Four Talks on Tensor Computations

4. Power Iterations, Symmetry, and Tensor Trains

Charles F. Van Loan

Cornell University

SCAN Seminar

November 17, 2014

What is this Lecture About?

Rayleigh Quotients!

The eigenvalues and eigenvectors of a symmetric matrix C can be defined by the quotient

$$\frac{x^T C x}{x^T x}$$

and the singular values and singular vectors of a general matrix A can be defined by the quotient

$$\frac{u^T A v}{\parallel u \parallel \parallel v \parallel}$$

These are called Rayleigh Quotients and they can be generalized to tensors.

What is this Lecture About?

Symmetry

Given a basis $\{\phi_i(\mathbf{r})\}_{i=1}^n$ of atomic orbital functions, we consider the following order-4 tensor:

$$\mathcal{A}(p,q,r,s) \;=\; \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_1)\phi_r(\mathbf{r}_2)\phi_s(\mathbf{r}_2)}{\|\mathbf{r}_1-\mathbf{r}_2\|} d\mathbf{r}_1 d\mathbf{r}_2.$$

This tensor has these symmetries:

$$\mathcal{A}(p,q,r,s) \;=\; \begin{cases} \mathcal{A}(q,p,r,s) & \text{(i)} \\ \mathcal{A}(p,q,s,r) & \text{(ii)} \\ \mathcal{A}(r,s,p,q) & \text{(iii)} \end{cases}$$

What is this Lecture About?

Data-Sparse Representations

The tensor $\mathcal{A}(1{:}n, 1{:}n, 1{:}n, 1{:}n, 1{:}n)$ involves $O(n^5)$ data but if

$$
\begin{aligned}
\mathcal{G}_1&: \quad n \times r \\
\mathcal{G}_2&: \quad r \times n \times r \\
\mathcal{G}_3&: \quad r \times n \times r \\
\mathcal{G}_4&: \quad r \times n \times r \\
\mathcal{G}_5&: \quad r \times n
\end{aligned}
$$

then the approximation

$$
\mathcal{A}(i_1, i_2, i_3, i_4, i_5) \approx
\sum_{k_1=1}^{r} \sum_{k_2=1}^{r} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)
$$

involves $O(nr^2)$ data. Cool even if $r = O(n)$.

Power Iterations

Rayleigh Quotient: Matrix Case

The Big Connection

The singular values and singular vectors of a general matrix A are related to the stationary values and vectors of

$$\psi_A(u, v) \; = \; \frac{u^T A v}{\| \, u \, \| \| \, v \, \|}.$$

Rayleigh Quotient: Matrix Case

Definition

$$\psi_A(u,v) = \frac{\displaystyle\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} A(i_1,i_2)u(i_1)v(i_2)}{\| u \|_2 \| v \|_2}$$

$$= \frac{u^T A v}{\| u \|_2 \| v \|_2}$$

$$= \frac{v^T A^T u}{\| u \|_2 \| v \|_2}$$

$$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2}, u \in \mathbb{R}^{n_1}, v \in \mathbb{R}^{n_2}$$

Rayleigh Quotient: Tensor Case

Definition

$$\psi_{\mathcal{A}}(u,v,w) \;=\; \frac{\displaystyle\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} \mathcal{A}(i_1,i_2,i_3)\cdot u(i_1)\,v(i_2)\,w(i_3)}{\parallel u \parallel \parallel v \parallel \parallel w \parallel}$$

$$= \; u^T \mathcal{A}_{(1)}(w \otimes v) \,/\, (\parallel u \parallel\parallel v \parallel\parallel w \parallel)$$

$$= \; v^T \mathcal{A}_{(2)}(w \otimes u) \,/\, (\parallel u \parallel\parallel v \parallel\parallel w \parallel)$$

$$= \; w^T \mathcal{A}_{(3)}(v \otimes u) \,/\, (\parallel u \parallel\parallel v \parallel\parallel w \parallel)$$

$$\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3},\; u \in \mathbb{R}^{n_1},\; v \in \mathbb{R}^{n_2},\; w \in \mathbb{R}^{n_3}$$

Rayleigh Quotient: Matrix Case

The Gradient

If

$$\psi_A(u,v) \;=\; \frac{\displaystyle\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} A(i_1,i_2)u(i_1)v(i_2)}{\parallel u \parallel_2 \parallel v \parallel_2}$$

and u and v are unit vectors, then

$$\nabla\psi_A(u,v) \;=\; \begin{bmatrix} Av \;-\; \psi_A(u,v)u \\ A^T u \;-\; \psi_A(u,v)v \end{bmatrix}$$

