

# スパース表現を探す —辞書学習におけるサンプル複雑度と アルゴリズム—

坂田綾香<sup>A</sup>, 樺島祥介<sup>B</sup>

<sup>A</sup> 統計数理研究所, <sup>B</sup> 東京工業大学

# 目次

## (前半)

1. 辞書学習の導入と先行研究の紹介
2. 辞書学習の応用事例
3. 辞書学習のサンプル複雑度とは

## (後半)

4. 既存の辞書学習のアルゴリズム
5. Bayes推定を用いた辞書学習のアルゴリズム

# 目次

## (前半)

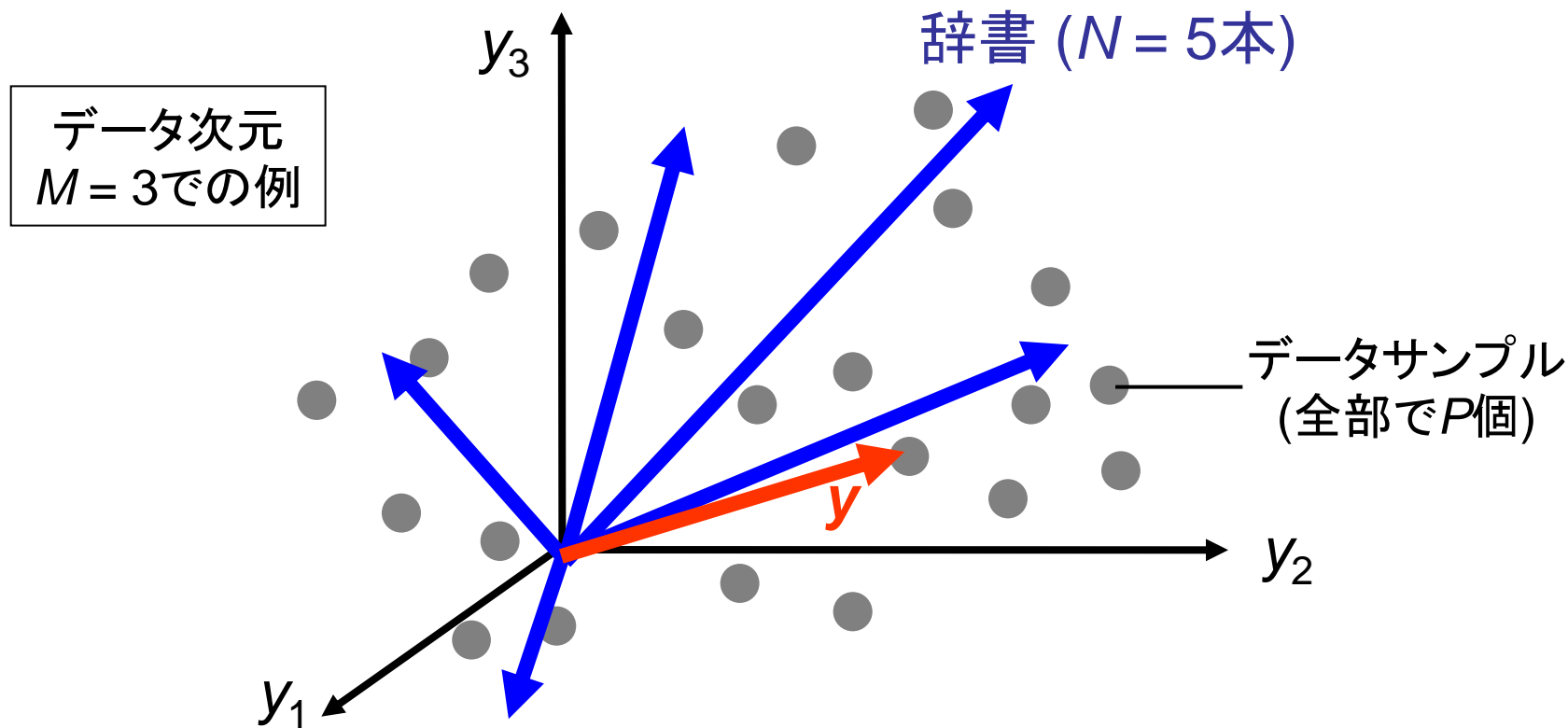
1. 辞書学習の導入と先行研究の紹介
2. 辞書学習の応用事例
3. 辞書学習のサンプル複雑度とは

## (後半)

4. 既存の辞書学習のアルゴリズム
5. Bayes推定を用いた辞書学習のアルゴリズム

# 「辞書」とは

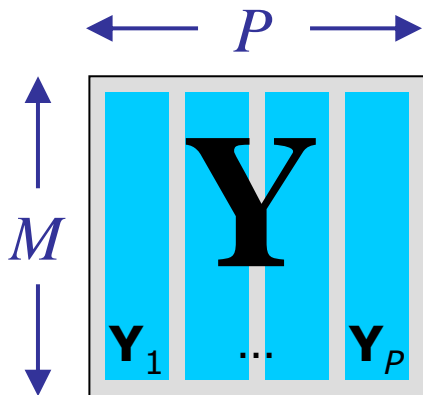
## スパース表現のための“基底”



全ての基底は使わない  $\longrightarrow$  スパース表現

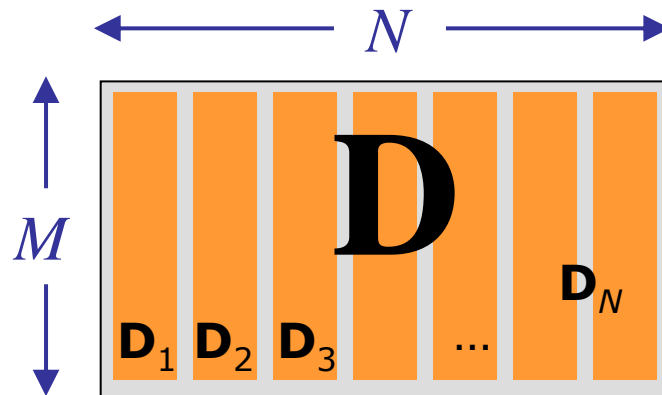
# 行列表記

$M$ 次元、 $P$ 個のデータ

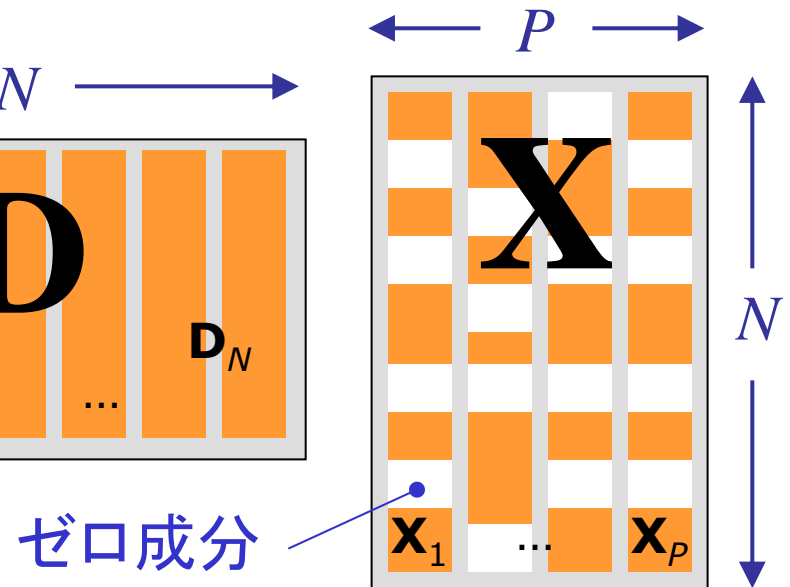


=

$M$ 次元、 $N$ 個の基底(辞書)



$P$ 個のデータごとの  
スパース表現



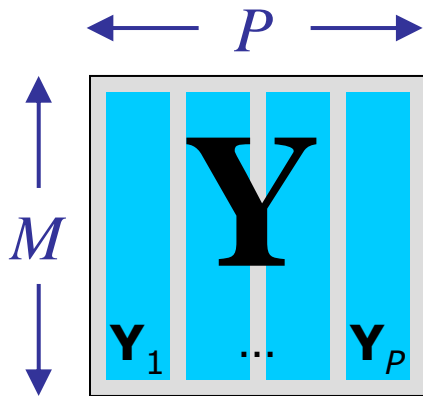
データ次元 < 辞書コラムの数 のとき、  
 $D$ を過完備辞書と呼ぶ。

# 過完備辞書を用いる目的

- 特徴を抽出する
  - データが持つ傾向を辞書として表現する。
- ノイズ耐性を得る
  - ゼロ成分が存在していることで、もとの信号とノイズを分離することが容易になる。
- 圧縮表現を得る
  - 冗長な成分をゼロ成分と見なすことで圧縮された表現が得られる。

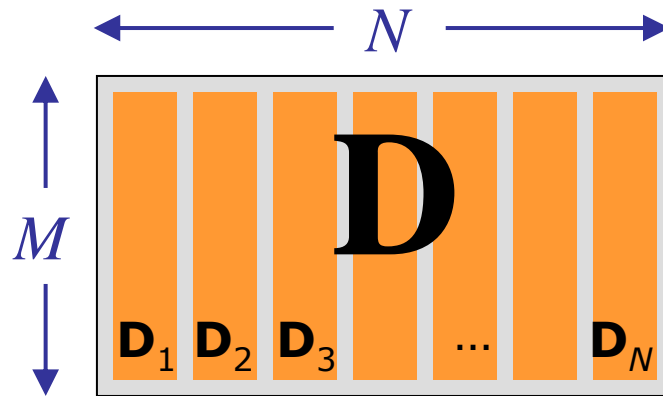
# 辞書学習

$M$ 次元、 $P$ 個のデータ

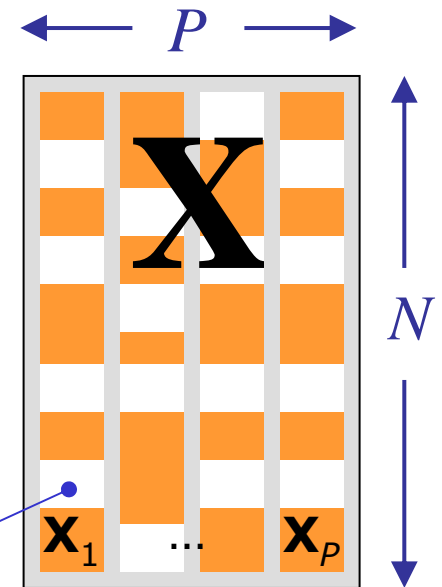


=

$M$ 次元、 $N$ 個の基底(辞書)



