Information Compression #8
Accelerating data analysis: Application of sparse modeling

Youhei Yamaji
FCS Group, GREEN, National Institute for Materials Science

1. Utilizing low-rank nature
-Recommendation system for materials science
-Restricted Boltzmann machine for many-body systems
2. Sparse modeling

Computational
Science
Alliance
The University of Tokyo

Low-Rank Nature

& Interpolation Based on Low-Rank Nature

Low-Rank Nature

(Naïve assumption) Recommendation system

Synthesizable (合成可能) or not

A. Seko, H. Hayashi, H. Kashima, & I. Tanaka, Phys. Rev. Materials 2, 013805 (2018).

Simon Funk, *Netflix Update: Try This At Home*.
http://sifter.org/~simon/journal/20061211.html

Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

(Physical assumption) Neural network for many-body wave function

Many-body wave function

G. Carleo & M. Troyer, Science 355, 602 (2017).
Y. Nomura, A. S. Darmawan, Y. Yamaji, & M. Imada, Phys. Rev. B 96, 205152 (2017).

Recommendation System by Collaborative Filtering

As a review,
Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Problem:
Recommend *objects* to *users* based on the rating matrix

Example of rating matrix

	Madmax: Fury Road	Pacific Rim	Les Miserables	Skyfall	Creed	Logan
Alice	5			3		
Bob		2			3	
Charlie	3		4			4
Damon	5	5		3		
Eddie			2		5	3

-Number of non-zero entries is small
-Sparseness is not clear *a priori* (演繹的に), at least, for me
Should be validated *a posteriori* (帰納的に)

Data sets for the rating matrix

From Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Dataset	# user	# object	# training score	# test score	density
Movielens	6,040	3,706	982,089	18,120	4.61%
Netflix	480,189	17,770	99,072,112	1,408,395	1.18%
small1	2,917	167	9,734	138	2.00%
small2	4,802	178	702,956	138	82.24%
small3	56	178	9,809	138	98.40%

-Number of non-zero entries is really small!

Singular Value Decomposition of Partially Unknown Matrix

As a review,
Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Mathematical formulation: Input

Rating matrix

$$V = \begin{bmatrix} 5 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 3 & 0 \\ 3 & 0 & 4 & 0 & 0 & 4 \\ 5 & 5 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 & 5 & 3 \end{bmatrix} \qquad V \in \mathbb{R}^{n \times m}$$

Indicator

$$I = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \qquad I \in \{0, 1\}^{n \times m}$$

Singular Value Decomposition of Partially Unknown Matrix

As a review,
Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Mathematical formulation

Rating matrix

$$V = \begin{bmatrix} 5 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 3 & 0 \\ 3 & 0 & 4 & 0 & 0 & 4 \\ 5 & 5 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 & 5 & 3 \end{bmatrix} \qquad V \in \mathbb{R}^{n \times m}$$

Infer the SVD of the *complete* rating matrix from V

$$\widetilde{V} \simeq U^T M \qquad U \in \mathbb{R}^{f \times n}, \quad M \in \mathbb{R}^{f \times m}$$

-Rank f approximation

Singular Value Decomposition of Partially Unknown Matrix

As a review,
Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

Mathematical formulation

Minimize the cost function

$$E = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} I_{ij}[V_{ij} - p(U_i, M_j)]^2$$

$$+ \frac{k_u}{2}\sum_{i=1}^{n}\|U_i\|_2^2 + \frac{k_m}{2}\sum_{j=1}^{m}\|M_j\|_2^2$$

-L_2 regularization

Feature vectors

$$U_i = (U_{1i}, U_{2i}, \ldots, U_{fi})^T$$

$$M_j = (M_{1j}, M_{2j}, \ldots, M_{fj})^T$$

A prediction function for the rating (You need to choose)

$$p(U_i, M_j) = \begin{cases} 1 & \text{if } U_i^T M_j < 0 \\ 1 + U_i^T M_j & \text{if } 0 < U_i^T M_j < 4 \\ 5 & \text{if } 4 < U_i^T M_j \end{cases}$$

Singular Value Decomposition of Partially Unknown Matrix

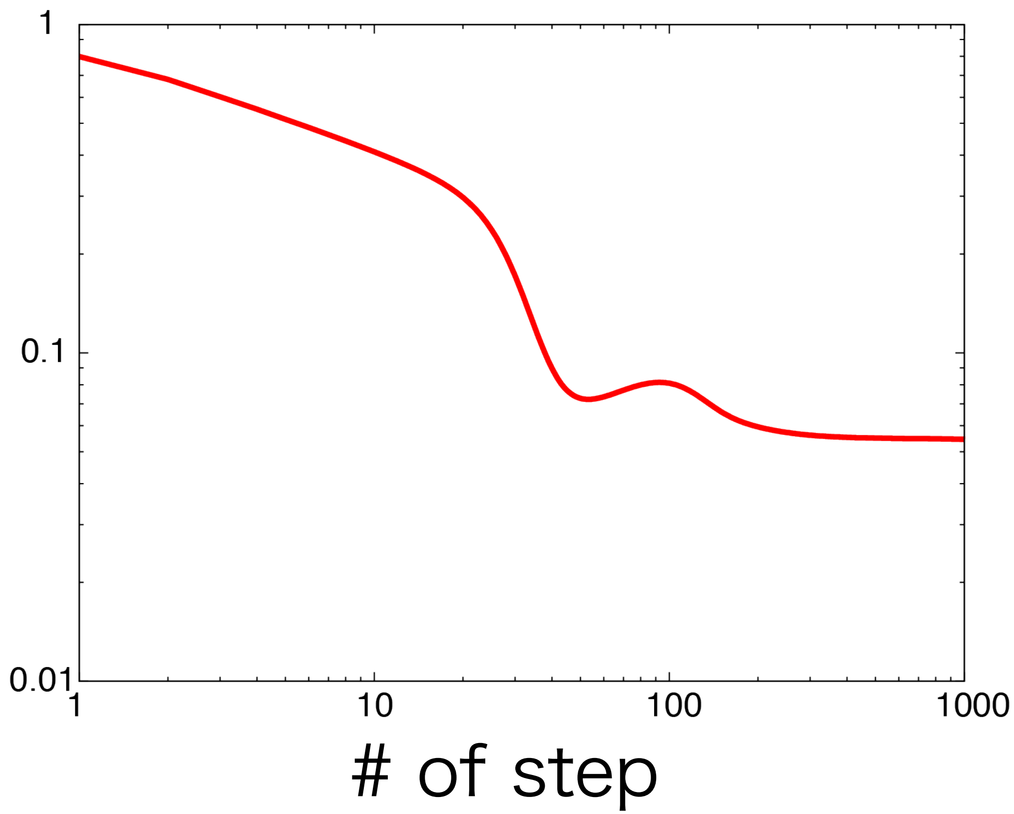
$$\mathrm{RMSE}(p, V) = \sqrt{\frac{\sum_{i,j} I_{ij}[V_{ij} - p(U_i, M_j)]^2}{\sum_{i,j} I_{ij}}}$$

Optimization:
Steepest descent

$$U_i \leftarrow U_i - \mu \nabla_U E$$
$$M_j \leftarrow M_j - \mu \nabla_M E$$

Parameters

$$f = 3$$
$$\mu = 0.1$$
$$k_u = k_m = 0.1$$



Singular Value Decomposition of Partially Unknown Matrix

Original rating matrix

$$V = \begin{bmatrix} 5 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 3 & 0 \\ 3 & 0 & 4 & 0 & 0 & 4 \\ 5 & 5 & 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 & 5 & 3 \end{bmatrix}$$

A predicted rating matrix

$$p(U_i, M_j) = \begin{bmatrix} 4.92 & 4.39 & 2.29 & 2.96 & 4.93 & 3.23 \\ 2.57 & 2.02 & 1.07 & 1.78 & 2.94 & 1.60 \\ 3.01 & 3.82 & 3.92 & 2.05 & 2.55 & 3.94 \\ 4.98 & 4.91 & 3.11 & 3.00 & 4.64 & 3.85 \\ 4.55 & 3.85 & 2.01 & 2.78 & 4.92 & 2.98 \end{bmatrix}$$

-$f > 2$ is enough (for Netflix, $f \sim 100$-1000)

-Should be validated (but the data set is too small)

Application of Recommendation System to Materials Science

A goal of materials science:
Synthesize a compound with desired functions

Difficulty:
Process of synthesis is complicated and costs much time and resources

→ *Virtual screening* is desirable

Virtual screening by recommendation system

A recommendation system tells us whether a given chemical composition (化学組成) is synthesizable or not

A. Seko, H. Hayashi, H. Kashima, & I. Tanaka, Phys. Rev. Materials 2, 013805 (2018).

-Crystal structure is another important problem

Application of Recommendation System to Material Science

A. Seko, H. Hayashi, H. Kashima, & I. Tanaka, Phys. Rev. Materials 2, 013805 (2018).

Example: Ternary compounds (３元化合物) $A_a B_b X_x$

-Generating possible chemical compositions

TABLE I. Elements included in the datasets of known CRCs. Rare-earth elements shown in parentheses are excluded in the candidate quaternary and quinary compositions.

