Structured Matrix Approximations via Tensor Decompositions

Misha E. Kilmer, Arvind K. Saibaba

Tufts University, North Carolina State University

February 26, 2022

This material is partly based upon work supported by the National Science Foundation under Grant No. DMS-1439786 and by the Simons Foundation Grant No. 5073.

Block-Structured matrices

Block-structured matrices (e.g., block Toeplitz/block Hankel) arise in many applications:

1. Signal processing
2. Finite difference discretizations of PDEs
3. Geostatistical/Spatiotemporal statistical applications
4. Image deblurring

Block-Structured matrices

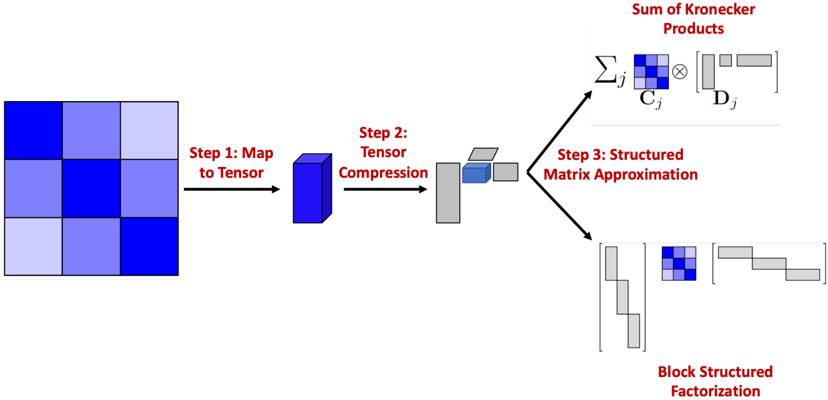
Block-structured matrices (e.g., block Toeplitz/block Hankel) arise in many applications:

1. Signal processing
2. Finite difference discretizations of PDEs
3. Geostatistical/Spatiotemporal statistical applications
4. Image deblurring

Our approach: Use tensor decompositions to

1. provide a unified approach for handling structured matrices
2. leverage inherent multidimensional structure, and
3. produce accurate and efficient matrix approximations

The talk in one slide

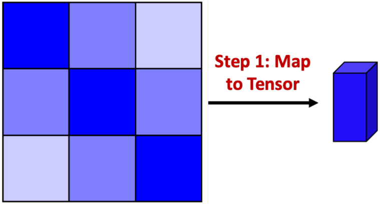


Applications:

1. System identification
2. Space-time covariance matrices

Extensions to multilevel structure

Step 1: Mapping matrices to tensors



Consider a block matrix $\mathbf{A} \in \mathbb{R}^{(\ell m)\times(nq)}$ with $\ell \times q$ blocks of size $m \times n$ each.

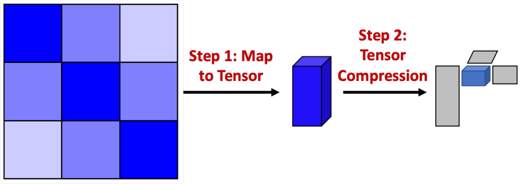
Idea: We identify

- the unique set of blocks $(\mathbf{A}_1, \dots, \mathbf{A}_p)$.
- the locations of the blocks and frequency of appearance, in a data structure $\mathcal{E}$.

Construct a 3D tensor: $\mathcal{T}_{\mathcal{E}}[\mathbf{A}] \in \mathbb{R}^{m\times p\times n}$

Advantage of our approach: treat all the structured matrices in the same framework.

