Gentlest ascent dynamics for calculating first excited state and exploring energy landscape of Kohn-Sham density functionals

Chen Li, Jianfeng Lu, and Weitao Yang

Citation: The Journal of Chemical Physics **143**, 224110 (2015); doi: 10.1063/1.4936411
View online: https://doi.org/10.1063/1.4936411
View Table of Contents: http://aip.scitation.org/toc/jcp/143/22
Published by the American Institute of Physics

Gentlest ascent dynamics for calculating first excited state and exploring energy landscape of Kohn-Sham density functionals

Chen Li,[1] Jianfeng Lu,[1,2,a)] and Weitao Yang[1,3,b)]
[1]*Department of Chemistry, Duke University, Durham, North Carolina 27708, USA*
[2]*Departments of Mathematics and Physics, Duke University, Durham, North Carolina 27708, USA*
[3]*Key Laboratory of Theoretical Chemistry of Environment, School of Chemistry and Environment, South China Normal University, Guangzhou 510006, China*

(Received 23 February 2015; accepted 12 November 2015; published online 9 December 2015)

We develop the gentlest ascent dynamics for Kohn-Sham density functional theory to search for the index-1 saddle points on the energy landscape of the Kohn-Sham density functionals. These stationary solutions correspond to excited states in the ground state functionals. As shown by various examples, the first excited states of many chemical systems are given by these index-1 saddle points. Our novel approach provides an alternative, more robust way to obtain these excited states, compared with the widely used ΔSCF approach. The method can be easily generalized to target higher index saddle points. Our results also reveal the physical interest and relevance of studying the Kohn-Sham energy landscape. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4936411]

I. INTRODUCTION

Kohn-Sham density functional theory (KS-DFT)[1] has achieved great success in ground state calculations over the past decades.[2] Moreover, its time-dependent variant, the time-dependent density functional theory (TDDFT), has also been developed as an efficient method to compute excitation energies.[3–5] Compared with wavefunction based methods for excited state calculations, such as configuration interaction (CI),[6] multireference CI (MR-CI),[7] complete active space self-consistent field (CASSCF),[8] and equation-of-motion coupled cluster (EOM-CC)[9–11] theories, density functional based methods are known for their tremendously lower computational cost, and hence their applicability to medium or large systems.

Besides TDDFT, attempts have also been made to obtain the excited state energies using the time-independent DFT.[12–27] It has been shown that every stationary solution of the ground-state functional represents an excited state.[28] Although the validity has only been shown for the exact functional, with implicit assumption of the v-representability of the excited state densities, it implies some meaning to the stationary solutions of approximate ground-state functionals whose energies are above the ground state. The excitation energies are then approximated by the energy difference between these stationary solutions (excited states) and the ground state.

To obtain these excited states, the ΔSCF method has been developed,[29,30] which modifies the standard self-consistent field iteration by adopting non-Aufbau occupations at each iteration: molecular orbital higher in energy is occupied, with orbitals unoccupied below in energy. At convergence, the method targets the non-Aufbau solutions of the ground state functional. Compared with TDDFT, the ΔSCF approach has much smaller computational cost, because it inherits the favorable computational scaling of the ground state DFT. Moreover, it has been recently shown that the ΔSCF approach can yield reasonable results, with accuracy comparable to the TDDFT calculations.[27,31–33]

Despite its advantage of straightforward modification of existing implementations of SCF, the ΔSCF method is known to have difficulties in converging to the excited states of the Kohn-Sham equations. The conventional SCF procedures are designed to effectively find the energy minimum, corresponding to the ground state. In the case of searching for non-Aufbau solutions, the naive modification of the SCF procedure can lead to a collapse to the ground state, or simply failure to converge if forcing the non-Aufbau occupation. To alleviate this problem, the constricted DFT was proposed previously, which restricts the excited state orbitals to a unitary transformation of the ground state with non-vanishing contribution from the virtual space.[22] Recently, there have been techniques developed, including the maximum overlap method (MOM),[34] which prevents collapse to the ground state, and the metadynamics related method,[35] which avoids reconvergence to the same state.

Although the above mentioned techniques improve upon the ΔSCF method, it is worth noticing that they rely on a proper initial guess that locates sufficiently close to the targeted state. Yet it is not always easy to construct such an initial guess, especially for complicated molecules which do not provide us with much chemical intuition. It is thus desirable to develop a more robust method that guarantees convergence to the right solution without much dependence on the initial guess. In this work, we will achieve this by taking a quite different point of view. Instead of trying to modify the conventional SCF procedure, our starting point is the notion of the energy landscape of the ground state Kohn-Sham DFT, defined on the configuration space of single Slater determinant.

The excited states, as the stationary solutions of the ground state Kohn-Sham density functional, then correspond

a)jianfeng@math.duke.edu
b)weitao.yang@duke.edu

to the critical points of the Kohn-Sham energy landscape. In particular, as will be shown in this paper, most of the first excited states correspond to index-1 saddle points (i.e., the stationary points with one and only one energy descending direction) of the energy functional. Therefore, finding these excited states amounts to locate index-1 saddle point of the energy landscape of the Kohn-Sham density functional.

The understanding of these critical points on the Kohn-Sham energy landscape, to the best of our knowledge, has not been explored to design robust algorithms to search for them. As we will demonstrate in this work, this new perspective opens a door to applying various methods, developed for locating transition states for chemical reactions,[36–44] for the search of excited states in ground state Kohn-Sham functionals. In this work, we will apply the recently developed gentlest ascent dynamics (GAD)[45] to Kohn-Sham DFT for finding excited states. While other methods of this kind are also possible, we choose GAD here as it offers a clear set of dynamical equations that converge to the index-1 saddle points and it is also connected to the well known steepest descent approach for Kohn-Sham DFT.[46–48]

The GAD was proposed to locate the index-1 saddle points of energy functions in the context of rare events in molecular dynamics.[45,49] Let us briefly summarize the idea of the GAD with comparison to the standard steepest descent dynamics (SDD) that targets the energy minimum. Given an energy function $E(x)$ on $\mathbb{R}^n$, SDD has the following form:

$$\frac{dx}{dt} = -\nabla E(x). \tag{1}$$

One can show that $E(x(t))$ is a decreasing function of t if $x(t)$ is a solution to Eq. (1); and $x(t)$ converges to a local minimum of the energy function E. In contrast, the gentlest ascent dynamics involve two coupled equations as the following:[45]

$$\frac{dx}{dt} = -\nabla E(x) + 2\frac{(\nabla E, v)}{(v,v)}v, \tag{2}$$

$$\frac{dv}{dt} = -\nabla^2 E(x)v + \frac{(v, \nabla^2 E v)}{(v,v)}v. \tag{3}$$

Here, Eq. (3) involves an auxiliary vector v, which attempts to find the direction that corresponds to the smallest eigenvalue of the Hessian $\nabla^2 E(x)$; and Eq. (2) is similar to Eq. (1), but with an added term that flips the sign of the v-direction in the minus energy gradient. This treatment enables the GAD to converge to the index-1 saddle points of the energy function.

As Kohn-Sham density functional theory is defined for Slater determinants, we have to modify the GAD formulation to take into account the orthonormality constraints of the orbitals, as will be discussed in detail below. This leads to a robust method for finding index-1 saddle points, which include first excited states for many systems as shown in this work. The GAD is not limited to index-1 saddle points; the generalization to higher index saddle points is straightforward and can be used for higher excited states.

