Array signal processing for radio-astronomy imaging and future radio telescopes

Amir Leshem
School of Engineering
Bar-Ilan University

Faculty of EEMCS
Delft University of Technology

Joint work with Alle-Jan van der Veen, Chen Ben-David, Ronny Levanda, A. Boonstra

Some data and simulations were provided by S. van der Toll, S. Wijnholds, Huib Intema, Huub Rottgering and T. Clarke

April 2009
What is parametric imaging?
References


Presentation overview

- The matrix measurement equation
  - Planar array
  - Non co-planar array
  - Polarization
  - Calibration*
  - Spatially varying calibration*

- Some ideas related to parametric imaging
  - LS-MVI Acceleration and hardware implementation
  - Joint imaging and calibration through SDP
  - L1 optimization*

- Examples
- Conclusions
High dynamic range

- 1 dimensional images – for simplicity
- Dirty images only!
- Theoretical dynamic range within the image 1,000,000:1
- Weakest source is only 10x noise RMSE!
- Sources are extended off the grid

- One source is moved through FOV
High dynamic range - Array beam

![Graph showing high dynamic range behavior in an array beam](image)
High dynamic range - sources

Source integrated power profile

Power $\sigma_{\text{noise}}$

$\theta^\circ$

$10^0$ $10^1$ $10^2$ $10^3$ $10^4$ $10^5$

$0$ $20$ $40$ $60$ $80$ $100$ $120$ $140$ $160$ $180$
Three 1-D dirty images:
Classical, AAR and AAR with 3% RMSE calibration error on each array element in each direction
Signals are stacked into vectors \( \mathbf{x}(t) = \left[ x_1(t), \ldots, x_p(t) \right]^T \)

Location of antenna \( i \) at time \( t \) is \( \mathbf{r}_i(t) \)

Baseline between antennas \( i, j \) is \( \mathbf{r}_i(t) - \mathbf{r}_j(t) \)
Imaging - the correlation process

\[
R_{k,\omega} = E\left[ x_{\omega}(t_k)x_{\omega}(t_k)^* \right] = \frac{1}{N} \sum_{n=0}^{10s/T_s} x_{\omega}(t_k + nT_s)x_{\omega}(t_k + nT_s)^*
\]

\( (R_{k,\omega})_{i,j} \) corresponds to the visibility \( V\left[ r_i(t_k), r_j(t_k) \right] \) at frequency \( \frac{\omega}{2\pi} \)
Coordinate system

Correlations are measured only for baselines $\mathbf{r}_i(t) - \mathbf{r}_j(t)$
Imaging equation – Planar array

\[ V_f(u, v) = \iint A^2(l, m)I_f(l, m)e^{-2\pi j(ul+vm)} dldm \]

- We limit ourselves to single frequency.
- \( A(l, m) \) is the "known" amplitude response of the antennas.
- \( I_f(l, m) \) is the intensity at location \((l, m)\).
- \( V_f(u, v) \) is the visibility function.
Imaging equation: Non-coplanar array

\[ V_f(u, v, w) = \int \int \frac{1}{\sqrt{1-l^2-m^2}} A^2(l,m,n) I_f(l,m) e^{-2\pi j (ul+vm+wn)} dldm \]
The discrete measurement equation

\[
V_{ijk} = V \left[ \mathbf{r}_i(t_k), \mathbf{r}_j(t_k) \right] = \sum_{l=1}^{d} I(s_l) e^{-j s_l^T (\mathbf{r}_i - \mathbf{r}_j)}
\]

- \( I(\cdot) \) is the brightness image (‘map’) of interest
- \( s_\ell \) is the unit direction vector of the \( \ell \)-th source (assuming discrete source model)
- \( V_{ijk} \) is the measured correlation \((R_k)_{ij}\) between antennas \(i\) and \(j\) at time \(t_k\)

**Classical Fourier-based imaging**

Given many samples of \( V_{ijk} \), we can compute \( I(s_\ell) \) via an inverse Fourier transform:

