Inverse Problems in Astrophysics

J.-L. Starck
CEA, IRFU, Service d'Astrophysique, France
jstarck@cea.fr
http://jstarck.free.fr

Inverse Problems in Astrophysics

- **Part 1: Introduction inverse problems and image deconvolution**
- Part 2: Introduction to Sparsity and Compressed Sensing
- Part 3: Wavelets in Astronomy: from orthogonal wavelets and to the Starlet transform.
- Part 4: Beyond Wavelets
- Part 5: Inverse problems and their solution using sparsity: denoising, deconvolution, inpainting, blind source separation.
- Part 6: CMB & Sparsity
- Part 7: Perspective of Sparsity & Compressed Sensing in Astrophsyics

Inverse Problems in Astrophysics

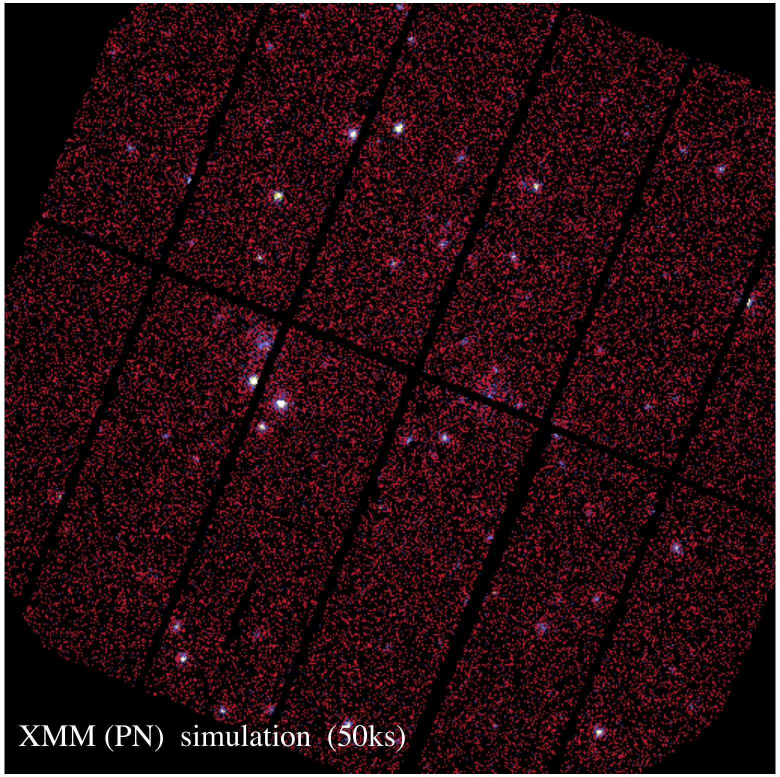
$$Y = HX + N$$

PB 1: find X knowing Y,H and the statistical properties of the noise N
Ex: Astronomical image deconvolution
Weak lensing

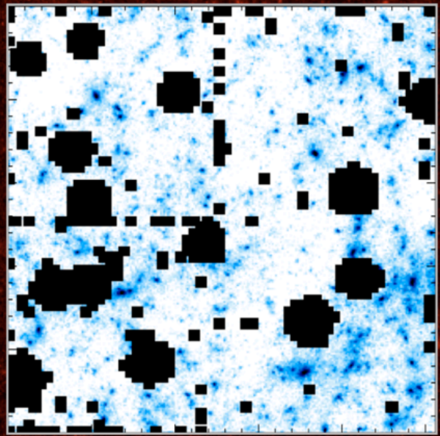
PB 2: find X and H knowing Y and the statistical properties of the noise N
Ex: Blind deconvolution

Ill posed problem, i.e. not an unique and stable solution ==> Regularization

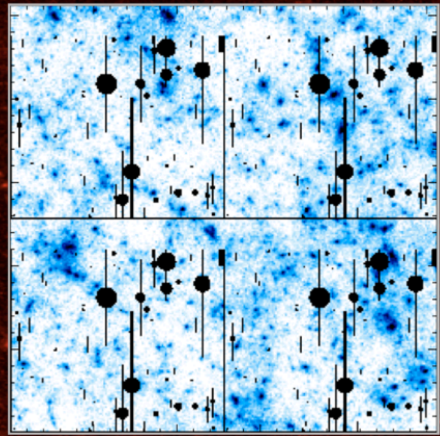
$$||Y - HX||^2$$ with some constraints on X


XMM (PN) simulation (50ks)

Masked masks

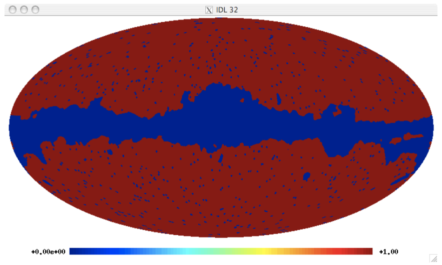


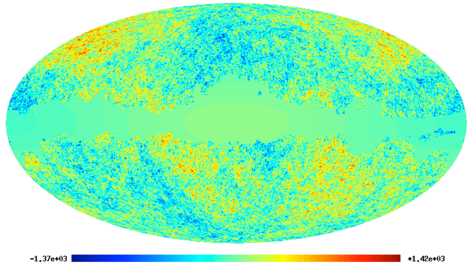
Mask pattern of CFHTLS survey on 1° x 1° field



Mask pattern of Subaru survey on 1° x 1° field

MISSING DATA





- Power estimation estimation.
- Gaussianity test, isotropy test, etc

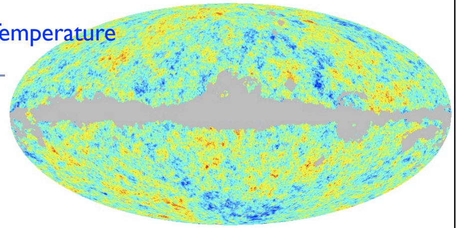
ISW Reconstruction

- Previously: Cross-Correlate
  <Tg>

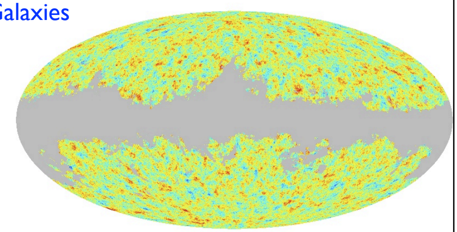
$$T_{\mathrm{CMB}}^{\mathrm{obs}} = T_{\mathrm{primordial}} + \alpha T_{\mathrm{ISW}}$$

- Reconstruct part of Temperature map due to ISW
  - Reconstruct large scale secondary anisotropies
  - Due to one or several galaxy distributions in foreground
  - Recover primordial T at large scales
- Detection tricky ➔ Reconstruction complex problem

