Information Compression #5
Basics of sparse modeling

Youhei Yamaji

Department of Applied Physics, The University of Tokyo

1. Underdetermined problems and sparseness
2. Compression sensing
3. Bayesian statistics connecting two problems

Computational
Science
Alliance
The University of Tokyo

Classification of Information Compression in Linear Algebra by Memory Costs

(1) A matrix can be stored

-SVD for dense matrix
-Compressed sensing, for example

(2) Although a matrix cannot be stored, vectors can be stored

-SVD for sparse matrix
-Krylov subspace method

(3) A vector cannot be stored

-Matrix product/tensornetwork states

Underdetermined Problems and Sparseness

Underdetermined Problem

If $n < N$, the linear system is underdetermined (劣決定):

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

$\vec{y} \in \mathbb{R}^n$	known vector/vector of measurement
$\vec{x} \in \mathbb{R}^N$	unknown vector/signal of interest
A	*n* x *N* measurement matrix

Without any *prior* knowledge (事前知識) about the signal, we can not solve this linear system/reconstruct the signal.

When we can minimize $\|\vec{y} - A\vec{x}\|_2^2 = (\vec{y} - A\vec{x})^t(\vec{y} - A\vec{x})$ without any prior knowledge, we will find many minima:

-Not unique
-Overfitting (過学習) (will be explained in the 8th lecture)

Underdetermined Problem

If $n < N$, the linear system is underdetermined (劣決定):

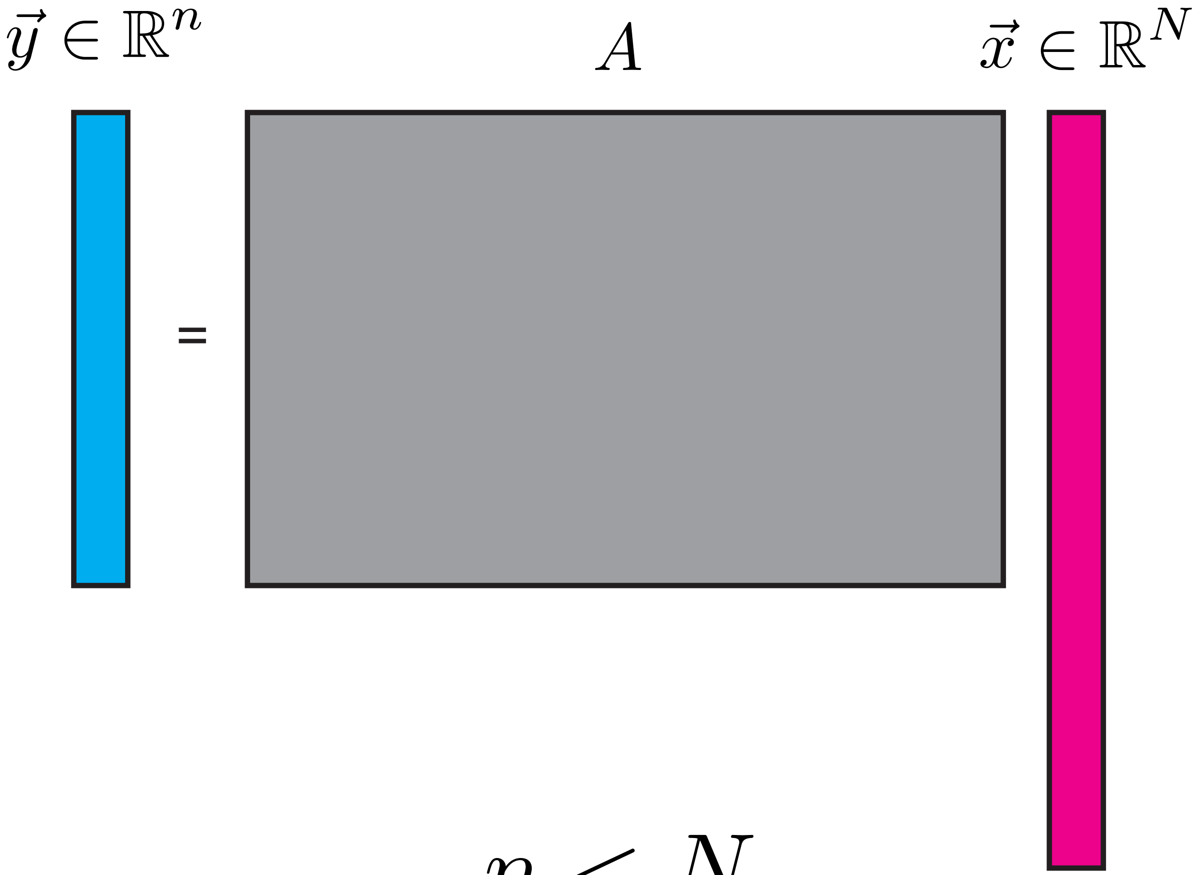
$$\vec{y} = A\vec{x}$$

$\vec{y} \in \mathbb{R}^n$	known vector/vector of measurement
$\vec{x} \in \mathbb{R}^N$	unknown vector/signal of interest
A	n x N measurement matrix

Without any *prior* knowledge (事前知識) about the signal, we can not solve this linear system/reconstruct the signal.