Cations	Li,Na,K,Rb,Cs,Be,Mg,Ca,Sr,Ba,Zn,Cd,Hg, B,Al,Sc,Y,La,Ga,In,Tl,Si,Ge,Sn,Pb,P,As,Sb,Bi Ti,Zr,Hf,V,Nb,Ta,Cr,Mo,W,Mn,Tc,Re, Fe,Ru,Os,Co,Rh,Ir,Ni,Pd,Pt,Cu,Ag,Au, (Ce,Pr,Nd,Pm,Sm,Eu,Gd,Tb,Dy,Ho,Er,Tm,Yb,Lu)
Anions	C,N,O,S,Se,Te,F,Cl,Br,I

-Synthesis is reported → non-zero entry 1

	Ternary	Quaternary	Quinary
ICSD	9,313	7,742	1,321
ICDD	2,369 (9,278)	2,647 (7,864)	639 (1,326)
SpMat	2,708 (10,461)	3,066 (8,141)	1,169 (1,893)
ICDD+SpMat	4,134 (12,573)	4,961 (11,307)	1,616 (2,562)
Candidates	7,405,200	1,188,038,460	23,104,706,560

Application of Recommendation System to Material Science

A. Seko, H. Hayashi, H. Kashima, & I. Tanaka, Phys. Rev. Materials 2, 013805 (2018).

Example: Ternary compounds (３元化合物) $\mathrm{A}_a\mathrm{B}_b\mathrm{X}_x$

-Schematic rating matrix

$$
\begin{array}{c}
\begin{array}{cccccccccccc}
\text{Li,O} & \text{Li,O} & \text{Li,O} & \text{Li,O} & \text{Li,O} & & \text{In,O} & \text{In,O} & \text{In,O} & \text{In,O} & \text{In,O} \\
(1,1,1) & (1,1,2) & (1,2,2) & (1,3,3) & (1,4,3) & & (1,1,2) & (1,1,3) & (1,2,4) & (2,2,5) & (3,3,8)
\end{array} \\
\left(\begin{array}{ccccccccccc}
0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 1 & 1 & 1 & 0 & \cdots & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & 0
\end{array}\right)
\begin{array}{l}
\text{Na} \\ \text{Mg} \\ \text{Al} \\ \\ \text{Cu} \\ \text{Zn} \\ \text{Ga}
\end{array}
\end{array}
$$

Interpolation by Using Low-Rank Nature

(Naïve assumption) Recommendation system

Synthesizable (合成可能) or not

A. Seko, H. Hayashi, H. Kashima, & I. Tanaka, Phys. Rev. Materials 2, 013805 (2018).

Simon Funk, *Netflix Update: Try This At Home*.
http://sifter.org/~simon/journal/20061211.html

Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

(Physical assumption)
Neural network for many-body wave function

Many-body wave function

G. Carleo & M. Troyer, Science 355, 602 (2017).
Y. Nomura, A. S. Darmawan, Y. Yamaji, & M. Imada, Phys. Rev. B 96, 205152 (2017).

Restricted Boltzmann Machine (RBM)

RBM: A simple artificial neural network

-A learning machine well mimicking smooth functions

RBM for L qubits wave function

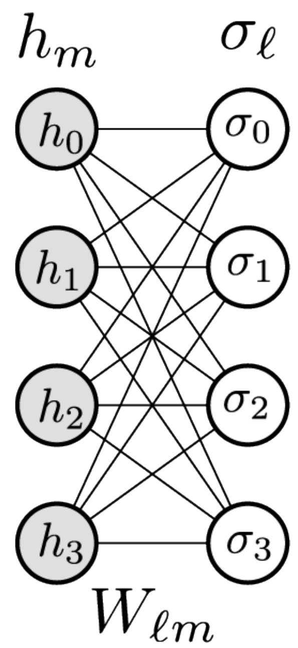
$$x = (\sigma_0, \sigma_1, \ldots, \sigma_{L-1})$$
$$\sigma_\ell = \pm 1$$

$$\mathcal{N}(x) = \sum_{\{h_k\}=\pm 1} \exp\left(\sum_\ell a_\ell \sigma_\ell + \sum_{\ell,m} \sigma_\ell W_{\ell m} h_m + \sum_m b_m h_m \right)$$

-Variational parameter optimized by
 enforced learning (cf.) natural gradient)

$\{a_\ell, W_{\ell m}, b_m\} \in \mathbb{R}^{L+\alpha L^2+\alpha L}$

$\rightarrow$ RBM is positive definite



RBM: Sparse representation

$$2^L \rightarrow L + \alpha L^2 + \alpha L$$

Restricted Boltzmann Machine (RBM)

G. Carleo & M. Troyer, Science 355, 602 (2017).

Application of RBM to 1D/2D Heisenberg model

$$\hat{H} = -J \sum_{\ell} \left(\hat{S}^x_\ell \hat{S}^x_{\ell+1} + \hat{S}^y_\ell \hat{S}^y_{\ell+1} - \hat{S}^z_\ell \hat{S}^z_{\ell+1} \right)$$

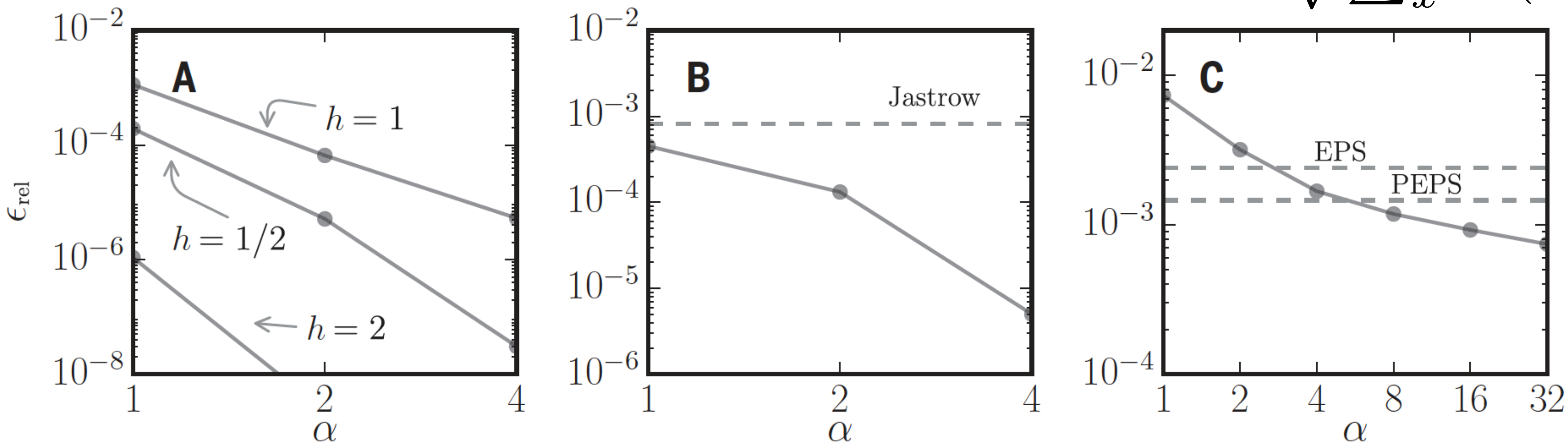
-Equivalent to antiferromagnetic Heisenberg model

$$x = (\sigma_0, \sigma_1, \ldots, \sigma_{L-1}) \qquad \sigma_\ell = \pm 1 \rightarrow \uparrow, \downarrow$$

$$\sigma_\ell = \pm 1$$

Minimize cost function

$$E = \langle \phi | \hat{H} | \phi \rangle \qquad |\phi\rangle = \frac{\sum_x \mathcal{N}(x) |x\rangle}{\sqrt{\sum_x \mathcal{N}(x)^2}}$$



Sparse Modeling
& Its Application to STM

Underdetermined Problem (Revisited)

