# Unification of algorithms for minimum mode optimization

Yi Zeng,[1,a)] Penghao Xiao,[2,a)] and Graeme Henkelman[2,b)]
[1]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, USA*
[2]*Department of Chemistry and the Institute for Computational and Engineering Sciences, University of Texas at Austin, Austin, Texas 78712, USA*

(Received 6 November 2013; accepted 6 January 2014; published online 30 January 2014)

Minimum mode following algorithms are widely used for saddle point searching in chemical and material systems. Common to these algorithms is a component to find the minimum curvature mode of the second derivative, or Hessian matrix. Several methods, including Lanczos, dimer, Rayleigh-Ritz minimization, shifted power iteration, and locally optimal block preconditioned conjugate gradient, have been proposed for this purpose. Each of these methods finds the lowest curvature mode iteratively without calculating the Hessian matrix, since the full matrix calculation is prohibitively expensive in the high dimensional spaces of interest. Here we unify these iterative methods in the same theoretical framework using the concept of the Krylov subspace. The Lanczos method finds the lowest eigenvalue in a Krylov subspace of increasing size, while the other methods search in a smaller subspace spanned by the set of previous search directions. We show that these smaller subspaces are contained within the Krylov space for which the Lanczos method explicitly finds the lowest curvature mode, and hence the theoretical efficiency of the minimum mode finding methods are bounded by the Lanczos method. Numerical tests demonstrate that the dimer method combined with second-order optimizers approaches but does not exceed the efficiency of the Lanczos method for minimum mode optimization.  [http://dx.doi.org/10.1063/1.4862410]

## I. INTRODUCTION

An important challenge in chemical and materials science is the simulation of the dynamics of systems over long time scales. Most chemical reactions cannot be simulated directly with traditional molecular dynamics (MD) simulations due to their limited accessible time scales. However, using the harmonic approximation to transition state theory,[1,2] which is generally valid for solid state systems at moderate temperature, any reaction rate can be determined from the reactant and transition states. Once these states are located, the dynamics of rare-event systems can be evolved by a kinetic Monte Carlo algorithm over time scales much longer than is possible with MD. Thus the challenge of studying such chemical reactions can be transformed into the task of searching for the saddle points (transition states) connected to a given minimum (reactant). Due to the high dimensionality and expensive force evaluation of chemical systems, great efforts have been made in developing efficient saddle point searching algorithms. A family of these algorithms, called minimum-mode following algorithms, employs the following evolution equation:

$$\dot{x} = F(x) - 2(\hat{\tau}^{\mathsf{T}} F(x))\hat{\tau}, \tag{1}$$

where $x$ is the position vector, $-F(x)$ is the gradient of the potential energy surface $V(x)$, and $\hat{\tau}$ is the unit vector along the minimum curvature mode direction. We denote $\mathbf{H} = \nabla\nabla V$ and $\hat{\tau}$ as the unit eigenvector of $\mathbf{H}$ with the minimum eigenvalue. When $\mathbf{H}$ has only one negative eigenvalue, the above equation reverses the force along that eigenvector and converges to a first-order saddle point. The efficiency of the algorithm relies on an efficient update of $\hat{\tau}$.

In this paper, we focus on the $\hat{\tau}$ updating algorithms that avoid calculating the Hessian matrix $\mathbf{H}$, since it is typically too expensive to calculate directly for large chemical systems of interest. We examine several existing methods for estimating $\hat{\tau}$, including the Lanczos[3] method as used in the activation relaxation technique (ART-nouveau),[4] the dimer method,[5–7] Raleigh-Ritz minimization[8] as used in the hybrid eigenvector following method,[9] the shifted power iteration method as used in the gentlest ascent dynamics method,[10] and the locally optimal block preconditioned conjugate gradient (LOBPCG) method.[11] Here, these methods are unified into the same mathematical framework so that their relative theoretical efficiencies can be understood.

