

Attractive Electron–Electron Interactions within Robust Local Fitting Approximations

Patrick Merlot,^[a] Thomas Kjærgaard,^[a] Trygve Helgaker,^[a] Roland Lindh,^[b] Francesco Aquilante,^[b,c] Simen Reine,^{*[a]} and Thomas Bondo Pedersen^{*[a]}

An analysis of Dunlap's robust fitting approach reveals that the resulting two-electron integral matrix is not manifestly positive semidefinite when local fitting domains or non-Coulomb fitting metrics are used. We present a highly local approximate method for evaluating four-center two-electron integrals based on the resolution-of-the-identity (RI) approximation and apply it to the construction of the Coulomb and exchange contributions to the Fock matrix. In this pair-atomic resolution-of-the-identity (PARI) approach, atomic-orbital (AO) products are expanded in auxiliary functions centered on the two atoms associated with each product. Numerical tests indicate that in 1% or less of all Hartree–Fock and Kohn–Sham calculations, the indefinite integral matrix causes nonconvergence in the self-consistent-field iterations. In these cases, the two-electron contribution to the total energy becomes negative, meaning that the electronic interaction is effectively attractive, and the total energy is dramatically lower

than that obtained with exact integrals. In the vast majority of our test cases, however, the indefiniteness does not interfere with convergence. The total energy accuracy is comparable to that of the standard Coulomb-metric RI method. The speed-up compared with conventional algorithms is similar to the RI method for Coulomb contributions; exchange contributions are accelerated by a factor of up to eight with a triple-zeta quality basis set. A positive semidefinite integral matrix is recovered within PARI by introducing local auxiliary basis functions spanning the full AO product space, as may be achieved by using Cholesky-decomposition techniques. Local completion, however, slows down the algorithm to a level comparable with or below conventional calculations. © 2013 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23284

Introduction

Density-fitting or resolution-of-the-identity (RI) techniques, the successors of early approximation schemes for evaluating two-electron integrals in atomic-orbital (AO) basis,^[1–3] were pioneered by Whitten,^[4] Baerends et al.,^[5] Sambe and Felton,^[6] and Dunlap et al.^[7,8] in the 1970s. The work in the 1990s by Feyereisen et al.^[9,10] and by Ahlrichs and coworkers^[11–14] was instrumental in turning the RI approach into the mainstream approximation it is today. For theoretical models, where two-electron integral evaluation and transformation constitute a major computational bottleneck—including, in particular, non-hybrid Kohn–Sham (KS) density-functional theory (DFT),^[15,16] second-order Møller–Plesset (MP2) theory,^[17] and second-order coupled-cluster^[18] theory—application of RI techniques results in computational acceleration by as much as one to two orders of magnitude for small- and medium-sized (on the order of 100 atoms) molecules, see Weigend et al.,^[19] Weigend,^[20] and Hättig and Weigend.^[21] Recently, RI methods have also gained popularity for a range of self-energy approximations using either Gaussian-type orbitals (GTOs)^[22,23] or numerical AOs.^[24] For most molecular electronic-structure models, RI techniques reduce the prefactor rather than the computational scaling with respect to molecule size. Notable exceptions are the Laplace-transform scaled-opposite-spin MP2 theory^[25] and the direct random-phase approximation (RPA),^[26] where the scaling is reduced from $O(N^5)$ to $O(N^4)$, with N a measure of molecule size, for each grid point in the numerical integration.

The RI approximation may also be used in conjunction with reduced-scaling methods for large molecules, including reduced-scaling exchange algorithms,^[27–30] local MP2 theory,^[31,32] and local coupled-cluster theory.^[33] In its most accurate form, the RI method scales cubically with system size and further local approximations are required (in the RI algorithm itself) to reduce scaling. Several local RI approaches have been suggested, as discussed in the Theory section. These approaches share a common feature: the accuracy of the two-electron integrals is reduced. Fortunately, much of the accuracy can be recovered by using Dunlap's robust formulation.^[7,34–37] A fully robust-variational RI approach was used by Reine et al.^[30] for Coulomb and exchange contributions in

[a] P. Merlot, T. Kjærgaard, T. Helgaker, S. Reine, T. B. Pedersen
Centre for Theoretical and Computational Chemistry, Department of
Chemistry, University of Oslo, P.O. Box 1033, Blindern, N-0315 Oslo, Norway
E-mail: simen.reine@kjemi.uio.no and t.b.pedersen@kjemi.uio.no

[b] R. Lindh, F. Aquilante
Department of Chemistry–Ångström, Theoretical Chemistry Programme,
Uppsala University, P.O. Box 518, S-75120 Uppsala, Sweden

[c] F. Aquilante
Center for Biomolecular Nanotechnologies @UNILE, Italian Institute of
Technology (IIT), Via Barsanti, I-73010 Arnesano (LE), Italy

Contract/grant sponsor: CoE Centre for Theoretical and Computational
Chemistry; contract/grant number: 179568/V30; contract/grant sponsor:
The Danish Council for Independent Research|Natural Sciences;
Contract/grant sponsor: Swedish Research Council (VR).

© 2013 Wiley Periodicals, Inc.

Hartree–Fock (HF) theory and hybrid KS DFT, whereas a partially robust correction was used by Sodt and Head-Gordon.^[29] Although the robust approximation^[35] provides superior accuracy, we demonstrate theoretically here that it exhibits one serious flaw: it is not manifestly positive semidefinite. As a consequence, electronic interactions may become effectively attractive. Numerical tests using a highly local pair-atomic resolution-of-the-identity (PARI) approach show that the indefinite approximate integral matrix may cause severe convergence problems or even convergence to a spurious state with a negative two-electron energy with a total electronic energy far below the exact ground-state energy.

Theory

RI methods approximate the four-center two-electron integrals $(ab|cd)$, in Mulliken notation,

$$(f|g) = \int_{\mathcal{R}^6} f(\mathbf{r}_1) \frac{1}{r_{12}} g(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \quad r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|, \quad (1)$$

by expanding each AO pair product $|ab\rangle$ between two AO basis functions $\chi_a(\mathbf{r})$ and $\chi_b(\mathbf{r})$ in atom-centered auxiliary basis functions $\Phi_\alpha(\mathbf{r})$:

$$|ab\rangle \approx |\widetilde{ab}\rangle = \sum_\alpha |\alpha\rangle c_\alpha^{ab} \quad (2)$$

where tilde denotes a fitted quantity. In contrast to the analogous expansion of molecular orbitals (MOs) in a basis of AOs, we have the advantage of knowing the target functions $|ab\rangle$. Therefore, the fitting basis $\{|\alpha\rangle\}$ does not need to be complete in the sense of spanning the full space of square-integrable functions on R^3 for the fit to be exact. Clearly, it is sufficient that the AO products and the fitting basis span the same subspace, $\text{span}\{|ab\rangle\} = \text{span}\{|\alpha\rangle\}$.

Robust fitting

Rather than simply replacing all AO products $|ab\rangle$ with their fitted approximations $|\widetilde{ab}\rangle$, Dunlap^[35] has advocated the robust expression

$$(ab|cd) \approx (\widetilde{ab}|\widetilde{cd}) = (\widetilde{ab}|cd) + (ab|\widetilde{cd}) - (\widetilde{ab}|\widetilde{cd}), \quad (3)$$

which guarantees that the integral error is bilinear in the error of the fitted AO products,

$$\Delta G_{ab,cd} = (ab|cd) - (\widetilde{ab}|\widetilde{cd}) = (ab - \widetilde{ab}|cd - \widetilde{cd}). \quad (4)$$

Since $\Delta \mathbf{G}$ is symmetric positive semidefinite, the Coulomb (Hartree) energy computed using the approximate integrals is a lower bound to the result computed using the exact integrals.^[7]

$$E_{\text{H}} - \widetilde{E}_{\text{H}} = \frac{1}{2} \sum_{abcd} D_{ab} \Delta G_{ab,cd} D_{cd} \geq 0. \quad (5)$$

Similar inequalities hold for energy contributions that can be written as the trace of a (positive or negative) semidefinite

matrix multiplied by the integral matrix. Important examples are the exact-exchange energy and the direct part of the coupled-cluster correlation energy, which includes the direct-MP2 and direct-RPA expressions as special cases. (The term “direct” indicates that only Coulomb contributions are treated at the correlated level, whereas the exchange contributions are left at the uncorrelated HF level.) Inequalities of the type in eq. (5) have been exploited to optimize auxiliary basis sets for the calculation of Coulomb energies,^[11,12,20] exchange energies,^[38,39] and correlation energies.^[14,19,40–42]

The above properties of the robust integral expression indicate that HF and KS energies computed using the RI approximation are less than those obtained with exact integrals. This is not fully guaranteed in practice, of course, as the density matrix obtained from self-consistent field (SCF) iterations in an RI-based calculation generally differs from that obtained in the corresponding SCF calculation with the exact integrals.

