ON AN ITERATIVE METHOD FOR BASIS PURSUIT WITH APPLICATION TO ECHO CANCELLATION WITH SPARSE IMPULSE RESPONSES¹

Pratik Shah^{*}, Steven L. Grant^{2*}, Jacob Benesty⁺

^{*}Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, <u>pratik.shah@mail.mst.edu</u>, <u>sgrant@mst.edu</u>

⁺INRS-EMT, University of Quebec, QC H5A 1K6, Montreal, Canada, <u>benesty@emt.inrs.ca</u>

ABSTRACT

Basis pursuit has been shown to be an effective method of solving inverse problems with a small amount of data when the system to be determined has a sparse representation. Adaptive filters fall under this general category of problems. Here, we use the echo cancellation context to introduce a method of solving the basis pursuit problem with an iterative method based on the proportionate normalized affine projection algorithm (PAPA). Earlier, it has been shown that PAPA can be derived from a basis pursuit perspective. Here we refine the assumptions made in those derivations and show that an iterative form of PAPA yields the same results as basis pursuit without resorting to the simplex method. The resulting algorithm has extremely fast convergence for adaptive filters with very sparse impulse responses. Simulations using the new iterative approach are also presented.

Index Terms— Basis pursuit, adaptive filters, echo cancellation, sparse solutions

1. INTRODUCTION

Basis pursuit (BP) [1] has been shown to be an effective method of solving inverse problems with a small amount of data when the system to be determined has a sparse representation. This is especially attractive to adaptive filtering problems in that it promises the possibility of fast convergence when the solution the adaptive filter seeks is sparse. One such application is network echo cancellation where a number of approaches exploiting its sparse nature have been discussed extensively in the literature, in particular the proportionate normalized least-mean-square (PNLMS) and the proportionate affine projection algorithm (PAPA) [2,3]. One feature of BP is that it requires the minimization of an L_1 norm with equality constraints, i.e.,

$$\min \left\| \mathbf{h} \right\|_{1} \text{ subject to } \mathbf{d} = \mathbf{X}^{T} \mathbf{h} .$$
 (1)

Typically this is accomplished using the simplex method [4]. Here, we propose an iterative approach which is a bit more flexible and amenable to the adaptive filtering context.

Recently, the link between PNLMS, PAPA, and basis pursuit has been shown [5]. There, the authors assume the a posteriori PAPA scaling matrix to be approximately equal to the a priori scaling matrix. This assumption establishes the link between the adaptive filtering algorithms and basis pursuit. However, this is not always an accurate approximation because often these two matrices can be significantly different. We propose an iterative method to improve the approximation. The signal model is described in section 2. Section 3 describes the link between PAPA and BP, section 4 presents the iterative method, and section 5 shows the simulation results and a discussion based on these results. The paper is concluded by section 6.

2. SIGNAL MODEL

The observed or desired signal is given by

$$d(n) = \mathbf{x}^{T}(n)\mathbf{h} + v(n), \qquad (2)$$

where *n* is the discrete time index

$$\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \dots & h_{L-1} \end{bmatrix}^t \tag{3}$$

is the *L*-tap impulse response of the system to be identified, the superscript T denotes the transpose of a vector or a matrix,

$$\mathbf{x}(n) = \left[x(n) \ x(n-1) \ \dots \ x(n-L+1) \right]^T$$
(4)

is a vector containing the *L* most recent samples of the zero mean input at time *n*, and v(n) is the zero mean additive white Gaussian noise which is independent of the input. In the affine projection algorithm (APA) and PAPA [3] it is typical to expand equation (2) to consider, say, *M* samples at a time where $1 \le M \le L$, thus,

$$\mathbf{d}(n) = \mathbf{X}^{T}(n)\mathbf{h} + \mathbf{v}(n)$$
(5)

¹ This work was partly performed under the Wilkens Missouri Endowment

² Formerly Steven L. Gay

where

$$\mathbf{X}(n) = \begin{bmatrix} \mathbf{x}(n) & \mathbf{x}(n-1) & \dots & \mathbf{x}(n-M+1) \end{bmatrix}, \quad (6)$$

$$\mathbf{d}(n) = \left[d(n) \ d(n-1) \ \dots \ d(n-M+1) \right]^T, \tag{7}$$

and

$$\mathbf{v}(n) = \left[v(n) \ v(n-1) \ \dots \ v(n-M+1)\right]^{T}.$$
 (8)