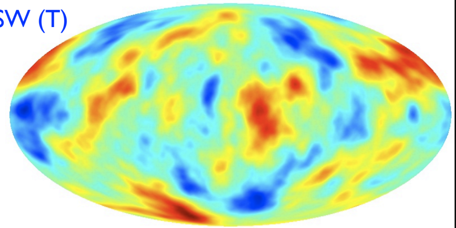
Temperature

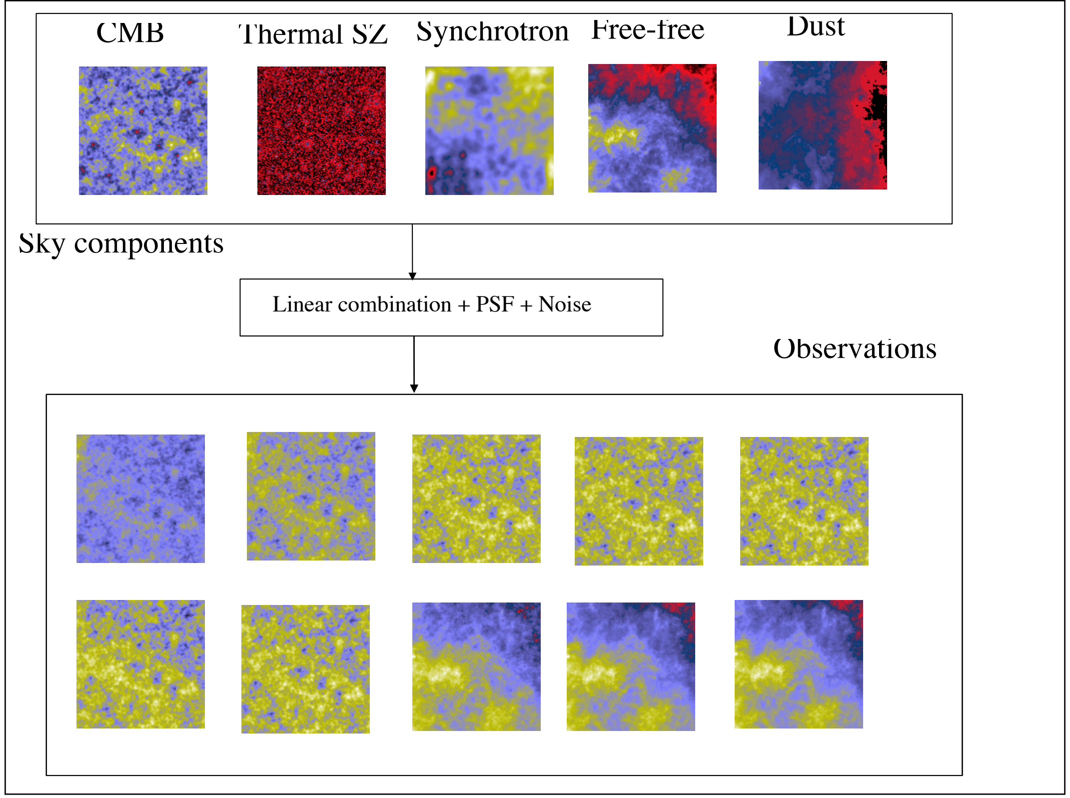


Galaxies



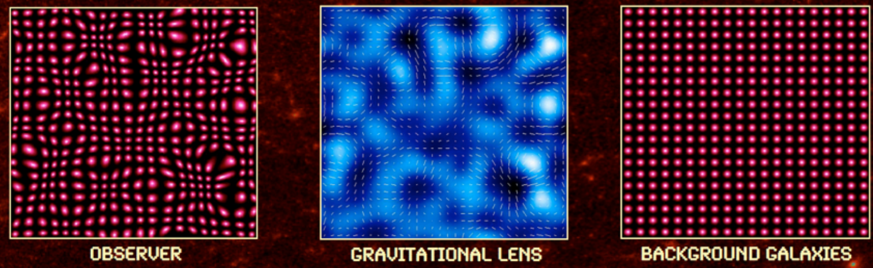
ISW (T)

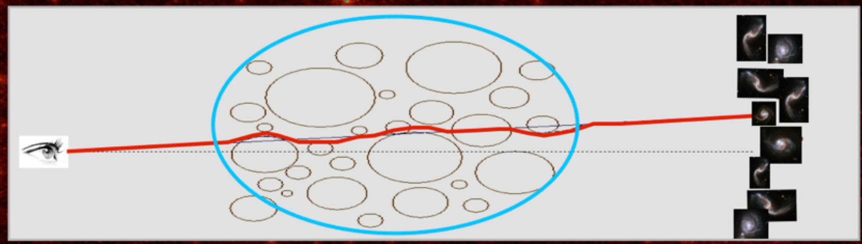



CMB
Thermal SZ
Synchrotron
Free-free
Dust
Sky components
Linear combination + PSF + Noise
Observations

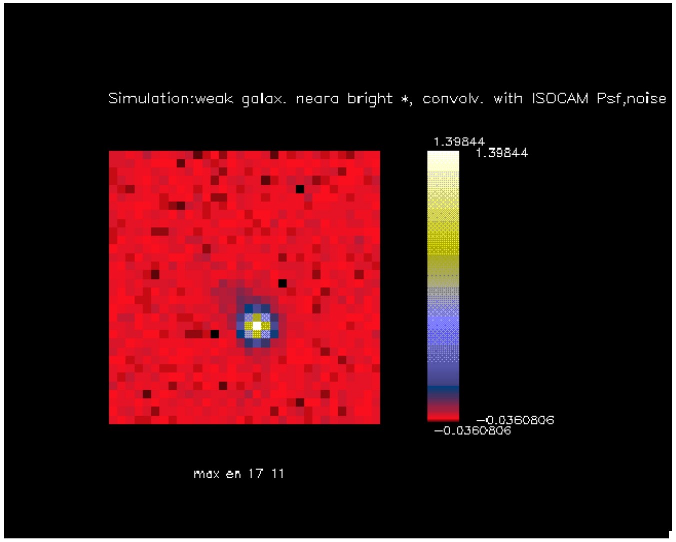
Strong Gravitational Lensing

Weak Gravitational Lensing





GRAVITATIONAL LENS

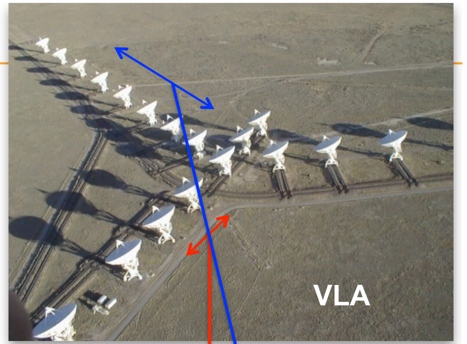

Simulation:weak galax. neara bright *, convolv. with ISOCAM Psf,noise
1.39844
1.39844
−0.0360806
−0.0360806
max en 17 11

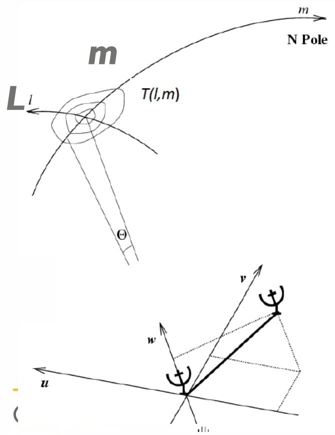
Multi-element interferometer

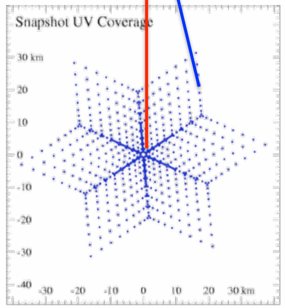
N antennas/telescopes

$\frac{N(N-1)}{2}$ independent baselines

1 projected baseline
= 1 sample in the Fourier « u,v » plane

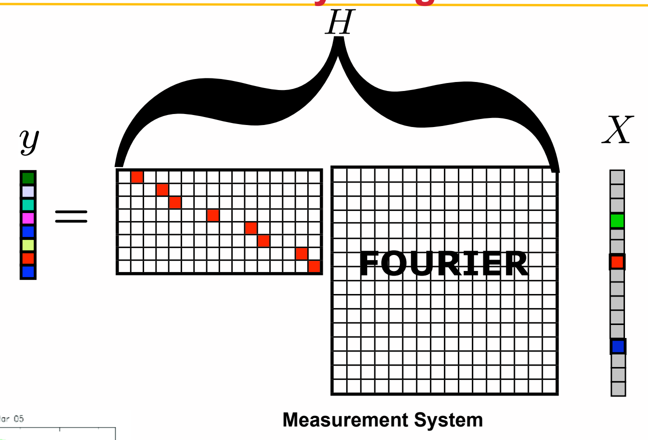


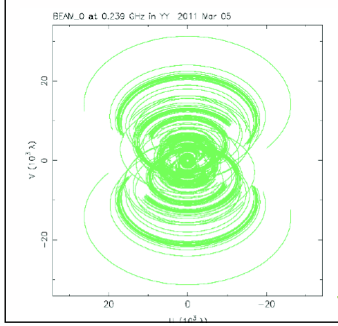




$$V(u,v) = \int\int T(l,m) e^{-i2\pi(ul+vm)} dl dm$$

Radio-Interferometry Image Reconstruction



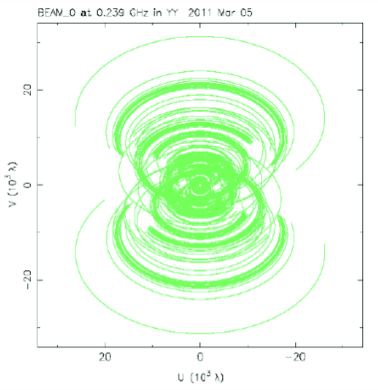


$$Y = HX + N$$

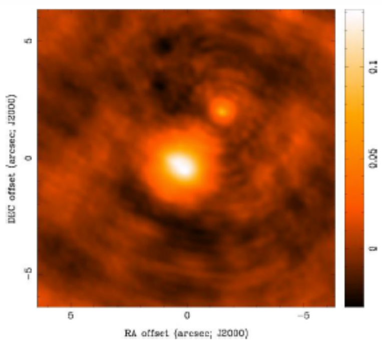
Fourier domain

Snapshot (u,v) coverage

Image domain



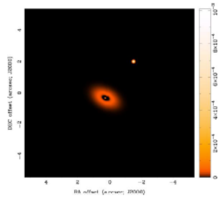
$\sim FT^{-1}$



Reconstructed image =
« true » sky * PSF =

Dirty image

discontinuous sampling of the (Fourier) (u,v) plane



True sky

Deconvolution

The image formation is expressed in the convolution integral

$$\begin{aligned} Y(x,y) &= \int_{x_1=-\infty}^{+\infty}\int_{y_1=-\infty}^{+\infty} h(x-x_1, y-y_1)X(x_1,y_1)dx_1dy_1 + N(x,y) \\ &= (h * X)(x,y) + N(x,y) = HX + N \end{aligned}$$

where Y is the data, H the point-spread-function (PSF), and X is the solution.

