Convergence rate of Bayesian tensor estimator and its minimax optimality

†‡鈴木　大慈

† 東京工業大学大学院情報理工学研究科 数理・計算科学専攻
‡JST さきがけ

2015 年 8 月 4 日
ERATO 感謝祭 SeasonII@一橋講堂

Taiji Suzuki: Convergence rate of Bayesian tensor estimator and its minimax optimality. The 32nd International Conference on Machine Learning (ICML2015), *JMLR Workshop and Conference Proceedings* 37:pp. 1273–1282, 2015.

Outline

本日のお題

低ランクテンソル推定
→ ベイズ推定量の性質

緩い仮定での最適な収束レート

$$\tilde{O}\left(\frac{d(M_1 + \cdots + M_K)}{n}\right)$$

例: 推薦システム

	映画 A	映画 B	映画 C	···	映画 X
ユーザ 1	4	8	*	···	2
ユーザ 2	2	*	2	···	*
ユーザ 3	2	4	*	···	*
⋮					

(e.g., Srebro et al. (2005a), NetFlix (Bennett & Lanning, 2007))

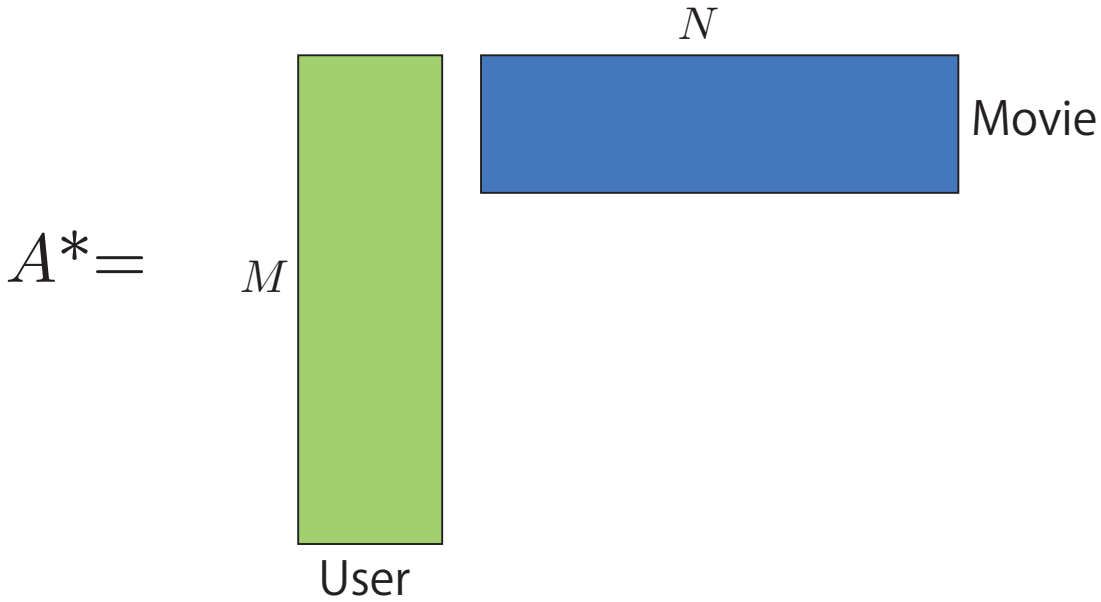
例: 推薦システム

ランク 1 と仮定する

	映画 A	映画 B	映画 C	···	映画 X
ユーザ 1	4	8	4	···	2
ユーザ 2	2	4	2	···	1
ユーザ 3	2	4	2	···	1
⋮					

(e.g., Srebro et al. (2005a), NetFlix (Bennett & Lanning, 2007))

例: 推薦システム



$$A^*_{ij} = \sum_{r=1}^{d} U_{ir} V_{jr}$$

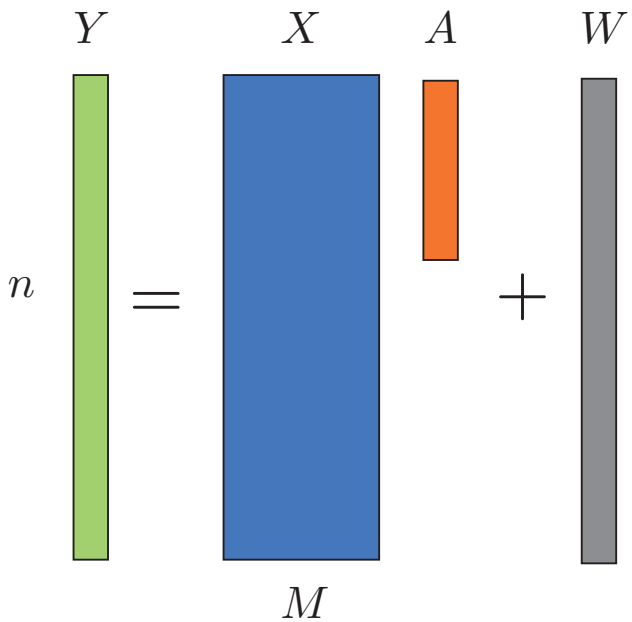
→ 低ランク行列補完:

- 低ランク行列の Rademacher Complexity: Srebro et al. (2005a).
- Compressed sensing: Candès and Tao (2009); Candès and Recht (2009).

例: 縮小ランク回帰

- 縮小ランク回帰 (Anderson, 1951; Burket, 1964; Izenman, 1975)
- マルチタスク学習 (Argyriou et al., 2008)

普通の回帰

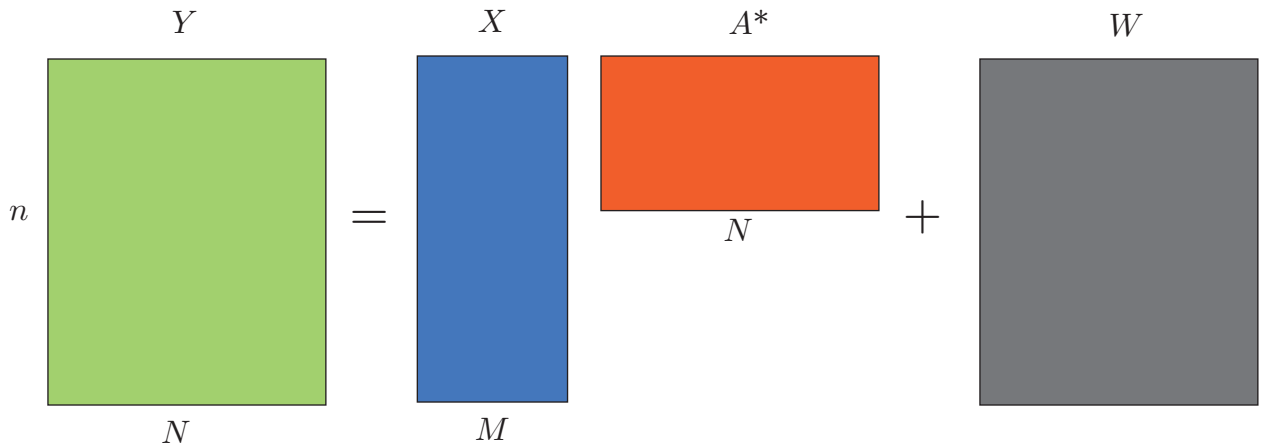


スパース推定では A がスパース (多くの成分が 0).