Positive definiteness

The appealing properties of the robust integral expression eq. (3) are valid regardless of how the coefficients c_α^{ab} of eq. (2) are determined. However, these schemes come with an important caveat: *the robust integral matrix $\widetilde{G}_{ab,cd} = (\widetilde{ab}|\widetilde{cd})$ is not manifestly positive semidefinite.* Negative eigenvalues imply that the total two-electron interaction may become effectively attractive with an unfortunate but formally valid density matrix during SCF iterations. In the worst case, such eigenvalues may lead to convergence to a spurious state corresponding to clustering of electrons around the most highly charged nuclei and a total energy far below (by thousands or even millions of hartrees) the exact electronic ground-state energy. Although strictly speaking in numerical accord with eq. (5) (since the approximate energy is below the exact one), such a result is not covered by that inequality as the converged density matrices obtained with exact and approximate integrals are fundamentally different. Less dramatic but more insidious consequences of an indefinite integral matrix in SCF calculations are convergence difficulties and nonconvergence. We note that small negative integral eigenvalues should not cause severe problems in noniterative schemes—in particular, if an inequality of the type in eq. (5) is rigorously valid.

To show that the robust integral expression in eq. (3) is not manifestly positive semidefinite, we multiply from left and right by an arbitrary real vector \mathbf{x} , yielding

$$\sum_{abcd} x_{ab} (\widetilde{ab}|\widetilde{cd}) x_{cd} = (\widetilde{X}|\widetilde{X}) + (X|\widetilde{X}) - (\widetilde{X}|\widetilde{X}), \quad (6)$$

$$|X\rangle = \sum_{ab} x_{ab} |ab\rangle, \quad |\widetilde{X}\rangle = \sum_{ab} x_{ab} |\widetilde{ab}\rangle \quad (7)$$

While the first two terms in eq. (6) are not necessarily nonnegative, the last term is nonnegative, $(\widetilde{X}|\widetilde{X}) \geq 0$, but it is subtracted. Overall, therefore, the robust integral matrix is only conditionally positive semidefinite.

Let us see how negative eigenvalues may arise in the robust integral matrix. Using eq. (4) and introducing $|\Delta X\rangle = |X - \tilde{X}\rangle$, we may recast eq. (6) in the form

$$(\tilde{X}|X) + (X|\tilde{X}) - (\tilde{X}|\tilde{X}) = (X|X) - (\Delta X|\Delta X). \quad (8)$$

If \mathbf{x} is an eigenvector of the exact integral matrix, this expression indicates that the eigenvalues of the exact integral matrix are shifted downwards in the robust RI approximation by roughly $(\Delta X|\Delta X)$. Although $(\Delta X|\Delta X)$ should be small in a good auxiliary basis, nonzero vectors \mathbf{x} generally do exist such that $(X|X) < (\Delta X|\Delta X)$, meaning that the robust integral matrix has at least one negative eigenvalue. Suppose, for example, that the AO product basis is linearly dependent in the Coulomb metric and let \mathbf{x} be a nonzero eigenvector of the exact integral matrix with a zero eigenvalue. Then $(X|X) = 0$ and, unless the fit is exact, $(\Delta X|\Delta X) > 0$, illustrating how negative eigenvalues of the robust integral matrix may arise.

Coulomb-metric fitting

Fortunately, the fitting coefficients can be determined in such a way that the robust integral matrix becomes manifestly positive semidefinite. With a given fitting basis, the most accurate results are obtained by minimizing the integral errors. Noting that the integral error matrix satisfies the Cauchy–Schwarz inequality,

$$|\Delta G_{ab,cd}| \leq \sqrt{\Delta G_{ab,ab} \Delta G_{cd,cd}}, \quad (9)$$

the smallest errors are obtained by requiring that the fitting coefficients minimize $\Delta G_{ab,ab}$ for each AO product ab . This approach may be viewed as a least-squares minimization of the norm of the fitting error $|ab - \tilde{ab}\rangle$ measured in the Coulomb metric, leading to the following linear equations for the fitting coefficients:

$$\langle \alpha | ab \rangle = \langle \alpha | \tilde{ab} \rangle = \sum_{\beta} \langle \alpha | \beta \rangle c_{\beta}^{ab}, \quad \forall \alpha, ab. \quad (10)$$

Apart from being robust, this fitting procedure gives a variational energy in, for example, HF and KS theories.^[35] Moreover, since $\langle ab | cd \rangle = \langle \tilde{ab} | \tilde{cd} \rangle = \langle ab | cd \rangle$ by virtue of eq. (10), the robust integral expression in eq. (3) can be reduced to

$$\langle \tilde{ab} | \tilde{cd} \rangle = \langle ab | cd \rangle. \quad (11)$$

Thus, the robust integral matrix becomes manifestly positive semidefinite when the fitting coefficients are determined by the Coulomb-metric fitting scheme of eq. (10).

An alternative approximate integral expression is given by

$$\langle \tilde{ab} | \tilde{cd} \rangle = \frac{1}{2} \langle \tilde{ab} | cd \rangle + \frac{1}{2} \langle ab | \tilde{cd} \rangle. \quad (12)$$

This expression and those in eqs. (3) and (11) are equivalent when the fitting coefficients are determined by eq. (10). On the other hand, with any other fitting scheme, these approxi-

mations differ and only eq. (3) is robust. We shall refer to the expressions in eqs. (11) and (12) as the nonrobust two-center (NR-2) and the nonrobust three-center (NR-3) approximations, respectively (bearing in mind that they become robust with Coulomb-metric fitting).

Local fitting

While it is well established that Coulomb-metric fitting provides superior accuracy in total energies compared with alternative schemes,^[7,10] the solution of the fitting equations in eq. (10) scales cubically with system size and represents a bottleneck for large systems. Although subquadratic scaling may be achieved by the preconditioned conjugate-gradient method,^[43] a linear-scaling scheme is needed. The fundamental problem is the long range of the Coulomb operator, which induces a slow decay of the coefficients c_{α}^{ab} with increasing distance between the AOs a and b and the fitting function α .^[44,45] Although the slow decay may be circumvented by using a locally complete auxiliary basis,^[46,47] the analysis points directly to two approaches for asymptotic linear scaling. One is to deploy a local metric in place of the Coulomb metric for computing the fitting coefficients;^[30,44] the other is to assign a local fitting domain (a subset of the auxiliary basis) in the vicinity of each target AO product $|ab\rangle$.^[27,29,48–50] These approaches may be combined, using both a local metric and local fitting domains, as done by Baerends et al.^[5] and Kryukov et al.^[51]

Whether a local metric or local domains are used, highest accuracy is obtained with the robust integral expression in eq. (3) which, however, is not positive semidefinite. The NR-3 expression is less accurate than the robust expression and also indefinite. Using local domains, Sodt and Head-Gordon^[29] applied the NR-3 scheme to the exchange contribution, reporting energy errors comparable with those obtained using Coulomb-metric fitting. These authors do not discuss the lack of positive semidefiniteness but, with a small domain radius of 4 bohr, the error for a graphite sheet with 50 carbon atoms and 18 hydrogen atoms is about $-5 mE_h$ per atom in the SV(P) orbital basis and TZPP auxiliary basis (optimized for exchange), which is of opposite sign and 2–3 orders of magnitude greater than that obtained with a domain radius above 6 bohr (see Table I of Sodt and Head-Gordon^[29]). Most likely, the large oppositely signed error arises from negative eigenvalues of the integral matrix.

