DENOISING RADIO INTERFEROMETRIC IMAGES BY SUBSPACE CLUSTERING

Nezihe Merve Gürel[⋆], *Paul Hurley*[⋆], *Matthieu Simeoni*[⋆†]

[⋆]IBM Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland
[†]Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{gur, 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. Furthermore, only the end image is cleaned, reducing flexibility substantially. We propose instead a method to remove the noise explicitly before imaging. To this end, we developed an algorithm that first decomposes the sensor signals into 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. From telescopic data to least-squares image estimates, far higher accuracy with low computation cost results without the need for long observation time.

Index Terms— radio interferometry, singular spectrum analysis, graph signal processing, Laplacian matrix, LOFAR

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 [1, 2, 3].

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), and beamforming (to favour certain regions over others). In addition, when data is heavily corrupted, it is thrown away [4], wasting resources.

This work proposes to explicitly denoise the antenna signals directly in real-time using a subspace algorithm. The ultimate goal is to obtain accurate output from the end of the processing chain, while also facilitating additional use cases by increasing reliability at each stage. This flexibility means 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. This appears to be the first proposal for such an explicit step, and it is certainly the first at antenna level.

Noise structurally different from the true[1] signal can usually be separated from observations 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 antenna measurements at a *specific-time instance*. The measurements represent the same snapshots taken from different perspectives, e.g. antenna positions. Therefore, given these distinct structures of the true signal and noise, subspace methods seem promising for distinguishing these two components.

To the best of our knowledge, denoising via subspace methods has been studied only when noise is far lower than the true signal. For example, Cadzow's method successfully performs noise reduction for high [6] as well as for moderate SNR signals [7] but works poorly in a low SNR regime. Singular Spectrum Analysis (SSA) is likewise powerful in denoising high SNR signals [8]. In [9], 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. Structured component extraction techniques in the framework of SSA, on the other hand, heavily relies on frequency-domain analysis [10] and performs well for high SNR signals. In [11], the author proposes to denoise directly the sky images by using subspace techniques, assuming a simple noise model corrupting the image.

The outline of the rest of the paper is as follows. We start by introducing the data model and subspace-based denoising in Section 2. Section 3 then describes the proposed method containing the SSA and clustering via spectral methods. There we propose a method to remove the noise from the radio astronomical observations. Section 4 shows denoising performance through experiment. We conclude in Section 5.

Notation: Scalars are denoted by italics, vectors by bold lower-case and matrices by bold upper-case, respectively. Let also $\circ$ denote the point-wise product between the elements of two vectors (matrices). We finally denote the number of antennas by L, number of sky sources by Q and the set of antennas and sources by $\mathcal{L} = \{1, 2, \dots, L\}$ and $\mathcal{Q} = \{1, 2, \dots, Q\}$, respectively.

2. PRELIMINARIES

2.1. Data Model

Radio interferometers measure electromagnetic radiation from space which are used to deduce a sky image. Hierarchically, antennas at various locations on the ground record radio waves coming from sources in space, and then send those to respective stations. Those geographically close to one another are grouped in stations and beamformed together. Interferometers first estimate cross-correlation between the time series measurements that is called visibilities, and sky image is estimated by a specified imaging technique.

Based on the assumptions introduced by [12], let us denote the series emitted by the source coming from direction $\mathbf{r}_q \in \mathbb{S}^2$ at time t by $\hat{s}(t, \mathbf{r}_q)$, and the intensity by $I_q \in \mathbb{R}$. The baseband representa-

[1]In time series analysis, true refers to the absence of noise

tion of $\hat{s}(t, \mathbf{r}_q)$ at a center frequency f_0 is then:

$$s_q(t, \mathbf{r}_q) = \hat{s}(t, \mathbf{r}_q)e^{j2\pi f_0 t} \tag{1}$$

where $s : \mathbb{R} \times \mathbb{S}^2 \to \mathbb{C}$ and $\hat{s}(t, \mathbf{r}_q) \sim \mathbb{C}\mathcal{N}(0, \mathrm{I}_q)$.

