Sparse Iterative Adaptive Approach with Application to Source Localization

William Rowe and Jian Li
Department of Electrical and Computer Engineering
University of Florida
Gainesville, FL 32611, USA
Email: wrowe001@ufl.edu

Petre Stoica
Department of Information Technology
Division of Systems and Control
Uppsala University
Uppsala, Sweden

***Abstract*—The iterative adaptive approach (IAA) is a spectral estimation algorithm that provides high resolution estimates with as little as a single snapshot. However, IAA is not a sparsity promoting algorithm which might be desirable in specific applications. In this work, we present two approaches for producing sparse IAA estimates. We examine the performance using a source localization example.**

I. Introduction

The estimation of the direction of arrival (DOA) of a plane wave impinging on an array is one of the classical array processing problems. There are many different estimation methods available for this specific problem such as delay-and-sum, Capon's method, or multiple signal classification (MUSIC) [1], [2]. The delay-and-sum method suffers from poor resolution, but can be used with a single snapshot. Capon's method and MUSIC require a good estimate of the array covariance matrix, which requires multiple snapshots [2]. When only a single snapshot is available a spectral estimation algorithm called the Iterative Adaptive Approach (IAA) can provide high resolution estimates [3] (IAA does require gridding of the parameter space while MUSIC does not).

IAA is an iterative method based on a weighted least squares approach that is non-parametric and user parameter free. It typically returns a dense estimate of the impinging signal power spectrum. Here we define a dense vector to be one where the majority of the elements are non-zero and conversely a sparse vector to be one where the majority of the elements are zero. Sparsity promoting algorithm such as Basis Pursuit (BP) or Sparse Bayesian Learning (SBL) have been of significant interest, but these algorithms commonly contain a user parameter that must be tuned carefully [3].

In this work we explore two methods of altering the IAA algorithm to output a sparse vector instead of a dense vector. The first method is a hard threshold detector based on the formulation of IAA. The second approach is a soft thresholding approach which modifies the IAA algorithm. Both methods are able to output sparse vectors while maintaining the robustness of the IAA algorithm.

We begin our discussion with a brief review of the IAA algorithm and the core implementation. We then present the hard thresholding and soft thresholding methods for the IAA algorithm. We follow this with a numerical simulation and comparison to orthogonal matching pursuit (OMP). We then conclude this work with a brief summary and final thoughts.

In this work a vector will be denoted by a bold face lower case letter $\mathbf{x}$ and a matrix will be represented by a bold face upper case letter $\mathbf{X}$. The diag$(\cdot)$ operator maps a vector to the diagonal elements of a square matrix and also maps the diagonal elements of a square matrix to a vector in the opposite case. $(\cdot)^H$ represents a conjugate transpose operation for a vector or matrix.

II. IAA

The Iterative Adaptive Approach is a spectral estimation technique that is based on a weighted least squares minimization. The method was originally proposed for source localization, but has found other applications in imaging, pulse compression, and missing data estimation [3], [4]. IAA assumes that the following general signal model is valid for the data:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} + \mathbf{e}. \tag{1}$$

Here $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is our measured data vector, and $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \ldots, \mathbf{a}(\theta_K)]$ is our steering matrix. Here K represents the grid size in the frequency domain and a steering vector of $\mathbf{A}$ can be written as

$$\mathbf{a}(\theta_k) = [e^{-j(2\pi f/c)x_1 \sin(\theta_k)}, \ldots, e^{-j(2\pi f/c)x_N \sin(\theta_k)}]^T.$$

Here f represents the carrier wave frequency, c is the speed of propagation, and x_n is the position of the nth element. $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_K]^T$ is a vector containing the unknown amplitude and phase at each point in the frequency domain; finally, $\mathbf{e} \in \mathbb{C}^{N\times 1}$ is the noise vector.

