Approximating matrices with multiple symmetries: with an application to quantum chemistry

Joseph Vokt, Senior, Computer Science
jpv52@cornell.edu
Joint-work with
Professor Charles Van Loan, Computer Science
cv@cs.cornell.edu

Themes for this talk

- Symmetry
- Low-rank approximation
- Preserving symmetry for low-rank approximation
- Bridging the gap from matrix to tensor

ERI Tensor Symmetry: Background

- The electron repulsion integral (ERI) tensor is defined by

$$\mathcal{A}(i,j,k,\ell) = \int d\mathsf{x}_1 d\mathsf{x}_2 \chi_i(\mathsf{x}_1)^* \chi_j(\mathsf{x}_1) \frac{1}{r_{12}} \chi_k(\mathsf{x}_2)^* \chi_\ell(\mathsf{x}_2)$$

- $\chi(\mathsf{x})$ is a molecular orbital that describes an electron's motion at coordinate $\mathsf{x} = (r, \theta, \phi, \omega)$ where ω is the spin coordinate
- This tensor describes two-body Coulomb interaction in molecular electronic structure theory
- Source: Szabo and Ostlund's *Modern Quantum Chemistry*

ERI Tensor Symmetry: Background

- For real orbitals, we have the following symmetries:

$$
\begin{aligned}
&\mathcal{A}(i,j,k,\ell) = \mathcal{A}(k,\ell,i,j) = \mathcal{A}(j,i,\ell,k) = \mathcal{A}(\ell,k,j,i) \\
&= \mathcal{A}(j,i,k,\ell) = \mathcal{A}(\ell,k,i,j) = \mathcal{A}(i,j,\ell,k) = \mathcal{A}(k,\ell,j,i)
\end{aligned}
$$

- For our purposes, we will be primarily concerned with

$$
\mathcal{A}(i,j,k,\ell) = \mathcal{A}(j,i,k,\ell) = \mathcal{A}(i,j,\ell,k) = \mathcal{A}(k,\ell,i,j)
$$

ERI Tensor Symmetry: Problem Statement

Compute μ, the molecular orbital integral transformation

$$\begin{aligned} \mu &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n} \mathcal{A}(i,j,k,\ell) v_i v_j v_k v_\ell \\ &= (v \otimes v)^T A (v \otimes v) \end{aligned}$$

where $v \in \mathbb{R}^n$ and $\mathcal{A} \in \mathbb{R}^{n\times n\times n\times n}$ is the fourth-order ERI tensor and A is the tensor unfolding.

ERI Tensor Symmetry: Motivation

- Applications of ERI tensors
  - *Ab initio* protein folding simulation
  - Rational therapeutic drug design
  - Engineering nano-scale devices
- The more electrons, the higher the dimensionality d
  - d-body problem at the heart of quantum chemistry

ERI Tensor Symmetry: Goal

- Goal: a low-rank approximation which preserves the original structure *without* using the SVD, while minimizing ERI evaluations
  - Why low-rank?
    - Data sparsity
  - Why preserve structure?
    - Numerical motivation: good to maintain data sparsity
    - Physical motivation: must not violate the Pauli principle
  - Why not use the SVD?
    - The SVD of an $n^2 x n^2$ matrix takes $O(n^6)$ flops
  - Why minimize ERI evaluations?
    - ERI evaluations are expensive to compute

Centrosymmetry: A tale of two symmetries

A persymmetric matrix is symmetric about its antidiagonal:

$$A = \begin{bmatrix} 1 & 2 & 3 & \boxed{4} \\ 5 & 6 & \boxed{7} & 3 \\ 8 & \boxed{9} & 6 & 2 \\ \boxed{10} & 8 & 5 & 1 \end{bmatrix}.$$

Centrosymmetry: A tale of two symmetries

A matrix $A \in \mathbb{R}^{n\times n}$ is persymmetric if $A^T = E_n A E_n$ where $E_n \in \mathbb{R}^{n\times n}$ is the n-by-n exchange permutation, e.g.,

$$E_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Centrosymmetry: A tale of two symmetries

A centrosymmetric matrix is a symmetric persymmetric matrix $A = E_n A E_n$ e.g.,

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 6 & 3 \\ 3 & 6 & 5 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$

Centrosymmetry: Block Structure

If $A \in \mathbb{R}^{n \times n}$ is blocked as

$$A = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{21} & A_{22} \end{array} \right] \qquad A_{ij} \in \mathbb{R}^{m \times m}$$

and A is centrosymmetric, where n is even, then

$$A = \left[\begin{array}{cc} A_{11} & A_{12} \\ E_m A_{12} E_m & E_m A_{11} E_m \end{array} \right]$$

Centrosymmetry: Block Diagonalization

If we define Q_E as

$$Q_E = \frac{1}{\sqrt{2}} \left[\begin{array}{c|c} I_m & I_m \\ E_m & -E_m \end{array} \right] = \left[\begin{array}{c|c} Q_+ & Q_- \end{array} \right]$$

then Q_E is orthogonal and

$$Q_E^T A Q_E = \left[\begin{array}{cc} A_{11} + A_{12} E_m & 0 \\ 0 & A_{11} - A_{12} E_m \end{array} \right]$$

is the block diagonalization of A.

