Cholesky decomposition of two-electron integrals.

a) Background information.

In quantum chemistry the two-electron integrals play a central role. Given a set of (typically atom centered) orbitals $\varphi_p(\mathbf{r}), p = 1,2,...,M$ these integrals are defined as

$$V(p,q,r,s) = \int_{all\,space} \varphi_p(\mathbf{r}_1)\varphi_q(\mathbf{r}_2)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\varphi_r(\mathbf{r}_1)\varphi_s(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2$$

If we have M orbitals (M is typically of order 100 … 500 and upwards) we would have M^4 two-electron integrals. There are several different type of notations to denote these two-electron integrals. The above so-called spatial V(p, q, r, s) integrals are expressed in '1212' notation and they satisfy the following permutational symmetry:

$$V(p,q,r,s) = V(r,q,p,s) = V(p,s,r,q) = V(r,s,p,q)$$
$$= V(q,p,s,r) = V(s,p,q,r) = V(q,r,s,p) = V(s,r,q,p)$$

This is easily verified from the definition of the integrals. Let me note here that I have assumed that the orbitals are real, which is the case for our applications. The set of two-electron integrals can be viewed as a matrix (see below), and it is positive definite. Therefore it can be decomposed using Cholesky decomposition. If we define the matrix $A(pq\,|\,rs) = V(p,q,r,s) \rightarrow A(i,j)$, where

$$i = 1...M^2 \leftarrow (p,q)$$
$$j = 1,...M^2 \leftarrow (r,s)$$

where *i* and *j* are compound labels, such that we compress the 4-index array into a two-dimensional square matrix. The matrix A can be subjected to Cholesky decomposition

$$A(i,j) = \sum_K L(i,K)L(j,K)$$

I find it more convenient to write

$A(pq,rs) = \sum_K L(pq,K)L(rs,K)$, such that the original indices are more clearly visible. In principle the sum *K* over the Cholesky vectors has dimension $N = M^2$, and little would be gained by Cholesky decomposition. In practical calculations (see Koch, de Meras and Pedersen, J. Chemical Phys. 118, p 9481-9484, a must read reference), the dimension of the Cholesky vectors is closer to 10 M. The remaining Cholesky vectors have a very small norm and can be neglected, without loss of significant accuracy. Very important aspects of the Cholesky algorithm are

1) To calculate a particular Cholesky vector only a particular column of **A** is needed, together with the already calculated Cholesky vectors. These integrals (matrix elements of A) are calculated as needed. If the Cholesky decomposition has converged, the remaining matrix elements of **A** do not need to be calculated. This presents a huge savings.
2) The accuracy of the Cholesky decomposition is well controlled by a single parameter Δ, and requiring that the "remainder" of the matrix to be decomposed has all of its diagonal elements less than Δ.

3) Integrals are always calculated in batches, for example if one selects four atoms P, Q, R, S one would calculate all integrals $V(p,q,r,s), p \in P, q \in Q, r \in R, s \in S$ together. For this reason we would also calculate a set of Cholesky vectors, corresponding to the batches of integrals. We can discuss later on. Importantly, the order in which Cholesky vectors are calculated does not matter much. Pivoting is needed, but is not particularly critical. This means we can effectively use a batched algorithm. I imagine that also the Cholesky label *K* is processed in batches.

For us, the usefulness of the Cholesky representation for the two electron integrals does not end at the level of calculating the integrals. We want to use the Cholesky decomposed vectors directly in subsequent quantum chemistry calculations. Let me give some examples of potential savings due to the Cholesky representation.

a) Consider the following term that arises for example in Coupled Cluster (CC) calculations (arguably the most accurate method in quantum chemistry):

$$R(p,q,i,j) = \sum_{r,s} V(p,q,r,s) t(r,s,i,j) \quad \forall p,q,i,j$$

where *i, j* represent so-called occupied orbitals, O in number. The dimension O equals the number of electrons in the system, and this is much less than the total number of orbitals *M*. If we would calculate the above term this would require on the order of M^4O^2 operations. Let us now introduce the Cholesky decomposition, and calculate the term in stages:

$$R(p,q,i,j) = \sum_{r,s,K} L(pq,K) L(rs,K) t(r,s,i,j) \quad \forall p,q,i,j$$

$$I(K,i,j) = \sum_{r,s} L(rs,K) t(r,s,i,j) \quad \forall K,i,j, \quad M^2NO^2 \text{ operations}$$

$$R(p,q,i,j) = \sum_{K} L(pq,K) I(K,i,j) \quad \forall pq,i,j, \quad M^2NO^2 \text{ operations}$$

Hence instead of M^4O^2 we require $2M^2NO^2$ operations. Please note that the stepwise calculation of the term is vital. If we assume $N \approx 10M$, $M = 500$, this presents a savings of a factor of 50.

