Stochastic effective core potentials, improving efficiency using a spin-dependent core definition

Jonas Feldt, Antoine Bienvenu, Roland Assaraf

Stochastic effective core potentials, improving efficiency using a spin-dependent core definition

Jonas Feldt,[a*] Antoine Bienvenu[a], and Roland Assaraf[a]

Numerically cheap single-core subsamplings have been used to build improved estimators for molecular properties in the variational Monte Carlo framework[1]. The resulting estimators depend only on the valence electron positions and can be thought of as an exact effective core potential for the total energy. We are proposing a spin-dependent core definition which enables exploiting these single-core subsamplings (or sidewalks) not only to decrease the variance of the estimators but also to restrict the main variational Monte Carlo dynamics to the valence region. This results mainly in a simplification of the algorithm and additionally in a gain in efficiency as illustrated on alkane chains and silicon clusters. An evaluation of the efficiency on transition metal systems is done using cobalt clusters, a gain of up to two orders of magnitude is achieved compared to a standard all-electron calculation.

1 Introduction

Quantum Monte Carlo (QMC) methods employ a stochastic approach to solve the Schrödinger equation. Freedom in the choice of the wave function allows to treat equally dynamic and static correlation which is exploited for the investigation of materials and excited states.[2] The N^{3-4} scaling of the computational cost with the system size N is very favourable compared to deterministic quantum chemistry methods. Among recent developments, the cost of multideterminant expansions and optimizing has been greatly reduced[3,4] to the point where the optimization of a geometry and all parameters of the wave function scale the same as the computation of the total energy[5].

One of the remaining challenges to establish QMC methods as highly accurate and cost effective methods is the steep scaling with the atomic number Z. The consequence is that empirical effective core potentials (ECPs) are widely used which introduce a bias that cannot be easily judged a priori. For example the Burkatzki-Filippi-Dolg potentials have been parametrized for Hartree-Fock and do not take into account the correlation energy.[6] It has been demonstrated that this bias is even larger for excited states properties.[7] Furthermore, the unfavourable scaling with the effective nuclear charge Z_{eff} remains, $Z_{\text{eff}}^{6.5}$ for the forces in diffusion Monte Carlo (DMC).[8]

Developments addressing this challenge have been mostly focusing on improving the sampling and reducing the correlation factor. For instance in variational Monte Carlo (VMC) the dynamic can be carried out very efficiently in spherical coordinates.[9] In DMC different grids based on a spatial discretization can be used for core and valence electrons. This enables to perform small moves adapted to the core electrons and large moves adapted to the valence electrons. This results in an acceleration of about one order of magnitude for very heavy elements ($Z = 118$) and achieves a scaling of Z^5 for all-electron calculations .[10] Adapting the moves to the core and valence electrons can be done also without discretizing the space i.e. in the usual framework of the (overdamped) Langevin Dynamics (drift and diffusion process) performed in Diffusion Monte Carlo methods and Variational Monte Carlo. For that purpose two time steps are introduced an optimized (one small one for the core electrons and a large for the valence electrons)[11], in this last reference the correlation factor was reduced by factor 2–4 in Variational Monte Carlo (up to the Neon atom).

Recently, a core-subsampling approach (sidewalks on the core electrons) was introduced to reduce not only the correlation factor but also the fluctuations coming from the core region[1]. The cost of these sidewalks scales linearly with the system size N if we use the locality of the information in the core region (e.g. the atomic orbitals are highly local). This cost is negligible compared to the $\mathcal{O}(N^3)$ cost for the main walk.

Such an algorithm is equivalent to an on-the-fly construction of an *exact* effective core potential, because the resulting estimator is independent of the positions of the core electrons. We are proposing to use these sidewalks or subsamplings not only to remove the fluctuations coming from the core region but also alleviate the main walk to treat only the valence region. This can be done without loss of ergodicity thanks to a spin-dependent definition of the core-valence separation[11] and the use of the final configuration of the sidewalk to advance the main walk. These updates improve to some degree the variance but mainly the ergodicity, the correlation factor within the main walk and the computational time.

