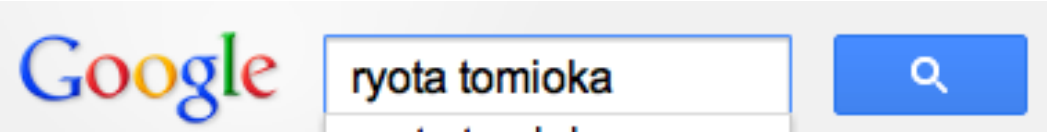
凸最適化に基づく
テンソル分解アルゴリズム

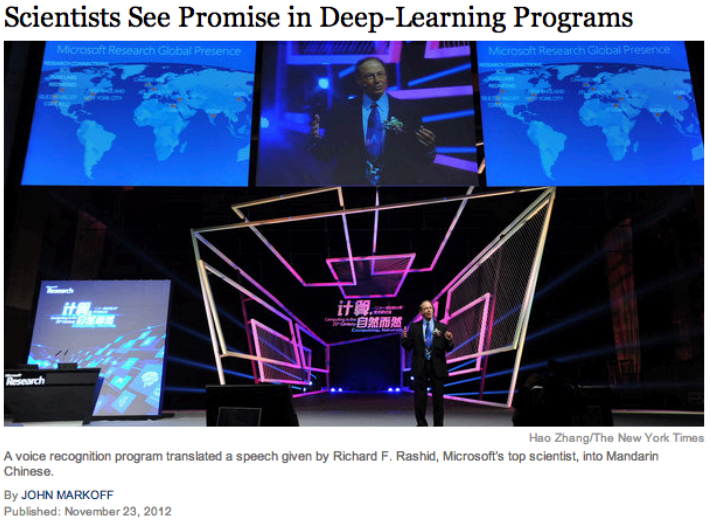
冨岡 亮太
tomioka@mist.i.u-tokyo.ac.jp

共同研究者：鈴木大慈、林浩平、鹿島久嗣
東京大学 大学院情報理工学系研究科 数理情報学専攻

自己紹介



- 専門：
  - 統計的機械学習 / データマイニング / 数理計画
- 機械学習？
  - 多種多量のデータを解析するための方法論
    - 例：ビジネス、医療、バイオ



Scientists See Promise in Deep-Learning Programs

Hao Zhang/The New York Times

A voice recognition program translated a speech given by Richard F. Rashid, Microsoft's top scientist, into Mandarin Chinese.

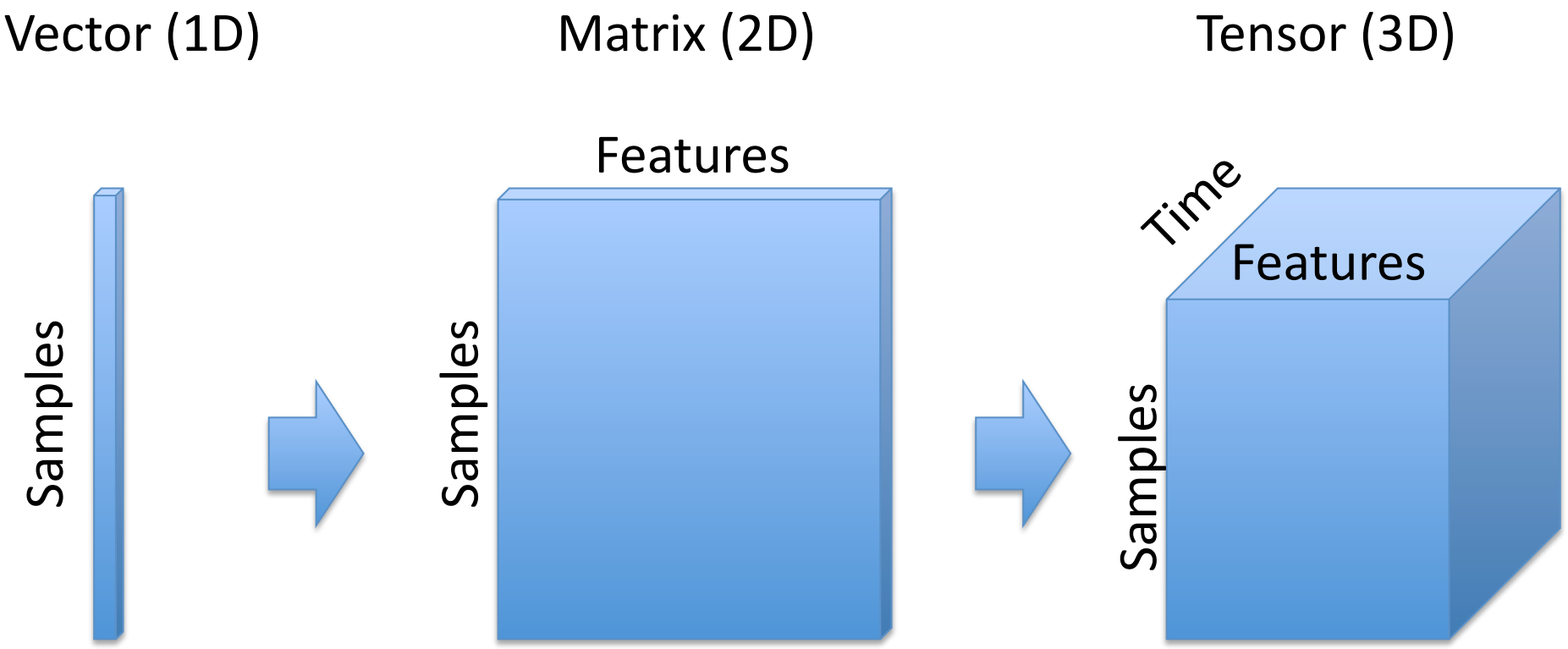
By JOHN MARKOFF
Published: November 23, 2012