An example of prior knowledge: Sparseness (疎性)

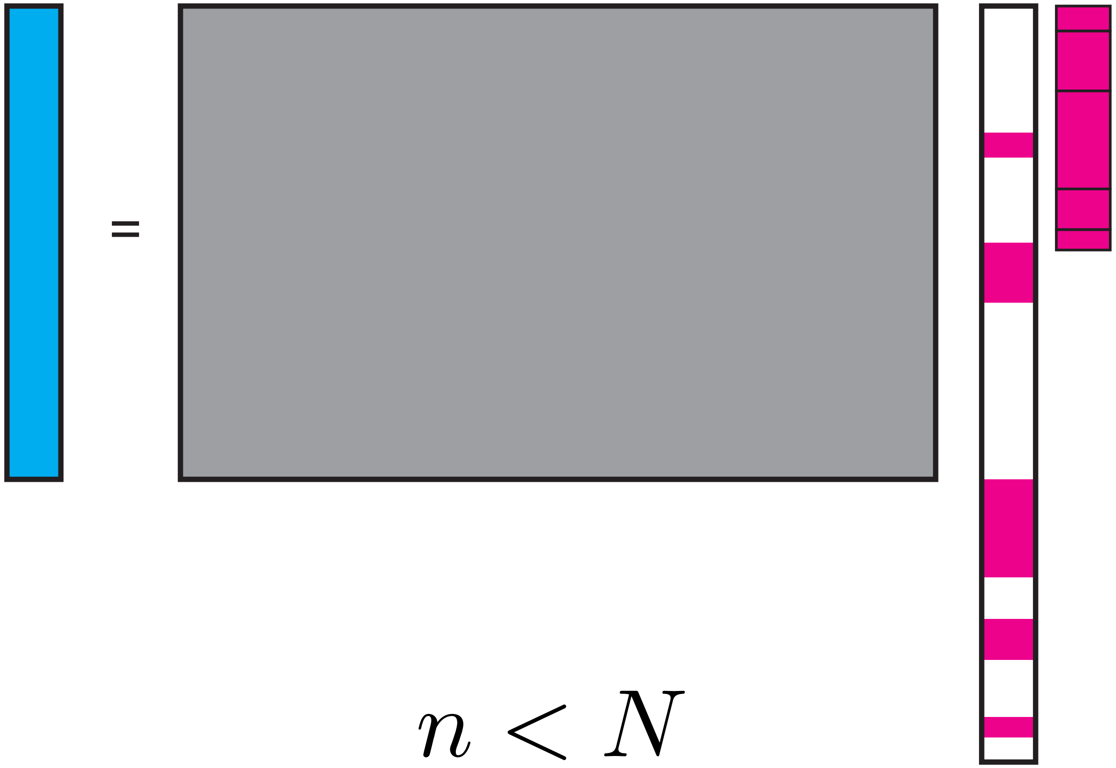
If we know that the unknown vector has as many as k non-zero entries and $k < n$, we can reconstruct the unknown vector.



$$n < N$$

We can not reconstruct *x*

Suppose we do not know where are *k* non-zero entries.
However, suppose we know the sparseness.



$$n < N$$

$$k < n$$

We can reconstruct *x*

Sparse Underdetermined Problem

Signal reconstruction by using L_0-norm

$\|\vec{x}\|_0$ L_0-norm: Number of non-zero entries of the vector

Minimizing $\|\vec{x}\|_0$ under the constraint $\vec{y} = A\vec{x}$

given $\vec{y} \in \mathbb{R}^n$ known vector/vector of measurement

A n x N measurement matrix

Nearly impossible/not practical

Sparse Underdetermined Problem

Signal reconstruction by using L_1-norm

L_1-norm: $$\|\vec{x}\|_1 = \sum_{\ell=1}^{n} |x_\ell|$$

Minimizing $\|\vec{x}\|_1$ under the constraint $\vec{y} = A\vec{x}$

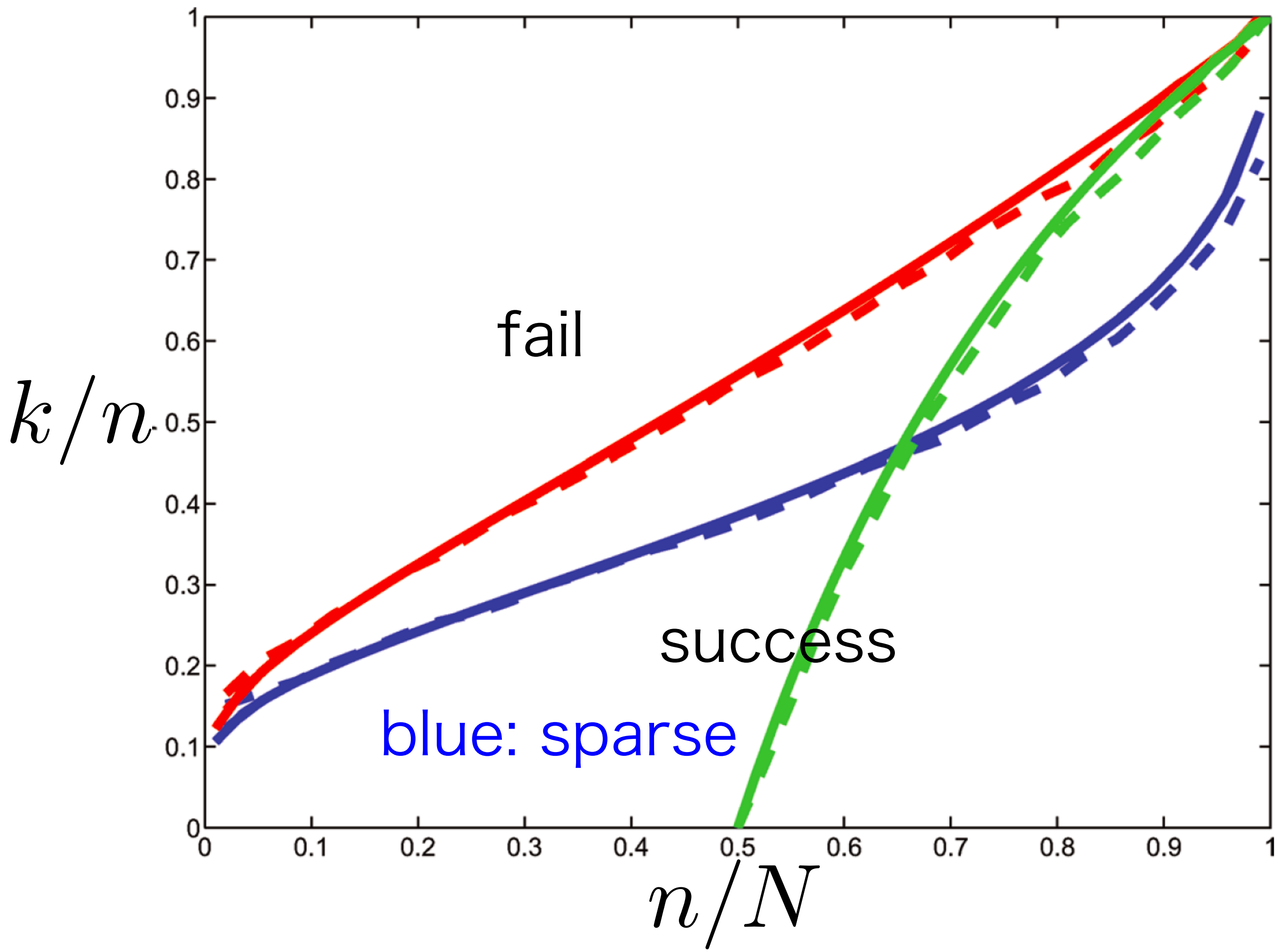
given
$\vec{y} \in \mathbb{R}^n$ known vector/vector of measurement
A n x N measurement matrix

