New Transition State Optimization and Reaction Path Finding Algorithm with Reduced Internal Coordinates

New Transition State Optimization and Reaction Path Finding Algorithm with Reduced Internal Coordinates

By Xiaotian Yang,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

Doctor of Philosophy (2021)

Hamilton, Ontario (Department of Chemistry & Chemical Biology)

TITLE: New Transition State Optimization and Reaction Path Finding Algorithm with Reduced Internal Coordinates

AUTHOR: Xiaotian Yang (McMaster University)

SUPERVISOR: Dr. Paul Ayers

NUMBER OF PAGES: xvi, 167

Sorbonne Université

Doctor of Philosophy (2021)

Paris, France (Laboratoire de Chimie Théorique)

TITRE: New Transition State Optimization and Reaction Path Finding Algorithm with Reduced Internal Coordinates

AUTEUR: Xiaotian Yang (Sorbonne Université)

SUPERVISEUSE: Dr. Julia Contreras-García

NOMBRE DE PAGES: xvi, 167

Abstract

Geometry optimization is a fundamental step in the numerical modelling of chemical reactions. Many thermodynamic and kinetic properties are closely related to the structure of the reactant, product, and the transition states connecting them. Different from the reaction and product, which are local minima on the potential energy surface, a transition state is the first-order saddle point with only one negative curvature. Over years, many methods have been devised to tackle the problem. Locating stable structures is relatively easy with a reliable algorithm and high accuracy. One can follow the gradient descent direction to pursuit the local minimum until convergence is reached. But for the transition state, the determination is more challenging as either the up-hill or down-hill direction is allowed in the process.

Motivated by the difficulty, many well-designed optimization algorithms are elaborated specifically to stress the problem. The performance of geometry optimization is affected by various aspects: the initial guess structure, the coordinate system representing the molecule, the accuracy of the initial Hessian matrix, the Hessian update schemes, and the step-size control of each iteration. In this thesis, we propose a new geometry optimization algorithm considering all the important components. More specifically, in Chapter 2, a new set of robust dihedral and redundant internal coordinates is introduced to effectively represent the molecular structures, as well as a computational efficient transformation method to generate a guess structure. In Chapter 3 and 5, a sophisticated robust algorithm is presented and tested to solve intricate transition state optimization problems. In Chapter 4, a new algorithm to exploring reaction pathways based on redundant

internal coordinates is illustrated with real chemical reactions. Last but not least, in Chapter 6, a systematic test to explore the optimal methods in each procedure is presented. A well-performed combination of optimization methods is drawn for generic optimization purposes.

All the methods and algorithms introduced in this thesis is included in our forthcoming open-source Python package named *GOpt*. It's a general-purpose library that can work in conjunction with major quantum chemistry software including *Gaussian*. More features are under development and await to be released in the coming update.

Acknowledgements

Ph.D. is a long journey, a bittersweet experience that I would cherish for the rest of my life. Time passes but the memory lasts. We say "goodbye" to the past, farewell our old friends to their new adventures; we say "greetings" to tomorrow, opening our arms to our new friends by our sides. Time flies. It feels like yesterday, I just landed at Pearson airport, passionate about my brand-new life in a foreign country. And now, I am sitting in front of my laptop, recalling every moment we share together. I have been expecting my graduation for a long time, but when it actually comes, I wish I could have spent more time with you all.

First and foremost, I'd like to thank Dr. Ayers and Dr. Contreras-García. They are great supervisors, not only in science but in life as well. I have tried my best to become as enthusiastic, optimistic, and considerate as them. None of my research could be completed without their help. I sincerely appreciate all the guidance and insights. The pandemic leads us to a down, isolated situation but they helped raise me out of the depression and stress.

I am also very grateful to all my committee members, Dr.Dumont, Dr.Daze, Dr.Savin, Dr.Giner, Dr.Heidar-Zadeh, Dr.Vogt-Geisse. Scheduling a defence between different timezone and hemispheres is not an easy task. Thank you for taking your valuable time reading my thesis, attending my defence, and guiding my research and project. It's a great honor for me to have you all on my committee.

No words are enough to describe my thanks to my dear friends David, Xiao, and Matt. As an international student living here alone, they offered me the most support and company. Winter in Canada is cold and chill. But they drove away

the long night with warmth and light. I am really lucky and grateful to have them in my life.

I'd also like to deliver my gratitude to all the past and current members of Ayers Group and Contreras Group. Every one of them is very friendly and open-minded. The diversified and inclusive atmosphere makes the research a delightful and pleasant experience.

The last but not least, I'd like to thank my parents. They have been supportive all the time and encourage me to pursue my dream. I barely express "love" to them before but here, "Dad and Mom, I love you."

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Xiaotian YANG, declare that this thesis titled, "New Transition State Optimization and Reaction Path Finding Algorithm with Reduced Internal Coordinates" and the work presented in it are my own. I confirm that:

The thesis I am submitting is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those regulations concerning disciplinary actions that may result from plagiarism. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Chapter 1

Introduction

1.1 Introduction

At its most fundamental level, chemistry is the study of how chemical bonds cleave and form to create new substances, along with the properties of these substances. The detailed sequence of steps by which a new substance is created is called the reaction mechanism. Key structures on the reaction path include the starting structure (the reactant), the final structure (the product), stable structures along the way (reactive intermediates), and first-order saddle points (transition states between stable structures along the path). The lowest-energy pathway that connects the reactant to the product is called the intrinsic reaction coordinate or minimum-energy reaction path [1]. The reaction path reveals, in atomistic detail, how the reactant transforms into the product. Some reaction paths are relatively simple for chemists to guess, or relatively easy to determine computationally. But this is not always the case: there are many reactions where it is difficult, both conceptually and computationally, to find key transition states, much less to fully characterize the reaction pathway. The goal of this thesis is to develop new computational methods to find transition states and location chemical reaction pathways

that work even for the most difficult reactions. This is especially important for reactions that are inaccessible (e.g., astrochemistry)[2, 3], dangerous (e.g., decomposition of high-energy materials)[4, 5], or unhealthy (e.g., metabolism of toxic substances) to experimentalists.[6, 7] Our specific goal is to leverage recent advances in computer hardware and software, new innovations in quantum chemistry, and new algorithms we shall develop to extend the range of chemical reactions for which detailed mechanistic computational studies can be performed.

The characteristics of a chemical reaction are largely determined by the molecular structures associated with the reactant, the product, the transition state, and the path connecting them. Therefore, locating the stationary points on the molecular potential surface is the first step towards successful numerical modeling. Mathematically, reactants, products, and reactive intermediates are local minima on the potential energy surface. Two local minima are connected by a stationary point which is a maximum along the reaction path but a minimum in all other directions. This saddle point is called the transition state (TS) between the two local minima.[8] Once all the important stationary points on the potential surface have been located, one can model the whole reaction process, including the mechanism(s) of the reaction and its kinetic and thermodynamic properties (reaction rate, equilibrium constant, exothermicity, etc.).[9] For multistep reactions, the existence of intermediate(s) complicates the reaction mechanism. In addition, there may be multiple possible reaction paths, wherein different intermediate structures connect the same reactants and products. In these complicated scenarios, having a complete minimum-energy path showing how reactants and products are connected by various sequences of structures is especially useful, as it provides researchers with atomistic detail about the reaction mechanism. This can be useful,

for example, for designing better catalysts.[10]

In computational studies of reaction mechanisms, three sorts of structure optimizations occur: minimization (for reactants, products, and reactive intermediates), saddle-point optimization (for transition states), and pathfinding (for the reaction coordinate). Each optimization is typically treated as a separate problem, and over the years researchers have developed many methods for each task. The effectiveness and efficiency of these algorithms are affected by many factors, among them the choice of the coordinate system, the initial guess structure(s), the initial Hessian, the Hessian update method, stepsize control methods, etc..[11–13]

Finding a local minimum on the potential surface is considered an easy task. One may simply follow the gradient descent direction until a minimum is reached, since a structure with lower energy is always preferred. For a transition state, the structure needs to be the maximum in only one dimension and a minimum in all others, so it is impossible to know whether a step should increase or decrease the energy without further (nonlocal) information about the structure. Researchers have designed multiple optimization algorithms to address the difficulty of transition-state optimization.[11, 14–17] Disappointed by the speed and robustness of current approaches, we developed a new set of algorithms to (a) effectively generating initial guess structures for transition-states, (b) optimize transition-states using chemical information about key internal coordinates, and (c) find reaction pathways in a more robust way. All these features and algorithms are included in our forth coming quantum chemistry software `GOpt`.

1.2 The Potential Energy Surface

Within the Born-Oppenheimer approximation, the electronic energy of a molecule is determined by its geometric structure, which is defined by the relative positions of its constituent atoms. To obtain the total energy of a system, one needs to solve the Schördinger equation,

$$\mathcal{H}\left|\Psi\right\rangle = \mathcal{E}\left|\Psi\right\rangle \tag{1.1}$$

where $\mathcal{H}$ is the quantum-mechanical operator for the energy, the Hamiltonian. The explicit expression of $\mathcal{H}$ is[18]

$$\begin{aligned}\mathcal{H} = &-\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \\ &+\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}\end{aligned} \tag{1.2}$$

In the Eqn1.2, M_A is the ratio of mass between nucleus A and one electron. Z_A is the nuclear charge of atom A. The first and second terms in $\mathcal{H}$ are the kinetic energy operators for the electrons and the nuclei respectively. The third term is the potential of the electron-nuclei attraction and the last two terms represent the Coulomb repulsion between electrons and between nuclei.

1.2.1 The Born-Oppenheimer Approximation

Because atomic nuclei are much more massive than electrons, it is sensible to assume the electrons adapt instantaneously to the relatively slow motions of the nuclei. That is, from the viewpoint of the electrons, the nuclei are clamped in well-defined positions. This is the basis for the Born-Oppenheimer approximation, and it leads to the concept of a molecular potential energy surface. As the nuclear

positions are assumed fixed, the nuclear kinetic energy (the second term in $\mathcal{H}$) is zero and the nuclear-nuclear repulsion (the last term in $\mathcal{H}$) is constant. The remaining terms define the electronic Hamiltonian, $\mathcal{H}_{elec}$,

$$\mathcal{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{1.3}$$

The electronic energy and wavefunction are determined by solving the electronic Schördinger equation

$$\mathcal{H}_{elec}\Psi_{elec} = \mathcal{E}_{elec}\Psi_{elec} \tag{1.4}$$

Note that the electronic wavefunction and energy change depending on the nuclear positions:

$$\Psi_{elec} = \Psi_{elec}(\{r_i\};\{R_A\}) \tag{1.5}$$

$$\mathcal{E}_{elec} = \mathcal{E}_{elec}(\{R_A\}) \tag{1.6}$$

It is common to add the nuclear interaction term to the electronic energy to obtain the potential energy surface on which the nuclei move, $U(\{R_A\})$,

$$U(\{R_A\}) = \mathcal{E}_{elec}(\{R_A\}) + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{1.7}$$

If the nuclei are assumed to be classical, then they are treated as classical point-particles moving on the potential energy surface. If the nuclei are assumed to be quantum, then the potential energy surface defines the potential in the nuclear Schrödinger equation. Because $\mathcal{E}_{tot}$ depends on the nuclear positions, one needs to repeatedly solve the electronic Schördinger equation. This task is normally

handled by quantum chemistry software.

The energy of a molecule with M atoms is a function of $3M-6$ free variables. When the positions of each nucleus are specified by their Cartesian coordinates, $\{X_A, Y_A, Z_A\}$, the molecular structure is defined by $3M$ coordinates. The true potential energy surface is only $3M-6$-dimensional, however, because of translation invariance (typically specified by the location of the center-of-mass, $(X_{com}, Y_{com}, Z_{com})$) and rotational invariance (typically specified by three Euler angles, $\{\alpha, \beta, \gamma\}$). This leaves a total of $3M-6$ degrees of freedom ($3M-5$ for a linear molecule).

Characterization of the Potential Energy surface

The potential energy surface is a function that indicates the relative stability of different arrangements of the atomic nuclei. Just like a geographical landscape, a potential energy surface has peaks, valleys, and pathways that connect them.

Valleys on the potential energy surface represent stable structures like reactants, products, and reactive intermediates. These structures are usually associated with the local minima at the bottom of the associate valley on the potential energy surface. Since any change of nuclear coordinates away from a local minima increases the energy, these are stable structures. Local minima are connected by paths on the potential energy surface. The most interesting paths are minimum-energy pathways (MEP), which specify the lowest-energy way to transform one structure to another; these *reaction paths* are parameterized by reaction coordinates. The highest-energy point on a reaction path is the transition state of that reaction. Mathematically, a transition state is the 1^{st} order saddle point on the potential energy surface. That is, a transition-state structure is the maximum in

one direction (tangent to the reaction coordinate) and the minimum in all other directions. Given the energies and energy-derivatives of the reactant, product, and transition state, one can easily estimate the thermodynamic and kinetic properties of the reaction using the (free) energy differences between structures.

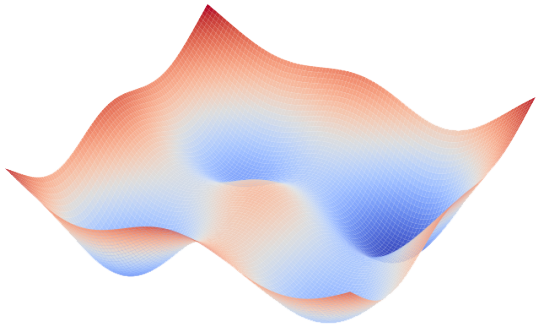


FIGURE 1.1: A four-well two-dimensional potential energy surface

In some reactions, there are pathways linking the same reactant and product structures.[19, 20] In these cases, paths with similar energy represent competing reaction mechanisms. The relative importance of mechanisms can be ascertained from the energy profile of the pathways. This is especially important for studies of chemical synthesis. Catalysts can be designed by preferentially lowering the barrier(s) of any of the feasible reaction pathways.

Mathematical Characterization of the Potential Energy surface

The potential energy surface, $U(\{R_A\})$, is a function which, given a specification of the molecular geometry, returns a real number. This real number will usually be substantially below zero, since it takes energy to dissociate a molecule into atoms, and the energy of a molecule where all the atoms are infinitely far apart

is the sum of the atomic energies, which are themselves negative (with magnitude equal to the energy required to remove all the electrons from the molecule). The potential energy surface is positive, then, only when two or more atomic nuclei are extremely close together.