The aim is to find an estimate of \mathbf{h} with an adaptive filter

$$\hat{\mathbf{h}}(n) = \left[\hat{h}_0(n) \ \hat{h}_1(n) \ \dots \hat{h}_{L-1}(n)\right]^T \tag{9}$$

such that the estimation error given by $\|\hat{\mathbf{h}}(n) - \mathbf{h}\|_2^2$ is bounded by an upper bound of ε , which is a small positive number and $\|\cdot\|_2^2$ is the square of the L_2 -norm.

The BP approach to this problem (in the noise-free case) is thus,

$$\min \left\| \hat{\mathbf{h}}(n) \right\|_{1} \text{ subject to } \mathbf{d}(n) = \mathbf{X}^{T}(n) \hat{\mathbf{h}}(n).$$
(10)

It has been shown that if $\hat{\mathbf{h}}(n)$ and $\mathbf{X}(n)$ meet certain conditions, BP provides the sparsest solution for $\hat{\mathbf{h}}(n)$ [1]. The condition on $\hat{\mathbf{h}}(n)$ is that it be *k*-sparse, that is only *k* of its elements are non-zero. The conditions on $\mathbf{X}(n)$ are that it has *M* columns where *M* is some multiple of *k*. Typically, for our problems, *M* is from 10 to 25 times *k*. The matrix $\mathbf{X}^{T}(n)$ must also satisfy the uniform uncertainty principle (UUP) which states that for any *k*sparse vector **f**,

$$0.8\frac{M}{L}\left\|\mathbf{f}\right\|_{2}^{2} \leq \left\|\mathbf{X}^{T}\left(n\right)\mathbf{f}\right\|_{2}^{2} \leq 1.2\frac{M}{L}\left\|\mathbf{f}\right\|_{2}^{2}.$$
 (11)

This can be roughly interpreted as stating that any k-sparse vector **f** shouldn't be either too much in the null space or too much in the signal space of $\mathbf{X}^{T}(n)$. Empirically, we have observed that for Gaussian random excitation $\mathbf{X}^{T}(n)$ satisfies the UUP.

When the above mentioned conditions are met one can use the simplex method to solve (10). The impressive result is that the solution is found using only M samples of filter outputs where typically M is much less than L, the filter length. This demonstrates the power of the sparseness constraint.

3. LINK BETWEEN BP AND PAPA

The affine projection algorithm is

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^{T}(n)\hat{\mathbf{h}}(n-1), \qquad (12)$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n) \left[\mathbf{X}^{T}(n) \mathbf{X}(n) \right]^{-1} \mathbf{e}(n) .$$
(13)

Defining the projection matrix,

$$\mathbf{P}(n) = \mathbf{I} - \mathbf{X}(n) \left[\mathbf{X}^{T}(n) \mathbf{X}(n) \right]^{-1} \mathbf{X}^{T}(n), \qquad (14)$$

the APA of (12) and (13) can be expressed as

$$\mathbf{h}(n) = \mathbf{P}\mathbf{h}(n-1) + \mathbf{h}(n), \qquad (15)$$

where

$$\tilde{\mathbf{h}}(n) = \mathbf{X}(n) \left[\mathbf{X}^{T}(n) \mathbf{X}(n) \right]^{-1} \mathbf{d}(n).$$
(16)

The details are shown in [5]. The vector $\mathbf{\tilde{h}}(n)$ is the minimum- L_2 -norm solution of the linear system of M equations, $\mathbf{d}(n) = \mathbf{X}^T(n)\mathbf{\tilde{h}}(n)$. That is,

$$\min \left\| \tilde{\mathbf{h}}(n) \right\|_{2}^{2} \text{ subject to } \mathbf{d}(n) = \mathbf{X}^{T}(n) \tilde{\mathbf{h}}(n).$$
(17)

