第21回学習物理領域セミナー ＋ 第73回DLAP

行列分解問題へのベイズ統計学的アプローチ

茨城大学
Ibaraki University

MLPhYs

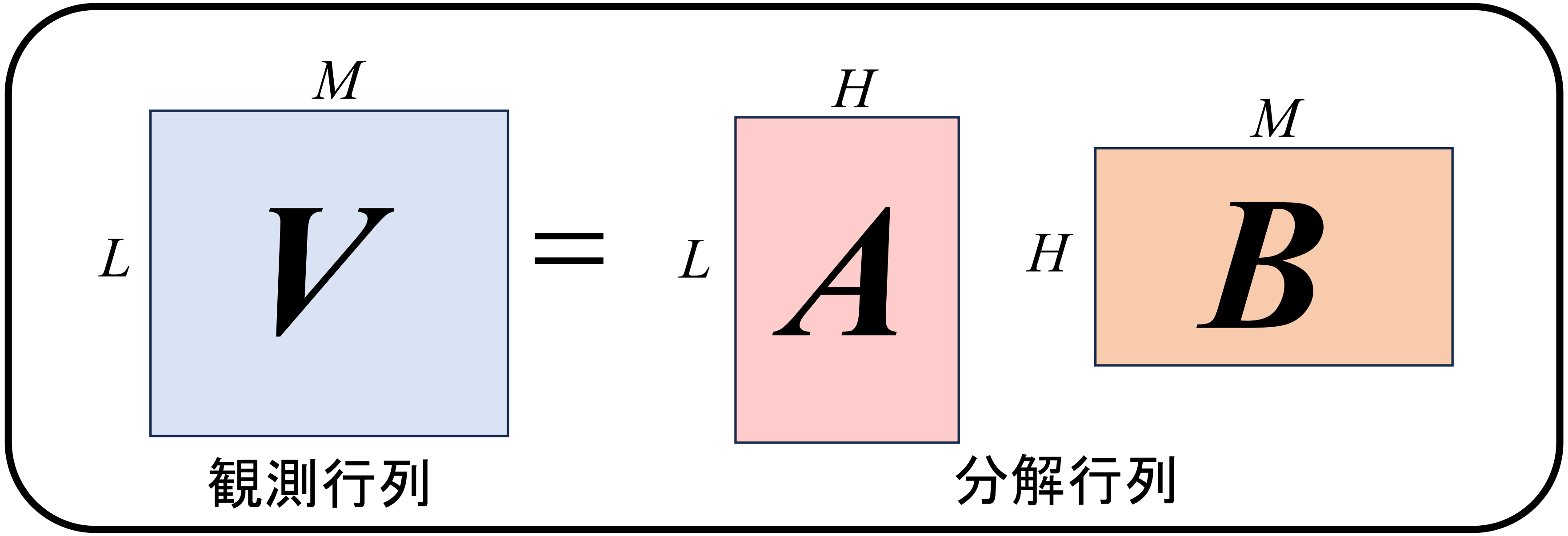
2025/11/20

茨城大学　竹田 晃人

目次

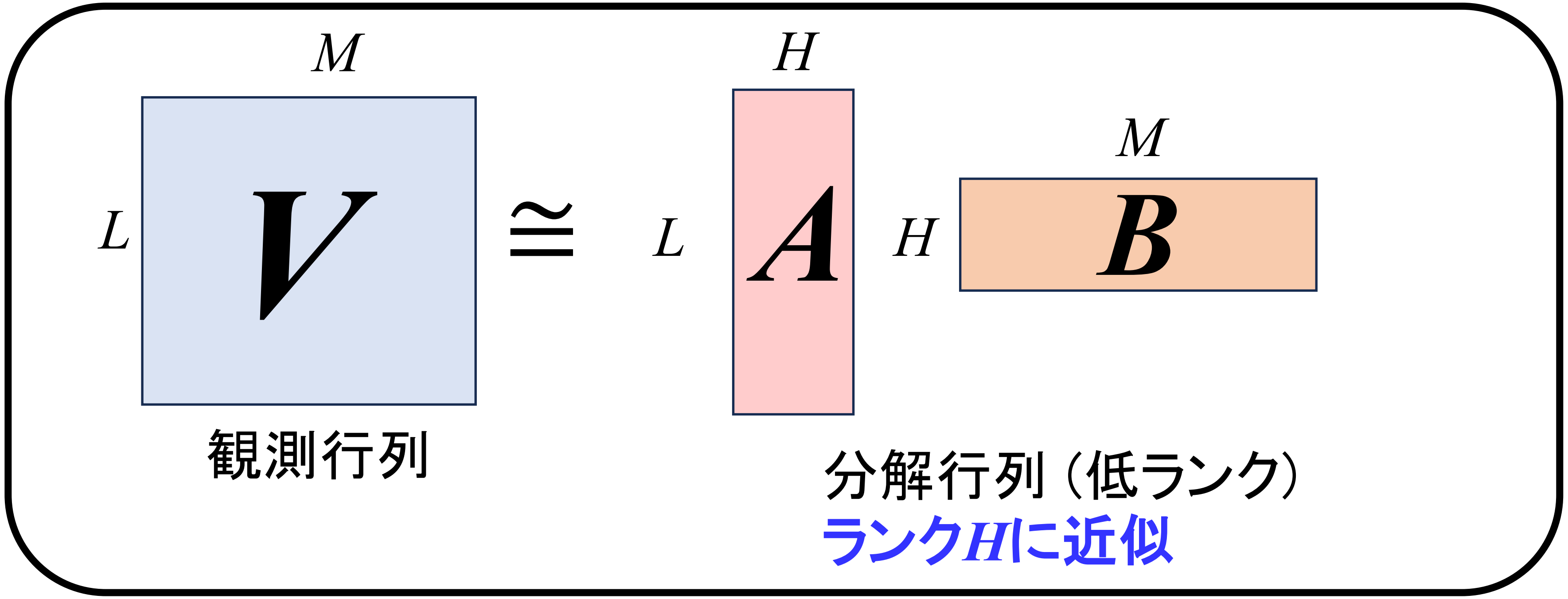
- 本研究の内容は，以下の方々との共同研究に基づきます．
  川澄亮太氏（群馬大学)，玉井智貴氏，遠藤優介氏（(元)茨城大学)
- セミナー当日に使用したファイルから一部（版権関係）を
  削除していますので御了承ください．また，細かい修正も行っています．

[partA] 行列分解問題



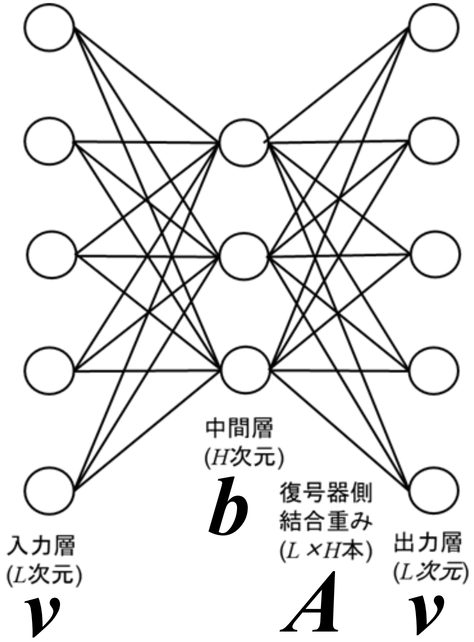
- 既知の観測行列Vから行列積を構成する未知の分解行列AとBを決定
- 適切な問題設定の下で分解行列は観測行列の特徴量を含む
  → 機械学習の問題で重要　　例えば 特異値分解（SVD）や主成分分析(PCA)も行列分解
- 機械学習の実問題でも頻出（例については後述）
- ニューラルネットワークとも関係

- **低ランク行列分解**
  - Hが小さい場合（Vから少数の特徴量を抽出）
    ※行列分解では一般に低ランク性を課すことが多い
  - 特異値分解（SVD）を使うことで低ランク近似が可能
  - 原情報の圧縮に利用可能（画像データの低次元圧縮等）



[partA] 低ランク行列分解とオートエンコーダ

オートエンコーダ（自己符号器）　Hinton-Salakhutdinov (2006)



- 入力情報と出力情報が一致するように学習するニューラルネットワーク
- 中間層次元 (H) < 入力層次元＝出力層次元 (L)
- 学習が進むと，中間層には入力情報の次元圧縮情報が格納

→ 次元削減器

$$\boldsymbol{v} = f(\boldsymbol{A}\boldsymbol{b})$$

$\boldsymbol{v}$	f	$\boldsymbol{A}$	$\boldsymbol{b}$
出力層（入力層）	非線形関数	復号器側結合重み	中間層

- $f(x) = x$（線形関数）
- 入力情報をM個使う

入力情報=出力情報(M個, L次元)
$\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_M$

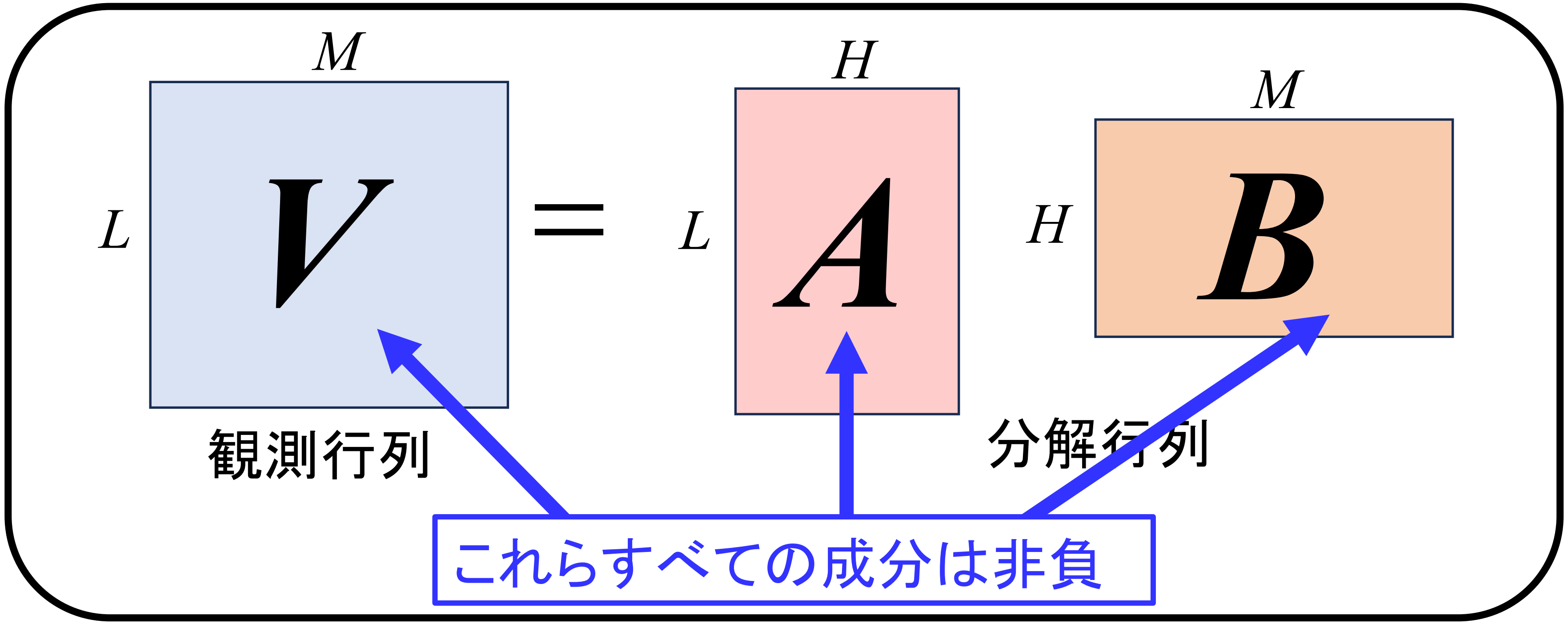
復号器側結合重み
$\boldsymbol{A}$

中間層の情報(M個, H次元)
$\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_M$

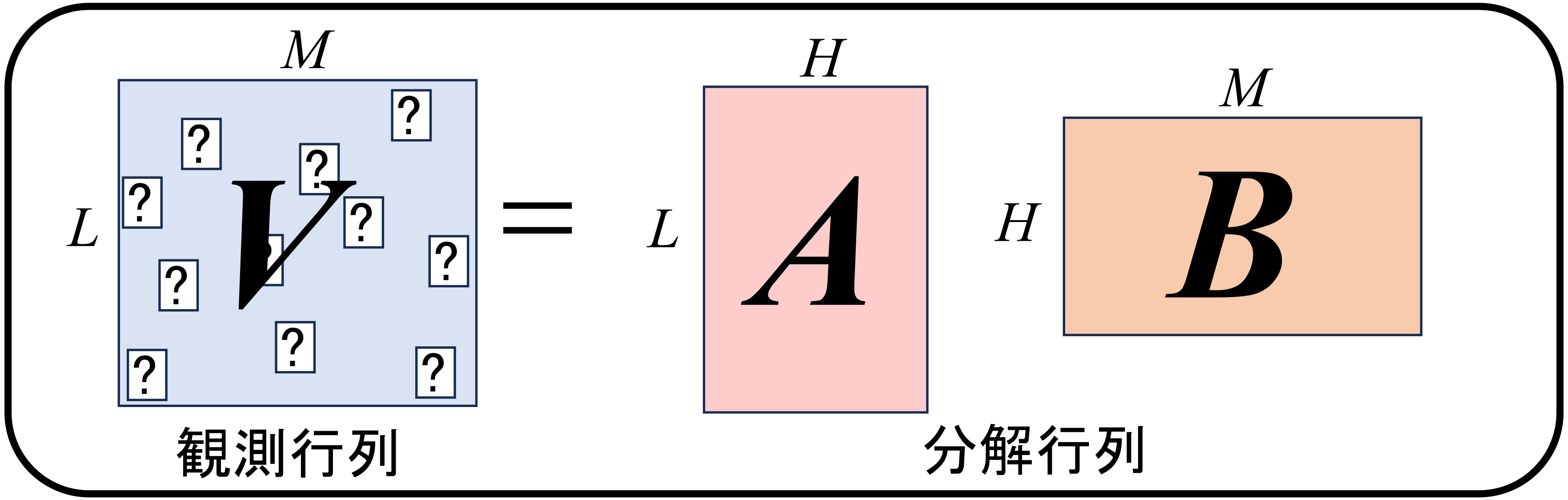
$$\underset{L \times M}{\boldsymbol{V}} \cong \underset{L \times H}{\boldsymbol{A}}\ \underset{H \times M}{\boldsymbol{B}}$$

低ランク行列分解はある種のオートエンコーダ

- **非負値行列分解**
  - $\boldsymbol{V}, \boldsymbol{A}, \boldsymbol{B}$の成分が非負
  - データが非負である場合（画素値，信号振幅，評価値…）の特徴抽出方法
  - $\boldsymbol{A}, \boldsymbol{B}$を交互に求めるアルゴリズム (Lee-Seung (2000) 等)
  - 音源分離，文書等のクラスタリング，バイオインフォマティックス等応用多数

