A Sharp Blockwise Tensor Perturbation Bound for Orthogonal Iteration

Department of Statistics
University of Wisconsin-Madison
Madison, WI 53706, USA

Department of Statistics
University of Wisconsin-Madison
Madison, WI 53706, USA

Department of Statistics
Columbia University
New York, NY 10027, USA

Department of Statistics
University of Wisconsin-Madison
Madison, WI 53706, USA

Abstract

HS

k k q

1. Introduction

$\mathbb{R}$

$\mathbb{R}$

$\mathbb{O}$

1:

2:

$\mathbf{U}$ $\mathbf{U}$

2. Notation and Preliminaries

$\mathbb{R}$

$\mathbb{O}$

$\mathbb{O}$ $\mathbf{U}$

$\mathbb{O}$

$\mathbb{R}$ $\mathbb{R}$

$\mathbb{R}$

$\mathbb{R}$

$\mathbb{R}$

$\mathbb{R}$

$\mathbb{R}$

$\mathbb{R}$

$[\![$ $]\!]$

$\mathbb{R}$

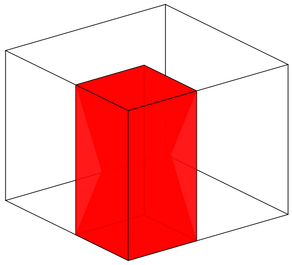
$\mathbb{R}$ $\mathbb{R}$

$[\![\quad]\!]$

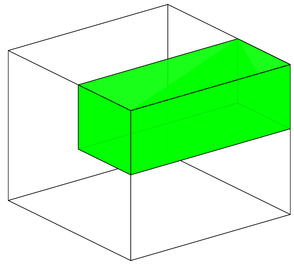
$\mathbf{V}$ $\mathbb{R}$
$\mathbf{V}$

$\mathbf{V}$ $\mathbb{R}$
$\mathbf{V}$

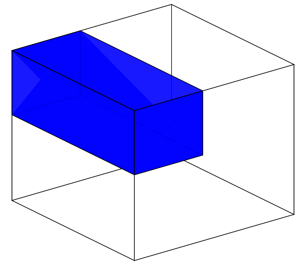
$\mathbf{V}$ $\mathbb{R}$
$\mathbf{V}$ $\mathbb{R}$
$\mathbf{V}$ $\mathbf{V}$



(a) τ



(b) τ



(c) τ

$\mathbf{U}$ $\mathbb{O}$

3. Illustration of Perturbation Bounds for $d = 3$ Asymmetric Case

---

$\mathbb{R}$ $\mathbb{O}$

1:

2:

$\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$

Consider the perturbation model with $\mathbb{R}$. Suppose . Define the block-wise errors as in and denote the initialization errors of as , . Assume the initialization error and the signal strength satisfy

and

Let $\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$ be the estimator of after steps in Algorithm 1. Then with inputs , , the mode- singular subspace updates in Algorithm 2 after iterations satisfy

and the -step tensor estimation satisfies

Moreover, when for some , the outputs of the estimated mode- singular subspaces of Algorithm 2 satisfy

$$— —— —— ——$$

and the output of tensor reconstruction satisfies

$$[\![\quad]\!]$$

Our theory relies on a lower bound assumption of the least singular value: , which is in the same vein as the classical matrix perturbation theory (Davis and Kahan, 1970; Wedin, 1972). Moreover, in the existing results on perturbation analysis for Canonical-Polyadic (CP) decomposition, e.g., Theorem 5.1 of Anandkumar et al. (2014a) and Theorem 1 of Anandkumar et al. (2014b), one assumes . Since , defined in can be seen as a counterpart of in Tucker decomposition.

In Theorem 1, we assume the initialization is warm in the sense that the maximum error is upper bounded by a constant. The constant in this upper bound is chosen for convenience and can be replaced by any fixed constant less than . Our perturbation bound applies to HOOI with any initialization as long as this condition holds, although the original HOOI algorithm was proposed with the initialization scheme named HOSVD, i.e., . Next, we briefly discuss two specific initialization schemes: HOSVD for tensor PCA/SVD (Richard and Montanari, 2014; Zhang and Xia, 2018) and diagonal-deletion SVD for tensor completion (Xia et al., 2020). For convenience of presentation, we focus on the setting .