We are interested in measuring the signal power $p_k = |\alpha_k|^2$ for $k = 1, 2, \ldots, K$. The signal power matrix $\mathbf{P}$ is defined as a diagonal $K \times K$ matrix with the p_k values on the diagonal. Using p_k then the noise and interference covariance matrix for the kth grid point can be defined as

$$\mathbf{Q}_k = \mathbf{R} - p_k \mathbf{a}_k \mathbf{a}_k^H, \tag{2}$$

where $\mathbf{R} \triangleq \mathbf{A}\mathbf{P}\mathbf{A}^H$. IAA minimizes the following function

978-1-4673-3146-3/13/$31.00 ©2013IEEE

with respect to the α_k values:

$$\sum_{k=1}^{K} ||\mathbf{y} - \alpha_k \mathbf{a}_k||^2_{\mathbf{Q}_k^{-1}}, \tag{3}$$

where $||\mathbf{x}||^2_{\mathbf{Q}_k^{-1}} = \mathbf{x}^H \mathbf{Q}_k^{-1} \mathbf{x}$. This is a complicated cost function since the $\mathbf{Q}_k$ depends on the α_k values. Also notice that the problem decouples for each α_k, hence each α_k can be solved separately for a given $\mathbf{Q}_k$. Therefore, IAA takes an iterative approach to solve this problem.

We start with some initial estimate of $\hat{\boldsymbol{\alpha}}$ (typically from a matched filter/delay-and-sum). Then $\widehat{\mathbf{P}} = \text{diag}(|\hat{\boldsymbol{\alpha}}|^2)$ and $\widehat{\mathbf{R}} = \mathbf{A}\widehat{\mathbf{P}}\mathbf{A}^H$. The optimal solution then for a given $\widehat{\mathbf{Q}}_k^{-1}$ is

$$\hat{\alpha}_k = \frac{\mathbf{a}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{y}}{\mathbf{a}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{a}_k} = \frac{\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}}{\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k}. \tag{4}$$

The equivalence in (4) can be shown using (2) and the matrix inversion lemma [3]. Using this estimate of $\boldsymbol{\alpha}$ then the matrix $\hat{\mathbf{P}}$ is updated and the process is repeated until some stopping criterion is met. For a more in-depth treatment on the derivation of IAA the reader is encouraged to read [3]. We present the generic IAA algorithm in Table I. Here a subscript of (i) represents an iteration marker and ϵ is the threshold of our convergence criterion. When ϵ is small then $\boldsymbol{\alpha}_{(i)} \approx \boldsymbol{\alpha}_{(i-1)}$ (where $\boldsymbol{\alpha}_0$ is our initialization value) and we say that IAA has converged, while β is an iteration threshold which sets a maximum number of iterations. In this paper we choose ϵ to be 10^{-3} and $\beta = 20$. For the basic implementation described in this section, we choose w_k to be 0 for all k.

III. Sparse IAA

Our main focus in this work is to consider how to promote sparsity in the IAA estimate. We will consider two methods to do that. The first method we consider is a hard thresholding that is applied to the IAA output. The second approach uses a soft thresholding and a modification to the IAA cost function.

A. Hard Thresholding

Consider the scenario where IAA has converged which implies that $\boldsymbol{\alpha}_{(i)} \approx \boldsymbol{\alpha}_{(i-1)}$. Then we assume that IAA has minimized (3). Since $\widehat{\mathbf{Q}}_k$ has converged to a stable value the residuals in (3) are whitened. Then for each grid point k, we have a white Gaussian noise (WGN) detector problem with unknown variance ($\hat{\sigma}^2_{\mathrm{k}}$) where our measurement is $\tilde{\mathbf{y}} = \widehat{\mathbf{Q}}_k^{-1/2}\mathbf{y}$ and our whitened noise and interference is $\tilde{\mathbf{w}} = \widehat{\mathbf{Q}}_k^{-1/2}(\sum_{j=1, j\neq k}^{K} \alpha_j \mathbf{a}_j + \mathbf{e})$. We can use a generalized likelihood ratio test (GLRT) to determine a threshold for deciding between $\mathcal{H}_0 : \tilde{\mathbf{y}} = \tilde{\mathbf{w}}$ and $\mathcal{H}_1 : \tilde{\mathbf{y}} = \widehat{\mathbf{Q}}_k^{-1/2}\hat{\alpha}_k \mathbf{a}_k + \tilde{\mathbf{w}}$. The GLRT test for this problem is given by

$$\begin{aligned} \gamma \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} & \frac{||\mathbf{y}||^2_{\widehat{\mathbf{Q}}_k^{-1}}}{||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\widehat{\mathbf{Q}}_k^{-1}}}, \\ 0 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} & T(\mathbf{y}, k), \end{aligned}$$

where

$$T(\mathbf{y}, k) = \gamma ||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\widehat{\mathbf{Q}}_k^{-1}} - ||\mathbf{y}||^2_{\widehat{\mathbf{Q}}_k^{-1}}. \tag{5}$$