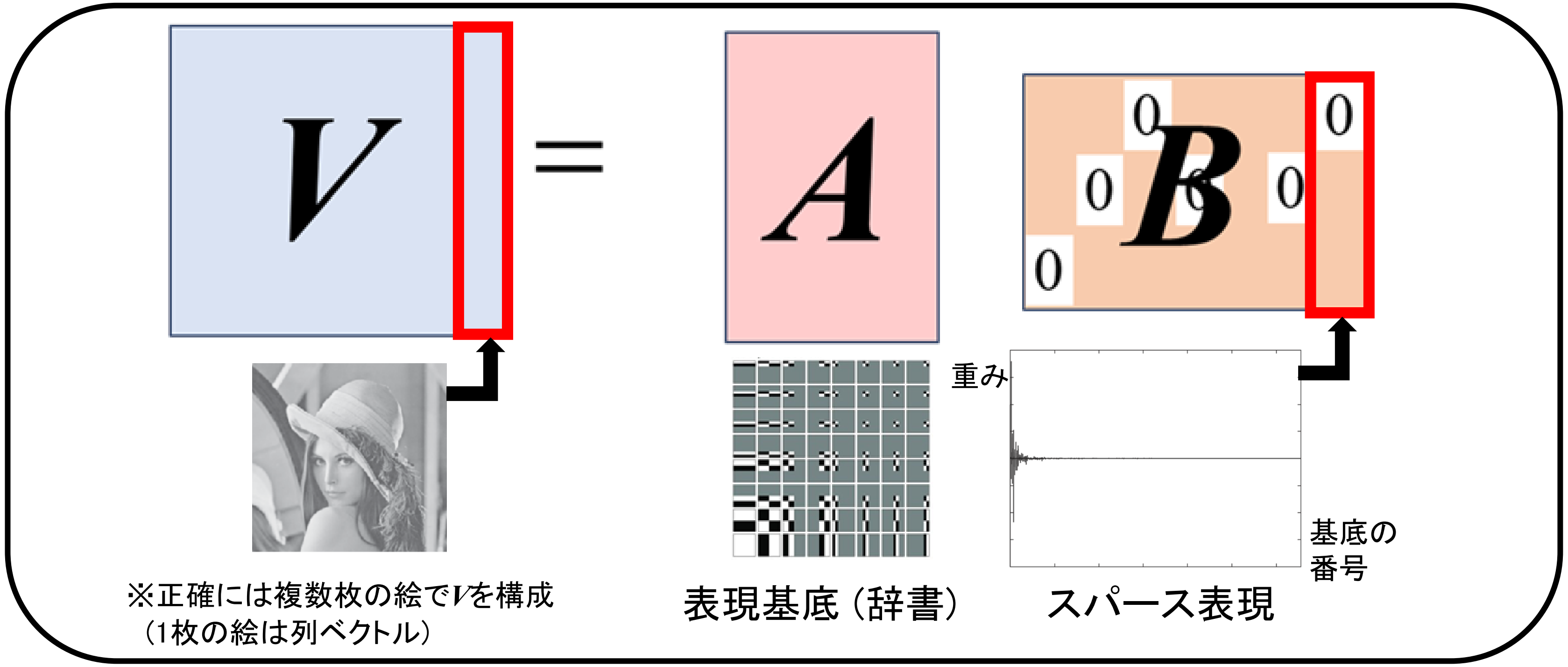


- **行列補完問題**
  - Vを全て観測できないが，Vの未観測成分が知りたい
  - Vの観測可能成分のみから分解行列 A,B を計算(推定)し，それを基にVの未観測成分を補完
  - 推薦システム(協調フィルタリング)，画像修復等に利用可能



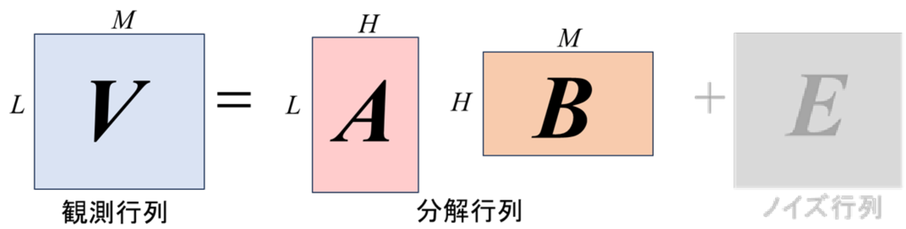
[partA] 行列分解問題の種類：スパース行列分解

- **スパース行列分解**
  - 片側の分解行列がスパース（ゼロ成分を多く含む）
  - 観測行列の潜在的なスパース表現(B)とその基底(A, 辞書)を求める
    スパース学習では重要，神経科学でも重要(スパースコーディング)
  - アルゴリズムは多数提案（例：MOD, K-SVD, sparsePCA）



[partB] ベイズ統計学の形式による行列分解

ベイズ推定の形式で分解行列の性質を調べる



分解行列解A, Bの事後分布　尤度(=V,A, Bの関係)　事前分布

$$P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V}) \propto P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B})$$

- 尤度でよく使われる設定：V にガウスノイズ行列Eが加わる

$$P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B}) \propto \exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{V}-\boldsymbol{A}\boldsymbol{B}\|^2_{\mathrm{FRO}}\right) \qquad \|\boldsymbol{X}\|_{\mathrm{FRO}} = \sqrt{\sum_{ij}(X_{ij})^2}$$

フロベニウスノルム

- 事前分布に関する設定：分解行列A, Bの性質を決定（簡単なのはガウス分布）

目的

- 分解行列A,Bの推定

← 事後分布の最大解（MAP解）を求める：難しい

← 周辺化事後分布を計算（着目成分以外を積分し除く）し各行列要素を推定：難しい

・上記のガウスノイズの尤度には, A,Bの4次項（4体相互作用)が入る

・事前分布が複雑な形だと, さらに周辺分布を求めるのが困難

$$P(A_{ij}\,|\,\boldsymbol{V}) = \int dA_{(ij\text{以外})}d\boldsymbol{B}P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V})$$

周辺化事後分布

- 行列分解問題の解の典型的な性質の解析

これらは簡単にはできない → 様々な解析技術が必要

ベイズ推定による行列分解の解析手法はいくつか知られている

(※ほぼ, ランダム系に対する統計力学的手法が由来)

(1) サンプリングによる手法 (マルコフ連鎖モンテカルロ等)

確率分布からのサンプリング

(2) レプリカトリック

自由エネルギーのランダム平均の手法

(3) 近似メッセージ伝搬法 (AMP)

Bethe近似法を発展させた局所物理量(周辺分布)の評価法

(4) Thouless–Anderson–Palmer (TAP) 方程式

元々はスピングラスの局所物理量・自由エネルギーの評価形式

(5) 変分ベイズ法

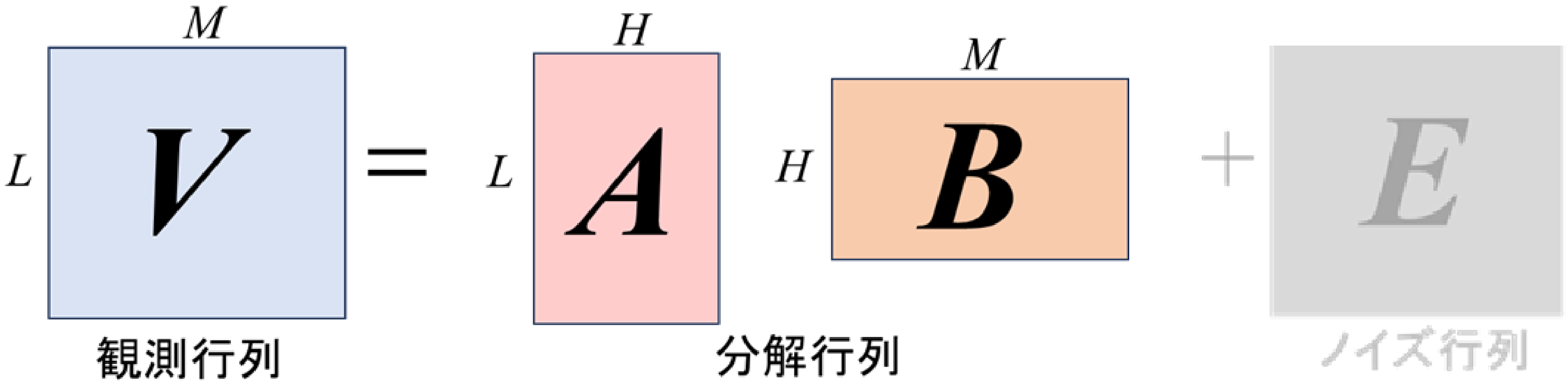
変分法による自由エネルギーの評価形式

[partB] (1) サンプリングによる行列分解

◆ **サンプリングによる手法** (マルコフ連鎖モンテカルロ等)

実際の分解行列解の求解を確率分布のサンプリングにより実施

(ベイズ推定による分解行列解の解析的表現を最後まで求めない)



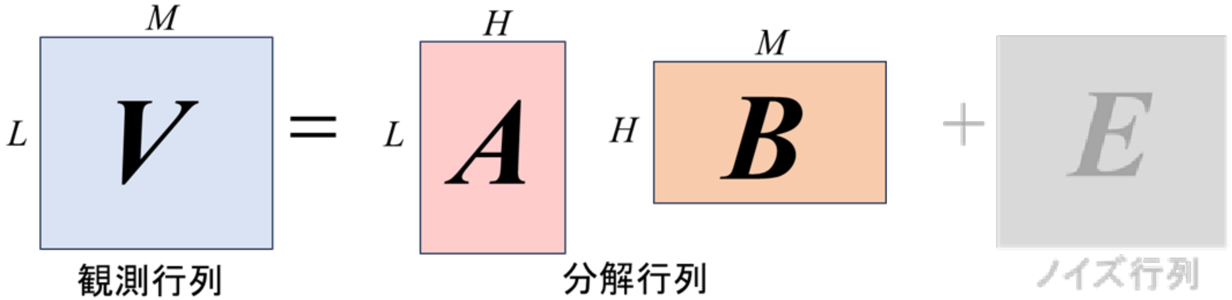
$$P(\color{red}{\boldsymbol{A},\boldsymbol{B}}\,|\,V) \propto P(V|A,\boldsymbol{B})P(A)P(\boldsymbol{B})$$

分解行列解$\color{red}{\boldsymbol{A},\boldsymbol{B}}$の事後分布　尤度(=$V$,$A$, $\boldsymbol{B}$の関係)　事前分布

左辺の事後分布に従う具体的な分解行列$\boldsymbol{A},\boldsymbol{B}$をサンプリングにより生成

[partB] (1) サンプリングによる行列分解

Salakhutdinov–Mnih (2008)



階層的ベイズモデル

$$P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V}) \propto P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A}\,|\,\mu_A,C_A)P(\boldsymbol{B}\,|\,\mu_B,C_B)$$

事後分布　尤度　事前分布（パラメータ含む）

- 尤度（ノイズ行列Eがガウス分布）:

$$P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B}) \propto \exp\left(-\frac{1}{\sigma^2}\|\boldsymbol{V}-\boldsymbol{AB}\|^2_{\mathrm{FRO}}\right)$$

- 事前分布（多次元ガウス分布，パラメータ付）

$$P(\boldsymbol{A}|\boldsymbol{\mu}_A,\boldsymbol{C}_A) = \prod^{N} \mathcal{N}(\boldsymbol{a}_i|\boldsymbol{\mu}_A,\boldsymbol{C}_A^{-1})$$

$$P(\boldsymbol{B}|\boldsymbol{\mu}_B,\boldsymbol{C}_B) = \prod_{i}^{N} \mathcal{N}(\boldsymbol{b}_i|\boldsymbol{\mu}_B,\boldsymbol{C}_B^{-1})$$

- 事前分布のパラメータ（平均と共分散行列）の分布（ガウス–ウイシャート分布）

$$P(\boldsymbol{\mu}_A,\boldsymbol{C}_A) = \mathcal{N}(\boldsymbol{\mu}_A|\boldsymbol{\mu}_0,(\beta_0\boldsymbol{C}_A)^{-1})\mathcal{W}(\boldsymbol{C}_A|\boldsymbol{W}_0,\nu_0)$$

$$P(\boldsymbol{\mu}_B,\boldsymbol{C}_B) = \mathcal{N}(\boldsymbol{\mu}_B|\boldsymbol{\mu}_0,(\beta_0\boldsymbol{C}_B)^{-1})\mathcal{W}(\boldsymbol{C}_B|\boldsymbol{W}_0,\nu_0)$$

$$\boldsymbol{\mu}_A,\boldsymbol{C}_A \longrightarrow \boldsymbol{\mu}_B,\boldsymbol{C}_B \longrightarrow \boldsymbol{A} \longrightarrow \boldsymbol{B}$$

の順でサンプルリングを繰り返す → 分解行列解
（マルコフ連鎖モンテカルロ法など）

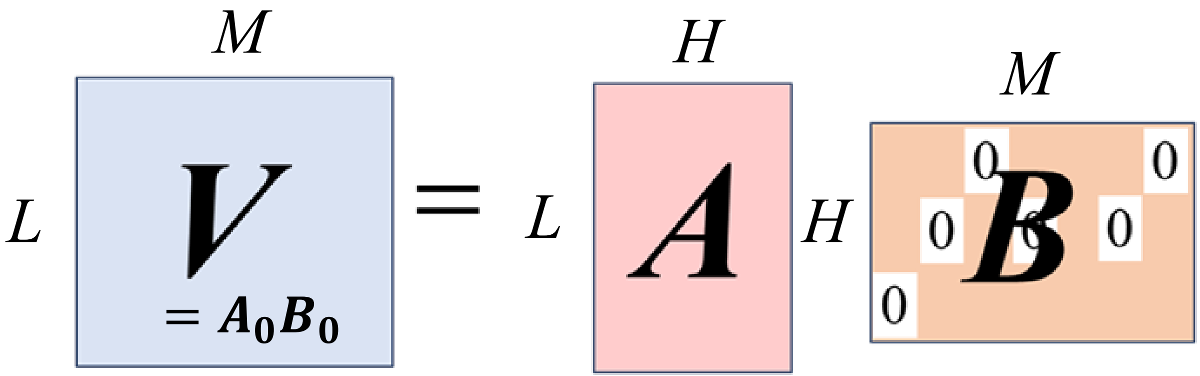
応用例（手法のテスト）：Netflixの映画レーティング予測（行列補完問題）

◆レプリカトリック

行列分解問題を自由エネルギー形式に落とし,
ランダム平均をレプリカトリックにより実行し，分解行列解の典型的な性質を評価

Sakata-Kabashima (2013)

スパース行列分解の解の典型的な性能を調べる



以下の最適化問題の求解でスパース行列分解を実施した場合の典型性能解析
(用意した分解行列解(埋め込み解) $\boldsymbol{A}_0, \boldsymbol{B}_0$ が実際に求めた解と一致するか？)

$$\min_{\boldsymbol{A},\boldsymbol{B}} \|\boldsymbol{V}-\boldsymbol{AB}\|_{\mathrm{FRO}}^2 \quad \text{subj. to} \quad \|\boldsymbol{A}\|_{\mathrm{FRO}}^2 = LH \quad \|\boldsymbol{B}\|_0 = HM\theta$$

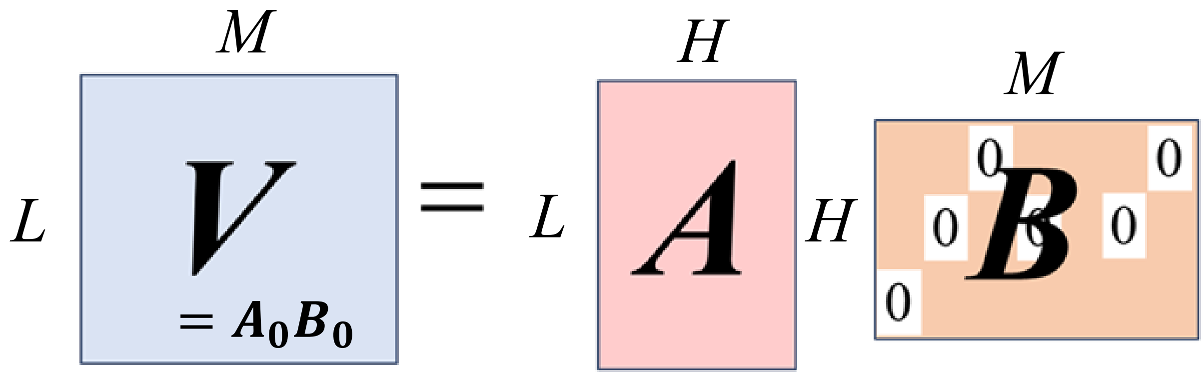
$\|\boldsymbol{B}\|_0 = \sum_{ij} \mathbf{1}(B_{ij} \neq 0)$ 行列 ℓ_0ノルム
(非ゼロ成分数)

非ゼロ成分
の割合

[partB] (2) レプリカトリック

ベイズによる定式化と分配関数

$$\min_{\boldsymbol{A},\boldsymbol{B}} \| \boldsymbol{A}_0\boldsymbol{B}_0 - \boldsymbol{A}\boldsymbol{B} \|_{\mathrm{FRO}}^2 \quad \text{subj. to} \quad \| \boldsymbol{A} \|_{\mathrm{FRO}}^2 = LH \quad \| \boldsymbol{B} \|_0 = HM\theta$$



$$P(\boldsymbol{A},\boldsymbol{B} \mid \boldsymbol{A}_0,\boldsymbol{B}_0) \propto P(\boldsymbol{A}_0,\boldsymbol{B}_0 \mid \boldsymbol{A},\boldsymbol{B}) P(\boldsymbol{A}) P(\boldsymbol{B})$$

