Nearest Kronecker Product

Approximate Matrix Factorization

John Daly, 27 July 2021

Nearest Kronecker Product (NKP) Matrix Factorization

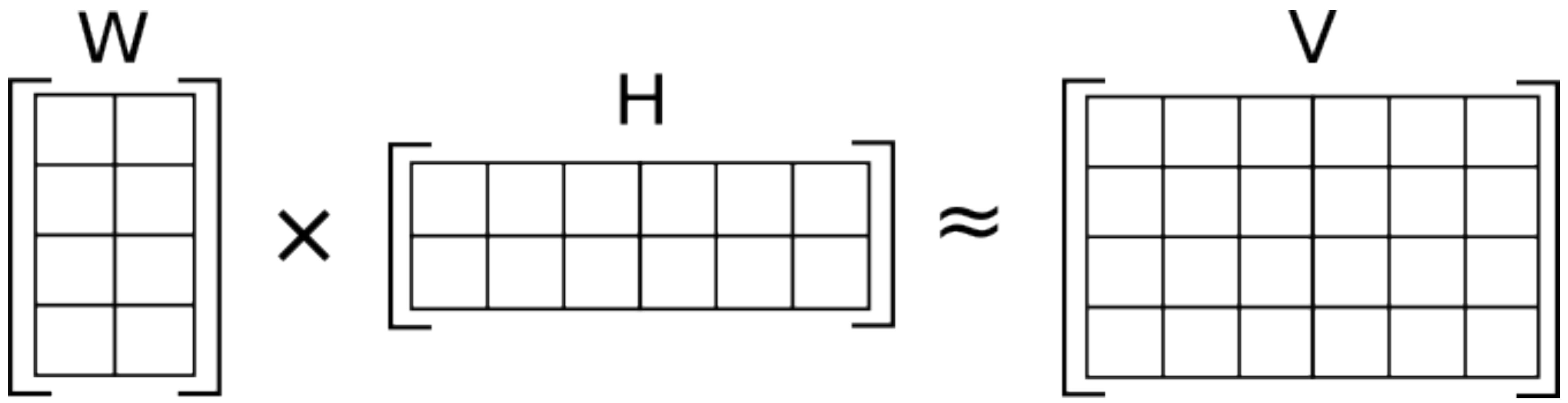
Why did I think this would be an interesting problem to explore?

- I am completely out of my mind!
  - Umm… not much else to add here
- NKP is an up-and-coming approach to matrix approximation
  - Effectively compresses large and dense matrices
  - Works as a pre-conditioner for dense linear algebra
  - Speeds up orthogonal linear projection in artificial neural networks
  - Used for image classification, compression, and denoising
- NKP may be a useful abstraction for LARC
  - Basis of block recursive matrix and tensor representations
  - Good indicator of the expected amount of matrix compression
  - Quantifies tradeoff in size and accuracy of the compressed format

But First, Non-Negative Matrix Factorization (NMF)