Practical

Sparse Underdetermined Problem

Signal reconstruction by using L_1-norm

D. L. Donoho, A. Maleki, and A. Montanari, PNAS 106, 18914 (2009).

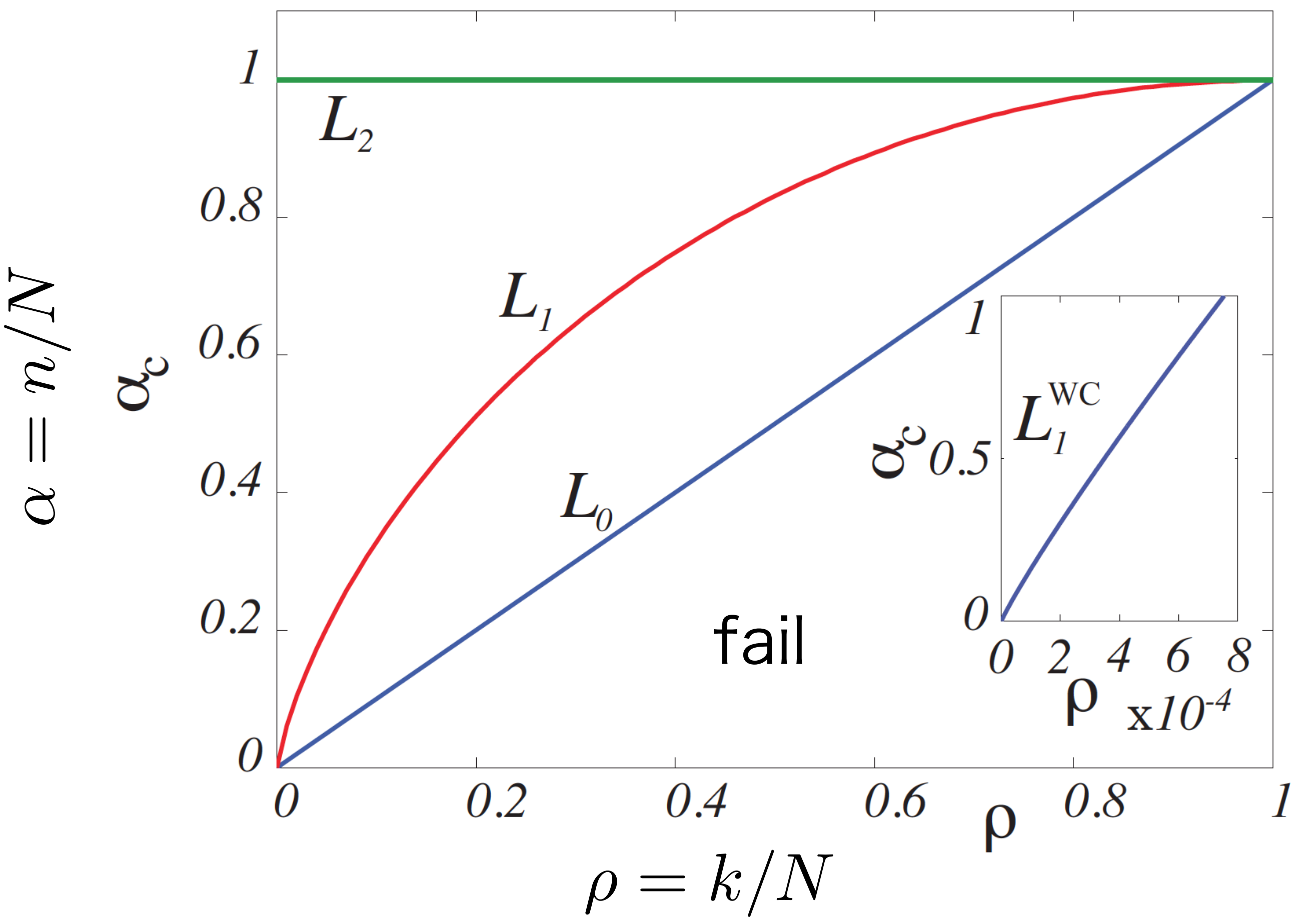


Sparse Underdetermined Problem

Signal reconstruction by using L_n-norm

Y. Kabashima, T. Wadayama, & T. Tanaka, ISIT 2010

$A^T A$: random matrix ensemble and full rank



Where Does Sparseness Come from?