Key chemical structures correspond to stationary points on the potential energy surface, that is, places where the gradient of the potential energy is zero: $\nabla U(\{R_A\}) = 0$. Generalizing to arbitrary choices for the coordinate system used to specify the molecular geometry, we introduce the vector-notation, $\mathbf{g}(\mathbf{x})$, as shorthand for the gradient of the potential:

$$\mathbf{g}(\mathbf{x}) = \nabla U(\mathbf{x}) = \begin{bmatrix} \frac{\partial U(\mathbf{x})}{\partial x_1} \\ \frac{\partial U(\mathbf{x})}{\partial x_2} \\ \vdots \end{bmatrix} \tag{1.8}$$

At a given structure $\mathbf{x}$ with potential $U(\mathbf{x})$, the gradient $\mathbf{g}(\mathbf{x})$ is the negative of the force exerted on the nuclei,

$$\mathbf{F}(\mathbf{x}) = -\mathbf{g}(\mathbf{x}) \tag{1.9}$$

In order to distinguish between stable molecular structures (minima) and transition states (first-order saddle points) on the potential energy surface, one uses the second derivative matrix, or Hessian, of the potential energy function, $\nabla\nabla^T U(\{R_A\})$. The notation $\nabla\nabla^T$ denotes the outer product of the gradient operators. Again,

we introduce a matrix-notation for the Hessian, $\mathbf{H}(\mathbf{x})$,

$$\mathbf{H}(\mathbf{x}) = \nabla\nabla^T U(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 U}{\partial x_1^2} & \frac{\partial^2 U}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 U}{\partial x_2 \partial x_1} & \frac{\partial^2 U}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{1.10}$$

The Hessian matrix is symmetric and describes the curvature of the potential energy surface for the specified molecular structure. The eigenvalues' signs specify whether a structure is in a valley (all eigenvalues are positive), near a first-order saddle point (one and only one negative eigenvalue), or at a higher-order saddle point (which is usually chemically irrelevant, as such points do not lie along minimum-energy pathways between stable structures). First-order saddle points are transition-states between stable molecular structures: if one starts in the direction of the eigenvector associated with the negative eigenvalue (which defines the negative-curvature direction) and then follows the steepest descent gradient pathway, one locates the reactant and product structures associated with the initializing transition state. The steepest-descent path one follows is a minimum energy pathway, and is often called the intrinsic reaction coordinate.

1.2.2 Numerical Calculations on Potential Energy Surfaces

The potential energy surface is a function of coordinates specifying the molecular geometry, $\mathbf{x}$. With a known initial structure x_0, one can estimate the potential of

nearby points, x, by Taylor expansion,[21]

$$\begin{aligned} U(\mathbf{x}) = U(\mathbf{x}_0) &+ \nabla U(\mathbf{x_0}) \cdot (\mathbf{x} - \mathbf{x}_0) \\ &+ \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \cdot \nabla\nabla^T U(\mathbf{x_0}) \cdot (\mathbf{x} - \mathbf{x}_0) \\ &+ ... \end{aligned} \tag{1.11}$$

Due to the computational expense associated with computing and using higher-order derivatives, Eqn 1.11 is normally truncated after the second-order derivative:

$$E(\mathbf{x}) \approx E(\mathbf{x}_0) + \mathbf{g}_0^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \mathbf{H}_0 \Delta\mathbf{x} \tag{1.12}$$

where $\Delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_0)$ and $\mathbf{g}_0$ and $\mathbf{H}_0$ are the gradient and the Hessian of the potential energy surface, evaluated at $\mathbf{x}_0$, respectively.

Traditional geometry optimization methods require the analytic calculation of energy and the gradient at each iteration; for most quantum chemistry methods the gradient can be computed relatively cheaply after the electronic wavefunction and energy have been determined.[22] For example, in the steepest-descent algorithm for determining local minima structures, one repeatedly takes small steps in the gradient-descent direction until one reaches a local minimum. The steepest-descent method does not work for transition states, because one needs to know the Hessian eigenvalues to determine in which direction the energy will be minimized, and in which directions it will be maximized. If the analytic Hessian is available, Newton's method is an effective strategy for optimizing both minima and transition-states. Unfortunately, the analytic computation of the Hessian is significantly more expensive than the analytic computation of the energy and gradient, so approximate Hessians are often used. The accuracy of approximate Hessians is

strongly affected by the coordinate system one uses to specify the molecular geometry; it is favorable to choose a coordinate system in which the coupling between coordinates (as indicated, for example, by off-diagonal elements in the Hessian) is relatively small.

1.3 Coordinate System

While any coordinate system which uniquely specifies the positions of the atoms in the system will suffice for geometry optimization, in practice, certain choices give a better computational performance.[23–25]

1.3.1 Cartesian Coordinates

Conceptually, the simplest coordinate system is to use the Cartesian coordinates of the atoms, $\{(X_\alpha, Y_\alpha, Z_\alpha)\}$. For a molecule with M atoms, there are $3M$ Cartesian coordinates. In popular quantum chemistry packages, Cartesian coordinates are used to compute the energy and its derivatives, so using Cartesian coordinates is the most straightforward choice. However, direct use of the Cartesian coordinates has several drawbacks. Most importantly, the relative positions of atoms in Cartesian coordinates are highly coupled. A simple change in a single Cartesian coordinate for one atom changes the bond distances and bond angles between that atom and all of its neighbors. Conversely, a simple change in the interatomic distance between two atoms tends to change the Cartesian coordinates not only of the two atoms involved in the bond, but also of all the other atoms connected to those atoms. The highly-coupled nature of molecular motions in Cartesian coordinates is reflected in the Hessian matrix, which has large off-diagonal elements. The large

number of nonzero elements in the Hessian matrix and the relatively large changes in the coordinates, gradient, and Hessian that occur after simple chemical changes make the Hessian difficult to approximate.

1.3.2 Internal Coordinates

Building on chemical intuition, one can specify the molecule's geometry with internal coordinates (bond lengths, bond angles, and dihedral angles). These coordinates are more descriptive and intuitive at characterizing molecular structures, and because they depend only on relative atomic positions, they automatically impart rotational and translation invariance. Internal coordinates are less coupled, so there are fewer off-diagonal elements in the Hessian, making it easier to approximate.

For a nonlinear molecule with M atoms, only $3M - 6$ independent internal coordinates are needed to fully define the structure. However, the number of internal coordinates one can specify is far higher. For example, for a molecule with three atoms, one can specify three bond angles and three bond lengths. There are many ways to remove the redundant coordinates: one can use three bond lengths, one bond length and two angles, or two bond lengths and one angle. It is unclear what the best choice will be. The redundancy problem becomes more severe with increasing molecule size, as it is not uncommon that the number of internal coordinates is an order of magnitude larger than $3M - 6$.[26, 27]

The inherent redundancy of the internal coordinates can be removed automatically or manually, by explicit construction. The most common manual approach is to define a set of non-redundant internal coordinates by constructing a Z-matrix.[28] In a Z-matrix, each atom's position is specified by one bond length,

one bond angle, and one dihedral angle. This gives $3M$ coordinates. The extra redundancy is removed by defining one atom as a reference atom, and not specifying any of its three coordinates relative to other atoms. A second atom's position is defined with a reference bond (one interatomic distance to an atom, typically the reference atom), but no bond angle or dihedral angle. A third atom's position is defined with a second reference bond and a bond angle (typically defined as the angle between the two reference bonds), but no dihedral coordinate.

The Z-matrix strategy performs seamlessly in many cases, though the performance can be sensitive to the specific bond lengths, angles, and dihedrals included. The transformation between Cartesian coordinates and the Z-matrix internal coordinates is likewise straightforward: bond lengths, angles, and dihedrals can be determined with straightforward trigonometry and, because the Z-matrix is nonredundant, any change in Z-matrix can be realized by a corresponding change in Cartesian coordinates. The Z-matrix strategy, however, performs poorly for cyclic molecules, because one of the bonds in the ring will be missing.[29, 30] For example, in ozone, which is a bent molecule, there is an obvious choice for the two bonds and the one bond angle that should be included in the Z-matrix. However, for isoozone, which is an equilateral triangle structure, picking the correct bonds and angles is ambiguous, and the molecular symmetry of the structure is not respected. Therefore, for a cyclic molecule, the bond which closes the ring is missing, the elongation and contraction of this bond can only be described using the other bonds and angles in the ring. These bonds and angles are therefore tightly coupled, and the Hessian matrix has a significant off-diagonal structure.

The problem of rings, along with other problems associated with arbitrary user choices that need to be made when constructing a Z-matrix, leads to the idea of

explicitly using redundant internal coordinates.[11] A set of primitive redundant internal coordinate is formed by including *all* chemically-sensible bonds, angles, and dihedrals, along with out-of-plane bends. The number of redundant internal coordinates in a system is larger than the $3M-6$ degree of freedom. To reduce the dimensionality of the redundant internal space to the desired $3M-6$, one takes suitable linear combinations of the redundant internal coordinates. One popular way to do this is to generate delocalized internal coordinates, but there are other choices.[26, 31]

1.3.3 Transformation between Cartesian coordinates and redundant internal coordinates

The Cartesian coordinates are normally used in quantum chemistry software packages to compute the energy and its derivatives. However, as mentioned in the previous section, internal coordinates are more suitable for geometry optimization. Therefore, one must be able to interconvert Cartesian and internal coordinates, along with the gradient and Hessian in these coordinate systems. At each optimization step, the gradient and (approximate) Hessian in internal coordinates will be used to compute a displacement of the internal coordinates, which then needs to be transformed back to Cartesian space so that the energy, gradient, and possibly Hessian can be computed for the next step.

The key tool in these transformations is the Wilson **B** matrix, which is the Jacobian of the transformation from Cartesian to internal coordinates, with elements,[32]

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \tag{1.13}$$

With Wilson **B** matrix, one can convert an infinitesimal change in Cartesian space to its corresponding change in redundant internal coordinates.

$$\delta\mathbf{q} = \mathbf{B} \cdot \delta\mathbf{x} \tag{1.14}$$

For most molecules with more than a few atoms, the number of redundant internal coordinates is far greater than $3M$, so **B** is rectangular and singular (because internal coordinates, but not Cartesian coordinates, are invariant to molecular translation and rotation). To compute the change in Cartesian coordinates introduced by an infinitesimal change in internal coordinates, the Moore-Penrose pseudo-inverse, $\mathbf{B}^+$, is used

$$\mathbf{B}^+ \cdot \delta\mathbf{q} = \delta\mathbf{x} \tag{1.15}$$

The gradient and Hessian can be converted between the internal, $(\mathbf{g}_q, \mathbf{H}_q)$, and Cartesian, $(\mathbf{g}_x, \mathbf{H}_x)$, coordinate systems using:

$$\mathbf{g}_x = \mathbf{B}^T\mathbf{g}_q \tag{1.16}$$

$$\mathbf{g}_q = \left(\mathbf{B}^T\right)^+\mathbf{g}_x \tag{1.17}$$

$$\mathbf{H}_x = \mathbf{B}^T\mathbf{H}_q\mathbf{B} + \mathbf{K} \tag{1.18}$$

$$\mathbf{H}_q = \left(\mathbf{B}^T\right)^+\Big(\mathbf{H}_x - \mathbf{K}\Big)\mathbf{B}^+ \tag{1.19}$$

where **K** is the matrix including the second derivatives of the internal coordinates with respect to Cartesian coordinates,

$$k_{jk} = \sum_{i=1}^{N_{int}} \Big[\mathbf{g}_q\Big]_i \frac{\partial^2 q_i}{\partial x_j \partial x_k} = \sum_{i=1}^{N_{int}} \Big[\mathbf{g}_q\Big]_i \frac{\partial b_{ij}}{\partial x_k} \tag{1.20}$$

Unlike the (explicit) transformation from Cartesian to internal coordinates, the transformation from internal coordinates to Cartesian coordinates cannot be expressed as a simple formula, and various iterative methods are used. Suppose the optimization starts from an initial structure $\mathbf{x}_0$, and its corresponding internal representation $\mathbf{q}_0$. The target structure we wish to converge is denoted similarly as $\mathbf{x}_{target}$ with internal $\mathbf{q}_{target}$. The first step towards the target is computed as

$$\mathbf{s}_0 = \mathbf{q}_{target} - \mathbf{q}_0 \tag{1.21}$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{B}^+\mathbf{s}_0 \tag{1.22}$$

$$\mathbf{q}_1 = \mathbf{q}(\mathbf{x}_1) \tag{1.23}$$

$$\mathbf{s}_1 = \Delta\mathbf{q} = \mathbf{q}_{target} - \mathbf{q}_1 \tag{1.24}$$

At the k^{th} iteration, the new structure $\mathbf{x}_k$ is computed

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{B}^+\mathbf{s}_{k-1} \tag{1.25}$$

$$\mathbf{s}_k = \mathbf{q}_{target} - \mathbf{q}_k \tag{1.26}$$

until $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ are sufficiently close together. However, this fixed-point iteration method does not always work. Typically, but not always, this failure is due to (nearly) linear bond angles.

We propose a different, robust, method for converting structures from internal coordinates to Cartesian coordinates. The strategy is based on the idea that only a $(3M-6)$-dimensional manifold within the N_int-dimensional internal-coordinate space correspond to physically realizable molecular geometries, and therefore it is only points on this manifold that have Cartesian-coordinate representations. Our

am is to choose the Cartesian structure, $\mathbf{x}$, on this manifold, $\mathbf{q}(\mathbf{x})$, that is closest to the target set of internal coordinates $\mathbf{q}_{target}$,

$$\underbrace{\min}_{x} \Big(\mathbf{q}_{target} - \mathbf{q}(\mathbf{x})\Big)^T \mathbf{W} \Big(\mathbf{q}_{target} - \mathbf{q}(\mathbf{x})\Big) \tag{1.27}$$

Here $\mathbf{W}$ is a positive-definite diagonal matrix with weight w_i for each internal coordinates. Eq.1.27 minimizes the weighted-squared deviation between the optimized and target structures. By default $\mathbf{W}$ matrix is the identity matrix, but sometimes it is beneficial to prioritize certain internal coordinates over others (e.g., in a constrained optimization).

1.4 Numerical Methods for Optimization

Newton-Raphson Method

Starting from the initial structure on the potential energy surface with coordinates $\mathbf{x}_0$ and energy U_0, the energy of nearby structures can be estimated through Taylor expansion

$$E(\mathbf{x}) = E(\mathbf{x}_0) + \mathbf{g}_0^T(\mathbf{x} - \mathbf{x_0}) + \frac{1}{2}(\mathbf{x} - \mathbf{x_0})^T \mathbf{H}_0 (\mathbf{x} - \mathbf{x_0}) + ... \tag{1.28}$$

Higher-order derivatives of energy are expensive to compute, so they are not routinely available in quantum-chemistry programs. Therefore the Taylor series is usually truncated after the second derivative. The gradient can also be expanded as a Taylor series,

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}_0 + \mathbf{H}_0(\mathbf{x} - \mathbf{x_0}) + ... \tag{1.29}$$

Truncating after the Hessian matrix, Eqn.1.29 becomes

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}_0 + \mathbf{H}_0 \Delta x \tag{1.30}$$

where $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x_0}$ is the step. The preceding equations hold for any coordinate system.