- “dirty image”: \( I_D(\mathbf{s}) := \sum_{i,j,k} V_{ijk} e^{j \mathbf{s}^T (\mathbf{r}_i - \mathbf{r}_j)} =: \sum_\ell I(s_\ell) B(\mathbf{s} - s_\ell) = I * B \)
- “dirty beam”: \( B(\mathbf{s} - s_\ell) := \sum_{i,j,k} e^{j (\mathbf{s} - s_\ell)^T (\mathbf{r}_i - \mathbf{r}_j)} \)

Every point source excites a beam \( B(\cdot) \) centered at its location \( s_\ell \)
Matrix formulation of the imaging equation

Recall

\[(R_k)_{i,j} \equiv V_{ijk} = \sum_{\ell=1}^{d} l(s_\ell) e^{-j s^T_\ell (r_i - r_j)} = \sum_{\ell=1}^{d} e^{-j s^T_\ell r_i} l(s_\ell) e^{-j s^T_\ell r_j}\]

- In "our" notation this translates to

\[R_k = A_k B A_k^H\]

where

\[A_k = \begin{bmatrix} a_k(s_1), \ldots, a_k(s_d) \end{bmatrix}, B = \begin{bmatrix} l(s_1) & 0 \\ \vdots & \ddots \\ 0 & l(s_d) \end{bmatrix}\]

- \[a_k(s) = \begin{bmatrix} e^{-j s^T r_1(t_k)} \\ \vdots \\ e^{-j s^T r_p(t_k)} \end{bmatrix}\]
The non co-planar array case

\[ R_k = A_k B A_k^H, \]

\[ R_k \equiv R(t_k), \quad A_k = [a_k(l_1, m_1), \ldots, a_k(l_d, m_d)] , \]

\[ a_k(l, m) = \begin{bmatrix}
    e^{-2\pi j(u_{10}(t_k)l + v_{10}(t_k)m + w_{10}(t_k)n)} \\
    \vdots \\
    e^{-2\pi j(u_{p0}(t_k)l + v_{p0}(t_k)m + w_{p0}(t_k)n)}
\end{bmatrix} \]

\[ B = \text{diag} \left[ \frac{I(l_1, m_1)}{\sqrt{1 - l_1^2 - m_1^2}}, \ldots, \frac{I(l_d, m_d)}{\sqrt{1 - l_d^2 - m_d^2}} \right]. \]
Matrix formulation of the imaging equation

\[ R_k = \Gamma(t_k)A_kA_k^H\Gamma(t_k)^H + \mathbf{R}_{nn}(t_k) \]

- \( \Gamma(t_k) \) is a self-calibration slowly time varying pxp matrix
- \( \mathbf{R}_{nn}(t_k) \) is the noise covariance matrix. Typically AWGN but when interference exists it dominates
When imaging polarized sources each element receives the two polarizations.

The steering vector at epoch \( k \) is now replaced by two orthogonal vectors describing circularly left and right polarization components

\[
\mathbf{a}_{k,L}(s), \mathbf{a}_{k,R}(s)
\]

For all \( s_1, s_2 \) we have

\[
\mathbf{a}_{k,L}(s_1) \perp \mathbf{a}_{k,R}(s_2)
\]

For a quasi mono-chromatic source the polarization can be described by the parameters

\[
\begin{bmatrix}
    b_{11} & b_{12} \\
    b_{21} & b_{22}
\end{bmatrix}
\]

which are the coherence between right and left circular polarization components.
The sources cross polarization coefficients $b_{11}, b_{12}, b_{21}, b_{22}$ are connected to the Stokes parameters $I, Q, U, V$ through the matrices

$$
\begin{bmatrix}
    b_{11} \\
    b_{12} \\
    b_{21} \\
    b_{22}
\end{bmatrix} = 
\begin{bmatrix}
    1 & 0 & 0 & 1 \\
    0 & 1 & j & 0 \\
    0 & 1 & -j & 0 \\
    1 & 0 & 0 & -1
\end{bmatrix} \begin{bmatrix}
    I \\
    Q \\
    U \\
    V
\end{bmatrix}
$$