In Fourier space we have:

$$\hat{Y}(u,v) = \hat{h}(u,v)\hat{X}(u,v) + \hat{N}(u,v)$$

We want to determine X knowing h and X. The main difficulties are the existence of:

- a cut-off frequency of the point spread function.
- the noise.

It is in fact an **ill posed problem**, there is not an unique solution.

Fourier-quotient method

A solution can be obtained by computing the Fourier transform of the deconvolved object $\hat{O}$ by a simple division between the image $\hat{I}$ and the PSF $\hat{P}$

$$\hat{\tilde{X}}(u,v) = \frac{\hat{Y}(u,v)}{\hat{h}(u,v)} = \hat{X}(u,v) + \frac{\hat{N}(u,v)}{\hat{h}(u,v)}$$

This method, sometimes called *Fourier-quotient method* is very fast. We only need to do a Fourier transform and an inverse Fourier transform.

For frequencies close the frequency cut-off, the noise term becomes important, and the noise is amplified. Then **in the presence of noise, this method cannot be used.**

Least-square solution

It is easy to verify that the minimization of $\| Y(x,y) - h(x,y) * X(x,y) \|^2$ lead to the solution:

$$\hat{\tilde{X}}(u,v) = \frac{\hat{h}^*(u,v)\hat{Y}(u,v)}{\mid \hat{h}(u,v) \mid^2}$$

which is defined on if $\hat{h}(u,v)$ is different from zero. The problem is general ill-posed and we need to introduce a *regularization* in order to find an unique and stable solution.

Tikhonov regularization

Tikhonov regularization consists of minimizing the term:

$$J_T(X) = \parallel Y - HX \parallel^2 + \lambda \parallel FX \parallel^2$$

where f corresponds to a high-pass filter. This criterion contains two terms. The first, $\parallel Y - HX \parallel^2$, expresses fidelity to the data Y, and the second, $\lambda \parallel FX \parallel^2$, expresses smoothness of the restored image.

λ is the regularization parameter and represents the trade-off between fidelity to the data and the smoothness of the restored image.

The solution is obtained directly in Fourier space

$$\hat{\tilde{X}}(u,v) = \frac{\hat{h}^*(u,v)\hat{Y}(u,v)}{\mid \hat{h}(u,v) \mid^2 + \lambda \mid \hat{f}(u,v) \mid^2}$$

Generalization

This method can be generalized, and we write:

$$\hat{\tilde{X}}(u,v) = \hat{W}(u,v)\frac{\hat{I}(u,v)}{\hat{h}(u,v)}$$

and W must satisfy the following conditions:

1. $\mid \hat{W}(u,v) \mid \leq 1,$ for any $\nu > 0$
2. $\lim_{(u,v)\rightarrow(0,0)} \hat{W}(u,v) = 1$ for any (u,v) such that $\hat{h}(u,v) \neq 0$.
3. $\hat{W}(u,v)/\hat{h}(u,v)$ bounded for any (u,v)

Any function sastifying these three conditions defines a regularized linear solution.

Most Used Windows

$\nu = \sqrt{u^2 + v^2}$

- Truncated window function: $\hat{W}(u,v) = \begin{cases} 1 & \text{if } \mid \hat{h}(u,v) \mid \geq \sqrt{\epsilon} \\ 0 & otherwise \end{cases}$ where ϵ is the regularization parameter.
- Rectangular window: $\hat{W}(u,v) = \begin{cases} 1 & \text{if } \mid \nu \mid \leq \Omega \\ 0 & otherwise \end{cases}$ where Ω defines the band-width.
- Triangular window: $\hat{W}(u,v) = \begin{cases} 1 - \frac{\nu}{\Omega} & \text{if } \mid \nu \mid \leq \Omega \\ 0 & otherwise \end{cases}$
- Hanning Window: $\hat{W}(u,v) = \begin{cases} \cos(\frac{\pi\nu}{\Omega}) & \text{if } \mid \nu \mid \leq \Omega \\ 0 & otherwise \end{cases}$
- Gaussian Window: $\hat{W}(u,v) = \begin{cases} \exp(-4.5\frac{\nu^2}{\Omega^2}) & \text{if } \mid \nu \mid \leq \Omega \\ 0 & otherwise \end{cases}$

Most Used Windows

Linear regularized methods have several advantages:

- very fast
- the noise in the solution can easily be derived from the noise in the data and the window function. For example, if the noise in the data is Gaussian with a standard deviation σ_d, the noise in the solution if $\sigma_s^2 = \sigma_d^2 \sum W_k^2$. This noise estimation does however not take into account the errors relative to the inaccurate knowledge of the PSF, which limits its interest in practice.

Linear regularized methods presents also several drawbacks

- Creation of Gibbs oscillations in the neighborhood of the discontinuities contained in the data. The visual quality is therefore degraded.
- No a priori information can be used. For example, negative values can exist in the solution, while in most cases, we know that it must positive.
- As the window function is a low-pass filter, the resolution is degraded. There is trade-off between the resolution we want to achieve and the noise level in the solution. Other methods, such wavelets-based methods, do not have such a constraint.

CLEAN decomposes an image into a set of diracs. We get

- a set
$$\delta_c = \{A_1\delta(x - x_1, y - y_1), \ldots, A_n\delta(x - x_n, y - y_n)\}$$
- a residual R.

The deconvolved image is:

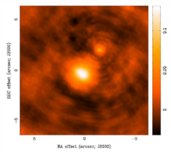
$$X(x, y) = \delta_c * B(x, y) + R(x, y)$$

where B is the clean beam.

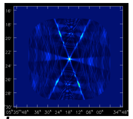
CLEAN

- *Optimal on point sources*
- *Iterative PSF subtraction from the dirty map*

Basic Algorithm

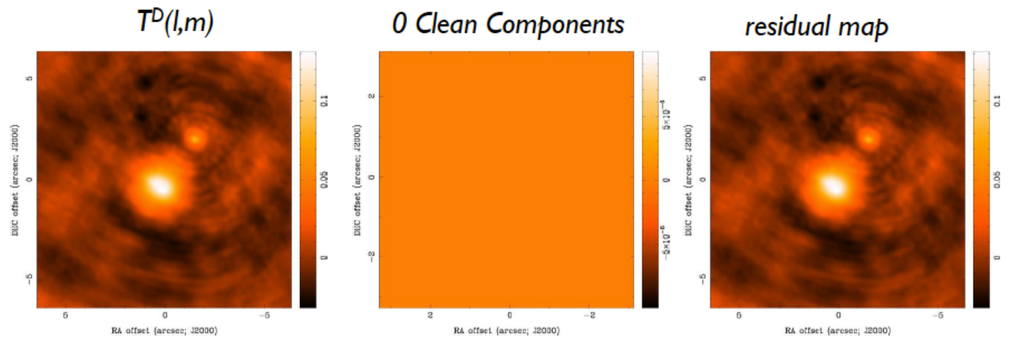


initialize:
i) *residual map* = *dirty map*
ii) *Clean Component* list = 0



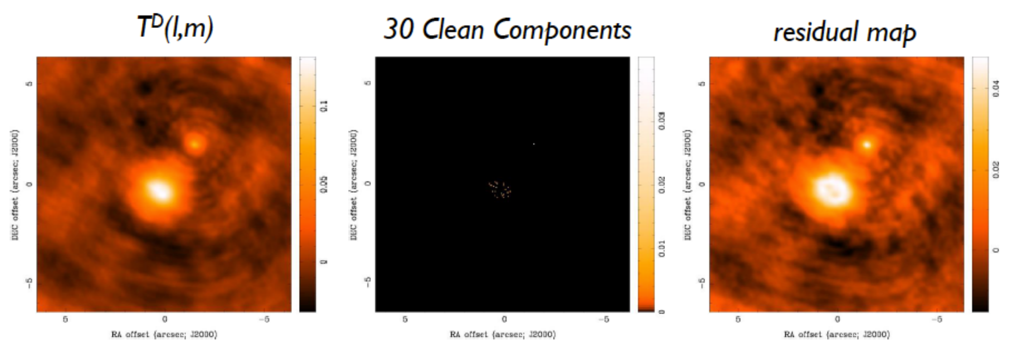
1. *identify the highest peak in the residual map as a point source*
2. *subtract a fraction of this peak from the residual map using a scaled dirty beam s(l,m) x gain*
3. *add this point source location and amplitude to the Clean Component list*
4. *goto step 1 (an iteration) unless stopping criterion reached*