事後分布　　　　　　尤度　　　　　　事前分布

- 尤度（ボルツマン分布として設定，ゼロ温度で最適化問題）

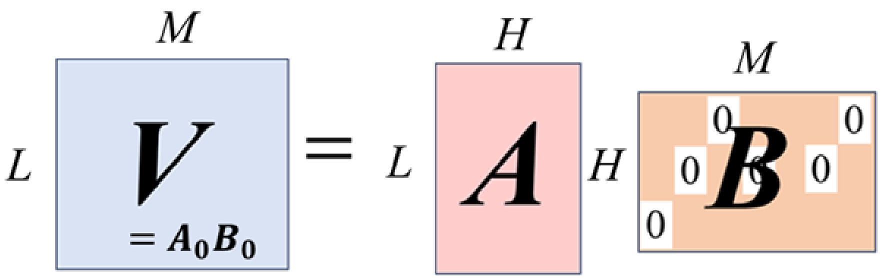
$$P(\boldsymbol{A},\boldsymbol{B} \mid \boldsymbol{A}_0,\boldsymbol{B}_0) \propto \exp\left(-\frac{\beta}{2H} \| \boldsymbol{A}\boldsymbol{B} - \boldsymbol{A}_0\boldsymbol{B}_0 \|_{\mathrm{FRO}}^2 \right)$$

- 事前分布（$\boldsymbol{A}$は正規化条件，$\boldsymbol{B}$は非ゼロ成分数を固定する制約）

$$P(\boldsymbol{A}) = \delta\left(\| \boldsymbol{A} \|_{\mathrm{FRO}}^2 - LH \right) \qquad P(\boldsymbol{B}) = \delta\left(\| \boldsymbol{B} \|_0 - HM\theta \right)$$

[partB] (2) レプリカトリック

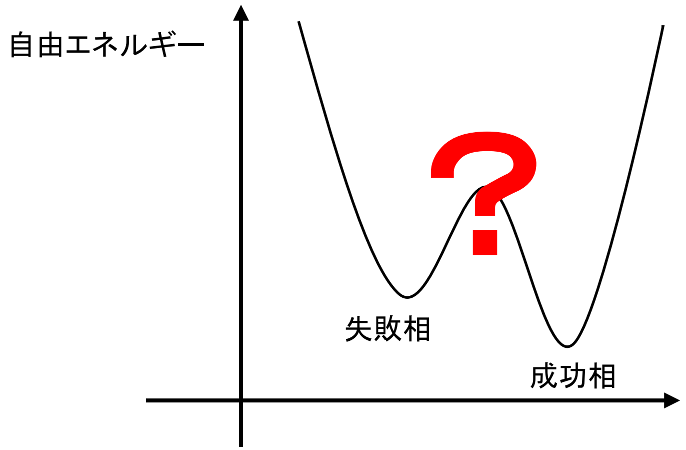
- 分配関数 (ベイズ統計学としては周辺尤度)



$$
\begin{aligned}
Z &= P(\boldsymbol{A}_0, \boldsymbol{B}_0) \\
&= \int d\boldsymbol{A} \int d\boldsymbol{B}\ P(\boldsymbol{V}|\boldsymbol{A}, \boldsymbol{B}) P(\boldsymbol{A}) P(\boldsymbol{B}) \\
&= \int d\boldsymbol{A} \int d\boldsymbol{B} \exp\left(-\frac{\beta}{2H}\|\boldsymbol{A}\boldsymbol{B} - \boldsymbol{A}_0\boldsymbol{B}_0\|_{\mathrm{FRO}}^2\right) \delta(\|\boldsymbol{A}\|_{\mathrm{FRO}}^2 - LH)\delta(\|\boldsymbol{B}\|_0 - HM\theta)
\end{aligned}
$$

- 最適化の解の典型的性質 → 復元成功相と失敗相の自由エネルギー比較
  ※自由エネルギーの正解行列 $\boldsymbol{A}_0, \boldsymbol{B}_0$ での平均操作の際にレプリカトリックが必要

- 自由エネルギーの解析から…
  埋め込み解を取り出せる条件が議論可能
  例: 次元 M を増加 → 埋め込み解が取り出せる M の閾値が存在
  (成功相が失敗相より自由エネルギーが低くなる)

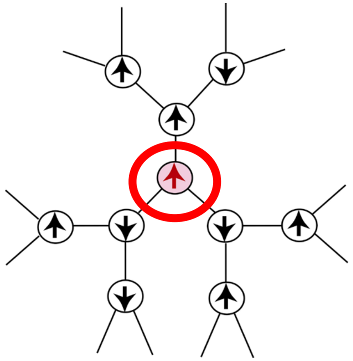


◆ 近似メッセージ伝搬法 (AMP)

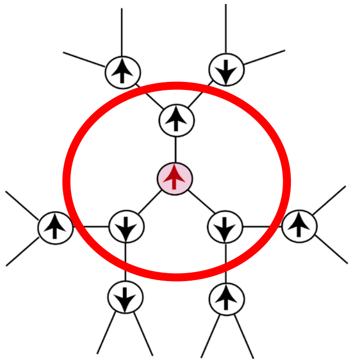
ベイズ行列分解のアルゴリズム（および性能評価法）をメッセージ伝搬法に基づき構成

> 分解行列A,Bの特定の行列要素の周辺分布を計算し，行列要素を推定
> （物理的には1体問題化のための計算手法）

・**平均場近似** ＝ 着目している変数以外を外場（平均場）として近似
・**Bethe近似** ＝ 近接サイトの影響まで厳密に計算し，それ以外の影響は外場(cavity場)
→ 相互作用がスパースな系で特に有効な近似
・**メッセージ伝搬法** ＝ Bethe近似に基づき，周辺分布（物理的には局所磁化等）を求めるアルゴリズム
・**近似メッセージ伝搬法 (AMP)**
＝ 密な相互作用を持つ系で，ガウス近似を利用しメッセージ伝搬法を精密化・高速化したもの



平均場近似
着目変数以外は
外場(平均場)
として表現

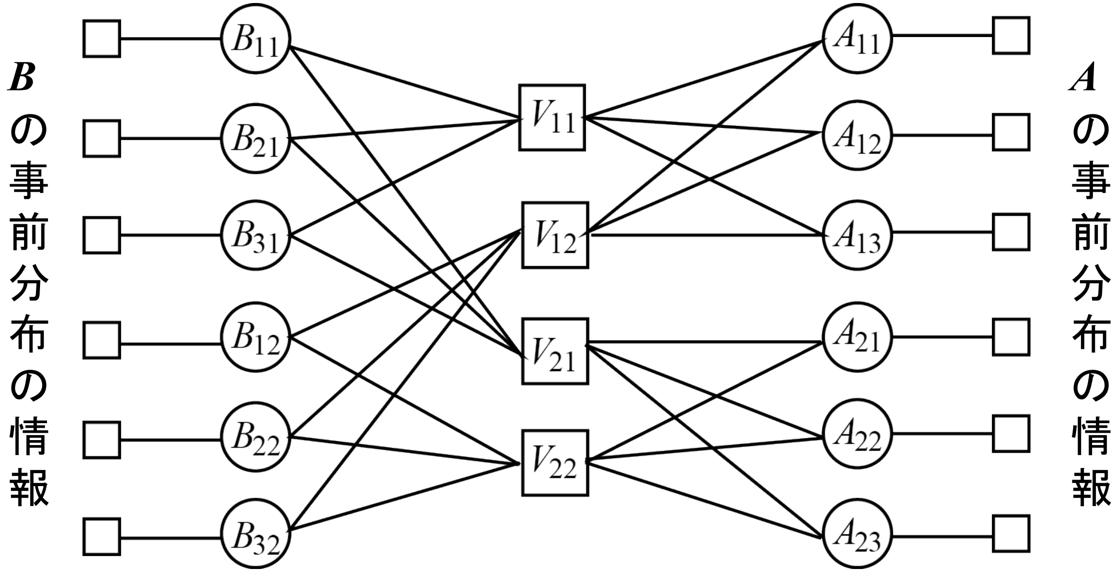


Bethe近似
着目変数の近接サイトまでは厳密
それ以外は外場（cavity場）

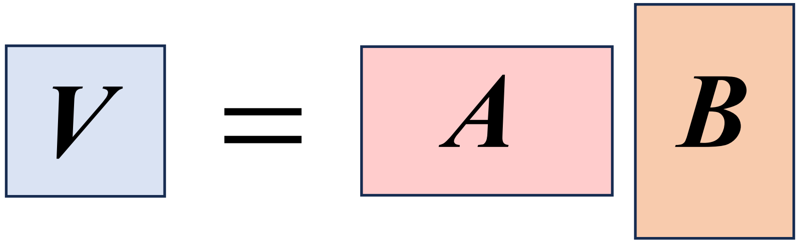
[partB] (3) 近似メッセージ伝搬法 (AMP)

Kabashima–Krzakala–Mézard–Sakata–Zdeborová (2016)

一般の行列分解の問題について，AMP (正確にはgeneralized AMP) で行列を分解する手法を提案



行列分解を表すファクターグラフ，このグラフ上をメッセージ伝搬



$$
\begin{aligned}
&\textbf{Vの分散} & V_{\mu l}^t &= \frac{1}{N}\sum_j \left[c_{jl}(t)s_{\mu j}(t) + c_{jl}(t)\hat{f}_{\mu j}^2(t) + \hat{x}_{jl}^2(t)s_{\mu j}(t)\right] \\
&\textbf{Vの期待値} & \omega_{\mu l}^t &= \frac{1}{\sqrt{N}}\sum_j \hat{x}_{jl}(t)\hat{f}_{\mu j}(t) - g_{\mathrm{out}}\left(\omega_{\mu l}^{t-1}, y_{\mu l}, V_{\mu l}^{t-1}\right) \\
& & &\quad \times \frac{1}{N}\sum_j \left[\hat{f}_{\mu j}(t)\hat{f}_{\mu j}(t-1)c_{jl}(t) + \hat{x}_{jl}(t)\hat{x}_{jl}(t-1)s_{\mu j}(t)\right] \\
&\textbf{Bの分散} & \left(\Sigma_{il}^t\right)^{-1} &= \frac{1}{N}\sum_\mu \left\{-\partial_\omega g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\left[\hat{f}_{\mu i}^2(t) + s_{\mu i}(t)\right] - g_{\mathrm{out}}^2\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)s_{\mu i}(t)\right\} \\
&\textbf{Bの期待値} & T_{il}^t &= \Sigma_{il}^t\left\{\frac{1}{\sqrt{N}}\sum_\mu g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\hat{f}_{\mu i}(t) - \hat{x}_{il}(t)\frac{1}{N}\sum_\mu \hat{f}_{\mu i}^2(t)\hat{\partial}_\omega g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\right. \\
& & &\quad \left. -\hat{x}_{il}(t-1)\frac{1}{N}\sum_\mu s_{\mu i}(t)g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)g_{\mathrm{out}}\left(\omega_{\mu l}^{t-1}, y_{\mu l}, V_{\mu l}^{t-1}\right)\right\} \\
&\textbf{Aの分散} & \left(Z_{\mu i}^t\right)^{-1} &= \frac{1}{N}\sum_l \left\{-\partial_\omega g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\left[\hat{x}_{il}^2(t) + c_{il}(t)\right] - g_{\mathrm{out}}^2\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)c_{il}(t)\right\} \\
&\textbf{Aの期待値} & W_{\mu i}^t &= Z_{il}^t\left\{\frac{1}{\sqrt{N}}\sum_l g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\hat{x}_{il}(t) - \hat{f}_{\mu i}(t)\frac{1}{N}\sum_l \hat{x}_{il}^2(t)\hat{\partial}_{og} g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)\right. \\
& & &\quad \left. -\hat{f}_{\mu i}(t-1)\frac{1}{N}\sum_l c_{il}(t)g_{\mathrm{out}}\left(\omega_{\mu l}^t, y_{\mu l}, V_{\mu l}^t\right)g_{\mathrm{out}}\left(\omega_{\mu l}^{t-1}, y_{\mu l}, V_{\mu l}^{t-1}\right)\right\} \\
& & \hat{x}_{il}(t+1) &= f_X\left(\Sigma_{il}^t, T_{il}^t\right), c_{il}(t+1) = f_c\left(\Sigma_{il}^t, T_{il}^t\right), \\
& & \hat{f}_{\mu i}(t+1) &= f_F\left(Z_{\mu i}^t, W_{\mu i}^t\right), s_{\mu i}(t+1) = f_s\left(Z_{\mu i}^t, W_{i\mu}^t\right)
\end{aligned}
$$

(細かい記号の定義は原論文を参照)

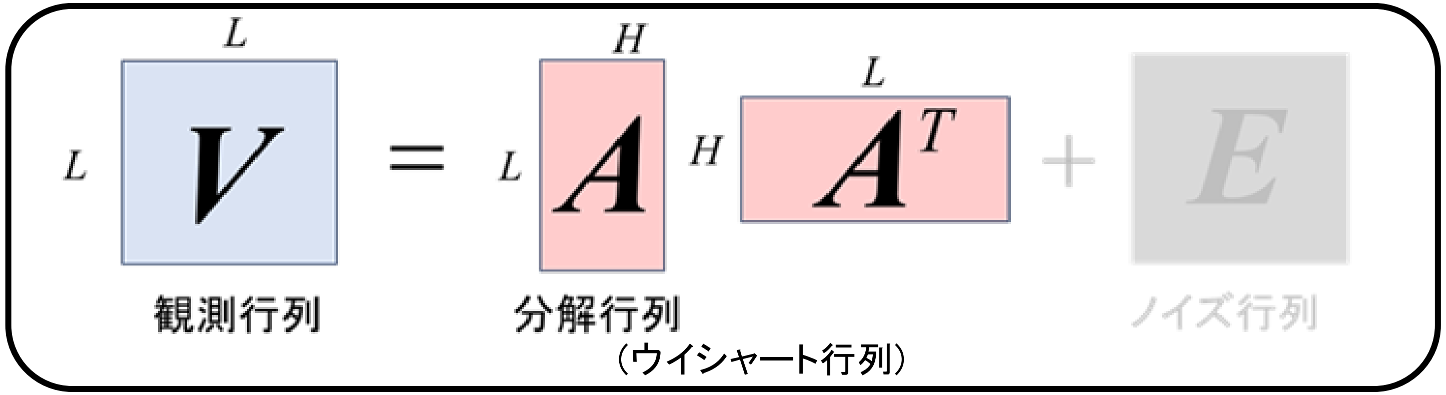
- AMPに状態発展法(State Evolution, SE)を組み合わせることで，行列分解問題の典型性能解析が可能
  → スパース行列分解，行列補完問題… 等様々な行列分解問題が解析できる

◆ Thouless-Anderson-Palmer(TAP)方程式

・スピングラスのThouless-Anderson-Palmer(TAP)形式(局所磁化の記述形式)による行列分解の解析
- ・AMPによる行列分解解析手法には問題が指摘 (厳密ではなく近似)
- ・TAP自由エネルギーをPlefka-Georges-Yedidia(PGY)展開 = 相互作用係数展開で評価
  → AMPは展開の2次までの近似，一般には3次以上も考慮する必要あり

※この方法で独立成分分析(ICA)を解析中 (Endo-Takeda, STATPHYS29)

Maillard-Krzakala-Mézard-Zdeborová (2022)
$B = A^T$となる対称行列 (以下のAがガウス型ランダム行列→ $A\,A^T$がウイシャート行列) の行列分解問題



