Numerical Linear Algebra Issues Underlying a Tensor Network Computation

Charles F. Van Loan
Department of Computer Science
Cornell University

Joint work with Stefan Ragnarsson, Center for Applied Mathematics, Cornell University

NSF Workshop on Future Directions in Tensor-Based Computation and Modeling, February 20-21, 2009.

The Rayleigh Quotient

One way to find the smallest eigenvalue and eigenvector of a symmetric matrix H is to minimize

$$r(a) = \frac{a^T H a}{a^T a}$$

$$a = a_{min},\ \lambda = r(a_{min}) \qquad \Rightarrow \qquad Ha = \lambda a$$

Modeling Electron Interactions

Have d "sites" (grid points) in physical space.

The goal is compute a wave function, an element of a 2^d Hilbert space.

The Hilbert space is a product of d 2-dimensional Hilbert spaces. (A site is either occupied or not occupied.)

A (discretized) wavefunction is a d-tensor, 2-by-2-by-2-by-2...

The H-Matrix

$$H = \sum_{i,j=1}^{d} t_{ij}\, H_i^T H_j \;+\; \sum_{i,j,k,\ell=1}^{d} v_{ijk\ell} H_i^T H_j^T H_k H_\ell$$

where

$$H_p = I_{2^{p-1}} \otimes A \otimes I_{2^{d-p}}$$

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$T = (t_{ij}) = T(1{:}d\,,1{:}d)$$

$$V = (v_{ijk\ell}) = V(1{:}d\,,1{:}d\,,1{:}d,\;1{:}d)$$

$O(d^4)$ parameters describe 4^d entries

$t_{ij} = t_{ji}$

$v_{ijk\ell} = v_{k\ell ij}$
$v_{ijk\ell} = v_{jik\ell} = v_{ij\ell k}$

$$\otimes \cdots \otimes \cdots \otimes \cdots \otimes \cdots \otimes$$

If $d = 40$ and

$$H_p \;=\; I_{2^{p-1}} \otimes A \otimes I_{2^{d-p}} \qquad A \;=\; \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

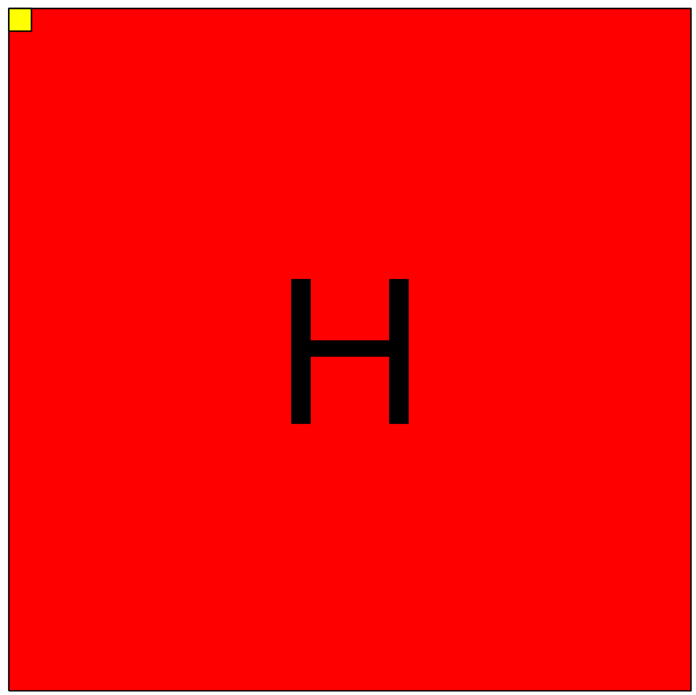
then $H_5^T\, H_9^T\, H_{20}\, H_{27} \;=$

$$I_{16} \otimes A^T \otimes I_8 \otimes A^T \otimes I_{1024} \otimes A \otimes I_{64} \otimes A \otimes I_8$$

Huge n from modest d via Kronecker Products

Perspective

The Google
Matrix $\Rightarrow$



2^{40}-by-2^{40}

$\uparrow\uparrow\uparrow$ KP Methodologies $\uparrow\uparrow\uparrow$

KP Singular Value Decomposition:

$$H \;=\; \sigma_1(B_1 \otimes C_1) \;+ \cdots +\; \sigma_R(B_R \otimes C_R)$$

KP Approximation:

$$H \;\approx\; B \otimes C$$

Flatten and Approximate:

$$\left(v_{ijk\ell}\right) \;=\; \begin{bmatrix} V_{11} & \cdots & V_{1d} \\ \vdots & \ddots & \vdots \\ V_{d1} & \cdots & V_{dd} \end{bmatrix} \;\approx\; \text{Sym} \otimes \text{Sym}$$

The Saga Continues...

