

# Constrained Imaging for Radio Astronomy

Ahmad Mouri Sardarabadi  
Faculty Electrical Eng.,  
Mathematics and Computer Science  
Delft University of Technology  
Mekelweg 4  
2628 CD Delft  
The Netherlands

Alle-Jan van der Veen  
Faculty Electrical Eng.,  
Mathematics and Computer Science  
Delft University of Technology  
Mekelweg 4  
2628 CD Delft  
The Netherlands

**Abstract**—We show that the imaging problem for radio astronomy is not only bounded below (the image is non-negative), but there is also an upper bound. We show that the tightest upper bound is the MVDR dirty image. We propose using active-set methods to solve the imaging problem and show that this algorithm is strongly related to sequential source removal techniques like the CLEAN. Also recent studies show the relation between non-negativity and sparsity which means that the proposed optimization also benefits from sparsity in an automatic way.

**Index Terms**—radio astronomy, imaging, deconvolution, active-set, constrained optimization, non-negativity

## I. INTRODUCTION

Many image deconvolution algorithms for radio astronomy, such as CLEAN [1], could be classified as sequential source removal techniques. However as the complexity of the telescopes is increasing, parametric methods and sparse reconstruction techniques are becoming more popular [2], [3]. As we will show here, with the correct formulation, finding the solution to a parametric method could lead to an active-set algorithm that is strongly related to sequential source removing techniques and hence benefit from both approaches.

One of the major complications with imaging for radio astronomy is that the deconvolution problem becomes ill-posed as the pixels become smaller [4]. To cope with this issue any a-priori knowledge that might be available should be used. The fact that the pixel values are non-negative could be used to improve the image estimates. An example of this approach is the non-negative least squares (NNLS) [5]. New studies show that some of the classical techniques are promoting sparsity. For example [6] shows the relation between the non-negativity constraint and sparsity. This makes it worthwhile to revisit these techniques.

In this paper we show that the pixel values are bounded also from above and then we will extend the idea behind NNLS and formulate the multichannel imaging as an optimization problem with both lower and upper bounds. Here we will use the active-set method because we can show a relation between this approach and sequential source removal techniques like the CLEAN. We will also demonstrate that the minimum variance distortionless response (MVDR) dirty image as defined by [7] forms the tightest upper bound on the imaging problem.

The automatic promotion of sparsity through non-negativity, the relation between the active-set methods and sequential source removal techniques combined with the MVDR dirty image and the flexibility in the choice of the cost function makes this approach a good candidate for practical use.

In Sec. II we will describe the data model for a multichannel telescope and in Sec. III we state the imaging problem and introduce the bounds on the image values. Sec. IV gives a brief description of the active-set method and describes the relation between the CLEAN algorithm and the constrained least squares (CLS), finally we will show the simulation results in the Sec. V.

## II. DATA MODEL AND PROBLEM DEFINITION

We have a system with  $p$  antennas which are observing the sky. The received signal is split into  $K$  subbands for which the output of the system can be modeled as

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{n}_k$$

where  $\mathbf{A}_k$  is a  $p \times m$  matrix containing the array response vectors of the sources (without loss of generality we assume that columns of  $\mathbf{A}_k$  are normalized),  $\mathbf{x}_k$  is a  $m \times 1$  vector representing the signals from the sky and  $\mathbf{n}_k$  is a  $p \times 1$  vector modeling the noise. We assume that the noise and the sky sources are independent such that the system can be modeled using covariance matrices

$$\mathbf{R}_k = \mathcal{E}\{\mathbf{y}_k \mathbf{y}_k^H\} = \mathbf{A}_k \mathbf{R}_x \mathbf{A}_k^H + \mathbf{R}_{n,k}$$

where  $^H$  is the Hermitian transpose and  $\mathbf{R}_x$  and  $\mathbf{R}_{n,k}$  are covariance of sky and noise signals, respectively. By assuming that sky sources are independent, we can model  $\mathbf{R}_x = \text{diag}(\boldsymbol{\sigma})$  where  $\sigma_i$  represents the power of the  $i$ th source and we assume that it is white over all subbands (If the source positions were known, the imaging problem would be equivalent as finding and estimate for  $\boldsymbol{\sigma}$ ). Vectorizing both sides we have

$$\mathbf{r}_k = \text{vect}(\mathbf{R}_k) = (\mathbf{A}_k^* \circ \mathbf{A}_k) \boldsymbol{\sigma} + \mathbf{r}_{n,k}$$

where  $^*$  is the complex conjugate,  $\circ$  is the Khatri-Rao product and  $\mathbf{r}_{n,k} = \text{vect}(\mathbf{R}_{n,k})$ . The multichannel model for the whole

system can now be given as

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_K \end{bmatrix} = \mathbf{A}\boldsymbol{\sigma} + \begin{bmatrix} \mathbf{r}_{n,1} \\ \vdots \\ \mathbf{r}_{n,K} \end{bmatrix}. \quad (1)$$

where

$$\mathbf{A} = \begin{bmatrix} (\mathbf{A}_1^* \circ \mathbf{A}_1) \\ \vdots \\ (\mathbf{A}_K^* \circ \mathbf{A}_K) \end{bmatrix}.$$