New York Times 2012/11/24

- 今回のミニシンポジウムとの関係
  - テンソルの構造を持つデータの解析

テンソル型データ

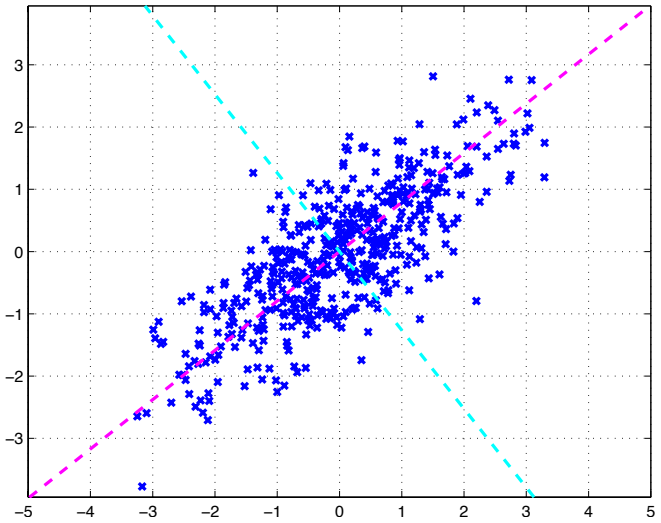
- 行列型データの拡張として捉えられる



More dimensions: space, conditions, etc.

テンソル型データで何をしたいか

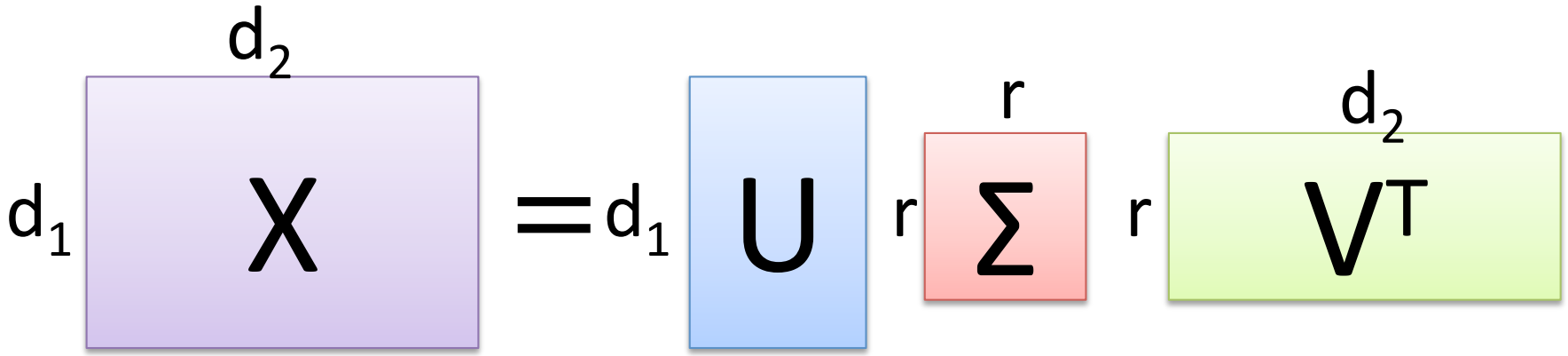
- ノイズに埋もれた低ランク構造を復元したい。
  - 主成分分析の多次元拡張



- 欠損値を復元したい。
  - 例：いくつかのセンサーが壊れている
  - レコメンデーション
    （e.g., Amazon, Facebook）

Tucker decomposition (HOSVD) / CP decomposition

復習：特異値分解 (SVD)



where U, V: Orthogonal ($U^TU=I$, $V^TV=I$)

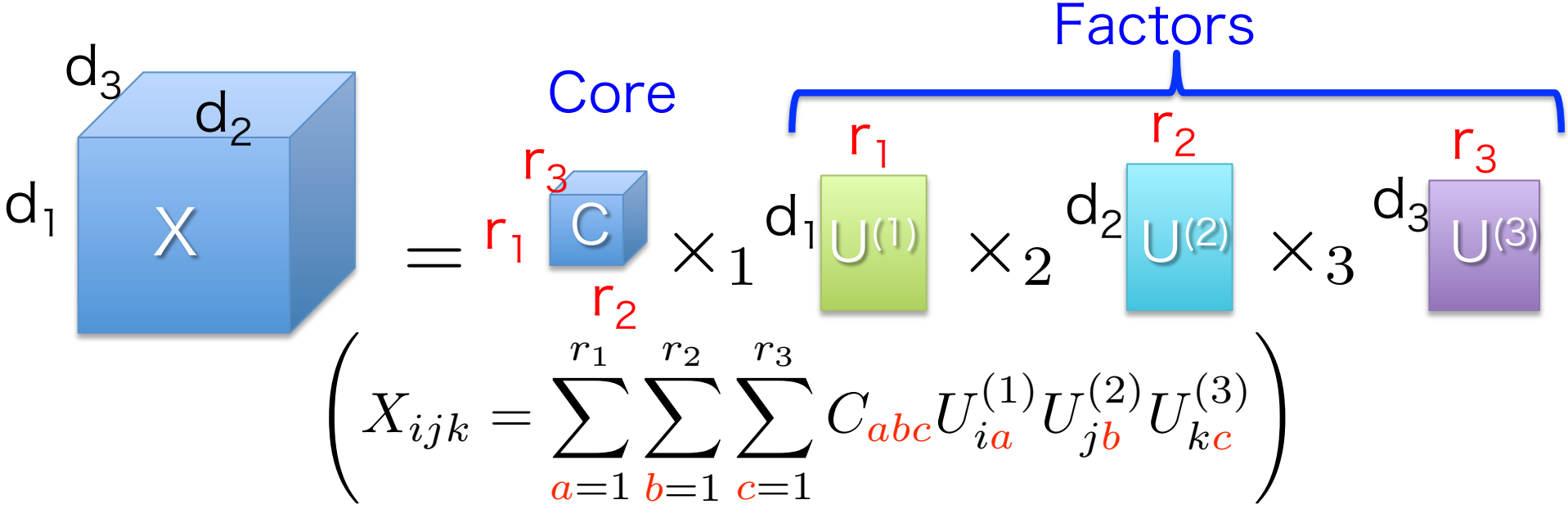
$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

σ_j: *j*th largest singular value

r: rank (number of non-zero singular values)

- Note: $r \leq \min(d_1, d_2)$
- Can be computed efficiently even for very large matrices (see Liberty et al. “Randomized algorithms for the low-rank approximation of matrices.”PNAS, 2007)

Tucker 分解 [Tucker 66]



