Sparse modeling in Astronomy

Makoto Uemura (Hiroshima University)

@Universidad de Concepcion, Aug. 2016

2004
Universidad de Concepcion

2005~
Hiroshima University, Japan
1.5-m “Kanata” Telescope

Sparse modeling project

Collaborations with information scientists

Let's start with a demo.

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

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

A linear problem: estimating **x** from **y**

Can be solved if $M \geqq N$ (ex. using least-square method)

No unique solution if $M<N$

Experiment: set 100 elements of **x**, 50x100 matrix A, and generate 50 elements of **y**.

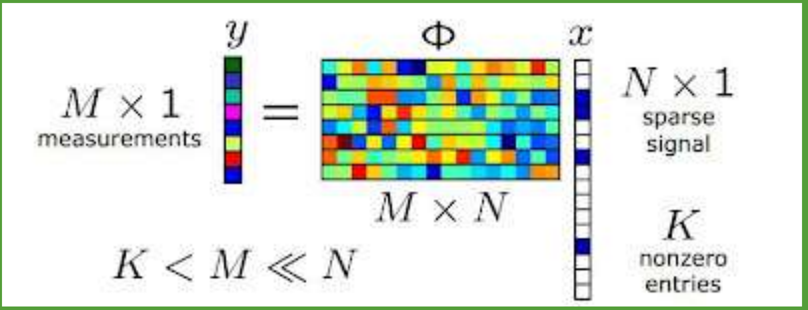
Then, estimate **x** from A and **y**.

Demo.

More variables can be estimated with less data

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

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



Sparse vectors can be reconstructed by l1-norm minimization

Compressed sensing

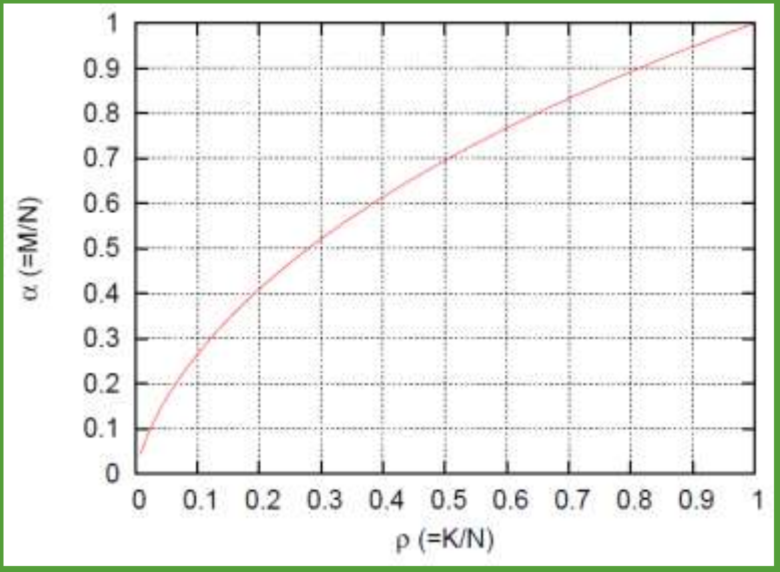
Condition for perfect reconstructions. (Candes & Tao 2006, Donoho 2006)

LASSO

sparse regression（Tibshirani 1996）

Even if *x* is not sparse, sparse modeling can estimate x when *x’ (=Bx)* is sparse.

Condition for perfect reconstruction of sparse vectors:
Sparsity (K/N) v.s. Data size (M/N)



Outline

- Estimating a sparse vector from a small data set.
  - Estimation of power spectra of periodic variables
    - *Kato & Uemura, 2012, PASJ, 64, 122*
  - Radio interferometer
    - *Honma, Akiyama, Uemura, & Ikeda, 2014, PASJ, 66, 95*
  - Doppler tomography
    - *Uemura, Kato, Nogami, & Mennickent, 2015, PASJ, 67, 22*
- Variable selection from data
  - The peak magnitude of Type-Ia supernovae
    - *Uemura, Kawabata, Ikeda, Maeda, 2015, PASJ, 67, 55*
- Summary

Estimation of power spectra of periodic variables

Estimation of periods from data.

The data is not uniformly sampled. → complicate window function → aliases

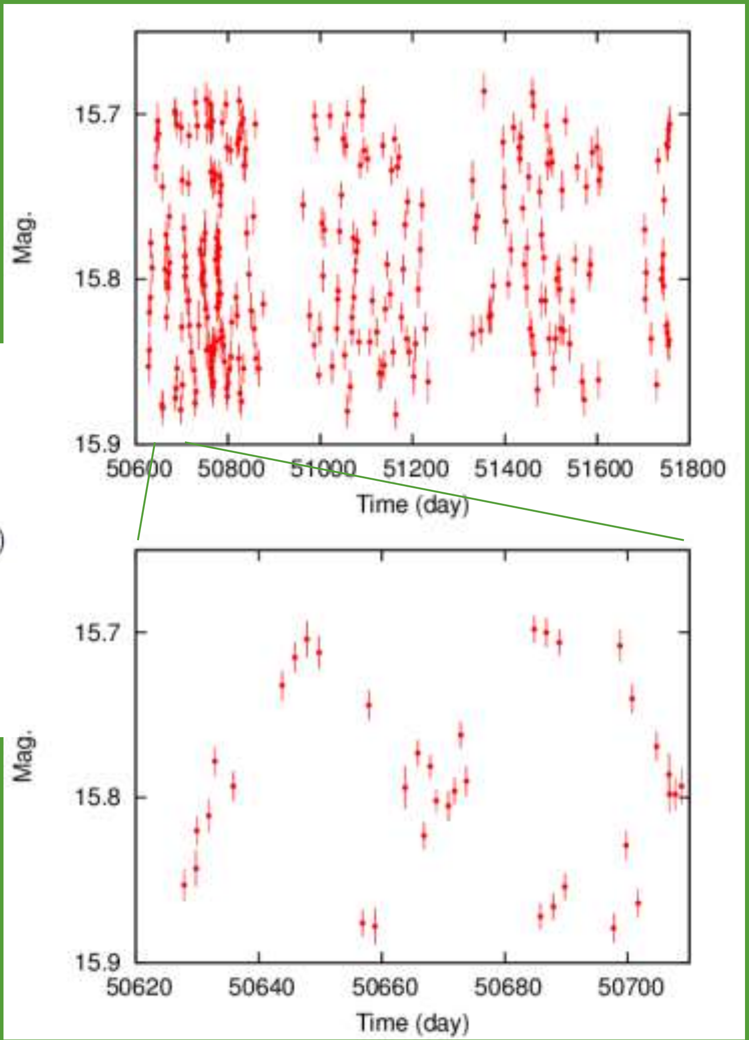
Power spectra should be “sparse” if the object only has a few periods.

$$\boldsymbol{y} = \mathcal{F}\boldsymbol{x}$$

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

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

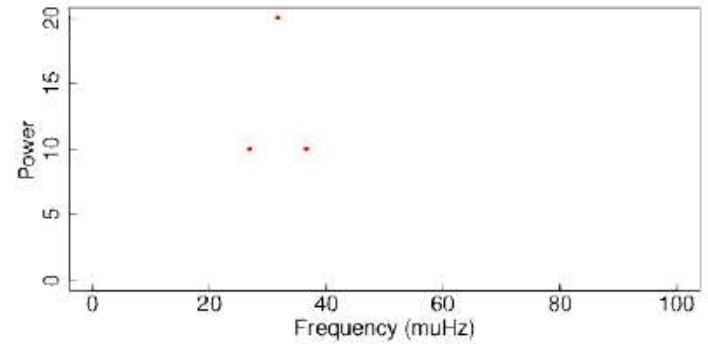
Example of non-uniformly sampled light curves



