Sketching and alternating projections for low-rank nonnegative matrix and tensor decompositions

Matveev S. A. [1,2], Sultonov A. O. [1], Budzinskiy S. S. [2]

[1]Lomonosov MSU

[2]Marchuk INM RAS, Moscow

Outline

1. Problem setup, short history
2. Test cases for nonnegative matrices/functions
3. Randomized SVD and complexity reduction
4. First observations for tensors in Tucker and TT-formats

Problem setup

Let

- $A \in \mathbb{R}^{m \times n}$, $A_{i,j} \geq 0$,
- $A \approx A_r, \text{rank} A^r = r$,
- $||A - A^r||_F \ll 1$.

Can we guarantee

$$A^r_{i,j} \geq 0?$$

Problem setup, NMF approach

Let

- $A \in \mathbb{R}^{m \times n}$, $A_{i,j} \geq 0$,
- $A \approx A^{(r)}, \text{rank} A^{(r)} = r$,
- $||A - A^{(r)}||_F \ll 1$.

Way I

Construct a nonnegative matrix factorization

$$A = U \cdot V^T, U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r},$$

Subject to

$$U_{i,j}, V_{i,j} \geq 0.$$

Problems:

1. nonnegative rank is often greater than usual($\text{rank}_+) \geq \text{rank}$
2. NP-hard task
3. Iterative methods require a lot of computing and requre significant efforts in theory
4. Nonnegative factors are not always necessary

Problem setup, NMF approach

Way I: NMF

NMF construction task

$$A = U \cdot V^T, U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r},$$

$$U_{i,j}, V_{i,j} \geq 0.$$

1. nonnegative rank is often greater than usual$(\text{rank}_+) \geq \text{rank}$
2. NP-hard task
3. Iterative methods require a lot of computing and requre significant efforts in theory
4. Nonnegative factors are not always necessary

References:

1. Gillis N. Nonnegative matrix factorization. – Society for Industrial and Applied Mathematics, 2020.
2. Tyrtyshnikov E. E., Shcherbakova E. M. Methods for nonnegative matrix factorization based on low-rank cross approximations //Computational Mathematics and Mathematical Physics. – 2019. – T. 59. – №. 8.

Problem setup, LRNMF

Way II: LRNMF

Construct low-rank nonnegative matrix factorization

$$A = U \cdot V^T, U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r},$$

$$U_{i,j}, V_{i,j} : A^{(r)}_{i,j} \geq 0$$

1. Nonnegative factors are **not** necessary U, V
2. Thus, we seek to get element from intersection of two sets: (I) low-rank matrices ($\leq r$) and (II) nonnegative matrices
3. Set I ($M_{\geq 0}$) is convex
4. Set II ($M_{\text{rank}A \leq r}$): low-rank matrices ($\leq r$) is smooth
5. For both $M_{\geq 0}$ and $M_{\text{rank}A \leq r}$ there are projection operators.

Hence, let us use alternating projections.

Alternating projections, 2020

$$\Pi_{\mathcal{M}_{\leq r}}(X) = \mathrm{SVD}_r(X) = U_r \Sigma_r V_r^T = \arg \min_{Y \in \mathcal{M}_{\leq r}} \|X - Y\|_F,$$

$$\Pi_{\mathbb{R}_+^{m \times n}}(X) = \max(X, 0) = \{\max(x_{ij}, 0)\} = \arg \min_{Y \in \mathbb{R}_+^{m \times n}} \|X - Y\|_F.$$

Song G. J., Ng M. K. Nonnegative low rank matrix approximation for nonnegative matrices //Applied Mathematics Letters. – 2020. – Vol. 105.