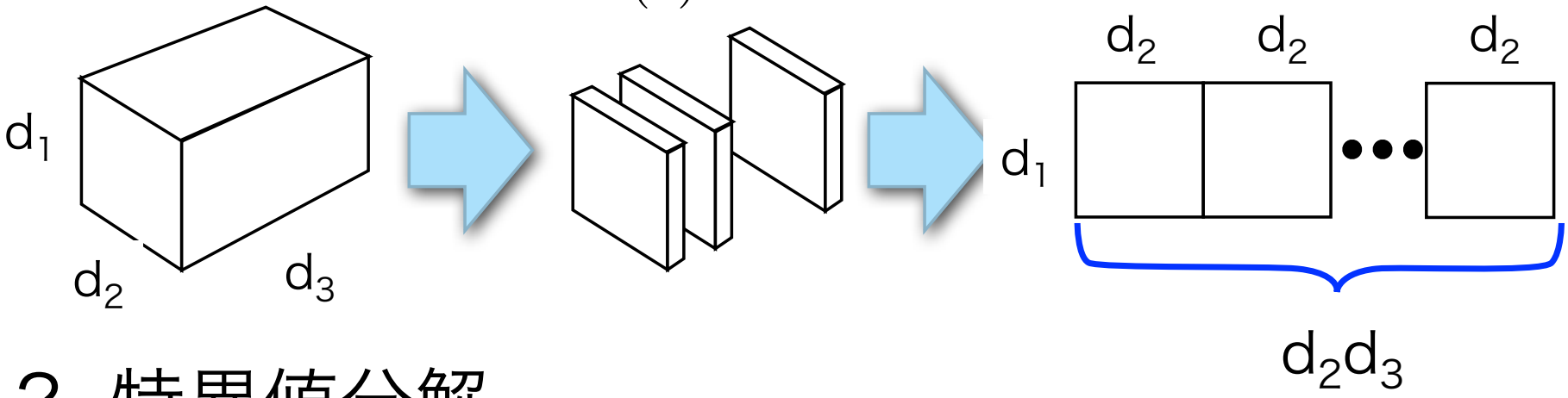
$$\left(X_{ijk} = \sum_{a=1}^{r_1} \sum_{b=1}^{r_2} \sum_{c=1}^{r_3} C_{abc} U^{(1)}_{ia} U^{(2)}_{jb} U^{(3)}_{kc} \right)$$