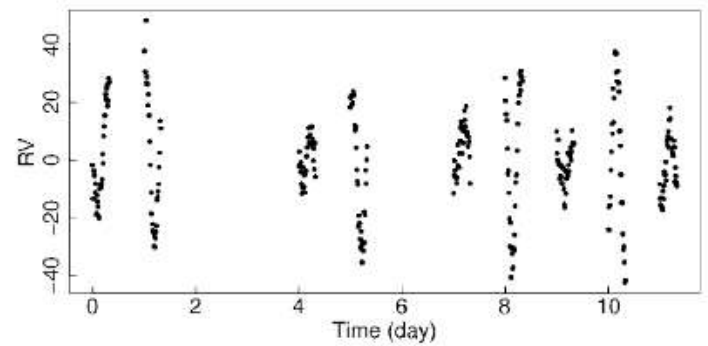
Data from OGLE (collaboration with Dr. Ita)

Experiment using artificial data

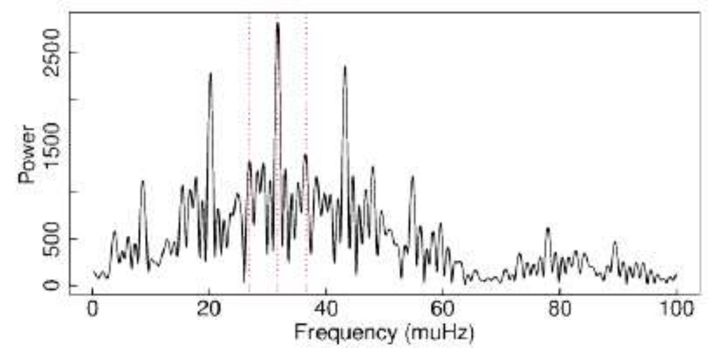
Assumed signals



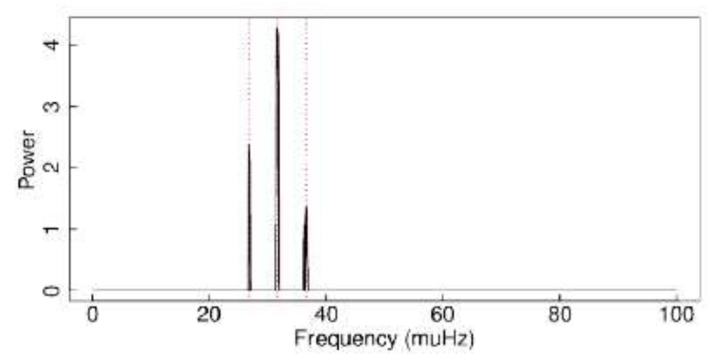
Simulated light curve



Power spectrum by
the standard Fourier transform

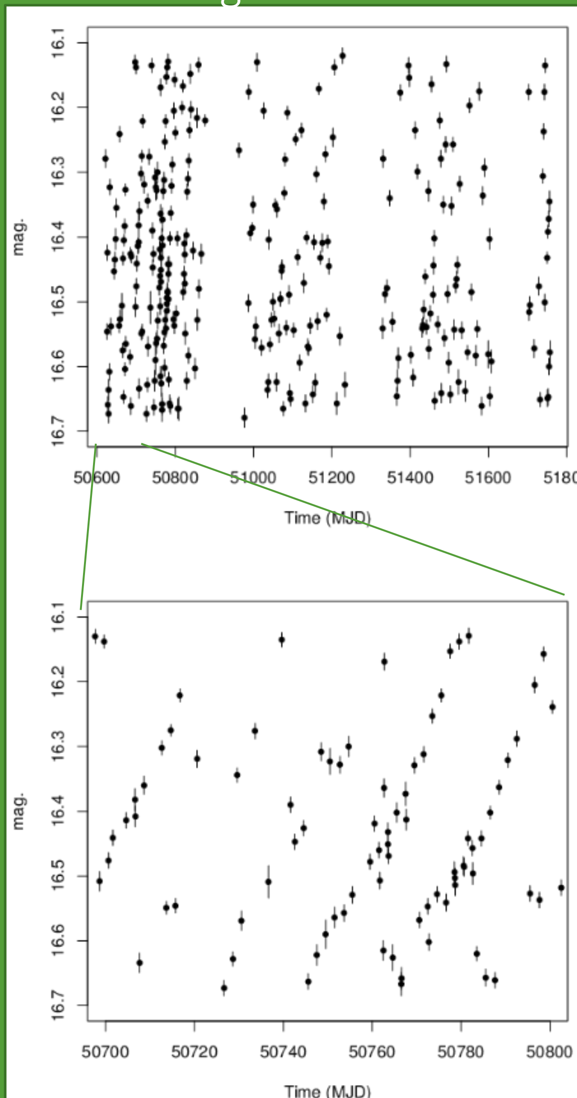


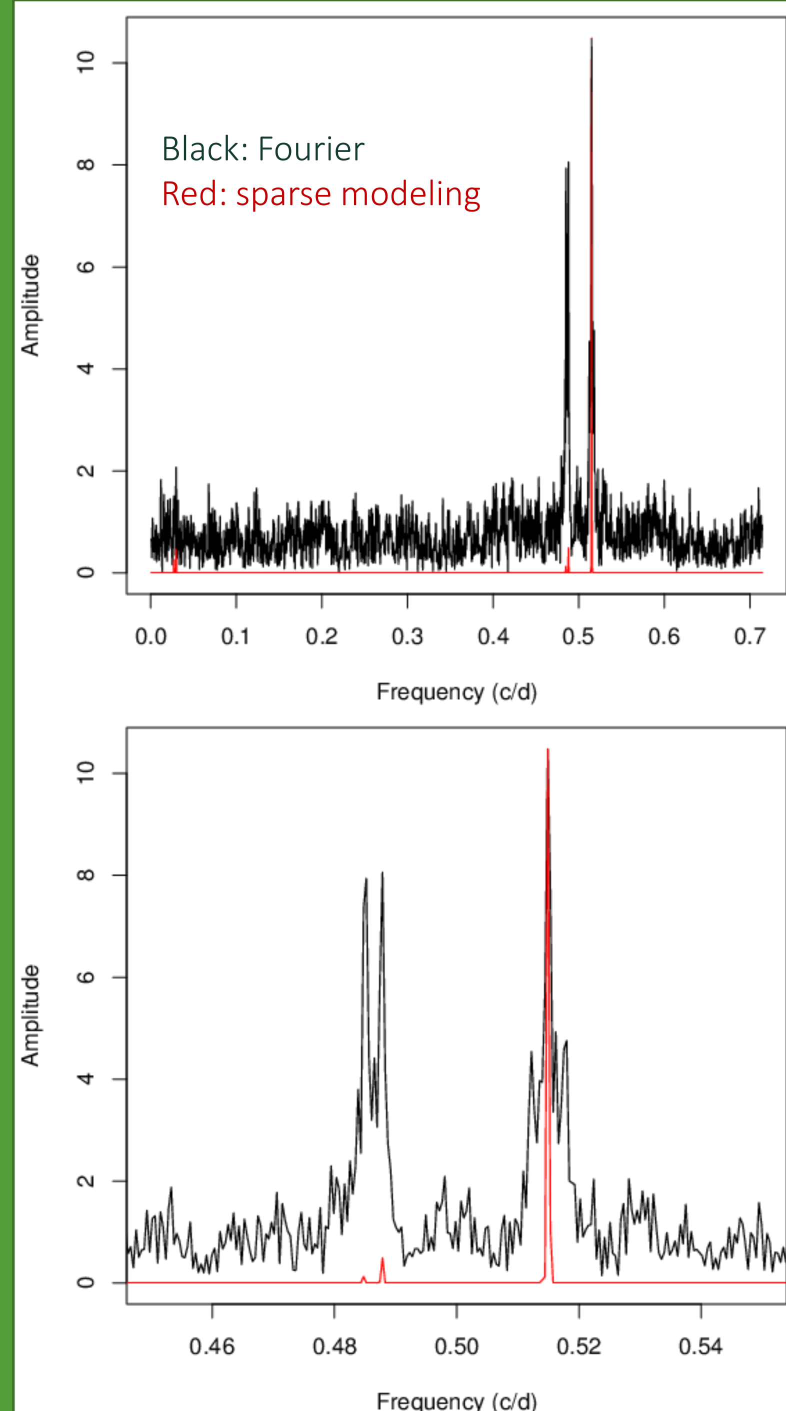
Power spectrum by
the sparse modeling



Using real data

Light curve



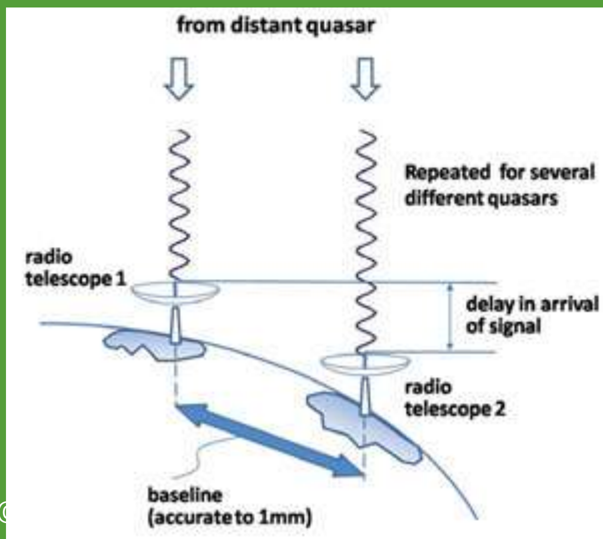


(Data from OGLE, collaboration with Dr. Ita)

Outline

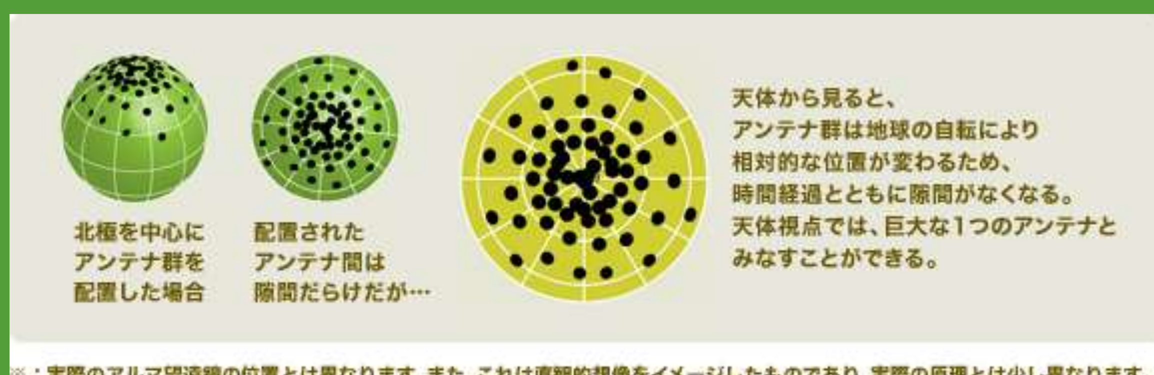
- Estimating a sparse vector from a small data set.
  - Estimation of power spectra of periodic variables