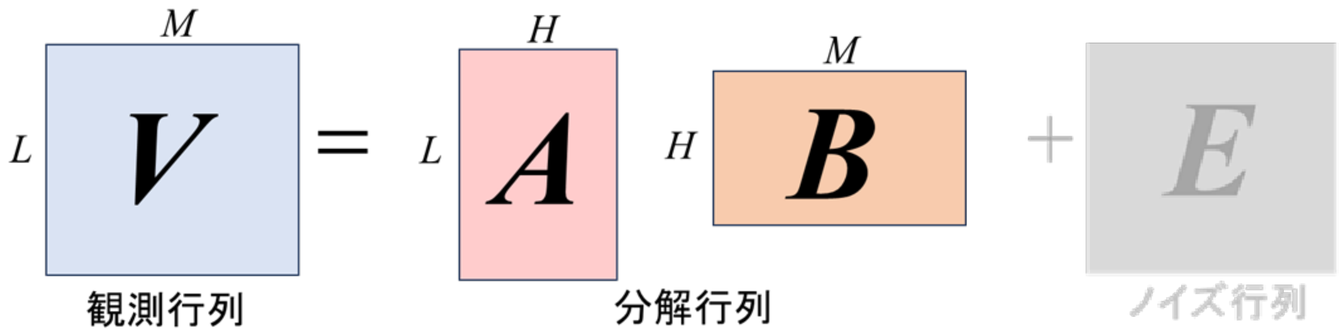
ウイシャート行列のノイズ除去問題を解く
(厳密解が別な方法で得られる → Harish-Chandra-Itzykson-Zuber行列積分とブラウン運動の議論)
→ PGYの2次近似 (AMP) は厳密解から少しずれている

◆ 変分ベイズ法

変分自由エネルギー形式により，分解行列解の統計的性質を変分法で評価

Ilin-Raiko (2010), Nakajima-Sugiyama (2011)

解析の流れ:



$$P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V}) \propto P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B})$$

分解行列解$\boldsymbol{A},\boldsymbol{B}$の事後分布　尤度(=$\boldsymbol{V},\boldsymbol{A}$, $\boldsymbol{B}$の関係)　事前分布

⬆

事後分布をそのままの形で $\boldsymbol{A},\boldsymbol{B}$ の推定に使うのは面倒

→事後分布 $P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V})$ を推定しやすい形に近似する

$$P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V}) \propto P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B})$$

Kullback–Leibler情報量を利用（確率分布間の距離）
（物理的には自由エネルギーに対応）

$$\mathrm{KL}(q(x)\,||\,p(x)) = \int dx q(x) \log \frac{q(x)}{p(x)}$$

$p(x),\ q(x)$: 確率分布

事後分布 $P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V})$ をよく近似する変分試行関数 $r(\boldsymbol{A},\boldsymbol{B})$ を求めたい
→ 次の関数に関する最小化問題を解く必要がある

$$\underset{r(\boldsymbol{A},\boldsymbol{B})}{\arg\min}\, \mathrm{KL}(r(\boldsymbol{A},\boldsymbol{B})\,||\,P(\boldsymbol{A},\boldsymbol{B}\,|\,\boldsymbol{V}))$$

$$\rightarrow \underset{r(\boldsymbol{A}),r(\boldsymbol{B})}{\arg\min}\, \mathrm{KL}(r(\boldsymbol{A})r(\boldsymbol{B})\,||\,P(\boldsymbol{V}\,|\,\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B}))$$

$\boldsymbol{A},\boldsymbol{B}$ の試行関数は独立とする（近似している）

[partB] (5) 変分ベイズ法

$$\mathrm{KL}(r(\boldsymbol{A})r(\boldsymbol{B}) \,||\, P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B})\,)$$

$$= \int d\boldsymbol{A}d\boldsymbol{B}\; r(\boldsymbol{A})r(\boldsymbol{B})\log\frac{r(\boldsymbol{A})r(\boldsymbol{B})}{P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})P(\boldsymbol{A})P(\boldsymbol{B})}$$

変分関数の停留条件を求めることで，試行関数が満たす関係式を書く

$$\mathrm{KL}((r(\boldsymbol{A}) + \delta r(\boldsymbol{A}))r(\boldsymbol{B})) - \mathrm{KL}(r(\boldsymbol{A})r(\boldsymbol{B})) = 0$$

$$\Longrightarrow \quad r(\boldsymbol{A}) = P(\boldsymbol{A})\exp\left(\int d\boldsymbol{B}\, r(\boldsymbol{B})\log P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})\right)$$

$$\mathrm{KL}(r(\boldsymbol{A})(r(\boldsymbol{B}) + \delta r(\boldsymbol{B}))) - \mathrm{KL}(r(\boldsymbol{A})r(\boldsymbol{B})) = 0$$

$$\Longrightarrow \quad r(\boldsymbol{B}) = P(\boldsymbol{B})\exp\left(\int d\boldsymbol{A}\, r(\boldsymbol{A})\log P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})\right)$$

ここまでは一般的な尤度や事前分布
指数関数の中に試行関数を含む(行列要素の)多重積分が残る

[partB] (5) 変分ベイズ法

以降は尤度と事前分布を以下のように仮定:

$$P(V|A,B)=\frac{1}{\sqrt{(2\pi\sigma^2)^{LH}}}\exp\left(-\frac{1}{2\sigma^2}\|V-AB\|_{\mathrm{FRO}}^2\right)$$

ノイズはガウス分布

$$P(A)=\frac{1}{\sqrt{(2\pi)^{LH}}}\exp\left(-\frac{1}{2}\mathrm{Tr}AA^T\right)=\prod_l\prod_h\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}a_{lh}^2\right)$$

$$P(B)=\frac{1}{\sqrt{(2\pi)^{MH}}}\exp\left(-\frac{1}{2}\mathrm{Tr}BB^T\right)=\prod_m\prod_h\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}b_{mh}^2\right)$$

A,Bはガウス事前分布

前述の試行関数の条件にこれらを代入→各試行関数が多次元ガウス分布

$$r(A)=P(A)\exp\left(\int dB\, r(B)\log P(V|A,B)\right) \Longrightarrow r(A)\propto\exp\left(-\frac{1}{2}\mathrm{Tr}(A-\hat{A})\hat{\Sigma}_A^{-1}(A-\hat{A})^T\right)$$ の形

$$r(B)=P(B)\exp\left(\int dA\, r(A)\log P(V|A,B)\right) \Longrightarrow r(B)\propto\exp\left(-\frac{1}{2}\mathrm{Tr}(B-\hat{B})\hat{\Sigma}_B^{-1}(B-\hat{B})^T\right)$$ の形

[partB] (5) 変分ベイズ法

試行関数の事後分布の平均と共分散行列 $\hat{A}, \hat{B}, \hat{\Sigma}_A, \hat{\Sigma}_B$ が満たす関係式を試行関数の関係式より導出

$$P(V|A,B) = \frac{1}{\sqrt{(2\pi\sigma^2)^{LH}}} \exp\left(-\frac{1}{2\sigma^2}\|V - AB\|_{\mathrm{FRO}}^2\right)$$

$$r(A) \propto \exp\left(-\frac{1}{2}\mathrm{Tr}(A-\hat{A})\hat{\Sigma}_A^{-1}(A-\hat{A})^T\right)$$

$$r(B) \propto \exp\left(-\frac{1}{2}\mathrm{Tr}(B-\hat{B})\hat{\Sigma}_B^{-1}(B-\hat{B})^T\right)$$

→

$$r(A) = P(A)\exp\left(\int dB\, r(B) \log P(V|A,B)\right)$$

$$r(B) = P(B)\exp\left(\int dA\, r(A) \log P(V|A,B)\right)$$

↓

Aの事後平均 $\hat{A} = \dfrac{1}{\sigma^2} V\hat{B}\hat{\Sigma}_A$

Bの事後平均 $\hat{B} = \dfrac{1}{\sigma^2} \hat{\Sigma}_B \hat{A}^T V$

Aの事後共分散 $\hat{\Sigma}_A = \sigma^2\left(\hat{B}\hat{B}^T + M\hat{\Sigma}_B + \sigma^2 I_H\right)^{-1}$

Bの事後共分散 $\hat{\Sigma}_B = \sigma^2\left(\hat{A}^T\hat{A} + L\hat{\Sigma}_A + \sigma^2 I_H\right)^{-1}$

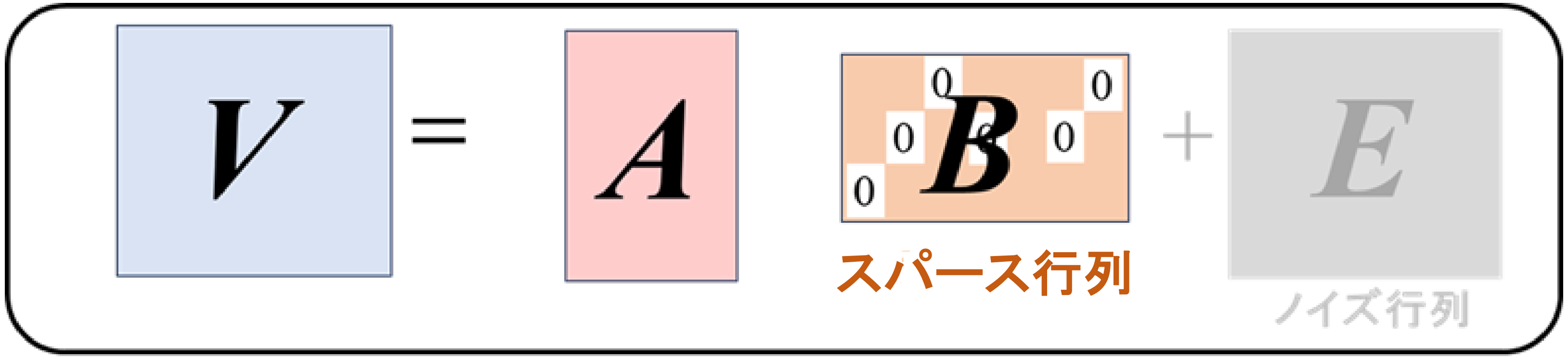
$\hat{A}, \hat{B}, \hat{\Sigma}_A, \hat{\Sigma}_B$について閉じた式 → 解くことで試行関数が求まり，行列分解が可能

応用例：変分ベイズ解を利用し，ノイズあり観測行列のランク推定の精度を特異値により議論

Nakajima-Tomioka-Sugiyama-Babacan (2015)

[partC] スパース行列分解の変分ベイズ推定

変分ベイズ法の形式で，スパース行列分解の性質を調べる



- 分解行列の事前分布が両方ともガウス分布（スパース行列分解ではない）
  前述の通り，変分ベイズ法を利用して分解行列が推定可能

- 分解行列の事前分布がガウス分布とラプラス分布

→ この場合はスパース行列分解問題になる

Kawasumi-KT (2018, 2023)　近似の下で，分解行列解の表現を書き下せる

→ スパース行列分解アルゴリズムとして使える

[partC] ラプラス事前分布と正則化

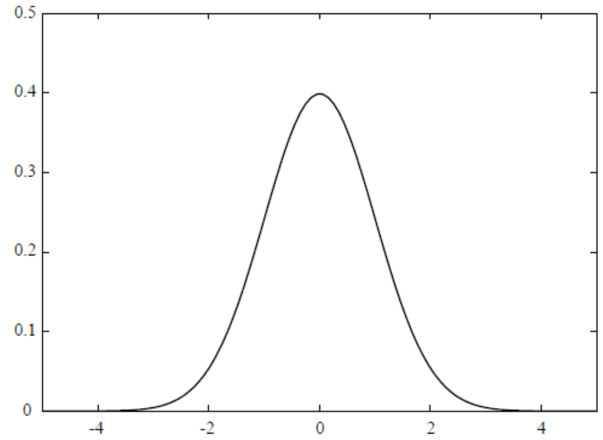
$$P([a_{lh}] = \boldsymbol{A}) \propto \prod_{l=1}^{L} \prod_{h=1}^{H} \exp\left(-\frac{a_{lh}^2}{2}\right)$$

$\boldsymbol{A}$の事前分布 (ガウス)
→ ℓ_2正則化

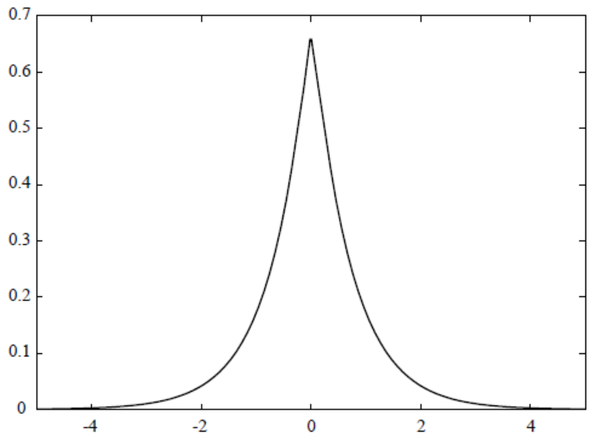
$$P([b_{hm}] = \boldsymbol{B}) \propto \prod_{h=1}^{H} \prod_{m=1}^{M} \exp\left(-\frac{|b_{hm}|}{k}\right)$$

$\boldsymbol{B}$の事前分布 (ラプラス)
→ ℓ_1正則化

k: 正則化係数 (後で重要)



ガウス分布



ラプラス分布
(ゼロ付近の密度が高い)

事前分布をラプラス分布にする
= ℓ_1正則化
このとき, 推定変数をスパースにできる
(推定結果にゼロ成分を多く含ませることが可能)

理由 : 事前分布の形状よりも幾何学的な考察が必要
(圧縮センシングなどが分かりやすい)

[partC] ガウス事前分布→ラプラス事前分布の問題点

尤度と事前分布を以下のように仮定

$$P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B}) = \frac{1}{\sqrt{(2\pi\sigma^2)^{LH}}}\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{V}-\boldsymbol{AB}\|_{\mathrm{FRO}}^2\right)$$ ノイズはガウス分布

$$P(\boldsymbol{A}) = \frac{1}{\sqrt{(2\pi)^{LH}}}\exp\left(-\frac{1}{2}\mathrm{Tr}\boldsymbol{AA}^T\right) = \prod_l\prod_h \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}a_{lh}^2\right)$$ $\boldsymbol{A}$はガウス事前分布

$$P(\boldsymbol{B}) \propto \prod_m\prod_h \exp\left(-\frac{1}{k}|b_{mh}|\right)$$ $\boldsymbol{B}$はラプラス事前分布

試行関数の条件は一般の分布で成り立つが
試行関数が多次元ガウス分布であることは言えない（試行関数の形は不明）

$$r(\boldsymbol{A}) = P(\boldsymbol{A})\exp\left(\int d\boldsymbol{B}\, r(\boldsymbol{B})\log P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})\right) \Longrightarrow \cancel{r(\boldsymbol{A}) \propto \exp\left(-\frac{1}{2}\mathrm{Tr}(\boldsymbol{A}-\hat{\boldsymbol{A}})\hat{\boldsymbol{\Sigma}}_A^{-1}(\boldsymbol{A}-\hat{\boldsymbol{A}})^T\right)}$$ の形

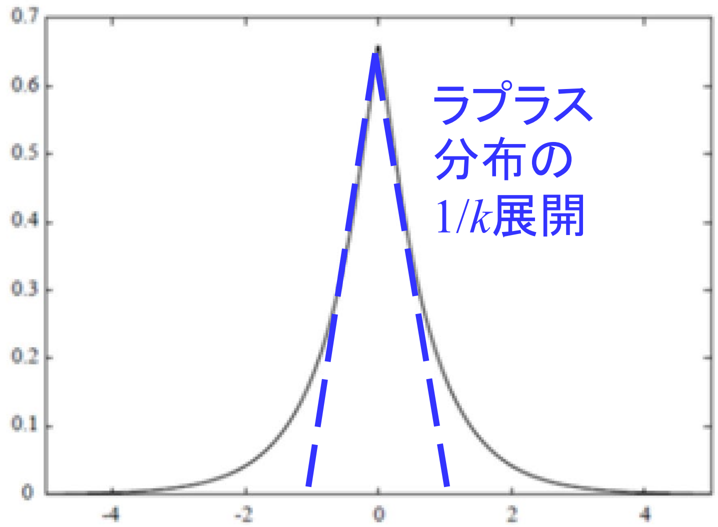
$$r(\boldsymbol{B}) = P(\boldsymbol{B})\exp\left(\int d\boldsymbol{A}\, r(\boldsymbol{A})\log P(\boldsymbol{V}|\boldsymbol{A},\boldsymbol{B})\right) \Longrightarrow \cancel{r(\boldsymbol{B}) \propto \exp\left(-\frac{1}{2}\mathrm{Tr}(\boldsymbol{B}-\hat{\boldsymbol{B}})\hat{\boldsymbol{\Sigma}}_B^{-1}(\boldsymbol{B}-\hat{\boldsymbol{B}})^T\right)}$$ の形