Centrosymmetry: Block Factorizations

- When A is positive definite, diagonal blocks are too
- Stable LDL^T factorizations:

$$P_+(A_{11} + A_{12}E_m)P_+^T = L_+D_+L_+^T$$
$$P_-(A_{11} - A_{12}E_m)P_-^T = L_-D_-L_-^T$$

- Obtaining both factorizations takes $2(\frac{1}{3}(\frac{n}{2})^3) = \frac{1}{4}(\frac{n^3}{3})$ flops
- Centrosymmetric A lets you solve $Ax = b$ in $\frac{1}{4}$ of the work as compared to non-symmetric A

Centrosymmetry: Low-rank approximation

Let

$$Y_+ \;=\; Q_+ P_+^T L_+ \;=\; [\, y_+^{(1)} \mid \cdots \mid y_+^{(m_+)} \,]$$

and

$$Y_- \;=\; Q_- P_-^T L_- \;=\; [\, y_-^{(1)} \mid \cdots \mid y_-^{(m_-)} \,]$$

Centrosymmetry: Low-rank approximation

$$
\begin{aligned}
A &= Q_E \left[\begin{array}{cc} A_{11} + A_{12} E_m & 0 \\ 0 & A_{11} - A_{12} E_m \end{array} \right] Q_E^T \\
&= Q_+ (A_{11} + A_{12} E_m) Q_+^T \; + \; Q_- (A_{11} - A_{12} E_m) Q_-^T \\
&= Y_+ D_+ Y_+^T \; + \; Y_- D_- Y_-^T \\
&= \sum_{k=1}^{m_+} d_+^{(k)} y_+^{(k)} [y_+^{(k)}]^T \; + \; \sum_{k=1}^{m_-} d_-^{(k)} y_-^{(k)} [y_-^{(k)}]^T
\end{aligned}
$$

- By terminating these sums early, we obtain low rank approximations that are also centrosymmetric.

ERI Tensor Symmetry: Block Structure

- Unfold $\mathcal{A}(i, j, k, \ell)$ as $[A_{k,\ell}]_{i,j}$

$$A = \mathcal{A}_{[1,3]\times[2,4]}$$

- $\mathcal{A}(i, j, k, \ell)$ is the (i, j) entry of the (k, ℓ) block of A
- Assume $n = 3$ for the following examples

ERI Tensor Symmetry: Block Structure

- If $\mathcal{A}(i,j,k,\ell) = \mathcal{A}(j,i,k,\ell)$, then $[A_{k,\ell}]_{i,j} = [A_{k,\ell}]_{j,i}$
- Symmetric blocks: each 3x3 block matrix is symmetric along its diagonal

1	2	3	19	20	21	37	38	39
2	4	5	20	22	23	38	40	41
3	5	6	21	23	24	39	41	42
7	8	9	25	26	27	43	44	45
8	10	11	26	28	29	44	46	47
9	11	12	27	29	30	45	47	48
13	14	15	31	32	33	49	50	51
14	16	17	32	34	35	50	52	53
15	17	18	33	35	36	51	53	54

ERI Tensor Symmetry: Block Structure

- If $\mathcal{A}(i,j,k,\ell) = \mathcal{A}(i,j,\ell,k)$, then $[A_{k,\ell}]_{i,j} = [A_{\ell,k}]_{i,j}$
- Block symmetry: the 9x9 matrix is symmetric at the 3x3 block matrix level

1	4	7	10	13	16	19	22	25
2	5	8	11	14	17	20	23	26
3	6	9	12	15	18	21	24	27
10	13	16	28	31	34	37	40	43
11	14	17	29	32	35	38	41	44
12	15	18	30	33	36	39	42	45
19	22	25	37	40	43	46	49	52
20	23	26	38	41	44	47	50	53
21	24	27	39	42	45	48	51	54

ERI Tensor Symmetry: Block Structure

- If $\mathcal{A}(i,j,k,\ell) = \mathcal{A}(k,\ell,i,j)$, then $[A_{k,\ell}]_{i,j} = [A_{i,j}]_{k,\ell}$
- Perfect shuffle permutation symmetry: entry (i,j) in the (k,ℓ) block is entry (k,ℓ) in the (i,j) block