Limiting the main walk to the valence electrons results in a simplified algorithm,also because there is no more codependency of the parameters of this method (time steps and size of the sidewalks) in the core and valence regions. In particular the optimal size of the sidewalks can be now obtained automatically for any cluster using a single atom calculation (eqn 4) avoiding human time consuming optimizations. Besides efficiency comparisons using alkanes chains and silicon clusters, we apply the method on cobalt clusters to evaluate its efficiency on a transition metal system.

2 Algorithm

With X being a random variable, for example the potential or the local energy, we are partitioning the variance based on the variance decomposition theorem

$$V(X) = \mathbb{E}(V(X|\Omega)) + V(\mathbb{E}(X|\Omega)) \tag{1}$$

[a] *Laboratoire de Chimie Théorique - UMR7616, Sorbonne Université & CNRS, 75005 Paris, France*

* E-mail: jfeldt.theochem@gmail.com

into a contribution from the core electrons $\mathbb{E}(V(X|\Omega))$ and the remaining part, which stems from the valence electrons.[1] We remind that $\mathbb{E}(X)$ stands for the expectation value of X (interpreted as a statistical average in the Monte Carlo framework) and V stands for the variance $V(X) \equiv \mathbb{E}(X^2) - \mathbb{E}(X)^2$ which measures the statistical fluctuations.

The condition Ω is a constraint in real space which allows the core electrons alone to move within their core region while the valence electrons are frozen. Performing an average under this constraint produces the conditional expectation value $\mathbb{E}(X|\Omega)$ and the conditional variance $V(X|\Omega) = \mathbb{E}(X^2|\Omega) - \mathbb{E}(X|\Omega)^2$. $\mathbb{E}(X|\Omega)$ is an improved estimator of X because it has the same expectation value (thanks to the law of total expectation $\mathbb{E}(\mathbb{E}(X|\Omega)) = \mathbb{E}(X)$) and a reduced variance $V(\mathbb{E}(X|\Omega) < V(X)$. As a by-product, using the estimator $\mathbb{E}(X|\Omega)$ does not change the accuracy. Note that $\mathbb{E}(X|\Omega)$ does not depend on the core electrons positions since they are averaged, it depends only on the valence positions.

$\Omega_c^{(i)}$ being the constraint allowing only the core electrons of the atom i to move, the improved estimator

$$\tilde{X} = X + \lambda \sum_i \left(\mathbb{E}(X|\Omega_c^{(i)}) - X \right) \tag{2}$$

eliminates the fluctuations of the core electrons completely when the core regions are independent[1] (while not modifying the expectation value since we add a zero-expectation value term). The coefficient λ can be determined to minimize the variance [*]. The conditional expectation values $\mathbb{E}(X|\Omega_c^{(i)})$ are evaluated by sidewalks with a number of M_s steps on the core electrons, using the ergodic theorem

$$\mathbb{E}(X|\Omega_c^i) = \lim_{M_s \to \infty} \frac{1}{M_s} \sum_{k=1}^{M_S} X^k \tag{3}$$

where X^k is the value of X for the k^{th} step of the sidewalk.

One natural idea would be to use the sidewalks not only to lower the variance of the estimators but also to advance the main walk, moving the core electrons of the main walker to the last positions of the core sidewalk. With such updating scheme the sidewalk can be considered as a part of the main walk and as such can be called a "subwalk". In this context $\Omega_c^{(i)}$ would freeze beside the valence electrons the first $i-1$ cores in their new configuration and all remaining cores (excepted i) in their old configuration. This approach should make it possible to restrict the main walk to only the valence electrons. In practice the process would not be ergodic with the definition of the core region in Ref. 1. In this reference the core region was the largest nucleus-centred ball containing n_c closest electrons (regardless their spin) to the nucleus. If there is no exchange between the valence and the core region the total spin of the core region is frozen. Hence, we will apply this idea with a different definition of the core and valence regions.

We introduce a spin-dependent core constraint, i.e. two core regions, one for the α electrons and one for the β electrons. If we order the α electrons by their distance to the nucleus, the first α valence electron is on the nucleus-centered sphere which defines the boundary of the α core region (see Fig. 1). Of course the same definition applies to the β core region.

Now we define equivalently the α (respectively β) valence region to be outside the (nucleus-centred) ball with the radius defined by the last α (respectively β) core electron (see Fig. 1). This core-valence separation definition is equivalent to the one introduced in reference[11].

