Sparse and Low-Rank Tensor Decomposition: Supplementary Material

Anonymous Author(s)
Affiliation
Address
email

1 Leurgans' Algorithm

For the sake of completeness, we present Leurgans' algorithm for tensor decomposition. The algorithm essentially uses simultaneous diagonalization (Lemma 2.2) at its core.

Algorithm 2 Leurgans' algorithm for tensor decomposition

1: **Input:** Tensor $\boldsymbol{X}$
2: Generate contraction vectors $a, b \in \mathbb{R}^{n_3}$, $c, d \in \mathbb{R}^{n_1}$ uniformly randomly distributed on the unit sphere.
3: Compute mode 3 contractions X_a^3 and X_b^3 respectively.
4: Compute eigen-decomposition of $M_1 := X_a^3(X_b^3)^\dagger$ and $M_2 := (X_b^3)^\dagger X_a$. Let U and V denote the matrices whose columns are the eigenvectors of M_1 and M_2^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_1 and M_2.) The eigenvectors, thus arranged are denoted as $\{u_i\}_{i=1,\dots,r}$ and $\{v_i\}_{i=1,\dots,r}$.
5: Compute mode 1 contractions X_c^1 and X_d^1 respectively.
6: Compute eigen-decomposition of $M_3 := X_c^1(X_d^1)^\dagger$ and $M_4 := (X_c^1)^\dagger X_d^1$. Let $\tilde{V}$ and $\tilde{W}$ denote the matrices whose columns are the eigenvectors of M_3 and M_4^T respectively corresponding to the non-zero eigenvalues, in sorted order. (Let r be the (common) rank of M_3 and M_4.)
7: Simultaneously reorder the columns of $\tilde{V}, \tilde{W}$, also performing simultaneous sign reversals as necessary so that the columns of V and $\tilde{V}$ are equal, call the resulting matrix W with columns $\{w_i\}_{i=1,\dots,r}$.
8: Solve for λ_i in the linear system
$$\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i.$$
9: **Output:** Decomposition $\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i \, u_i \otimes v_i \otimes w_i$.

2 Proofs

Since our algorithmic approach reduces the tensor decomposition problem to that of sparse and low-rank matrix decomposition, some of the proofs of the lemmas below reuse existing results. Rather than reproving the intermediate results here, we simply refer the reader to the appropriate references.

Our first lemma establishes that given information about two contractions of a tensor, one may recover the the tensor via linear algebraic operations.

Lemma 2.2. *[4, 19] Suppose we are given an order 3 tensor $\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i \, u_i \otimes v_i \otimes w_i$ of size $n_1 \times n_2 \times n_3$ satisfying the conditions of Assumption 1.1. Suppose the contractions X_a^3 and X_b^3 are computed with respect to unit* *vectors $a, b \in \mathbb{R}^{n_3}$ distributed independently and uniformly on the unit sphere $\mathbb{S}^{n_3-1}$ and consider the matrices M_1 and M_2 formed as:*

$$M_1 = X_a^3(X_b^3)^\dagger \qquad M_2 = (X_b^3)^\dagger X_a^3.$$

Then the eigenvectors of M_1 (corresponding to the non-zero eigenvalues) are $\{u_i\}_{i=1,\ldots,r}$, and the eigenvectors of M_2^T are $\{v_i\}_{i=1,\ldots,r}$.

Proof. Suppose we are given an order 3 tensor $\boldsymbol{X} = \sum_{i=1}^r \lambda_i \, u_i \otimes v_i \otimes w_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. From the definition of contraction (2), it is straightforward to see that

$$X_a^3 = U D_a V^T \qquad D_a = \text{diag}(\lambda_1 a^T w_1, \ldots, \lambda_r a^T w_r)$$

$$X_b^3 = U D_b V^T \qquad D_b = \text{diag}(\lambda_1 b^T w_1, \ldots, \lambda_r b^T w_r).$$