- それぞれのモード（次元）が異なるランクを持つ
- 直交変換の不定性
  - 通常はコアが all-orthogonal となるようにする。

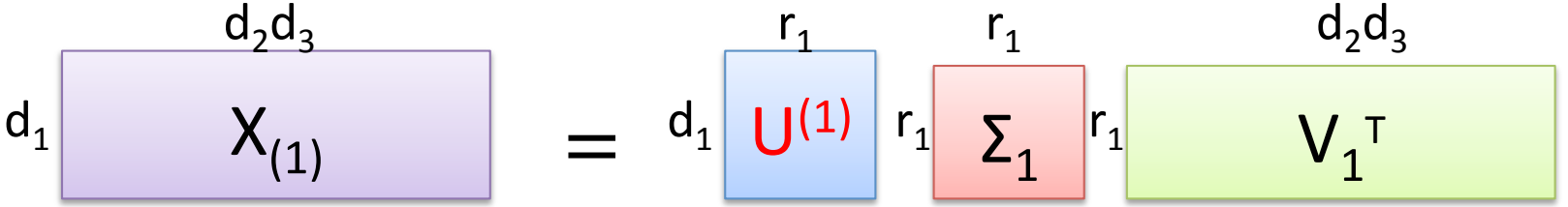
Tucker 分解の計算

1. テンソルの展開（行列化）

Mode-1 unfolding $\boldsymbol{X}_{(1)}$



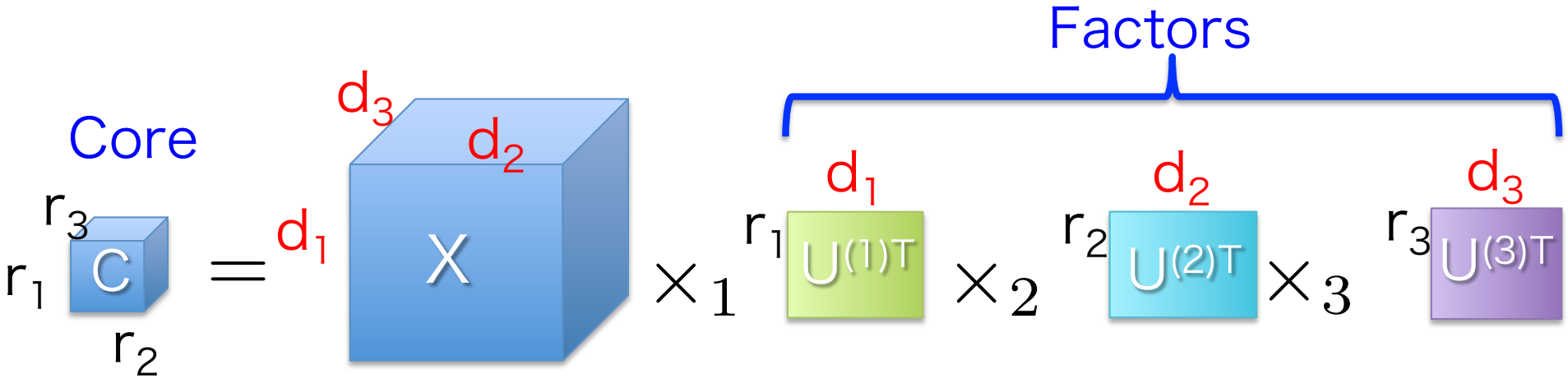
2. 特異値分解



3. すべてのモードに関して繰り返す

コアの計算

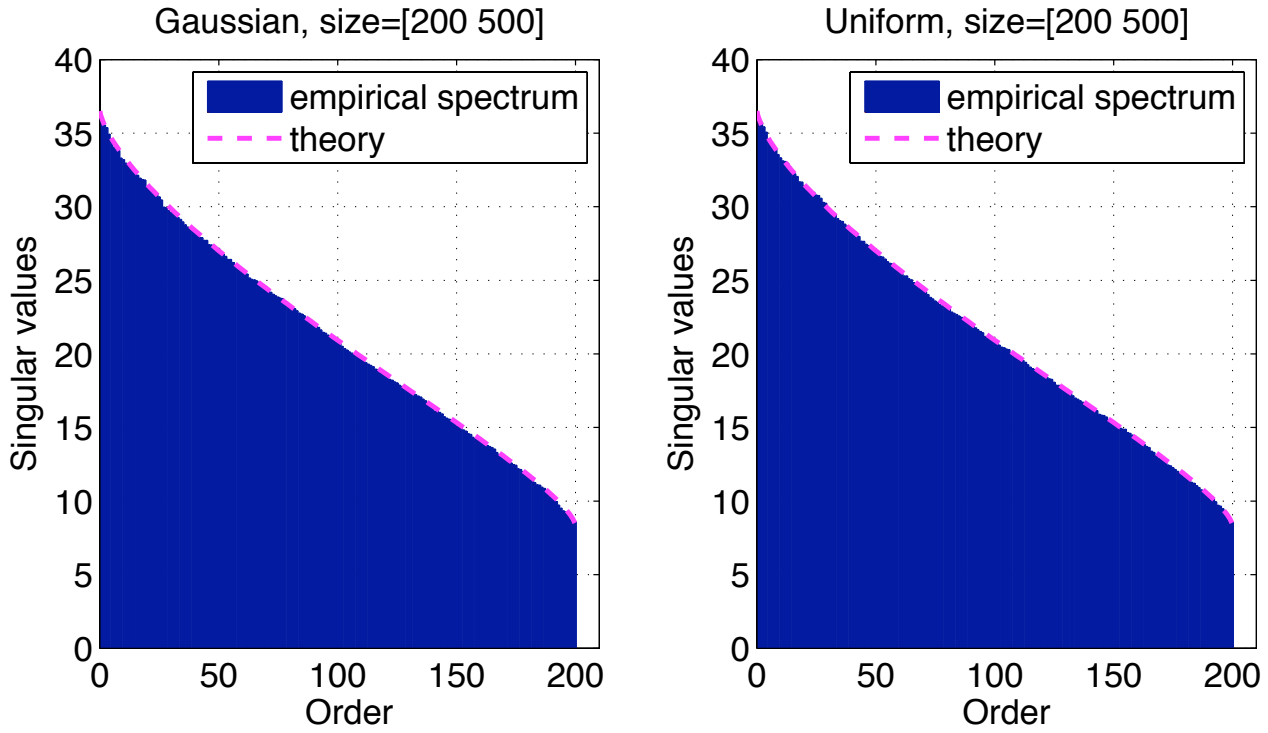
$U^{(1)}, U^{(2)}, U^{(3)}$ を計算したあと、



$$C_{abc} = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \sum_{k=1}^{d_3} X_{ijk} U^{(1)}_{ia} U^{(2)}_{jb} U^{(3)}_{kc}$$

ランダムテンソル理論？

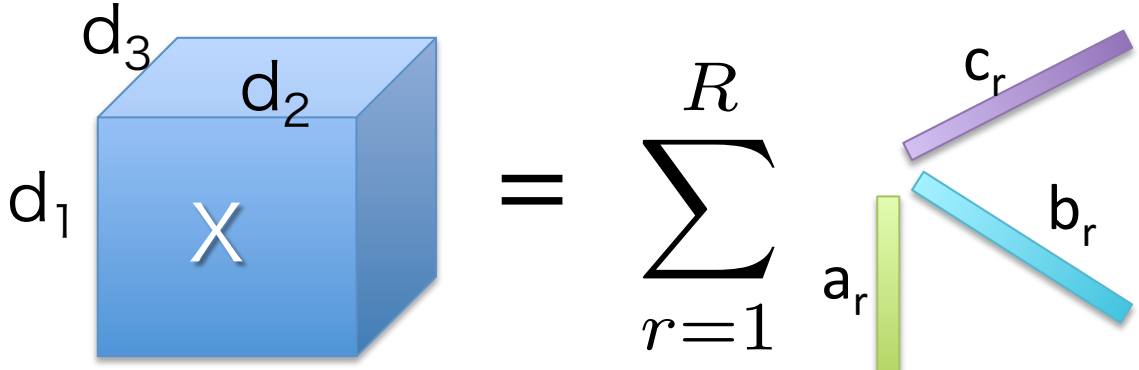
- Marchenko-Pastur 分布 (ランダム行列の特異値の分布)



テンソルのコアについて似たようなことが言えるか？

CP分解

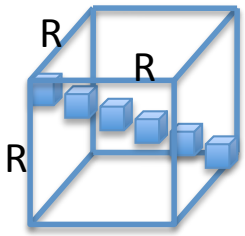
- CP = CANDECOMP [Carroll & Chang 70] / PARAFAC [Harshman 70]



$$\left(X_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \right)$$

最小のRは
ランクと
呼ばれる。

Tucker分解の特別な場合（コアが対角）と見なせる　C =



CP分解の性質

- ランクの判定はNP完全 [Håstad 90]
  – 実際には適当なランクで近似することが多い。
- CP分解はある条件のもとでスケーリング と置換の自由度を除いてユニーク

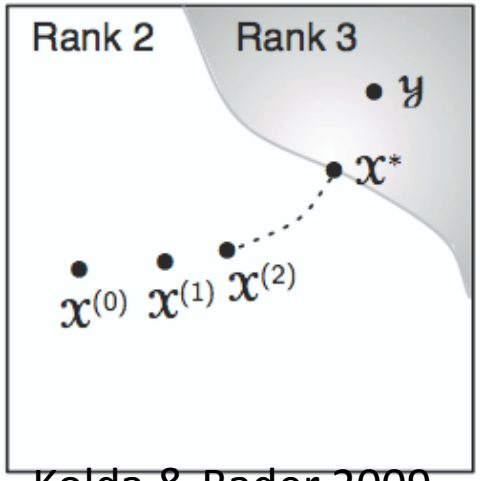
$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2.$$

k-rank: 行列の任意のk本の列が線形独立となる最大のk

CP分解の性質（続き）

X is rank 3

$$\mathcal{X} = a_1 \circ b_1 \circ c_2 + a_1 \circ b_2 \circ c_1 + a_2 \circ b_1 \circ c_1$$



