Accurate and Fast Geometry Optimization with Time Estimation and Method Switching

Satoshi Imamura,* Akihiko Kasagi, and Eiji Yoshida

Computing Laboratory, Fujitsu Limited

E-mail: s-imamura@fujitsu.com

Abstract

Geometry optimization is an important task in quantum chemical calculations to analyze the characteristics of molecules. A top concern on it is a long execution time because time-consuming energy and gradient calculations are repeated across several to tens of steps. In this work, we present a scheme to estimate the execution times of geometry optimization of a target molecule at different accuracy levels (i.e., the combinations of ab initio methods and basis sets). It enables to identify the accuracy levels where geometry optimization will finish in an acceptable time. In addition, we propose a gradient-based method switching (GMS) technique that reduces the execution time by dynamically switching multiple methods during geometry optimization. Our evaluation using 46 molecules in total shows that the geometry optimization times at 20 accuracy levels are estimated with a mean error of 29.5%, and GMS reduces the execution time by up to 42.7% without affecting the accuracy of geometry optimization.

1 Introduction

Geometry optimization is a process to find the atomic coordinates that minimize the energy of a target molecule. It is an important basis task in quantum chemical calculations because the optimized geometries are used to analyze the molecular characteristics and structures.^{1–3} In geometry optimization, a stationary point on a potential energy surface (PES) is explored by iteratively calculating the energy and gradients of a molecule while changing its atomic coordinates step by step.

With the Taylor series, the energy at a point x on a PES, E(x), is represented in a quadratic approximation with respect to a near point x_0 ,

$$E(x) = E(x_0) + \boldsymbol{G}^T(x_0)\Delta x + \frac{1}{2}\Delta x^T \boldsymbol{H}(x_0)\Delta x,$$

where $G(x_0)$ is the gradient vector (dE/dx) at x_0 , $\Delta x = x - x_0$, and $H(x_0)$ is the Hessian matrix (d^2E/dx^2) at x_0 . By differentiating the equation with respect to coordinates, the gradients at x, G(x), is represented in a quadratic approximation as

$$\boldsymbol{G}(x) = \boldsymbol{G}(x_0) + \boldsymbol{H}(x_0)\Delta x.$$

As G(x) becomes zero at a stationary point on a PES, the displacement to the stationary point, Δx , is given by

$$\Delta x = -\boldsymbol{H}(x_0)^{-1}\boldsymbol{G}(x_0).$$

Solving this equation is called the Newton-Raphson step, which is a core part of geometry optimization. $G(x_0)$ is obtained by differentiating $E(x_0)$ with respect to the coordinates. On the other hand, $H(x_0)$, which is hard to calculate exactly, is commonly approximated with a quasi-Newton method such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.⁴

In every step of geometry optimization, energy and gradient calculations at current coordinates are performed. For both of them, a wide variety of ab initio methods with different accuracy and computational costs are available, such as Hartree-Fock method (HF),⁵ density function theory (DFT),⁶ Møller-Plesset perturbation theory (MP),⁷ configuration interaction theory (CI),⁸ and coupled cluster theory (CC).⁹ There is basically a trade-off between an accuracy and computational cost among them, which means that more accurate methods require higher computational costs.

To improve the efficiency of geometry optimization, various approaches have been proposed. Chaudhuri and Freed extended the *improved virtual orbital-complete active space* configuration interaction (IVO-CASCI) method to enable geometry optimization and vibrational frequency calculation.¹⁰ It achieved a comparable or higher accuracy compared to *con*figuration interaction singles (CIS) and complete active space self-consistent field (CASSCF) with a lower computational cost. Park implemented the analytical gradient theory for the adaptive sampling CI SCF (ASCI-SCF) method.¹¹ It achieved a good accuracy with large active spaces by approximating gradients depending on the sampled determinants. Warden et al. examined several *focal-point methods* combining MP methods with coupled cluster singles, doubles, and perturbative triples [CCSD(T)] to achieve a high accuracy with a lower computational cost.¹² Sahu et al. enabled geometry optimization and vibrational spectra calculation for proteins by combining the molecular tailoring approach (MTA) with DFT and utilizing large-scale parallelization on supercomputers.¹³ Khire et al. also applied MTA to enable the PES construction of medium-sized molecules at the CCSD(T)/aug-cc-pVTZ level.¹⁴ Ahuja et al. applied a reinforcement learning approach that produces a correction term for the quasi-Newton step with BFGS to improve the convergence of geometry optimization.¹⁵ Delgado et al. proposed a variational quantum algorithm to perform geometry optimization using a quantum computer.¹⁶ It minimizes a general cost function in a variational scheme by simultaneously optimizing both the ansatz parameters and Hamiltonian parameters. It achieved a good agreement to the full configuration interaction (FCI) method in a noise-less quantum computer simulation.