Good choice of basis set

An example: Electron standing wave

Sparseness in Fourier space

J. E. Hoffman, K. McElroy, D.-H. Lee, K. M. Lang, H. Eisaki, S. Uchida, and J. C. Davis, Science 16, 1148 (2002).

How Do We Store Data? -Before going into sparseness

An Example: Gray scale pictures

L x *M* integer matrices

Based on coarse-grained basis set

$$f(x,y) \simeq \widetilde{f}(x,y)$$

$$\widetilde{f}(x,y) = \sum_{\ell=0}^{L-1} \sum_{m=0}^{M-1} a_{\ell m} \Theta_{\ell m}(x,y)$$

“Pixel bases”

$$\Theta_{\ell m}(x,y) = \begin{cases} 1 & (\epsilon\ell < x < \epsilon(\ell+1) \text{ and } \epsilon m < y < \epsilon(m+1)) \\ 0 & (\text{ohterwise}) \end{cases}$$

$$a_{\ell m} \in \mathbb{Z}, \ \ [0, 255]$$

How Do We Store Data? -Before going into sparseness

An Example: Gray scale pictures

L x M integer matrices

Based on coarse-grained basis set

$$\Theta_{\ell m}(x,y) = \theta_\ell(x)\theta_m(y)$$

$$\theta_\ell(x) = \begin{cases} 1 & (\epsilon\ell < x < \epsilon(\ell+1)) \\ 0 & (\text{ohterwise}) \end{cases}$$

$$(A)_{\ell m} = a_{\ell m}$$

"SVD bases"

$$A = U\Sigma V^\dagger$$

$$v_\lambda(y) = \sum_{m=0}^{M-1} V_{m\lambda}\theta_m(y)$$

$$u_\lambda(x) = \sum_{\ell=0}^{L-1} U_{\ell\lambda}\theta_\ell(x)$$

Good bases but not flexible

Simple Orthogonal Basis Sets

Rectangular function

$$\theta_\ell(x) = \begin{cases} 1 & (\epsilon\ell < x < \epsilon(\ell+1)) \\ 0 & (\mathrm{ohterwise}) \end{cases}$$

*Please be careful about the difference
between Heaviside step function and rectangular function

Plane wave (平面波)

$$e^{+i\vec{k}\cdot\vec{r}} = \cos(\vec{k}\cdot\vec{r}) + i\sin(\vec{k}\cdot\vec{r})$$

Wavelet

Transformation by Plane Wave

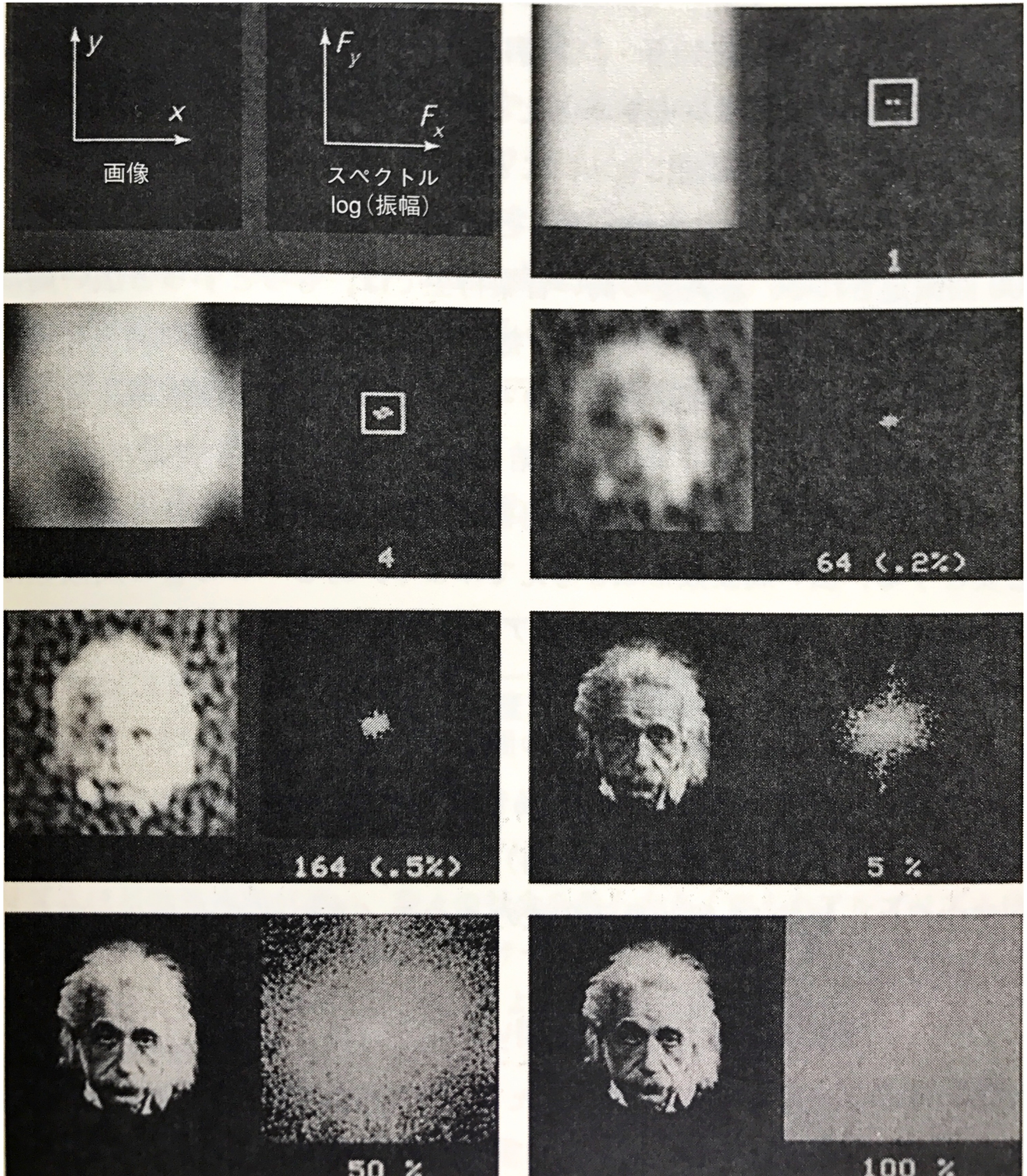
Fourier transformation

$$f(x,y) = \sum_{\ell_x}\sum_{\ell_y} F(\ell_x,\ell_y) e^{+i\frac{2\pi\ell_x}{\Lambda}x + i\frac{2\pi\ell_y}{\Lambda}y}$$

$$x, y \in [0, \Lambda]$$

Inverse Fourier transformation

$$F(\ell_x,\ell_y) = \frac{1}{\Lambda^2}\int_0^{\Lambda} dx \int_0^{\Lambda} dy f(x,y) e^{-i\frac{2\pi\ell_x}{\Lambda}x - i\frac{2\pi\ell_y}{\Lambda}y}$$