Complexity $O(mn^2 + mn)$

Algorithm 1: Exact alternating projections (SVD)

Data: Initial approximation $Y^{(0)} \in \mathbb{R}^{m \times n}$ of rank r, number of iterations s

for $i = 1, \ldots, s$ **do**

  $X^{(i)} \leftarrow \max(Y^{(i-1)}, 0)$;

  $[U_r, \Sigma_r, V_r] \leftarrow \mathrm{SVD}_r(X^{(i)})$;

  $Y^{(i)} \leftarrow U_r \Sigma_r V_r^T$;

end

return $Y^{(s)}$

Test function

Possible functions/data:

- Images (BW, RGB, multispectral)
- Solutions of PDEs and related models in higher dimensions (Chemical master equation, Fokker-Planck equation, population balance equations etc)
- Probability distributions

In all cases low-rank matrices and tensors are useful.

Test function

Solution of the two-component Smoluchowski equtaion

$$\frac{\partial n(v_1, v_2, t)}{\partial t} = -n(v_1, v_2, t) \int_0^\infty \int_0^\infty K(u_1, u_2; v_1, v_2) n(u_1, u_2, t) du_1 du_2$$
$$+ \frac{1}{2} \int_0^{v_1} \int_0^{v_2} K(v_1 - u_1, v_2 - u_2; u_1, u_2) n(v_1 - u_1, v_2 - u_2, t) n(u_1, u_2, t) du_1 du_2,$$

For case of constant kernel function

$$K(u_1, u_2; v_1, v_2) \equiv K,$$

and initial condition

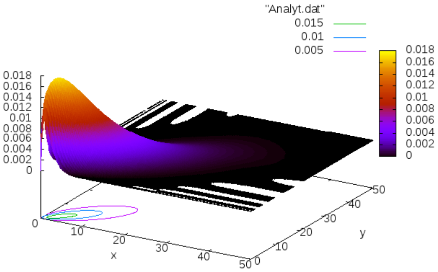
$$n(v_1, v_2, t = 0) = \sqrt{K} abe^{-av_1 - bv_2},$$

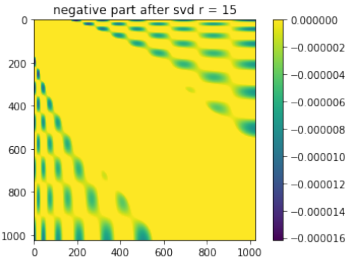
there is a known analytical solution(it has good *low-rank* approximations!)

$$n(v_1, v_2, t) = \sqrt{K} \frac{abe^{-av_1 - bv_2}}{(1 + \sqrt{K}t/2)^2} I_0 \left(2\sqrt{\frac{abv_1 v_2 \sqrt{K}t}{\sqrt{K}t + 2}} \right),$$

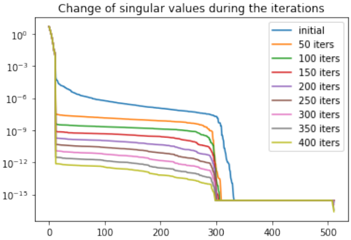
Test function

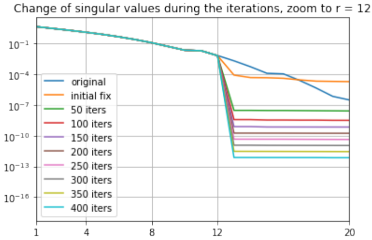
Matveev S. A. et al. Tensor train versus Monte Carlo for the multicomponent Smoluchowski coagulation equation //Journal of Computational Physics. – 2016. – T. 316.





How does it work?





Can we go faster? Yes, via tangent projections

$$T_Y\mathcal{M}_r = \left\{ U_r A^T + B V_r^T \ : \ A \in \mathbb{R}^{n\times r}, \quad B \in \mathbb{R}^{m\times r} \right\},$$

$$\Pi_{T_Y\mathcal{M}_r}(X) = U_r U_r^T X + (I - U_r U_r^T) X V_r V_r^T.$$

$$\tilde{\Pi}_{\mathcal{M}_{\leq r}}(X) = \mathrm{SVD}_r(\Pi_{T_Y\mathcal{M}_r}(X)).$$

Song G., Ng M. K., Jiang T. X. Tangent Space Based Alternating Projections for Nonnegative Low Rank Matrix Approximation //arXiv preprint arXiv:2009.03998. – 2020.