例: 縮小ランク回帰

- 縮小ランク回帰 (Anderson, 1951; Burket, 1964; Izenman, 1975)
- マルチタスク学習 (Argyriou et al., 2008)

縮小ランク回帰



A^* は 低ランク

高次元データでの問題意識

- 協調フィルタリング
- コンピュータビジョン
- 音声認識
- ゲノムデータ
- 金融データ

次元 $d = 10000$ の時，サンプルサイズ $n = 1000$ で推定ができるか？
どのような条件があれば推定が可能か？

- 何らかの**低次元性**を利用.

高次元データでの問題意識

- 協調フィルタリング
- コンピュータビジョン
- 音声認識
- ゲノムデータ
- 金融データ

次元 $d = 10000$ の時，サンプルサイズ $n = 1000$ で推定ができるか？
どのような条件があれば推定が可能か？

- 何らかの**低次元性**を利用.

→ 今日は**低ランク性**を扱う.

Outline

低ランク行列推定

低ランク行列推定の基本形

$$Y_i = \langle X_i, A^* \rangle + W_i$$

$A^*, X_i \in \mathbb{R}^{M \times N}$: 行列, $\langle X_i, A^* \rangle := \mathrm{Tr}[X_i^\top A^*]$
$W_i \sim \mathcal{N}(0, \sigma^2)$: 観測ノイズ

観測量: $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$
推定したいもの: A^*

低ランク行列推定

低ランク行列推定の基本形

$$Y_i = \langle X_i, A^* \rangle + W_i$$

$A^*, X_i \in \mathbb{R}^{M \times N}$: 行列, $\langle X_i, A^* \rangle := \mathrm{Tr}[X_i^\top A^*]$
$W_i \sim \mathcal{N}(0, \sigma^2)$: 観測ノイズ

観測量: $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$
推定したいもの: A^*

$M \times N \gg n$: 高次元データ (不良設定問題)
$\rightarrow A^*$ は 低ランク であるとして推定する.

ランク正則化付きリスク最小化

正則化付き推定法の基本的な考え方:

$$\min_{A \in \mathbb{R}^{M \times N}} \quad \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle X_i, A \rangle)^2 + \mathrm{Pen}(\mathrm{rank}(A)).$$

しかし,

- rank 関数は行列 A に対して凸関数ではない.
- ランク制約を満たす行列の集合は凸集合を成さない.

→ 凸緩和 (トレースノルム正則化)

トレースノルム正則化

$$\min_{A\in\mathbb{R}^{M\times N}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle X_i, A\rangle)^2 + \lambda\|A\|_{\mathrm{tr}},$$

$$\|A\|_{\mathrm{tr}} := \mathrm{Tr}(\sqrt{A^\top A}) = \sum_{i=1}^{d}\sigma_i(A).$$

※トレースノルムは $\{A \mid \|A\|_{\mathrm{sp}} \le 1\}$ 上でランク関数の凸包絡.

これは，特異値への ℓ_1 正則化.
→ 特異値がスパース $=$ 低ランク.

(Srebro et al., 2005b; Argriou et al., 2008; Argyriou et al., 2008)

Lasso:

$$\min_{\beta\in\mathbb{R}^{M}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i^\top\beta)^2 + \lambda\|\beta\|_{\ell_1}.$$

トレースノルム正則化の理論

真の行列 A^* は $M \times N$ のランク d 行列.

ノイズあり:

- スパース推定の一般論: Negahban et al. (2012).
- 行列補完, Spikiness 条件: Negahban and Wainwright (2012).
- 行列補完+マルチタスク学習: Rohde and Tsybakov (2011).
- 対称行列の行列補完: Koltchinskii (2012).

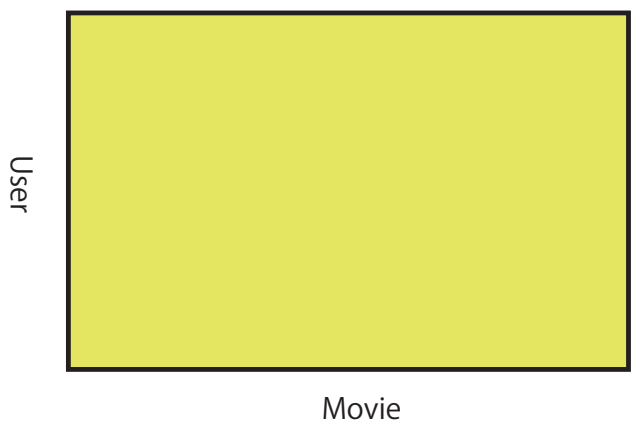
$$\|\hat{A} - A^*\|_2^2 = O_p\left(\frac{d(M+N)\log(MN)}{n}\right).$$

ただし，$\|A\|_2 := \sqrt{\frac{1}{MN}}\|A\|_{\mathrm{F}}$.

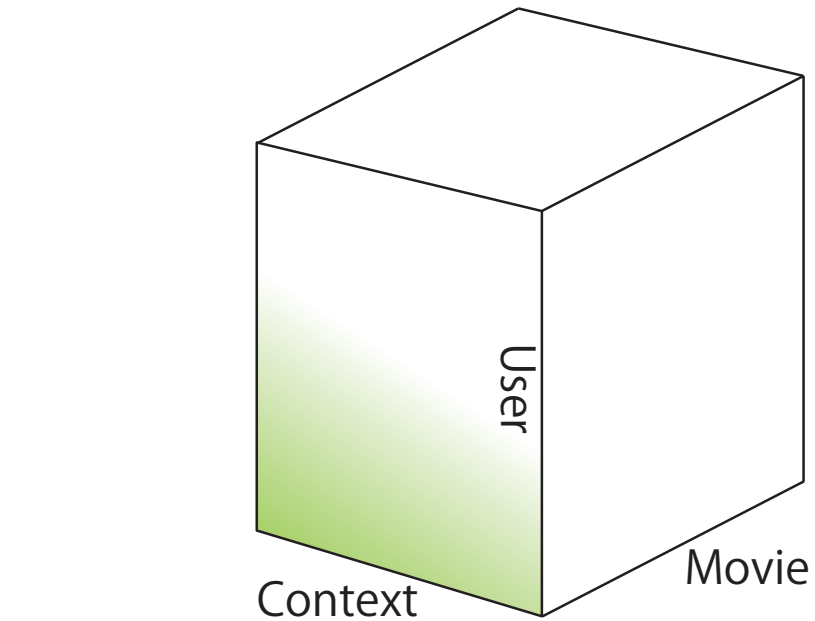
基本的に観測デザインへの条件（制限強凸性）が必要.