Or, what I learned from reading wikipedia

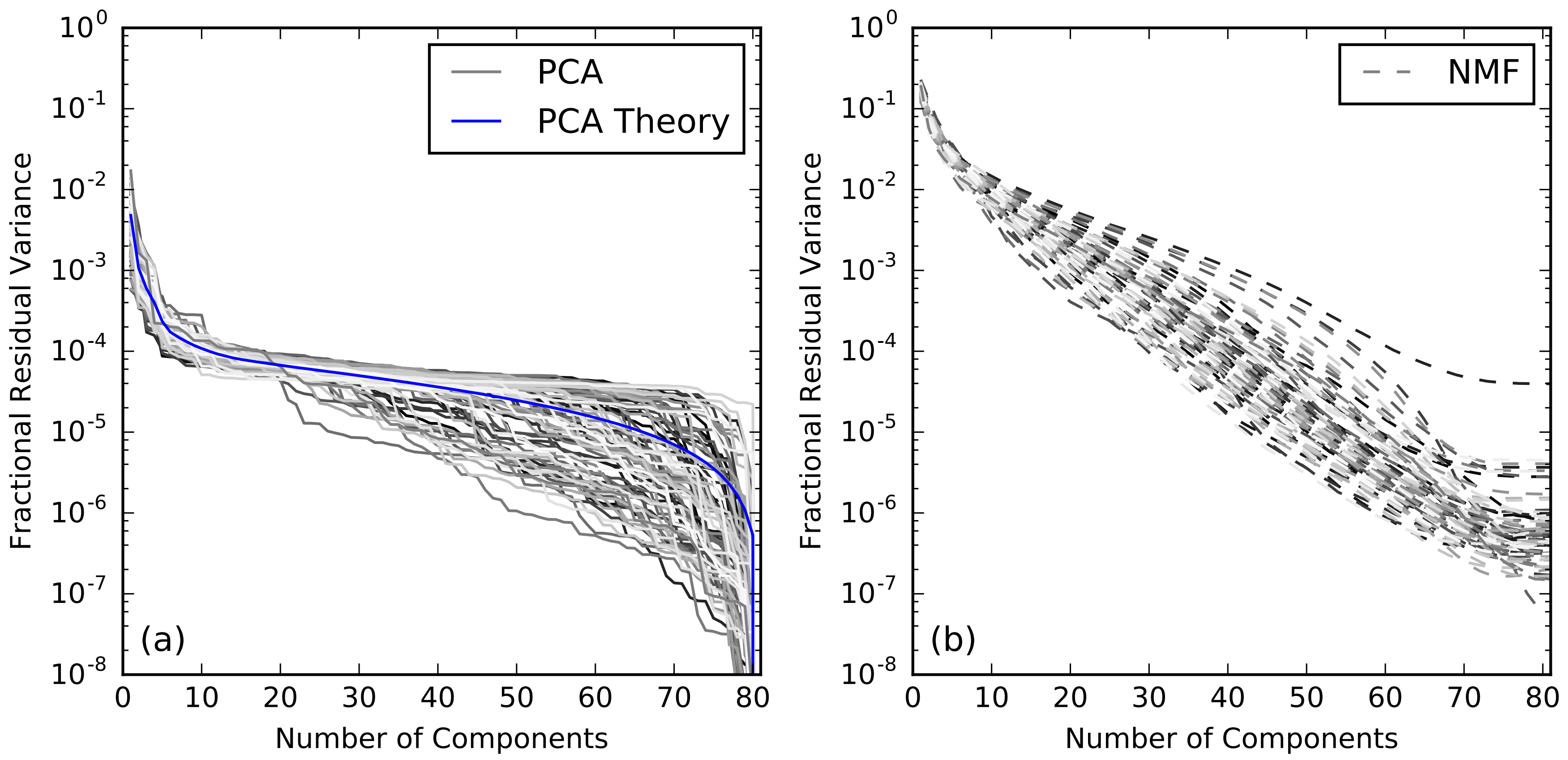
- For three matrices with positive elements, write $\mathbf{V}$ in terms of reduced rank matrices $\mathbf{W}$ and $\mathbf{H}$



- Accuracy based on some kind of matrix norm, for example

$$\phi(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|_F^2$$

- In general NP-hard to solve, but accurate factorizations may
  - Infer "latent structure" in data
  - Be easier to compute with
- NMF compares well to principal component analysis (PCA)



The Nearest Kronecker Product Problem

The SVD Method of Van Loan and Pitsianis

- Suppose $M \in \mathfrak{R}^{m \times n}$ is given with $m = m_1 m_2$ and $n = n_1 n_2$ then the NKP problem is to find $A \in \mathfrak{R}^{m_1 \times n_1}$ and $B \in \mathfrak{R}^{m_2 \times n_2}$ the minimize the Frobenius norm

$$\phi(A, B) = \|M - A \otimes B\|_F = \|\tilde{M} - \text{vec}(A)\,\text{vec}(B)^T\|_F$$