Step 2: Tensor Compression

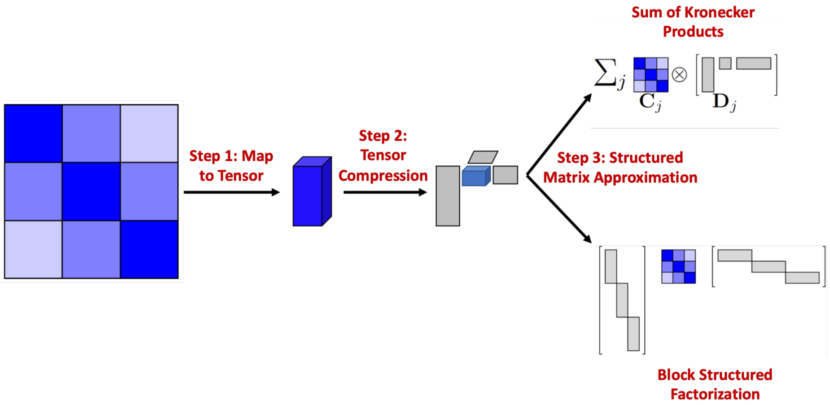


- Tucker format
  1. Higher Order Singular Value Decomposition (HOSVD), Sequentially Truncated HOSVD, Higher Order Orthogonal Iteration
  2. Randomized Algorithms for Tucker decomposition

- CP format
  1. Alternating least squares

Kolda, Bader, SIAM Review, 2009. Cichocki, Foundations and Trends in Machine Learning, 2016. Minster, Saibaba, Kilmer, SIMODS, 2020.

Step 3: Mapping compressed tensors to matrices

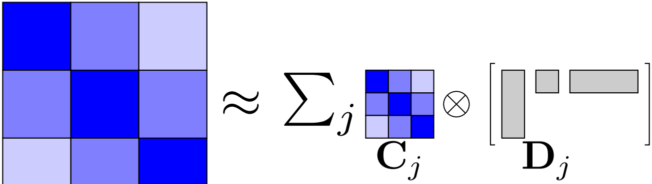


Recovering structured matrix approximations

Suppose we have the compressed tensor in Tucker form

$$\mathcal{T}_{\mathcal{E}}[\mathbf{A}] \approx \widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}] := [\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W}]$$

with rank (r_1, r_2, r_3) we can approximate



Sum of Kronecker products

$$\mathcal{M}_{\mathcal{E}}[\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]] = \sum_{j=1}^{r_2} \mathbf{C}_j \otimes (\mathbf{U}\mathtt{sq}(\mathcal{G}_{:,j,:})\mathbf{W}^\top).$$

Here $\mathbf{C}_j = \sum_{k=1}^{p} \mathbf{E}_k \otimes v_{kj}$ has the same structure as $\mathbf{A}$

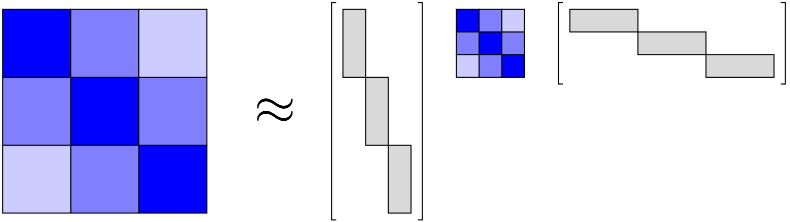
Similar expressions can be derived when CP decomposition is used.

Recovering structured matrix approximations

Suppose we have the compressed tensor in Tucker form

$$\mathcal{T}_{\mathcal{E}}[\mathbf{A}] \approx \widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}] := [\mathcal{G}; \mathbf{U}, \mathbf{V}, \mathbf{W}]$$

with rank (r_1, r_2, r_3) we can approximate



Block-structured format

$$\mathcal{M}_{\mathcal{E}}[\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]] = (\mathbf{I} \otimes \mathbf{U})\mathbf{M}(\mathbf{I} \otimes \mathbf{W}^\top).$$

Here $\mathbf{M} \in \mathbb{R}^{(r_1 \ell) \times (r_3 q)}$ has the same structure as $\mathbf{A}$

Similar expressions can be derived when CP decomposition is used.