Key molecular structure like the reactions, products, (quasi)stable reactive intermediates, and transition states are all stationary points on the potential energy surface with zero gradient, $\mathbf{g}(\mathbf{x}) = 0$.

$$\mathbf{0} = \mathbf{g}_x = \mathbf{g}_0 + \mathbf{H}_0 \Delta x \tag{1.31}$$

Rearranging this equation, one can estimate the change in molecular structure that is needed to obtain a stationary point by either solving a system of linear equations or inverting the Hessian matrix,

$$\mathbf{H}_0 \Delta\mathbf{x} = -\mathbf{g}_0 \tag{1.32}$$

$$\Delta\mathbf{x} = -\mathbf{H}_0^{-1}\mathbf{g}_0 \tag{1.33}$$

The step $\Delta\mathbf{x}$ in Eqn.1.33 is called the Newton step. One can update the structure $\mathbf{x}_0$ with the Newton step to $\mathbf{x}_1 = \mathbf{x}_0 + \Delta\mathbf{x}$. If the the quadratic approximation is exact (i.e., all the higher-order derivatives of the potential energy are zero), the gradient of $\mathbf{g}(x_1)$ will be exactly zero and the optimization is finished. This is not the case for real molecules, so one needs to iterate until convergence is reached. This iterative process of solving the linear equation is called *Newton–Raphson* method.

The *Newton–Raphson* method is based on the hypothesis that the objective function at the initial point is well-approximated by a paraboloid. It converges rapidly when the higher-order derivatives are small and is the basis for other, more-sophisticated, numerical optimization algorithm.[33–36] For smooth objective functions and sufficiently small $\Delta\mathbf{x}$, the higher-order terms in the Taylor series are always negligible. Conversely, if $\Delta\mathbf{x}$ is too large, higher-order terms are nonnegligible and Newton step from 1.33 is no longer reliable. Since higher-order derivatives are rarely practical for molecular structure optimizations, the optimization is instead performed by imposing constraints that ensure the step does not exceed the region where the quadratic approximation is robust. One common approach is to set a trust radius such that $\|\Delta\mathbf{x}\| < \tau$. The τ is set to ensure the proper behaviour of the Newton step.[37, 38] The trust radius is updated based on the performance of each iteration. Another solution is to guarantee the decrease of gradient in each iteration by performing a line-search, i.e., selecting α so that the step $\mathbf{x}_i = \mathbf{x}_{i-1} + \alpha\Delta\mathbf{x}$ minimizes the norm of the gradient.

1.4.1 Quasi-Newton method

Newton's method is accurate and efficient where the quadratic approximation to the potential energy surface is accurate. It also requires the analytic computation of the Hessian matrix. Unlike energy and gradient evaluation, which can be evaluated using the unperturbed (zeroth-order) electronic wavefunction, computation of the Hessian matrix requires computing the first-order electronic wavefunction, and is thus relatively computationally demanding. Therefore, for large molecules, it's computationally prohibitive to evaluate the Hessian to take Newton's step at every optimization iteration. Quasi-Newton methods avoid this cost by using an

approximate Hessian. Specifically, in quasi-Newton methods, the gradients from previous iterations are used to approximate the Hessian matrix of the new configuration.

At the beginning of the optimization, the first Hessian matrix used in the procedure can be computed analytically through *ab initio* methods, or approximated by either semi-empirical or numerical methods.[39, 40] It's also possible to introduce a Hessian with a low accurate molecular mechanical force field[41–43] or even use a pure identity matrix with a scaling factor (though that may lead to slow convergence or outright convergence failures).[44]

After each iteration, a new step $\mathbf{s}_n$ is taken. With the new structure $\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{s}_n$, one can obtain the energy E_{n+1} and gradient $\mathbf{g}_{n+!}$ through a standard computation. The updated Hessian needs to ensure the gradient changes match the step taken under the quadratic approximation for local potential.

$$\Delta \mathbf{g}_n \approx \mathbf{H}_{n+1} \Delta \mathbf{x}_n \tag{1.34}$$

$\mathbf{y}$, denoted as secant condition, is defined as

$$\mathbf{y} = \Delta \mathbf{g}_n = (\mathbf{g}_{n+1} - \mathbf{g}_n) \tag{1.35}$$

Enforcing the secant condition $\mathbf{y}$ is the key to many different quasi-Newton update methods.

Broyden-Fletcher-Goldfarb-Shanno (BFGS)

BFGS is the most famous and widely used quasi-Newton update,[45–49]

$$\mathbf{H}^{new} = \mathbf{H}^{old} + \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}\Delta\mathbf{x}} - \frac{\left(\mathbf{H}^{old}\Delta\mathbf{x}\right)\left(\mathbf{H}^{old}\Delta\mathbf{x}\right)^T}{\left(\Delta\mathbf{x}^T\right)\mathbf{H}^{old}\Delta\mathbf{x}} \tag{1.36}$$

Given a positive-definite Hessian matrix, the BFGS method preserves its positive-definite nature, ensuring that the update step is always in an energy-decreasing direction. This feature makes BFGS ideal for minimization, such as locating reactants and products, but ineffective for transition states, where accomodating negative eigenvalues of the Hessian is essential.

Symmetric-Rank-1 (SR1)

SR1 is a simple and straightforward rank-one update method,[50]

$$\mathbf{H}^{new} = \mathbf{H}^{old} + \frac{\left(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\right)\left(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\right)^T}{\left(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\right)^T\Delta\mathbf{x}} \tag{1.37}$$

The SR1 method does not guarantee a positive-definite matrix. This feature makes SR1 a better candidate for transition state optimization than BFGS. There is also a drawback to this method. When the value of $\mathbf{H}^{old}\Delta\mathbf{x}$ is very close to $\mathbf{y}$, a numerical ill-conditioning occurs due to division by zero. To circumvent this problem, one can check the value of $\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}$ first and conduct the Hessian update only when the difference is not negligible.

Powell-symmetric-Broyden (PSB)

PSB is a rank-two Hessian update method. It has the advantage of an accurate rank-two level adjustment without constraining the update to be positive-definite,[51]

$$\mathbf{H}^{new} = \mathbf{H}^{old} + \frac{\Big(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\Big)\Delta\mathbf{x}^T + \Delta x\Big(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\Big)^T}{\|\Delta\mathbf{x}\|^2} - \Big(\frac{\Delta\mathbf{x}\cdot\Big(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\Big)}{\|\Delta\mathbf{x}\|^4}\Big)\Delta\mathbf{x}\mathbf{x}^T \tag{1.38}$$

Bofill

Both SR1 and PSB are proper candidates for geometry optimization, including transition states. Bofill proposed a mixed method combining the SR1 and PSB methods.[52]

$$\mathbf{H}^{Bofill} = \phi\mathbf{H}^{SR1} + (1-\phi)\mathbf{H}^{PSB}$$
$$\phi = \frac{\Big\|(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x})^T\Delta\mathbf{x}\Big\|^2}{\|\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x}\|^2\|\Delta x\|^2} \tag{1.39}$$

The Bofill update is considered among the best quasi-Newton methods for transition state optimization.

The Hessian matrix for transition state needs to have exact one negative eigenvalue; its corresponding eigenvector is directed along the reaction coordinate. Knowing the information about bond-breaking and forming can also improve the performance and effectiveness of the optimization process. In Chapter3, we introduce a new optimization algorithm that exploits chemical intuition to ensure an appropriate negative-eigenvalue direction. Specifically, from the difference between

the reactant and product, the algorithm can recognize the key internal coordinates. These reduced coordinates are used to ensure that the Hessian matrix has the correct eigenstructure.

1.5 Iteration strategies

During the optimization process, one starts from the initial structure $\mathbf{x}_0$ and iteratively changes the structure to approach a stationary point on the potential energy surface. The iterative process is terminated when certain stopping protocols (e.g., a sufficiently small value for the gradient) are satisfied. Each successive candidate structure is determined using the gradient $\mathbf{g}_i$ and (approximate) Hessian $\mathbf{H}_i$ of the current structure $\mathbf{x}_i$, possibly together with information from previous points $\mathbf{x}_{i-1}, \mathbf{x}_{i-2}...\mathbf{x}_0$.

1.5.1 Line Search method

In the linear search method, a direction vector $\mathbf{p}_i$ is chosen. The algorithm searches along the specified direction until a point with sufficiently lower energy or smaller gradient magnitude is found. This simplifies the multidimensional optimization problem into a one-dimensional optimization,

$$\underbrace{\min}_{\alpha>0} f(\mathbf{x}_i + \alpha \mathbf{p_i}) \tag{1.40}$$

Accurately minimizing this expression is expensive and unnecessary because too many steps are taken when the structure is still far from the minimum. Convergence can be guaranteed as long as a sufficient decrease in the objective function is

achieved. After a new point $\mathbf{x}_{i+1}$ is found, a new $\mathbf{p}_{i+1}$ is selected, and the algorithm is repeated.[44]

Steepest Descent

When choosing a optimization direction, the easiest choice is the direction of $-\nabla U(\mathbf{x}_i)$, namely, $-\mathbf{g}_i$. This is the direction along with the energy of the system decreases most rapidly. Following the gradient descent direction, the energy change of the system is approximated by the Taylor's expansion up to the second order derivatives,

$$\Delta U \approx f(\mathbf{x}_i + \alpha \mathbf{p}_i) - f(\mathbf{x}_i) = \alpha \mathbf{p}_i^T \nabla f_i + \frac{1}{2}\alpha^2 \mathbf{p}_i^T (\nabla^2 f_i)\mathbf{p}_i \tag{1.41}$$

To minimize the value of ΔU is equivalent to

$$\underbrace{min}_{\mathbf{p}_i} \mathbf{p}_i^T \nabla f_i = \underbrace{min}_{\mathbf{p}_i} \|\mathbf{p}_i\| \|\nabla f_i\| cos\theta \tag{1.42}$$

where $\mathbf{p}_i$ is a unit vector, $\|\mathbf{p}_i\| = 1$. Eqn. 1.42 is minimized when the $\cos\theta = -1$.

Linear search along the steepest descent direction is an effective way to lower the energy of the system, especially because no information about the Hessian matrix is required. Pictorially, the direction of the move is perpendicular to the contour of the energy surface. Steepest descent works well when the contours of the objective function are well scaled, but this is frequently untrue in chemical applications, where the eigenvalues of the Hessian matrix span several orders of magnitude. Steepest-descent, then, should only be used as a last result in molecular structure optimizations.[53]

Newton direction

The Newton direction is derived from Eqn.1.33, based on the local quadratic approximation. For minimization, the linear search method requires a positive-definite Hessian matrix to fulfill the descent requirement. However, iteratively taking the full Newton step does not always converge. Instead, the Newton direction is used, and then scaled by the factor α using a line-search protocol to ensure sufficient reduction in the energy (or the gradient norm).

Using the Newton step gives relatively fast and robust convergence. The limitation is the computation cost for the Hessian matrix in each iteration. Without explicitly compute the Hessian matrix, one can either use finite-differences to estimate the Hessian (e.g., the truncated Newton method) or use approximations to the Hessian matrix (quasi-Newton methods).

1.5.2 Trust-Region Methods

To compute a more accurate and consistent step for optimization, a model function is normally taken to estimate the potential energy locally, near the current structure. For example, the newton step from 1.33 is premised on the quadratic energy model. The model is accurate for steps close to the current structure $\mathbf{x}_i$. If the step is too long, the higher-order derivatives omitted in the quadratic model become non-negligible and the approximations in the local model become inaccurate [44, 54]. To ensure a proper step size, it is important to limit the maximum length of the step at each iteration,

$$\|\Delta x\| \leq \tau \tag{1.43}$$

where τ is the trust radius set according to the accuracy of the current model. When a step surpasses the trust radius, the step is deemed risky and scaled back accordingly. Specifically, when the step from Eqn.1.33 is larger than τ, one minimizes the energy subject to the constraint on the step-size. Introducing $\tilde{\lambda}$ as the Lagrange multiplier for the step-size constraint, the constrained Newton step is obtained by solving the linear system,

$$(\mathbf{H} + \tilde{\lambda}\mathbf{I})\Delta\mathbf{x} = -\mathbf{g} \tag{1.44}$$

When minimizing the objective function, $\tilde{\lambda}$ is set to be positive. Conversely, when maximizing the function, $\tilde{\lambda}$ is negative. Expanding the Hessian matrix with spectral theorem,

$$\mathbf{H} = \sum_{i=1}^{N_{int}} \lambda_i \chi_i \chi_i^T \tag{1.45}$$

the solution for optimizaiton step can be expressed as[52, 55]:

$$\Delta\mathbf{x}(\tilde{\lambda}) = \sum_{i=1}^{N_{int}} \frac{-1}{\lambda_i + \tilde{\lambda}\mathrm{sgn}(\lambda_i)} \chi_i \chi_i^T \mathbf{g} \tag{1.46}$$

where $\tilde{\lambda}$ is determined by solving the nonlinear equation,

$$\left\|\Delta\mathbf{x}(\tilde{\lambda})\right\| = \tau \tag{1.47}$$

and $\mathrm{sgn}(\lambda_i)$ denotes the sign of each eigenvalue. This shift ensures eigenvalues retain their sign but the step length is reduced. The τ from previous iteration need to be updated with respect to the accuracy of local quadratic approximation. A conventional method is comparing the real energy difference between two

structures with approximated energy change from the model

$$\rho = \frac{E(\mathbf{x}+\Delta\mathbf{x}) - E(\mathbf{x}))}{\mathbf{g}^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}\mathbf{H}\mathbf{x}} \tag{1.48}$$

When the two energies are close, $\rho \approx 1$ and the quadratic model is considered accurate, so τ is increased in the next iteration. If the quadratic model is poor, the trust radius is decreased. In intermediate situations, the trust radius is retained for the next iteration.

The preceding method applies the same scaling $\tilde{\lambda}$ for both positive and negative eigenvalues. There is also revised version where separate values are used for negative-curvature and positive-curvature respectively,

$$\Delta\mathbf{x} = \frac{-1}{\lambda_1 - \lambda_n}\chi_1\chi_1^T\mathbf{g} + \sum_{i=2}^{N_{int}} \frac{-1}{\lambda_i + \lambda_p}\chi_i\chi_i^T\mathbf{g} \tag{1.49}$$

λ_p and λ_n must obviously be selected to ensure that $\lambda_1 - \lambda_n < 0$ and $\lambda_i + \lambda_p > 0$.