In the above decompositions, $U \in \mathbb{R}^{n_1 \times r}$, $V \in \mathbb{R}^{n_2 \times r}$, and the matrices $D_a, D_b \in \mathbb{R}^{r \times r}$ are diagonal and non-singular almost surely (since a, b are random). Now,

$$\begin{aligned} M_1 &:= X_a^3(X_b^3)^\dagger \\ &= U D_a V^T (V^\dagger)^T D_b^{-1} U^\dagger \\ &= U D_a D_b^{-1} U^\dagger \end{aligned} \tag{7}$$

and similarly we obtain

$$M_2^T = V D_b^{-1} D_a V^\dagger. \tag{8}$$

Since we have $M_1 U = U D_a D_b^{-1}$ and $M_2^T V = V D_b^{-1} D_a$, it follows that the columns of U and V are eigenvectors of M_1 and M_2^T respectively (with corresponding eigenvalues given by the diagonal matrices $D_a D_b^{-1}$ and $D_b^{-1} D_a$). ☐

Theorem 3.1. *Suppose $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$, where $\boldsymbol{X} = \sum_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i$, has rank $r \leq n_1$ and such that the factors satisfy Assumption 1.1. Suppose $\boldsymbol{Y}$ has support Ω and the following condition is satisfied.*

$$\mu\left(\Omega^{(3)}\right) \zeta\left(U, V\right) \leq \frac{1}{6} \qquad\qquad \mu\left(\Omega^{(1)}\right) \zeta\left(V, W\right) < \frac{1}{6}.$$

Then Algorithm 1 succeeds in exactly recovering the component tensors, i.e. $(\boldsymbol{X}, \boldsymbol{Y}) = (\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}})$ whenever ν_k are picked so that $\nu_3 \in \left(\frac{\zeta(U,V)}{1-4\zeta(U,V)\mu\left(\Omega^{(3)}\right)}, \frac{1-3\zeta(U,V)\mu\left(\Omega^{(3)}\right)}{\mu\left(\Omega^{(3)}\right)}\right)$ and $\nu_1 \in \left(\frac{\zeta(V,W)}{1-4\zeta(V,W)\mu\left(\Omega^{(1)}\right)}, \frac{1-3\zeta(V,W)\mu\left(\Omega^{(1)}\right)}{\mu\left(\Omega^{(1)}\right)}\right)$. Specifically, choice of $\nu_3 = \frac{(3\zeta(U,V))^p}{\left(\mu\left(\Omega^{(3)}\right)\right)^{1-p}}$ and $\nu_1 = \frac{(3\zeta(V,W))^p}{\left(\mu\left(\Omega^{(1)}\right)\right)^{1-p}}$ for any $p \in [0, 1]$ in these respective intervals guarantees exact recovery.

Proof. Since $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$, we have $Z_a^3 = X_a^3 + Y_a^3$. By Lemma 2.1 X_a^3 is a low-rank matrix with row space span(U) and column space span(V). Hence the incoherence parameter for X_a^3 is precisely $\zeta\left(U, V\right)$. Since $\boldsymbol{Y}$ is sparse with support Ω, Y_a^3 is sparse with support $\Omega^{(3)}$. By assumption, $\mu\left(\Omega^{(3)}\right) \zeta\left(U, V\right) \leq \frac{1}{6}$. By [10, Theorem 2], the convex relaxation (6) with the prescribed regularization parameter exactly recovers the unique low-rank and sparse components, i.e. X_a^3, Y_a^3. Similarly, the procedure repeated with respect to the contraction vector b recovers X_b^3. By Lemma 2.2, step 6 of Algorithm 1 exactly recovers the U, V. The same procedure repeated with contractions along the first mode with respect to c, d ensures recovery of V, W. Note that in Step 12 of Algorithm 1, the linear system is full rank (since the factors are linearly independent by Assumption 1.1), overdetermined and thus has a unique and correct solution. ☐

Lemma 3.2. *We have:*

$$\mu\left(\Omega^{(k)}\right) \leq \textit{deg}(\boldsymbol{Y}), \textit{ for all } k.$$

Proof. Given a tensor $\boldsymbol{Y}$ with support Ω, the sparsity pattern of Y_a^3 is contained within $\Omega^{(3)}$. By the definition of the degree of $\boldsymbol{Y}$, we have $\deg(Y_a^3) \leq \deg(\boldsymbol{Y})$. By [10, Proposition 3] the result follows. ☐