B. B. Hubbard,
“Ondes et ondelettes”
(ウェーブレット入門)

Compressed Sensing

An Important Application: Compressed Sensing for MRI

Sparsifying Transformation

M. Lustig, D. Donoho, and J. M. Pauly,
Magnetic Resonance in Medicine 58, 1182 (2007).

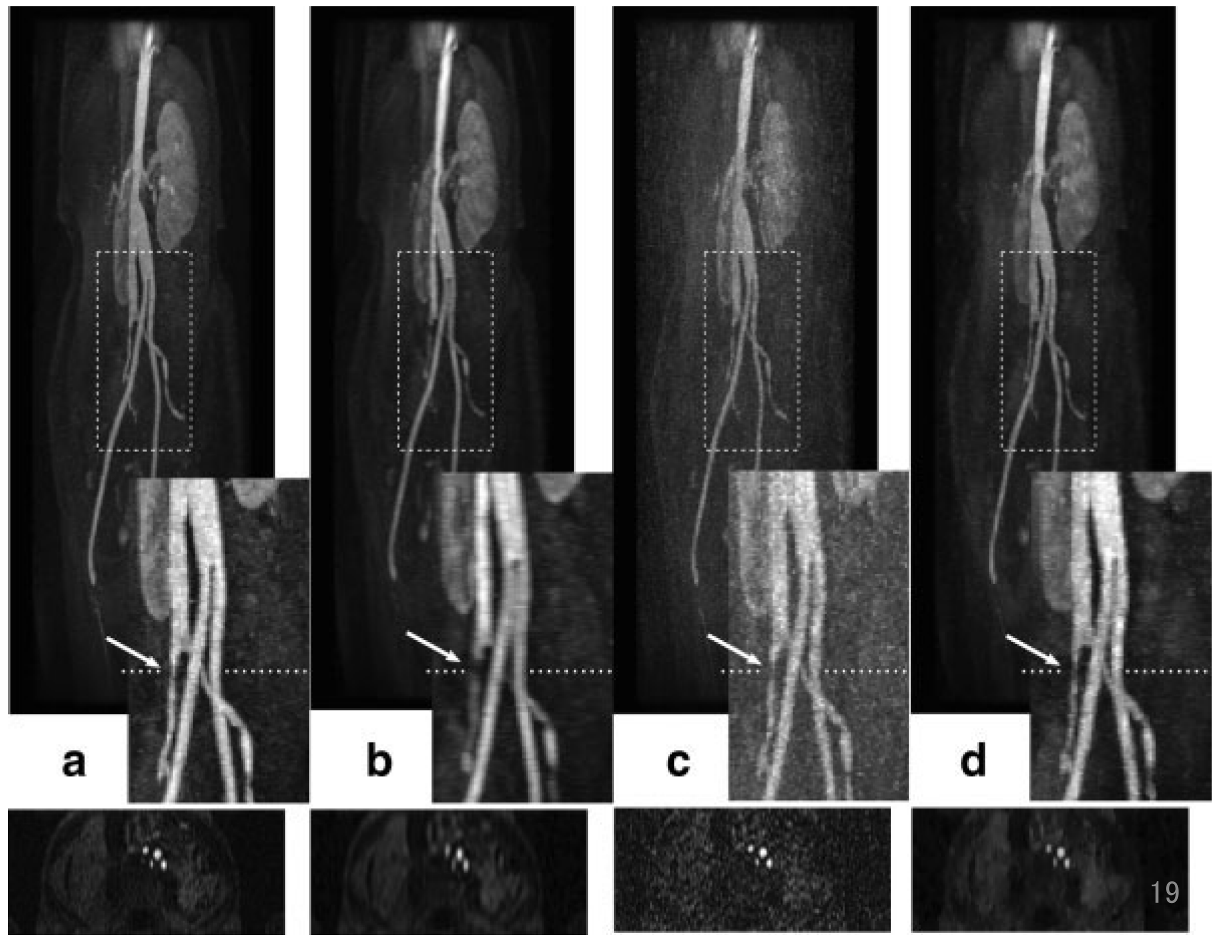
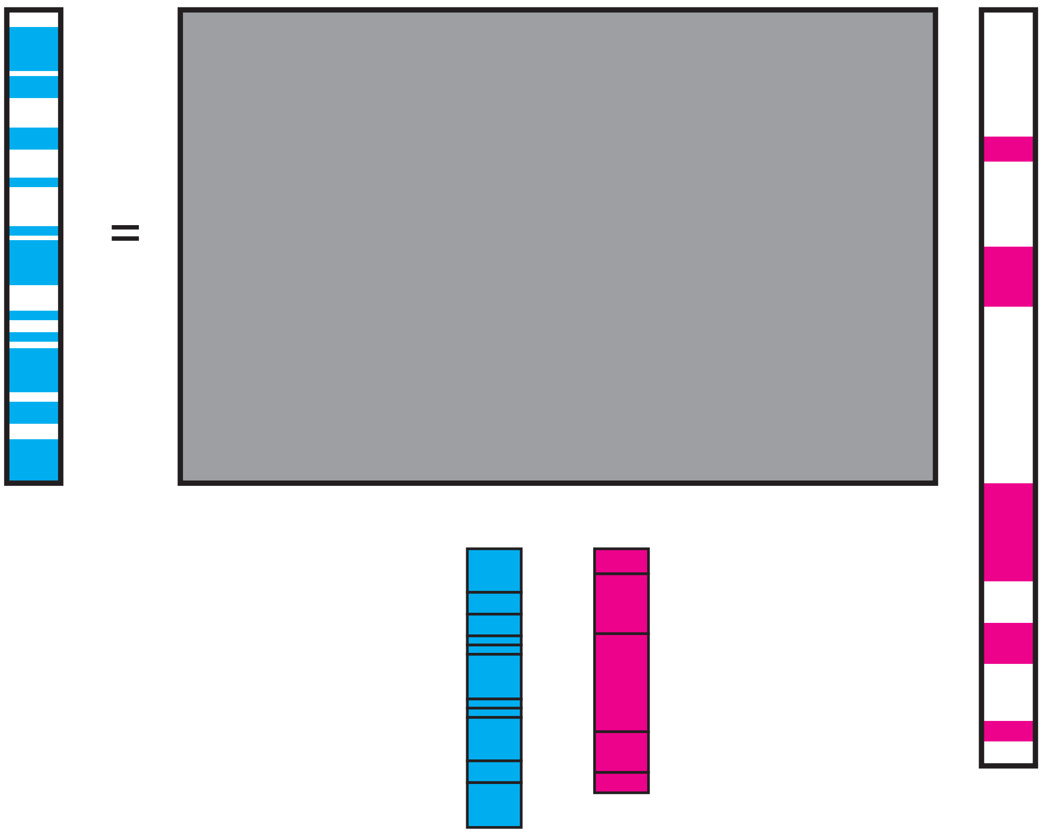