Outline

テンソルデータ



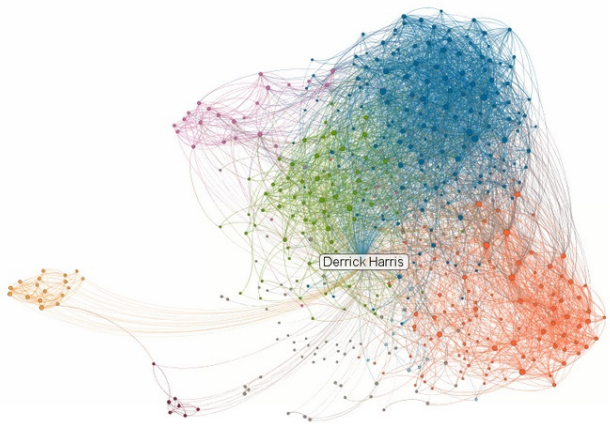
行列

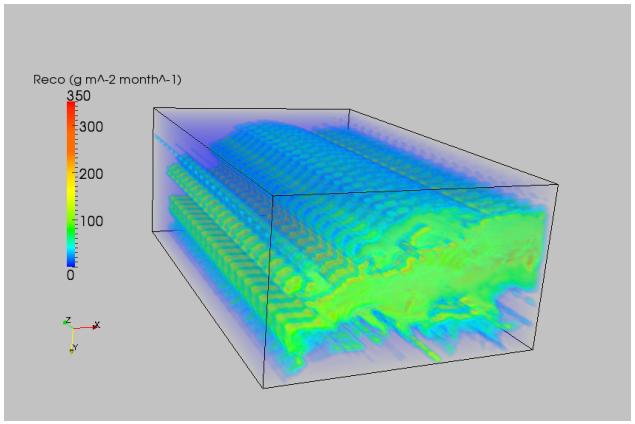


テンソル (多次元アレイ)

Applications

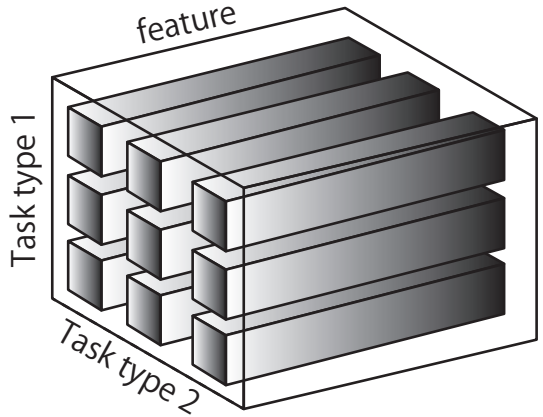
- 推薦システム
- 関係データ
- マルチタスク学習
- 時空間データ解析 (空間 (2D) × 時間)
- 動画像処理 (画像 (2D) × 時間)



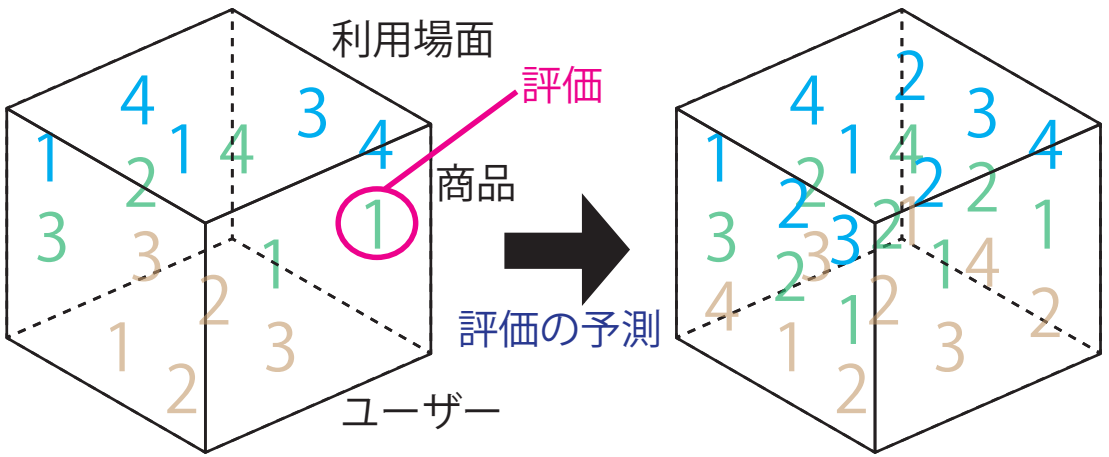


Applications

- マルチタスク学習



- 推薦システム（テンソル補完）



問題設定: 回帰

回帰の問題設定:

$$Y_i = \langle X_i, \mathcal{A}^* \rangle + W_i.$$

$\mathcal{A}^*, X_i \in \mathbb{R}^{M_1 \times \dots M_K}$: テンソル.
$\langle X_i, \mathcal{A}^* \rangle := \sum_{j_1,\dots,j_K} X_{i,(j_1,\dots,j_K)} \mathcal{A}^*_{j_1,\dots,j_K}$.
$W_i \sim \mathcal{N}(0, \sigma^2)$: 観測ノイズ.

仮定: $\mathcal{A}^*$ は "低ランク".

問題設定: 回帰

回帰の問題設定:

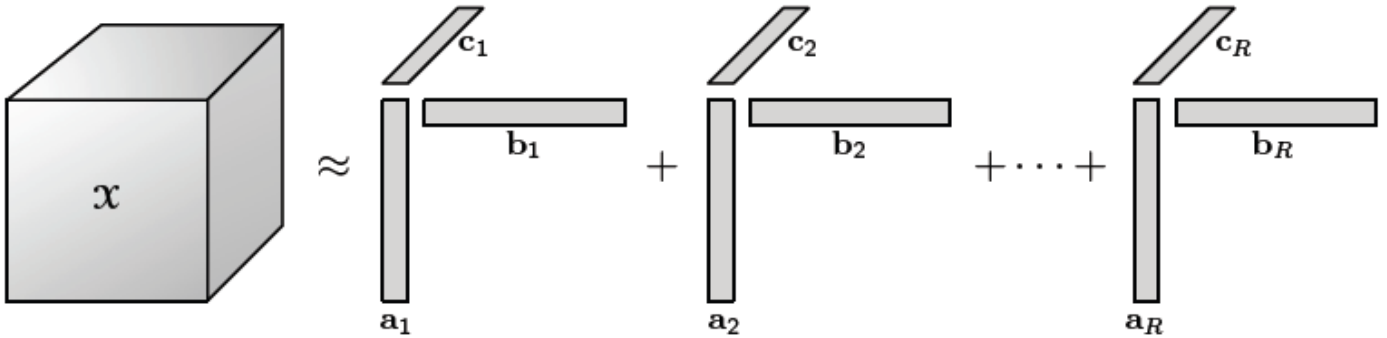
$$Y_i = \langle X_i, \mathcal{A}^* \rangle + W_i.$$

$\mathcal{A}^*, X_i \in \mathbb{R}^{M_1 \times \dots M_K}$: テンソル.
$\langle X_i, \mathcal{A}^* \rangle := \sum_{j_1,\dots,j_K} X_{i,(j_1,\dots,j_K)} \mathcal{A}^*_{j_1,\dots,j_K}$.
$W_i \sim \mathcal{N}(0, \sigma^2)$: 観測ノイズ.

仮定: $\mathcal{A}^*$ は "低ランク".

E.g., $X_i = e_{j_1} \otimes e_{j_2} \otimes \cdots \otimes e_{j_K}$ でテンソル補完問題.