The in-depth evaluation of geometry optimization has also been conducted. Cremer et al. compared the accuracy of geometry optimization with several MP and CC methods within large correlation consistent basis sets.¹⁷ Their evaluation showed that the CCSD(T)/cc-pVQZ levels achieve a very high accuracy. Bálint and Jäntschi compared the 39 combinations of various methods and basis sets to analyze the relationship between them and to determine which to use under different circumstances.¹⁸ Shajan et al. compared various open-source geometry optimization implementations via their opensource interface.³ They demonstrated that *internal coordinates*, which represent molecular structures with bond lengths, bond angles, and torsion angles, achieved the better convergence than Cartesian coordinates, and the choice of the initial Hessian and Hessian update method in quasi-Newton approaches also contribute to the convergence.

Recently, surrogate models that predict PESs at the DFT level have been studied intensively to reduce the computational cost of geometry optimization. Río et al. ¹⁹ and Yang et al. ²⁰ presented active learning methods with a Gaussian process regression (GPR) model and neural network (NN) model, respectively. In an active learning process, DFT is executed to calculate accurate energy and gradients when the model prediction uncertainty is high, and surrogate models are updated with the new data. Laghuvarapu et al. proposed a NN model that predicts a molecular energy as the sum of energy contributions from bonds, angles, non-bonds, and dihedrals.²¹ Born and Kästner extended a GPR model to internal coordinates and demonstrated that the convergence of geometry optimization is improved compared to a GPR model based on Cartesian coordinates.²²

A top concern on geometry optimization is a long execution time because time-consuming energy and gradient calculations are repeated across several to tens of steps. Even if a surrogate model as introduced above is used for geometry optimization, ab initio calculations are still necessary to collect training data and complement the model prediction uncertainty. The times required for energy and gradient calculations at each step depend on methods, basis sets, and the size of molecules. High accuracy levels (e.g., CCSD(T) with large basis sets) are generally preferred in various calculations, such as rotational constants, vibrational frequencies, and chemical reactions.^{12,14} However, geometry optimization at such a high accuracy level cannot finish in a practical time for medium- or large-sized molecules. When the various sizes of molecules are required to be optimized, it is too arduous to manually

	HF	MP2	CCSD	$\operatorname{CCSD}(T)$
cc-pV5Z	1h	20h	91h	114h
cc- $pVQZ$	6m	2h	$\mathbf{9h}$	72h
cc- $pVTZ$	50s	10m	34m	$9\mathrm{h}$
cc- $pVDZ$	24s	42s	2m	55m
STO-3G	10s	5s	27s	$1\mathrm{m}$

Table 1: The estimated execution times of the geometry optimization of benzene at 20 accuracy levels.

select a practical accuracy level for each of them.

In this work, we present a scheme to estimate the execution times of geometry optimization of a target molecule at different accuracy levels. It enables to identify the accuracy levels where the geometry optimization of a target molecule finishes in an acceptable time and select an appropriate level from them. For instance, Table 1 showing the estimated times for benzene tells us that geometry optimization at the CCSD/cc-pVQZ level will finish in one night, whereas that at the CCSD(T)/cc-pV5Z level will take around five days. Our evaluation demonstrates that the execution times at 20 accuracy levels are estimated with a mean error of 29.5% for 16 molecules used by Puzzarini et al.,²³ and an appropriate accuracy level can be selected for each of the various sizes of 30 molecules in Baker set²³ in addition to the 16 molecules based on the estimated times and a target time.

In addition, we propose a dynamic method switching technique to reduce the execution time of geometry optimization. It uses light-weight methods at a first few steps and then appropriately switches to more accurate methods for the following steps, based on the norms of gradients obtained from the pre-executed geometry optimization at the lowest accuracy level (e.g., HF/STO-3G). Our evaluation shows that it reduces the execution time by a geometric mean of 22.2% (up to 42.7%) across 16 molecules in the Puzzarini set without any influence on the accuracy.