$P$ 個のデータごとの  
スパース表現



ゼロ成分

データ $Y$ から  
辞書 $D$ とスパース表現 $X$ を学習することを  
辞書学習と呼ぶ。

# 辞書学習は行列分解の一種

<https://sites.google.com/site/igorcarron2/matrixfactorizations>

## Welcome to The Advanced Matrix Factorization Jungle

g+1 67



[ A living document on the state of the art matrix factorization algorithms and their numerous implementations ]

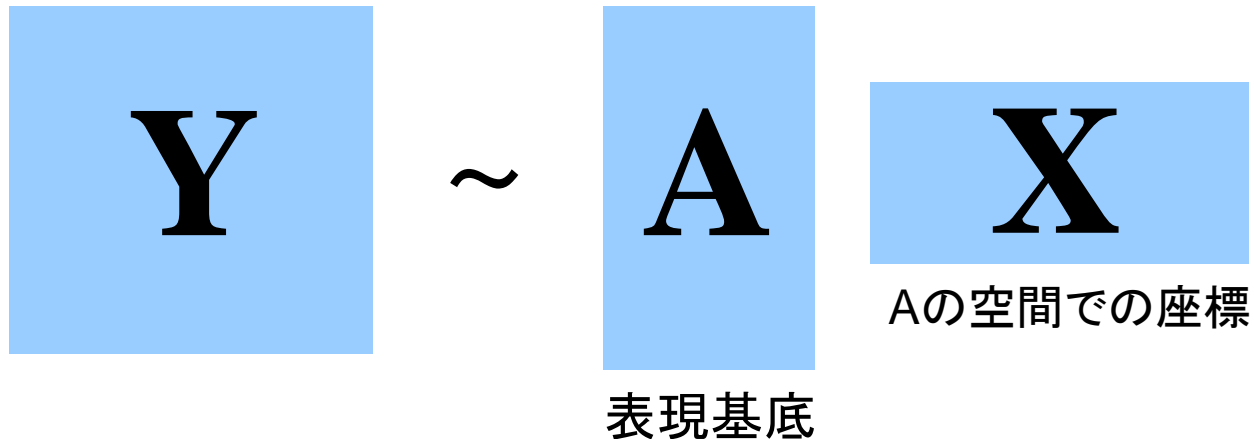
### Table of Content

- Introduction
  - Why this page ?
  - Contributions
  - Heuristics and Phase transitions
  - Notations
- Randomized Algorithms ( Randomized Numerical Linear Algebra (RandNLA))
- Kernel Factorizations
- Spectral Clustering,  $A = DX$  with unknown D and X, solve for **sparse X** and  $X_i = 0$  or  $1$
- K-Means / K-Median:  $A = DX$  with unknown D and X, solve for  $XX^T = I$  and  $X_i = 0$  or  $1$
- Subspace Clustering,  $A = AX$  with unknown X, solve for **sparse/other conditions on X**
- Graph Matching:  $A = XBX^T$  with unknown X, B solve for **B and X as a permutation**
- NMF:  $A = DX$  with unknown D and X, solve for elements of **D, X positive**
- Generalized Matrix Factorization,  $W * L - W * UV'$  with W a known mask, U, V unknowns solve for U, V and **L lowest rank possible**
- Matrix Completion,  $A = H * L$  with H a known mask, L unknown solve for **L lowest rank possible**
- Stable Principle Component Pursuit (SPCP)/ Noisy Robust PCA,  $A = L + S + N$  with L, S, N unknown, solve for **L low rank, S sparse, N noise**



# 行列分解問題

データを行列の積として近似する問題の総称

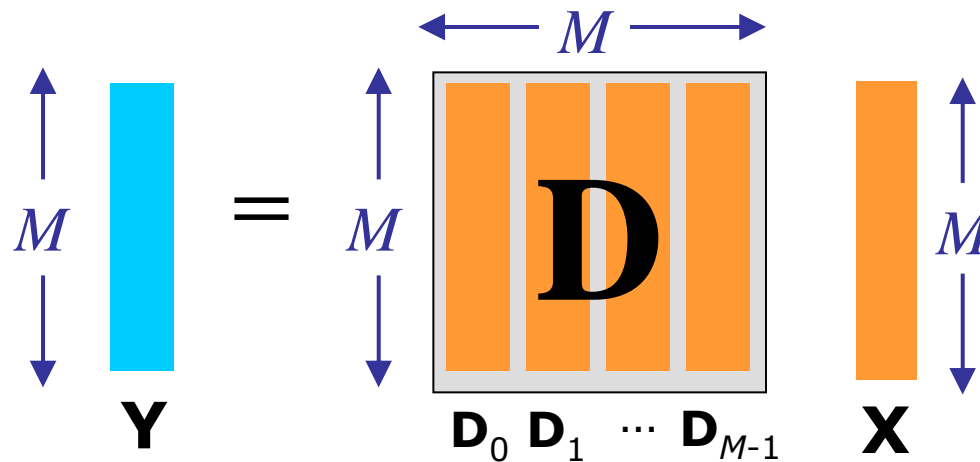


- 個別の問題ごとに、行列の性質を仮定する。
  - 主成分分析(PCA) ... Aの列は互いに直交する
  - 非負因子行列分解(NMF) ... A, Xの要素が非ゼロ
  - 辞書学習(DL)...Xがスパース

# スプース基底に関する研究

## 1. 1960年代: フーリエ基底による信号処理

– 1965 FFT



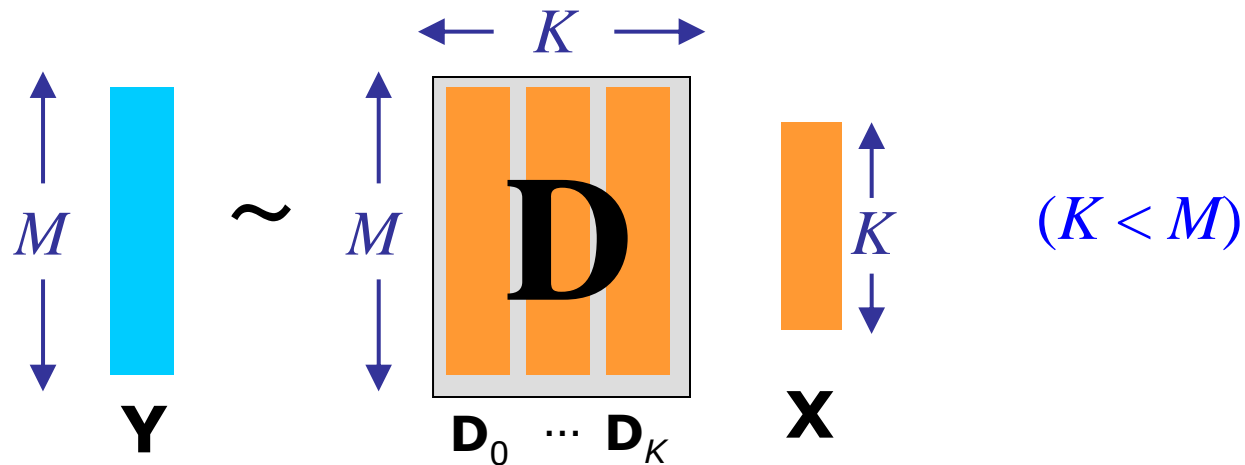
$$\mathbf{D}_n = \exp(iny), \quad \mathbf{x} = \mathbf{D}^T \mathbf{y}$$

# スパース基底に関する研究

## 1. 1960年代: フーリエ基底による信号処理

– 1965 FFT

フーリエ基底を用いた圧縮  
(under complete基底)



$$\mathbf{D}_n = \exp(in\mathbf{y}), \quad \mathbf{x} = \mathbf{D}^T \mathbf{y}$$

# スパース基底に関する研究

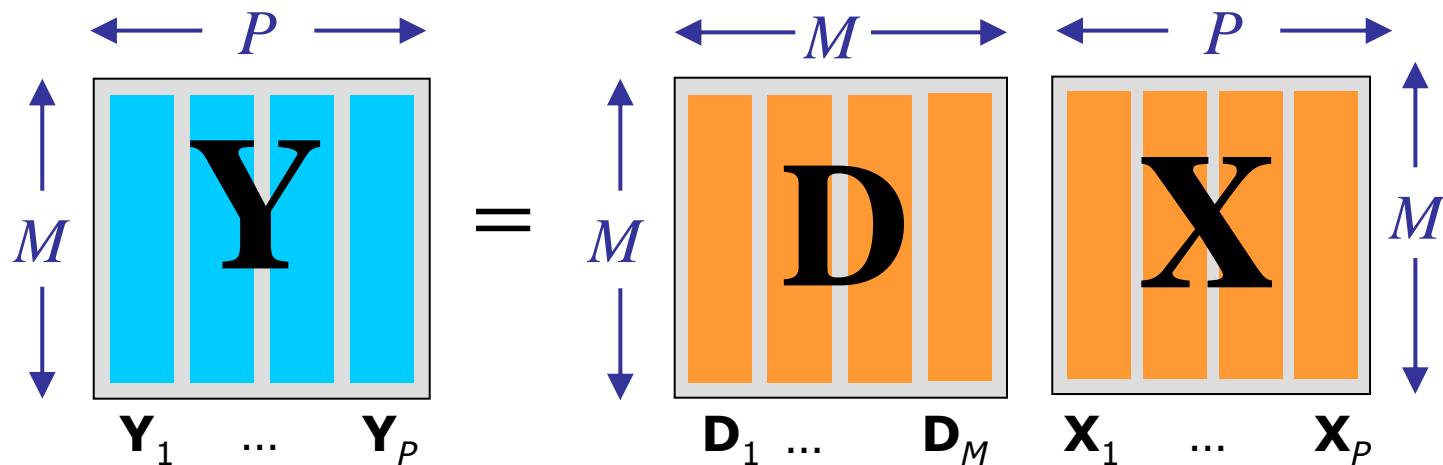
## 2. 1970, 1980年代:主成分分析(PCA)

$$\text{主成分: } \mathbf{D}_1 = \arg \max_{\mathbf{a}: \|\mathbf{a}\|_2=1} \sum_{l=1}^P (\mathbf{a}^T \mathbf{Y}_l)^2 = \arg \max_{\mathbf{a}: \|\mathbf{a}\|_2=1} \mathbf{a}^T \Sigma \mathbf{a}$$