This implies that it will be sufficient to deconvolve the parameters $b_{11}, b_{12}, b_{21}, b_{22}$ and then recover the Stokes parameter.
The matrix form of the (calibrated) measurement equation now becomes (assuming N point sources)

\[ \mathbf{A}_k = \left[ \mathbf{a}_{k,L}(s_1), \mathbf{a}_{k,R}(s_1), \ldots, \mathbf{a}_{k,L}(s_N), \mathbf{a}_{k,R}(s_N) \right] \]

\[ \mathbf{B} = \begin{bmatrix} b_{11}^1 & b_{12}^1 \\ b_{21}^1 & b_{22}^1 \\ \vdots & \vdots \\ b_{11}^N & b_{12}^N \\ b_{21}^N & b_{22}^N \end{bmatrix} \]

\[ \mathbf{R}_k = \mathbf{A}_k \mathbf{B} \mathbf{A}_k^H + \sigma^2 \mathbf{I}, \quad k=1,\ldots,K \]
Self calibration with polarized sources

\[ R_k = \Gamma_k A_k B A_k^H \Gamma_k^H + \sigma^2 I, \quad k=1,\ldots,K \]

\[ \Gamma_k = \text{diag}(\Gamma_{k,1},\ldots,\Gamma_{k,p}) \]

\[ \Gamma_{k,i} = \begin{bmatrix} (\gamma_{11})_{k,i} & (\gamma_{12})_{k,i} \\ (\gamma_{21})_{k,i} & (\gamma_{22})_{k,i} \end{bmatrix} \]
The “CLEAN” algorithm (initially Hogbom 1974)

\[
\ell = 0, \quad \gamma = 0.1 \cdots 0.5 \text{ ("loop gain")}
\]

while \(I_D\) is not noise-like:

\[
\begin{align*}
\mathbf{s}_\ell & = \arg \max I_D(\mathbf{s}) \\
\lambda_\ell & = \frac{I_D(\mathbf{s}_\ell)}{B(0)} \\
I_D(\mathbf{s}) & := I_D(\mathbf{s}) - \gamma \lambda_\ell B(\mathbf{s} - \mathbf{s}_\ell) \\
\ell & = \ell + 1
\end{align*}
\]

\[
I(\mathbf{s}) = I_D(\mathbf{s}) + \sum_\ell \gamma \lambda_\ell B_{\text{synth}}(\mathbf{s} - \mathbf{s}_\ell)
\]

- This is a successive cancellation type algorithm
- Often combined with calibration refinement (SELF-CAL)
- Alternative possibilities (e.g., Max Entropy). MLE?
Deconvolution in the (u,v) plain

- Clark proposed to estimate several CLEAN components in a single dirty image and subtract them.
- Cotton-Schwab proposed to perform the subtraction in the (u,v) domain. This allows for non-grid points:
  - We will always consider the Cotton-Schwab approach.
  - Perform non-grid estimation whenever possible.
  - We will discuss extension of the Clark approach.
Deconvolution

Classical Fourier imaging

\[ I_D(s) = \sum_{k=1}^{K} a_k^H(s) R_k a_k(s) \]

Super-resolution MVDR based imaging

Pseudo-spectrum

\[ I'_D(s) = \sum_{k=1}^{K} w_k^H(s) R_k w_k(s) \]

MVDR criterion

\[ \hat{w}_k(s) = \arg \min_w I'_D(s) \]
such that \[ w_k^H(s)a_k(s) = 1 \]

Solution

\[ \hat{w}_k(s) = \beta_k R_k^{-1} a_k(s) \]
where \[ \beta_k = \frac{1}{a_k^H(s) R_k^{-1} a_k(s)} \]

\[ I'_D(s) = \sum_{k=1}^{K} \frac{1}{a_k^H(s) R_k^{-1} a_k(s)} \]

April 2009
We would like to maintain the Gaussian receiver noise isotropic. To that end we add to the MVDR equations the requirement
\[ \|w\| = 1 \]
Still we want a solution of the form \( w_k(s) = \alpha R_k^{-1}a_k(s) \)
We obtain that
\[
w = \frac{R_k^{-1}a_k(s)}{a_k(s)R_k^{-2}a_k(s)}
\]
and the angular spectrum becomes
\[
I_D'(s) = \sum_{k=1}^{K} \frac{a_k^H(s)R_k^{-1}a_k(s)}{a_k^H(s)R_k^{-2}a_k(s)}
\]
Since the covariance matrices at different epochs represent independent observations it is reasonable to assume that the overall covariance of the data is given by

\[
\bar{\mathbf{R}} = \begin{bmatrix}
\mathbf{R}_1 & & \\
& \ddots & \\
& & \mathbf{R}_K
\end{bmatrix}
\]