Consider now the vector of source signals $\mathbf{s}(t) = [s(t, \mathbf{r}_q)]_{q=\mathcal{Q}}$ and also the vector of antenna signals governed by antennas denoted by $\mathbf{x}(t) = [x_l(t)]_{l=\mathcal{L}}$. We then have the following matrix product [13]:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t, \mathbf{r}_q) + \mathbf{n}(t) \tag{2}$$

where $\mathbf{n}(t) \in \mathbb{C}^L$ is the additive white Gaussian noise (AWGN) at the antennas and independently drawn from the distribution $\mathbb{C}\mathcal{N}(0, \sigma_n^2 \mathbf{I}_L)$. Let $\mathbf{A} \in \mathbb{C}^{L \times Q}$ be the antenna steering matrix with each column given by individual antenna steering vector

$$\mathbf{a_q} = [e^{-j2\pi\langle \mathbf{r_q}, \mathbf{p_l}\rangle}]_{l=\mathcal{L}}.$$

Consider now M stations, each composed of L antennas with respective beamforming matrix $\mathbf{W} \in \mathbb{C}^{M \times L}$ with individual beamforming vectors $\mathbf{w}_m$ in each column[2]. The beamformed output vector is thus:

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t).$$

Current radio astronomical imaging techniques takes cross-covariance matrix so-called visibilities as input to infer a sky estimate. The input to imager is given by $\Sigma = \mathbb{E}[\mathbf{y}(t)\mathbf{y}(t)^\dagger]$ where $\dagger$ stands for the Hermitian transpose. The various techniques for mapping from the visibilities to the sky image are throughly studied in the literature, ranging from CLEAN [14] to A(W)-projection [15, 16].

2.2. Subspace-based Denoising

Subspace-based methods embed the observed data into higher dimensional vector space and split the entire space into disjoint subspaces. Desired features can be extracted within the subspace of interest, provided they are independent. This is achieved by applying a Singular Value Decomposition (SVD) to a trajectory matrix, typically with rows and columns consisting of sub-series of the initial series governed by the observations, i.e. Hankel matrix. There are alternative structures for the trajectory matrix such as Toeplitz, block Hankel, block Toeplitz, Covariance and circulant [17, 18]. The choice of Hankel structure is however favoured due to the existence of inverse embedding operator. Followed by a proper split of the singular-triples, a low-rank approximation is then used to reconstruct the respective trajectories. Finally, the trajectories are mapped back from higher dimensional space to the initial one (inverse embedding).

Let $\mathbf{x} \in \mathbb{C}^L$ be a vector of corrupted samples. The usual methodology for learning high SNR signals can be summarized as follows.

1. Form a trajectory matrix $\mathbf{X} \in \mathbb{C}^{W \times V}$ from $\mathbf{x}$, i.e. $f(\mathbf{x}) = \mathbf{X}$, where $f : \mathbb{C}^L \to \mathbb{C}^{W \times V}$, W and V are pre-set positive valued integers determined by the choice of trajectory type so-called window size.
2. Apply SVD to $\mathbf{X}$.
3. Perform a low-rank approximation from the leading singular values of $\mathbf{X}$ and call it $\hat{\mathbf{X}} \in \mathbb{C}^{\mathrm{W} \times \mathrm{V}}$.
4. Map from the trajectory to the original vector space back, i.e., $f^{-1}(\hat{\mathbf{X}}) = \hat{\mathbf{x}}$.

[2]We refer to [13] for a detailed information about the beamforming techniques and the choice of $\mathbf{w}_m$.

Remark: The above subspace-based learning algorithm works well under two main constraints. First, true signal and noise lie in disjoint subspaces, that is, they majorly have distinct tendencies in the vector space so as to split the entire vector space into two disjoint subspaces. Secondly, the observed signal is high SNR, and hence only low valued singular values are associated to the noise. Low SNR thus raises many problems in identifying the underlying vector space as the noise subspace is no longer governed by singular vectors with the smallest eigenvalues. Instead the singular values of the noise subspace are typically spread over the full range in the SVD. Basic SSA as well as other subspace-based techniques are thus not sufficient in such a scenario as they can distinguish subspaces by analysing only the spectrum of singular values [8]. It is then essential to employ more sophisticated algorithms by incorporating the singular vectors into the subspace analysis.