II. METHODOLOGY

Let the Kohn-Sham energy functional E be an explicit functional of the Kohn-Sham orbitals, $E = E[\{\psi_i^\sigma\}]$ with each orbital function $\psi_i^\sigma \in L^2(\mathbb{R}^3)$ (assumed to be real for simplicity). To apply the gentlest ascent dynamics to KS-DFT and also for the simplicity of notations, instead of working with multiple orbital functions, it is helpful to view the energy functional E evaluated on $\mathbf{\Psi}$ in the space $L^2(\mathbb{R}^3;\mathbb{R}^N)$, namely, at each spatial point $\boldsymbol{r}$, $\mathbf{\Psi}(\boldsymbol{r})$ takes value in $\mathbb{R}^N$. To be more specific, $\mathbf{\Psi}(\boldsymbol{r})$ stands for the collection of occupied orbitals, which are organized as a column vector

$$\mathbf{\Psi}(\boldsymbol{r}) = \begin{pmatrix} \mathbf{\Psi}^\alpha(\boldsymbol{r}) \\ \mathbf{\Psi}^\beta(\boldsymbol{r}) \end{pmatrix}, \tag{4}$$

where $\mathbf{\Psi}^\sigma(\boldsymbol{r}) = \left(\psi_1^\sigma(\boldsymbol{r}), \psi_2^\sigma(\boldsymbol{r}), \ldots, \psi_{N^\sigma}^\sigma(\boldsymbol{r})\right)^{\mathrm{T}}$ for $\sigma = \alpha, \beta$. Here, N^α and N^β indicate the number of the spin-up and spin-down electrons, where $N = N^\alpha + N^\beta$ is the total number of electrons. In this work, we only consider excitations that maintain the spin multiplet, i.e., N^α and N^β are fixed.

The auxiliary perturbation direction $\mathbf{\Phi}(\boldsymbol{r})$ of the electron orbitals is treated similarly, $\mathbf{\Phi}(\boldsymbol{r}) = \begin{pmatrix} \mathbf{\Phi}^\alpha(\boldsymbol{r}) \\ \mathbf{\Phi}^\beta(\boldsymbol{r}) \end{pmatrix}$, with $\mathbf{\Phi}^\sigma(\boldsymbol{r}) = \left(\varphi_1^\sigma(\boldsymbol{r}), \varphi_2^\sigma(\boldsymbol{r}), \ldots, \varphi_{N^\sigma}^\sigma(\boldsymbol{r})\right)^{\mathrm{T}}$.

We denote $\langle \cdot, \cdot \rangle$ as the inner product of $\mathbf{\Psi}$ in the space of $L^2(\mathbb{R}^3;\mathbb{R}^N)$, while using $(\cdot,\cdot)$ for the standard inner product in $L^2(\mathbb{R}^3)$,

$$\langle \mathbf{\Psi}, \mathbf{\Phi} \rangle \equiv \sum_\sigma \sum_{i=1}^{N_\sigma} (\psi_i^\sigma, \varphi_i^\sigma) = \sum_\sigma \sum_{i=1}^{N_\sigma} \int \psi_i^\sigma(\boldsymbol{r}) \phi_i^\sigma(\boldsymbol{r}) d\boldsymbol{r}. \tag{5}$$

The gradients of the Kohn-Sham energy functional with respect to the orbitals are denoted by $\frac{\partial E}{\partial \mathbf{\Psi}}$ and $\frac{\partial^2 E}{\partial \mathbf{\Psi}^2}$. The first gradient acting on $\mathbf{\Phi}$ gives a number

$$\frac{\partial E}{\partial \mathbf{\Psi}}[\mathbf{\Phi}] \equiv \left\langle \frac{\partial E}{\partial \mathbf{\Psi}}, \mathbf{\Phi} \right\rangle \equiv \sum_{i\sigma} \left(\frac{\delta E}{\delta \psi_i^\sigma}, \varphi_i^\sigma \right), \tag{6}$$

where $\frac{\delta E}{\delta \psi_i^\sigma}$ is the standard functional derivative of the Kohn-Sham functional. The second gradient $\frac{\partial^2 E}{\partial \mathbf{\Psi}^2}$ acting on $\mathbf{\Phi}$ gives a vector in $L^2(\mathbb{R}^3;\mathbb{R}^N)$, with its component being

$$\left(\frac{\partial^2 E}{\partial \mathbf{\Psi}^2}[\mathbf{\Phi}]\right)_i^\sigma \equiv \sum_{\sigma' j} \frac{\delta^2 E}{\delta \psi_i^\sigma \delta \psi_j^{\sigma'}} [\varphi_j^{\sigma'}]. \tag{7}$$

For the Kohn-Sham energy functional $E[\mathbf{\Psi}]$, it is assumed that the components of $\mathbf{\Psi}$ are orthonormal. Thus, we define $\mathcal{W}$ as the admissible class of $\mathbf{\Psi}$s,

$$\mathcal{W} \equiv \{\mathbf{\Psi} \mid O_{ij}^\sigma(\mathbf{\Psi}, \mathbf{\Psi}) = \delta_{ij}, \forall\, \sigma, i, j\}. \tag{8}$$

Here, $O_{ij}^\sigma(\mathbf{\Psi}, \mathbf{\Psi}) = (\psi_i^\sigma, \psi_j^\sigma)$ represents the standard inner product between the ith and jth element of vector $\mathbf{\Psi}$.

Direct substitution of Eqs. (6) and (7) into Eqs. (2) and (3) does not give us the right dynamics, since it will violate the orthonormality constraints of the orbitals $\mathbf{\Psi}$. More specifically, consider orbitals $\mathbf{\Psi} \in \mathcal{W}$ and a perturbation direction $\mathbf{\Phi}$ with $\|\mathbf{\Phi}\|^2 = \langle \mathbf{\Phi}, \mathbf{\Phi} \rangle = 1$, $\mathbf{\Psi} + t\mathbf{\Phi}$ is not necessarily in $\mathcal{W}$ and we have

$$\begin{aligned} O_{ij}^\sigma(\mathbf{\Psi} + t\mathbf{\Phi}, \mathbf{\Psi} + t\mathbf{\Phi}) &= (\psi_i^\sigma + t\varphi_i^\sigma, \psi_j^\sigma + t\varphi_j^\sigma) \\ &= \delta_{ij} + tU_{ij}^\sigma + t^2 V_{ij}^\sigma, \end{aligned} \tag{9}$$

where $U_{ij}^\sigma = (\psi_i^\sigma, \varphi_j^\sigma) + (\psi_j^\sigma, \varphi_i^\sigma)$ and $V_{ij}^\sigma = (\varphi_i^\sigma, \varphi_j^\sigma)$.

Hence, to guarantee that the perturbed orbitals satisfy the orthonormality constraints to the first order of t, we restrict

the choice of $\mathbf{\Phi}$ such that $(\psi_i^\sigma, \varphi_j^\sigma) = 0, \forall i, j$. In other words, $\mathbf{\Phi}$ lies in the virtual space with respect to $\mathbf{\Psi}$. Note that this restriction is stronger than requiring $\boldsymbol{U} = 0$, while this does not over-restrict the possible perturbation directions, thanks to the symmetry of the Kohn-Sham energy functional with respect to rotations of the orbitals (see the supplementary material[50] for details).