Another popular method is the rational function optimization(RFO) method.[56–59] In the RFO method, the quadratic approximation is replaced by a rational function model. This shift allows higher order derivatives to be approximated through Padé approximation. The energy change is expressed as

$$\Delta E(\mathbf{x}) = \frac{\mathbf{g}^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\mathbf{H}\Delta\mathbf{x}}{1 + \Delta\mathbf{x}^T\mathbf{S}\Delta\mathbf{x}} \tag{1.50}$$

To minimize the energy change, the equation is rewritten as a system of linear equations

$$\begin{bmatrix} \mathbf{H}^{old} & \mathbf{g}^{old} \\ (\mathbf{g}^{old})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ 1 \end{bmatrix} = 2\Big(\Delta E\Big) \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} \tag{1.51}$$

where $\mathbf{S}$ is normally chosen to be a scalar times the identity matrix, $\mathbf{S} = \xi\mathbf{I}$ After solving the Eqn.1.51, the smallest eigenvalue is a proper candidate for minimization tasks, while the second smallest eigenvalue is more suitable for transition state optimization.

The RFO method permits a strategy for 1.49, wherein one solves two separate generalized eigenvalue problems of negative-curvature and positive-curvature respectively for the Lagrange multipliers:

$$\begin{bmatrix} \lambda_1 & \mathbf{x}_1^T\mathbf{g} \\ \mathbf{g}^T\mathbf{x}_1 & 0 \end{bmatrix} \mathbf{v}_n = \lambda_n \begin{bmatrix} \xi & 0 \\ 0 & 1 \end{bmatrix} \mathbf{v}_n \tag{1.52}$$

$$\begin{bmatrix} \lambda_2 & 0 & 0 & \dots & \mathbf{x}_2^T\mathbf{g} \\ 0 & \lambda_3 & 0 & \dots & \mathbf{x}_3^T\mathbf{g} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{3N-6} & \mathbf{x}_{3N-6}^T\mathbf{g} \\ \mathbf{g}^T\mathbf{x}_2 & \mathbf{g}^T\mathbf{x}_3 & \dots & \mathbf{g}^T\mathbf{x}_{3N-6} & 0 \end{bmatrix} \mathbf{v}_p = \lambda_p \begin{bmatrix} \xi & 0 & 0 & \dots & 0 \\ 0 & \xi & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \xi & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \mathbf{v}_p \tag{1.53}$$

The absolute value of the largest eigenvalue from 1.52 is assigned to λ_n while the absolute value of the smallest eigenvalue of 1.53 is allocated to λ_p. In this case, the minimization is taken place in all the other space while the negative eigenvector is left to ascend the energy barrier towards the transition state on the potential energy surface.

1.6 Transition State Optimization

There are two main tasks for molecular geometry optimization: minimization and saddle point optimization. When conducting minimization, procedures are relatively straightforward. One can take a step in the energy descent direction until a local minimum is attained. Saddle point optimization is more complicated because, at a stationary point on the PES, the structure is the maximum in some direction(s) but minimum in the other direction(s). When taking a step towards the desired saddle point, the energy can go either uphill or downhill. It's also possible to observe an increase in the magnitude of the gradient when taking a good step towards the transition state. Without extra information about the landscape, one's ability to optimize to a desired saddle point is quite limited.

To approach the transition state of interest, three categories of methods are generally used. The most common one is to generate a guess based on chemical intuition. The guess structure is expected to be close to the target transition state structure. The numerical optimization procedures can effectively optimize the guess structure to the saddle point from the input geometry. Other methods try to automate the searching procedure by exploiting the information from the reaction, product, or both.

1.6.1 Single-Ended Methods

One common method is to start the geometry optimization from one end of the reaction, typically the reactant or product. The path to the transition state requires the initial structure to go uphill on the potential energy surface. Beginning

from a near minimum point, every direction is an energy-ascending direction. Selecting a proper direction to drive the energy higher is crucial to the success of the algorithm.

The coordinate-driving method supposes that the reaction can be characterized by one coordinate. The algorithm takes steps in that direction and, at each step, constrained optimization is performed to minimize all the other coordinates. If the minimum energy path can be parameterized using the selected coordinate, then this procedure recovers the minimum-energy path and the highest point along the path provides a good guess for the transition state guess structure. Coordinate-driving is ineffective when a reaction cannot be simply characterized. For example, when the path is curved so that the value of the coordinate-driving coordinate does not increase or decrease monotonically, the coordinate-driving method fails.[60–63] The coordinate-driving method also presupposes significant information about the reaction mechanism, which may not be accessible in all cases.

To extend the applicability of coordinate driving, one can select more additional internal coordinates as the driven coordinates (multi-dimensional coordinate driving) or construct compound coordinates. For example, in a hydrogen-transfer reaction like isomerization from HCN to CNH, the bond angle $\measuredangle_{HCN}$ changing from 0° to 180° is a sensible choice. If the reaction involving an atom moving directly from one donor to the acceptor, such as the atom transfer reaction from $AB \cdots C$ to $A \cdots BC$, the difference between to two bonds $q = \mathbf{R}_{BC} - \mathbf{R}_{AB}$ is a more descriptive driving coordinate.

Another simple but effective method is the direction-of-least-ascent, which is often called the hill-climbing method.[64, 65] This method leads the structure to go uphill towards the direction with the least energy ascending. It performs well

for small molecules with a simple mechanism. When the molecules are large, the least-ascent direction normally characterizes a conformational change (e.g., a methyl rotation) instead of the desired chemical transformation. Therefore, for large molecules, it is more reliable to follow the eigenvector of the Hessian matrix that corresponds to the reaction path of interest.[66, 67]

The dimer method is an algorithm for this eigenvector following technique.[68–71] In the dimer method, two points are kept at a fixed small distance. In each iteration, the curvature is calculated by finite difference between the points' energies along the axis. The direction of the next step is determined by rotating the dimer to align with the lowest curvature direction. Then the dimer is displaced in this direction. This allows the lowest-eigenvalue direction to be traced without explicit computation of the Hessian. The cost of the dimer method is mainly determined by the rotation of the dimers for the lowest eigenvalue mode.

Gradient-extremal path is an alternative method to construct a path from one stationary point to another stationary point on the potential energy surface.[72–74] At each stationary point, the gradient $\mathbf{g}$ is a eigenvector of the Hessian matrix $\mathbf{H}(s)\mathbf{g}(s) = 0 \cdot \mathbf{g}(s)$. The gradient extremal path, $\mathbf{x}(s)$, is defined by the eigenvalue condition,

$$\mathbf{H}(s)\mathbf{g}(s) = \lambda(s)\mathbf{g}(s) \tag{1.54}$$

where s is the arc length in the path and $\mathbf{g}(s)$ and $\mathbf{H}(s)$ are the gradient and Hessian at $x(s)$, respectively. Starting from any point on the potential energy surface, one can follow Eqn.1.54 to generate a path towards a stationary point on the surface. Because gradient-extremal paths pass through stationary points, paths starting from different initial structures intersect at the stationary structures. There are

also some drawbacks in the gradient-extremal method. It has the tendency to form a circuitous path on the surface. The path generated from gradient-extremal is also dependent on coordinate system. Finally, in each step, an analytical calculation for Hessian matrix is required, so direct implementation of the gradient-extremal path is computationally costly.

The reduced-gradient-following(RGF) method is also devised for stationary points searching.[75–77] In RGF, the direction of gradient in each iteration is fixed by a constant. Like the gradient-extremal path, RGF paths starting from different directions intersect at the stationary points of the potential energy surface. Unlike gradient-extremal paths, RGF paths do not require computing the Hessian at each step.

All these methods have some intrinsic drawbacks as the algorithm is searching for a one-dimensional parameterized path for the transition state. The performance is good when the molecule is small. But when the system is large, these methods are no longer reliable because of the high dimensionality of the potential energy surface. Multi-dimensional surface walking is considered the most advantageous choice when dealing with more complicated system.[78–81] One can select several key internal coordinates involving in the reaction as the reduced space to generate a reduced potential energy surface. The reduced surface is characterized by the key internal coordinates while keeping all the other coordinates minimized. However, multi-dimensional surface walking is computationally costly (with cost growing exponentially with the number of coordinates retained), and selecting the right key coordinates is crucial to the success of the algorithm. It is not that difficult to identify processes where six or more key coordinates may be required.

1.6.2 Double-Ended Methods

Double-ended methods (more generally, multi-ended methods) exploit the information from the reactant and product (and, more generally, reactive intermediates) to provide a more comprehensive description of the reaction and the transition-state region. In double-ended methods, the reaction path is initially represented by a sequence of points that interpolate between the reactant and the product. During the optimization process, each point is optimized starting from its initial guess structure towards a point on the minimum-energy reaction path.

One of the most popular double-ended method is nudged elastic band (NEB) method.[82–90] In NEB method, the reaction path consists of several equally-spaced points. All the points are connected by a virtual spring with zero equilibrium length.

$$V^{spring} = \frac{1}{2}\left(\|\mathbf{x}_i - \mathbf{x}_{i-1}\|^2 + \|\mathbf{x}_{i+1} - \mathbf{x}_i\|^2\right) \tag{1.55}$$

At each point, the gradient of the point is the sum of potential surface and the spring

$$\mathbf{g} = \mathbf{g}^{spring} + \mathbf{g}^{PES} \tag{1.56}$$

where $\mathbf{g}^{spring} = \frac{dV^{spring}}{ds}$ and $\mathbf{g}^{PES}$ is the surface potential gradient. The NEB method projects out the perpendicular components of spring gradient and the parallel components of the surface potential gradient during the optimization. The

force at each point is thus

$$\mathbf{g}^{NEB} = \mathbf{g}_{\parallel}^{spring} + \mathbf{g}_{\perp}^{PES} \tag{1.57}$$

$$\mathbf{g}_{\parallel}^{spring} = \tau\tau^{T}\mathbf{g}^{spring} \tag{1.58}$$

$$\mathbf{g}_{\perp}^{PES} = (\mathbf{I} - \tau\tau^{T})\mathbf{g}^{PES} \tag{1.59}$$

where τ is the unit vector tangent to the reaction path. The $\mathbf{g}_{\parallel}^{spring}$ component in $\mathbf{g}^{NEB}$ is solely for maintaining equal distance between points while the $\mathbf{g}_{\perp}^{PES}$ is used to drag points downhill to the optimal position in the perpendicular space. This separation decouples the interference of spring force from the relaxation process, circumventing the corner-cutting problem of the (non-nudged) elastic band method for curved reaction paths.

For some reactions, the energy of the system changes rapidly without enough restoring force acting on it. This situation causes a kinky reaction path, which slows down the convergence. To solve this problem, a switch function is introduced into NEB method to gradually increase the perpendicular components of the spring force.

$$\mathbf{g}^{\tilde{NEB}} = \mathbf{g}^{NEB} + \mathbf{f}(\phi)\Big(\mathbf{g}^{spring} - \mathbf{g}_{\parallel}^{spring}\Big) \tag{1.60}$$

where the switch function, $\mathbf{f}(\phi)$, change from 0 to 1 as the angle of the path change from 0° to 90°.

$$\mathbf{f}(\phi) = \frac{1}{2}(1 + \cos(\pi\cos(\phi))) \tag{1.61}$$

The drawback of the NEB method is it couples neighbouring points on the reaction path, resulting in a slow optimization process.

In the string method(SM), the reaction path is set to be a smooth one-dimensional

curve connecting the reactant and product.[91–96] The curve is parameterized by the reaction progress, $\mathbf{x}(t)$, where t is normalized to be 0 for the reactant and 1 for the product. The initial guess string is generated as an interpolation from the reactant to the product. The string path is expected to fit the minimum energy path(MEP) of the reaction after the optimization. This requires the tangent unit vector τ of the path to be parallel to the gradient

$$\tau = \frac{d\mathbf{x}(t)}{dt} \propto \mathbf{g} \tag{1.62}$$

The force of each point on the path when the point is not on the MEP is

$$\mathbf{F} = -\mathbf{g}_{\perp} = (\mathbf{I} - \tau\tau^{T})\mathbf{g}^{PES} \tag{1.63}$$

Following the Eqn.1.63 the force gradually evolves the initial guess towards the desired MEP. Since it is impossible to optimize an infinite number of points on the path, practical implementations of the string method represent the string path with discrete states and connect the states by an interpolation curve, typically a cubic spline. The objective of the optimization is to minimize $\sum_i \left\|\mathbf{g}_{\perp}^{(i)}\right\|$ at each state along the whole string path. One can adopt various numerical optimization methods such as the steepest-decent or Quasi-Newton method to update the state.

Unlike the NEB method, there's no direct spring energy between neighbouring points. Less-coupled states make the string method converge more smoothly. However, without the nudged force, the states on the string must be redistributed to ensure adequate coverage near the transition state region. Also, an extra "kinking" force is required to straighten out the path, as the path sometimes doubles-back on itself, especially in regions where the gradient is small.

In the growing string method (GSM), the number of states representing the path increases steadily.[97–100] Starting from both reactant and product, states are added to the string until the paths meet and a complete reaction path is constructed. A parameterized density function and an indicator function are defined to keep track of the newly added points and the spacing. For each growing-string step, the density function and indicator function evolve to ensure a uniform distribution of states. The optimization is carried separately between the two segments of the united path. The highest energy point interpolated on the united path is the optimal transition state guess for further optimization. To reduced the computational burden, low-level computational methods are used to sketch the string path with a relative efficient optimization algorithm. Then a more advanced ansatz is applied for a more accurate result. The GSM method works well in some cases but, as molecule size increases, it becomes increasingly probable that the reactant-starting-string and the product-starting-string never meet.

1.7 Summary

Effectively modelling chemical reactions requires accurate determination of key molecular structures (reactant, product, intermediates, transition states), their energies, and the reaction path between them. With these pieces of information, the reaction mechanism can be explored in detail, and key thermodynamic and kinetic properties of the chemical reaction can be modelled.

In this chapter, we reviewed molecular geometry optimization. While geometry optimization is an old problem, it is also difficult, so it is still an active area for

research. First, a good initial guess structure is important for geometry optimization. It's relatively easy to generate a good starting point for minimization to find the optimal geometry for a reactant, product, or reactive intermediate. It's more challenging to find a good guess for transition states. In chapter.2, a new efficient and unambiguous method is introduced. The generated structure is an interpolation between the structures of the reactant and the product. Besides, a new set of robust dihedrals are also implemented to facilitate a more comprehensive representation of molecules during the optimization process.

The most efficient geometry optimization methods use the Taylor expansion of the potential energy surface to provide local information about the surface near a given molecular structure. When the third- and higher-order derivatives are omitted, the update step is called the Newton step or quasi-Newton step, depending on whether the Hessian is calculated exactly or approximated. Transition state optimization requires the Hessian matrix to have one exact negative eigenvalue. We present a new hybrid method where reaction information is used to identify key internal coordinates. The entries of the Hessian associated with the reaction space are then explicitly evaluated through finite differences, while the remaining elements are updated using traditional quasi-Newton methods. This new approach combines the advantage of accurate Hessian information with a fast update method. In Chapter.3, systematic testing proves this strategy is promising for effectively converge transition state optimization.

In Chapter.5, comprehensive testing is conducted to measure the performance of our new approach. Different initial guess structures are generated by a certain amount of perturbation from the transition state structure. The difficulty increases as the perturbation change from 0.1 to 0.4 atomic units. The result shows our new

approach is better than traditional approaches even when the initial guess is poor.