Error in the matrix approximation

Let $\mathbf{A} \in \mathbb{R}^{(\ell m)\times(qn)}$ and let $\mathcal{T}_{\mathcal{E}}[\cdot]$ and $\mathcal{M}_{\mathcal{E}}[\cdot]$ be the matrix-to-tensor and tensor-to-matrix mappings respectively.

Theorem (Kilmer, S.)

Let $\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}] \approx \mathcal{T}_{\mathcal{E}}[\mathbf{A}]$ be a tensor approximation computed using any appropriate method. Then the error in the matrix approximation satisfies

$$\|\mathbf{A} - \mathcal{M}_{\mathcal{E}}[\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]]\|_F = \|\mathcal{T}_{\mathcal{E}}[\mathbf{A}] - \widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]\|_F.$$

Error in the matrix approximation

Let $\mathbf{A} \in \mathbb{R}^{(\ell m)\times(qn)}$ and let $\mathcal{T}_{\mathcal{E}}[\cdot]$ and $\mathcal{M}_{\mathcal{E}}[\cdot]$ be the matrix-to-tensor and tensor-to-matrix mappings respectively.

Theorem (Kilmer, S.)

Let $\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}] \approx \mathcal{T}_{\mathcal{E}}[\mathbf{A}]$ be a tensor approximation computed using any appropriate method. Then the error in the matrix approximation satisfies

$$\|\mathbf{A} - \mathcal{M}_{\mathcal{E}}[\widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]]\|_F = \|\mathcal{T}_{\mathcal{E}}[\mathbf{A}] - \widehat{\mathcal{T}}_{\mathcal{E}}[\mathbf{A}]\|_F.$$

Main message:

1. The error in the tensor approximation equals error in matrix approximation in the Frobenius norm
2. The error is independent of the particular format/tensor decomposition that is used
3. The resulting matrix approximations are efficient to store and easy to work with

Tests from SuiteSparse Collection

Each matrix is of size $(\ell n) \times (\ell n)$ and is block tridiagonal.

Name	ℓ	n	Target rank r	Relative Error	Compression
`pde2961`	63	47	20	8.36×10^{-10}	0.4470
`t2d_q4`	99	99	5	2.93×10^{-15}	0.034
`t2d_q9`	99	99	5	2.93×10^{-15}	0.034
`fv2`	99	99	5	2.28×10^{-15}	0.034
`chem_master1`	201	201	5	1.79×10^{-15}	0.030
`ecology1`$^{(*)}$	500	1000	5	6.10×10^{-15}	0.009

We report the name of the matrix, the number of block rows ℓ, the size of each block n, the target rank used, the relative error and the compression ratio. $(*)$ used the leading principal submatrix of size 500000×500000.

System Identification

Consider the linear time invariant system

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k \\ \boldsymbol{y}_k &= \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{D}\boldsymbol{u}_k \end{aligned} \qquad k = 0, 1, \dots$$

In the impulse response case, we are given data of the form of *Markov parameters*

$$\boldsymbol{h}_j = \begin{cases} \boldsymbol{D} & j = 0 \\ \boldsymbol{C}\boldsymbol{A}^{j-1}\boldsymbol{B} & j = 1, 2, \dots, \end{cases}$$

Goal

Given the Markov parameters $\{\boldsymbol{h}_k\}$ recover the system matrices $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ (up to a similarity transformation).

Up to a similarity transformation $(\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}, \boldsymbol{T}\boldsymbol{B}, \boldsymbol{C}\boldsymbol{T}^{-1}, \boldsymbol{D})$.