[partC] ラプラス事前分布の下での近似

近似を置かないと解析が進まないので、以下の近似を置く

- 試行関数の成分独立性（平均場近似） $r(A)=\prod_l\prod_h r_a(a_{lh})$ $\quad r(\boldsymbol{B})=\prod_h\prod_m r_b(b_{hm})$
- 事後分布の共分散行列（$H\times H$行列）は対角成分のみ持つ（分散成分のみ）
- ラプラス事前分布のパラメータ（ℓ_1 正則化係数）について、$k\to\infty$ を仮定し，$1/k$展開（1次まで）

$$P(\boldsymbol{B})\propto\prod_m\prod_h\exp\left(-\frac{1}{k}|b_{mh}|\right)\simeq 1-\frac{1}{k}\sum_{mh}|b_{mh}|+O\left(\frac{1}{k^2}\right)$$



※$r(\boldsymbol{B})$の1次，2次モーメントを計算したいが，$\boldsymbol{B}$の多重積分ができない
→平均場近似と$1/k$展開で1個の積分まで落ちる

$$r(\boldsymbol{B})=P(\boldsymbol{B})\exp\left(\int dA\, r(A)\log P(V|A,\boldsymbol{B})\right)$$

※展開により確率分布がill-definedになる可能性があるので注意が必要 → 後の話とも関係

[partC] 近似で得られる解析解

以上の試行分布の仮定と事前分布の近似を，
試行分布の一般的な関係式に代入し，事後平均と事後分散を計算

$$r(A) = P(A)\exp\left(\int dB\, r(B)\log P(V|A,B)\right) \qquad r(B) = P(B)\exp\left(\int dA\, r(A)\log P(V|A,B)\right)$$

Kawasumi-KT (2018)

***A*の事後平均** $$\bar{a}_{lh} = \sum_{h'=1}^{H}\sum_{m=1}^{M} v_{lm}(\hat{\Sigma}_{\mathbf{A}l}^{-1})_{hh'}\bar{b}_{h'm}$$

***A*の事後分散** $$(\Sigma_{\mathbf{A}l})_{hh} = \sigma^2(\hat{\Sigma}_{\mathbf{A}l}^{-1})_{hh}$$

***B*の事後平均** $$\bar{b}_{hm} = \sum_{h'=1}^{H}\sum_{l=1}^{L} v_{lm}(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{hh'}\bar{a}_{lh'} \boxed{- \frac{1}{kZ_B}\sum_{h'=1}^{H}\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\mathrm{erf}(\omega_{h'm})}$$

***B*の事後分散** $$(\Sigma_{\mathbf{B}m})_{hh} = \sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{hh} \boxed{- \frac{1}{kZ_B}\sum_{h'=1}^{H}\sqrt{\frac{2}{\pi\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h'}}}\{\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\}^2\exp(\omega_{h'm})} \boxed{-\left(\frac{1}{kZ_B}\right)^2\left\{\sum_{h'=1}^{H}\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\mathrm{erf}(\omega_{h'm})\right\}^2}$$

$1/k$展開による
補正項

[partC] 近似で得られる解析解

前ページで定義していない変数

$$\omega_{hm} := \frac{1}{\sqrt{2\sigma^2(\hat{\Sigma}_{B_m}^{-1})_{hh}}} \sum_{h'=1}^{H} \sum_{l=1}^{L} v_{lm} (\hat{\Sigma}_{B_m}^{-1})_{hh'} \bar{a}_{lh'}$$

$$(\hat{\Sigma}_{A_l})_{hh'} := \sigma^2 \delta_{hh'} + \sum_{m=1}^{M} \left((\Sigma_{B_m})_{hh} \delta_{hh'} + \bar{b}_{hm} \bar{b}_{h'm} \right)$$

$$(\hat{\Sigma}_{B_m})_{hh'} := \sum_{l=1}^{L} \left((\Sigma_{A_l})_{hh} \delta_{hh'} + \bar{a}_{lh} \bar{a}_{lh'} \right)$$

$$S_B := \sum_{h=1}^{H} \sum_{m=1}^{M} \left\{ \sqrt{\frac{2\sigma^2(\hat{\Sigma}_{B_m}^{-1})_{hh}}{\pi}} \exp(-\omega_{hm}^2) + \left(\sum_{h'=1}^{H} \sum_{l=1}^{L} v_{lm} (\hat{\Sigma}_{B_m}^{-1})_{hh'} \bar{a}_{lh'} \right) \operatorname{erf}(\omega_{hm}) \right\}$$

$$\boxed{Z_B = 1 - \frac{S_B}{k}}$$

行列*B*の(近似)事後分布に関する正規化因子
以降の話で重要

[partC] 解析解の性質

Aの事後平均 $$\bar{a}_{lh} = \sum_{h'=1}^{H}\sum_{m=1}^{M} v_{lm}(\hat{\Sigma}_{\mathbf{A}l}^{-1})_{hh'}\bar{b}_{h'm}$$

Aの事後分散 $$(\Sigma_{\mathbf{A}l})_{hh} = \sigma^2(\hat{\Sigma}_{\mathbf{A}l}^{-1})_{hh}$$

1/k展開による
補正項

Bの事後平均 $$\bar{b}_{hm} = \sum_{h'=1}^{H}\sum_{l=1}^{L} v_{lm}(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{hh'}\bar{a}_{lh'} \boxed{- \frac{1}{kZ_B}\sum_{h'=1}^{H}\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\mathrm{erf}(\omega_{h'm})}$$

Bの事後分散 $$(\Sigma_{\mathbf{B}m})_{hh} = \sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{hh} \boxed{- \frac{1}{kZ_B}\sum_{h'=1}^{H}\sqrt{\frac{2}{\pi\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h'}}}\{\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\}^2\exp(\omega_{h'm})}$$
$$\boxed{-\left(\frac{1}{kZ_B}\right)^2\left\{\sum_{h'=1}^{H}\sigma^2(\hat{\Sigma}_{\mathbf{B}m}^{-1})_{h'h}\mathrm{erf}(\omega_{h'm})\right\}^2}$$

- 青い部分は全てkZ_Bを分母に含む
- これらの項がない場合($kZ_B \to \infty$)は，Bの事前分布が一様分布の場合

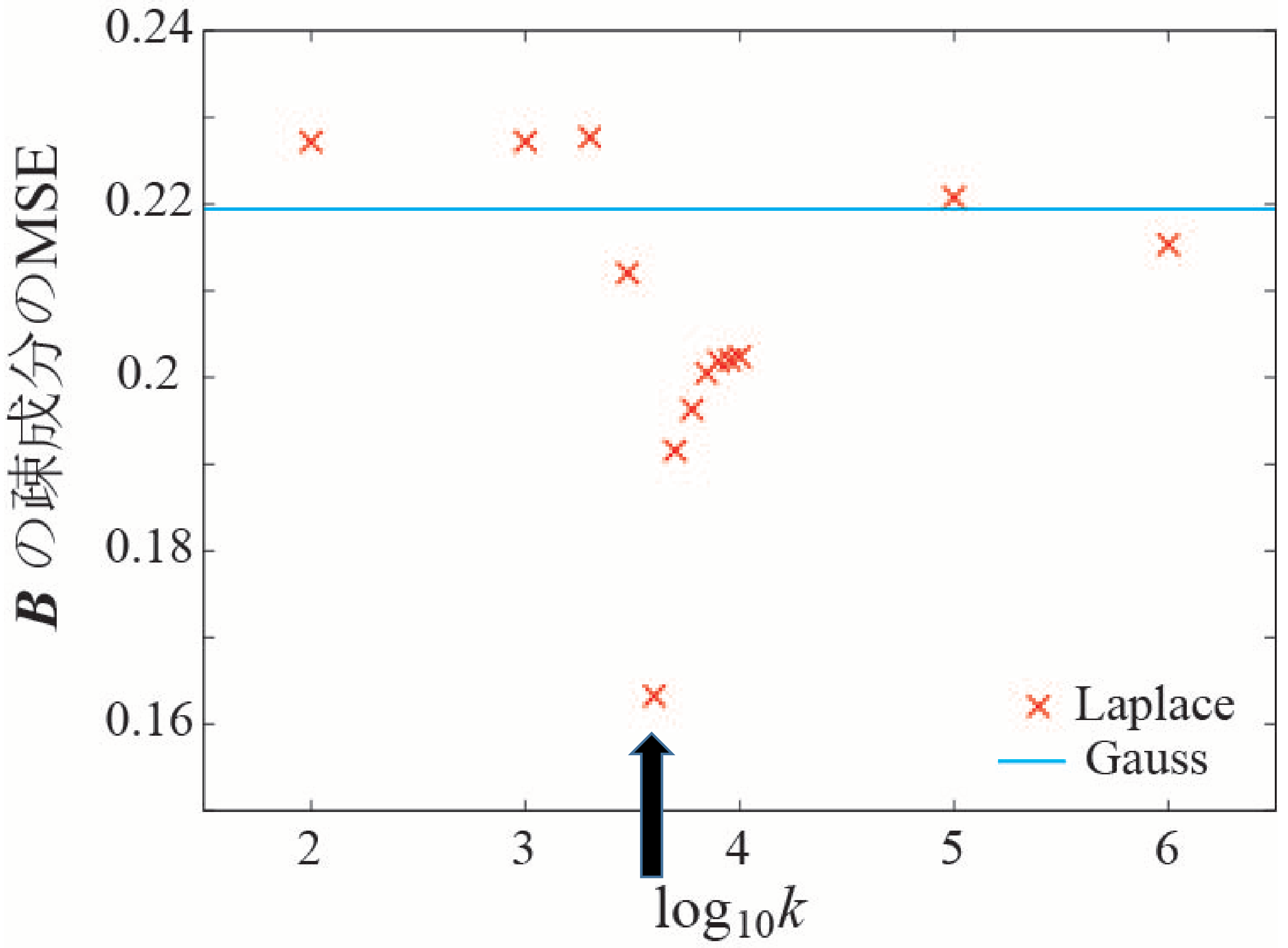
  ➡ 青い部分はラプラス分布による補正項
- $kZ_B \to 0$ では補正項の寄与が大きくなる → 以降の話と関係

[partC] 解析解をアルゴリズムとして利用した結果

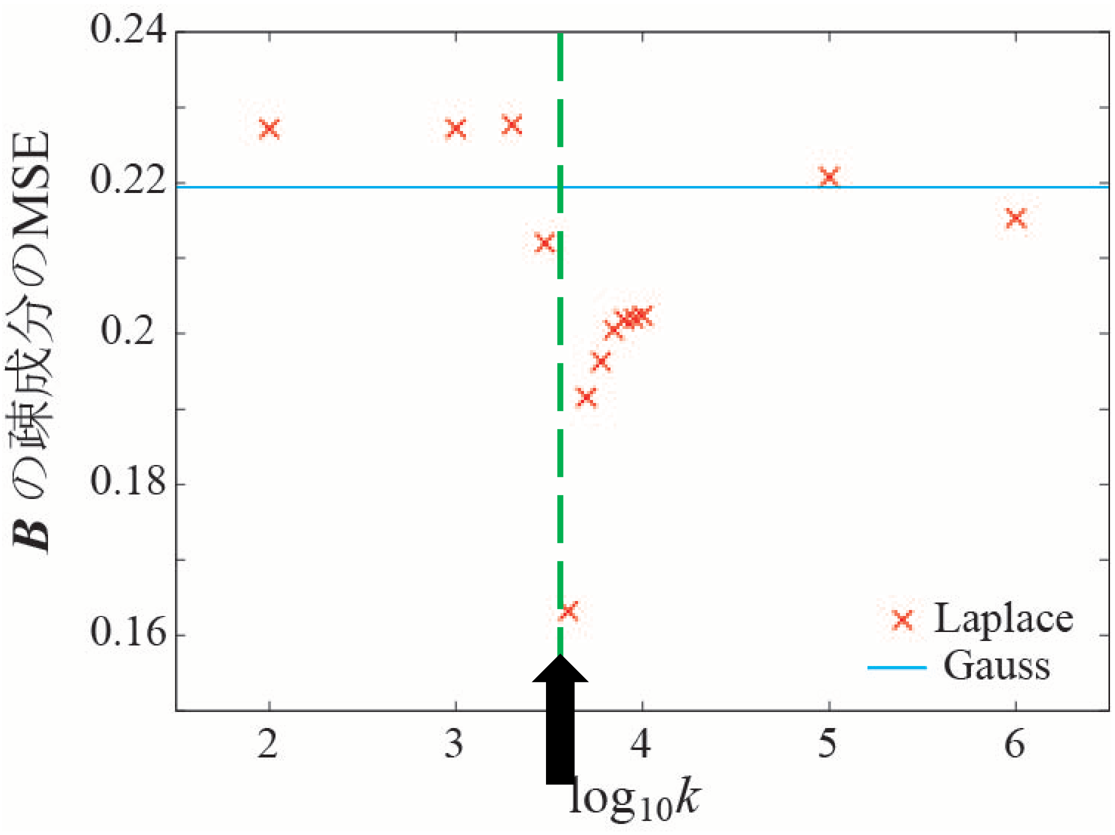
前記の解を
逐次代入型アルゴリズムとして
数値実験で評価

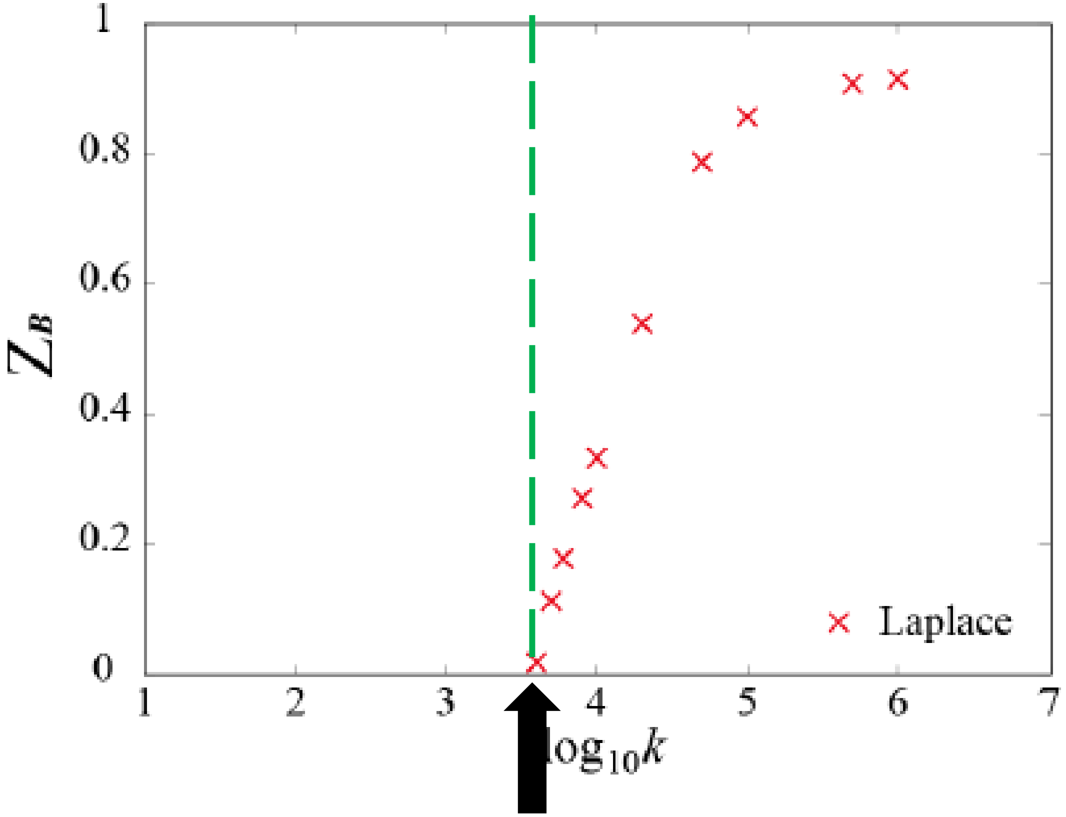
- スパース正解行列$\boldsymbol{B}$の
  非ゼロ成分比は0.5
- ラプラス分布のパラメータ k は
  固定し，様々な値で実験

特定の k の値で
ラプラス事前分布による
スパース行列の再構成が
ガウス事前分布より良好



[partC] 結果に関する考察(正規化因子の零点)