Kolda & Bader 2009

Y is rank 2

$$\mathcal{Y} = \alpha(a_1 + \frac{1}{\alpha}a_2) \circ (b_1 + \frac{1}{\alpha}b_2) \circ (c_1 + \frac{1}{\alpha}c_2) - \alpha a_1 \circ b_1 \circ c_1$$

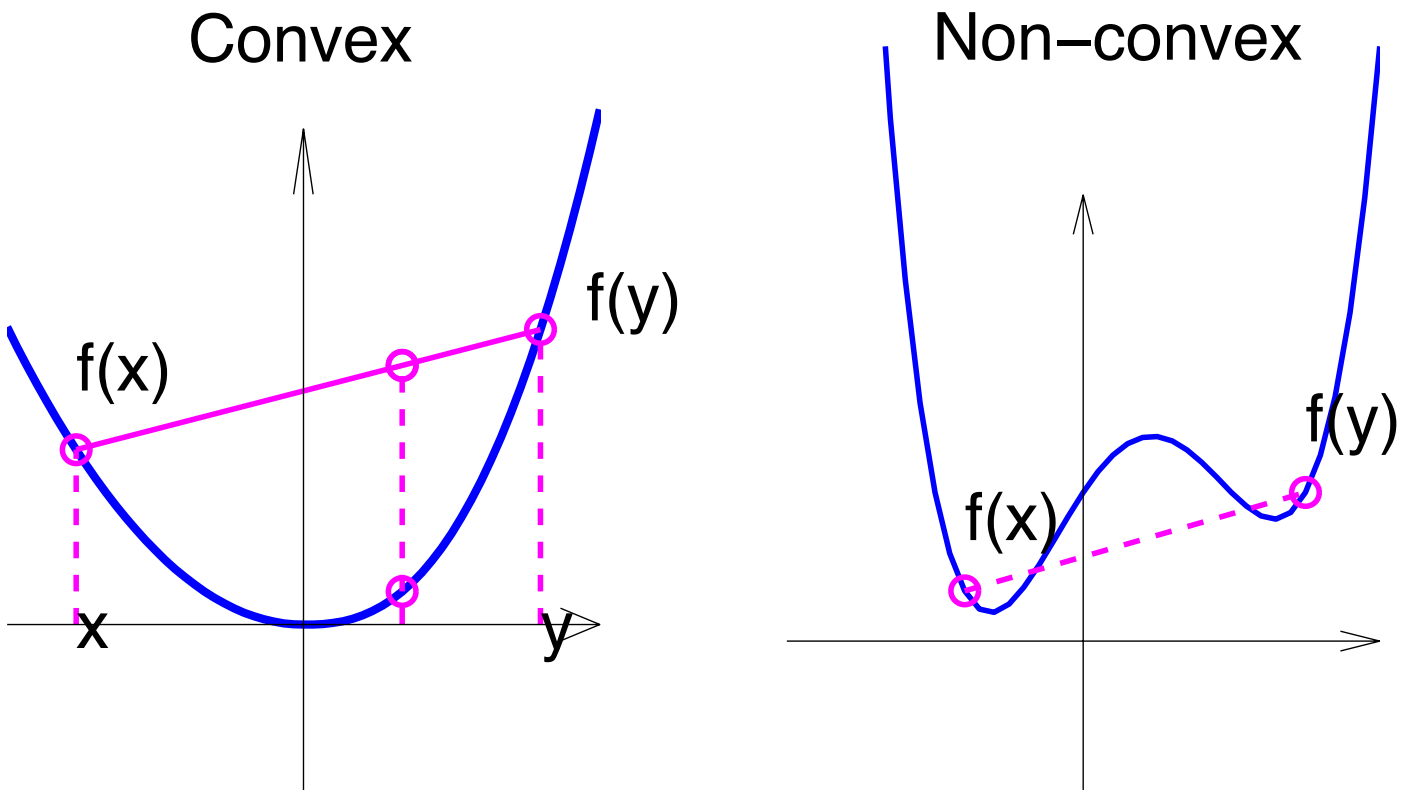
$$\|\mathcal{X} - \mathcal{Y}\|_F \to 0 \quad (\alpha \to \infty)$$

従来のテンソル分解法の問題

- ノイズや欠損値にどう対応するか
  - 単純にSVDはできない
  - 繰り返し最適化法は局所最適
- ランクをどう選ぶか
  - ランクが高すぎると
    - ノイズの影響を受けやすい（オーバーフィット）
    - 計算量多い（計算量トレードオフ）
  - 何ら仮定なしに欠損値の復元は不可能
    - 低ランク性はひとつの有効な仮定