FIG. 10. Reconstruction from 5-fold accelerated acquisition of first-pass contrast enhanced abdominal angiography. (a) Reconstruction from a complete data set. (b) LR (c) ZF-w/dc (d) CS reconstruction from random undersampling. The patient has a aorto-bifemoral bypass graft. This is meant to carry blood from the aorta to the lower extremities. There is a high-grade stenosis in the native right common illiac artery, which is indicated by the arrows. In figure parts (a) and (d) flow across the stenosis is visible, but it is not on (b) and (c).

Compressed Sensing

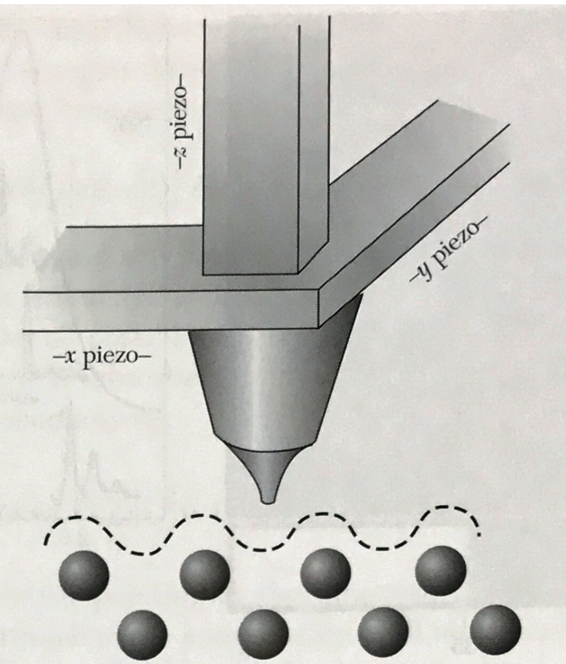
Reconstruction of the signal with undersampled measurements
→ Acceleration of measurements

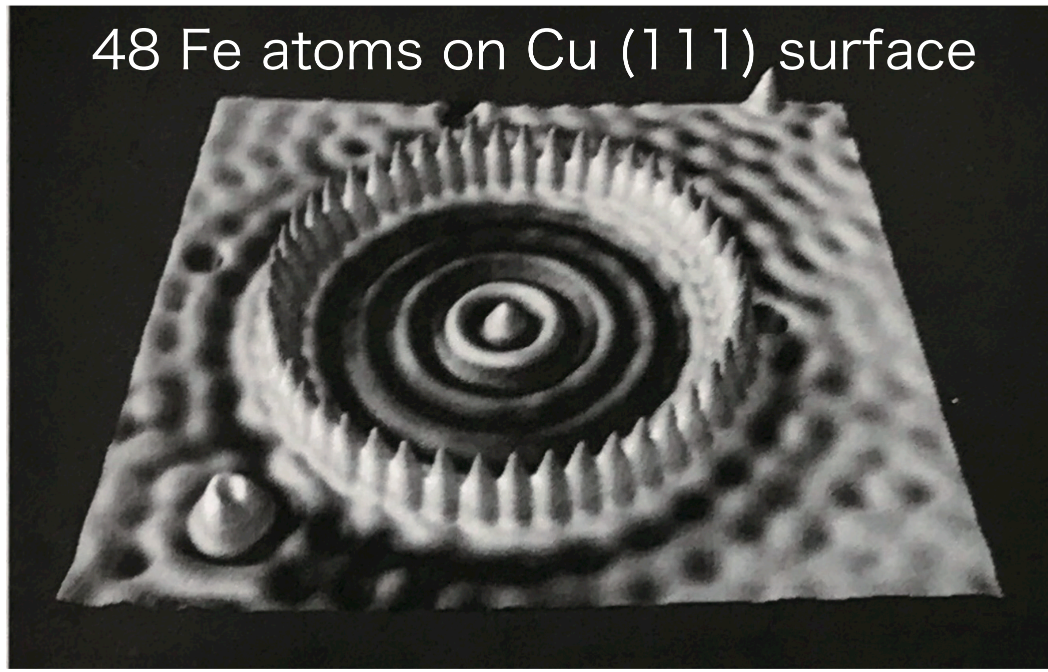


An Example of Compressed Sensing

Standing wave of electrons

Scanning tunneling microscope (STM)