When it is especially difficult to find an initial guess for a transition state, it is helpful to first find a (approximate) reaction path on the potential energy surface. In Chapter.5, we introduced a new double-ended approach to generate the reaction path by sequential bisection. The optimization technique is the same as the one in Chapter.3 except the key coordinates are selected as the reaction vector. Unlike the transition state optimization looking for a saddle point, at each iteration of the path point, a local minimum in the hyperplane perpendicular to the reaction path is found.

The key innovations of this thesis are released as a free and open-source software package, GOpt, written in Python 3. GOpt is designed to work with Gaussian, but is easily adapted to in quantum chemistry software method that prints energies and gradients in a commonly accessible format. In addition to the methodological innovations in GOpt, GOpt is distinguished from other packages by its reliance on modern software engineering practices, including comprehensive documentation and complete testing of code correctness, quality, and readability. GOpt is designed to be used as a Python library by other Python packages, and its API is designed to facilitate this usage.

1.8 References

[1] K. Fukui. The path of chemical reactions-the IRC approach. *Accounts of chemical research* 14(12) (1981), 363–368.

[2] M. Biczysko, J. Bloino, and C. Puzzarini. Computational challenges in Astrochemistry. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 8(3) (2018), e1349.

[3] V. Aquilanti, C. Caglioti, P. Casavecchia, G. Grossi, A. Lombardi, F. Palazzetti, and F. Pirani. The astrochemical observatory: Computational and theoretical focus on molecular chirality changing torsions around O–O and S–S bonds. In: *AIP Conference Proceedings*. Vol. 1906. 1. AIP Publishing LLC. 2017, 030010.

[4] M. M. Qasim, B. Moore, L. Taylor, L. Gorb, J. Leszczynski, and P. Honea. Structural characteristics and reactivity relationships of nitroaromatic and nitramine explosives–a review of our computational chemistry and spectroscopic research. *International Journal of Molecular Sciences* 8(12) (2007), 1234–1264.

[5] C. J. Cramer, J. A. Bumpus, A. Lewis, and C. Stotts. Characterization of high explosives and other energetic compounds by computational chemistry and molecular modeling. *Journal of chemical education* 84(2) (2007), 329.

[6] M. Li, J. Sun, D. Han, B. Wei, Q. Mei, Z. An, X. Wang, H. Cao, J. Xie, and M. He. Theoretical investigation on the contribution of HO, SO4-and CO3-radicals to the degradation of phenacetin in water: Mechanisms, kinetics, and toxicity evaluation. *Ecotoxicology and Environmental Safety* 204 (2020), 110977.

[7] Y. Ji, H. Wang, Y. Gao, G. Li, and T. An. A theoretical model on the formation mechanism and kinetics of highly toxic air pollutants from halogenated formaldehydes reacted with halogen atoms. *Atmospheric Chemistry & Physics Discussions* 13(7) (2013).

[8] D. Wales. *Energy landscapes, Cambridge molecular science*. 2003.

[9] D. McQuarrie. Statistical Mechanics (Harper Collins, New York, 1976). ().

[10] D. J. Durand and N. Fey. Computational ligand descriptors for catalyst design. *Chemical reviews* 119(11) (2019), 6561–6594.

[11] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* 17(1) (1996), 49–56.

[12] V. Bakken and T. Helgaker. The efficient optimization of molecular geometries using redundant internal coordinates. *The Journal of chemical physics* 117(20) (2002), 9160–9174.

[13] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *The Journal of chemical physics* 96(4) (1992), 2856–2860.

[14] H. B. Schlegel. Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry* 3(2) (1982), 214–218.

[15] G. H. Jóhannesson and H. Jónsson. Optimization of hyperplanar transition states. *The Journal of Chemical Physics* 115(21) (2001), 9644–9656.

[16] A. Heyden, A. T. Bell, and F. J. Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method

and partitioned rational function optimization method. *The Journal of chemical physics* 123(22) (2005), 224101.

[17] P. N. Plessow. Efficient transition state optimization of periodic structures through automated relaxed potential energy surface scans. *Journal of chemical theory and computation* 14(2) (2018), 981–990.

[18] I. N. Levine, D. H. Busch, and H. Shull. *Quantum chemistry*. Vol. 6. Pearson Prentice Hall Upper Saddle River, NJ, 2009.

[19] G. N. Simm, A. C. Vaucher, and M. Reiher. Exploration of reaction pathways and chemical transformation networks. *The Journal of Physical Chemistry A* 123(2) (2018), 385–399.

[20] N. Schneider, M. Finger, C. Haferkemper, S. Bellemin-Laponnaz, P. Hofmann, and L. H. Gade. Multiple Reaction Pathways in Rhodium-Catalyzed Hydrosilylations of Ketones. *Chemistry–A European Journal* 15(43) (2009), 11515–11529.

[21] J. Greenstadt. On the relative efficiencies of gradient methods. *Mathematics of Computation* 21(99) (1967), 360–367.

[22] H. B. Schlegel. Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1(5) (2011), 790–809.

[23] J. Baker. Techniques for geometry optimization: A comparison of Cartesian and natural internal coordinates. *Journal of computational chemistry* 14(9) (1993), 1085–1100.

[24] J. Baker and F. Chan. The location of transition states: A comparison of Cartesian, Z-matrix, and natural internal coordinates. *Journal of computational chemistry* 17(7) (1996), 888–904.

[25] H. B. Schlegel. A comparison of geometry optimization with internal, cartesian, and mixed coordinates. *International Journal of Quantum Chemistry* 44(S26) (1992), 243–252.

[26] J. Baker, A. Kessi, and B. Delley. The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of chemical physics* 105(1) (1996), 192–212.

[27] G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay. The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces. *Journal of the American Chemical Society* 114(21) (1992), 8191–8201.

[28] M. S. Gordon and J. Pople. Approximate Self-Consistent Molecular-Orbital Theory. VI. INDO Calculated Equilibrium Geometries. *The Journal of Chemical Physics* 49(10) (1968), 4643–4650.

[29] J. Baker and W. J. Hehre. Geometry optimization in cartesian coordinates: The end of the Z-matrix? *Journal of computational chemistry* 12(5) (1991), 606–610.

[30] J. Baker. Techniques for geometry optimization: A comparison of Cartesian and natural internal coordinates. *Journal of computational chemistry* 14(9) (1993), 1085–1100.

[31] J. Baker, D. Kinghorn, and P. Pulay. Geometry optimization in delocalized internal coordinates: An efficient quadratically scaling algorithm for large molecules. *The Journal of chemical physics* 110(11) (1999), 4986–4991.

[32] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1980.

[33] R. Fletcher. Practical methods of optimization john wiley & sons. *New York* 80 (1987), 4.

[34] P. E. Gill, W. Murray, and M. H. Wright. *Practical optimization*. SIAM, 2019.

[35] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.

[36] L. Scales. *Introduction to non-linear optimization*. Macmillan International Higher Education, 1985.

[37] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.

[38] R. Fletcher. Practical methods of optimization john wiley & sons. *New York* 80 (1987), 4.

[39] I. Prigogine and S. A. Rice. *New methods in computational quantum mechanics*. Vol. 198. John Wiley & Sons, 2009.

[40] T. Bredow and K. Jug. Theory and range of modern semiempirical molecular orbital methods. *Theoretical Chemistry Accounts* 113(1) (2005), 1–14.

[41] H. B. Schlegel. Estimating the Hessian for gradient-type geometry optimizations. *Theoretica chimica acta* 66(5) (1984), 333–340.

[42] J. M. Wittbrodt and H. B. Schlegel. Estimating stretching force constants for geometry optimization. *Theochem* 398(399) (1997), 55–61.

[43] T. H. Fischer and J. Almlof. General methods for geometry and wave function optimization. *The Journal of Physical Chemistry* 96(24) (1992), 9768–9774.

[44] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[45] R. Fletcher. A new approach to variable metric algorithms. *The computer journal* 13(3) (1970), 317–322.

[46] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation* 24(109) (1970), 23–26.

[47] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics* 6(1) (1970), 76–90.

[48] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation* 24(111) (1970), 647–656.

[49] D. F. Shanno and P. C. Kettler. Optimal conditioning of quasi-Newton methods. *Mathematics of Computation* 24(111) (1970), 657–664.

[50] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[51] M. J. Powell. A hybrid method for nonlinear equations. *Numerical methods for nonlinear algebraic equations* (1970).

[52] J. M. Bofill. Updated Hessian matrix and the restricted step method for locating transition structures. *Journal of Computational Chemistry* 15(1) (1994), 1–11.

[53] J. C. Meza. Steepest descent. *Wiley Interdisciplinary Reviews: Computational Statistics* 2(6) (2010), 719–722.

[54] J. E. Dennis Jr and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM, 1996.

[55] T. Helgaker, P. Jorgensen, and J. Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2014.

[56] P. Culot, G. Dive, V. Nguyen, and J. Ghuysen. Theor. Chim. Acta (1992).

[57] D. T. Nguyen and D. A. Case. On finding stationary states on large-molecule potential energy surfaces. *The Journal of Physical Chemistry* 89(19) (1985), 4020–4026.

[58] J. Baker. An algorithm for the location of transition states. *Journal of Computational Chemistry* 7(4) (1986), 385–395.

[59] E. Besalú and J. M. Bofill. On the automatic restricted-step rational-function-optimization method. *Theoretical Chemistry Accounts* 100(5-6) (1998), 265–274.

[60] U. Burkert and N. L. Allinger. Pitfalls in the use of the torsion angle driving method for the calculation of conformational interconversions. *Journal of computational chemistry* 3(1) (1982), 40–46.

[61] I. H. Williams and G. M. Maggiora. Use and abuse of the distinguished-coordinate method for transition-state structure searching. *Journal of Molecular Structure: THEOCHEM* 89(3-4) (1982), 365–378.

[62] P. Scharfenberg. Theoretical analysis of constrained minimum energy paths. *Chemical Physics Letters* 79(1) (1981), 115–117.

[63] M. J. Rothman and L. L. Lohr Jr. Analysis of an energy minimization method for locating transition states on potential energy hypersurfaces. *Chemical Physics Letters* 70(2) (1980), 405–409.

[64] G. Crippen and H. Scheraga. Minimization of polypeptide energy: XI. The method of gentlest ascent. *Archives of biochemistry and biophysics* 144(2) (1971), 462–466.

[65] D. Poppinger. On the calculation of transition states. *Chemical Physics Letters* 35(4) (1975), 550–554.

[66] C. J. Cerjan and W. H. Miller. On finding transition states. *The Journal of chemical physics* 75(6) (1981), 2800–2806.

[67] J. Simons, P. Joergensen, H. Taylor, and J. Ozment. Walking on potential energy surfaces. *The Journal of Physical Chemistry* 87(15) (1983), 2745–2753.

[68] G. Henkelman and H. Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of chemical physics* 111(15) (1999), 7010–7022.

[69] J. Kästner and P. Sherwood. Superlinearly converging dimer method for transition state search. *The Journal of chemical physics* 128(1) (2008), 014106.

[70] C. Shang and Z.-P. Liu. Constrained Broyden minimization combined with the dimer method for locating transition state of complex reactions. *Journal of Chemical Theory and Computation* 6(4) (2010), 1136–1144.

[71] A. Heyden, A. T. Bell, and F. J. Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method

and partitioned rational function optimization method. *The Journal of chemical physics* 123(22) (2005), 224101.

[72] J. Pancir. Calculation of the least energy path on the energy hypersurface. *Collection of Czechoslovak Chemical Communications* 40(4) (1975), 1112–1118.

[73] M. Basilevsky and A. Shamov. The local definition of the optimum ascent path on a multi-dimensional potential energy surface and its practical application for the location of saddle points. *Chemical Physics* 60(3) (1981), 347–358.

[74] D. K. Hoffman, R. S. Nord, and K. Ruedenberg. Gradient extremals. *Theoretica chimica acta* 69(4) (1986), 265–279.

[75] W. Quapp, M. Hirsch, O. Imig, and D. Heidrich. Searching for saddle points of potential energy surfaces by following a reduced gradient. *Journal of computational chemistry* 19(9) (1998), 1087–1100.

[76] M. Hirsch and W. Quapp. Improved RGF method to find saddle points. *Journal of computational chemistry* 23(9) (2002), 887–894.

[77] R. Crehuet, J. M. Bofill, and J. M. Anglada. A new look at the reduced-gradient-following path. *Theoretical Chemistry Accounts* 107(3) (2002), 130–139.

[78] K. K. Irikura and R. D. Johnson. Predicting unexpected chemical reactions by isopotential searching. *The Journal of Physical Chemistry A* 104(11) (2000), 2191–2194.

[79] J. M. Bofill and J. M. Anglada. Finding transition states using reduced potential-energy surfaces. *Theoretical Chemistry Accounts* 105(6) (2001), 463–472.

[80] J. M. Anglada, E. Besalú, J. M. Bofill, and R. Crehuet. On the quadratic reaction path evaluated in a reduced potential energy surface model and the problem to locate transition states. *Journal of Computational Chemistry* 22(4) (2001), 387–406.

[81] S. K. Burger and P. W. Ayers. Dual grid methods for finding the reaction path on reduced potential energy surfaces. *Journal of Chemical Theory and Computation* 6(5) (2010), 1490–1497.

[82] G. Henkelman, B. P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics* 113(22) (2000), 9901–9904.

[83] G. Henkelman and H. Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of chemical physics* 113(22) (2000), 9978–9985.

[84] S. A. Trygubenko and D. J. Wales. A doubly nudged elastic band method for finding transition states. *The Journal of chemical physics* 120(5) (2004), 2082–2094.

[85] J.-W. Chu, B. L. Trout, and B. R. Brooks. A super-linear minimization scheme for the nudged elastic band method. *The Journal of chemical physics* 119(24) (2003), 12708–12717.

[86] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras. Adaptive nudged elastic band approach for transition state calculation. *The Journal of chemical physics* 117(10) (2002), 4651–4658.

[87] I. F. Galván and M. J. Field. Improving the efficiency of the NEB reaction path finding algorithm. *Journal of computational chemistry* 29(1) (2008), 139–143.

[88] D. Sheppard, R. Terrell, and G. Henkelman. Optimization methods for finding minimum energy paths. *The Journal of chemical physics* 128(13) (2008), 134106.

[89] D. R. Alfonso and K. D. Jordan. A flexible nudged elastic band program for optimization of minimum energy pathways using ab initio electronic structure methods. *Journal of computational chemistry* 24(8) (2003), 990–996.

[90] N. González-Garcia, J. Pu, À. González-Lafont, J. M. Lluch, and D. G. Truhlar. Searching for saddle points by using the nudged elastic band method: an implementation for gas-phase systems. *Journal of chemical theory and computation* 2(4) (2006), 895–904.

[91] E. Weinan, W. Ren, and E. Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *Journal of Chemical Physics* 126(16) (2007), 164103.