Core and valence regions overlap because the space between the last core and the first valence electron can be explored by both core and valence electrons. With these definitions performing sidewalks in the core regions and a main walk restricted to the valence region should be now ergodic, because these regions have different radii which can evolve and exchange, ensuring for example the sampling of the spin in a given volume of the space. Note that with this new definition the core electrons have always the same indices (in particular we do not need anymore to reorder the electrons at the beginning of each sidewalk).

In summary, one iteration of the algorithm consists of the following steps:

1. Independent sidewalks for all cores.
2. Update of the full system after each core sidewalk.
3. Main walk only for the valence electrons .
4. Update of the full system with new valence positions.
5. Computation of the improved estimator $\tilde{X}$.

The update of the configuration after each core sidewalk is a small modification of the algorithm presented in Ref. 1 but it has several consequences. The first is that we do not need to move the core electrons in the main walk as stated in step 3. This implies a further simplification of the dynamics which consisted in moving one electron at a time using two drift and diffusion processes with different time steps (a small one for the core region and a large one for the valence region) 1. Only one (large) time step for the valence region can be now used, avoiding the small time step move which was only efficient in the core region.

In this modified dynamics the treatment of core and valence electrons is now similar except that we do many moves per iteration for the core electrons and only one move for the valence electrons. With the new algorithm the estimator (eqn (2)) is formally the same except that $\Omega_c^{(i)}$ for two different indices i represents different frozen configurations (i.e. the core electrons are not the same after one update). Note that if different core regions are independent the core updates do not modify this estimator and its variance, only the different definition of the core does.

The combination of all sidewalks (including updates) constitutes in itself an efficient (i.e. quickly decorrelating) move for all electrons. Hence the sidewalks are not only used to reduce the variance but also correlation.

[*] Writing $\tilde{X} = X + \lambda C$, the variance can be expanded as $V(\tilde{X}) = V(X) + 2\lambda \ \mathrm{cov}(X,C) + \lambda^2 V(C)$ which is a quadratic function of λ easy to minimize after computing the parameters (two variances and one covariance).

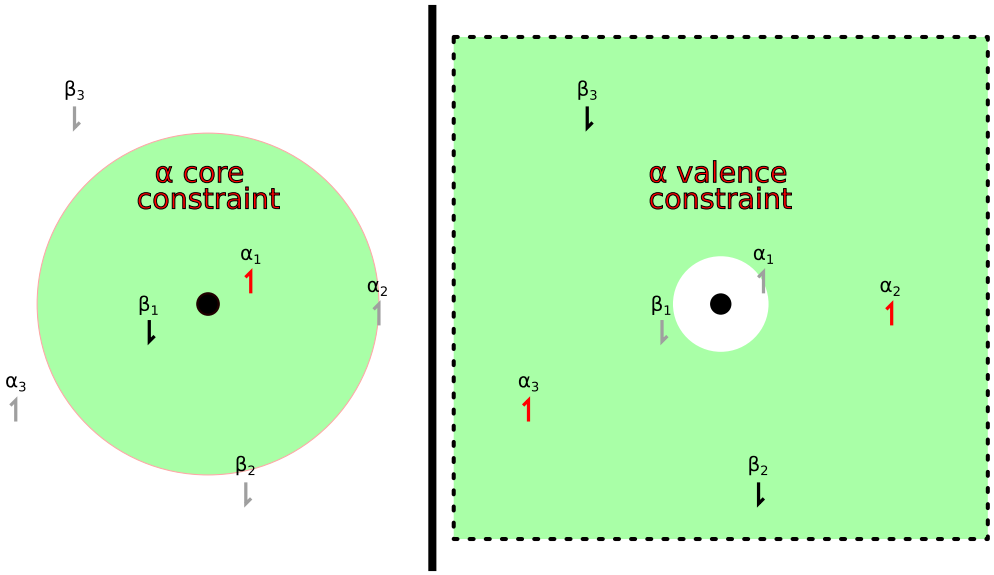

Fig. 1 Representation of the α core and valence constraint for a nucleus surrounded by 3 α and 3 β electrons. The electrons are labeled according to their distance to the nucleus. The constrained electrons are shown in red and can move freely within the green area, left for the core electrons and right for the two α valence electrons. The frozen electrons within the core sidewalk (left) and the valence main walk (right) are shown in gray. The valence space is infinite but here only shown within the dashed frame. Equivalent constraints apply to the β electrons.

