圧縮センシングの理論と最近の~~進展~~応用

小渕　智之

京都大学大学院情報学研究科

2025/02/06

構成

- イントロダクション
- 定式化・アルゴリズム・緩和
- 性能評価の理論
- 質量分析への応用

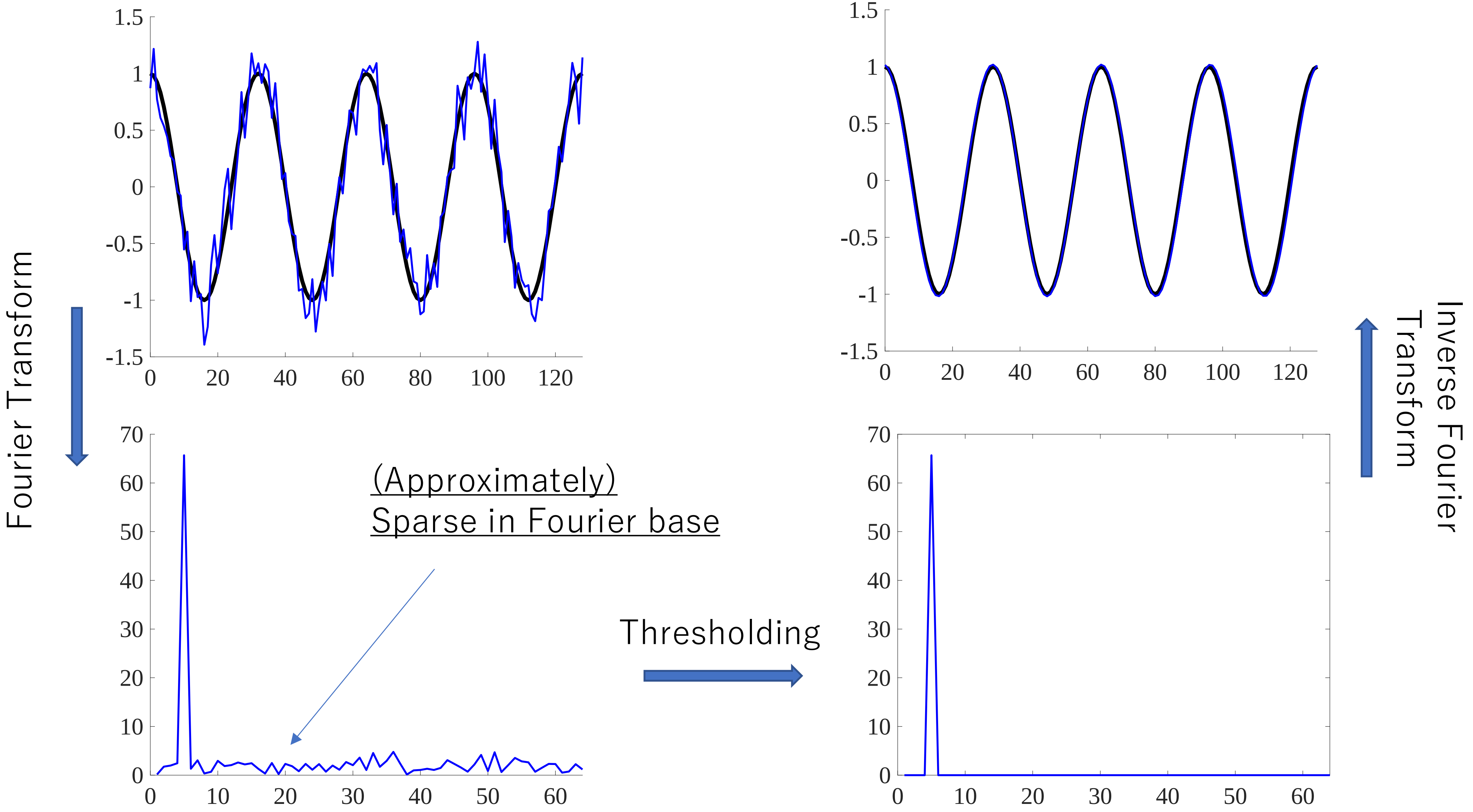
イントロダクション

スパースな信号とは何か

- 定義：ある基底で表現したときに，たくさんの要素がゼロになる信号
- この性質を「スパース（疎）性」と呼ぶ.
- スパース性を利用して情報処理に役立てることが出来る.

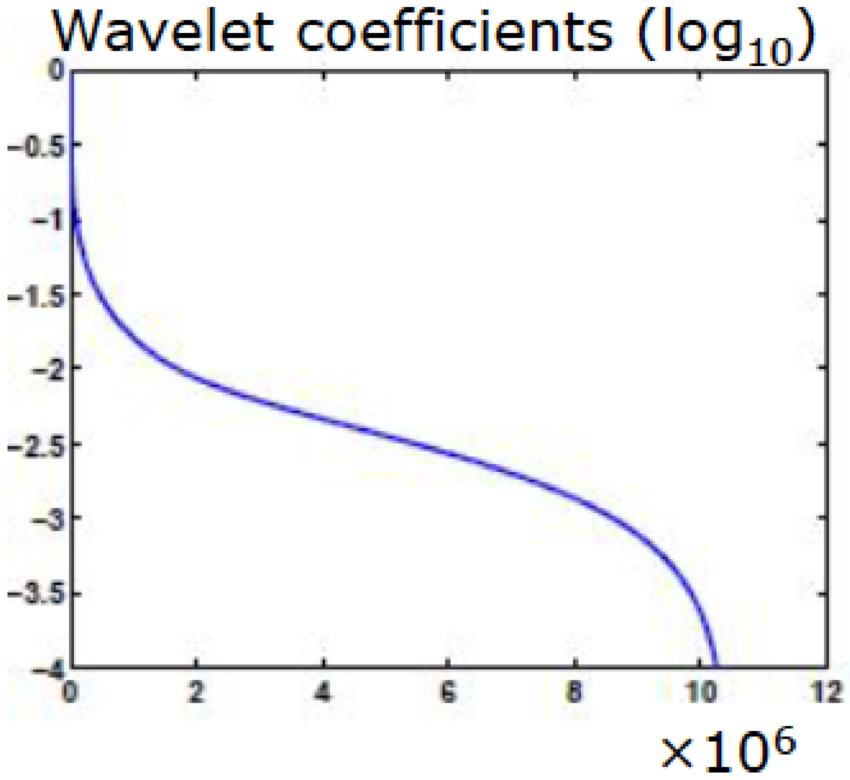
例：ノイズ除去

$$y = \cos(2\pi * 4 * x/128) + 0.2 * \xi$$



例：データ圧縮

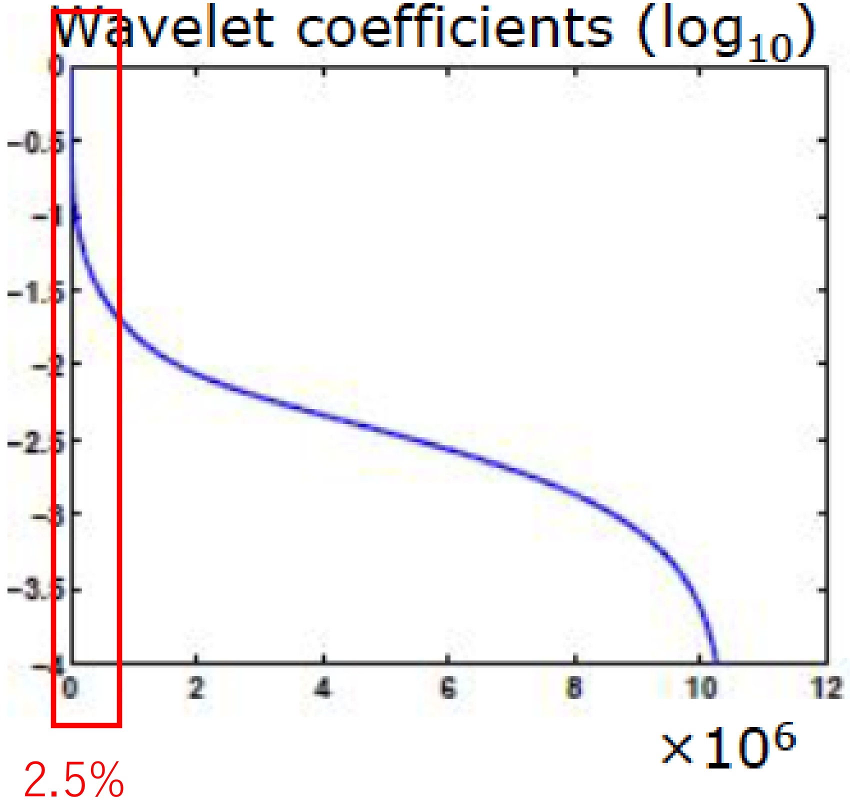
Wavelet
Transform



Romberg and Wakin, 2007

例：データ圧縮

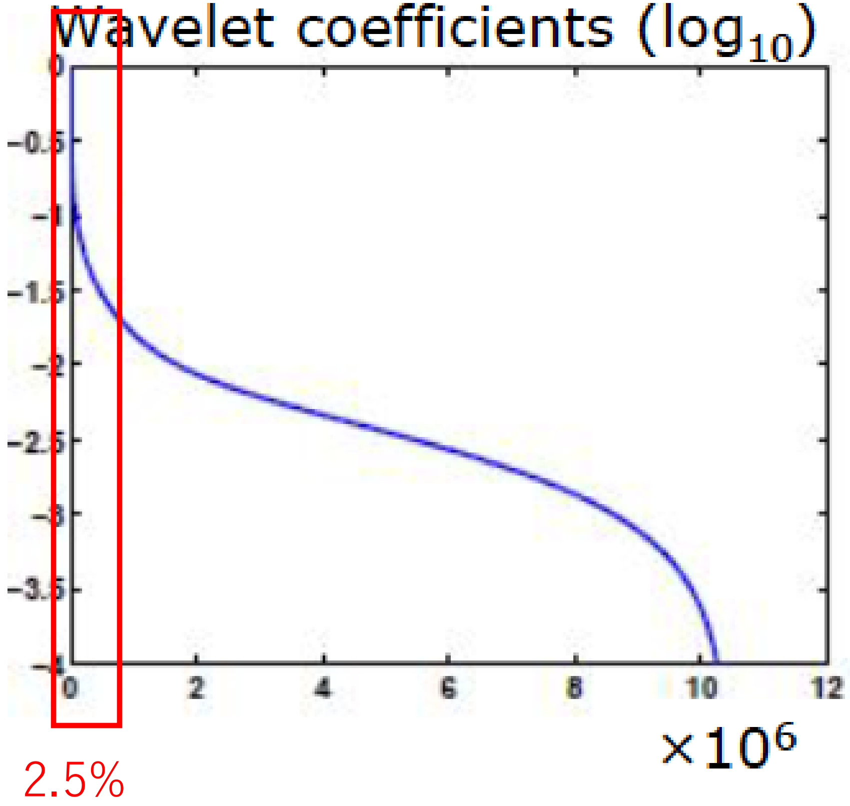
Wavelet
Transform



Romberg and Wakin, 2007

例：データ圧縮

Inverse
Wavelet
Transform



Romberg and Wakin, 2007

例：データ圧縮

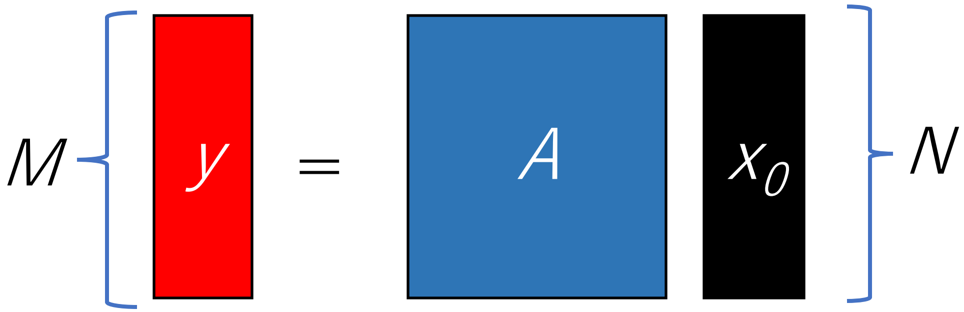
Original

Reconstructed

Romberg and Wakin, 2007

圧縮センシングとは何か

- “常識”より少ない回数の観測で，信号復元を可能にする技術
  - 常識＝ナイキスト・シャノンのサンプリング定理
    - 要は，変数の数≤観測数
- 特に線形観測の場合の応用が多い
  - 反射波地震探査
  - トモグラフィー（X線CT, MRI）
  - シングルピクセルカメラ
  - ノイズ除去（画像・音声）
  - 無線通信
  - グループテスト



線形観測：
観測結果yと既知の観測方法Aから，元の信号x_0を推定

例：トモグラフィー

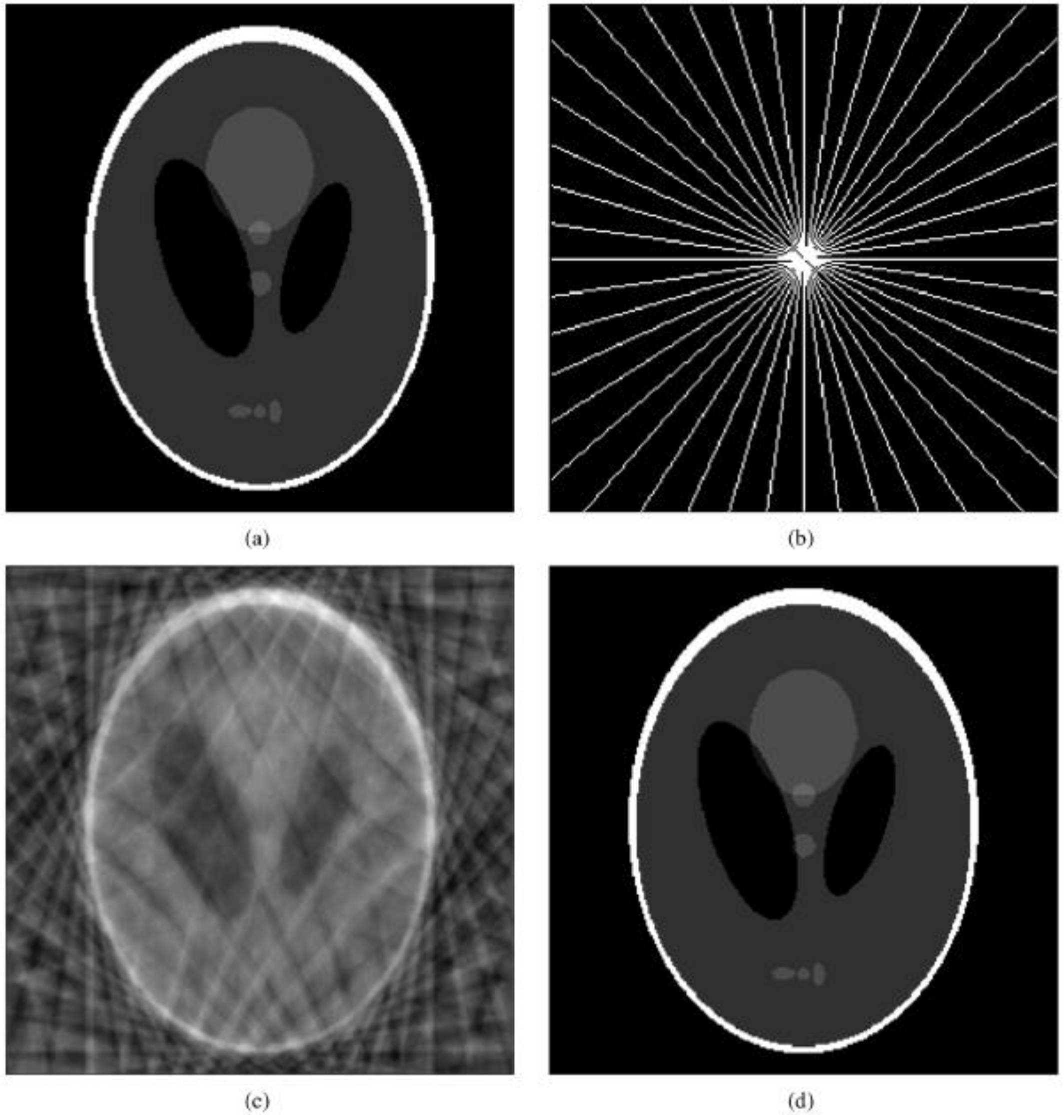
- 画像$f(x_1, x_2)$の空間フーリエ変換を測定

$$\hat{f}(k_1, k_2) = \sum_{x_1, x_2} f(x_1, x_2) e^{-2\pi i (k_1 x_1 + k_2 x_2)},$$

- 観測された$\hat{f}(k_1, k_2)$から原画像$f(x_1, x_2)$を推定.
    - 特徴的な周波数を持つ画像なら，フーリエ空間で疎.

