A FAST COMPRESSIVE SENSING METHOD WITH APPLICATION TO NETWORK ECHO CANCELLATION

Pratik Shah, Steven L. Grant *, and Jacob Benesty†*

*Missouri University of Science and Technology, Rolla, Missouri
E-mail: sgrant@mst.edu, pratik.shah@mail.mst.edu
+INRS-EMT, University of Quebec, Montreal, Canada
E-mail: benesty@emt.inrs.ca

ABSTRACT

Compressive sensing methods have been effectively used for sparse system identification. Many methods have been proposed to exploit this sparsity to reduce the amount of data required for identification. Most though, have high computational complexity. Recently, an iterative method based on the proportionate affine projection algorithm with row action projections (iPAPA-RAP) has been shown to have good convergence properties with relatively low complexity. Here, we present extensions of that algorithm that significantly speed convergence and as a result lower overall computational complexity. The main improvement is the addition of a zero attractor step with a variable scale factor. Significantly, this scale factor is made to be a function of the sparsity of the estimated system parameters. This greatly improves the convergence behavior of the resulting algorithm. It is compared with iteratively reweighted least-squares (IRLS) and l_0 – zero attracting projections (l_0-ZAP). Results show that the proposed algorithm converges faster with lower overall complexity.

Index Terms**—**PAPA; sparse; IRLS; compressed sensing; adaptive filter; ZiPR.

1. INTRODUCTION

Compressive sensing is a technique that allows the identification of sparse systems with limited data points. Existing algorithms such as IRLS [1] and l_0-*ZAP* [2] provide efficient signal recovery options in such scenarios. Recently, sparse system recovery methods have been applied to the problem of network echo cancellation. A link has been established between basis pursuit and the proportionate normalized least mean squares (PNLMS) [3] and the affine projection algorithm (PAPA) in [4] and [5]. Further work along the same lines has resulted in an iterative algorithm that provides rapid convergence [6]. A lower complexity version of the algorithm has been successfully applied to solve the problem of network echo cancellation [7]. A common feature of all the algorithms has been the minimization of an l_p norm of the echo cancellation coefficients with equality constraints, i.e.,

$$\min \|\mathbf{h}\|_p \text{ subject to } \mathbf{d} = \mathbf{X}^T\mathbf{h}, \tag{1}$$

where $\|\bullet\|_p$ denotes the l_p norm of the N-length vector $\mathbf{h}$, and

$$\|\mathbf{h}\|_p = \left(\sum_{i=0}^{N-1} |h_i|^p\right)^{1/p}. \tag{2}$$

IRLS and the algorithms of [5] and [6] minimize the l_1 norm of the vector whereas l_0-ZAP minimizes an approximation of the l_0 norm of the vector. The work presented in this paper is an extension of [7]. A zero attractor step similar to that of [2] is added to the iterative algorithm to increase the speed of convergence. In addition to the zero-attractor step, an arbitrary p-norm of the coefficient vector is minimized. The use of the arbitrary norm also results in improved convergence. We call the resulting algorithm zero attracting iterated PAPA with RAP, (ZiPR).

The signal model is described in Section 2. Section 3 describes the modifications to the iterative algorithm, specifically the zero-attractor step and the use of arbitrary norms. In Section 4, results of applying this method to the network echo cancellation problem are presented where convergence and computational complexity are compared with IRLS and l_0-ZAP. The last section concludes the paper with scope for future work.