- Van Loan and Pitsianis show how to solve this using the singular value decomposition (SVD) of a permuted version of M

$$\phi(A,B) = \left\| \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \\ m_{51} & m_{52} & m_{53} & m_{54} \\ m_{61} & m_{62} & m_{63} & m_{64} \end{bmatrix} - \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \right\|_F = \left\| \begin{bmatrix} m_{11} & m_{21} & m_{12} & m_{22} \\ m_{31} & m_{41} & m_{32} & m_{42} \\ m_{51} & m_{61} & m_{52} & m_{62} \\ m_{13} & m_{23} & m_{14} & m_{24} \\ m_{33} & m_{43} & m_{34} & m_{44} \\ m_{53} & m_{63} & m_{54} & m_{64} \end{bmatrix} - \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{21} & b_{12} & b_{22} \end{bmatrix} \right\|_F$$

- This reduces the NKP problem to a rank-1 approximation problem that is solved by computing the SVD from $U^T \tilde{M} V = \Sigma$ and then setting

$$\text{vec}(A) = \sqrt{\sigma_1}\, U(:\,, 1) \qquad \text{vec}(B) = \sqrt{\sigma_1}\, V(1, :)$$

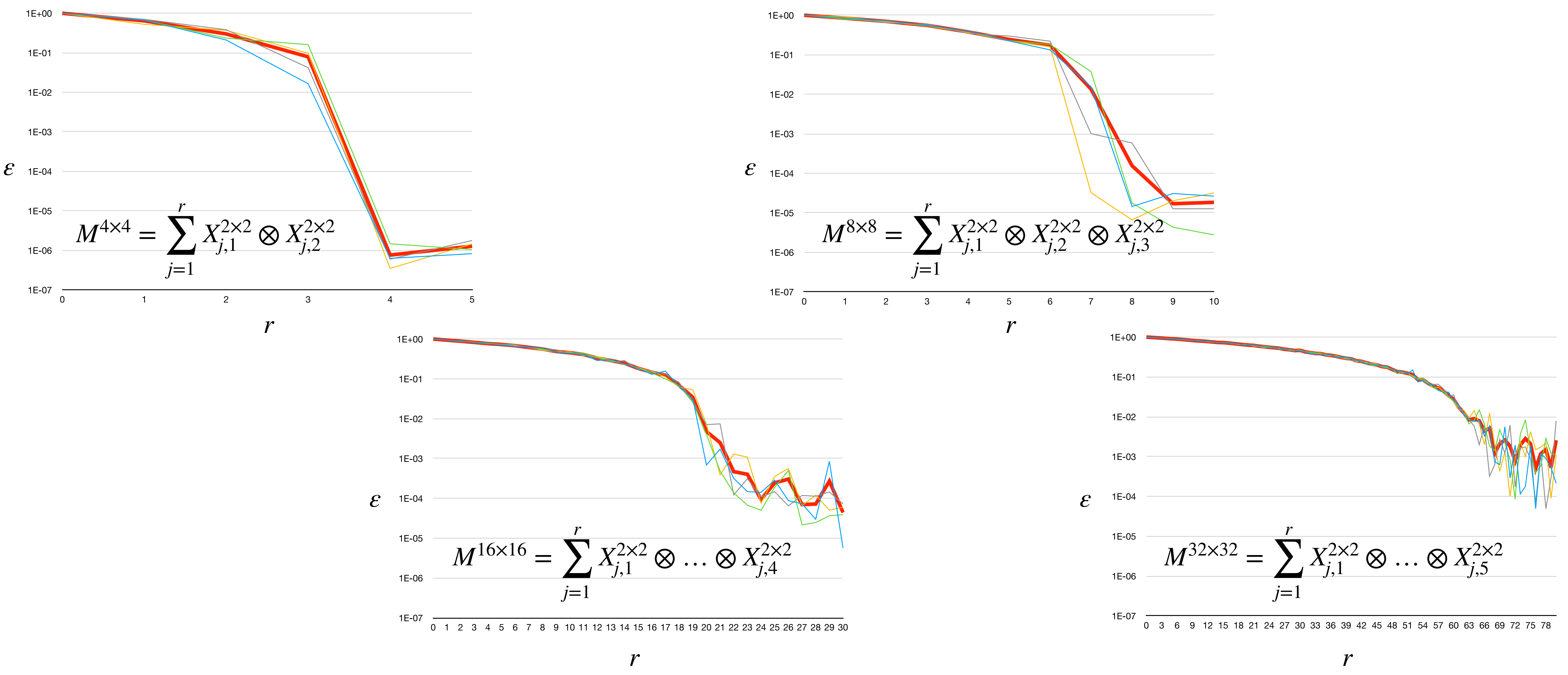
