Beyond the CLEAN approach in image synthesis.

Astroinformatics 2014

Pablo E. Román[1] and Takeshi Asahi[2] and Simón Casassus[3]

[1]Laboratorio de Astroinformática
Center of Mathematical Modeling.
[2]Modeling in Scientific Imaging and
Visualization Laboratory (MOTIV)
Center of Mathematical Modeling.
[3]Departamento de Astronomía.
Universidad de Chile.

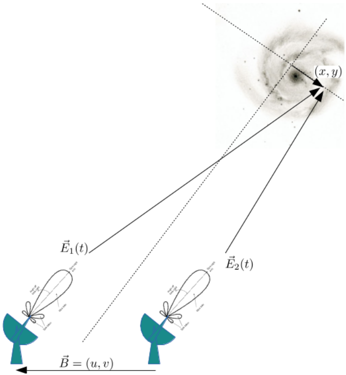
August 28, 2014

Table of contents

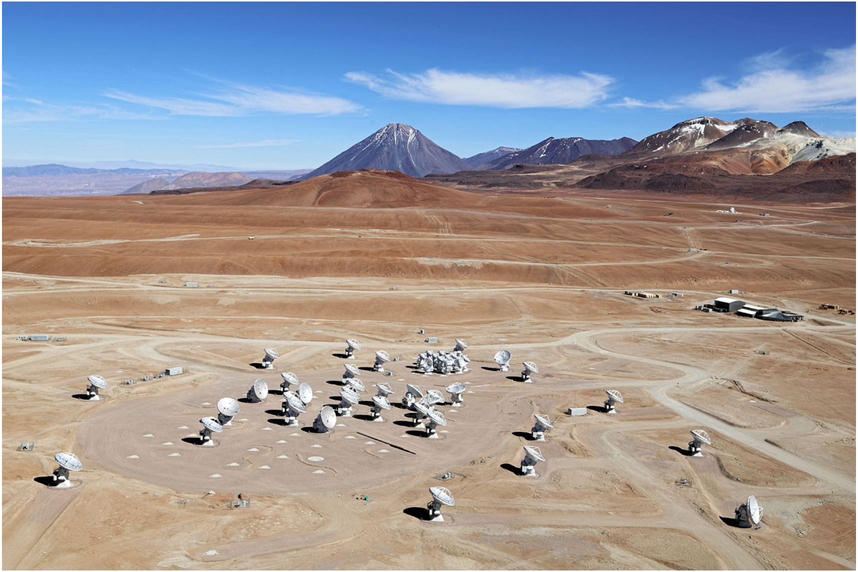
1. Introduction
   - Radio Interferometry
   - The problem
2. Current solution
   - CLEAN ALGORITHM (1978)
   - OTHER ALGORITHMS
   - Our Approach

Radio Interferometry

- **Measurements**: time integrated correlation $V_{i,j} =< \vec{E}_i^* \vec{E}_j >$ between pairs of antennas (i, j) called VISIBILITY.
- **Van Cittert-Zernike theorem**: Fourier transform of the power emission (Image intensity) in a sky position corresponds to VISIBILITY [Taylor et al., 1999].
- $V(u, v) = \int_{\mathbb{R}^2} \mathcal{A}(x, y) I(x, y) e^{-2\pi i(ux+vy)} dxdy$
- **Theoretical Resolution**: λ/B

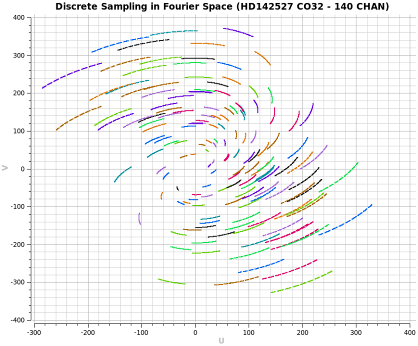


Radio interferometer (ALMA)