Let us now derive the correct gradients of the energy functional under constraints by considering $L[\mathbf{\Psi} + t\mathbf{\Phi}] = E[\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi})]$ as the targeting function and its Taylor expansion up to t^2. Here, $\hat{\boldsymbol{F}}$ denotes the orthonormalization operator

$$\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi}) = \boldsymbol{A}(\mathbf{\Psi} + t\mathbf{\Phi}), \tag{10}$$

with the (t-dependent) matrix $\boldsymbol{A}$ chosen such that $\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi}) \in \mathcal{W}$. To determine $\boldsymbol{A}$, we calculate

$$\begin{aligned}
\delta_{ij} &= O_{ij}^\sigma\big(\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi}), \hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi})\big) \\
&= O_{ij}^\sigma\big(\boldsymbol{A}(\mathbf{\Psi} + t\mathbf{\Phi}), \boldsymbol{A}(\mathbf{\Psi} + t\mathbf{\Phi})\big) \\
&= \Big(\sum_k A_{ik}^\sigma(\psi_k^\sigma + t\varphi_k^\sigma), \sum_l A_{jl}^\sigma(\psi_l^\sigma + t\varphi_l^\sigma)\Big) \\
&= \sum_{kl} A_{ik}^\sigma A_{jl}^\sigma(\delta_{kl} + t^2 V_{kl}^\sigma),
\end{aligned} \tag{11}$$

where we have used that $U_{kl}^\sigma = (\psi_k^\sigma, \varphi_l^\sigma) + (\psi_l^\sigma, \varphi_k^\sigma) = 0$. In matrix form, the above equation reads

$$\boldsymbol{A}^\sigma(\boldsymbol{I}^\sigma + t^2\boldsymbol{V}^\sigma)\boldsymbol{A}^{\sigma,\mathrm{T}} = \boldsymbol{I}^\sigma. \tag{12}$$

Hence, $\boldsymbol{A}$ can be chosen as

$$\boldsymbol{A}^\sigma = (\boldsymbol{I}^\sigma + t^2\boldsymbol{V}^\sigma)^{-1/2} = (\boldsymbol{I}^\sigma - \frac{1}{2}t^2\boldsymbol{V}^\sigma) + O(t^3) \tag{13}$$

where in the last equality, we have used Taylor expansion and kept terms up to the second order. It then follows:

$$\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi}) = (\mathbf{\Psi} + t\mathbf{\Phi}) - \frac{1}{2}t^2\boldsymbol{V}\mathbf{\Psi}, \tag{14}$$

up to the second order in t, where $\boldsymbol{V} = \begin{bmatrix} V^\alpha & \mathbf{0} \\ \mathbf{0} & V^\beta \end{bmatrix}$. The last term in Eq. (14) arises to satisfy the orthonormality constraint to the second order, which will lead to extra terms in the Hessian of the energy functional as discussed below.

Now we write the constrained energy functional L perturbed at an approachable $\mathbf{\Psi}$ (note $L[\mathbf{\Psi}] = E[\mathbf{\Psi}]$) in terms of Taylor expansion up to second order as the following formal expression:

$$\begin{aligned}
L[\mathbf{\Psi} + t\mathbf{\Phi}] &= E[\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi})] \\
&= L[\mathbf{\Psi}] + t\frac{\partial L}{\partial \mathbf{\Psi}}[\mathbf{\Phi}] + \frac{1}{2}t^2\langle\mathbf{\Phi}, \hat{\boldsymbol{H}}[\mathbf{\Phi}]\rangle.
\end{aligned} \tag{15}$$

Here, $\frac{\partial L}{\partial \mathbf{\Psi}}$ is the effective gradient operator and $\hat{\boldsymbol{H}} \equiv \frac{\partial^2 L}{\partial \mathbf{\Psi}^2}$ is the Hessian for the Lagrangian, which we will refer to as the Hessian in the following derivations for simplicity. To obtain their explicit expressions, we substitute Eq. (14) into $E[\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi})]$ and it follows, up to second order of t,

$$\begin{aligned}
E[\hat{\boldsymbol{F}}(\mathbf{\Psi} + t\mathbf{\Phi})] &= E[\mathbf{\Psi} + t\mathbf{\Phi} - \frac{1}{2}t^2\boldsymbol{V}\mathbf{\Psi}] \\
&= E[\mathbf{\Psi}] + t\frac{\partial E}{\partial \mathbf{\Psi}}[\mathbf{\Phi}] + \frac{1}{2}t^2\Big(\Big\langle\mathbf{\Phi}, \frac{\partial^2 E}{\partial \mathbf{\Psi}^2}\mathbf{\Phi}\Big\rangle \\
&\quad - \frac{\partial E}{\partial \mathbf{\Psi}}[\boldsymbol{V}\mathbf{\Psi}]\Big).
\end{aligned} \tag{16}$$

Comparing Eqs. (15) and (16), we obtain

$$\frac{\partial L}{\partial \mathbf{\Psi}} = \frac{\partial E}{\partial \mathbf{\Psi}} \tag{17}$$

and

$$\langle\mathbf{\Phi}, \hat{\boldsymbol{H}}[\mathbf{\Phi}]\rangle = \Big\langle\mathbf{\Phi}, \frac{\partial^2 E}{\partial \mathbf{\Psi}^2}\mathbf{\Phi}\Big\rangle - \frac{\partial E}{\partial \mathbf{\Psi}}[\boldsymbol{V}\mathbf{\Psi}]. \tag{18}$$

The second term on the right hand side of Eq. (18) can be rewritten as

$$\begin{aligned}
\frac{\partial E}{\partial \mathbf{\Psi}}[\boldsymbol{V}\mathbf{\Psi}] &= \Big\langle\frac{\partial E}{\partial \mathbf{\Psi}}, \boldsymbol{V}\mathbf{\Psi}\Big\rangle = \sum_{i\sigma}(2\hat{h}_s^\sigma\psi_i^\sigma, \sum_j V_{ij}^\sigma\psi_j^\sigma) \\
&= 2\sum_{\sigma ij}(\varphi_i^\sigma, \varphi_j^\sigma)(\hat{h}_s^\sigma\psi_i^\sigma, \psi_j^\sigma) = 2\langle\mathbf{\Phi}, \hat{\mathbf{\Gamma}}\mathbf{\Phi}\rangle,
\end{aligned} \tag{19}$$

where $(\hat{\mathbf{\Gamma}}\Phi)_{i\sigma} = \sum_j(\hat{h}_s^\sigma\psi_i^\sigma, \psi_j^\sigma)\varphi_j^\sigma = \sum_j \gamma_{ij}^\sigma\varphi_j^\sigma$. Here, $\gamma_{ij}^\sigma \equiv (\hat{h}_s^\sigma\psi_i^\sigma, \psi_j^\sigma)$ and $\hat{h}_s^\sigma$ is the one-electron effective Hamiltonian of spin σ. Substituting Eq. (19) into Eq. (18), we have

$$\langle\mathbf{\Phi}, \hat{\boldsymbol{H}}[\mathbf{\Phi}]\rangle = \Big\langle\mathbf{\Phi}, \Big(\frac{\partial^2 E}{\partial \mathbf{\Psi}^2} - 2\hat{\mathbf{\Gamma}}\Big)\mathbf{\Phi}\Big\rangle, \tag{20}$$

which implies the Hessian is

$$\hat{\boldsymbol{H}} = \frac{\partial^2 E}{\partial \mathbf{\Psi}^2} - 2\hat{\mathbf{\Gamma}}. \tag{21}$$