テンソルのランク: CP-ランク



CP-分解 (Canonical Polyadic Decomp.)
(Hitchcock, 1927; Hitchcock, 1927)

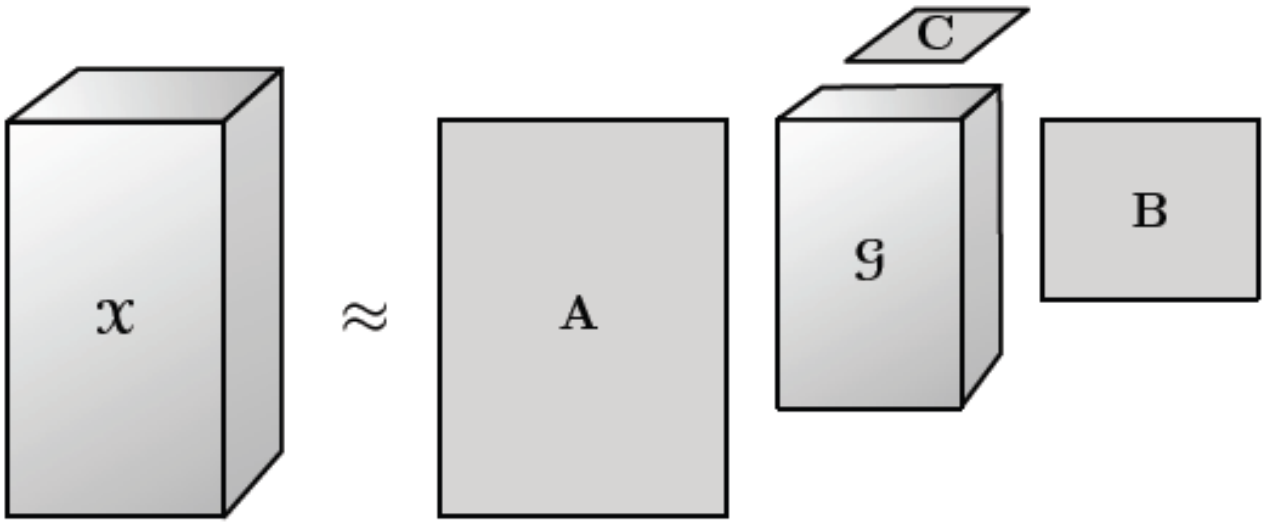
(figure is from Kolda and Bader (2009))

$$\mathcal{X}_{ijk} = \sum_{r=1}^{d} a_{ir} b_{jr} c_{kr} =: [[A, B, C]].$$

CP-分解はテンソルの CP ランクを定義する.

- CP-分解は **NP**-困難.
- CP-ランクはテンソルの辺の長さより大きくなる可能性がある.
- 直交分解が存在するとは限らない (対称テンソルに限っても).

テンソルのランク: Tucker-ランク



Tucker-分解 (Tucker, 1966)

(figure is from Kolda and Bader (2009))

$\mathcal{X}_{ijk} = \sum_{l=1}^{r_1} \sum_{m=1}^{r_2} \sum_{n=1}^{r_3} g_{lmn} a_{il} b_{jm} c_{kn} =: [[G; A, B, C]]$.

- G はコアテンソルとよばれている.
- Tucker-ランク $= (r_1, r_2, r_3)$
- r_1, r_2, r_3 はテンソルの辺の長さより大きくはならない.

凸最適化によるテンソル推定方法

凸最適化によるアプローチ:

$$\min_{\mathcal{A}\in\mathbb{R}^{M_1\times M_2\times\cdots\times M_K}} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle X_i, \mathcal{A}\rangle)^2 + \mathrm{pen}(\mathcal{A}).$$

- 和型 Schatten-1 ノルム (Tomioka et al., 2011)
- 畳み込み型 Schatten-1 ノルム (Tomioka & Suzuki, 2013)
- Square deal (Mu et al., 2014)

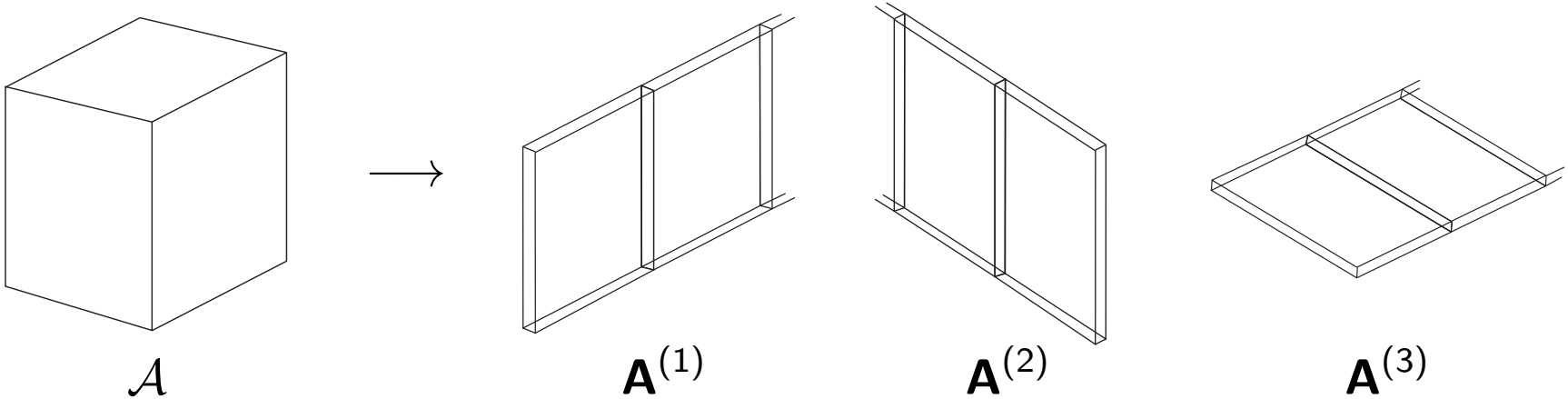
和型 Schatten-1 ノルム

$$\|\mathcal{A}\|_{S_1/1} := \sum_{k=1}^{K} \|\mathbf{A}^{(k)}\|_{\mathrm{Tr}}$$

和型 Schatten-1 ノルム正則化

$$\widehat{\mathcal{A}} = \arg\min_{\mathcal{A}} \|\mathcal{Y} - \mathcal{A}\|_F^2 + \lambda_n \|\mathcal{A}\|_{S_1/1}.$$

Tucker-ランクが小さなテンソルを推定.



畳み込み型 Schatten-1 ノルム

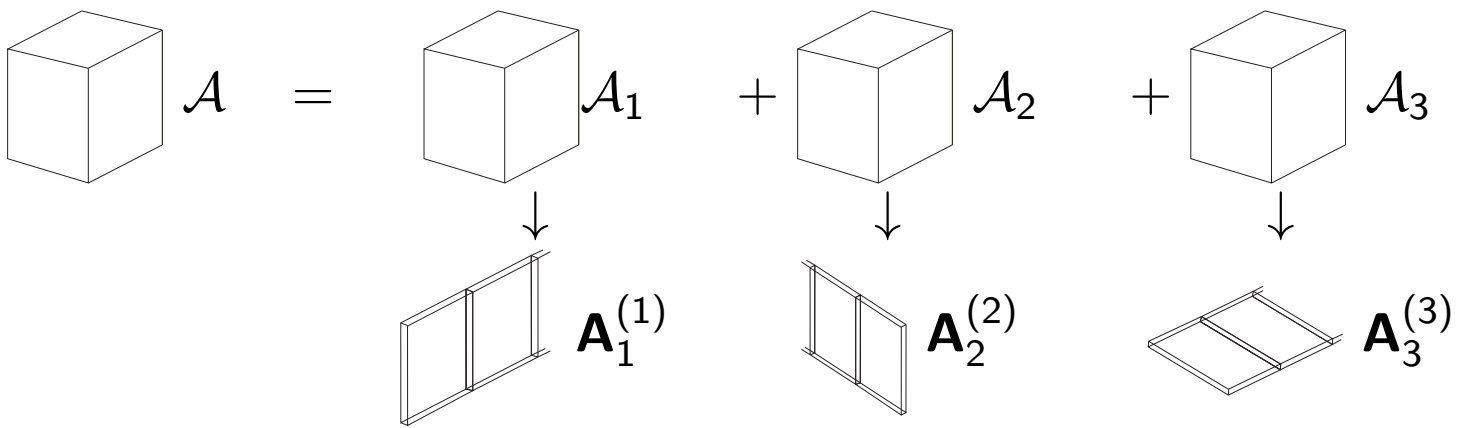
$$\|\mathcal{A}\|_{\underline{S_1/1}} := \inf_{\mathcal{A}=\mathcal{A}_1+\mathcal{A}_2+\cdots+\mathcal{A}_K} \sum_{k=1}^{K} \|\mathbf{A}_k^{(k)}\|_{\mathrm{Tr}}$$