Choosing a vector \( \bar{\mathbf{w}} \) for \( \bar{\mathbf{R}} \) yields \( \bar{\mathbf{w}} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_K^T]^T \) where

\[
\mathbf{w}_k = \mathbf{R}_k^{-1} \mathbf{a}_k (s)
\]

In this case the MVDR dirty image becomes:

\[
I_{\text{MVDR}} (s) = \frac{1}{\sum_{k=1}^K \mathbf{a}_k^H (s) \mathbf{R}_k^{-1} \mathbf{a}_k (s)}
\]
The AAR dirty image

\[ I_{AAR}(s) = \frac{\sum_{k=1}^{K} a_k^H(s) R_k^{-1} a_k(s)}{\sum_{k=1}^{K} a_k^H(s) R_k^{-2} a_k(s)} \]
Dirty images

(a) Classical
(b) MVDR
(c) AAR
Bandwidth synthesis

• Extension to bandwidth synthesis is different than the classical imaging due to non-linear structure
• Spectral index of sources can be included as extra parameter (not shown below)

\[
I_{\text{AAR}}(s) = \frac{\sum_{\omega} \sum_{k=1}^{K} a_{k,\omega}^H (s) R_{k,\omega}^{-1} a_{k,\omega} (s)}{\sum_{\omega} \sum_{k=1}^{K} a_{k,\omega}^H (s) R_{k,\omega}^{-2} a_{k,\omega} (s)}
\]
Classical beamforming based CLEAN

$$\begin{align*}
\max_{\mathbf{s}} & \quad \sum_{k=1}^{K} \mathbf{w}_k^H(\mathbf{s}) \mathbf{R}_k \mathbf{w}_k(\mathbf{s}) \\
\mathbf{R}_k & = \sum_{n=1}^{N} b_n \mathbf{a}_k(s_n) \mathbf{a}_k^H(s_n) + \sigma^2 \mathbf{I} \\
\mathbf{I}_D(\mathbf{s}) & = \sum_{k=1}^{K} \mathbf{a}_k^H(\mathbf{s}) \mathbf{R}_k \mathbf{a}_k(\mathbf{s}) \\
\hat{\mathbf{s}} & = \arg \max_{\mathbf{s}} \mathbf{I}_D(\mathbf{s}) \quad \hat{\alpha} = \max \mathbf{I}_D(\hat{\mathbf{s}}) \\
\mathbf{R}_k & = \mathbf{R}_k - \gamma \hat{\alpha} \mathbf{a}_k(\mathbf{s}) \mathbf{a}_k^H(\mathbf{s})
\end{align*}$$
\[ \hat{s} = \arg \max_s l_{AAR}(s) \quad \hat{\alpha} = \max \ l_{AAR}(\hat{s}) \]

\[ R_k = R_k - \gamma \hat{\alpha} a_k(s) a_k^H(s) \]
Least Squares power estimator to remove bias

\[ \lambda^{(n)} = \arg \min_{\alpha} \sum_{k=1}^{K} \left\| R^{(n)}_k - \alpha a(s^{(n)})a^H(s^{(n)}) \right\|_F^2 \]

The LS-MVI algorithm

Measure $\mathbf{R}_1, \ldots, \mathbf{R}_K$

Set $\mathbf{R}_k^{(0)} = \mathbf{R}_k, \gamma = 0.1$

Until convergence criterion

Source location estimator

$$\mathbf{s}^{(n)} = \arg \max_{\mathbf{s}} I_{\text{AAR}}^{(n)}(\mathbf{s})$$

Power estimator

$$\lambda^{(n)} = \arg \min_{\alpha} \sum_{k=1}^{K} \left\| \mathbf{R}_k^{(n)} - \alpha \mathbf{a}(\mathbf{s}^{(n)})\mathbf{a}^H(\mathbf{s}^{(n)}) \right\|_F^2$$

