Density-Fitting Approximations to the Electron Repulsion Integrals

C. David Sherrill

School of Chemistry and Biochemistry
Georgia Institute of Technology

Created on 19 July 2010

Here we will follow some of the presentation from Werner, Manby, and Knowles.[1] Density fitting is a way to approximate the usual two-electron integrals,

$$(pq|rs) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_1)\frac{1}{r_{12}}\phi_r(\mathbf{r}_2)\phi_s(\mathbf{r}_2), \tag{1}$$

where we have assumed real orbitals. We may also consider this electron repulsion integral as the repulsion between two generalized electron densities,

$$(pq|rs) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \rho_{pq}(\mathbf{r}_1)\frac{1}{r_{12}}\rho_{rs}(\mathbf{r}_2), \tag{2}$$

where $\rho_{pq}(\mathbf{r}) = \phi_p(\mathbf{r})\phi_q(\mathbf{r})$ and $\rho_{rs}(\mathbf{r}) = \phi_r(\mathbf{r})\phi_s(\mathbf{r})$. The densities may be approximated using an auxiliary basis set as

$$\overline{\rho}_{pq}(\mathbf{r}) = \sum_P^{N_{fit}} d_P^{pq}\chi_P(\mathbf{r}), \tag{3}$$

where d_P^{pq} are the fitting coefficients. This is the origin of the name "density fitting." There are various methods for obtaining these fitting coefficients. If one minimizes the following (positive definite) functional with a $1/r_{12}$ weight factor,[2,3,4]

$$\Delta_{pq} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\left[\rho_{pq}(\mathbf{r}_1) - \overline{\rho}_{pq}(\mathbf{r}_1)\right]\left[\rho_{pq}(\mathbf{r}_2) - \overline{\rho}_{pq}(\mathbf{r}_2)\right]}{r_{12}}, \tag{4}$$

then, as shown by Dunlap $et\ al.$,[3,4] this minimizes the error in the electric field and leads to the following fitting coefficients:

$$d_Q^{pq} = \sum_P (pq|P)[\mathbf{J}^{-1}]_{PQ}, \tag{5}$$

where

$$(pq|P) = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \phi_p(\mathbf{r}_1)\phi_q(\mathbf{r}_1)\frac{1}{r_{12}}\chi_P(\mathbf{r}_2), \tag{6}$$

and

$$J_{PQ} = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \chi_P(\mathbf{r}_1)\frac{1}{r_{12}}\chi_Q(\mathbf{r}_2). \tag{7}$$

Thus we have replaced the usual four-index, two-electron integrals $(pq|rs)$, with

$$
\begin{aligned}
(pq|rs) &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \rho_{pq}(\mathbf{r}_1) \frac{1}{r_{12}} \rho_{rs}(\mathbf{r}_2), &\text{(8)}\\
&= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \sum_Q d_Q^{pq} \chi_Q(\mathbf{r}_1) \frac{1}{r_{12}} \phi_r(\mathbf{r}_2)\phi_s(\mathbf{r}_2),\\
&= \sum_Q d_Q^{pq}(Q|rs),\\
&= \sum_{PQ} (pq|P)[\mathbf{J}^{-1}]_{PQ}(Q|rs).
\end{aligned}
$$

Note that the expression above looks superficially like a resolution of the identity (RI). Indeed, such expansions have also been called RI approximations. However, it should be pointed out that this is not a standard RI expression (and hence the name density fitting may be more appropriate).

As mentioned above, there are a number of possible choices for the fitting coefficients d_P^{pq} in equation 3, depending on the functional form which is minimized. Instead of equation 4, which uses the Coulomb operator $g(\mathbf{r}_1,\mathbf{r}_2) = 1/r_{12}$, one may use similar expressions which use an overlap[5,6,7] $g(\mathbf{r}_1,\mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2)$, an anti-Coulomb operator $g(\mathbf{r}_1,\mathbf{r}_2) = -|\mathbf{r}_1 - \mathbf{r}_2|$ (which is optimal for representing the potential caused by ρ_{pq}),[8] or other choices. One disadvantage of the Coulomb metric is that the fitting coefficients d_P^{pq} decay slowly with respect to the distance between $|pq\rangle$ and $|P\rangle$ (and the defining equations can even diverge in the case of periodic boundary conditions); the overlap metric does not have these difficulties, but it is not as accurate.[9] To overcome these difficulties, Jung, Sodt, Gill, and Head-Gordon[9] have investigated the attenuated Coulomb operator, $g(\mathbf{r}_1,\mathbf{r}_2) = erfc(\omega|\mathbf{r}_1-\mathbf{r}_2|)/|\mathbf{r}_1-\mathbf{r}_2|$, which can be tuned between the Coulomb and overlap metrics depending on the size of ω. (Note: these authors also mention that the number of significant quantities $\sum_P (pq|P)[\mathbf{J}^{-1/2}]_{PQ}$ (see below) grows quadratically with system size for any of the three metrics they considered, and thus the sqare-root formulation appears to be unsuitable for linear-scaling algorithms for large systems).