b) In so-called Hartree-Fock calculations, which is often the starting point in quantum chemistry for more advanced calculations like CC, we need to calculate a term

$$f(p,q) += \sum_{r,s} V(p,q,r,s) * D(r,s), \quad M^4 \text{ operations}$$

Using Cholesky decomposition we can proceed as follows

$$f(p,q) += \sum_{r,s,K} L(p,q,K) L(r,s,K) * D(r,s)$$

$$I(K) = \sum_{r,s,K} L(r,s,K) * D(r,s), \quad M^2N \text{ operations}$$

$$\text{f(p,q)} += \sum_{r,s,K} L(p,q,K) I(K), M^2N \text{ operations}$$

again reducing the cost by a factor of about M/20. In order to make this fully effective we will also have to use that the Cholesky vectors are sparse (just like the two-electron integrals themselves). This means that not all elements $L(p,q,K)$ are needed. Many of them are vanishingly small (for a given K) and we wish to neglect the small elements. Also the matrix D is sparse and this can also aid the effectiveness of the calculation. The Hartree-Fock equations contain another term, that illustrate yet another important aspect

$$f(s,q)+=\sum_{r,s} V(p,q,r,s)*D(r,p), \quad M^4 \text{ operations}$$

Now we would use a *different* Cholesky decomposition of the two electron integrals

$$f(s,q)+=\sum_{r,s,K} J(s,q,K)J(r,p,K)*D(r,p)$$

$$I(K)=\sum_{r,s,K} J(r,p,K)*D(r,p), \quad M^2N \text{ operations}$$

$$\text{f(s,q)+=}\sum_{r,s,K} J(s,q,K)I(K), \; M^2N \text{ operations}$$

So in practice we would decompose the two-electron integrals twice to get both the L, and the J-type Cholesky decomposition. There is more symmetry in these equations that can be exploited easily (see below). For this term sparsity is *very* important.

c) A third application of Cholesky decomposition is a little bit harder to explain. That is, the original problem takes time to understand. It concerns the transformation of the two-electron integrals from the original Atomic orbital basis to the Molecular orbital basis. Using the Cholesky decomposition of the two-electron integrals it is much easier to design an efficient parallel algorithm to perform the transformation. This has always been a bottleneck in conventional calculations. The algorithm is as follows

$$I(a,b,K)=\sum_{p,q} C(a,p)C(b,q)L(p,q,K)$$

$$V(a,b,c,d)=\sum_{K} I(a,b,K)I(c,d,K)$$

The parallelization simply consists of a parallelization over the Cholesky index K.

b) Concrete plans for the Project

The purpose of the project is to create efficient algorithms, that work in parallel to create the Cholesky decomposition over the two-electron integrals, in various ways. In a subsequent step we can then address some of the applications. This is less important however, than getting the basics of the algorithm right.

To get started we will use a simple in-core Cholesky decomposition of the two-electron integrals. The purpose is to get some simple programs to do this, that we can use to check more involved implementations. Also it will allow us to gradually build an out-of-core Cholesky program using sparsity and pivoting, and working in parallel eventually.

The first topic of investigation concerns different schemes to decompose the two-electron integrals.

We can use Cholesky decomposition in the following forms:

$$A(pq \mid rs) = V(p,q,r,s), \quad M^2$$

$$A_+(p \le q \mid r \le s) = V(p,q,r,s) + V(q,p,r,s), \ \frac{1}{2}M(M+1)$$

$$A_-(p < q \mid r < s) = V(p,q,r,s) - V(p,q,r,s), \frac{1}{2}M(M-1)$$

$$B(pr \mid qs) = V(p,q,r,s)$$

$$B_+(p \le r \mid q \le s) = V(p,q,r,s)$$

Here the vertical bar indicates the matrix rows and columns. If we have M orbitals, the dimensions of the matrices are as indicated. It can be expected that the length of the Cholesky decomposition varies with the precise definition of the matrices used in the Cholesky decomposition. This would be determined by a threshold that determines the accuracy of the decomposition. This also implies that we should use a pivoting algorithm to do the Cholesky decomposition. This has to be an integral part of the algorithm. Initially we might assume the integrals are simply given as the four-dimensional array. We can fill in the matrix A, as indicated and then perform Cholsky decomposition.

In a subsequent version of the algorithm be should use that the integrals are processed in batches. Hence we get only one batch of integrals into memory (chosen appropriately), and we process all integrals in the batch, updating the desired Cholesky vectors, before proceeding to the next batch. Later on we can calculate the integrals as needed, rather than getting them from disk. Only then will we start saving time, as not all integrals need to be calculated. The final aspect of the algorithm concerns parallelization. It is presumably a good idea to design the parallel algorithm before we write the advanced algorithms, such that we minimize the work. Once we have determined which Cholesky decompositions are useful, we might see if we can calculate the K- and J-type Cholesky vectors together, such that we minimize the recalculation of integrals. This is probably difficult, and is perhaps not so important.

The final implementation (or anything beyond the prototype) would use the Nwchem Quantum chemistry code, and the parallel machinery associated with it.

That is it for now.