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

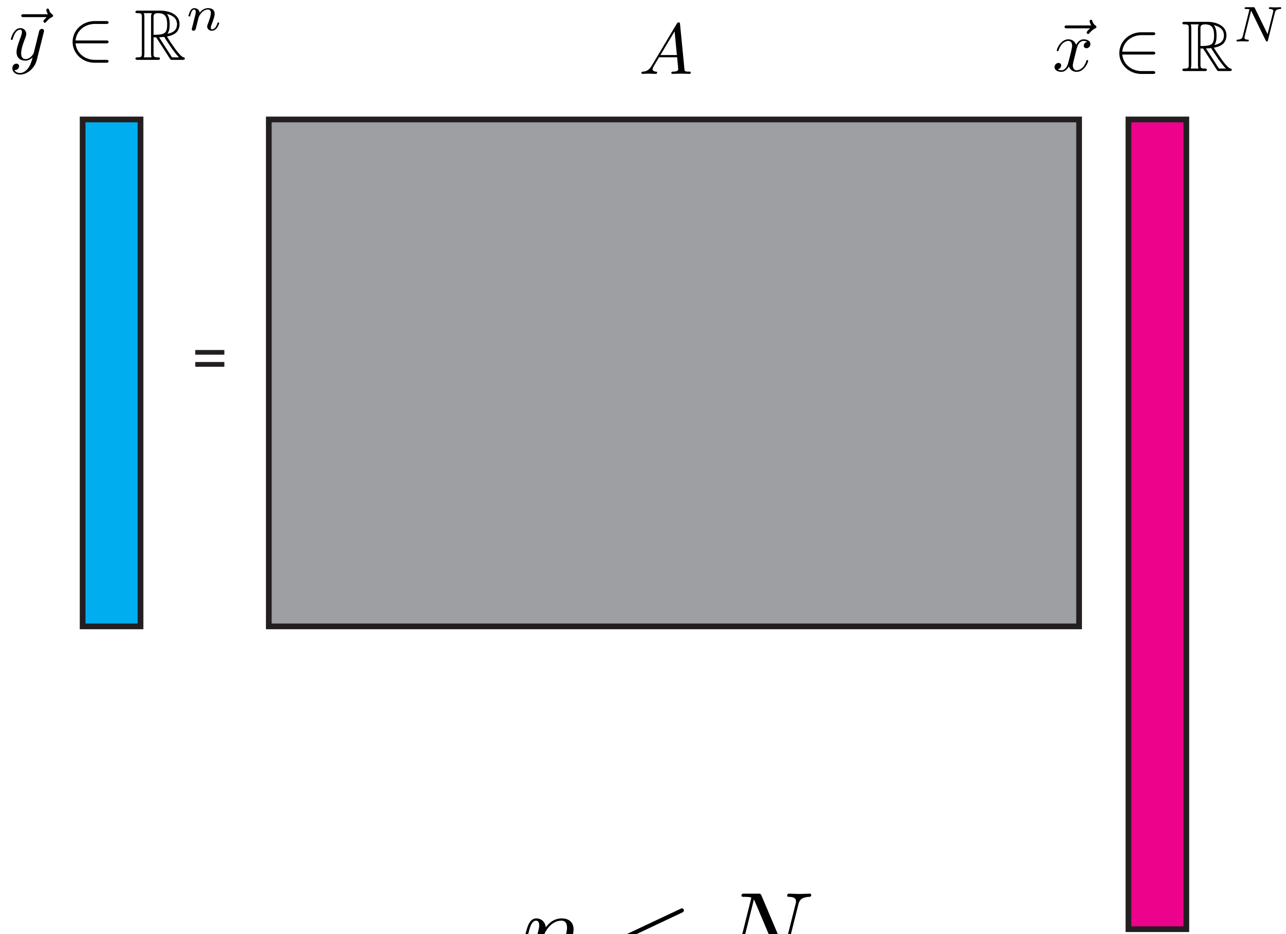
$$\vec{y} = A\vec{x}$$

$\vec{y} \in \mathbb{R}^n$	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)

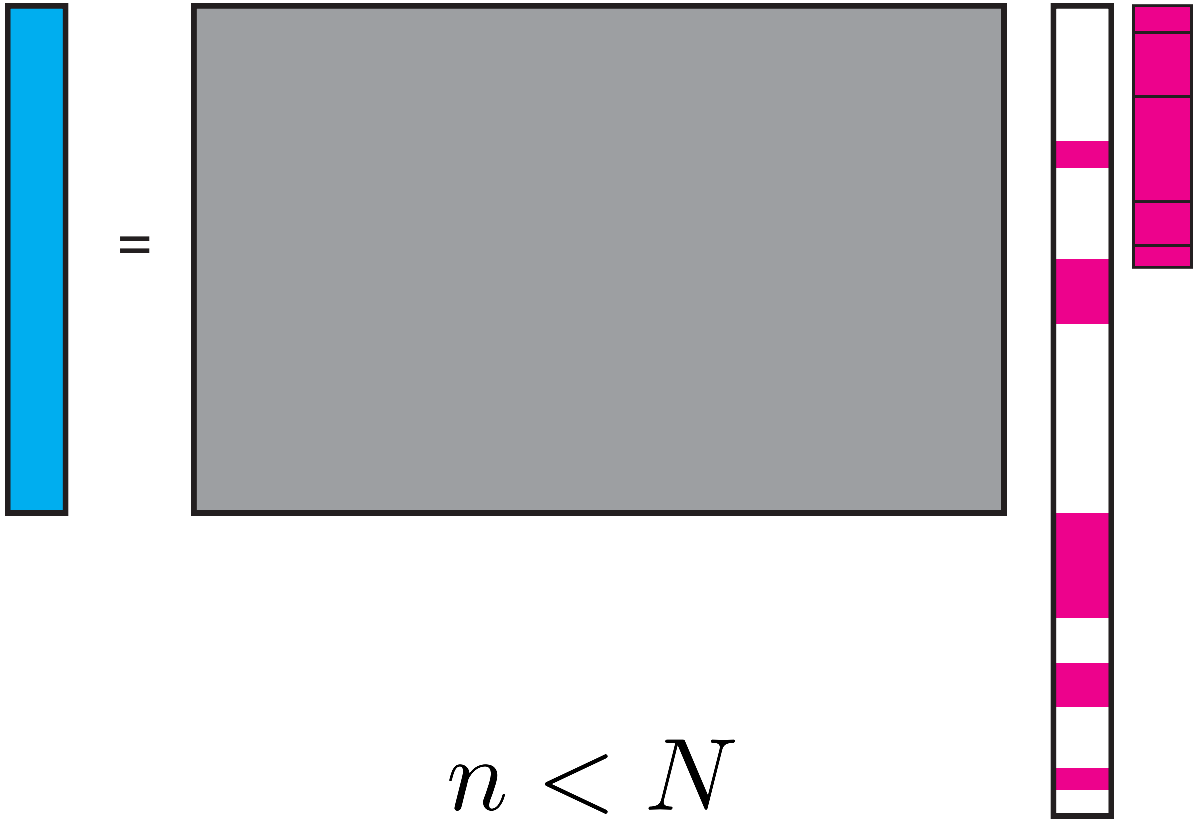
If we know that the unknown vector has as many as k non-zero entries and $k < n$, we can reconstruct the unknown vector.



$$n < N$$

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 (Revisited)

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 (Revisited)

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

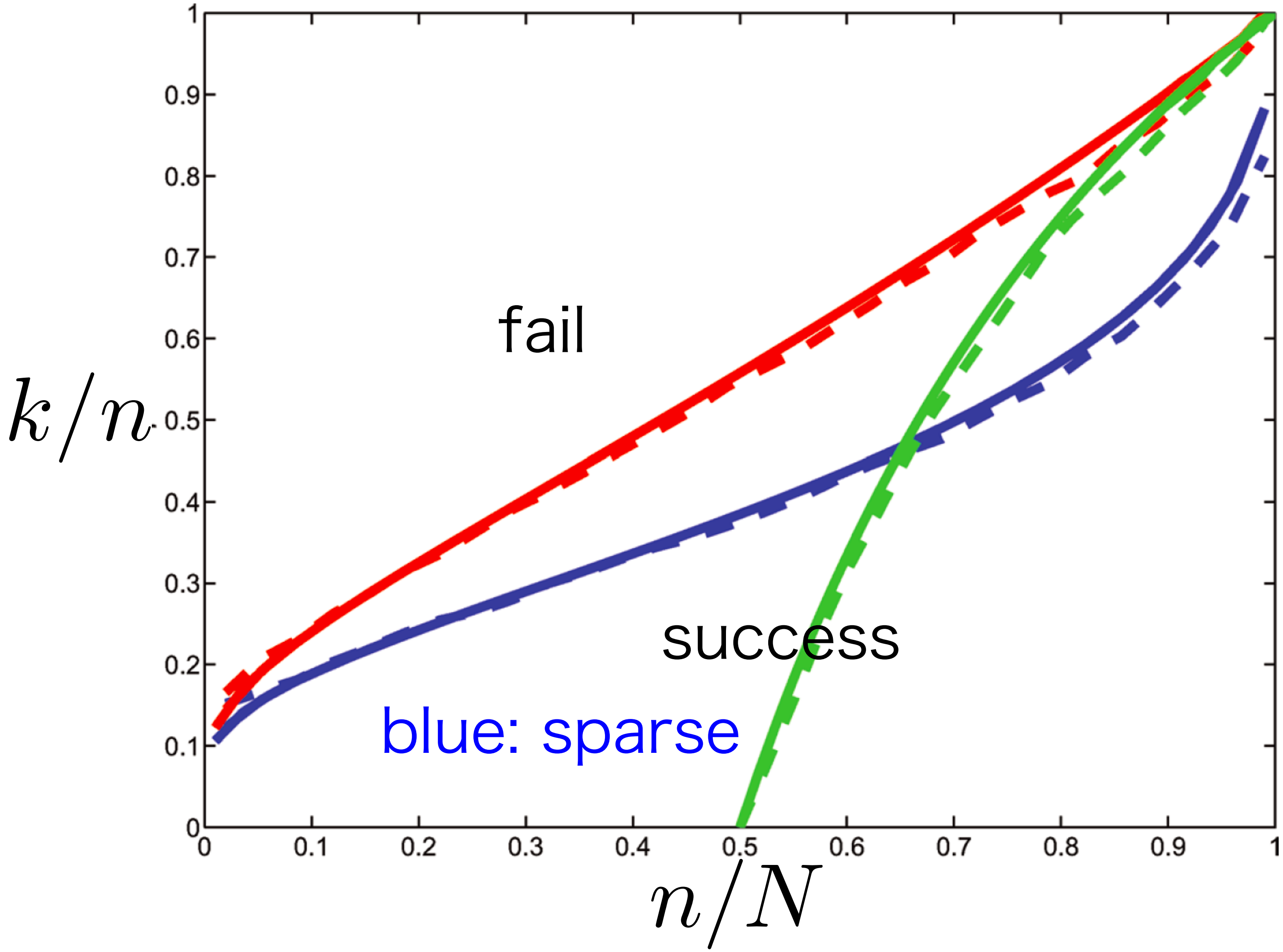
A n x N measurement matrix

Practical

Sparse Underdetermined Problem (Revisited)