In density fitting, one normally uses atom-centered Gaussians as the auxiliary basis set, and the size of the auxiliary basis in density fitting is usually about 3-4 times the size of the standard basis. One small disadvantage is that different basis set work better depending on the type of density being fit. There are Coulomb fitting basis sets for use in evaluating Coulomb integrals for Hartree-Fock theory or Density Functional theory.[10,11,12] Somewhat larger basis sets are required to fit exchange integrals, but these also work well to fit Coulomb integrals; such "JK-fit" basis sets have been introduced by Weigend.[13] Unfortunately, these JK-fit basis sets are not optimal for Hartree-Fock procedures; basis sets optimized to fit distributions like ρ_{ia} (where i and a are occupied and virtual orbitals, respectively) have been developed. Auxiliary basis sets have been developed for use in conjunction with correlation-consistent basis sets for density fitted MP2 and CC2 computations.[14,15]

A closely related approach is the pseudospectral approximation, which uses a real-space grid instead of atom-centered Gaussians as the auxiliary basis set.[16,17] The Cholesky decomposition,[18,19,20,21,22] when applied to electron repulsion integrals, also expresses four-index integrals as sums of products of three-index integrals, although it avoids the use of the Coulomb metric matrix $\mathbf{J}$.

One big advantage of density fitting techniques (or other techniques expressing four-index integrals in terms of three-index integrals) is that the storage requirements are greatly reduced. It is much more likely that the three-index integrals will fit in memory than it is that the four-index integrals will fit in memory. Even if neither set fits in memory, there will be much less I/O for reading the three-index integrals from disk than there would be for reading the four-index integrals. Of course, a disadvantage of using the three-index integrals is that, unless there is a favorable factorization, the four-index integrals must be constructed from the three-index integrals according to an equation like $(pq|rs) = \sum_{PQ}(pq|P)[\mathbf{J}^{-1}]_{PQ}(Q|rs)$. However, this floating-point intensive procedure is usually faster than the I/O one would have to do to read the four-index integrals. To reduce the cost of constructing the four-index integrals from the three-index integrals, it can be advantageous to use a symmetric expression which splits the matrix $\mathbf{J}^{-1}$ into a product $\mathbf{J}^{-1/2}\mathbf{J}^{-1/2}$:

$$(pq|rs) = \sum_{PQR} (pq|P)[\mathbf{J}^{-1/2}]_{PQ}[\mathbf{J}^{-1/2}]_{QR}(R|rs). \tag{9}$$

This form is advantageous because it factorizes:

$$(pq|rs) = \sum_{Q} \left\{ \sum_{P}(pq|P)[\mathbf{J}^{-1/2}]_{PQ} \right\} \left\{ \sum_{R}[\mathbf{J}^{-1/2}]_{QR}(R|rs) \right\} = \sum_{Q} b_{pq}^{Q} b_{rs}^{Q}, \tag{10}$$

where

$$b_{pq}^{Q} = \sum_{P}(pq|P)[\mathbf{J}^{-1/2}]_{PQ}. \tag{11}$$

As pointed out by Werner $et\ al.$,[1] we can reduce the time required to compute the three-index integrals using the Scharz inequality, i.e.,

$$(\mu\nu|P) \leq (\mu\nu|\mu\nu)^{1/2}(P|P)^{1/2}. \tag{12}$$

For relatively little computational expense, we can pre-compute the "diagonal" elements of the 4-index and 2-index electron repulsion integrals $(\mu\nu|\mu\nu)$ and $(P|P)$, and use these to estimate the size of $(\mu\nu|P)$. If $(\mu\nu|P)$ is predicted to be sufficiently small, then we can avoid computing it explicitly. Typically, integrals are computed by shells (e.g., a d shell includes all 5 or 6 components of the d function, etc.). In that case, we can decide whether or not to compute a "shell triplet" $(\mu_s\nu_s|P_s)$, where μ_s is a shell of functions μ, etc. To do this, we can estimate the largest integral $(\mu\nu|P)$ that would belong to the shell triplet $(\mu_s\nu_s|P_s)$, using the maximum integral $(\mu\nu|\mu\nu)_{max}$

corresponding to the shell pair μ_s, ν_s ($\mu \in \mu_s, \nu \in \nu_s$) and the maximum integral $(P|P)_{\max}$ for $P \in P_s$. If the product $(\mu\nu|\mu\nu)_{\max}^{1/2}(P|P)_{\max}^{1/2}$ is sufficiently small, we can skip computing the entire shell triplet $(\mu_s\nu_s|P_s)$.