1	4	7	4	25	28	7	28	40
2	5	8	12	26	29	15	33	41
3	6	9	19	27	30	22	37	42
2	12	15	5	26	33	8	29	41
10	13	16	13	31	34	16	34	43
11	14	17	20	32	35	23	38	44
3	19	22	6	27	37	9	30	42
11	20	23	14	32	38	17	35	44
18	21	24	21	36	39	24	39	45

ERI Tensor Symmetry: Block Structure

- With all eight symmetries

1	2	3	2	7	8	3	8	12
2	4	5	7	9	10	8	13	14
3	5	6	8	10	11	12	14	15
2	7	8	4	9	13	5	10	14
7	9	10	9	16	17	10	17	19
8	10	11	13	17	18	14	19	20
3	8	12	5	10	14	6	11	15
8	13	14	10	17	19	11	18	20
12	14	15	14	19	20	15	20	21

ERI Tensor Symmetry: Block Diagonalization

- Define $Q = [Q_{\text{sym}} | Q_{\text{skew}}]$ such that
  - Q_{sym} and Q_{skew} are sparse orthonormal bases for vectorized symmetric and skew-symmetric matrices respectively
- Then Q is orthogonal and

$$Q^T A Q = \begin{bmatrix} A_{\text{sym}} & 0 \\ 0 & A_{\text{skew}} \end{bmatrix}$$

is the block diagonalization of A.

ERI Tensor Symmetry: Block Diagonalization

When $n = 3$, $Q \in \mathbb{R}^{9\times 9}$

$$Q_9 = \frac{1}{\sqrt{2}} \left[\begin{array}{cccccc|ccc}
\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\
0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\
0 & 0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right] = [Q_{\text{sym}} | Q_{\text{skew}}]$$

$$Q_9(:,4) \equiv \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad Q_9(:,7) \equiv \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

ERI Tensor Symmetry: Block Factorizations

- When A is positive semi-definite, A_{sym} and A_{skew} are too
- Stable symmetric-pivoting LDL^T factorizations:

$$P_{\text{sym}} A_{\text{sym}} P_{\text{sym}}^T = L_{\text{sym}} D_{\text{sym}} L_{\text{sym}}^T$$
$$P_{\text{skew}} A_{\text{skew}} P_{\text{skew}}^T = L_{\text{skew}} D_{\text{skew}} L_{\text{skew}}^T$$

- Compute a lazy-evaluation, symmetric-pivoting, rank-revealing LDL^T factorization on each block
- Without rank-revealing LDL^T
  - $n^6/24$ flops for one factorization
  - $n^6/12$ flops total
- With rank-revealing LDL^T, where r is the rank of A
  - $r^2(n(n+1)/2) = r^2(n^2/2 + n/2)$ for one factorization
  - $r^2 n^2$ flops total

ERI Tensor Symmetry: Low-rank approximation

Let

$$Y_{\text{sym}} \;=\; Q_{\text{sym}} P_{\text{sym}}^T L_{\text{sym}} \;=\; [\, y_{\text{sym}}^{(1)} \mid \cdots \mid y_{\text{sym}}^{(m_{\text{sym}})} \,]$$

and

$$Y_{\text{skew}} \;=\; Q_{\text{skew}} P_{\text{skew}}^T L_{\text{skew}} \;=\; [\, y_{\text{skew}}^{(1)} \mid \cdots \mid y_{\text{skew}}^{(m_{\text{skew}})} \,]$$

ERI Tensor Symmetry: Low-rank approximation

$$
\begin{aligned}
A &= Q \begin{bmatrix} A_{\mathrm{sym}} & 0 \\ 0 & A_{\mathrm{skew}} \end{bmatrix} Q^T \\
&= Q_{\mathrm{sym}} A_{\mathrm{sym}} Q_{\mathrm{sym}}^T \;+\; Q_{\mathrm{skew}} A_{\mathrm{skew}} Q_{\mathrm{skew}}^T \\
&= Y_{\mathrm{sym}} D_{\mathrm{sym}} Y_{\mathrm{sym}}^T \;+\; Y_{\mathrm{skew}} D_{\mathrm{skew}} Y_{\mathrm{skew}}^T \\
&= \sum_{k=1}^{m_{\mathrm{sym}}} d_{\mathrm{sym}}^{(k)} y_{\mathrm{sym}}^{(k)} [y_{\mathrm{sym}}^{(k)}]^T \;+\; \sum_{k=1}^{m_{\mathrm{skew}}} d_{\mathrm{skew}}^{(k)} y_{\mathrm{skew}}^{(k)} [y_{\mathrm{skew}}^{(k)}]^T \\
&= \sum_{k=1}^{m_{\mathrm{sym}}} d_{\mathrm{sym}}^{(k)} B_k \otimes B_k + \sum_{k=1}^{m_{\mathrm{skew}}} d_{\mathrm{skew}}^{(k)} C_k \otimes C_k
\end{aligned}
$$

- B_k, C_k is the $n \times n$ reshaping of $y_{\mathrm{sym}}^{(k)}, y_{\mathrm{skew}}^{(k)}$ respectively
- By terminating these sum early, we obtain low-rank approximations with ERI tensor symmetry.