- MRIなどで顕著な応用
    - MRI：脳・脊椎・四肢・子宮・前立腺などの撮像で威力を発揮.
        - 現在の撮像にかかる時間は30~60分←できるだけ短くしたい.

例：トモグラフィー

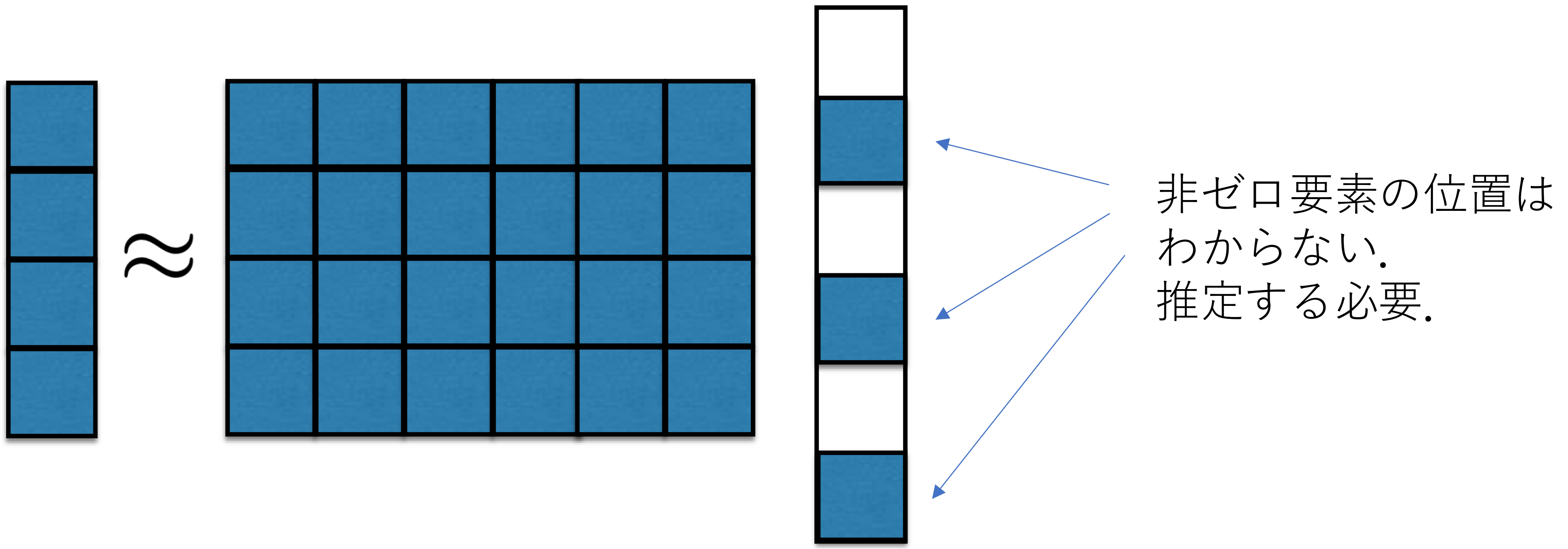


- トモグラフィーのシミュレーション
  - 左上：元画像（Shepp-Logan Phantom, 頭部を模したテスト用の標準画像）
  - 右上：22方向からの2D FTサンプリング（それぞれの方向で512点）
  - 左下：擬似逆行列による再構成(従来法)
  - 右下：スパース性を利用した再構成
- スパース性を利用した方法で完全な再構成が実現.
- 従来観測の1/50の観測量でOK!

E. Candes, J. Romberg, and T. Tao, IEEE Trans. IT, 52, 489-502 (2006)

圧縮センシング=劣決定線形方程式

- 数学的には，圧縮センシングとは劣決定線形方程式
  - 条件数より変数の数が多い.
  - そのままでは解けない（解が無数にある）ので，何らかの仮定が必要
- スパース性の仮定=いくつかの変数は零
  - 非ゼロ要素数$K \leq$観測数Mなら原理的に解ける.
  - 非ゼロ要素の位置をどう推定・計算するかが鍵



定式化・アルゴリズム・緩和

定式化

- 素直な定式化：ℓ_0 ノルム最小化

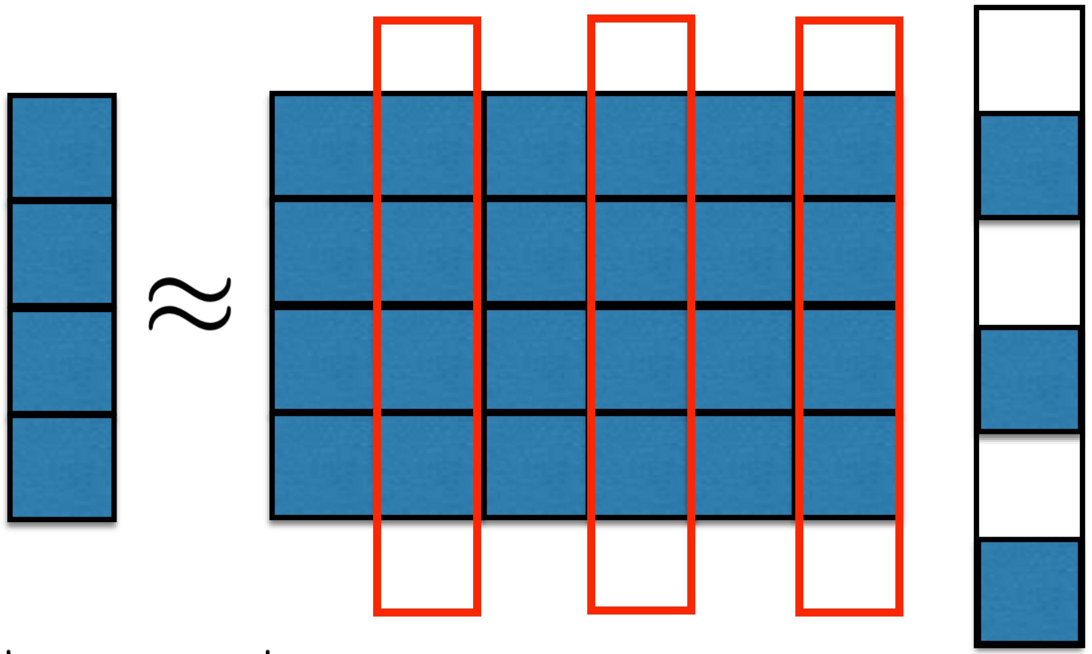
$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{x}\|_0 \ \ \mathrm{s.t.}\ \boldsymbol{y} = A\boldsymbol{x}$$

$\|\boldsymbol{x}\|_0$=# of nonzero components of $\boldsymbol{x}$

  - $\boldsymbol{y} = A\boldsymbol{x}$を満たす中で最もスパースなものを選ぶ.
  - 真の信号の非ゼロ要素数$K = \|\boldsymbol{x}_0\|_0$が$K < M$を満たすなら, それを必ず計算出来る.
  - 計算複雑度は高い.

定式化

- 素直な定式化：ℓ_0 ノルム最小化
    - $K = 1,2,\ldots$に対して　${}_N C_K$通りの全ての変数の組み合わせを試みる.
      →$K = \|\boldsymbol{x}_0\|_0$までやれば正解が得られる.
      ←指数関数的計算量(if $K = O(N)$)
    - $N = 6, M = 4, K = 3$の例



$$\boldsymbol{y} \approx x_2 \boldsymbol{a}_2 + x_4 \boldsymbol{a}_4 + x_6 \boldsymbol{a}_6$$

定式化

- 誤差許容での素直な定式化：ℓ_0 ノルム正則化つき最小二乗法

$$\hat{\boldsymbol{x}} = \arg\min \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 \ \text{ s.t. } \|\boldsymbol{x}\|_0 \leq K$$

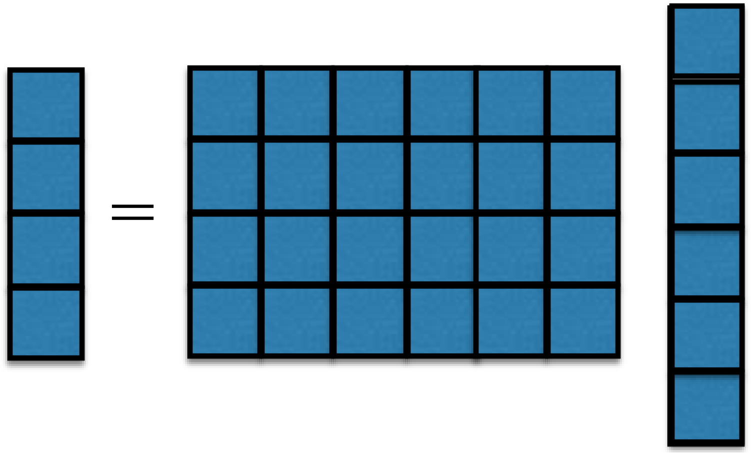
  - スパース性をコントロールしながら，誤差が一番低いものを探す.
  - ノイズ有り状況にも適用可
  - 計算量的に困難なのは同じ.

現実的な定式化・アルゴリズム

- 貪欲法（Greedy Algorithm）
  - Orthogonal Matching Pursuit (OMP)
    - Varieties: LS-OMP, MP, Weak-MP
  - Thresholding Algorithm
  - Advanced methods: CoSaMP, Iterative Hard Thresholding (IHT)
- 緩和法(Relaxation Method)
  - コスト関数を別のより取り扱いやすい関数で近似
  - Basis Pursuit (BP)(=$\boldsymbol{\ell_1}$緩和)
  - Iterated-Reweighted-Least-Squares (IRLS) (=$\boldsymbol{\ell_p}$緩和)
- 確率的推論
  - ベイズ(疎性を導く事前分布)
  - その他
    - 統計力学的方法(コスト関数に付随したボルツマン分布を利用)

貪欲法（Greedy Algorithm）

- 基本方針
  - $\boldsymbol{y}$を表現するための列$\boldsymbol{a}_i$をAから選び出してくる.

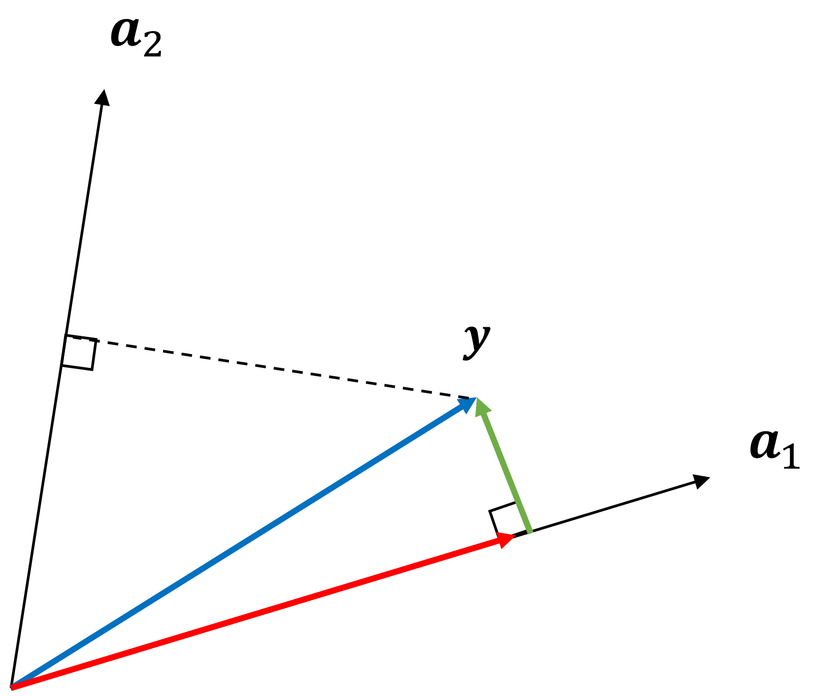


$$\boldsymbol{y} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_N\boldsymbol{a}_N$$

  - $\boldsymbol{y}$を近似にするのに良さそうな列$\boldsymbol{a}_i$を順番に(Greedy manner)選んでくる
    - 選ばれた列の集合をサポート(support)と呼ぶ.

1列による近似

- 近似＝射影：$\hat{x}_i = \arg\min(\boldsymbol{y} - x_i \boldsymbol{a}_i)^2 = \frac{\boldsymbol{a}_i \cdot \boldsymbol{y}}{|\boldsymbol{a}_i|^2}$
- 近似の良さ＝残差の小ささ： $\epsilon(i) = (\boldsymbol{y} - \hat{x}_i \boldsymbol{a}_i)^2 = \boldsymbol{y}^2 - \frac{(\boldsymbol{a}_i \cdot \boldsymbol{y})^2}{|\boldsymbol{a}_i|^2}$



Orthogonal Matching Pursuit(OMP)

- Input: $\boldsymbol{y}, A, \epsilon_0$
- Initialization:

$$k = 0, \hat{\boldsymbol{x}}^0 = \boldsymbol{0}, \boldsymbol{r}^0 = \boldsymbol{y}, S^0 = \phi$$

- Iterate the following:
  - $k = k + 1$
  - Sweep: $\forall i \notin S^{k-1}, \epsilon(i) = \min_{x_i}\left(x_i \boldsymbol{a}_i - \boldsymbol{r}^{k-1}\right)^2 = \left(\boldsymbol{r}^{k-1}\right)^2 - \left(\boldsymbol{a}_i \cdot \boldsymbol{r}^{k-1}\right)^2 / \boldsymbol{a}_i^2$
  - Update Support: $\hat{\imath} = \arg\min_{i \notin S^{k-1}} \epsilon(i), S^k = S^{k-1} \cup \{\hat{\imath}\}$
  - Update Solution and Residual

$$\hat{\boldsymbol{x}}^k = \operatorname{argmin}_{\boldsymbol{x}}\left(\boldsymbol{y} - A_{S^k}\boldsymbol{x}_{S^k}\right)^2, \boldsymbol{r}^{\mathrm{k}} = \boldsymbol{y} - A_{S^k}\hat{\boldsymbol{x}}^k$$

  - Stopping Rule: If $\left|\boldsymbol{r}^{\mathrm{k}}\right| < \epsilon_0$, then stop.