- $\boldsymbol{B}$がスパース行列となるには, $\boldsymbol{B}$の分布の正規化因子である Z_B がほぼ0となるような k を選べばよい
- 十分大きな k では $Z_B \to 1$ となる (このとき $kZ_B \to \infty$)

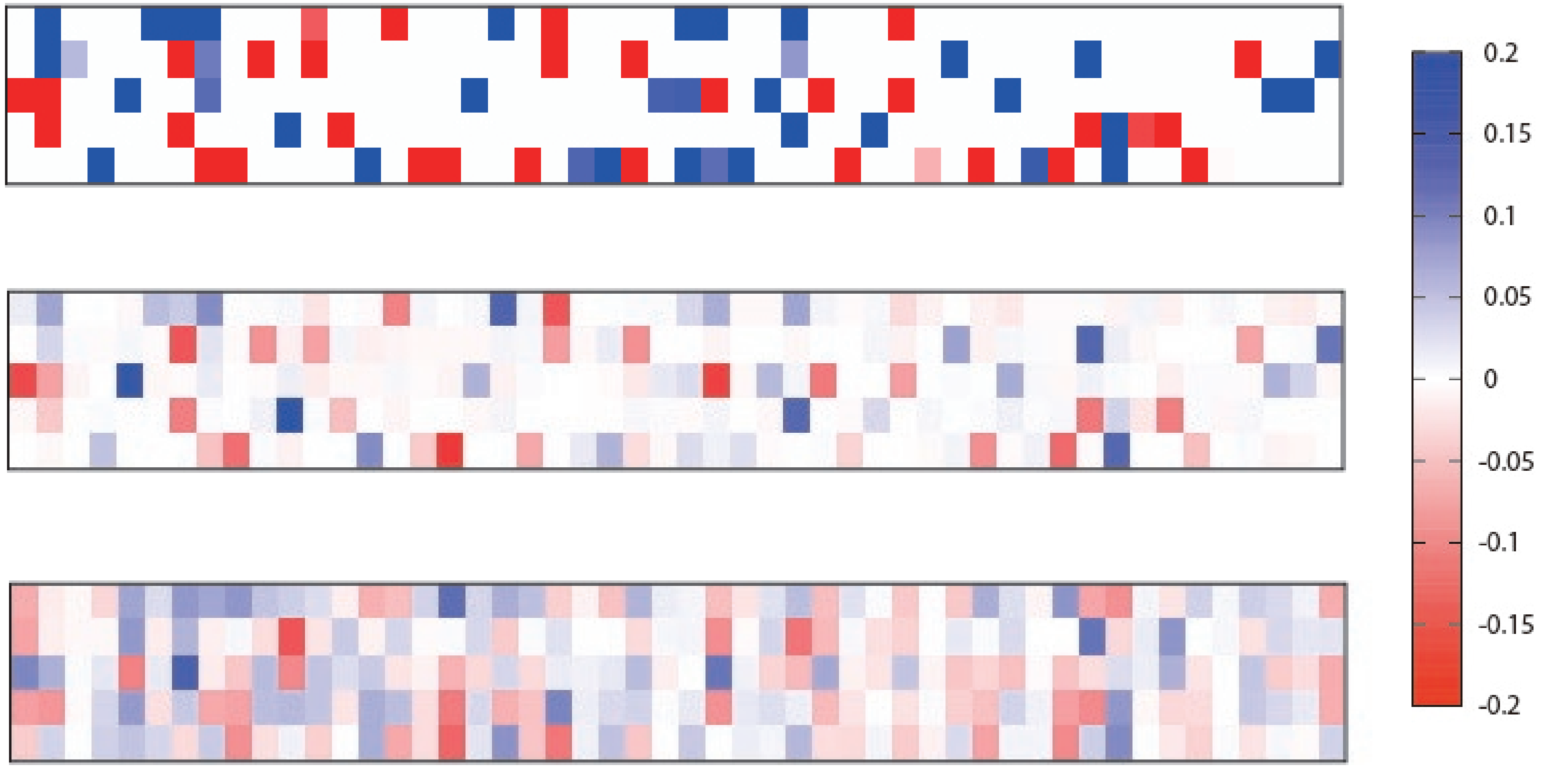
[partC] 復元したスパース行列の例

最適なkの値付近で，$\boldsymbol{B}$の復元結果を可視化 ($\boldsymbol{B}$: 5 × 50行列)

用意した
スパース正解行列$\boldsymbol{B}$

ラプラス事前分布で
求めた$\boldsymbol{B}$

ガウス事前分布で
求めた$\boldsymbol{B}$

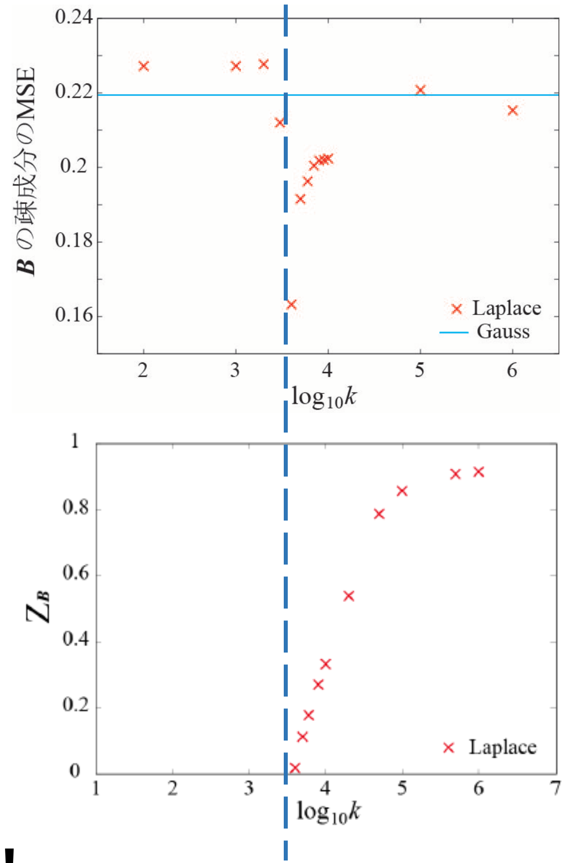


ラプラス事前分布で求めた$\boldsymbol{B}$は，用意したスパース行列に近い(<u>が，少し余計な成分が見える</u>)

※ 行列サイズが大きくなると，誤差が大きくなってしまう…

[partC] ラプラス事前分布の正則化係数について

事前分布のパラメータ k の適切な値
= Z_B（近似事後分布の正規化因子）の k に関する零点
（実験からの知見，理論的には理由は明確ではない）



問題点：
与えられた行列 V に対し，Z_B のゼロ点は解析的には不明
行列分解の数値実験を行わないと，Z_B のゼロ点の位置は不明

➡ 最適な k を知るには，k を変えて実験する必要あり，面倒！

➡ スパース行列の解を数値的に求めながら，
最適な k を適応的に探索することを考える

[partC] 正則化係数の自動調整

スパース行列分解のアルゴリズム実行時に,
ラプラス分布のパラメータ (正則化係数) k を以下のように更新

$$k^{(n)} = (1-\varepsilon)k^{(n-1)} + \varepsilon S_B^{(n-1)}$$

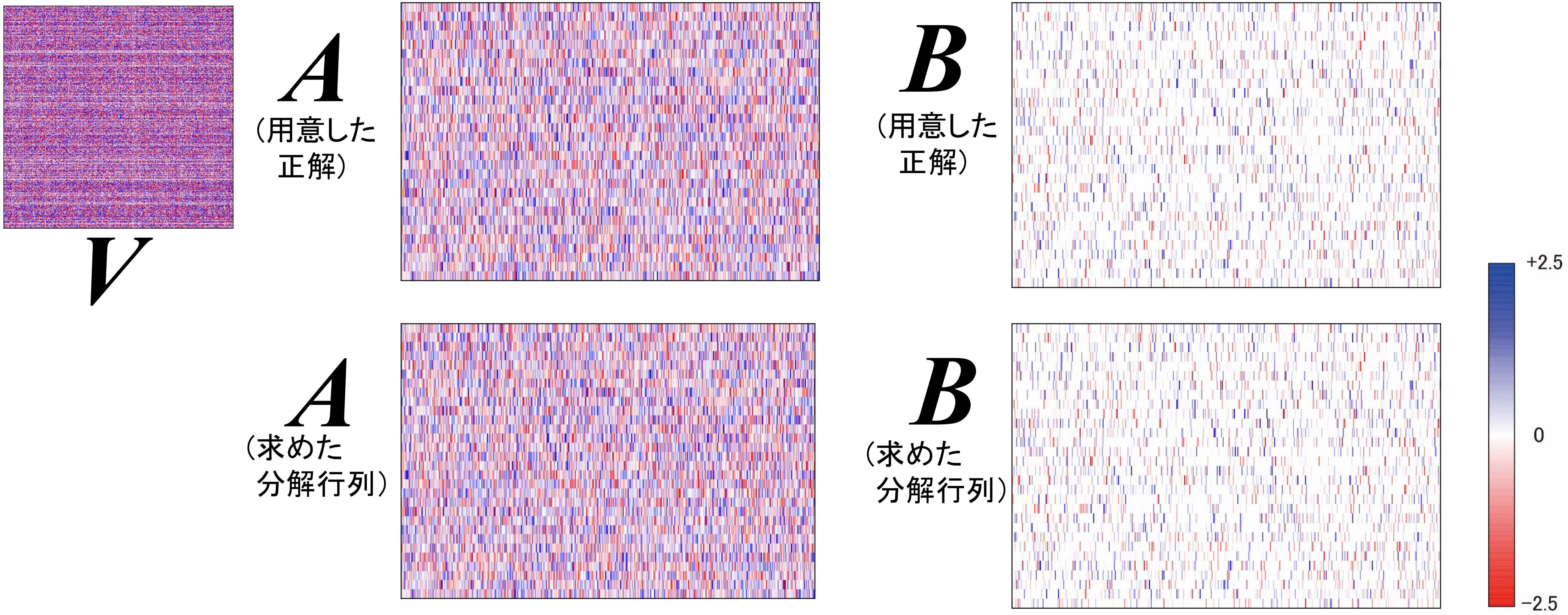
括弧付き添字:アルゴリズムの更新ステップ数

ε: 忘却係数 →アルゴリズムを収束させるためのパラメータ $(0 < \varepsilon < 1)$
εは小さく取る, 大きすぎるとアルゴリズムが不安定化

kが収束すれば, ステップ数$n \to \infty$ のときに $k \approx S_B$
$Z_B = 1 - \dfrac{S_B}{k}$ なので, このとき$Z_B \to 0$ となりラプラス分布による補正項が効く

[partC] 自動調整下でのスパース行列分解の結果

元の行列と分解行列解との比較例 ($L = 500$, $M = 500$, H（中間次元）$= 30$, $\boldsymbol{B}$の非ゼロ成分比 $= 0.2$)

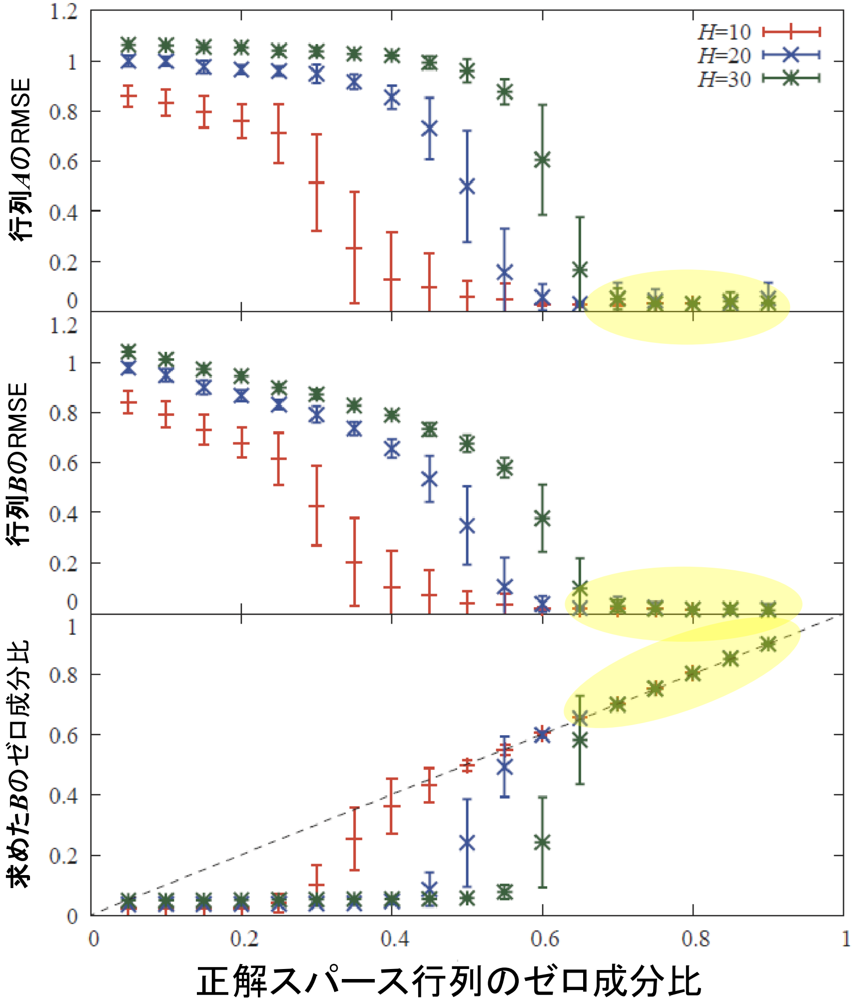


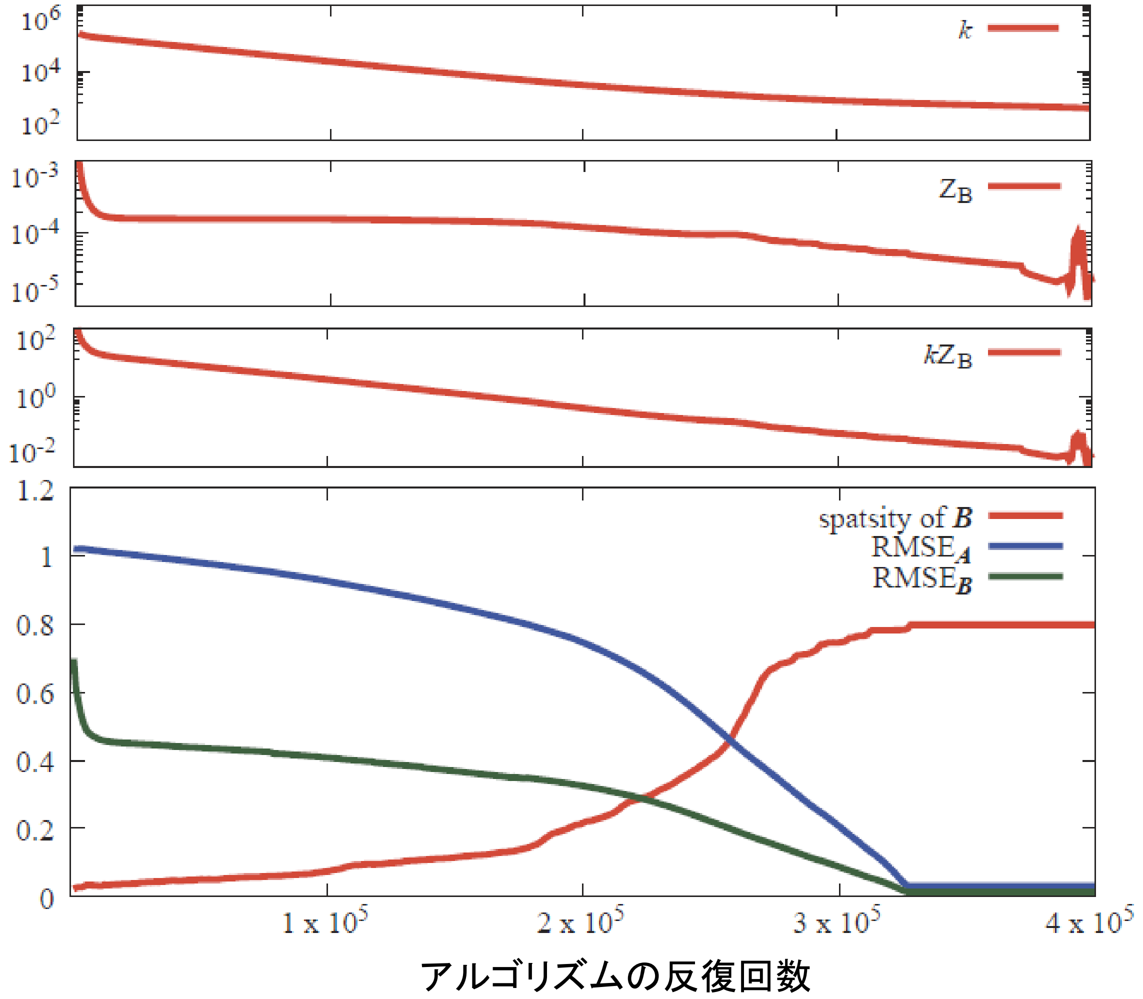
行列サイズが大きくても
用意した正解と分解行列がほぼ一致，差はほとんどない → 適応的な k の決定の結果