    - *Kato & Uemura, 2012, PASJ, 64, 122*
  - Radio interferometer
    - *Honma, Akiyama, Uemura, & Ikeda, 2014, PASJ, 66, 95*
  - Doppler tomography
    - *Uemura, Kato, Nogami, & Mennickent, 2015, PASJ, 67, 22*
- Variable selection from data
  - The peak magnitude of Type-Ia supernovae
    - *Uemura, Kawabata, Ikeda, Maeda, 2015, PASJ, 67, 55*
- Summary

Image reconstruction of radio interferometer



Radio interferometer: high angular resolution using multiple radio telescopes (of different sites = VLBI).

Data = complex visibility ← 2D Fourier transform of the intensity

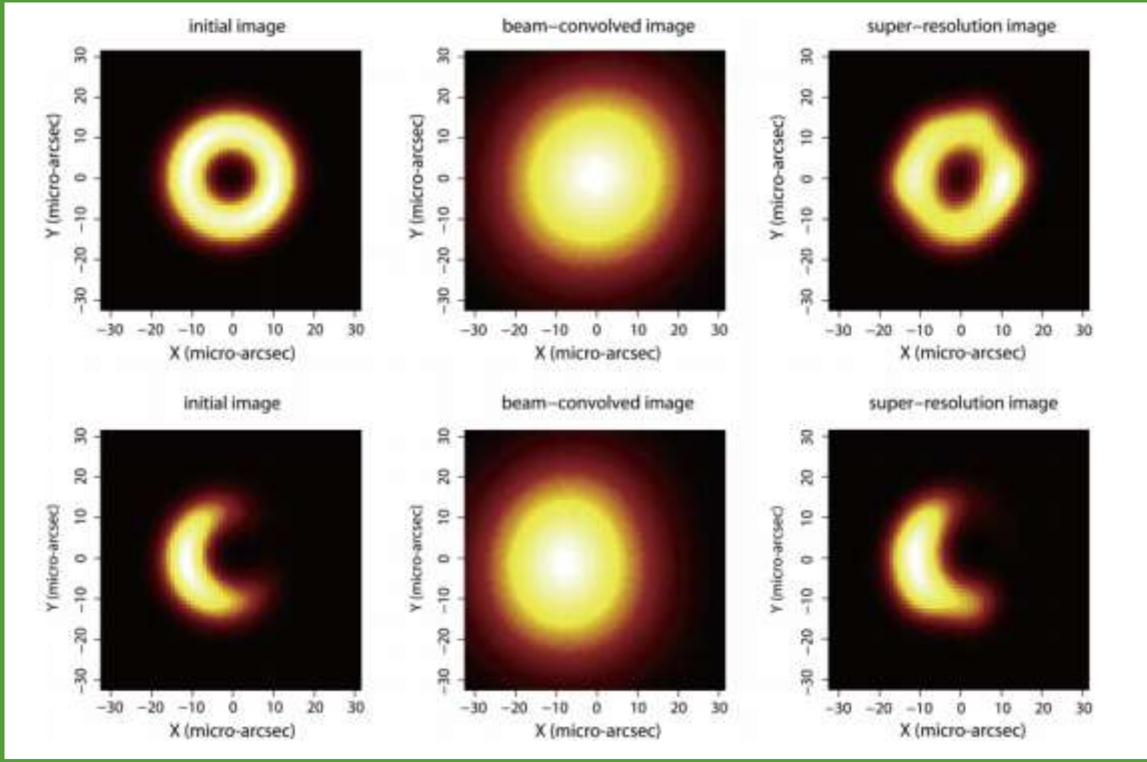


Experiments for the black hole shadow with EHT

Assumption: Radio sources are sparse in the map.