◆ Computational cost: $O(MNK_0)$, ($K_0 (= O(1))$:final support size)

◆ c.f.) Exact enumeration: $O(MN^{K_0}K_0^2)$

緩和法(Relaxation Method)

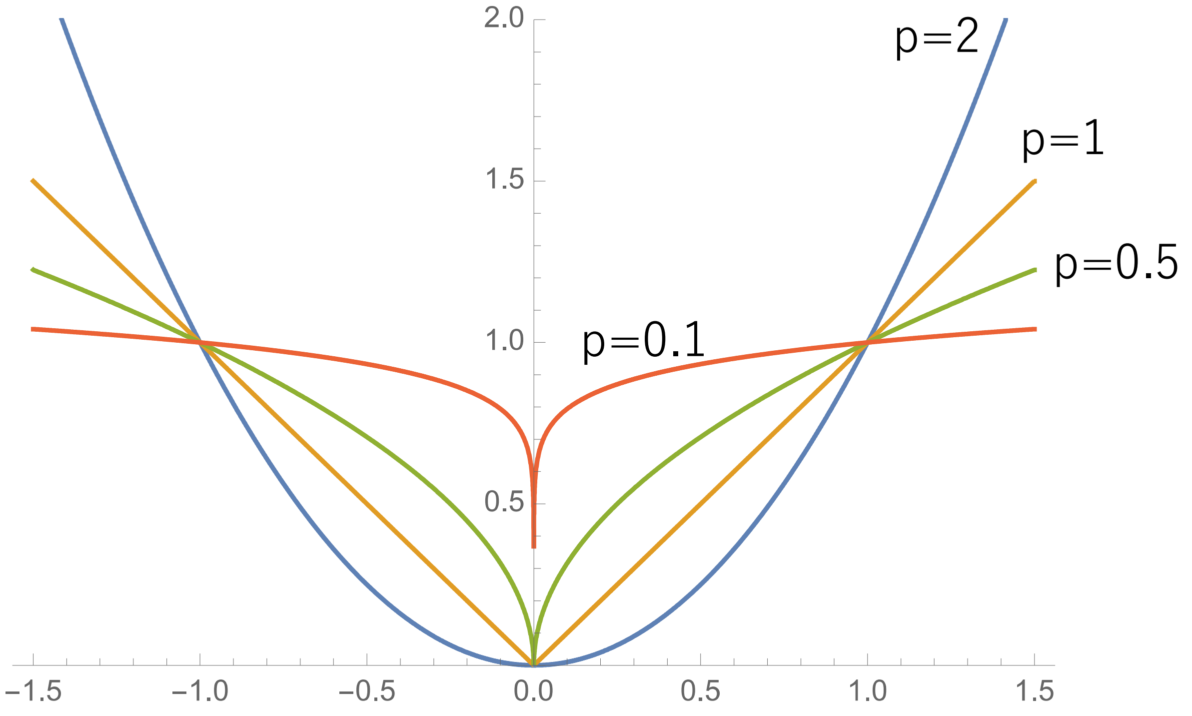
- ℓ_0コストが使いづらいのは不連続だから.
  ←連続関数で近似すればいいのでは？
  →緩和（Relaxation）

- $\ell_0$norm to $\ell_p$norm

$$\|\boldsymbol{x}\|_0 = \lim_{p\to+0}\sum_i^N |x_i|^p$$

$$\Rightarrow \|\boldsymbol{x}\|_p = \left(\sum_i^N |x_i|^p\right)^{1/p}$$

$|x|^p$ profile



凸緩和(Convex Relaxation)

- $p \leq 1$では解がスパースになる.
  - 原点における特異性が鍵
- $p \geq 1$では凸性がある.
  - 最適化が簡単
  - 解が一意

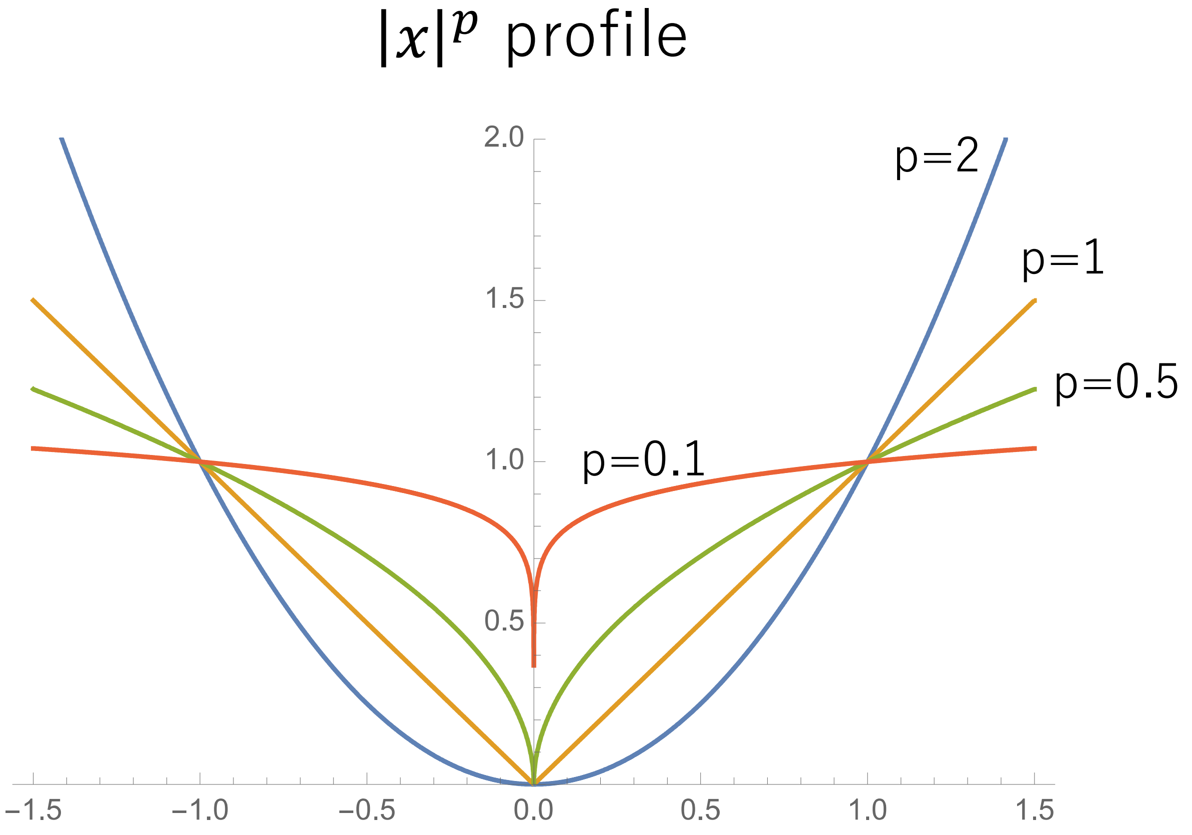
- $p = 1$がいいとこ取り.

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

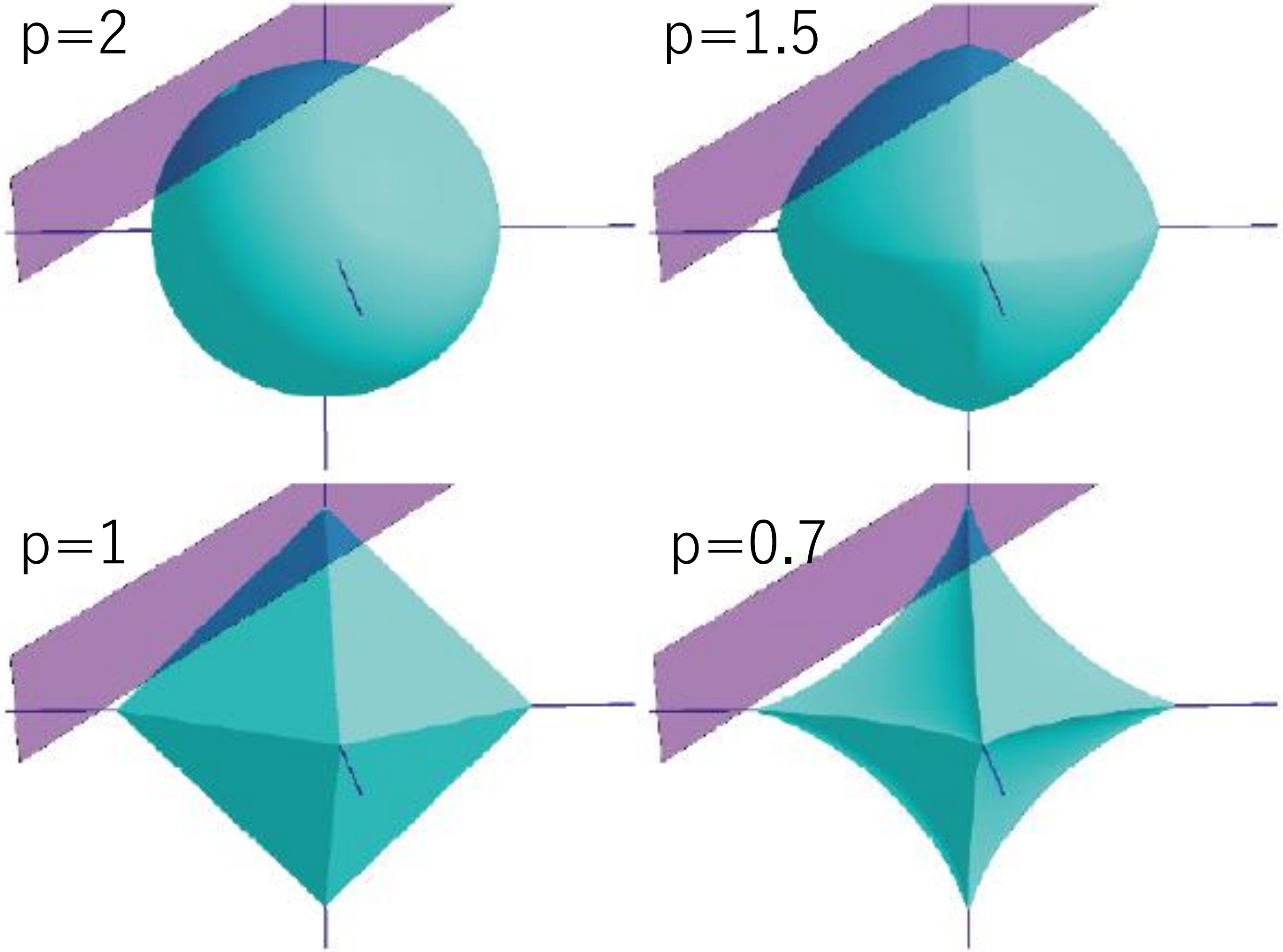
- 解くべき緩和問題

$$\hat{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \mathrm{s.t.}\ \boldsymbol{y} = A\boldsymbol{x}$$

←LassoとかBasis pursuitと呼ばれる.



$p \leq 1$ では解がスパース

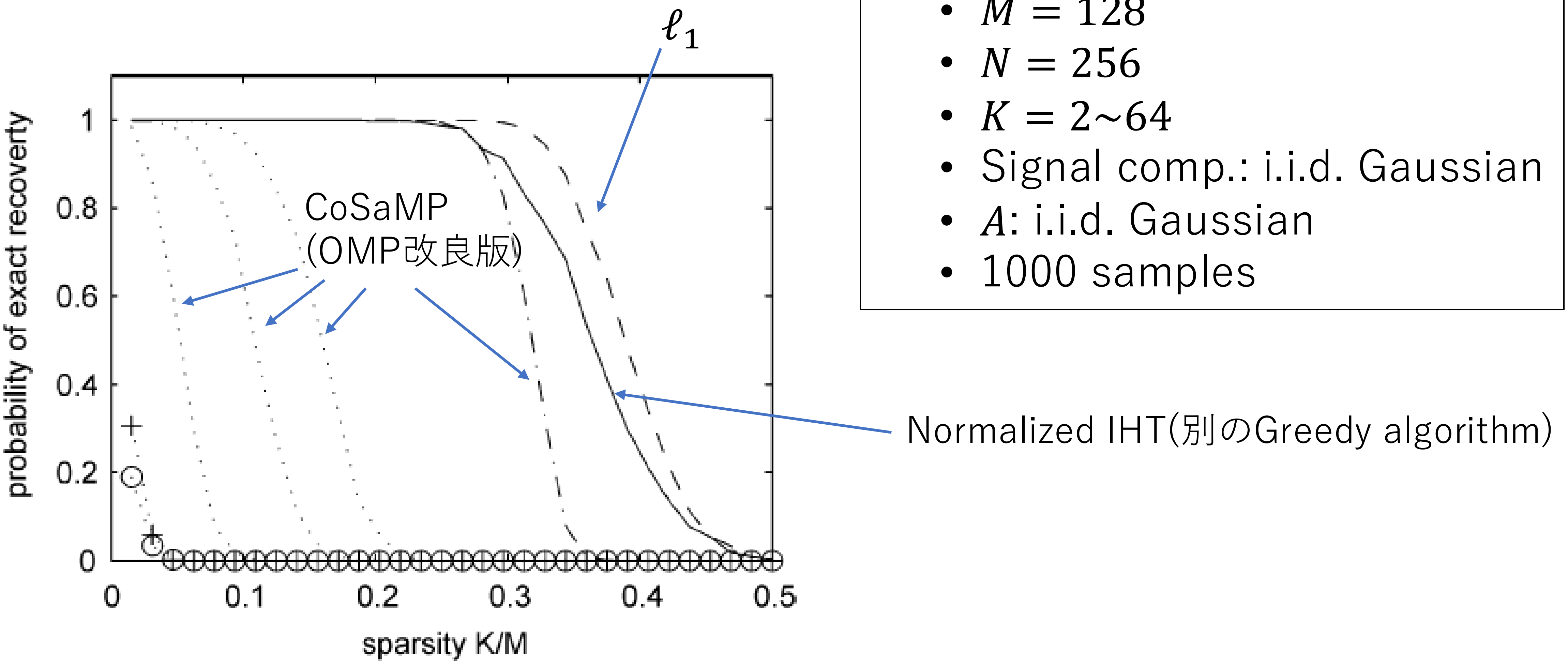


Purple plane:
$\boldsymbol{y} = A\boldsymbol{x}$

Green figures:
ℓ_p ball

M. Elad: Sparse and Redundant Representations (2010)

Performance Comparison



Blumensath-Davis, 2010

Experimental condition

- $M = 128$
- $N = 256$
- $K = 2 \sim 64$
- Signal comp.: i.i.d. Gaussian
- A: i.i.d. Gaussian
- 1000 samples

性能評価の理論

性能評価の理論について

- 論点：推定解と真の解の一致性
  - アルゴリズムの収束性については今回は言及しない.

- 基礎的方法：Mutual Coherenceを用いる方法
  - OMP, ℓ_1, Thresholding methodの完全復号の十分条件
- 発展的方法：
  - 制限等長性（Restricted isometry property, RIP）による方法 (Candès-Tao, 2005)
  - 積分幾何による方法（Donoho-Tanner, 2006）
  - 情報統計力学による方法
    - レプリカ法（Kabashima-Wadayama-Tanaka, 2009）
    - 状態発展法（Donoho-Maleki-Montanari, 2009）

情報統計力学的定式化

- 基本戦略：ベイズ推定の枠組みを利用
  - 事前分布($\boldsymbol{\ell_p}$の定式化)

$$P_\beta(\boldsymbol{x}) \propto e^{-\beta\|\boldsymbol{x}\|_p^p}$$

  - データの生成仮定(ノイズ無しを仮定)

$$P(\boldsymbol{y}|\boldsymbol{x_0}, A) = \delta(\boldsymbol{y} - A\boldsymbol{x}_0)$$

  - 事後分布

$$P(\boldsymbol{x}|\boldsymbol{y}, A) = P(\boldsymbol{x}|\boldsymbol{x_0}, A) = \frac{e^{-\beta\|\boldsymbol{x}\|_p^p}\,\delta(A\boldsymbol{x} - A\boldsymbol{x}_0)}{Z_\beta(\boldsymbol{x_0}, A)}$$

$\beta \to \infty$で$\boldsymbol{\ell_p}$ノルム最小化解におけるデルタ関数的な分布になる.
Zは規格化定数（分配関数）で

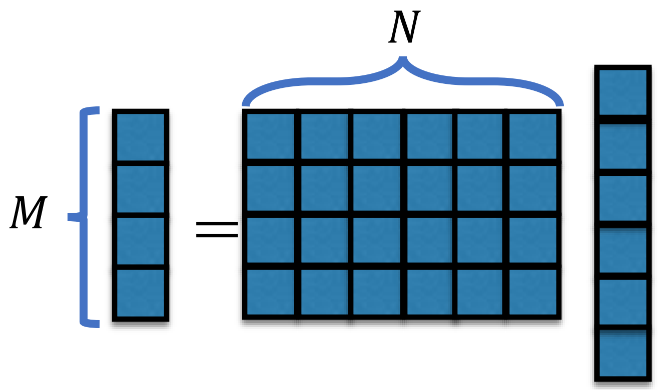
$$Z_\beta(\boldsymbol{x}_0, A) = \int d\boldsymbol{x}\, e^{-\beta\|\boldsymbol{x}\|_p^p}\,\delta(A\boldsymbol{x} - A\boldsymbol{x}_0)$$

レプリカ法による方法

- 典型的性質は平均自由エネルギー(≒積率母関数)から計算できる：
  $f = -\left(\frac{1}{N\beta}\right)\left[\log Z_\beta(\boldsymbol{x}_0, A)\right]_{A,\boldsymbol{x}_0}$, $[\cdots]_{A,\boldsymbol{x}_0}$: $A, \boldsymbol{x}_0$に関する平均
  - $\log Z$ の平均$[\cdots]_{A,\boldsymbol{x}_0}$を計算するのは難しい．←レプリカ法
- レプリカ法：
  - レプリカ恒等式
    $[\log Z(\boldsymbol{J})] = \lim_{n\to 0}\frac{1}{n}\log[Z^n(\boldsymbol{J})]$，$[\cdots] = \int d\boldsymbol{J}P(\boldsymbol{J})(\cdots)$
  - 手順
    - nが自然数だと仮定して$[Z^n(\boldsymbol{J})]$を評価
    - 得られた表現の解析接続を用いて自然数から実数に表現を拡張
    - $\lim_{n\to 0}$を取る．

レプリカ法による方法

- 解析のための単純化仮定
  - 高次元極限 $N, M \to \infty, \alpha = \frac{M}{N} = O(1)$
    - 最近はこれをproportional limitと呼ぶことが多い．
  - 非ゼロ要素数もNに比例：$\rho = K/N$
  - A: ランダム行列 ← 最重要
    - 変数間の相関が極限で切れ，$N \to \infty$で有効的に一体問題になる．
    - 現実的ではない．
      - よくある観測行列は，フーリエ，ウェーブレット，アダマールなど．
      - 後半の質量分析への応用では，帯行列．
    - が，ある種の universality(Donoho-Tanner universality)が成り立ち，右回転不変アンサンブルの行列なら，同じ再構成限界を与える．
      - フーリエやウェーブレット行列からランダムにサンプルしてつくった観測行列なら同じ結果になる．



レプリカ法による方法

- 結果(at $\frac{|\boldsymbol{x}_0|^2}{N} \to \rho$)

$$\lim_{\beta\to\infty} f = \mathrm{Extr}_{\Theta}\left\{ \begin{array}{c} \dfrac{\alpha(Q-2m+\rho)}{\chi} + \hat{m}m - \dfrac{1}{2}\hat{Q}Q + \dfrac{1}{2}\hat{\chi}\chi \\ +(1-\rho)\int Dz\, F_p\left(\sqrt{\hat{\chi}}z; \hat{Q}\right) + (\rho)\int Dz\, F_p\left(\sqrt{\hat{m}^2+\hat{\chi}}z; \hat{Q}\right) \end{array} \right\},$$

Extr_x: extermization w.r.t. x.