畳み込み型 Schatten1 ノルム正則化

$$\widehat{\mathcal{A}} = \arg\min_{\mathcal{A}} \quad \|\mathcal{Y} - \mathcal{A}\|_F^2 + \lambda_n \|\mathcal{A}\|_{\underline{S_1/1}},$$

$$\text{s.t.} \quad \mathcal{A} = \sum_{k=1}^{K} \mathcal{A}_k, \|\mathbf{A}_k^{(k')}\|_{s_\infty} \leq \frac{\alpha}{K}\sqrt{N/n_{k'}} \ (\forall k' \neq k).$$

ランクの小さな方向を見つける.



収束レート解析

$\mathcal{A}^*$ のモード k 展開のランクが r_k であるとする: $r_k = \mathrm{rank}(\mathbf{A}^{*(k)})$.
$N = M_1 \times \cdots \times M_K$.

収束レート

- 和型 (Tomioka et al., 2011)
  $\frac{1}{N}\|\|\widehat{\mathcal{A}} - \mathcal{A}^*\|\|_F^2 \leq \frac{C}{n}\left(\frac{1}{K}\sum_{k=1}^K(\sqrt{M_k} + \sqrt{N/M_k})\right)^2 \left(\frac{1}{K}\sum_{k=1}^K \sqrt{r_k}\right)^2$
- 畳み込み型 (Tomioka & Suzuki, 2013)
  $\frac{1}{N}\|\|\widehat{\mathcal{A}} - \mathcal{A}^*\|\|_F^2 \leq \frac{C}{n}\left(\max_k(\sqrt{M_k} + \sqrt{N/M_k})\right)^2 \min_k r_k$
- Square deal (Mu et al., 2014):
  $\frac{1}{M}\|\hat{A} - A^*\|_2^2 \leq C\frac{\min\{\prod_{k\in I_1} r_k, \prod_{k\in I_2} r_k\}}{n}\left(\prod_{k\in I_1} M_k + \prod_{k\in I_2} M_k\right)$,
  where I_1 and I_2 are any disjoint decomposition of index set $\{1, \ldots, K\}$.

- これらは最適ではない.
- ある種の強凸性の仮定が必要.

ベイズ推定

ベイズ法で低ランクテンソル推定を行った時の統計的性質を調べる.

- 一般の回帰の設定で解析.
- より少ない条件で最適レートを達成.

事前分布

ランク d' と $U^{(k)}$ に事前分布 (データを見る前の確からしさ) を置く.

- **CP-ランクが d' のテンソル上の事前分布:**
  ランク d' なるテンソル $\mathcal{A}$ の分解
  $\mathcal{A} = [[U^{(1)}, \ldots, U^{(K)}]]$ $(U^{(k)} \in \mathbb{R}^{d' \times M_k})$ と分解されるとして, 事前分布を次のように入れる :

$$\pi(U^{(1)}, \ldots, U^{(K)} | d') \propto \exp\left\{ -\frac{d'}{2\sigma_{\mathrm{p}}^2} \sum_{k=1}^{K} \mathrm{Tr}[{U^{(k)}}^{\top} U^{(k)}] \right\}.$$

- **ランクの事前分布:**

$$\pi(d') \propto \xi^{d'} \qquad (\sum_{d'} \pi(d') = 1),$$

  ただし $1 > \xi > 0$ はある定数.

事後分布

データ $D_n = \{(X_i, Y_i)\}_{i=1}^n$ を観測した後の確からしさ.

$$\underbrace{\pi(\mathcal{A}, d'|D_n)}_{\text{事後分布}} \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \langle X_i, \mathcal{A}\rangle)^2\right] \times \pi(\mathcal{A}|d')\pi(d')$$

尤度（データへの当てはまり） × 事前分布

ベイズ推定量は Gibbs サンプリングで計算可能:

- 条件付き事後分布 $\pi(U^{(k)}|\{U^{(k')}\}_{k'\neq k}, d', D_n)$ は正規分布.
- 条件付き事後分布からのサンプリングを $k = 1, \ldots, K$ とサイクリックに繰り返す.
- ランク d' の事後確率は周辺尤度を Gibbs サンプリングを通して計算.

関連研究

- 確率テンソルの推定 (多項分布): Zhou et al. (2013)
- 離散カテゴリ，判別分析 (条件付き多項分布): Yang and Dunson (2013)
- テンソル積空間上でのノンパラメトリック確率密度関数推定: Shen and Ghosal (2014)
- 低ランク行列のベイズ推定の収束レート (回帰): Alquier (2013)
- テンソルのベイズモデルおよび計算方法: Xiong et al. (2010); Xu et al. (2013); Rai et al. (2014)

用語の準備

- $\mathcal{A}^*$ の *max-norm*:

$$\|\mathcal{A}^*\|_{\max,2} := \min_{\{U^{(k)}\}} \{\max_{i,k} \|U^{(k)}_{:,i}\| \mid \mathcal{A}^* = [[U^{(1)}, \ldots, U^{(K)}]],\ U^{(k)} \in \mathbb{R}^{d \times M_k}\}.$$

- 固定デザイン・ランダムデザインのリスク:

固定デザイン:

$$\|\mathcal{A} - \mathcal{A}^*\|_n^2 := \frac{1}{n} \sum_{i=1}^{n} \langle X_i, \mathcal{A} - \mathcal{A}^* \rangle^2.$$

ランダムデザイン:

$$\|\mathcal{A} - \mathcal{A}^*\|_{L_2}^2 := \mathrm{E}_{X \sim P_X}[\langle X, \mathcal{A} - \mathcal{A}^* \rangle^2].$$

固定デザインとランダムデザインの違い

固定デザイン: $\|\mathcal{A} - \mathcal{A}^*\|_n^2 := \left(\frac{1}{n}\sum_{i=1}^n \langle X_i, \mathcal{A} - \mathcal{A}^*\rangle^2\right)^{\frac{1}{2}}$.

$$\underbrace{\begin{bmatrix} 4 & 8 & * \\ 2 & * & 2 \\ 2 & 4 & * \end{bmatrix}}_{\text{true value}} + \underbrace{W}_{\text{noise}} = \begin{bmatrix} 5.5 & 9 & * \\ 1 & * & 3 \\ 4 & 6 & * \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 8 & * \\ 2 & * & 2 \\ 2 & 4 & * \end{bmatrix}$$