This paper is structured as follows. In Sec. II the Lanczos method is presented with the essential idea of the Krylov subspace behind the algorithm. Another widely used numerical scheme, the dimer method, is presented in Sec. III. We show that the dimer method searches for the lowest eigenvector of the Hessian within the same Krylov subspace as the Lanczos algorithm. In Sec. IV we present the power iteration method with Rayleigh quotient shift. This method is shown to be a special restarted case of the Lanczos algorithm for which the convergence rate is significantly slower in high dimensional space. In Sec. V we numerically compare the efficiency of the Lanczos algorithm, the dimer method coupled with three different optimizers, and the shifted power iteration method. Finally, we conclude in Sec. VI that the performance of methods such as the dimer, which are limited to a subspace of the same Krylov subspace as the Lanczos method, does

a)Y. Zeng and P. Xiao contributed equally to this work.
b)Author to whom correspondence should be addressed. Electronic mail: henkelman@cm.utexas.edu

 

not exceed Lanczos efficiency for finding the lowest curvature mode.

## II. LANCZOS ALGORITHM

The Lanczos algorithm is a specialized Arnoldi iteration method of eigenvalue calculations for symmetric matrices.[3] Before discussing the Lanczos algorithm, we first restate the eigenvalue problem as a minimization problem in Theorem 1, and then review the concept of the Krylov subspace and present the Lanczos algorithm based on Krylov subspace projection and search. In this section, we assume that the smallest eigenvalue has multiplicity 1.

**Theorem 1.** *Given a symmetric matrix* $\mathbf{H} \in \mathbb{R}^{m\times m}$, $v$ *is the eigenvector associated with the smallest eigenvalue* $\lambda$, *if and only if* $v$ *solves the minimization problem*,

$$\min_{b\in\mathbb{R}^m\setminus\{0\}} \frac{b^{\mathsf{T}}\mathbf{H}b}{b^{\mathsf{T}}b}. \tag{2}$$

*Proof.* Since $\mathbf{H}$ is a symmetric matrix, all eigenvectors $v_1, v_2, \cdots, v_m$ form an orthogonal basis of the space $\mathbb{R}^m$ and all eigenvalues of $\mathbf{H}$ are real numbers. We can write $b = \sum_i a_i v_i$, then we have,

$$\frac{b^{\mathsf{T}}\mathbf{H}b}{b^{\mathsf{T}}b} = \frac{\sum_i \lambda_i a_i^2}{\sum_i a_i^2}. \tag{3}$$

This function takes its minimum value when $b$ is equivalent to the eigenvector $v$. Thus, $v$ solves the minimization problem.

The other side of the statement follows from the uniqueness of the eigenvector corresponding to the smallest eigenvalue. □

*Remark 1. Such an optimal solution* $v$ *must be unique also due to the uniqueness of the eigenvector associates with the smallest eigenvalue.*

Having transformed the eigenvalue problem to a minimization problem, the minimization problem can be solved as follows. We first solve the optimization problem in a low dimension subspace, which can be done more easily than in the original space $\mathbb{R}^m$. Then we consider the optimal solution in a space with one more dimension to find a better solution. As we move to higher and higher dimensional subspaces, this optimal solution will converge to the true solution. Here the low dimensional subspace we will use is the Krylov subspace, $\mathcal{K}_n$, which is defined as in Remark 2.

*Remark 2. The nth order Krylov subspace* $\mathcal{K}_n$ *generated by a square matrix* $\mathbf{H} \in \mathbb{R}^{m\times m}$ *and a nonzero vector* $v \in \mathbb{R}^m$ *is defined as*

$$\mathcal{K}_n = \text{span}\{v, \mathbf{H}v, \mathbf{H}^2v, \cdots, \mathbf{H}^{n-1}v\}. \tag{4}$$

*When n is smaller than the rank r of the matrix* $\mathbf{H}$, $\mathcal{K}_n$ *is a n-dimensional space; when n is greater or equal to r,* $\mathcal{K}_n$ *is an r-dimensional space. The product* $\mathbf{H}v$ *is calculated by the following approximation:*

$$\mathbf{H}v = \frac{F(x) - F(x+v)}{||v||} + \mathcal{O}(\|v\|^2). \tag{5}$$

To solve the minimization problem in the Krylov subspace, a set of orthogonal basis is utilized. This basis can be obtained by the Gram-Schmidt process iteratively as shown in Refs. 12 and 13. We define such a basis for a $n$-dimensional Krylov subspace $\mathcal{K}_n$ as $\mathbf{Q}_n = [q_1, q_2, \cdots, q_n] \in \mathbb{R}^{m\times n}$. Therefore, any vector $b \in \mathcal{K}_n$ can be represented by, $b = \mathbf{Q}_n r$, where $r \in \mathbb{R}^n$. Then the minimization problem projected on a Krylov subspace $\mathcal{K}_n$ can be solved, as shown in Theorem 2.