$\Theta = \{\chi, Q, m, \hat{\chi}, \hat{Q}, \hat{m}\}$,

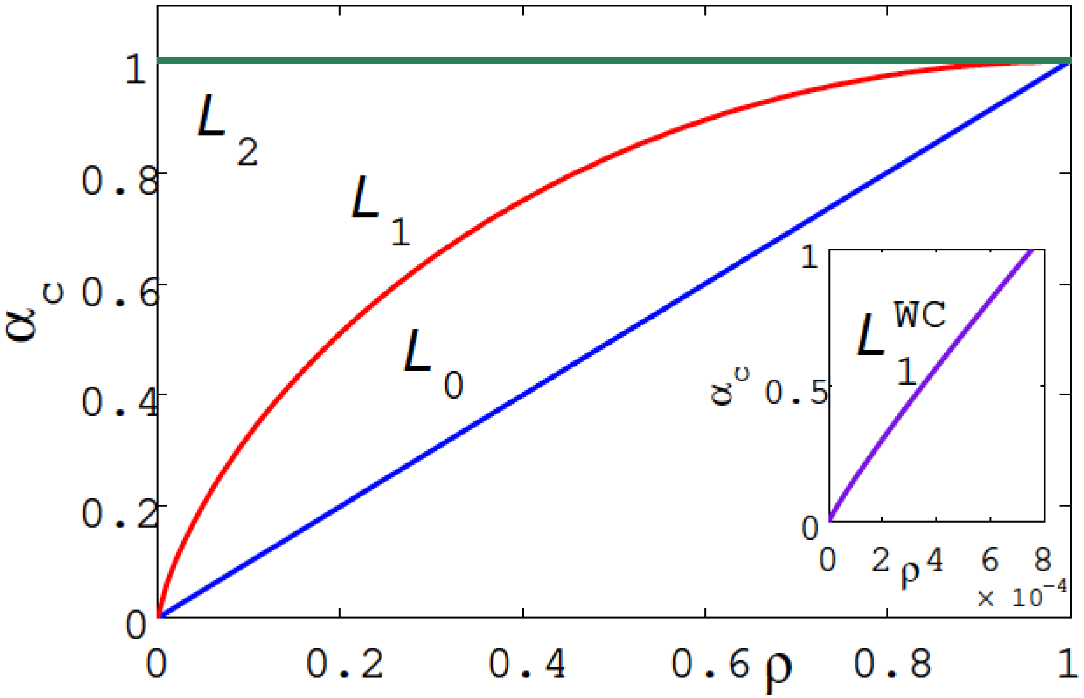
$$F_p\left(h; \hat{Q}\right) = \lim_{\epsilon\to+0}\left\{\min_x\left(\frac{\hat{Q}}{2}x^2 - hx + |x|^{p+\epsilon}\right)\right\}.$$

- 平均二乗誤差ϵ(Mean Square Error, MSE)などがここから計算できる.

$$\epsilon = Q - 2m + \rho$$

レプリカ法による方法

- 結果(at $\frac{|\boldsymbol{x}_0|^2}{N} \to \rho$)



- 完全再構成限界＝相境界
  - 境界上（下）でMSE=0(>0)
  - ℓ_1の場合は，Donoho-Tanner相転移と呼ばれる（Donoho-Tanner, 2006）.
- ℓ_pで$p < 1$の場合の転移点もℓ_0の場合に等しいことが解析から示唆.

Kabashima-Wadayama-Tanaka, 2009

質量分析への応用

Contents

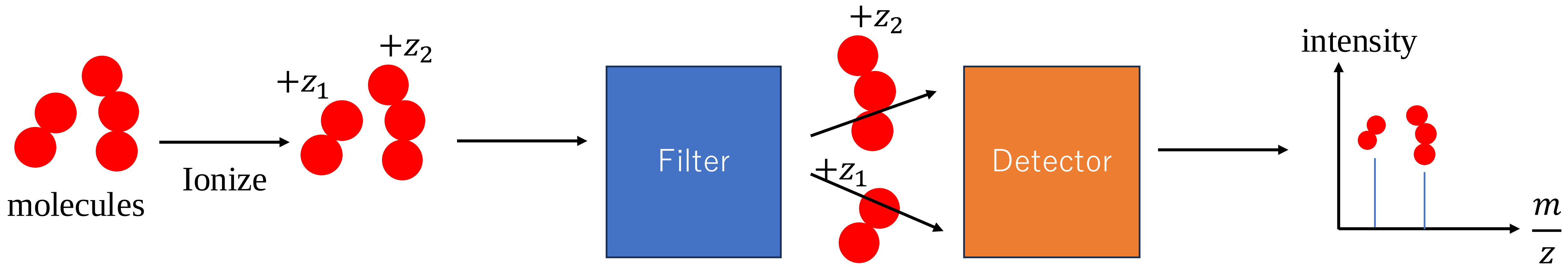
- Research Background
  - Mass Spectrometer (MS)
- Research Purpose and Proposed Method
  - Purpose: Better Resolution of MS w/o Additional Cost/Time/Instrument
  - Method:
    Low Resolution Measurement (Band Measurement) + Sparsity-Based Inference (lasso)
- Validation of Proposed Method
  - Model Selection
  - Theoretical Analysis
  - Simulation Study
  - Real Data Analysis

Mass Spectrometry

- Mass Spectrometry (MS):
  A method to analyze chemical structure/composition of sample molecule(s) from spectrum of mass-to-charge ratio (m/z) ← Mass Spectrum.
  - Utilized in various disciplines
    - Medical Sciences
    - Earth Sciences
    - Archaeology
    - Organic Chemistry
    - Bio Sciences
    - …

Mass Spectrometry

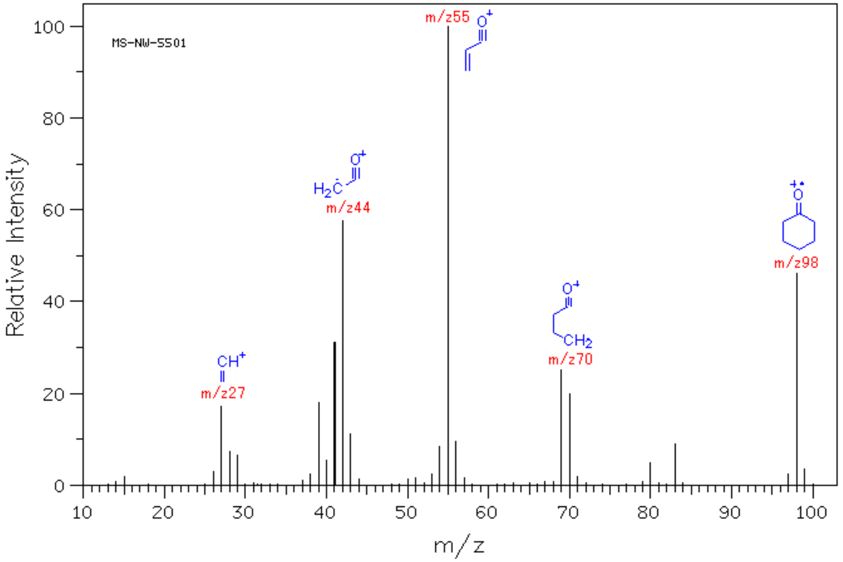
- Standard operations in MS
  - Prepare a vacuum under high voltage in device
  - Ionize the sample
    - Ions fly due to the electrostatic force in the device
  - Flying ions are separated by some actions according to their m/z.
  - Each separated ion is detected → A mass spectrum (m/z vs intensity) is obtained.



Mass Spectrometry

- Standard operations in MS
  - Prepare a vacuum under high voltage in device
  - Ionize the sample
    - Ions fly due to the electrostatic force in the device
  - Flying ions are separated by some actions according to their m/z.
  - Each separated ion is detected → A mass spectrum (m/z vs intensity) is obtained.

A Time-of-Flight (ToF) type spectrometer (From Wikipedia)
https://commons.wikimedia.org/w/index.php?curid=433732

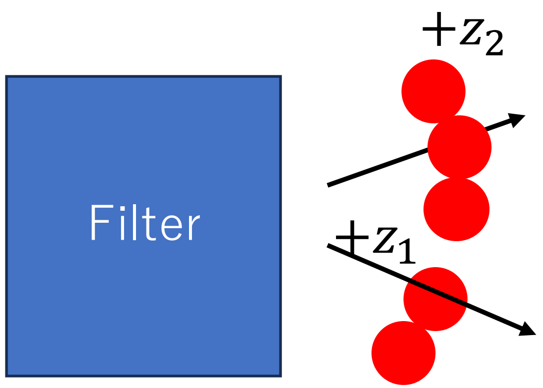


A spectrum example (From Chem-Station)
https://www.chem-station.com/yukitopics/ms.htm

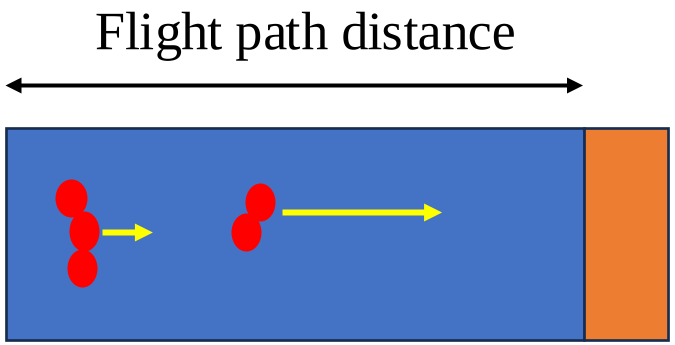
Physical Mechanism of Filter

(inverse of) mass-to-charge ratio

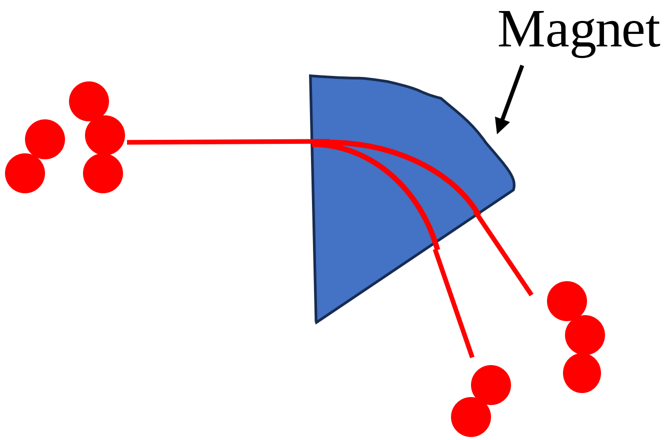
- Newtonian dynamics: $\frac{1}{2}mv^2 = zeV \Rightarrow v = \sqrt{\frac{z}{m}2eV}$
  - m: Mass
  - V: Voltage
  - e: Elementary charge
  - z: Charge number



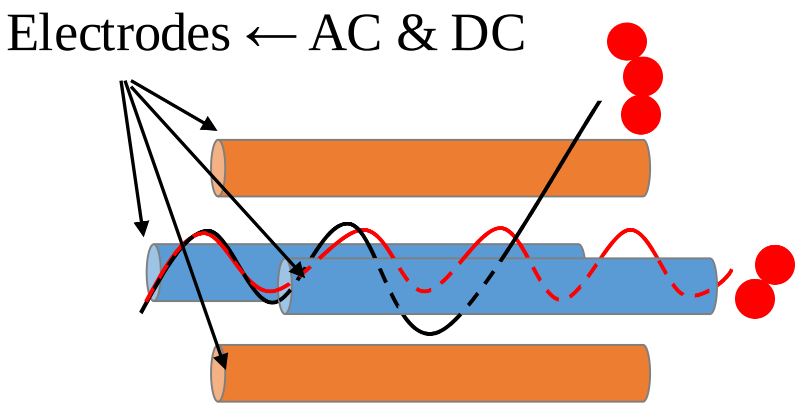
- Different types of filters



- Time of flight (ToF)



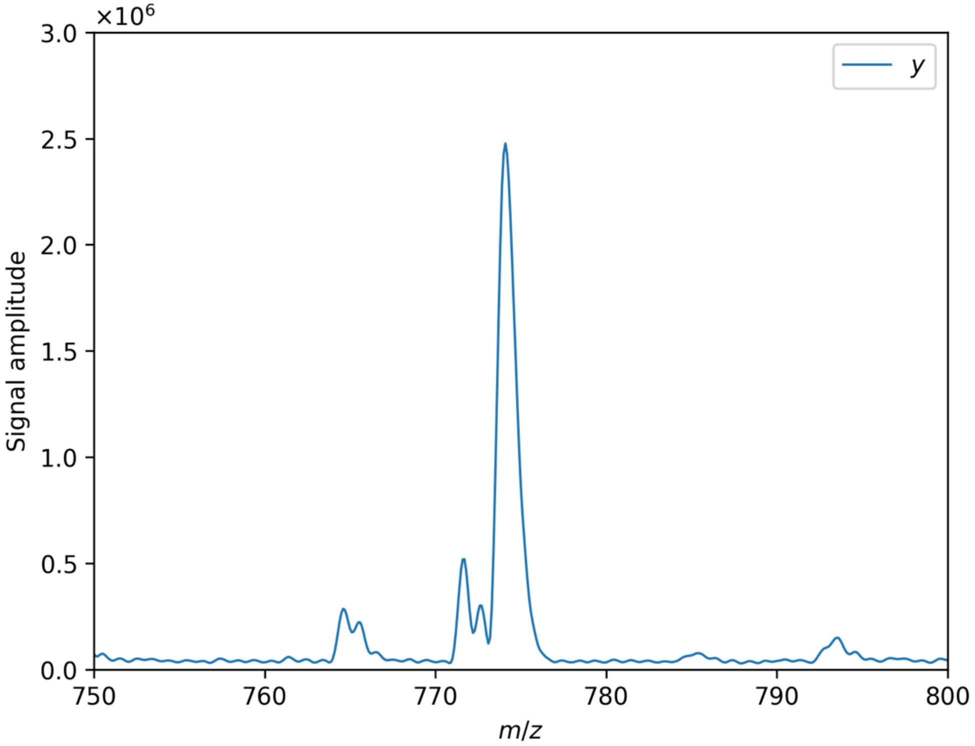
- Magnetic Sector

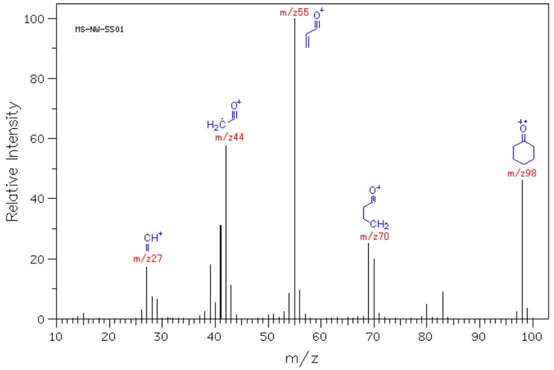


- Quadrupole mass filter (QMF)

Example Data

- A raw data from QMF





- Experimental settings:
  - Filter: QMF (LCMS-8060)
  - Step size: 0.1 m/z
  - Measurement mode: Unit
    - Related to filter & detector resolution
  - Number of scans: 300
    - Scan time: 1.00 sec/scan
  - Measurement target: Synthetic peptide (EELNAISGPNEFAR, monoisotopic mass: 1545.742)
  - Solvent: H_2O (49.5%), CH_3CN(50%), CH_3COOH(0.5%)
  - Sample concentration: 200 fmol/μL

- How to obtain the spectrum from the output? ← some processing (e.g. thresholding)
  - This tends to limit the resolution
    → empirical resolution ≈ 0.5 m/z ← lower than the device resolution limit