Signal reconstruction by using L_1-regularization

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

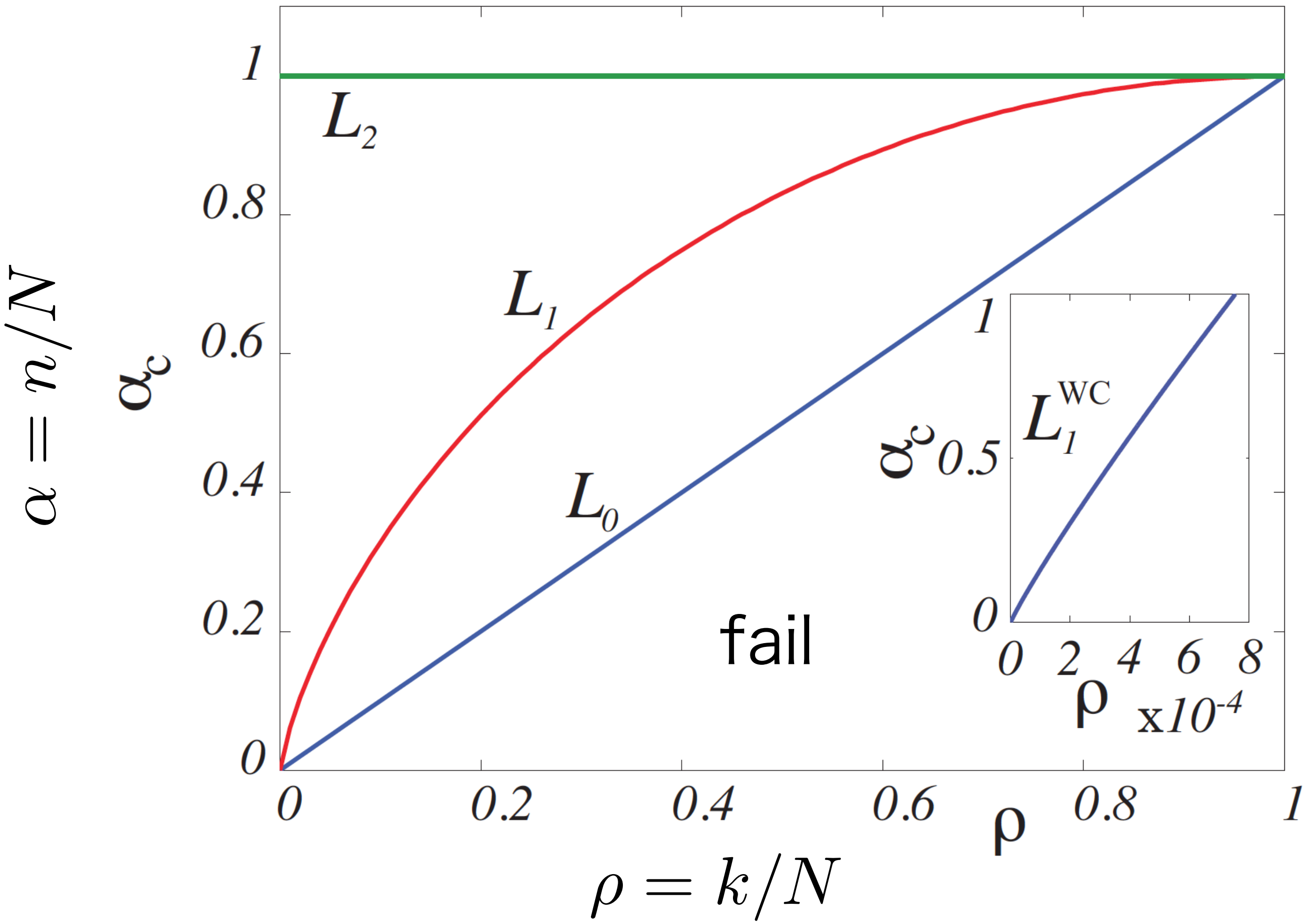


Sparse Underdetermined Problem (Revisited)

Signal reconstruction by using L_n-regularization

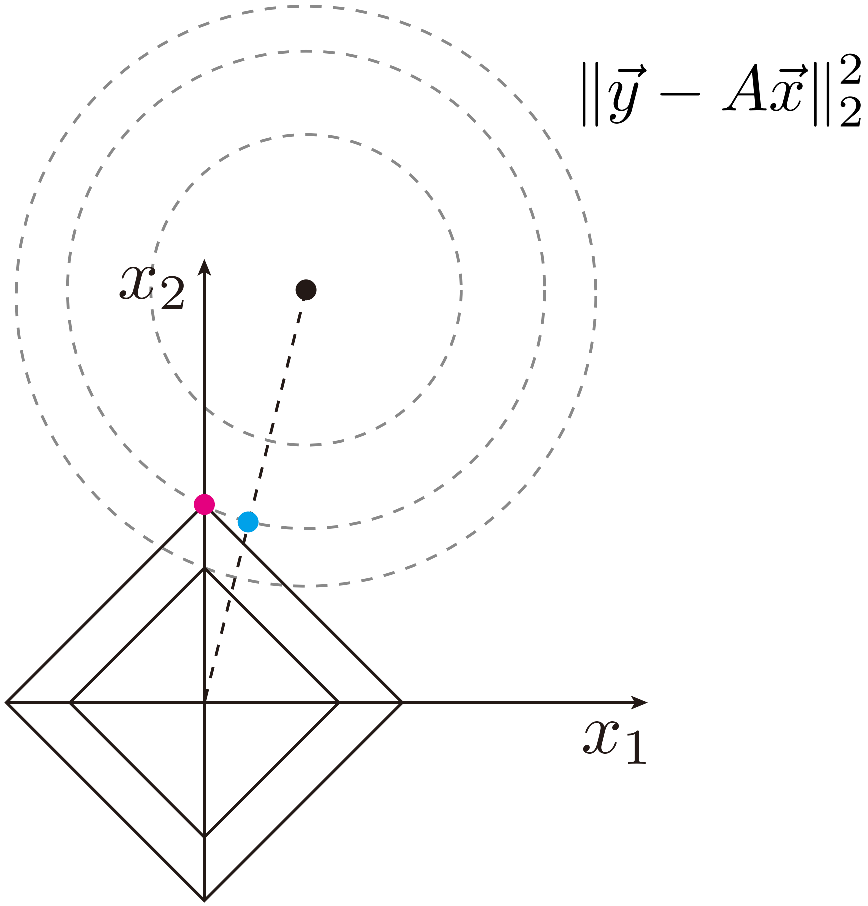
Y. Kabashima, T. Wadayama, & T. Tanaka, ISIT 2010

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

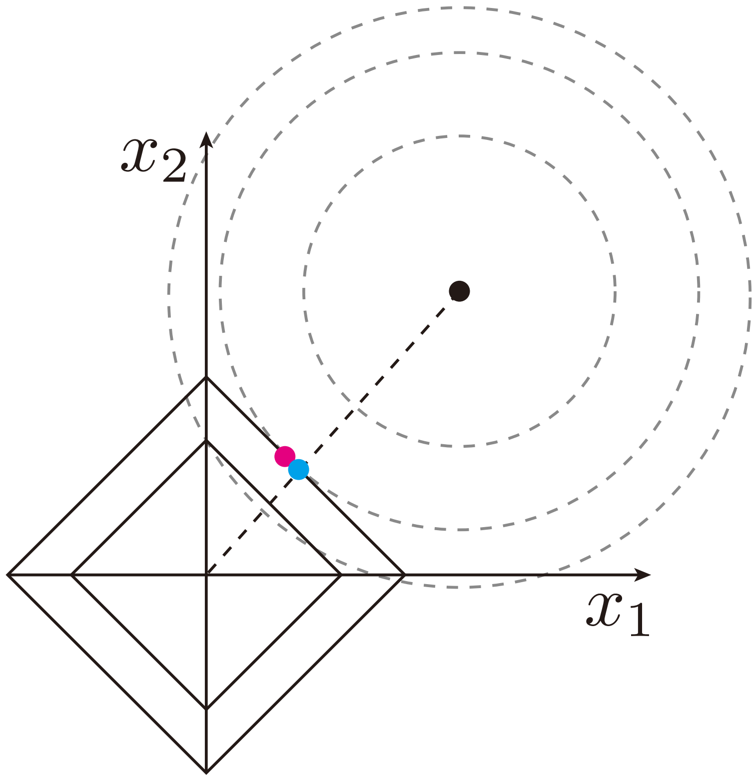


What Regularization Does?