2. SIGNAL MODEL

The observed or desired signal is given by

$$d(n) = \mathbf{x}^T(n)\mathbf{h} + v(n) \tag{3}$$

where n is the discrete time index,

$$\mathbf{h} = \left[h_0 \; h_1 \; \dots h_{N-1}\right]^T \tag{4}$$

is the N-tap impulse response of the system to be identified. The superscript ^{T} denotes the transpose of a vector or a matrix.

$$\mathbf{x}(n) = \left[x(n)\ x(n-1)\ \ldots\ x(n-N+1)\right]^T \tag{5}$$

is a vector containing the N most recent samples of the zero mean input at time sample n. The signal *v(n)* is the zero mean additive white Gaussian noise which is independent of the input. In the affine projection algorithm (APA) and PAPA [4] it is typical to expand equation (3) to consider, say, *M* samples at a time where $1 \le M \le N$, thus,

$$\mathbf{d}(n) = \mathbf{X}^T(n)\mathbf{h} + \mathbf{v}(n), \tag{6}$$

where

$$\mathbf{X}(n) = \left[\mathbf{x}(n)\ \mathbf{x}(n-1)\ \ldots \mathbf{x}(n-M+1)\right], \tag{7}$$

$$\mathbf{d}(n) = \left[d(n)\ d(n\text{-}1)\ \ldots\ d(n-M+1)\right]^T, \tag{8}$$

and

$$\mathbf{v}(n) = \left[v(n)\ v(n\text{-}1)\ \ldots\ v(n-M+1)\right]^T. \tag{9}$$

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

$$\hat{\mathbf{h}}(n) = \left[\hat{h}_0(n)\ \hat{h}_1(n)\ \ldots \hat{h}_{N-1}(n)\right]^T \tag{10}$$

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

3. PROPOSED ALGORITHM

The iterative PAPA algorithm proposed in [6] was made efficient by implementing it using row action projections (RAP) in [7]. The l_1 norm used in [6] and [7] forces the sparsest solution provided certain conditions are met [8]. Here, this approach is generalized by using a PAPA that minimizes the l_p norm as derived in [5]. In addition, we increase the convergence speed of the iterated PAPA-RAP at the price of slightly higher computational complexity by moving the normalization step inside the RAP loop. The resulting algorithm is shown in Algorithm 1 where we have used a particular $p = q$ norm.

We propose adding a zero attraction step to the algorithm to further enhance the rate of convergence. Typically, zero attractor algorithms are derived by adding a scaled version of the negative gradient of the l_p norm of the coefficients to the coefficient update. Consider a particular $p = \rho$ and the cost function,

$$J(\mathbf{h}) = \|\mathbf{h}\|_\rho. \tag{11}$$

Taking the gradient with respect to $\mathbf{h}$,

$$\nabla_{\mathbf{h}} J(\mathbf{h}) = \left(\sum_{i=0}^{N-1} |h_i|^\rho\right)^{\frac{1}{\rho}-1} sign(\mathbf{h}) \circ |\mathbf{h}|^{\rho-1}, \tag{12}$$

where the symbol $\circ$ represents the vector element-by-element or Hadamard product. The zero attractor update, $\mathbf{r}$ then is just

Algorithm 1: Iterated PAPA- RAP

A. *Initialization*:
   $i = 0,\ \mu = 1, I = I_{\text{init}}, J = J_{\text{init}}$
   $\hat{\mathbf{h}} = [\tau, \tau, \ldots, \tau]^T$ where $\tau = 1e^{-6}$
B. *Loop*: while $i < I$
   C. *Initialization*: $j = 0$
   D. *Loop*: while $j < J$
      E. $\mathbf{G} = \dfrac{diag\{|\hat{\mathbf{h}}|\}^{2-q}}{\sum |\hat{h}_i|^{2-q}}$
      F. *Initialization*: $m = 0$
      G. *Loop*: while $m < M$
         a) $\beta = \mu / \left[\mathbf{x}^T(n-m)\mathbf{G}\mathbf{x}(n-m) + \delta\right]$
         b) $e = d(n-m) - \mathbf{x}^T(n-m)\hat{\mathbf{h}}$
         c) $\hat{\mathbf{h}} = \hat{\mathbf{h}} + \beta \mathbf{G}\mathbf{x}(n-m)e$
         d) $m = m + 1$
   H. $j = j + 1$
I. $i = i + 1$