※  $\Sigma = \mathbf{Y}^T \mathbf{Y}$

=  $\Sigma$ の第1固有ベクトル

第 $k$ 成分:  $\mathbf{D}_k = \Sigma$ の第 $k$ 固有ベクトル



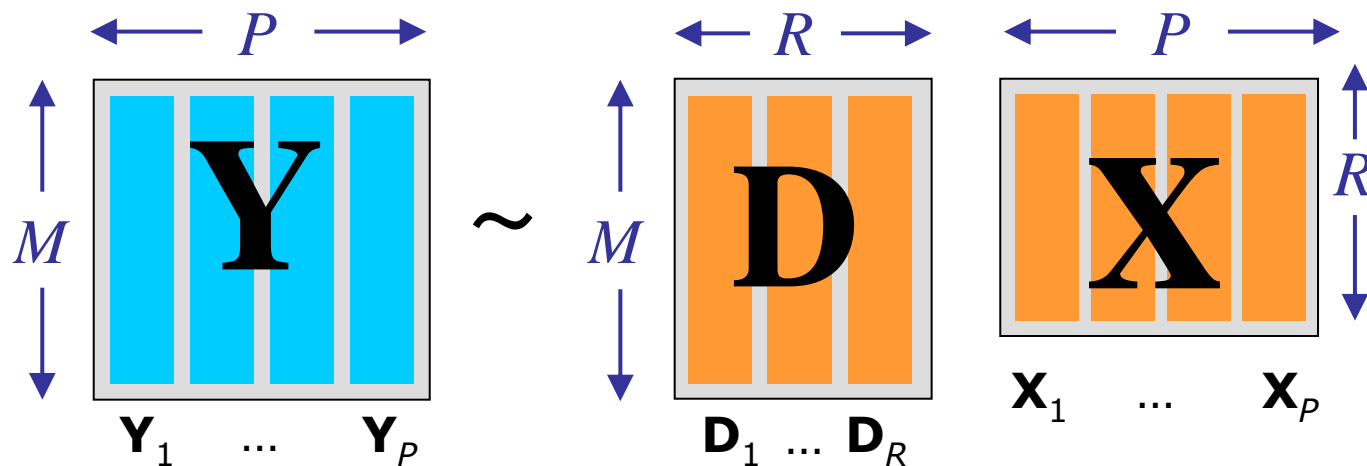
# スパース基底に関する研究

## 2. 1970, 1980年代:主成分分析(PCA)

$$\begin{aligned} \text{主成分: } \mathbf{D}_1 &= \arg \max_{\mathbf{a}: \|\mathbf{a}\|_2^2=1} \sum_{l=1}^P (\mathbf{a}^T \mathbf{Y}_l)^2 = \arg \max_{\mathbf{a}: \|\mathbf{a}\|_2^2=1} \mathbf{a}^T \Sigma \mathbf{a} \\ &= \Sigma \text{の第1固有ベクトル} \end{aligned} \quad \text{※ } \Sigma = \mathbf{Y}^T \mathbf{Y}$$

第 $k$ 成分:  $\mathbf{D}_k = \Sigma$ の第 $k$ 固有ベクトル

PCAに基づく圧縮(under complete基底)



# スパース基底に関する研究

## 3. 過完備基底の提案

- Simoncelli et al. (1992)

- 基底の直交性を破り、冗長にする
- Wavelet変換の並進・回転不変性の欠如を補う目的

- Nason and Silverman (1995)

- Stationary Wavelet Transform

# スパース基底に関する研究

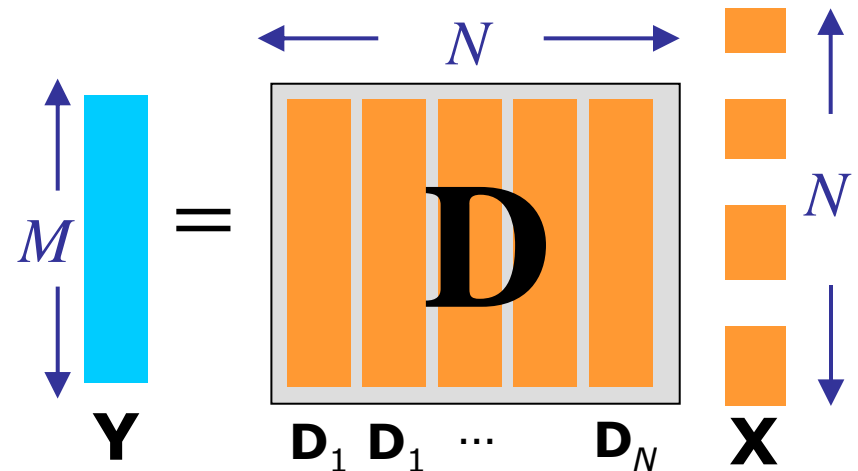
## 4. 「変換」から「基底選択」へ

- 変換によるスパース表現ではなく、  
固定された辞書の選択によりスパース表現を得る

- Chen et al. (1994)

- 基底選択問題をL1最小化として定式化
- L1最小化に対してBasis Pursuitアルゴリズムを提案

$$\min \|\mathbf{x}\|_1, \text{ subject to } \mathbf{y} = \mathbf{D}\mathbf{x}$$



# スパース基底に関する研究

## 5. 辞書の学習

- Olshausen and Field (1997)

- 視覚野における情報表現の疎性について

- エッジやラインなどの、少数の基本的性質が画像の本質である。
- 視覚野において、少数の性質を抽出するコーディングが行われていると考えられる。



- 基底行列の線形和として、高次元の情報を表す方法を提案。ただし和の数は少ないとする(スパース性)
- 最尤推定に基づく学習を通して、入力データから**基底を学習**。



# スパース基底に関する研究

## Olshausen-Fieldの問題設定

画像  $I(\mathbf{x})$  を

$$I(\mathbf{x}) = \sum_i a_i \phi_i(\mathbf{x}_i)$$

として表す基底  $\{\phi_i(\mathbf{x})\}$ 、スパースベクトル  $\{a_i\}$  を知りたい

## 方法

$$E(I, a | \phi) = \sum_{\mathbf{x}_i} \left[ I(\mathbf{x}) - \sum_i a_i \phi_i(\mathbf{x}_i) \right]^2 + \lambda \underbrace{\sum_i S(a_i)}_{\text{スパース正則化}}$$

として

$$\phi^* = \arg \min_{\phi} \left\langle \min_a E(I, a | \phi) \right\rangle \quad \swarrow I \text{ 平均}$$

# スパース基底に関する研究

## 6. 辞書学習に対するアルゴリズムの開発

- Method of Optimal Direction (Engan et al. 1999)
- K – SVD (Aharon et al. 2006)

## 7. 辞書学習に対するサンプル複雑度の解析

- Ahron et al. (2006)
- Vainsencher et. al.(2011)

# 目次

## (前半)

1. 辞書学習の導入と先行研究の紹介
2. 辞書学習の応用事例
3. 辞書学習のサンプル複雑度とは

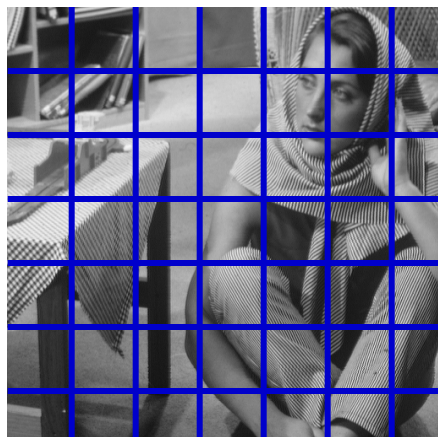
## (後半)

4. 既存の辞書学習のアルゴリズム
5. Bayes推定を用いた辞書学習のアルゴリズム

# 過完備辞書による画像のスパース表現

[Elad (2010)]

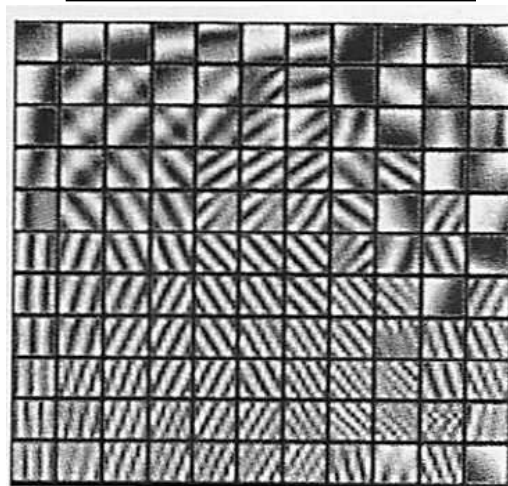
データ集合



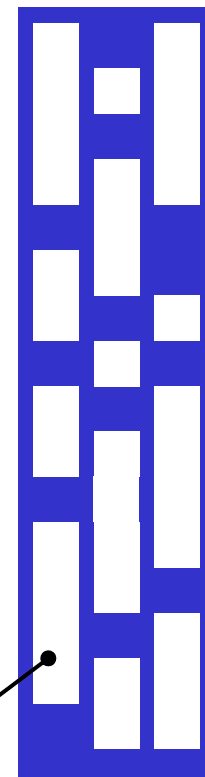
$2.5 \times 10^5$  個のパッチ

~

Dictionary



スパース表現



ゼロ成分

4つのDictionary成分の重ね合わせとして全パッチを表現。

# 辞書学習によるノイズ除去

[Elad and Aharon (2006)]

$$\min_{\mathbf{Z}, \mathbf{D}, \mathbf{X}} \left\{ \lambda \|\mathbf{Y} - \mathbf{Z}\|_2^2 + \sum_{ij} \mu_{ij} \|x_{ij}\|_0 + \|\mathbf{D}\mathbf{X} - \mathbf{Z}\|_2^2 \right\}$$