Molecule Set	Molecules
Alkane (10)	$ C_nH_{2n+2} (n = 1, 2,, 8, 10, 12)$
Small (18)	$ \begin{vmatrix} \text{LiH}, \text{O}_2, \text{N}_2, \text{H}_2\text{O}, \text{BeH}_2, \text{NH}_3, \text{CO}_2, \text{HCl}, \text{CH}_4, \text{C}_2\text{H}_2, \text{C}_2\text{H}_4, \text{C}_2\text{H}_6, \text{C}_3\text{H}_4, \\ \text{C}_3\text{H}_6, \text{C}_3\text{H}_8, \text{C}_4\text{H}_6, \text{C}_4\text{H}_8, \text{C}_4\text{H}_{10} \end{vmatrix} $
Baker ²³ (30)	 water, ammonia, ethane, acetylene, allene, hydroxysulfane, benzene, methylamine, ethanol, acetone, disilyl-ether, 1,3,5-trisilacyclohexane, benzaldehyde, 1,3-difluorobenzene, 1,3,5-trifluorobenzene, neopentane, furan, naphthalene, 1,5-difluoronaphthalene, 2-hydroxybicyclopentane, ACHTAR10, ACANIL01, benzidine, pterin, difuropyrazine, mesityl-oxide, histidine, dimethylpentane, caffeine, menthone
$Puzzarini^{24}$ (16)	$ \begin{vmatrix} \text{HF, N}_2, \text{CO, F}_2, \text{H}_2\text{O}, \text{HCN, HNC, CO}_2, \text{NH}_3, \text{CH}_4, \text{C}_2\text{H}_2, \text{HOF, HNO,} \\ \text{N}_2\text{H}_2, \text{C}_2\text{H}_4, \text{H}_2\text{CO} \end{vmatrix} $

Table 2: Four molecule sets

2 Methods

2.1 Time Estimation

The execution time of geometry optimization, T_{go} , is represented as

$$T_{go} = (T_e + T_g) \times S,\tag{1}$$

where T_e is an energy calculation time, T_g is a gradient calculation time, and S is the number of optimization steps. Hence, the estimation of T_e , T_g , and S is necessary to estimate T_{go} .

2.1.1 Estimation of T_e and T_g

The computational costs of ab initio methods basically depend on the number of basis functions, N. For instance, the general computational costs of HF, MP2, CCSD, and CCSD(T) are $O(N^3)$, $O(N^5)$, $O(N^6)$, and $O(N^7)$, respectively.²⁵ However, the actual T_e and T_g of each method depend on its implementation and a machine configuration where it is executed. Therefore, to estimate T_e and T_g , we use a linear regression model represented as

$$log_{10}(T_{est}) = m \cdot log_{10}(N) + c,$$
 (2)

where *m* is a regression coefficient corresponding to the exponent part of a computational cost $O(N^m)$, and *c* is an intercept. The same model is used to estimate both T_e and T_g . Note that the T_e and T_g of CCSD and CCSD(T) also strongly depend on the number of iterations in CCSD calculation; thus, the above model estimates the time taken per iteration for CCSD and CCSD(T).

In this work, we target four methods (HF, MP2, CCSD, and CCSD(T)) and five basis sets (STO-3G, cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z) implemented in PySCF.²⁶ An energy time model and gradient time model are fitted for each of the 20 accuracy levels using the execution times measured on our server (see Section 3.1 for our experimental setup). T_e and T_g with STO-3G are measured for ten alkane molecules listed in Table 2, while those with cc-pV{D,T,Q,5}Z are measured for 18 *small* molecules. This model fitting procedure should be available for other methods and basis sets.

Table 3 shows m, c, and the coefficient of determination, R^2 , of the fitted energy and gradient time models. We can see that almost all the models are well fitted with a R^2 of over 0.75. The exceptions are the energy time models for HF/cc-pVTZ and MP2/cc-pVTZ, and the gradient time models for CCSD(T)/cc-pV{Q,5}Z. The former two cases are due to a sudden increase in T_e for HF and MP2 with around 250 basis functions. In the latter two cases, the T_g of CCSD(T) does not scale well to the number of basis functions due to the non-optimized implementation in PySCF.²⁷ Moreover, we can also see that m is basically larger for higher accuracy levels. Since m corresponds to the exponent part of computational cost, $O(N^m)$, this observation is in a good agreement with the computational costs of the four methods.

To estimate the T_e and T_g of CCSD and CCSD(T), the number of iterations in CCSD

Table 3: The coefficients and R^2 of time estimation models.

	HF			MP2			CCSD			CCSD(T)		
Basis set	m	c	\mathbb{R}^2	$\mid m$	c	\mathbb{R}^2	$\mid m$	c	\mathbb{R}^2	m	c	\mathbb{R}^2
STO-3G	0.67	-1.19	0.91	0.60	-1.03	0.90	1.40	-2.55	0.91	1.03	-2.04	0.96
cc- $pVDZ$	0.84	-1.48	0.90	0.77	-1.29	0.87	1.21	-2.43	0.79	1.60	-2.99	0.96
cc- $pVTZ$	1.10	-2.00	0.70	1.21	-2.14	0.75	2.49	-4.95	0.95	2.54	-4.78	0.98
cc- $pVQZ$	2.40	-4.80	0.92	2.66	-5.28	0.93	3.60	-7.48	0.98	3.04	-5.79	0.98
cc-pV5Z	3.27	-6.88	0.97	3.60	-7.52	0.97	4.27	-9.12	0.98	3.77	-7.47	0.99

