Towards More Accurate Radio Telescope Images

Nezihe Merve Gürel[1,2], Paul Hurley[2], Matthieu Simeoni[2,3]
[1] Department of Computer Science, ETH Zürich
[2] Cognitive Computing and Industry Solutions, IBM Research
[3] School of Computer and Communication Sciences, EPFL
nezihe.guerel@inf.ethz.ch, {pah,meo}@zurich.ibm.com

Abstract

Radio interferometry usually compensates for high levels of noise in sensor/antenna electronics by throwing data and energy at the problem: observe longer, then store and process it all. We propose instead a method to remove the noise explicitly before imaging. To this end, we developed an algorithm that first decomposes the instances of antenna correlation matrix, the so-called visibility matrix, into additive components using Singular Spectrum Analysis and then cluster these components using graph Laplacian matrix. We show through simulation the potential for radio astronomy, in particular, illustrating the benefit for LOFAR, the low frequency array in Netherlands. Least-squares images are estimated with far higher accuracy with low computation cost without the need for long observation time.

1. Introduction

In radio astronomy, signals are inherently weak. Noise in antennas is thus significant (far stronger than the signal), and a low signal-to-noise ratio (SNR) results. Relative to the brightest source in the sky, the SNR is usually on the order of -30 dB or less [11, 14, 18].

To date, this thermal noise is circumvented through collecting large amounts of data (long observation times) together with large numbers of antennas (whose role is noise resilience as well as spatial resolution). In addition, when data is heavily corrupted, it is thrown away [16], wasting resources.

This work proposes to explicitly denoise the visibility matrix using a subspace algorithm. The ultimate goal is to obtain accurate output in the end of the processing chain and thus one may choose to observe for a shorter time, or to use less antennas and still obtain better quality images by reducing substantially the cost of building a phased-array.

Typically, noise structure is different from the true pattern underlaying in the visibility matrix, hence the noise can usually be separated from the visibility matrix by subspace methods. In this paper, we specifically study the case where thermal noise follows an additive white Gaussian distribution and no correlation exists among antennas. Our intuition behind employing subspace methods can then be justified as follows. Consider the visibility matrix samples at a *specific-time instance*. These samples represent the same snapshots taken at different time instances, which differs merely by a negligible phase difference among each other. That is, the underlaying structure in the visibility matrix samples is essentially similar yet corrupted by different noise. Therefore, given this similarity throughout the visibility matrix samples and corruption by different noise, subspace methods seem promising for distinguishing these two components.

Related Work: To the best of our knowledge, denoising via subspace methods has extensively been studied when noise is far lower than the true signal. For example, Cadzow's method successfully performs noise reduction for high [3] as well as for moderate SNR signals [4] but works poorly in a low SNR regime. Singular Spectrum Analysis (SSA) is likewise powerful in denoising high SNR signals [5]. In [15], authors propose a subspace based identification method in the presence of small disturbance for linear time-varying channels, which also potentially function noise reduction for high SNR phased-array signals. This work, on the other hand, is built upon [7, 9, 8, 6].

Problem Statement: Consider L closely located antennas where sample taken by i'th antenna at time t_n is denoted by $x_i(t_n)$. The entries of visibility matrix instance $\mathbf{\Sigma}_{i,j}(t_n)$ is estimated by the cross product between i, j'th antennas, i.e., $x_j(t_n)x_i(t_n)^\dagger$ where † stands for the Hermitian transpose, also i and j denote row and column indices, respectively.

Radio interferometers measure electromagnetic radiation from space which are used to deduce a sky image. Interferometers first estimate cross-product between the time series measurements $\mathbf{\Sigma}_{i,j}(t_n)$, that is called visibilities, average them through a certain time interval and form visibility matrix such that $\mathbf{\Sigma} = 1/N \sum_{n=1}^{N} \mathbf{\Sigma}(t_n)$ for N time samples, and finally sky image is estimated from the visibility matrix by a specified imaging technique. The various techniques for mapping from the visibility matrix to the sky image are throughly studied in the literature, ranging from CLEAN [10] to A(W)-projection [13, 2].

2. Proposed Method

In this section, we propose a method to denoise visibility matrix instances separately by exploiting the similarity in between, even under significant noise corruption. The approach decomposes the low SNR visibility matrix instances into additive components. It then learns, by spectral clustering, the representative components for the noise by detecting outliers.

2.1. Decomposition via Singular Spectrum Analysis

SSA has attracted much attention by permitting us to create algorithms with high learning capability to extract meaningful features from a given data set [15]. Opening up its role regarding noise reduction, SSA reconstructs noise from the residual and uninterpretable feature space under high SNR constraint. The methodology for decomposition part can be given as follows.