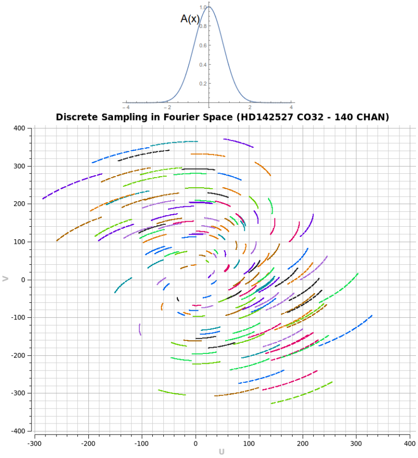
BIG PROBLEM: Sparse Sampling

- $V(u,v) = \int_{\mathbb{R}^2} \mathcal{A}(x,y) I(x,y) e^{-2\pi i(ux+vy)} dxdy$
- **Points in uv plane**: [Distance/wavelenght].
- **Sampling variation**:
  - Earth rotation (≈phase+radial)
  - Frequency (≈radial)
- **Real valued image**: $\bar{V}(-z) = V(z),\ z = (u,v)$
- **Far from regular: NO SAMPLING THEOREM.**



Sparse Sampling in Interferometry

1. $V(u,v) = \mathcal{F}[\mathcal{A} \cdot I](u,v)$
2. **V function is a random variable**: is a Random Field.
3. **We want**: estimation for $E(I)$.
4. **We want**: estimation for $E(I^2)$.
5. **Our data**: (Simplified version)
   - Sampling points $\{z_k = (u_k, v_k)\}_{k=1}^N$, $N \sim 10^4 - 10^6$
   - Sampled function values at those points $\{V_k^o\}_{k=1}^N$
   - Variance estimation of measurement $\{\sigma_k\}_{k=1}^N$

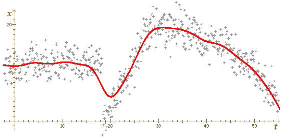


In short: Main issues

- **Large variance in measurement**: Visibilities deviations near the order of the value $V_k \approx \sigma_k$
- **Heteroscedastic variance**: σ_k depends on time, frequency, and pairs of antennas.
- **Irregular sampling**: Traditional methods works on regular mesh (Shannon-Whitaker Interpolation), FFT uniform convergence. Even in near from regular meshes interpolation, some formulas has been proposed (Marsvati, et al. 2001). This case is far from traditional theory.
- **Image variance computation**: A regular image of 1024x1024 pixels should require to compute a variance matrix of 10^{12} entries! An instrument without errors measurements?
- **Images has never been observed**: Protoplanetary disk are practically invisible to visible light and central star is invisible to millimeter frequency.
- **Missing low frequencies**: uv hole.

A comment ...

Solving the inverse problem is equivalent to “interpolate” in the complex plane.



THE CLEAN ALGORITHM

CASA
Common Astronomy Software Applications

MIRIAD

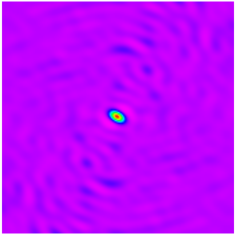
DIFMAP

...

CLEAN: The dirty beam/image concept

$$
\begin{aligned}
S(u,v) &= \sum_{k=1}^{N} \omega_k \delta(u_k, v_k) \\
V(u,v) &\to S(u,v)V(u,v) \\
A(x_j,y_j)I^D(x_j,y_j) &= \mathcal{F}[S(u,v)V(u,v)] \\
&= \mathcal{F}[S] * \mathcal{F}[V] = B^D * (AI) \\
&(= \sum_{k=1}^{N} V(u_k,v_k) e^{2\pi i (x_j u_k + y_j v_k)}) \\
B^D &= \sum_{k=1}^{N} \omega_k Cos(2\pi(x u_k + y v_k))
\end{aligned}
$$

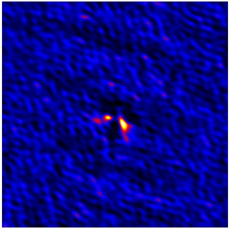
B^D



CLEAN: The dirty beam/image concept

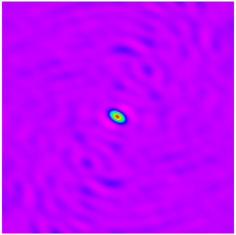
$$
\begin{aligned}
S(u,v) &= \sum_{k=1}^{N} \omega_k \delta(u_k, v_k) \\
V(u,v) &\rightarrow S(u,v)V(u,v) \\
A(x_j,y_j)I^D(x_j,y_j) &= \mathcal{F}[S(u,v)V(u,v)] \\
&= \mathcal{F}[S] * \mathcal{F}[V] = B^D * (AI) \\
&(= \sum_{k=1}^{N} V(u_k,v_k) e^{2\pi i (x_j u_k + y_j v_k)}) \\
B^D &= \sum_{k=1}^{N} \omega_k Cos(2\pi(xu_k + yv_k))
\end{aligned}
$$