**Theorem 2.** $\mathbf{Q}_n r$ *solves the minimization problem in the Krylov subspace* $\mathcal{K}_n$

$$\min_{b\in\mathcal{K}_n\setminus\{0\}} \frac{b^{\mathsf{T}}\mathbf{H}b}{b^{\mathsf{T}}b}, \tag{6}$$

*if r is the eigenvector corresponding to the smallest eigenvalue of the matrix* $\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n$.

*Proof.* Since any vector $b \in \mathcal{K}_n$ can be written as, $b = \mathbf{Q}_n r_b$

$$\frac{b^{\mathsf{T}}\mathbf{H}b}{b^{\mathsf{T}}b} = \frac{(\mathbf{Q}_n r_b)^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n r_b}{(\mathbf{Q}_n r_b)^{\mathsf{T}}(\mathbf{Q}_n r_b)} = \frac{r_b^{\mathsf{T}}(\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n)r_b}{r_b^{\mathsf{T}}r_b}. \tag{7}$$

By Theorem 1, the eigenvector $r$ associated with the smallest eigenvalue of the matrix $\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n$ solves this minimization problem. Therefore, the vector $\mathbf{Q}_n r$ solves the original optimization problem in the space $\mathcal{K}_n$. □

*Remark 3. By construction of the orthogonal basis* $\mathbf{Q}_n$, *the matrix* $\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n$ *is an upper Hessenberg matrix, e.g., an upper triangular matrix plus a nonzero first subdiagonal. It is also symmetric since* $\mathbf{H}$ *is symmetric and* $\mathbf{Q}_n$ *is an orthogonal basis. These two properties confirm that the matrix* $\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n$ *is tridiagonal.*

Finally, the eigenvalue problem of an unknown matrix $\mathbf{H}$ is transformed to an iterative series of calculations to find the smallest eigenvalue of a known low dimensional matrix $\mathbf{Q}_n^{\mathsf{T}}\mathbf{H}\mathbf{Q}_n$, which can be done efficiently, for example, by the QR algorithm. In theory, this scheme is guaranteed to converge if $n$ grows to the rank of the matrix $\mathbf{H}$, but in practice, the convergence will be faster than this bound.[14,15]

## III. DIMER METHOD

The dimer method is another iterative algorithm for minimum mode finding.[5] With improvements in the implementation,[6,7,16,17] the dimer method has become widely used in calculations of chemical reaction rates especially with forces evaluated from self-consistent electronic structure methods.

We note that the Raleigh-Ritz optimization method used in hybrid eigenvector following, as developed by the Wales group,[9] is based upon the same finite-difference gradient of the Rayleigh quotient that is used in the dimer method. So

while the methods are described using different language and have some minor differences in their implementation, they are equivalent for the purposes of this analysis. The same approach was also used previously by Voter to construct a bias potential for the acceleration of MD in his hyperdynamics method.[18]

In this section, we present the dimer method within the same theoretical framework as the Lanczos algorithm. Numerical comparisons have been previously made between these two algorithms,[17] but now, under this mathematical framework, we can compare their relative theoretical efficiency.

In the dimer method, the minimum curvature mode is determined by rotating a pair of images separated by a small distance $||v||$ according to the torque acting on the dimer. The torque is the force difference divided by the distance, and thus has the same form as $\mathbf{H}v$ in Eq. (5). Rotating along the torque direction is the mechanism by which the dimer method finds the minimum curvature mode in a specific subspace of the Krylov subspace.

At each iteration, the new direction $\tau$ that minimizes $\frac{\tau^{\mathsf{T}}\mathbf{H}\tau}{\tau^{\mathsf{T}}\tau}$ is found in the plane spanned by $\{v, \mathbf{H}v\}$, assuming the simplest case in which the SD direction is taken for the rotation plane. A second direction $\Theta$ is then constructed perpendicular to $v$ to form an orthogonal basis set $\mathbf{Q}_2 = [v, \Theta]$, reducing the optimization problem to two dimensions,

$$\Theta = \mathbf{H}v - (v^{\mathsf{T}}\mathbf{H}v)v, \tag{8}$$

$$\mathbf{A} = \mathbf{Q}_2^{\mathsf{T}}\mathbf{H}\mathbf{Q}_2 = \begin{pmatrix} v^{\mathsf{T}}\mathbf{H}v & v^{\mathsf{T}}\mathbf{H}\Theta \\ \Theta^{\mathsf{T}}\mathbf{H}v & \Theta^{\mathsf{T}}\mathbf{H}\Theta \end{pmatrix}. \tag{9}$$

