Geometry optimization speedup through a geodesic approach to internal coordinate optimization

Eric D. Hermes, Khachik Sargsyan, Habib N. Najm, and Judit Zádor*

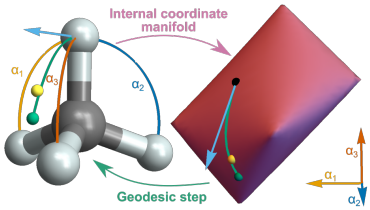
Combustion Research Facility, Sandia National Laboratories, Livermore, CA 94551-0969 USA

E-mail: jzador@sandia.gov

Abstract

We present a new geodesic-based method for geometry optimization in a basis of redundant internal coordinates. This method realizes displacements along internal coordinates by following the geodesic generated by the displacement vector on the internal coordinate manifold. Compared to the traditional Newton method approach to taking displacements in internal coordinates, this geodesic approach substantially reduces the number of steps required to reach convergence on a molecular structure minimization benchmark. This new geodesic method can in principle be implemented in any existing optimization code, and only requires the implementation of derivatives of the Wilson B-matrix and the ability to solve a relatively inexpensive ordinary differential equation.

Graphical TOC Entry



A methane molecule with three bending angles labeled and a 3D representation of the internal coordinate manifold comprising these angles. A displacement vector is used to generate a geodesic curve on this manifold, which is also superimposed on the methane molecule.

Geometry optimization is a crucial first step in the computational modeling of molecules, solids, and other atomic systems. The most obvious and direct way to optimize molecular geometries involves direct optimization of the Cartesian positions of each atom in the molecule. This approach can be very inefficient, as large amplitude molecular motions would require many rectilinear steps in a Cartesian coordinate space in order to preserve local molecular properties. An alternative approach commonly used for molecules is to take curvilinear steps along internal coordinates such as bond distances, bending angles, and dihedral angles, as this enables direct optimization of chemically relevant features.[1–3]

The Cartesian coordinate vector of an n-atom molecule $\mathbf{x} \in \mathbb{R}^{3n}$ encodes the geometry of a molecule as the Cartesian positions of each atom in that molecule. The internal coordinate vector $\mathbf{q} \in \mathbb{R}^m$ encodes the geometry of a molecule in a set of m local coordinates, typically consisting of bond distances, bending angles, and dihedral angles.[4] These internal coordinates cannot represent net translation or rotation of the molecule, so in general only $3n-6$ internal coordinates are required to fully specify the geometry of non-linear molecules. When m is greater than its minimum possible value of $3n-6$, it is said that the internal coordinate representation is redundant.[2,3] Even in a redundant internal coordinate basis, the set of all valid internal coordinate vectors $\mathbf{q}$ only spans a $(3n-6)$-dimensional space due to correlations between redundant internal coordinates. As described by Zhu *et. al.*,[5] the space of valid internal coordinates can be considered a $(3n-6)$-dimensional manifold embedded in a larger m-dimensional space. It is necessary to ensure that all geometry optimization steps in a redundant internal coordinate basis stay on the $(3n-6)$-dimensional manifold, i.e. the steps correspond to valid internal coordinates. This means both that displacement vectors $\mathbf{\Delta q} \in \mathbb{R}^m$ must be tangent to the internal coordinate manifold and that new structures obtained during optimization must be found in a way that accounts for curvature of the manifold.

One way to ensure that the displacement vector $\mathbf{\Delta q}$ lies tangent to the manifold is to temporarily switch to a minimal local coordinate system in which only valid displacement

vectors are possible. The delocalized internal coordinate approach defines a new structure $\mathbf{p} \in \mathbb{R}^{3n-6}$ as a linear transformation of the redundant internal coordinates $\mathbf{p} = \mathbf{U}^T\mathbf{q}$. The matrix $\mathbf{U}$ is the $(m \times (3n-6))$ matrix of left singular vectors of Jacobian matrix $\mathbf{B}$, also known as the Wilson B-matrix,[4,6]