Generalizing the NKP Problem

The answer is “far from obvious” on this one

- Consider the problem of NKP approximation of $M^{2^k \times 2^k}$ in terms of a linear combination of Kronecker products
$$\phi(X) = \|M - \sum_{j=1}^{r} X_{j,1} \otimes \ldots \otimes X_{j,k}\|_F$$
- What is the minimum value of r required to exactly represent the value of M, i.e., to cause the value of $\phi(X)$ to be zero?
  - Van Loan simply states that it is “a much more difficult problem”
  - Tyrtyshnikov says the problem “requires quite intricate algorithms and still needs adequate theory” and bounds r by the approximation error ε as $r = \mathcal{O}(\log \varepsilon^{-1} \cdot \log \log \varepsilon^{-1})$
- To get a feel for the problem, perform numerical simulation to compute NKP approximations for a set of randomly generated dense matrices M

Numerical Simulation of NKP

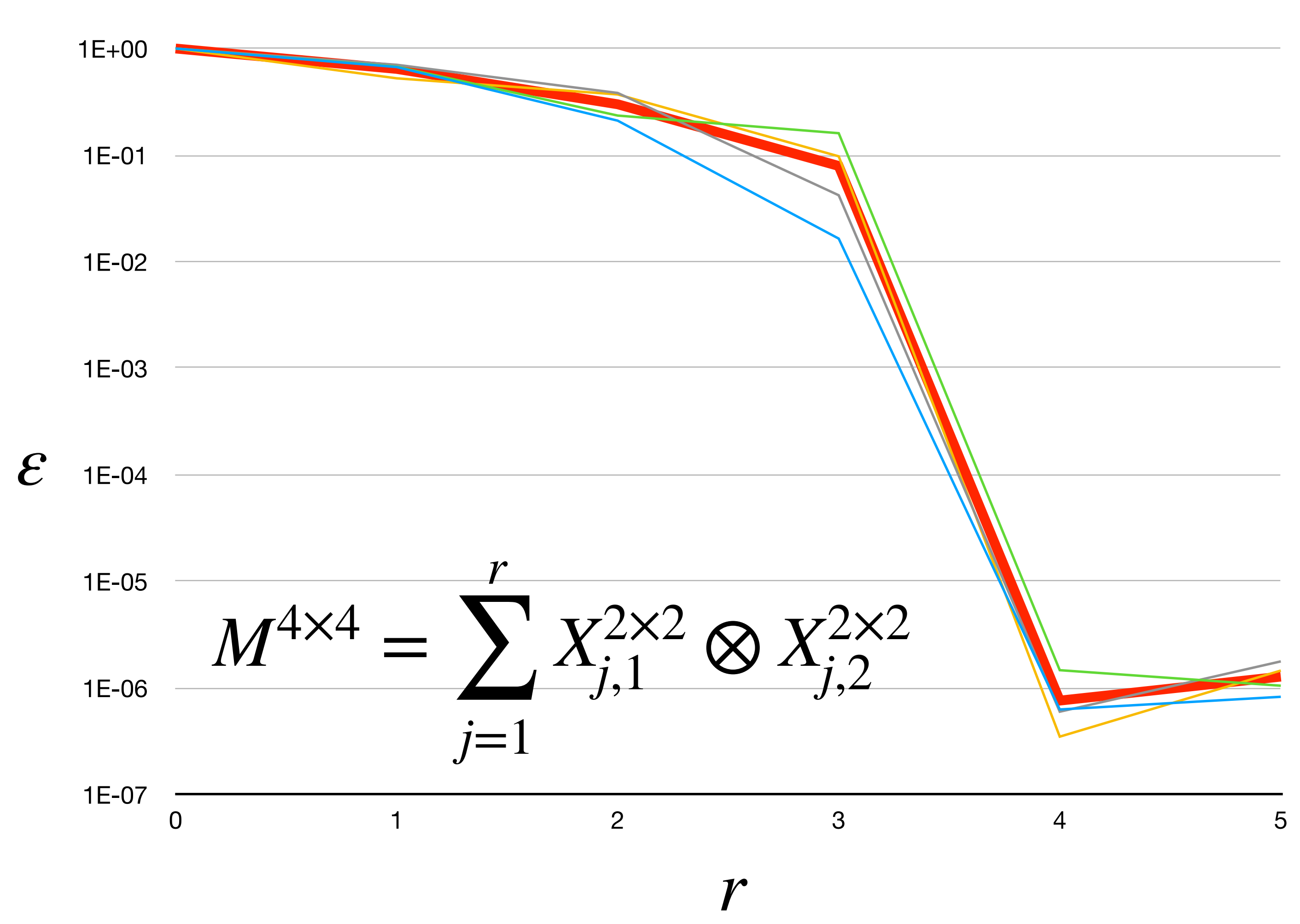
Results of using scipy *minimize* to estimate approximations