CLEAN RUNNING



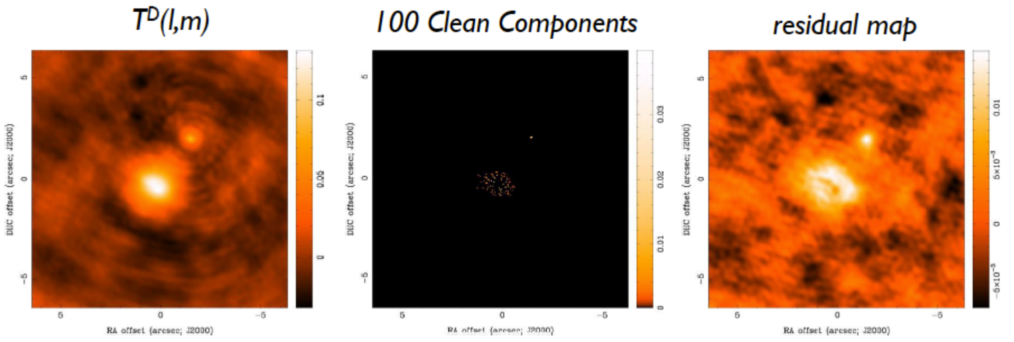
Stolen from D. Wilner presentation

CLEAN RUNNING



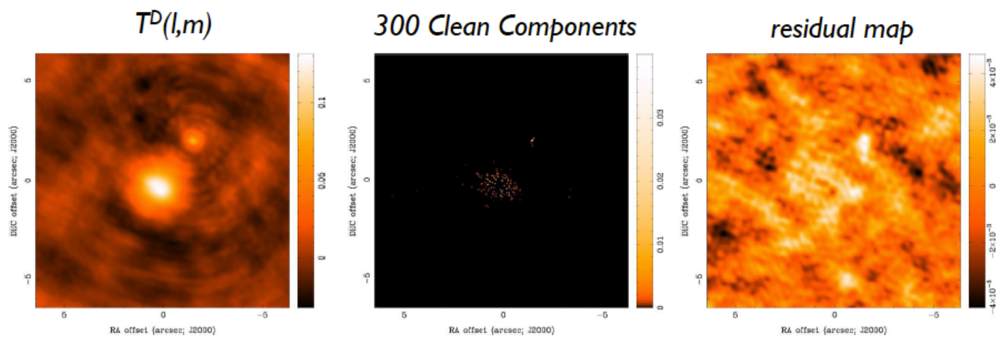
Stolen from D. Wilner presentation

CLEAN RUNNING



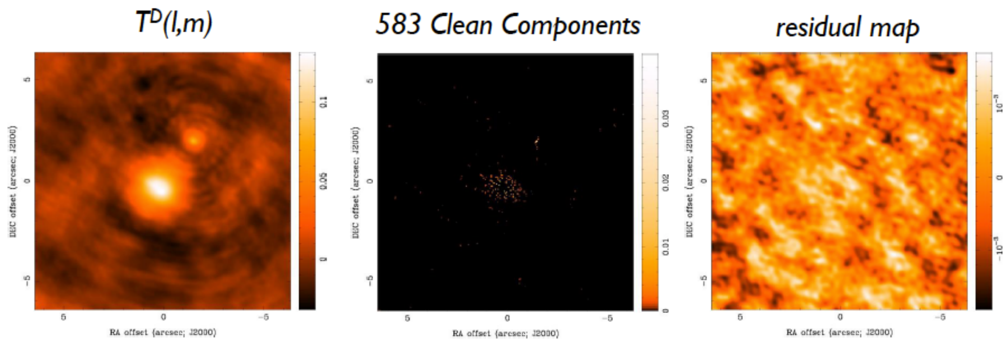
Stolen from D. Wilner presentation

CLEAN RUNNING



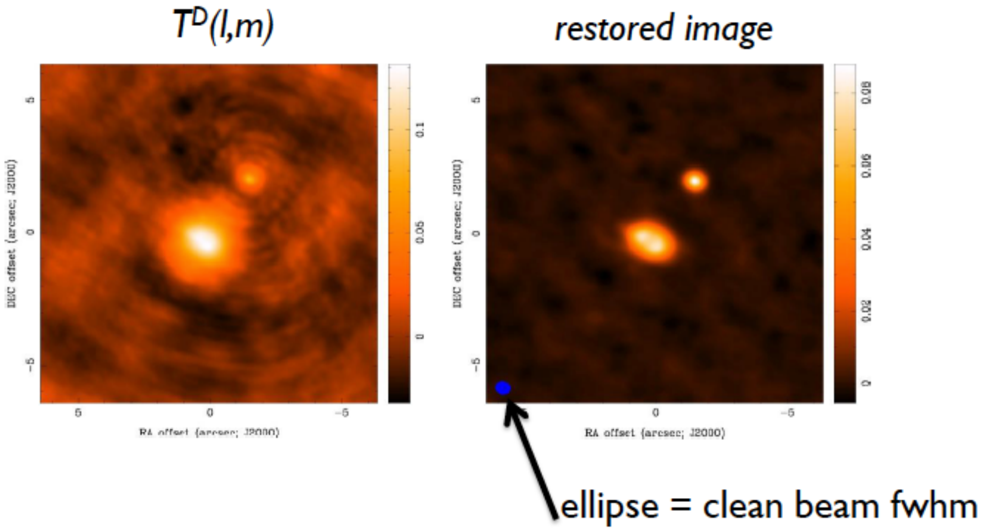
Stolen from D. Wilner presentation

CLEAN RUNNING



Stolen from D. Wilner presentation

CLEAN RUNNING



Stolen from D. Wilner presentation

Bayesian methodology

The Bayesian approach consists to construct the conditional probability density relationship:

$$p(X/Y) = \frac{p(Y/X)p(X)}{p(Y)}$$

The Bayes solution is found by maximizing the right part of the equation. The maximum likehood solution (ML) maximizes only the density $p(Y/X)$ over X:

$$ML(X) = \max_X p(Y/X)$$

The maximum-a-posteriori solution (MAP) maximizes over X the product $p(Y/X)p(X)$ of the ML and a prior:

$$MAP(X) = \max_X p(Y/X)p(X)$$

$p(Y)$ is considered as a constant value which has no effect in the maximization processus, and is neglected. The ML solution is equivalent to the MAP solution assuming an uniform density probability for $p(X)$.

Log-Likehood Function

$$MAP(X) = \max_X p(Y/X)p(X)$$

It is generally useful in practice log-likehood function, and we minimize:

$$J(X) = \min_X - \log p(Y/X)p(X)$$

$$J(X) = \min_X - \log p(Y/X) - \log p(X)$$

Maximum Likehood with Gaussian Noise

The probability $p(Y/X)$ is

$$p(Y/X) = \frac{1}{\sqrt{2}\pi\sigma_n} \exp - \frac{(Y - HX)^2}{2\sigma_N^2}$$

and maximizing $p(X/Y)$ is equivalent to minimize

$$J(X) = \frac{\| Y - HX \|^2}{2\sigma_n^2}$$

Using the steepest descent minimization method, a typical iteration is

$$X^{n+1} = X^n + \gamma(Y - H^t X^n)$$

The solution can also be found directly using the FFT by

$$\hat{X}(u,v) = \frac{\hat{h}^*(u,v)\hat{Y}(u,v)}{\hat{h}^*(u,v)\hat{h}(u,v)}$$

Wiener

If the object and the noise are assumed to follow Gaussian distributions with zero mean and variance respectively equal to σ_X and σ_N, then Bayes solution leads to the Wiener filter solution

$$\hat{X}(u,v) = \frac{\hat{h}^*(u,v)\hat{Y}(u,v)}{\mid \hat{h}(u,v) \mid^2 + \frac{\sigma_N^2(u,v)}{\sigma_X^2(u,v)}}$$

Maximum Likehood with Poisson noise

$$p(Y/X) = \prod_k \frac{(HX)_k^{Y_k} \exp -(HX)_k}{Y_k!}$$

The maximum can be computed by derivating the logarithm:

$$\frac{\partial \ln p(Y/X)}{\partial X} = 0$$

which leads to the result (assuming the PSF is normalized to the unity)

$$\frac{Y}{H^t X} H^t = 1$$

Multiplying both side by X_k

$$X_k = [\frac{Y_k}{(HX)_k} H^t] X_k$$

and using the Picard iteration leads to

$$X_k^{n+1} = [\frac{Y}{HX^n} H^t]_k X_k^n$$

it is the Richardson-Lucy algorithm.