$$\mathbf{B} = \frac{\partial \mathbf{q}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{U} & \mathbf{U}' \end{bmatrix} \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}^T \\ \mathbf{V}'^T \end{bmatrix}, \tag{1}$$

where $\mathbf{U}$ are the aforementioned left singular vectors, $\mathbf{S}$ is the diagonal matrix of non-zero singular values, $\mathbf{V}$ are the right singular vectors, and $\mathbf{U}'$ and $\mathbf{V}'$ are respectively the left and right singular vectors spanning the null space of $\mathbf{B}$. Projecting the coordinates, gradient, and exact or approximate Hessian from the full redundant internal coordinate basis into the delocalized internal coordinate basis enables the use of standard geometry optimization algorithms such as rational function optimization (RFO)[7] or quasi-Newton BFGS.[8–11] Regardless of which optimization algorithm is chosen, the result is a displacement vector $\Delta\mathbf{p}$ in the delocalized internal coordinate space which is tangent to the manifold by construction. This displacement vector can be projected back into the full redundant internal coordinate space through the relation $\Delta\mathbf{q} = \mathbf{U}\Delta\mathbf{p}$.

Even if $\Delta\mathbf{q}$ is a locally valid displacement vector at point $\mathbf{q}_0$, this does not guarantee that $\mathbf{q}_0 + \Delta\mathbf{q}$ is a valid point, as the internal coordinate manifold may be curved. This problem is traditionally solved by updating the geometry to be the point on the internal coordinate manifold which is closest to $\mathbf{q}_0 + \Delta\mathbf{q}$. This can be accomplished with Newton's root-finding method, in which a series of rectilinear displacements are taken in the Cartesian coordinate basis according to the equation

$$x^i_{(k+1)} = x^i_{(k)} + \left(\mathbf{B}^+_{(k)}\right)^i_\lambda \left(q^\lambda_0 + \Delta q^\lambda - q^\lambda_{(k)}\right), \qquad i = 1, \ldots, 3n, \tag{2}$$

where we have used the Einstein summation convention, $\mathbf{q}_{(k)}$ and $\mathbf{x}_{(k)}$ are respectively the

internal and Cartesian coordinates at iteration k, and $\mathbf{B}^{+}_{(k)}$ is the Moore-Penrose pseudo-inverse of the Jacobian matrix evaluated at $\mathbf{x}_{(k)}$.[1,2,12] In equation 2 and below, Latin indices correspond to quantities represented in Cartesian coordinates while Greek indices correspond to quantities represented in internal coordinates. The converged Cartesian coordinates $\mathbf{x}_{(k)}$ obtained from equation 2 are then used to calculate the new internal coordinates $\mathbf{q}_{\text{Newton}}$, which correspond to the point on the internal coordinate manifold closest to $\mathbf{q}_0+\Delta\mathbf{q}$. Though each iteration of equation 2 consists of a rectilinear displacement in Cartesian coordinates, the Newton method results in a curvilinear displacement, as the point $\mathbf{q}_{\text{Newton}}$ necessarily lies on the internal coordinate manifold.

This approach is computationally facile and generally converges in only a few iterations, with the greatest cost being the evaluation and inversion of the Jacobian matrix. However, when the manifold has a high degree of redundancy or coupling between coordinates, such as in systems with rings, equation 2 may fail to converge. In this scenario, one potential solution is to iterate equation 2 only a single time, which is equivalent to taking the rectilinear Cartesian displacement $\mathbf{x}_0+\mathbf{B}^{+}_{0}\Delta\mathbf{q}$.[12] This fallback approach can have substantial deleterious effects on optimization performance, as these rectilinear displacements tend to perturb bond distances when modifying bending angles or dihedral angles. Even when equation 2 does converge, it cannot fully account for the changing coupling between internal coordinates during the displacement because it does not explicitly consider the curvature of the manifold at any point.