- *(Tensor Denoising) Suppose we observe a tensor $\mathbb{R}$ and aim to recover from . To this end, we can apply HOOI by inputting . When has i.i.d. entries, Theorem 1 in Zhang and Xia (2018) showed if one initializes by HOSVD, as long as , the initialization condition — holds with high probability. Zhang and Xia (2018) also showed the signal strength requirement is essential, which means HOSVD is a proper initialization in the tensor denoising model.*

- *(Tensor Completion) Suppose we observe a set of entries, selected uniformly at random and indexed by , from a noisy tensor . Denote as*

$$\text{otherwise}$$

Suppose has i.i.d. entries. Then, it is easy to check that is an unbiased estimator of , where is the sampling ratio. Xia et al. (2020) proposed to apply HOOI on to estimate . They proposed to set as the leading singular vectors of with diagonal deletion (i.e., zero the diagonal values of) and showed that — holds with high probability when , where is the cardinality of . At the same time, they proved that HOSVD requires to achieve the same initialization performance and may not be an ideal initialization scheme for tensor completion.

In addition, the random initialization is also widely considered in the literature. For example, Anandkumar et al. (2014a) proposed to pick the best one among many random trials. It can be proved that if the number of random trials is large enough (usually polynomial in the dimension), one can find a trial such that the initialization is good enough (Anandkumar et al., 2014a).

The upper bound in includes two parts: a fixed quantity that represents the intrinsic estimation error, and another quantity that decays linearly to with respect to iteration index . The linear convergence of HOOI was observed in Ishteva et al. (2011), while Theorem 1 gives a rigorous proof for it. Note that HOOI can be viewed as a special alternating minimization method, which was shown to have asymptotic linear convergence rate in solving nonlinear least squares problems (Ruhe and Wedin, 1980). This fact also sheds light on the linear convergence of HOOI.

Our tensor perturbation bounds on singular subspace share the same spirit as the unilateral perturbation bounds on singular subspaces of matrix SVD in Cai and Zhang (2018). Consider the matrix perturbation setting mentioned in Section 2.1 with the additional assumption that is rank- and has SVD . Cai and Zhang (2018) showed that the upper bound of can be written as — —, which can be interpreted as the sum of first and second order perturbations. In Theorem 1, the upper bound of can be also written as — which can be interpreted as summation of the first, second, and third order perturbations. This phenomenon also generalizes to order- case in Theorem 3.

Due to the unilateral property, when the tensor dimension of each mode is at different order, the estimation error rate of singular subspace in each mode can vary significantly. For example in the tensor denoising setting, where $\mathbb{R}$ is a multilinear rank- tensor and is a random tensor with i.i.d. standard normal entries. Let and suppose , . Consider , then by random matrix theory (Vershynin, 2010), we can show —, and with high probability. Thus when —, Theorem 1 immediately implies, with high probability for some . So we can see the perturbation of depends on . Also as decreases, for different , different order perturbations in could dominate in the perturbation bound of . For example, when , Theorem 1 yields

for constants . More details about the application of HOOI perturbation bounds in order- tensor denoising and numerical studies for this unilateral property of singular subspace perturbation can be found in Sections 6.1 and 7.1, respectively.

It is worth mentioning that the power iteration in Algorithm 1 plays an important role for refining tensor reconstruction. Without power iteration, the estimator $\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$ with is called truncated HOSVD (T-HOSVD) in the literature (De Lathauwer et al., 2000b). It is not hard to show for some . Since and may be much larger than , we can see the power iteration can greatly improve the accuracy for tensor reconstruction, and this echos the findings in literature in tensor denoising setting (Zhang and Xia, 2018).

$[\![\quad]\!]$

Suppose $\mathbb{R}$ is a general order- tensor and $\mathbb{O}$ are general matrices with orthonormal columns. For any subset , we further define projections of on as follows,

Then,

$[\![\quad]\!]$

Consider perturbation model , we have the following deterministic lower bound for reconstructing ,

$$\overline{}$$

When tensor order is fixed, combining Theorem 1 and 2, we have shown that HOOI with good initialization is optimal for tensor reconstruction in the class . At the same time, from , we see the error rate of tensor reconstruction is optimal even after one iteration of HOOI i.e., and more iterations can improve the coefficient in front of . This suggests that in some applications where running HOOI until convergence is prohibitive, we can just run it for one iteration to get a fairly good reconstruction. See more in Section 7.2 about a numerical comparison of HOOI and one-step HOOI.