I^D

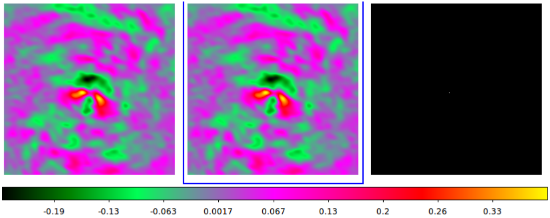


CLEAN ALGORITHM [Högbom, 1974]

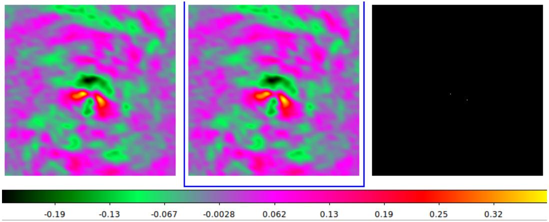
1. $I^R \leftarrow I^D$, $I^M \leftarrow 0$
2. $k = Argmax_k\{I^R(x_k, y_k)\}$
3. $I^M \leftarrow I^M + \lambda_k \delta(x - x_k, y - y_k)$
4. $I^R \leftarrow I^R - B^D * \lambda_k \delta(x - x_k, y - y_k)$
5. If $|I^R| > \epsilon$ Then goto (2)
6. Return $I^R + B^C * I^M$



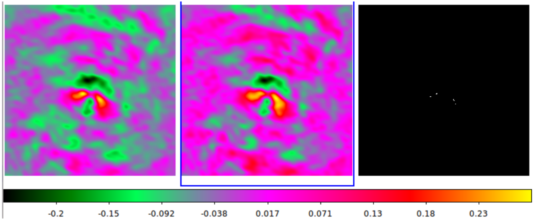
CLEAN ITER 0



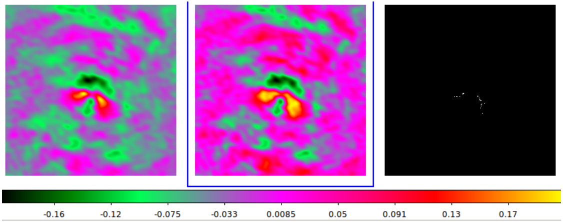
CLEAN ITER 1



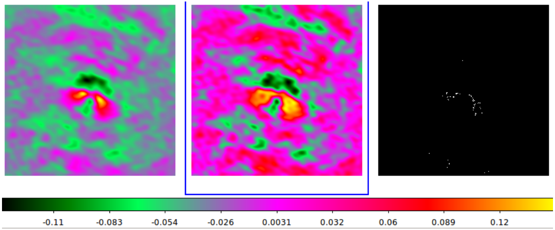
CLEAN ITER 10



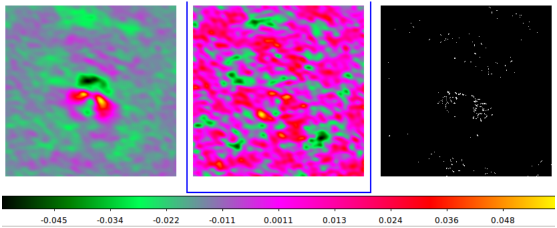
CLEAN ITER 30



CLEAN ITER 100



CLEAN ITER 1000



Some observations

- State-of-the-art by nearly 40 years!!!
- Very efficient and could be assisted
- No statistical assessment
- Restored images are evaluated by physics argumentation
- Many other algorithms has been tested, but CLEAN remains the most used.
- Example: **1000 component representation of a 1Mpix image!**

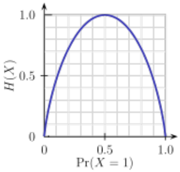
CLEAN INTERPRETATION: First sparse algorithm

- Matching pursuit algorithm
- Maximize total intensity (Norm-1)
- Compressed sensing
- Suboptimal solution
- Efficient and sparse solution
- Millions of variables reduced to less than hundred

Algorithm Matching Pursuit
Input: Signal: $f(t)$, dictionary D.
Output: List of coefficients: (a_n, g_{γ_n}).
Initialization:
$R_1 \leftarrow f(t)$;
$n \leftarrow 1$;
Repeat:
find $g_{\gamma_n} \in D$ with maximum inner product $|\langle R_n, g_{\gamma_n} \rangle|$;
$a_n \leftarrow \langle R_n, g_{\gamma_n} \rangle$;
$R_{n+1} \leftarrow R_n - a_n g_{\gamma_n}$;
$n \leftarrow n + 1$;
Until stop condition (for example: $\|R_n\| < \text{threshold}$)

