Sparse modeling for image reconstruction in Astronomy

Makoto Uemura

(Hiroshima University)

HIRES2014@ESO Garching

Outline

- What is the “sparse modeling”?
  - Example: l1-norm reconstruction
- Applications
  1. Period analysis of variable stars
     - ✓ Kato & Uemura (2012)
  2. Image reconstruction of radio interferometer
     - ✓ Honma, Akiyama, Uemura, & Ikeda (2014)
  3. Doppler tomography
     - ✓ Uemura, Kato, Nogami, & Mennickent (in prep.)
  4. Gamma-ray Compton camera
     - ✓ Ikeda, Uemura+ (2014)
- Summary

Linear inverse problem

Data — Model (known) — Model parameters

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

$$\boldsymbol{y} = A\boldsymbol{x}$$

M=50 — N=100

- Estimate ***x*** from ***y*** (inverse-problem)
- M=50, N=100 (ill-posed problem)

Data | Model (known) | Model parameters

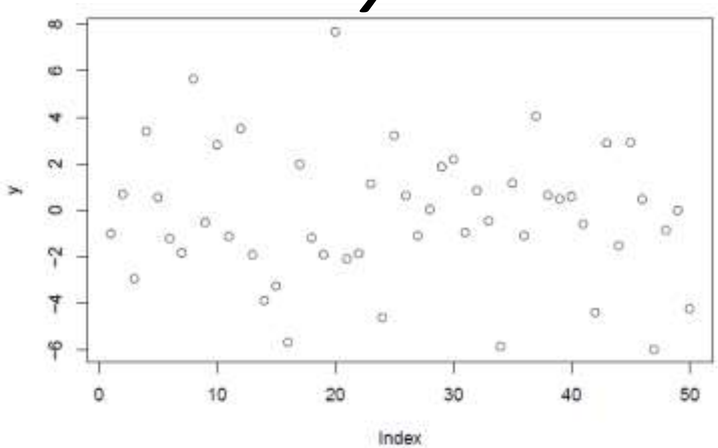
$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

$$\boldsymbol{y} = A\boldsymbol{x}$$

M=50

N=100

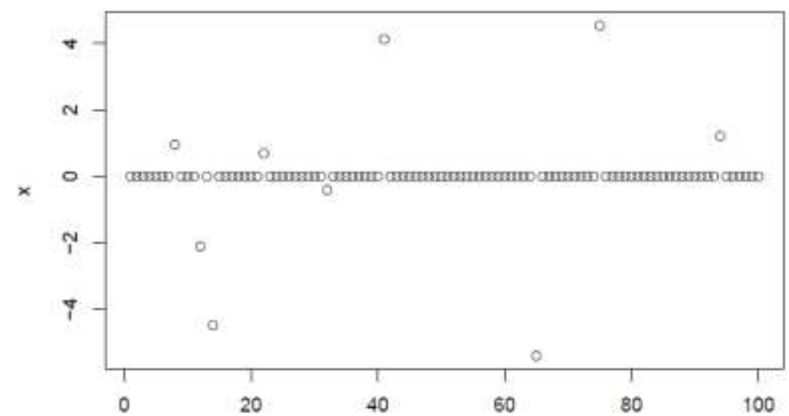
y



A

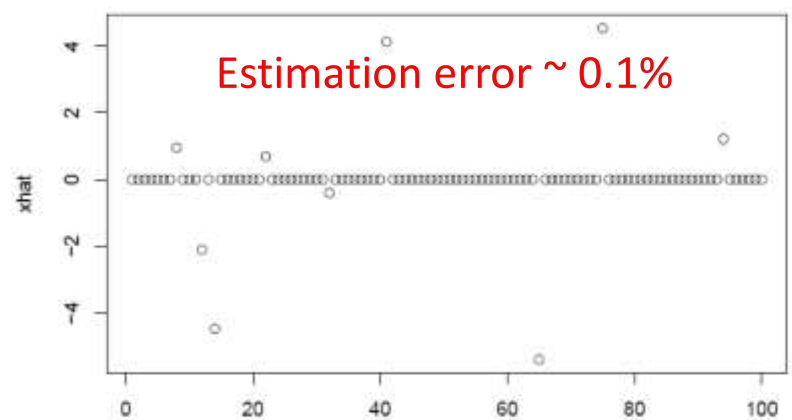
(random matrix)

X



$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|$$



Reconstruction of sparse vectors

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$

$$\|\boldsymbol{x}\|_1 = \sum_i |x_i|$$

- Sparse vectors can be reconstructed using l1-norm minimization
  - Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 96)
  - Compressed sensing (Candes+06; Donoho 06)