Based on the constrained gradient and Hessian, the gentlest ascent dynamics for KS-DFT is given by

$$\frac{d\mathbf{\Psi}}{dt} = -\frac{\partial L}{\partial \mathbf{\Psi}} + 2\langle\frac{\partial L}{\partial \mathbf{\Psi}}, \mathbf{\Phi}\rangle\mathbf{\Phi} + \mathbf{\Lambda}\mathbf{\Psi}, \tag{22}$$

$$\frac{d\mathbf{\Phi}}{dt} = -\hat{\boldsymbol{H}}\mathbf{\Phi} + \mu\mathbf{\Phi} + \boldsymbol{K}\mathbf{\Psi}. \tag{23}$$

Here, μ is a Lagrange multiplier for the normalization constraint $\langle\mathbf{\Phi}, \mathbf{\Phi}\rangle = 1$, which attributes $\mathbf{\Phi}$ with the meaning of a direction; $\mathbf{\Lambda}$ and $\boldsymbol{K}$ are Lagrange multiplier matrices, for the orthonormality constraint of $\mathbf{\Psi}$ and the constraint that $\mathbf{\Phi}$ resides on the virtual space, respectively. Here, the last constraint prevents the components of $\mathbf{\Phi}$ to fall onto the space spanned by the orbitals $\mathbf{\Psi}$.

As a comment, Eqs. (22) and (23) can be easily modified to target saddle points of higher indices. Taking index-2 saddle points, for example, instead of having only one auxiliary perturbation direction, the GAD dynamics would require two of them, which detect the directions of the smallest two eigenvalues of the Hessian and flip the force direction in these two modes. Therefore, the formula is given by

$$\frac{d\mathbf{\Psi}}{dt} = -\frac{\partial L}{\partial \mathbf{\Psi}} + 2\langle\frac{\partial L}{\partial \mathbf{\Psi}}, \mathbf{\Phi_1}\rangle\mathbf{\Phi_1} + 2\langle\frac{\partial L}{\partial \mathbf{\Psi}}, \mathbf{\Phi_2}\rangle\mathbf{\Phi_2} + \mathbf{\Lambda}\mathbf{\Psi}, \tag{24}$$

$$\frac{d\mathbf{\Phi_1}}{dt} = -\hat{\boldsymbol{H}}\mathbf{\Phi_1} + \mu_{11}\mathbf{\Phi_1} + \boldsymbol{K_1}\mathbf{\Psi}, \tag{25}$$

$$\frac{d\mathbf{\Phi_2}}{dt} = -\hat{\boldsymbol{H}}\mathbf{\Phi_2} + \mu_{22}\mathbf{\Phi_2} + \mu_{12}\mathbf{\Phi_2} + \boldsymbol{K_2}\mathbf{\Psi}. \tag{26}$$

Here, $\mathbf{\Phi_1}$ and $\mathbf{\Phi_2}$ are the auxiliary perturbation directions; μ_{11} and μ_{22} are the Lagrange multipliers for the normalization constraints of $\langle\mathbf{\Phi_i}, \mathbf{\Phi_i}\rangle = 1$; μ_{12} is the Lagrange multiplier for the constraint of $\langle\mathbf{\Phi_1}, \mathbf{\Phi_2}\rangle = 0$ so that the two directions are not parallel; $\boldsymbol{K_i}$ is the Lagrange multiplier matrix for the constraint that $\mathbf{\Phi_i}$ resides on the virtual space.

The GAD equations for higher index saddle points can be constructed similarly. As can be seen, the implementation of these dynamics is essentially the same as the GAD that targets index-1 saddle points. In this paper, we focus on the latter (Eqs. (22) and (23)) as a proof of concept of the GAD in the Kohn-Sham context.

In terms of the orbital functions, after some algebra and with the right hand side of Eqs. (22) and (23) multiplied by $\frac{1}{2}$ for the formal elegancy (this does not affect the dynamics), they reduce to[50]

$$\frac{d\psi_i^\sigma(\boldsymbol{r})}{dt} = -\hat{h}_s^\sigma \psi_i^\sigma(\boldsymbol{r}) + 2\sum_{j\sigma'}(\hat{h}_s^\sigma \psi_j^{\sigma'}, \varphi_j^{\sigma'})\varphi_i^\sigma(\boldsymbol{r}) + \sum_j \lambda_{ij}^\sigma \psi_j^\sigma(\boldsymbol{r}), \tag{27}$$

$$\frac{d\varphi_i^\sigma(\boldsymbol{r})}{dt} = -\hat{h}_s^\sigma \varphi_i^\sigma(\boldsymbol{r}) - \sum_{\sigma'} \hat{V}^{\sigma\sigma'}[2\sum_j \psi_j^{\sigma'}\varphi_j^{\sigma'}]\psi_i^\sigma(\boldsymbol{r}) + \sum_j \gamma_{ij}^\sigma \varphi_j^\sigma(\boldsymbol{r}) + \mu\varphi_i^\sigma(\boldsymbol{r}) + \sum_j \kappa_{ij}^\sigma \psi_j^\sigma(\boldsymbol{r}). \tag{28}$$

Here, $\hat{V}^{\sigma\sigma'}$ is the derivative of the Kohn-Sham effective potential with respect to the electron density, given by

$$\hat{V}^{\sigma\sigma'}[f] = \int \frac{\delta v_{\text{eff}}^\sigma(\boldsymbol{r})}{\delta \rho^{\sigma'}(\boldsymbol{r}')} f(\boldsymbol{r}')d\boldsymbol{r}', \tag{29}$$

where $v_{\text{eff}}^\sigma(\boldsymbol{r})$ is the effective one-electron potential.

A few remarks are in order. Eq. (27) without the second term on the right hand side, i.e.,

$$\frac{d\psi_i^\sigma(\boldsymbol{r})}{dt} = -\hat{h}_s^\sigma \psi_i^\sigma(\boldsymbol{r}) + \sum_j \lambda_{ij}^\sigma \psi_j^\sigma(\boldsymbol{r}), \tag{30}$$

is just the well-known steepest descent dynamics for KS-DFT, which had been designed and applied to find the local minima, including the ground state.[46–48] The gentlest ascent dynamics swaps the force direction in the softest mode such that the dynamics converges to the index-1 saddle point instead. This is potentially a useful way to scan the landscape of the KS density functionals.

For practical implementation of the GAD for the KS-DFT, the orbital $\boldsymbol{\Psi}$ and perturbation direction $\boldsymbol{\Phi}$ are projected onto atomic bases, and each orbital is represented by a vector in the coefficient matrix. Consequently, Eqs. (27) and (28) become matrix equations. Here, we implement an $O(N^4)$ algorithm to compute the GAD equations (the computational bottleneck is given by evaluation of $\hat{V}^{\sigma\sigma'}$), see the supplementary material for details.[50] Although the implementation can be improved in various ways, for example, an $O(N^3)$ algorithm can be achieved by using resolution of identity and density fitting techniques,[51,52] it suffices for our purpose of presenting the idea of GAD and testing on small systems. Starting with appropriate initial guesses, the equations are evolved using a projected Euler method: the Lagrange multipliers are dropped in the ordinary differential equations (ODEs); instead, to impose the constraints, at each time step, the ψ_i^σ's and φ_i^σ's are transformed using the Gram-Schmidt procedure to satisfy the constraints. The dynamics is evolved till it converges (the Kohn-Sham energy $E[\boldsymbol{\Psi}]$ is used to track convergence).