ERI Tensor Symmetry: Integral transformation

$$
\begin{aligned}
\mu &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathcal{A}(i,j,k,l) v_i v_j v_k v_l \\
&= (v \otimes v)^T A (v \otimes v) \\
&= (v \otimes v)^T \left(\sum_{k=1}^{m_{\text{sym}}} d_{\text{sym}}^{(k)} B_k \otimes B_k + \sum_{k=1}^{m_{\text{skew}}} d_{\text{skew}}^{(k)} C_k \otimes C_k\right)(v \otimes v) \\
&= \sum_{k=1}^{m_{\text{sym}}} d_{\text{sym}}^{(k)} \cdot (v \otimes v)^T (B_k \otimes B_k)(v \otimes v)^T \\
&+ \sum_{k=1}^{m_{\text{skew}}} d_{\text{sym}}^{(k)} \cdot (v \otimes v)^T (B_k \otimes B_k)(v \otimes v)^T \\
&= \sum_{k=1}^{m_{\text{sym}}} d_{\text{sym}}^{(k)} \cdot (v^T B_k v)^2 + \sum_{k=1}^{m_{\text{skew}}} d_{\text{skew}}^{(k)} \cdot (v^T C_k v)^2
\end{aligned}
$$

ERI Tensor Symmetry: Integral transformation

- A_{skew} factorization isn't necessary
- Let C_k be the $n \times n$ reshaping of $y_{\text{skew}}^{(k)} \in \mathbb{R}^{n^2}$
- By construction $C_k = -C_k^T$ is skew-symmetric, so $v^T C_k v = v^T C_k^T v = -v^T C_k v = 0$
- The second summation is zero

MATLAB Implementation

- IntegralTransformation
  - Calls StructLDLT on the tensor unfolding
  - Does the integral transformation
- StructLDLT
  - Computes diagonal block A_{sym} (and A_{skew} optionally)
    - Uses QsymT (next slide)
  - Calls LazyLDLT on A_{sym} (and on A_{skew} optionally)
  - Returns Y_{sym} (and Y_{skew} optionally)
    - Uses Qsym
- LazyLDLT
  - Evaluate the diagonal of A [n^2 ERIs]
  - While d(j) is greater than the zero threshold [r loops]
    - Search d(j:n) for largest diagonal element
    - Swap d(k) and d(j)
    - Update pivot vector, permute rows of L and A
    - Evaluate the next subcolumn of A [$O(n^2)$ ERIs]
    - Compute d(j) and column j of L

MATLAB Implementation: QsymT

- Goal: find a way to compute $Q_{\text{sym}}^T v$, $v \in \mathbb{R}^{n^2}$ by utilizing the structure of Q_{sym}^T
  - Sparse matrix-multiply?

MATLAB Implementation: QsymT

- No matrix-multiplication necessary

$$Q_{sym}^T \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{bmatrix} = Q_{sym}^T vec \begin{bmatrix} v_1 & v_4 & v_7 \\ v_2 & v_5 & v_8 \\ v_3 & v_6 & v_9 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_5 \\ v_9 \\ (v_2 + v_4)/\sqrt{2} \\ (v_3 + v_7)/\sqrt{2} \\ (v_6 + v_9)/\sqrt{2} \end{bmatrix}$$

MATLAB Implementation: Qsym

$$Q_{sym}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_4 \\ v_5 \\ v_4 \\ v_2 \\ v_6 \\ v_5 \\ v_6 \\ v_3 \end{bmatrix} = vec\begin{bmatrix} v_1 & v_4 & v_5 \\ v_4 & v_2 & v_6 \\ v_5 & v_6 & v_3 \end{bmatrix}$$

Summary

- Centrosymmetry and ERI tensor symmetry
    - Block structure
    - Block diagonalization
    - Block factorizations
    - Low-rank approximation with original symmetry
- Structure leads to reduced work
    - $O(r^2 n^2)$ algorithm for computing computing a molecular integral transformation

Future work

- Parallel implementation
  - Exploit block structure
- Sixth order tensor approximation
  - First three indices are super-symmetric, last three indices are super-symmetric
  - Key problem is finding a sparse orthogonal basis to project onto
  - Group theory approach looks promising
- d-order tensor approximation
- Thanks to Charles Van Loan, the McNair Scholars program, and David Bindel