- Sparse modeling: effective use of the sparsity of information
  - “Initiative for High-Dimensional Data-Driven Science through Deepening of Sparse Modeling” P.I. M. Okada (U. Tokyo), JSPS grant “Scientific Research of Innovative Areas”, 2013-2018
  - “SparseAstro project” P.I. J. L. Starck (CEA), European Research Council grant, 2009-2014

Period analysis of variable stars

(Kato & Uemura, 2012, PASJ, 64, 122)

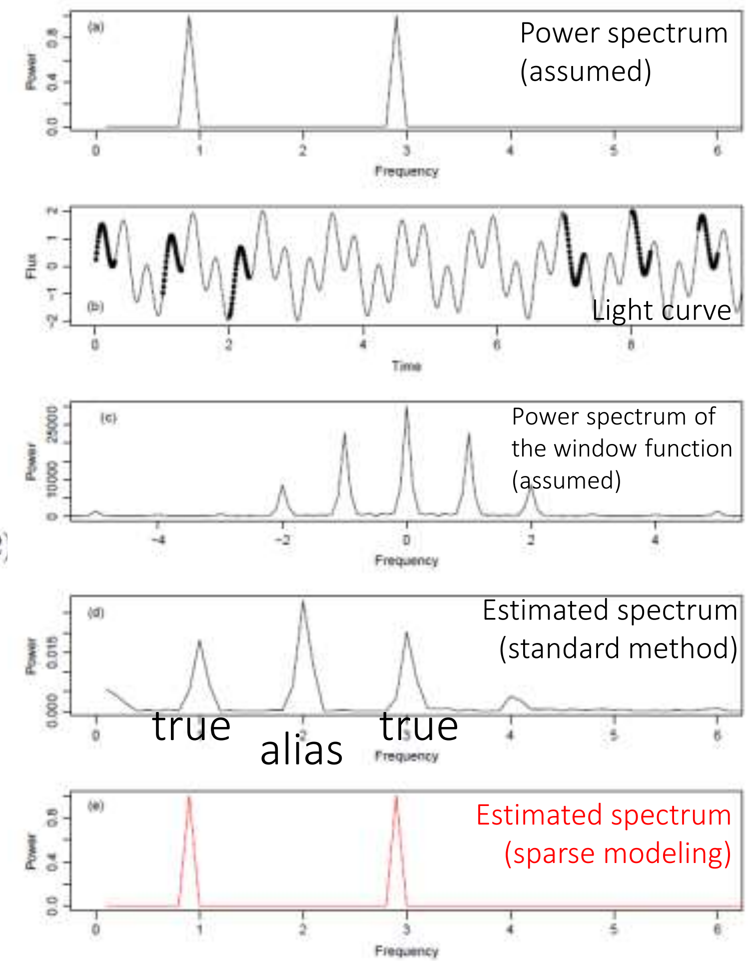
- Estimate spectra in the frequency domain from the light curve
- Un-evenly sampled data -> aliases
- Spectra are "sparse" in the case of pulsating variables"

Light curve

Spectrum

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} \cos(t_1\nu_1) & \cdots & \cos(t_1\nu_N) & \sin(t_1\nu_1) & \cdots & \sin(t_1\nu_N) \\ \cos(t_2\nu_1) & \cdots & \cos(t_2\nu_N) & \sin(t_2\nu_1) & \cdots & \sin(t_2\nu_N) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(t_M\nu_1) & \cdots & \cos(t_M\nu_N) & \sin(t_M\nu_1) & \cdots & \sin(t_M\nu_N) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \\ b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix} \quad (2)$$

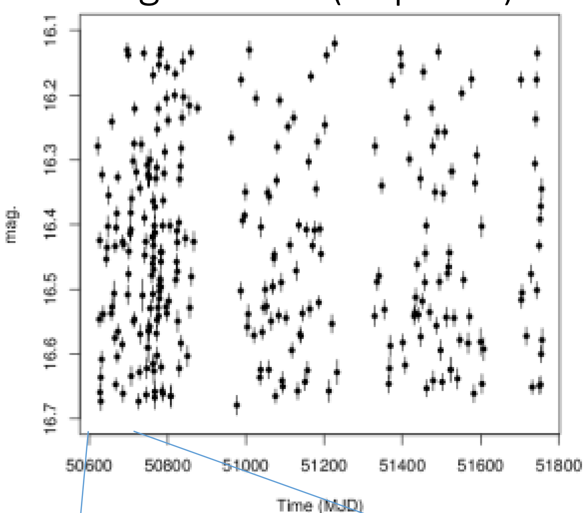
$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \mathcal{F}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$