To our knowledge, indefiniteness has not been discussed in the literature, perhaps because the fully robust integral approximation is not widely used with local fitting. Another reason may be that indefiniteness does not present serious difficulties in practice and thus has remained unnoticed. In the following, we introduce the highly local PARI method where, presumably, problems associated with indefiniteness should be the most severe.

The PARI method

If local AO basis functions such as GTOs or Slater-type orbitals are used, the AO pair products $|ab\rangle$ are local in nature. The

Table 1. Comparison of exact and approximate integral matrices for H₂ ($R = 1.4a_0$) and C₂H₄ (planar, $R_{CC} = 2.52032a_0$, $R_{CH} = 2.06844a_0$) with the 6-31G/df-def2 basis sets. For each integral matrix, the largest integral error $\|\Delta\mathbf{G}\|_{\max}$ (in μE_h), the lowest integral eigenvalue γ_{\min} (in μE_h), the number of negative eigenvalues N_{neg} , and the energy error (in μE_h) at convergence $\Delta E = E_{\text{approximate}} - E_{\text{exact}}$ are listed. The addition of two-center functions to the df-def2 auxiliary basis set is denoted '+2C(δ)', where δ is the integral target accuracy (in E_h).

Integrals	H ₂				C ₂ H ₄			
	$\ \Delta\mathbf{G}\ _{\max}$	γ_{\min}	$N_{\text{neg}}^{[a]}$	ΔE	$\ \Delta\mathbf{G}\ _{\max}$	γ_{\min}	$N_{\text{neg}}^{[b]}$	ΔE
exact	0.0	43.7	0	0.0	0.0	0.0	0	0.0
RI	2689.1	15.4	0	-31.2	1546.0	0.0	0	-41.3
PARI	8384.3	-957.6	2	-51.6	8384.3	-1469.8	117	-219.6
PARI + 2C(10^{-4})	50.2	13.8	0	-17.8	89.6	-162.0	115	-63.2
PARI + 2C(10^{-6})	0.0	43.7	0	0.0	1.0	-0.8	77	-0.8
PARI + 2C(10^{-8}) ^[c]	0.0	43.7	0	0.0	0.0	-0.0	20	0.0
PARI + 2C(10^{-10})	0.0	43.7	0	0.0	0.0	0.0	0	0.0
NR-2 ^[d]	18628.0	5.5	0	686.7	22301.8	0.0	0	-83440.0
NR-3 ^[e]	9073.2	-271.1	2	324.1	11182.9	-435.6	100	-28464.8
dPtOE ^[f]	481.6	58.3	0		388.0	-209.3	31	

[a] Out of a total of 10 eigenvalues. [b] Out of a total of 351 eigenvalues. [c] For C₂H₄ the values are: $\|\Delta\mathbf{G}\|_{\max} = 9.0 \times 10^{-3} \mu E_h$, $\gamma_{\min} = -5.3 \times 10^{-3} \mu E_h$, and $\Delta E = -2.3 \times 10^{-3} \mu E_h$. [d] Equation (11). [e] Equation (12). [f] Diagonal blocks of the PARI integral matrix replaced by exact integrals. The energy has not been computed for dPtOE.

idea of using local fitting for approximating these products is an old one, predating the RI approximation—it was used, for instance, in the axial expansion of Boys and Shavitt,^[1] with auxiliary basis functions positioned on the line connecting the two parent atoms. The use of auxiliary basis functions centered exclusively on the parent atoms has also been exploited by Baerends et al.^[5] in nonhybrid KS DFT. In their approach, the electron density $\rho(\mathbf{r})$ is decomposed into pair-atomic densities $\rho^{AB}(\mathbf{r})$ in the manner,

$$\rho(\mathbf{r}) = \sum_{AB} \rho^{AB}(\mathbf{r}), \quad (13)$$

where each pair-atomic density is expanded in auxiliary functions centered only on the parent atoms *A* and *B*. Essentially, the same idea was used more recently by Giese and York.^[52] In the PARI scheme, we exploit locality in a manner similar to that of Baerends et al. except that we approximate the electron integrals $\langle ab|cd \rangle$ individually for each $|cd \rangle$ rather than its total Coulomb interaction $\langle ab|\rho \rangle$. Also, we apply the robust Dunlap correction.

In the PARI method, we approximate each four-center two-electron integral $\langle ab|cd \rangle$ using the robust integral expression in eq. (3), $\langle ab|cd \rangle^{\text{PARI}} = \langle ab|cd \rangle + \langle ab|cd \rangle - \langle ab|\tilde{c}\tilde{d} \rangle$ with $|\tilde{ab} \rangle$ and $|\tilde{cd} \rangle$ expanded in atom-centered auxiliary basis functions $|\alpha \rangle$ and $|\gamma \rangle$ centered only on the parent atoms *A* and *B* for $|\tilde{ab} \rangle$ and *C* and *D* for $|\tilde{cd} \rangle$, respectively:

$$|\tilde{ab} \rangle = \sum_{\alpha \in A \cup B} c_{\alpha}^{ab} |\alpha \rangle, \quad |\tilde{cd} \rangle = \sum_{\gamma \in C \cup D} c_{\gamma}^{cd} |\gamma \rangle. \quad (14)$$

The PARI fitting coefficients are determined from the equations

$$\sum_{\beta \in A \cup B} \langle \alpha|\beta \rangle c_{\beta}^{ab} = \langle \alpha|ab \rangle, \quad \forall \alpha \in A \cup B, \quad (15)$$

thereby minimizing the integral errors $\Delta G_{ab,ab}^{\text{PARI}} = \langle ab - \tilde{ab}|ab - \tilde{ab} \rangle$. Involving only local quantities, the number

of equations is equal to the number of nonzero AO products which, in turn, scales linearly with system size. The PARI fitting problem thus scales linearly. More generally, RI and PARI expansion coefficients can be obtained by minimizing the residual interaction in different metrics^[30,44]—for example, in the Coulomb metric used above or in the overlap metric. The minimization may furthermore be constrained to conserve charge^[5,7] and possibly higher order electric multipole moments.^[52]

The small number of auxiliary functions used for each AO pair in the PARI scheme compared with the global Coulomb-metric scheme makes it reasonable to expect relatively large integral errors. Moreover, negative eigenvalues of the PARI integral matrix are most likely larger (numerically) than those obtained with local fitting domains. We thus anticipate that large auxiliary basis sets are needed in PARI calculations. Therefore, before describing the PARI calculation of Coulomb and exchange contributions, we discuss how Cholesky-decomposition (CD) techniques may be invoked to control the accuracy of PARI integrals while maintaining locality.

Unless qualified further, the term "PARI" is in the following taken to mean an unconstrained fitting in the Coulomb metric according to eq. (15), whereas the term "RI" implies an unconstrained global Coulomb-metric fitting.

Control of integral accuracy by Cholesky decomposition

The CD of two-electron repulsion integrals offers a viable alternative path to the construction of auxiliary basis sets, where the constituent functions are constructed with the objective of minimizing integral errors. Following its introduction in quantum chemistry by Beebe and Linderberg in 1977^[53] and subsequent investigations in the 1980s,^[54–56] the CD method was revived by Koch et al. in 2003^[57]. CD-based approaches are reviewed in the recent articles by Pedersen et al.^[46] and Aquilante et al.,^[47] to which we refer for details. In the context of PARI, the most interesting aspect of CD-based auxiliary basis sets is their inherent locality.^[46,47,58]

The minimized values of the diagonal integral errors $\Delta G_{ab,ab}^{\text{PARI}}$ depend on the auxiliary basis. If the auxiliary functions on A and B span the same space as the product set $\{|ab\rangle\}$ for all AB , then $\Delta G_{ab,ab}^{\text{PARI}} = 0$ and the PARI scheme becomes exact. In practice, for $A \neq B$, it is difficult to span $|ab\rangle$ accurately with only atom-centered functions since, in general, a large, diffuse auxiliary basis is then needed. For diagonal AA pairs, atomic CD auxiliary functions span the AO product space with only a limited number of atom-centered functions.^[58] We note that the use of atomic CD auxiliary basis sets in conjunction with the PARI approximation is equivalent to the limited-expansion-of-diatomic-overlap (LEDO) method of Billingsley and Bloor,^[3] with on-site linear dependence removed in advance (by the atomic CD).