Rayleigh Quotient: Tensor Case

The Gradient

If

$$\psi_{\mathcal{A}}(u,v,w) \;=\; \frac{\displaystyle\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} \mathcal{A}(i_1,i_2,i_3)\cdot u(i_1)\, v(i_2)\, w(i_3)}{\parallel u \parallel \parallel v \parallel \parallel w \parallel}$$

and u, v, and w are unit vectors, then

$$\nabla\psi_{\mathcal{A}} \;=\; \left[\begin{array}{c} \mathcal{A}_{(1)}(w\otimes v) \;-\; \psi_{\mathcal{A}}(u,v,w)u \\ \mathcal{A}_{(2)}(w\otimes u) \;-\; \psi_{\mathcal{A}}(u,v,w)v \\ \mathcal{A}_{(3)}(v\otimes u) \;-\; \psi_{\mathcal{A}}(u,v,w)w \end{array}\right]$$

Rayleigh Quotient: Matrix Case

The SVD Connection

If $\nabla \psi_A(u, v) = 0$, then

$$\begin{aligned} Av &= \psi_A(u, v) \cdot u \\ A^T u &= \psi_A(u, v) \cdot v \end{aligned}$$

Thus, the stationary values (vectors) for ψ_A define the singular values (singular vectors) for A.

Rayleigh Quotient: Tensor Case

The SVD Connection

If $\nabla \psi_A(u, v, w) = 0$, then

$$\begin{aligned} A_{(1)} \cdot (w \otimes v) &= \psi_A(u, v, w) \cdot u \\ A_{(2)} \cdot (w \otimes u) &= \psi_A(u, v, w) \cdot v \\ A_{(3)} \cdot (v \otimes u) &= \psi_A(u, v, w) \cdot w \end{aligned}$$

We **define** the stationary values (vectors) for $\psi_{\mathcal{A}}$ to be the singular values (singular vectors) for $\mathcal{A}$.

Rayleigh Quotient: Matrix Case

The Gradient Equations

$$\begin{aligned} Av &= \psi_A(u,v)\cdot u \\ A^T u &= \psi_A(u,v)\cdot v \end{aligned}$$

Power Method Step

$$\begin{aligned} \tilde{u} &= Av, & u &\leftarrow \tilde{u}/\parallel \tilde{u} \parallel \\ \tilde{v} &= A^T u. & v &\leftarrow \tilde{v}/\parallel \tilde{v} \parallel \\ \sigma &= \psi_A(u,v) \end{aligned}$$

Same as power method applied to $A^T A$.

In the limit, $\sigma u v^T$ converges to the closest rank-1 to A.

Rayleigh Quotient: Tensor Case

The Gradient Equations

$$
\begin{aligned}
A_{(1)} \cdot (w \otimes v) &= \psi_A(u, v, w) \cdot u \\
A_{(2)} \cdot (w \otimes u) &= \psi_A(u, v, w) \cdot v \\
A_{(3)} \cdot (v \otimes u) &= \psi_A(u, v, w) \cdot w
\end{aligned}
$$

Higher-Order Power Method Step

$$
\begin{aligned}
\tilde{u} &= \mathcal{A}_{(1)}(w \otimes v), \quad u_1 = \tilde{u}/\| \tilde{u} \| \\
\tilde{v} &= \mathcal{A}_{(2)}(w \otimes u), \quad u_2 = \tilde{v}/\| \tilde{v} \| \\
\tilde{w} &= \mathcal{A}_{(3)}(v \otimes u), \quad u_3 = \tilde{w}/\| \tilde{w} \| \\
\sigma &= \psi_{\mathcal{A}}(u, v, w)
\end{aligned}
$$

In the limit, $\sigma u \circ v \circ w$ tends to converge to the closest rank-1 to $\mathcal{A}$.

Power Iterations with Symmetry

Rayleigh Quotients: Symmetric Matrix Case

Definition

If $C \in \mathbb{R}^{n \times n}$ is symmetric, then

$$\begin{aligned} \phi_C(x) &= \frac{x^T C x}{x^T x} \\ &= \left(\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} C(i_1, i_2) x(i_1) x(i_2) \right) \Big/ \left(x^T x \right) \end{aligned}$$

is a Rayleigh quotient.

Gradient

If $x \in \mathbb{R}^n$ is a unit vector, then

$$\nabla \phi_C = 2 \left(C x - \phi_C(x) \cdot x \right).$$

Rayleigh Quotients: Symmetric Matrix Case

The Eigenvalue Connection

If $\nabla\phi_C(x) \;=\; 0$ then

$$C\,x \;=\; \phi_C(x)\cdot x$$

Thus, the stationary values (vectors) of ϕ_C define the eigenvalues (eigenvectors) of C.

This has many implications for both analysis and algorithms...

Rayleigh Quotients: Symmetric Matrix Case

Power Method Step

$$\begin{aligned} \tilde{x} &= Cx \\ x &= \tilde{x}/\| \tilde{x} \| \\ \lambda &= x^T C x \end{aligned}$$

To formulate a tensor analog we need to be clear on the notion of a symmnetric tensor

Tensor Transposition: The Order-3 Case

Six possibilities...

