Gradients in the ab initio scalar zeroth-order regular approximation (ZORA) approach

J.H. van Lenthe a, S. Faas a,*, J.G. Snijders b

a Theoretical Chemistry Group, Debye Institute, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands
b Theoretical Chemistry, Materials Science Centre, Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 22 March 2000; in final form 7 July 2000

Abstract

We discuss ways to obtain analytical gradients within the scalar zeroth-order regular approximation (ZORA) to the Dirac–Fock equation within an ab initio context. Simply employing the relativistic density within the non-relativistic gradient package is in error by $10^{-5}$. We introduce a new strictly atomic scheme which in addition to yielding exact gradients is also computationally inexpensive and avoids the gauge invariance problems that plague molecular ZORA approaches. We show that the total and orbital energies produced with the scaled version of this method are generally, i.e. except for very short interatomic distances, very close to the full molecular scaled ZORA results. Equilibrium geometries from full molecular scaled ZORA and strictly atomic ZORA are shown to be within 0.01 Å from Dirac–Fock.

1. Introduction

In recent years, the ZORA [1–4] as well as Douglas–Kroll methods [5,6] have been established as cost-effective alternatives to the Dirac–Fock method in single-point calculations. Equilibrium geometries of small, say 2 or 3 atomic, molecules can be obtained by performing a number of single-point calculations. To enable geometry optimisations on larger molecules, the availability of analytical energy gradients is essential. Gradients for the four-component Dirac–Fock method are currently being developed within the DIRAC program package [7].

The simplest way to obtain approximate energy gradients for scalar relativistic wavefunctions is to feed the relativistic density to a non-relativistic gradient package, i.e. all expressions stay the same as in the non-relativistic case. The only thing that is different with respect to a non-relativistic calculation is that the density is obtained from the relativistically calculated wavefunction. This approach, known as a ‘poor man’s method’ (B.A. Hess, personal communication; Ref. [8]), has also successfully been applied using the Douglas–Kroll method [9,10]. We will investigate the error introduced by this approach.

Recently two different ways [11,12] have been developed to obtain analytical energy gradients for the ZORA method within the context of density functional theory (DFT). In both papers, changes are made to the ZORA kinetic energy operator, which is orbital and geometry dependent, to facilitate the computation of analytical energy gradients.

The ZORA (MP) method due to van Wullen [11] uses a model potential (MP) in the ZORA correc-
tions. One of the conditions that this MP should satisfy is that it contains no contributions from ‘dis-
tant’ nuclei. This is meant to eliminate the ZORA gauge invariance problem. However, the final form
of the model potential still contains the full nuclear potential ($V_{\text{nuc}}$) and is therefore not able to meet this
requirement. Even though $V_{\text{nuc}}$ is counteracted by the Coulomb repulsion of the different electron clouds, a
small gauge invariance problem remains [12].

In order to obtain analytical gradients, van Lenthe et al. [12] use a fixed, i.e. orbital independent, poten-
tial to evaluate the ZORA corrections. They use the ZORA ESA method for calculating the molecular
binding energies to avoid gauge invariance problems. Still, terms of order $(1/c^2)$ are being neglected.

In the present work we propose a new, strictly atomic, scalar (scaled) ZORA scheme. We eliminate
all geometry and orbital dependencies in the ZORA kinetic energy operator. For this new ZORA scheme
the expression for the gradients is equal to that in a non-relativistic calculation. Therefore the use of a
non-relativistic gradient package is valid. Moreover, the use of these strictly atomic ZORA corrections
circumvents the problem of gauge invariance in molecular calculations. To enable scaled ZORA gra-
dients, we define a strictly atomic scaling procedure.

The quality of the atomic scalar (scaled) ZORA method is investigated by test calculations on the
molecules: Au$_2$, CsI, AuH and AgH, for which we compare with the full Coulomb option presented
earlier [13,14] and Dirac–Fock results [7]. The CsI molecule has been chosen because the strictly atomic
option might be expected to be inferior for ionic molecules. For both methods we compare numerical
and analytical gradients.