The stationary points of the Kohn-Sham density functional theory satisfy

$$\frac{\partial E}{\partial \boldsymbol{\Psi}} = 0. \tag{31}$$

The stability of the stationary solutions is determined by the positiveness of the Hessian, given in Eq. (21). Let ϵ denote the smallest eigenvalue of the Hessian. The orbitals $\boldsymbol{\Psi}$ is a local minimum if ϵ is positive; it is a saddle point if ϵ is negative. If ϵ is zero, it can be viewed as a degenerate local minimum with a saddle point. To numerically evaluate ϵ and hence the stability of the Kohn-Sham stationary solutions, we can evolve the dynamics Eq. (28) while fixing ψ_i^σ, until all the φ_i^σ's converges. This is essentially a power method for finding the smallest eigenvalue of a matrix.[53] As discussed above, in practice, the Lagrange multipliers are dropped and replaced by the Gram-Schmidt procedures at each time step. The $\boldsymbol{\Phi}$ consisting of the converged φ_i^σ's is the direction corresponding to the smallest eigenvalue of the Hessian ϵ_{evol}, and furthermore, $\frac{1}{2}\epsilon_{\text{evol}} = \frac{1}{2}\langle \boldsymbol{\Phi}, \hat{\boldsymbol{H}}[\boldsymbol{\Phi}]\rangle$. In the orbital language, this becomes

$$\frac{1}{2}\epsilon_{\text{evol}} = \sum_{\sigma i}(\varphi_i^\sigma, \hat{h}_s^\sigma \varphi_i^\sigma) + \sum_{\sigma\sigma' i}(\varphi_i^\sigma, \hat{V}^{\sigma\sigma'}[2\sum_j \psi_j^{\sigma'}\varphi_j^{\sigma'}]\psi_i^\sigma) - \sum_{\sigma ij}\gamma_{ij}^\sigma(\varphi_i^\sigma, \varphi_j^\sigma). \tag{32}$$

$\frac{1}{2}\epsilon_{\text{evol}}$ can also be validated by comparing it with a central finite difference approximation (where t is chosen to be small):

$$\frac{1}{2}\epsilon_{\text{diff}} = \frac{1}{2t^2}\big(E[\hat{\boldsymbol{F}}(\boldsymbol{\Psi} + t\boldsymbol{\Phi})] - E[\hat{\boldsymbol{F}}(\boldsymbol{\Psi} - t\boldsymbol{\Phi})]\big). \tag{33}$$

For saddle points, to determine their indices, i.e., the number of descending directions, we can also directly diagonalize the projected Hessian onto the virtual space. This is similar in spirit to the SCF stability analysis in the literature.[54–57] Alternatively, the dynamics of Eq. (28) can be extended to calculate more than one eigenvalue, which is analogous to the subspace iteration method in numerical linear algebra,[53] for which we will not go into details.

To diagonalize the projected Hessian, we define a projection operator $\hat{\boldsymbol{P}} = \begin{bmatrix} \hat{\boldsymbol{P}}^\alpha & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{P}}^\beta \end{bmatrix}$, where

$$\hat{\boldsymbol{P}}^\sigma = \begin{bmatrix} \hat{p}^\sigma & 0 & \cdots & 0 \\ 0 & \hat{p}^\sigma & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \hat{p}^\sigma \end{bmatrix}, \tag{34}$$

with $\hat{p}^\sigma = \sum_i |\psi_i^\sigma\rangle\langle\psi_i^\sigma|$. Then, $\hat{\boldsymbol{I}} - \hat{\boldsymbol{P}}$ is the projection operator which eliminates the occupied orbital space components. The eigenvalues of $(\hat{\boldsymbol{I}} - \hat{\boldsymbol{P}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{I}} - \hat{\boldsymbol{P}})$ then correspond to the curvature of ascending or descending directions when the stationary solution $\boldsymbol{\Psi}$ is perturbed. We note that there exist multiple eigenvectors with zero eigenvalue, which form a span of $\hat{\boldsymbol{P}}$. Nevertheless, the smallest nonzero eigenvalue should give exactly ϵ.

III. RESULTS AND DISCUSSIONS

We select several atoms and small molecules, on which our gentlest ascent dynamics is tested. The functional used is LDA with VWN5[58] as the correlation functional. Other approximate functionals lead to similar results, since the GAD algorithm is not functional dependent. The results on BLYP[59,60] and B3LYP[60,61] functionals as representative of generalized gradient approximations (GGAs) and hybrid functionals can be found in the supplementary material.[50] For all the DFT calculations, the Gaussian atomic basis set 6–31++G** is used except for H and He, for which aug-cc-pVDZ is used. The basis used for the calculation of equation-of-motion coupled cluster singles and doubles (EOM-CCSD) is aug-cc-pVTZ. The structures of the molecules are taken from the G2-97 set[62] (except H_2 whose bondlength here is chosen to be 1 Å, which is slightly stretched from its ground state geometry, but close to its excited state structure), corresponding to their ground state energy minima. All the DFT calculations are done with an in–house built quantum chemistry software package QM^{4}D,[63] while the EOM-CCSD calculation is performed using the GAUSSIAN package.[64] In the following, if not specially noted, the GAD equations are referred to as Eqs. (22) and (23) that target the index-1 saddle point.

In Table I, we compare the energies obtained by the non-Aufbau ΔSCF (for which we will refer to as ΔSCF for simplicity in the following context) and gentlest ascent dynamics. Here, we focus on the total energy comparison between ΔSCF and GAD, rather than the excitation energy of these two methods compared with other approaches. For the latter, we refer the reader to the previous literature.[27,31–33] We also note that for singlet excitations of closed shell systems, both GAD and ΔSCF can only capture the spin broken-symmetry state, which is a mixture of the spin pure singlet and triplet state. In ΔSCF, usually the spin purification procedure is needed to obtain the purified singlet excitation energy.[29,32] Yet, this is not the focus of our paper. As can be seen from Table I, our GAD approach gives essentially the same energies compared to the conventional non-Aufbau ΔSCF for these atoms and molecules. They may slightly differ since the two approaches are implemented differently, yet such difference in energy is way below the chemical accuracy. For some molecules such as HF, the ΔSCF fails to converge. Shown in Figure 1 is the comparison of the convergence in the case of HF for both the GAD and the ΔSCF procedure. As can be seen, while the ΔSCF fails to converge and the energy oscillates around a certain value, the gentlest ascent dynamics almost monotonically directs itself to the desired state. The GAD yields a more robust tool for finding the first excited state.

We also remark that in the case of Li atom, the GAD is able to find two excited states, with $(1s)^2(2p)^1$ corresponding to the first excited state and $(1s)^1(2s)^2$ corresponding to a core excitation which is much higher in energy. As will be shown in Table IV, both of these states are index-1 saddle points.

Let us stress that similar to the steepest descent dynamics, the GAD only guarantees local convergence, i.e., it is crucial to construct appropriate initial guess so as to converge to the desired state. Although using a random guess may still converge to some state, it might take many iterations to converge, and furthermore, the converged state might not be the excited state we aim to find. Therefore, it is more advantageous to start with an initial guess that is close by the desired state.

For the systems studied here, we use the occupied and virtual KS-orbitals of the ground state to form an initial guess. For example, in the Li case, the ground state configuration is $(1s)^2(2s)^1$, with a pair of electrons occupying the $1s$ orbitals and an unpaired α electron occupying the $2s$ orbital. If we target the first excited state, the paired $1s$ orbitals and the

TABLE I. Comparison of index-1 saddle point energies obtained by non-Aufbau ΔSCF and gentlest ascent dynamics.