As an alternative to the Newton approach, we suggest a new method for realizing a displacement vector $\Delta\mathbf{q}$ based on geodesics of the internal coordinate manifold. Geodesics are curves which trace the shortest path between two points on a manifold. In our application, the geodesic is determined from the starting geometry $\mathbf{q}_0$ and a vector which is tangent to the geodesic, which we choose to be $\Delta\mathbf{q}$. The orientation of $\Delta\mathbf{q}$ determines the trajectory of the geodesic $\mathbf{q}(\tau)$, where τ is the dimensionless geodesic parameter, while the magnitude $\|\Delta\mathbf{q}\|$ determines the distance along the trajectory to travel. The trajectory can be found

by solving the geodesic equation,

$$\ddot{q}^{\lambda} + \Gamma^{\lambda}_{\mu\nu}\dot{q}^{\mu}\dot{q}^{\nu} = 0, \qquad \lambda = 1, \ldots, m \tag{3}$$

where Newton's dot notation is used to refer to derivatives with respect to τ and $\Gamma^{\lambda}_{\mu\nu}$ are the Christoffel symbols of the second kind for the internal coordinates (see the supplementary material for more details).[13] Equation 3 is solved for the initial conditions $\mathbf{q}(0) = \mathbf{q}_0$ and $\dot{\mathbf{q}}(0) = \Delta\mathbf{q}$. We are interested in finding a point that is a distance $\|\Delta\mathbf{q}\|$ from $\mathbf{q}_0$, so we integrate equation 3 until $\tau = 1$ and choose our new geometry $\mathbf{q}_{\text{geodesic}}$ to be $\mathbf{q}(1)$, as this equation generates trajectories of constant speed.

Equation 3 cannot be solved directly, as the internal coordinates $\mathbf{q}$ are calculated from the Cartesian coordinates $\mathbf{x}$ and are therefore not independent variables. Instead, we solve the geodesic equation in the Cartesian coordinate basis,

$$\ddot{x}^{i} + \left(\mathbf{B}^{+}\right)^{i}_{\lambda} \frac{\partial^2 q^{\lambda}}{\partial x^{k} \partial x^{l}} \dot{x}^{k}\dot{x}^{l} = 0, \qquad i = 1, \ldots, 3n \tag{4}$$

where $\mathbf{x}(0) = \mathbf{x}_0$ are the Cartesian coordinates corresponding to $\mathbf{q}_0$ and $\dot{\mathbf{x}}(0) = \mathbf{B}^{+}\Delta\mathbf{q}$. The point $\mathbf{x}(1)$ obtained from this differential equation is used to calculate the new internal coordinates $\mathbf{q}(1)$. Though equation 4 depends on the second derivative of $\mathbf{q}$ with respect to $\mathbf{x}$, this quantity is not prohibitively onerous to implement for commonly-used internal coordinate types, and it has sparse structure that can be exploited to accelerate the summation over indices k and l. These second derivatives can be evaluated numerically from the Jacobian matrix,[4] analytically,[14,15] or through automatic differentiation.[16] Equation 4 can be solved using an off-the-shelf ODE solver such as LSODA[17] or CVODE[18] using a standard order reduction strategy.

Following a geometry step, it is typical for optimization algorithms to update an approx-

imate Hessian matrix $\mathbf{H}$ in order to satisfy the secant condition,

$$H_{\lambda\mu}\left(\mathbf{q}_1 - \mathbf{q}_0\right)^{\mu} = \left(\mathbf{g}_1 - \mathbf{g}_0\right)_{\lambda}, \qquad \lambda = 1, \ldots, m, \tag{5}$$

where $\mathbf{g}$ is the gradient vector in the internal coordinate basis. In order for the approximate curvature to lie in the tangent space of the manifold at the new point $\mathbf{q}_1$, this secant condition must be modified to

$$H_{\lambda\mu}\left(\dot{\mathbf{q}}(1)\right)^{\mu} = \left(\mathbf{g}_1 - \tilde{\mathbf{g}}_0\right)_{\lambda}, \qquad \lambda = 1, \ldots, m, \tag{6}$$