2. Theory

The ab initio scalar ZORA Fock equation is given by

$$
\left( \frac{1}{2} \vec{p}^2 - \frac{1}{1 - (V_r/2c^2)} \vec{p} + V_{\text{nuc}} + J_{\phi\phi} - K_{\phi\phi} \right) \phi_j
= \epsilon_r^{\text{ZORA}} \phi_j
$$

with $c$ the speed of light and $J$, $K$ and $p$ the Coulomb, exchange and momentum operators, re-
respectively. The only difference with respect to the Schrödinger equation is the kinetic energy operator
($T^{\text{ZORA}}$), which has become geometry and potential dependent through $V_r = V_{\text{nuc}} + J_{\phi\phi}$.

The matrix elements of the ZORA kinetic energy operator are calculated as the non-relativistic kinetic
energy plus a relativistic ZORA correction [14].

$$
(T^{\text{ZORA}})_{\mu\nu} = \frac{1}{2} \left( \phi_\mu | p^2 | \phi_\nu \right) + \left( \phi_\mu | T^{\text{corr}} | \phi_\nu \right)
= \frac{1}{2} \left( \phi_\mu | p^2 | \phi_\nu \right)
- \frac{1}{2} \sum_{\lambda} \left( \phi_\mu | \vec{p} \phi_\lambda \right) S^{-1}_{\lambda\lambda} \left( \phi_\lambda | \vec{p} \phi_\nu \right)
+ \frac{1}{2} \sum_{\lambda,\kappa} \left( \phi_\mu | \vec{p} \phi_\lambda \right) S^{-1}_{\kappa\kappa} \left( \phi_\kappa | \vec{p} \phi_\nu \right)
\times \left( \phi_\lambda \left| 1 - \frac{V_r}{2c^2} \right| \phi_\nu \right) ^{-1} S^{-1}_{\lambda\lambda} \left( \phi_\lambda \phi_\nu \right)
$$

(2)

where $S^{-1}_{\mu\nu}$ is the inverted metric of the non-orthogo-
nal basis $\{ \phi \}$.

From Eq. (2) it can be concluded that the terms that are neglected in geometry optimisations using
the full Coulomb ZORA option are terms of order $(V_r/2c^2)$. For valence orbitals $\langle (V_r/2c^2) \rangle$ is negli-
gible causing the last two terms in Eq. (2) to approx-
imately cancel each other. Taking the derivative of just the first term is therefore for the valence orbitals
a good approximation. Remember that the relativistic effect on the valence electrons is mainly indirect
through a change in core density which is accounted for. The only integrals contributing to the gradient
are those involving orbitals on different centres. For core orbitals, where $\langle (V_r/2c^2) \rangle$ is non-negligible,
these are small.

We now modify the ZORA kinetic energy operator is such a way that the use of non-relativistic
gradient operator becomes exact. Important for the new atomic ZORA scheme is the fact that kinetic
energy (Eq. (2)) is split in two parts: (1) the non-relativistic kinetic energy, and (2) the ZORA corrections.
to it. This makes it possible to make additional modifications to the original ZORA corrections while still obtaining the Schrödinger equation in the non-relativistic limit.

In the new ZORA scheme we want to have strictly atomic ZORA corrections to the kinetic energy operator leading to ZORA corrections that are independent of geometry and (molecular) density. They are therefore constant throughout the SCF procedure and geometry optimisation. This implies, among other things, the use of a Coulomb matrix \( V_{\text{atom}} \) with only intra-atomic elements in combination with an approximate density, which is a superposition of densities of the constituent neutral atoms. Hence

\[
\frac{\partial V_{\text{atom}}}{\partial X_i} = 0
\]

where \( X_i \) is an arbitrary molecular co-ordinate. Note that the core electrons are the most important electrons for ZORA and these do not influence electrons on other atoms.