Apart from the optimality of our perturbation bound in tensor reconstruction, it is also interesting to study whether the perturbation bounds in , for singular subspaces are optimal or not and we leave it as an interesting future work.

4. A Blockwise Perturbation Bound of Higher-order Orthogonal Iteration for Tensor Decomposition

$\mathbf{V}$ $\mathbb{R}$
$\mathbf{V}$

Consider the perturbation model with $\mathbb{R}$, symmetric index groups and blockwise errors in . Suppose . Denote the initialization errors of as , .

Assume the initialization error and the signal strength satisfy

and

Let $\mathbf{U}$ $\mathbf{U}$ *be the estimator of after steps in Algorithm 1. Then with inputs , , , the mode- singular subspace updates in Algorithm 1 after iterations satisfy*

and the -step tensor estimation satisfies

Moreover, when for some , the outputs of estimated mode- singular subspace of Algorithm 1 satisfy

for where –, and the output of tensor reconstruction satisfies

It is easy to check – based on and the requirement of the signal strength . So we have in the upper bounds of and . However, in many practical applications, such as tensor denoising to be introduced in Section 6.1, tensor order is fixed and . In this case could be much smaller than – and the scale of can be very close to .

Compared with the perturbation bounds of power iteration for supersymmetric CP-low-rank decomposition (Anandkumar et al., 2014a, Theorem 5.1), our Theorem 3 covers more general symmetric and partial symmetric multilinear low-rank decomposition settings. Also in Theorem 5.1 of Anandkumar et al. (2014a), the tensor reconstruction error bound of power iteration is given in terms of tensor spectral norm, which does not improve upon the guarantee by the trivial estimator . On the other hand, the tensor reconstruction error of in Theorem 3 is given in Hilbert-Schmidt norm and can be significantly better than the guarantee for as in most of the applications.

We note that in Theorem 3, the constants in our condition and perturbation bounds and scales exponentially w.r.t. the tensor order . We think this exponential dependence on is not sharp. In fact, in Theorem 1 of the recent work Luo and Zhang (2021), they show the dependence on in and can be reduced to poly .

We provide a sketch on how to prove and . The rest of the results and follow easily from , by plugging in . The idea is to develop the recursive error bounds of , i.e., the estimate of at iteration , based on the error bound of , i.e., the estimate at iteration . The argument can be divided into three steps. It is worth mentioning that all three steps involves complex tensor algebra and this makes the proof even more difficult.

First, we denote

. In HOOI procedure, the update for the mode- singular subspace satisfies

here . To give an upper bound for , we aim to give an upper bound for

by using , . The main idea to bound is to introduce **U U** *in each mode multiplication, expand the mode products, then write the whole term into summation of many small terms.*

After getting an upper bound for , we use induction to prove the following claim, *One technical difficulty is to deal with the sequential updating of singular subspaces in HOOI and we use the induction idea again to tackle it. Tools we use in this step include the singular subspace bound in (Luo et al., 2020, Theorem 5).*

The final and most challenging step involves upper bounding the tensor reconstruction error $\mathbf{U}$ $\mathbf{U}$ *by the unified quantity . By decomposing onto the estimated singular subspaces, we can show that*

$$\mathbf{U} \qquad \mathbf{U}$$

$$\mathbf{U} \qquad \mathbf{U}$$

By definition, $\mathbf{U}$ $\mathbf{U}$ *. We further show*

Here is a quantity that depends on . The main challenge to prove (a) is that is not the left singular subspace of . So to leverage the SVD property of , we have to project onto and , then use the subspace perturbation bounds established before.