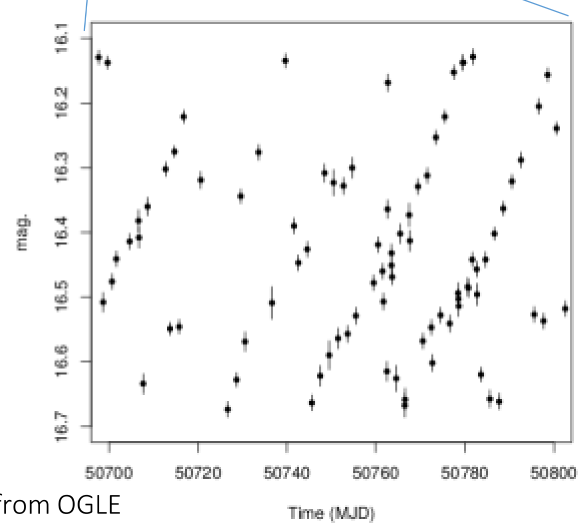


No aliases, no white noise, sharp signals.

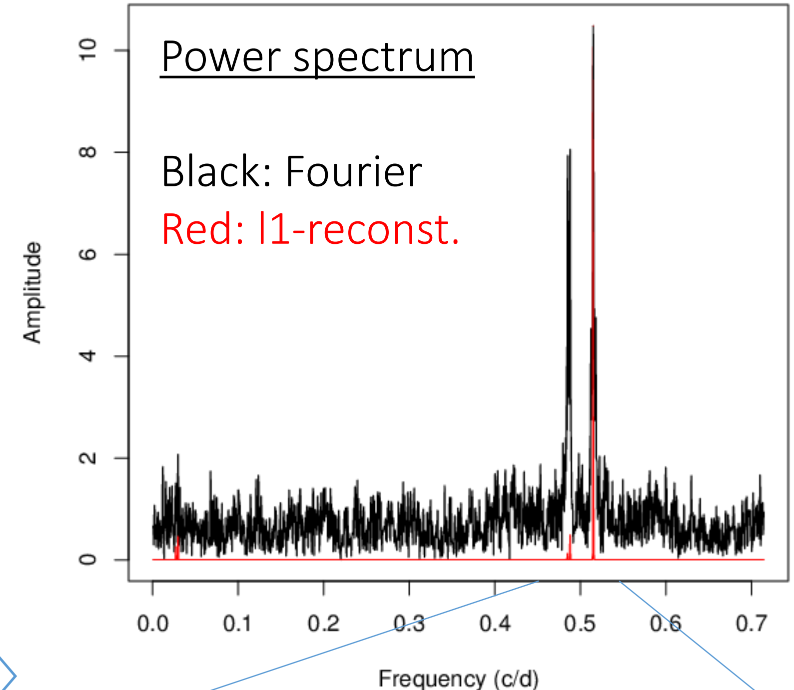
Period analysis: Example

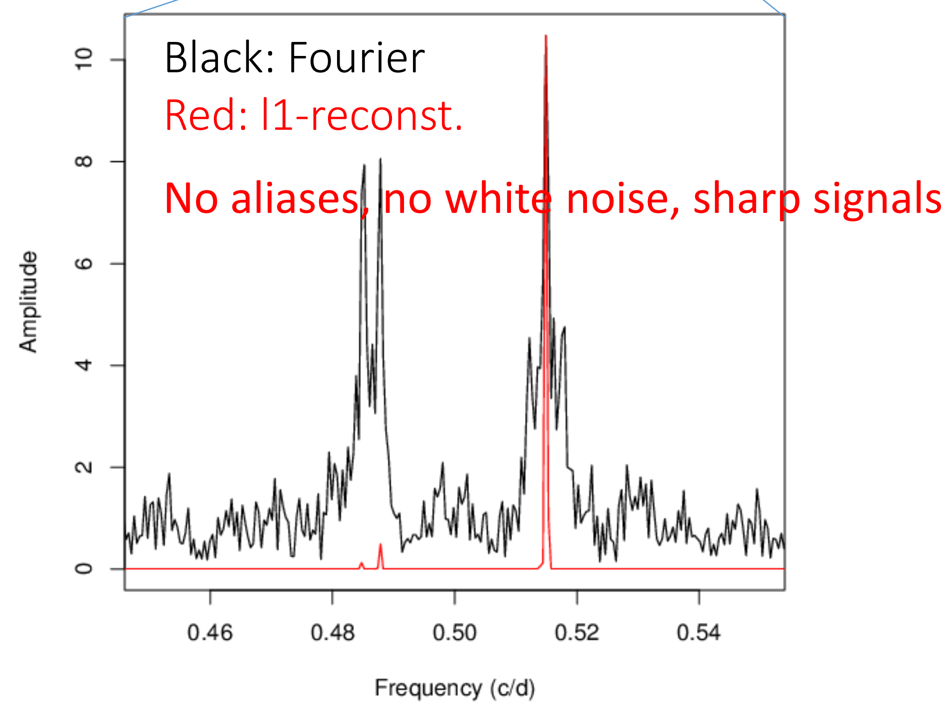
Light curve (Cepheid)