3 Numerical Results

Simulations have been carried out for alkane chains and silicon clusters to compare the performance of the improved algorithm in this work which uses updates and a spin-dependent core definition with the previous algorithm[1] without updates and a purely spatial core definition. Due to the updates the estimator in eqn (2) is different from the previous estimator. The computational cost can be compared using the expression $\zeta = Vct$ with the variance V, the correlation factor c and the computational time of a single step t. The gain in computational efficiency G is defined as the inverse reduction of the computational cost. We carried out simulations for small model systems (CH_4 and Si_1) fitting the convergence of the variance with M_s to estimate the reduction for very long sidewalks. This reduction in the limit of large M_s is transferable to large systems where the optimal M_s^* itself is increasingly large. The correlation factor is also transferable in this limit[1]. Therefore, we can estimate the gain in the computational efficiency (i.e. reduction of the cost) in the asymptotic limit of many atoms G^∞ as 38 (alkanes) and 774 (Si clusters) compared to a regular all-electron VMC simulation. Compared to the previous algorithm it is an improvement by a factor of 3.7 (alkanes) and 3.9 (Si clusters).

We can further analyze the gain G^∞ by separating according to the definition of the computational cost into contributions due to the variance G_V^∞, the correlation factor G_c^∞ and the computational time G_t^∞. We begin with the estimator eqn (2) with a fixed value of $\lambda = 1$. The correlation factor was already small with the previous algorithm (1.8 for carbon and 1.5 for silicon), consequently we observe only an additional gain of 10% (alkanes) and 19% (Si clusters). This supports an improvement in ergodicity coming from the core updates but also, the new optimal time step for the valence dynamic is larger especially for the silicon clusters. However, we observe also a small loss in the variance by about 10% (alkanes) and 4% (silicons). To understand it, we carried out additional simulations without updates but with a spin-dependent core definition. Because we observe the same loss we attribute this loss to the new core definition. It turns out that contrary to the previous core definition the value $\lambda = 1$ is not optimal for the variance. This is a signature of a diminished core-valence separation coming from the new definition of these regions. Indeed $\lambda = 1$ is the optimal value to lower the variance when the valence and the core regions are independent, since in this limit eqn (2) cancels fully the fluctuations coming from the cores of the molecule. Here optimizing λ recovers the loss in the variance and we even observe a small gain of about 26% (alkanes) and 11% (silicons). This reduced core separation should not be too surprising, since in rare cases a valence electron of one spin can be closer to the nucleus than a core electron of the opposite spin. Optimizing λ however decreases the correlation factor by 8% (alkanes) or 21% (silicons) in comparison to Ref. 1. Nevertheless, this is compensated for by the gain in the variance. We will focus in the following on the estimator with optimal λ.

The appreciable effect on the overall gain is in the computational time t which is improved by a factor of 2.9 (alkanes) and 3.4 (Si clusters). The cubic scaling with the system size (in the asymptotic limit of a large number of atoms) is the same for the new algorithm but the prefactor is reduced. On one hand, the computational time for the main walk is reduced because only the valence electrons need to be taken into account. On the other hand, an additional cost (scaling $\mathcal{O}(N^3)$) is added to the sidewalk which stems from the Sherman-Morrison update of the full configuration at the end of each sidewalk. Nevertheless, this is more efficient because we require less updates for more electrons: instead of one update for each single electron in the main walk we are carrying out only one update for all electrons of a core

Fig. 2 The gain G for alkane chains comparing different algorithms.

Fig. 3 The gain G for silicon clusters comparing different algorithms.

at once. The formulas for the update of n_c electrons at once are shown in Appendix A. Also, we are exploiting the locality within the core subsystems for an efficient update and the $O(N^2)$ cost for the update of one core comes with a very low prefactor.