[92] A. Samanta and E. Weinan. Optimization-based string method for finding minimum energy path. *Communications in Computational Physics* 14(2) (2013), 265–275.

[93] P. Y. Ayala and H. B. Schlegel. A combined method for determining reaction paths, minima, and transition state geometries. *The Journal of chemical physics* 107(2) (1997), 375–384.

[94] S. K. Burger and W. Yang. Sequential quadratic programming method for determining the minimum energy path. *The Journal of chemical physics* 127(16) (2007), 164107.

[95] S. K. Burger and W. Yang. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *The Journal of chemical physics* 124(5) (2006), 054109.

[96] E. F. Koslover and D. J. Wales. Comparison of double-ended transition state search methods. *The Journal of chemical physics* 127(13) (2007), 134102.

[97] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of chemical physics* 120(17) (2004), 7877–7886.

[98] A. Goodrow, A. T. Bell, and M. Head-Gordon. Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *The Journal of chemical physics* 129(17) (2008), 174109.

[99] A. Goodrow, A. T. Bell, and M. Head-Gordon. Transition state-finding strategies for use with the growing string method. *The Journal of chemical physics* 130(24) (2009), 244108.

[100] A. Goodrow, A. T. Bell, and M. Head-Gordon. A strategy for obtaining a more accurate transition state estimate using the growing string method. *Chemical Physics Letters* 484(4-6) (2010), 392–398.

Chapter 2

Generating Initial Guesses for Transition States with Redundant Internal Coordinates and Robust Dihedrals

2.1 Abstract

A new set of robust dihedral indicators are designed to circumvent the problem of ill-defined geometry changes associated with the dihedral angle with collinear bonds. Using the robust internal coordinates, an interpolation algorithm is used to generate a high-quality initial guess for the transition-state structure using only the reactant and product structures. A comprehensive assessment confirms the robustness and efficiency of this procedure for guessing transition-state structures.

2.2 Introduction

The performance of geometry optimization methods is sensitive to the coordinate system that is used to specify the molecular geometry.[1–3] Using atoms' Cartesian coordinates is the most straightforward choice, and is implemented by default in many modern quantum chemistry software packages including Gaussian,[4] Psi4,[5] and HORTON.[6] However, during the molecular geometry changes, atoms' Cartesian coordinates are tightly coupled together, which makes this set of coordinates inefficient for geometry optimization.[7] For example, during a simple bond-breaking process represented by Cartesian coordinates, one may keep one group unchanged while moving the others away from the reaction site. To accomplish the stretch, all atoms in the second group have to change their $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ coordinates synchronously to maintain the same relative position. This coupling only becomes more extensive, and more difficult to decode into chemical insight, for more complicated reaction mechanisms.

Therefore, for geometry optimization, it is usually more efficient and intuitive to optimize using internal coordinates including interatomic distances, the angle between bonds, and the dihedral/torsion angles for rotation around bonds. Internal coordinates have direct chemical interpretation and are more weakly coupled, so the second-derivative (Hessian) of the potential energy surface is more diagonally dominant when using internal coordinates than when using Cartesian coordinates. However, there are many different ways to choose internal coordinates, and in some systems (especially heavily-branched molecules and molecules with rings), no single intuitive choice for the internal coordinates exists, and defining a sensible set of internal coordinates using, e.g., a Z-matrix,[8] becomes difficult. Especially in such

cases, it is helpful to use redundant internal coordinates, wherein all chemically intuitive interatomic bonds, bond angles, and dihedrals are included.[9–11] This resolves the difficulty of making an arbitrary choice of internal coordinates, but introduces two new problems:

- An arbitrary change in redundant internal coordinates generally does not correspond to a physically realizable change in molecular geometry. For example, in the ozone molecule, anytime the bond angles are changed so that their sum is not 180 degrees, the structure is not physically realizable.

- When three or more atoms are collinear, a dihedral angle is ill-defined because all choices for the dihedral angle describing rotation about one of the collinear bonds give the same molecular geometry. This causes numerical ill-conditioning of the derivatives of the potential energy surface with respect to the dihedral angle for systems with (nearly) collinear bonds.

Here we present solutions to these issues. First, we provide a specific method for constructing redundant internal coordinates. Then we discuss how the ill-conditioning of the dihedral angle can be removed by using an alternative specification that is robust for near-linear bonds. We then present a method, based on projecting points from the high-dimensional redundant internal-coordinate space to the $(3N-6)$-dimensional manifold of physically-realizable molecular structures, that maps nonrealizable changes in redundant internal coordinates to the closest-possible physically-realizable change. This robust algorithm makes it possible to perform computational studies using sets of redundant internal coordinates that

are far from physically realizable, which allows larger step-sizes in geometry optimization. It also allows one to generate good guesses for transition-state geometries by interpolating between the reactant and product structures in redundant internal coordinates.

2.3 Methodology

2.3.1 Normal Redundant Internal Coordinates

The geometry of a molecule with N atoms can be described by $3N$ Cartesian coordinates, $\{X_i\}_{i=1}^{3N}$. It can also be characterized by internal coordinates: bond lengths, bond angles, dihedral angles, etc., $\{q_i\}_{i=1}^{M_{int}}$. Each internal coordinate can be calculated directly from the Cartesian coordinates:

$$q_{AB}^{bond} = \|\mathbf{R}_{AB}\| \tag{2.1}$$

$$q_{ABC}^{angle} = \cos^{-1}\left(\frac{\mathbf{R}_{BA}\cdot\mathbf{R}_{BC}}{\|\mathbf{R}_{BA}\|\|\mathbf{R}_{BC}\|}\right) \tag{2.2}$$

$$q_{ABCD}^{dihed} = \cos^{-1}\left(\frac{\mathbf{R}_{BA}\times\mathbf{R}_{BC}\cdot\mathbf{R}_{CB}\times\mathbf{R}_{\mathbf{CD}}}{\|\mathbf{R}_{BA}\times\mathbf{R}_{BC}\|\|\mathbf{R}_{CB}\times\mathbf{R}_{CD}\|}\right) \tag{2.3}$$

where $\mathbf{R}_{AB}$ is the interatomic vector in Cartesian coordinates,

$$\mathbf{R}_{AB} = (x_B - x_A, y_B - y_A, z_B - z_A) \tag{2.4}$$

Since the transformation from Cartesian to internal is a non-linear transformation, the inverse cannot be expressed as a matrix. However, small changes in Cartesian coordinates can be mapped into small changes in internal coordinates by a linear transformation. Specifically, the Jacobian matrix of the transformation

is called the Wilson **B** matrix, with elements:[12]

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \tag{2.5}$$

Owing to the redundancy of the internal coordinates, the **B** matrix is rectangular with M rows and 3N columns; typically $M \gg 3N$. For a change in Cartesian coordinates δx, the corresponding change in internal coordinates can be expressed as:

$$\delta \mathbf{q} = \mathbf{B} \cdot \delta \mathbf{x} \tag{2.6}$$

Since the internal coordinates are invariant to molecular translations and rotations but the Cartesian coordinates are not, the Wilson **B** matrix is always singular. Therefore the Moore-Penrose pseudoinverse, $\mathbf{B}^+$, is used for the inverse transformation,

$$\delta \mathbf{x} = \mathbf{B}^+ \cdot \delta \mathbf{q} \tag{2.7}$$

Given a set of Cartesian coordinates $\{\mathbf{x}_0\}$, we can easily construct a set of internal coordinates $\{\mathbf{q}_0\}$ through 2.1 - 2.3. However, given an arbitrary set of redundant internal coordinates, $\{\mathbf{q}_1\}$, it may not be possible to reconstruct a corresponding set of Cartesian coordinates. If $\{\mathbf{q}_1\}$ is close to the known $\mathbf{x}_0(\mathbf{q}_0)$, one can use 2.7 to estimate the Cartesian structure,

$$\mathbf{x}_1 \approx \mathbf{x}_0 + \mathbf{B}^+(\mathbf{q}_1 - \mathbf{q}_0) \tag{2.8}$$

If the internal coordinates corresponding to $\mathbf{x}_1$ are not close enough to $\mathbf{q}_1$, then we can iterate this procedure by setting $\mathbf{x}_0 = \mathbf{x}_1$ and $\mathbf{q_0} = \mathbf{q}(\mathbf{x}_1)$ and reevaluate Eq. (2.8) until eventually convergence,[10, 11, 13–26] which occurs where $\mathbf{x}_0$ and

$\mathbf{x}_1$ are sufficiently close to each other. This fixed-point iteration method is used in most geometry optimization software, but it does not always converge.

2.3.2 Robust Redundant Internal Coordinates

Inspired by the method for selecting internal coordinates in the `Dalton` program,[27] we specify a protocol to define a set of redundant internal coordinates.

Interatomic Distance

Five types of interatomic distances are considered.

1. Regular (covalent) bonds are defined between all pairs of atoms, α and β, whose distance is less or equal than the 1.3 times the sum of their covalent radii:

$$R_{\alpha\beta} \leq 1.3 * (r_{\alpha}^{cov} + r_{\beta}^{cov}) \tag{2.9}$$

2. Hydrogen bonds are designated between hydrogen atoms covalently bonded to one atom with strong electronegativity, $\mathbf{X}$ = N, O, F, P, S, Cl, and located in the peripheral area of another strong electronegative atom, $\mathbf{Y}$ = N, O, F, P, S, Cl. The H-Y distance is required to be less than 0.9 times the sum of their van der Waals radii and the angle between X-H-Y must be larger than 90°.

$$R_{HY} \leq 0.9 * (r_{H}^{vdW} + r_{Y}^{vdW}) \tag{2.10}$$

$$\measuredangle_{XHY} > 90° \tag{2.11}$$

3. Interfragment bonds are included when the system has more than one fragment. Atoms connected by a regular bond are assigned to the same fragment group in the system. Between different fragments, interfragment bonds are added.

   (a) If each fragment is a single atom, then the interatomic distance is included.

   (b) If one fragment is an atom and the other fragment is polyatomic, three coordinates are added, including at least two inter-fragment bonds.

   (c) If both fragments are polyatomic, six internal coordinates, including at least two inter-fragment bonds, will be added.

   In cases (b) and (c), at least two interfragment bonds are necessary to specify the relative positions of the fragments. By default, the two shortest interfragment bonds are selected. Additional interfragment bonds are added when atoms in different fragments are closer than 2 Å or closer than 1.3 times the shortest interfragment distance.

$$R_{XY} \leq \max(1.3 * R_{inter}^{min}, 2\text{Å}) \tag{2.12}$$

   In some cases (e.g., two sheet-like molecules stacked on top of each other), the number of interfragment bonds becomes prohibitively large. To avoid this, the total number of bonds is not allowed to exceed the number of non-hydrogen atoms in the fragments.

4. To describe rotations of functional groups linked by long linear chains, a special linear-chain bond is defined. Specifically, for a molecule with a long-chain structure, the distance between the first and last atoms of the chain is added.

5. Auxiliary bonds are added between any two atoms that are closer than 2.5 times the sum of their covalent radii is counted. Most auxiliary bonds describe Urey-Bradley (1-3) interactions.[28] Unlike the aforementioned bond types, auxiliary bonds are not used when constructing bond angles and dihedrals.

$$R_{XY} \leq 2.5 * (r_x^{cov} + r_y^{cov}) \tag{2.13}$$

Bond Angles

For every atom α that connects two other atoms β, γ by non-auxiliary bonds, the angle $\measuredangle_{\beta\alpha\gamma}$ is counted as an essential internal coordinate.

Conventional Dihedral Angles

Dihedral $\measuredangle_{\alpha\beta\gamma\delta}$ is defined as the angle between two planes, the first defined by the positions of atoms $\alpha\beta\gamma$ and the second defined by the positions of atoms $\beta\gamma\delta$. The dihedral angle can therefore be computed as the angle between the normal vectors of these planes. The normal vectors can be obtained from the cross products of the unit vectors in that plane,

$$\hat{n}_{\alpha\beta\gamma} = \frac{\hat{R}_{\beta\gamma} \times \hat{R}_{\beta\alpha}}{\left\| \hat{R}_{\beta\gamma} \times \hat{R}_{\beta\alpha} \right\|} \tag{2.14}$$

where $\hat{R}_{\beta\alpha}$ denotes the unit vector of bond linking atom β and α,

$$\hat{R}_{\beta\alpha} = \frac{R_\alpha - R_\beta}{\|R_\alpha - R_\beta\|} \tag{2.15}$$

The conventional dihedral angle is computed as angle between the normal vectors from each plane.

$$\measuredangle_{\alpha\beta\gamma\delta} = cos^{-1}\left(\hat{n}_{\alpha\beta\gamma} \cdot \hat{n}_{\beta\gamma\delta}\right) \tag{2.16}$$

If one were to include all possible dihedrals in the system, an enormous number of internal coordinates would be obtained. To reduce the unnecessary redundancy, we first restrict the dihedrals to exclude auxiliary bonds. Then, given a (non-auxiliary) bond $R_{\beta\gamma}$, among all the atoms connected to β, we select α as the one with the most bonded neighbours. Any atoms that are connected to γ are included in the dihedrals $\alpha\beta\gamma*$. Symmetrically, δ is selected to be the most bonded atom among γ's neighbour atoms. All dihedrals $*\beta\gamma\delta$ are appended to the internal coordinates set.

The above description doesn't include every situation. Sometimes, the dihedral is consist of planes $\alpha\beta\gamma$ and $\delta\beta\gamma$ where α and δ are both bonded to atom β. This kind of *improper dihedral* is used to describe puckering motions for the central atom in near-planar structures. For this situation, we will include the dihedral, denoted as $\alpha\beta\gamma\delta$, if the sum of the angles $\measuredangle_{\alpha\beta\gamma}$, $\measuredangle_{\alpha\beta\delta}$, and $\measuredangle_{\gamma\beta\delta}$ is greater than 345°.

Robust Dihedral indicators

There are still numerical issues associated with torsions around bonds for which the $\alpha\beta\gamma$ angle or the $\beta\gamma\delta$ angle is nearly 180 degrees. In such cases, a small change

in the position of the terminal atom can cause an enormous change in the dihedral angle. To circumvent the problem, we developed two new robust dihedral descriptors to replace the traditional dihedral angle in our implementations. Specifically, we use the cosine angle between the two terminal bonds $\alpha\beta$ and $\gamma\delta$ bonds and the volume of the parallelepiped enclosed by $\alpha\beta\gamma\delta$.

$$\hat{R}_{\beta\alpha} \cdot \hat{R}_{\gamma\delta} \tag{2.17}$$

$$\hat{R}_{\beta\gamma} \cdot \left(\hat{R}_{\beta\alpha} \times \hat{R}_{\gamma\delta}\right) \tag{2.18}$$

These robust dihedral descriptors prevent the failure of redundant internal coordinates in the geometry optimization algorithms because, when the position of an atom is changed by a small amount, the corresponding changes in the robust dihedral descriptors is also small. To evaluate the performance of the robust dihedral indicators, a comprehensive test is performed in the next section.