Data from OGLE

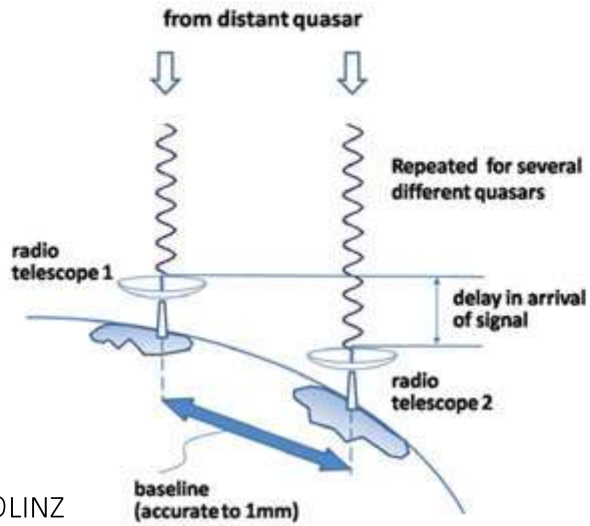




Radio interferometer

(Honma, Akiyama, Uemura, & Ikeda, 2014, PASJ, 66, 95)

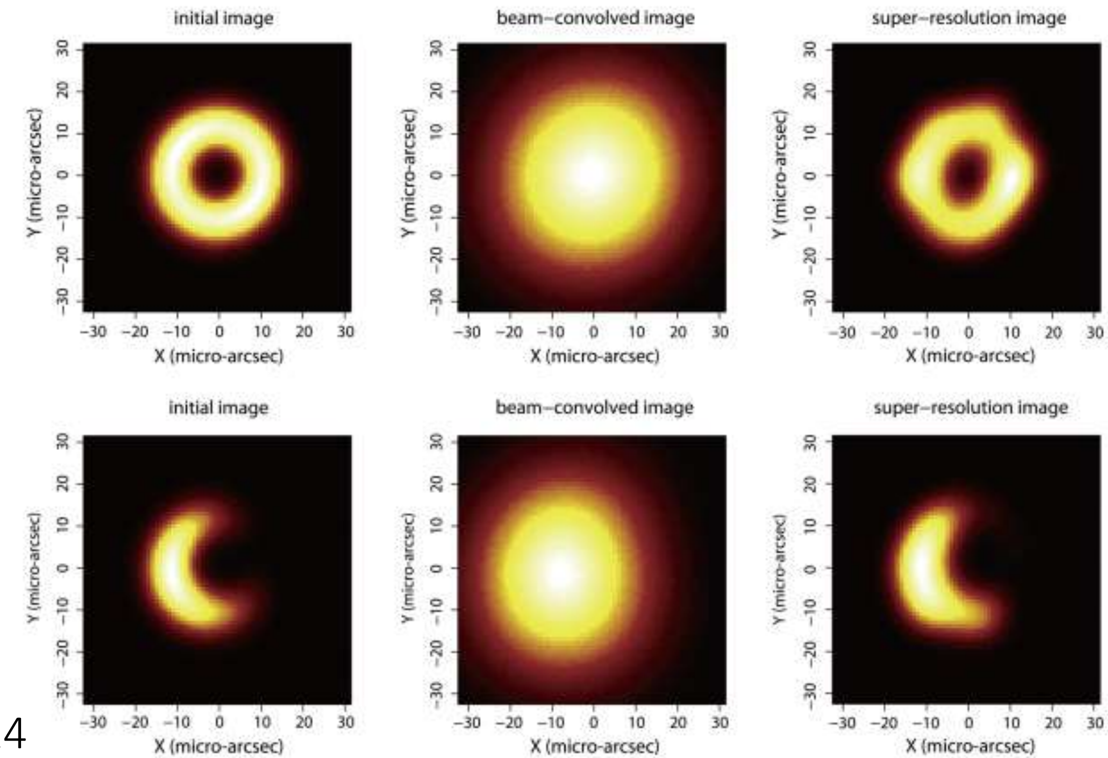
- Radio map ← 2D Fourier transform → Visibilities (Data)
- Radio maps are sparse in the case of compact sources.
- Examples: No. of data = 976, No. of pixels = 64x64=4096



©LINZ

Super-resolution

Assumed image | Beam-convolved | l1 reconstruction



Honma, MU+14

Sparsity in images

- Find bases for the sparse representation.