Here, calculating $\mathbf{H}\Theta$ requires a second force evaluation, $F(x + \Theta)$. The $2 \times 2$ matrix $\mathbf{A}$ can then be diagonalized. The eigenvector $\tau$, which is expressed as $\tau = r_1 v + r_2\Theta$, is then the starting point of the next iteration. Note that

$$\mathbf{H}\tau = r_1\mathbf{H}v + r_2\mathbf{H}\Theta, \tag{10}$$

can be obtained without any further force evaluation since $\mathbf{H}v$ and $\mathbf{H}\Theta$ are already known, as pointed out in Ref. 7. The minimization is repeated in a sequence of two-dimensional spaces: span$\{v, \mathbf{H}v\}$, span$\{\tau, \mathbf{H}\tau\}, \cdots$, where span$\{v, \mathbf{H}v\}$ is the Krylov subspace $\mathcal{K}_2$. Also because

$$\mathbf{H}\tau = (r_1 - r_2(v^{\mathsf{T}}\mathbf{H}v))\mathbf{H}v + r_2\mathbf{H}^2v, \tag{11}$$

$\tau$ and $\mathbf{H}\tau$ are in the Krylov subspace $\mathcal{K}_3 = \text{span}\{v, \mathbf{H}v, \mathbf{H}^2v\}$. After the $n$th iteration, $n$ force evaluations have been made, which is the same number as in the Lanczos method. However, each two-dimensional space considered in the dimer method is a subspace of $\mathcal{K}_{n+1}$. Therefore, the dimer method convergence is theoretically limited by that of the Lanczos algorithm.

In previous descriptions of the dimer method, the above procedure was done by finding a unit vector $\hat{r} \in \mathbb{R}^2$ to minimize $\hat{r}^{\mathsf{T}}\mathbf{A}\hat{r}$, which is exactly the dimer energy in Ref. 5. The connection between minimizing the dimer energy and solving the eigenvector problem of Eq. (9) can be seen by expanding $\hat{r}$ by the unit eigenvectors of $\mathbf{A}$,

$$\hat{r}^{\mathsf{T}}\mathbf{A}\hat{r} = (a_x\hat{v}_x + a_y\hat{v}_y)^{\mathsf{T}}\mathbf{A}(a_x\hat{v}_x + a_y\hat{v}_y) = a_x^2\lambda_x + a_y^2\lambda_y, \tag{12}$$

where $\lambda_x$ and $\lambda_y$ are eigenvalues of the matrix $\mathbf{A}$.

**Theorem 3.** *The steepest descent dimer method is equivalent to the Lanczos algorithm with restarts every two steps; its theoretical efficiency must be lower than the Lanczos algorithm.*

When the rotation plane is determined by algorithms based on the previous search direction $\Theta$, as in the conjugate gradient (CG) algorithm, the dimer method is more efficient but remains under the limit of the Lanczos method. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm for updating $\Theta$, as well as the limited memory version (L-BFGS),[19] is even more efficient. When the initial Hessian is set as a constant times the identity matrix (the standard case in practice), BFGS/L-BFGS is also searching for the lowest mode in the subspace of $\mathcal{K}_{n+1}$ at iteration $n$. The LOBPCG method performs the minimization in a three-dimensional subspace, with one extra direction that is a linear combination of the previous directions.[11] The search space of LOBPCG is therefore still a subspace of the Krylov subspace, and its theoretical efficiency also cannot exceed the Lanczos limit.

## IV. POWER ITERATION METHOD WITH A RAYLEIGH SHIFT

Another method for finding the lowest curvature mode of the Hessian is the power iteration method, which has been employed in some recent saddle point searching algorithms.[10,20] In this section, we present the motivation and mechanism of the power iteration method with a Rayleigh shift, and prove the convergence of this method. Similar to the dimer method, we demonstrate that the search space of this method is contained in the Krylov subspace $\mathcal{K}_n$ for the minimization problem of Eq. (2).