If $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then there are $6 = 3!$ possible transpositions identified by the notation $\mathcal{C}^{<[i\ j\ k]>}$ where $[i\ j\ k]$ is a permutation of $[1\ 2\ 3]$:

$$\mathcal{B} = \left\{ \begin{array}{c} \mathcal{C}^{<[1\ 2\ 3]>} \\ \mathcal{C}^{<[1\ 3\ 2]>} \\ \mathcal{C}^{<[2\ 1\ 3]>} \\ \mathcal{C}^{<[2\ 3\ 1]>} \\ \mathcal{C}^{<[3\ 1\ 2]>} \\ \mathcal{C}^{<[3\ 2\ 1]>} \end{array} \right\} \quad \Longrightarrow \quad \left\{ \begin{array}{c} b_{ijk} \\ b_{ikj} \\ b_{jik} \\ b_{jki} \\ b_{kij} \\ b_{kji} \end{array} \right\} = c_{ijk}$$

for $i = 1{:}n_1,\ j = 1{:}n_2,\ k = 1{:}n_3$.

Symmetric Tensors

The Order-3 Definition: $\mathcal{C} \in \mathbb{R}^{n \times n \times n}$

$$\mathcal{C}(i,j,k) = \begin{cases} \mathcal{C}(i,k,j) \\ \mathcal{C}(j,i,k) \\ \mathcal{C}(i,k,i) \\ \mathcal{C}(k,i,j) \\ \mathcal{C}(k,j,i) \end{cases}$$

Symmetric Tensors

Rank-1 Symmetric Tensors

If $x \in \mathbb{R}^n$, then

$$\mathcal{C} \;=\; x \circ x \circ x$$

is a symmetric rank-1 tensor. This is obvious since

$$\mathcal{C}(i_1, i_2, i_3) \;=\; x(i_1)x(i_2)x(i_3).$$

Note that

$$\text{vec}(x \circ x \circ \circ x) \;=\; x \otimes x \otimes x$$

Symmetric Tensors

Symmetric Rank

An order-3 symmetric tensor $\mathcal{C}$ has symmetric rank r if there exists $x_1, \ldots, x_r \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}^r$ such that

$$\mathcal{C} \;=\; \sum_{k=1}^{r} \sigma_k \cdot x_k \circ x_k \circ x_k$$

and no shorter sum of symmetric rank-1 tensors exists. Symmetric rank is denoted by $\text{rank}_S(\mathcal{C})$.

Note, there may be a shorter sum so

$$\mathcal{C} \;=\; \sum_{k=1}^{\tilde{r}} \tilde{\sigma}_k \cdot \tilde{x}_k \circ \tilde{y}_k \circ \tilde{z}_k$$

Rayleigh Quotients: Symmetric Tensor Case

Definition

$$\phi_{\mathcal{C}}(x) \;=\; \frac{\displaystyle\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} \mathcal{C}(i_1,i_2,i_3)\cdot x(i_1)\,x(i_2)\,x(i_3)}{\parallel x \parallel^3}$$

where $\mathcal{C} \in \mathbb{R}^{n\times n\times n}$ is symmetric and $x \in \mathbb{R}^N$.

Alternative Formulation

$$\phi_{\mathcal{C}}(x) \;=\; \frac{x^T \mathcal{C}_{(k)}(x \otimes x)}{\parallel x \parallel^3} \qquad k \text{ doesn't matter}$$

Easy to get the gradient of $\phi_{\mathcal{C}}$...

Rayleigh Quotients: Symmetric Tensor Case

Gradient of $\phi_{\mathcal{C}}$

If x is a unit vector, then

$$\nabla \phi_{\mathcal{C}}(x) \;=\; \mathcal{C}_{(1)}(x \otimes x) \;-\; \phi_{\mathcal{C}}(x) \cdot x$$

Idea for improving x:

$$\tilde{x} = \mathcal{C}_{(1)}(x \otimes x), \quad x = \tilde{x} / \| \tilde{x} \|$$

Rayleigh Quotients: Symmetric Tensor Case

Symmetric Higher-Order Power Method

Initialize unit vector x.

Repeat

$$\tilde{x} \;=\; \mathcal{C}_{(1)}(x \otimes x)$$

$$x = \tilde{x}/\parallel \tilde{x} \parallel$$

Sample Convergence Result

If the order of $\mathcal{C}$ is even and M is a square unfolding, then the iteration converges if M is positive definite.

The SVD - SymEig Connection

The "Sym" of a Matrix

$$\text{sym}(A) \;=\; \left[\begin{array}{cc} 0 & A \\ A^T & 0 \end{array}\right] \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$$

The SVD of A Relates to the EVD of sym(A)