A simple strategy for spanning AO products locally is to include functions positioned on the line between the two atoms in question.^[1] In the same spirit, CD may be deployed to select exactly those AO products that are needed for each atom pair.^[46] Suppose that we aim for an absolute error of at most $\delta \geq 0$ in every integral and suppose that a given atom-centered auxiliary basis has one or more pairs for which $\Delta G_{ab,ab}^{\text{PARI}} > \delta$. For each atom pair, we Cholesky decompose the corresponding diagonal block of the integral-error matrix with threshold δ , thereby obtaining the linearly independent subset of the AO products $|ab\rangle$ that is not spanned by the atom-centered auxiliary basis. These two-center functions are then added to the atom-centered auxiliary basis and the AO products are re-expanded as

$$|\widetilde{ab}\rangle = \sum_{\alpha \in A \cup B + AB} c_{\alpha}^{ab} |\alpha\rangle, \quad (16)$$

where AB denotes the added set of two-center functions, compared to eq. (14). Upon solving the (unconstrained) fitting equations, all elements of the PARI integral-error matrix satisfy the inequality $|\Delta G_{ab,cd}^{\text{PARI}}| \leq \delta$, implying that errors in total energies and molecular properties can be controlled through the integral target accuracy δ . Moreover, the PARI integral matrix is expected to approach positive semidefiniteness smoothly as δ is decreased.

Approaching local completeness by adding two-center functions carries a computational penalty and makes the auxiliary basis geometry dependent. As a result, discontinuities are introduced into the potential-energy surface, making geometry optimizations difficult to perform. These discontinuities decrease with the target accuracy, eventually becoming unimportant. Calculations by Aquilante et al.^[59] for full CD of the exact integral matrix, which also involve two-center functions, indicate that rather loose thresholds may be sufficient.

We finally mention that the PARI integral approximation becomes essentially identical to Foerster's^[60] elimination of redundant orbital products (in the Coulomb metric) if only two-center functions were included in the auxiliary basis set.

The PARI-J and PARI-K methods

An existing RI code is straightforwardly adapted to the PARI scheme for evaluating the Coulomb contribution to the Fock/

KS matrix (referred to here as the PARI-J method) since similar contributions appear in the RI and PARI schemes. For $J_{ab} = (ab|\rho) = \sum_{cd} (ab|cd)D_{cd}$, we get the PARI-J approximation by replacing all occurrences of $(ab|cd)$ by $(ab|cd)^{\text{PARI}}$:

$$J_{ab}^{\text{PARI}} = (ab|\tilde{\rho}) + (\widetilde{ab}|\rho - \tilde{\rho}) = \sum_{\gamma} (ab|\gamma)c_{\gamma} + \sum_{\alpha \in A \cup B} c_{\alpha}^{ab} \{g_{\alpha} - \tilde{g}_{\alpha}\}, \quad (17)$$

$$g_{\alpha} = \sum_{cd} (\alpha|cd)D_{cd}, \quad \tilde{g}_{\alpha} = \sum_{\gamma} (\alpha|\gamma)c_{\gamma}, \quad c_{\gamma} = \sum_{cd} c_{\gamma}^{cd} D_{cd}. \quad (18)$$

The contributions $(ab|\tilde{\rho})$ and g_{α} are calculated in both the RI and PARI methods but used with different coefficients c_{γ} . The added complexity of the PARI approximation is the contribution involving \tilde{g}_{α} ; however, these integrals are also needed for the RI fitting equations. We, therefore, expect similar efficiencies of the RI-J and PARI-J methods, although the latter will be faster for large systems because of the linear complexity of the PARI fitting equations.

In the PARI-K scheme, we apply the PARI method to the exchange contribution to the Fock/KS matrix, replacing $K_{ab} = \sum_{cd} (ac|bd)D_{cd}$ by the fitted elements

$$K_{ab}^{\text{PARI}} = \sum_{cd} (ac|bd)^{\text{PARI}} D_{cd}. \quad (19)$$

Compared with regular RI schemes, the locality of PARI-K fits may greatly improve efficiency by reducing the time spent on transformations and contractions. To minimize cost, we calculate the three-center integrals only once per SCF cycle (exploiting permutational symmetry), aiming to perform all contractions in an efficient manner, while keeping memory and disk-storage requirements to a minimum. To calculate three-center integrals only once per SCF iteration, we split the pair-product expansions in two parts:

$$|\widetilde{ab}\rangle = |\underline{ab}\rangle + |\overline{ab}\rangle \equiv \sum_{\alpha \in A} c_{\alpha}^{ab} |\alpha\rangle + \sum_{\beta \in B \neq A} c_{\beta}^{ab} |\beta\rangle, \quad (20)$$

where underline indicates that the functions are restricted to that atom only. Exploiting the full permutation symmetry of the integrals, the PARI-K algorithm for symmetric and antisymmetric matrices is outlined in Figure 1. Nonsymmetric matrices may be treated by decomposing these into their symmetric and antisymmetric parts.

Results

Most of the results reported here have been obtained with a development version of LSDALTON^[61] compiled with Intel ifort version 11.1 and linked with the MKL library version 11.1. The calculations with two-center functions in the auxiliary basis were performed with a development version of MOLCAS^[62] compiled with gfortran version 4.1.2. All calculations were carried out on a single core of a 2.66 GHz Intel Xeon X5355 Quad Core processor.

Calculate $\langle\alpha|\beta\rangle$
 Calculate fitting coefficients c_α^{ab}
 Loop A
 Construct density-contracted coefficients $d_{\alpha,d}^a = \sum_b c_\alpha^{ab} D_{bd}$
Three-center contributions
 Loop C
 Loop $D \geq C$
 Calculate $\langle\alpha|cd\rangle$
 $K_{ac}^{\text{PARI}} \pm \sum_{\alpha,d} d_{\alpha,d}^a \langle\alpha|cd\rangle$
 $X_{\alpha,c}^a \pm \sum_d \langle\alpha|cd\rangle D_{ad}$
 If $C \neq D$
 $K_{ad}^{\text{PARI}} \pm \sum_{\alpha,c} d_{\alpha,c}^a \langle\alpha|cd\rangle$
 $X_{\alpha,d}^a \pm \sum_c \langle\alpha|cd\rangle D_{ac}$
 End if
 End loop D
 Loop $B \neq A$
 $K_{bc}^{\text{PARI}} \pm \sum_{\alpha,a} c_\alpha^{ab} X_{\alpha,c}^a$
 End loop B
 End loop C
 Loop $B \neq A$
 $K_{bd}^{\text{PARI}} \pm \sum_{\alpha,a} c_\alpha^{ab} X_{\alpha,d}^a$
 End loop B
Two-center contributions
 Loop $C \leq A$
 Get $\langle\alpha|\gamma\rangle$
 Contract $\langle\alpha|\tilde{c}d\rangle = (1 + \delta_{AC})^{-1} \sum_\gamma \langle\alpha|\gamma\rangle c_\gamma^{cd}$
 $K_{ac}^{\text{PARI}} \pm \sum_{\alpha,d} d_{\alpha,d}^a \langle\alpha|\tilde{c}d\rangle$
 $K_{ad}^{\text{PARI}} \pm \sum_{\alpha,c} d_{\alpha,c}^a \langle\alpha|\tilde{c}d\rangle$
 $\tilde{X}_{\alpha,c}^a \pm \sum_d \langle\alpha|\tilde{c}d\rangle D_{ad}$
 $\tilde{X}_{\alpha,d}^a \pm \sum_c \langle\alpha|\tilde{c}d\rangle D_{ac}$
 Loop $B \neq A$
 $K_{bc}^{\text{PARI}} \pm \sum_{\alpha,a} c_\alpha^{ab} \tilde{X}_{\alpha,c}^a$
 Loop B
 End loop C
 Loop $B \neq A$
 $K_{bd}^{\text{PARI}} \pm \sum_{\alpha,a} c_\alpha^{ab} \tilde{X}_{\alpha,d}^a$
 Loop B
 End loop A
 Symmetrize or anti-symmetrize \mathbf{K}^{PARI}

Figure 1. Outline of the PARI-K algorithm.