(a) Energy time models

(b) Gradient time models

	HF			MP2			CCSD			CCSD(T)		
Basis set	m	c	\mathbb{R}^2	$\mid m$	c	\mathbb{R}^2	m	С	\mathbb{R}^2	$\mid m$	c	\mathbb{R}^2
STO-3G	2.31	-3.37	0.97	2.75	-4.58	0.99	2.03	-3.74	0.96	3.54	-5.39	0.94
cc- $pVDZ$	2.26	-4.02	0.97	3.21	-5.68	0.90	1.98	-3.82	0.88	4.49	-7.32	0.98
cc- $pVTZ$	1.75	-3.32	0.97	3.63	-6.62	0.99	3.29	-6.39	0.99	4.52	-8.02	0.88
cc-pVQZ	2.03	-3.90	0.96	3.65	-6.62	1.00	3.83	-7.59	1.00	4.45	-8.24	0.75
cc-pV5Z	2.83	-5.62	0.98	3.75	-6.81	1.00	4.01	-8.06	1.00	3.29	-5.89	0.56

calculation is necessary in addition to the time taken per iteration estimated with the regression models. Under the assumption that the number of iterations does not differ significantly with different basis sets, the number of iterations is obtained from the pre-executed energy calculation at the CCSD/STO-3G level for a target molecule. The time overhead of this pre-execution is negligible compared to geometry optimization with a higher accuracy level, because the energy calculation at the CCSD/STO-3G level is performed only once. For instance, the energy calculation at the CCSD/STO-3G level takes only three seconds for benzene, whereas the geometry optimization at the CCSD/Cc-pVDZ level takes 257 seconds.

2.1.2 Estimation of S

The number of steps at an accuracy level, S_{level} , is estimated with the number of steps obtained from the pre-executed geometry optimization at the HF/STO-3G level, $S_{HF/STO-3G}$. Figure 1 plots $S_{HF/STO-3G}$ versus S_{level} at each of the 20 accuracy levels for 16 molecules in Puzzarini set listed in Table 2. Note that the results at the CCSD(T)/cc-pV{Q,5}Z



Figure 1: The number of steps at the HF/STO-3G level, $S_{HF/STO-3G}$, versus that at each of the 20 accuracy levels, S_{level} , for 16 molecules in Puzzarini set. The dotted gray lines show the region where the difference between $S_{HF/STO-3G}$ and S_{level} is within three.

levels for the molecules with more than three atoms are not included, because geometry optimization does not finish in a practical time. We can see that $S_{HF/STO-3G}$ is a good estimator of S_{level} because the difference between them is within three in almost all the results. There are only three exceptions out of 308 results: CCSD(T)/cc-pV{D,T}Z for HOF and CCSD(T)/cc-pVTZ for HNO. The time overhead of the pre-executed geometry optimization at the HF/STO-3G level is negligible compared to geometry optimization at a higher accuracy level. For instance, the geometry optimization of benzene at the HF/STO-3G level takes only 13 seconds, while that at the MP2/cc-pVTZ level takes 556 seconds.

2.2 Gradient-based Method Switching (GMS)

To reduce T_{go} at a selected accuracy level, we propose a novel technique that dynamically switches multiple ab initio methods during geometry optimization. Its main concept is to save time by using light-weight methods at a few first steps where the selected accuracy level is unnecessary for energy and gradient calculations.

We investigate how T_{go} is affected by using light-weight methods at a few first steps. With



Figure 2: The geometry optimization time, T_{go} , normalized with respect to CCSD (upper) and the number of steps, S, (lower) when HF or MP2 is used at the x first steps before switching to CCSD for caffeine with STO-3G.

the assumption that the CCSD/STO-3G level is selected, Figure 2 shows T_{go} normalized with respect to CCSD and the number of steps, S, when HF or MP2 is used at a few first steps for caffeine/STO-3G. The x-axis indicates the number of first steps where HF or MP2 is used before switching to CCSD. When HF is used before CCSD (HF->CCSD), T_{go} is reduced by using HF only at the first step. Otherwise, T_{go} is increased due to the increase of S. On the other hand, when MP2 is used before CCSD (MP2->CCSD), T_{go} is minimized without the increase of S by using MP2 at the first four steps. From these results, we obtain the following two observations: (1) T_{go} can be reduced by using light-weight methods at the appropriate number of first steps. (2) The appropriate number of first steps is different depending on light-weight methods.