where $\dot{\mathbf{q}}(1)$ is obtained from the solution to equation 3 and $\tilde{\mathbf{g}}_0$ is the gradient vector at point $\mathbf{q}_0$ which has been parallel transported along the geodesic to the point $\mathbf{q}_1$.[19] Parallel transport is the process of translating vectors that are tangent to a manifold along a curvilinear trajectory on that manifold (such as a geodesic) in such a way that the vectors remain both tangent to the manifold along the entire trajectory and self-parallel along infinitesimal displacements. For more details on how $\tilde{\mathbf{g}}_0$ is determined, see the supplementary material. In the Hessian update scheme of our geodesic approach, the raw displacement $\mathbf{q}_1 - \mathbf{q}_0$ is replaced by $\dot{\mathbf{q}}(1)$, and the initial gradient vector $\mathbf{g}_0$ is replaced by its parallel transported equivalent $\tilde{\mathbf{g}}_0$.

An illustration comparing the geodesic and Newton stepping methods is presented in figure 1. In this figure, the purple surface represents the manifold of valid internal coordinates in a methane molecule with all internal coordinates fixed except for three bending angles, as depicted in figure 1c. Though this system has three free bending angle coordinates, only two degrees of freedom remain due to the coupling between the angular coordinates. Figures 1a and 1b depict the entire manifold in a basis of the three free bending angles from two different persepctives. In this basis, the internal coordinate manifold takes the form of an octahedron with smoothed edges. The geodesic approach follows the curvature of the manifold to find the new point $\mathbf{q}_{\text{geodesic}}$. In contrast, the Newton method converges