Positive semidefiniteness

To illustrate the problems arising from an indefinite integral matrix, we used MOLCAS to calculate the ground state of ethylene (planar, $R_{CC} = 2.52032a_0$, $R_{CH} = 2.06844a_0$). The conventional calculation converges in 10 iterations to an energy of $-78.575129E_h$ at the BLYP/cc-pVTZ level of theory. A change of the sign of all two-electron integrals, thereby obtaining a negative semidefinite integral matrix, leads to convergence in nine iterations to an energy of $-297.455623E_h$ with essentially all electrons on the carbon atoms, as expected. A shift of all diagonal integrals ($ab|ab$) by either $-0.1E_h$ or $-0.01E_h$ results in an indefinite integral matrix, with convergence in 389 iterations to $-1208098.3E_h$ and no convergence in 400 iterations, respectively. This simple example illustrates the problems associated with an indefinite two-electron integral matrix: convergence to a spurious state or failure to converge.

The testing of a symmetric real matrix for positive semidefiniteness can be done by CD, which fails if the matrix has negative eigenvalues, or by brute-force diagonalization. We have chosen the latter approach, with the entire integral matrix stored in main memory, thereby limiting the size of molecules and basis sets that can be tested for positive semidefiniteness.

The results for H_2 and C_2H_4 are reported in Table 1, where we compare exact and approximate integral matrices at the BLYP level of theory with the 6-31G orbital basis^[63] and df-def2 auxiliary basis set.^[12] Even for these small systems, the PARI integral matrix has 2 negative eigenvalues (out of 10) for H_2 and 117 (out of 351) for C_2H_4 . The lowest PARI eigenvalue is about $-1mE_h$ for both systems, even though the largest PARI integral error is only 3–4 times greater than the largest RI error. In spite of negative eigenvalues, the PARI calculations converge to total energies with errors only 2–5 times greater than the corresponding RI errors.

Upon the addition of two-center functions to the auxiliary basis, integral and energy errors are reduced and eventually disappear. Concurrently, the number and magnitude of negative eigenvalues are reduced as the integral matrix approaches positive semidefiniteness. However, very accurate integrals (with target accuracy $10^{-10}E_h$) are needed to make the PARI integral matrix positive semidefinite for C_2H_4 . The NR-2 integral matrix is (by construction) positive semidefinite but with large integral and energy errors. In the NR-3 scheme, even though the largest integral and energy errors are increased relative to the PARI scheme, the magnitude of the most negative eigenvalue is reduced. Since the largest PARI errors are in the diagonal integrals, we have tested whether indefiniteness can be cured by substituting exact integrals in the diagonal blocks of the PARI integral matrix. This approach is denoted dPtoE in Table 1. Evidently, as this diagonal substitution works only for H_2 , the dPtoE approach is not a general cure but we note that the number and magnitude of negative eigenvalues are significantly decreased relative to the PARI scheme for C_2H_4 .

Importantly, the addition of two-center functions as needed to achieve a given target accuracy does not alter the complexity of PARI fitting, which remains asymptotically linear—see Figure 2, where the CPU time required to set up and solve the

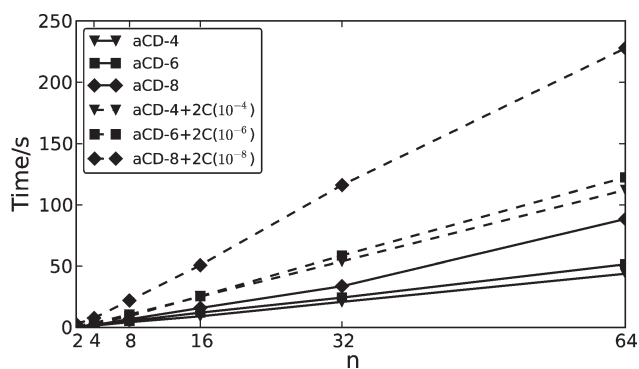


Figure 2. CPU time required to set up and solve the PARI fitting equations with and without two-center functions added to the auxiliary basis set. Sample calculations performed for linear carbon chains C_nH_2 with the cc-pVTZ basis set and aCD-type auxiliary basis sets on a single core of a 2.66 GHz Intel Xeon X5355 Quad Core processor.

fitting equations with and without two-center functions is plotted against system size. As a model system to demonstrate scaling, we have chosen linear chains of carbon atoms with a single hydrogen atom at each end—that is, C_nH_2 with a C–C distance of 1.40 Å and a C–H distance of 1.09 Å. Here, we use

the cc-pVTZ basis set^[64] and atom-centered auxiliary basis sets denoted aCD-4, aCD-6, and aCD-8 generated by atomic CD^[65] with decomposition thresholds 10^{-4} , 10^{-6} , and 10^{-8} E_h , respectively. The auxiliary basis sets aCD-4+2C(10^{-4}), aCD-6+2C(10^{-6}), and aCD-8+2C(10^{-8}) were obtained from these by adding two-center functions to ensure target accuracies of 10^{-4} , 10^{-6} , and 10^{-8} E_h , respectively. The linear scaling of the PARI fitting calculations is clearly displayed in Figure 2, but we note that the prefactor is increased by a factor of two or more when two-center functions are added to the auxiliary basis. The number of two-center functions added for pairs involving only carbon atoms is listed in Table 2 for each target accuracy δ .

Table 2. C_nH_m , cc-pVTZ basis set. Number of two-center functions added to atom pairs involving only carbon atoms. R_{CC} denotes distance between carbon atoms in a given pair. The number of two-center functions added to carbon-carbon pairs with distances different from those given is zero.						
Auxiliary Basis set	$R_{CC}/\text{\AA}$					
	0.0	1.4	2.8	4.2	5.6	7.0
df-cc-pVTZ + 2C(10^{-4})	86	57	10	1	0	0
df-cc-pVTZ + 2C(10^{-6})	137	170	78	22	1	0
df-cc-pVTZ + 2C(10^{-8})	199	282	169	55	12	1
aCD-4 + 2C(10^{-4})	0	7	0	0	0	0
aCD-6 + 2C(10^{-6})	0	58	30	5	0	0
aCD-8 + 2C(10^{-8})	0	147	82	24	3	0

The number of two-center functions required to achieve a given accuracy depends strongly on the one-center auxiliary basis. With aCD auxiliary basis sets, two-center functions are most urgently needed for pairs involving bonded carbon pairs ($R_{CC} = 1.4$ \AA) but become important also for other pairs ($R_{CC} > 1.4$ \AA) as δ decreases. Significantly more two-center functions are needed (also at larger distances) with the df-cc-pVTZ auxiliary basis—in particular, many two-center functions are needed for diagonal AO pairs ($R_{CC} = 0.0$ \AA). This behaviour reflects the fact that the aCD sets are constructed with high integral accuracy in mind.^[65]

In spite of the negative eigenvalues of the PARI integral matrix, all SCF calculations presented here converge. Therefore, although the PARI energy errors are somewhat larger than the corresponding RI errors, further investigations of accuracy and efficiency are of interest.

Fitting errors and timings

We now examine the fitting errors obtained for the Coulomb and exact-exchange contributions to the KS matrix by using the RI and PARI schemes in restricted B3LYP calculations on selected closed-shell molecules. We also consider the speed-ups obtained with the RI and PARI schemes by comparing calculations performed with and without density fitting.