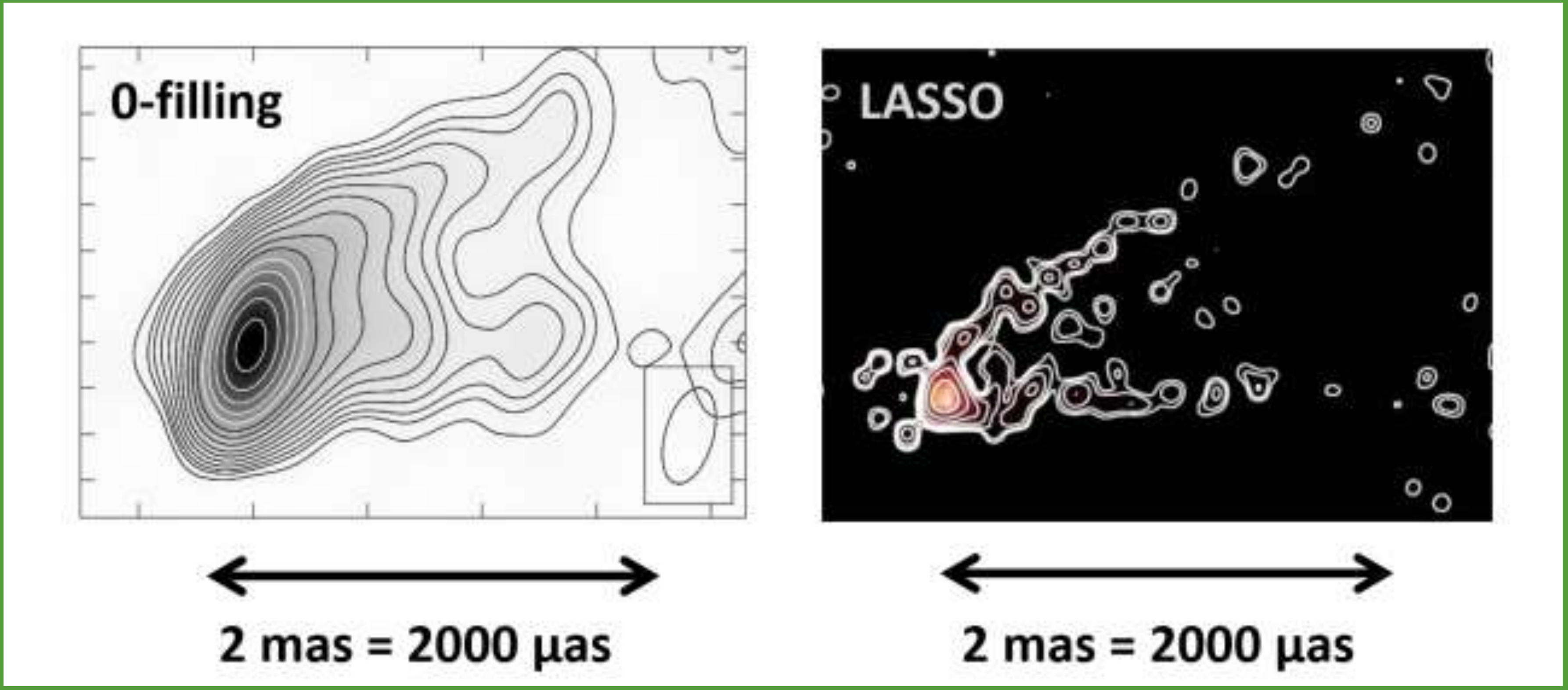
Super-resolution

Honma+14



Application to the real data of M87

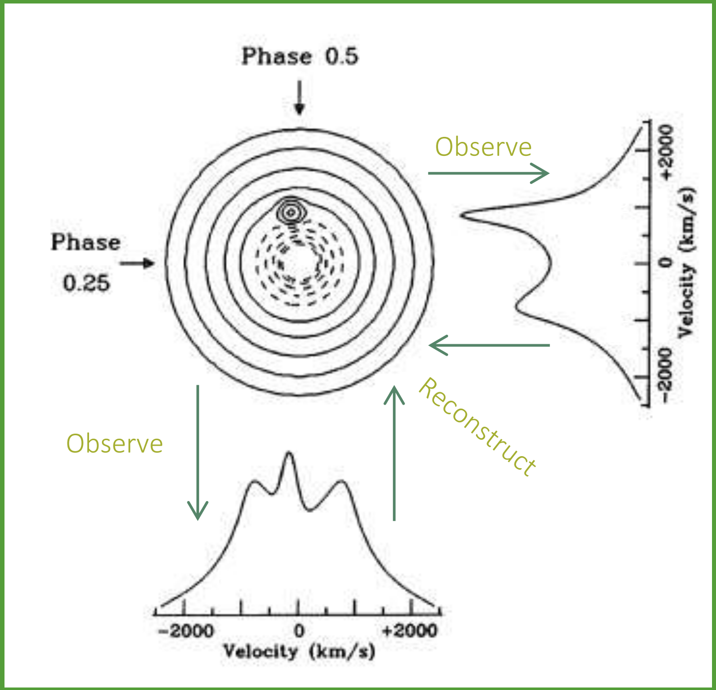
Honma+15

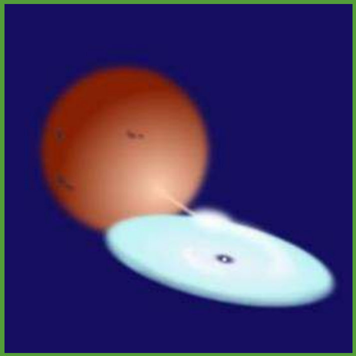


Outline

- Estimating a sparse vector from a small data set.
  - Estimation of power spectra of periodic variables
    - *Kato & Uemura, 2012, PASJ, 64, 122*
  - Radio interferometer
    - *Honma, Akiyama, Uemura, & Ikeda, 2014, PASJ, 66, 95*
  - Doppler tomography
    - *Uemura, Kato, Nogami, & Mennickent, 2015, PASJ, 67, 22*
- Variable selection from data
  - The peak magnitude of Type-Ia supernovae
    - *Uemura, Kawabata, Ikeda, Maeda, 2015, PASJ, 67, 55*
- Summary

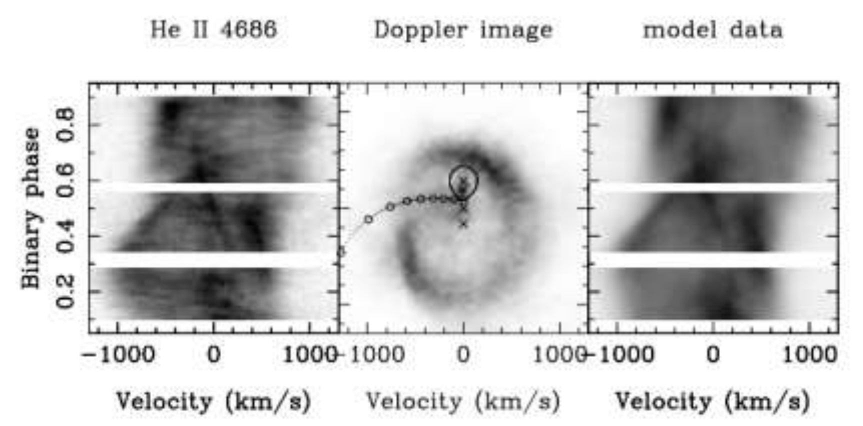
Doppler tomography





Doppler Tomography = Reconstruction of emission-line intensity map on the velocity space from time-series spectra

- A similar method to medical X-ray CT



IP Peg (Harlaftis+99)

Structure of the Doppler tomography

Linear problem if the disk is geometrically and optically thin.