Eigensystem Realization Algorithm

Form the block-Hankel matrix $\boldsymbol{\mathcal{H}}_s$ defined as

$$\boldsymbol{\mathcal{H}}_s = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \dots & \boldsymbol{h}_s \\ \boldsymbol{h}_2 & \boldsymbol{h}_3 & \dots & \boldsymbol{h}_{s+1} \\ \vdots & \vdots & \iddots & \vdots \\ \boldsymbol{h}_s & \boldsymbol{h}_{s+1} & \dots & \boldsymbol{h}_{2s-1} \end{bmatrix} \in \mathbb{R}^{(ms)\times(ns)} \tag{1}$$

Assume $d \ll s$, such that $\mathsf{rank}(\boldsymbol{\mathcal{H}}_s) = d \leq \min\{sm, sn\}$.

Kung, 1978. Juang and Pappa, 1985.

Eigensystem Realization Algorithm

Form the block-Hankel matrix $\boldsymbol{\mathcal{H}}_s$ defined as

$$\boldsymbol{\mathcal{H}}_s = \begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \dots & \boldsymbol{h}_s \\ \boldsymbol{h}_2 & \boldsymbol{h}_3 & \dots & \boldsymbol{h}_{s+1} \\ \vdots & \vdots & \iddots & \vdots \\ \boldsymbol{h}_s & \boldsymbol{h}_{s+1} & \dots & \boldsymbol{h}_{2s-1} \end{bmatrix} \in \mathbb{R}^{(ms)\times(ns)} \tag{1}$$

Assume $d \ll s$, such that $\mathsf{rank}(\boldsymbol{\mathcal{H}}_s) = d \le \min\{sm, sn\}$.

Algorithm: Given target rank $r \le d$

1. Compute the reduced-SVD $\boldsymbol{\mathcal{H}}_s \approx \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^\top$
2. Partition the left singular vectors

$$\boldsymbol{U}_r = \begin{bmatrix} \boldsymbol{\Upsilon}_f \\ * \end{bmatrix} = \begin{bmatrix} * \\ \boldsymbol{\Upsilon}_l \end{bmatrix}$$

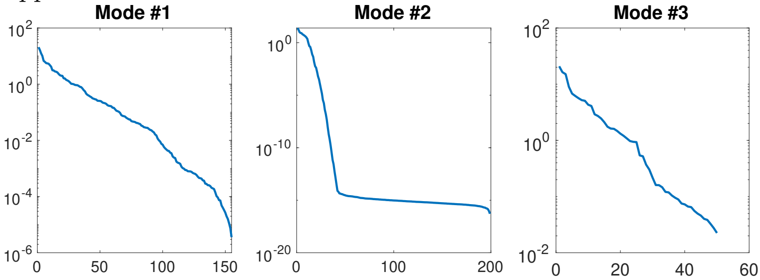
3. Compute $\boldsymbol{A}_r = \boldsymbol{\Sigma}_r^{-1/2} \boldsymbol{\Upsilon}_f^\dagger \boldsymbol{\Upsilon}_l \boldsymbol{\Sigma}_r^{1/2}$. Recover $\boldsymbol{B}_r, \boldsymbol{C}_r$ from the SVD.

Kung, 1978. Juang and Pappa, 1985.

Numerical Results: Power systems

System	System Size	Inputs n	Outputs m	Target rank
Power System	155	50	155	75

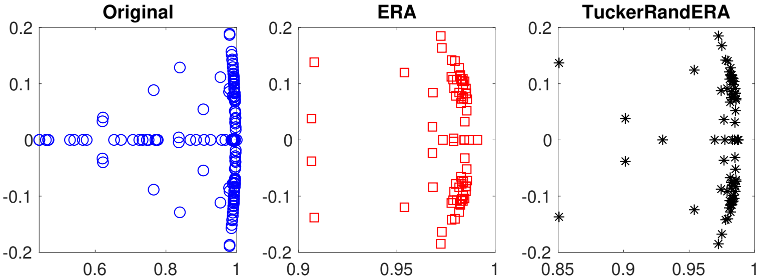
Our approach: use the tensor-based framework for block-Hankel matrices.



Decay of the singular values in each mode. Based on this decay we chose the target rank $(56, 50, 30)$.