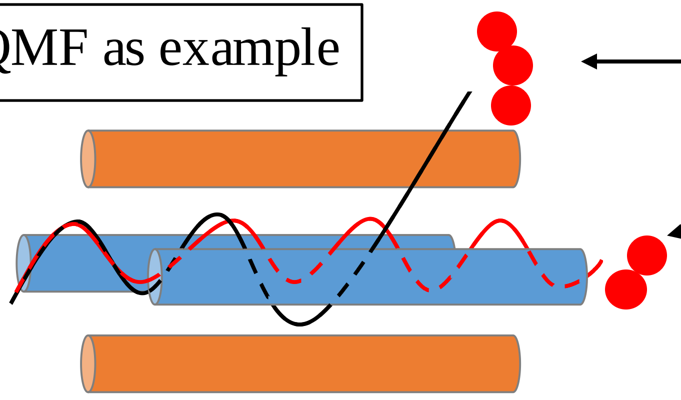
<u>The QMF (or general MS)'s ability tends to be overlooked.</u>

Research Purpose and Ideas

- Purpose: Enhancing the sensitivity and accuracy of spectrometers
    - w/o increasing the observation cost/time
    - w/o device alteration
- Ideas
    - Employ low-resolution observations
        - Equipped by default in most of standard mass spectrometers like QMF
    - Reconstruct high-resolution spectrum
      from the low-resolution output by statistical methods

Mathematical Model of Measurement

QMF as example



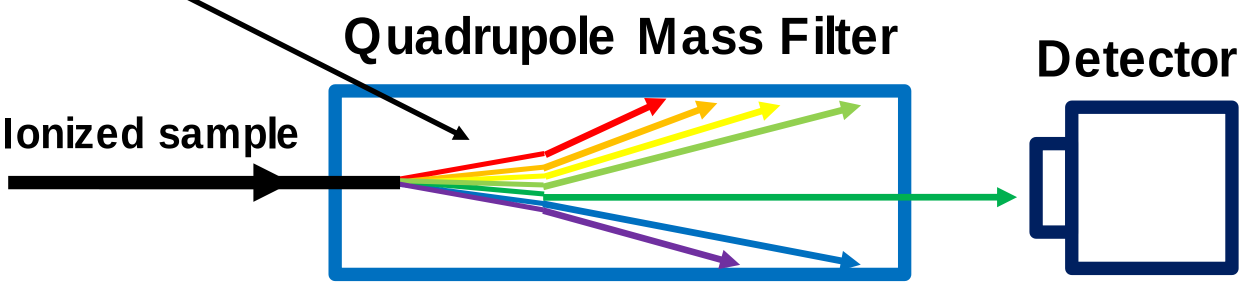
Other non-focused m/z values are discarded

Only single m/z value is selected
→high resolution

Achieved by tuning current intensity

A schematic

Many ions are discarded without being observed



Mathematical Model of Measurement

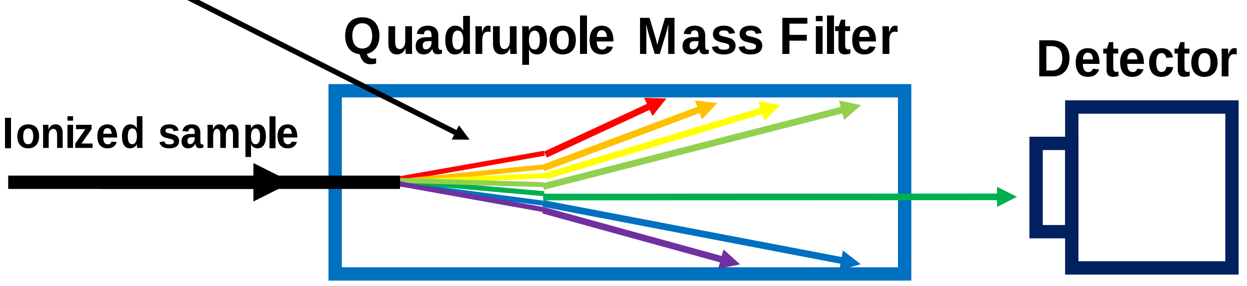
A linear observation model

$$\boldsymbol{y} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{pmatrix} + \text{noise} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{pmatrix} + \text{noise}$$

A schematic

Unit matrix represents the high-resolution observation

Many ions are discarded without being observed



Mathematical Model of Measurement in Low Resolution

A linear observation model

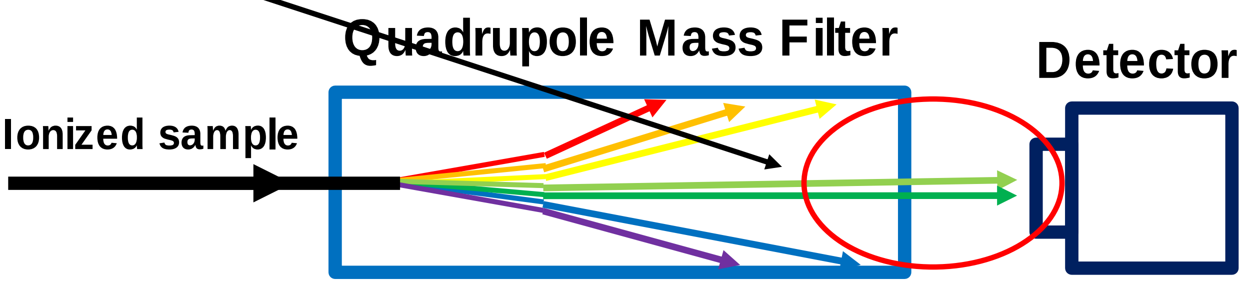
Output becomes a mixture of signals

$$\boldsymbol{y} = \overbrace{\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}^{w} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{pmatrix} + \text{noise} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_2 + \beta_3 \\ \beta_3 + \beta_4 \\ \beta_4 + \beta_5 \\ \beta_5 + \beta_6 \\ \beta_6 + \beta_7 \\ \beta_7 \end{pmatrix} + \text{noise}$$

A schematic

Observation matrix w/ band width w ($w = 2$ above)

Lower resolution measurement ← termed Band Measurement (BM)



Lasso for Demixing Signals

$$\|\boldsymbol{\beta}\|_2^2 = \sum_i^N |\beta_i|^2, \|\boldsymbol{\beta}\|_1 = \sum_i^N |\beta_i|$$

- Mass spectrum is essentially sparse ← Sparsity assumption can be utilized!

- Two (equivalent) formulations of Lasso for noisy case
  - Normal form

$$\widehat{\boldsymbol{\beta}}_r = \text{argmin}_{\boldsymbol{\beta}}\{\|\boldsymbol{y} - \boldsymbol{X\beta}\|_2^2\} \text{ s.t. } \|\boldsymbol{\beta}\|_1 \leq r$$

  - Lagrange form

$$\widehat{\boldsymbol{\beta}}_\lambda = \text{argmin}_{\boldsymbol{\beta}}\left\{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\}$$

    - This case is focused hereafter.

The combination use of BM and lasso = BM-lasso ← out proposed method

Issues in BM-lasso

$$\widehat{\boldsymbol{\beta}}_\lambda = \mathrm{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}$$

- Model Selection: How to choose λ?
- Validation of the proposed method
  - Theoretical Analysis
    - Does BM-lasso really outperform the conventional method?
  - Simulation Study
    - Detailed check in a wider parameter region
  - Real Data Analysis
- ~~Algorithm~~ (not focused today)

Model Selection

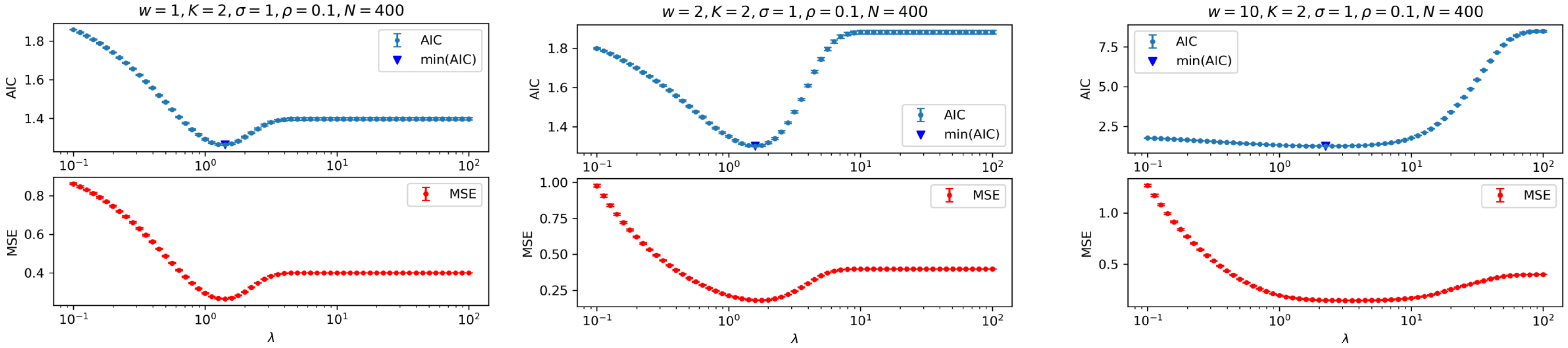
- A common criterion is to minimize the prediction error (PE) $\mathbb{E}\frac{1}{N}\|\boldsymbol{y}_{\mathrm{new}}-\hat{\boldsymbol{y}}\|_2^2$ $(\hat{\boldsymbol{y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}})$
- How to compute PE?
  - Cross validation (CV) cannot be applied in the current case
    - Observation matrix is clearly not from i.i.d.
  - Akaike's Information Criterion (AIC)
    - An estimator of PE
    - For lasso:

$$\mathrm{AIC}=\frac{1}{N}\left(\|\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\|_2^2+2\sigma^2\|\hat{\boldsymbol{\beta}}\|_0\right)\ (\approx \mathrm{PE})$$

$$\|\boldsymbol{x}\|_0 = \#\text{ of nonzero components of }\boldsymbol{x}$$

Model Selection: Simulation Result

- Simulation on fully artificial data



- True Signal: $\boldsymbol{\beta}^* \in \mathbb{R}^N, P_{\beta^*}(\beta_i^*) = (1-\rho)\delta(\beta_i^*) + \rho\delta(\beta_i^* - K)$
- Noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$
- $\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|_2^2$: MSE of estimation

- AIC-minimum model ≈ MSE-minimum model → AIC is reasonable for model selection

Validation: Theoretical Analysis

- To show the superiority of BM-lasso, we perform a statistical mechanical analysis
- Define Hamiltonian and Boltzmann distribution (w/ inverse temperature γ):

$$\mathcal{H}(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

$$P_\gamma(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y}) = \frac{1}{Z} e^{-\gamma\mathcal{H}(\boldsymbol{\beta}|\boldsymbol{X}_w,\boldsymbol{y})} = \frac{1}{Z}\prod_{\mu=1}^{N} \phi_\mu\left(y_\mu \mid \boldsymbol{\beta}, \boldsymbol{x}_w^{(\mu)}\right) \prod_{i=1}^{N} \psi_i(\beta_i)$$

  - Factorized form and potential functions
    - $\phi_\mu\left(y_\mu \mid \boldsymbol{\beta}, \boldsymbol{x}_w^{(\mu)}\right) = e^{-\frac{\gamma}{2}\left(y_\mu - \left(\boldsymbol{x}_w^{(\mu)}\right)^\top \boldsymbol{\beta}\right)^2}$, $\boldsymbol{x}_w^{(\mu)}$: μth row vector of $\boldsymbol{X}_w$
    - $\psi_i(\beta_i) = e^{-\gamma\lambda|\beta_i|}$

Validation: Theoretical Analysis

- To show the superiority of BM-lasso, we perform a statistical mechanical analysis
- Define Hamiltonian and Boltzmann distribution (w/ inverse temperature γ):

$$\mathcal{H}(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

$$P_\gamma(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y}) = \frac{1}{Z} e^{-\gamma\mathcal{H}(\boldsymbol{\beta}|\boldsymbol{X}_w,\boldsymbol{y})} = \frac{1}{Z}\prod_{\mu=1}^{N} \phi_\mu\left(y_\mu \mid \boldsymbol{\beta}, \boldsymbol{x}_w^{(\mu)}\right)\prod_{i=1}^{N}\psi_i(\beta_i)$$

  - Factorized form and potential functions
    - $\phi_\mu\left(y_\mu \mid \boldsymbol{\beta}, \boldsymbol{x}_w^{(\mu)}\right) = e^{-\frac{\gamma}{2}\left(y_\mu - \left(\boldsymbol{x}_w^{(\mu)}\right)^\top \boldsymbol{\beta}\right)^2}$, $\boldsymbol{x}_w^{(\mu)}$:μth row vector of $\boldsymbol{X}_w$
    - $\psi_i(\beta_i) = e^{-\gamma\lambda|\beta_i|}$
- The ground state = lasso estimator $\widehat{\boldsymbol{\beta}}_\lambda$
  - Quantities of interest (QOIs) can be computed from the Boltzmann distribution:
    E.g. MSE: $\lim_{N\to\infty}\mathbb{E}\left[\frac{1}{N}\left\|\widehat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}^*\right\|_2^2\right] = \lim_{N\to\infty}\lim_{\gamma\to\infty}\mathbb{E}\left[\mathrm{Tr}_{\boldsymbol{\beta}} P_\gamma(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y})\frac{1}{N}\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\|_2^2\right]$

Validation: Theoretical Analysis

- How to compute the average over the Boltzmann distribution?
  ← Use Belief Propagation (BP) and Density Evolution (DE)

Validation: Theoretical Analysis

- How to compute the average over the Boltzmann distribution?
  ← Use Belief Propagation (BP) and Density Evolution (DE)
- BP: A message-passing algorithm to compute the marginal distributions $P_i(\beta_i|\boldsymbol{y})$

Validation: Theoretical Analysis

- How to compute the average over the Boltzmann distribution?
  ← Use Belief Propagation (BP) and Density Evolution (DE)

- BP: A message-passing algorithm to compute the marginal distributions $P_i(\beta_i|\boldsymbol{y})$

- DE: An algorithm to compute the distribution of the marginal distribution

$$\mathcal{P}_i\big(P_i(\cdot)\big) = \int d\boldsymbol{y}\, P(\boldsymbol{y}) \prod_{\beta_i} \delta\big(P_i(\beta_i) - P_i(\beta_i|\boldsymbol{y})\big)$$

Validation: Theoretical Analysis

- How to compute the average over the Boltzmann distribution?
  $\leftarrow$ Use Belief Propagation (BP) and Density Evolution (DE)
- BP: A message-passing algorithm to compute the marginal distributions $P_i(\beta_i|\boldsymbol{y})$
- DE: An algorithm to compute the distribution of the marginal distribution

$$\mathcal{P}_i\left(P_i(\cdot)\right) = \int d\boldsymbol{y}\, P(\boldsymbol{y}) \prod_{\beta_i} \delta\left(P_i(\beta_i) - P_i(\beta_i|\boldsymbol{y})\right)$$

- Once $\left\{\mathcal{P}_i\left(P_i(\cdot)\right)\right\}_i$ is given, we can compute QOIs
  E.g. $\mathbb{E}\left[\hat{\beta}_i\right] = \lim_{\gamma\to\infty} \mathbb{E}\left[\mathrm{Tr}_{\boldsymbol{\beta}} P_\gamma(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y})\beta_i\right] = \lim_{\gamma\to\infty} \int DP_i\, \mathcal{P}_i\left(P_i(\cdot)\right) \int d\beta_i P_i(\beta_i)\beta_i$

Validation: Theoretical Analysis

- How to compute the average over the Boltzmann distribution?
  ← Use Belief Propagation (BP) and Density Evolution (DE)
- BP: A message-passing algorithm to compute the marginal distributions $P_i(\beta_i|\boldsymbol{y})$
- DE: An algorithm to compute the distribution of the marginal distribution