ノイズを含む画像

ノイズ除去された画像

スパース制約

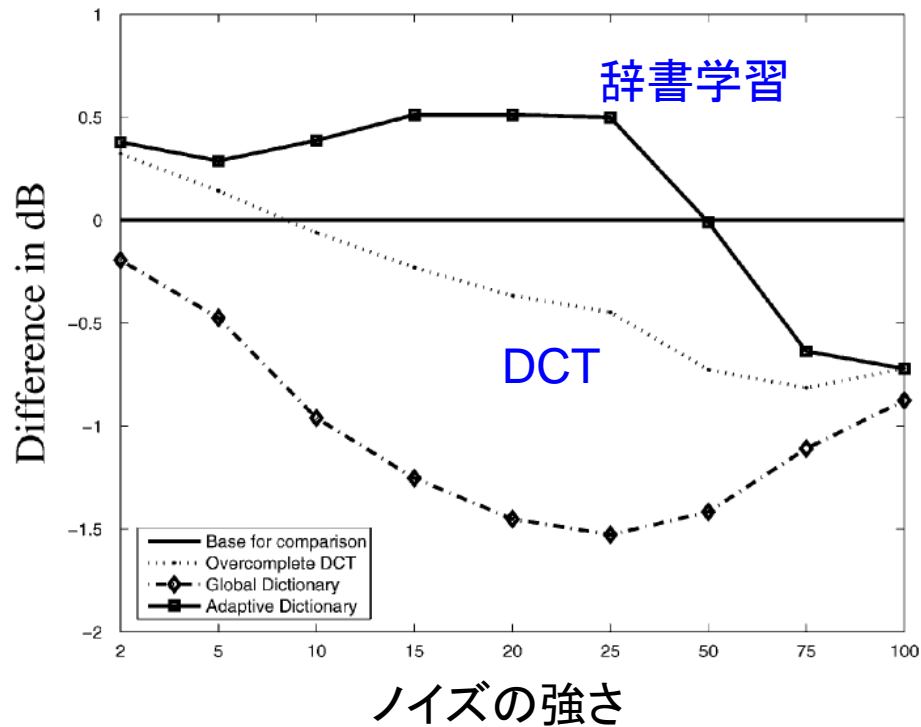
辞書



# 辞書学習によるノイズ除去

[Elad and Aharon (2006)]

$\sigma = 20$



# 目次

## (前半)

1. 辞書学習の導入と先行研究の紹介
2. 辞書学習の応用事例
3. 辞書学習のサンプル複雑度とは

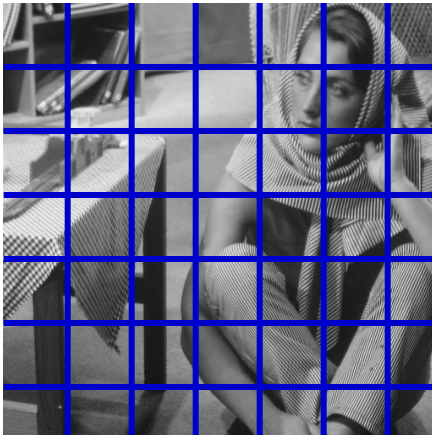
## (後半)

4. 既存の辞書学習のアルゴリズム
5. Bayes推定を用いた辞書学習のアルゴリズム

# Overcomplete dictionaryによるスパース表現

[Elad (2010)]

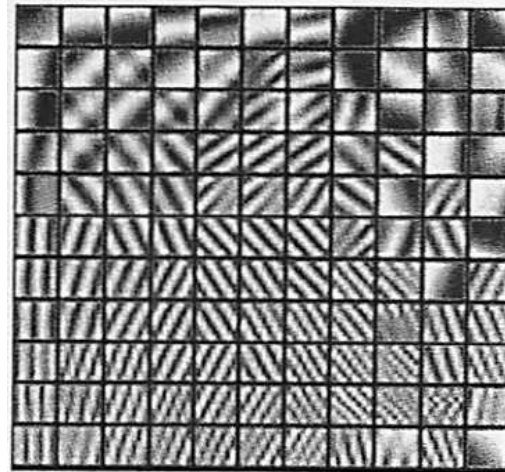
データ集合



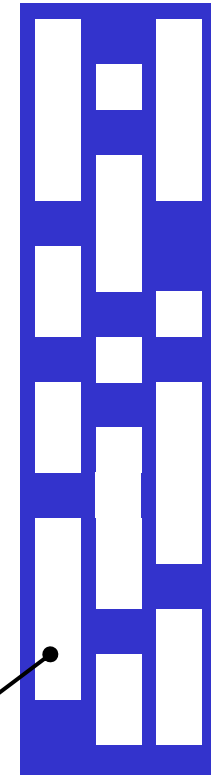
$2.5 \times 10^5$  個のパッチ

~

Dictionary



スパース表現



ゼロ成分

どのくらいサンプルがあれば、dictionaryを決定可能？



# Dictionary同定のための条件 [Aharon et. al. (2006)]

## 1. Support/Spark condition: $\|X^0_i\|_0 = k < \sigma(D^0)/2$

- $\sigma(D^0)$  (spark)...最小の線形従属なコラムの数  
( $D^0 \in \mathbb{R}^{M \times N}$  がランダム行列のとき、 $\sigma(D^0) = M+1$ ).

## 2. Richness condition:

- 同じ $D^0$ コラムの組み合わせ( ${}_N C_k$ 通り)を持つ、 $k+1$ 個のサンプルが存在すること。(したがって、 $P > (k+1) {}_N C_k$ )

## 3. Non-degeneracy condition:

- 同じ $D^0$ コラムの組み合わせを持つ $k+1$ サンプルのランクは  $k$ .
- 異なる $D^0$ コラムの組み合わせを持つ $k+1$ サンプルのランクは $k+1$ .

# Dictionary同定のための条件 [Aharon et. al. (2006)]

## 1. Support/Spark condition: $\|X^0_i\|_0 = k < \sigma(D^0)/2$

- $\sigma(D^0)$  (spark)...最小の線形従属なコラムの数  
( $D^0 \in \mathbb{R}^{M \times N}$  がランダム行列のとき、 $\sigma(D^0) = M+1$ ).

## 2. Richness condition:

- 同じ $D^0$ コラムの組み合わせ( ${}_N C_k$ 通り)を持つ、 $k+1$ 個のサンプルが存在すること。(したがって、 $P > (k+1) {}_N C_k$ )

## 3. Non-degeneracy condition:

- 同じ $D^0$ コラムの組み合わせを持つ $k+1$ サンプルのランクは  $k$ .
- 異なる $D^0$ コラムの組み合わせを持つ $k+1$ サンプルのランクは $k+1$ .

# Dictionary同定のための条件 [Aharon et. al. (2006)]

- Aharon et al. (2006)は
  - 諸条件を満たしたうえで、 $P > P_c \sim \exp(O(N))$ ならば、Dictionaryを一意に同定できることが数学的に証明可。
    - $P > P_c \sim \exp(O(N))$ は十分条件。
- 実際は  $P_c \sim 2N(k+1) \sim O(N^2)$ で十分なのでは、と考察しているが、数学的証明は難しい。  
(後に[Vainsencher et. al.(2011)]により証明される。)
- 一方で、 $MN + NP_\rho$ 個の未知変数に対して既知のデータ数は $MP$ 。
  - $M \sim O(N)$ のとき、 $P \sim O(N)$ で良いのでは?

# 研究動機

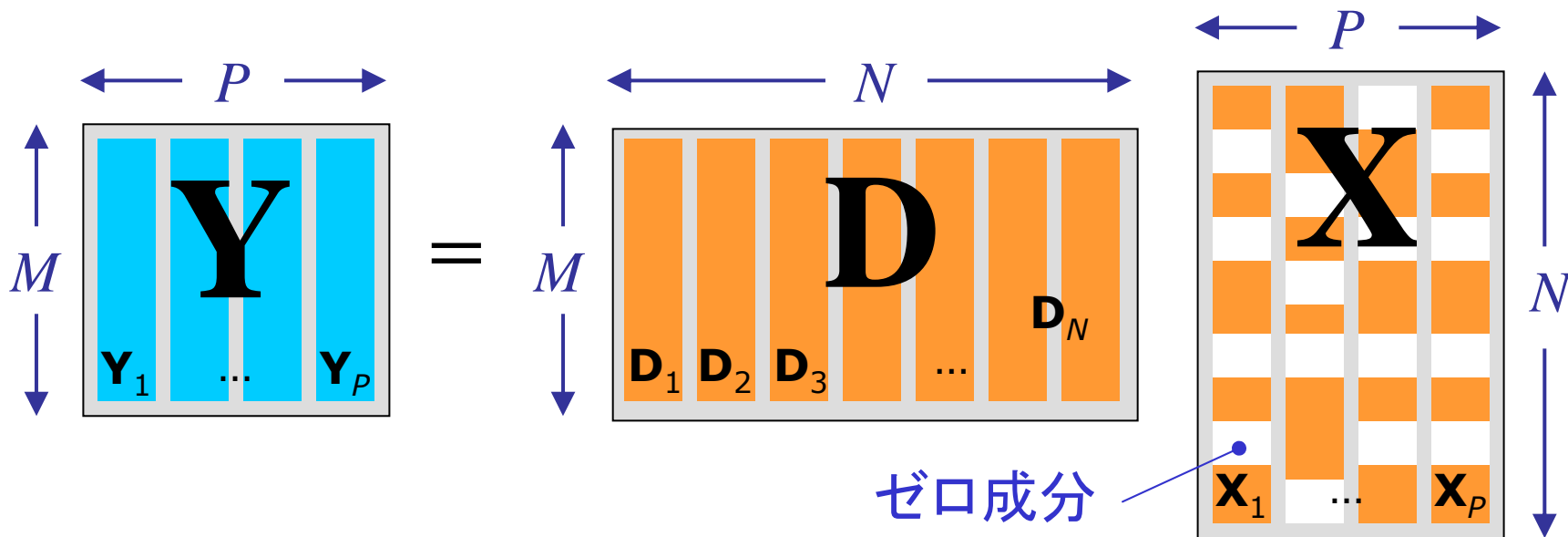
- 統計力学的アプローチから  $P_c$  を見積もる。
- 特に大自由度極限(熱力学的極限)  $N, M, P \rightarrow \infty$  における DL の典型的な振る舞いを調べる。
- Aharon らの planted solution シナリオを採用し, sample complexity の典型的な値を評価する。

# 制約付き二乗誤差最小化による辞書学習

$M$ 次元、 $P$ 個のデータ

$M$ 次元、 $N$ 個の基底(辞書)