Minster, Saibaba, Kar, Chakrabortty, SIMAX, 2021.

Numerical Results: Power systems



Computational runtime (seconds) and accuracy (Hausdorff distance)

s	Size	ERA	RandERA	TuckerERA	Error
100	15500×5000	68	1.83	0.79	0.05
200	31000×10000	–	3.95	1.52	0.015
700	108500×35000	–	15.10	6.62	0.01
1000	155000×50000	–	20.21	10.38	0.01

Multilevel approximations

Structured matrices may have recursive structure. Examples:

- Block-Toeplitz with Toeplitz Blocks
- Triply block Toeplitz

Suppose $\mathbf{A}$ has L levels of structure. Write

$$\mathbf{A} = \sum_{i_1=1}^{p_1} \cdots \sum_{i_L=1}^{p_L} \mathbf{E}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{E}_{i_L}^{(L)} \otimes \sqrt{\eta_{i_1}^{(1)} \cdots \eta_{i_L}^{(L)}} \mathbf{A}^{(i_1,\dots,i_L)},$$

where

- the matrices $\mathbf{A}^{(i_1,\dots,i_L)}$ are the $m \times n$ non-redundant blocks at level L
- the matrices $\mathbf{E}_k^{(j)} \in \mathbb{R}^{\ell_j \times q_j}$ represent mapping matrices at level j

Multilevel approximations

Structured matrices may have recursive structure. Examples:

- Block-Toeplitz with Toeplitz Blocks
- Triply block Toeplitz

Suppose $\mathbf{A}$ has L levels of structure. Write

$$\mathbf{A} = \sum_{i_1=1}^{p_1} \cdots \sum_{i_L=1}^{p_L} \mathbf{E}_{i_1}^{(1)} \otimes \cdots \otimes \mathbf{E}_{i_L}^{(L)} \otimes \sqrt{\eta_{i_1}^{(1)} \cdots \eta_{i_L}^{(L)}} \mathbf{A}^{(i_1,\dots,i_L)},$$

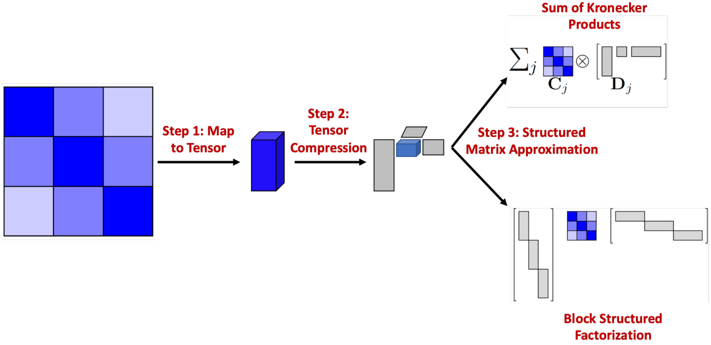
where

- the matrices $\mathbf{A}^{(i_1,\dots,i_L)}$ are the $m \times n$ non-redundant blocks at level L
- the matrices $\mathbf{E}_k^{(j)} \in \mathbb{R}^{\ell_j \times q_j}$ represent mapping matrices at level j

Remarks:

- We work with tensors (and decompositions) of order $L+2$
- We can extend our approach to handle arbitrary number of levels and different structures at each level
- Many connections to Tensor Train and Matrix Product Operators.

Contributions



- A new, unified approach for structured matrix approximations that leverages tensor decompositions
- Extensions to multilevel structures possible
- Applications: System identification, spacetime covariances, image deblurring

Thank you!

Preprint: M.E. Kilmer and A.K. Saibaba, Structured Matrix Approximations via Tensor Decompositions. arXiv preprint:
`https://arxiv.org/abs/2105.01170`

This material is partly based upon work supported by the National Science Foundation under Grant No. DMS-1439786 and by the Simons Foundation Grant No. 5073.