Wavefunction-related computations lead to

$$\min_{a \in \mathrm{I\!R}^N} \frac{a^T H a}{a^T a}$$

However a is SO BIG that it cannot be stored explicitly.

The Curse of Dimensionality

Constrain for Tractability

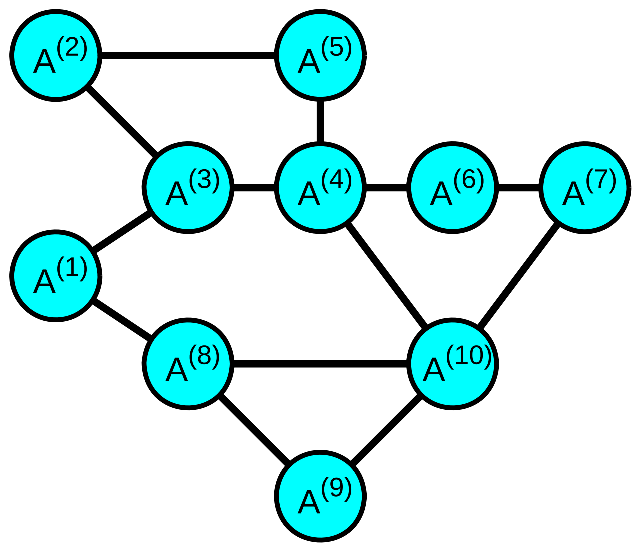
Minimize

$$r(a) = \frac{a^T H a}{a^T a}$$

subject to the contraint that a is a *tensor network*.

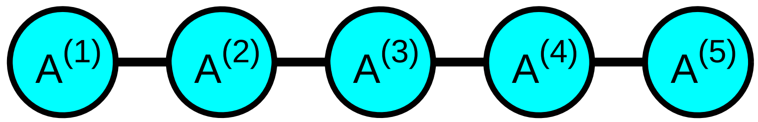
What is a Tensor Network?

A tensor network is a tensor of high dimension that is built up from many sparsely connected tensors of low-dimension.



Nodes are tensors and the edges are contractions.