### A. Derivation of the shifted power iteration method

The power iteration method is an iterative eigenvector computation method which can be described as

$$v_{n+1} = \frac{\mathbf{H}v_n}{\|\mathbf{H}v_n\|}, \tag{13}$$

where $\mathbf{H}$ is a square matrix with eigenvalues $\lambda_1 < \lambda_2 <= \lambda_3 <= \ldots < \lambda_m$ and $|\lambda_1| < |\lambda_m|$. In this scheme, $v_n$ will converge to the eigenvector associated with eigenvalue $\lambda_m$ as $n \to \infty$.[13]

We already know that all eigenvalues of our Hessian matrix $\mathbf{H}$ are on the real axis. Moreover, if $\lambda$ is an eigenvalue of $\mathbf{H}$ associated with the eigenvector $v$, and $\mathbf{I}$ is the identity matrix, then for any constant $a \in \mathbb{R}$, $a - \lambda$ is the eigenvalue of a new matrix $a\mathbf{I} - \mathbf{H}$ with eigenvector $v$. As a result, we can linearly shift the desired eigenvalue to become the eigenvalue with greatest absolute value without changing the eigenvector. Therefore, the power iteration method will converge to our desired eigenvector if we find such a shift.

To pick an appropriate shift, $a$, we use the current maximum absolute value of Rayleigh quotients at each iteration plus a small increment,

$$a_n = \max\left\{\left|\frac{v_1^{\mathsf{T}}\mathbf{H}v_1}{v_1^{\mathsf{T}}v_1}\right|, \left|\frac{v_2^{\mathsf{T}}\mathbf{H}v_2}{v_2^{\mathsf{T}}v_2}\right|, \cdots \left|\frac{v_n^{\mathsf{T}}\mathbf{H}v_n}{v_n^{\mathsf{T}}v_n}\right|\right\} + \log n. \tag{14}$$

The $\log n$ term is added to prevent the case where $a_n - \lambda_1 = -(a_n - \lambda_m)$, i.e. $a_n = 0.5^*(\lambda_1 + \lambda_m)$, even though this scenario is unlikely. With the dynamical update of the shift constant $a_n$ according to Eq. (14), the modified power iteration method can be described,

$$v_{n+1} = \frac{(a_n\mathbf{I} - \mathbf{H})v_n}{\|(a_n\mathbf{I} - \mathbf{H})v_n\|}. \tag{15}$$

**Theorem 4.** *The iterative algorithm shown in Eq. (15) will converge to the eigenvector associated with the smallest eigenvalue of* **H**, *if this eigenvalue has multiplicity* 1.

*Proof.* Let $\lambda_1$ be the smallest eigenvalue and $\lambda_2$ the second smallest one, by our assumption smallest eigenvalue has multiplicity 1, $\lambda_1 < \lambda_2$.

The Rayleigh quotient is bounded by the maximal absolute value of eigenvalues, which we assume to be $L$. Let $v$ be the true eigenvector we want to obtain, then the convergence rate of the $v_n$ depends on the ratio of two eigenvalues which are with largest absolute values.[13] The convergence rate is,

$$\|v_n - v\| = \mathcal{O}\left(\prod_{k=1}^{n}\frac{L + \log k - \lambda_1}{L + \log k - \lambda_2}\right) \to 0, \tag{16}$$

as $n \to \infty$, which proves the convergence of the algorithm. □

### B. Krylov subspace of the shifted power iteration method

While the convergence of the power iteration method with a Rayleigh shift is guaranteed in principle, the convergence is slow in practice. The resulting $v_n$ at the $n$th iteration from this method is located in the Krylov subspace $\mathcal{K}_{n+1}$, which is defined in Remark 2. We will prove this statement in Lemma 1 in order to conclude that the shifted power iteration method will always converge slower than the Lanczos algorithm.

**Lemma 1.** *For any* $n > 0$, $v_n \in \mathcal{K}_{n+1}$, *where* $\mathcal{K}_{n+1} = \text{span}\{v_0, \mathbf{H}v_0, \cdots, \mathbf{H}^n v_0\}$.

*Proof.* We will prove this by induction.

When $n = 1$, $v_1 = (a_1\mathbf{I} - \mathbf{H})v_0 = a_1 v_0 + \mathbf{H}v_0 \in \mathcal{K}_2$.