The ZORA corrections are calculated separately for each atom. This means that all matrix elements involved in the calculation of \( T_{\text{core}} \), i.e. all matrix elements in the last two terms of Eq. (2), are set to zero for functions situated on different atoms. Together with the use of an atomic Coulomb matrix \( V_{\text{atom}} \) this amounts to the evaluation of \( \langle \phi_{\mu} | T_{\text{core}} | \phi_{\nu} \rangle \) in an atomic calculation once and forever. This correction might then be stored for each atom and basis set in a library and added as a block diagonal correction to the one-particle Hamiltonian in a molecular calculation. The resulting ZORA correction matrix is block diagonal over the atoms and depends only on intra-atomic terms.

The ZORA methods suffer from a lack of gauge invariance. If one shifts the potential with a constant value the energies should be shifted with the same constant. This is not the case for the ZORA Hamiltonian. In molecular calculations this causes an incorrect geometry dependence of the ZORA corrections. This can be seen if one considers a diatomic molecule AB. The charge distribution of atom B results in a constant shift of the potential for e.g. the core electrons of atom A. These core electrons do not respond correctly, i.e. by shifting their energy by the same amount, to this shift due to the presence of the full molecular potential in the denominator of the ZORA kinetic energy operator. Scaled ZORA solves the gauge invariance problem almost completely but still a small problem remains, e.g. if 1 Hartree is added to the potential in a Xenon calculation the shift in total energy is in error by 0.004 Hartree for ZORA and by 0.00004 Hartree for scaled ZORA [15]. The strictly atomic scheme avoids this problem completely because the kinetic energy operator working on the electrons of A does not contain contributions from atom B.

The scaled ZORA method is defined by solving Eq. (1) self-consistently and scaling the orbital energies afterwards by

\[
\varepsilon_i^{\text{scaled}} = \frac{\varepsilon_i}{1 + \left| \phi_i \right| \frac{c^2}{(2\varepsilon^2 - V_i)} \frac{\hat{p}}{2} \left| \phi_i \right|}
\]

where \( \phi_i \) is the molecular orbital with orbital energy \( \varepsilon_i \).

The strictly atomic scaling procedure is performed for each atom seperately. One builds the closed-shell-type Fock matrix using the atomic density and diagonalises it. This choice of Fock matrix can be made even for open-shell atoms [16]. Note that the open shells are usually located in the valence region where the effect of the scaling is very small anyway. The resulting orbitals are used to calculate the scaling factors. The scaling correction matrix, defined as the diagonal matrix of orbital energies minus the diagonal matrix of scaled orbital energies, back transformed to AO basis, is added to the 1-electron operator. This procedure is identical to that proposed for correlated scaled ZORA calculations [15]. In the one electron basis this results in a one electron correction matrix of the form

\[
\psi\left(\varepsilon^{\text{scaled}} - \varepsilon\right)\psi^\dagger
\]

where \( \psi \) are the eigenvectors of the above-mentioned atomic Fock matrix and \( \varepsilon^{\text{scaled}} \) and \( \varepsilon \) are the corresponding diagonal eigenvalue matrices. This re-
sults in a scaling correction that is block diagonal over the atoms and again only dependent on intraatomic terms. For one atom the combination of the atomic scaling correction and the ZORA kinetic energy correction leads to a Hamiltonian that if used in a Hartree–Fock calculation reproduces both the scaled orbital energies and the scaled ZORA total energy.

3. Results

The quality of the new atomic scalar scaled ZORA scheme is investigated in calculations on a covalent molecule (Au₃) and an ionic one (CsI). The results are compared with the scaled scalar ZORA full Coulomb method presented earlier. The basis sets used in the calculations are based upon non-relativistic sets due to Poirier et al. [17] and Huzinaga et al. [18] and are augmented with tight s and p functions and diffuse d (and f) functions. The sets are available upon request.