As the GLRT detector for a WGN detector with known amplitude and unknown variance is not a constant false alarm rate detector [5], we choose $\gamma = 1 + \frac{\rho}{M}$ where M is the number of real valued measurements in $\mathbf{y}$. This selection of γ is motivated by the general information criterion (GIC) in model order selection which for the sinusoidal estimation problem is of the form:

$$\ln \hat{\sigma}^2 + \frac{\rho\nu}{M} \approx \ln\left(\hat{\sigma}^2 (1 + \frac{\rho\nu}{M})\right), \tag{6}$$

when $M \gg \rho\nu$ [6] where ν corresponds to the model order. Selection of $\rho = 2$ is motivated by the Akaike Information Criteria (AIC) and selection of $\rho = \ln M$ is motived by the Bayesian Information Criteria (BIC) [6]. We set $\nu = 1$ because at each test point k, there is only one real valued parameter under test ($\hat{\sigma}^2_{\mathrm{k}}$).

This detector though would be quite computationally expensive since it utilizes the inverse of each $\widehat{\mathbf{Q}}_k$. We can reduce this to only requiring the inverse of $\widehat{\mathbf{R}}$, which would be precalculated from the last iteration of IAA. To show this note that

$$||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\widehat{\mathbf{Q}}_k^{-1}} = \mathbf{y}^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{y} - \frac{|\mathbf{a}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{y}|^2}{\mathbf{a}_k^H \widehat{\mathbf{Q}}_k^{-1} \mathbf{a}_k},$$

and that

$$\begin{aligned} \widehat{\mathbf{Q}}_k^{-1} &= \widehat{\mathbf{R}}^{-1} + \frac{\widehat{\mathbf{R}}^{-1} \mathbf{a}_k \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} |\alpha_k|^2}{1 - |\alpha_k|^2 \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k} \\ &= \widehat{\mathbf{R}}^{-1} + \frac{\widehat{\mathbf{R}}^{-1} \mathbf{a}_k \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} |\alpha_k|^2}{\mu_k}. \end{aligned} \tag{7}$$

Here we have defined $\mu_k = 1 - |\alpha_k|^2 \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k$. Then we can write

$$\begin{aligned} ||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\mathbf{Q}_k^{-1}} &= \mathbf{y}^H \widehat{\mathbf{R}}^{-1} \mathbf{y} + \frac{|\hat{\alpha}_k|^2 |\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mu_k} \\ & - \frac{|\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mu_k \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k}. \end{aligned}$$

Our GLRT test function $T(\mathbf{y}, k)$ then becomes

$$\begin{aligned} T(\mathbf{y}, k) &= \frac{\rho}{M} \mathbf{y}^H \widehat{\mathbf{R}}^{-1} \mathbf{y} + \frac{\rho}{M} \frac{|\hat{\alpha}_k|^2 |\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mu_k} \\ & -(1 + \frac{\rho}{M}) \frac{|\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mu_k \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k}, \\ &= \frac{\rho}{M} \mathbf{y}^H \widehat{\mathbf{R}}^{-1} \mathbf{y} - (1 + \frac{\rho}{M}) \frac{|\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}|^2}{\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k} \kappa, \end{aligned} \tag{8}$$

where $\kappa = \frac{1}{\mu_k}(1 - \frac{\rho}{\rho+M} |\hat{\alpha}_k|^2 \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k)$. Equation (8) is a more computationally efficient version of (5), since it reuses $\widehat{\mathbf{R}}^{-1}$. Our hard threshold detector has the following form:

$$\begin{cases} \hat{\alpha}_k = 0 & \text{if } T(\mathbf{y}, k) \geq 0, \\ \hat{\alpha}_k = \hat{\alpha}_k & \text{if } T(\mathbf{y}, k) < 0. \end{cases}$$