2.3.3 Mapping between Internal coordinates and Cartesian coordinates

Based on the protocol we proposed in the previous section, the number of robust internal coordinates M_{int} will be far greater than 3N - 6. Therefore, a randomly selected value for the internal coordinates is exceedingly unlikely to correspond to a realizable molecular structure. The feasible structures of molecule define a 3N - 6 Manifold inside the M-dimensional space of internal coordinates.

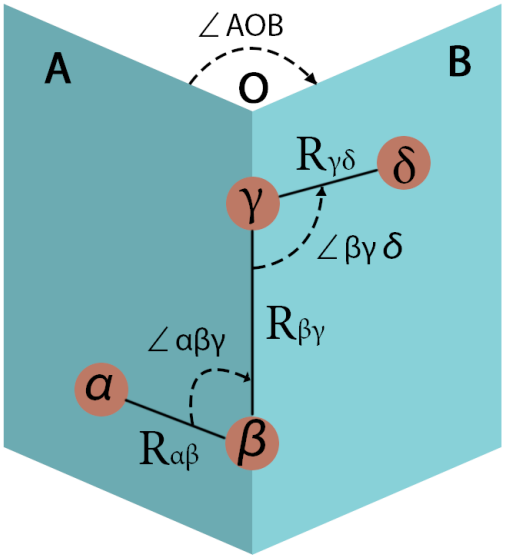


FIGURE 2.1: Illustration of internal coordinates consist of 4 atoms

In order to map every point from the M-dimenstional space onto the $3N-6$-dimensional manifold consistently, a mapping scheme $f : \mathbb{R}^M \mapsto \mathbb{R}^{3N-6}$ is constructed as follow. Given a set of redundant internal coordinates of interest, $\mathbf{q}^{(target)}$, the closest point on the 3N - 6 $\mathbf{q}(\mathbf{x})$ manifold, measured by following cost function, is selected,

$$\mathbf{x}(\mathbf{q}^{(target)}) = \arg\underbrace{\min}_{\mathbf{x}}(\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)})^T\mathbf{W}(\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)}) \tag{2.19}$$

where $\mathbf{W}$ is a weight matrix with only diagonal elements. It is used to weight the importance of each internal coordinates when conducting the manifold mapping. Typically, $\mathbf{W}$ is a identity matrix. However, alternative weightings are sometimes desired. For example, if some internal coordinates are selected to be frozen during an optimization, a large weight forces that internal coordinate to be retained. Similarly, when conventional dihedrals are included in the system, its weight is

adjusted to $\sin^2 \measuredangle_{\alpha\beta\gamma} \sin^2 \measuredangle_{\beta\gamma\delta}$ to mitigate problems with the near-collinear bonds. In the objective function, cosine functions are applied to angles with periodicity and square functions are deployed as the main penalty function for bonds:

$$\text{cost}_{bond} = \Big(q(x) - q^{(target)}\Big)^2 \tag{2.20}$$

$$\text{cost}_{angle} = \Big(\cos(\theta(x)) - \cos\Big(\theta^{(target)}\Big)\Big)^2 \tag{2.21}$$

For the conventional dihedral,

$$\text{cost}_{dihed} = \Big(\cos(\phi(x)) - \cos\Big(\phi^{(target)}\Big)\Big)^2 + \Big(\sin(\phi(x)) - \sin\Big(\phi^{(target)}\Big)\Big)^2 \tag{2.22}$$

The complete objective function to be minimized is thus,

$$\begin{aligned} f(\mathbf{q}) = & \sum_{bonds} \Big(q_{\alpha\beta}(x) - q_{\alpha\beta}^{(target)}\Big)^2 + \sum_{robust} \Big(q_{\alpha\beta\gamma\delta}(x) - q_{\alpha\beta\gamma\delta}^{(target)}\Big)^2 \\ & + \sum_{angle} \Big(\cos(\theta_{\alpha\beta\gamma}) - \cos\Big(\theta_{\alpha\beta\gamma}^{(target)}\Big)\Big)^2 \\ & + \sum_{dihedral} \Big(\cos(\phi(x)) - \cos\Big(\phi^{(target)}\Big)\Big)^2 + \Big(\sin(\phi(x)) - \sin\Big(\phi^{(target)}\Big)\Big)^2 \end{aligned} \tag{2.23}$$

To generate a sensible initial guess structure, a double-ended method is proposed in conjunction with redundant internal coordinates and the manifold projection method. We start with Cartesian coordinates of the reactant and product. Following the protocols introduced in the last section, the redundant internal coordinates are generated for both structures. We make a line segment that interpolates from the reactant to the product. Without further information, the midpoint of the segment is selected as the guess structure for the transition-state optimization,

$$\mathbf{q}^{guess} = (1-p)\mathbf{q}^{reactant} + p\mathbf{q}^{product} \tag{2.24}$$

In general, the guess in the redundant coordinates is not physically realizable, so the manifold projection method is applied to locate the nearest Cartesian representation. We observe that the final structure is, in most cases, an excellent initial guess for transition-state optimization.

$$\mathbf{C}^{guess}(p) = \underbrace{\min}_{x} |\mathbf{q}(x) - [(1-p)\mathbf{q}^{reactant} + p\mathbf{q}^{product}]|_w^2 \tag{2.25}$$

$$\mathbf{x}^{guess}(p) = \arg\underbrace{\min}_{\mathbf{x}} |\mathbf{q}(\mathbf{x}) - [(1-p)\mathbf{q}^{reactant} + p\mathbf{q}^{product}]|_w^2 \tag{2.26}$$

where p is the fractional variable determining the interpolation ratio. When $p = 0$, the guess structure is equivalent to the reactant, while $p = 1$, the guess is the same as the product. The choice of p value indicates the resemblance of the guess to either of the two known structures. Without further information about the reaction mechanism, it is sensible to select $p = 0.5$ as an impartial starting value. The redundant internal coordinates corresponding to $p = 0.5$ are rather far from the manifold of realizable molecular structures. For this strategy, then, it is important that the robust manifold projection technique is being used to find the molecular structure, because the fixed-point iteration method commonly used in the literature is only robust when the redundant internal coordinates are nearly realizable.

2.4 Results and Discussion

2.4.1 Testing Protocol

A set of 32 chemical reactions involving various mechanisms is constructed. It includes proton or hydrogen transfer reactions, pericyclic reactions, Diels-Alder reactions, intramolecular reactions, S_N2 reaction, free radical reactions, etc.. Detailed reaction information is presented in Table 2.1.

All calculations are computed at the Hartree-Fock level with the 6-31+G basis, except for reactions 21 and 25, where the 6-31+G(d,p) basis was used. (For these reactions, exceptionally slow convergence was obtained with the 6-31+G basis set.) We are not asserting that this computational method is accurate, but it defines a potential energy surface with well-defined barriers, and is thus suitable for testing transition-state optimization methods. The reactant and product structures used in the double-ended method are obtained from an intrinsic reaction coordinate(IRC) process from the known, pre-optimized, TS.[29] The transition state optimization update is performed by *GOpt*, a geometry optimization package developed by us along with *Gaussian 16* for energy and gradient evaluation at each iteration.

2.4.2 Overview of result

When constructing the guess structure, the union of the redundant internal coordinates from the reactant and product structures are used. Robust dihedrals and conventional dihedrals are deployed separately to form the respective initial

guesses, as well as in the following optimization iterations. All other parameters are set to be the same. The purpose of this study is to evaluate the efficiency, robustness, and consistency of the two sets of dihedrals, and to establish the reactant/product averaging technique as a reasonable strategy for generating transition-state guesses.

2.4.3 Result and Discussion

Table2.1 presents our computational data. First, we establish that the mean between the reactant and product structures works as an initial guess for the transition-state optimization. As a first indication, we evaluate whether the transition-state guess has one, and only one, negative eigenvalue. Both definitions for the dihedral coordinate work well here: the average number of negative eigenvalues are 0.97 and 1.03 respectively, within the same error range of ± 0.03 (see Table 2.2). This isn't a significant difference.

When used in optimization, the advantage of the robust dihedrals becomes clear. Transition-state optimizations with conventional dihedrals fail for reactions 24, 25, 26, 27, and 30, all of which have ill-defined conventional dihedrals at the initial geometry. When both methods work, usually the same number of iterations (13 cases) are required, with conventional dihedrals converging more rapidly in 9 cases and robust dihedrals converging more rapidly in 5 cases. The additional iterations required by the robust dihedrals can be justified by the fact that each conventional dihedral is replaced by two robust dihedrals: as the dimensionality (i.e., the number of internal coordinates) is higher when robust dihedrals are used, it is sensible for the number of iterations to increase also. Note, however, that in

difficult cases (e.g., reaction 21 before we added polarization functions to the basis set), the robust dihedrals work significantly better (36 vs. 71 iterations).

For reactions where the conventional dihedral fails, you can eliminate the problematic coordinate. Table 2.3 shows the number of iterations required if this is done. This is not suitable as a general-purpose strategy because by eliminating the defective collinear dihedral, the ability to describe non-collinear structures is removed and the optimization is constrained to collinear structures. This works for these reactions (where the true transition-state has a collinear dihedral, but does not work in general. For these reactions, the constrained transition-state optimization (with omitted dihedrals) tends to converge a bit more rapidly when conventional dihedrals are used.

2.5 Conclusion

In this chapter, we proposed a set of protocols to generate the redundant internal coordinates. First, five types of bonds are built, including covalent bonds, hydrogen bonds, interfragment bonds, long-distance bonds for linear chains, and auxiliary (Urey-Bradley) bonds. Second, the cosine of all the bond angles between all pairs of non-auxiliary bonds are added; the cosine is used because it naturally includes the periodicity of the bond angle. Adding dihedral angles for all (nonauxiliary) bonds leads to a prohibitively large number of dihedral coordinates. In GOpt, for every non-auxiliary bond $\beta\gamma$, the atoms α is selected as the atom bonded to β with the most bonds connected to other atoms. Then all possible $\alpha\beta\gamma$ are included. Symmetrically, the dihedrals defined by $*\beta\gamma\delta$ are also added to the system where δ is set to be the atom with the most bonded atom. we also

introduced a new set of robust dihedral indicators to deal with the troublesome linear system in geometry optimization. Based on the tests, the robust dihedral has shown promising results compared to the traditional dihedrals, which fail for 5 of the 32 reactions considered, and perform poorly in one other case. However, for cases where conventional dihedrals work well, using robust dihedrals tends to increase the computational cost slightly, by about 10%.

The other issue associated with redundant internal coordinates is that converting Cartesian to/from redundant internal coordinates is not unique. Notably, most values of the redundant internal coordinates do not correspond to any permissible molecular structure. We avoid this using a manifold projection method to find the Cartesian structure whose redundant internal representation is as close as possible to the specified redundant internal coordinates.

As an initial application of the proposed techniques for generating redundant internal coordinates, we generated initial guesses for transition-state optimizations from the average of reactant and product structures in internal coordinates. This procedure provides excellent initial guesses for the transition-state geometry for the reactions considered: the Hessian of the initial-guess structure has exactly one negative eigenvalue in 28 of 32 reactions and the optimization is 100% successful and requires on average less than 8 iterations when robust dihedrals are used. When conventional dihedrals are used, the number of iterations is about 10% less, but the optimizer fails for 5 reactions.

In conclusion, we recommend the average of the reactant and product structures in internal coordinates as a suitable initial guess for geometry optimization. We also recommend using our new robust dihedral descriptor in most cases, as

avoiding the human intervention required to remedy failed optimizations is generally more salient than the slight increase in computer time associated with the robust dihedral descriptors.

TABLE 2.1: The number of iterations and negative eigenvalues for generated guess structures. An entry of – indicates that the calculation failed due to ill-defined dihedrals.

index	Reaction	Num. of neg. eigval.		Num. of opt. iter.	
		Conv. Dihed	Spe. Dihed	Conv. Dihed	Spe. Dihed
1	$C_4H_6 + C_2H_4$	1	1	6	6
2	$C_5H_6 + C_2H4$	1	1	4	4
3	$C_4H_4Si + C_2H_4$	1	1	4	4
4	$C_6H_8O + C_2H_4$	1	1	5	5
5	$C_4H_5N + C_2H_4$	1	1	5	5
6	C_4H_6	1	1	4	4
7	C_6H_8	1	1	3	3
8	C_8H_8	1	1	4	4
9	$C_{12}H_{18}$	1	1	6	6
10	$N_2O + C_2H_4$	1	1	5	5
11	$N_3 + C_2H_4$	1	1	12	11
12	$N2C_2 + C_2H_4$	1	1	5	6
13	$ONC + C_2H_4$	1	1	8	11
14	$N_2C_H + C_2H_4$	1	1	6	5
15	$HF + C_2H_4$	1	1	6	10
16	$C_2H_4 + H_2$	1	1	9	10
17	$HCN + H_2$	1	1	6	6
18	$HNC + H_2$	1	1	11	7
19	$C_2H_6 + SiH_2$	1	1	5	5
20	$HONS$	1	1	5	5
21	$HNCS$	1	1	6	6
22	$C_3H_4O_2$	1	1	6	7
23	C_6H_8	1	1	6	8
24	$CH_3F + Cl^-$	1	1	-	10
25	$CH_3Cl + F^-$	1	2	-	8
26	$CH_3F + F^-$	0	1	-	11
27	$CH_3OH + F^-$	1	1	-	6
28	$CH_3OH + \cdot OOH$	1	1	10	11
29	$CH_3OH + \cdot CH_3$	2	1	9	11
30	$HF + \cdot CH_3$	1	2	-	7
31	$N_2O + \cdot H$	1	1	9	13
32	$H_2O + \cdot CH_3$	0	0	14	16

TABLE 2.2: Average performance of conventional dihedrals and robust dihedrals

Stats \ Methods	Conv. Dihed	Special Dihed
Average Neg. Eigenvalues	0.97	1.03
Average iteration	7.15	7.38
Convergence rate	84.4%	100%
Median iteration	6	6

TABLE 2.3: Optimization iteration needed after removing ill-defined conventional dihedrals

Index	Reaction	Conventional	Special
24	$CH_3F + Cl^-$	6	10
25	$CH_3Cl + F^-$	5	8
26	$CH_3F + F^-$	8	11
27	$CH_3OH + F^-$	7	6
30	$HF + \cdot CH_3$	20	25

2.6 Reference

[1] J. Baker. Techniques for geometry optimization: A comparison of Cartesian and natural internal coordinates. *Journal of computational chemistry* 14(9) (1993), 1085–1100.

[2] J. Baker and F. Chan. The location of transition states: A comparison of Cartesian, Z-matrix, and natural internal coordinates. *Journal of computational chemistry* 17(7) (1996), 888–904.

