Cholesky Decomposition Techniques in Quantum Chemical Implementations
Outline

- What is MOLCAS?
- A crash course in Cholesky Decomp. (CD)
- LK Exchange
- 1C-CD
- Analytic CD gradients
- aCD on-the-fly RI auxiliary basis set
- AcCD auxiliary basis sets
- Some showcases
What is MOLCAS?
The hallmark:

a-CASSCF/MS-CASPT2/ANO is and will be our protocol of choice

Typical applications:

- Chemical reactions
- Photo Chemistry
- Heavy Element Chemistry
Chemical reactions:

“Chemiluminescence of 1,2-dioxetane. Mechanism uncovered”
Photo Chemistry:

“Intramolecular triplet-triplet energy transfer in oxa- and aza-dipipi-methane photosensitized systems”
Heavy Element Chemistry:

“Agnostic interaction in the methylidene metal dihydride complexes $H_2MCH_2$ ($M = Y, Zr, Nb, Mo, Ru, Th, or U$)"
CD technique is trivial and involve no more than elementary vector manipulations.

In an essence CD of 2-electron integrals is a truncated version of a standard Gram-Schmidt orthogonalization in a Coulumbic metric.

Let me demonstrate!
The Gram-Schmidt is formulated as
\[ l = i - \sum_{K=1}^{l-1} K \langle K | i \rangle \]
Or in matrix form
\[ \langle j | l \rangle = \langle j | i \rangle - \sum_{K=1}^{l-1} \langle j | K \rangle \langle K | i \rangle \]
and
\[ \langle l | l \rangle = \langle l | i \rangle \]
\[ \langle i | l \rangle = \langle i | i \rangle - \sum_{K=1}^{l-1} \langle i | K \rangle \langle K | i \rangle \]
Imagine the identity

\[ V_{ij} = \sum_{kl} V_{ik} (V^{-1})_{kl} V_{lj} \]

Transform some index to the GS basis

\[ V_{ij} = \sum_{K} V_{iK} (V^{-1/2})_{K} (V^{-1/2})_{K} V_{Kj} = \sum_{K} L_{i}^{K} L_{j}^{K} \]

Finally expressions

\[ L_{i}^{l} = \left( V_{ii} - \sum_{K=1}^{l-1} L_{i}^{K} L_{i}^{K} \right)^{-1/2} = L_{i}^{l} \]

\[ L_{j}^{l} = \left( V_{ji} - \sum_{K=1}^{l-1} L_{j}^{K} L_{j}^{K} \right) / L_{i}^{l} \]
The GS procedure lends itself to a single parameter controlled truncation of the GS basis.

A list of all 

\[ V_{ii} = V_{ii} - \sum_{k=1}^{n} V_{ik} V_{ik} \]

is stored and updated as we include new GS basis functions. If all remaining \( V_{ii} \) are below the threshold then terminate!
In DF-RI-CD the 2-electron integrals are expressed as
\[ \langle ij | kl \rangle = \sum_{J} L_{ij}^{J} L_{kl}^{J} \]
\[ \langle ij | kl \rangle = V_{ij,kl} \]
\[ \langle l | J \rangle = V_{lj} \]

For DF/RI we have
\[ L_{ij}^{J} = \sum_{l} \langle ij | l \rangle (V^{-1/2})_{lj} \]

While for CD we have
\[ L_{ij}^{J} = \langle ij | J \rangle (V_{JJ})^{-1/2} \]
\[ L_{jj} = \left( V_{jj} - \sum_{K=1}^{J-1} L_{jK}^{2} \right)^{1/2} \]
\[ L_{ij} = \left( V_{ij} - \sum_{K=1}^{J-1} L_{iK}^{2} L_{jK} \right) / L_{jj} \]
### Comparison

<table>
<thead>
<tr>
<th></th>
<th><strong>DF/RI-version</strong></th>
<th><strong>CD-version</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auxiliary Basis set</td>
<td>External</td>
<td>Internal (num.)</td>
</tr>
<tr>
<td>Gradient</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Method-dependent?</td>
<td>Yes mostly</td>
<td>No</td>
</tr>
<tr>
<td>Parameter dependent?</td>
<td>Not directly</td>
<td>Yes</td>
</tr>
<tr>
<td>Could be exact?</td>
<td>Not automatically</td>
<td>Yes</td>
</tr>
</tbody>
</table>
The LK approach for Exchange

- **CD localization of the occupied orbitals**
  The CD localization scheme is a non-iterative procedure.

- **Error bounded screening**
  Reformulation, ERI matrix in AO basis is positive definite and satisfy the Cauchy-Schwarz inequality.
LK Scaling

Linear Glycines / cc-pVDZ
Cholesky Localization

Cholesky localization is not perfect – do we care?

![Graph showing CPU time vs. number of basis functions for different methods]

- Canonical
- Cholesky MOs
- Boys
- Pipek-Mezey
Q: Will a CD procedure which exclude all 2-center products as potential auxiliary basis function retain an acceptable accuracy?
Test set: 21 reactions, B3LYP 6-31G structures
1-Center CD vs. full CD

RMS Error / SVP

Method

kcal/mol

SVWN
B3LYP
MP2
Observations

- 1-Center approximation *does not* degrade the CD accuracy significantly!
- for 1C-CD the decomposition time is **3-4 times faster than the full CD** with the same threshold.