B. Soft Thresholding

IAA iteratively minimizes (3), which includes no penalty on $|\alpha_k|^2$ to help induce a sparse result. For this reason we consider the following modified cost function

$$\sum_{k=1}^{K} ||\mathbf{y} - \alpha_k \mathbf{a}_k||^2_{\mathbf{Q}_k^{-1}} + \sum_{k=1}^{K} w_k |\alpha_k|^2. \tag{9}$$

This problem decouples for $k = 1, 2, \ldots, K$ for a given $\mathbf{Q}_k$. Then for a given k we seek to minimize with respect to α_k

$$||\mathbf{y} - \alpha_k \mathbf{a}_k||^2_{\mathbf{Q}_k^{-1}} + w_k |\alpha_k|^2, \quad w_k \geq 0. \tag{10}$$

We choose w_k to mimic the GIC such that

$$w_k = \frac{||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\widehat{\mathbf{Q}}_k^{-1}}}{M|\hat{\alpha}_k|^2}, \tag{11}$$

where $\hat{\alpha}_k$ is the most recent estimate of α_k. This mimics GIC because when the $\hat{\alpha}_k \approx \alpha_k$ then (10) becomes

$$||\mathbf{y} - \hat{\alpha}_k \mathbf{a}_k||^2_{\widehat{\mathbf{Q}}_k^{-1}}(1 + \frac{1}{M}) = \hat{\sigma}^2(1 + \frac{1}{M}), \tag{12}$$

which resembles (6) when $\nu = 1$ and $\rho = 1$.

The solution to (10) is

$$\alpha_k = \frac{\mathbf{a}_k^H \mathbf{Q}_k^{-1} \mathbf{y}}{\mathbf{a}_k^H \mathbf{Q}_k^{-1} \mathbf{a}_k + w_k}. \tag{13}$$

Just as in IAA the values of $\mathbf{Q}_k$ depend on the α_k so we use the same iterative approach to solve this problem by replacing $\mathbf{Q}_k$ with $\widehat{\mathbf{Q}}_k$. We also want to remove the usage of the $\widehat{\mathbf{Q}}_k$ term and rely only on $\widehat{\mathbf{R}}$. Using the same substitution as (7) the solution can be rewritten as

$$\hat{\alpha}_k = \frac{\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{y}}{\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1} \mathbf{a}_k + w_k \mu_k}. \tag{14}$$

We note that w_k can also be re-written in terms of $\widehat{\mathbf{R}}$ by

$$\begin{aligned} w_k \;=\; & \frac{1}{M|\hat{\alpha}_k|^2}\Bigg(\mathbf{y}^H \widehat{\mathbf{R}}^{-1}\mathbf{y} \\ & - \frac{|\mathbf{a}_k^H \widehat{\mathbf{R}}^{-1}\mathbf{y}|^2}{\mathbf{a}_k^H \mathbf{R}^{-1}\mathbf{a}_k}\left(\frac{1 - |\hat{\alpha}_k|^2 \mathbf{a}_k^H \widehat{\mathbf{R}}^{-1}\mathbf{a}_k}{\mu_k}\right)\Bigg). \end{aligned} \tag{15}$$

The IAA algorithm is then modified to use (14) instead of the original $\hat{\alpha}_k$ update formula.

This method produces sparse results, but it is sensitive to grid size selection. If the K is chosen to be too large then the output is biased downward significantly. Using a grid size approximately 10 times finer than the expect resolution provides adequate results in our empirical results. Furthermore, the w_k term includes the value of $\hat{\alpha}_k$ in the denominator which may result in computational problems in implementation when a $\hat{\alpha}_k$ goes to zero. Therefore, if the w_k takes on a NaN value due to a division by zero, then $\hat{\alpha}_k$ is replaced with zero. Furthermore, since many $\hat{\alpha}_k$ go to zero, $\widehat{\mathbf{R}}$ can become ill-conditioned. If the condition number of $\widehat{\mathbf{R}}$ approaches machine epsilon, then we terminate the algorithm and use the last estimate of $\hat{\boldsymbol{\alpha}}$.