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - A\boldsymbol{x}||_2^2 + \lambda||\boldsymbol{x}||_1$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - A'\boldsymbol{x}||_2^2 + \lambda||B\boldsymbol{x}||_1$$

$$A' = AB^{-1}\boldsymbol{x}'$$

- Typical bases
  - Wavelet
  - Differential = total variation (TV)

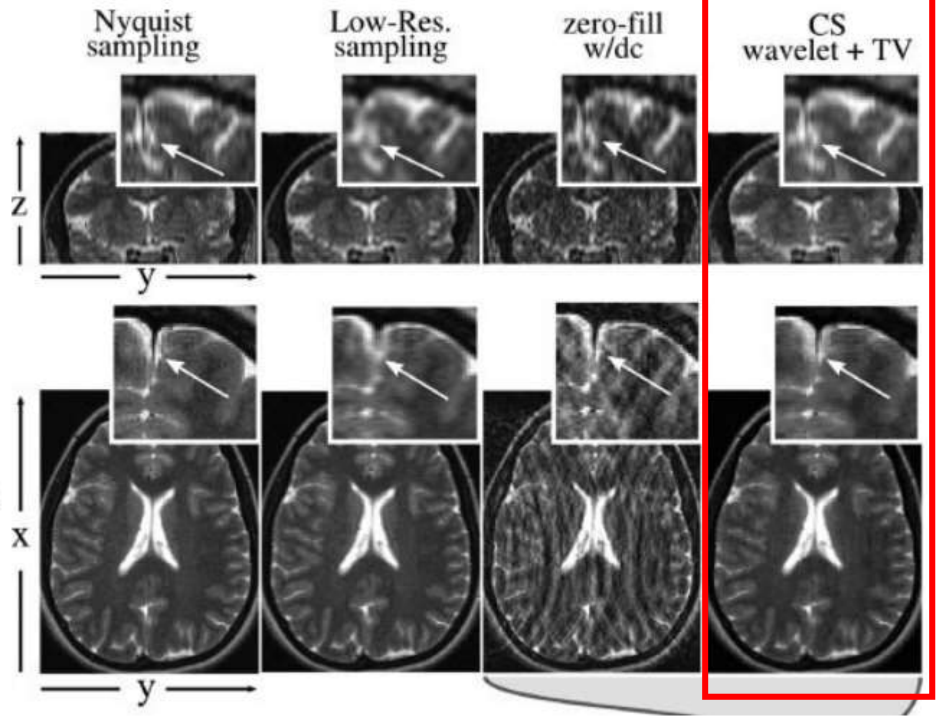
- Example: MRI image reconstruction

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{y} - A\boldsymbol{x}||_2^2 + \lambda_1||W\boldsymbol{x}||_1 + \lambda_2 D_1(\boldsymbol{x}) + \lambda_3||\boldsymbol{x}||_1$$

Matrix for total variation (TV)

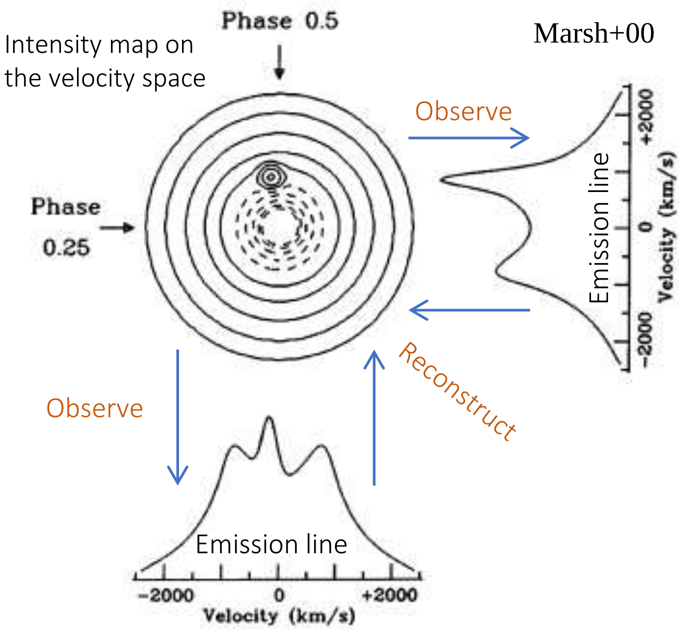
$$D_1 = \begin{pmatrix} +1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & +1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & +1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & +1 \end{pmatrix}$$