*In the 1C-CD approximation a fixed auxiliary basis set is used, hence we can compute analytic derivatives!*
This is now a trivial matter! Use RI (and LK) technology!

<table>
<thead>
<tr>
<th>Molecule/Basis set</th>
<th>( \mathcal{X}^a )</th>
<th>( N^b )</th>
<th>1C-CD ( \delta = 10^{-2} )</th>
<th>1C-CD ( \delta = 10^{-4} )</th>
<th>Integral-direct</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ (Gly)(_1)</td>
<td>10</td>
<td>95</td>
<td>4.1 (6)</td>
<td>4.6 (7)</td>
<td>8.5 (8)</td>
</tr>
<tr>
<td>cc-pVDZ (Gly)(_{10})</td>
<td>73</td>
<td>734</td>
<td>371 (209)</td>
<td>573 (221)</td>
<td>1257 (666)</td>
</tr>
<tr>
<td>cc-pVDZ (Gly)(_{20})</td>
<td>143</td>
<td>1444</td>
<td>1578 (593)</td>
<td>2159 (993)</td>
<td>2298 (2298)</td>
</tr>
<tr>
<td>cc-pVDZ (Gly)(_{30})</td>
<td>213</td>
<td>2154</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cc-pVTZ (Gly)(_1)</td>
<td>10</td>
<td>220</td>
<td>27.8 (21)</td>
<td>41.3 (23)</td>
<td>95.5 (45)</td>
</tr>
<tr>
<td>cc-pVTZ (Gly)(_{10})</td>
<td>73</td>
<td>1678</td>
<td>2738 (871)</td>
<td>6007 (993)</td>
<td>(4970)</td>
</tr>
</tbody>
</table>

\(^a\) Number of atoms. \(^b\) Number of basis functions.
### 1C-CD vs. Conv.: Energies and Bond Distances

**TABLE II.** 1C-CD DFT/BLYP results at different decomposition thresholds $\delta$, employing the SVP basis set. Accuracy $|\Delta E|$ (in mhartrees) of computed total energies and maximum deviation $|\Delta r|$ (in pm) of bond distances compared to conventional calculations. All calculations have been performed without point group symmetry restrictions.

| Molecule | $|\Delta E|$ | $|\Delta r|$ |
|----------|--------------|--------------|
| $N_2$    | $\delta=10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ |
|          | 3.65         | 1.48         | 0.015      | 0.009      | 0.0009     | 0.0298     | 0.0001     | 0.0001     |
| $C_2H_2$ | 4.35         | 0.08         | 0.006      | 0.003      | 0.1615     | 0.0246     | 0.0107     | 0.0112     |
| $C_2H_6$ | 4.04         | 1.33         | 0.021      | 0.009      | 0.7536     | 0.0683     | 0.0219     | 0.0276     |
| $C_6H_6$ | 7.76         | 1.54         | 0.061      | 0.075      | 0.0252     | 0.0297     | 0.0051     | 0.0054     |
| $CH_3OH$ | 3.91         | 1.32         | 0.039      | 0.012      | 0.1248     | 0.0067     | 0.0019     | 0.0017     |
| HCOCl    | 4.83         | 1.25         | 0.053      | 0.026      | 0.1183     | 0.0148     | 0.0073     | 0.0071     |
| SO$_2$   | 5.33         | 1.38         | 0.041      | 0.010      | 0.0133     | 0.0099     | 0.0021     | 0.0008     |
Q: Given the accuracy of the 1C-CD approach, could it be used to design general DF/RI auxiliary basis sets which are method-free?

Use *atomic* CD technique to design the aCD RI basis sets. Plug them into your RI code! aCD/RI

**aCD/RI:** 1C-CD quality results without the recursive nature of CD
Accuracy of aCD RI basis sets

On-the-fly CD of the atomic two-electron integral matrix
Baker test set of reactions (cc-pVTZ)

<table>
<thead>
<tr>
<th>Method</th>
<th>Aux. basis</th>
<th>Total Energy Max Error kcal mol⁻¹</th>
<th>Activation Energy Max Error kcal mol⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP</td>
<td>aCD-3</td>
<td>1.516</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>aCD-4</td>
<td>0.062</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>aCD-5</td>
<td>0.062</td>
<td>0.062</td>
</tr>
<tr>
<td>MP2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>aCD-3</td>
<td>1.406</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>aCD-4</td>
<td>0.187</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>aCD-5</td>
<td>0.187</td>
<td>0.047</td>
</tr>
</tbody>
</table>

<sup>a</sup> MP2<sup>a</sup> RI basis set is 2-4 times faster than 1C-CD!
The CD approach does not automatically eliminate the redundancy in the primitive space product space! But we can do that!