Given $v_n \in \mathcal{K}_{n+1}$, we can write $v_n = \sum_{i=0}^{n} c_i \mathbf{H}^i v_0$. Then,

$$v_{n+1} = (a_n\mathbf{I} - \mathbf{H})v_n = (a_n\mathbf{I} - \mathbf{H})\sum_{i=0}^{n} c_i\mathbf{H}^i v_0 \in \mathcal{K}_{n+2}. \tag{17}$$

Thus, the statement holds for any general $n > 0$. □

Since the eigenvalue problem can be taken as a minimization problem as shown in Theorem 1, the power iteration solution is within the same Krylov subspace as the Lanczos algorithm, with the same number of iterations. Therefore, the convergence rate of the power iteration method is limited by the Lanczos algorithm.

## V. NUMERICAL TEST

In Secs. I– IV, we have compared the theoretical efficiencies of the Lanczos algorithm, the dimer method, and the shifted power iterative method. We proved that the convergence rates of the later two methods are bounded by the Lanczos algorithm due to the restriction of a smaller search space than Lanczos at each step. In this section, we conduct a numerical comparison to demonstrate our results in practice.

The convergence rates of the algorithms are compared for Lennard-Jones clusters with 38 atoms. Geometry configurations are chosen randomly near saddle points where the existence of one negative eigenvalue of the Hessian is guaranteed. A random direction is used as an initial guess for each of the minimum-mode searches. More details of the benchmark

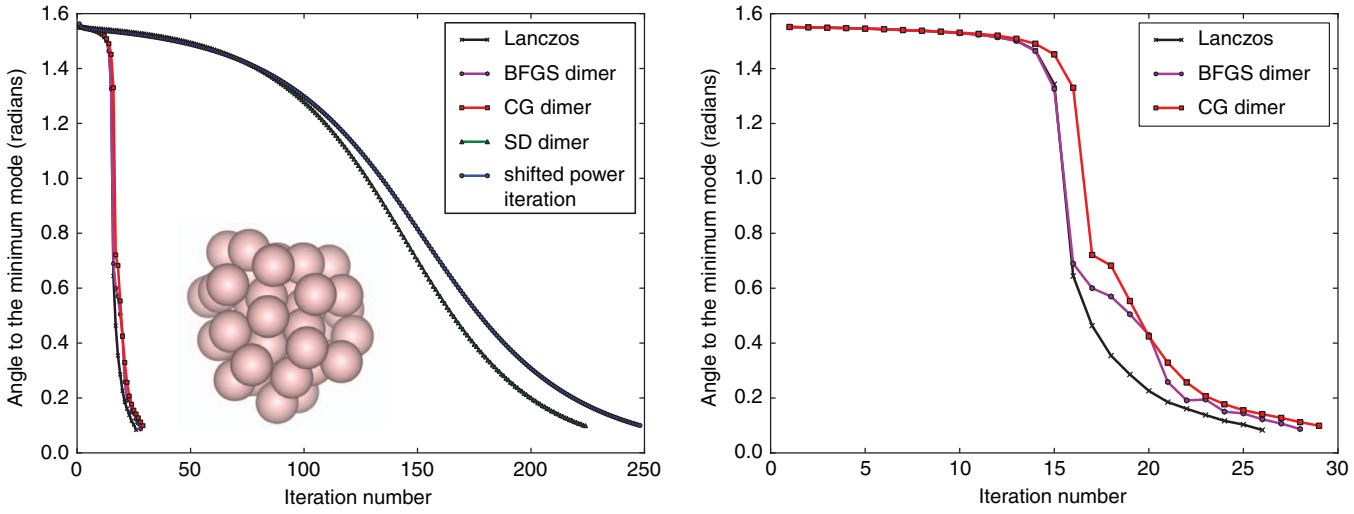


FIG. 1. The angle, in radians, towards the true minimum mode as a function of iteration number (left) and a zoom in of the region of relevance for the Lanczos, BFGS dimer, and CG dimer methods (right).

TABLE I. Steps to convergence.

| Method | $\bar{N}$ | Max N | Min N |
|---|---|---|---|
| Lanczos | 25 | 54 | 13 |
| BFGS dimer | 27 | 65 | 13 |
| CG dimer | 29 | 80 | 13 |

system are discussed elsewhere.[21,22] The Lanczos method is implemented with full reorthogonalization, which is faster by one step on average than without reorthogonalization. The dimer method is implemented with three optimizers for determining the rotation plane: SD, CG, and BFGS. The initial Hessian for BFGS is taken as $\alpha\mathbf{I}$, where $\alpha$ is set to be 60 eV/Å$^2$ and $\mathbf{I}$ is the identity matrix. Other $\alpha$ values tested did not give significantly better results. When the rotation direction from the BFGS becomes almost perpendicular (within 3°) of the SD direction, the BFGS is restarted with the initial Hessian. No parameters are needed for the other methods. All the methods are implemented in the TSASE software.[23,24]

The angle between the estimated lowest mode and the true minimum mode is plotted at each iteration in a typical run in Fig. 1. Clearly the Lanczos method is the fastest, while the shifted power iteration and the SD dimer are significantly slower. The CG dimer and the BFGS dimer are marginally slower than Lanczos. The similar convergence trends of these three methods indicate some commonality between them.