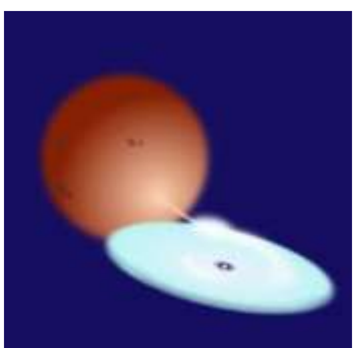
Example of MRI reconstruction Lustig+07



Doppler tomography with TVM

(Uemura, Kato, Nogami, & Mennickent, 2014, PASJ, submitted)





$$\hat{\boldsymbol{x}} = argmin \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} \right\|_2^2 + \lambda f(\boldsymbol{x})$$

Data — Observation Matrix — image

- Total Variation Minimization (TVM)

$$TV(\boldsymbol{x}) = \sum \sqrt{(\Delta^h \boldsymbol{x})^2 + (\Delta^v \boldsymbol{x})^2}$$

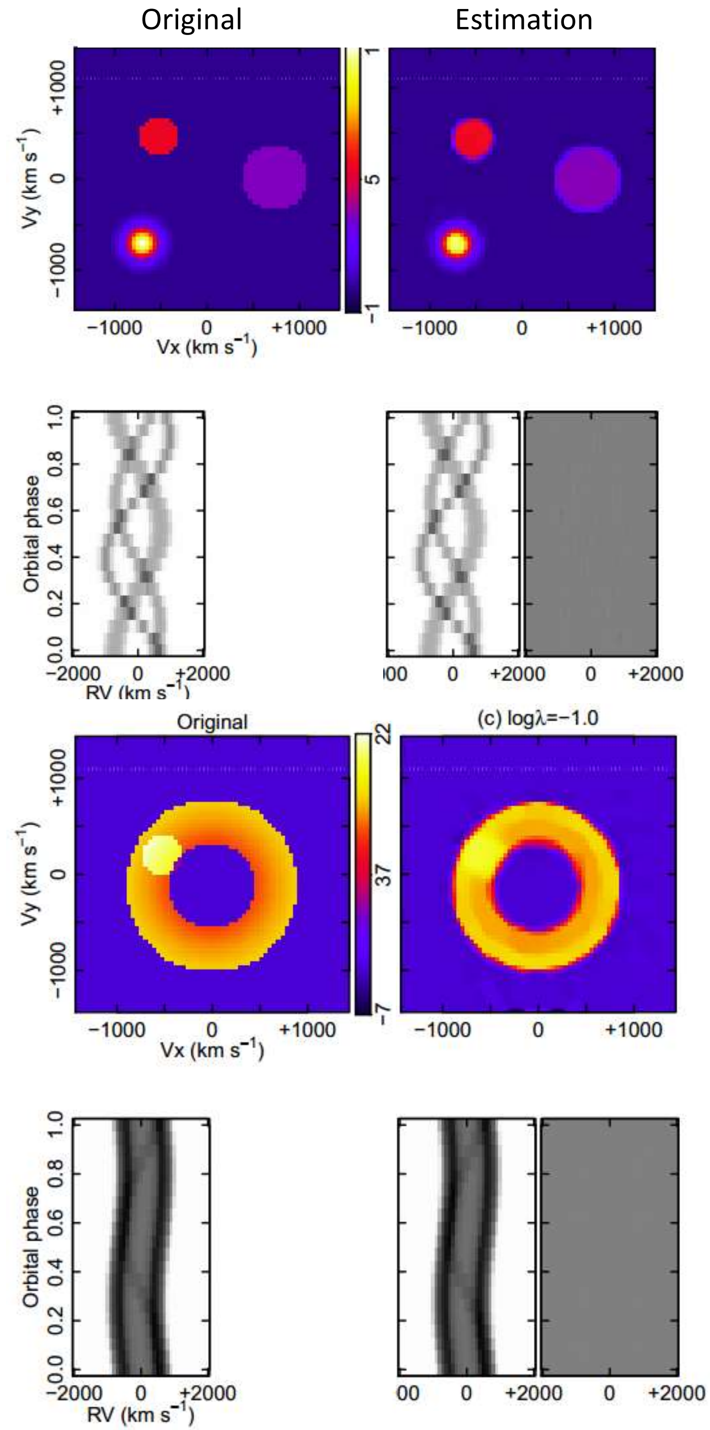
  - Δx: differential operator = $x_{i+1} - x_i$
  - Sparse in the gradient domain

Uemura+14

Doppler tomography: Simulations

- No. of input data = 820
  - Δphase = 0.05, ΔRV=100 km/s
- No. of pixels = 64x64=4096