Unexpected Results

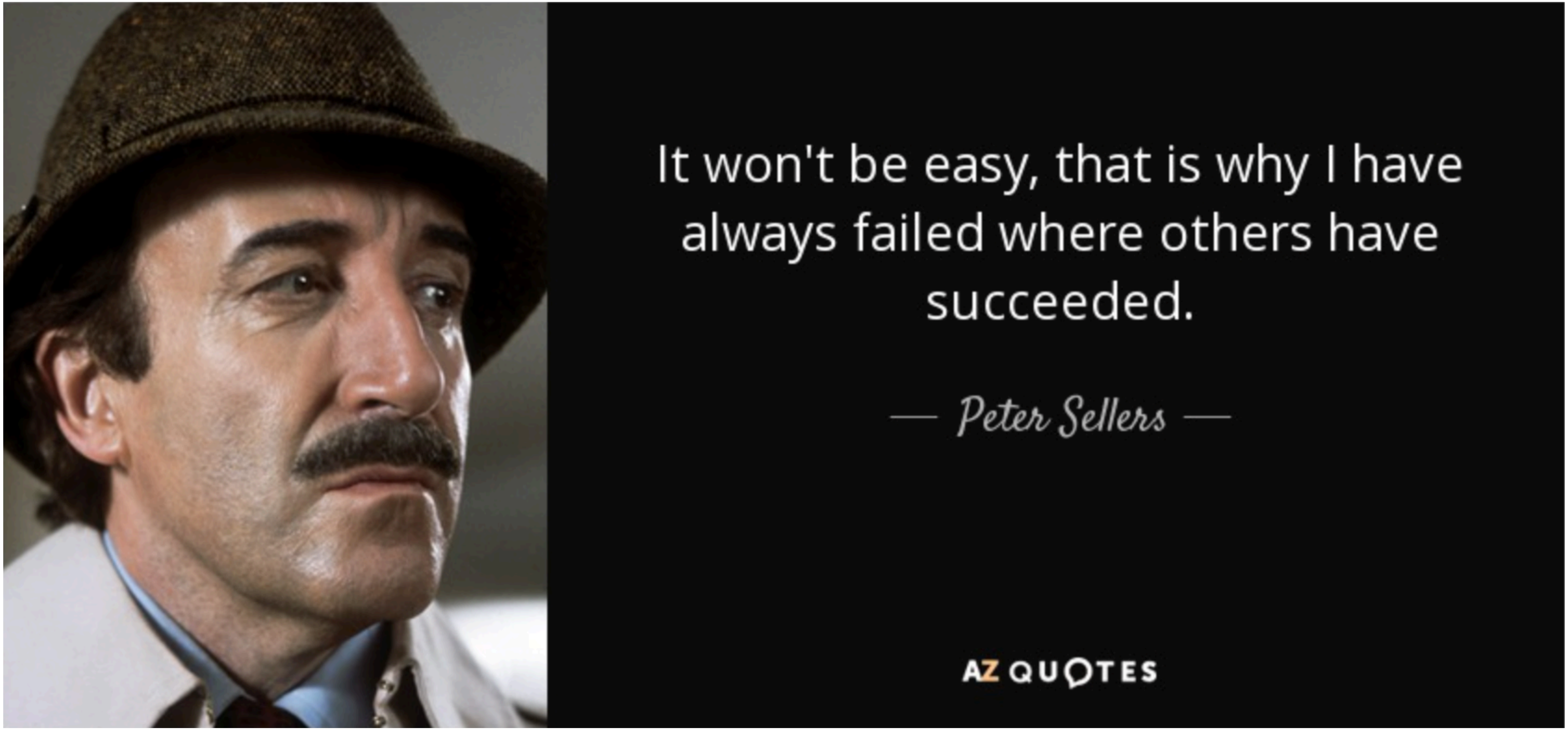
Why $r = 4$ instead of $r = 2$?