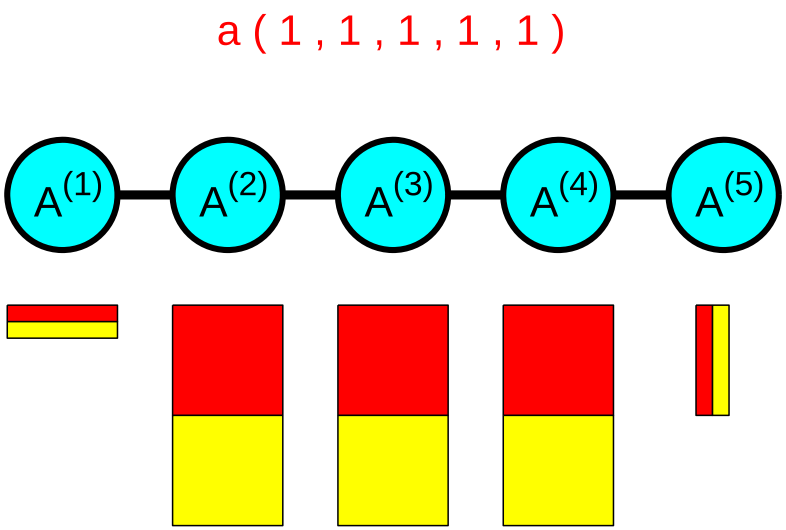
A 5-Site Linear Tensor Network



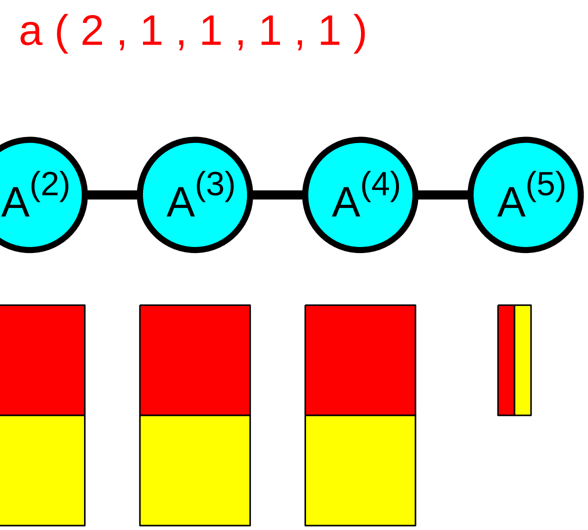
$$
\begin{aligned}
A^{(1)} &: 2 \times m \\
A^{(2)} &: m \times m \times 2 \\
A^{(3)} &: m \times m \times 2 \\
A^{(4)} &: m \times m \times 2 \\
A^{(5)} &: m \times 2
\end{aligned}
$$

m is a parameter, typically around 100.

If $a(1{:}2\,,1{:}2\,,1{:}2\,,1{:}2\,,1{:}2)$ is 5-site LTN then...

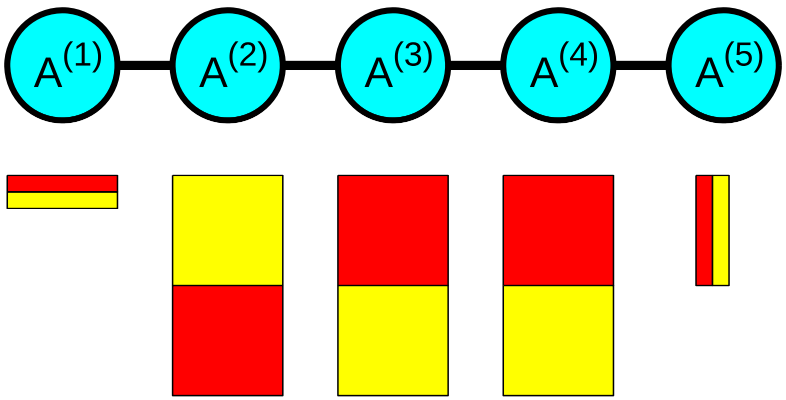


If $a(1{:}2\,,1{:}2\,,1{:}2\,,1{:}2\,,1{:}2)$ is 5-site LTN then...

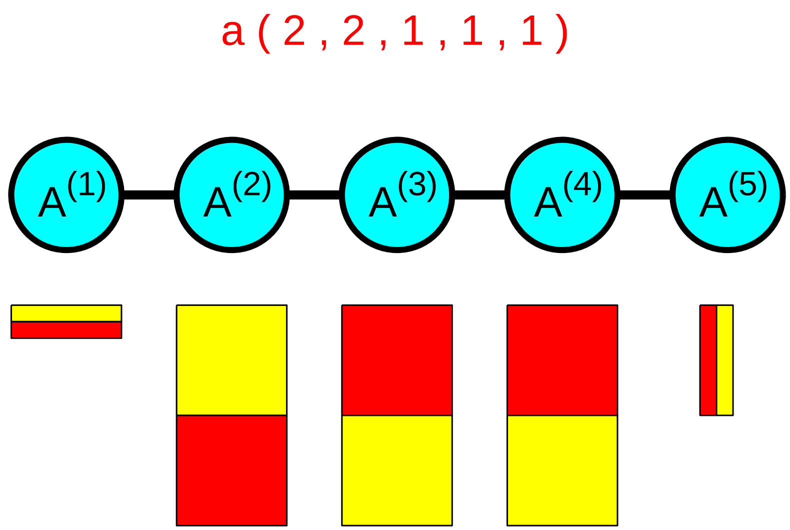


If $a(1{:}2\,,1{:}2\,,1{:}2\,,1{:}2\,,1{:}2)$ is 5-site LTN then...

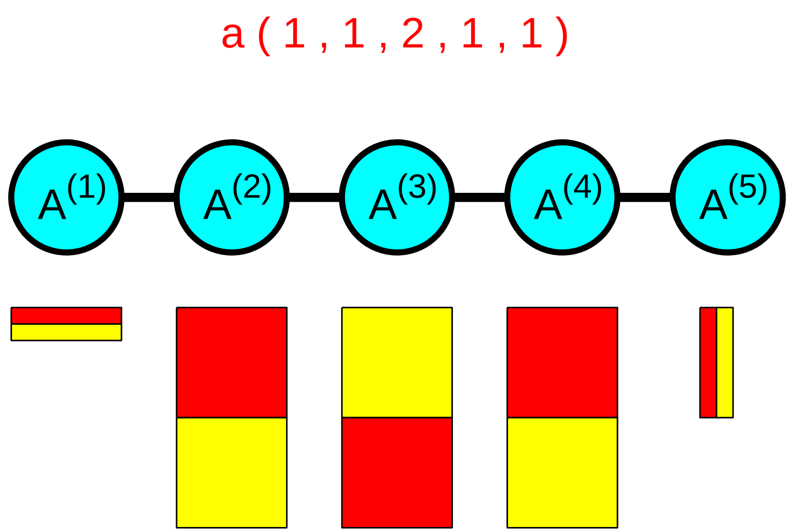
a (1 , 2 , 1 , 1 , 1)