Lemma 3.3. *We have*

$$\zeta(U,V) \leq 2inc(\boldsymbol{X}) \qquad \zeta(V,W) \leq 2inc(\boldsymbol{X}).$$

Proof. From [10, Proposition 4], we have $\zeta(U,V) \leq 2\max\{\beta(\text{span}(U)), \beta(\text{span}(V))\}$. Similarly, we have $\zeta(V,W) \leq 2\max\{\beta(\text{span}(V)), \beta(\text{span}(W))\}$. The result follows by applying the definition of inc$(\boldsymbol{X})$. □

Corollary 3.4. *Let $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$, with $\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i$ and rank $r \leq n_1$, the factors satisfy Assumption 1.1 and incoherence inc$(\boldsymbol{X})$. Suppose $\boldsymbol{Y}$ is sparse and has degree deg$(\boldsymbol{Y})$. If the condition*

$$inc(\boldsymbol{X})deg(\boldsymbol{Y}) < \frac{1}{12}$$

holds then Algorithm 1 successfully recovers the true solution, i.e. . $(\boldsymbol{X},\boldsymbol{Y}) = (\hat{\boldsymbol{X}},\hat{\boldsymbol{Y}})$ when the parameters

$$\begin{aligned}\nu_3 &\in \left(\frac{2inc_3(\boldsymbol{X})}{1-8deg_3(\boldsymbol{Y})inc_3(\boldsymbol{X})}, \frac{1-6deg_3(\boldsymbol{Y})inc_3(\boldsymbol{X})}{deg_3(\boldsymbol{Y})}\right)\\ \nu_1 &\in \left(\frac{2inc_1(\boldsymbol{X})}{1-8deg_1(\boldsymbol{Y})inc_1(\boldsymbol{X})}, \frac{1-6deg_1(\boldsymbol{Y})inc_1(\boldsymbol{X})}{deg_1(\boldsymbol{Y})}\right).\end{aligned}$$

Specifically, a choice of $\nu_3 = \frac{(6inc_3(\boldsymbol{X}))^p}{(2deg_3(\boldsymbol{Y}))^{1-p}}$, $\nu_1 = \frac{(6inc_1(\boldsymbol{X}))^p}{(2deg_1(\boldsymbol{Y}))^{1-p}}$ for any $p \in [0,1]$ is a valid choice that guarantees exact recovery.

Proof. Follows immediately from Lemma 3.2, Lemma 3.3, and the conditions of Theorem 3.1 being satisfied. □

We consider, for the sake of simplicity, tensors of uniform dimension, i.e. $\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z} \in \mathbb{R}^{n\times n\times n}$. We define the *random sparsity* model to be one where each entry of the tensor $\boldsymbol{Y}$ is non-zero independently and with identical probability ρ. We make no assumption about the mangitude of the entries of $\boldsymbol{Y}$, only that its non-zero entries are thus sampled.

Lemma 3.5. *Let $\boldsymbol{X} = \sum_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i$, where $u_i, v_i, w_i \in \mathbb{R}^n$ are uniformly randomly distributed on the unit sphere $\mathbb{S}^{n-1}$. Then the incoherence of the tensor $\boldsymbol{X}$ satisifies:*

$$inc(\boldsymbol{X}) \leq c_1\sqrt{\frac{\max\{r, \log\, n\}}{n}}$$

for some constants c_1, c_2, with probability exceeding $1 - c_2 n^{-3}\log\, n$.

Proof. Since u_i are picked uniformly randomly on the unit sphere, the subspace span(U) is a uniformly random subspace. Equivalently, span(U) = span$(\tilde{U})$ for some random matrix $\tilde{U}$ which is uniformly distributed with respect to the set of partial isometries in $\mathbb{R}^{n\times k}$. By [7, Lemma 2.2] we have that

$$\beta(\text{span}(U)) \leq c_1\sqrt{\frac{\max\{r, \log\, n\}}{n}}$$

with probability exceeding $1 - k_0 n^{-3}\log\, n$ for some constant k_0. The same results hold for the incoherences of span(V), span(W). By the definition of inc$(\boldsymbol{X})$, we have the required result. required result. □