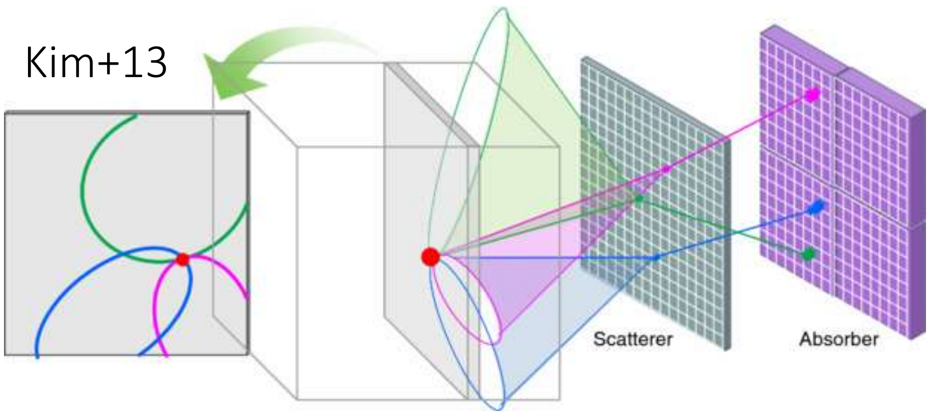
- Good reconstruction of both flat-top, sharp-edged profile and smooth, Gaussian profile



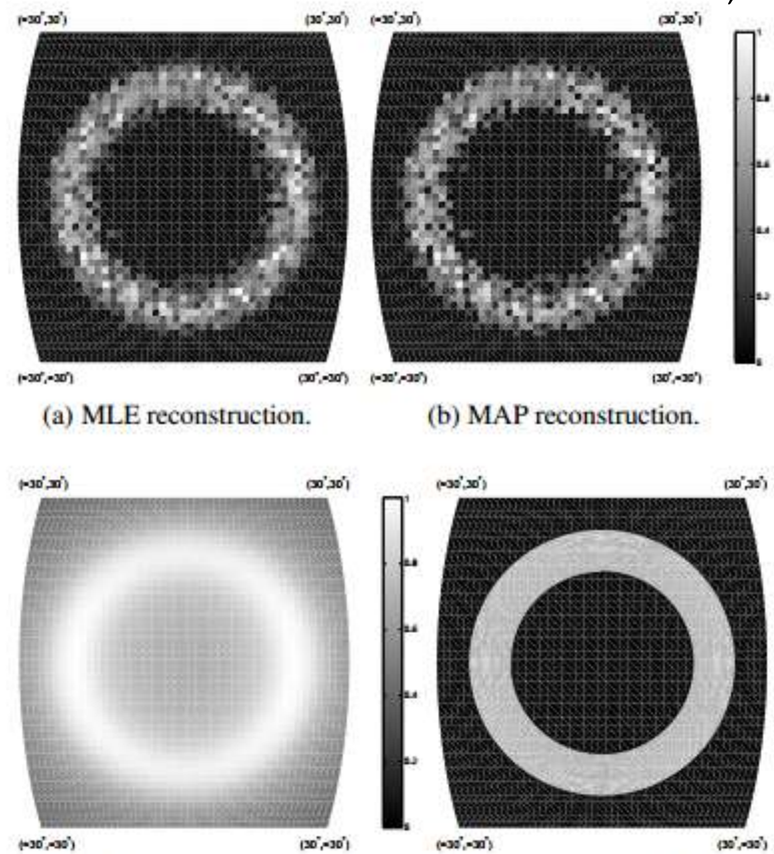
Gamma-ray Compton camera

(Ikeda, Uemura, et al. 2014, NIMPA, 760, 46)

Kim+13



Ikeda, MU+14



(a) MLE reconstruction. (b) MAP reconstruction.

(c) Back projection reconstruction. (d) True source distribution.

- Back-projection
  - drawing Compton-cones event by event
- Bin mode estimation
  - Reconstruction using the simulated response and Dirichlet prior

$$\hat{\rho}_{\mathrm{MAP}} = \arg\max_{\rho} \left[\log P(\rho \,|\, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_N) \right] = \arg\max_{\rho} \left[\log \pi(\rho) + L(\rho) \right],$$

$$L(\rho) = \sum_{t=1}^{N} \log q(\boldsymbol{v}_t) = \sum_{t=1}^{N} \log \sum_{\boldsymbol{u}} p(\boldsymbol{v}_t \,|\, \boldsymbol{u})\ \rho(\boldsymbol{u}).$$

$$\pi_\beta(\rho) = \frac{\Gamma(M\beta)}{\Gamma(\beta)^M} \prod_{\boldsymbol{u}} \rho(\boldsymbol{u})^{\beta-1}$$