1st Step: Decomposition

At this step, the noisy visibility matrix $\mathbf{\Sigma}(t_n)$ is analysed by performing an SVD. The SVD of $\mathbf{\Sigma}(t_n)$ is given by the product of three matrices $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_W)$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_W)$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_W)$ such that $\mathbf{\Sigma}(t_n) = \mathbf{U}\Lambda\mathbf{V}^T$ where $\mathbf{u}_n$ and $\mathbf{v}_n$ denote nth Empirical Orthogonal Function (EOF) as a sequence of elements of the singular-vector corresponding to nth singular value for $n \in \{1, 2, ..., W\}$.

Let singular-values be in decreasing order of magnitude $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_W \geq 0$. The collection of eigentriples can then be formed as $(\mathbf{u}_n, \lambda_n, \mathbf{v}_n)$, $n \in \{1, 2, ..., W\}$. The SVD of the visibility matrix instance can be written as a sum of rank-one bi-orthogonal elementary matrices X_w, $w \leq W$ as $\mathbf{\Sigma}(t_n) = \mathbf{X}_1 + ... + \mathbf{X}_W$.

2.2. Clustering via Graph Laplacian Matrix

Spectral clustering has recently emerged in machine learning, pattern recognition and computer vision as a promising modern clustering algorithm [1]. The methodology is essentially forming a distance matrix from a predetermined similarity measure between components. Then a matrix of leading eigenvectors is derived from the distance matrix, which is used by the k-means algorithm to cluster the components.

The algorithmic steps can be summarised as follows [17, 12, 1].

2nd step: Forming Graph Laplacian Matrix

Form an affinity matrix $\mathbf{X}^A \in \mathbb{R}^{W \times W}$ where each (i,j)-element is described by $\mathbf{X}^A_{i,j} = \frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{\sqrt{\|\mathbf{X}_i\|\|\mathbf{X}_j\|}}$.

A graph Laplacian matrix $\mathbf{L}$ can be formed as $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{X}^A\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{D}$ is a diagonal matrix defined by $\mathbf{D}_{j,j} = \sum_i \mathbf{X}^A_{i,j}$.

3rd step: Eigendecomposition

Define $\mathbf{Y} \in \mathbb{R}^{W \times k}, d \in \{1, 2, ..., W-1\}$ to be a matrix governed by k leading eigenvectors of $\mathbf{L}$ in its columns, and re-normalise its rows to have unit length.

4th step: Cluster via K-means

Given that each row of $\mathbf{Y}$ represents a point in $\mathbb{R}^k$, seperate the components $\mathbf{X}_w$ for $w = \{1, 2, ..., W\}$ into k clusters using k-means clustering, which is followed by assigning the component $\mathbf{X}_w$ to a distinct cluster that w'th row of $\mathbf{Y}$ belongs to.

The AWGN assumption implies that noise components have arbitrary structures and hence the respective vertices are very often far from each other. Instead, actual cluster of our interest has multiple components within. We therefore detect the outliers, i.e., noise components, to subtract from the noisy visibility instance.

Let $\mathcal{I}_{\text{noise}}$ denote the set of indices where noise components like in. The denoised visibility matrix $\hat{\mathbf{\Sigma}}$ can then be computed as $\hat{\mathbf{\Sigma}}(t_n) = \mathbf{\Sigma}(t_n) - \sum_{i \in \mathcal{I}_{\text{noise}}} \{\mathbf{X}_w\}_i$.

The above steps are repeated and the denoised visibility matrix is estimated by $\hat{\mathbf{\Sigma}} = 1/N \sum_{n=1}^{N} \hat{\mathbf{\Sigma}}(t_n)$, and finally, the denoised image is inferred using $\hat{\mathbf{\Sigma}}$.

3. Experimental Results

We now show how effectively the scheme performs in noise reduction through simulation on LOw Frequency ARray by illustrating how much the least-squares images has been cleared up. Fig.1 shows how remarkably close the denoised image is to the true image, the one without the presence of noise. The residual obtained after removing the true least-squares is almost clear, showing it captures it really well.

4. Conclusions

We showed the possibility of significant noise reduction in radio interferometers where the algorithm was shown to perform beyond what we even anticipated.

We inferred that a detailed analysis on the singular spectrum permits us to successfully eliminate the vast majority of noise from the observations followed by the spectral clustering. A more sophisticated analysis based on SSA helped us to build a rigorous framework to construct our learning algorithm.