Three sets of calculations have been carried out. First, we perform B3LYP calculations on the 359 closed-shell molecules in the G3 benchmark set, using the 6-31G orbital basis and

the df-def2 auxiliary basis, denoting these calculations G3/6-31G. Next, we perform B3LYP calculations on the same molecules using the larger cc-pVTZ basis with the df-cc-pVTZ auxiliary basis (the triple-zeta valence-plus-polarization basis of Weigend^[38]) denoting these calculations G3/cc-pVTZ. Since no 6-31G basis exists for atoms with $31 \leq Z \leq 36$, molecules with these atoms were not considered. The G3/6-31G set, therefore, contains 323 closed-shell molecules. Likewise, no cc-pVTZ basis exists for potassium and no df-cc-pVTZ basis exists for $Z = 2, 3, 4, 10, 11, 12, 18, 31 - 35$; the G3/cc-pVTZ set, therefore, contains only 299 closed-shell molecules. In addition to the G3/6-31G and G3/cc-pVTZ calculations, we have performed calculations on 18 molecules previously considered by Peach et al.^[66] in a study of excitation energies and by Reine et al.^[30] in a study of integral density fitting; to these, we added benzene with a C—C distance of 1.395 \AA and a C—H distance of 1.0996 \AA . This resulting set contains molecules that are somewhat larger (up to 42 atoms and 1004 AOs) than those in the G3 set (up to 26 atoms and 322 AOs). For these 19 molecules, we have carried out restricted B3LYP calculations in the cc-pVTZ orbital basis with the df-cc-pVTZ auxiliary basis, denoting the calculations M19/cc-pVTZ.

In Table 3, we have listed the mean errors, standard deviations, and maximum absolute errors for the G3/6-31G, G3/cc-pVTZ,

Table 3. Mean errors ($\bar{\mu}$), standard deviations (σ), and maximum absolute errors (μ_{\max}) in μE_h for the G3/6-31G, G3/cc-pVTZ, and M19/cc-pVTZ sets calculated using the RI approximation for Coulomb (RI-J) and HF exchange (RI-K) and the PARI approximation for Coulomb (PARI-J) and HF exchange (PARI-K) at the B3LYP level of theory. Details about the three benchmark sets are given in the text.

		RI-J	PARI-J	RI-K	PARI-K
G3/6-31G ^[a]	$\bar{\mu}$	-20	-72	8	20
	σ	8	48	3	10
	μ_{\max}	35	202	11	21
G3/cc-pVTZ ^[b]	$\bar{\mu}$	-12	-22	9	14
	σ	4	10	3	7
	μ_{\max}	17	59	10	42
M19/cc-pVTZ	$\bar{\mu}$	-10	-20	9	14
	σ	2	5	4	3
	μ_{\max}	5	14	6	11

[a] PARI SCF optimizations of hexafluorobenzene and chloro-pentafluorobenzene did not converge. [b] PARI SCF optimizations of 3-butyn-2-one and 2-butyne did not converge.

pVTZ, and M19/cc-pVTZ calculations. The fitting errors are small—typically about $10 \mu E_h$. The Coulomb errors are larger than the exchange errors (and of opposite sign), reflecting the 20% weight of exact exchange in B3LYP. As expected, the PARI errors are larger than (but comparable to) the RI errors: about a factor of three in the 6-31G/df-def2 basis and less than a factor of two in the cc-pVTZ/df-cc-pVTZ basis.

While these errors are encouraging, we emphasize that two PARI calculations in each of the G3/6-31G and G3/cc-pVTZ sets failed to converge for four molecules, as indicated in Table 3. The nonconvergence is caused by negative eigenvalues of the

Table 4. Total BLYP energy errors (in μE_h) with two-center functions added to the auxiliary basis set to guarantee the specified integral accuracy δ (in E_h). Results are reported for the molecules for which the SCF iterations did not converge using one-center auxiliary functions only (see Table 3); 'nc' indicates nonconvergence.

Set	Molecule	δ		
		10^{-4}	10^{-6}	10^{-8}
G3/6-31G	hexafluorobenzene	-378.7	-9.5	-0.03
	chloro-pentafluorobenzene	-336.1	-9.0	-0.05
G3/cc-pVTZ	3-butyn-2-one	nc	-5.3	-0.07
	2-butyne	-18.8	-4.9	-0.05

integral matrix and is remedied by adding two-center functions to the auxiliary basis, see Table 4. An integral accuracy of $1 \mu E_h$ ensures convergence in all cases and $100 \mu E_h$ is sufficient in three cases. The exception is 3-butyn-2-one at the BLYP/cc-pVTZ level of theory (we have only implemented two-center fitting functions for PARI-J, hence we cannot use B3LYP) with the df-cc-pVTZ auxiliary basis set with two-center functions added to guarantee an integral accuracy of $100 \mu E_h$.

For 3-butyn-2-one from the G3/cc-pVTZ set, Figure 3 gives a summary of our findings regarding integral accuracy, $\|\Delta G\|_{\max}$,

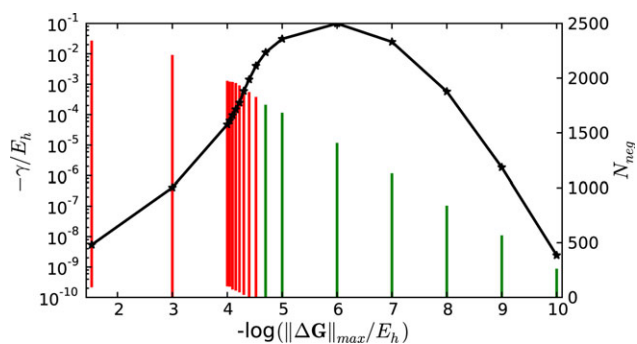


Figure 3. Results obtained for 3-butyn-2-one at the BLYP/cc-pVTZ level of theory with the df-cc-pVTZ auxiliary basis set with and without two-center functions added. The vertical lines show the extent of the negative PARI integral eigenvalues (smallest and largest) at a given integral accuracy measured by the maximum PARI integral error (the calculation without two-center functions is the one with least accurate integrals). The color of the vertical lines indicates divergence (red) or convergence (green) of the SCF iterations. The curve shows the total number of negative eigenvalues (on the right axis).

the number of negative integral eigenvalues N_{neg} , the range of the negative eigenvalues γ , and SCF convergence. Figure 4 shows the SCF iterations for the conventional calculation, the RI calculation, and selected PARI calculations with and without two-center functions added to the auxiliary basis. Using one-center auxiliary functions only, the largest negative PARI integral eigenvalue is $-25.5 \text{ m}E_h$ and the optimization diverges. In the first four SCF iterations, the absolute energy error (relative to the conventional energy at the same iteration) is in the range $10\text{--}100 \mu E_h$. Subsequently, the Coulomb energy decreases significantly at the fifth iteration and becomes negative at the sixth iteration, leading to a dramatic decrease in

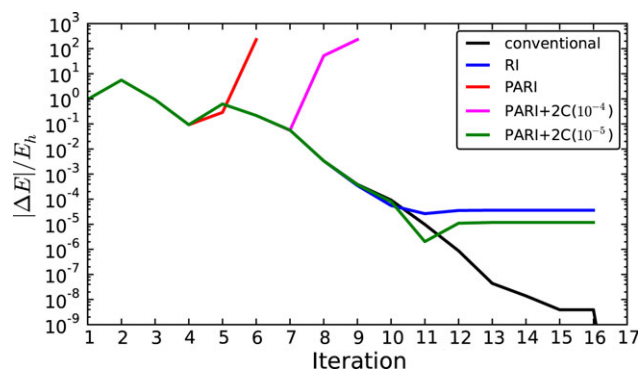


Figure 4. SCF iterations for 3-butyn-2-one at the BLYP/cc-pVTZ level of theory with the df-cc-pVTZ auxiliary basis set with and without two-center functions added. The absolute energy error on the vertical axis is defined relative to the energy at convergence of the conventional calculation. The PARI and PARI + $2C(10^{-4})$ calculations are aborted after iterations 6 and 9, respectively, due to negative Coulomb energy caused by negative eigenvalues of the approximate two-electron integral matrix.

the total energy as depicted in Figure 4. At this point, the electronic interactions are effectively attractive and the calculation is aborted.