The two slowest methods were not considered for further study, but for the three competitive methods, 200 minimum mode searches were run at different cluster geometries for a more statistically significant comparison. A summary of the results is presented in Table I. The convergence criteria is that the angle to the true minimum mode is smaller than 0.14, which corresponds to an overlap (dot product of unit vectors) greater than 0.99. We did not observe any case in which the dimer method converges faster than Lanczos, although in some cases they converge at the same rate. Typically, the BFGS dimer is faster than the CG dimer when a reasonable initial Hessian value, $\alpha$, is chosen. These numerical results are consistent with our theoretical conclusions.

## VI. CONCLUSION

In summary, we have presented three classes of minimum mode searching algorithms, the Lanczos algorithm, dimer method, and the shifted power method, under the same mathematical framework of minimization in the Krylov subspace. With a theoretical understanding of these methods, we can see the dimer and shifted power methods are searching in a subspace of the Krylov subspace for which the Lanczos method explicitly finds the minimum curvature mode. This leads to the conclusion that with the same number of evaluations of the potential gradient, the Lanczos algorithm will theoretically converge no slower than the other two classes of methods. The result of this research can provide theoretical guidance for any future improvements to methods for finding the minimum curvature mode. Key to methods that can outperform the Lanczos algorithm will be the determination of subspaces that are outside of the Krylov subspace.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation. Y.Z. is supported by a Graduate Research Fellowship under Grant No. 1122374 and G.H. by Grant No. CHE-1152342. We gratefully acknowledge the Institute for Pure and Applied Mathematics at the University of California in Los Angeles for hosting our participation in the long program "Materials for a Sustainable Energy Future" from Sept. 9 to Dec. 13, 2013. We thank Keith Promislow for helpful discussions.

[1] C. Wert and C. Zener, Phys. Rev. **76**, 1169 (1949).
[2] G. H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).
[3] C. Lanczos, *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators* (United States Government Press Office, 1950).
[4] R. Malek and N. Mousseau, Phys. Rev. E **62**, 7723 (2000).
[5] G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[6] A. Heyden, A. T. Bell, and F. J. Keil, J. Chem. Phys. **123**, 224101 (2005).
[7] J. Kästner and P. Sherwood, J. Chem. Phys. **128**, 014106 (2008).
[8] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
[9] L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).
[10] W. E and X. Zhou, Nonlinearity **24**, 1831 (2011).
[11] J. Leng, W. Gao, C. Shang, and Z.-P. Liu, J. Chem. Phys. **138**, 094110 (2013).
[12] Y. Saad, *Numerical Methods for Large Eigenvalue Problems* (SIAM, 1992), Vol. 158.
[13] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra* (SIAM, 1997), Vol. 50.
[14] Y. Saad, SIAM J. Numer. Anal. **17**, 687 (1980).
[15] J. Kuczynski and H. Wozniakowski, SIAM J. Matrix Anal. Appl. **13**, 1094 (1992).
[16] A. Poddey and P. E. Blöchl, J. Chem. Phys. **128**, 044107 (2008).
[17] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson, J. Chem. Phys. **121**, 9776 (2004).
[18] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
[19] J. Nocedal, Math. Comput. **35**, 773 (1980).
[20] A. Samanta and W. E, J. Chem. Phys. **136**, 124104 (2012).
[21] S. T. Chill, J. Stevenson, V. Ruehle, C. Shang, P. Xiao, D. Wales, and G. Henkelman, "Benchmarks for characterization of minima, transition states and pathways in atomic systems," J. Chem. Phys. (unpublished).
[22] See http://optbench.org/ for the optimization benchmarks.
[23] See https://wiki.fysik.dtu.dk/ase/ for information about the ASE project.
[24] See http://theory.cm.utexas.edu/henkelman/code/ to obtain the TSASE code.