ランダムデザイン: $\|\mathcal{A} - \mathcal{A}^*\|_{L_2}^2 := \mathrm{E}_{X\sim P_X}[\langle X, \mathcal{A} - \mathcal{A}^*\rangle^2]^{\frac{1}{2}}$.

$$\underbrace{\begin{bmatrix} 4 & 8 & * \\ 2 & * & 2 \\ 2 & 4 & * \end{bmatrix}}_{\text{true value}} + \underbrace{W}_{\text{noise}} = \begin{bmatrix} 5.5 & 9 & * \\ 1 & * & 3 \\ 4 & 6 & * \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 8 & 4 \\ 2 & 4 & 2 \\ 2 & 4 & 2 \end{bmatrix}$$

ベイズ事後分布の収束 (posterior 1)

$\mathcal{A}^*$: CP-ランク d.
$\|\mathcal{A}^*\|_{\max,2} := \min_{\{U^{(k)}\}}\{\max_{i,k} \|U^{(k)}_{:,i}\| \mid \mathcal{A}^* = [[U^{(1)}, \ldots, U^{(K)}]],\ U^{(k)} \in \mathbb{R}^{d \times M_k}\}$. (c.f., max-norm (Srebro & Shraibman, 2005))
$L = \|\mathcal{A}^*\|_{\max,2}$ とする.

Theorem

ある $n, \{M_k\}_k$ とは関係ない定数 C が存在して,

$$\mathrm{E}_{Y_{1:n}|X_{1:n}}\left[\frac{1}{2\sigma^2}\int \|\mathcal{A}-\mathcal{A}^*\|_n^2 \mathrm{d}\Pi(\mathcal{A}|X_{1:n}, Y_{1:n})\right]$$
$$\leq C\frac{d(\sum_{k=1}^K M_k)}{n}(\tfrac{L}{\sigma_{\mathrm{p}}} \vee 1)^2 \log\left(\frac{\sigma_{\mathrm{p}}^K}{\xi} K\sqrt{n(\textstyle\sum_{k=1}^K M_k(\tfrac{L}{\sigma_{\mathrm{p}}} \vee 1)^2)^K}\right).$$

- このレートはデザイン行列に 強凸性を何も仮定せず に得られる.
- ランク d は事前に知っている必要はない. 自動的に調節.

ベイズ事後分布の収束 (posterior 1)

$\mathcal{A}^*$: CP-ランク d.
$\|\mathcal{A}^*\|_{\max,2} := \min_{\{U^{(k)}\}}\{\max_{i,k} \|U^{(k)}_{:,i}\| \mid \mathcal{A}^* = [[U^{(1)}, \ldots, U^{(K)}]],\ U^{(k)} \in \mathbb{R}^{d \times M_k}\}$. (c.f., max-norm (Srebro & Shraibman, 2005))
$L = \|\mathcal{A}^*\|_{\max,2}$ とする.

Theorem

ある $n, \{M_k\}_k$ とは関係ない定数 C が存在して,

$$\mathrm{E}_{Y_{1:n}|X_{1:n}}\left[\frac{1}{2\sigma^2}\left\|\int \mathcal{A}\mathrm{d}\Pi(\mathcal{A}|X_{1:n}, Y_{1:n}) - \mathcal{A}^*\right\|_n^2\right]$$
$$\leq C\frac{d(\sum_{k=1}^K M_k)}{n}(\tfrac{L}{\sigma_{\mathrm{p}}} \vee 1)^2 \log\left(\tfrac{\sigma_{\mathrm{p}}^K}{\xi} K\sqrt{n(\textstyle\sum_{k=1}^K M_k(\tfrac{L}{\sigma_{\mathrm{p}}} \vee 1)^2)^K}\right).$$

- このレートはデザイン行列に 強凸性を何も仮定せず に得られる.
- ランク d は事前に知っている必要はない. 自動的に調節.

ベイズ事後分布の収束 (posterior 2)

$\|\mathcal{A}^*\|_{\max,2} \leq R\sigma_{\mathrm{p}}$ を仮定.
(真のテンソルは事前分布の台に含まれている)

Theorem

ある $n, \{M_k\}_k$ とは関係ない定数 C が存在して,

$$\mathrm{E}_{Y_{1:n},X_{1:n}}\left[\frac{1}{2\sigma^2}\int \|\mathcal{A}-\mathcal{A}^*\|_n^2 \mathrm{d}\Pi(\mathcal{A}|\|\mathcal{A}\|_{\max,2} \leq R, X_{1:n}, Y_{1:n})\right]$$
$$\leq C\frac{d(\sum_{k=1}^K M_k)}{n}(1 \vee R^2)\log\left(K\frac{\sigma_{\mathrm{p}}^K}{\xi}\sqrt{nR^K}\right).$$

log 内の値が改善されている.

ランダムデザインにおけるベイズ事後分布の収束

$\|A^*\|_{\max,2} \leq R\sigma_{\mathrm{p}}$ を仮定.
(真のテンソルは事前分布の台に含まれている)

Theorem

ある $n, \{M_k\}_k$ とは関係ない定数 C が存在して,

$$\mathrm{E}_{Y_{1:n},X_{1:n}}\left[\frac{1}{2\sigma^2}\int \|\mathcal{A}-\mathcal{A}^*\|_{L_2}^2 \mathrm{d}\Pi(\mathcal{A}|\|\mathcal{A}\|_{\max,2} \leq R, X_{1:n}, Y_{1:n})\right]$$
$$\leq C\frac{d(\sum_{k=1}^K M_k)}{n}\left(1 \vee R^{2(K+1)}\right)\log\left(K\frac{\sigma_{\mathrm{p}}^K}{\xi}\sqrt{nR^K}\right).$$

ランダムデザインでも ほぼ同じ収束レートを達成.

ミニマックス最適性

- ランク d，max-ノルム R 以下の集合:

$$\mathcal{H}_d(R) := \left\{ \mathcal{A} \in \mathbb{R}^{M_1 \times \cdots \times M_K} \mid \text{CP-ランク } d, \|\mathcal{A}\|_{\max,2} \leq R \right\}.$$

- $X_i = e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_K}$ で，$(i_1, \ldots, i_K)$ は一様分布.

Theorem

$\mathcal{H}_d(R)$ 上のテンソル推定量のミニマックス最適リスクは以下で抑えられる:

$$\min_{\widehat{\mathcal{A}}} \max_{\mathcal{A}^* \in \mathcal{H}_d(R)} \mathrm{E}[\|\widehat{\mathcal{A}} - \mathcal{A}^*\|_{L_2}^2] \gtrsim \min \left\{ \sigma^2 \frac{d(M_1 + \cdots + M_K)}{n}, R^{2K} \right\}.$$

※ 先の収束レートは log 項を除いてミニマックス最適レートを達成.