Constraints

We assume now that there exists a general operator, $\mathcal{P}_{\mathcal{C}}(.)$, which enforces a set of constaints on a given object X, such that if X satisfies all the constraints, we have:

$$X = \mathcal{P}_{\mathcal{C}}(X)$$

The main used constraints are:

- Positivity: the object must be positive. $\mathcal{P}_{C_p}(X(x,y)) = \begin{cases} X(x,y) & \text{if } X(x,y) \geq 0 \\ 0 & otherwise \end{cases}$
- Support constraint: the objects belongs to a given spatial domain $\mathcal{D}$.

$$\mathcal{P}_{C_s}(X(x,y)) = \begin{cases} X(x,y) & \text{if } (x,y) \in \mathcal{D} \\ 0 & otherwise \end{cases}$$

- Band-limited: the Fourier transform of the object belongs to a given frequency domain. For instance, if F_c is the cut-off frequency of the instrument, we want to impose the object to be band-limited: $\mathcal{P}_{C_f}(\hat{X}_\nu) = \begin{cases} \hat{X}_\nu & \text{if } \nu < F_c \\ 0 & otherwise \end{cases}$

These constraints can be incorporated easily in the basic iterative scheme.

Iterative Regularized Methods

- Landweber:

$$X^{n+1} = \mathcal{P}_C[X^n + \mu H^t(Y - HX^n)]$$

- Richardon Lucy Method:

$$X^{n+1} = \mathcal{P}_C[X^n[\frac{Y}{HX^n}H^t]]$$

- Tikhonov: Tikhonov solution:

$$\nabla(J_T(X)) = H^tHX + \mu F^t * FX - H^tY$$

and applying the following iteration:

$$X^{n+1} = X^n - \gamma\nabla(J_T(X))$$

The constraint Tikhonov solution is therefore obtained by:

$$X^{n+1} = \mathcal{P}_C[X^n - \gamma\nabla(J_T(X))]$$

Maximum Entropy Method (MEM)

In the absence of any information on the solution X except its positivity, a possible course of action is to derive the probability of X from its entropy, which is defined from information theory. Then if we know the entropy E of the solution, we derive its probability by

$$\mathrm{p}(X) = \exp(-\lambda E(X))$$

Given the data, the most probable image is obtained by maximizing $\mathrm{p}(X|Y)$. We need to minimize

$$\log \mathrm{p}(X|Y) = -\log \mathrm{p}(Y|X) + \lambda E(X) - \log \mathrm{p}(Y)$$

The last term is a constant and can be omitted.

MEM and Gaussian Noise

Then, in the case of Gaussian noise, the solution is found by minimizing

$$J(X) = \sum_{pixels} \frac{(Y - HX)^2}{2\sigma^2} + \lambda E(X) = \frac{\chi^2}{2} + \lambda E(X)$$

which is a linear combination of two terms: the entropy of the signal, and a quantity corresponding to χ^2 in statistics measuring the discrepancy between the data and the predictions of the model. λ is a parameter that can be viewed alternatively as a Lagrangian parameter or a value fixing the relative weight between the goodness-of-fit and the entropy E.

Information Theory

The main idea of information theory (Shannon, 1948) is to establish a relation between the received information and the probability of the observed event

- The information is a decreasing function of the probability. This implies that the more information we have, the less will be the probability associated with one event.
- Additivity of the information. If we have two independent events E_1 and E_2, the information $\mathcal{I}(E)$ associated with the happening of both is equal to the addition of the information of each of them.

$$\mathcal{I}(E) = k \ln(p)$$

where k is a constant. Information must be positive, and k is generally fixed at -1.

Other Entropy Functions

- Burg (1967)

$$E_b(X) = - \sum_{pixels} \ln(X)$$

- Frieden (1975)

$$E_f(X) = - \sum_{pixels} X \ln(X)$$

- Gull and Skilling (1984)

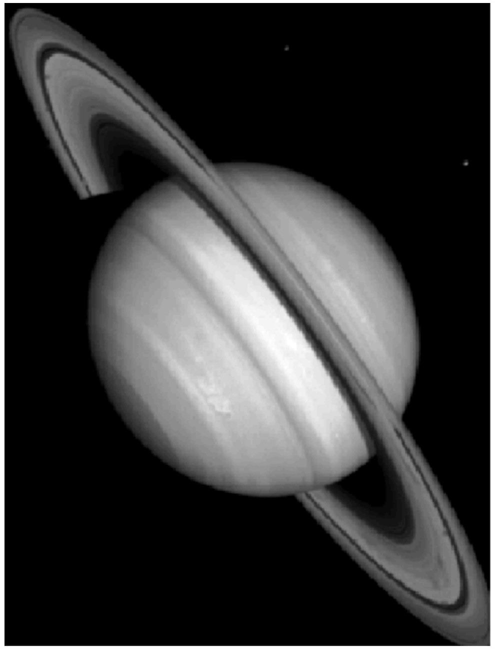
$$E_g(X) = \sum_{pixels} X - M - X \ln(X|M)$$

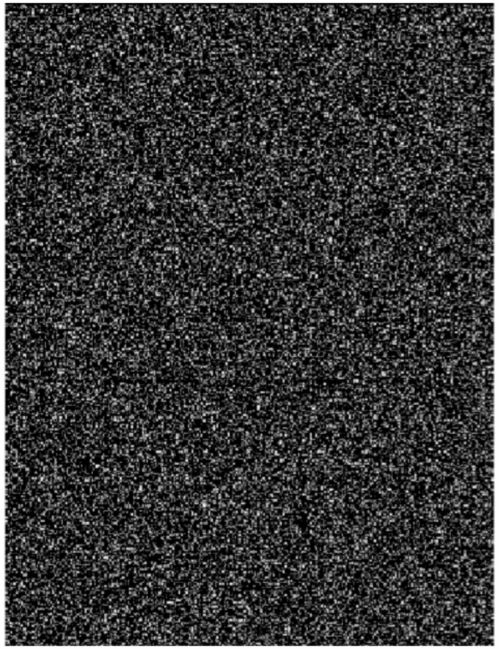
The last definition of the entropy has the advantage of having a zero maximum when X equals the model M, usually taken as a flat image.

Problems

- The entropy is maximum for a flat image, and decreases when when we have some fluctuations.
- The results varied strongly with the background level (Narrayan, 1986).
- Adding a value at a given pixel of a flat image does't furnish the same information that subtracting it. A consequence of this is that absorption features (under the background level) are poorly reconstructed (Narrayan, 1986).
- Gull and Skilling entropy presents the difficulty of estimating a model. Furthermore it has been shown (Bontekoe et al, 1994) that the solution was dependent on this choice.
- a value of λ which is too large gives a resulting image which is too regularized with a large loss of resolution. A value which is too small leads to a poorly regularized solution showing unacceptable artifacts.

Which image contains more information ?





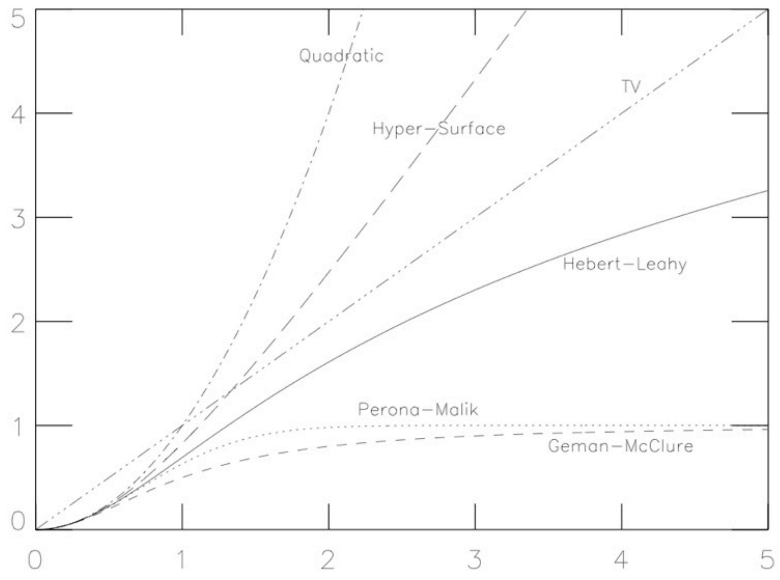
Penalized Gradients

Generally, functions ϕ are chosen with a quadratic part which ensures a good smoothing of small gradients (Green, 1990), and a linear behavior which cancels the penalization of large gradients (Bouman and Sauer, 1993):

1. $\lim_{t \to 0} \frac{\phi^{'}(t)}{2t} = 1$, smooth faint gradiants.
2. $\lim_{t \to \infty} \frac{\phi^{'}(t)}{2t} = 0$, preserve strong gradiants.
3. $\frac{\phi^{'}(t)}{2t}$ is strictly decreasing.