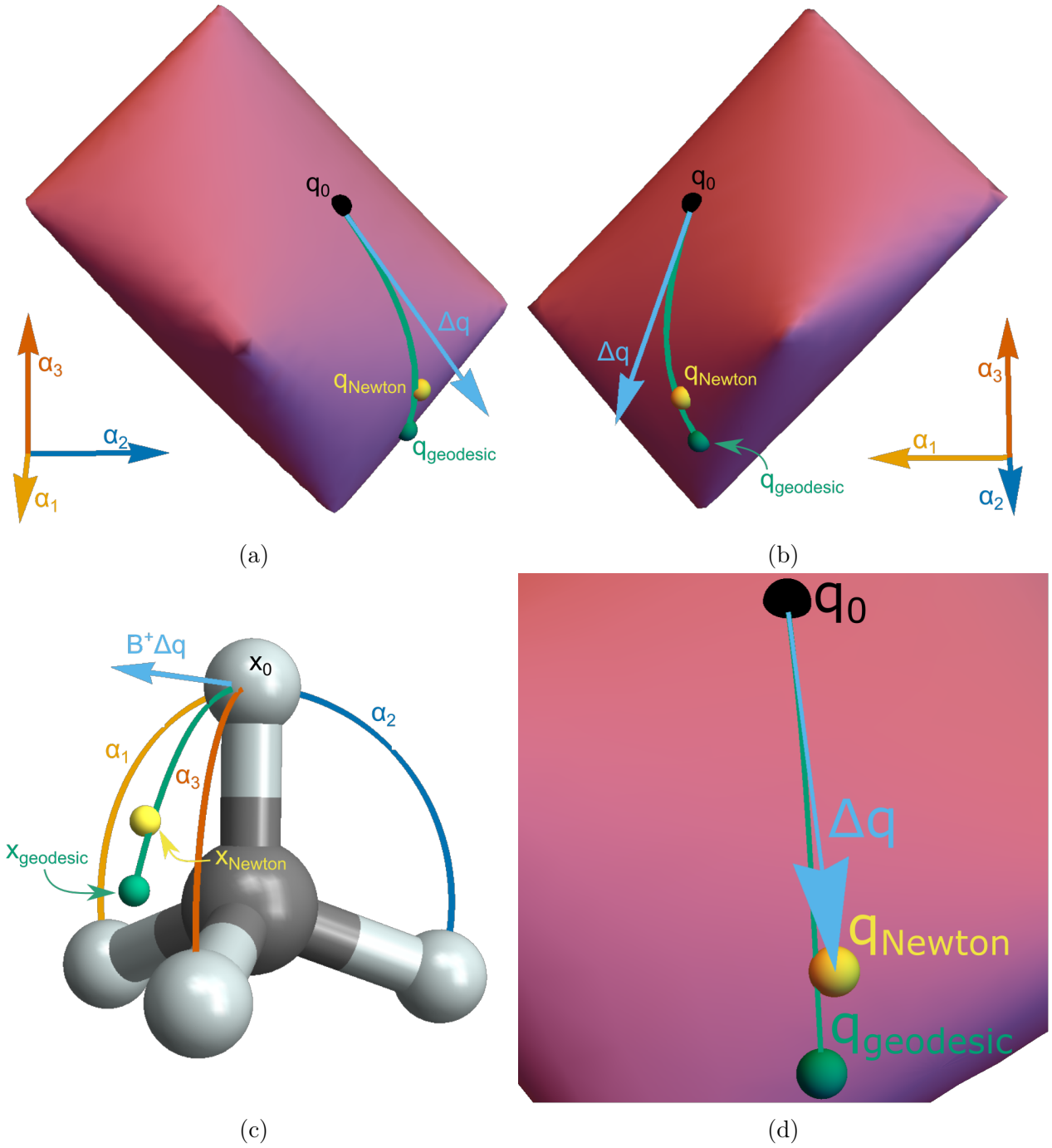


Figure 1: (a), (b) The internal coordinate manifold of a methane molecule with all internal coordinates fixed except three bending angles, from two different perspectives. Labeled are the initial structure $\mathbf{q}_0$ (black), the displacement vector $\Delta\mathbf{q}$ (light blue), the final structure of the Newton method $\mathbf{q}_{\text{Newton}}$ (yellow), and the final structure of the geodesic method $\mathbf{q}_{\text{geodesic}}$ (green). (c) A real-space representation of the same methane molecule with the three free bending angles labeled α_1 (orange), α_2 (dark blue), and α_3 (red). Additionally, the Cartesian equivalents of the initial structure, displacement vector, final Newton structure, and final geodesic structure are also labeled. (d) A zoomed-in perspective of the manifold in the region around the displacement, which shows more clearly that the point $\mathbf{q}_{\text{Newton}}$ does not lie on the geodesic curve.

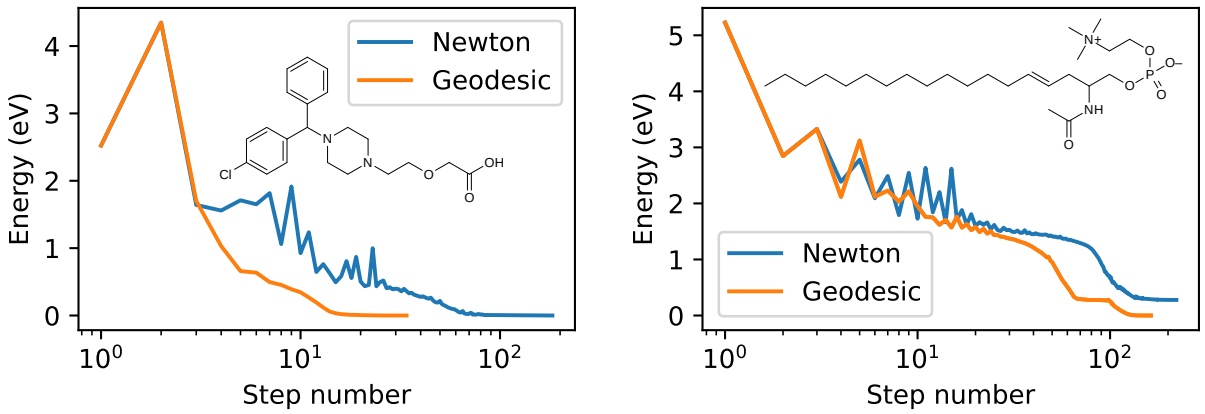
to the point $\mathbf{q}_{\text{Newton}}$ on the manifold which is closest to $\mathbf{q}_0 + \Delta\mathbf{q}$. Figure 1d shows the same manifold, but rotated and zoomed to better illustrate the difference between the Newton and geodesic stepping methods. From figure 1d, it is clear that $\mathbf{q}_{\text{Newton}}$ does not lie on the geodesic curve. This is to be expected, as the Newton stepping method is not aware of the curvature of the manifold, unlike the geodesic method which follows the curvature of the manifold by construction. Though it is clear from this figure that the Newton and geodesic methods result in different structures, it is not obvious which of the two stepping methods is better for geometry optimization.