Next, we are looking at the results of simulations for alkanes up to 40 carbon atoms shown in Fig. 2 and for silicon clusters up to 24 atoms in Fig. 3 (red lines). This corresponds to about 350 electrons for the largest systems. The maximal gain for a given system is obtained by determining the optimal number of steps M_s^* in the core sidewalk which is a balance between the reduction of the variance and the additional computational cost. A simple formula to determine M_s^* is shown in eqn (4) in the computational details (see Sec. 4). For comparison we present also the results of the previous algorithm [1] (blue lines). It can be seen that the gain of the new algorithm is always at least as good as the previous gain. We observe similar results for alkane chains up to 20 carbon atoms and silicon clusters up to 5 atoms. For larger systems one can see an increasingly larger gain with the improved algorithm. For the largest systems studied here we gain a factor of 1.9 ($C_{40}H_{82}$) and 2.0 (Si_{24}).

The improvement for the alkanes comes mostly from the reduced computational time t and additionally for more than 30 carbon atoms the updating scheme reduces the variance (in the optimal regime). This suggests that the convergence towards the asymptotic limit of many atoms is accelerated. For the silicon clusters the situation is different. Because of the large number of core electrons $n_c = 10$, the number of electrons in the main walk is strongly reduced (and consequently the computational time t_0, see eqn (4)). Therefore, the optimal sidewalk length M_s^* is shorter which results in a reduced optimal gain in the variance multiplied by a factor of 0.35 (one atom) to 0.7 (24 atoms). With increasing system size this gain will approach its asymptotic value. The reduced optimal gain in the variance is counterbalanced by an even larger gain in the computational time t of up to 4.1 times which leads to an overall improvement of a factor 2 for 24 atoms.

Last, we are evaluating the efficiency of the new algorithm (compared to regular all-electron calculations) for transition metals on the example of clusters of cobalt atoms (hexagonal $P6_3/mmc$ space group [12]). In the asymptotic limit of large clusters the gain G^∞ is given by about 2270 including a gain in the variance of about 130. The efficiency for clusters consisting of up to 20 cobalt atoms is shown in Fig. 4. The gain is still far from its asymptotic value. Nevertheless, we can obtain already an improvement of the efficiency by one order of magnitude for only four cobalt atoms and by a factor of 64 for the largest system Co_{20}. The efficiency for all three systems studied here is compared in Table 1 both for the asymptotic gain G_∞ and for a gain for twenty atoms of either carbon, silicon or cobalt G_{20}. The gain in the asymptotic limit is increasing drastically with Z by two orders of magnitudes going from $Z = 6$ to 27. The gain for a medium-sized system of 20 atoms increases as well with Z but at a slower rate (a factor 6.4 from $Z = 6$ to $Z = 27$).

Fig. 4 The gain G for cobalt clusters comparing with a regular all-electron calculation.

Table 1 The gain in efficiency (compared to regular all-electron calculations) for the asymptotic limit of many atoms G_∞ and twenty atoms G_{20} and the atomic number Z as well as the number of core electrons n_c for the alkanes, silicon and cobalt clusters.

	C	Si	Co
Z	6	14	27
n_c	2	10	18
G_{20}	10	40	64
G_∞	38	774	2270

4 Computational Details

The main walk and the sidewalks are carried out with a drift and Brownian diffusion (overdamped Langevin process). Two time steps are used, τ_c and τ, for the core and the valence region respectively. These two parameters and the number of iterations M_s of any one-core sidewalk have to be optimized for a minimal cost $\zeta = Vct$ (we remind that V is the variance, c the correlation factor and t the computational time for one iteration). These three parameters can be derived from a small number of simulations on very small systems (e.g. isolated atoms or CH_4 for alkanes). The time steps are directly transferable to larger systems, and the variance gain as a function of M_s is also transferable [1]. For a given chemical element the reduction of the variance is a linear function of $1/M_s$ leading to $r_V = \tilde{V}/V = r_\infty + a/M_s$. These transferability considerations allow defining a simple formula for the optimal choice for M_s (see appendix C of Ref. 1)

$$M_s^* = \sqrt{\frac{a}{r_\infty}\frac{t_0}{t_c}} \tag{4}$$

where t_0 is the computational time of one iteration excluding the sidewalks, i.e. the time to do a single walk on all the valence electrons and all the updates (it is scaling as $O(N^3)$). t_c is the CPU time for all the core sidewalks (of the same element) with $M_s = 1$ (scaling as $O(N)$). This formula is valid with the assumption that the correlation factor in the main walk (when sampling the improved estimator) does not depend on M_s, which is obviously true for sufficiently large M_s. Without the updating scheme the correlation factor converges too slowly [1] to apply eqn (4) for small systems and tedious optimizations of M_s have to be performed. We have checked that with the updating scheme this is no longer the case and eqn (4) can be applied to a small number of atoms (in all our calculations $M_s^* > 15$, see figures 5 and 6).