Summary

- We can obtain better reconstructions using the sparsity of information (sparse modeling).
- Applications
  - Period analysis: determination of periods without aliases
    - ✓ Kato & Uemura (2012)
  - Radio interferometer: super-resolution
    - ✓ Honma, Akiyama, Uemura, & Ikeda (2014)
  - Doppler tomography: reconstructions of sharp-edged structures
    - ✓ Uemura, Kato, Nogami, & Mennickent (2014, submitted.)
  - Compton camera: high contrast images
    - ✓ Ikeda, Uemura+ (2014)

Backup slides

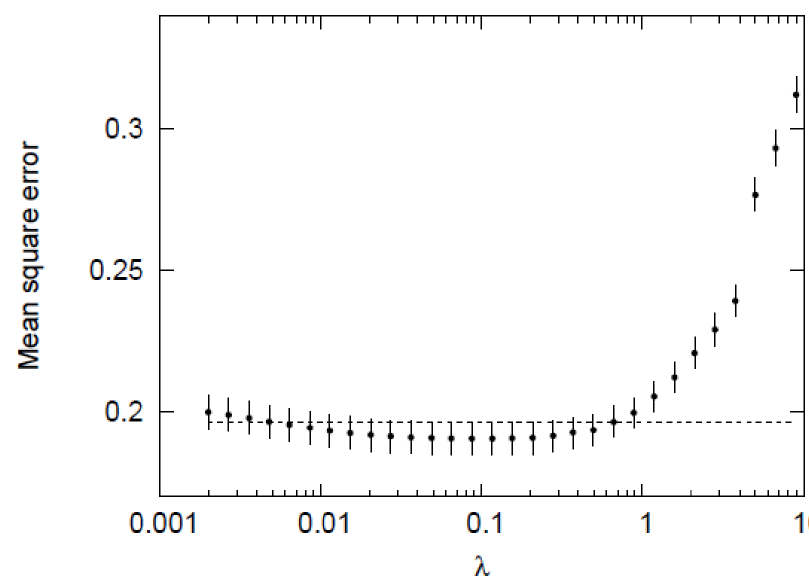
Model selection: How can we determine λ?

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$

- Akaike information criteria (AIC)
  - $\mathrm{AIC} = -2\log L + 2K$

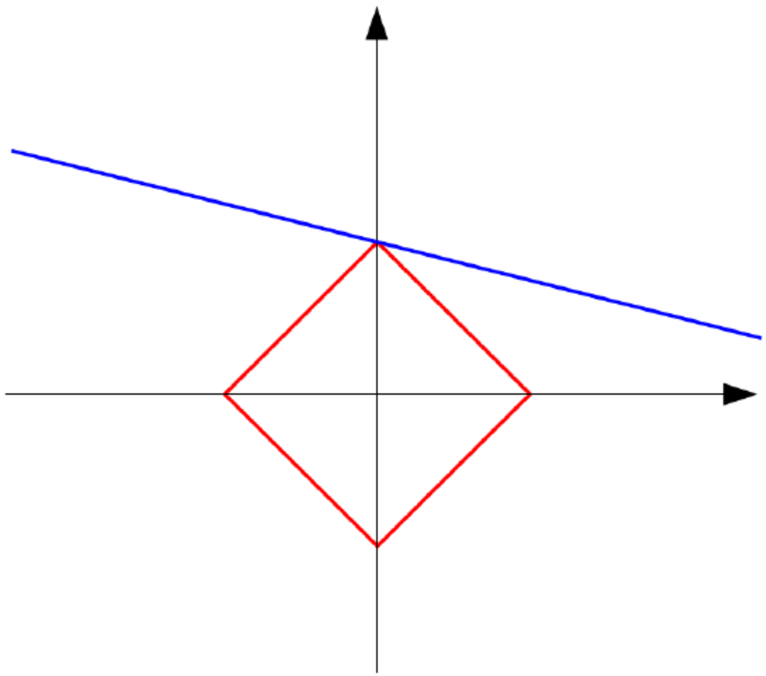
    Log-likelihood The number of parameters
  - Balance between the goodness of fit and the complexity of the model
  - Model with minimal AIC = best model
- Cross-validation
  1. Data → training set & validation set
  2. Optimization of the model using the training set
  3. Validation (with MSE, for example) of the model using validation set
  4. Model with minimal MSE = best model

Example of cross-validation curve



L1-norm minimization

Why does it select a sparse solution ?



How “sparse” do we allow?