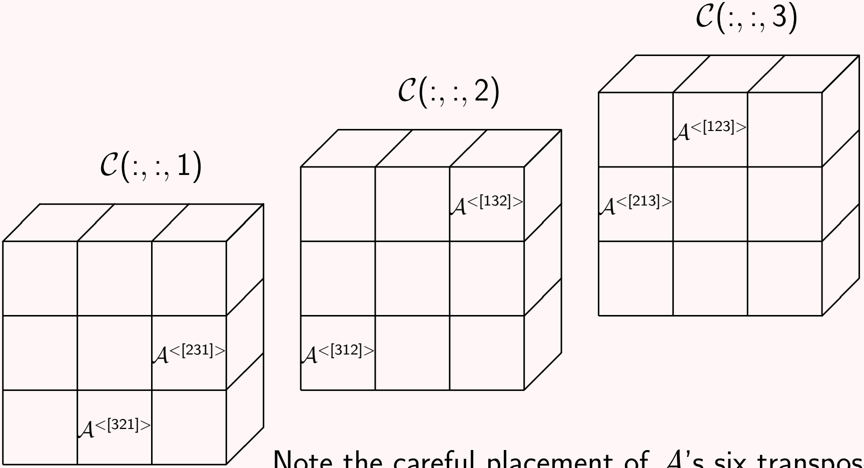
If $A = U \cdot \text{diag}(\sigma_i) \cdot V^T$ is the SVD of $A \in \mathbb{R}^{n_1 \times n_2}$, then for $k = 1{:}\text{rank}(A)$

$$\left[\begin{array}{cc} 0 & A \\ A^T & 0 \end{array}\right] \left[\begin{array}{c} u_k \\ \pm v_k \end{array}\right] \;=\; \pm\sigma_k \left[\begin{array}{c} u_k \\ \pm v_k \end{array}\right]$$

where $u_k = U(:, k)$ and $v_k = V(:, k)$.

The above SVD-related power method was essentially traditional power method applied to finding the largest eigenvector of sym(A).

An Order-3 Example...



Note the careful placement of $\mathcal{A}$'s six transposes

Eigenvalue Rayleigh Quotients For Symmetric Tensors

Equivalence

If $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, then applying the Higher-Order Power Method to $\mathcal{A}$, is "equivalent" to applying the Symmetric Higher order Power Method to $\text{sym}(\mathcal{A})$.

Exiting the former with σ, u_1, u_2 and u_3 is essentially the same as exiting the latter with σ and

$$x = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

Eigenvalue Rayleigh Quotients For Symmetric Tensors

Theorem

If

$$\left\{ \sigma, \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \right\}$$

is a stationary pair for sym($\mathcal{C}$) then so are

$$\left\{ \sigma, \begin{bmatrix} u_1 \\ -u_2 \\ -u_3 \end{bmatrix} \right\}, \quad \left\{ -\sigma, \begin{bmatrix} u_1 \\ -u_2 \\ u_3 \end{bmatrix} \right\}, \quad \left\{ -\sigma, \begin{bmatrix} u_1 \\ u_2 \\ -u_3 \end{bmatrix} \right\}$$

Recall that for matrices...

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_k \\ \pm v_k \end{bmatrix} = \pm\sigma_k \begin{bmatrix} u_k \\ \pm v_k \end{bmatrix}$$

Symmetric Tensors: Interesting Aside about Rank

Fact

If $\mathcal{C} \in \mathbb{C}^{n\times\cdots\times n}$ is an order-d symmetric tensor, then with probability 1

$$\text{rank}_S(\mathcal{C}) = \begin{cases} f(d,m)+1 & \text{if } (d,n) = (3,5),(4,3),(4,4), \text{ or } (4,5) \\ f(d,n) & \text{otherwise} \end{cases}$$

where

$$f(d,n) = \text{ceil}\left(\frac{\binom{n+d-1}{d}}{n} \right)$$

Symmetric Tensor Rank is "more tractable" than General Tensor Rank.

Symmetric Tensors: Interesting Aside about Rank

Interesting Possible Connection

Easy:

$$d!\ \mathrm{rank}(\mathcal{A}) \ \leq \ \mathrm{rank}_S(\mathrm{sym}(\mathcal{A}))$$

Equality is hard or perhaps not true. But if it could be established, then we would have new insight into the tensor rank problem.

More on Tensor Symmetry

The Two-Electron Integral Tensor (TEI)

Given a basis $\{\phi_i(\mathbf{r})\}_{i=1}^{n}$ of atomic orbital functions, we consider the following order-4 tensor:

$$\mathcal{A}(p,q,r,s) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_1)\phi_r(\mathbf{r}_2)\phi_s(\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|} d\mathbf{r}_1 d\mathbf{r}_2.$$

The TEI tensor has these symmetries:

$$\mathcal{A}(p,q,r,s) \ = \begin{cases} \mathcal{A}(q,p,r,s) & \text{(i)} \\ \mathcal{A}(p,q,s,r) & \text{(ii)} \\ \mathcal{A}(r,s,p,q) & \text{(iii)} \end{cases}$$

We say that $\mathcal{A}$ is "((12)(34))-symmetric".