$P$ 個のデータごとの  
スパース表現

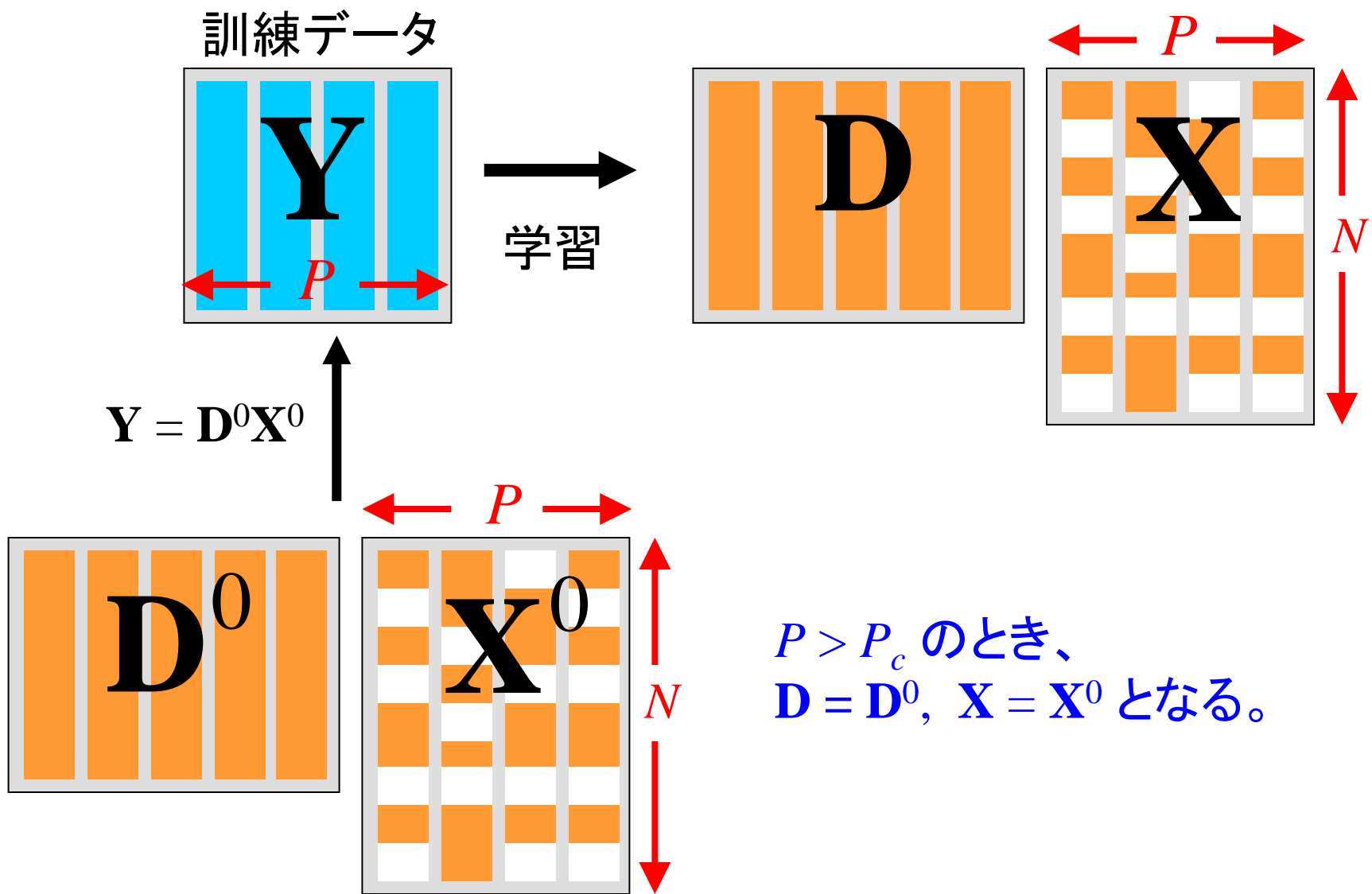


$$\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|^2, \text{ subject to } \|\mathbf{D}\|^2 = MN, \|\mathbf{X}\|_0 = NP\theta$$

辞書を規格化

$\mathbf{X}$ の非ゼロ成分数

# Planted solution scenario



# 制約付き二乗誤差最小化による辞書学習

- 事後分布

$$P_{\beta}(\mathbf{D}, \mathbf{X} | \mathbf{D}^0, \mathbf{X}^0) = \frac{\exp\left(-\beta \|\mathbf{DX} - \mathbf{D}^0 \mathbf{X}^0\|^2 / 2N\right)}{Z_{\beta}(\mathbf{D}^0, \mathbf{X}^0)} \\ \times \delta(\|\mathbf{D}\|^2 - NM) \delta(\|\mathbf{X}\|_0 - NP\theta)$$

➤  $\beta \rightarrow \infty$  で  $\|\mathbf{D}^0 \mathbf{X}^0 - \mathbf{DX}\|^2$  の最小化が実現する。

- 求めたいもの

➤  $\mathbf{D}$  と  $\mathbf{D}^0$  の類似度、 $\mathbf{X}$  と  $\mathbf{X}^0$  の類似度の平均値

➤  $\mathbf{D}$ 、 $\mathbf{X}$  の分散

➤  $\|\mathbf{DX} - \mathbf{D}^0 \mathbf{X}^0\|^2$  の平均値

# DLのP依存性を評価する

①  $\mathbf{D}$ と $\mathbf{D}^0$ ,  $\mathbf{X}$ と $\mathbf{X}^0$ の平均二乗誤差 (一成分当たり)

$$\text{MSE}_D = \frac{1}{MN} \left[ \left\| \langle \mathbf{D} \rangle - \mathbf{D}_0 \right\|^2 \right]_0 = 2 \left( 1 - \frac{1}{MN} \left[ \langle \mathbf{D} \rangle \cdot \mathbf{D}_0 \right]_0 \right) \equiv 2(1 - m_D)$$

$P(\mathbf{D}, \mathbf{X} | \mathbf{D}^0, \mathbf{X}^0)$ による  
 $\mathbf{D}$ ,  $\mathbf{X}$ 平均

$P_0(\mathbf{D}^0, \mathbf{X}^0)$ による  
 $\mathbf{D}_0$ ,  $\mathbf{X}_0$ 平均

$$\begin{aligned} \text{MSE}_X &= \frac{1}{NP} \left[ \left\| \langle \mathbf{X} \rangle - \mathbf{X}_0 \right\|^2 \right]_0 \\ &= \rho - \frac{2}{NP} \left[ \langle \mathbf{X} \rangle \cdot \mathbf{X}_0 \right]_0 + \frac{1}{NP} \left[ \langle \mathbf{X} \rangle^2 \right]_0 \equiv \rho - 2m_X + Q_X \end{aligned}$$

$\text{MSE}_D = 0$ ,  $\text{MSE}_X = 0$ なら、 $\langle \mathbf{D} \rangle$ ,  $\langle \mathbf{X} \rangle$ は $\mathbf{D}^0$ ,  $\mathbf{X}^0$ に収束



# DLの*P*依存性を評価する

## ② **D**と**X**の分散（一成分当たり）

$$\chi_D = \frac{\beta}{MN} \sum_{\mu,i} \left( \left[ \langle D_{\mu i}^2 \rangle \right]_0 - \left[ \langle D_{\mu i} \rangle^2 \right]_0 \right) = \beta \left( 1 - \frac{1}{MN} \sum_{\mu,i} \left[ \langle D_{\mu i} \rangle^2 \right]_0 \right)$$

$$\chi_X = \frac{\beta}{NP} \sum_{il} \left( \left[ \langle X_{il}^2 \rangle \right]_0 - \left[ \langle X_{il} \rangle^2 \right]_0 \right)$$

# DLのP依存性を評価する

## ③ エネルギー密度 (一成分当たり)

$$f = \text{extr}_{\Omega, \hat{\Omega}} \left\{ - \underbrace{\alpha}_{M/N} \left( \frac{\hat{Q}_D - \hat{\chi}_D \chi_D}{2} - \hat{m}_D m_D + \frac{\hat{\chi}_D + \hat{m}_D^2}{2\hat{Q}_D} \right) - \underbrace{\gamma}_{P/N} \left( \frac{\hat{Q}_X Q_X - \hat{\chi}_X \chi_X}{2} - \hat{m}_X m_X + \lambda \theta + \left\langle \left\langle \phi(h; \hat{Q}_X, \lambda) \right\rangle \right\rangle_h \right) + \frac{\alpha \gamma (Q_X - 2m_D m_X + \rho)}{2(1 + Q_X \chi_D + \chi_X)} \right\}$$

$$- \Omega = \{m_D, \chi_D, Q_X, m_X, \chi_X\}, \hat{\Omega} = \{\hat{Q}_D, \hat{m}_D, \hat{\chi}_D, \hat{Q}_X, \hat{m}_X, \hat{\chi}_X, \lambda\}$$

$m_D, m_X, Q_X, \chi_D, \chi_X$ は、 $f$ の停留点として与えられる。

# DLのP依存性を評価する:まとめ

## 平均二乗誤差と分散

$$\Omega^* = \arg \operatorname{extr}_{\Omega} f(\Omega, \hat{\Omega}) \longrightarrow \begin{matrix} \text{MSE}_{D'} & \text{MSE}_{X'} \\ \chi_{D'} & \chi_X \end{matrix}$$

$\{m_D, \chi_D, Q_X, m_X, \chi_X\}$

## エネルギー

$$f(\Omega^*, \hat{\Omega}^*) = \frac{1}{MP} \left[ \left\| \langle \mathbf{DX} \rangle - \mathbf{D}^0 \mathbf{X}^0 \right\|^2 \right]_0$$

5変数の連立方程式を解き、 $\Omega^*$ を求めればよい。

→ 複数個の解が存在する。

# 解の分類

## 解①

$$\begin{aligned} & \bullet m_D = 1 \\ & \bullet m_X = \rho \\ & \bullet Q_X = \rho \end{aligned} \left\{ \begin{array}{l} \bullet \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \bullet \chi_D < \infty, \chi_X < \infty \longrightarrow f = 0 \end{array} \right.$$

## 解②

$$\begin{aligned} & \bullet m_D = 0 \\ & \bullet m_X = 0 \\ & \bullet Q_X \in \mathbb{R}_+ \end{aligned} \left\{ \begin{array}{l} \bullet \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \bullet \chi_D < \infty, \chi_X < \infty \longrightarrow f \neq 0 \end{array} \right.$$

# 解の分類

## 解①: 成功解(S)

- $MSE_D = 0$
- $MSE_X = 0$

$$\left\{ \begin{array}{l} \cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f = 0 \end{array} \right.$$