凸最適化に基づく大域最適な推定法
（かつランク自動決定）を提案

凸関数・非凸関数

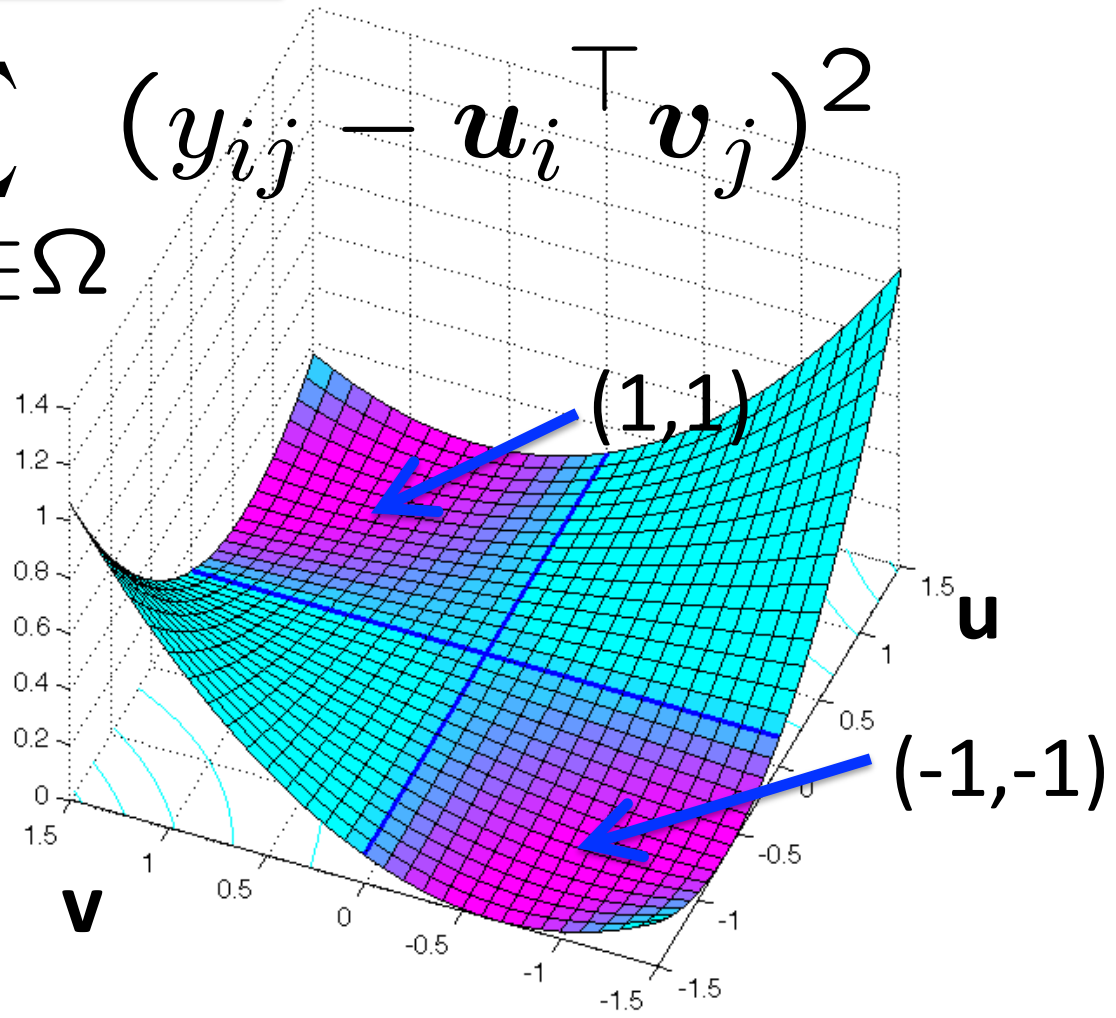


低ランク分解が局所最適になるイメージ

最適化問題

Non-convex!

$$\underset{\boldsymbol{U},\boldsymbol{V}}{\text{minimize}} \sum_{(ij)\in\Omega} (y_{ij} - \boldsymbol{u}_i{}^\top \boldsymbol{v}_j)^2$$



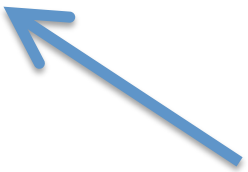
ランク制約？

最適化問題

Still non-convex!

$$\underset{\boldsymbol{W}}{\text{minimize}} \quad \sum_{(ij)\in\Omega} (y_{ij} - w_{ij})^2,$$

$$\text{subject to} \quad \text{rank}(\boldsymbol{W}) \leq r$$



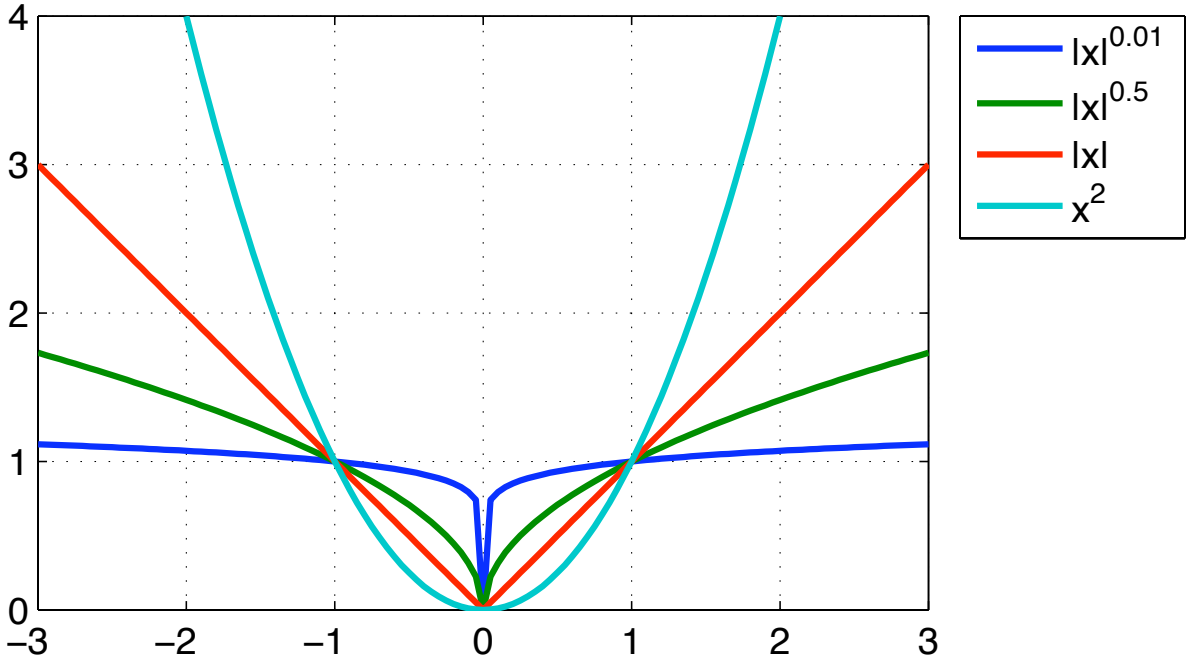
W=UV^T を言い換えただけ

ランクの凸緩和

Schatten p-ノルム $\|\boldsymbol{W}\|_{S_p}^p := \sum_{j=1}^{r} \sigma_j^p(\boldsymbol{W})$
(のp乗)

$\sigma_j(\boldsymbol{W})$: jth largest singular value

$$\|\boldsymbol{W}\|_{S_p}^p \xrightarrow{p \to 0} \mathrm{rank}(\boldsymbol{W})$$



p=1 は最もタイトな凸緩和
(trace norm / nuclear norm とも呼ばれる)