The array response matrix may depend on more parameters than just the direction of arrival, we will assume that the array is calibrated for those parameters. Using the model above, we will define the imaging problem in the next section.

### III. IMAGING PROBLEM

We can formulate the imaging problem as following:

Given  $N$  samples of  $\mathbf{y}$  for each of the  $K$  subbands, such that we have

$$\hat{\mathbf{r}}_k = \text{vect}(\hat{\mathbf{R}}_k) = \text{vect}\left(\frac{1}{N} \sum_{n=0}^{N-1} \mathbf{y}_k[n] \mathbf{y}_k[n]^H\right)$$

and an array response  $\tilde{\mathbf{A}}$  for the pixels in the image, we want to estimate the power of each pixel, which we stack in a vector  $\boldsymbol{\sigma}_p$ . We have introduced  $\tilde{\mathbf{A}}$  and  $\boldsymbol{\sigma}_p$  to emphasize the difference between the true (and unknown) position of the sources and the position of each pixel in the image.

As an example we will start with the Least Squares (LS), this will help us explain some of the ideas, however we are not limited to any particular cost function. Finding the best approximation for LS cost function means minimizing  $\|\hat{\mathbf{r}} - \mathbf{A}\boldsymbol{\sigma}_p\|_2^2$ . It is straightforward to show that the solution for this problem is given by any  $\boldsymbol{\sigma}_p$  that satisfies

$$\mathbf{H}\boldsymbol{\sigma}_p = \mathbf{s}. \quad (2)$$

where we define  $\mathbf{H} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})$  as the convolution matrix and  $\mathbf{s} = \tilde{\mathbf{A}}^H \hat{\mathbf{r}}$  as the (matched filter) dirty image. In a majority of situations (2) is ill-posed and has an infinite number of solutions which makes some kind of regularization unavoidable.

We will now formulate our regularization as a set of constraints on the image. A first constraint comes from the fact that we are estimating the variance (power density) on each pixel which means that  $\boldsymbol{\sigma}_p \geq \mathbf{0}$ . This is the non-negativity constraint that has been studied for example in [5]. However by closer inspection of the  $i$ th pixel on the dirty image we will see that for the  $k$ th subband we have

$$s_{i,k} = \tilde{\mathbf{a}}_{i,k}^H \hat{\mathbf{R}}_k \tilde{\mathbf{a}}_{i,k} = \sigma_{p,i} + \tilde{\mathbf{a}}_{i,k}^H \hat{\mathbf{R}}_{r,k} \tilde{\mathbf{a}}_{i,k} \quad (3)$$

where  $\tilde{\mathbf{a}}_{i,k}$  is the  $i$ th column of  $\tilde{\mathbf{A}}_k$  and  $\hat{\mathbf{R}}_{r,k}$  is the contribution of all other pixels and the noise and is per definition positive-semidefinite. In fact  $\tilde{\mathbf{a}}_{i,k}^H \hat{\mathbf{R}}_{r,k} \tilde{\mathbf{a}}_{i,k}$  could be used as a slack variable. This means that  $\sigma_{p,i} \leq s_{i,k}$  or  $\boldsymbol{\sigma}_p \leq \mathbf{s}$ . However there are some cases where some pixels in the dirty image are not positive, this is the result of a common practice in radio

astronomy that uses the calibration information, and subtracts an estimate of the noise covariance matrix  $\hat{\mathbf{R}}_n$  from the data and uses  $\hat{\mathbf{R}}_0 = \hat{\mathbf{R}} - \hat{\mathbf{R}}_n$  to form the dirty image. To avoid this scenario one should make sure that  $\hat{\mathbf{R}}_0$  remains positive-semidefinite when such an operation takes place.