No self-occultation

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

Data — Observation Matrix — image

- Maximum Entropy Method (MEM)
  - Standard method to date

$$S = -\sum^{M} p_i \ln \frac{p_i}{q_i}.$$

$$q_i = \frac{D_i}{\sum_{j=1}^{M} D_j},$$

  - Best for real Doppler maps?
    - Hot spot and/or shock region may have sharp edges, making entropy low

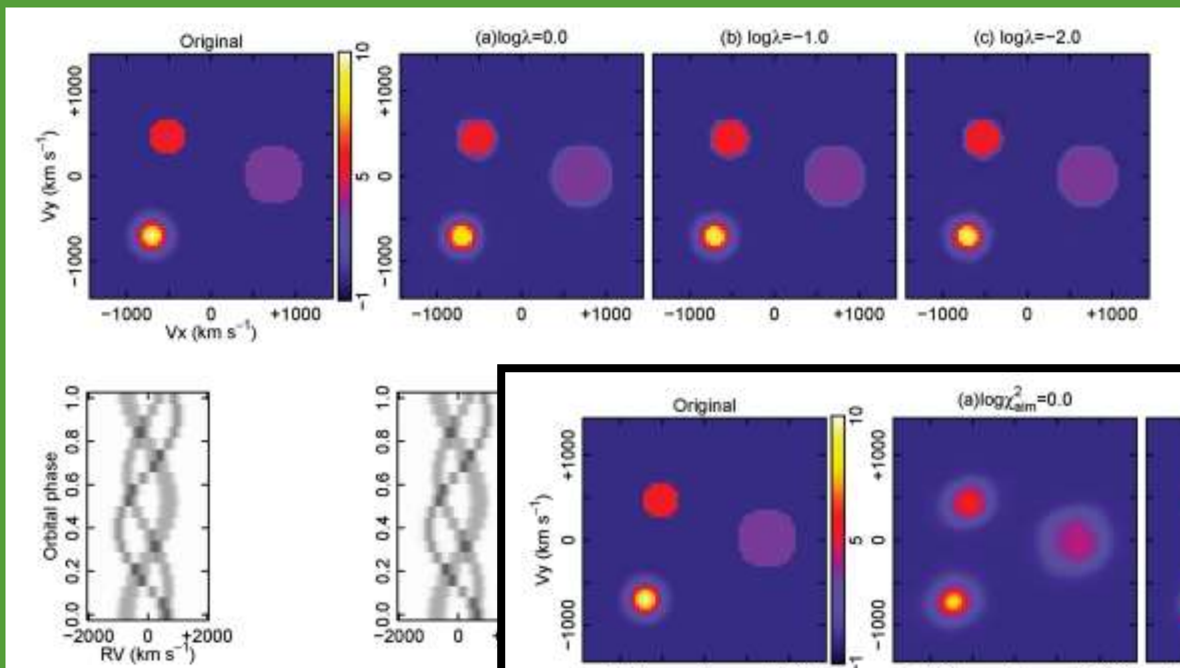
- Total Variation Minimization (TVM)
  - Simple prior
  - Regularization:

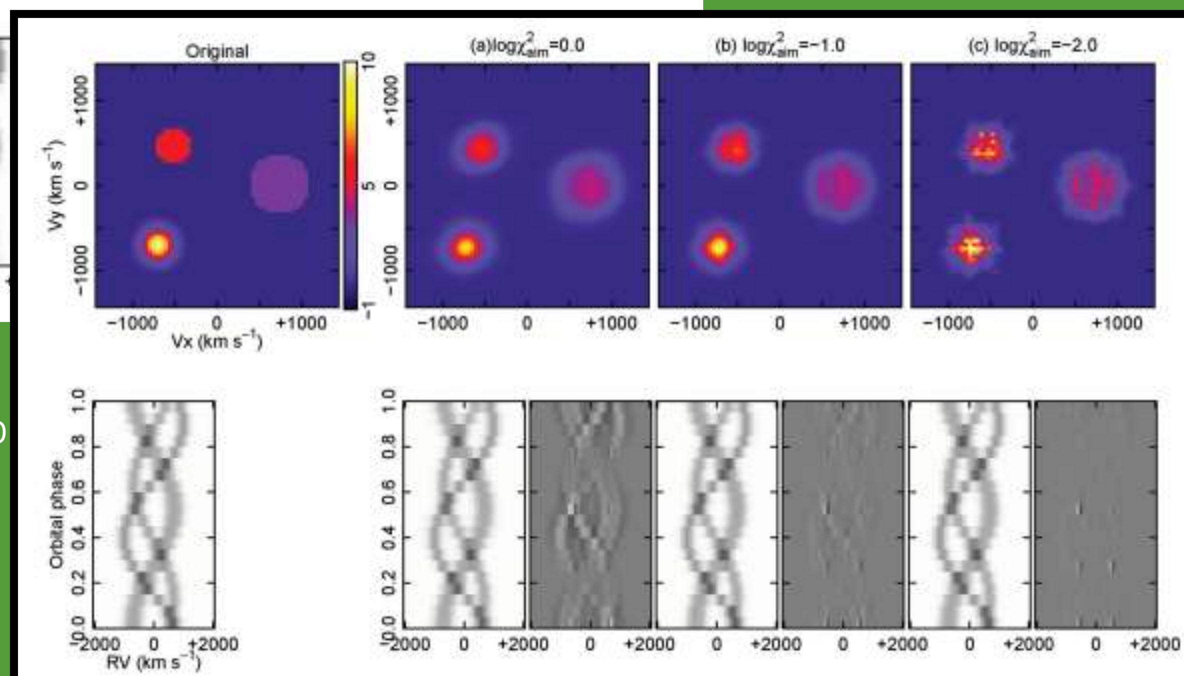
$$TV(\boldsymbol{x}) = \sum \sqrt{(\Delta^h \boldsymbol{x})^2 + (\Delta^v \boldsymbol{x})^2}$$