3. PROPOSED METHOD

In this section, we propose a method to denoise phased-arrays by exploiting the known structure of the antenna signals even under significant noise corruption. The approach decomposes the low SNR signals into additive components. It then learns, by spectral clustering, the representative components for the noise and true signal.

3.1. Problem Statement

Consider now the vector of antenna measurements at a *specific-time instance* t_n governed by L antennas on the ground and call it the antenna spatial series. Opening up the antenna steering matrix $\mathbf{A}$, Eqns.1 and 2 yield the following

$$\mathbf{x}(t_n) = e^{j2\pi f_0 t_n} \sum_q \hat{s}(t_n, \mathbf{r}_q)[e^{-j2\pi\langle \mathbf{r}_q, \mathbf{p}_l\rangle}]_{l=\mathcal{L}} + \mathbf{n}(t_n) \tag{3}$$

where $\mathbf{n}(t_n)$ is assumed to be AWGN and summation is over all the sources within the field of view.

The SNR between the sum of source signals and the noise is usually below 5 dB in real-life applications [20, 21]. An analysis on the spectrum of singular values only is thus not sufficient to remove noise from the observations due to the low SNR. AWGN moreover implies that the vectors spanning noise subspace have random fluctuations only, instead of meaningful characteristics, i.e., trend, harmonics, or sum of such meaningful components. However, the true signal has a structured behaviour that can be explained as follows.

Let $\mathbf{r} \in \mathbb{S}^2$ be a reference position in sky and the relative position of the source q, denoted by $\Delta\mathbf{r}_q$, be given by $\mathbf{r}_q - \mathbf{r}$. The signal coming from the reference direction to the antennas are given by

$$\mathbf{s}_r(t) = e^{j2\pi f_0 t}[e^{-j2\pi\langle \mathbf{r}, \mathbf{p}_l\rangle}]_{l=\mathcal{L}}.$$

Therefore, Eqn.3 can be re-written as

$$\mathbf{x}(t_n) = \mathbf{s}_r(t_n) \circ \sum_q e^{j2\pi f_0 t_n}\hat{s}(t_n, \mathbf{r}_q)[e^{-j2\pi\langle \Delta\mathbf{r}_q, \mathbf{p}_l\rangle}]_{l=\mathcal{L}} + \mathbf{n}(t_n).$$

A closer look at the above equation reveals that true signal can mathematically be expressed as a weighted and rotated sum of the reference signal $\mathbf{s}_r(t)$. Thus, the components decomposed from true spatial series will be similar vectors with a phase delay and a fixed rotation in between. In addition, sources that are close to each other induce the antenna spatial series in similar way hence they are very likely to fall into the same component. Each individual component associated with the true series is thus induced by one or more distinct sources.

Given that the entire vector space is resolved sufficiently well, it only remains to determine the set of components associated with the true spatial series by exploiting the similarity feature between the true components that noise components do not have. Our method thus consists of two stages. First, we employ the subspace algorithm SSA to decompose the observed signal into its components. We then learn the representative of the components to extract true series from the corrupted measurements by employing spectral clustering.

3.2. 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 [9]. Opening up its role regarding noise reduction, SSA reconstructs noise from the residual and uninterpretable feature space under high SNR constraint. We will however follow the theory of SSA only to achieve a proper decomposition of the observed signal.