Can we go faster? Yes, via tangent projections

Song G., Ng M. K., Jiang T. X. Tangent Space Based Alternating Projections for Nonnegative Low Rank Matrix Approximation //arXiv preprint arXiv:2009.03998. – 2020.

Complexity $O(mnr)$

Algorithm 2: Alternating projections via tangent spaces (Tangent)

Data: Initial approximation $Y^{(0)} \in \mathbb{R}^{m \times n}$ of rank r, number of iterations s

$[U_r, \Sigma_r, V_r] \leftarrow \mathrm{SVD}_r(Y^{(0)})$;

for $i = 1, \ldots, s$ **do**

$X^{(i)} \leftarrow \max(Y^{(i-1)}, 0)$;

$G_1 \leftarrow U_r^T X^{(i)} \in \mathbb{R}^{r \times n}, \quad G_2 \leftarrow (I - U_r U_r^T) X^{(i)} V_r \in \mathbb{R}^{m \times r}$;

$[Q_1, R_1] \leftarrow \mathrm{QR}(G_1^T), \quad [Q_2, R_2] \leftarrow \mathrm{QR}(G_2)$;

$Z \leftarrow \begin{bmatrix} G_1 V_r & R_1^T \\ R_2 & 0 \end{bmatrix} \in \mathbb{R}^{2r \times 2r}$;

$[\tilde{U}_r, \Sigma_r, \tilde{V}_r] \leftarrow \mathrm{SVD}_r(Z)$;

$U_r \leftarrow \begin{bmatrix} U_r & Q_2 \end{bmatrix} \tilde{U}_r, \quad V_r \leftarrow \begin{bmatrix} V_r & Q_1 \end{bmatrix} \tilde{V}_r$;

$Y^{(i)} \leftarrow U_r \Sigma_r V_r^T$;

end

return $Y^{(s)}$

Can we use sketching?

Answer: Yes! Just replace the original SVD by randomized SVD (and check various sketchings)

Matveev S. A., Budzinskiy S. Sketching for low-rank nonnegative matrix approximation: a numerical study //arXiv preprint arXiv:2201.11154. – 2022.

Algorithm 3: Alternating projections via HMT

Data: Initial approximation $Y^{(0)} \in \mathbb{R}^{m \times n}$ of rank r, co-range sketch size $k \geq r$, number of power method iterations p, test matrix generator $\textsc{TestMatrix}$, number of iterations s

for $i = 1, \ldots, s$ **do**

  $X^{(i)} \leftarrow \max(Y^{(i-1)}, 0)$;

  $\Psi \leftarrow \textsc{TestMatrix}(n, k) \in \mathbb{R}^{n \times k}$;

  $Z_1 \leftarrow X^{(i)} \Psi \in \mathbb{R}^{m \times k}$;

  $[Q, R] \leftarrow \mathrm{QR}(Z_1)$;

  for $j = 1, \ldots, p$ **do**

    $Z_2 \leftarrow Q^T X^{(i)} \in \mathbb{R}^{k \times n}$;

    $[Q, R] \leftarrow \mathrm{QR}(Z_2^T)$;

    $Z_1 \leftarrow X^{(i)} Q \in \mathbb{R}^{m \times k}$;

    $[Q, R] \leftarrow \mathrm{QR}(Z_1)$;

  end

  $Z_2 \leftarrow Q^T X^{(i)} \in \mathbb{R}^{k \times n}$;