	Configuration	E_{GAD} (a.u.)	E_{diff}[a] (10^{-8} a.u.)
H	2s	−0.127 664 22	0.1
He	$(1s)^1(2s)^1$	−2.076 104 93	0.2
Li	$(1s)^2(2p)^1$	−7.279 291 90	122
	$(1s)^1(2s)^2$	−5.229 653 96	0.5
Be	$(1s)^2(2s)^1(2p)^1$	−14.321 785 75	81
H_2	$(1\sigma)^1(1\sigma^*)^1$	−0.795 607 78	0.0
Li_2	$(\sigma_{1s})^2(\sigma^*_{1s})^2(\sigma_{2s})^1(\sigma^*_{2s})^1$	−14.673 647 10	0.5
CO	1st	−112.167 761 66	48
OH	gs[b]	−75.157 207 74	...[c]
	4th[d]	−74.844 085 40	149
HF	1st	−99.416 976 46	...[e]
H_2O	1st	−75.598 200 55	0.5

[a] $E_{diff} = E_{GAD} - E_{\Delta SCF}$, where $E_{\Delta SCF}$ is the energy by non-Aufbau ΔSCF.
[b] The ground state of OH is degenerate; see the analysis in the main text.
[c] ΔSCF cannot converge if forcing the occupation number of (11101) for β electrons.
[d] As shown in Table IV, the 4th excited state is an index-1 saddle point.
[e] ΔSCF oscillates in energy and does not converge.

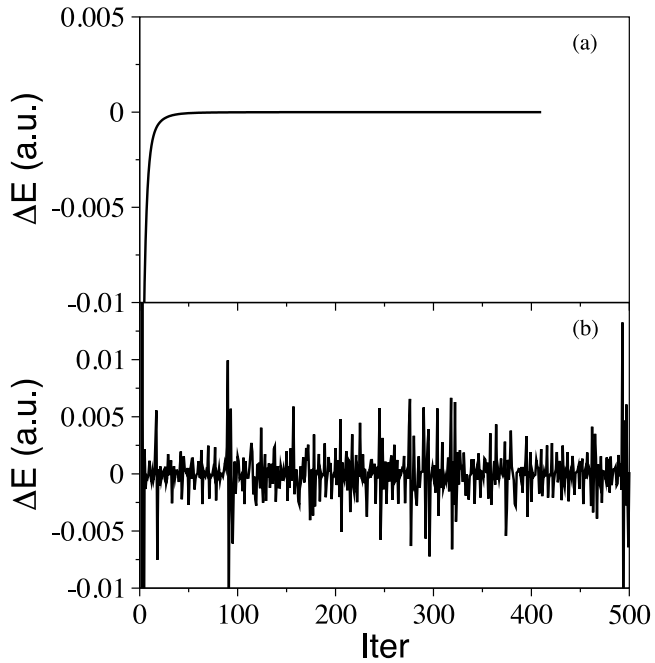

FIG. 1. The iteration process of HF using (a) gentlest ascent dynamics and (b) non-Aufbau ΔSCF. The plotted function ΔE is the energy difference between two neighboring iterations.

lowest unoccupied molecular orbital (LUMO) of α electron ($2p$ orbital) are used as initial guess, which is close to the true first excited state (they are different due to orbital relaxation). If we aim at the core excited state, then the initial guess is formed by swapping the $1s$ and $2s$ orbitals of the β electron. This enables our gentlest ascent dynamics to converge to the nearby index-1 saddle point of $(1s)^1(2s)^2$. For other atoms and molecules, initial guesses are formed similarly. Using a fixed time stepping, it usually takes hundreds of iterations to find the desired excited state. We remark that the large number of steps is partially due to the stringent convergence criteria we take (10^{-10} a.u. as energy threshold). It is also possible to accelerate the convergence by using variable time stepping, which we would leave to future works.

Similar to the Li atom but more interesting is the case of the OH molecule, where the ground state is degenerate. By swapping the 4th (highest occupied) and 5th (lowest unoccupied) orbitals of the beta electrons at its ground state as initial guess of the molecular orbitals, the GAD converges to the same state as the ground state (this is verified by comparing the GAD converged energy and density with the ground state–they are the same within numerical errors). Furthermore, once the ground state Hessian is diagonalized, as shown in Table IV, the smallest nonzero eigenvalue is given by a tiny positive number, which suggests that it is likely to be zero within numerical error. Therefore, the energy landscape around the ground state of OH molecule is rather flat, and its ground state can be viewed as overlapping with the first excited state. Note that because of this, the standard non-Aufbau ΔSCF will not converge to this state, showing again the robustness of the GAD method. In addition, it is also worth noticing that the 4th excited state of OH, which corresponds to the transition of 5th (highest occupied) to 6th (lowest unoccupied) of the alpha electron from the ground state, is an index-1 saddle point, which we will make further comment in the sequel.

Shown in Table II is the comparison between ϵ_{diff} and ϵ_{evol} of some example atoms and molecules, evaluated at their ground states as well as their excited states. As can be seen, ϵ_{evol} essentially agree with ϵ_{diff}. For the ground states, they give positive eigenvalues, indicating that these states are local minima as expected. For excited states, ϵ's are negative, which also reasonably suggests that there are descending directions at these solutions.

We emphasize one unique feature of the energy landscape of the Kohn-Sham energy functional, in the local density approximation of exchange-correlation at least, indicated by our numerical results. For the perhaps more familiar free energy landscape of chemical systems, it is expected that a saddle point separates two attraction basins associated with different local minima. Hence, both local minima have energy lower than the saddle point, which implies that the first excited state would correspond to a local minimum, rather than an index-1 saddle point. However, in the Kohn-Sham case, if we start at an index-1 saddle point $\mathbf{\Psi}$ and follow the descending mode given by $\mathbf{\Phi}$, no matter which direction we follow (in the sense of perturbing the state as $\hat{F}(\mathbf{\Psi} + t\mathbf{\Phi})$ or $\hat{F}(\mathbf{\Psi} - t\mathbf{\Phi})$), following the energy descending direction, we end up in the same ground state. Therefore, the index-1 saddle point and the ground state, linked by the forward and backward energy descending directions, form a loop. This is very different from the usual picture that the index-1 saddle point bridges the reactant and product states as a transition state in chemical reactions. This is perhaps caused by the fact that the Kohn-Sham energy functional is defined on a curved configuration space (mathematically speaking, a Grassmannian manifold) rather than a flat one.

Another interesting feature of the Kohn-Sham energy landscape is that higher excited states may not necessarily correspond to saddle points of higher indices, although many of them are.[65] This is illustrated by the example of the OH in Table III, where we present the four lowest-lying excited states of OH of the doublet symmetry (5 α electrons and 4 β electrons). As can be seen, the 2nd and 3rd excited states are index-2 saddle points, whereas there is an index-1 saddle point lying above them. Similar situation is observed for Li atom excitations, where the core excitation ($1s^12s^2$) is an index-1 saddle point, but definitely not the first excited state (the first excited state is $1s^22p^1$ since it is lower in energy). All these

TABLE II. Comparison of $\frac{1}{2}\epsilon_{\text{diff}}$ of Eq. (33) and $\frac{1}{2}\epsilon_{\text{evol}}$ of Eq. (32) at stationary solutions of some example systems.