To identify the appropriate number of first steps using light-weight methods, we focus on the norm of gradients, ||G||, calculated at each optimization step. It is a useful metric



Figure 3: The error in the norm of gradients, $||\boldsymbol{G}||$, from that calculated with CCSD(T) for 18 small molecules with STO-3G.



Figure 4: The norm of gradients calculated at the HF/STO-3G level, $||\mathbf{G}||_{HF/STO-3G}$, for caffeine.

to know the calculation accuracy required at each step for two reasons: $||\boldsymbol{G}||$ can be calculated from the first step, and it decreases gradually as atomic coordinates get closer to the stationary ones. Hence, we evaluate the accuracy of $||\boldsymbol{G}||$ calculation with HF, MP2, and CCSD by comparing with CCSD(T). Figure 3 plots the error in $||\boldsymbol{G}||$ from that calculated with CCSD(T), $||\boldsymbol{G}||_{CCSD(T)}$, for 18 small molecules listed in Table 2 with STO-3G. We can see that more accurate methods generally achieve lower errors. Therefore, we use the maximum error of each method shown above the graph as a threshold to use it during geometry optimization. $||\boldsymbol{G}||$ at each step is obtained from the pre-executed geometry optimization of a target molecule at the HF/STO-3G level. Figure 4 shows $||\boldsymbol{G}||$ calculated at each step at the HF/STO-3G level, $||\boldsymbol{G}||_{HF/STO-3G}$, for caffeine. The maximum errors in $||\boldsymbol{G}||$ calculation of HF, MP2, and CCSD evaluated in Figure 3 are shown with horizontal dotted lines. We implement the gradient-based method switching (GMS) technique that selects a method used at each step by checking whether $||\boldsymbol{G}||_{HF/STO-3G}$ exceeds the corresponding maximum error. For instance, when CCSD(T) with an arbitrary basis set is selected as an accuracy level for caffeine, a method at each step is selected as [HF, MP2, MP2, CCSD, CCSD, CCSD(T), ...]. As discussed in Section 2.1.2, the time overhead of the pre-executed geometry optimization at the HF/STO-3G level is negligible compared to that with a higher accuracy level.

2.3 Whole Procedure

In this section, we summarize the whole procedure to estimate T_{go} for a target molecule and perform geometry optimization at a selected accuracy level with our proposed GMS technique.

2.3.1 Advance preparation

The following two steps are required to be performed only once in advance for an experimental setup.

(a) Data collection: First, T_e and T_g at all accuracy levels are measured for the molecule sets listed in Table 2 as learning data for the time estimation models. The ten alkane molecules and 18 small molecules are used for STO-3G and the other larger basis sets, respectively. The Cartesian coordinates of all the 28 molecules optimized with *composite/CBS-Q* are obtained from CCCBDB.²⁸ Second, the maximum errors in ||G|| calculation of all methods are evaluated for the 18 small molecules with STO-3G, as shown in Figure 3. The values of $||G||_{CCSD(T)}$ used as baselines are listed in the Supporting Information. In this work using the four methods and five basis sets, the whole data collection takes

31 hours in total.

(b) Time estimation model fitting: With T_e and T_g measured in the step (a) and the numbers of basis functions, N, of the 28 molecules, the linear regression models shown in Equation 2 are fitted to estimate T_e and T_g at all accuracy levels, as shown in Table 3.

2.3.2 Geoemetry optimization of a target molecule

(1) **Pre-executions:** For the T_{go} estimation and GMS, two pre-executions are necessary for a target molecule. First, the number of steps, $S_{HF/STO-3G}$, and the norm of gradients at each step, $||G||_{HF/STO-3G}$, are obtained from the geometry optimization at the HF/STO-3G level. Second, if CCSD or CCSD(T) is included in target methods, the number of iterations in CCSD calculation is obtained from the energy calculation at the CCSD/STO-3G level. For benzidine which is the largest in Baker set, the geometry optimization at the HF/STO-3G level takes 154 seconds, and the energy calculation at the CCSD/STO-3G level takes 30 seconds.

(2) Time estimation and accuracy level selection: T_e and T_g at all accuracy levels are estimated with the energy and gradient time models fitted in the step (b), the number of basis functions, N, of the target molecule, and the number of CCSD iterations obtained in the step (1). Then, T_{go} at all accuracy levels are calculated based on Equation 1 with $S_{HF/STO-3G}$ obtained in the step (1) and the estimated T_e and T_g . After that, an accuracy level where the estimated T_{go} is acceptable can be selected.