DF-MP2

The most time-consuming step of an MP2 computation is actually the $\mathcal{O}(N^5)$ integral transformation (the energy evaluation scales as $\mathcal{O}(N^4)$. The overall integral transformation is

$$(ia|jb) = \sum_{\mu\nu\rho\sigma} C_\mu^i C_\nu^a C_\rho^j C_\sigma^b (\mu\nu|\rho\sigma) \tag{13}$$

where $(ia|jb)$ are the required integrals in the MO basis (note that we only need this particular subset of MO integrals, where i and j are occupied orbitals and where a and b are virtual orbitals), $(\mu\nu|\rho\sigma)$ are the AO-basis integrals, and C_μ^i is the contribution of AO basis function μ to molecular orbital i, etc. If there are N_{occ} occupied orbitals and N_{vir} virtual orbitals, then there are (ignoring permutational symmetry) $N_{occ}^2 N_{vir}^2$ of the required MO integrals. Thus, the transformation above would appear to require on the order of $N_{occ}^2 N_{vir}^2 N_{AO}^4$ floating-point operations. However, due to factorization, the actual cost is less:

$$(i\nu|\rho\sigma) = \sum_\mu C_\mu^i (\mu\nu|\rho\sigma) \quad N_{occ} N_{AO}^4 \tag{14}$$

$$(ia|\rho\sigma) = \sum_\nu C_\nu^a (i\nu|\rho\sigma) \quad N_{occ} N_{vir} N_{AO}^3 \tag{15}$$

$$(ia|j\sigma) = \sum_\rho C_\rho^j (ia|\rho\sigma) \quad N_{occ}^2 N_{vir} N_{AO}^2 \tag{16}$$

$$(ia|jb) = \sum_\sigma C_\sigma^b (ia|j\sigma) \quad N_{occ}^2 N_{vir}^2 N_{AO}. \tag{17}$$

Because N_{AO} is the largest quantity, the first step is the most time-consuming.

Now in the density-fitted approach, we can factorize the work more effectively:

$$b_{i\nu}^Q = \sum_\mu C_\mu^i b_{\mu\nu}^Q \quad N_{occ} N_{AO}^2 N_{aux} \tag{18}$$

$$b_{ia}^Q = \sum_\nu C_\nu^a b_{i\nu}^Q \quad N_{occ} N_{vir} N_{AO} N_{aux} \tag{19}$$

$$(ia|jb) = \sum_Q b_{ia}^Q b_{jb}^Q \quad N_{occ}^2 N_{vir}^2 N_{aux}. \tag{20}$$

4

The most expensive step of the DF transformation is the last one, with a cost of $\mathcal{O}(N_{occ}^2 N_{vir}^2 N_{aux})$, which is much less than $\mathcal{O}(N_{occ} N_{AO}^4)$ from the conventional transformation (even though $N_{aux} \approx 3N_{AO}$).

For a closed-shell molecule, the MP2 correlation energy may be evaluated as

$$E_{corr}^{MP2} = \sum_{ijab} (ia|jb) \frac{2(ia|jb) - (ib|ja)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{21}$$

The error caused by the DF approximation is very small as long as an appropriate DF auxiliary basis is used (and note that, as mentioned above, one typically uses different auxiliary basis sets for different tasks, such as Hartree-Fock or MP2). For van der Waals dimers, DF-MP2 reproduces conventional MP2 binding energies within a few hundredths of one kcal mol^{-1}, while providing speedups of around 1-5x.

References

[1] H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).

[2] J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).

[3] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, Int. J. Quantum Chem. Symp. **11**, 81 (1977).

[4] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979).

[5] E. J. Baerends, D. E. Ellis, and P. Ros, Chem. Phys. **2**, 41 (1973).

[6] O. Vahtras, J. Almlöf, and M. W. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).

[7] C. K. Skylaris, L. Gagliardi, N. C. Handy, A. G. Ioannou, S. Spencer, and A. Willetts, J. Mol. Struct. THEOCHEM **501**, 229 (2000).

[8] P. M. W. Gill, B. G. Johnson, J. A. Pople, and S. W. Taylor, J. Chem. Phys. **96**, 7178 (1992).

[9] Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon, Proc. Natl. Acad. Sci. USA **102**, 6692 (2005).

[10] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).

[11] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs, Theor. Chem. Acc. **97**, 119 (1997).

[12] F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).

[13] F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).

[14] F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).

[15] C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).

[16] T. J. Martinez and E. A. Carter, Pseudospectral methods applied to the electron correlation problem, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony, volume 2 of *Advanced Series in Physical Chemistry*, pages 1132–1165. World Scientific, Singapore, 1995.

[17] R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, and Y. Cao, J. Phys. Chem. A **103**, 1913 (1999).

[18] N. H. F. Beebe and J. Linderberg, Int. J. Quantum Chem. **12**, 683 (1977).

[19] I. Roeggen and E. Wisloff-Nilssen, Chem. Phys. Lett. **132**, 154 (1986).

[20] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).

[21] F. Aquilante, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **126**, 194106 (2007).

[22] F. Aquilante, L. Gagliardi, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **130**, 154107 (2009).