	Configuration[a]	$\frac{1}{2}\epsilon_{\text{diff}}$ (a.u.)	$\frac{1}{2}\epsilon_{\text{evol}}$ (a.u.)
H	1s	0.3064	0.306 4
	2s	−0.4401	−0.440 1
	3s	−0.3906	−0.390 6
He	$(1s)^2$	0.6251	0.625 1
	$(1s)^1(2s)^1$	−0.8701	−0.870 2
Li	$(1s)^2(2s)^1$	0.0805	0.080 5
	$(1s)^2(2p)^1$	−0.0298	−0.029 8
	$(1s)^1(2s)^2$	−2.401	−2.401
Be	$(1s)^2(2s)^2$	0.0581	0.058 1
	$(1s)^2(2s)^1(2p)^1$	−0.1078	−0.107 8
H_2	$(1\sigma)^2$	0.2178	0.217 8
	$(1\sigma)^1(1\sigma^*)^1$	−0.3139	−0.313 9
	$(1\sigma^*)^2$	−0.5097	−0.509 7
Li_2	$(\sigma_{1s})^2(\sigma^*_{1s})^2(\sigma_{2s})^2$	0.0230	0.023 0
	$(\sigma_{1s})^2(\sigma^*_{1s})^2(\sigma_{2s})^1(\sigma^*_{2s})^1$	−0.0292	−0.029 2
CO	gs	0.1893	0.189 3
	1st	−0.2411	−0.241 1
OH	gs[b]	0.0033	0.000 005
	2nd	−0.1327	−0.132 7
	3th	−0.3593	−0.359 3
	4th	−0.3939	−0.393 9
HF	gs	0.3073	0.307 3
	1st	−0.4572	−0.462 1
H_2O	gs	0.2188	0.218 8
	1st	−0.3228	−0.322 8
CH_2O	gs	0.1046	0.106 3
	1st	−0.1400	−0.139 2
CH_3CHO	gs	0.1208	0.121 8
	1st	−0.1559	−0.156 1

[a]When the number of electrons becomes large, the configuration is abbreviated as ground state (gs), first excited state (1st) or second excited state (2nd), and so on. All the excited states except CH_2O and CH_3CHO are obtained by the GAD; the index-1 saddle points (including but not limited to first excited states) are obtained by Eqs. (22) and (23), while the index-2 saddle points are obtained by Eqs. (24)–(26). For some larger molecules like CH_2O and CH_3CHO, here we compute ϵ_{diff} and ϵ_{evol} at their first excited states obtained by ΔSCF and the numbers show that they are also saddle points.
[b]First excited state is viewed as the same state as the ground state and omitted here.

TABLE III. Saddle point indices of the 4 lowest-lying excited states of OH and comparison of the excited state energies between GAD, TDDFT(TDLDA), and EOM-CCSD.

	Configuration	Saddle point index	Excitation energy (eV) GAD	TD-DFT	EOM-CCSD
1	$\beta: 4 \to 5$	0[a]	0	0.005	0.020
2	$\beta: 3 \to 5$	2	3.80	3.85	4.12
3	$\beta: 4 \to 6$[b]	2	8.01	6.61	7.69
4	$\alpha: 5 \to 6$[c]	1	8.85	7.39	8.45

[a]The ground state has degeneracy. The smallest eigenvalue of the Hessian is zero within numerical error. Here, we denote this state as index-0 saddle point for the ease of notation. Its excitation energy by the GAD is zero due to the degeneracy. Note that TDDFT and EOM-CCSD predict a tiny excitation energy. This is an artifact due to the ansatz of spin unrestriction. It is supposed to be degenerate with the ground state, yet calculated to be slightly higher in energy due to the fact that the orbitals of different spins do not share the same spatial function.
[b]TD-DFT gives a mixed excitation of $\beta: 4 \to 6$ (97%) and $\alpha: 5 \to 6$ (3%).
[c]TD-DFT gives a mixed excitation of $\beta: 4 \to 6$ (3%) and $\alpha: 5 \to 6$ (97%).

examples suggest that the Kohn-Sham energy landscape is much more complicated than expected, which is worth further exploration.

In Table IV, we further present the 5 smallest eigenvalues (including zero without counting its degeneracy) of the projected Hessian (multiplied by $\frac{1}{2}$ for comparison with $\frac{1}{2}\epsilon_{\text{evol}}$) for some of these systems. This is done by directly diagonalizing $\frac{1}{2}(\hat{\boldsymbol{I}} - \hat{\boldsymbol{P}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{I}} - \hat{\boldsymbol{P}})$ in the atomic basis. As can be seen, for each of the first excited states, there is only one negative eigenvalue, indicating that these solutions are all index-1 saddle points. Moreover, the smallest nonzero eigenvalues in Table III have perfect agreement with $\frac{1}{2}\epsilon_{\text{evol}}$'s in Table II, suggesting that the evolution approach is accurate in predicting the smallest nonzero eigenvalue of the Hessian.

TABLE IV. The smallest 5 eigenvalues (including zero without counting its degeneracy) of the projected Hessian (multiplied by $\frac{1}{2}$ for comparison with $\frac{1}{2}\epsilon_{\text{evol}}$) at stationary solutions of some example systems.

	Configuration	Eigenvalue (a.u.) 1	2	3	4	5
H	1s	0.0000	0.306 4	0.410 1	0.4101	0.4101
	2s	−0.4401	0.000 0	0.076 6	0.0766	0.0766
	3s	−0.3906	−0.056 7	0.000 0	0.0012	0.0014
He	$(1s)^2$	0×2[a]	0.625 1	0.742 7	0.8427	0.8427
	$(1s)^1(2s)^1$	−0.8702	0×2[a]	0.197 6	0.1976	0.1976
Li	$(1s)^2(2s)^1$	0×5[a]	0.080 5	0.080 5	0.0805	0.1123
	$(1s)^2(2p)^1$	−0.0298	0×5[a]	0.000 3	0.0009	0.0480
	$(1s)^1(2s)^2$	−2.401	0×5[a]	0.032 8	0.0328	0.0328
Be	$(1s)^2(2s)^2$	0×8[a]	0.058 1	0.058 1	0.0581	0.2027
	$(1s)^2(2s)^1(2p)^1$	−0.1078	0×8[a]	0.000 2	0.0004	0.0862
H_2	$(1\sigma)^2$	0×2[a]	0.217 7	0.357 0	0.3682	0.4113
	$(1\sigma)^1(1\sigma^*)^1$	−0.3139	0×2[a]	0.083 2	0.1758	0.4716
	$(1\sigma^*)^2$	−0.5097	−0.148 5	0×2[a]	0.1034	0.1266
Li_2	gs	0×18[a]	0.022 9	0.036 8	0.0368	0.0510
	1st	−0.0292	0×18[a]	0.021 1	0.0211	0.0534
CO	gs	0×98[a]	0.189 3	0.189 3	0.2612	0.3149
	1st	−0.2411	0×98[a]	0.000 005	0.0890	0.0898
OH	gs	0×41[a]	0.000 002	0.143 4	0.2383	0.2716
	2nd	−0.1327	−0.132 7	0×41[a]	0.2339	0.2339
	3th	−0.3593	−0.359 3	0×41[a]	0.1313	0.1313
	4th	−0.3939	0×41[a]	0.000 008	0.1237	0.1284
HF	gs	0×50[a]	0.307 3	0.307 3	0.3301	0.3301
	1st	−0.4621	0×50[a]	0.000 003	0.1284	0.1917
H_2O	gs	0×50[a]	0.218 8	0.240 2	0.2846	0.2892
	1st	−0.3228	0×50[a]	0.082 2	0.0877	0.1684

[a]There are degenerate zero eigenvalues, with the number showing the degeneracy.