TABLE I: IAA

Step 0 - Initialize: Calculate $\boldsymbol{\alpha}_{(0)} = \mathbf{A}^H \mathbf{y}$. and set $i = 1$.
Step 1: Calculate $\mathbf{P}_{(i)} = \text{diag}(\boldsymbol{\alpha}_{(i-1)})$ and $\mathbf{R} = \mathbf{A}\mathbf{P}_{(i)}\mathbf{A}^H$.
Step 2: Calculate w_k using (15).
Step 3: Calculate $\alpha_{k,(i)} = \frac{\mathbf{a}_k^H \mathbf{R}_{(i)}^{-1}\mathbf{y}}{\mathbf{a}_k^H \mathbf{R}_{(i)}^{-1}\mathbf{a}_k + w_k \mu_k}$.
Step 4: If $\lvert\lvert\boldsymbol{\alpha}_{(i)} - \boldsymbol{\alpha}_{(i-1)}\rvert\rvert_2^2 < \epsilon$ or $i \geq \beta$, then quit. Otherwise $i = i + 1$ and go to step 1.

IV. Examples

We will examine the performance of the methods using a plane wave source localization example. We assume we have a 12 element uniform linear array with $\lambda/2$ spacing where λ is the wavelength. We will estimate the DOA using 4 different sparse methods, the first being orthogonal matching pursuit (OMP) [7] where we have set the stopping threshold to 0.01. The second and third methods are the hard thresholding methods of IAA which we refer to as IAA-AIC and IAA-BIC respectively. The fourth method is the soft thresholding method IAA which we refer to as IAA-Soft. In our simulation, there are three incoming waves at $-25, 20, 30$ degrees respectively and all have an amplitude of 1. The phase of each wave is randomly distributed between 0 and 2π. We include additive circularly symmetric complex Gaussian noise to distort our signal. The steering matrix $\mathbf{A}$ was chosen such that each column corresponding to the integer direction of arrivals from -89 to 90 degrees (180 grid points).

Fig. 1 shows the target locations, the standard delay-and-sum estimate, the clairvoyant Capon estimate (this estimate is made using the true noise and interference covariance matrix), and the IAA estimate for only one snapshot when the signal-to-noise ratio (SNR) is 20 dB. The delay-and-sum method in this example is unable to resolve the two close point sources while IAA is similar to the clairvoyant Capon estimate. Fig. 1 also contains the sparse estimates for the IAA-AIC and IAA-BIC estimates. IAA-BIC is a more strict threshold and is more sparse than the AIC threshold which can be seen in Fig. 1. Fig. 2 shows the IAA-Soft and OMP results for the same data as Fig. 1 as well as the dense estimates for reference. OMP split the peaks at 20 and 30 degrees, which resembles the delay-and-sum peak, while the IAA-Soft correctly estimate the two sources.

We ran this simulation 1000 times at each specified SNR. The SNR was varied from 5 dB to 30 dB in steps of 2.5 dB. We chose to measure the performance via probability of detection and power estimation error. We declare a detection if energy is detected at grid points within 2 degrees of the true target locations. For example, the target at 20 would be declared detected if we detect energy at $18, 19, 20, 21,$ or 22 degrees. The maximum energy in each region is used as our power estimate. The absolute error squared of the power is measured for each target as well. The error term is averaged to determine the mean square error (MSE) for the power estimates at each SNR value. IAA converged on average in 10.2 iterations and IAA-Soft converged in 14.78 iterations.