Suppose now that $\mathbf{x}(t_n)$ is decomposed into its additive components such that

$$\mathbf{x}(t_n) = \mathbf{x}_1(t_n) + \mathbf{x}_2(t_n) + \ldots + \mathbf{x}_r(t_n) \quad (4)$$

where $\mathbf{x}_w(t_n) \in \mathbb{C}^L$ for $w = \{1, 2, \ldots, r\}$ can be assigned to disjoint clusters, i.e. noise and true signal.
The methodology for decomposition part can be given as follows.[3]
1st Step: Embedding
A projection matrix is formed from the sub-series of an observed series $\mathbf{x} \in \mathbb{C}^{\mathrm{L}}$ in rows and columns, i.e., Hankel matrix. Typically, the observed series $\mathbf{x}$ is mapped into a sequence of lagged vectors of size W, i.e., window length, by forming $V = L - W + 1$ lagged vectors

$$\mathbf{x}_i = (x_i,\ x_{i+1},\ ...,\ x_{i+W-1})^T,\ \ i = 1, ..., V.$$

The trajectory matrix of $\mathbf{x}$ is then given by

$$\mathbf{X} = [\mathbf{x}_1 : ... : \mathbf{x}_V] = \begin{pmatrix} x_1 & x_2 & \cdots & x_V \\ x_2 & x_3 & \cdots & x_{V+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_W & x_{W+1} & \cdots & x_L. \end{pmatrix}.$$

The window length W is the most critical parameter in the theory of SSA [18]. It depends on the properties of the observed series as well as the purpose of the analysis. Without loss of generality, the window length should be as large as possible and proportional to the period of the harmonics if any. In [17], the authors argue that in such mappings, i.e., Hankelisation, W should not be larger than half of the length of the observed series. Hence, so as to maximise spectral resolution, we set W to $L/2$.
2nd Step: Decomposition
At this step, the singular spectrum of the trajectory matrix $\mathbf{X}$ is analysed by performing an SVD. The SVD of $\mathbf{X}$ is given by the product of three matrices $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_W)$, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_W)$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_W)$ such that $\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}^T$ 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 $n \in \{1, 2, ..., W\}$.

[3] We drop the time indicator t_n for the sake of simplicity.

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 trajectory matrix can be written as a sum of rank-one bi-orthogonal elementary matrices $X_w, w \leq W$ as follows:

$$\mathbf{X} = \mathbf{X}_1 + ... + \mathbf{X}_W.$$

3rd step: Reconstruction
Averaging over the anti-diagonals (inverse of the *Hankelisation*) transfers the low-rank components into an additive series. Hence, we obtain the following expansion:

$$\mathbf{x} = \mathbf{x}_1 + ... + \mathbf{x}_W.$$

Another remark is that the disjointness property of the subspaces can be shifted to the vector space by reconstructing the components from a single rank in the trajectory space. Although basic SSA is usually concerned with the singular vectors of the trajectory, it is more convenient for our algorithm to analyse the vector components followed by the SVD decomposition in higher dimensional space.

3.3. Clustering via Graph Laplacian Matrix

Spectral clustering has recently emerged in machine learning, pattern recognition and computer vision as a promising modern clustering algorithm [22]. Combining the k-means and graph Laplacian, spectral clustering methods outperform algorithms [23]. 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 methodological principle behind spectral methods is to cluster data points by forming an undirected graph treating each data point (component) $\mathbf{x}_w$ as a vertex, edges are weighted according to the similarity between vertices. The goal can thus be seen as finding a partition of the graph such that the set of vertices with shorter edges among each other represent a cluster.
The algorithm can be summarised as follows [24, 25, 22].
4th step: Forming Graph Laplacian Matrix
Form an affinity matrix $\mathbf{X}^A \in \mathbb{R}^{L\times L}$ where each (i,j)-element is described by

$$\mathbf{X}^A_{i,j} = \frac{\|x_i - x_j\|}{\sqrt{\|x_i\|\|x_j\|}}.$$

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 $\mathbf{D}_{j,j} = \sum_i \mathbf{X}^A_{i,j}$.

There is no unique way to define the Graph Laplacian and the similarity measure that describes the affinity matrix. Our method uses the normalised graph Laplacian and correlation coefficient as a similarity measure.
5th step: Eigendecomposition
Define $\mathbf{Y} \in \mathbb{R}^{L\times k}, d \in \{1, 2, ..., W-1\}$ to be a matrix governed by k leading eigenvectors of $\mathbf{L}$ in its columns, and re-normalise its rows to have unit length.
6th 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 wth row of $\mathbf{Y}$ belongs to.