5. Perturbation Bounds of Power Iteration for Tensors with Partial Low Multilinear Rank Structure

$\mathbb{R}$

$\mathbb{R}$ $\mathbb{O}$

$\mathbb{R}$ $\mathbb{O}$

1:

2:

$\mathbf{U}$ $\mathbf{U}$

$\mathbf{V}$ $\mathbb{R}$ $\mathbf{V}$

$\mathbf{V}$ $\mathbb{R}$ $\mathbf{V}$

Consider the perturbation model with $\mathbb{R}$. Suppose . Define the blockwise errors as in and denote the initialization errors of as , . Assume the initialization error and the signal strength satisfy

$$-$$

and

Then with inputs , , the mode- singular subspace updates in Algorithm 3 after iterations satisfy

$$-$$

Moreover, when for some , the outputs of estimated mode-singular subspace of Algorithm 3 satisfy

$$-$$

The output of tensor reconstruction satisfies

6. Implications in Statistics and Machine Learning

Consider the tensor denoising problem and Algorithm 1 with inputs , , initialization and for some , where is the minimal singular value of each matricization of . Assume and . Then if , with probability at least , the output satisfy

and

for some constants .

$\mathbb{M}$ $\mathbb{M}$

$\overline{}$
J E

J E

Suppose has the tensor co-clustering/block structure , where is a multilinear rank- tensor. Assume with $\mathbb{O}$ *is the Tucker decomposition of* $^{-}$ $^{-}$*. Then*

with $^{-}$ $\mathbb{O}$ *for* .

$\mathbb{R}$ $\mathbb{R}$

$\mathbb{M}$

1:

2:

$\mathbb{M}$ $\mathbb{R}$

$\mathbf{\Pi}$ $\mathbb{M}$ $\mathbf{X}$ $\mathbb{R}$

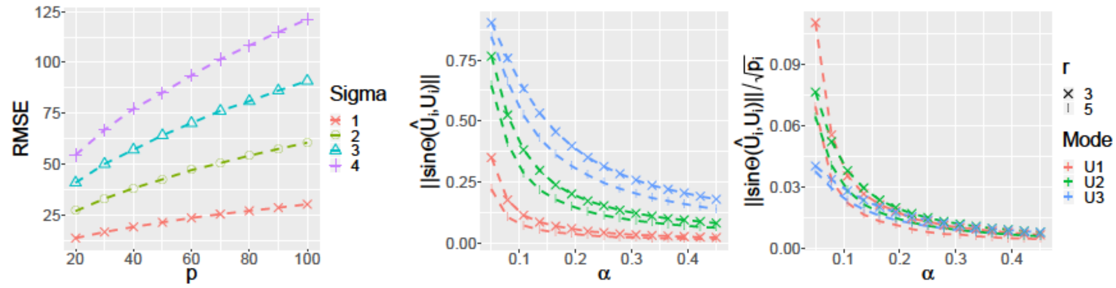
Consider the tensor co-clustering/block model “ ” and the Algorithm 4 with inputs , initializations and for some , where is the minimal singular value at each matricization of the core tensor parameter . Assume , , and holds. Then if

for sufficiently large constant , with probability at least for some , , satisfy

and we also have the following upper bound on cocluster recovery error,

Here are some constants depending only on and , is the second largest cocluster size at mode .

7. Numerical Studies



(a) Tensor Reconstruction

(b) Convergence of Singular Subspace

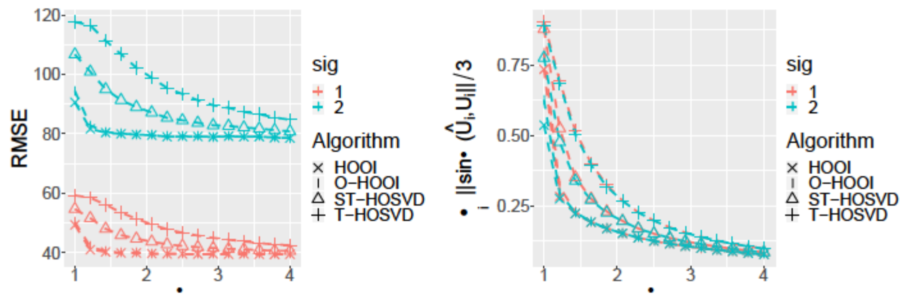
Figure 2: HOOI with good initialization. (a) Tensor reconstruction error $\|\widehat{\mathcal{T}} - \mathcal{T}\|_{\rm HS}$ for $p \in \{20, 30, \ldots, 100\}, r = 5, \sigma \in \{1, 2, 3, 4\}$ and $\lambda = 5\sqrt{pr}\sigma$; (b) Mode-$k$ singular subspace estimation with and without rescaling under $p_1 = 10, p_2 = 100, p_3 = 500$, $r \in \{3, 5\}, \sigma = 1$ and $\lambda = \alpha \cdot p_3 \frac{\sqrt{r_1}}{\sqrt{p_1}}$ with varying α