If $a(1{:}2\,,1{:}2\,,1{:}2\,,1{:}2\,,1{:}2)$ is 5-site LTN then...



If $a(1{:}2\,,1{:}2\,,1{:}2\,,1{:}2\,,1{:}2)$ is 5-site LTN then...



A length-2^d vector that is represented by $O(dm^2)$ numbers.

LTN(5,m): Scalar Definition

$$a(n_1, n_2, n_3, n_4, n_5) \;=\; \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \sum_{i_3=1}^{m} \sum_{i_4=1}^{m} \boldsymbol{-} * \blacksquare * \blacksquare * \blacksquare * \boldsymbol{|}$$

$$A^{(1)}(n_1, i_1) \;*\; A^{(2)}(i_1, i_2, n_2) \;*\; A^{(3)}(i_2, i_3, n_3) \;*\; A^{(4)}(i_3, i_4, n_4) \;*\; A^{(5)}(i_4, n_5)$$

LTN(5,m): Scalar Definition

$$a(n_1, n_2, n_3, n_4, n_5) \;=\; \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \sum_{i_3=1}^{m} \sum_{i_4=1}^{m}$$

$$A^{(1)}(n_1, i_1) \;*\; A^{(2)}(i_1, i_2, n_2) \;*\; A^{(3)}(i_2, i_3, n_3) \;*\; A^{(4)}(i_3, i_4, n_4) \;*\; A^{(5)}(i_4, n_5)$$

Since a contraction is a generalized matrix product, then shouldn't it be described that way?

The Block Vec Product $\odot$

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \odot \begin{bmatrix} G_1 \\ G_2 \\ G_2 \end{bmatrix} = \begin{bmatrix} F_1 G_1 \\ F_1 G_2 \\ F_1 G_3 \\ F_2 G_1 \\ F_2 G_2 \\ F_2 G_3 \end{bmatrix}$$

The Block Vec Product $\odot$

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \odot \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \odot \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} = \begin{bmatrix} F_1G_1H_1 \\ F_1G_1H_2 \\ F_1G_2H_1 \\ F_1G_2H_2 \\ F_2G_1H_1 \\ F_2G_1H_2 \\ F_2G_2H_1 \\ F_2G_2H_2 \end{bmatrix}$$

LTN(5,m): $\odot$ Definition

$$\begin{bmatrix} a(1,1,1,1,1) \\ a(2,1,1,1,1) \\ a(1,2,1,1,1) \\ \vdots \\ a(1,2,2,2,2) \\ a(2,2,2,2,2) \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

BlockVec of a Tensor

If

$$A \;=\; A(1{:}N_1, 1{:}N_2, 1{:}N_3, 1{:}N_4)$$

then

$$\text{blockVec}(A) \;=\; \begin{bmatrix} A(:,:,1,1) \\ A(:,:,2,1) \\ A(:,:,3,1) \\ A(:,:,1,2) \\ A(:,:,2,2) \\ A(:,:,3,2) \end{bmatrix} \qquad N_3 = 3,\;\; N_4 = 2$$

A “Canonical” Contraction

$$A(i, j, n_3, n_4, m_3, m_4, m_5) \;=\; \sum_k B(i, k, n_3, n_4) C(k, j, m_3, m_4, m_5)$$

$$A(:, :, n_3, n_4, m_3, m_4, m_5) \;=\; B(:, :, n_3, n_4) C(:, :, m_3, m_4, m_5)$$

$$\mathrm{BlockVec}(A) \;=\; \mathrm{BlockVec}(B) \odot \mathrm{BlockVec}(C)$$

General Contractions

$$\sum_k B(n_1, n_2, k, n_4) C(m_1, m_2, m_3, k, m_5)$$

$\Downarrow$ Transposition

$$\sum_k \tilde{B}(n_1, k, n_2, n_4) \tilde{C}(k, m_2, m_3, m_1, m_5)$$

$\Downarrow$ Canonical Contraction

$$\tilde{A}(n_1, m_2, n_2, n_4, m_3, m_1, m_5)$$

$\Downarrow$ Transposition

$$A(n_1, n_2, n_4, m_2, m_3, m_1, m_5)$$

The Path to High Performance?