In order to determine the difference in performance between the Newton and geodesic methods, we use a geometry optimization benchmark originally developed by Birkholz and Schlegel consisting of 20 molecules that have between 20 and 95 atoms.[12] Potential energies were evaluated using dftb+ with the DFTB3 parameterization.[20–24] Structure optimization was performed by Sella, an open source Python package primarily focused on saddle point optimization which is also capable of performing geometry minimization.[25,26] We note that because Sella is primarily intended to be used for saddle point optimization, the performance of its RFO minimization algorithm is likely lower than that of purpose-built minimization codes. The focus is therefore only on the relative performance of the Newton and geodesic stepping approaches, with all other aspects of the minimization algorithm held fixed. Of the original 20 molecules in the benchmark, one molecule was excluded due to a missing initial structure from the reference and another was excluded as DFTB3 lacks parameters for Aluminum. Scripts to reproduce these calculations can be found in the supplementary material.

The results in table 1 indicate that the geodesic approach requires fewer steps to reach convergence in all tested systems. For the molecules raffinose and sphingomyelin, the geodesic approach converges to a lower energy structure than the Newton approach while also requiring fewer steps to converge. The optimization trajectories of two of the molecules, cetirizine and sphingomyelin, are illustrated in figure 2. The largest component of the gradient for the

Table 1: Number of gradient evaluations required to converge for the standard and geodesic stepping methods. An asterisk indicates convergence to a higher-energy structure.

Species	Newton	Geodesic
Artemisinin	122	33
Avobenzone	292	90
Azadirachtin	255	243
Bisphenol A	270	89
Cetirizine	183	34
Codeine	389	108
Diisobutyl phthalate	175	59
Estradiol	151	47
Inosine	238	93
Maltose	208	104
Mg Porphyrin	86	15
Ochratoxin A	235	47
Penicillin V	168	55
Raffinose	325*	169
Sphingomyelin	221*	163
Tamoxifen	205	58
Vitamin C	160	60
Zn EDTA	136	50



(a) Cetirizine

(b) Sphingomyelin

Figure 2: Optimization trajectories for the Cetirizine (a) and Sphingomyelin (b) test systems using the Newton (blue) and geodesic (orange) methods. A log scale is used for the step number axis to better highlight early optimization steps.

structures in this test set tends to lie in bond-stretching coordinates, and so early stages of geometry optimization are dominated by bond stretch displacements. These bond stretch displacements tend to be rectilinear or nearly rectilinear, meaning the manifold has very low curvature in these directions, and so the two methods tend to take very similar steps near the beginning of optimization. After the bond stretching modes are largely relaxed, the larger amplitude angle bending and dihedral angle modes begin to dominate the optimization, and it is at this point that the Newton and geodesic methods begin to exhibit different performance characteristics. In this regime, the Newton method more frequently takes steps that result in an increase in the potential energy as evidenced by the many spikes in the optimization trajectories in figure 2. In contrast, the geodesic method is less likely to take steps that increase the potential energy and generally reaches convergence in fewer steps overall compared to the Newton method.

The improved performance of the geodesic approach as compared to the Newton approach is a consequence of the consideration of coupling between internal coordinates along the entire displacement trajectory. In Sella's primary application of saddle point optimization, preliminary results suggest a substantial increase in performance compared to other leading algorithms, which we intend to show in a future publication. We expect that incorporating the geodesic stepping approach into leading minimization codes should result in a noticeable performance boost. In particular, minimization algorithms that use a line-search will be able to interpolate the solution of the geodesic equation in order to find arbitrary intermediate points along the geodesic.