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

Doppler tomography with TVM

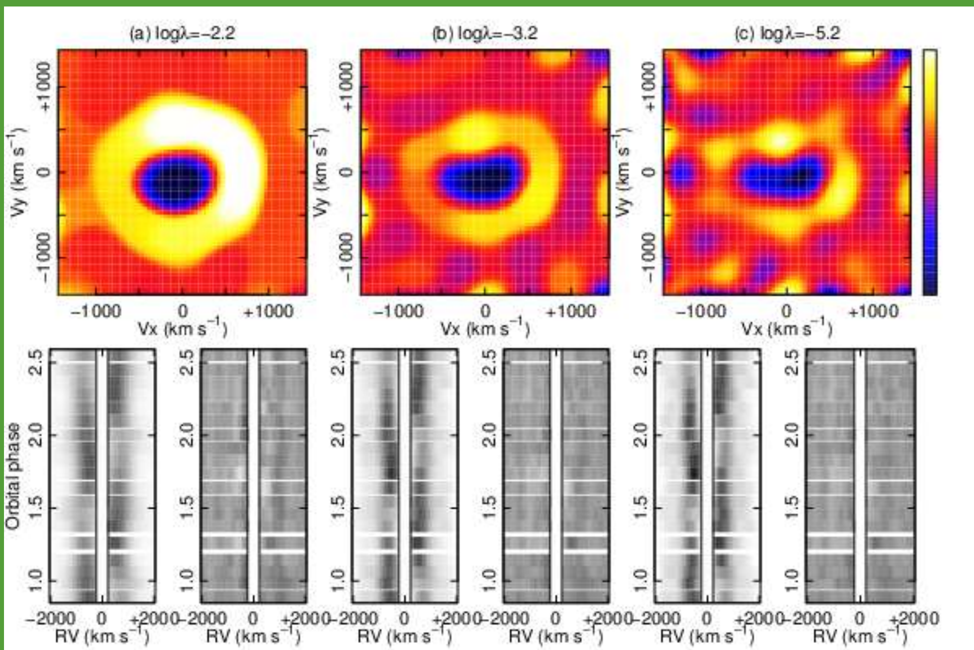
Uemura+15

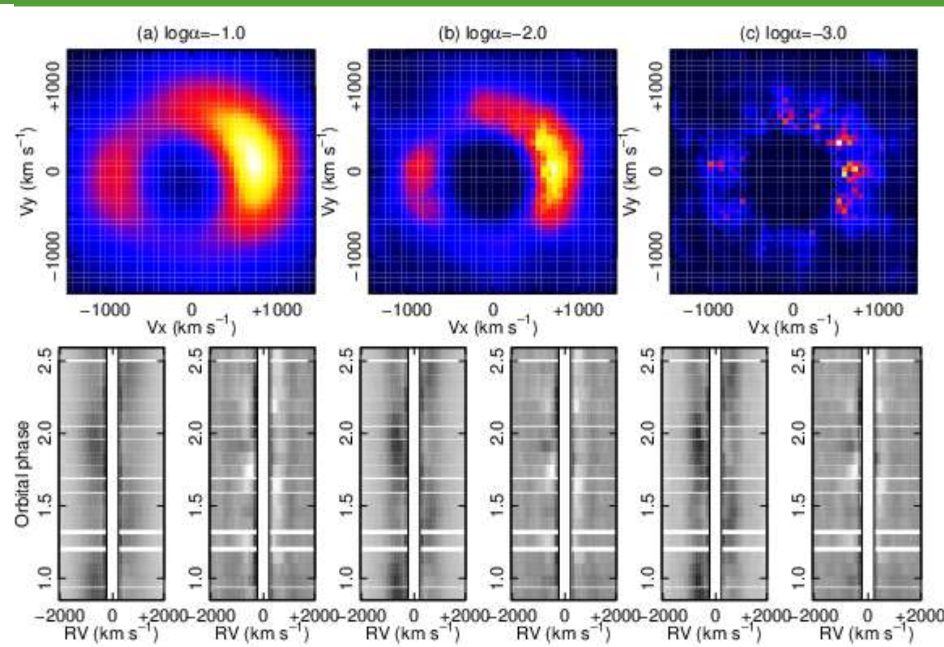




Reconstruction of both sharp

Example for the real data (WZ Sge)





TVM

- ✓ Elliptical structure of the entire disk
- ✓ Small, localized sources around the secondary star and a part of disk
- ✓ Small residuals between the data and model.

MEM (Spruit+88)

- ✓ Circular structure of the entire disk
- ✓ Two-armed spirals
- ✓ Rotating residuals

TVM can reconstruct locally confined sources which are not done with MEM

Outline

- Estimating a sparse vector from a small data set.
  - Estimation of power spectra of periodic variables
    - *Kato & Uemura, 2012, PASJ, 64, 122*
  - Radio interferometer
    - *Honma, Akiyama, Uemura, & Ikeda, 2014, PASJ, 66, 95*
  - Doppler tomography
    - *Uemura, Kato, Nogami, & Mennickent, 2015, PASJ, 67, 22*
- Variable selection from data
  - The peak magnitude of Type-Ia supernovae
    - *Uemura, Kawabata, Ikeda, Maeda, 2015, PASJ, 67, 55*
- Summary

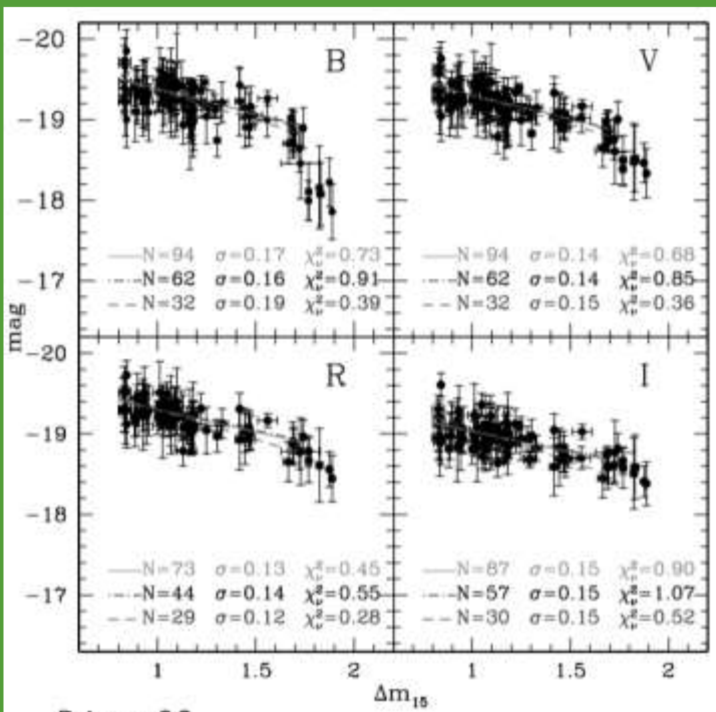
The peak luminosity of type Ia supernovae