The optimal time steps are shown in Table 2. We found that the spin-dependent core definition has a negligible effect on the optimal value of τ_c, the latter can be taken directly from Ref. 1. The optimal value of τ is larger, which is not surprising since it

Table 2 The optimal time steps for core and valence for alkane chains and silicon clusters

	Alkanes	Silicon clusters
τ_c	0.004	0.007
τ	0.8	1.8

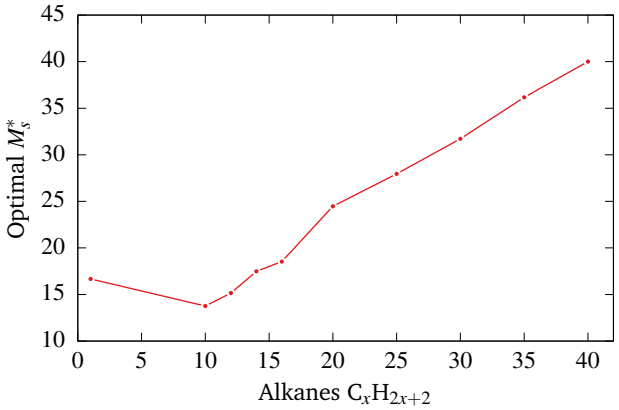

Fig. 5 The optimal sidewalk length M_s^* for alkane chains obtained from Eq. (4) .

is now exclusively used for the valence electrons. The computational cost to move the valence electrons is below 10% of the total CPU time for silicon clusters even for Si_{32}. This is because most of the variance comes from the core region which has to be consequently sampled much more extensively. With these time steps we found an acceptance probability for the valence move of about 0.5 for silicon and 0.66 for carbon.

The value of M_s^* for the alkanes are shown in Fig. 5 and for the silicon clusters in Fig. 6. One can observe that the linear scaling regime of M_s^* is quickly reached for about 70 electrons for both alkanes and silicon clusters at which point a simple linear extrapolation can be used for even larger systems. Additionally, the simplicity of eqn (4) allows to determine the optimal value M_s^* also for smaller systems where one cannot rely on the cubic scaling of t_c.

The wave function and the Jastrow factor have been taken from Ref. 1 and have been generated based on an SCF calculation carried out with Quantum Package.[13] The wave function for the cobalt clusters has been generated in the same manner. The very simple Jastrow factor ensures the electron-electron cusp condition. A Slater atomic orbital basis set[14] has been used with a TZP basis for the alkanes and SZ for the silicon and cobalt clusters which is expanded by a large sum of Gaussian functions for Quantum Package.

5 Conclusion

In this paper we are extending the stochastic ECP approach originally proposed to improve the estimator (removing fluctuations coming from core region using sidewalks) to the main dynamics itself: here the main walk moves only the valence electrons as we would expect in a complete ECP formalism, while the sidewalks focus exclusively on the core electrons. Key has been to update the system at the end of each core sidewalk: to maintain (and even improve) ergodicity we replaced the purely spatial core definition with a spin-dependent one. Compared to the previous algorithm the number of parameters (time steps) is reduced and the determination of the length of the sidewalk is extracted from calculations on a single atom which avoid a tedious optimization for large systems. The observed additional gain in efficiency (2-4) is growing with the system size as we observe an accelerated convergence towards the asymptotic limit of large systems. Tests include a transition metal (cobalt clusters). Large gains are observed with respect to traditional all-electron calculations (a factor 64 for 20 cobalt atoms) but are still small compared to the asymptotic gain (a factor 2270 for a very large cluster of cobalt atoms). This suggests that there is a large room of improvement, a keypoint of future developments would be to further improve the efficiency of the core sidewalks for the gain to converge much faster to the asymptotic limit. A natural idea would be for example to adapt this method to many shells. One of the most interesting perspective is the generalization to the diffusion Monte Carlo approach thanks to the efficient small time step dynamics in the core region and the following efficient updating of the full configuration of the electrons.