Lemma 3.6. *Suppose the entries of $\boldsymbol{Y}$ are sampled according to the random sparsity model, and*

$$\rho = O\left(\left(n^{\frac{3}{2}}\max(\log n, r)\right)^{-1}\right).$$

Then the tensor $\boldsymbol{Y}$ satisfies:

$$deg(\boldsymbol{Y}) \leq \frac{\sqrt{n}}{12c_1\max(\log n, r)}$$

with probability exceeding $1 - exp\left(-c_3\frac{\sqrt{n}}{\max(\log n, r)}\right)$ for some constant $c_3 > 0$.

Proof. To bound $\deg(\boldsymbol{Y})$ we must bound the degree of any matrix supported on $\Omega^{(k)}$ for $k = 1, 2, 3$. To this end we introduce the following sets of random variables:

- Let $B_{ijk} \sim \text{Bernoulli}(\rho)$ be the random variable such that $B_{ijk} = 1$ when $(i, j, k) \in \Omega$ and 0 otherwise.
- Let C be a matrix such that $C_{ij} = 1$ if $(i, j) \in \Omega^{(3)}$ and 0 otherwise.

We have that

$$C_{ij} \leq \sum_{k=1}^{n} B_{ijk}.$$

Hence, for any column of C (say j^{th} column), we have that the number of non-zeroes in the column (let us denote this by $\deg(C_j)$) is given by:

$$\deg(C_j) = \sum_{i=1}^{n} C_{ij} \leq \sum_{i=1}^{n}\sum_{k=1}^{n} B_{ijk}. \tag{9}$$

Since (9) is a sum of $i.i.d.$ Bernoulli random variables, we have by the (multiplicative form of) Chernoff-Hoeffding inequality:

$$\mathbb{P}\left(\deg(C_j) > 2n^2\rho\right) \leq \exp\left(-c_0 \frac{\sqrt{n}}{\max(\log n, r)}\right)$$

for some constant c_0. In other words,

$$\mathbb{P}\left(\deg(C_j) > \frac{\sqrt{n}}{12c_1 \max(\log n, r)}\right) \leq \exp\left(-c_0 \frac{\sqrt{n}}{\max(\log n, r)}\right).$$

The same argument applies for all the rows and columns of C, and thus the same bound applies. By taking a union bounds over these rows and columns we have that:

$$\mathbb{P}\left(\deg(C) > \frac{\sqrt{n}}{12c_1 \max(\log n, r)}\right) \leq 2n \exp\left(-c_2 \frac{\sqrt{n}}{\max(\log n, r)}\right)$$

for some constant c_2. Note that $\deg(C)$ is an upper bound on $\deg_3(\boldsymbol{X})$. In an identical manner, we can bound the degrees along the first and second mode, and taking union bounds over the three modes we get the result. $\square$

Corollary 3.7. *Let $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$ where $\boldsymbol{X}$ is low rank with random factors as per the conditions of Lemma 3.5 and $\boldsymbol{Y}$ is sparse with random support as per the conditions in Lemma 3.6. Provided $r \sim o\left(n^{\frac{1}{2}}\right)$, Algorithm 1 successfully recovers the correct decomposition, i.e. $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}) = (\boldsymbol{X}, \boldsymbol{Y})$ with probability exceeding $1 - n^{-\alpha}$ for some $\alpha > 0$.*

Proof. The result follows immediately from Lemma 3.5, Lemma 3.6, and Corollary 3.4. $\square$

References

[1] A. Anandkumar, R. Ge, D. Hsu, and S. M. Kakade, *A tensor approach to learning mixed membership community models*, The Journal of Machine Learning Research, 15 (2014), pp. 2239–2312.

[2] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, Tech. Rep. 1, 2014.

[3] C. Beckmann and S. Smith, *Tensorial extensions of independent component analysis for multisubject FMRI analysis*, NeuroImage, 25 (2005), pp. 294–311.

[4] A. Bhaskara, M. Charikar, A. Moitra, and A. Vijayaraghavan, *Smoothed analysis of tensor decompositions*, in Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, 2014, pp. 594–603.