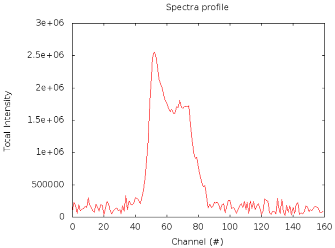
Maximum Entropy Methods (MEM)

- [Sutton and Wandelt, 2006]
- Measured visibilities V_k with variance $\sigma_k^2 = 1/\omega_k$
- Maximum likelihood problem derives a χ^2 minimization plus an entropy term.
- $\min_{[I_i]} \sum_k |(V_k - V^M(u_k, v_k))/\sigma_k|^2 + \lambda \sum_i I_i \log(I_i/M_i)$
- $I_i \geq 0$



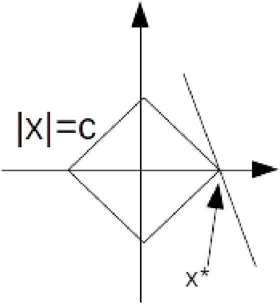
MEM Algorithm

- Optimization problem in millions of variables $\simeq$ number of pixels
- Several hour of processing in cluster
- Much better peak-to-noise
- Less artifact
- statistical basis



Compressed sensing (CS) based Algorithm

- Sparse representation: Most coefficient c_k are null.
- Assumption: sparsity and completeness.
- Norm L_1 approximate spike measure $|\cdot|_0$
- LASSO method, split optimization, convex method, matching pursuit.
- CLEAN basis: Set of Gaussian centered on pixel grid.



- $f(x) \sim \sum_{k \in S} c_k \phi_k(x)$
- $\min_c |c|_1 = \sum_{k \in S} |c_k|$
  $\sum_j |f^o(x_j) - \sum_{k \in S} c_k \phi_k(x_j)|_2^2 \leq \epsilon$

CS Algorithm

- [Carrillo et al., 2013] implementation for sparse problem in image synthesis.
- Better results using redundant basis.
- Empirical way to propose representation.
- How to guaranty sparsity feasibility?
- No statistical assessment.
- **Radio interferometric images are restricted to be physical.**

CS Algorithm

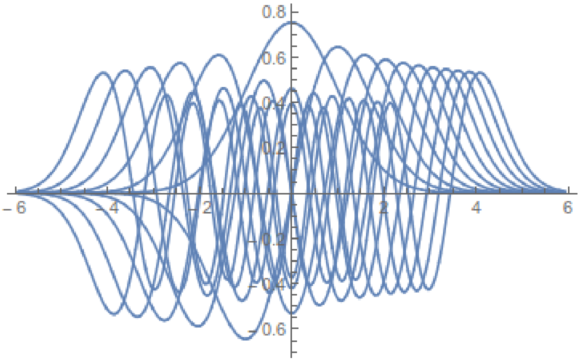
- Which basis to use?
- Grid could be replaced by coefficient space!
- How to guaranty sparsity feasibility?

Reason for CS

- **Sparse representation enable enhanced INDEXING.**
- **Sparse representation enable more representative FEATURES.** [Oja et al., 1999]
- **Sparse representation seems to be implemented in biological system such as vision.**[Lennie, 2003][Simoncelli, 2005]
- Grid could be replaced by coefficient space!

Our approach: Diagonalize the Fourier operator!

1. Gaussian Hermite functions $\phi_k(x) = \frac{1}{\sqrt{\sqrt{2\pi}2^k k!}} e^{-x^2/2} H_k(x)$
2. Fourier Eigenvector: $\mathcal{F}[\phi_k](x) = (-i)^k \phi_k(x)$
3. (Roman et al., in prep)