(3) Geometry optimization with GMS: The geometry optimization of the target molecule is performed with GMS at the accuracy level selected in the step (2). GMS selects a method used in each step by comparing $||G||_{HF/STO-3G}$ obtained in the step (1) and the maximum errors evaluated in the step (a).

3 Results and Discussion

In this section, we evaluate the estimation accuracy of T_{go} , the effectiveness of selecting an accuracy level based on the estimated T_{go} , and the time reduction by GMS. We first describe our experimental setup and then show the evaluation results.

3.1 Experimental Setup

In this work, we target 20 accuracy levels composed of four ab initio methods (HF, MP2, CCSD, and CCSD(T)) and five basis sets (STO-3G, cc-pVDZ, cc-pVTZ, cc-pVQZ, and cc-pV5Z) implemented in PySCF.²⁶ PySCF is a Python-based open-source quantum chemical calculation framework. We use the *geomopt* module in PySCF via an interface to *geomeT*- RIC^{29} with the default convergence criteria. The *LinearRegression* module in *scikit-learn*³⁰ is used to fit the energy and gradient time estimation models. A server containing two Xeon Gold 6240M processors and 384 GB DRAM is used for all experiments in this work.

For evaluation, we select 16 molecules used by Puzzarini et al.²⁴ and 30 molecules in Baker set,²³ as listed in Table 2. For the molecules in Puzzarini set, we obtain the Cartesian coordinates optimized with *composite/CBS-Q* from CCCBDB²⁸ and initialize all the bond distances to 1.0 Å while keeping bond angles. Moreover, the experimental Cartesian coordinates of the molecules in Puzzarini set are also obtained from CCCBDB and used to evaluate the accuracy of optimized coordinates. For the molecules in Baker set, we use the initial Cartesian coordinates provided by Shajan et al.³

3.2 Time Estimation Accuracy

First, we evaluate the estimation accuracy of T_{go} with Figure 5 plotting the measured versus estimated T_{go} at the 20 accuracy levels for the 16 molecules in Puzzarini set. Different colored dots show the results of different methods, and the black line indicates the exact match between the estimated and measured T_{go} . Note that the results at the CCSD(T)/cc-



Figure 5: The measured versus estimated T_{go} at the 20 accuracy levels for 16 molecules in Puzzarini set.

 $pV{Q,5}Z$ levels for the molecules with more than three atoms are not included in similar to Figure 1. We can see that T_{go} is estimated accurately in the most results. The mean absolute percentage error (MAPE) across all the results is 29.5%, which is sufficiently low to identify the accuracy levels where geometry optimization finishes in an acceptable time. However, T_{go} with CCSD(T) shown with red dots are under-estimated significantly in a lot of cases due to the low R^2 of the gradient time models for cc-pV{Q,5}Z as shown in Table 3. This is because the gradient calculation time of CCSD(T) does not scale well to the number of basis functions due to the non-optimized implementation in PySCF.²⁷ It is our future work to optimize it or to consider a better time estimation model for the gradient calculation with CCSD(T).

3.3 Accuracy Level Selection and Method Switching

Next, we evaluate the effectiveness of the estimated time-based accuracy level selection (ETALS) and gradient-based method switching (GMS) technique. We set a target T_{go} and select the highest accuracy level where the estimated T_{go} is below it in the following three



Figure 6: The RMSD of the optimized coordinates in Å and the execution time of geometry optimization, T_{go} , for 16 molecules in Puzzarini set with a target T_{go} of 1,000 seconds.

steps. (1) Without HF and STO-3G, the highest accuracy level meeting the target T_{go} is searched with the outer loop in the order of larger basis sets and the inner loop in the order of more accurate methods. (2) If the accuracy level is not found in the step 1, the largest basis set with HF meeting the target T_{go} is searched. (3) If the accuracy level is not found in the step 2, the most accurate method with STO-3G meeting the target T_{go} is searched. For instance, when T_{go} are estimated for benzene as shown in Table 1 and a target T_{go} is set to 300 seconds (5 minutes), CCSD/cc-pVDZ is selected.

Figure 6 shows the root mean square deviation (RMSD) of the optimized coordinates with respect to the experimental coordinates and T_{go} for the 16 molecules in Puzzarini set with a target T_{go} of 1,000 seconds. The HF/cc-pV5Z level is evaluated as a naive baseline (blue bars), where the geometry optimization of all the molecules finishes in the target T_{go} . The orange bars show that the accuracy levels selected with ETALS achieve the much lower coordinates RMSD than HF/cc-pV5Z in around 1,000 seconds for almost all the molecules.



Figure 7: Hartree energy, Eh, (upper) and the norm of gradients, ||G||, (lower) calculated during geometry optimization. The x-axis shows the elapsed time in seconds.