$$\mathbf{r} = -\alpha\lambda \nabla_{\mathbf{h}} J(\hat{\mathbf{h}}), \tag{13}$$

where α is a variable scale factor and λ is a fixed one. Applying equation (12) directly to equation (13) is problematic, because for $0 < \rho < 1$, as elements of $\hat{\mathbf{h}}$ approach zero, the gradient grows without bound. To prevent this we apply an expedient, replacing $sign(\hat{\mathbf{h}})$ element-wise with

$$f(\hat{h}_i) = \begin{cases} \hat{h}_i / \alpha, & |\hat{h}_i| \le \alpha \\ sign(\hat{h}_i), & |\hat{h}_i| > \alpha \end{cases}. \tag{14}$$

Thus, the update vector becomes,

$$r_i = \begin{cases} \dfrac{\kappa}{\alpha}\, sign(\hat{h}_i)|\hat{h}_i|^\rho, & |\hat{h}_i| \le \alpha \\ \kappa\ sign(\hat{h}_i)\ |\hat{h}_i|^{\rho-1}, & |\hat{h}_i| > \alpha \end{cases} \quad \text{for } 0 \le i < N, \tag{15}$$

where

$$\kappa = -\alpha\lambda \left(\sum_{i=0}^{N-1} |\hat{h}_i|^\rho\right)^{\frac{1}{\rho}-1}. \tag{16}$$

We have found that best results are achieved when the variable scale factor, α is set to a measure of the sparseness of the current coefficient estimate given by

Algorithm 2: ZiPR

A. *Initialization*:

$\mu = 1, I = I_{\text{init}}, J = J_{\text{init}}$

$\hat{\mathbf{h}} = [\tau, \tau, ..., \tau]^T$ where $\tau = 1e^{-6}$

B. $i = 0$

C. *Loop*: while $i < I$

D. *Initialization*: $j = 0$

E. *Loop*: while $j < J$

F. $\mathbf{G} = \dfrac{diag\{|\hat{\mathbf{h}}|\}^{2-q}}{\sum |\hat{h}_i|^{2-q}}$

G. *Initialization*: $m = 0$

H. *Loop*: while $m < M$

a) $e = d(n-m) - \mathbf{x}^T (n-m) \hat{\mathbf{h}}$

b) $\beta = \mu / [\mathbf{x}^T (n-m) \mathbf{G} \mathbf{x}(n-m) + \delta]$

c) $\hat{\mathbf{h}} = \hat{\mathbf{h}} + \beta \mathbf{G} \mathbf{x}(n-m) e$

d) $\alpha = -\log_{10} \left(\dfrac{N \|\hat{\mathbf{h}}\|_\infty - \|\hat{\mathbf{h}}\|_1}{\|\hat{\mathbf{h}}\|_\infty (N-1)} \right)$

e) $\kappa = -\alpha \lambda \left(\sum_{i=0}^{N-1} |\hat{h}_i|^\rho \right)^{\frac{1}{\rho}-1}$

f) for $k = 0, ..., N-1$

$$r_k = \begin{cases} \dfrac{\kappa}{\alpha} \operatorname{sign}\{\hat{h}_k\} |\hat{h}_k|^\rho, & |\hat{h}_k| \le \alpha \\ \kappa \, sign(\hat{h}_k) \, |\hat{h}_k|^{\rho-1}, & |\hat{h}_k| > \alpha \end{cases}$$

g) $\hat{\mathbf{h}} = \hat{\mathbf{h}} + \mathbf{r}$

h) $m = m + 1$

I. $j = j + 1$

J. $i = i + 1$

$$\alpha = -\log_{10} \left(\frac{N \|\hat{\mathbf{h}}\|_\infty - \|\hat{\mathbf{h}}\|_1}{\|\hat{\mathbf{h}}\|_\infty (N-1)} \right) \quad (17)$$