凸最適化に基づく低ランク行列補完

最適化問題　凸緩和

$$\underset{\boldsymbol{W}}{\text{minimize}} \quad \sum_{(ij)\in\Omega} (y_{ij} - w_{ij})^2,$$

$$\text{subject to} \quad \|\boldsymbol{W}\|_{S_1} \leq \tau$$

Schatten 1-ノルム　$\|\boldsymbol{W}\|_{S_1} = \sum_{j=1}^{r} \sigma_j(\boldsymbol{W})$

$\sigma_j(\boldsymbol{W})$: *j*th largest singular value

See Candes & Recht 09; Candes & Tao 10; Foygel & Srebro 11

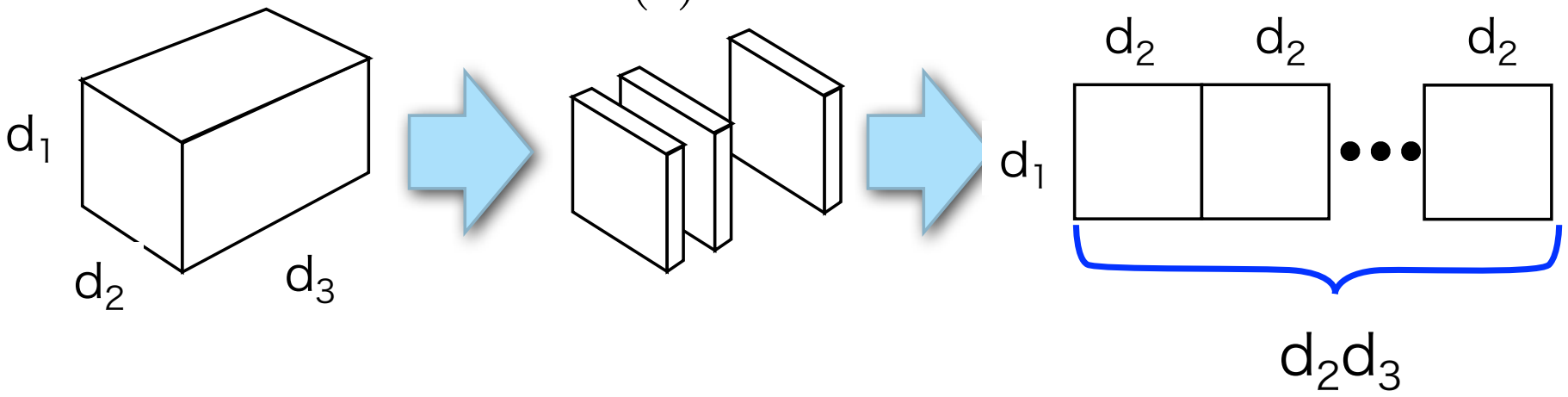
Tuckerランクの凸緩和

[Liu+09, Signoretto+10, Tomioka+10, Gandy+11]

$$\|\mathcal{W}\|_{S_1} := \frac{1}{L}\sum_{l=1}^{L} \|\boldsymbol{W}_{(l)}\|_{S_1}$$

Schatten 1-norm of the mode-l unfolding

Mode-1 unfolding $\boldsymbol{X}_{(1)}$

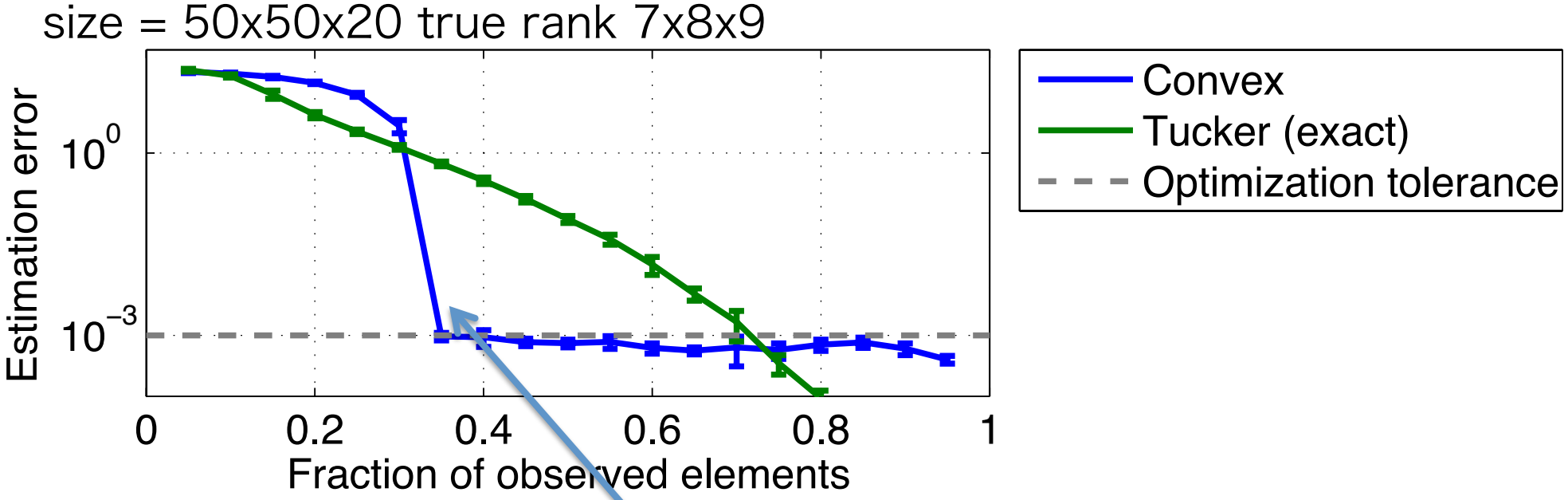


復元性能

最適化問題

$$\underset{\mathcal{X}}{\text{minimize}} \quad \|\mathcal{X}\|_{S_1},$$

$$\text{subject to} \quad \mathcal{X}_{ijk} = \mathcal{Y}_{ijk} \quad ((i,j,k) \in \Omega)$$



解析：問題設定

観測モデル

$\mathcal{W}^*$ ：真のテンソル ランク($r_1,\ldots,r_L$)

$$y_i = \langle \boldsymbol{\mathcal{X}}_i, \mathcal{W}^* \rangle + \epsilon_i \quad (i = 1, \ldots, M)$$