Table 1
Orbital energies for CsI in Hartree for selected orbitals computed with scaled ZORA and scaled strictly atomic ZORA at an interatomic distance of 7.2 Bohr

<table>
<thead>
<tr>
<th>Orbital</th>
<th>Full</th>
<th>Strictly atomic</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1σ</td>
<td>−1324.87580</td>
<td>−1324.87598</td>
<td>0.00019</td>
</tr>
<tr>
<td>2σ</td>
<td>−1224.39790</td>
<td>−1224.39700</td>
<td>0.00090</td>
</tr>
<tr>
<td>3σ</td>
<td>−211.98307</td>
<td>−211.98317</td>
<td>0.00010</td>
</tr>
<tr>
<td>4σ</td>
<td>−192.76441</td>
<td>−192.76392</td>
<td>0.00049</td>
</tr>
<tr>
<td>1π</td>
<td>−190.11767</td>
<td>−190.11778</td>
<td>0.00011</td>
</tr>
<tr>
<td>5σ</td>
<td>−190.11767</td>
<td>−190.11778</td>
<td>0.00011</td>
</tr>
<tr>
<td>6σ</td>
<td>−172.71405</td>
<td>−172.71352</td>
<td>0.00053</td>
</tr>
<tr>
<td>3π</td>
<td>−172.71399</td>
<td>−172.71347</td>
<td>0.00052</td>
</tr>
<tr>
<td>7σ</td>
<td>−45.77324</td>
<td>−45.77328</td>
<td>0.00004</td>
</tr>
<tr>
<td>8σ</td>
<td>−40.37334</td>
<td>−40.37312</td>
<td>0.00022</td>
</tr>
<tr>
<td>5π</td>
<td>−38.52342</td>
<td>−38.52346</td>
<td>0.00004</td>
</tr>
<tr>
<td>9σ</td>
<td>−38.52326</td>
<td>−38.52330</td>
<td>0.00004</td>
</tr>
<tr>
<td>10σ</td>
<td>−33.71991</td>
<td>−33.71968</td>
<td>0.00023</td>
</tr>
<tr>
<td>7π</td>
<td>−33.71956</td>
<td>−33.71934</td>
<td>0.00022</td>
</tr>
<tr>
<td>18</td>
<td>−27.77867</td>
<td>−27.77872</td>
<td>0.00004</td>
</tr>
<tr>
<td>19σ</td>
<td>−1.34242</td>
<td>−1.34241</td>
<td>0.00001</td>
</tr>
<tr>
<td>20σ</td>
<td>−0.75180</td>
<td>−0.75178</td>
<td>0.00002</td>
</tr>
<tr>
<td>21π</td>
<td>−0.71618</td>
<td>−0.71617</td>
<td>0.00001</td>
</tr>
<tr>
<td>21σ</td>
<td>−0.70829</td>
<td>−0.70828</td>
<td>0.00001</td>
</tr>
<tr>
<td>22σ</td>
<td>−0.27047</td>
<td>−0.27046</td>
<td>0.00001</td>
</tr>
<tr>
<td>23π</td>
<td>−0.26104</td>
<td>−0.26103</td>
<td>0.00001</td>
</tr>
</tbody>
</table>

Table 2
Total energies for the CsI molecule in Hartree for different interatomic distances (in Å) computed with scaled ZORA

<table>
<thead>
<tr>
<th>R (Å)</th>
<th>Full</th>
<th>Strictly atomic</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>−14872.4896423</td>
<td>−14872.4896465</td>
<td>0.0000041</td>
</tr>
<tr>
<td>3.0</td>
<td>−14872.6161193</td>
<td>−14872.6161295</td>
<td>0.0000103</td>
</tr>
<tr>
<td>3.2</td>
<td>−14872.6303926</td>
<td>−14872.6303786</td>
<td>0.0000140</td>
</tr>
<tr>
<td>3.4</td>
<td>−14872.6359015</td>
<td>−14872.6358699</td>
<td>0.00000317</td>
</tr>
<tr>
<td>3.6</td>
<td>−14872.6363432</td>
<td>−14872.6363012</td>
<td>0.00000420</td>
</tr>
<tr>
<td>4.5</td>
<td>−14872.6178190</td>
<td>−14872.6177611</td>
<td>0.00000579</td>
</tr>
<tr>
<td>9.0</td>
<td>−14872.5564637</td>
<td>−14872.5563985</td>
<td>0.00000652</td>
</tr>
<tr>
<td>18.0</td>
<td>−14872.5268146</td>
<td>−14872.5267488</td>
<td>0.00000658</td>
</tr>
</tbody>
</table>

Table 1 lists the orbital energies for selected orbitals calculated with the scaled ZORA and the scaled strictly atomic ZORA methods at an interatomic distance of 7.2 Bohr. We see that the error introduced ranges from 0.0009 Hartree in the core region to 0.00001 Hartree in the valence region.