This is done because the zero attractor step's function is to find the sparsest solution among a possibly infinite number of solutions to the underdetermined set of equations available in a single sample period. As the estimated coefficients become sparser, the coefficients are presumably nearing the optimal estimate and the scale factor should become smaller to decrease the coefficient estimation noise. The factor

$$S(\hat{\mathbf{h}}) = \frac{N \|\hat{\mathbf{h}}\|_\infty - \|\hat{\mathbf{h}}\|_1}{\|\hat{\mathbf{h}}\|_\infty (N-1)} \quad (18)$$

is similar to Hoyer's sparsity measure [9] where have replace the use of Hoyer's l_2 norm with an l_∞ norm. It is easy to see that

$$0 \le S(\hat{\mathbf{h}}) \le 1 \quad (19)$$

with $S(\hat{\mathbf{h}}) = 0$ when $\hat{\mathbf{h}}$ is least sparse and $S(\hat{\mathbf{h}}) = 1$ when it is most sparse. By taking the negative logarithm of $S(\hat{\mathbf{h}})$ as in (17) we have

$$0 \le \alpha < \infty \quad (20)$$

with $\alpha = 0$ when $\hat{\mathbf{h}}$ is most sparse and α very large when it is not sparse.

The combination of these steps, resulting in the proposed algorithm, ZiPR, is shown in Algorithm 2.

The proportionate steps of Algorithm 1 result in the filter estimate adapting along the axes in the coefficient space, while the zero attraction step forces the coefficients with small magnitudes to zero. Often, in an attempt to find the sparsest solution, coefficients with relatively small magnitude adapt very slowly. As a result, the convergence of the algorithm is constrained. To overcome this problem, we propose the use of norms l_q and l_ρ for the proportionate part and the zero attracting part, respectively to be: $1 < q \le 1.5$ and $0.5 < \rho \le 1$. The selection of q being slightly greater than unity allows the algorithm to adapt coefficients with small magnitudes; meanwhile, the selection of ρ for the zero attraction steps ensures a sparse solution.

4. SIMULATIONS AND RESULTS

Simulations were performed to compare the performance of the proposed algorithm with existing ones. The system to be identified, $\mathbf{h}$, was a normalized coefficient vector of length $N = 1024$ with $K = 30$ non-zero coefficients. The non-zero coefficients were obtained from a zero mean unit variance Gaussian distribution. The excitation signal was also zero mean, unit variance, and Gaussian distributed. The SNR was adjusted to 40 dB. The parameters used for the simulation were $\mu = 0.5$, $q = 1.3$, $\rho = 1$, $J = 1$, $\lambda = 0.01$, and $\delta = 0.1$. For all three algorithms, the projection order was $M = 200$. The performance of the ZiPR algorithm was compared with IRLS and l_0-ZAP. The parameters for these algorithms were adjusted to obtain quickest and deepest convergence.

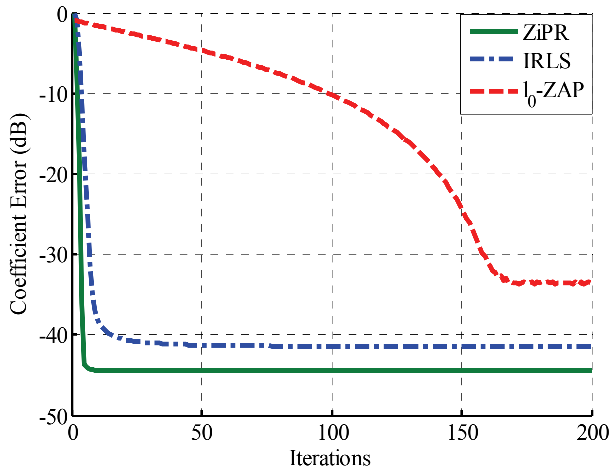

Figure 1: Comparison of convergence of l_0-ZAP, IRLS, and ZiPR versus iterations for a single sample period.