The ((12)(34))-Symmetric Multilinear Product where $X \in \mathbb{R}^{n \times n}$

$$\mathcal{B}(j_1, j_2, j_3, j_4)$$

$$=$$

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} \mathcal{A}(i_1, i_2, i_3, i_4) \cdot X(i_1, j_1) \cdot X(i_2, j_2) \cdot X(i_3, j_3) \cdot X(i_4, j_4)$$

and the tensor $\mathcal{B}$ has these symmetries...

$$\mathcal{B}(p, q, r, s) = \begin{cases} \mathcal{B}(q, p, r, s) \\ \mathcal{B}(p, q, s, r) \\ \mathcal{B}(r, s, p, q) \end{cases}$$

In other words $\mathcal{B}$ is also ((12)(34))-symmetric.

The Block Matrix Product Formulation

If $A = (A_{rs})$ and $B = (B_{rs})$ are n-by-n block matrices with n-by-n blocks and

$$\mathcal{A}(p,q,r,s) \quad \leftrightarrow \quad [\, A_{rs} \,]_{pq}$$

$$\mathcal{B}(p,q,r,s) \quad \leftrightarrow \quad [\, B_{rs} \,]_{pq}$$

then the ((12)(34))-symmetric multilinear product is equivalent to

$$B \;=\; (X \otimes X)^T A (X \otimes X)$$

The n^2-by-n^2 matrix A is special. It is positive definite with very low rank. And it has multiple symmetries...

A "Inherits" $\mathcal{A}$'s Structure

Since

$$\mathcal{A}(p,q,r,s) \;\leftrightarrow\; [\,A_{rs}\,]_{pq}$$

and

$$\mathcal{A}(p,q,r,s) \;=\; \begin{cases} \mathcal{A}(q,p,r,s) & \text{(i)} \\ \mathcal{A}(p,q,s,r) & \text{(ii)} \\ \mathcal{A}(r,s,p,q) & \text{(iii)} \end{cases}$$

it follows that

(i) The blocks of A are symmetric ($A_{rs}^T = A_{rs}$) because of (i).

(ii) A is symmetric as a block matrix ($A_{rs} = A_{sr}$) because (ii).

Time to Talk About Unfoldings

The tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times n \times n}$ can be unfolded several ways.

- We have depicted the $[1,3] \times [2,4]$ unfolding $A = \mathcal{A}_{[1,3]\times[2,4]}$ defined by

$$\mathcal{A}(i_1, i_2, i_3, i_4) \;\rightarrow\; A(i_1 + (i_3 - 1)n, i_2 + (i_4 - 1)n)$$

- Also of interest is the $[1,2] \times [3,4]$ unfolding $A = \mathcal{A}_{[1,2]\times[3,4]}$ defined by

$$\mathcal{A}(i_1, i_2, i_3, i_4) \;\rightarrow\; A(i_1 + (i_2 - 1)n, i_3 + (i_4 - 1)n)$$

If $\mathcal{A}$ is ((12)(34))-symmetric, these two unfoldings display different multiple symmetries...

The $[1,3]\times[2,4]$ Unfolding a ((12)(34))-Symmetric $\mathcal{A}$

If $A = \mathcal{A}_{[1,3]\times[2,4]}$, then (as we have seen) A is block symmetric with symmetric blocks.

$$A = \left[\begin{array}{ccc|ccc|ccc} 11 & 12 & 13 & 12 & 17 & 18 & 13 & 18 & 22 \\ 12 & 14 & 15 & 17 & 19 & 20 & 18 & 23 & 24 \\ 13 & 15 & 16 & 18 & 20 & 21 & 22 & 24 & 25 \\ \hline 12 & 17 & 18 & 14 & 19 & 23 & 15 & 20 & 24 \\ 17 & 19 & 20 & 19 & 26 & 27 & 20 & 27 & 29 \\ 18 & 20 & 21 & 23 & 27 & 28 & 24 & 29 & 30 \\ \hline 13 & 18 & 22 & 15 & 20 & 24 & 16 & 21 & 25 \\ 18 & 23 & 24 & 20 & 27 & 29 & 21 & 28 & 30 \\ 22 & 24 & 25 & 24 & 29 & 30 & 25 & 30 & 31 \end{array}\right].$$

The $[1,2] \times [3,4]$ Unfolding of a ((12)(34)) Symmetric $\mathcal{A}$

If $A = \mathcal{A}_{[1,2]\times[3,4]}$, then A is symmetric and (among other things) is "perfect shuffle" symmetric.

$$A = \left[\begin{array}{ccc|ccc|ccc} 11 & 12 & 13 & 12 & 14 & 15 & 13 & 15 & 16 \\ 12 & 17 & 18 & 17 & 19 & 20 & 18 & 20 & 21 \\ 13 & 18 & 22 & 18 & 23 & 24 & 22 & 24 & 25 \\ \hline 12 & 17 & 18 & 17 & 19 & 20 & 18 & 20 & 21 \\ 14 & 19 & 23 & 19 & 26 & 27 & 23 & 27 & 28 \\ 15 & 20 & 24 & 20 & 27 & 29 & 24 & 29 & 30 \\ \hline 13 & 18 & 22 & 18 & 23 & 24 & 22 & 24 & 25 \\ 15 & 20 & 24 & 20 & 27 & 29 & 24 & 29 & 30 \\ 16 & 21 & 25 & 21 & 28 & 30 & 25 & 30 & 31 \end{array}\right]$$

Each column reshapes into a 3x3 symmetric matrix, e.g., $A(:,)$ reshapes to

$$\left[\begin{array}{ccc} 11 & 12 & 13 \\ 12 & 14 & 15 \\ 13 & 15 & 16 \end{array}\right]$$

What is perfect shuffle symmetry?