An application to modern radio interferometers validated our claim and showed that the heuristic arguments are in agreement with simulation. Simulation results bear out that our approach offers good accuracy even with far less sampling time, potentially saving time and energy. A fundamental contribution was to show that even using less data, sky estimate can still be recovered very accurately.

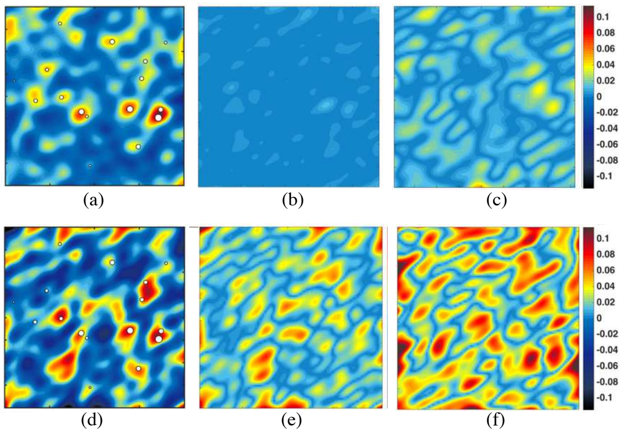

Figure 1. The white circles denote sources. Denoised least-square estimates of the sky are provided over 5ms and 0.5ms of observation, respectively in (a) and (d). The residuals after subtracting the true estimates are given in (b) and (e) in the presence and absence denoising, respectively for 5ms and 0.5ms of measurement duration. We observe that identification of true sources is far easier and artefacts are significantly reduced when denoising is used. When fewer time samples are used (10 times smaller in the case of Fig.1(d)), the performance remains good, illustrated by the residual Fig.1(e) which is vastly superior to the residual in the absence of noise in Fig.1(f). This suggests we can drastically reduce the observation time (and thus the amount of data), and still obtain good estimates. Such a result has potentially profound consequences for the power consumption, engineering, use cases and phased-array design.

References

[1] Y. W. A. Y. Ng, M. I. Jordan. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press, 2001.

[2] C. T. et al. Applying full polarization a-projection to very wide field of view instruments: An imager for lofar. *Astronomy Astrophysics*, 553(A105), may 2013.

[3] T. B. et al. Sparce sampling of signal innovations: Theory, algorithms and performance bounds. *IEEE Signal Processing Magazine, Special issue on Compressive Sampling*, 25(2):31–40, 2008.

[4] J. Gillard. Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Statistics and its Interface*, 3(3):335–343, 2010.

[5] N. Golyandina and A. Zhigljavsky. *Singular Spectrum Analysis for time series*. Springer Science & Business Media, 2013.

[6] N. M. Gürel. Denoising phased-array antenna signals in the presence of extreme noise. In *MSc Thesis, EPFL*, 2016.

[7] N. M. Gürel, P. Hurley, and M. Simeoni. Denoising radio interferometric images by subspace clustering. In *IEEE International Conference on Image Processing*, pages 2134–2138, 2017.

[8] N. M. Gürel, P. Hurley, and M. Simeoni. Method and systems to reduce noise in phased-array signals from receivers located at different locations. In *Patent numbers: CH9-2016-0057*, 2017.

[9] N. M. Gürel, P. Hurley, and M. Simeoni. On denoising crosstalk in radio interferometry. In *International BASP Frontiers Workshop*, pages 2134–2138, 2017.

[10] J. A. Hogbom. *Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines*, volume 15 of *0005-1098*. Pergamon, 1974.

[11] B. D. Jeffs. Signal processing for phased array feeds in radio astronomical telescopes. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):635–646, 2008.

[12] J. S. M. Meila. Learning segmentation by random walks. In *Advances in Neural Information Processing Systems*, pages 873–879. MIT Press, 2001.

[13] U. R. S. Bhatnagar and K. Golap. The wb a-projection and hybrid algorithms. *The Astrophysical Journal*, 770(2), 2013.

[14] J. B. S. J. Wijnholds. Calibratability by design for ska's low frequency aperture array. In *General Assembly and Scientific Symposium*. IEEE, 2014.

[15] M. Viberg. *Subspace-based Methods for the Identification of Linear Time-invariant Systems*, volume 31 of *0005-1098*. Pergamon, 1995.

[16] I. M. Vidal. Nonlinear least squares. http://www3.mpifr-bonn.mpg.de, April 2011.

[17] Y. Weiss. Segmentation using eigenvectors: A unifying view. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*. IEEE, 1999.

[18] H. I. Y. Kayano. A study on characteristics of em radiation from stripline structure. *Radio Science*, 46(5), oct 2011.