The addition of two-center functions to increase integral accuracy leads to smaller largest negative eigenvalues, but the PARI integral matrix remains indefinite even with an integral accuracy of 10^{-10} . Nevertheless, convergence is achieved with an integral accuracy of about $20 \mu E_h$ or better. The associated largest negative eigenvalues are $-0.2 \text{ m}E_h$ or smaller, and the total energy errors range from $-22 \mu E_h$ to $-0.3 \text{ n}E_h$. Interestingly, as depicted in Figure 3, the total number of negative eigenvalues passes through a maximum at an integral accuracy of $1 \mu E_h$ and does not appear to have direct influence on convergence behavior.

For the M19/cc-pVTZ set, timings and speed-ups relative to calculations without density fitting are given in Table 5. The PARI-J and RI-J timings are comparable, as expected from the algorithm employed. The addition of two-center fitting functions to achieve target integral accuracies of 10^{-4} , 10^{-6} , and $10^{-8} E_h$ slows down the PARI-J calculations by factors of 21, 56, and 78, respectively, for hexacene, because of the increased time spent on integral evaluation. Integrals involving two-center fitting functions include four-center two-electron integrals and require the evaluation of a disproportionate number of primitive Gaussian integrals. We have used an integral-direct algorithm in this work; a significant improvement is expected by caching some or all of the integrals involving two-center fitting functions.

The PARI-K timings show promising speed-ups relative to the standard Link^[67] algorithm. Table 5 indicates an increased efficiency for larger molecules, which can be improved by optimizing the prescreening at the atomic level of the algorithm. The PARI-K speed-ups may also be improved by using localized occupied MOs in a manner similar to Polly et al.^[27]

Exploration of various PARI parameterizations

In this section, we explore the impact of different local metrics when combined with additional constraints on the fitting

Table 5. Timings and speed-ups of different fitting approximations for the M19/cc-pVTZ set in the 3rd SCF iteration. For the *J*-engine Coulomb (*J*) and LinK exchange (*K*) calculations, timings are in seconds; for the density-fitted (RI-*J*, PARI-*J*, and PARI-*K*) calculations, speed-ups relative to the *J* and *K* calculations are given, respectively. Details about the benchmark set are given in the text.

	atoms	AOs	<i>J</i>	<i>K</i>	RI- <i>J</i>	PARI- <i>J</i>	PARI- <i>K</i>
HCl	2	48	0.2	0.3	4.5	3.6	1.6
CO	2	60	0.2	0.6	4.4	3.1	1.5
N ₂	2	60	0.2	0.6	4.6	3.8	1.6
H ₂ CO	4	88	0.8	2.3	6.5	5.2	2.2
polyacetylene oligomer (<i>n</i> = 3)	10	204	14	38	17	15	4.3
benzene (acene-0)	12	264	39	135	23	20	6.8
polyacetylene oligomer (<i>n</i> = 4)	14	292	41	118	21	20	4.8
polyacetylene oligomer (<i>n</i> = 5)	18	380	84	302	26	21	5.9
naphtalene (acene-1)	18	412	172	488	32	27	7.1
dipeptide	19	410	116	412	25	22	5.6
N-phenylpyrrole	20	456	214	650	32	30	6.8
4-(N,N-dimethylamino)benzotrile	21	470	223	812	31	28	7.9
polyacetylene oligomer (<i>n</i> = 2)	22	468	139	411	26	24	4.8
β-dipeptide	22	468	163	571	27	24	5.7
anthracene (acene-2)	24	560	419	1229	37	33	7.0
tripeptide	26	572	269	934	24	26	5.5
tetracene (acene-3)	30	708	780	2355	40	34	7.2
pentacene (acene-4)	36	856	1240	3669	41	38	6.7
hexacene (acene-5)	42	1004	1822	5440	44	40	6.5

equations. In Table 6, we compare errors in the B3LYP energy obtained with different versions of the PARI-*J* and PARI-*K* methods. Specifically, we perform PARI calculations on acenes in the Coulomb and overlap metrics, using unconstrained fitting, charge-constrained fitting, and charge–dipole-constrained fitting.

Broadly speaking, while the overlap metric attempts to reproduce the overall shape of the density, the Coulomb metric weighs more strongly regions that are important for the two-electron repulsion energy. Accordingly, the PARI method produces much smaller errors in the Coulomb metric than in the overlap metric—in particular, for unconstrained fitting.

When interpreting the errors in Table 6, we should bear in mind that exact exchange constitutes only 20% of the total exchange energy in the B3LYP model; when comparing Coulomb and exchange errors, the latter should be multiplied by five. When no constraints are imposed on the fitting, the Coulomb energy is better reproduced than the exact-exchange energy.

In the overlap metric, errors increase with system size and most PARI-*J* calculations fail to converge due to negative eigenvalues of the integral matrix. The increase in error is stronger for the Coulomb energy, for which long-range interactions are more important than for the exchange energy. Imposing charge conservation, the overlap metric errors are reduced to a level only 50% greater than those of the Coulomb metric.

Charge constraints improve the description of long-range interactions (at the expense of short-range interactions). Further long-range improvements may be expected from imposing conditions also on the dipole moment. However, dipole constraints tend to increase errors in the overlap metric. Moreover, the Coulomb errors increase by up to 50%. The best compromise between accurate short- and long-range interactions is, therefore, achieved by imposing charge constraints which, however, are not sufficient to cure the convergence failures.

Table 6. Errors in the B3LYP energy per non-hydrogen atom (in μE_h) in the first SCF iteration for selected acenes from the M19/cc-pVTZ molecules, using different varieties of the PARI approximation. In all calculations, the converged density obtained without density fitting was used as starting guess. The fitting coefficients are obtained using the overlap and Coulomb metrics. For each metric, the fitting has been obtained without constraints (free), with charge constraint (charge), and with charge and dipole constraints (dipole). SCF calculations that did not converge are in bold. Details about the benchmark set are given in the text.

		Coulomb metric			overlap metric		
		free	charge	dipole	free	charge	dipole
PARI- <i>J</i>	benzene	−23.3	−22.9	−24.7	−69.0	−32.2	−46.3
	naphtalene	−22.1	−21.7	−22.7	−98.7	−30.9	−40.8
	anthracene	−21.5	−21.1	−21.8	−120.7	−30.1	−37.8
	tetracene	−21.2	−20.8	−21.3	−135.5	−29.7	−36.2
PARI- <i>K</i>	benzene	12.8	12.8	13.0	26.3	18.6	18.2
	naphtalene	14.2	14.3	14.5	28.5	20.6	20.2
	anthracene	14.9	14.9	15.1	29.4	21.4	21.2
	tetracene	15.2	15.2	15.5	29.9	21.9	21.7

Conclusions

We have identified an inherent weakness of local RI approaches. When local fitting domains or non-Coulomb-fitting metrics are combined with robust corrections, the matrix representing the Coulomb operator is indefinite. Indefiniteness may lead to nonconvergence in SCF iterations or convergence to a spurious state of energy far below the true ground-state energy, even if integral errors are small. In addition to a general theoretical analysis, we have investigated the problem numerically by means of the PARI approximation to four-center two-electron integrals.

Involving only auxiliary functions centered on the parent atoms, PARI is a highly local fitting approach of linear complexity, including Dunlap's robust correction to ensure quadratic integral accuracy in the fitting errors. Algorithms for the Coulomb and exchange contributions to the Fock/KS matrix have been described and sample calculations on organic molecules demonstrate that, with a few exceptions (less than 1%), the indefinite PARI integral matrix does not cause convergence problems. The PARI and standard RI schemes provide similar speed-ups and comparable accuracy for the Coulomb contribution. The PARI errors are somewhat larger and the speed-ups less pronounced (up to a factor of eight) for the exchange contribution, which may be improved by exploiting locality in the occupied MO subspace. In a few cases, the indefinite PARI integral matrix gives rise to severe convergence problems, whose only known cure is local completeness of the auxiliary basis set, as demonstrated by calculations using two-center fitting functions derived by CD of the residual integral error for each pair of atoms. In our current implementation, the addition of two-center functions to the auxiliary basis is accompanied by a severe computational penalty.