The additive white Gaussian noise assumption on the noise distribution implies that noise components are random vectors and hence the respective vertices are very often far from each other. We then look for a cluster that has multiple components within, and moreover a fixed phase shift and rotation exists in between. The true signal can be reconstructed by summing these components. By repeating the above scheme for all data vectors, antenna signals can be estimated over both time and space.

4. EXPERIMENTAL RESULTS

We now show how effectively the scheme performs in noise reduction. A MATLAB simulation of the LOFAR radio astronomy interferometer was performed. Antennas geographically close to one another were grouped in stations, and their time-series were sampled and beamformed at station level. Individual beams were then correlated, and a least-squares estimate of the underlying sky (or dirty image in the nomenclature of radio astronomy) obtained [8, 13].

Antenna spatial series is formed at each station separately, and the algorithm was executed for each series over time.

4.1. Denoising the Antenna Series

Fig.1 shows an example of the true spatial series estimate at one time instance. The residual components outside of the true cluster are associated to the noise. The absolute resulting estimate validates that the method successfully captures the tendency of the true series [19]. Denoised time series by a single antenna, depicted in Fig.2, can then be reconstructed by using denoised spatial-series at consecutive time-instances.

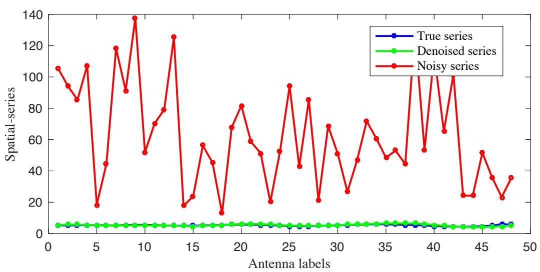

Fig. 1: True spatial series vs. denoised spatial series.

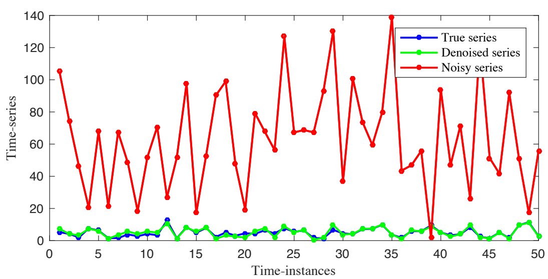

Fig. 2: True time series vs. denoised time series.

4.2. Improving Least Square Image Estimates

A key measure of effectiveness is how much the least-squares image estimate has been cleared up. Fig.3 shows how remarkably close the denoised image is to the true image, the one without the presence of noise. For this experiment, 12 core LOFAR stations were used.

When data is collected over 5ms, the least-squares estimate after denoising is as shown in Fig.3(a). The residual -difference between it and the true least-squares estimate- is almost clear, showing it captures it really well. In contrast, Fig.3(c), created without denoising in the loop, shows a much bigger residual.

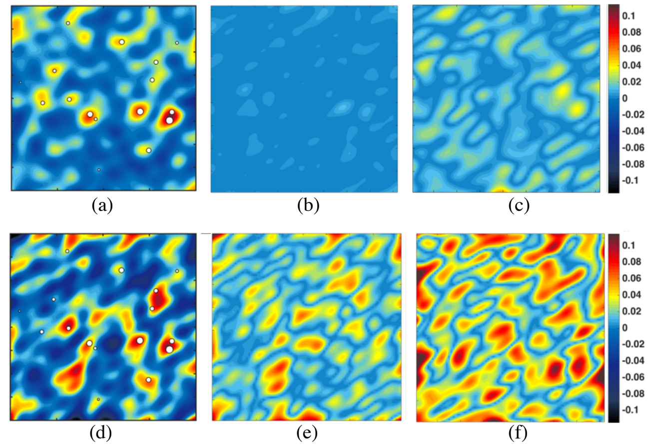

Fig. 3: The white circles denote sources. Denoised least-square estimates of the sky are provided over 5ms and 0.5ms of observation, respectively in (a) and (d). The residuals after subtracting the true estimates are given in (b) and (e) in the presence and absence denoising, respectively for 5ms and 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.3(d)), the performance remains good, illustrated by the residual Fig.3(e) which is vastly superior to the residual in the absence of noise in Fig.3(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. Note no rigid assumptions are imposed on the length of the observed spatial series. In practical applications, the procedure can be applied to series of length varying from a few dozen to thousands. In fact, the method seems promising for even short series length in related applications.