Replacing the minimum- L_2 -norm solution with a minimum- L_1 -norm solution we have the optimization problem,

min
$$\|\tilde{\mathbf{h}}(n)\|_{1}$$
 subject to $\mathbf{d}(n) = \mathbf{X}^{T}(n)\tilde{\mathbf{h}}(n),$ (18)

which is the BP problem of (10). Using this optimization problem and the method of Lagrange multipliers it is relatively straight forward [5] to show that

$$\tilde{\mathbf{h}}(n) = \mathbf{G}(n)\mathbf{X}(n) \left[\mathbf{X}^{T}(n)\mathbf{G}(n)\mathbf{X}(n)\right]^{-1} \mathbf{d}(n).$$
(19)

This leads to the coefficient update formula, $\hat{i}(x) = \hat{i}(x-1)$

$$\mathbf{h}(n) = \mathbf{h}(n-1) +$$

+
$$\mathbf{G}(n)\mathbf{X}(n)[\mathbf{X}^{T}(n)\mathbf{G}(n)\mathbf{X}(n)]^{-1}\mathbf{e}(n)$$
 (20)

where

and

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^{T}(n)\hat{\mathbf{h}}(n-1)$$
(21)

$$\mathbf{G}(n) = diag\left\{ \left| \hat{\mathbf{h}}(n) \right| \right\}.$$

But, (20) is intractable and there is no obvious solution for $\hat{\mathbf{h}}(n)$. In [5] this was resolved by using the simple expedient,

$$\mathbf{G}(n) \approx \mathbf{G}(n-1) \tag{23}$$

(22)

thus, removing the difficulty and arriving at the un-relaxed and un-regularized PAPA algorithm,

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{G}(n-1)\mathbf{X}(n) \Big[\mathbf{X}^{T}(n)\mathbf{G}(n-1)\mathbf{X}(n) \Big]^{-1} \mathbf{e}(n).$$
⁽²⁴⁾

3. ITERATIVE METHOD

We propose an iterative method to solve (21) and (20) that will avoid the expedient of (23). The idea is to use (21) and a relaxed and regularized version of PAPA and then repeatedly feedback the resulting coefficient vector. This results in the following algorithm:

Algorithm One

1) Initialize $\hat{\mathbf{h}}^{[0]} = \begin{bmatrix} \iota & \cdots & \iota \end{bmatrix}^T$ 2) For i = 1 through Ia. $\mathbf{e}^{[i]} = \mathbf{d}(n) - \mathbf{X}^T(n) \hat{\mathbf{h}}^{[i-1]}$ b. $\mathbf{G}^{[i-1]} = diag \left\{ \left| \hat{\mathbf{h}}^{[i-1]} \right| \right\}$

$$\hat{\mathbf{h}}^{[i]} = \hat{\mathbf{h}}^{[i-1]}$$

c.
$$+\mu \mathbf{G}^{[i-1]} \mathbf{X}(n) \Big[\mathbf{X}^{T}(n) \mathbf{G}^{[i-1]} \mathbf{X}(n) + \delta \mathbf{I} \Big]^{-1} \mathbf{e}^{[i]}$$

The algorithm is initialized with $\hat{\mathbf{h}}^{[0]} = \begin{bmatrix} \iota & \cdots & \iota \end{bmatrix}^T$, where $\iota \approx 10^{-20}$, with $\mu \approx 0.01$, and $\delta = \iota$. Using this method one obtains virtually the same result as when the simplex method is used to solve the optimization problem of (10). Again, the impressive aspect of this algorithm is that convergence is observed in only *M* sample periods.

One advantage of the iterative approach of algorithm one over the simplex method is that the iterations can easily be implemented using data from different sample periods and the complexity of the many iterations can be spread out over those samples. In addition, we may add "gear shifting" to the stepsize parameter, μ , to accelerate convergence and vary the iteration parameter, I, to lower the computational complexity once the algorithm has converged. These considerations suggest the following approach:

Algorithm Two

- 1) Initialize $\mu = 0.5$, $I = I_{init}$, $\hat{\mathbf{h}}(n) = \begin{bmatrix} t & \cdots & t \end{bmatrix}^T$ where $t = 10^{-20}$
- 2) let $\hat{\mathbf{h}}^{[0]} = \hat{\mathbf{h}}(n)$ and i = 0