IV. CONCLUDING REMARKS

In this paper, we have introduced the constrained gentlest ascent dynamics to obtain the first excited states of approximate density functionals as index-1 saddle points, which can be used as a robust alternative to compute the first excitation energies. It can be easily generalized to target higher excited states (or higher index saddle points). Furthermore, at stationary solutions, the underlying dynamics for the perturbation direction can be used to compute the smallest eigenvalue of the Hessian. This provides an efficient tool to explore the energy landscape. Although it may take hundreds of iterations to find the first excited state, it suffices for our purpose of proof of concept. Further modifications with line search and variable time stepping as well as other related techniques can be made to accelerate the convergence. Compared with the MOM method which takes less number of iterations but relies on the initial guess, our GAD method can be regarded as a more robust alternative that trades off the computational time in return for a guaranteed convergence.

ACKNOWLEDGMENTS

Support from the National Science Foundation (Grant No. CHE-13-62927) (C.L.) and (Grant Nos. DMS-1312659 and DMS-1454939) (J.L.) is gratefully appreciated. W.Y. appreciates the support as part of the Center for the Computational Design of Functional Layered Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0012575. The work of J.L. is also supported in part by the Alfred P. Sloan Foundation.

[1]W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2]A. D. Becke, J. Chem. Phys. **140**, 18A301 (2014).
[3]A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).
[4]E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
[5]E. K. U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985).
[6]W. Kutzelnigg, J. Mol. Struct.: THEOCHEM **181**, 33 (1988).
[7]R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, Mol. Phys. **35**, 771 (1978).
[8]K. P. Lawley and B. O. Roos, Adv. Chem. Phys. **69**, 399 (1987).
[9]K. Emrich, Nucl. Phys. A **351**, 379 (1981).
[10]H. Sekino and R. J. Bartlett, Int. J. Quantum Chem., Symp. **26**, 255 (1984).
[11]J. Geertsen, M. Rittby, and R. J. Bartlett, Chem. Phys. Lett. **164**, 57 (1989).
[12]A. K. Theophilou, J. Phys. C **12**, 5419 (1979).
[13]L. Fritsche, Phys. Rev. B **33**, 3976 (1986).
[14]E. K. U. Gross, L. N. Oliveira, and W. Kohn, Phys. Rev. A **37**, 2805 (1988).
[15]E. K. U. Gross, L. N. Oliveira, and W. Kohn, Phys. Rev. A **37**, 2809 (1988).
[16]A. Görling, Phys. Rev. A **59**, 3359 (1999).
[17]M. Levy and A. Nagy, Phys. Rev. Lett. **83**, 4361 (1999).
[18]P. W. Ayers and M. Levy, Phys. Rev. A **80**, 012508 (2009).
[19]P. W. Ayers, M. Levy, and A. Nagy, Phys. Rev. A **85**, 042518 (2012).
[20]N. Ferré and X. Assfeld, J. Chem. Phys. **117**, 4119 (2002).
[21]P.-F. Loos and X. Assfeld, Int. J. Quantum Chem. **107**, 2243 (2007).

[22]T. Ziegler, M. Seth, M. Krykunov, J. Autschbach, and F. Wang, J. Chem. Phys. **130**, 154102 (2009).
[23]T. Ziegler, M. Krykunov, and J. Cullen, J. Chem. Theory Comput. **7**, 2485 (2011).
[24]J. Cullen, M. Krykunov, and T. Ziegler, Chem. Phys. **391**, 11 (2011).
[25]T. Ziegler, M. Krykunov, and J. Cullen, J. Chem. Phys. **136**, 124107 (2012).
[26]F. A. Evangelista, P. Shushkov, and J. C. Tully, J. Phys. Chem. A **117**, 7378 (2013).
[27]I. Seidu, M. Krykunov, and T. Ziegler, J. Phys. Chem. A **119**(21), 5107 (2015).
[28]J. P. Perdew and M. Levy, Phys. Rev. B **31**, 6264 (1985).
[29]T. Ziegler, A. Rauk, and E. J. Baerends, Theor. Chim. Acta **43**, 261 (1977).
[30]U. von Barth, Phys. Rev. A **20**, 1693 (1979).
[31]C.-L. Cheng, Q. Wu, and T. Van Voorhis, J. Chem. Phys. **129**, 124112 (2008).
[32]T. Kowalczyk, S. R. Yost, and T. V. Voorhis, J. Chem. Phys. **134**, 054128 (2011).
[33]S. R. Yost, T. Kowalczyk, and T. Van Voorhis, J. Chem. Phys. **139**, 174104 (2013).
[34]N. A. Besley, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **130**, 124308 (2009).
[35]A. J. W. Thom and M. Head-Gordon, Phys. Rev. Lett. **101**, 193001 (2008).
[36]G. M. Crippen and H. A. Scheraga, Arch. Biochem. Biophys. **144**, 462 (1971).
[37]G. T. Barkema and N. Mousseau, Phys. Rev. Lett. **77**, 4358 (1996).
[38]A. Pederson, S. F. Hafstein, and H. Jónsson, SIAM J. Sci. Comput. **33**, 633 (2011).
[39]C. J. Cerjan and W. H. Miller, J. Chem. Phys. **75**, 2800 (1981).
[40]G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[41]R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson, J. Chem. Phys. **121**, 9776 (2004).
[42]A. Poddey and P. E. Blochl, J. Chem. Phys. **128**, 44107 (2008).
[43]L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).
[44]W. Gao, J. Leng, and X. Zhou, http://arxiv.org/pdf/1406.1991v1.pdf (2014).
[45]W. E and X. Zhou, Nonlinearity **24**, 1831 (2011).
[46]R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[47]M. C. Payne, J. D. Joannopoulos, D. C. Allan, M. P. Teter, and D. H. Vanderbilt, Phys. Rev. Lett. **56**, 2656 (1986).
[48]I. Štich, R. Car, M. Parrinello, and S. Baroni, Phys. Rev. B **39**, 4997 (1989).
[49]A. Samanta and W. E, J. Chem. Phys. **136**, 124104 (2012).
[50]See supplementary material at http://dx.doi.org/10.1063/1.4936411 for detailed derivations.
[51]B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, J. Chem. Phys. **71**, 3396 (1979).
[52]O. Vahtras, J. Almlof, and M. W. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).
[53]G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, 4th ed. (Johns Hopkins University Press, Baltimore, MD, 2013).
[54]R. Bauernschmitt and R. Ahlrichs, J. Chem. Phys. **104**, 9047 (1996).
[55]J. Čížek and J. Paldus, J. Chem. Phys. **47**, 3976 (1967).
[56]R. Seeger and J. A. Pople, J. Chem. Phys. **66**, 3045 (1977).
[57]G. Chambaud, B. Levy, and P. Millie, Theor. Chim. Acta **48**, 103 (1978).
[58]S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[59]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[60]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[61]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[62]L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).
[63]X. Hu, H. Hu, X. Zheng, X. Zeng, P. Wu, D. Peng, Y. Jin, L. Yu, and W. Yang, QM4D, http://www.qm4d.info.
[64]Gaussian 09, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
[65]G. M. J. Barca, A. T. B. Gilbert, and P. M. W. Gill, J. Chem. Phys. **141**, 111104 (2014).