凸正則化手法との比較

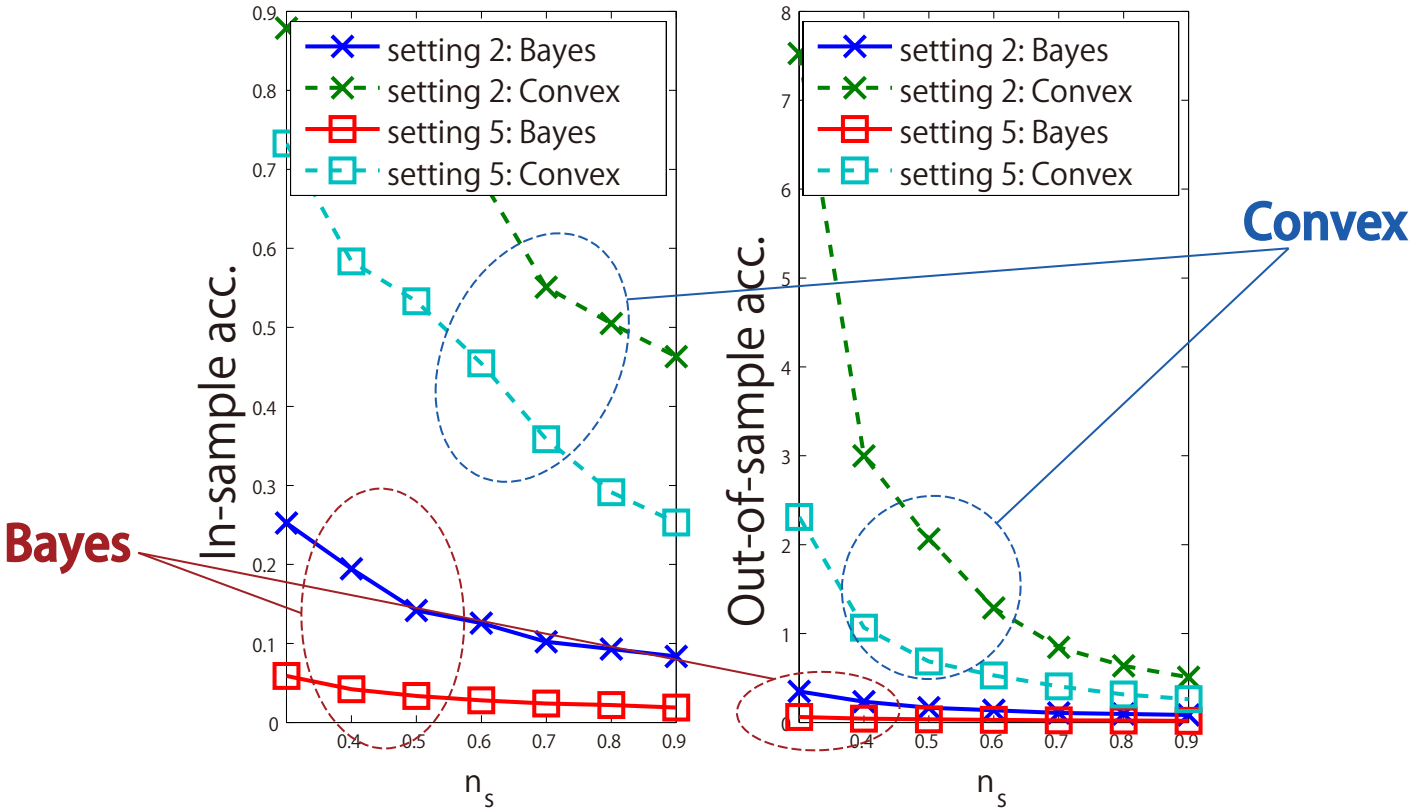


Figure: ベイズ推定法と凸正則化法 (和型 Schatten-1 ノルム) との比較.

誤差のスケール

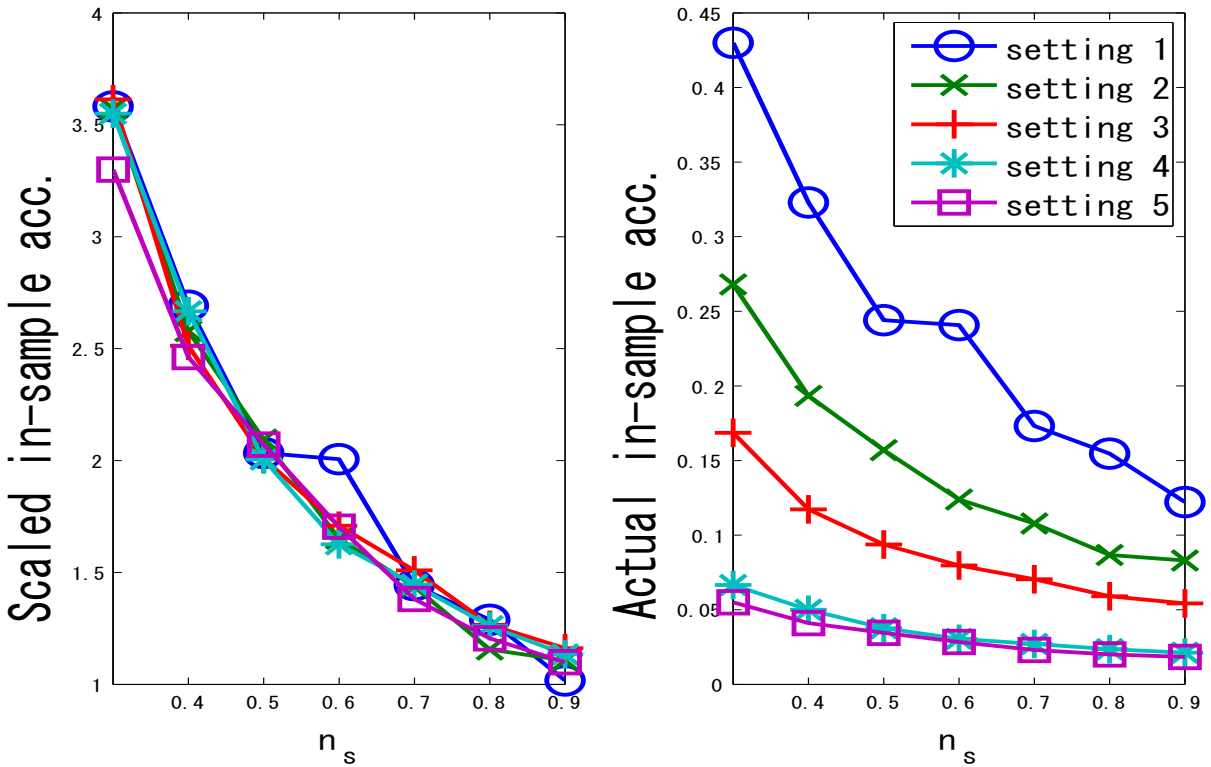


Figure: 固定デザイン: The scaled predictive accuracy (left) and the actual predictive accuracy (right) against the number of samples.

$$\text{scaled accuracy} = \frac{\text{actual accuracy}}{d(\sum_{k=1}^{K} M_k)}$$