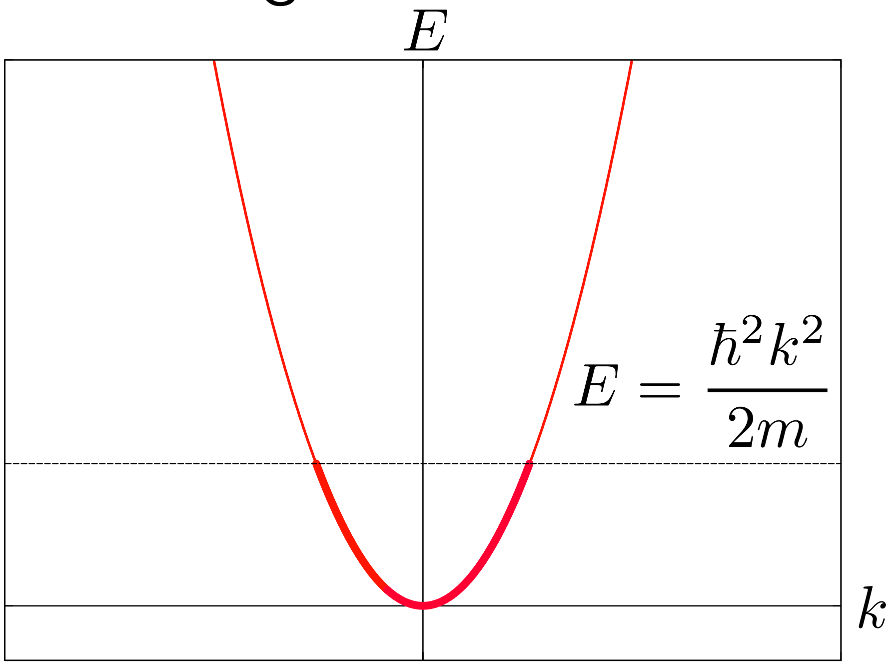
$$\frac{dI}{dV} \propto \sum_{\ell} |\psi_{\ell}(\vec{r})| \delta(E_{\mathrm{F}} + eV - E_{\ell})$$

From "Introduction to Solid State Physics (8th edition),"
C. Kittel (John Wiley & Sons, Inc)

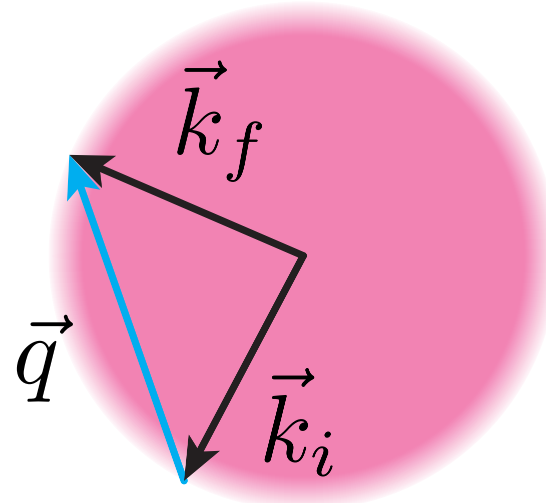
QPI by STM

Quansiparticle Interference (QPI)

Standing wave of electrons by impurities
→ changes in momentum of electrons



Scattering by impurities



-Pauli principle
-Fermi-Dirac distribution
-Conservation of energy

$$|e^{+i\vec{k}_i\cdot\vec{r}} + re^{+i\vec{k}_f\cdot\vec{r}}|^2 = 1 + r^2 + 2r\cos(\vec{q}\cdot\vec{r})$$

An Example of Compressed Sensing

Compressed sensing of QPI

Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa, and M. Okada, J. Phys. Soc. Jpn. 85, 093702 (2016).

Discrete Fourier transformation

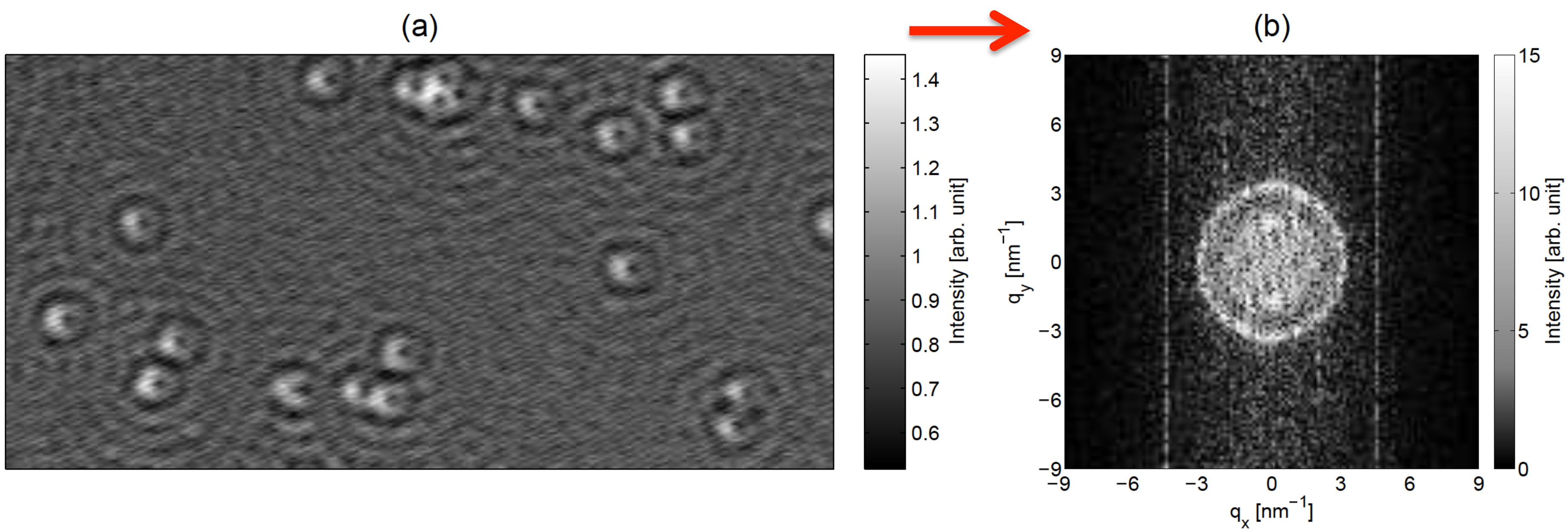


Fig. 1. (a) dI/dV map of Ag(111) surface. (b) FT of (a) obtained by conventional method.