Hermite functions

1. Solutions of the harmonic quantum oscillator
2. Also called "Shapelets" [P. Melchior et al., 2007]
3. $\phi_{n,m}(x,y) = \frac{1}{\sqrt{\pi 2^{n+m+1} n! m!}} H_n(x) H_m(y) e^{-\frac{1}{2}(x^2+y^2)}$
4. $\mathcal{F}[\phi_{nm}] = (-i)^{n+m}\phi_{nm} \rightarrow$ No FFT needed
5. First approach: $V^o(u_l, v_l) = \sum_{n=0,m=0}^{M} y_{nm}\phi_{nm}(u_l, v_l)$
6. LASSO: $Min_Y |V - AY|^2 + \alpha|Y|_1$, $[A_{l,(n,m)}] = [\phi_{nm}(u_l, v_l)]$.
7. Second approach: Overcomplete Dictionary and sparsity $\{\phi_{n,m,u_0,v_0,\sigma_u,\sigma_v,\theta} = \phi_{n,m}(R_\theta(u-u_0)/\sigma_u, (v-v_0)/\sigma_v)\}$
8. Five parameter $u_0, v_0, \sigma_u, \sigma_v, \theta$

Advantages of this basis

- GRID-LESS: No need of a grid for using FFT. Fourier transform is implemented as $\mathcal{F}[f(x)] = \mathcal{F}[\sum_{k\in S} c_k\phi_k(x)] = \sum_{k\in S}(-i)^k c_k\phi_k(x)$
- Diagonalize covariance matrix of coefficient represented in this basis.
- Coefficient of the representation would were uncorrelated variable if they are Gaussian....
- Variance of the coefficient representation are minimal overall other representation.
- It is a Gabor like basis. Biological vision usually represented in this family of basis [Gorea, 1991].
- **Sparcity by coefficient decay: If $f(x)$ decay at $\pm\infty$ as $e^{-x^2/a}$ then coefficient c_k decay asymptotically faster than any polynomial! $\rightarrow$ THE PRIMARY BEAM EFFECT!!!**

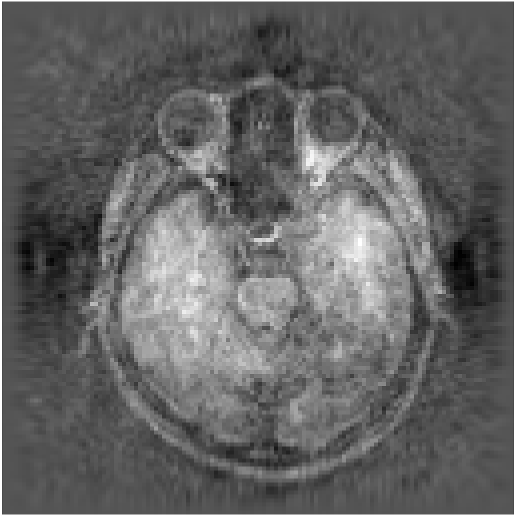
Some tests

Original

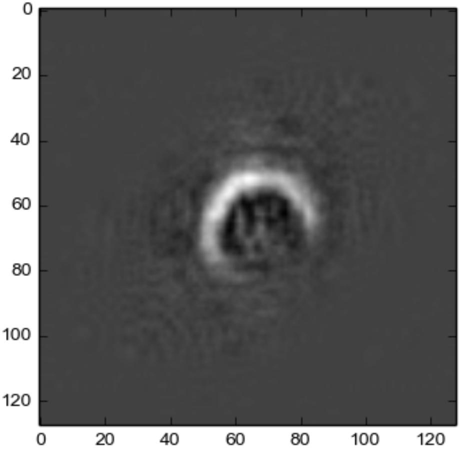
Interpolated in Fourier space

(n=100)

Some tests: Noise (proportional to intensity) + random sampling (10 % of total grid)



Some tests: 15000 visibility of 80000 ALMA sampling from HD142527 band9 continuum



Bibliography

R. E. Carrillo, J. D. McEwen, D. V. D. Ville, J.-P. Thiran, and Y. Wiaux. Sparsity averaging for compressive imaging. *IEEE Signal Process. Lett.*, 20(6):591–594, 2013.

A. Gorea. *"Representations of Vision: Trends and Tacit Assumptions in Vision Research"*. Cambridge University Press, 1991. ISBN 9780521412285. URL `http://books.google.cl/books?id=N5pOAAAAIAAJ`.

J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics Supplement*, 15: 417, June 1974.

P. Lennie. The cost of cortical computation. *Current Biology*, 13(6):493 – 497, 2003.

E. Oja, A. Hyvrinen, and P. Hoyer. Image feature extraction and denoising by sparse coding. *Pattern Analysis & Applications*, 2(2):104–110, 1999. ISSN 1433-7541. doi: 10.1007/s100440050021. URL `http://dx.doi.org/10.1007/s100440050021`.