- Compute a normal aCD basis set.
- Do a complementary CD elimination in the primitive product space. Keep essential products as exponents of the acCD basis.
- Do a least-square fit to the aCD basis set to get the contract coefficients of the acCD basis.
Tc ANO-RCC auxiliary basis set
\( \tau = 1.0 \times 10^{-4} \)

Tc-ANO: 21 s-fnctns
Tc-aCD: 231 s-fnctns
Tc-acCD: 32 s-fnctns

Total # of products
conv: 23871
aCD: \textbf{17272}
acCD: \textbf{3946}
The CD-RI hierarchy in MOLCAS

- CD(τ=0) – Conventional
- CD(τ)
- 1C-CD(τ)
- aCD(τ)/RI
- acCD(τ)/RI
- External aux. bfn/RI
CD developments: material published so far

- “Analytic derivatives for the Cholesky representation of the two-electron integrals” - **CD gradients**
- “Unbiased auxiliary basis sets for accurate two-electron integral approximations” - **CD-RI auxiliary basis sets**
- “Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): Application to the spin-state energetics of Co-III(diiminato) (NPh) - **CD-CASPT2**
- “Accurate ab initio density fitting for multiconfigurational self-consistent field methods“ - **CD-CASSCF**
- ”Quartic scaling evaluation of canonical scaled opposite spin second-order Moller-Plesset correlation energy using Cholesky decompositions” - **CD-MP2**
- “Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals“ - **CD-HF**
CD-CASPT2 Example: Relative energies of spin-states of Ferrous complex

K. Pierloot et al.

-CD(-6)-CASPT2/CASSCF(14-in-16)/ANO
-810 bfn (no symmetry) / 964 bfn (C\textsubscript{2} symmetry)

FIG. 1. (Color online) Examples of [Fe(L)(NHS\textsubscript{4})] complexes. (A) The cis-NH\textsubscript{3} complex. (B) the trans-CO complex.
CD-CASPT2 example

Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): Application to the spin-state energetics of Co-III(diiminato)(NPh)

Aquilante et al.
- CASSCF/CASPT2
- ANO-RCC-VTZP
- 869 bfn

*Figure 1. Model of the Co–imido complex used in the calculations.*
Highly Accurate CCSD(T) and DFT-SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer

Pitonak et al., CPC

- MP2 – 1648 bfn
- CCSD(T) – 920 bfn
Model for Biotin@Avidin: 257 atoms, 924 e\textsuperscript{-}

6-31G\* : 2091 bsf

Cholesky-MP2: 44 hr on 4 CPUs

0.1 mH accuracy in the total energy (\(\delta = 10^{-4}\))
1,3-DIPHENYLISOBENZOFURAN:
Photovoltaic material with singlet fission
Zdenek Havlas, Andrew Schwerin, and Jozef Michl

CAS(16el/14orb; 7a,7b)
ANO-L(C,O: 4s3p2d1f, H: 3s2p1d, Ryd(8,8,8)/[1,1,1])
Cholesky (Thrs= 1.0d-5)
35 atoms (O₁C₂₀H₁₄)
835 orbitals (419a, 416b)
Speed up: 48 h - > 15 h
**CHOLESKY DECOMPOSITION IN MOLCAS 7.0 (Z. Havlas et al.)**

(Ethylene)$_n$, stacked with distance 3Å

- **Basis:** cc-pVDZ
- **Active space:** $n \times (2,2)$, $\pi$ orbitals only
- **Processors:** AMD Athlon 64 X2 Dual Core Processor 4800+
- **CPU speed:** 2.4 GHz
Interested in more examples?

The Show Case Room

This web page is dedicated to scientific papers in which MOLCAS has been of significance. This we hope will inspire old and new MOLCAS users to see what is possible with the MOLCAS software. This web page contains four sections:

1. a dynamically updated list of all papers which quotes the official "MOLCAS paper" [MOLCAS: a program package for computational chemistry],
2. a list of MOLCAS hallmark application papers if you have a paper you think should be on the list email roland.lindh@teokem.lu.se, please!
3. a list of MOLCAS methods papers.

**MOLCAS reference papers since September 2004: search the ISI data base!**

1. "Molecular Basis of DNA Photodimerization: Intrinsic Production of Cyclobutane Cytosine Dimers."

2. "Cholesky decomposition based multiconfigurational second order perturbation theory (CD-CASPT2): Application to the spin state energetics of Co[III][diminato](NPH)."

   L. Serrano-Andrés, M. Merchán, A.C. Borin.

Done
Furture work!

- HF 1C-CD and RI gradients
- CASSCF 1C-CD and RI gradients
- MP2 1C-CD gradients
- Localized and linear scaling RI and CD
- Numerical problems with accurate RI/CD
- CD in the N-particle space
Summary

- DF, RI and CD are interrelated
- LK Exchange
- Gradients for CD
- $1C$-CD approximation is equivalent in performance and accuracy to DF/RI
- $1C$-CD approach can be used to derive “method-free” aCD RI auxiliary basis sets
- Much smaller acCD auxiliary basis sets can be derived from aCD basis sets without any further loss of accuracy.