Reconstruction by L_1-Regularization

Fourier transformation

$$\vec{f} = G\vec{F}$$

$$f(j_x, j_y) = \sum_{\ell_x} \sum_{\ell_y} F(\ell_x, \ell_y) e^{+i\frac{2\pi\ell_x}{L_x}j_x + i\frac{2\pi\ell_y}{L_y}j_y}$$

Only if $n = N$, we can perform inverse transformation:

$$F(\ell_x, \ell_y) = \frac{1}{L_x L_y} \sum_{j_x} \sum_{j_y} f(j_x, j_y) e^{-i\frac{2\pi\ell_x}{L_x}j_x - i\frac{2\pi\ell_y}{L_y}j_y}$$

To reconstruct F from undersampled measurements f

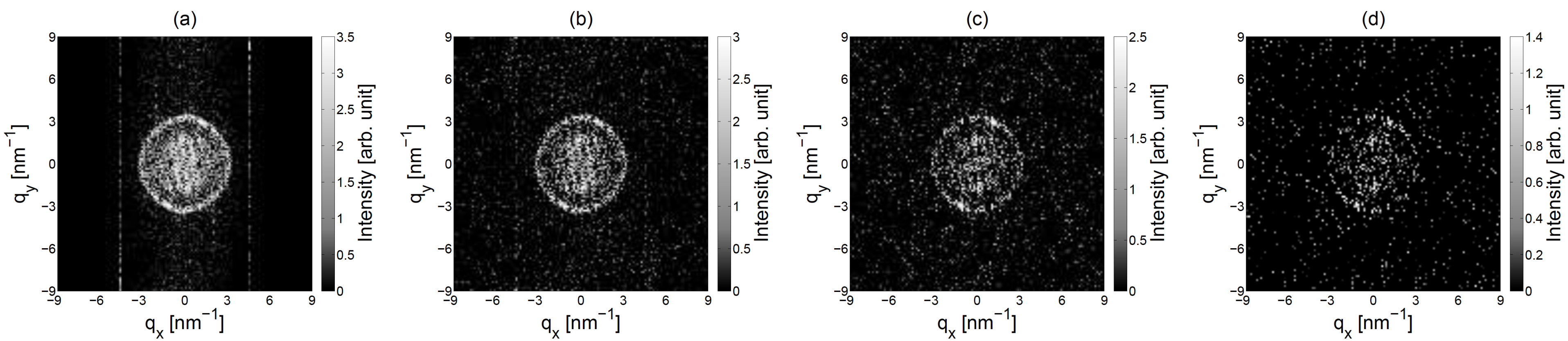
*Minimizing $\|\vec{f} - G\vec{F}\|_2^2 + \lambda\|\vec{F}\|_1$
by changing F with given undersampled f

*L_1-regularization or
least absolute shrinkage and selection operator (LASSO)

Results of Compressed Sensing for QPI

Reconstruction by LASSO

Y. Nakanishi-Ohno, M. Haze, Y. Yoshida, K. Hukushima, Y. Hasegawa, and M. Okada, J. Phys. Soc. Jpn. 85, 093702 (2016).



From full data | From 25% of data | 11% | 6%

Undersampled measurements

Bayesian Connects Underdetermined and Overdetermined Problems

Bayesian is *Elder*

Thomas Bayes
1710-1761
English statistician and Protestant minister

Standard Statistics taught in undergraduate level developed by
Ronald Fisher 1890-1962
Jerzy Neyman 1894-1981
Egon Pearson 1895-1980
Frequentist

Bayesian Statistics Connects Them

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

$n > N$: Overdetermined problem (優決定/過剰決定問題)

Minimizing $\|\vec{y} - A\vec{x}\|_2^2$
$\rightarrow$ Least square fitting

$n < N$: Underdetermined problem (劣決定問題)

Minimizing $\|\vec{y} - A\vec{x}\|_2^2 + \lambda\|\vec{x}\|_1$
$\rightarrow$ LASSO

Bayesian

Bayes' theorem

Sampling/data distribution
(条件付き確率)

$$p(\vec{x}|\vec{y}) = \frac{p(\vec{y}|\vec{x})p(\vec{x})}{p(\vec{y})}$$

prior distribution
(事前確率)
reflecting
-Domain knowledge
-Physical intuition

Posterior density
(事後確率)

$$p(\vec{y}) = \sum_{\vec{x}} p(\vec{y}|\vec{x})p(\vec{x})$$

Bayesian Statistics Connects Them

$n > N$: Overdetermined problem (優決定/過剰決定問題)

$$p(\vec{x}) \propto \text{constant}$$

$$p(\vec{y}|\vec{x}) \propto \exp[-\|\vec{y} - A\vec{x}\|_2^2/\lambda]$$

cf.) Okubo-sensei's
lecture on pseudo inverse

$n < N$: Underdetermined problem (劣決定問題)

$$p(\vec{x}) \propto \exp[-\|\vec{x}\|_1]$$

$$p(\vec{y}|\vec{x}) \propto \exp[-\|\vec{y} - A\vec{x}\|_2^2/\lambda]$$

$$\boxed{p(\vec{x}|\vec{y}) \propto p(\vec{y}|\vec{x})p(\vec{x})}$$

Maximum of the posterior probability $p(\vec{x}|\vec{y})$
$\rightarrow$ Estimation of x

Next Week

1st: Huge data in modern physics
2nd: Information compression in modern physics
3rd: Review of linear algebra
4th: Singular value decomposition and low rank approximation
5th: Basics of sparse modeling
6th: Basics of Krylov subspace methods
7th: Information compression in materials science
8th: Accelerating data analysis: Application of sparse modeling
9th: Data compression: Application of Krylov subspace method
10th: Entanglement of information and matrix product states
11th: Application of MPS to eigenvalue problems
12th: Tensor network representation
13th: Information compression by tensor network renormalization