- When $r = 2$ there are 16 x-values and 16 m-values, so why is there no exact solution for M in terms of $X_1 \otimes X_2$?
- Partitioning the equations into two groups of 8 there is no way to arrange them so that the groups have more than 8 variables in common
- This appears to be an inconsistent set of equations, but how do we prove inconsistency for a non-linear system?
- The Krull dimension is non-zero.




```
m[1] == x[1] × x[5] + x[9] × x[13]
m[2] == x[1] × x[6] + x[9] × x[14]
m[3] == x[2] × x[5] + x[10] × x[13]
m[4] == x[2] × x[6] + x[10] × x[14]
m[5] == x[1] × x[7] + x[9] × x[15]
m[6] == x[1] × x[8] + x[9] × x[16]
m[7] == x[2] × x[7] + x[10] × x[15]
m[8] == x[2] × x[8] + x[10] × x[16]
m[9] == x[3] × x[5] + x[11] × x[13]
m[10] == x[3] × x[6] + x[11] × x[14]
m[11] == x[4] × x[5] + x[12] × x[13]
m[12] == x[4] × x[6] + x[12] × x[14]
m[13] == x[3] × x[7] + x[11] × x[15]
m[14] == x[3] × x[8] + x[11] × x[16]
m[15] == x[4] × x[7] + x[12] × x[15]
m[16] == x[4] × x[8] + x[12] × x[16]
```

Generalizing the NKP Problem



Sum of Kronecker Product (SKP) Representations

Or, look at this paper that I stumbled across during lunch

Sum of Kronecker products representation and its Cholesky factorization for spatial covariance matrices from large grids