Optimized Level-3 BLAS

Multidimensional Transpose

Block Tensor Data Structures

The Saga Continues...

Minimize

$$r(a) = \frac{a^T H a}{a^T a}$$

subject to the constraint that

$$a = \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

Rephrased

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) \;=\; \frac{a^T H a}{a^T a}$$

where

$$a \;=\; \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

The Sweep Algorithm

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) = \frac{a^T H a}{a^T a}$$

where

$$a = \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

and only $A^{(1)}$ varies. (A small eigenproblem.)

The Sweep Algorithm

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) \;=\; \frac{a^T H a}{a^T a}$$

where

$$a \;=\; \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

and only $A^{(2)}$ varies. (A small eigenproblem.)

The Sweep Algorithm

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) \;=\; \frac{a^T H a}{a^T a}$$

where

$$a \;=\; \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

and only $A^{(3)}$ varies. (A small eigenproblem.)

The Sweep Algorithm

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) = \frac{a^T H a}{a^T a}$$

where

$$a = \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

and only $A^{(4)}$ varies. (A small eigenproblem.)

The Sweep Algorithm

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) \;=\; \frac{a^T H a}{a^T a}$$

where

$$a \;=\; \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{bmatrix}$$

and only $A^{(5)}$ varies. (A small eigenproblem.)

Componentwise Optimization

Can we do better?

Superpositioning

Given

$$A = \begin{bmatrix} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{bmatrix} \odot \cdots \odot \begin{bmatrix} A^{(d-1)}(:,:,1) \\ A^{(d-1)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} A^{(d)}(:,1) \\ A^{(d)}(:,2) \end{bmatrix}$$

$$B = \begin{bmatrix} B^{(1)}(1,:) \\ B^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} B^{(2)}(:,:,1) \\ B^{(2)}(:,:,2) \end{bmatrix} \odot \cdots \odot \begin{bmatrix} B^{(d-1)}(:,:,1) \\ B^{(d-1)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} B^{(d)}(:,1) \\ B^{(d)}(:,2) \end{bmatrix}$$

find

$$C = \begin{bmatrix} C^{(1)}(1,:) \\ C^{(1)}(2,:) \end{bmatrix} \odot \begin{bmatrix} C^{(2)}(:,:,1) \\ C^{(2)}(:,:,2) \end{bmatrix} \odot \cdots \odot \begin{bmatrix} C^{(d-1)}(:,:,1) \\ C^{(d-1)}(:,:,2) \end{bmatrix} \odot \begin{bmatrix} C^{(d)}(:,1) \\ C^{(d)}(:,2) \end{bmatrix}$$

so that $\| (A+B) - C \|_F = \min$

Something better that alternating least squares?

Another Constraint

Minimize

$$r(A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}) \; = \; \frac{a^T H a}{a^T a}$$

where

$$a \;=\; \left[\begin{array}{c} A^{(1)}(1,:) \\ A^{(1)}(2,:) \end{array}\right] \odot \left[\begin{array}{c} A^{(2)}(:,:,1) \\ A^{(2)}(:,:,2) \end{array}\right] \odot \left[\begin{array}{c} A^{(3)}(:,:,1) \\ A^{(3)}(:,:,2) \end{array}\right] \odot \left[\begin{array}{c} A^{(4)}(:,:,1) \\ A^{(4)}(:,:,2) \end{array}\right] \odot \left[\begin{array}{c} A^{(5)}(:,1) \\ A^{(5)}(:,2) \end{array}\right]$$

$\Uparrow$ $\Uparrow$ $\Uparrow$

Want these to have orthogonal columns.