1. L_1 will select a sparse solution



2. Dense solution



solution by L_2-regularization

solution by L_1-regularization

Bias-Variance Decomposition

Trade-off between bias and variance

C. M. Bishop,
Pattern Recognition and Machine Learning (Springer-Verlag, 2006)

Solving inverse problem: Minimizing $\|\vec{y} - A\vec{x}\|_2^2 + \lambda\|\vec{x}\|_q$

$(q = 1, 2)$

-Assume $y_i = \sum_j A_{ij} x_j^* + \epsilon_i$

$\{x_j^*\}$: true signal

$\{\epsilon_i\}$: gauss noise

$$\mathbb{E}_\epsilon[\sum_j (x_j - x_j^*)^2] = \mathbb{E}_\epsilon[\{\sum_j (\mathbb{E}_\epsilon[x_j] - x_j^*)\}^2] + \mathbb{E}_\epsilon[\{\sum_j (x_j - \mathbb{E}_\epsilon[x_j])\}^2]$$

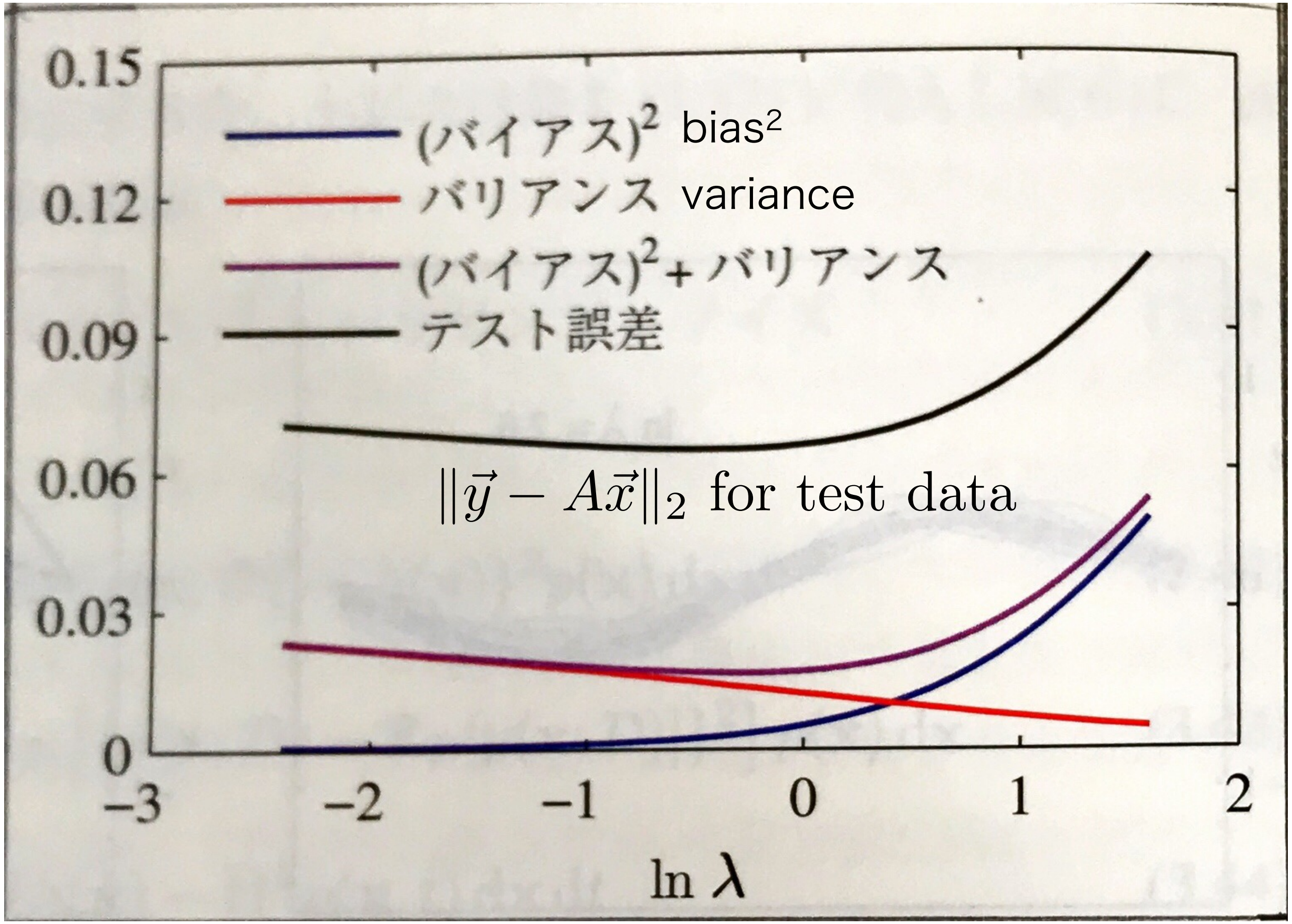
bias^2 variance

$$\mathbb{E}_\epsilon[f(\epsilon)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int d\epsilon p(\epsilon) \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) f(\epsilon)$$

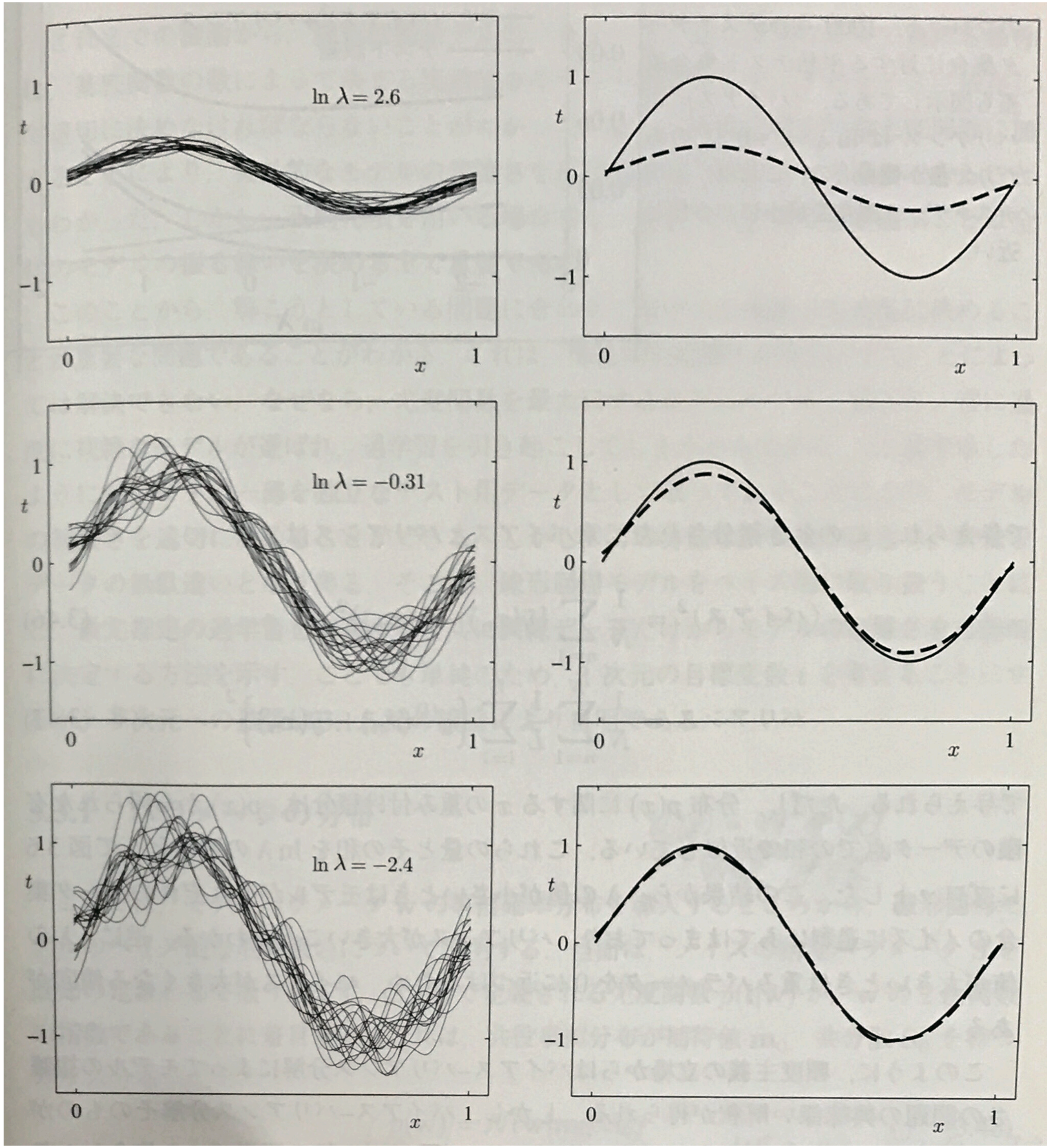
Bias-Variance Decomposition

Trade-off between bias and variance

C. M. Bishop,
Pattern Recognition and Machine Learning (Springer-Verlag, 2006)



Gaussian



Determination of λ

In general, it is difficult to calculate $\mathbb{E}_\epsilon[\sum_j (x_j - x_j^*)^2]$