3) while
$$i < I$$

a. $i = i + 1$
b. $\mathbf{e}^{[i]} = \mathbf{d}(n) - \mathbf{X}^{T}(n) \hat{\mathbf{h}}^{[i-1]}$
c. $\mathbf{G}^{[i-1]} = diag \left\{ \left| \hat{\mathbf{h}}^{[i-1]} \right| \right\}$
 $\hat{\mathbf{h}}^{[i]} = \hat{\mathbf{h}}^{[i-1]}$
d. $+\mathbf{G}^{[i-1]} \mathbf{X}(n) \left[\mathbf{X}^{T}(n) \mathbf{G}^{[i-1]} \mathbf{X}(n) + \delta \mathbf{I} \right]^{-1} \mathbf{e}^{[i]}$

- 4) let $\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}^{[I]}$ and increment n
- 5) If $\left\| \mathbf{e}^{[I]} \right\|_{2}^{2} < \tau$, let $\mu = 0.005$ and I = 1, otherwise $\mu = 0.5$ and $I = I_{init}$
- 6) goto step 2.

Step (5) is the gear shifting step. Here, if $\left\| \mathbf{e}^{[I]} \right\|_{2}^{2} < \tau$, where

 τ is a threshold typically set in the neighborhood of 0.005, we lower the stepsize and decrease the number of iterations, I. Otherwise we use the initial stepsize (set for fast convergence) and the initial number of iterations I_{init} which is typically set in the neighborhood of 10 to 20.

4. SIMULATIONS

In order to verify the performance of these algorithms, simulations were performed based on the signal model described in section 2. All the simulations were done at a sample rate of 8000 samples per second.

For algorithm one we set the number of iterations to 800, the overall length of the impulse response, L, is set to 100 and the signal to noise ratio (SNR) is set to 40 dB. The sparsity of the impulse response (where the sparsity of a vector is defined as the number of non-zero elements) is set to k = 3 and M is set to 60. The algorithm is applied to each sample period and compared to the simplex method applied to the optimization problem of (18) (again, for each sample period). The coefficient error is shown in figure 1 where there is an echo path change at n = 300. The two algorithms have roughly the same performance, thus verifying that algorithm one is a good replacement for the simplex method.

In the second simulation, algorithm two is compared to PAPA. The signal to noise ratio (SNR) is set to 40 dB, L is set to 100, and k is set to 3. The maximum iterations per sample, I_{init} is set to 12 and M is set to 30. Figure 2 shows the coefficient error of the two algorithms. Algorithm two converges considerably faster than PAPA.

In the third simulation a more realistic situation comparable to the network echo cancellation problem is considered. Here, the echo path impulse response is shown in Figure 3. The SNR is set to 40 dB, L is set to 512, M is set to 30, and I_{init} is set to 12. The comparison is with PAPA. Again, algorithm two has superior convergence and re-convergence performance to PAPA.

5. CONCLUSIONS

Basis pursuit has been shown to be an effective method of solving inverse problems with a small amount of data when the system to be determined has a sparse representation. Here, we use the echo cancellation context to introduce a method of solving the basis pursuit problem with an iterative method based on the PAPA. Earlier, it was shown that PAPA can be derived from a basis pursuit perspective. Here we refined the assumptions made in those derivations and showed that an iterative form of PAPA yields the same results as basis pursuit without resorting to the simplex method. The resulting algorithm has extremely fast convergence for adaptive filters with very sparse impulse responses. Simulations using the new iterative approach were presented.

6. REFERENCES

[1] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol 20, no. 1, pp. 33—62, 1998.

[2] D. L. Duttweiler, "Proportionate normalized least-meansquares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Process.*, vol. 8, pp. 508—518, Sept. 2000.

[3] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin: Springer-Verlag, 2001.

[4] G. Strang, *Linear Algebra and its Applications. Second edition*. New York: Academic Press, 1980.

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



Figure 1: A comparison of algorithm 1 and the simplex method to solve the BP problem. Here the filter length is 100, the sparseness is k = 3, and M = 60.



Figure 2: A comparison of algorithm 2 and PAPA. The filter length is 100, the sparseness is k = 3, and M = 30.







Figure 4: A comparison of algorithm 2 and PAPA. The echo path is that of figure 3 with an echo path length of 512 and M = 30.