誤差のスケール

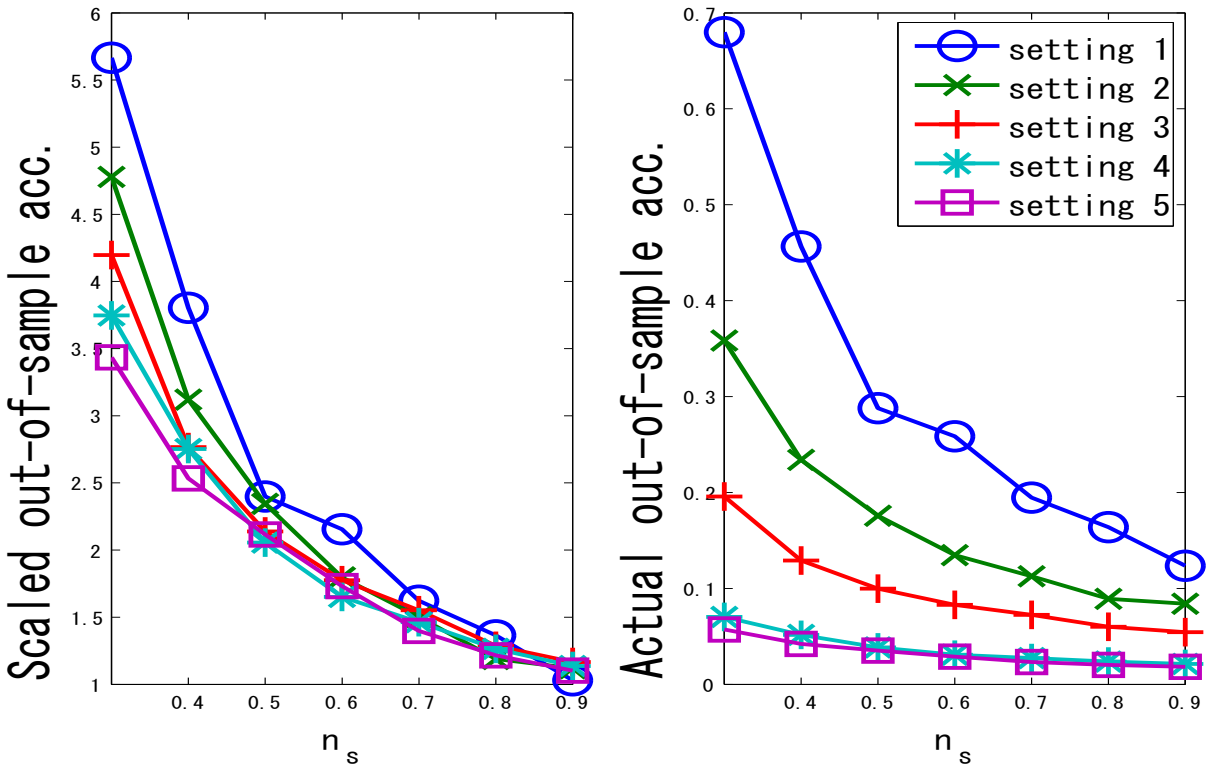


Figure: ランダムデザイン: The scaled predictive accuracy (left) and the actual predictive accuracy (right) against the number of samples.

$$\text{scaled accuracy} = \frac{\text{actual accuracy}}{d(\sum_{k=1}^{K} M_k)}$$

まとめ

- 低ランクテンソル推定における事後分布の収束レートを導出.
- 推定誤差のミニマックス最適レートを導出.
- 導出した上界は (ほぼ) ミニマックス最適であることを証明.

Posterior contraction rate

$$\|\widehat{\mathcal{A}} - \mathcal{A}^*\|_n^2 \leq C \frac{d(M_1 + \cdots + M_K)}{n} \log(K\sqrt{nR^K}).$$

Minimax rate

$$\min_{\widehat{\mathcal{A}}} \max_{\mathcal{A}^* \in \mathcal{H}_d(R)} \mathrm{E}[\|\widehat{\mathcal{A}} - \mathcal{A}^*\|_{L_2}^2] \gtrsim \min\left\{\sigma^2 \frac{d(M_1 + \cdots + M_K)}{n}, R^{2K}\right\}.$$

Alquier, P. (2013). Bayesian methods for low-rank matrix estimation: Short survey and theoretical study. *Algorithmic Learning Theory* (pp. 309–323). Springer-Verlag.

Anderson, T. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327–351.

Argriou, A., Evgeniou, T., & Pontil, M. (2008). Convex multi-task feature learning. *Machine Learning*, *73*, 243–272.

Argyriou, A., Micchelli, C. A., Pontil, M., & Ying, Y. (2008). A spectral regularization framework for multi-task structure learning. *Advances in Neural Information Processing Systems 20* (pp. 25–32). Cambridge, MA: MIT Press.

Bennett, J., & Lanning, S. (2007). The netflix prize. *Proceedings of KDD Cup and Workshop 2007*.

Burket, G. R. (1964). *A study of reduced-rank models for multiple prediction*, vol. 12 of *Psychometric monographs*. Psychometric Society.

Candès, E., & Tao, T. (2009). The power of convex relaxations: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, *56*, 2053–2080.

Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, *9*, 717–772.

Hitchcock, F. L. (1927). Multilple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, *7*, 39–79.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 248–264.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*, 455–500.

Koltchinskii, V. (2012). *Sharp oracle inequalities in low rank estimation* (Technical Report). arXiv:1210.1144.

Mu, C., Huang, B., Wright, J., & Goldfarb, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. *Proceedings of the 31th International Conference on Machine Learning* (pp. 73–81).

Negahban, S., Ravikumar, P., Wainwright, M. J., & Yu, B. (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science*, *27*, 538–557.

Negahban, S., & Wainwright, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, *13*, 1665–1697.

Rai, P., Wang, Y., Guo, S., Chen, G., Dunson, D., & Carin, L. (2014). Scalable Bayesian low-rank decomposition of incomplete multiway tensors. *Proceedings of the 31th International Conference on Machine Learning* (pp. 1800–1808).

Rohde, A., & Tsybakov, A. B. (2011). Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, *39*, 887–930.

Shen, W., & Ghosal, S. (2014). Adaptive bayesian density regression for high-dimensional data. arXiv:1403.2695.

Srebro, N., Alon, N., & Jaakkola, T. (2005a). Generalization error bounds for collaborative prediction with low-rank matrices. *Advances in Neural Information Processing Systems (NIPS) 17*.

Srebro, N., Rennie, J., & Jaakkola, T. (2005b). Maximum margin matrix factorization. *Advances in Neural Information Processing Systems 17* (pp. 1329–1336). MIT Press.

Srebro, N., & Shraibman, A. (2005). Rank, trace-norm and max-norm. *Proceedings of the 18th Annual Conference on Learning Theory* (pp. 545–560). Springer-Verlag.

Tomioka, R., & Suzuki, T. (2013). Convex tensor decomposition via structured schatten norm regularization. *Advances in Neural Information Processing Systems 26* (pp. 1331–1339). NIPS2013.

Tomioka, R., Suzuki, T., Hayashi, K., & Kashima, H. (2011). Statistical performance of convex tensor decomposition. *Advances in Neural Information Processing Systems 24* (pp. 972–980). NIPS2011.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

Xiong, L., Chen, X., Huang, T.-K., Schneider, J., & Carbonell, J. G. (2010). Temporal collaborative filtering with bayesian probabilistic tensor factorization. *Proceedings of SIAM Data Mining* (pp. 211–222).

Xu, Z., Yan, F., & Qi, Y. (2013). Bayesian nonparametric models for multiway data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *99*, 1. PrePrints.

Yang, Y., & Dunson, D. B. (2013). Bayesian conditional tensor factorizations for high-dimensional classification. arXiv:1301.4950.

Zhou, J., Bhattacharya, A., Herring, A., & Dunson, D. (2013). Bayesian factorizations of big sparse tensors. arXiv:1306.1598.