EFFICIENT ENERGY AND ELECTROSTATIC PROPERTIES CALCULATIONS AT THE MP2 THEORY LEVEL: A CASE STUDY OF DENSITY MATRIX-BASED COMPUTATIONAL QUANTUM CHEMISTRY

Grzegorz MAZUR

Department of Computational Methods in Chemistry
Jagiellonian University
ul. R. Ingardena 3, 30-060 Kraków, Poland
e-mail: `mazur@chemia.uj.edu.pl`

Marcin MAKOWSKI

Department of Theoretical Chemistry
Jagiellonian University
ul. R. Ingardena 3, 30-060 Kraków, Poland
e-mail: `makowskm@chemia.uj.edu.pl`

Jakub SUMERA, Krzysztof KOWALCZYK

Faculty of Chemistry
Jagiellonian University
ul. R. Ingardena 3, 30-060 Kraków, Poland

Communicated by Jacek Kitowski

Abstract. Wavefunction-less, density matrix-based approach to computational quantum chemistry is briefly discussed. Implementation of second-order Møller-Plesset Perturbation Method energy and dipole moment calculations within the new paradigm is presented. Efficiency and reliability of the method is analyzed.

Keywords: Computational chemistry, MP2, Laplace transform, linear scaling

1 INTRODUCTION

Significantly growing demand for high quality computational chemistry results for nano-materials and bio-chemical systems can be recently observed. The standard, wavefunction based computational quantum chemistry methods often cannot fulfill the expectations. The main reason is inherent high computational complexity of these methods directly stemming from the bad locality of molecular orbitals, which are the main building block of any wavefunction-based quantum-chemical formalism. This gives rise to the paradigm shift towards alternative, purely density matrix-based computational quantum chemistry methods which promise linear computational complexity. Supported by high-performance hardware and efficient algorithms, the new approach may provide viable toolkit for treating nano- and biosystems.

The aim of the paper is twofold. We highlight the crucial algorithmic aspects of density-matrix based approach, and study its efficiency and reliability. We choose the implementation of the second order Møller-Plesset perturbation method (MP2) as an illustrative example.

2 LAPLACE-TRANSFORM MP2

To keep the paper self-contained, we briefly introduce the fundamental elements of the formalism we implemented and analyzed.

The second order energy correction in the Møller-Plesset perturbation method (MP2) takes the form [1, 2]

$$E_2 = -\sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{1}$$

where i, j denote occupied orbitals, a, b denote virtual orbitals, ϵ stands for orbital energy and $(ia|jb)$ represents a two-electron integral in chemists' notation.

From the computational complexity point of view, the crucial step is the transformation of the two-electron integrals from the atomic orbital (AO) to the molecular orbital (MO) basis set

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma}^{N} C_{\mu i} C_{\nu a} C_{\lambda j} C_{\sigma b} (\mu\nu|\lambda\sigma) \tag{2}$$

where Greek letters denote AOs, C stands for the molecular orbital coefficient matrix and N represents the dimension of the Hilbert subspace spanned by the atomic orbitals. The complexity stems from the fact that the C matrix is dense. Hence, even though its AO basis counterpart shows sparse structure with only $\mathcal{O}(N^2)$ non-negligible elements, the integral tensor in the MO basis is dense. This results in the transformation having time complexity of $\mathcal{O}(N^5)$ and memory complexity of $\mathcal{O}(N^4)$, which basically precludes calculations for large systems.

The breakthrough in reducing computational complexity of MP2 calculations was made by Häser [3]. Noticing that the formula for energy correction is a Laplace transform

$$\mathcal{L}[f](x) = \int_0^\infty f(t)e^{-tx}\mathrm{d}t \tag{3}$$

of a constant function

$$\mathcal{L}[1](x) = \int_0^\infty e^{-tx}\mathrm{d}t = \frac{1}{x} \tag{4}$$

he proposed an alternative formulation of the energy correction

$$E_2 = -\int \sum_{ij}^{occ} \sum_{ab}^{virt} (ia|jb)[2(ia|jb) - (ib|ja)]e^{-t(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)}\mathrm{d}t. \tag{5}$$

Introducing energy-weighted molecular orbitals

$$|i'\rangle = e^{\frac{1}{2}t\epsilon_i}|i\rangle \qquad |a'\rangle = e^{-\frac{1}{2}t\epsilon_a}|a\rangle \tag{6}$$

the energy correction can be expressed as

$$E_2 = -\int \sum_{ij}^{occ} \sum_{ab}^{virt} (i'a'|j'b')[2(i'a'|j'b') - (i'b'|j'a')]\mathrm{d}t. \tag{7}$$

The main feature of the Laplace-transform-based MP2 energy expression is that it is invariant with respect to unitary transformations [3]. This allows for choosing a basis in which the complexity of the calculations is optimal. And since the complexity is directly related to the locality of the orbitals, the most natural choice is the atomic orbital basis set.