[partC] 自動調整下でのスパース行列分解の結果

スパース行列復元のアルゴリズムの性質

Kawasumi-KT (2023)





スパース性が高い領域では
正解スパース行列がほぼ復元できている

反復アルゴリズムを続けると
ある点で誤差がほぼゼロになる（ただし収束が遅い）

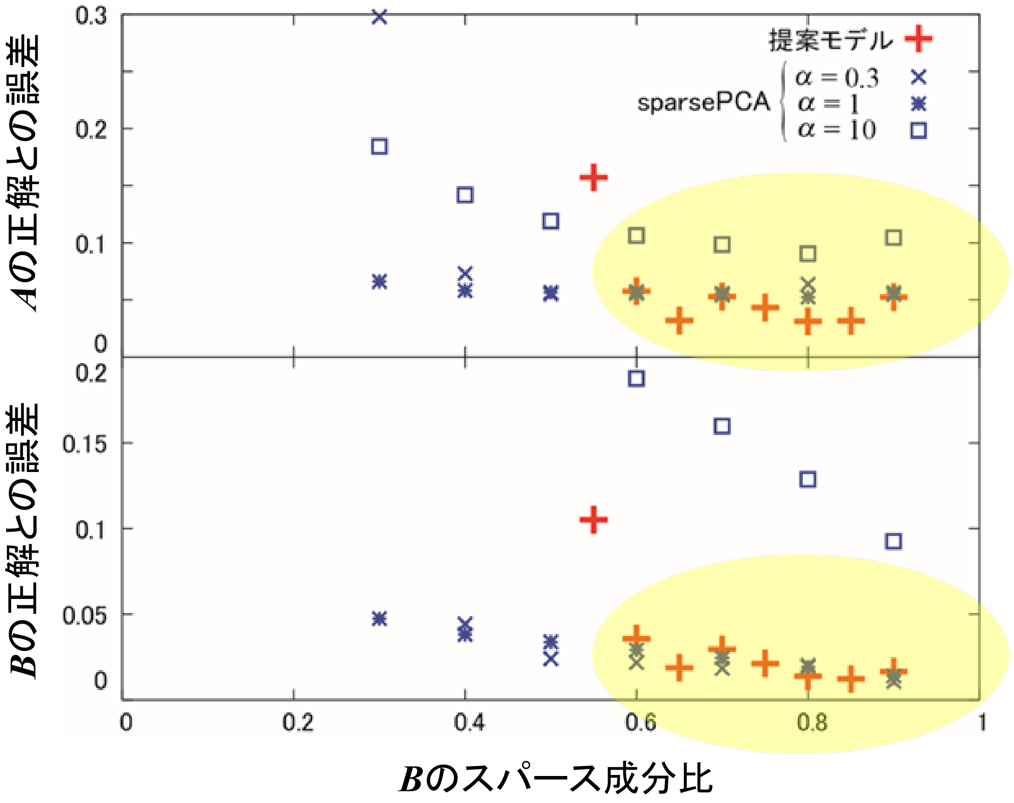
[partC] 自動調整下でのスパース行列分解の結果

既存手法との比較:

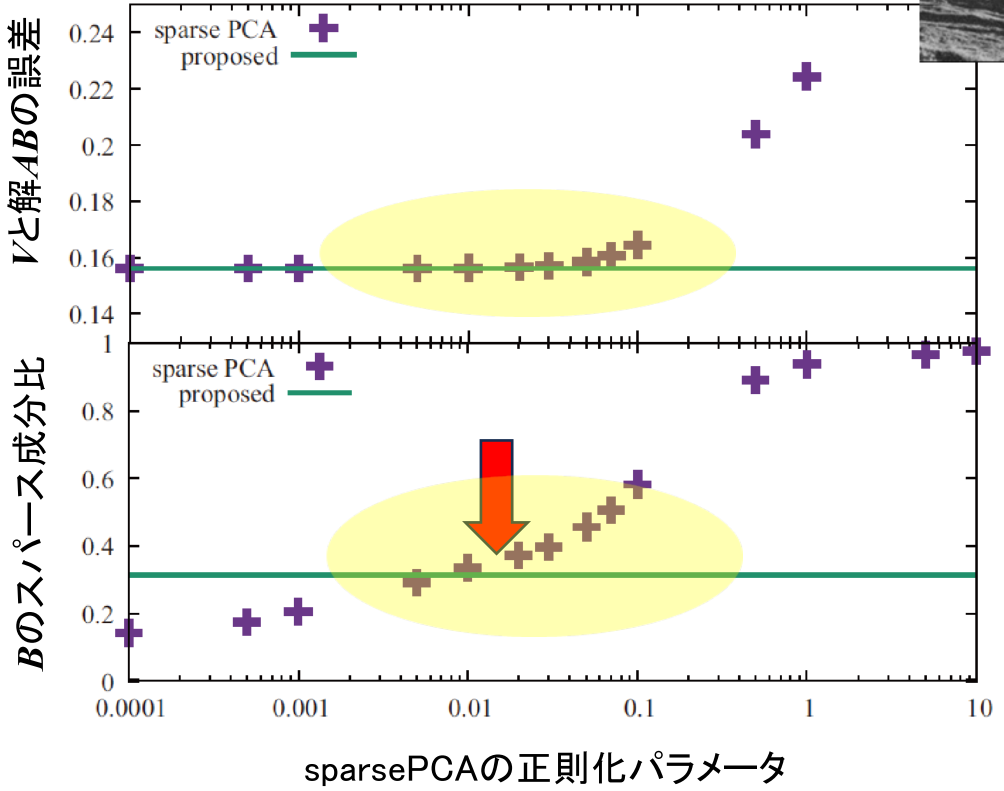
Kawasumi-KT (2023)

sparsePCA (Mairal et al. (2009)): $\min_{A,B}\left(\|V-AB\|_{\mathrm{FRO}}^{2}+\tilde{\alpha}\|B\|_{1}\right)$ を解くアルゴリズム

(A)人工データ:正解として与えた行列A,Bの求解



(B) 実画像のスパース行列分解



Tree:
The USC-SIPI
Image Database

スパース性が高い領域で正解行列の再現性が高い
sparsePCAと同程度かそれ以上

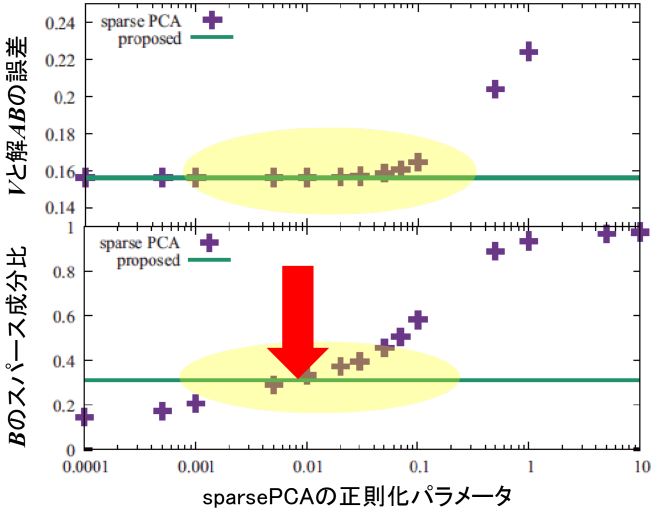
積ABの誤差が最低の領域内で
Bのスパース性が高い解を得ている？

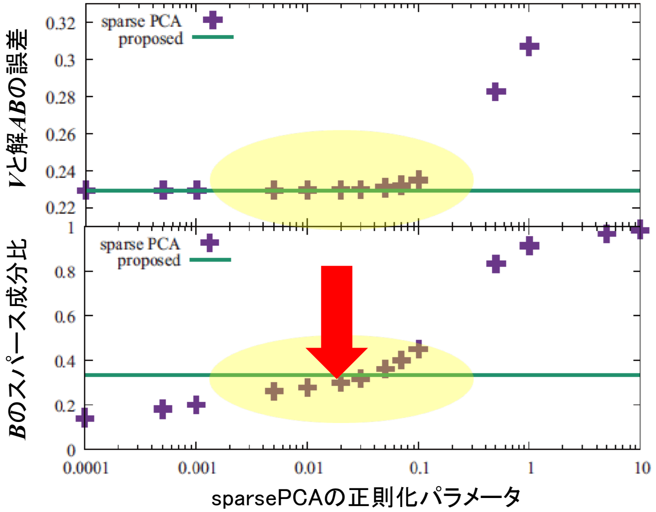
[partC] 自動調整下でのスパース行列分解の結果

既存手法 (sparsePCA) との比較　($H = 40$)

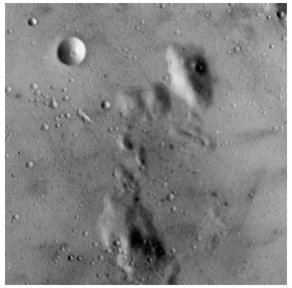
Kawasumi-KT (2023)

Tree

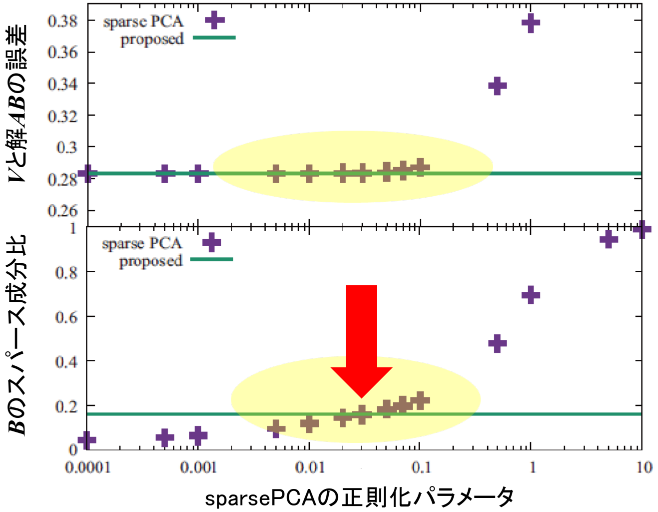


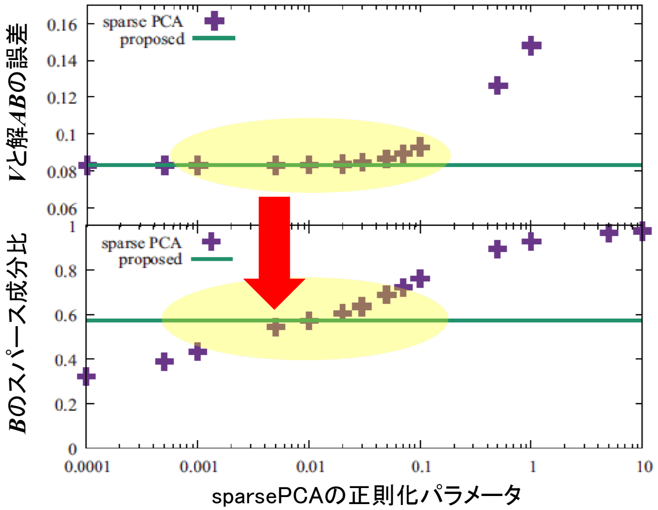


Moon surface



Aerial





Clock

どの画像でも
同じ傾向が見える

以上の画像データはThe USC-SIPI Image Databaseより引用

[partD] 変分ベイズ解の性質の解析？

変分ベイズ法で行列分解問題が解けていることが分かった
解の性質とアルゴリズムの動的挙動が解析的に知りたい…
→ スパース行列分解の変分ベイズ解は複雑なので面倒？

まず事前分布がガウス分布である行列分解モデルから　　Nakajima–Sugiyama (2011)

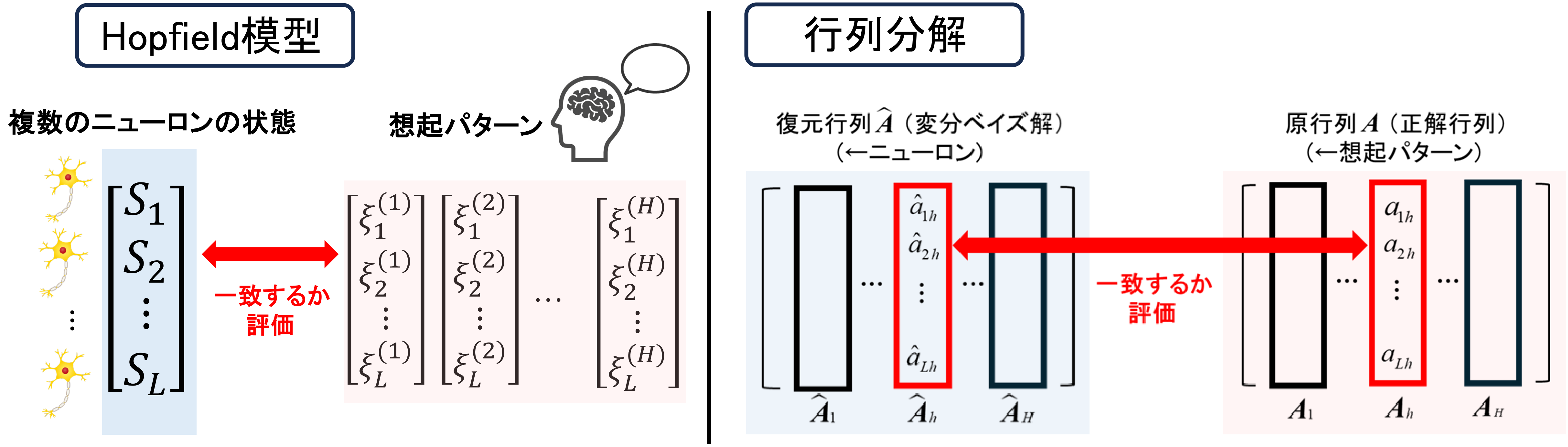
$$\hat{A} = \frac{1}{\sigma^2} V \hat{B} \hat{\Sigma}_A \qquad \hat{B} = \frac{1}{\sigma^2} \hat{\Sigma}_B \hat{A}^T V$$

$$\hat{\Sigma}_A = \sigma^2 \left(\hat{B}\hat{B}^T + M\hat{\Sigma}_B + \sigma^2 I_H \right)^{-1} \qquad \hat{\Sigma}_B = \sigma^2 \left(\hat{A}^T\hat{A} + L\hat{\Sigma}_A + \sigma^2 I_H \right)^{-1}$$

$\hat{A}, \hat{B}$: A, Bの事後平均　　$\hat{\Sigma}_A, \hat{\Sigma}_B$: A, Bの事後共分散

逐次代入で解くと分解行列解が得られる → 行列分解アルゴリズムとして利用可能

[partD] 行列分解と連想記憶(Hopfield模型)との関係



Hopfieldモデル(連想記憶モデル)と行列分解を上のように対応

- 正解行列の特定のベクトル → 想起したいパターン
- 復元行列の特定のベクトル → ニューロンの状態

Hopfield模型の古典的解析手法を行列分解に応用できるか？

- Amit–Gutfreund–Sompolinsky (1985 , 1987), 自由エネルギー形式, 想起可能性
- Shiino–Fukai (1993) → 学習則ベース, 想起可能性
- Amari–Maginu (1988) → 学習則ベース, 想起ダイナミクス