Fig. 6 The optimal sidewalk length M_s^* for silicon clusters obtained from Eq. (4).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

J. F. acknowledges the Deutsche Forschungsgemeinschaft (DFG) for financial support (Grant FE 1898/1-1).

A Updating the logarithmic gradient of the determinants

In Ref. 1 (appendix A) we described a method to update the determinants and its derivatives when a few electrons are moved, using a matrix D representing the logarithmic gradient of the Slater matrix with respect to the atomic orbitals coefficients. Now we also have to update the matrix D at the end of the sidewalk when a few (core) electrons have moved.

Let us first remind how to update the Slater determinant Φ when a few electrons are moved

$$\Phi = \det(XC) \tag{5}$$

where X is the $N \times p$ matrix of atomic orbitals and C the $p \times N$ matrix of coefficients, N and p being respectively the number of electrons and molecular orbitals. The drift, the local energy and many other possible quantities depending on the electron positions involve logarithmic derivatives of Φ with respect to X.

$$\partial_\lambda \ln \Phi(X) = \mathrm{tr}(D\,\partial_\lambda X) \tag{6}$$

where

$$D \equiv C(XC)^{-1} \tag{7}$$

eqn (6) can be seen as the application of the chain rule involving D, the rectangular matrix representing the logarithmic gradient of Φ with respect to X. Note that eqn (6) must be valid for any parameter λ and can thus be seen as the definition of D. If we are going to move only a few electrons within a core sidewalk, X is modified in X' which differs from X by a few lines, $D(X)$ must be replaced by $D' = D(X')$ with efficient formulas.

First, we define the operator P which applied on the left selects those lines, PX' are the lines which may differ from the lines of PX. We also define the operator Q^T which applied on the right of PX or PX' removes the zero columns (atomic orbitals which are zero because the electrons selected by P are out of range). P and Q can be written in terms of rectangular matrices containing zeroes and ones[4]. $\bar{X}$ is the matrix of atomic orbitals within a subsystem and a submatrix of X'

$$\bar{X} \equiv PX'Q^T.$$

We obtain the changed Slater determinant $\Phi(X')$

$$\begin{aligned}\Phi(X') &= \det(XC)\det(PX'C(XC)^{-1}P^T)\\ &= \det(XC)\det(PX'Q^TQC(XC)^{-1}P^T)\\ &= \det(XC)\det(\bar{X}\bar{C}) \qquad (8)\end{aligned}$$

where $\bar{C}$ is a submatrix of $D = C(XC)^{-1}$.

Eqn (8) performs an update of the determinant of a product of two matrices, using the determinant of a reduced matrix. The last expression of Φ depends on X' and we can write

$$X = P^TPX + (1-P^TP)X' \qquad (9)$$

where P^TP represents the projector on the space spanned by the lines which have been modified. We note that the final expression should not depend on PX. Introducing

$$\begin{aligned}\bar{\alpha} &\equiv (\bar{X}\bar{C})^{-1}\\ \bar{D} &\equiv \bar{C}(\bar{X}\bar{C})^{-1}\end{aligned}$$

the logarithmic derivative of eqn (8) is

$$\begin{aligned}\partial_\lambda \ln\Phi(X') &= \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(\bar{D}\,\partial_\lambda\bar{X}) + \mathrm{tr}(\bar{\alpha}\bar{X}\,\partial_\lambda\bar{C})\\ &= \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(\bar{D}\,\partial_\lambda\bar{X})\\ &- \mathrm{tr}(\bar{\alpha}\bar{X}QD\,\partial_\lambda X\,DP^T) \qquad (10)\end{aligned}$$

which should also be $\mathrm{tr}(D'\,\partial_\lambda X')$ so that D' can be obtained by identification. First using the cyclic property of the trace we obtain

$$\begin{aligned}\partial_\lambda \ln\Phi(X') = &\ \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(Q^T\bar{D}P\,\partial_\lambda X') -\\ &\ \mathrm{tr}(DP^T\bar{\alpha}\bar{X}QD\,\partial_\lambda X). \qquad (11)\end{aligned}$$

