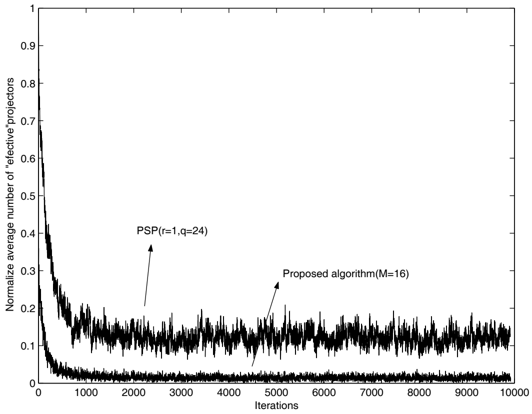

Figure 4: Normalized average number of projectors for the proposed algorithm ($M = 16$) and the PSP algorithm ($q = 24$, $r = 1$) under SNR=20 dB.

noise can be used to improve the convergence behavior of the algorithm when the noise is not white.

Acknowledgement

The authors would like to express their deep gratitude to Prof. Alejandro De la Plaza from University of Buenos Aires for his help in preparing the manuscript.

REFERENCES

[1] P. L. Combettes, J. Trussell"The Use of Noise Properties in Set Theoretic Estimation", *IEEE Transactions on Signal Processing*, vol. 39, no. 7, Jul. 1991, pp. 1630-1641.

[2] P. L. Combettes, "The Foundations of Set Theoretic Estimation", *Proceeding of the IEEE*, vol. 81, no. 2, Feb. 1993, pp. 182-208.

[3] P. L. Combettes, "Convex Set Theoretic Image Recovery by Extrapolated Iterations of Parallel Subgradient Projections", *IEEE Transactions on Image Processing*, vol.6, no. 4, Apr. 1997, pp.493-506.

[4] S. Dasgupta and Y.F. Huang, "Asymptotically Convergent Modified Recursive Least-Squares with Data-Dependent Updating and Forgetting Factor for Systems with Bounded Noise", *IEEE Transactions on Information Theory*, vol. IT-33, no. 3, May 1987, pp. 383-392.

[5] S. L. Gay and S. Tavathia, "The fast affine projection algorithm", *Proc. IEEE ICASSP*, 1995, pp. 3023-2026.

[6] S. Gollamudi, S. Nagaraj, S. Kapoor and Y. F. Huang, "Set-Membership Filtering and a Set-Membership Normalized LMS Algorithm with Adaptive Step Size", *IEEE Signal Processing Lett.*, vol. 5, no. 5, May 1998.

[7] Y.F. Huang, "A Recursive Estimation Algorithm Using Selective Updating for Spectral Analysis and Adaptive Signal Processing ", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 5, Oct. 1986, pp. 1331-1334.

[8] D. Luenberger, *Optimization by Vector Space Methods*, Wiley Professional Paperback Series, John Wiley and Sons, 1969.

[9] S. Nagaraj, S. Gollamudi, S. Kapoor and Y. F. Huang "BEACON: An Adaptive Set-Membership Filtering Technique with Sparse Updates", *IEEE Transactions on Signal Processing*, vol. 47, no. 11, Nov. 1999.

[10] H. J. Trusell and M. R. Civanlar, "The Feasible Solution in Signal Restoration", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, Dec. 1984, pp. 201-212.

[11] Simon Haykin. *Adaptive Filter Theory.* Prentice Hall. Fourth Edition, NJ, 2002.

[12] I. Yamada, K. Slavakis and K. Yamada, "An Efficient Robust Adaptive Filtering Algorithm Based on Parallel Subgradient Projection Techniques", *IEEE Transactions on Signal Processing*, vol. 50, no.5, May 2002, pp.1091-1101.

[13] M. Yukawa and I. Yamada, "Acceleration of Adaptive Parallel Projection Algorithms by Pairwise Optimal Weight Realization", *Proc. EUSIPCO 2004*, Sept. 6-10, Austria, pp. 713-716.