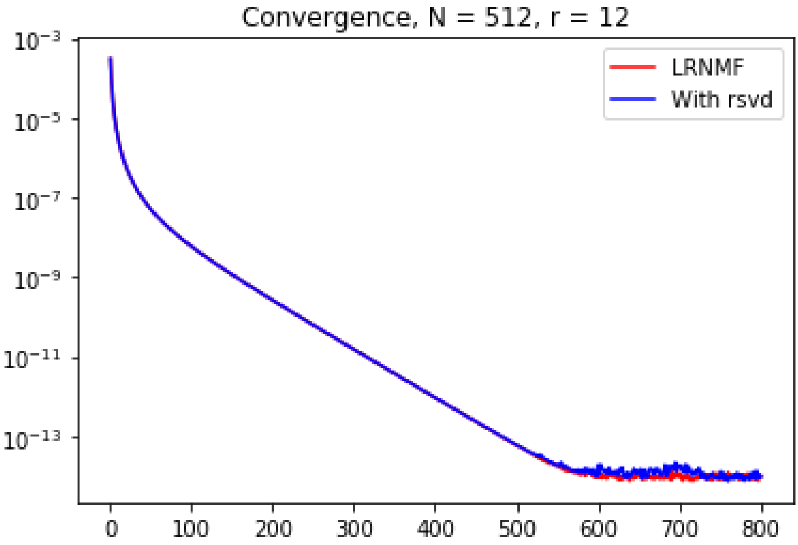
  $[U_r, \Sigma_r, V_r] \leftarrow \mathrm{SVD}_r(Z_2)$;

  $Y^{(i)} \leftarrow Q U_r \Sigma_r V_r^T$;

end

return $Y^{(s)}$

Convergence tests



FLOPs comparison

Method	Sketch	Flops (init/per iter)	$\|\|\varepsilon\|\|_F$ (init/res)
SVD	N/A	$2.3 \cdot 10^{10}/2.3 \cdot 10^{10}$	$2.39 \cdot 10^{-2}/2.70 \cdot 10^{-2}$
Tangent	N/A	$2.3 \cdot 10^{10}/6.5 \cdot 10^{7}$	$2.39 \cdot 10^{-2}/2.72 \cdot 10^{-2}$
HMT(0, 15)	$\mathrm{Rad}(0.2)$	$3.7 \cdot 10^{7}/5.8 \cdot 10^{7}$	$2.43 \cdot 10^{-2}/2.75 \cdot 10^{-2}$
Tropp(15, 25)	$\mathrm{Rad}(0.2)$	$1.2 \cdot 10^{7}/3.3 \cdot 10^{7}$	$2.43 \cdot 10^{-2}/2.87 \cdot 10^{-2}$
GN(40)	$\mathrm{Rad}(0.2)$	$1.2 \cdot 10^{7}/3.3 \cdot 10^{7}$	$8.47 \cdot 10^{-2}/1.83 \cdot 10^{-1}$

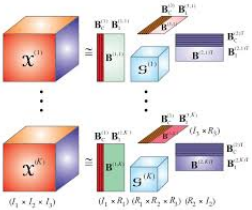
Test case: 1000 iterations with rank-10 LRNMF approximations for "Smoluchowski" matrix 1024×1024.

Tensors, Tucker format

Tucker format is a following decomposition for tensor $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$:

$$A = G \times_1 U_1 \cdots \times_d U_d,$$

where $G \in \mathbb{R}^{r_1 \times \cdots \times r_d}$, $U_1 \in \mathbb{R}^{r_1 \times n_1}, \ldots, U_d \in \mathbb{R}^{r_d \times n_d}$.



How to modify alternating projections?

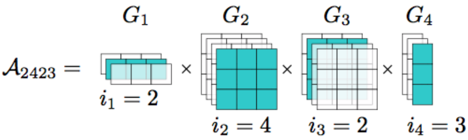
- HOSVD may replace basic SVD
- Randomized SVD can be easily incorporated and reduce complexity from $O(N^{d+1})$ to $O(N^d)$

Tensors, TT-format

Tensor Train is a following decomposition for tensor $A \in \mathbb{R}^{n_1 \times \dots \times n_d}$:
$A \in \mathbb{R}^{n_1 \times \dots \times n_d}$ as following

$$A(i_1, \dots, i_d) = \sum_{\alpha_0, \dots, \alpha_d} G_1(\alpha_0, i_1, \alpha_1) G_2(\alpha_1, i_2, \alpha_2) \dots G_d(\alpha_{d-1}, i_d, \alpha_d),$$