5. CONCLUSIONS

We showed the possibility of significant noise reduction for the radio astronomical data at the earlier stages of processing pipeline. 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. Error reduction at an early stage leads not only to very accurate sky estimates. Additional use-cases open up from the new information: accurate phased arrays and an accurate least-squares estimate. 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.

A rigorous bound on the complexity needs be derived. Although SVD is the most expensive operation of the procedure, computation of principal components is quite fast in modern computers. The method thus yields accurate results at a reasonable computational cost.

6. REFERENCES

[1] B. D. Jeffs et al., "Signal processing for phased array feeds in radio astronomical telescopes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 635–646, 2008.

[2] J. Bregman S. J. Wijnholds, "Calibratability by design for ska's low frequency aperture array," in *General Assembly and Scientific Symposium*. 2014, IEEE.

[3] H. Inoue Y. Kayano, "A study on characteristics of em radiation from stripline structure," *Radio Science*, vol. 46, no. 5, oct 2011.

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

[5] G. J. A. Harker et al., "Detection and extraction of signals from the epoch of reionization using higher order one-point statistics," *Monthly Notices of the Royal Astronomical Society*, vol. 393, no. 4, pp. 1449–1458, feb 2009.

[6] T. Blu et al., "Sparce sampling of signal innovations: Theory, algorithms and performance bounds," *IEEE Signal Processing Magazine, Special issue on Compressive Sampling*, vol. 25, no. 2, pp. 31–40, 2008.

[7] J. Gillard, "Cadzow's basic algorithm, alternating projections and singular spectrum analysis," *Statistics and its Interface*, vol. 3, no. 3, pp. 335–343, 2010.

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

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

[10] T. Alexandrov, "A method of trend extraction using singular spectrum analysis," *Statistical Journal*, vol. 7, no. 1, pp. 1–22, apr 2009.

[11] S. Yatawatta, "Subspace techniques for radio-astronomical data enhancement," *Astrophysics*, sep 2008.

[12] R. A. Perley, "Synthesis imaging in radio astronomy ii," *ASP Conf. Series,*, vol. 180, 1999.

[13] M. Simeoni, "Towards more accurate and efficient beamformed radio interferometry imaging," M.S. thesis, Ecole Polytechnique Federale de Lausanne, 2015.

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

[15] U. Rau S. Bhatnagar and K. Golap, "The wb a-projection and hybrid algorithms," *The Astrophysical Journal*, vol. 770, no. 2, 2013.

[16] C. Tasse et al., "Applying full polarization a-projection to very wide field of view instruments: An imager for lofar," *Astronomy and Astrophysics*, vol. 553, no. A105, may 2013.

[17] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *Journal of Data Science*, vol. 5, pp. 239–257, 2007.

[18] A. Shlemov N. Golyandina, "Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series," *Statistics and its interface*, vol. 8, no. 3, pp. 277–294, 2015.

[19] N. M. Gürel, "Denoising phased-array antenna signals in the presence of extreme noise," M.S. thesis, Ecole Polytechnique Federale de Lausanne, 2016.

[20] S. W. Ellingson, "Sensitivity of antenna arrays for long-wavelength radio astronomy," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 6, pp. 1855–1863, mar 2011.

[21] M.P. van Haarlem et al., "Lofar: The low-frequency array," *Instrumentation and Methods for Astrophysics*, vol. 556, no. A2, aug 2013.

[22] Y. Weiss A. Y. Ng, M. I. Jordan, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*. 2001, pp. 849–856, MIT Press.

[23] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, 2007.

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

[25] J. Shi M. Meila, "Learning segmentation by random walks," in *Advances in Neural Information Processing Systems*. 2001, pp. 873–879, MIT Press.