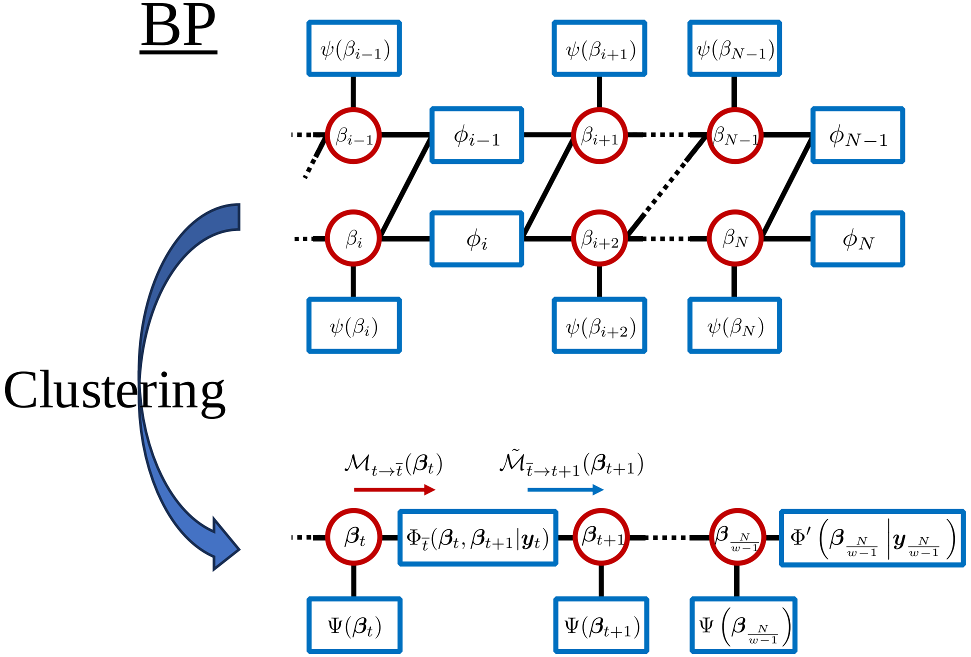
$$\mathcal{P}_i\left(P_i(\cdot)\right) = \int d\boldsymbol{y}\, P(\boldsymbol{y}) \prod_{\beta_i} \delta\left(P_i(\beta_i) - P_i(\beta_i|\boldsymbol{y})\right)$$

- Once $\left\{\mathcal{P}_i\left(P_i(\cdot)\right)\right\}_i$ is given, we can compute QOIs
  E.g. $\mathbb{E}\left[\hat{\beta}_i\right] = \lim_{\gamma\to\infty} \mathbb{E}\left[\mathrm{Tr}_{\boldsymbol{\beta}} P_\gamma(\boldsymbol{\beta} \mid \boldsymbol{X}_w, \boldsymbol{y})\beta_i\right] = \lim_{\gamma\to\infty} \int DP_i\, \mathcal{P}_i\left(P_i(\cdot)\right) \int d\beta_i P_i(\beta_i)\beta_i$
- BP and DE are not executable in general cases
  ← The interaction network (= variables dependence) should be tree

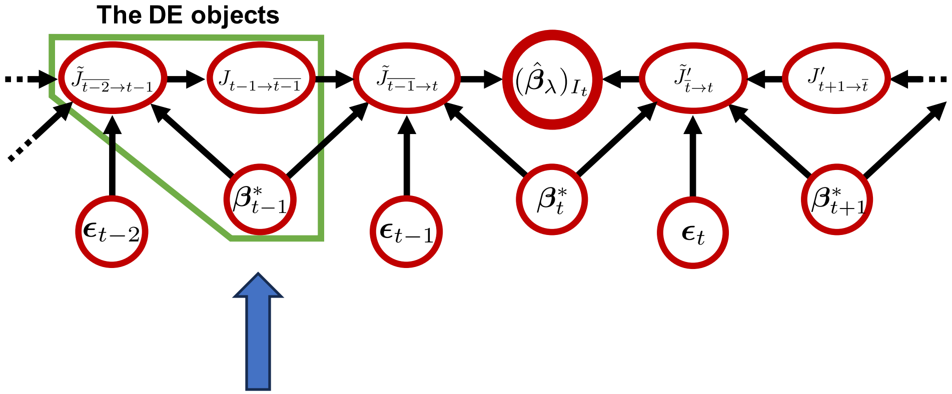
Validation: Theoretical Analysis

- Troubles in the BP and DE application
    - BP: Loops in the interaction network/graphical model ← Clustering
    - DE: Correlations among DE objects ← Conditional distribution considered

BP



DE

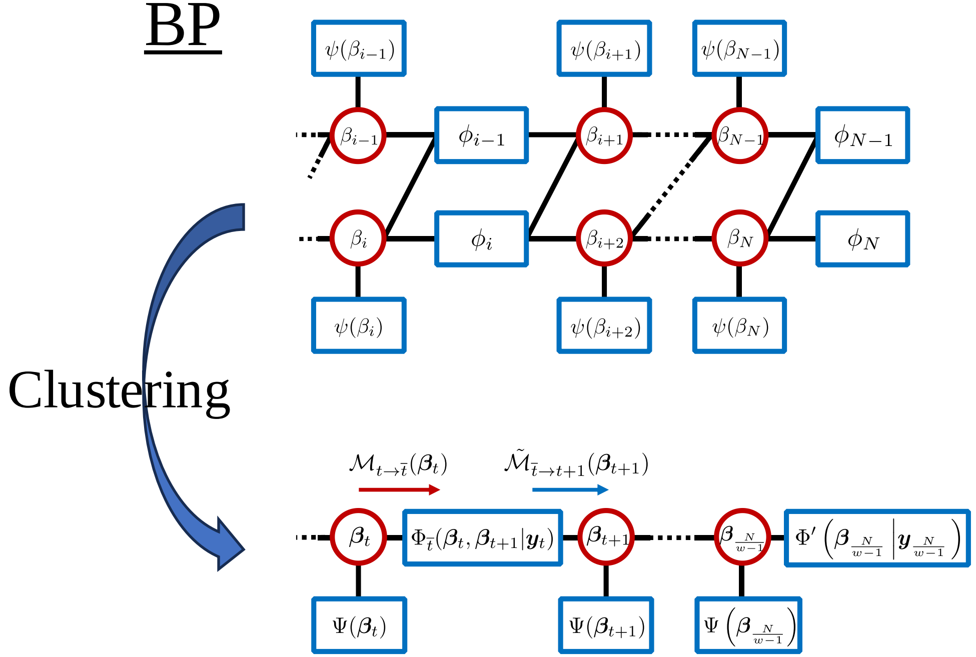


Consider $\mathcal{P}_{\overline{t-1}\to t}(J|\boldsymbol{\beta}_t^*)$ instead of $\mathcal{P}(J)$

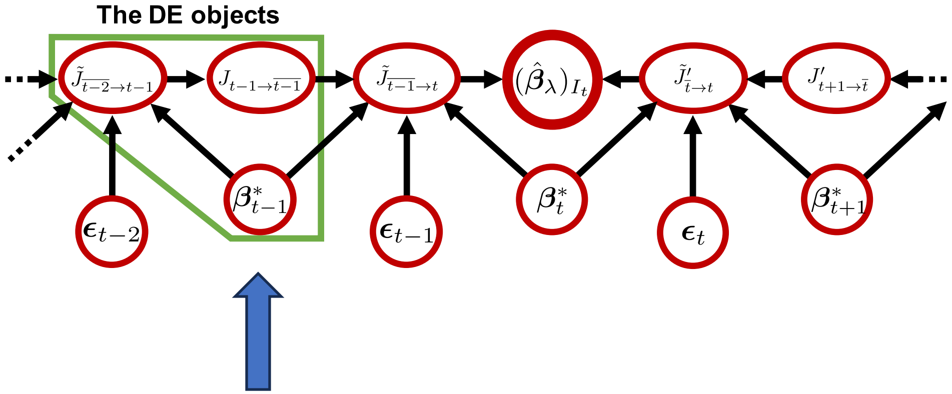
Validation: Theoretical Analysis

- Troubles in the BP and DE application
  - BP: Loops in the interaction network/graphical model ← Clustering
  - DE: Correlations among DE objects ← Conditional distribution considered

BP



DE

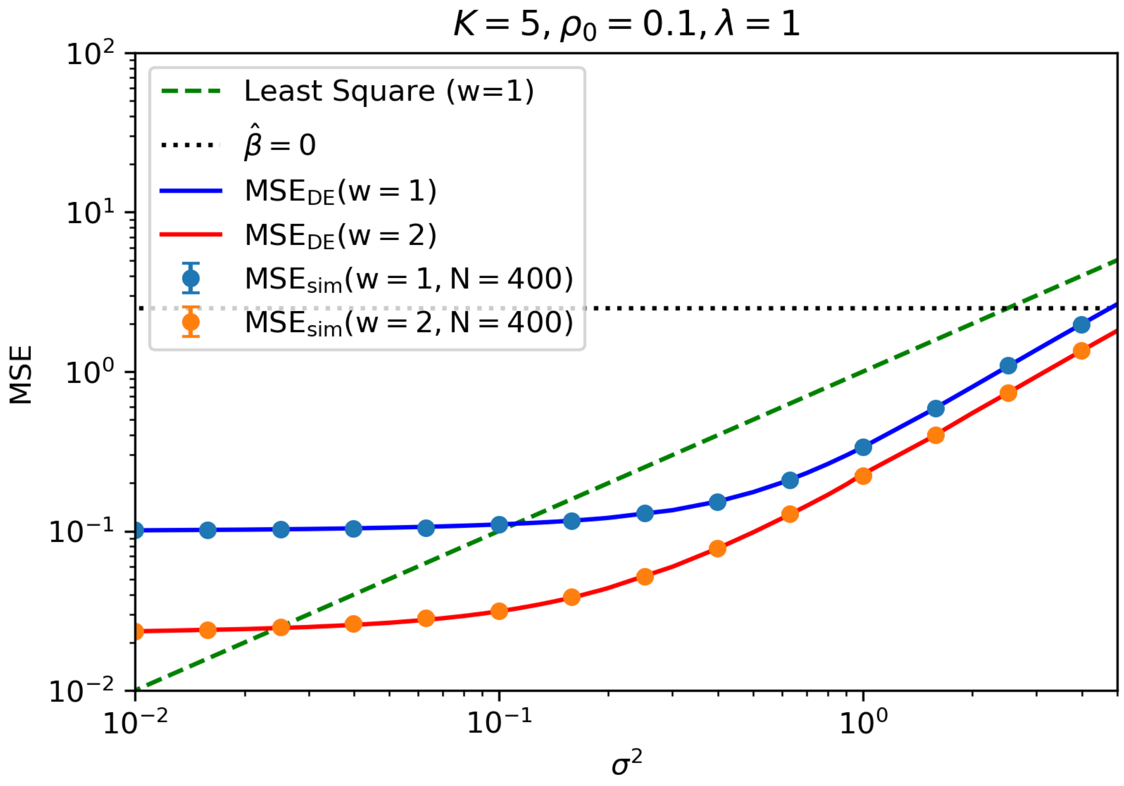


Consider $\mathcal{P}_{\overline{t-1}\to t}(J|\boldsymbol{\beta}_t^*)$ instead of $\mathcal{P}(J)$

- Representation problem in β and $\mathcal{P}$
  - β ← Discretizing and Bounding the DoD → L dim vector → Comp. Cost.= $O(L^{w-1})$
  - $\mathcal{P}$ ← Population of particles $\{J(\boldsymbol{\beta})\}_{i=1}^{N_{\text{pop}}}$

Validation: Theoretical Analysis

- Data model
  - $P_{\beta^*}(\beta_i^*) = (1-\rho)\delta(\beta_i^*) + \rho\delta(\beta_i^* - K)$
  - Noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$



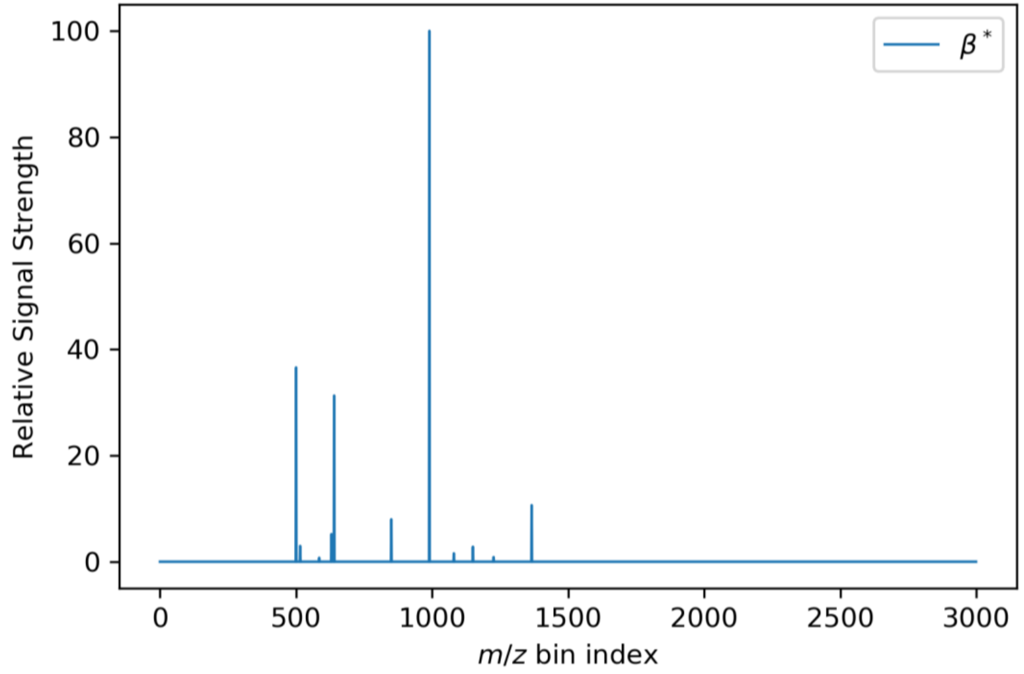
Markers: Simulation results at $N = 400$
Error bar (from 200 realizations) is smaller than marker

- BM-lasso at $w = 2$ is better than that at $w = 1$ at any SNR
- BM-lasso at $\lambda = \lambda_{\mathrm{AIC}}$ becomes always better than Least Square at $w = 1$ (not shown)

BM-lasso's superiority is theoretically checked!

Validation: Simulation Study

- Simulation on semi-artificial setting
  - Dataset: an actual mass spectra registered in Massbank (a public database for MS)
    - m/z range: 50-650
    - 26 mass spectra contained
      - represented as 3000-dimensional vectors $\boldsymbol{\beta}^*$ (converted from the peak list)
        - normalized as the maximum intensity to be 100
  - Noise: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_N)$ with $\sigma^2 = 1$
  - Observation: $\boldsymbol{y} = \boldsymbol{X}_w \boldsymbol{\beta}^* + \boldsymbol{\epsilon}$



Validation: Simulation Study

- Simulation on semi-artificial setting

TABLE II: Means and standard errors of the statistics with the actual mass spectra ($\lambda = \lambda_{AIC}$).

	$w=1$	$w=5$	$w=10$	$w=20$	$w=25$	$w=30$	$w=35$	$w=40$
λ_{AIC}	2.486(.077)	5.065(.177)	6.829(.286)	9.123(.468)	9.788(.559)	10.88(.56)	11.40(.58)	11.76(.47)
MSE	.0226(.0023)	.0090(.0012)	.0060(.0010)	.0050(.0008)	.0043(.0007)	.0044(.0007)	.0046(.0007)	.0054(.0010)
λ_{AIC}/w	2.486(.077)	1.013(.035)	0.683(.029)	0.456(.023)	0.392(.022)	0.363(.019)	0.326(.017)	0.294(.012)
F1 score	0.212	0.210	0.241	0.261	0.258	0.295	0.287	0.286

	True Peak	**True Non-Peak**
Estim. Peak	TP	FP
Estim. Non-Peak	FN	TN

$$\text{F1 Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

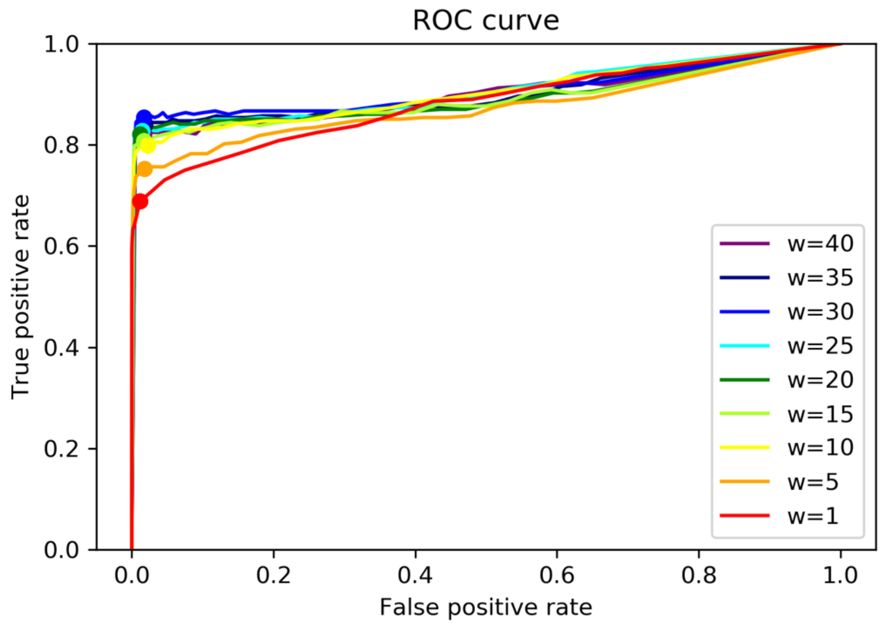
- A binary classifier evaluation index well used in medicine and pharmacy
  - Larger is better

Validation: Simulation Study

- Simulation on semi-artificial setting

TABLE III: Effect of the signal strength on the true positive rate ($\lambda = \lambda_{\text{AIC}}$). The numbers in parentheses represent the ratio of the number of detected peaks to the total number of peaks.