[partD] Hopfield模型の解析手法の利用:信号雑音分離

$$\hat{A} = \frac{1}{\sigma^2} V \hat{B} \hat{\Sigma}_A \qquad \hat{B} = \frac{1}{\sigma^2} \hat{\Sigma}_B \hat{A}^T V$$

$$\hat{\Sigma}_A = \sigma^2 \left(\hat{B}\hat{B}^T + M\hat{\Sigma}_B + \sigma^2 I_H \right)^{-1} \qquad \hat{\Sigma}_B = \sigma^2 \left(\hat{A}^T \hat{A} + L\hat{\Sigma}_A + \sigma^2 I_H \right)^{-1}$$

これらは行列要素に依存する量 → 行列分解の性質を表す大域的な量の性質が知りたい

連想記憶における信号雑音分離との類似性を利用
→ 上記を連則記憶の学習則とみなし, 信号(埋め込んだ正解行列)と雑音成分に分けたい

信号成分→正解との重なり
(正解行列と分解行列との
h 番目のベクトルの重なり)

$$\mu_h = \sum_{l=1}^{L} a_{lh} \hat{a}_{lh} \qquad \nu_h = \sum_{m=1}^{M} b_{hm} \hat{b}_{hm}$$

雑音成分→ガウス雑音
(正解行列と分解行列との
異なる行/列のベクトルの重なり)

$$\sum_{l=1}^{L} a_{lh'} \hat{a}_{lh} \sim N(0, \sigma_{ah'}^2) \qquad \sum_{m=1}^{M} b_{h'm} \hat{b}_{hm} \sim N(0, \sigma_{bh'}^2) \qquad (h' \neq h)$$

[partD] 信号雑音分離で得られる簡約式

A,*B* ともガウス事前分布の場合の変分ベイズ解は
信号雑音分離仮説が正しければ，次を満たすと予想．(変数8個の関係式)

Tamai-Kawasumi-KT, STATPHYS28

・正解との重なり

$$\mu_h = LM\frac{\hat{\Sigma}_{Ah}}{\sigma^2}(\nu_h + M\mu_h\hat{\Sigma}_{Bh}) \qquad \nu_h = LM\frac{\hat{\Sigma}_{Bh}}{\sigma^2}(\mu_h + L\nu_h\hat{\Sigma}_{Ah})$$

・干渉ノイズ分散

$$\sigma_a^2 = \left(LM\frac{\hat{\Sigma}_{Ah}}{\sigma^2}\right)^2\left(\sigma_b^2 + \left(\frac{\hat{\Sigma}_{Bh}}{L}\right)^2\sigma_a^2\right) \qquad \sigma_b^2 = \left(LM\frac{\hat{\Sigma}_{Bh}}{\sigma^2}\right)^2\left(\sigma_a^2 + \left(\frac{\hat{\Sigma}_{Ah}}{M}\right)^2\sigma_b^2\right)$$

・共分散行列の
対角成分

$$\hat{\Sigma}_{Bh} = \frac{\sigma^2/(LM)}{(1/L)(\sigma^2/(LM)) + L\hat{\Sigma}_{Ah} + \rho_A} \qquad \hat{\Sigma}_{Ah} = \frac{\sigma^2/(LM)}{(1/M)(\sigma^2/(LM)) + M\hat{\Sigma}_{Bh} + \rho_B}$$

・h列/h行ベクトルの
正規化定数

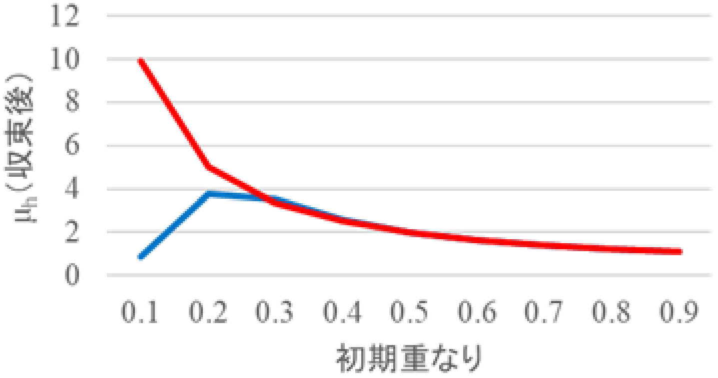
$$\rho_A = \left(LM\frac{\hat{\Sigma}_{Ah}}{\sigma^2}\right)^2\left(\nu_h^2 + (H-1)\sigma_b^2 + 2\frac{\hat{\Sigma}_{Bh}}{L}\mu_h\nu_h + \frac{\sigma^2}{M}\rho_B\right)$$

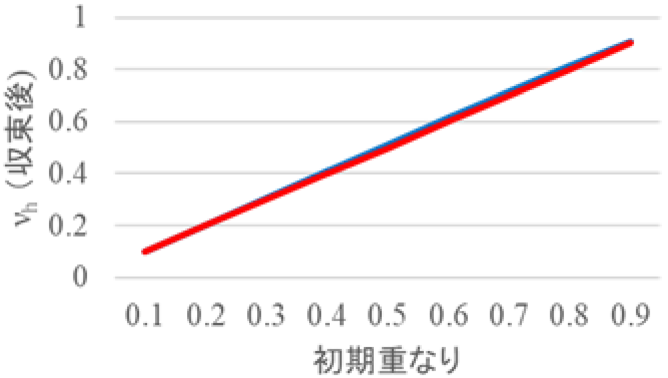
$$\rho_B = \left(LM\frac{\hat{\Sigma}_{Bh}}{\sigma^2}\right)^2\left(\mu_h^2 + (H-1)\sigma_a^2 + 2\frac{\hat{\Sigma}_{Ah}}{M}\mu_h\nu_h + \frac{\sigma^2}{L}\rho_A\right)$$

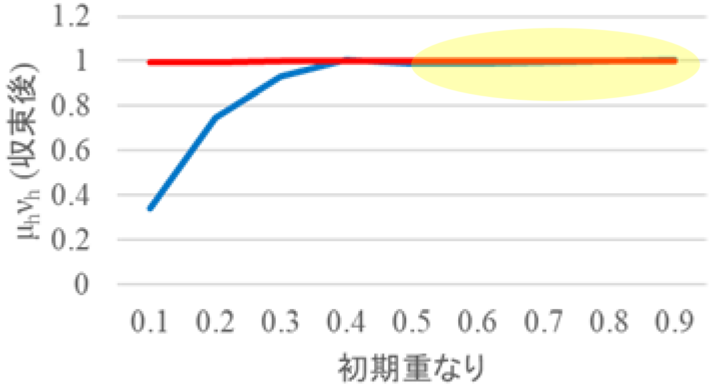
[partD] 簡約式の解と数値実験との比較

簡約式を逐次的に解くことで，簡約式の解となる重なりμ_h, ν_hの値を調べる．（μ_h, ν_hの重なりの初期値を変更）
実際の行列分解実験と比較： 簡約式（赤線━），数値実験（青線━）

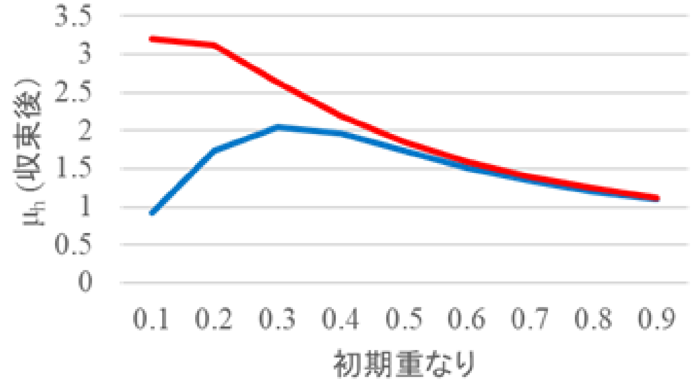
低ノイズ
（$\sigma \to$ 小）

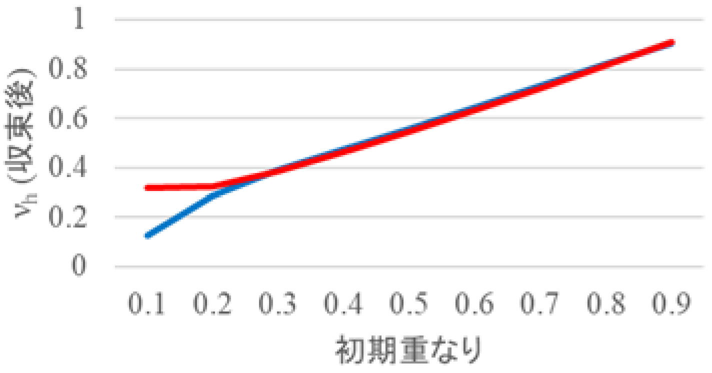


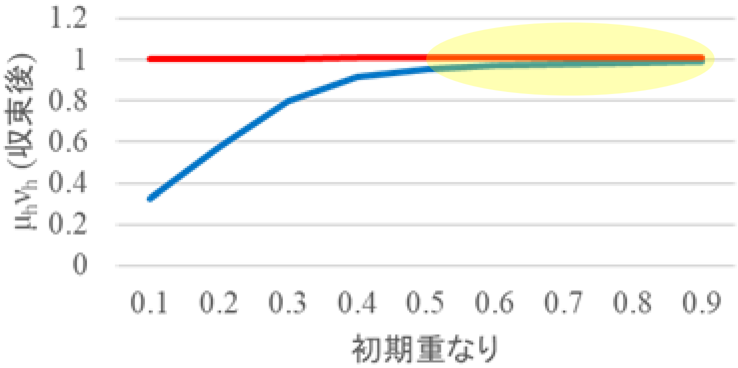




高ノイズ
（$\sigma \to$ 大）







収束後の積$\mu_h \nu_h$が1に近ければ収束解は用意した正解行列に近い

- μ_h, ν_hの初期値 ~ 1 → 正解行列付近に収束，簡約式は分解行列解の性質をほぼ正しく記述
- μ_h, ν_hの初期値 ~ 0 → 正解行列から遠くに収束→行列が回転（成分が混合）し正解から離れる

[partD] 自由エネルギー形式による簡約式の導出

これまでの解析は変分ベイズ解を用いたが，自由エネルギーから出発することも可能

変分ベイズ自由エネルギーを事後平均，事後共分散 $\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}, \hat{\boldsymbol{\Sigma}}_A, \hat{\boldsymbol{\Sigma}}_B$ で表現

$$
\begin{aligned}
2f \quad = \quad & \frac{\|\boldsymbol{V} - \widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^T\|_{\mathrm{FRO}}^2}{\sigma^2} - L\log|\widehat{\boldsymbol{\Sigma}}_A| - M\log|\widehat{\boldsymbol{\Sigma}}_B| + \mathrm{Tr}(\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}} + L\widehat{\boldsymbol{\Sigma}}_A) + \mathrm{Tr}(\widehat{\boldsymbol{B}}^T\widehat{\boldsymbol{B}} + M\widehat{\boldsymbol{\Sigma}}_B) \\
& - \frac{1}{\sigma^2}\mathrm{Tr}\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}}\widehat{\boldsymbol{B}}^T\widehat{\boldsymbol{B}} + \frac{1}{\sigma^2}\mathrm{Tr}(\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}} + L\widehat{\boldsymbol{\Sigma}}_A)(\widehat{\boldsymbol{B}}^T\widehat{\boldsymbol{B}} + M\widehat{\boldsymbol{\Sigma}}_B)
\end{aligned}
$$

大域的変数を導入し，自由エネルギーをこれらで表現しなおす

・正解行列と推定行列の重なり

$$\mu_h = \frac{1}{L}\sum_{l=1}^{L} a_{lh}\hat{a}_{lh} \qquad \nu_h = \frac{1}{M}\sum_{m=1}^{M} b_{hm}\hat{b}_{hm}$$

・干渉ノイズ

$$\sigma_{Ahh'}^2 = \left(\frac{1}{L}\sum_{l=1}^{L} a_{lh'}\hat{a}_{lh}\right)^2 \qquad \sigma_{Bhh'}^2 = \left(\frac{1}{M}\sum_{m=1}^{M} b_{h'm}\hat{b}_{hm}\right)^2$$

・正規化因子

$$\rho_{ah} = \frac{1}{L}\sum_{l=1}^{L} \hat{a}_{lh}^2 \qquad \rho_{bh} = \frac{1}{M}\sum_{m=1}^{M} \hat{b}_{hm}^2 \qquad e_h = \frac{1}{LM}\sum_{l=1}^{L}\sum_{m=1}^{M} \varepsilon_{lm}\hat{a}_{lh}\hat{b}_{hm}$$

[partD] 自由エネルギー形式による簡約式の導出

得られた自由エネルギーから，導入した変数間の関係式を導出（状態方程式）

$$\mu_h = \frac{M}{\sigma^2}\hat{\Sigma}_{Ah}\nu_h\left(1 - M\frac{\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}}{\sigma^2}\right)^{-1} \qquad \sigma^2_{ah'h} = \frac{1}{2}\left(\frac{M\hat{\Sigma}_{ah}}{\sigma^2}\right)^2\sigma^2_{bh'h}\left(1 \pm \sqrt{1 + \frac{4}{L\hat{\Sigma}_{ah}}\left(\frac{\sigma^2}{M\sigma^2_{bh'h}}\right)^2}\right) + \frac{\hat{\Sigma}_{ah}}{L}$$

$$\nu_h = \frac{L}{\sigma^2}\hat{\Sigma}_{Bh}\mu_h\left(1 - L\frac{\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}}{\sigma^2}\right)^{-1} \qquad \sigma^2_{bh'h} = \frac{1}{2}\left(\frac{L\hat{\Sigma}_{bh}}{\sigma^2}\right)^2\sigma^2_{ah'h}\left(1 \pm \sqrt{1 + \frac{4}{M\hat{\Sigma}_{bh}}\left(\frac{\sigma^2}{L\sigma^2_{ah'h}}\right)^2}\right) + \frac{\hat{\Sigma}_{bh}}{M}$$

$$\hat{\Sigma}_{Ah} = \frac{\sigma^2}{\sigma^2 + M\hat{\Sigma}_{Bh} + M\rho_{Bh}} \qquad \rho_{Ah} = \mu_h^2 + \frac{M}{\sigma^2}(\hat{\Sigma}_{Ah})^2\nu_h^2 + \left(\frac{L-H+1}{L}\hat{\Sigma}_{Ah} + \sum_{h'\neq h}\sigma^2_{Ah'h}\right)$$

$$\hat{\Sigma}_{Bh} = \frac{\sigma^2}{\sigma^2 + L\hat{\Sigma}_{Ah} + L\rho_{Ah}} \qquad \rho_{Bh} = \nu_h^2 + \frac{L}{\sigma^2}(\hat{\Sigma}_{Bh})^2\mu_h^2 + \left(\frac{M-H+1}{M}\hat{\Sigma}_{Bh} + \sum_{h'\neq h}\sigma^2_{Bh'h}\right)$$

$$e_h = \hat{\Sigma}_{Bh}\mu_h^2 + \frac{1}{\alpha}\hat{\Sigma}_{Ah}\nu_h^2 + \sigma^2\frac{\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}}{\sigma^2 - M\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}} - \frac{L+M}{2}\frac{\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}}{\sigma^2 - M\hat{\Sigma}_{Ah}\hat{\Sigma}_{Bh}}\sum_{h'\neq h}\sqrt{\sigma^2_{Ah'h}\sigma^2_{Bh'h}} \qquad (\alpha = M/L)$$

信号雑音分離による簡約式と対応付けられそうだが，若干ずれている
→ どの程度差異があるか数値的に検証の必要あり

本日のまとめ

$$\boldsymbol{V} = \boldsymbol{A}\ \boldsymbol{B}$$

(A) 行列分解問題の種類を紹介

(B) 行列分解問題とベイズ統計学との関連性を紹介
ベイズ推定を通じ，様々な統計力学的手法が解析に利用できる

(C) ラプラス事前分布の下でのスパース行列分解を解析
幾つかの近似の下で，変分ベイズ法での解析解の表式が得られる → アルゴリズム
正規化因子の零点から正則化係数を決めると，スパース行列が精度良く復元（理由？）

(D) ガウス事前分布下での行列分解の変分ベイズ解析解の性質を
Hopfield模型の信号雑音分離法を援用し解析
正解行列との初期重なりが高い → 簡約式での解の記述が妥当（重なりが低いとずれる？）

今後も調べるべき点はありそう…