Such functions are often called L_2-L_1 functions.

Penalized Gradients



Conclusions on Part 1

DECONVOLUTION METHODS IN ASTRONOMY

Wiener

Richardson Lucy method → Noise amplification

Maximum Entropy Method → Problem to restore point sources, bias, etc

CLEAN Method → Problem to restore extended sources

SIGNAL PROCESSING DOMAIN

Markov Random Field, TV

Deconvolution

The image formation is expressed in the convolution integral

$$\begin{aligned} Y(x,y) &= \int_{x_1=-\infty}^{+\infty} \int_{y_1=-\infty}^{+\infty} h(x-x_1, y-y_1) X(x_1,y_1) dx_1 dy_1 + N(x,y) \\ &= (h * X)(x,y) + N(x,y) = HX + N \end{aligned}$$

where Y is the data, H the point-spread-function (PSF), and X is the solution.

In Fourier space we have:

$$\hat{Y}(u,v) = \hat{h}(u,v)\hat{X}(u,v) + \hat{N}(u,v)$$

We want to determine X knowing h and X. The main difficulties are the existence of:

- a cut-off frequency of the point spread function.
- the noise.

It is in fact an **ill posed problem**, there is not an unique solution.

Inverse Problems in Astrophysics

- Part 1: Introduction inverse problems and image deconvolution
- **Part 2: Introduction to Sparsity and Compressed Sensing**
- Part 3: Wavelets in Astronomy: from orthogonal wavelets and to the Starlet transform.
- Part 4: Beyond Wavelets
- Part 5: Inverse problems and their solution using sparsity: denoising, deconvolution, inpainting, blind source separation.
- Part 6: CMB & Sparsity
- Part 7: Perspective of Sparsity & Compressed Sensing in Astrophsyics

Entering the 21th Century

==> **paradigm shift** in statistics/signal processing:

20th century

Shannon Nyquist sampling + band limited signals + linear l2 norm regularization

21st century

Compressed Sensing + sparse signals + non-linear l0-l1 norm regularization

Weak Sparsity or Compressible Signals

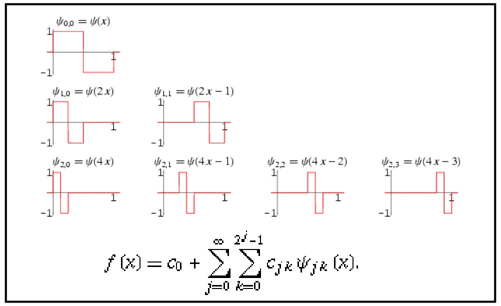
A signal *s* (*n* samples) can be represented as sum of weighted elements of a given dictionary

Dictionary (basis, frame)

$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

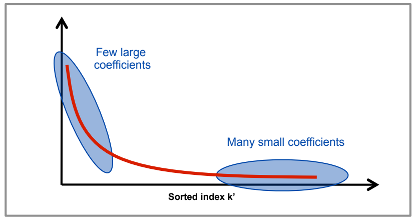
Atoms

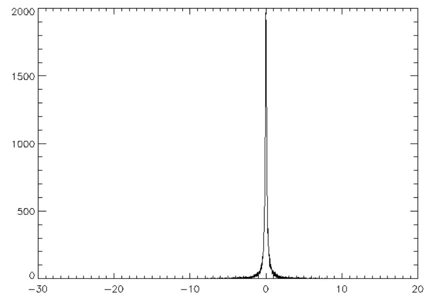
Ex: Haar wavelet



$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

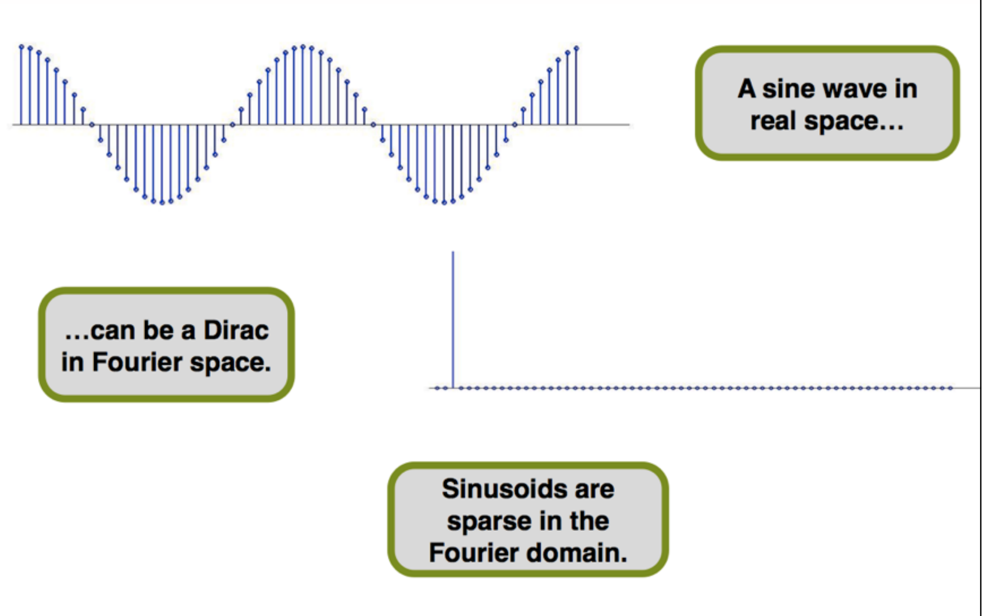
coefficients



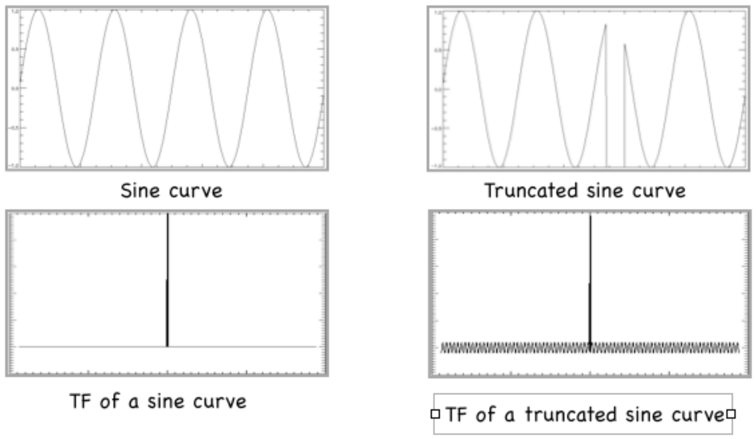


- Fast calculation of the coefficients
- Analyze the signal through the statistical properties of the coefficients
- Approximation theory uses the sparsity of the coefficients