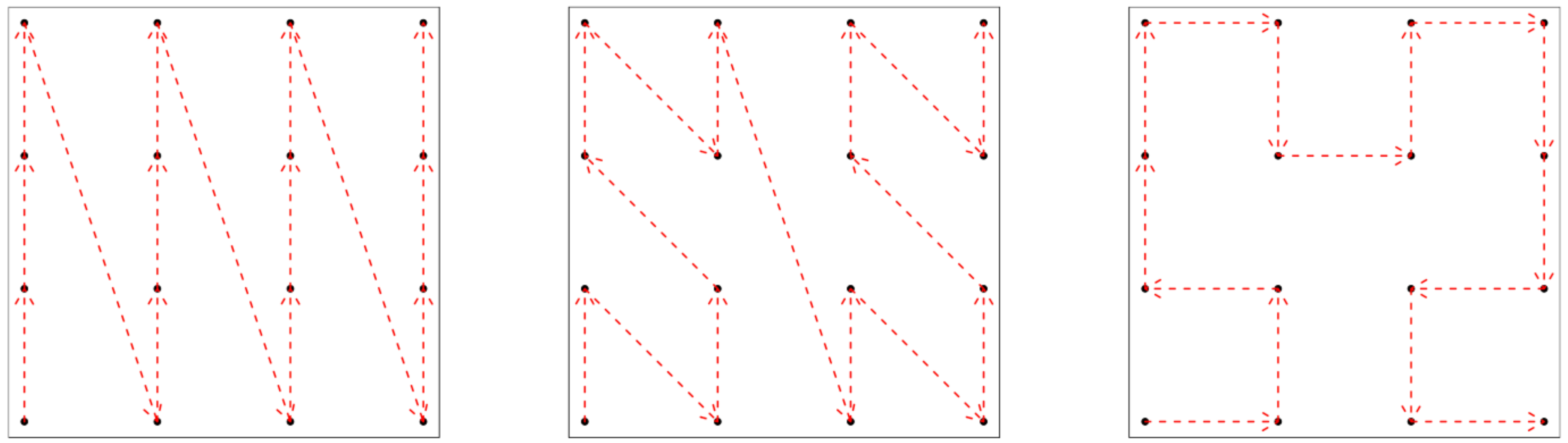
Jian Cao [a,*], Marc G. Genton [a], David E. Keyes [a], George M. Turkiyyah [b]

[a] CEMSE Division, Extreme Computing Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
[b] Department of Computer Science, American University of Beirut, Beirut, Lebanon

A B S T R A C T

The sum of Kronecker products (SKP) representation for spatial covariance matrices from gridded observations and a corresponding adaptive-cross-approximation-based framework for building the Kronecker factors are investigated. The time cost for constructing an n-dimensional covariance matrix is $O(nk^2)$ and the total memory footprint is $O(nk)$, where k is the number of Kronecker factors. The memory footprint under the SKP representation is compared with that under the hierarchical representation and found to be one order of magnitude smaller. A Cholesky factorization algorithm under the SKP representation is proposed and shown to factorize a one-million dimensional covariance matrix in under 600 seconds on a standard scientific workstation. With the computed Cholesky factor, simulations of Gaussian random fields in one million dimensions can be achieved at a low cost for a wide range of spatial covariance functions.

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}$$



(a) (b) (c)

Fig. 1. Illustrations of (a) y-major order, (b) z-curve order, and (c) Hilbert curve order.

References

Papers discussing NKP theory, implementation, and application

- Charles F. Van Loan and Nikos Pitsianis, “Approximation with Kronecker Products”, *Proceedings Linear Algebra for Large Scale and Real-Time Applications*, 1993.
- Charles F. Van Loan, “The Ubiquitous Kronecker Product”, *Journal of Computational and Applied Mathematics*, 2000.
- Wolfgang Hackbusch and Boris N. Khoromskij and Eugene E. Tyrtyshnikov, “Hierarchical Kronecker Tensor-Product Approximations”, *Journal of Numerical Mathematics 13(2)*, 2005.
- Kim Batselier and Ngai Wong, “A Constructive Arbitrary-Degree Kronecker Product Decomposition of Tensors”, *arXiv:1507.08805 [math.NA]*, 2016.
- Chencheng Cai and Rong Chen and Han Xiao, “KoPA: Automated Kronecker Product Approximation”, *arXiv:1912.02392 [math.ST]*, 2020.
- Will Pazner and Per-Olof Person, “Approximate Tensor-Product Preconditioners for Very High Order Discontinuous Galerkin Methods”, *arXiv:1704.04549 [math.NA]*, 2018.
- Xu Zhang and Felix X. Yu and Ruiqi Guo and Sanjiv Kumar and Shengjin Wang and Shih-Fu Chang, “Fast Orthogonal Projection Based on Kronecker Product”, *International Conference on Computer Vision (ICCV)*, 2015.