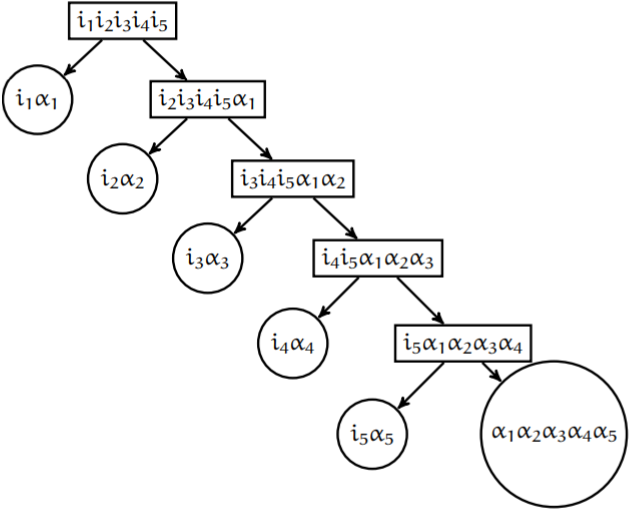
where $G_k \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$



How to modify alternating projections? The very same trick

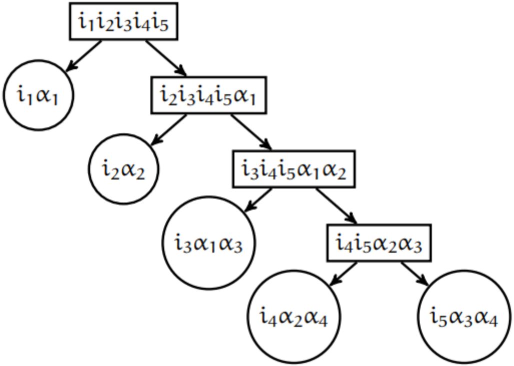
- TTSVD may also replace basic SVD :)
- Randomized SVD can be easily incorporated and reduce complexity from $O(N^{d+1})$ to $O(N^d)$

HO-SVD and TT-SVD



HOSVD graphical representation

HO-SVD and TT-SVD

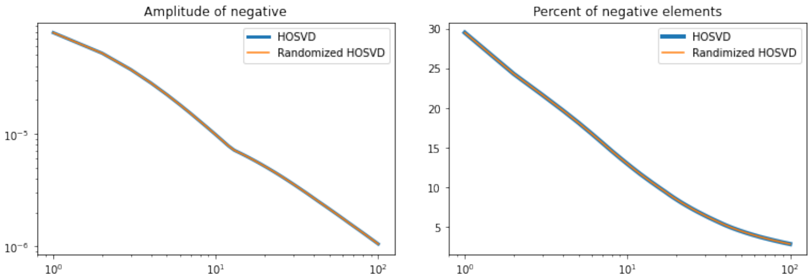


TTSVD graphical representation

Oseledets I. V., Tyrtyshnikov E. E. (2009). Breaking the curse of dimensionality, or how to use SVD in many dimensions. SIAM Journal on Scientific Computing, 31(5), 3744-3759.

Test case: mixture of three multidimensional Gaussians

Test case: tensor modes $118 \times 128 \times 138$.
Three random covariance matrices for Gaussians.
Tensor is generated as elementwise sum of these three tensors.



We see the very same results for this test but in TT-format.

1. Sketching and randimozed SVD work fine for alternating projections for low-rank nonnegative matrix factorizations.
2. Our observations can be also generalized for tensors in TT and Tucker formats
3. Main task (challenge?) for future: find a way of reduction of complexity for projector onto $M_{\geq 0}$ and $T_{\geq 0}$ sets

Thank you for attention![1].

[1]This work has been supported by Russian Science Foundation project 21-71-10072.
Preprint for matrix case is available Matveev S. A., Budzinskiy S. Sketching for low-rank nonnegative matrix approximation: a numerical study //arXiv preprint arXiv:2201.11154. – 2022.
Paper with tests for tensors is almost ready.