A function of the color (=interstellar extinction) and decay rate

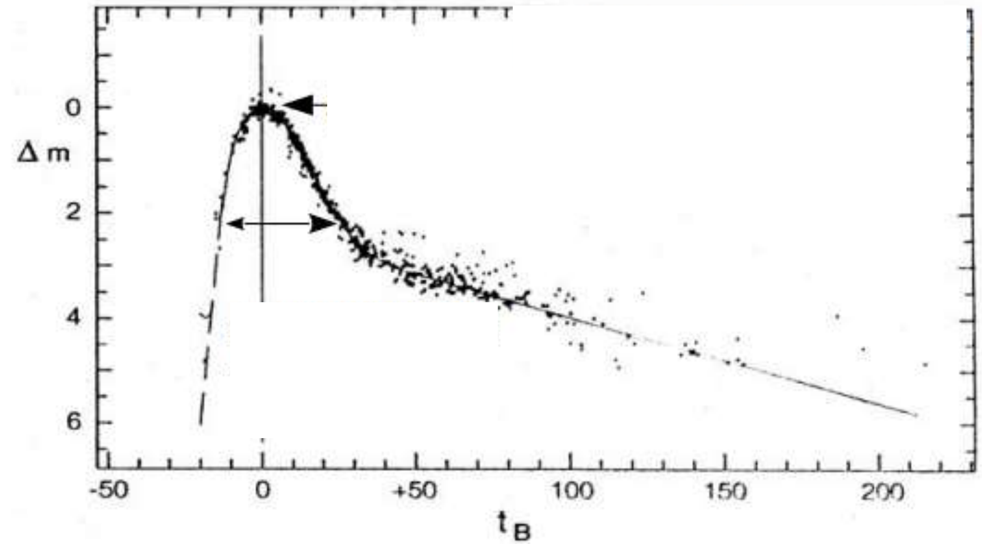
Phillips 93, Hamuy+96, Prieto+06

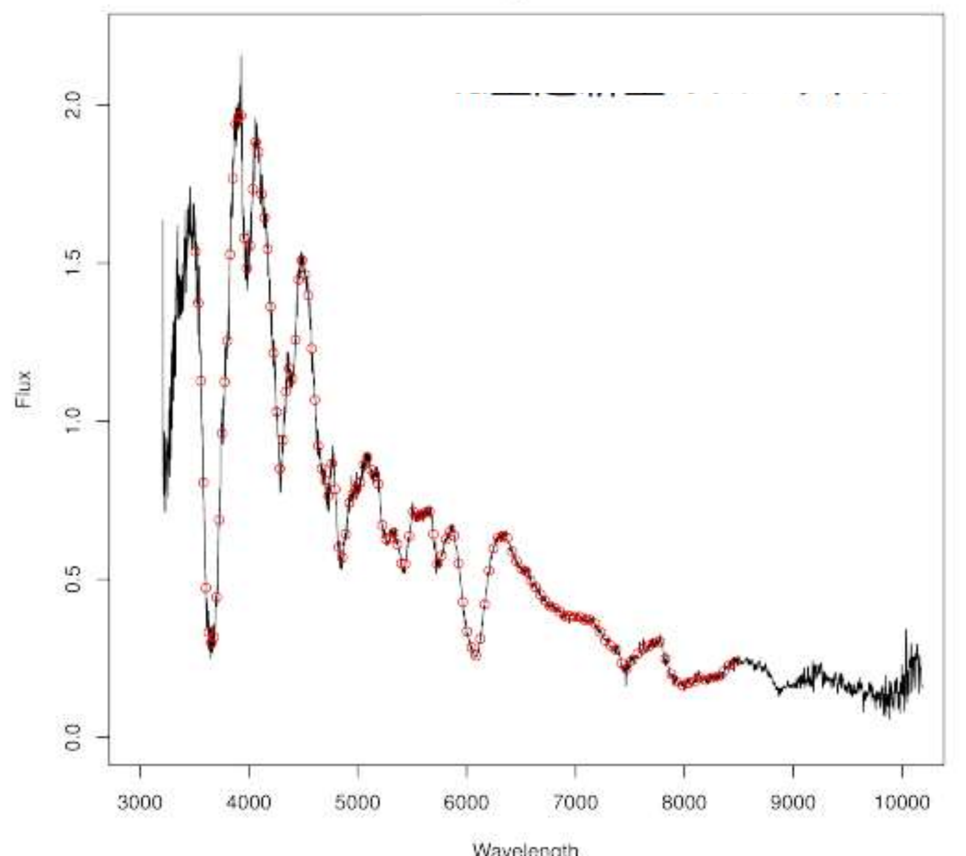
, and any others?

Search for the 3rd parameter using spectroscopic data.



Prieto+06

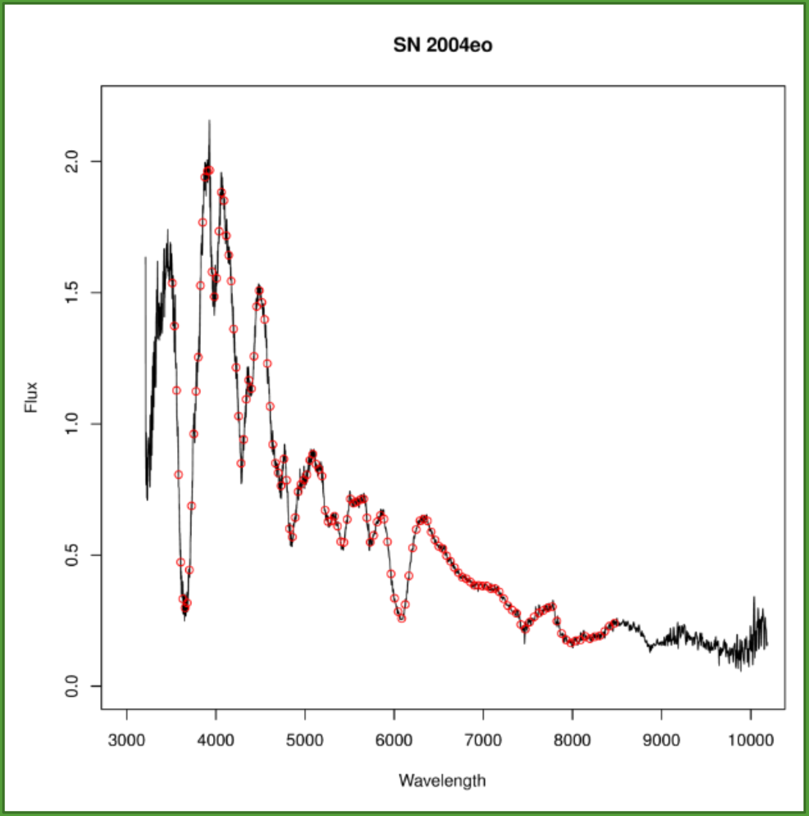




Which parameter determines the peak mag.

According to Silverman+12 (BSNIP III) ,

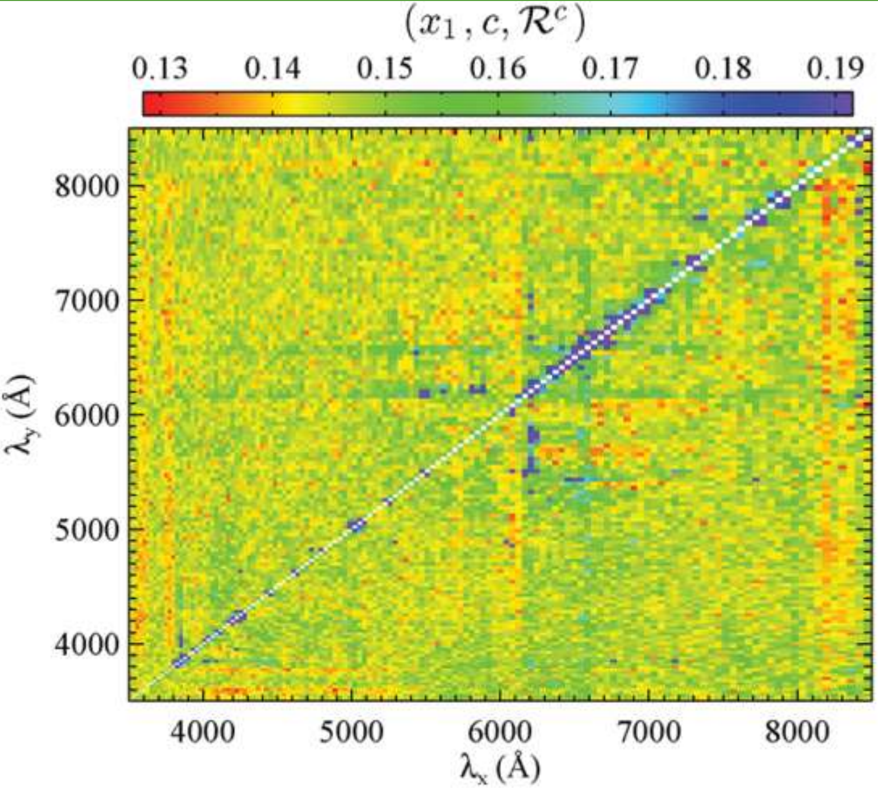
- Velocity of Si II 6355 (Blondin+11)
- Velocity Ca II H&K (Foley&Kasen 11)
- Depth of the blue side of S II “W” (Blondin+11)
- EW of Si II 4000 (Arsenijevic+08, Walker+11, Chotard+11, Nordin+11, Walker+11)
- EW of Fe II, Mg II (Nordin+11)
- EW of Si II 5972, 6355 （Hachinger+06, Nordin+11）
- Si II EW ratio EW(5972)/EW(6355), Ca II H&K flux ratio (Fr/Fb) (Nugent+95, Hachinger+06)
- SiS flux ratio Fr(S II “W”)/Fr (Si II 6355) (Bongard+06)
- SSi EW ratio EW(S II “W”)/EW(Si II 5972), SiFe EW ratio EW(Si II 5972)/EW(Fe II) (Hachinger+06)
- Search for a good variable using arbitrary flux ratio (Bailey+09)