Acknowledgments

The authors would like to acknowledge the NOTUR computing facilities which have been used to conduct the calculations in this article.

Keywords: resolution-of-the-identity · density fitting · indefinite two-electron integral matrix · convergence problems · Hartree–Fock · density-functional theory

How to cite this article: P. Merlot, T. Kjærgaard, T. Helgaker, R. Lindh, F. Aquilante, S. Reine, T. B. Pedersen, *J. Comput. Chem.* **2013**, *34*, 1486–1496. DOI: 10.1002/jcc.23284

- [1] S. F. Boys, I. Shavitt, A Fundamental Calculation of the Energy Surface for the System of Three Hydrogen Atoms; University of Wisconsin Report. *WIS-AF-13*, **1959**.
- [2] F. E. Harris, R. Rein, *Theor. Chim. Acta* **1966**, *6*, 73.
- [3] F. P. Billingsley, II, J. E. Bloor, *J. Chem. Phys.* **1971**, *55*, 5178.
- [4] J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496.
- [5] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41.
- [6] H. Sambe, R. H. Felton, *J. Chem. Phys.* **1975**, *62*, 1122.
- [7] B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 3396.

- [8] B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 4993.
- [9] M. Feyereisen, G. Fitzgerald, *Chem. Phys. Lett.* **1993**, *208*, 359.
- [10] O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*, 514.
- [11] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283.
- [12] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, *Theor. Chem. Acc.* **1997**, *97*, 119.
- [13] R. Bauernschmitt, M. Häser, O. Treutler, R. Ahlrichs, *Chem. Phys. Lett.* **1997**, *264*, 573.
- [14] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143.
- [15] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, 864.
- [16] L. J. Sham, W. Kohn, *Phys. Rev. A* **1965**, *140*, 1133.
- [17] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618.
- [18] O. Christiansen, H. Koch, P. Jørgensen, *Chem. Phys. Lett.* **1995**, *243*, 409.
- [19] F. Weigend, A. Köhn, C. Hättig, *J. Chem. Phys.* **2002**, *116*, 3175.
- [20] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- [21] C. Hättig, F. Weigend, *J. Chem. Phys.* **2000**, *113*, 5154.
- [22] R. Flores-Moreno, J. V. Ortiz, *J. Chem. Phys.* **2009**, *131*, 124110.
- [23] V. P. Vysotskiy, L. S. Cederbaum, *J. Chem. Phys.* **2010**, *132*, 044110.
- [24] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, M. Scheffler, *New J. Phys.* **2012**, *14*, 053020.
- [25] Y. Jung, R. C. Lochan, A. D. Dutoi, M. Head-Gordon, *J. Chem. Phys.* **2004**, *121*, 9793.
- [26] H. Eshuis, J. Yarkony, F. Furche, *J. Chem. Phys.* **2010**, *132*, 234114.
- [27] R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- [28] F. Aquilante, T. B. Pedersen, R. Lindh, *J. Chem. Phys.* **2007**, *126*, 194106.
- [29] A. Sodt, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 104106.
- [30] S. Reine, E. Tellgren, A. Krapp, T. Kjærgaard, T. Helgaker, B. Jansík, S. Høst, P. Salek, *J. Chem. Phys.* **2008**, *129*, 104101.
- [31] H.-J. Werner, F. R. Manby, P. J. Knowles, *J. Chem. Phys.* **2003**, *118*, 8149.
- [32] H.-J. Werner, F. R. Manby, *J. Chem. Phys.* **2006**, *124*, 054114.
- [33] M. Schütz, F. R. Manby, *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
- [34] B. I. Dunlap, *J. Mol. Struct. (Theochem)* **2000**, *501*, 221.
- [35] B. I. Dunlap, *J. Mol. Struct. (Theochem)* **2000**, *529*, 37.
- [36] B. I. Dunlap, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113.
- [37] B. I. Dunlap, N. Rösch, S. Trickey, *Mol. Phys.* **2010**, *108*, 3167.
- [38] F. Weigend, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.
- [39] F. Weigend, *J. Comput. Chem.* **2008**, *29*, 167.
- [40] C. Hättig, *Phys. Chem. Chem. Phys.* **2005**, *7*, 59.
- [41] A. Hellweg, C. Hättig, S. Höfener, W. Klopper, *Theor. Chem. Acc.* **2007**, *117*, 587.
- [42] J. G. Hill, J. A. Platts, *J. Chem. Phys.* **2008**, *128*, 044104.
- [43] V. D. Domínguez-Soria, G. Geudtner, J. L. Morales, P. Calaminici, A. M. Köster, *J. Chem. Phys.* **2009**, *131*, 124102.
- [44] Y. Jung, A. Sodt, P. M. W. Gill, M. Head-Gordon, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6692.
- [45] P. M. W. Gill, A. T. B. Gilbert, S. W. Taylor, G. Friesecke, M. Head-Gordon, *J. Chem. Phys.* **2005**, *123*, 061101.
- [46] T. B. Pedersen, F. Aquilante, R. Lindh, *Theor. Chem. Acc.* **2009**, *124*, 1.
- [47] F. Aquilante, L. Boman, J. Boström, H. Koch, R. Lindh, A. Sánchez de Merás, T. B. Pedersen, In *Challenges and Advances in Computational Chemistry and Physics*, Vol. 13, Chap. 13; R. Zalesny, M. G. Papadopoulos, P. G. Mezey, J. Leszczynski, Eds.; Springer Science+Business Media B.V., Springer Netherlands., **2011**; pp. 301–343.
- [48] R. T. Gallant, A. St-Amant, *Chem. Phys. Lett.* **1996**, *256*, 569.
- [49] A. Sodt, J. E. Subotnik, M. Head-Gordon, *J. Chem. Phys.* **2006**, *125*, 194109.
- [50] P. Salek, S. Høst, L. Thøgersen, P. Jørgensen, P. Manninen, J. Olsen, B. Jansík, S. Reine, F. Pawłowski, E. Tellgren, T. Helgaker, S. Coriani, *J. Chem. Phys.* **2007**, *126*, 114110.
- [51] M. Krykunov, T. Ziegler, E. van Lenthe, *Int. J. Quantum Chem.* **2009**, *109*, 1676.
- [52] T. J. Giese, D. M. York, *J. Chem. Phys.* **2011**, *134*, 194103.
- [53] N. H. F. Beebe, J. Linderberg, *Int. J. Quantum Chem.* **1977**, *12*, 683.
- [54] I. Røeggen, E. Wisløff-Nilssen, *Chem. Phys. Lett.* **1986**, *132*, 154.

- [55] D. W. O'Neal, J. Simons, *Int. J. Quantum Chem.* **1989**, *36*, 673.
- [56] S. Wilson, *Comput. Phys. Commun.* **1990**, *58*, 71.
- [57] H. Koch, A. Sánchez de Merás, T. B. Pedersen, *J. Chem. Phys.* **2003**, *118*, 9481.
- [58] F. Aquilante, L. Gagliardi, T. B. Pedersen, R. Lindh, *J. Chem. Phys.* **2009**, *130*, 154107.
- [59] F. Aquilante, R. Lindh, T. B. Pedersen, *J. Chem. Phys.* **2008**, *129*, 034106.
- [60] D. Foerster, *J. Chem. Phys.* **2008**, *128*, 034108.
- [61] LSDALTON, A linear scaling molecular electronic structure program, Release Dalton2011, 2011, Available at: <http://www.daltonprogram.org/>
- [62] F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-Å. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov, R. Lindh, *J. Comput. Chem.* **2010**, *31*, 224.
- [63] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
- [64] T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007.
- [65] F. Aquilante, R. Lindh, T. B. Pedersen, *J. Chem. Phys.* **2007**, *127*, 114107.
- [66] M. J. G. Peach, P. Benfield, T. Helgaker, D. J. Tozer, *J. Chem. Phys.* **2008**, *128*, 044118.
- [67] C. Ochsenfeld, C. A. White, M. Head-Gordon, *J. Chem. Phys.* **1998**, *109*, 1663.

Received: 30 December 2012

Revised: 1 March 2013

Accepted: 5 March 2013

Published online on 3 April 2013