Update $\mathbf{R}_k^{(n)}$

$$\mathbf{R}_k^{(n+1)} = \mathbf{R}_k^{(n)} - \gamma \lambda^{(n)} \mathbf{a}(\mathbf{s}^{(n)})\mathbf{a}^H(\mathbf{s}^{(n)})$$
Enforcing the non-negative definite constraint

All source powers are non-negative and residual covariance is also non-negative

\[
R_k - \sigma^2 I - \alpha a(s^{(n)}) a^H (s^{(n)}) \geq 0
\]
\[
\alpha \geq 0
\]

Solution is simple:
Solve the unconstrained problem and use bisection on \( \alpha \) finding sequentially over \( k \) \( \alpha \) that is good for all \( k \)
Other enhancements

- Clark type algorithm using SDP
- Cotton-Schwab with off-grid estimation
Complexity analysis

Main bottleneck: Recomputing the dirty image at each stage
Complexity of naive implementation is $M^2 p^2 K$ per iteration
where $M^2$ is the number of pixels in the image, $p$ number of antennas

We can reduce this to $M^2 pK$ per iteration
Still high compared to gridded CLEAN with $2M^2 \log_2 M$ operations for recomputing the dirty image, but manageable.
These techniques are easy to implement using parallel processors!
Dedicated accelerators might be a viable option!

Can use coarse grid and use local optimization to obtain accurate location estimate!
We do not need intermediate images just accurate parameter estimates!
Assume that \( a_k(s) \) is not completely known.
\( \bar{a}_k(s) \) is the nominal value.
\( C_k \) is a positive definite matrix describing the uncertainty at epoch \( k \)
\[
(a_k(s) - \bar{a}_k(s))^H C_k (a_k(s) - \bar{a}_k(s)) \leq 1
\]

The robust beamforming problem is given by

\[
\begin{align*}
\begin{bmatrix}
\rho, \hat{a}_1(s), \ldots, \hat{a}_K(s)
\end{bmatrix} &= \arg \max_{\rho, a_1, \ldots, a_K} \rho \\
\text{subject to} & \\
R_k - \sigma^2 I - \rho a_k a_k^H & \succeq 0 \\
\rho & \geq 0 \\
(a_k(s) - \bar{a}_k(s))^H C_k (a_k(s) - \bar{a}_k(s)) & \leq 1
\end{align*}
\]
Equivalent SDP formulation

\[ [\rho, \hat{a}_1(s), \ldots, \hat{a}_K(s)] = \arg \min_{\tau, a_1, \ldots, a_K} \tau \]
subject to

\[
\begin{bmatrix}
R_k - \sigma^2 I & a_k \\
a_k^H & \tau
\end{bmatrix} \succeq 0
\]

\[
\begin{bmatrix}
C_k & (a_k - \bar{a}_k(s)) \\
(a_k - \bar{a}_k(s))^H & 1
\end{bmatrix} \succeq 0
\]

\[ \rho = 1 / \tau \]
\[ \hat{\Gamma} = \arg \min_{\Gamma} \sum_{k=1}^{K} \sum_{n=1}^{N} \left\| \hat{a}_k(s^{(n)}) - \Gamma \bar{a}_k(s^{(n)}) \right\|^2 \]
Extensions

- Simplified Maximum likelihood (Leshem and van der Veen 2000)
- Maximum likelihood using EM (Lanterman)
- Robust beamforming techniques, e.g., diagonal loading
- Generalized LS-MVI using parametric source templates (template: shapelets, curvelets, ridgelets, wavelets…)
- Global optimization techniques: L1 and others

- Performance analysis
- Sensitivity to modeling assumptions
- Resolution limits
Simulated experiments

- East west array with 10 elements and longest baseline 1000λ.
  - 720 covariance matrices
- Fully calibrated array
Resolution example
Extended sources example
Extended sources example

(a) CLEAN
(b) LS_MVI
Conclusions

- Exciting science can be done with future instruments
- Science strongly depends on technology

- Great challenges:
  - Calibration is extremely challenging!
  - Image processing can take us beyond the equipment!!
  - Real time adaptive interference mitigation is crucial
  - Architectures and algorithms for extremely large arrays
    - Comm – 2-4 sensors
    - Radar - thousands of sensors
    - Radio astronomy – 10,000-100,000,000 elements