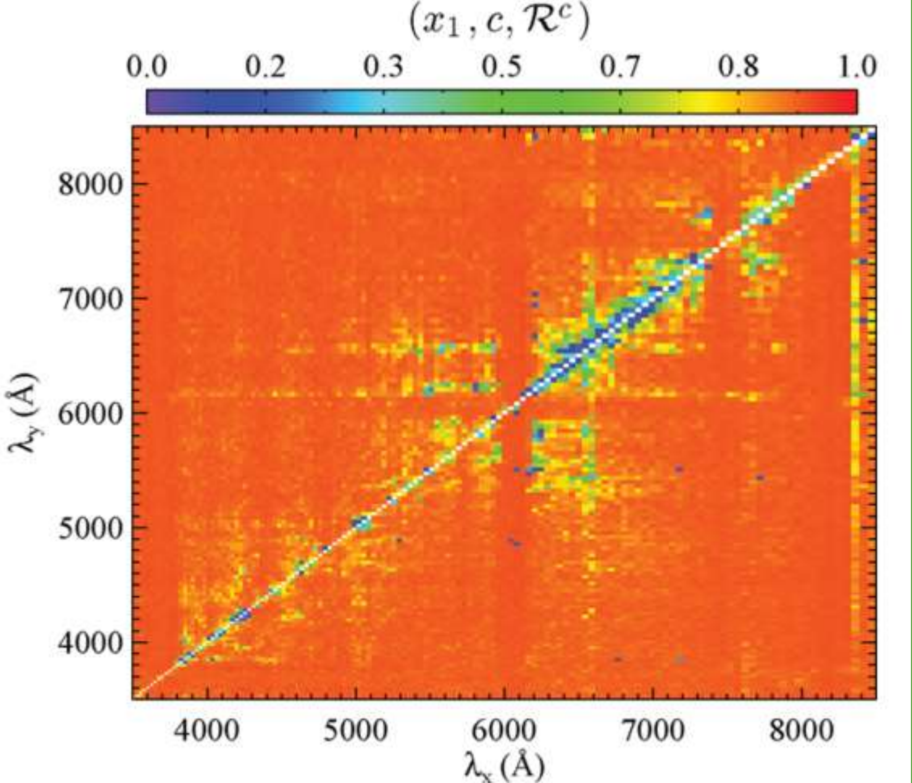
Models with arbitrary flux ratios

- Bailey+09, Silverman+12 (BSNIP III)
- Spectra of λ3500-8500 are re-binned into 134 bin, then calculate 134x133 flux ratios.
- Search for the flux ratio having high correlation with the peak magnitude.
- A set of R(3750/4550) , Lightcurve width (x1), & color (c) gives the best model.

Residuals



Correlation coefficients



A problem of variable selection

$$
\begin{pmatrix} MB_{\mathrm{SN1994S}} \\ MB_{\mathrm{SN1995E}} \\ \vdots \\ MB_{\mathrm{SN2008s1}} \end{pmatrix} = \begin{pmatrix} x1_{\mathrm{SN1994S}} & c_{\mathrm{SN1994S}} & EW_{\mathrm{SiII4000,SN1994S}} & FWHM_{\mathrm{SiII4000,SN1994S}} & 3535/3512_{\mathrm{SN1994S}} & 3558/3512_{\mathrm{SN1994S}} & \cdots & 8416/8472_{\mathrm{SN1994S}} \\ x1_{\mathrm{SN1995E}} & c_{\mathrm{SN1995E}} & EW_{\mathrm{SiII4000,SN1995E}} & FWHM_{\mathrm{SiII4000,SN1995E}} & 3535/3512_{\mathrm{SN1995E}} & 3558/3512_{\mathrm{SN1995E}} & \cdots & 8416/8472_{\mathrm{SN1995E}} \\ \vdots & & & & \vdots & & & \\ x1_{\mathrm{SN2008s1}} & c_{\mathrm{SN2008s1}} & EW_{\mathrm{SiII4000,SN2008s1}} & FWHM_{\mathrm{SiII4000,SN2008s1}} & 3535/3512_{\mathrm{SN2008s1}} & 3558/3512_{\mathrm{SN2008s1}} & \cdots & 8416/8472_{\mathrm{SN12008s1}} \end{pmatrix} \begin{pmatrix} c_{x1} \\ c_c \\ c_{EW} \\ c_{FWHM} \\ c_{3535/3512} \\ c_{3558/3512} \\ \vdots \\ c_{8416/8472} \end{pmatrix}
$$

- They use a linear model. The number of variables is larger than the number of data if we use arbitrary flux ratios.
- But, we want to know a model with a few variables. The coefficient vector should be sparse. We want to select the variables from the data. Let's use sparse modeling !
  - L1 minimization $\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$
  - Data from the UC Berkeley database of supernovae.

Model & Result

Variables

- Color
- Light curve width
- Continuum-normalized spectra in log scale
  - *Line information*
- Total flux-normalized spectra in log scale
  - *Local color information*
- Previously proposed flux ratios
- Total 276 candidates of variables

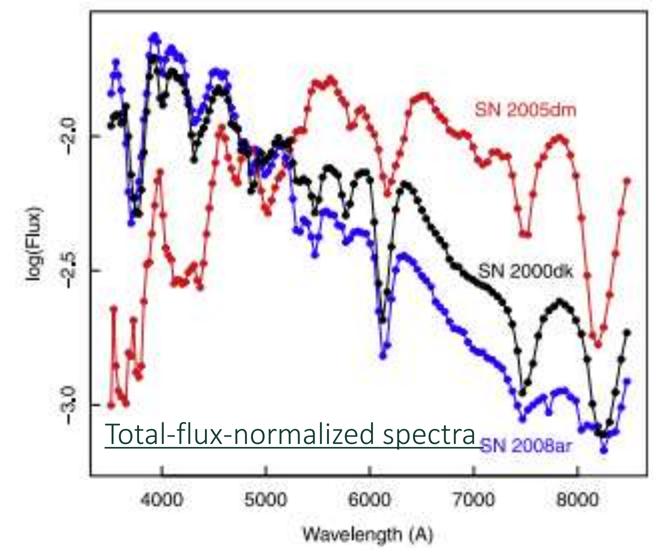
Data

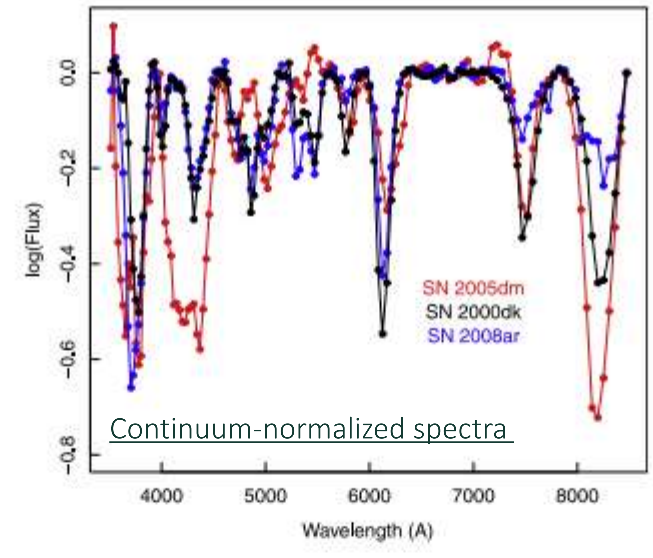
- From Berkely supernova database
- 78 objects.

Model selection

- L1 minimization with cross-validation

No additional variables. The classical model with color and decay rate is the best.





Summary

- Estimating a sparse vector from a small data set.
  - Estimation of power spectra of periodic variables
    - *Eliminating aliases using the sparsity of power spectra*
  - Radio interferometer
    - *Super-resolution using the sparsity of radio maps*
  - Doppler tomography
    - *Accurate reconstruction using the sparsity in the gradient domain (TVM)*
- Variable selection from data
  - The peak magnitude of Type-Ia supernovae
    - *Estimating the variables and number of variables from the data.*