7.1 Perturbation Bounds of HOOI with good initialization

In this simulation, we study the perturbation bounds of HOOI with randomly generated good initialization. Let $\mathcal{T} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$, where $\mathbf{U}_i \in \mathbb{R}^{p_i \times r}$ is generated uniformly at random from $\mathbb{O}_{p_i,r}$ and $\mathcal{S} \in \mathbb{R}^{r\times r\times r}$ is a diagonal tensor with diagonal values $\{i\lambda\}_{i=1}^r$. The initializations of $\mathbf{U}_i$ of Algorithm 1 are $\widetilde{\mathbf{U}}_i^{(0)} = \frac{1}{\sqrt{2}}\mathbf{U}_i + \frac{1}{\sqrt{2}}\mathbf{U}_i'$, where $\mathbf{U}_i' = \mathbf{U}_{i\perp}\mathbf{O}$ for some random orthogonal matrix $\mathbf{O} \in \mathbb{O}_{p_i - r, r}$. It is easy to check that $\|\sin\Theta(\mathbf{U}_i, \widetilde{\mathbf{U}}_i^{(0)})\| = \frac{\sqrt{2}}{2}$ for $i = 1, 2, 3$.

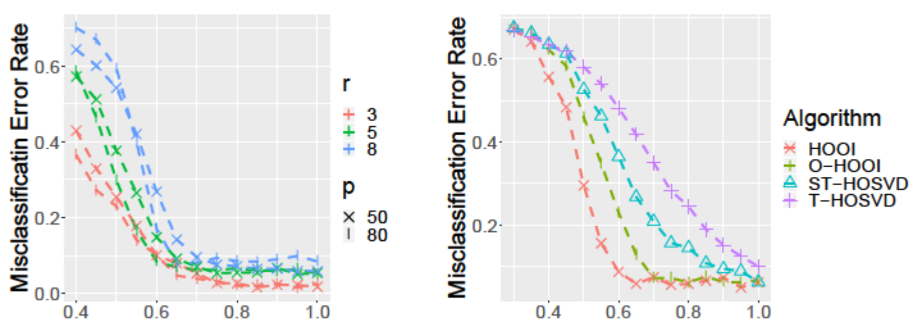
First for tensor reconstruction, let $p \in \{20, 30, \ldots, 100\}, r = 5$, $\sigma \in \{1, 2, 3, 4\}$ and $\lambda = 5\sqrt{pr}\sigma$. We can check that with high probability, $\|\mathcal{Z}\|_{\rm HS} \leqslant Cp^{\frac{3}{2}}\sigma$ and $\xi \leqslant C\sqrt{pr}\sigma$ for some $C > 0$ following the same proof as Theorem 5. In Figure 2(a), the RMSE of tensor reconstruction of HOOI is presented. We find as the perturbation results in Section 4 suggest, $\|\widehat{\mathcal{T}} - \mathcal{T}\|_{\rm HS}$ can be much smaller than $\|\mathcal{Z}\|_{\rm HS}$. This demonstrates the superior performance of the HOOI estimator compared to the trivial estimator $\widetilde{\mathcal{T}}$. At the same time, the RMSE for tensor reconstruction increases as p and σ become bigger and this matches our theoretical findings in Theorem 3 that the error bound of HOOI for $\|\widehat{\mathcal{T}} - \mathcal{T}\|_{\rm HS}$ is $O(\xi)$, which increases as p, σ increase.

Next we demonstrate the unilateral perturbation bounds for mode-k singular subspace estimation. Specifically, we consider $p_1 = 10, p_2 = 100, p_3 = 500$, $r \in \{3, 5\}, \sigma = 1$ and $\lambda = \alpha \cdot p_3 \frac{\sqrt{r}}{\sqrt{p_1}}$ with varying α. The errors of the mode-1, mode-2, mode-3 estimated singular subspaces with and without rescaling are provided in Figure 2(b). We can see from Figure 2(b) left panel the errors of estimated singular subspaces converge to different values depending on the corresponding mode size p_i, and a further rescaling of estimation error by $\sqrt{p_i}$ makes them roughly on the same level (see Figure 2(b) right panel). This matches the unilateral property of the singular subspace perturbation results in Remark 4 that when $\lambda = O(p_3 \frac{\sqrt{r}}{\sqrt{p_1}})$, $\|\sin\Theta(\widehat{\mathbf{U}}_k, \mathbf{U}_k)\| \leqslant C\frac{\sqrt{p_i}}{\lambda/\sigma}$ for some $C > 0$, and this upper bound increases linearly with respect to $\sqrt{p_i}$.