This result demonstrates that ETALS enables to select a high accuracy level for each molecule based on the estimated and target T_{go} . The reason why T_{go} exceeds the target T_{go} by 50% for the HF molecule is because the CCSD(T)/cc-pV5Z level is selected based on the underestimated T_{qo} due to the low R^2 of the gradient time model as shown in Table 3. In addition, the green bars show that GMS reduces T_{go} for almost all the molecules at the accuracy levels selected with ETALS without any influence on the coordinates RMSD. The maximum reduction is 42.7% for F2, where CCSD/cc-pV5Z is selected. Figure 7a compares the energy and ||G|| calculated during the geometry optimization of F2 at the CCSD/cc-pV5Z level between without and with GMS. The x-axis shows the elapsed time in seconds, and the dots represent optimization steps. We can see that GMS reduces the number of steps using CCSD, where the time per step is around 160 seconds, from eight to four by using HF at the first two steps and MP2 at the third step. Although the energies calculated with HF at the first two steps are significantly different from those with CCSD, the final energy and $||\boldsymbol{G}||$ converge to the comparable values. The geometric mean of the T_{go} reduction by GMS across the 16 molecules is 22.2%. This is not a drastic reduction, but GMS can be applied without any concern because it does not affect the accuracy of geometry optimization. We also conduct the same evaluation with target T_{go} of 100 and 300 seconds, where the geometric



Figure 8: The execution time of geometry optimization, T_{go} , for 30 molecules in Baker set with a target T_{go} of 300 seconds.

means of the T_{go} reduction are 10.3% and 15.8%, respectively.

Figure 8 plots T_{go} for the 30 molecules in Baker set with a target T_{go} of 300 seconds. The evaluation of coordinates RMSD is excluded because the experimental coordinates of almost all the molecules are not available. This graph shows that the appropriate accuracy levels are selected with ETALS so that geometry optimization finishes in around 300 seconds for almost all the molecules. As the various sizes of molecules are included in Baker set, the selected accuracy levels differ significantly depending on the molecular sizes. For instance, CCSD/ccpVQZ is selected for water including three atoms, while HF/STO-3G is selected for menthone including 29 atoms. For neopentane, 1_5-diffuoronaphthalene, and difuropyrazine, T_{go} are relatively long because CCSD/cc-pVDZ and HF/cc-pVTZ are selected based on the underestimated T_{go} due to the low R^2 of the corresponding energy time models as shown in Table 3. GMS reduces T_{go} for several small molecules at left side, where CCSD is mainly selected with ETALS, and the time is saved by using HF and MP2 at a few first steps. In contrast, the time reduction by GMS cannot be seen for large molecules at right side, because HF and MP2 are selected with ETALS. Unfortunately, GMS increases T_{go} for water and acetylene because the number of steps is increased by using HF or MP2 at a few first steps. Figure 7b shows the behavior of the geometry optimization of acetylene, where MP2/cc-pVQZ is selected with ETALS. While the number of steps is four without GMS, it is increased to six by using HF at the first step with GMS. Although the time increases by GMS can be avoided by conservatively setting higher thresholds to select light-weight methods, it would also decrease the time reduction by GMS. With the current thresholds shown in Figure 4, the time increases by GMS are observed only in the two cases out of 46 cases through Figure 6 and Figure 8.

4 Conclusion

In this work, we propose the scheme to estimate the geometry optimization times at different accuracy levels for a target molecule and the GMS technique that reduces the execution time by dynamically switching multiple methods during geometry optimization. They enable to identify the accuracy levels where geometry optimization will finish in an acceptable time and perform geometry optimization at a selected accuracy level in a shorter time than only using a single method. The evaluation using 46 molecules in total demonstrates that the geometry optimization times at 20 accuracy levels are estimated with a MAPE of 29.5%, and GMS reduces the execution time by up to 42.7% without affecting the accuracy of geometry optimization.

References

 Schlegel, H. B. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *Journal of Computational Chemistry* 2003, 24, 1514–1527.