K	$w = 1$	$w = 5$	$w = 10$	$w = 20$	$w = 30$	$w = 40$
0.0-0.5	0.0% (0/41)	7.3% (3/41)	7.3% (3/41)	17% (7/41)	27% (11/41)	17% (7/41)
0.5-1.0	5.3% (1/19)	5.3% (1/19)	47% (9/19)	42% (8/19)	53% (10/19)	42% (8/19)
1.0-1.5	18% (4/22)	45% (10/22)	59% (13/22)	72% (16/22)	77% (17/22)	68% (15/22)
1.5-2.0	20% (3/15)	73% (11/15)	93% (14/15)	93% (14/15)	93% (14/15)	93% (14/15)
2.0-3.0	55% (11/20)	85% (17/20)	90% (18/20)	100% (20/20)	95% (19/20)	100% (20/20)
3.0-5.0	96% (22/23)	100% (23/23)	100% (23/23)	100% (23/23)	100% (23/23)	96% (22/23)
5.0-100	100% (168/168)	100% (168/168)	100% (168/168)	100% (168/168)	100% (168/168)	100% (168/168)



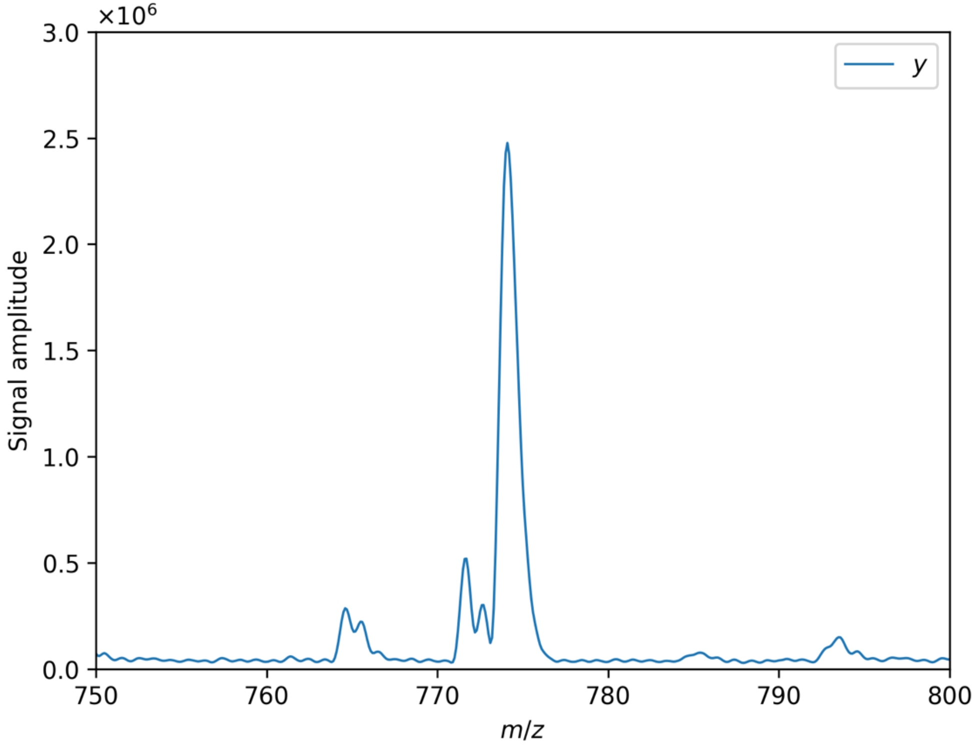
- ROC curve: TP vs FP
  - Each curve is drawn by sweeping λ
  - Circle: $\lambda = \lambda_{\text{AIC}}$
    - Upper left is better

BM-lasso significantly improves peak detection!

Validation: Real Data Analysis

$$\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{pmatrix} + \text{noise} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_2 + \beta_3 \\ \beta_3 + \beta_4 \\ \beta_4 + \beta_5 \\ \beta_5 + \beta_6 \\ \beta_6 + \beta_7 \\ \beta_7 \end{pmatrix} + \text{noise}$$

- Real data observed by QMF



- Experimental settings:
  - Filter: QMF (LCMS-8060)
  - Step size: 0.1 m/z
  - Measurement mode: Unit
    - Related to filter & detector resolution
  - Number of scans: 300
    - Scan time: 1.00 sec/scan
  - Measurement target: Synthetic peptide (EELNAISGPNEFAR, monoisotopic mass: 1545.742)
  - Solvent: H_2O (49.5%), CH_3CN(50%), CH_3COOH(0.5%)
  - Sample concentration: 200 fmol/μL

- A gap from theory & simulation: Is observation matrix $\boldsymbol{X}$ really band matrix?
  - The value of X should be related to the ion transmission rate (device-dependent)
    - Need to estimate from data

Validation: Real Data Analysis

- Observation matrix estimation of a QMF
  - The key ideas
    - Observe the well-known material ← The true signal $\boldsymbol{\beta}^*$ can be known
    - Estimate $\boldsymbol{X}$ by solving $\widehat{\boldsymbol{X}} = \mathrm{argmin}_{\boldsymbol{X}} \{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\|_2^2\}$
      - Restricting the form of $\boldsymbol{X}$ to be against overfitting

$$\boldsymbol{X} = \begin{pmatrix} f(-J) & f(-J+1) & f(-J+2) & \cdots & 0 & 0 & 0 \\ 0 & f(-J) & f(-J+1) & \cdots & 0 & 0 & 0 \\ 0 & 0 & f(-J) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & f(J) & 0 & 0 \\ 0 & 0 & 0 & \cdots & f(J-1) & f(J) & 0 \\ 0 & 0 & 0 & \cdots & f(J-2) & f(J-1) & f(J) \end{pmatrix} \in \mathbb{R}^{N-2J \times N}.$$

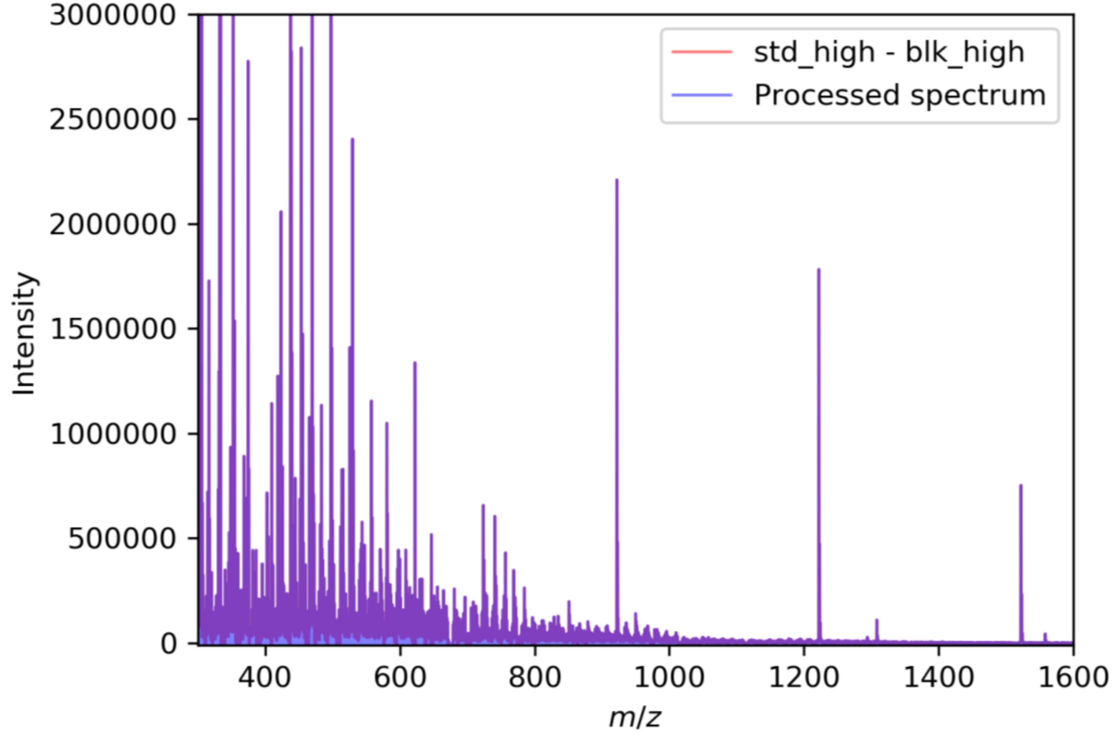
      - $\boldsymbol{f} = \big(f(-J), f(-J+1), \ldots, f(J)\big)$: Band vector ← $\mathrm{argmin}_{\boldsymbol{f}} \{\|\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{f})\boldsymbol{\beta}^*\|_2^2\}$
        - Band width J becomes a hyper parameter
          - should be larger than the true band width (← device property)
          - should be small enough to avoid overfitting

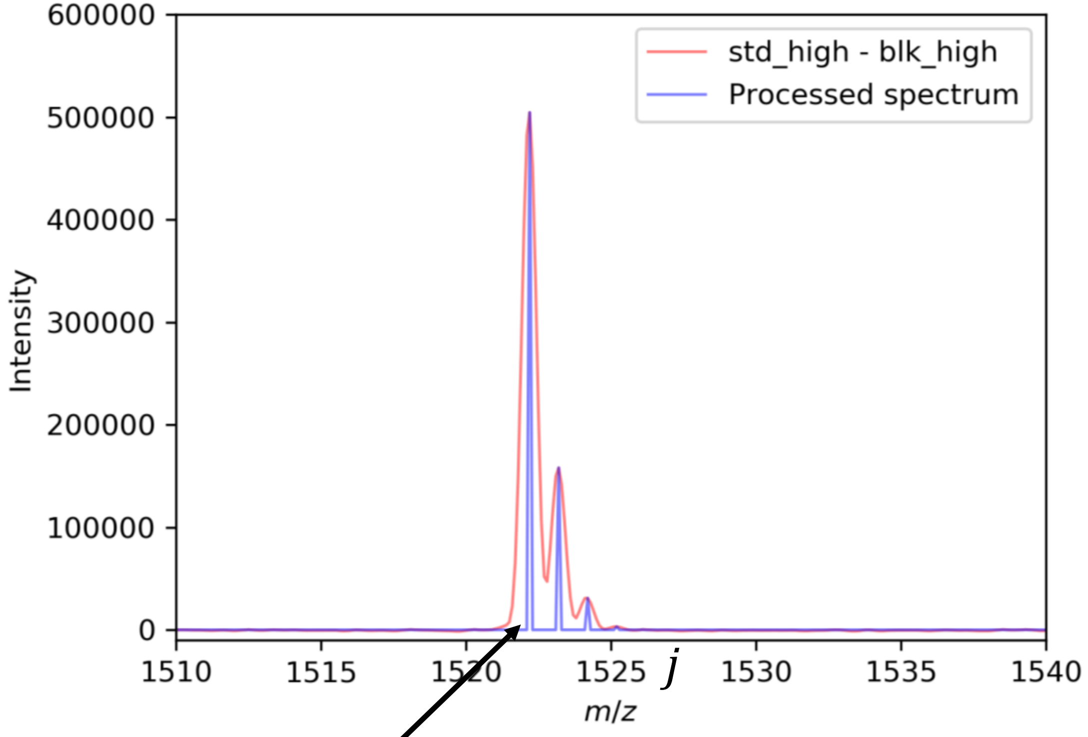
Validation: Real Data Analysis

- Observation matrix estimation of a QMF
  - Experimental settings:
    - Filter: QMF (LCMS-8060)
    - Step size: 0.1 m/z
    - Measurement mode: High, Unit, Low
    - Number of scans: 500
    - Measurement target: ESI-L low concentration tuning mix
      - Reference sample used for calibration of QMF
      - Peak locations: 922,1222,1522 m/z
    - Solvent: $\mathrm{H_2O}$ (50%), $\mathrm{CH_3CN}(49.5\%)$, $\mathrm{CH_3COOH}(0.5\%)$
  - Preprocessing:
    Subtract the blank observation (=observation w/o target) from the observation w/ target
  - Band width: $J = 50$

Validation: Real Data Analysis

- Observation matrix estimation of a QMF

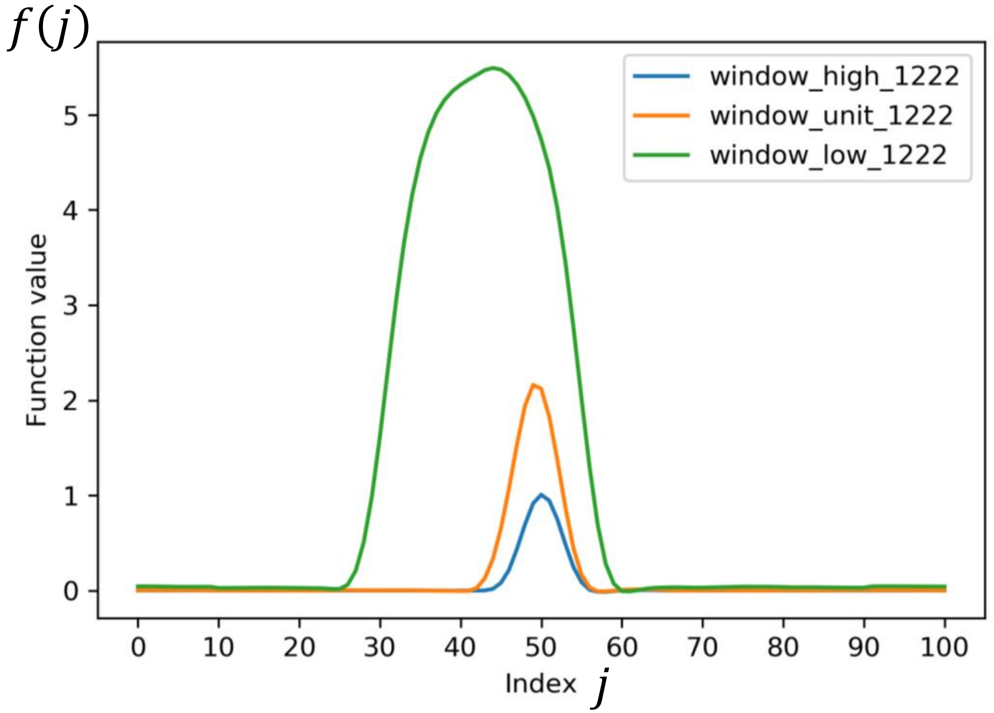


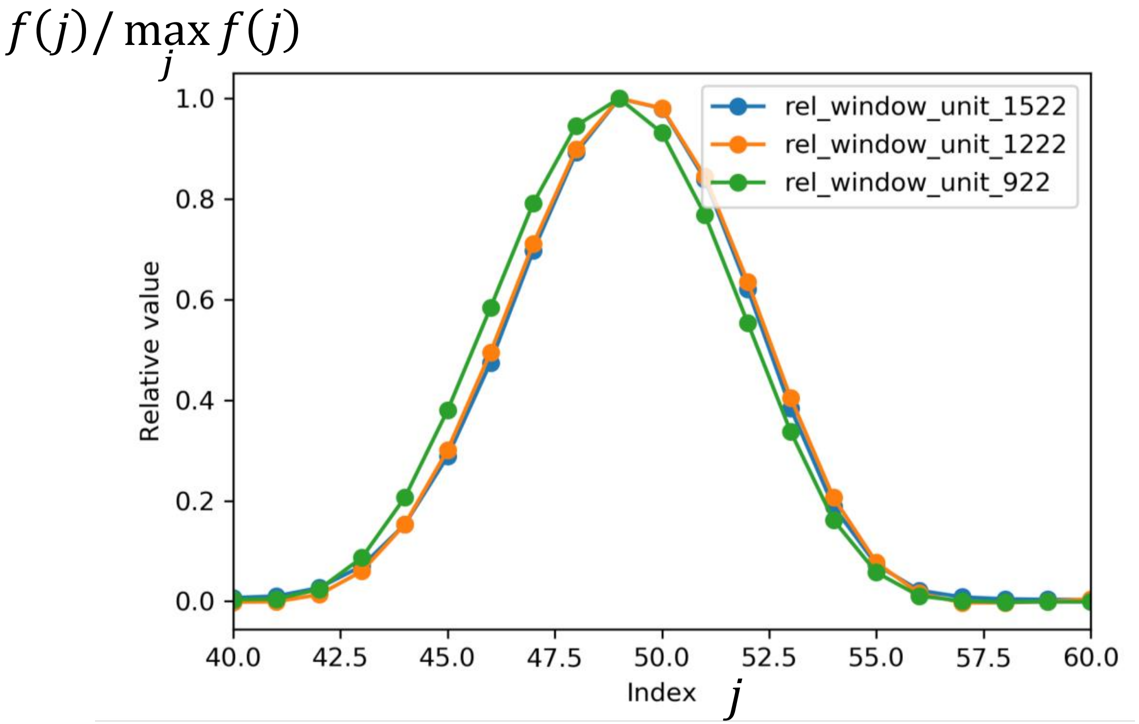


Blue peaks: $\boldsymbol{\beta}^*$
Red curve: $\boldsymbol{y}$

Validation: Real Data Analysis

- Observation matrix estimation of a QMF





- High: FWHM ≈ 0.6 m/z, peak height ≈ 1
- Unit: FWHM ≈ 0.7 m/z, peak height ≈ 2
- Low: FWHM ≈ 2.3 m/z, peak height≥ 5