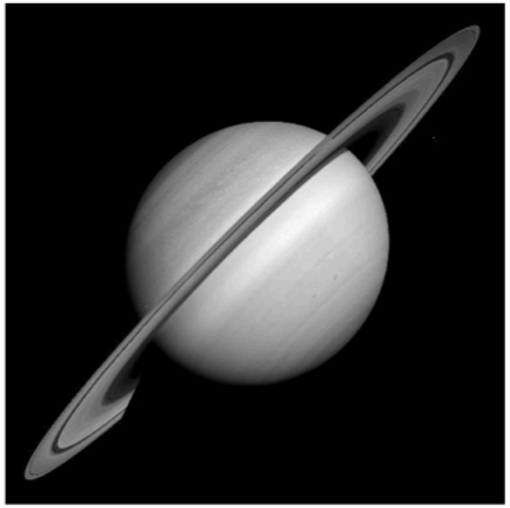
Strict Sparsity: k-sparse signals



Minimizing the l_0 norm

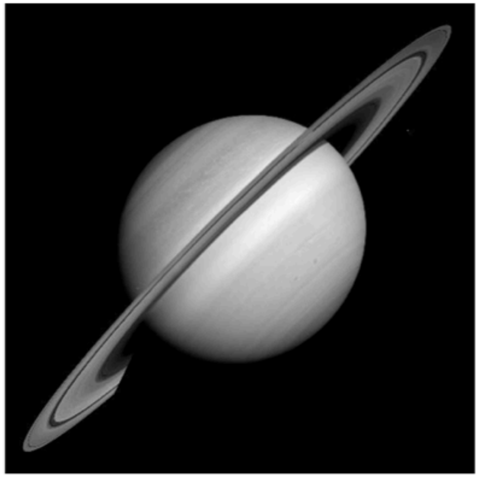


with $0^0 = 0$, $$\| \alpha \|_0 = \sum_k \alpha_k^0 = \# \{\alpha_k \neq 0\}$$



The top 1% of the coefficients concentrate only 8.66% of the energy. Not sparse…

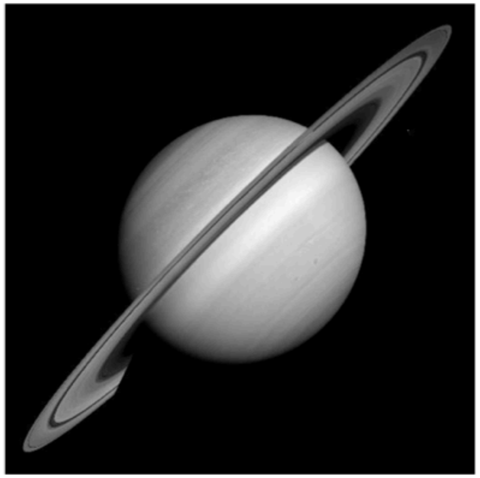
1% largest coefficients in real space (the others are set to 0)



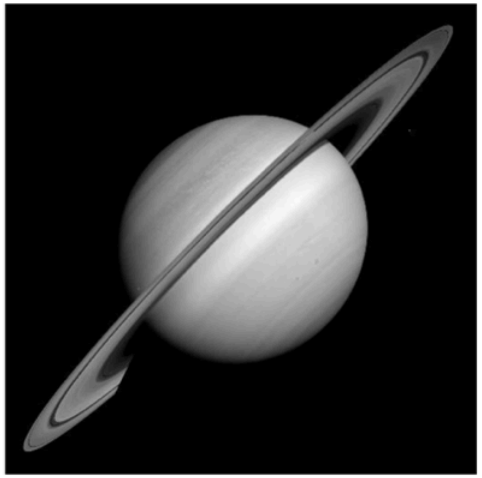
1% largest coefficients in wavelet space
(the others are set to 0)

The wavelet coefficients encode edges and large scale information.

Wavelet transform



1% of the wavelet coefficients concentrate 99.96% of the energy: This can be used as a *prior.*

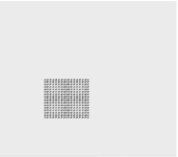
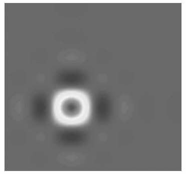
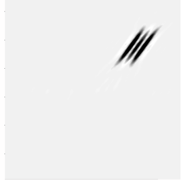


Reconstruction, after throwing away 99% of the wavelet coefficients

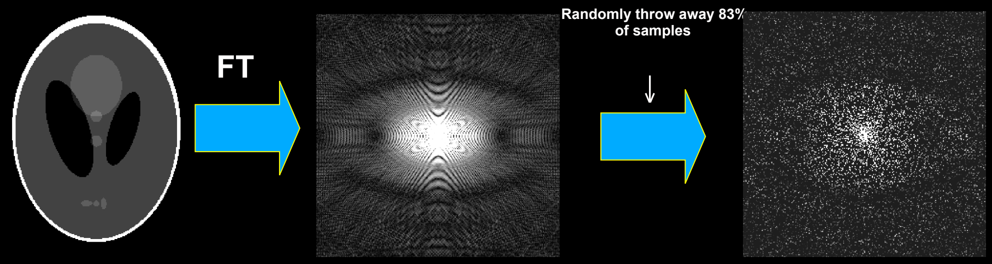
Sparsity Model 1: we consider a dictionary which has a fast transform/reconstruction operator:

$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

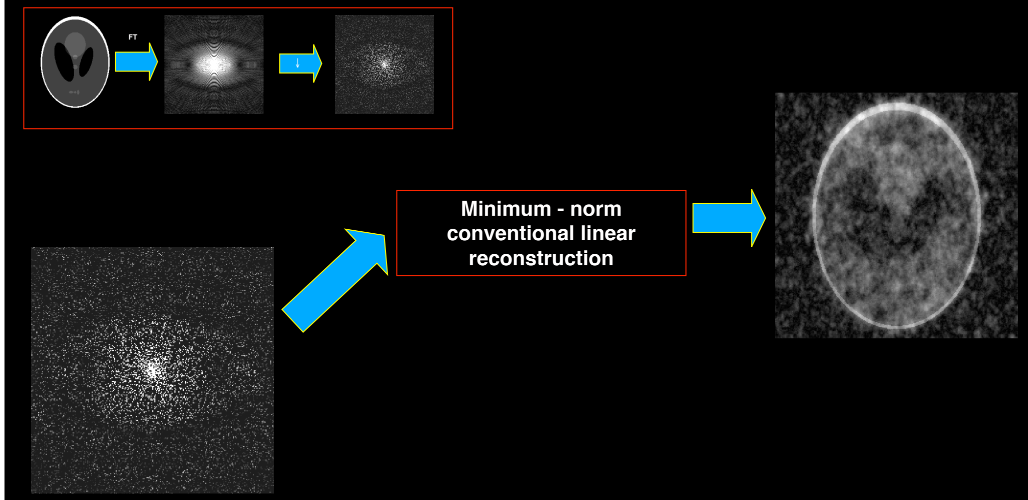
Local DCT	Stationary textures Locally oscillatory	
Wavelet transform	Piecewise smooth Isotropic structures	
Curvelet transform	Piecewise smooth, edge	

A Surprising Experiment*



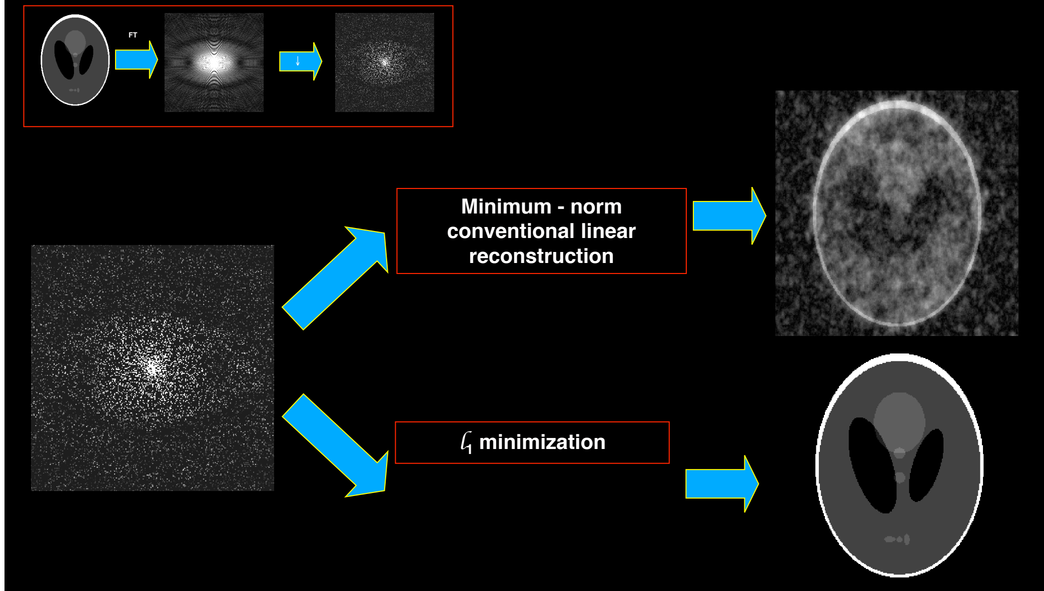
* E.J. Candes, J. Romberg and T. Tao.

A Surprising Result*



* E.J. Candes, J. Romberg and T. Tao.