## 解②: 失敗解(F)

- $MSE_D > 0$
- $MSE_X > 0$

$$\left\{ \begin{array}{l} \cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f \neq 0 \end{array} \right.$$

# 解の分類

## 解①: 成功解(S)

- $MSE_D = 0$
- $MSE_X = 0$

$$\cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0$$

$$\cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f = 0$$

この解が唯一の安定解として存在するとき、学習成功。

## 解②: 失敗解(F)

- $MSE_D > 0$
- $MSE_X > 0$

$$\cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0$$

$$\cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f \neq 0$$

# 解の分類

## 解①: 成功解(S)

- $MSE_D = 0$
- $MSE_X = 0$

$$\cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0$$

$$\cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f = 0$$

この解が唯一の安定解として存在するとき、学習成功。

## 解②: 失敗解(F)

- $MSE_D > 0$
- $MSE_X > 0$

$$\cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0$$

$$\cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f \neq 0$$

この解が存在するとき、 $D^0 X^0 = DX$ を満たす  $D \neq D^0, X \neq X^0$  がたくさん存在。

# ① 成功解の $\gamma$ 依存性

- $\text{MSE}_D = 0$
  - $\text{MSE}_X = 0$
- $\left\{ \begin{array}{l} \cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \dots 1 < \gamma < \gamma_S \text{で存在} \\ \cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f = 0 \\ \dots \alpha > \theta_{\text{eff}}^S(\theta, \rho), \gamma_S < \gamma \text{で存在} \end{array} \right.$

$$\alpha > \theta_{\text{eff}}^S(\theta, \rho) = \theta + (1 - \rho) \sqrt{\frac{2}{\pi}} u \exp\left(-\frac{u^2}{2}\right)$$

$$\gamma > \gamma_S(\alpha, \theta, \rho) = \frac{\alpha}{\alpha - \theta_{\text{eff}}^S(\theta, \rho)}$$

$u$  は次のように決められる。

$$\int_u^{+\infty} dz \frac{\exp(-z^2/2)}{\sqrt{2\pi}} = \frac{\theta - \rho}{2(1 - \rho)}$$



## ② 失敗解の $\gamma$ 依存性

- $\text{MSE}_D > 0$
  - $\text{MSE}_X > 0$
- $\left\{ \begin{array}{l} \cdot \chi_D = \infty, \chi_X = \infty \longrightarrow f = 0 \\ \dots 0 < \gamma < \gamma_F \text{ で存在。} \\ \cdot \chi_D < \infty, \chi_X < \infty \longrightarrow f \neq 0 \\ \dots \alpha > \theta_{\text{eff}}^F(\theta, \rho), \gamma_F < \gamma \text{ で存在。} \end{array} \right.$

$$\alpha > \theta_{\text{eff}}^F(\theta) = \theta + \sqrt{\frac{2}{\pi}} v \exp\left(-\frac{v^2}{2}\right)$$

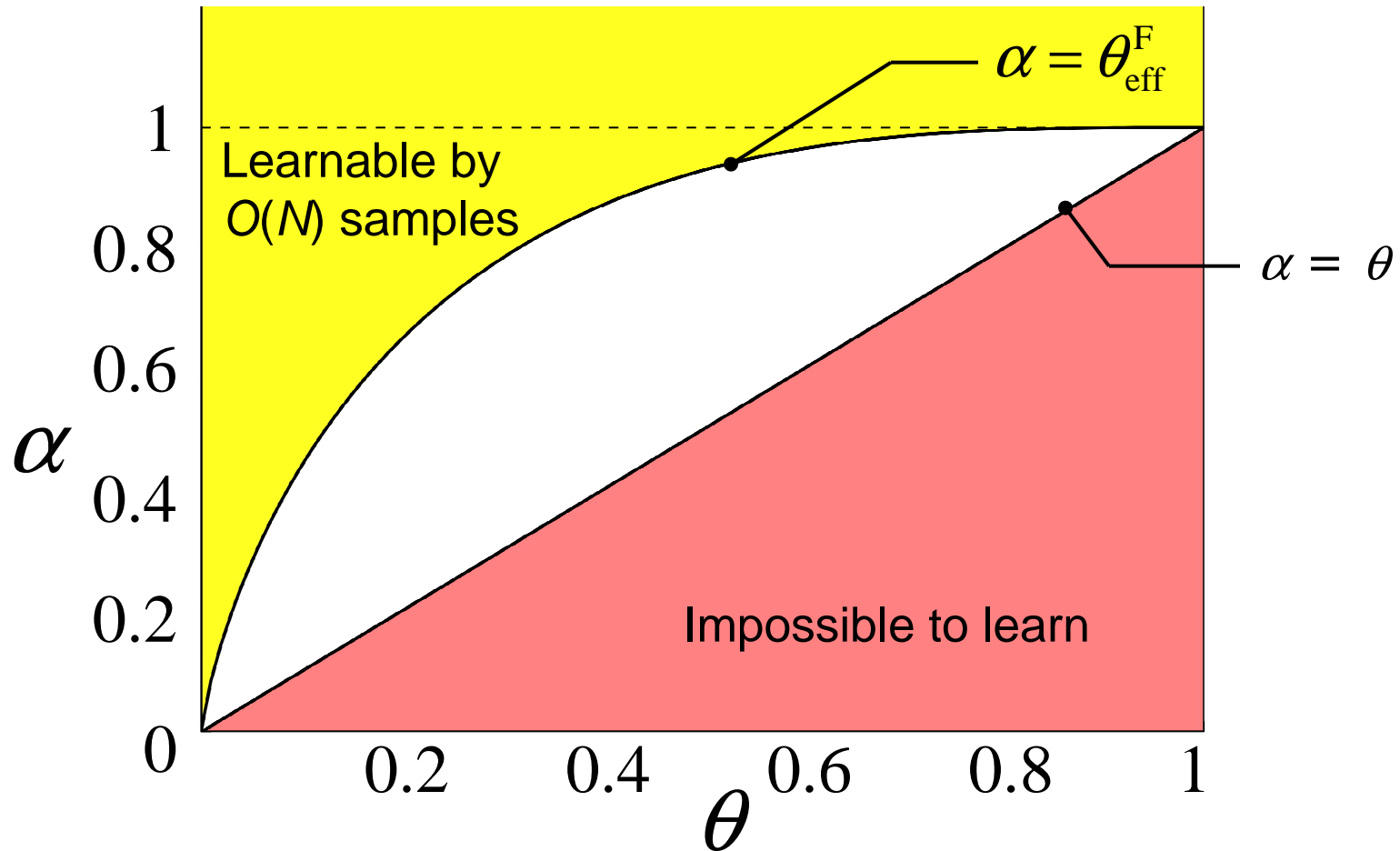
$$\gamma > \gamma_F(\alpha, \theta) = \frac{a}{\left(\sqrt{\alpha} - \sqrt{\theta_{\text{eff}}^F}\right)^2}$$

$v$  は次のように決められる。

$$\int_v^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) = \frac{\theta}{2}$$

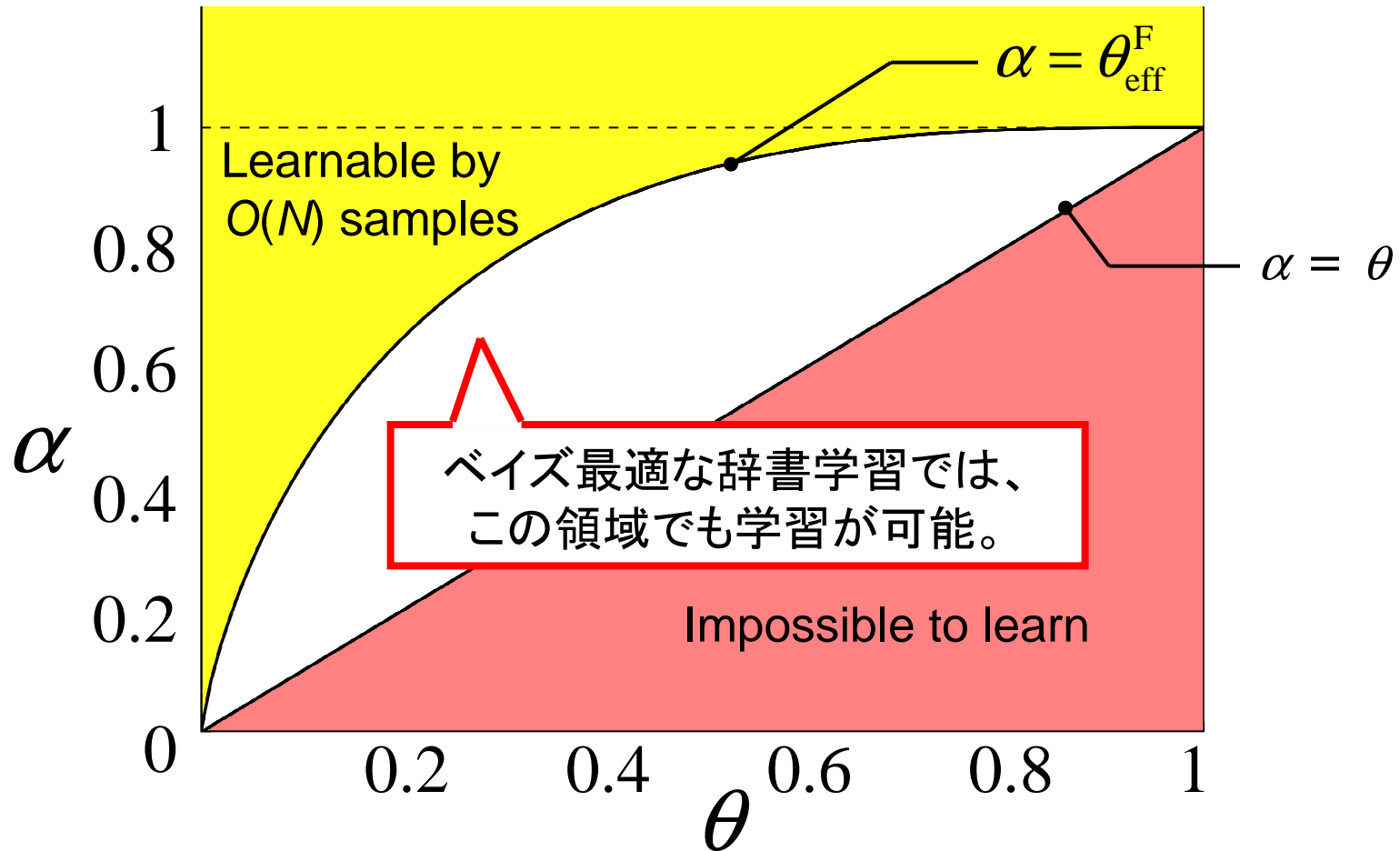
# $\alpha - \theta$ 平面の相図

$\alpha > \theta_{\text{eff}}^F$ では、 $(\alpha, \theta_{\text{eff}}^F)$ から決まる $\gamma_F$ を用いて  
 $P > N\gamma_F$ のときにplanted solutionを同定できる。