Now we can formulate our imaging problem as

$$\begin{aligned} \hat{\boldsymbol{\sigma}}_p &= \arg \min_{\boldsymbol{\sigma}_p} f(\boldsymbol{\sigma}_p) \\ \text{subject to } &\mathbf{0} \leq \boldsymbol{\sigma}_p \leq \mathbf{s}. \end{aligned} \quad (4)$$

Here we will limit ourselves to LS cost functions for  $f(\boldsymbol{\sigma}_p)$ .

A common set of algorithms that deals with inequality constraints is the active-set method [8, pp.186-189]. We will show that a simple implementation of CLEAN is closely related to solving (4) with this class of algorithms.

#### MVDR Dirty Image

In this section we will discuss the MVDR dirty image and how it could be used to improve the active-set methods. It has been argued that using the MVDR dirty image will improve the performance of the CLEAN algorithm [7]. We will show the validity of this argument for any cost function optimized with the active-set method. We derive the MVDR dirty image as the answer to the following question: What is the tightest upper bound we could set for our optimization problem?

To answer this question we will revisit (3) and write it in a more general way.

$$\begin{aligned} \tilde{s}_{i,k} &= \mathbf{w}_{i,k}^H \hat{\mathbf{R}}_k \mathbf{w}_{i,k} \\ &= \mathbf{w}_{i,k}^H \tilde{\mathbf{a}}_{i,k} \sigma_{p,i} \tilde{\mathbf{a}}_{i,k}^H \mathbf{w}_{i,k} + \mathbf{w}_{i,k}^H \hat{\mathbf{R}}_{r,k} \mathbf{w}_{i,k} \\ &= \sigma_{p,i} + \mathbf{w}_{i,k}^H \hat{\mathbf{R}}_{r,k} \mathbf{w}_{i,k}. \end{aligned}$$

To make the third equality hold we have to demand  $\mathbf{w}_{i,k}^H \tilde{\mathbf{a}}_{i,k} = 1$ . For the matched filter dirty image we have chosen  $\mathbf{w}_{i,k} = \tilde{\mathbf{a}}_{i,k}$  without putting any demands on the error term. Now the question is how to choose  $\mathbf{w}$  that the second term in the most right hand expression is minimized. Note that  $\hat{\mathbf{R}}_{r,k}$  is not changed and the term we want to minimize remains non-negative.

A simple linear minimization with equality constrain shows that

$$\mathbf{w} = \frac{1}{\tilde{\mathbf{a}}^H \hat{\mathbf{R}}^{-1} \tilde{\mathbf{a}}} \hat{\mathbf{R}}^{-1} \tilde{\mathbf{a}}$$

is the solution to this problem and the dirty image found in this way is called the MVDR dirty image (the subscripts  $i$  and  $k$  are dropped here). We will recommend using the MVDR dirty image on any cost function we choose to optimize.

### IV. ACTIVE-SET ALGORITHM

Active-set methods is a name given to a range of algorithms that are capable of solving optimization problems with inequality constraints. Without going into much details we will give a brief description of the active-set method by closely following [8].

The main idea is the use of KKT criteria [9] to divide the problem with inequality constraints into smaller steps with equality constraints. The KKT criteria states that at a minimum

of a function  $f(\boldsymbol{\sigma}_p)$  subject to inequality constraint  $\mathbf{G}\boldsymbol{\sigma}_p \leq \mathbf{b}$  the following holds:

$$-\mathbf{g} = \mathbf{G}^T \boldsymbol{\eta} \quad (5)$$

$$\boldsymbol{\eta} \geq \mathbf{0} \quad (6)$$

where in our case

$$\mathbf{G} = \begin{bmatrix} -\mathbf{I} \\ \mathbf{I} \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{0} \\ \mathbf{s} \text{ (or } \tilde{\mathbf{s}}) \end{bmatrix},$$

$\boldsymbol{\eta}$  is a vector of KKT multipliers and  $\mathbf{g}$  is the gradient of  $f(\boldsymbol{\sigma}_p)$ . At each step the pixels are divided into two sets, one set is equal to the boundary and forms the equality constraints (active-set) and the other set is used to minimize the cost function (free-set). At the end of each step it is decided to continue with the current equality set or to update it. If moving along the direction of descend makes a pixel value to violate one of the boundaries, that pixel is set equal to the corresponding boundary and the corresponding pixel is added to the active-set. However when we decide that an answer to the current sub-problem is sufficiently approximated, we estimate KKT multipliers and remove the pixel corresponding to smallest negative multiplier from the active-set and add it to the free-set. The reason for this follows directly from KKT. At the solution of an optimization problem with inequality constraints, all the multipliers must be positive, a negative value means that in that direction the cost function could have been decreased without violating any constraints.

In the following section we will describe the relation between CLEAN and LS.