A Surprising Result*



E.J. Candes

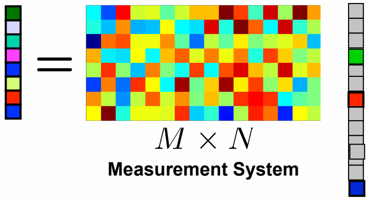
Compressed Sensing

* E. Candès and T. Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? ", IEEE Trans. on Information Theory, 52, pp 5406-5425, 2006.
* D. Donoho, "Compressed Sensing", IEEE Trans. on Information Theory, 52(4), pp. 1289-1306, April 2006.
* E. Candès, J. Romberg and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, Feb. 2006.

A non linear sampling theorem

"Signals with exactly K components different from zero can be recovered perfectly from ~ K log N incoherent measurements"

Replace samples with few linear projects: Y = H X



$$K < M \leq N$$

Reconstruction via non linear processing:

$$\min_X \|X\|_1 \quad \text{s.t.} \quad Y = HX$$

Compressed Sensing Reconstruction

Measurements: $y_k = \left\langle X, H_k \right\rangle$

Reconstruction via non linear processing: $\min_X \|X\|_1 \;\; \text{s.t.} \;\;\; Y = HX$

In practice, X is sparse in a given **dictionary**: $X = \Phi\alpha$

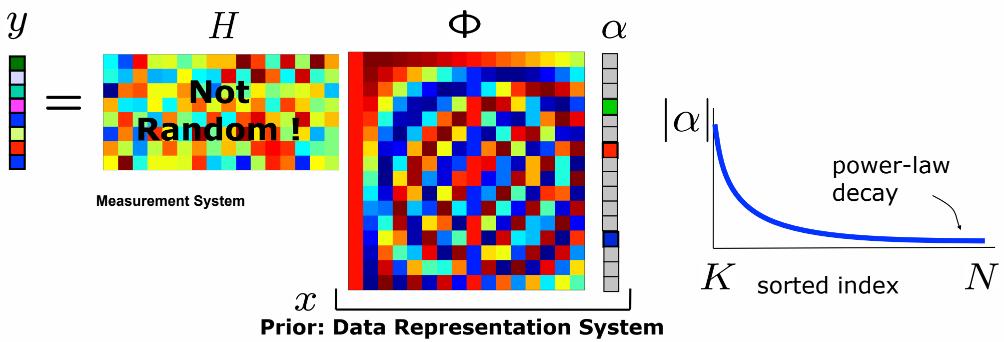
and we need to solve: $\min_\alpha \|\alpha\|_1 \;\; \text{s.t.} \;\;\; Y = H\Phi\alpha$

The mutual incoherence is defined as $\mu_{H,\Phi} = \sqrt{N} \max_{i,k} \left| \left\langle H_i, \Phi_k, \right\rangle \right|$

the number of required measurements is : $m \geq C\mu^2_{H,\Phi} K \log n$

Soft Compressed Sensing Definition

$$Y = HX = H\Phi\alpha$$

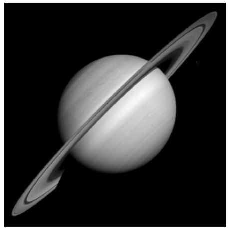


Mutual coherence: $\mu_{H,\Phi} = \max_{i,k} \left| \left\langle H_i, \Phi_k, \right\rangle \right|$

Mutual coherence the degree of similarity between the sparsity and measurement systems.

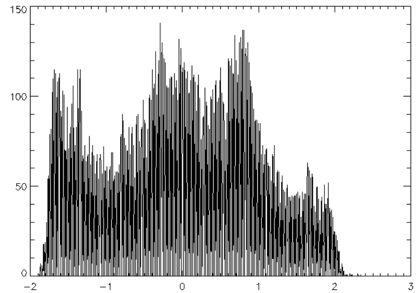
Reconstruction via non linear processing: $\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad Y = H\Phi\alpha$

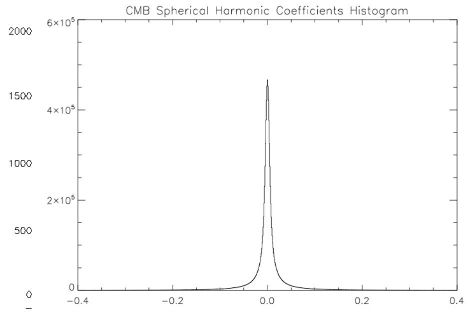
Weak Sparsity or Compressible Signals



Direct Space

Curvelet Space





How to measure sparsity ?

$$\text{with } 0^0 = 0, \quad \| \alpha \|_0 = \sum_k \alpha_k^0 = \# \{\alpha_k \neq 0\}$$

Formally, the sparsest coefficients are obtained by solving the optimization problem:

(P0) Minimize $\|\alpha\|_0$ subject to $s = \phi\alpha$

It has been proposed (*to relax and*) to replace the l_0 norm by the l_1 norm (Chen, 1995):

(P1) Minimize $\|\alpha\|_1$ subject to $s = \phi\alpha$

It can be seen as a kind of convexification of (P0).

It has been shown (Donoho and Huo, 1999) that for certain dictionary, if there exists a highly sparse solution to (P0), then it is identical to the solution of (P1).

==> Link the sparsity and the sampling through the Compressed Sensing.

INVERSE PROBLEMS AND SPARSE RECOVERY

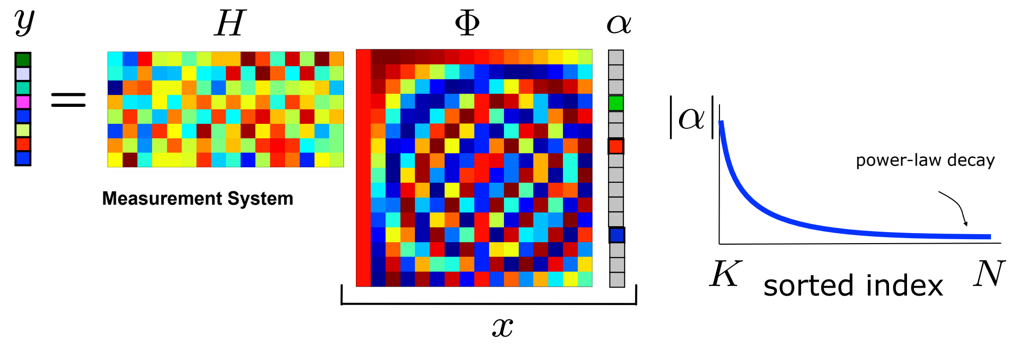
$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse

- **Denoising**
- **Deconvolution**
- **Component Separation**
- **Inpainting**
- **Blind Source Separation**
- **Minimization algorithms**
- **Compressed Sensing**

$$\min_{\alpha} \|\alpha\|_p^p \quad \text{subject to} \quad \|Y - H\Phi\alpha\|^2 \leq \epsilon$$

Very efficient recent methods now exist to solve it (proximal theory)



Denoising using a sparsity model

$$Y = X + N$$

Denoising using a sparsity prior on the solution:

X is sparse in Φ, i.e. $X = \Phi\alpha$ where most of α are negligible.

$$\tilde{\alpha} \in \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + t \parallel \alpha \parallel_p^p, \quad 0 \leq p \leq 1.$$

p=0

$$\tilde{\alpha} \in \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + \frac{t^2}{2} \parallel \alpha \parallel_0$$

==> Solution via Iterative **Hard** Thresholding

$$\tilde{\alpha}^{(t+1)} = \mathrm{HardThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu\Phi^T(Y - \Phi\tilde{\alpha}^{(t)})), \mu = 1/\left\|\Phi\right\|^2.$$

$$\tilde{\alpha}_{j,k} = \mathrm{HardThresh}_t(\alpha_{j,k}) = \begin{cases} \alpha_{j,k} & \text{if } \mid \alpha_{j,k} \mid \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

1st iteration solution:

$$\tilde{X} = \Phi \ \mathrm{HardThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y)$$

Exact for Φ orthonormal.

***p*=1**

$$\tilde{\alpha} = \arg\min_{\alpha} \frac{1}{2} \| Y - \Phi\alpha \|^2 + t \| \alpha \|_1$$

==> Solution via iterative **Soft** Thresholding

$$\tilde{\alpha}^{(t+1)} = \mathrm{SoftThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu\Phi^T(Y - \Phi\tilde{\alpha}^{(t)})), \mu \in (0, 2/\left\|\Phi\right\|^2).$$

$$\tilde{\alpha}_{j,k} = \mathrm{SoftThresh}_t(\alpha_{j,k}) = \mathrm{sign}(\alpha_{j,k})(| \alpha_{j,k} | - t)_+$$

1st iteration solution:

$$\tilde{X} = \Phi\ \mathrm{SoftThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y)$$

Exact for Φ orthonormal.