Determination of λ
-Cross validation
-Information criterion
-Bootstrap

cf.) Akaike information criterion

L_1 regularization for Overdetermined Problems: Model Selection

Overdetermined problems

$$\vec{y} = A\vec{x} \qquad (n > N)$$

known $\vec{y} \in \mathbb{R}^n$
unknown $\vec{x} \in \mathbb{R}^N$
known $A \in \mathbb{R}^{n \times N}$

→ Model selection

Minimization of $\|\vec{y} - X\vec{\beta}\|_2^2 + \lambda\|\vec{\beta}\|_1$
to find a fitting model for data $\vec{y} \in \mathbb{R}^n$

$$(n > p)$$

known $\vec{y} \in \mathbb{R}^n$
unknown $\vec{\beta} \in \mathbb{R}^p$
known $X \in \mathbb{R}^{n \times p}$

L_1 regularization for Overdetermined Problems: Model Selection

Example of data

Table 2.1 *Crime data: Crime rate and five predictors, for $N = 50$ U.S. cities.*

city	funding	hs	not-hs	college	college4	crime rate
1	40	74	11	31	20	478
2	32	72	11	43	18	494
3	57	70	18	16	16	643
4	31	71	11	25	19	341
5	67	72	9	29	24	773
⋮	⋮	⋮	⋮	⋮		
50	66	67	26	18	16	940

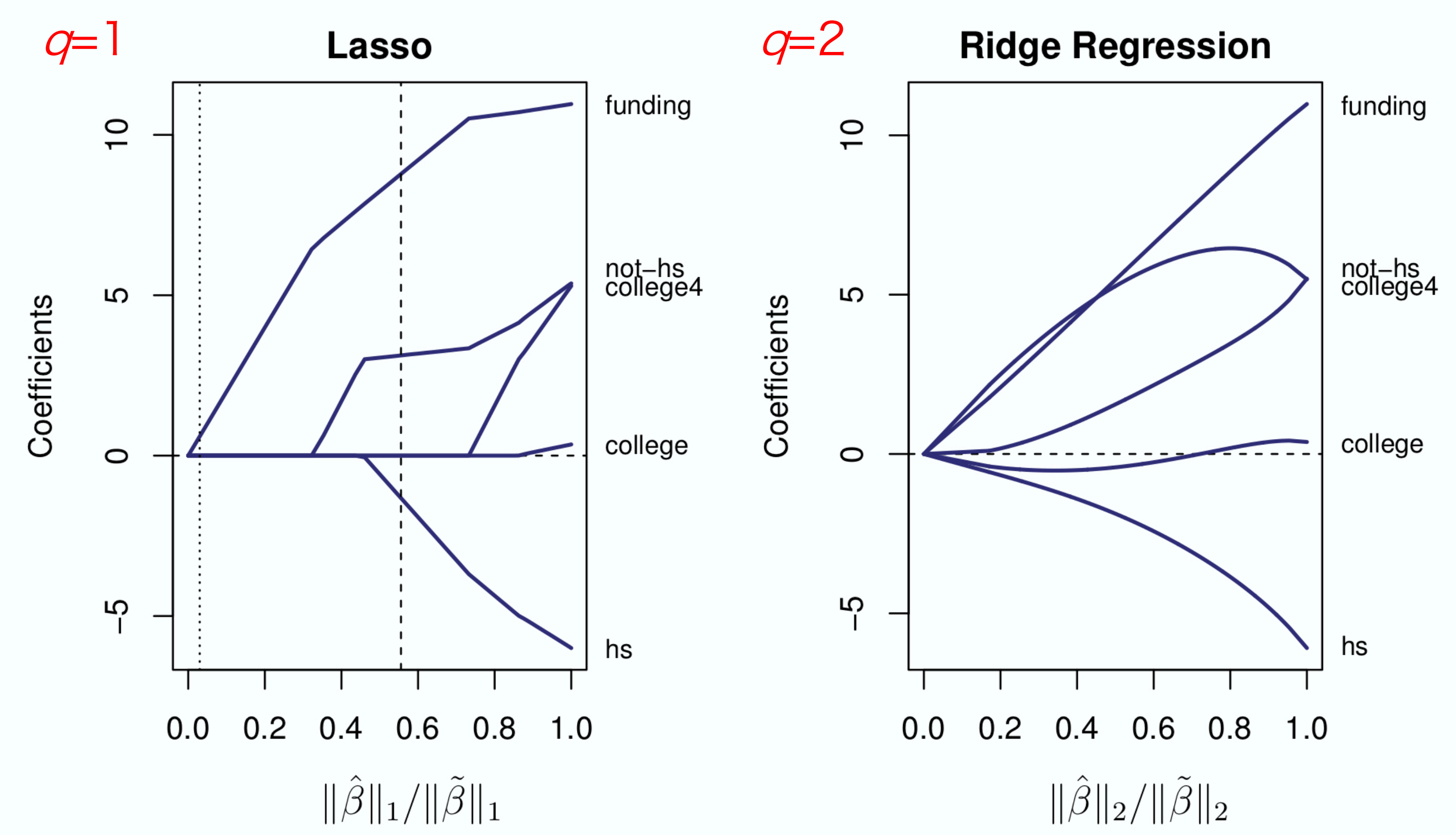
$X \in \mathbb{R}^{50\times 5}$ $\vec{y} \in \mathbb{R}^{50}$

G. S. Thomas, The Rating Guide to Life in America's Small Cities (Prometheus books, 1990)

T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalization* (CRC Press, 2015)

L_1 regularization for Overdetermined Problems: Model Selection

T. Hastie, R. Tibshirani, and M. Wainwright,
Statistical Learning with Sparsity: The Lasso and Generalization (CRC Press, 2015)



$\hat{\beta}$: Optimized by Lq regularization (q=1,2)

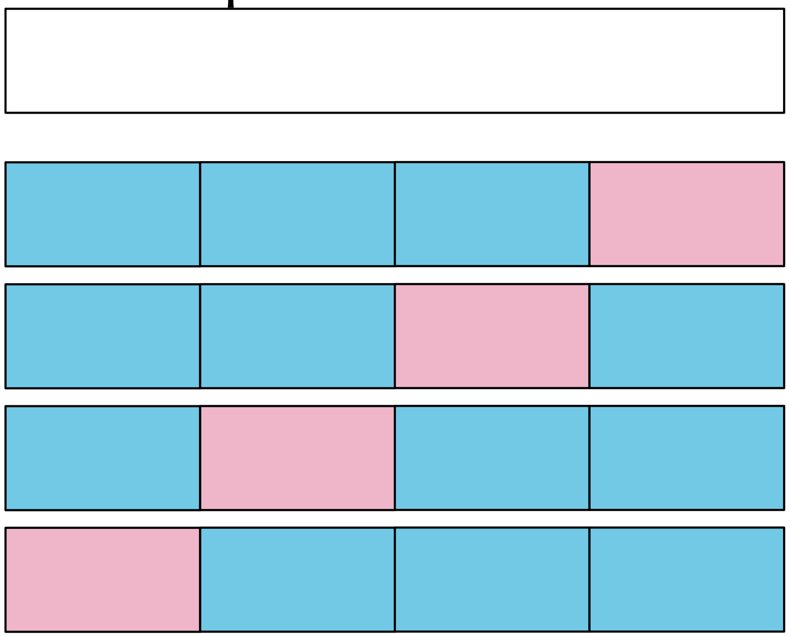
$\tilde{\beta}$: Obtained by least square fitting

How to Choose λ: An Example Cross Validation (CV)