In Table 2 the total energies for the CsI molecule at various inter-atomic distances identical both methods are given. The error introduced by the strictly atomic approximation is very small and always less than 0.00007 Hartree. Near the equilibrium geometry (3.52 Å) the error is even somewhat smaller.

The same comparison can be made for the Au₃ molecule (Table 3). Again we see that the agreement between the methods is very good. Only at very small intermolecular distances the deviations is as large as −0.054 Hartree.
CSI molecules at a few inter-atomic distances. The deviations, between analytical and numerical gradients, for the non-relativistic and strictly atomic ZORA approaches are of order $10^{-6}$ or less. It is likely that the error is in the numerical procedure.

The deviation between the numerical and analytical gradients for the full Coulomb method is of order $10^{-3}$. Comparable errors are found using this method in combination with the Douglas–Kroll approach [20].

The equilibrium bondlengths of the molecules CSI, Au₂, AuH and AgH is calculated with the strictly atomic and full Coulomb ZORA options (Table 5). We see that, for scaled ZORA, both methods are in close agreement for all molecules. This means that the new scaling procedure is able to reproduce the effect of the original one. Moreover we see that the strictly atomic ZORA results are, just as the full molecular scaled ZORA results, within 0.01 Å from the Dirac–Fock results obtained in the same (large component) basis set. It should be noted that the full Coulomb results are obtained from single-point calculations. This is because the geometry optimisations do not converge. Very close to the equilibrium the geometry optimisation algorithm fails due to the fact that the calculated hessian does not correspond to the total energy curve. The deviation between scaled and unscaled results, which are caused by the gauge invariance problem connected to the full Coulomb ZORA approach, has dissapeared in case of the strictly atomic option.

### 4. Conclusions

We investigated the error introduced by the use of the relativistic density in the non-relativistic gradient expressions, i.e. the ‘poor man’s’ method, in a full Coulomb ZORA calculation. We showed that this error is of order $e^{-5}$ ($\approx 5e^{-5}$). This was supported by numerical evidence. Full Coulomb ZORA geometry calculations can be performed if one relaxes the convergence criteria a bit. Hessian-related convergence problems suggest that this method can be useful in finding optimal geometries but not for saddle-point optimisations.

A new, strictly atomic (scaled) ZORA scheme has been implemented in the GAMESS-UK package [19]. All geometry and orbital dependence has been removed from the ZORA kinetic energy operator, resulting in the first ZORA implementation that avoids the disadvantages of gauge invariance and enables exact energy gradients. The new ZORA scheme has been compared to the full Coulomb ZORA method in calculations on the total and orbital energies of the CSI and Au₂ molecules. The method is shown to be quite accurate, while being computationally much cheaper. The ZORA corrections have to be calculated only once on basis of a density obtained from a number of atomic SCF calculations. The method is able to give accurate relativistic results at almost non-relativistic cost.

In calculations on the bondlength of the molecules CSI, Au₂, AuH and AgH we showed that the strictly...
The atomic option is as accurate as the full Coulomb one. Moreover it was demonstrated that the scaled full Coulomb and both the scaled and unscaled strictly atomic ZORA methods are within 0.01 Å of Dirac–Fock.

The new ZORA scheme circumvents the gauge invariance problems that is present in the original ZORA methods. Consequently, scaled and unscaled results, for the strictly atomic option, are in very good agreement.

Acknowledgements

We thank U. Wahlgren for communicating his Douglas–Kroll results and Alf Christian Hennum for performing the Dirac–Fock [7] calculations.

References

[19] M.F. Guest et al., GAMESS-UK is a package ab initio programs, The package is derived from the original GAMESS code due to M. Dupuis, D. Spangler, J. Wendoloski, NRCC Software Catalog, vol. 1, Program No. QG01 (GAMESS), 1980.