Perfect Shu□e Symmetry

An n^2-by-n^2 matrix A has perfect shuffle symmetry if

$$A = \Pi_{n,n} A \Pi_{n,n}$$

where

$$\Pi_{n,n} \; = \; I_{n^2}(:, v), \qquad v \; = \; [\, 1{:}n{:}n^2 \mid 2{:}n{:}n^2 \mid \; \cdots \; \mid n{:}n{:}n^2 \,].$$

e.g.,

$$\Pi_{3,3} \; = \; \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

The Perfect Shu□e Π_{nn}

Because $\Pi_{n,n}$ is symmetric it has just two eigenvalues: $+1$ and -1.

$$\Pi_{3,3} = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \mathbf{x_{12}} \\ x_{22} \\ x_{32} \\ x_{13} \\ x_{23} \\ x_{33} \end{bmatrix} = \pm \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ \mathbf{x_{21}} \\ x_{22} \\ x_{23} \\ x_{31} \\ x_{32} \\ x_{33} \end{bmatrix}$$

If $\Pi_{n,n}x = x$, then $\texttt{reshape}(x, n, n)$ is symmetric.

If $\Pi_{n,n}x = -x$, then $\texttt{reshape}(x, n, n)$ is skew-symmetric.

Solution Framework:

1. Closed-form block diagonalization with a special highly-structured orthogonal $Q = [Q_1\ Q_2]$:

$$Q^T A Q \;=\; \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

2. Compute half-sized rank-revealing LDL factorizations:

$$P_i A_i P_i^T \;\approx\; L_i D_i L_i^T, \quad i = 1, 2$$

3. Set

$$\tilde{A} \;=\; Y_1 D_1 Y_1^T \;+\; Y_2 D_2 Y_2^T$$

where $Y_i = Q_i P_i^T L_i$.

4. $B \;\approx\; \tilde{B} \;=\; (X \otimes X)^T \tilde{A} (X \otimes X)$

Tensor Trains

Tensor Train: An Example

Given the "carriages"...

$$
\begin{aligned}
\mathcal{G}_1 &: \quad n_1 \times r_1 \\
\mathcal{G}_2 &: \quad r_1 \times n_2 \times r_2 \\
\mathcal{G}_3 &: \quad r_2 \times n_3 \times r_3 \\
\mathcal{G}_4 &: \quad r_3 \times n_4 \times r_4 \\
\mathcal{G}_5 &: \quad r_4 \times n_5
\end{aligned}
$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$
\mathcal{A}(i_1, i_2, i_3, i_4, i_5)
$$

$$
=
$$

$$
\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)
$$

Think of a graph where the nodes are low-order tensors and the edges are contractions.

Tensor Train: An Example

Given the "carriages"...

$$
\begin{aligned}
\mathcal{G}_1 &: \quad n_1 \times r_1 \\
\mathcal{G}_2 &: \quad r_1 \times n_2 \times r_2 \\
\mathcal{G}_3 &: \quad r_2 \times n_3 \times r_3 \\
\mathcal{G}_4 &: \quad r_3 \times n_4 \times r_4 \\
\mathcal{G}_5 &: \quad r_4 \times n_5
\end{aligned}
$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$

$$=$$

$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\mathcal{G}_1: \quad n_1 \times r_1$$

$$\mathcal{G}_2: \quad r_1 \times n_2 \times r_2$$

$$\mathcal{G}_3: \quad r_2 \times n_3 \times r_3$$

$$\mathcal{G}_4: \quad r_3 \times n_4 \times r_4$$

$$\mathcal{G}_5: \quad r_4 \times n_5$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$

$$=$$

$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\mathcal{G}_1: \quad n_1 \times r_1$$
$$\mathcal{G}_2: \quad r_1 \times n_2 \times r_2$$
$$\mathcal{G}_3: \quad r_2 \times n_3 \times r_3$$
$$\mathcal{G}_4: \quad r_3 \times n_4 \times r_4$$
$$\mathcal{G}_5: \quad r_4 \times n_5$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1}\sum_{k_2=1}^{r_2}\sum_{k_3=1}^{r_3}\sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1)\cdot\mathcal{G}_2(k_1, i_2, k_2)\cdot\mathcal{G}_3(k_2, i_3, k_3)\cdot\mathcal{G}_4(k_3, i_4, k_4)\cdot\mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Tensor Train: An Example

Given the "carriages"...

$$\begin{array}{ll} \mathcal{G}_1: & n_1 \times r_1 \\ \mathcal{G}_2: & r_1 \times n_2 \times r_2 \\ \mathcal{G}_3: & r_2 \times n_3 \times r_3 \\ \mathcal{G}_4: & r_3 \times n_4 \times r_4 \\ \mathcal{G}_5: & r_4 \times n_5 \end{array}$$