#### A. LS and CLEAN

In this part we will give the expressions for the gradient and the Hessian of the LS cost function and then show its relation with the CLEAN algorithm using some of the ideas from the active-set algorithm. We write out the vector norm from the LS cost function as an inner product and take the derivatives with respect to the unknown parameters

$$f(\boldsymbol{\sigma}_p) = \frac{1}{2} \mathbf{z}^H \mathbf{z}$$

where  $\mathbf{z} = \hat{\mathbf{r}} - \tilde{\mathbf{A}}\boldsymbol{\sigma}_p$  is the error vector and its Jacobian is

$$\frac{\partial \mathbf{z}}{\partial \boldsymbol{\sigma}_p^T} = -\tilde{\mathbf{A}}$$

and so the gradient of  $f(\boldsymbol{\sigma}_p)$  becomes

$$\mathbf{g} = -\tilde{\mathbf{A}}^H (\hat{\mathbf{r}} - \tilde{\mathbf{A}}\boldsymbol{\sigma}_p) = \mathbf{H}\boldsymbol{\sigma}_p - \mathbf{s}. \quad (7)$$

It is then straightforward to find the Hessian which is equal to  $\mathbf{H}$  as defined in (2). Now that the gradient and the Hessian are found we can use them to find the constrained solution for the LS cost function.

Let the  $(\cdot)_{\mathcal{X}}^{(n)}$  be a vector or a square matrix with rows and/or columns selected from the set  $\mathcal{X}$  at step  $n$ , then if the solution of the least squares problem

$$\mathbf{H}_{\mathcal{F}} \boldsymbol{\sigma}_{p,\mathcal{F}} = \mathbf{g}_{\mathcal{F}} \quad (8)$$

where  $\mathcal{F}$  is the free-set, satisfies all the constraints, we use it as the solution of the sub-problem at step  $n$  and continue to update the free/active-set by estimating the KKT multipliers, otherwise we will use it as the direction of descend. The first order approximation of multipliers follows directly from KKT and is given by

$$\boldsymbol{\eta}_{\mathcal{L}} = \mathbf{g}_{\mathcal{L}} \quad (9)$$

$$\boldsymbol{\eta}_{\mathcal{U}} = -\mathbf{g}_{\mathcal{U}} \quad (10)$$

where  $\mathcal{L}$  is the part of active-set corresponding to the lower bound and  $\mathcal{U}$  to the upper bound. These steps are repeated while  $\boldsymbol{\eta}$  contains negative values or the maximum number of iterations is reached.

Now we describe the relation between CLEAN and CLS. The CLEAN algorithm can be describe as follows:

- 1) start from an empty cleaned image;
- 2) find the brightest source in the sky and add it to the cleaned image;
- 3) update the dirty image by applying the convolution matrix on the cleaned image and subtracting it from the dirty image;
- 4) if a bright source can be detected in the new dirty image go to step 2 otherwise terminate.

Now we will relate these steps to the active-set algorithm as described above.

An empty image is an acceptable starting point because the trivial solution is a feasible point. Because we start from trivial solution  $\mathcal{F}$  and  $\mathcal{U}$  are empty, this means that the KKT multipliers can be estimated as  $\boldsymbol{\eta} = -\mathbf{s}$  and its smallest negative value is the largest pixel on the dirty image. Adding this pixel, with index  $i$ , to the free-set and solving (8) will give us  $\hat{\sigma}_{p,i} = s_i$ . This value does not violate the boundary and is set as the solution for this iteration. However this pixel is immediately added to active-set again. Now that we updated  $\hat{\boldsymbol{\sigma}}_p$  we need to calculate the gradient at the current point before we proceed. Using (7) we see that except for a sign difference this is exactly what is done in step 3. So far the two algorithm have been identical.

It is straightforward to show that the second iteration is also identical, however unless the upper bound is changed, the third iteration will be different. This is because at the end of second iteration, the pixel value no longer has to be equal to the upper bound, and it will still be in the free-set when we start the third iteration. We could update the upper bound for each iteration based on the current best estimates. However this will make the two algorithms identical only to the point where  $\boldsymbol{\eta}_{\mathcal{U}}$  starts to correct for the overestimated values in the upper bound.