(a) Tensor Reconstruction (b) Singular Subspace Estimation

Figure 3: Comparison of HOOI, one-step HOOI (O-HOOI), truncated HOSVD (T-HOSVD), sequentially truncated HOSVD (ST-HOSVD) in tensor denoising under $p = 100$, $r = 5$, $\sigma \in \{1, 2\}$, $\lambda = \alpha \cdot p^{\frac{3}{4}}\sigma$ with $\alpha \in [1, 4]$. (a) Tensor reconstruction; (b) Averaged singular subspace estimation



(a) HOOI for Cocluster Recovery (b) Comparison in Cocluster Recovery

Figure 4: Tensor cocluster recovery under $\sigma = 1$, $\lambda = \alpha \cdot \frac{r^{3/2}}{p^{3/4}}\sigma$ with varying α. (a) HOOI on tensor cocluster recovery under $p \in \{50, 80\}$, $r \in \{3, 5, 8\}$. (b) Comparison of HOOI, one-step HOOI (O-HOOI), truncated HOSVD (T-HOSVD), sequentially truncated HOSVD (ST-HOSVD) in cocluster recovery under $p = 80, r = 5$.

8. Conclusion and Discussion

Acknowledgments

Appendix A. Tensor Perturbation Bounds for HOOI in Asymmetric Case

Consider the perturbation model with ℝ *and* , .
Define *as the set of all possible index sets with* *elements from* *and* . *For* , *let* . *Now we define the blockwise errors as*

$\mathbf{V}$ ℝ
$\mathbf{V}$

for

Denote the initialization errors of *as* , . *Assume the initialization error and the signal strength satisfy*

and

where .

Then with inputs , , , *the estimated mode-* *singular subspaces updates in Algorithm 1 after* *iterations satisfy*

Moreover, when , *the outputs of estimated mode-* *singular subspaces of Algorithm 1 satisfy* *for* , *where* —, *and the output of tensor reconstruction satisfies*