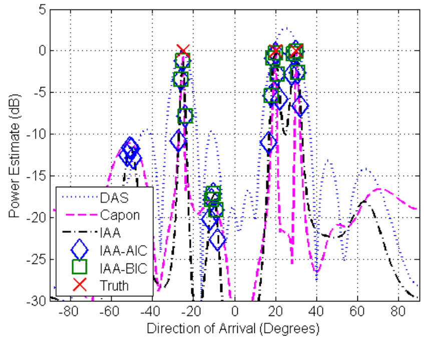

Fig. 1: Angular Power Estimates (IAA-AIC, IAA-BIC)

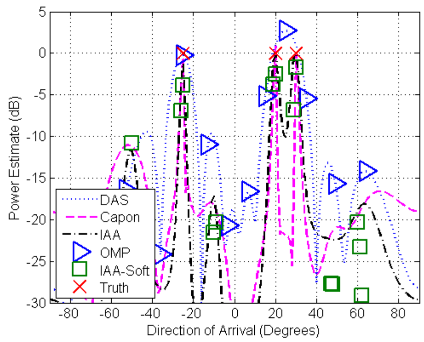

Fig. 2: Angular Power Estimates (OMP, IAA-Soft)

IAA-Soft exited from iterations due to an ill-conditioned $\widehat{\mathbf{R}}$ (on approximately iteration 13) in 49 of the 11,000 trials.

Fig. 3 shows the measured probabilities of detection. IAA-AIC and IAA-BIC offer higher detection rates compared to the IAA-Soft and OMP results. OMP is most likely suffering here due to the closely spaced sources at 20 and 30 degrees and the single snapshot. In Fig. 4 it can be seen that OMP offers a better MSE of the power estimate than IAA-Soft, but offers slightly worse performance than IAA-AIC and IAA-BIC. IAA-Soft provides sparse results, but tends to be biased downwards which results in a larger MSE in the power estimation. IAA-BIC provides the sparsest results while maintain good MSE and detection performance.

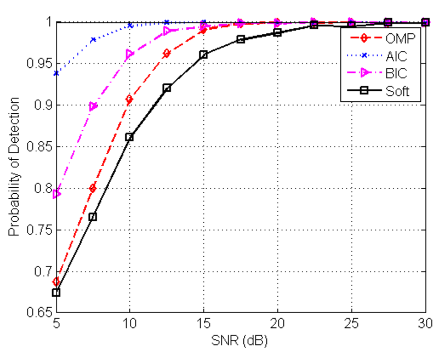

Fig. 3: Probability of Detection

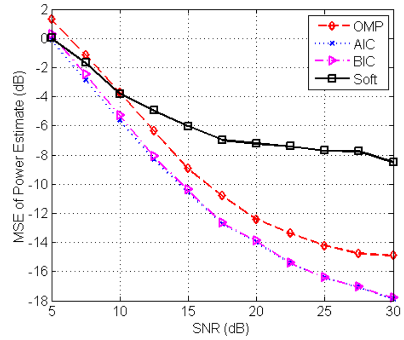

Fig. 4: MSE of Power Estimate (dB)

V. Conclusions

In this work we have reviewed IAA and presented three methods of altering IAA to promote sparsity. From our empirical results IAA-BIC provides the best level of sparsity while maintaining a high detection rate, good sparsity level, and low MSE of the power estimates. The method performs well with a single snapshot of data. The methods were compared to orthogonal matching pursuit as a reference for measuring performance.

Acknowledgment

This work was supported in part by the Swedish Research Council (VR), by the Office of Naval Research (ONR) under Grant No. N00014-12-1-0381, NSF CCF-1218388, and the SMART fellowship program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distributed reprints for Governmental purposes notwithstanding any copyright notation thereon.

References

[1] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY: Wiley, 2002.

[2] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[3] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, “Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares,” *IEEE Transactions of Aerospace and Electronic Systems*, vol. 46, pp. 425–443, Jan. 2010.

[4] J. Karlsson, W. Rowe, L. Xu, G. Glentis, and J. Li, “Fast missing-data IAA with application to notched spectrum SAR,” *IEEE Transactions on Aerospace and Electronic Systems*, to be published.

[5] S. Kay, *Fundamentals of Statistical Signal Processing Volume II: Detection Theory*. Upper Saddle River, New Jersey: Prentic Hall, 1998.

[6] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” *IEEE Signal Processing Magazine*, vol. 21, pp. 36–47, Jul. 2004.

[7] Y. Pati, R. Rezaifar, and P. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” *1993 Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 40–44, Nov. 1993.