We define the train" $\mathcal{A}(1{:}n_1, 1{:}n_2, 1{:}n_3, 1{:}n_4, 1{:}n_5)$...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1}\sum_{k_2=1}^{r_2}\sum_{k_3=1}^{r_3}\sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1)\cdot\mathcal{G}_2(k_1, i_2, k_2)\cdot\mathcal{G}_3(k_2, i_3, k_3)\cdot\mathcal{G}_4(k_3, i_4, k_4)\cdot\mathcal{G}_5(k_4, i_5)$$

$O(nr^2)$ vs $O(n^5)$

Computing a Tensor Train Representation

Main Idea

A sequence of unfoldings is produced.

The unfoldings get narrower and narrower.

A rank-revealing SVD $U(\Sigma V^T) = UZ$ is computed each time.

The "carriages" are reshaped U-matrices.

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \ = \ U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \ = \ U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \ = \ U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \;=\; U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \;=\; U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

3(a) Rank-revealing SVD: $\texttt{reshape}(Z_2, [r_2 n_3, n_4 n_5]) \;=\; U_3 Z_3.$

$\mathcal{G}_3 = \texttt{reshape}(U_3, [r_2, n_3, r_3]).$

Computing a Tensor Train Representation

1(a) Rank-revealing SVD: $\texttt{reshape}(A, [n_1\ ,\ n_2 n_3 n_4 n_5]) \;=\; U_1 Z_1.$

$\mathcal{G}_1 = \texttt{reshape}(U_1, [n_1, r_1]).$

2(a) Rank-revealing SVD: $\texttt{reshape}(Z_1, [r_1 n_2, n_3 n_4 n_5]) \;=\; U_2 Z_2.$

$\mathcal{G}_2 = \texttt{reshape}(U_2, [r_1, n_2, r_2]).$

3(a) Rank-revealing SVD: $\texttt{reshape}(Z_2, [r_2 n_3, n_4 n_5]) \;=\; U_3 Z_3.$

$\mathcal{G}_3 = \texttt{reshape}(U_3, [r_2, n_3, r_3]).$

4(a) Rank-revealing SVD: $\texttt{reshape}(Z_3, [r_3 n_4, n_5]) \;=\; U_4 Z_4.$

$\mathcal{G}_4 = \texttt{reshape}(U_4, [r_3, n_4, r_4]).$

$\mathcal{G}_5 = \texttt{reshape}(Z_4, [r_4, n_5]).$

Computing a Tensor Train Representation

Why Do We Emerge With...

$$\mathcal{A}(i_1, i_2, i_3, i_4, i_5)$$
$$=$$
$$\sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} \sum_{k_4=1}^{r_4} \mathcal{G}_1(i_1, k_1) \cdot \mathcal{G}_2(k_1, i_2, k_2) \cdot \mathcal{G}_3(k_2, i_3, k_3) \cdot \mathcal{G}_4(k_3, i_4, k_4) \cdot \mathcal{G}_5(k_4, i_5)$$

The reason is because...

1. The rows and columns of an unfolding can be "multi-indexed"

2. The product of two matrices can sometimes be viewed as a contraction between two tensors.

Multi-Indexing Rows and Columns

```
n = [2 3 2 2 3];
A = randn(n);
B = reshape(A,[n(1)*n(2)*n(3) , n(4)*n(5) ]
```

$$
\begin{array}{cc}
 & \begin{array}{cccccc} (1,1) & (2,1) & (1,2) & (2,2) & (1,3) & (2,3) \end{array} \\
B = & \left[\begin{array}{cccccc}
a_{11111} & a_{11121} & a_{11112} & a_{11122} & a_{11113} & a_{11123} \\
a_{21111} & a_{21121} & a_{21112} & a_{21122} & a_{21113} & a_{21123} \\
a_{12111} & a_{12121} & a_{12112} & a_{12122} & a_{12113} & a_{12123} \\
a_{22111} & a_{22121} & a_{22112} & a_{22122} & a_{22113} & a_{22123} \\
a_{13111} & a_{13121} & a_{13112} & a_{13122} & a_{13113} & a_{13123} \\
a_{23111} & a_{23121} & a_{23112} & a_{23122} & a_{23113} & a_{23123} \\
a_{11211} & a_{11221} & a_{11212} & a_{11222} & a_{11213} & a_{11223} \\
a_{21211} & a_{21221} & a_{21212} & a_{21222} & a_{21213} & a_{21223} \\
a_{12211} & a_{12221} & a_{12212} & a_{12222} & a_{12213} & a_{12223} \\
a_{22211} & a_{22221} & a_{22212} & a_{22222} & a_{22213} & a_{22223} \\
a_{13211} & a_{13221} & a_{13212} & a_{13222} & a_{13213} & a_{13223} \\
a_{23211} & a_{23221} & a_{23212} & a_{23222} & a_{23213} & a_{23223}
\end{array}\right]
\begin{array}{c}
(1,1,1) \\ (2,1,1) \\ (1,2,1) \\ (2,2,1) \\ (1,3,1) \\ (2,3,1) \\ (1,1,2) \\ (2,1,2) \\ (1,2,2) \\ (2,2,2) \\ (1,3,2) \\ (2,3,2)
\end{array}
\end{array}
$$