[5] S. Bhojanapalli and S. Sanghavi, *A new sampling technique for tensors*, arXiv preprint arXiv:1502.05023, (2015).

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, Journal of the ACM, 58 (2011), pp. 11–37.

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

[8] R. B. CATTELL, *Parallel proportional profiles and other principles for determining the choice of factors by rotation*, Psychometrika, 9 (1944), pp. 267–283.

[9] V. CHANDRASEKARAN, S. SANGHAVI, P. A. PARRILO, AND A. S. WILLSKY, *Rank-sparsity incoherence for matrix decomposition*, SIAM Journal on Optimization, 21 (2011), pp. 572–596.

[10] Y. CHEN, H. XU, C. CARAMANIS, AND S. SANGHAVI, *Robust matrix completion and corrupted columns*, in Proceedings of the 28th International Conference on Machine Learning (ICML-11), L. Getoor and T. Scheffer, eds., New York, NY, USA, 2011, ACM, pp. 873–880.

[11] N. GOYAL, S. VEMPALA, AND Y. XIAO, *Fourier PCA and robust tensor decomposition*, in Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, 2014, pp. 584–593.

[12] C. J. HILLAR AND L.-H. LIM, *Most tensor problems are NP-hard*, Journal of the ACM, 60 (2013), pp. 45:1–45:39.

[13] D. HSU, S. KAKADE, AND T. ZHANG, *Robust matrix decomposition with sparse corruptions*, Information Theory, IEEE Transactions on, 57 (2011), pp. 7221–7234.

[14] B. HUANG, C. MU, D. GOLDFARB, AND J. WRIGHT, *Provable models for robust low-rank tensor completion*, Pacific Journal of Optimization, 11 (2015), pp. 339–364.

[15] A. KRISHNAMURTHY AND A. SINGH, *Low-rank matrix and tensor completion via adaptive sampling*, in Advances in Neural Information Processing Systems, 2013.

[16] J. B. KRUSKAL, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Applicat., 18 (1977).

[17] V. KULESHOV, A. CHAGANTY, AND P. LIANG, *Tensor factorization via matrix factorization*, arXiv.org, (2015).

[18] S. LEURGANS, R. ROSS, AND R. ABEL, *A decomposition for three-way arrays*, SIAM Journal on Matrix Analysis and Applications, 14 (1993), pp. 1064–1083.

[19] Q. LI, A. PRATER, L. SHEN, AND G. TANG, *Overcomplete tensor decomposition via convex optimization*, in IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Cancun, Mexico, Dec. 2015.

[20] N. MESGARANI, M. SLANEY, AND S. A. SHAMMA, *Discrimination of speech from non-speech based on multiscale spectro-temporal modulations*, Audio, Speech and Language Processing, IEEE Transactions on, 14 (2006), pp. 920–930.

[21] C. MU, B. HUANG, J. WRIGHT, AND D. GOLDFARB, *Square deal: Lower bounds and improved relaxations for tensor recovery*, preprint arXiv:1307.5870, 2013.

[22] P. NETRAPALLI, U. NIRANJAN, S. SANGHAVI, A. ANANDKUMAR, AND P. JAIN, *Non-convex robust PCA*, in Advances in Neural Information Processing Systems, 2014.

[23] N. RAO, P. SHAH, AND S. WRIGHT, *Forward-backward greedy algorithms for signal demixing*, in Signals, Systems and Computers, 2013 Asilomar Conference on, IEEE, 2014.

[24] P. SHAH, N. RAO, AND G. TANG, *Optimal low-rank tensor recovery from separable measurements: Four contractions suffice*, arXiv.org, (2015).

[25] G. TANG AND P. SHAH, *Guaranteed tensor decomposition: A moment approach*, International Conference on Machine Learning (ICML 2015), (2015), pp. 1491–1500.

[26] R. TOMIOKA, K. HAYASHI, AND H. KASHIMA, *Estimation of low-rank tensors via convex optimization*, preprint arXiv:1010.0789, 2011.

[27] M. YUAN AND C.-H. ZHANG, *On tensor completion via nuclear norm minimization*, preprint arXiv:1405.1773, 2014.