$$[\![\quad]\!]$$

Appendix B. Additional Proofs

$\mathbf{U}$ $\mathbb{O}$

$\mathbf{U}$ $\mathbb{O}$

$\mathbf{U}$ $\mathbb{O}$

$\mathbb{R}$

$\mathbf{U}$ $\mathbb{O}$

■

$\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$ $\mathbf{U}$

$\mathbf{U}$ $\mathbf{U}$

$\mathbf{U}$ $\mathbf{U}$

$\llbracket$ $\rrbracket$ $\mathbf{U}$ $\mathbf{U}$

$\blacksquare$

$\mathbb{R}$

$\blacksquare$

U U

U U U U

U U

U U U

U U

U U

U U

U U

U U U

U U U

U U

U U

U U U

U U

U U

— U U — U

—

—

U U U

U

$$\frac{\mathbf{U}}{\mathbf{U}} \qquad \frac{\mathbf{U}}{\mathbf{U}}$$

$$\frac{\mathbf{U} \qquad \mathbf{U}}{\mathbf{U} \qquad \mathbf{U}}$$

U **U**

U **U** **U** **U** **U** **U**

U **U** **U** **U** **U**

U **U** **U**

U

U **U** **U** **U**

U **U**

U **U** **U** **U**

U **U**

$\mathbf{X}$ $\mathbf{X}$

$\mathbf{U}$ $\mathbf{U}$

$\mathbf{U}$ $\mathbf{U}$

$\mathbf{U}$ $\mathbf{U}$

References

Annals of Statistics

IEEE Journal of Selected Topics in Signal Processing

Artificial Intelligence and Statistics

Journal of Machine Learning Research

arXiv preprint arXiv:1402.5180

arXiv preprint arXiv:2007.09024

Available online, January

arXiv preprint arXiv:1807.06693

Numerical linear algebra with applications

Advances in Mathematics

Journal of Multivariate Analysis

Matrix analysis

The Annals of Statistics

The Annals of Statistics

Journal of the American Statistical Association

Biometrika

The Annals of Statistics

Conference on Learning Theory

The Journal of Machine Learning Research

arXiv preprint arXiv:1811.12804

Journal of Machine Learning Research

IEEE Transactions on Information Theory

Conference on Learning Theory

IEEE Signal Processing Magazine

SIAM Journal on Numerical Analysis

Journal of Chemometrics: A Journal of the Chemometrics Society

Linear Algebra and its Applications

SIAM journal on Matrix Analysis and Applications

SIAM journal on Matrix Analysis and Applications

Psychometrika

SIAM Journal on Matrix Analysis and applications

SIAM journal on matrix analysis and applications

arXiv preprint arXiv:1706.06516

Journal of Machine Learning Research

Advances in Neural Information Processing Systems

The Annals of Statistics

SIAM Journal on Matrix Analysis and Applications

Linear Algebra and its applications

Geometric aspects of functional analysis

SIAM Journal on Matrix Analysis and Applications

GAMM-Mitteilungen

Advances in Neural Information Processing Systems

Journal of Fourier analysis and applications

arXiv preprint arXiv:2002.11255

IEEE Transactions on Information Theory

Journal of the ACM (JACM)

Conference on Learning Theory

Numerical Algorithms

SIAM Journal on Matrix Analysis and Applications

International Conference on Algorithmic Learning Theory

arXiv preprint arXiv:2002.04457

arXiv preprint arXiv:1909.06503

arXiv preprint arXiv:1807.02884

SIAM Journal on Matrix Analysis and Applications

SIAM review

2008 Eighth IEEE international conference on data mining

High Dimensional Probability VII

Applied multiway data analysis

Psychometrika

45th Annual IEEE Symposium on Foundations of Computer Science

arXiv preprint arXiv:2002.06524

The Annals of Statistics

Biometrika

Journal of Statistical Software

2010 IEEE International Conference on Image Processing

arXiv preprint arXiv:1702.07449

Advances in Neural Information Processing Systems

IEEE transactions on Neural Networks

arXiv preprint arXiv:2005.10743

arXiv preprint arXiv:2104.12031

arXiv preprint arXiv:2008.01312

Advances in Neural Information Processing Systems

The quarterly journal of mathematics

SIAM Journal on Matrix Analysis and Applications

Linear Algebra and its Applications

SIAM Journal on Scientific Computing

SIAM Journal on Scientific Computing

IEEE transactions on signal processing

Annales de l'Institut Henri Poincaré, Probabilités et Statistiques

The Annals of Statistics

Advances in Neural Information Processing Systems

The Annals of Statistics

SIAM review

SIAM Journal on Scientific Computing

Journal of Multivariate Analysis

Advances in Neural Information Processing Systems

Proceedings of the 2007 SIAM International Conference on Data Mining

IEEE Transactions on Signal Processing

Available online, January

Linear algebra and its applications

Proceedings of the 2009 SIAM International Conference on Data Mining

SIAM Journal on Mathematics of Data Science

Journal of the American Statistical Association

Proceedings of the 2010 SIAM International Conference on Data Mining

Psychometrika

Pacific Journal Of Optimization

SIAM Journal on Scientific Computing

arXiv preprint arXiv:1011.3027

Random Structures & Algorithms

IEEE Transactions on Knowledge and Data Engineering

SIAM Journal on Matrix Analysis and Applications

Advances in Neural Information Processing Systems

SIAM Journal on Matrix Analysis and Applications

BIT Numerical Mathematics

Mathematische Annalen

Advances in Neural Information Processing Systems

Foundations of Computational Mathematics

Journal of Machine Learning Research

The Annals of Statistics, to appear

2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)

Linear and Multilinear Algebra

2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)

IEEE Transactions on Image Processing

Biometrika

Foundations of Computational Mathematics

IEEE Transactions on Information Theory

The Annals of Statistics

Journal of the American Statistical Association

IEEE Transactions on Information Theory

SIAM Journal on Mathematics of Data Science

SIAM Journal on Matrix Analysis and Applications

Advances in Neural Information Processing Systems

Journal of the American Statistical Association

arXiv preprint arXiv:2010.02482