ガウス雑音

最適化問題　　データ尤度　　正則化項

$$\hat{\mathcal{W}} = \underset{\mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_K}}{\operatorname{argmin}} \left(\frac{1}{2} \| \boldsymbol{y} - \mathfrak{X}(\mathcal{W}) \|_2^2 + \lambda_M \| \mathcal{W} \|_{S_1} \right)$$

正則化定数

$$\left(D = \prod_{l=1}^{L} d_k \right)$$

観測作用素　$\mathfrak{X} : \mathbb{R}^D \to \mathbb{R}^M$

$$\mathfrak{X}(\mathcal{W}) = (\langle \boldsymbol{\mathcal{X}}_1, \mathcal{W} \rangle, \ldots, \langle \boldsymbol{\mathcal{X}}_M, \mathcal{W} \rangle)^\top$$

解析結果（ランダムガウスデザイン）

1. サンプル数Mに関する条件

$$\frac{\#\text{samples}(M)}{\#\text{variables}(D)} \geq c_1 \underbrace{\|\boldsymbol{d}^{-1}\|_{1/2}\|\boldsymbol{r}\|_{1/2}}_{\text{正規化ランク}} \approx \frac{r}{d}$$

2. 正則化定数λ_Mに関する条件

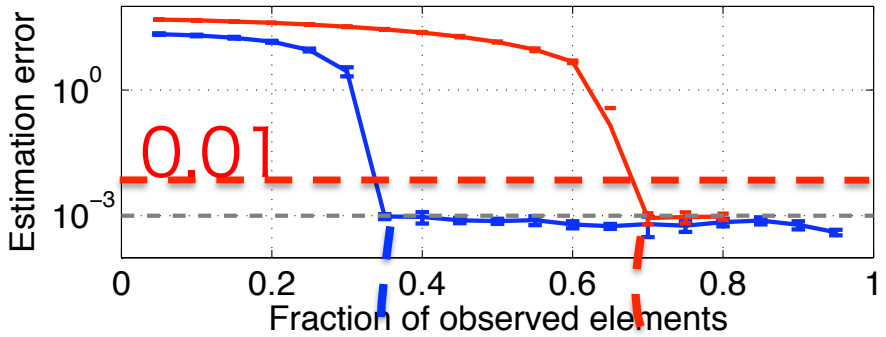
$$\lambda_M \geq c_0\sigma \sum_{l=1}^{L} \left(\sqrt{d_l} + \sqrt{D/d_l}\right) /(L\sqrt{M})$$

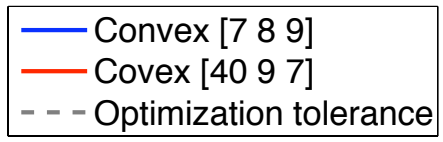
$$\frac{\left\|\left|\hat{\mathcal{W}} - \mathcal{W}^*\right|\right\|_F^2}{N} \leq O_p\left(\frac{\sigma^2\|\boldsymbol{d}^{-1}\|_{1/2}\|\boldsymbol{r}\|_{1/2}}{M}\right)$$

$$\|\boldsymbol{d}^{-1}\|_{1/2} := \left(\frac{1}{L}\sum_{l=1}^{L}\sqrt{1/d_l}\right)^2, \quad \|\boldsymbol{r}\|_{1/2} := \left(\frac{1}{L}\sum_{l=1}^{L}\sqrt{r_l}\right)^2$$

テンソル補完性能

size = 50x50x20 true rank 7x8x9 or 40x9x7





$$\frac{\#\text{samples}(M)}{\#\text{variables}(D)}$$

観測ノイズなし

ディスカッション

- ランダムガウスデザイン（＝Xの要素が独立同一なガウス乱数）
  - 解析が容易（ランダム行列の最大特異値）
  - テンソル補完の状況とは異なる
  - それにも関わらず理論と実験はよく一致
- 理論はかなり悲観的
  - 必要なサンプル数

$$M = O(rd^{L-1}) \gg O(rdL + r^L)$$

まとめ

- Tucker 分解 (=HOSVD)
  - ランクの判定は容易
  - 特異値 -> コアテンソル
  - 不定性、局所最適
- CP 分解
  - ランクの判定はNP完全
  - 分解がユニークになる場合がある
  - 局所最適
- 凸最適化にもとづくHOSVD
  - 大域最適解が求まる
  - 解の統計的な性質の理論解析が可能
  - 計算量的には改善が必要

参考文献

- Kolda & Bader (2009) Tensor decompositions and applications. SIAM Review, 51(3):455–500.
- Tucker (1966) Some mathematical notes on three-mode factor analysis. Psychometrika, 31(3):279–311.
- Candès & Recht. Exact matrix completion via convex optimization. Found. Comput. Math., 9(6):717–772, 2009.
- Candès & Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory,* 56(5):2053–2080, 2010.
- Foygel & Srebro. Concentration-based guarantees for low-rank matrix reconstruction. *Arxiv preprint arXiv:1102.3923,* 2011.
- Gandy, Recht, & Yamada (2011) Tensor completion and low-n-rank tensor recovery via convex optimization. Inverse Problems, 27:025010.
- Liu, Musialski, Wonka, & Ye. (2009) Tensor completion for estimating missing values in visual data. In Prof. ICCV.
- Signoretto, de Lathauwer, & Suykens (2010) Nuclear norms for tensors and their use for convex multilinear estimation. Tech Report 10-186, K.U.Leuven.
- Recht, Fazel, & Parrilo (2010) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review, 52(3):471–501.
- Tomioka, Hayashi, & Kashima (2011) Estimation of low-rank tensors via convex optimization. Technical report, arXiv:1010.0789, 2011.
- Tomioka, Suzuki, Hayashi, & Kashima (2011) Statistical performance of convex tensor decomposition. Advances in NIPS 24. 2011, Granada, Spain.

Singular value shrinkage

$$\mathrm{softth}(\boldsymbol{W}) = \underset{\boldsymbol{Z}\in\mathbb{R}^{d_1\times d_2}}{\mathrm{argmin}} \left(\frac{1}{2}\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2 + \lambda\|\boldsymbol{Z}\|_{S_1} \right)$$

$$= \boldsymbol{U}\max(\boldsymbol{S}-\lambda, 0)\boldsymbol{V}^\top$$

where $W=USV^\top$