# $\alpha - \theta$ 平面の相図

$\alpha > \theta_{\text{eff}}^F$ では、 $(\alpha, \theta_{\text{eff}}^F)$ から決まる $\gamma_F$ を用いて  
 $P > N\gamma_F$ のときにplanted solutionを同定できる。



# ベイズ最適な学習則

- 平均自乗誤差(MSE)を定義

$$\text{MSE}_D = \frac{1}{MN} \sum_{\mu i} \left\langle (D_{\mu i}^0 - \hat{D}_{\mu i}(\mathbf{Y}))^2 \right\rangle_{\mathbf{D}^0, \mathbf{X}^0, \mathbf{Y}}$$

$$\text{MSE}_X = \frac{1}{NP} \sum_{il} \left\langle (X_{il}^0 - \hat{X}_{il}(\mathbf{Y}))^2 \right\rangle_{\mathbf{D}^0, \mathbf{X}^0, \mathbf{Y}}$$

- $\hat{\mathbf{D}}(\mathbf{Y}), \hat{\mathbf{X}}(\mathbf{Y})$  は任意の学習則を用いて $\mathbf{Y}$ から推定した解
- $\langle \dots \rangle$  は次の同時分布による平均

$$P(\mathbf{D}^0, \mathbf{X}^0, \mathbf{Y}) = \delta\left(\mathbf{Y} - \frac{\mathbf{D}^0 \mathbf{X}^0}{\sqrt{N}}\right) P_0(\mathbf{D}^0) P_0(\mathbf{X}^0; \rho)$$

# ベイズ最適な学習則

- MSE はベイズ最適な学習則により最小化される。

– この学習則による推定値は以下の通り。

$$\hat{\mathbf{D}}_i^{\text{ODL}}(\mathbf{Y}) = \frac{\sqrt{M} \langle \mathbf{D}_i \rangle}{\|\langle \mathbf{D}_i \rangle\|_2} \quad (i = 1, \dots, N), \quad \hat{\mathbf{X}}^{\text{ODL}}(\mathbf{Y}) = \langle \mathbf{X} \rangle$$

- $\mathbf{D}_i$  は  $\mathbf{D}$  の  $i$  番目のコラム
- $\langle \dots \rangle$  は事後分布  $P(\mathbf{D}, \mathbf{X} | \mathbf{Y}) = P(\mathbf{D}, \mathbf{X}, \mathbf{Y}) / P(\mathbf{Y})$  による  $\mathbf{D}$ ,  $\mathbf{X}$  平均.
  - つまり、推定の際にモデルの真の分布が分かっている。
  - このとき、レプリカ対称解は安定。[Y. Iba (1999)]

# ベイズ最適な学習則の解析

真の値と推定値の重なり  $m_D$ ,  $m_X$  を定義する

$$m_D = \frac{1}{MN} \sum_{\mu i} \left\langle D_{\mu i}^0 \hat{D}_{\mu i}^{\text{ODL}}(\mathbf{Y}) \right\rangle_{\mathbf{D}^0, \mathbf{X}^0, \mathbf{Y}}$$
$$m_X = \frac{1}{NP} \sum_{il} \left\langle X_{il}^0 \hat{X}_{il}^{\text{ODL}}(\mathbf{Y}) \right\rangle_{\mathbf{D}^0, \mathbf{X}^0, \mathbf{Y}}$$

$\text{MSE}_D = 2(1 - m_D)$ ,  $\text{MSE}_X = 2(\rho - m_X)$  となるので

- $m_D = 1, m_X = \rho$  :  $\mathbf{D}^0$  と  $\mathbf{X}^0$  の学習に成功。
- $m_D = 0, m_X = 0$  : 学習失敗。

レプリカ法により  $m_D, m_X$  のパラメータ依存性を明らかにする。

# ベイズ最適な学習則の解析

- レプリカ法による  $m_D$ ,  $m_X$  の表式

$$m_D = \frac{\hat{m}_D}{1 + \hat{m}_D}$$

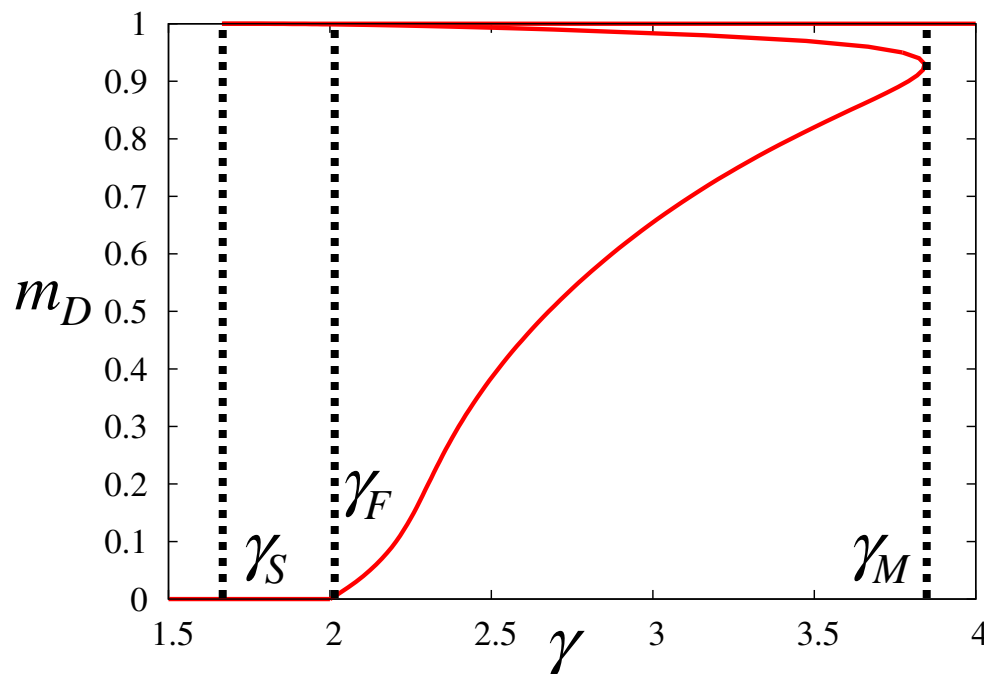
$$m_X = \int Dz dX^0 P_0(X^0) \left( \frac{\Xi_X^+ \sqrt{\hat{m}_X} z + \hat{m}_X X^0}{\Xi_X (1 + \hat{m}_X \sigma_X^2)} \right)^2$$

$$\hat{m}_X = \frac{\alpha m_D}{\rho \sigma_X^2 - m_D m_X}, \quad \hat{m}_D = \frac{\gamma m_X}{\rho \sigma_X^2 - m_D m_X}$$

$$\alpha = M / N$$
$$\gamma = P / N$$

$$\Xi_X^+ = \frac{\rho}{\sqrt{1 + \hat{m}_X \sigma_X^2}} \exp\left( \frac{\sigma_X^2 (\sqrt{\hat{m}_X} z + \hat{m}_X X^0)^2}{2(1 + \hat{m}_X \sigma_X^2)} \right), \quad \Xi_X = (1 - \rho) + \Xi_X^+$$

# $m_D$ の $\gamma$ 依存性 ( $\alpha = 0.5, \rho = 0.2$ )



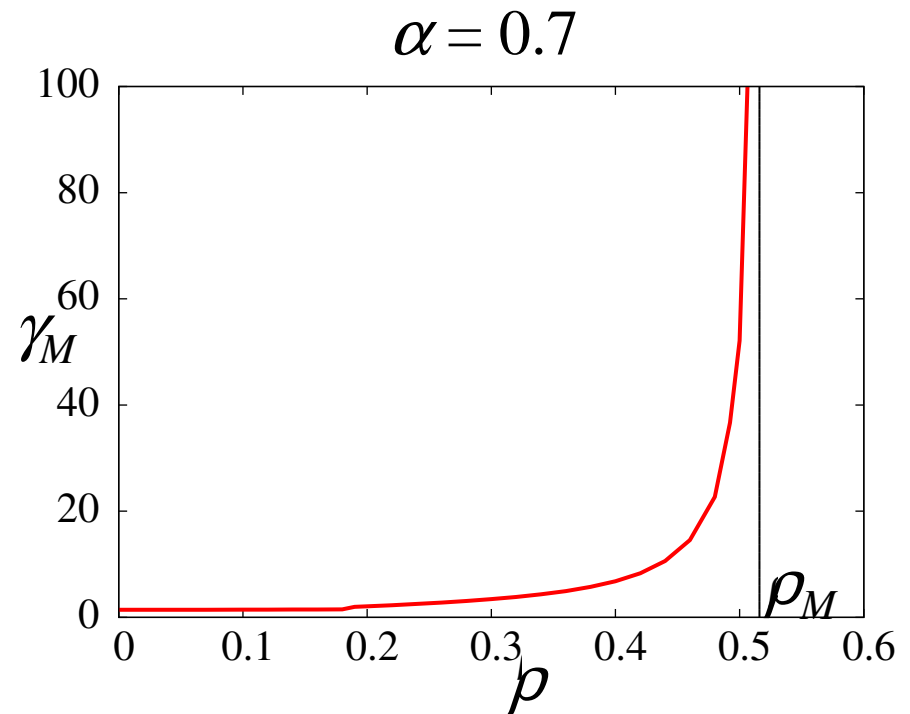
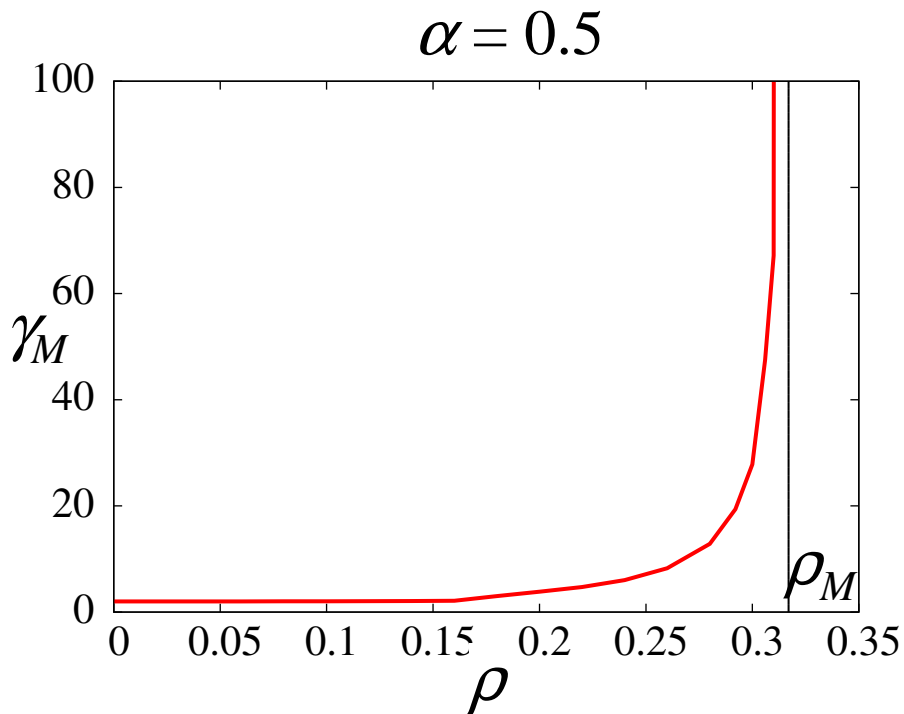
※  $\alpha$  = 圧縮率、  
 $\rho$  = 非ゼロ要素の割合