- (2) Schlegel, H. B. Geometry optimization. WIREs Computational Molecular Science 2011, 1, 790–809.
- (3) Shajan, A.; Manathunga, M.; Götz, A. W.; Jr., K. M. M. Geometry Optimization: A Comparison of Different Open-Source Geometry Optimizers. *Journal of Chemical Theory and Computation* **2023**, *19*, 7533–7541.
- (4) Fischer, T. H.; Almlof, J. General methods for geometry and wave function optimization. The Journal of Physical Chemistry 1992, 96, 9768–9774.
- (5) Baerends, E. J.; Ellis, D. E.; Ros, P. Self-consistent molecular Hartree—Fock—Slater calculations I. The computational procedure. *Chemical Physics* **1973**, *2*, 41–51.
- (6) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. Phys. Rev. 1965, 140, A1133–A1138.
- (7) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. Phys. Rev. 1934, 46, 618–622.
- (8) David Sherrill, C.; Schaefer, H. F. The Configuration Interaction Method: Advances in Highly Correlated Approaches; Advances in Quantum Chemistry; Academic Press, 1999; Vol. 34; pp 143–269.
- (9) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* 2007, 79, 291–352.
- (10) Chaudhuri, R. K.; Freed, K. F. Geometry optimization using improved virtual orbitals: A complete active space numerical gradient approach. *The Journal of Chemical Physics* 2007, *126*, 114103.
- (11) Park, J. W. Second-Order Orbital Optimization with Large Active Spaces Using Adaptive Sampling Configuration Interaction (ASCI) and Its Application to Molecular Ge-

ometry Optimization. Journal of Chemical Theory and Computation **2021**, 17, 1522–1534.

- (12) Warden, C. E.; Smith, D. G. A.; Burns, L. A.; Bozkaya, U.; Sherrill, C. D. Efficient and automated computation of accurate molecular geometries using focal-point approximations to large-basis coupled-cluster theory. *The Journal of Chemical Physics* **2020**, *152*, 124109.
- (13) Sahu, N.; Khire, S. S.; Gadre, S. R. Combining fragmentation method and highperformance computing: Geometry optimization and vibrational spectra of proteins. *The Journal of Chemical Physics* **2023**, *159*, 44309.
- (14) Khire, S. S.; Gurav, N. D.; Nandi, A.; Gadre, S. R. Enabling Rapid and Accurate Construction of CCSD(T)-Level Potential Energy Surface of Large Molecules Using Molecular Tailoring Approach. *The Journal of Physical Chemistry A* 2022, *126*, 1458– 1464.
- (15) Ahuja, K.; Green, W. H.; Li, Y.-P. Learning to Optimize Molecular Geometries Using Reinforcement Learning. *Journal of Chemical Theory and Computation* 2021, 17, 818– 825.
- (16) Delgado, A.; Arrazola, J. M.; Jahangiri, S.; Niu, Z.; Izaac, J.; Roberts, C.; Killoran, N. Variational quantum algorithm for molecular geometry optimization. *Physical Review A* 2021, 104, 052402.
- (17) Cremer, D.; Kraka, E.; He, Y. Exact geometries from quantum chemical calculations. Journal of Molecular Structure 2001, 567-568, 275–293.
- (18) Bálint, D.; Jäntschi, L. Comparison of Molecular Geometry Optimization Methods Based on Molecular Descriptors. *Mathematics* 2021, 9, 2855.

- (19) del Río, E. G.; Mortensen, J. J.; Jacobsen, K. W. Local Bayesian optimizer for atomic structures. *Physical Review B* 2019, 100, 104103.
- (20) Yang, Y.; Jiménez-Negrón, O. A.; Kitchin, J. R. Machine-learning accelerated geometry optimization in molecular simulation. *Journal of Chemical Physics* 2021, 154.
- (21) Laghuvarapu, S.; Pathak, Y.; Priyakumar, U. D. BAND NN: A Deep Learning Framework for Energy Prediction and Geometry Optimization of Organic Small Molecules. *Journal of Computational Chemistry* 2020, 41, 790–799.
- (22) Born, D.; Kastner, J. Geometry Optimization in Internal Coordinates Based on Gaussian Process Regression: Comparison of Two Approaches. *Journal of Chemical Theory and Computation* **2021**, *17*, 5955–5967.
- (23) Baker, J. Techniques for geometry optimization: A comparison of cartesian and natural internal coordinates. *Journal of Computational Chemistry* **1993**, *14*.
- (24) Puzzarini, C.; Heckert, M.; Gauss, J. The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. molecules containing first-row atoms. *The Journal of Chemical Physics* 2008, *128*, 194108.
- (25) Solving quasiparticle band spectra of real solids using neural-network quantum states.
 Communications Physics 2021, 4, 106.
- (26) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; Mc-Clain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework. WIREs Computational Molecular Science 2018, 8, e1340.
- (27) Sun, Q. Geometry optimization using CCSD(T)/meta-GGA. https://github.com/ pyscf/pyscf/issues/226, Last checked: February, 2024.

- (28) Russell, J. D. NIST Computational Chemistry Comparison and Benchmark Database: NIST Standard Reference Database Number 101 Release 22, May 2022. http:// cccbdb.nist.gov/.
- (29) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *The Journal of Chemical Physics* 2016, 144, 214108.
- (30) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011, 12, 2825–2830.

TOC Graphic