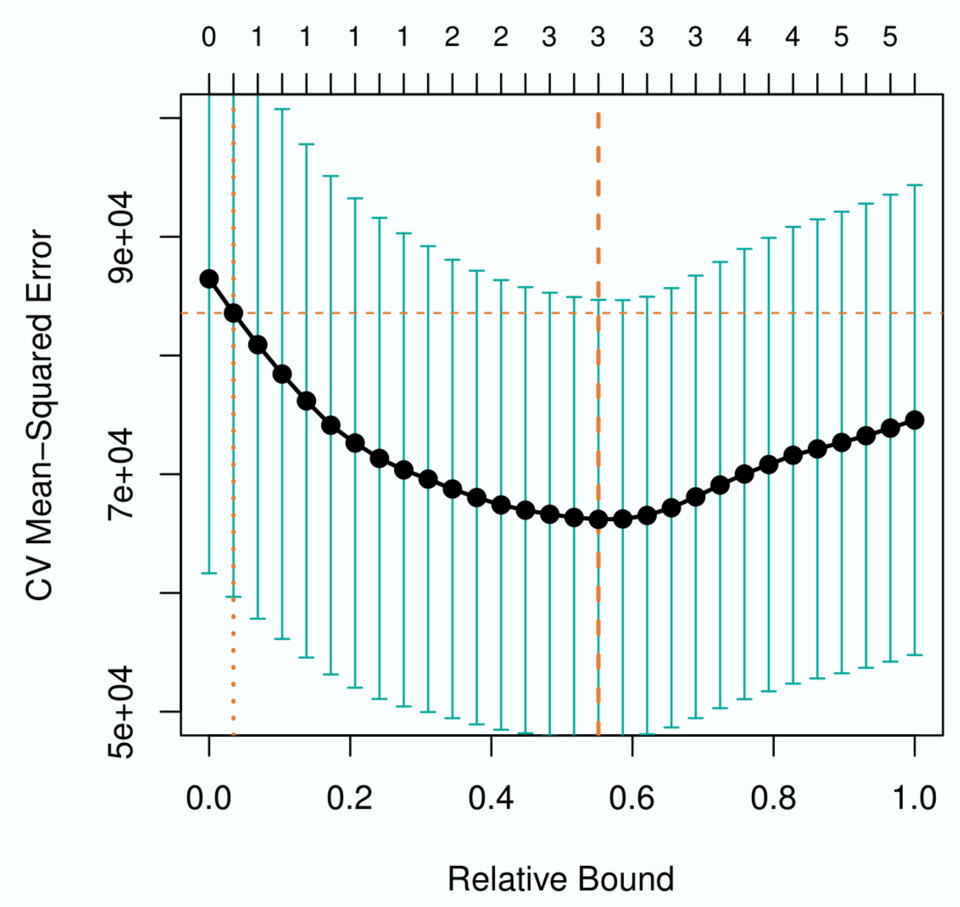
K fold cross validation

-Divided into training and test
-Optimize β by training data and estimate test error for each set
-Average test errors over K sets

An example: 4-fold cv



10-fold cv for crime rate

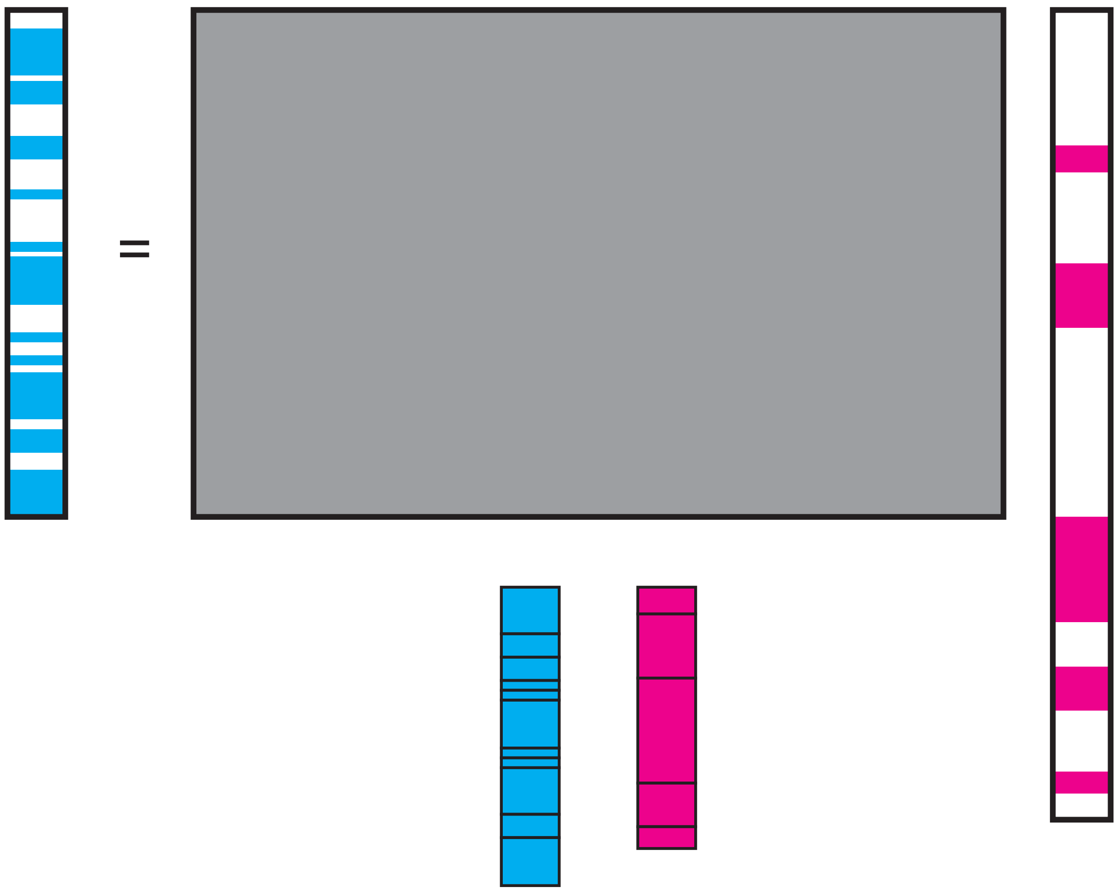


$$\|\hat{\beta}\|_1 / \|\widetilde{\beta}\|_1$$

λ chosen to minimize cv error

Compressed Sensing (Revisited)

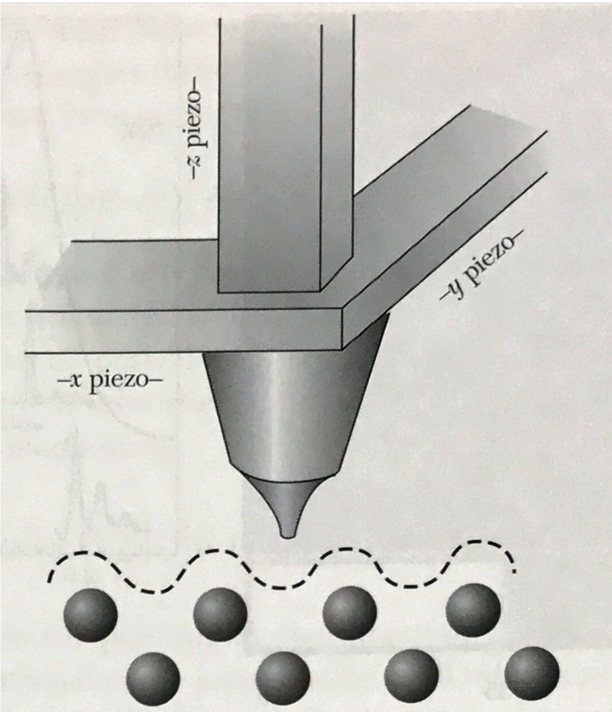
Reconstruction of the signal with undersampled measurements
→ Acceleration of measurements

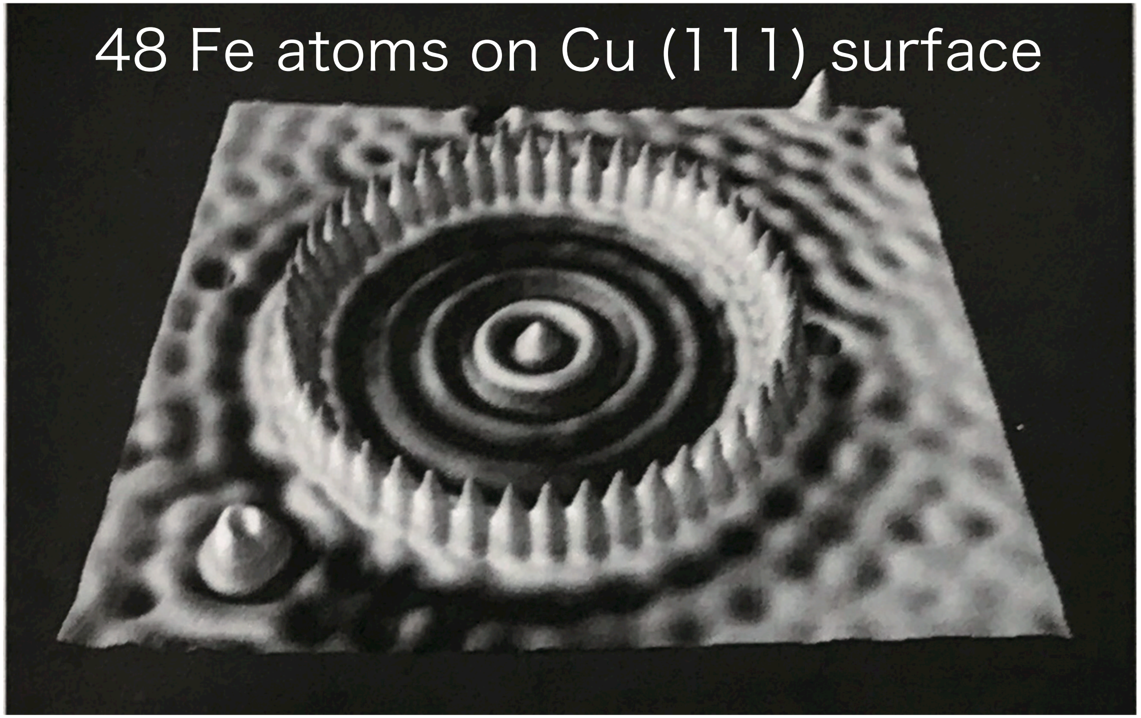


An Example of Compressed Sensing

Standing wave of electrons

Scanning tunneling microscope (STM)