We have shown that the the CLEAN algorithm is the repetition of the first iteration of the active-set method with

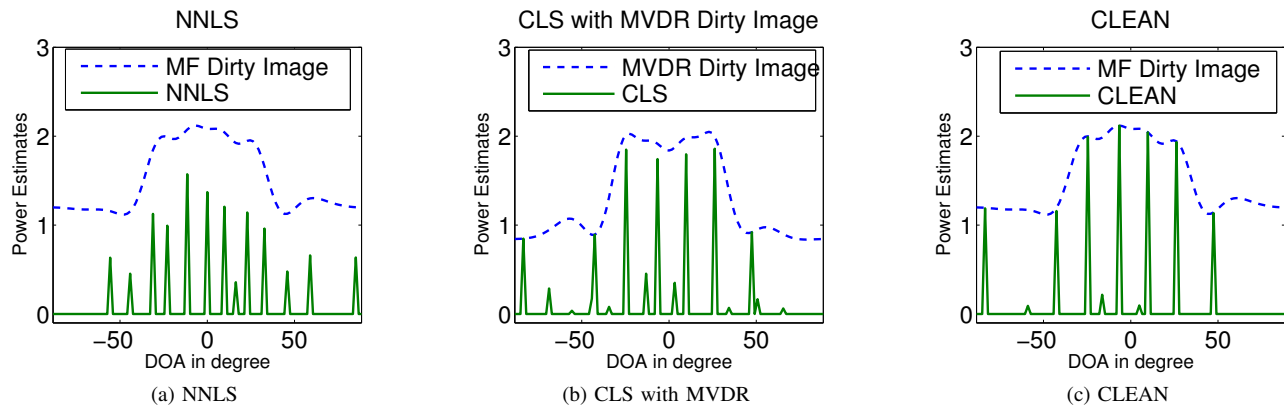


Fig. 1: Simulation Results

LS cost function and that the updated dirty image in step 3 of CLEAN, up to a sign, is the gradient as defined by (7). It is then worth noting that we have found mathematical motivations for each step of the CLEAN algorithm which otherwise are not so clear.

## V. SIMULATIONS

We use simulations to verify some of the claims we have made in the previous sections. For our simulations we will use a uniform linear array (ULA) with  $p = 7$  antennas. We place four sources with equal powers,  $\mathbf{R}_x = \mathbf{I}$ , at exactly one angular resolution distance ( $\approx 16^\circ$ ). Because the sky sources are usually very weak, we add white Gaussian noise to obtain an SNR of  $-10$ dB. We will choose our pixels to be one tenth of the angular resolution of the array ( $\approx 1.6^\circ$ ) which will lead to 109 pixels. Because of the low SNR we use a large number of samples  $N = 10K$ . In order to show the effect of noise and to make sure that we don't introduce negative pixels, we use  $\hat{\mathbf{R}}_0 = \hat{\mathbf{R}} - 0.9\mathbf{R}_n$ .

Figure 1 illustrates the results of the simulations for (a) NNLS, (b) constrained LS (CLS) with MVDR dirty image as upper bound and (c) the result of CLEAN. As expected, adding an upper bound greatly improve the LS estimates by comparing the results of CLS to NNLS. CLEAN and CLS have similar results, but because of the MVDR dirty image, the noise is slightly better suppressed in the case of CLS.

## VI. CONCLUSIONS

We showed that the image deconvolution problem has physical bounds that should be used to improve the estimation process. We have found the tightens upper bound on the image values to be the MVDR dirty image. Also starting from a parametric model we arrived at the active-set algorithm that shows strong similarity with the sequential source removal techniques. We have shown that there is a mathematical reasoning, based on the KKT criteria, that explains why at each iteration the strongest source in the residual image should be used in the optimization process.

## REFERENCES

- [1] J. A. Högbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, pp. 417–426, Jun. 1974.
- [2] S. Wijnholds and A.-J. van der Veen, "Data driven model based least squares image reconstruction for radio astronomy," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 2704–2707.
- [3] R. Levanda and A. Leshem, "Synthetic aperture radio telescopes," *Signal Processing Magazine, IEEE*, vol. 27, no. 1, pp. 14–29, 2010.
- [4] S. Wijnholds and A.-J. van der Veen, "Fundamental imaging limits of radio telescope arrays," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 613–623, 2008.
- [5] D. Briggs, "High fidelity deconvolution of moderately resolved sources," Ph.D. dissertation, The New Mexico Institute of Mining and Technology, 1995.
- [6] A. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *Information Theory, IEEE Transactions on*, vol. 54, no. 11, pp. 4813–4820, 2008.
- [7] A. Leshem and A.-J. van der Veen, "Radio-astronomical imaging in the presence of strong radio interference," *Information Theory, IEEE Transactions on*, vol. 46, no. 5, pp. 1730–1747, 2000.
- [8] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1981.
- [9] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed. University of California Press, Berkeley, CA, USA, 1950, pp. 481–492.