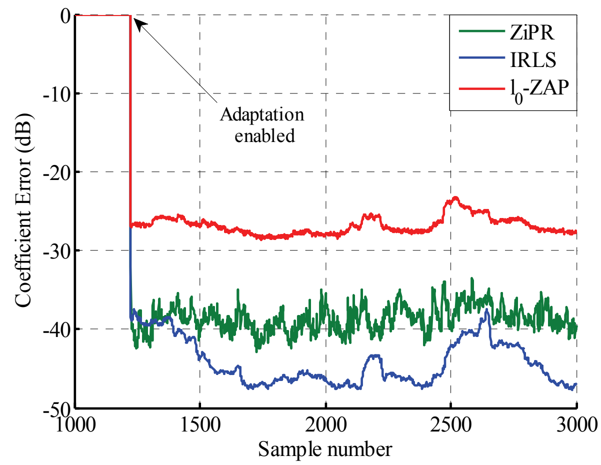

Figure 2: Comparison of convergence of l_0-ZAP, IRLS, and ZiPR in a network echo cancellation scenario with white noise excitation.

Figure 1 shows the comparison of coefficient error versus iteration number *for a single sample period* for all three algorithms. The coefficient error was calculated as

$$CE_{\text{dB}} = 10 \bullet \log_{10} \left\| \mathbf{h}\text{-}\hat{\mathbf{h}} \right\|_2^2 . \tag{21}$$

The l_0-ZAP algorithm converged only to -35 dB coefficient error in 165 iterations and did not converge any further; it suffers from a much slower initial convergence rate compared to the others. The convergence of IRLS was observed to be faster than l_0-ZAP; it converged to -40 dB coefficient error in 15 iterations. The final converged level of ZiPR was slightly better than that of IRLS. The striking result from the figure is the rapid initial convergence of ZiPR. This algorithm converged to less than -40 dB coefficient error in as few as 4 iterations.

In order to compare the computational complexity of the three algorithms, we consider the number of multiplications per iteration as well as the number of iterations required to converge to their final converged levels. The number of multiplications per iteration required by each algorithm is shown in the Table 1. The exponentiation function is considered to be a table look-up and thus is considered to add to the complexity the equivalent to one multiplication.

Although l_0-ZAP requires few multiplications per iteration, it required almost 68 million multiplications to fully converge. Also, the final convergence level of l_0-ZAP is not as low as the other two algorithms. IRLS required 735 million multiplications to converge while ZiPR required only 32 million. In the case of this experiment, since ρ was set to 1, the complexity of the algorithm further reduces by a factor of $2N$. Thus, ZiPR converged to a level 3 dB lower than IRLS and required only 27 million multiplications, about 25 times faster than IRLS and 2.5 times faster than l_0-ZAP.

Table 1: Number of multiplications per iteration for l_0-ZAP, IRLS and ZiPR

IRLS	l_0-ZAP	ZiPR
$2MN + M^2(N+1) + M^3$	$2N(M+1)$	$(13N+7)MJ$

Network impulse response identification is a natural sparse identification problem. Simulations with zero-mean unit-variance white Gaussian noise excitation were performed. For the adaptive filter application we execute Algorithm 2 once each sample period omitting the initialization step after the first sample. The number of iterations for each algorithm was set to the number of iterations it required to converge to its lowest coefficient error. Thus, this number was 165, 15 and 4 for l_0-ZAP, IRLS and ZiPR, respectively. A 1024-tap, 13-sparse network impulse response was used. The non-zero taps with magnitude greater than -40 dB of the largest tap were considered while measuring this sparsity. The true number of non-zero taps was 975. The results are shown in Figure 2. l_0-ZAP converged only to about -25 dB in terms of coefficient error. IRLS converged to 40 dB coefficient error and consistently stays at or below that level. ZiPR converged to -40 dB in one sample period and stayed around that level for the rest of the simulation. There is a marginal difference between the performance of IRLS and ZiPR but the trade-off is acceptable given the large difference in their computational complexities.