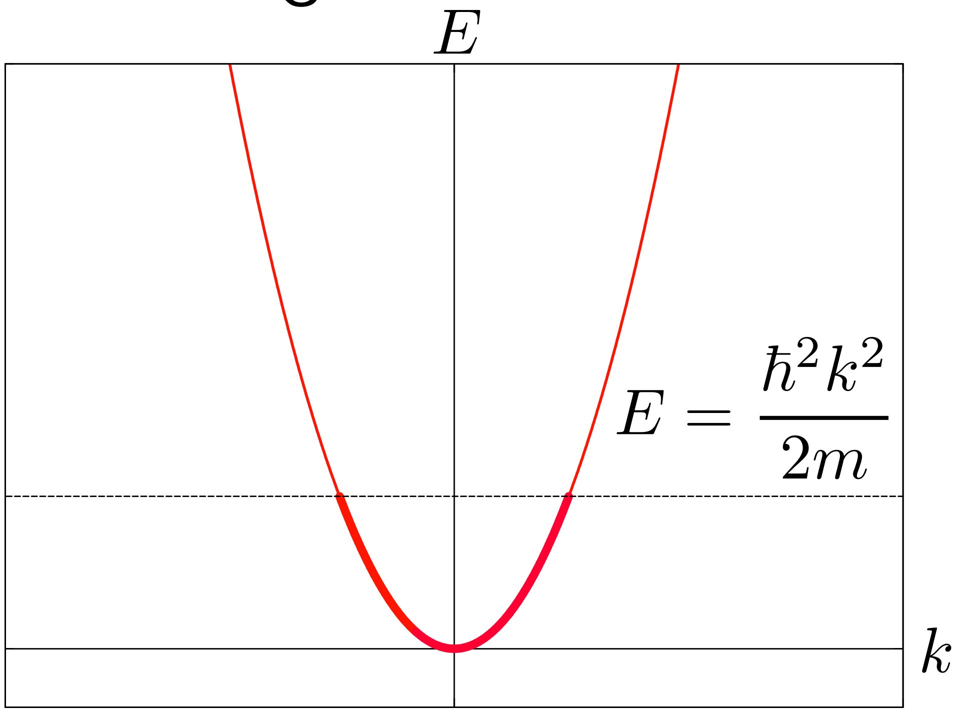
$$\frac{dI}{dV} \propto \sum_{\ell} |\psi_\ell(\vec{r})| \delta(E_{\mathrm{F}} + eV - E_\ell)$$

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

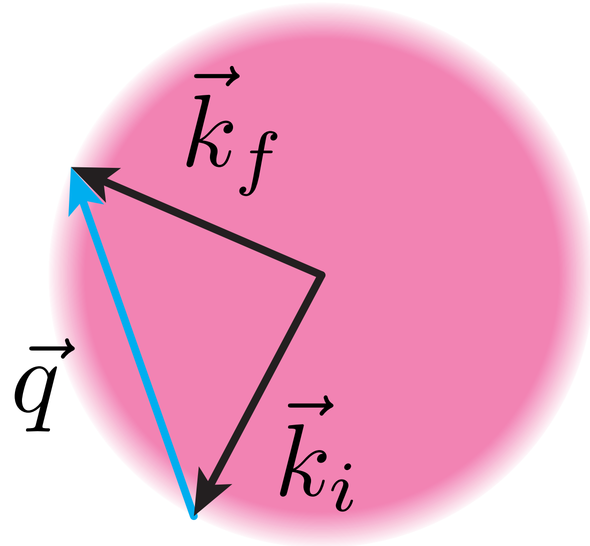
QPI by STM

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

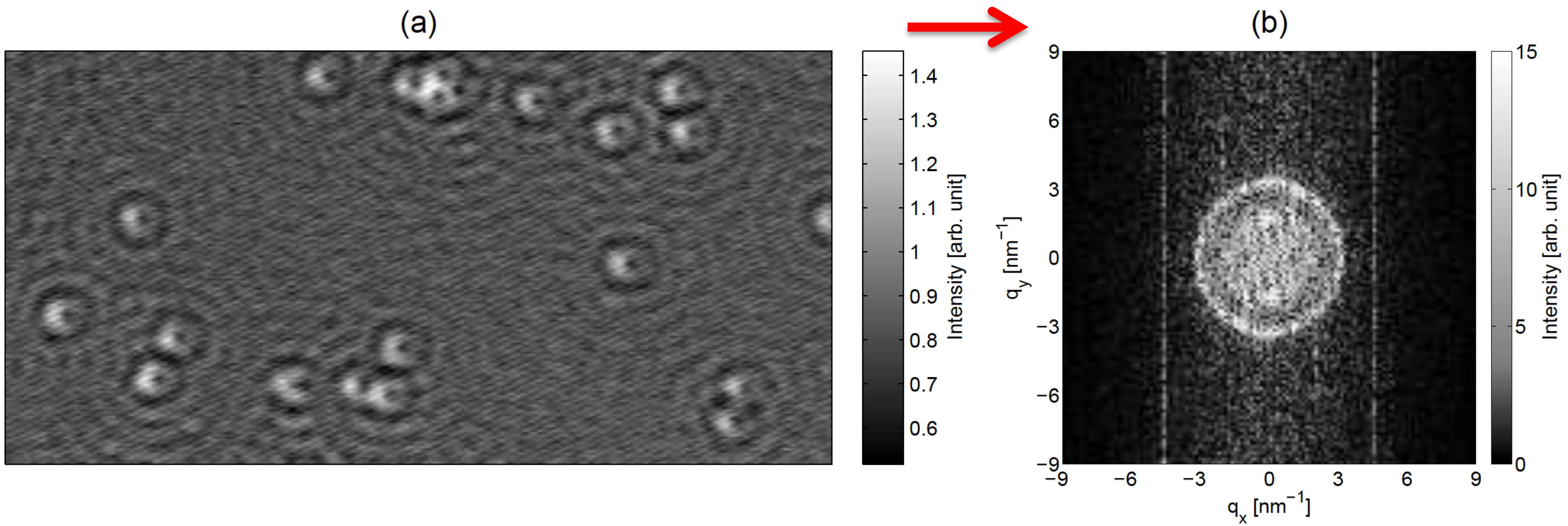


Fig. 1. (a) dI/dV map of Ag(111) surface. (b) FT of (a) obtained by conventional method.

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

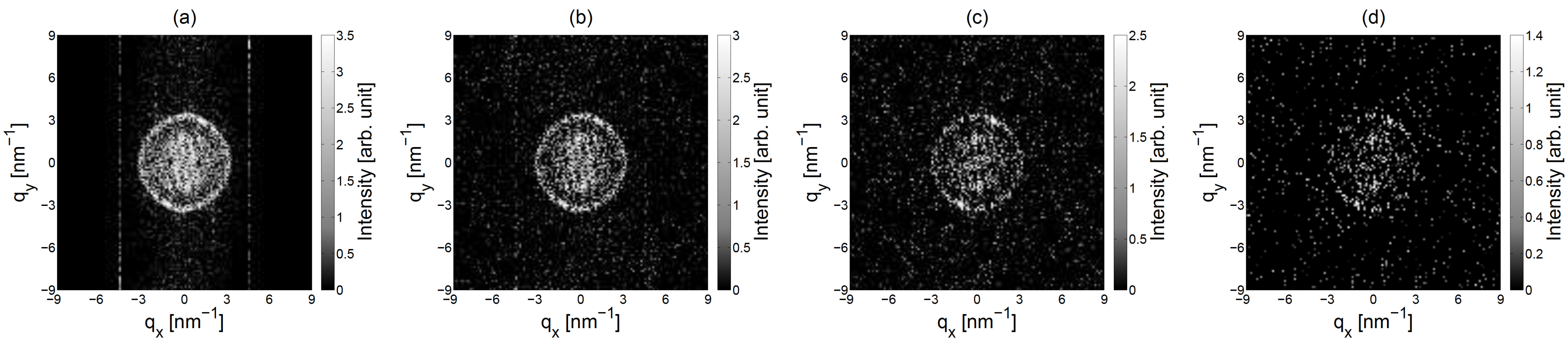
*Minimizing $\|\vec{f} - G\vec{F}\|_2 + \lambda\|\vec{F}\|_1$
by changing F with given undersampled f

*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

cf.) Random projection

Important References

-Simon Funk, *Netflix Update: Try This At Home*.
http://sifter.org/~simon/journal/20061211.html

-Chih-Chao Ma, *A Guide to Singular Value Decomposition for Collaborative Filtering*.

-Pythob library: scikit-learn
http://scikit-learn.org/stable/

-For applications of neural network, please see references cited in
Y. Nomura, A. S. Darmawan, Y. Yamaji, & M. Imada,
Phys. Rev. B 96, 205152 (2017).

-Regularization
C. M. Bishop,
Pattern Recognition and Machine Learning (Springer-Verlag, 2006)

-Lasso and sparse modeling
T. Hastie, R. Tibshirani, and M. Wainwright,
Statistical Learning with Sparsity: The Lasso and Generalization
(CRC Press, 2015)

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