Matrix Multiplication with Multi-indexing

Suppose B is $(n_1 n_2 n_3) \times r$ and C is $r \times (n_4 n_5)$

$A = BC$ as a Matrix Product

$$A(i,j) \;=\; \sum_{k=1}^{r} B(i,k)C(k,j)$$

$1 \le i \le n_1 n_2 n_3, \;\; 1 \le j \le n_4 n_5$

$A = BC$ as a tensor contraction

$$A(i_1,i_2,i_3,j_1,j_2) \;=\; \sum_{k=1}^{r} B(i_1,i_2,i_3,k)C(k,j_1,j_2)$$

$1 \le i_1 \le n_1, \;\; 1 \le i_2 \le n_2, \;\; 1 \le i_3 \le n_3, \;\; 1 \le j_1 \le n_4, \;\; 1 \le j_2 \le n_5$

Summary of Lecture

Key Words

- Symmetric tensors have the property that the value of $\mathcal{A}(\mathbf{i})$ does not depend upon the order the indices. They arise in many settings and special properties.
- Setting to zero the gradient of a multilinear Rayleigh quotient leads to the idea of eigenvalues and singular value of tensors.
- Higher-order power methods based on the Rayleigh quotient can be used to find nearest rank-1 approximations to either general or symmetric tensors
- When tensors with multiple symmetries are unfolded, the unfoldings have multiple symmetries that can be exploited.
- Data sparse approximations are critical to the development of tensor methods that scale with d. Tensor Trains are one approach in this direction.

References

P. Comon, G.H. Golub, L-H. Lim, and B. Mourain 2008). "Symmetric Tensors and Symmetric Tensor Rank," *SIAM J. Matrix Anal. Appl.*, 30, 1254–1279.

E. Kofidis and P. Regalia (2001). "Tensor Approximation and Signal Processing Applications," in *Structured Matrices in Mathematics, Computer Science, and Engineering I,* V. Olshevsky (ed), American Mathematical Society, Providence, RI, 103–133.

E. Kofidis and P.A. Regalia (2002). "On the Best Rank-1 Approximation of Higher-Order Super-Symmetric Tensors," *SIAM J. Matrix Analysis and Applications*, 23, 863-884.

L.-H. Lim (2005). "Singular Values and Eigenvalues of Tensors: a Variational Approach," *Proc. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Process*, 129-132.

J. Morton and L.-H. Lim (2009). "Principal Cumulant Component Analysis," Technical report, Stanford University.

References

L. Qi (2005). "Eigenvalues of a Real Supersymmetric Tensor," *Journal of Symbolic Computation*, 40, 1302–1324.

L. Qi (2007). "Eigenvalues and Invariants of Tensors," *J. Mathematical Analysis and Applications*, 325, 1363–1377.

T. Zhang and G. H. Golub (2001). "Rank-one Approximation to High order Tensors," *SIAM Journal on Matrix Analysis and Applications,* 23, 534–550.

References

W. Hackbusch and B.N. Khoromskij (2007). "Tensor-product Approximation to Operators and Functions in High Dimensions," *Journal of Complexity,* 23, 697–714.

I.V. Oseledets and E.E. Tyrtyshnikov (2008). "Breaking the Curse of Dimensionality, Or How to Use SVD in Many Dimensions," *SIAM J. Sci. Comput. 31*, 3744–3759.

W. Hackbusch and S. Kuhn (2009). "A New Scheme for the Tensor Representation," *J. Fourier Anal. Appl. 15*, 706–722.

I.V. Oseledets, D.V. Savostyanov, and E.E. Tyrtyshnikov (2009). "Linear Algebra for Tensor Problems," *Computing 85*, 169-188.

I. Oseledets and E. Tyrtyshnikov (2010). "TT-Cross Approximation for Multidimensional Arrays," *Lin. Alg. Applic. 432*, 70–88.

G. Beylkin and M.J. Mohlenkamp (2002). "Numerical Operator Calculus in Higher Dimensions," *Proceedings of the National Academy of Sciences,* 99(16), 10246–10251.

G. Beylkin and M.J. Mohlenkamp (2005). "Algorithms for Numerical Analysis in High Dimensions," *SIAM J. Sci. Comp. 26*, 2133–2159.

References

S. Ragnarsson and C. F. Van Loan, *Block tensors and symmetric embeddings*, Linear Algebra Appl., 438, 853–874, 2013.

S. Ragnarsson and C. F. Van Loan, *Block tensor unfoldings*, SIAM J. Matrix Anal. Appl., 33(1), 149–169, 2012.