- $\gamma > \gamma_S = \alpha / (\alpha - \rho)$  で  $m_D = 1$ ,  $m_X = \rho$  の解が現れる。  
→ Sample complexity は  $P_c = N\gamma_S$ .
- $\gamma > \gamma_M$  で  $m_D = 1$ ,  $m_X = \rho$  以外の解が消える。  
しかし、 $\gamma_M$  は  $\rho \rightarrow \rho_M(\alpha)$  で発散する。



# $\gamma_M$ の $\alpha, \rho$ 依存性

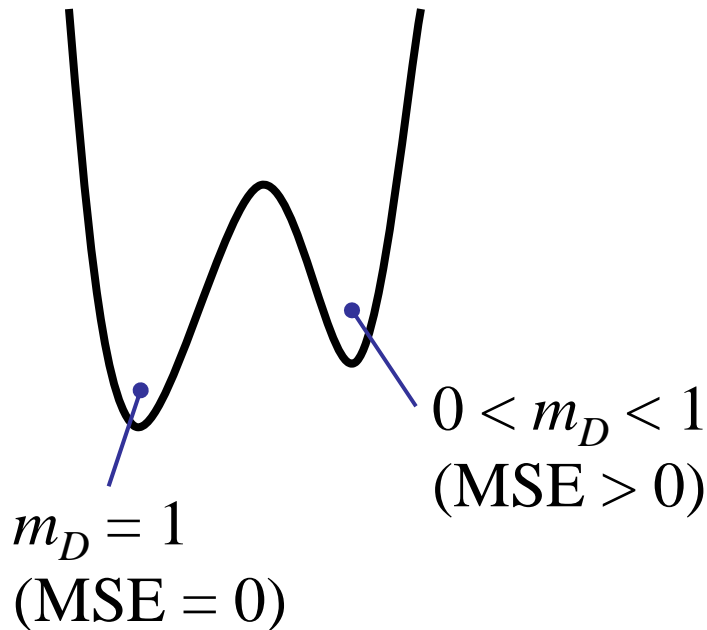
※  $\alpha$  = 圧縮率、 $\rho$  = 非ゼロ要素の割合



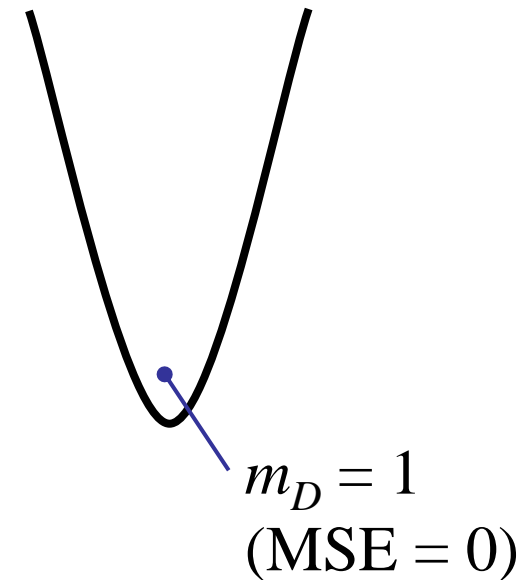
- $\alpha$  と  $\rho$  の差が広がるにつれて  $\gamma_M$  は増加し、 $\rho > \rho_M(\alpha)$  で発散する。

# $\gamma_M$ の意味

$$\gamma_S < \gamma < \gamma_M$$



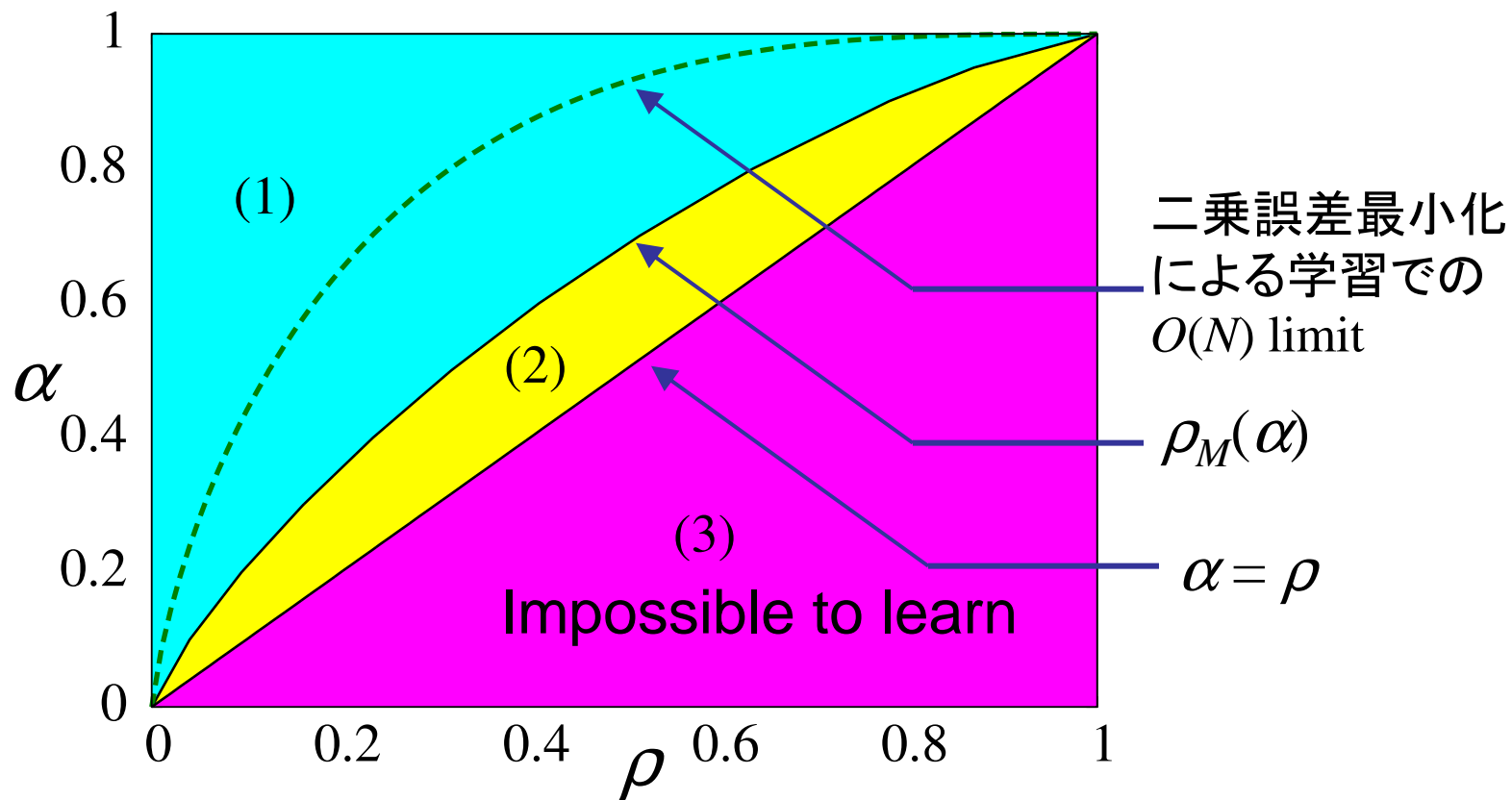
$$\gamma > \gamma_M$$



$\gamma > \gamma_M$  で  $m_D = 1$  が大域的安定解となる。

→  $\gamma_M$  が有限の時、局所探索アルゴリズムにより学習が可能。

# $\alpha - \rho$ 平面上の相図



- (1)+(2): サンプル複雑度は  $P_c = N\gamma_S$ ,  $\gamma_S = \alpha/(\alpha - \rho)$ .
- (1):  $\gamma_M$  が有限 ( $\gamma > \gamma_M$  では  $m_D = 1$  が大域的安定解となる)

# まとめ

辞書学習に対して、ベイズ最適な学習則を用いた場合のサンプル複雑度の解析を行った。

- 非ゼロ要素の割合  $\rho < \text{圧縮率 } \alpha$  のとき、  
原理的にはサンプル数  $P > P_c = \gamma_s N$  で辞書学習が達成可。
  - サンプル複雑度は  $O(N)$  である。
  - 先行研究の上界  $O(N^2)$  を改善した。
- 特に  $\rho < \rho_M(\alpha)$  の場合、  
学習成功状態が大域的な安定解となるため、  
多項式時間で解を得られる可能性が示唆された。

# Dictionary Learningのアルゴリズム

Dictionary learningにおける理論限界:  $P > P_c \sim O(N)$



理論限界を実現するアルゴリズムを構成したい。

## 必要なこと

- 既存アルゴリズムの性能評価
  - Method of Optimal Direction, K – SVD
- 新しいアルゴリズムの開発、改善
  - Belief propagation