Factoring Block Vector Products

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \odot \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} = \begin{bmatrix} Q_1 R \\ Q_2 R \end{bmatrix} \odot \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

$$= \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \odot \begin{bmatrix} RG_1 \\ RG_2 \end{bmatrix}$$

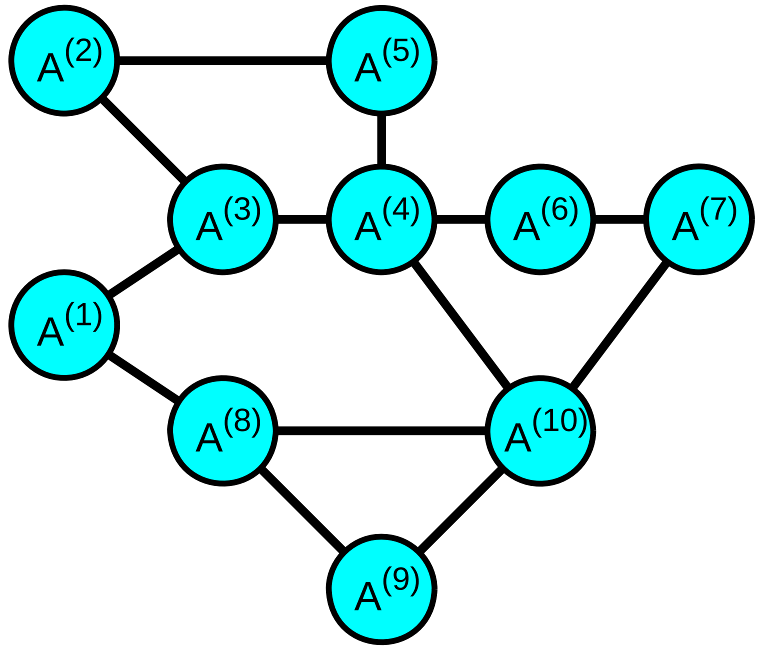
$$= \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \odot \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} S$$

$\uparrow\uparrow\uparrow$ Product QR/SVD $\uparrow\uparrow\uparrow$

Contractions that involve long sequences of matrix-matrix multiplications are numerically dangerous *unless* orthogonal matrices are involved.

Envision the maintenance of block vec orthogonality via product QR and product SVD.

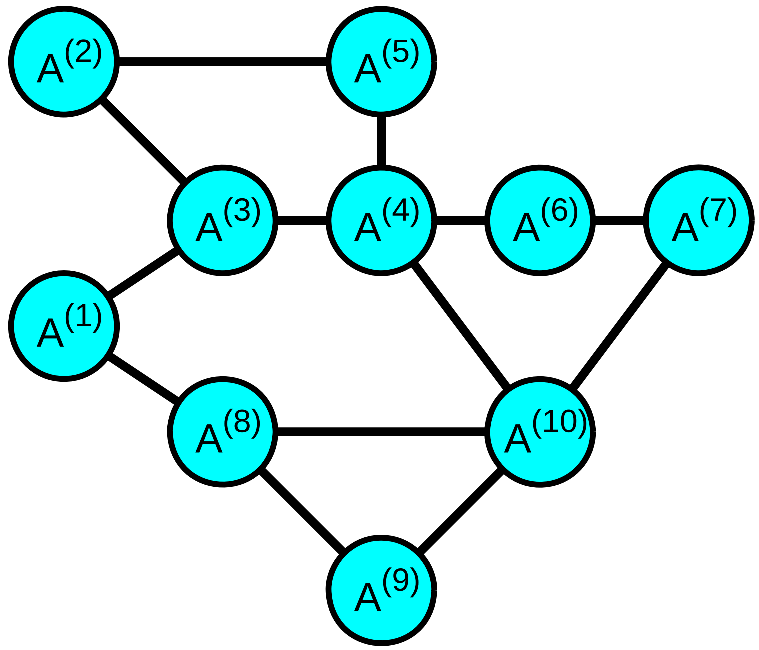
More Complicated Tensor Networks



At each site there is a tensor. Its dimension is $k+1$ where k is the number of site neighbors. E.g.,

$$A^{(2)} = A^{(2)}(1{:}m, 1{:}m, 1{:}2)$$
$$A^{(4)} = A^{(4)}(1{:}m, 1{:}m, 1{:}m, 1{:}m, 1{:}2)$$

Dealing With Complexity



Domain decomposition ideas?

Low-Rank Nodes?

Summary

The tensor network paradigm in quantum chemistry is a great venue to promote the idea of tensor-based computational thinking:

- Data structures. How do we lay out a tensor network in memory?
- Identifying important kernel operations and developing tensor BLAS
- Low rank representations to handle intermediate contractions.
- Nearness problems and Multilinear Optimization.