Introducing orbital-energy-weighted pseudo-density matrices

$$\underline{P}_{\mu'\mu} = \sum_i^{occ} C_{\mu'i}C_{\mu i}e^{\varepsilon_i t} \tag{8}$$

$$\overline{P}_{\nu'\nu} = \sum_a^{virt} C_{\nu'a}C_{\nu a}e^{-\varepsilon_a t} \tag{9}$$

where the C matrix represents the canonical molecular orbitals, and defining transformed AO orbitals as

$$|\underline{\mu}\rangle = \sum_\nu^N \underline{P}_{\mu\nu}|\nu\rangle \qquad |\overline{\mu}\rangle = \sum_\nu^N \overline{P}_{\mu\nu}|\nu\rangle \tag{10}$$

yields the workhorse of the Laplace-transform Atomic Orbital MP2 (LT-AO MP2)

method [3]

$$E_2 = -\int_0^\infty e_2(t)\mathrm{d}t = -\int_0^\infty \sum_{\mu\nu\lambda\sigma}^N (\underline{\mu}\overline{\nu}|\underline{\lambda}\overline{\sigma})[2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)]\mathrm{d}t. \qquad (11)$$

The formula can be efficiently implemented owing to the sparsity of the pseudo-density matrices, strictly following the sparsity pattern of the density matrix. Namely, the number of non-zero elements of a density matrix in an insulator is scaling linearly with the size of the system. Moreover, the non-zero elements may occur only in a relatively narrow diagonal band. The sparsity, complemented with standard pre-screening techniques results in a significant reduction in number of the transformed two-electron integrals which have to be calculated and stored in memory. This, in turn, allows for a relatively straightforward implementation of calculations having both time and memory complexity of $\mathcal{O}(N^2)$, where N stands for the number of orbitals. Furthermore, stricter prescreening should lead to the linearly scaling method. Low complexity should be achievable despite the fact that the integration in Equation (11) has to be performed numerically. (Analytical integration would bring us back to the canonical MP2 method.)

While the LT-AO MP2 formulation by Häser allows to efficiently calculate the energy correction, it has some drawbacks. The main one is that the definition of the pseudo-density matrices involves the canonical MO coefficients and orbital energies. This prevents complementing the LT-AO MP2 method with linearly-scaling, diagonalization-free Hartree-Fock calculations. To make things worse, implementation of even basic energy-derivative based molecular properties computations involves dense matrix of MO coefficients derivatives, significantly degrading the efficiency.

The issue has been recently solved by Súrjan, who proved [4] that the LT-AO MP2 pseudo-density matrices may be expressed purely in terms of the density and Fock matrices, without any reference to the MO coefficients or orbital energies. The resulting expressions [4]

$$\underline{\boldsymbol{P}}(t) = e^{t\boldsymbol{P}\boldsymbol{F}}\boldsymbol{P} \qquad \overline{\boldsymbol{P}}(t) = e^{-t\boldsymbol{Q}\boldsymbol{F}}\boldsymbol{Q} \qquad (12)$$

where $\boldsymbol{P}$ and $\boldsymbol{Q}$ denote the electron and hole density matrices, are the basis of density matrix-based LT-AO MP2 formalism.

This formulation, while keeping all the advantages of the original LT-AO MP2 method, allows for efficient calculations of the energy-derivatives based molecular properties like dipole moment or polarizability.

3 IMPLEMENTATION

Calculations of the LT-AO MP2 energy and dipole moment have been implemented in the Niedoida computational chemistry package [5].

The main aspect of the algorithm is an efficient representation of the transformed two-electron integral tensor. In our implementation, a four-level tree is used. The integral values are stored in contiguous memory regions attached to the leaves of the tree. Such a structure allows for optimal memory consumption, and at the same time maintains efficiency of the algebraic operations performed on the integrals. The process of tree construction involves multi-level prescreening procedure, which allows for skipping branches containing only negligibly small integrals.