Results for colored noise excitation are shown in Figure 3. All parameters are the same as previously. Here, l_0-ZAP performed about the same, but IRLS performed much worse – by a margin of about 20 dB. ZiPR, on the other hand showed the best performance, degrading only by about 5 to 10 dB from the white noise case.

The computational complexity of ZiPR as described above is 27 million multiplications per sample period. This is rather large for network echo cancellation, so we make the following observations:

1. ZiPR's convergence is achieved in a single sample period rather than in thousands of sample periods as with conventional adaptive filters.
2. ZiPR may be easily modified to distribute its complexity over many sample periods. Thus, there is a possible trade-off in convergence and performance that is not available with conventional adaptive filters.
3. Most network echo cancellers are implemented at IP network access points where hundreds to thousands of cancellers are implemented in a concentrated unit. Thus dynamic resource allocation techniques as described in [10] may be used to concentrate computational resources only to those channels which currently need convergence.

5. CONCLUSION

This work extends the iterative adaptive algorithms defined in [6] and [7] by combining a zero attraction step and considering norms other than just l_1, thus providing faster convergence at reduced overall complexity. In addition, a key parameter in the zero attractor step was made a function of the sparseness of the current coefficient estimate. The algorithm is compared to both IRLS and l_0-ZAP and found to have superior convergence performance and lower computational complexity.

6. ACKNOWLEDGMENTS

This work was partly funded by the Wilkens Missouri Endowment.

7. REFERENCES

[1] B. Rao, K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," IEEE Trans. Signal Process., vol. 47, pp. 187—200, 1999.

[2] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filter framework," IEEE Journal of Selected Topics in Signal Process., vol. 4, no. 2, pp. 409—420, Apr. 2010.

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

[4] 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.

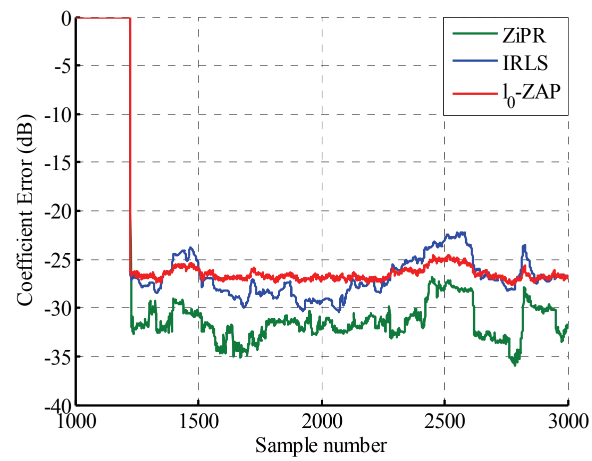

Figure 3: Comparison of convergence of l_0-ZAP, IRLS, and ZiPR in a network echo cancellation scenario with colored noise excitation.

[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.

[6] P. Shah, S. Grant, and J. Benesty, "On an iterative method for basis pursuit with application to echo cancellation with sparse impulse responses," in *Proc. IEEE ICASSP,* 2012, pp. 177—180.

[7] P. Shah, S. Grant, and J. Benesty, "An efficient iterative method for basis pursuit adaptive filters for sparse systems," Proc. APSIPA Annual Summit and Conf., 2012.

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

[9] P.O. Hoyer, "Non-negative sparse coding," Neural Networks for Signal Processing, 2002, Proceedings of the 12th IEEE Workshop on , vol., no., pp. 557- 565, 2002

[10] Gansler, Tomas; Benesty, J.; Sondhi, M.; Gay, S.L., "Dynamic resource allocation for network echo cancellation," Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on , vol.5, no., pp.3233,3236 vol.5, 2001