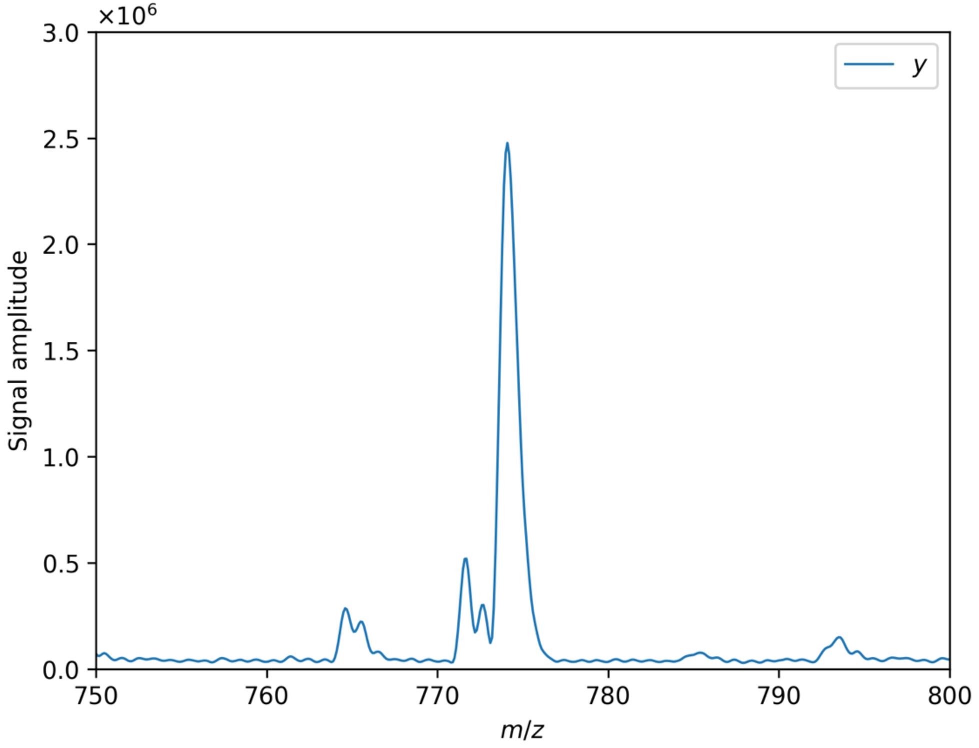
(full width at half maximum)

- m/z dependence seems to be weak

These findings are consistent w/ empirical knowledge and QMF's instructions

Validation: Real Data Analysis

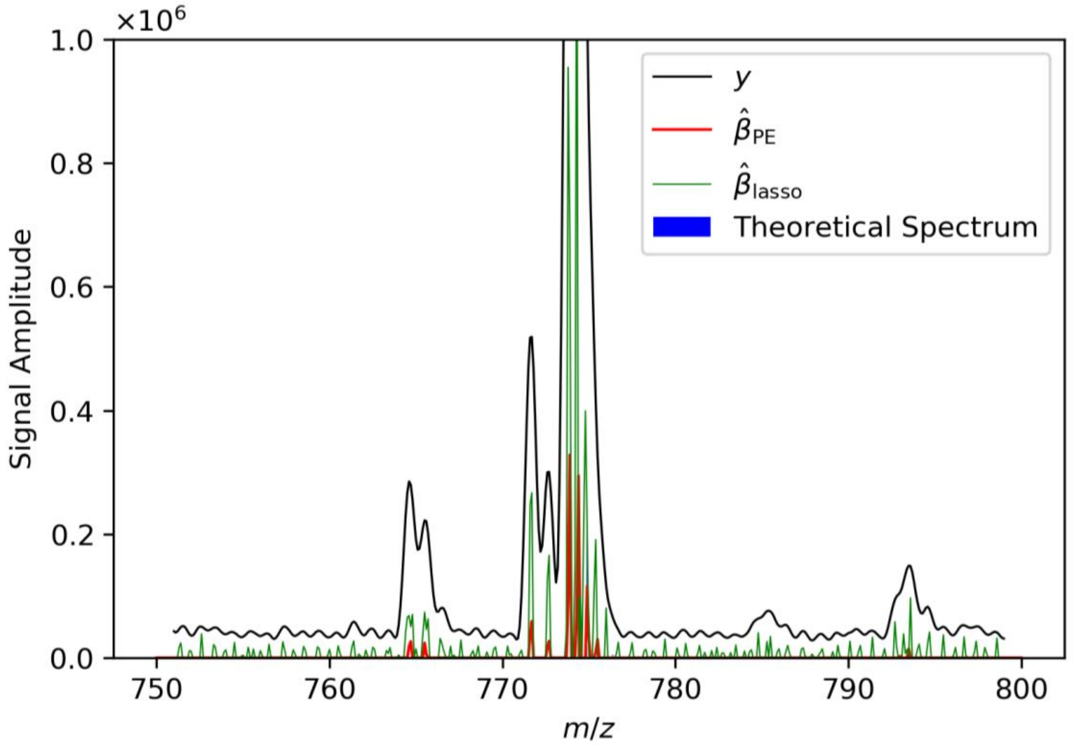
- Real data observed by QMF

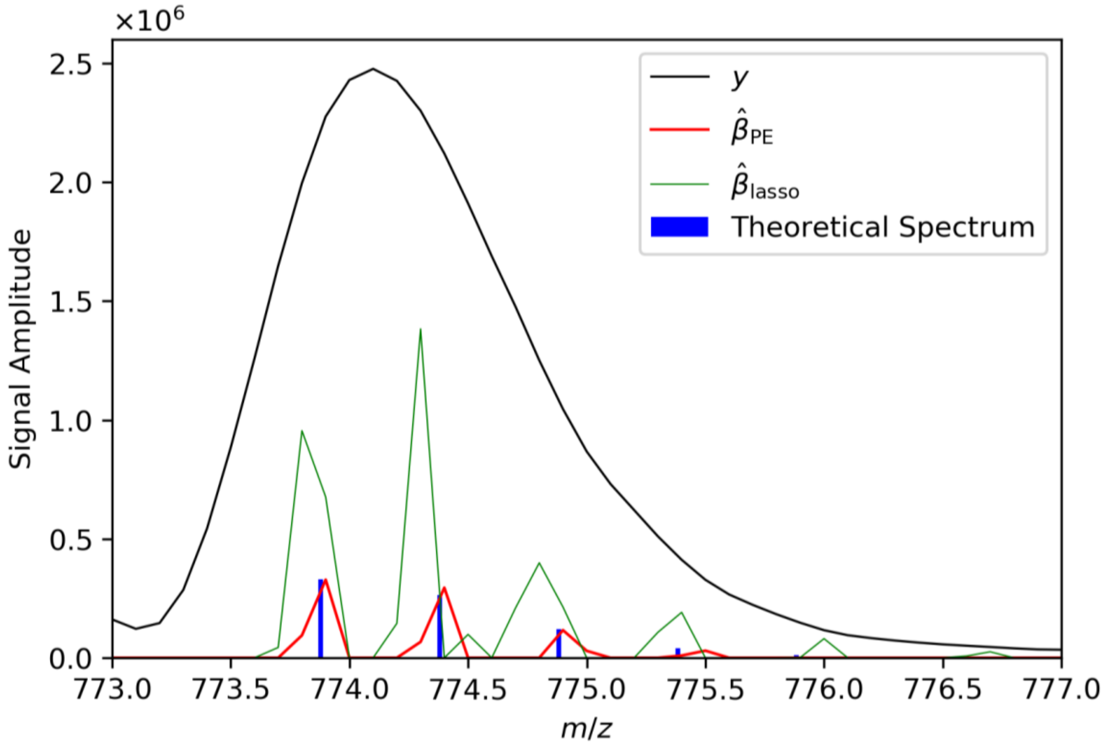


- Experimental settings:
  - Filter: QMF (LCMS-8060)
  - Step size: 0.1 m/z
  - Measurement mode: Unit
    - Related to filter & detector resolution
  - Number of scans: 300
    - Scan time: 1.00 sec/scan
  - Measurement target: Synthetic peptide (EELNAISGPNEFAR, monoisotopic mass: 1545.742)
  - Solvent: $\mathrm{H_2O}$ (49.5%), $\mathrm{CH_3CN}$(50%), $\mathrm{CH_3COOH}$(0.5%)
  - Sample concentration: 200 fmol/μL
- $\boldsymbol{X}$: Estimated using the reference sample

Validation: Real Data Analysis

- Real data observed by QMF



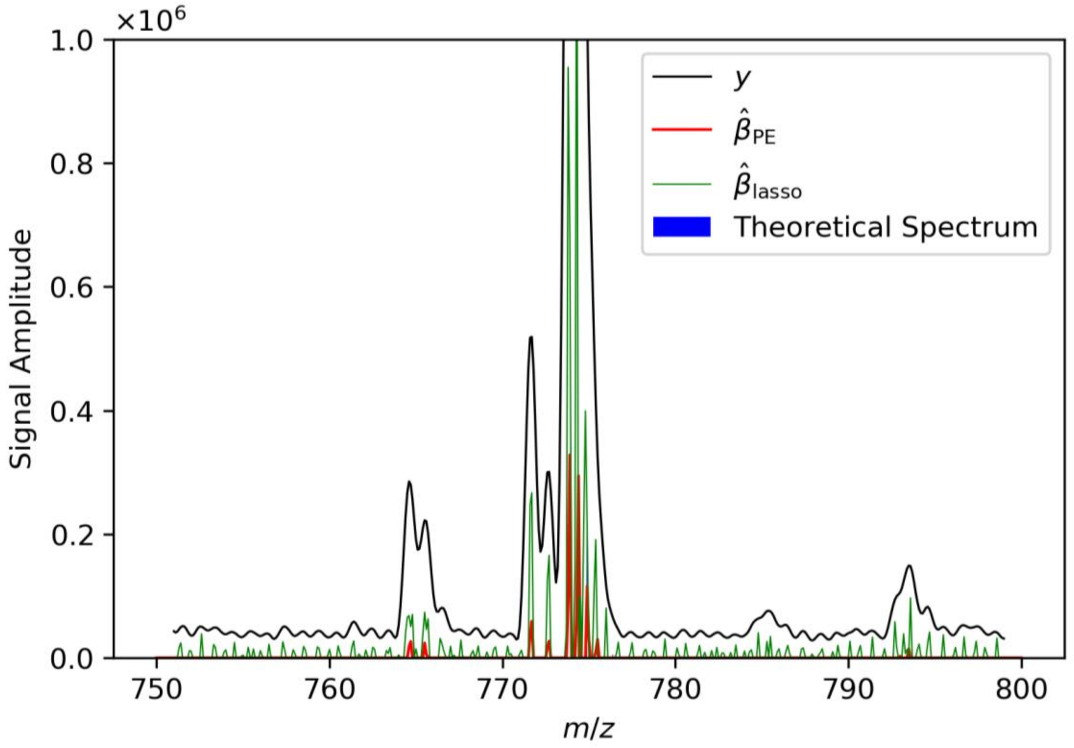


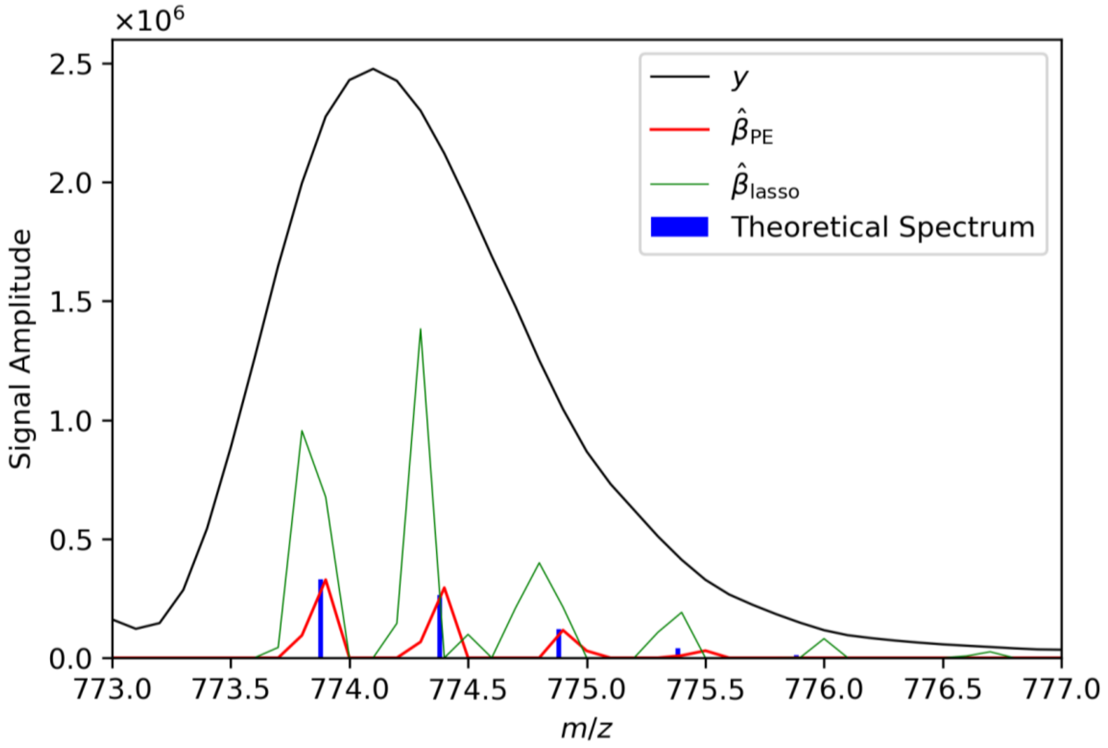
- Finer and reasonable peaks are obtained by BM-lasso
  - The peak at 773.8 m/z
    - $773.8 = \frac{1545.742+2}{2} = (\text{mass} + \text{protonation } (+2H^+ \text{ for ionization}))/(z = 2)$

$\leftarrow$ Consistent result with conventional method & interpretation

Achievement

- Real data observed by QMF





- The peaks are at $\approx 773.8, 774.3, 774.8, 775.3$
  - Isotopes seem to be detected
  - The location difference is $0.5 \rightarrow$ Charge number $z = 2$ is directly suggested

←BM-lasso enables these by exceeding the conventional resolution limit!

Summary

- To enhance the sensitivity and accuracy of spectrometers, we proposed BM-lasso and provided theoretical, numerical, and experimental evidences supporting its superiority.
    - Key ideas
        - Band measurement (BM): Utilizing the low-resolution measurement
        - Sparse Modelling: Utilize the sparsity of the spectrum
            - Lasso is an efficient implementation of this
- Achievements
    - Exceeding the conventional resolution limit of QMF
        - Isotopes and the charge number become directly detectable
    - The method is applicable to many spectrometers other than QMF
    - Technical advance in theoretical treatment:
      A successful example of DE in high-order Markov chain
      with non-Gaussian state distribution

IEEE Trans. Sig. Proc., 72, 1724 (2024)
or
techrxiv.23625177.v1

Patent

- A Japanese patent application is submitted based on the presented result (特開2023-032197)

Schematic