To improve efficiency, the computational procedure was parallelized. We adopted Single Program Multiple Data (SPMD) model implemented by means of the Message Passing Interface (MPI). The scheduling of subtasks was implemented according to [6].

4 RESULTS

In order to verify applicability of the LT-AO MP2 method, we performed a series of calculations for selected model systems, for which canonical MP2 results are available. We assessed both the accuracy and computational complexity of the method.

The numerical integration was performed in a transformed coordinate system. This way it was possible to make the integration range finite and at the same time to reduce the variation of the integrand, allowing for good accuracy even with very low number of quadrature points. Specifically, all LT-AO MP2 calculations reported here were performed using the coordinate transformation defined by [7]

$$t(r) = \frac{\sum_k a_k r^k}{(1-r)^m}.$$ (13)

The optimal parameter values $m = 2$, $k = 3 \dots 6$, $a_k = 2.64, -6.09, 6.07, -2.7$ were found by minimizing the integration error of exponential function $e^{-\alpha x}$ with varying value of the α factor. As suggested in [7], we used the Euler-Maclaurin quadrature. All LT-AO MP2 calculations reported here employed 10-point grid.

The integral prescreening was performed according to the Häser protocol [3]. The threshold used for skipping negligible individual integral contributions was set to 10^{-8} a.u. The accuracy was calculated with respect to the reference results from canonical MP2 method as implemented in the GAMESS package [8].

The first system under study is a linear chain of water molecules. While not very realistic, such a model is well suited (and often used) for testing methods applicable to large systems. The other testing set consists of selected nucleobases and their pairs, which allows for direct evaluation of applicability of LT-AO MP2 to calculations for molecules of importance in biochemistry.

The results collected in Figure 1, Tables 1 and 2 show that even with modest integration grid size the error introduced by the LT-AO MP2 approximation is well within the bounds of so called chemical accuracy.
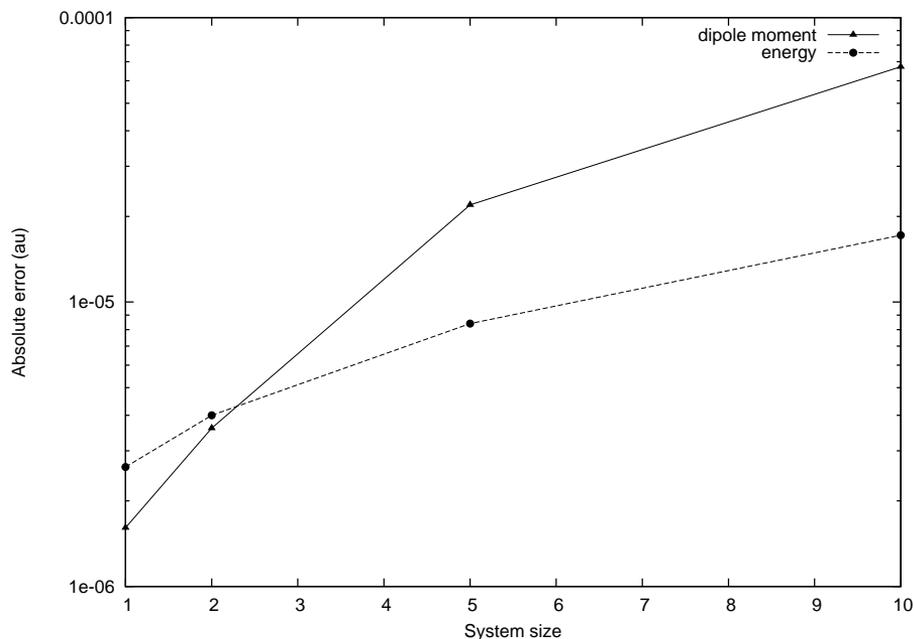
Fig. 1. Accuracy of LT-AO MP2 energy and dipole moment calculations for linear chains of water molecules in the 6-31G** basis set. System size denotes the number of molecules in chain.