Since this expression should not depend on $P\,\partial_\lambda X$ we have

$$\begin{aligned}\partial_\lambda \ln\Phi(X') = &\ \mathrm{tr}(D\,\partial_\lambda X') + \mathrm{tr}(Q^T\bar{D}P\,\partial_\lambda X') -\\ &\ \mathrm{tr}(DP^T\bar{\alpha}\bar{X}QD\,\partial_\lambda X') \qquad (12)\end{aligned}$$

which can be also found using eqn (9). By identification we finally find the (Sherman Morrison) formula to update the logarithmic gradient

$$D' = D - DP^T\bar{\alpha}\bar{X}QD + Q^T\bar{D}P. \qquad (13)$$

In the second term we select a few columns (DP^T) and lines (QD) of D and build the product with the small rectangular subsystem matrix $\bar{\alpha}\bar{X}$. The cost for the update comes with a very small prefactor because we are exploiting the locality of the subsystem (thanks to the matrix Q which selects a few orbitals). A bit more efficient formula can obtained as follows

$$\begin{aligned}QD' &= QD - \bar{D}\bar{X}QD + \bar{D}P\\ QD'P^T &= QDP^T - \bar{D}\bar{X}QDP^T + \bar{D}.\end{aligned}$$

Using that $\alpha\bar{X}QDP^T = P^TP$ we can decompose D' as follows

$$D' = D(1-P^TP) - DP^T\bar{\alpha}\bar{X}QD(1-P^TP) + Q^T\bar{D}P \qquad (14)$$

B Updating a Jastrow-Slater Function

We define a Jastrow-Slater function

$$\Psi = J\Phi = e^U\Phi \qquad (15)$$

where Φ is a Slater determinant and J the Jastrow factor. We need to update U when a few electrons move but also the spatial derivatives of U which are involved in the drift. This is trivial since the new value of U can be written as

$$U' = U + (U' - U) \qquad (16)$$

If U is a sum of pairwise interactions the only term to be computed is $U'-U$ which involves only a few pairs of electrons. The computational cost scales as $O(N)$ which can be reduced to $O(1)$ if the pairwise interactions are short range. This was the case for the silicon and cobalt clusters because we used a cutoff ($r_{\mathrm{cutoff}} = 7.07$ au). The same discussion holds for the spatial derivatives of U.

References

1 J. Feldt and R. Assaraf, *J. Chem. Theory Comput.*, 2021, **17**, 1380–1389.

2 J. Feldt and C. Filippi, in *Quantum Chemistry and Dynamics of Excited States*, John Wiley & Sons, Ltd, 2020, pp. 247–275.

3 B. K. Clark, M. A. Morales, J. McMinis, J. Kim and G. E. Scuseria, *J. Chem. Phys.*, 2011, **135**, 244105.

4 C. Filippi, R. Assaraf and S. Moroni, *J. Chem. Phys.*, 2016, **144**, 194105.

5 R. Assaraf, S. Moroni and C. Filippi, *J. Chem. Theory Comput.*, 2017, **13**, 5273–5281.

6 M. Burkatzki, C. Filippi and M. Dolg, *J. Chem. Phys.*, 2007,

126, 234105.
7 A. Scemama, M. Caffarel, A. Benali, D. Jacquemin and P.-F. Loos, *Results Chem.*, 2019, **1**, 100002.
8 J. Tiihonen, R. C. Clay and J. T. Krogel, *J. Chem. Phys.*, 2021, **154**, 204111.
9 C. J. Umrigar, *Phys. Rev. Lett.*, 1993, **71**, 408–411.
10 K. Nakano, R. Maezono and S. Sorella, *Phys. Rev. B*, 2020, **101**, 155106.
11 D. Bressanini and P. J. Reynolds, *J. Chem. Phys.*, 1999, **111**, 6180–6189.
12 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. a. Persson, *APL Mater.*, 2013, **1**, 011002.
13 Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Véril, M. Caffarel, P.-F. Loos, E. Giner and A. Scemama, *J. Chem. Theory Comput.*, 2019, **15**, 3591–3609.
14 E. V. Lenthe and E. J. Baerends, *J. Comput. Chem.*, 2003, **24**, 1142–1156.