Acknowledgement

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences and Biosciences Division, as part of the Computational Chemistry Sciences Program (Award Number: 0000232253).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

We would like to thank Dr. Laura McCaslin for useful discussions leading to improved readability and comprehensibility of this work.

Supporting Information Available

An extended derivation of equation 4 as well as scripts to reproduce the data presented in table 1 and figure 2 can be found in the supporting information.

References

(1) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. *Journal of the American Chemical Society* **1979**, *101*, 2550–2560.

(2) Pulay, P.; Fogarasi, G. Geometry optimization in redundant internal coordinates. *The Journal of Chemical Physics* **1992**, *96*, 2856–2860.

(3) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* **1996**, *17*, 49–56.

(4) Wilson, E.; Decius, J.; Cross, P. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Books on Chemistry; Dover Publications, 2012.

(5) Zhu, X.; Thompson, K. C.; Martínez, T. J. Geodesic interpolation for reaction pathways. *The Journal of Chemical Physics* **2019**, *150*, 164103.

(6) Baker, J.; Kessi, A.; Delley, B. The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of Chemical Physics* **1996**, *105*, 192–212.

(7) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. Search for stationary points on surfaces. *The Journal of Physical Chemistry* **1985**, *89*, 52–57.

(8) Broyden, C. G. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. *IMA J. Appl. Math.* **1970**, *6*, 76–90.

(9) Fletcher, R. A new approach to variable metric algorithms. *Comput. J.* **1970**, *13*, 317–322.

(10) Goldfarb, D. A family of variable-metric methods derived by variational means. *Math. Comput.* **1970**, *24*, 23–26.

(11) Shanno, D. F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **1970**, *24*, 647–656.

(12) Birkholz, A. B.; Schlegel, H. B. Exploration of some refinements to geometry optimization methods. *Theoretical Chemistry Accounts* **2016**, *135*, 84.

(13) Spivak, M. *A Comprehensive Introduction to Differential Geometry*; A Comprehensive Introduction to Differential Geometry v. 3; Brandeis University, 1970.

(14) Hollman, D. S.; Schaefer, H. F. Arbitrary order El'yashevich–Wilson B tensor formulas for the most frequently used internal coordinates in molecular vibrational analyses. *The Journal of Chemical Physics* **2012**, *137*, 164103.

(15) McCaslin, L. M. From Basis Sets to Force Fields: Improving Methods for High-Accuracy Quantum Chemical Calculations of Small Molecules. Ph.D. thesis, The University of Texas at Austin, 2016.

(16) Bradbury, J.; Frostig, R.; Hawkins, P.; Johnson, M. J.; Leary, C.; Maclaurin, D.; Necula, G.; Paszke, A.; VanderPlas, J.; Wanderman-Milne, S. et al. JAX: composable transformations of Python+NumPy programs. 2018; `http://github.com/google/jax`.

(17) Petzold, L. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal on Scientific and Statistical Computing* **1983**, *4*, 136–148.

(18) Cohen, S. D.; Hindmarsh, A. C.; Dubois, P. F. CVODE, A Stiff/Nonstiff ODE Solver in C. *Computers in Physics* **1996**, *10*, 138–143.

(19) Gabay, D. Minimizing a differentiable function over a differential manifold. *Journal of Optimization Theory and Applications* **1982**, *37*, 177–219.

(20) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics* **2020**, *152*, 124101.

(21) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *Journal of Chemical Theory and Computation* **2013**, *9*, 338–354.

(22) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *Journal of Chemical Theory and Computation* **2014**, *10*, 1518–1537.

(23) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *The Journal of Physical Chemistry B* **2015**, *119*, 1062–1082.

(24) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the

DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *Journal of Chemical Theory and Computation* **2015**, *11*, 332–342.

(25) Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Accelerated Saddle Point Refinement through Full Exploitation of Partial Hessian Diagonalization. *Journal of Chemical Theory and Computation* **2019**, *15*, 6536–6549.

(26) Hermes, E. D. Sella. 2021; `https://doi.org/10.5281/zenodo.4747052`.