Nucleobase	Error (10^{-3} a.u.)
Adenine (A)	1.8
Cytosine (C)	0.3
Guanine (G)	1.7
Thymine (T)	1.0
A–T	7.0
C–G	2.0

Table 1. Absolute error of LT-AO MP2/6-31G** energy calculations with respect to the canonical MP2 results

Nucleobase	Error (10^{-4} a.u.)		
	x	y	z
Adenine (A)	1.5	0.2	0.5
Cytosine (C)	4.0	0.0	1.6
Guanine (G)	4.7	4.8	4.1
Thymine (T)	3.3	2.8	0.9
A–T	2.0	1.8	4.0
C–G	5.7	0.4	0.7

Table 2. Absolute error of LT-AO MP2/6-31G** dipole moment calculations with respect to the canonical MP2 results

Basis set	Complexity	
	LT-AO MP2	Canonical MP2
STO-3G	2.81	5.54
6-31G**	2.60	5.44

Table 3. Overall computational complexity of energy calculations for linear chains of water molecules

Basis set	Complexity	
	LT-AO MP2	Canonical MP2
STO-3G	2.82	4.73
3-21G	2.49	4.80
6-31G**	2.65	–

Table 4. Overall computational complexity of dipole moment calculations for linear chains of water molecules. (For canonical MP2 calculations in the 6-31G** basis not enough data-points were collected to reliably fit the complexity.)

Efficiency data in Tables 3 and 4 show that the computational complexity of the LT-AO MP2 method is more than two orders lower with respect to the canonical counterpart. For actual computational time, the crossover in the case of linear chains of water molecules was observed at the system size of roughly 20 molecules.

5 CONCLUSIONS

We have shown that the density matrix-based quantum chemistry stands up to the expectations with respect to improved computational complexity, at the same time maintaining adequate accuracy.

While the new paradigm introduces some algorithmic issues which require further studies, it is already possible to obtain quadratic scaling with the system size at the MP2 level, and linear scaling seems achievable.

REFERENCES

[1] MØLLER, C.—PLESSET, M. S.: Note on an Approximation Treatment for Many-Electron Systems. Phys Rev, Vol. 46, 1934, p. 618.

[2] HELGAKER, T.—JORGENSEN, P.—OLSEN, J.: Molecular Electronic-Structure Theory. Wiley 2000.

[3] HÄSER, M.: Moller-Plesset (MP2) Perturbation Theory for Large Molecules. Theor Chim Acta, Vol. 87, 1993, p. 147.

[4] SURJAN, P. R.: The MP2 Energy As a Functional of the Hartree-Fock Density Matrix. Chem Phys Lett, Vol. 406, 2005, p. 318.

[5] MAZUR, G.—MAKOWSKI, M.: Development and Optimization of Computational Chemistry Algorithms. Comput Informat, Vol. 28, 2009, p. 115.

[6] MAZUR, G.—MAKOWSKI, M.—BRELA, M.: Effective Resource Allocation in Parallel Quantum-Chemical Calculations. Comput Informat, Vol. 30, 2011, p. 761.

[7] AYALA, P. Y.—SCUSERIA, G. E.: Linear Scaling Second-order Moller-Plesset Theory in the Atomic Orbital Basis for Large Molecular Systems. Vol. 110, 1999, No. 8, p. 3660.

[8] SCHMIDT, M. W.—BALDRIDGE, K. K.—BOATZ, J. A.—ELBERT, S. T.—GORDON, M. S.—JENSEN, J. H.—KOSEKI, S.—MATSUNAGA, N.—NGUYEN, K. A.—SU, S.—WINDUS, T. L.—DUPUIS, M.—MONTGOMERY, J. A.: General Atomic and Molecular Electronic Structure System. J Comput Chem, Vol. 14, 1993, p. 1347.

Grzegorz MAZUR is an Assistant Professor at the Department of Computational Methods in Chemistry of Jagiellonian University in Cracow (Poland). He received his Ph. D. in chemistry in 2001 from the same university. His current research interests include theoretical description of the excited states properties and optimization of quantum chemistry algorithms.

Marcin MAKOWSKI is an Assistant Professor at the Department of Theoretical Chemistry of Jagiellonian University in Cracow (Poland). He received his M. Sc. degree in 2001 and his Ph. D. in 2004, both in chemistry, from the same university, and his B. Sc. degree in computer science from the University of Mining and Metallurgy in Cracow in 2004. His current research interests include theoretical molecular spectroscopy, linear scaling methods in quantum chemistry and optimization of quantum chemistry algorithms. He participates in several research projects related with the development of theoretical chemistry formalisms and numerical methodologies that allow to calculate electronic structure of large molecular systems efficiently.

Jakub SUMERA has recently received his M. Sc. in chemistry. His field of interest includes theoretical and computational chemistry.

Krzysztof KOWALCZYK has recently received his M. Sc. in chemistry. His field of interest includes theoretical and computational chemistry.