[3] H. B. Schlegel. A comparison of geometry optimization with internal, cartesian, and mixed coordinates. *International Journal of Quantum Chemistry* 44(S26) (1992), 243–252.

[4] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, et al. *Gaussian 16*. 2016.

[5] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2(4) (2012), 556–565.

[6] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, K. Boguslawski, M. Chan, Y. Zhao, T. Kim, S. Vandenbrande, D. Yang, C. Gonzlez-Espinoza, et al. Horton 2.0. 0, 2015. *Accessed: January* 25 (2018).

[7] H. B. Schlegel. Geometry optimization. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1(5) (2011), 790–809.

[8] W. Hehre and L. Radom. Schleyer, P. vR; Pople, JA. *Ab initio molecular orbital theory* (1986).

[9] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* 17(1) (1996), 49–56.

[10] M. von Arnim and R. Ahlrichs. Geometry optimization in generalized natural internal coordinates. *The Journal of chemical physics* 111(20) (1999), 9183–9190.

[11] J. Baker, A. Kessi, and B. Delley. The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of chemical physics* 105(1) (1996), 192–212.

[12] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1980.

[13] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *The Journal of chemical physics* 96(4) (1992), 2856–2860.

[14] G. Fogarasi, X. Zhou, P. W. Taylor, and P. Pulay. The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces. *Journal of the American Chemical Society* 114(21) (1992), 8191–8201.

[15] J. Baker and P. Pulay. Geometry optimization of atomic microclusters using inverse-power distance coordinates. *The Journal of chemical physics* 105(24) (1996), 11100–11107.

[16] J. Baker, D. Kinghorn, and P. Pulay. Geometry optimization in delocalized internal coordinates: An efficient quadratically scaling algorithm for large molecules. *The Journal of chemical physics* 110(11) (1999), 4986–4991.

[17] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *The Journal of chemical physics* 96(4) (1992), 2856–2860.

[18] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* 17(1) (1996), 49–56.

[19] J. Baker and P. Pulay. Efficient geometry optimization of molecular clusters. *Journal of Computational Chemistry* 21(1) (2000), 69–76.

[20] Ö. Farkas and H. B. Schlegel. Geometry optimization methods for modeling large molecules. *Journal of Molecular Structure: THEOCHEM* 666 (2003), 31–39.

[21] Ö. Farkas and H. B. Schlegel. Methods for geometry optimization of large molecules. I. An O (N 2) algorithm for solving systems of linear equations for the transformation of coordinates and forces. *The Journal of chemical physics* 109(17) (1998), 7100–7104.

[22] F. Eckert, P. Pulay, and H.-J. Werner. Ab initio geometry optimization for large molecules. *Journal of computational chemistry* 18(12) (1997), 1473–1483.

[23] S. R. Billeter, A. J. Turner, and W. Thiel. Linear scaling geometry optimisation and transition state search in hybrid delocalised internal coordinates. *Physical Chemistry Chemical Physics* 2(10) (2000), 2177–2186.

[24] B. Paizs, J. Baker, S. Suhai, and P. Pulay. Geometry optimization of large biomolecules in redundant internal coordinates. *The Journal of Chemical Physics* 113(16) (2000), 6566–6572.

[25] K. N. Kudin, G. E. Scuseria, and H. B. Schlegel. A redundant internal coordinate algorithm for optimization of periodic systems. *The Journal of Chemical Physics* 114(7) (2001), 2919–2923.

[26] K. Doll, R. Dovesi, and R. Orlando. Analytical Hartree–Fock gradients with respect to the cell parameter: systems periodic in one and two dimensions. *Theoretical Chemistry Accounts* 115(5) (2006), 354.

[27] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, et al. The D alton quantum chemistry program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4(3) (2014), 269–284.

[28] H. C. Urey and C. A. Bradley Jr. The vibrations of pentatonic tetrahedral molecules. *Physical review* 38(11) (1931), 1969.

[29] K. Fukui. The path of chemical reactions-the IRC approach. *Accounts of chemical research* 14(12) (1981), 363–368.

Chapter 3

A Robust Algorithm for geometry optimization and transition state search with Reduced Internal Coordinates

3.1 Abstract

A robust algorithm for geometry optimization in redundant internal coordinates, with robust dihedral descriptors that prevent failures due to collinear bonds, is proposed. One of the salient advantages of this method is the separation of a subset of key internal coordinates, corresponding to the bond-breaking and bond-forming processes. Quasi-Newton updates are used at each iteration, but when the elements of the Hessian matrix associated with key coordinates are judged to be inaccurate, they are approximated with finite differences of gradients. Optimization steps are controlled using a trust radius, and a Hessian modification scheme is used to ensure that the Hessian matrix has a chemically sensible eigenstructure.

A set of 32 reactions consisting of various reaction types are used to assess the performance of our algorithm and compare it to the popular *Berny* algorithm. All the energy and gradient evaluation are computed in Gaussian[1]. Compared with *Berny* algorithm, our new approach exhibits more robust and consistent results. The new algorithm and all the advanced features will be accessible in our coming open-source Python package *GOpt*.

3.2 Introduction

To effectively model a chemical reaction, one needs the structures of the reactant, product, and the transition state (TS) connecting them. Mathematically, these structures correspond to stationary points on the potential energy surface(PES) where the reactant and product are local minima and the TS is the first-order saddle point. It's relatively easy to locate the reactant and product as, given a sufficiently accurate initial guess structure, one may follow the gradient downhill to the minimum. However, finding the TS is a more challenging task, both because it is more difficult to form an initial guess for the TS and because it is a saddle point, so it is a maximum along one direction and a minimum in the others.

Many sophisticated geometry optimization methods have been developed. There are three popular types of transition-state optimization methods. First, one can perform a straightforward optimization starting from a guess structure. This method is effective when the guess structure is relatively close to the real transition state. The performance is highly dependent on the quality of the guess structure and the researcher's chemical intuition. The other two methods can be divided

into two categories: single-ended methods and double-ended methods according to whether there is one, or two, starting point structures.

The most popular optimization method is the *Berny* algorithm.[2] It starts from a provided initial guess structure and an initial Hessian matrix. The Berny algorithm displaces the guess along the direction with the negative eigenvalue. The Berny algorithm is regarded as one of the most effective algorithms for geometry optimization. It is the default option in Gaussian software for geometry optimization. Like all methods based on a single initial structure, the Berny algorithm is sensitive to the initial structure chosen, and fails to perform robustly when the initial Hessian does not have one, and only one, negative eigenvalue.

Single-ended methods normally start from the reactant or the product. An ascending direction is selected to drive the energy uphill towards the transition state. The simplest method is the coordinate driving.[3] One coordinate is selected as the dominant coordinate to propel the reaction process. More advanced methods such as least-ascent,[4–7] and dimer-methods[8–11] are introduced to utilize the eigenvector information. Gradient-extremal-method[12–14] and reduced-gradient-following[15–17] are also effective methods in generating pathways passing through stationary points. These methods do not give the lowest-energy-path but they potentially discover more, and different, transition states.

Two-ended methods, on the other hand, do not directly rely on the initial guess structure. These methods start with the input reactant and product coordinates. Many families of methods, such as the nudged elastic band (NEB) methods,[18–26] string methods (SM),[27–32] and growing string methods (GSM),[33–36] have been developed. Among these methods, Synchronous Transit-Guided Quasi-Newton

(STQN) is one of the most renowned. STQN uses a linear synchronous transit(LST)[37] or quadratic synchronous transit(QST)[38] method to connect the two end points. A minimum-energy path is obtained on the PES. The interpolated highest energy point is taken as the initial guess for TS optimization. It's a relatively expensive approach since multiple energy and gradient evaluations are needed.

Though continual progress has been made for various computational methods, TS optimization is still challenging. Optimization Failures frequently occur. Among the most common failures are (1) failures due to the underlying quantum chemistry software, typically because converging the self-consistent field (SCF) calculation becomes problematic, (2) unphysical structures obtained during the optimization process, and (3) inaccurate Hessian matrix approximations leading to poor convergence or convergence to an undesired structure. We will address the latter two of these issues in this work.

Many factors contribute to the success of a TS optimization, such as the choice of coordinates, the selection of initial Hessian matrix, the Newton or Quasi-Newton update method, and the step-size control strategy..[39–41] Motivated by the problems of existing methods, we herein propose a robust and efficient algorithm, the *GOpt* algorithm.

In *GOpt*, we adapted the redundant internal coordinates from the work of Pulay.[42] We specify a set of redundant internal coordinates with bonds, angles, and robust dihedrals to determine the molecular structure. The redundancy of internal coordinates is eliminated by forming a set of $3N - 6$ delocalized coordinates through linear combination. The method implemented in *GOpt* is similar to the one proposed by Baker.[43, 44]

3.3 Methodology

3.3.1 Overview

Geometry optimization is difficult because the number of stationary points grows exponentially as the number of atoms increases. This is especially problematic for transition states, where a specific transition state connecting the reactant and product structure is desired, and not another, quite possibly nearby, transition state associated with a different chemical transformation or conformation change. At a mathematical level, then, geometry optimization is nearly intractable. Yet chemists are frequently able, through intuition and experience, to suggest plausible molecular structures for reactants, products, and transition states.

To build a mathematical formulation for these chemical insights, we note that chemists' intuition is guided by the realization that during a chemical reaction, typically only a few key internal coordinates change significantly. These coordinates are typically interatomic distances associated with the formation and fracture of chemical bonds and/or the opening or closing of bond angles. These key chemical coordinates define a reduced-dimensionality potential energy surface (all other coordinates are minimized over or thermally-averaged). A system with M internal coordinates can then be characterized, mechanistically, with many fewer key coordinates; these key coordinates are commonly called the reduced coordinates, because they can be used to parameterize a reduced-dimensionality potential energy surface. The most accurate way to effectively identify the key internal coordinates is to allow the researcher running the software to specify them explicitly. However, for large datasets, this may be impractical, and then key internal coordinates can be identified based on the changes between the structure of the reactants and the

products.

To use this intuition in a practical geometry optimization method, we treat the key internal coordinates and the non-key internal coordinates separately. After selected the K key internal coordinates, the remaining non-key coordinates are determined. The non-key coordinates are then reduced to form a nonredundant set of 3N – 6 – K coordinates, all of which are linear combinations of the original redundant internal coordinate set. Combining the key internal basis and the non-key internal basis, a reduced-internal transformation matrix, $\mathbf{V}$ is obtained.

During the optimization process, we map the molecular structure from Cartesian coordinates to the redundant internal coordinates, then to reduced internal coordinates. Using the $\mathbf{V}$ matrix, the conversion of the gradient and Hessian to nonredundant reduced+nonreduced coordinates is straightforward, and an optimization step can be determined. Because determining the Hessian is expensive, but having accurate values for the Hessian is most important for the block associated with key coordinates, the key-coordinate-Hessian is approximated with a finite-difference approximation, and the eigenstructure of the key-coordinate and the non-key-coordinate blocks of the Hessian are forced to have appropriate eigenstructure.

With these revisions, a quasi-Newton algorithm for both geometry minimization and transition-state finding becomes straightforward. The cost is superficially more than a typical quasi-Newton method because additional gradient calculations are needed for the finite-difference updates of the key-coordinate portions of the Hessian, but these updates are infrequently required and relatively affordable, and their impact on increasing the convergence rate is important.

3.3.2 Selection of redundant internal coordinates

The details of selecting redundant internal coordinates in **GOpt** is fully described in Chapter 2. Here, we briefly recap its salient features. Interatomic bonds, bond angels, and dihedrals are used to fully describe the molecular structure. Five types of bonds are built including covalent bonds, hydrogen bonds, interfragment bonds, long-distance bonds for linear chains, and auxiliary bonds. All these bonds are measured in atomic unit. Bond angles are measured between any two non-auxiliary bonds. The value is measure in cosine function rather to provide more robust performance for nearly collinear structure and to directly include the periodicity of the bond angle. One of the major improvements is the selection of dihedrals compared to normal redundant internal coordinate schemes. In **GOpt**, for every non-auxiliary bond $\beta\gamma$, the atoms α is selected as the atom bonded to β with the most bond connections to other atoms. Then all possible $\alpha\beta\gamma*$ dihedrals are included. Symmetrically, the dihedrals defined by $*\beta\gamma\delta$ are also added to the system where δ is set to be the atom with the most bonds. Conventional dihedral angles fail when three atoms in the dihedral are located near one line. A small change in the Cartesian coordinates of these atoms may result in an enormous shift in the dihedral angle. To circumvent the problem, we proposed the robust dihedral indicators as substitutes. They are defined as,

$$\hat{R}_{\beta\alpha} \cdot \hat{R}_{\gamma\delta} \tag{3.1}$$

$$\hat{R}_{\beta\gamma} \cdot (\hat{R}_{\beta\alpha} \times \hat{R}_{\gamma\delta}) \tag{3.2}$$

where

$$\hat{R}_{\beta\alpha} = \frac{R_\alpha - R_\beta}{|R_\alpha - R_\beta|} \tag{3.3}$$

is the unit vector along the direction of $\beta\alpha$ bond. It's worth noting the geometrical meaning of these indicators. Eqs 3.1 represents the cosine angle of the two ending bonds included in the dihedral. Eqs 3.2 computes the volume of the parallelepiped enclosed by $\alpha\beta\gamma\delta$. These indicators can effectively prevent the collapse of redundant internal coordinates in the geometry optimization process. Unlike conventional dihedral angles, when a small perturbation is imposed on one atom, the corresponding changes in the robust dihedral descriptors are small as well, ensuring the consistency of the coordinate transformation.

3.3.3 Coordinate transformations

Cartesian coordinates are the direct and straightforward representation for the positions of the atoms in a molecular structure, but they are neither numerically efficient nor chemically intuitive. Conversely, internal coordinates are effective for geometry optimization and chemically intuitive, but they are less facile for computing energy and its derivatives in quantum chemistry software. A robust transformation method for interconverting Cartesian and redundant internal coordinates is therefore essential. Because the transformation from Cartesian to internal is not linear, the best approximation for the transformation is by Wilson **B** matrix, which is the Jacobian matrix of the transformation from the $3N_{atoms}$ Cartesian coordinates to the M_{int} internal coordinates, with entries:

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \tag{3.4}$$

These elements can be obtained from trigonometry. Usually, the **B** matrix is rectangular because there are usually far more internal coordinates than the $3N$ Cartesian coordinates, where N denotes the number of atoms. Since only $3N-6(5)$ degrees of freedom are needed to fully specify a molecule's structure, the **B** matrix is also singular, with only $3N-6(5)$ nonzero singular values. The corresponding nonsingular vectors are called the delocalized internal coordinates. By rearranging eqn. 3.4, we obtained the matrix format,

$$\mathbf{B}\delta x = \delta q \tag{3.5}$$

This equation expresses the change of internal coordinates induced by a small change in Cartesian coordinates. Because **B** matrix is rectangular, there isn't a unique inverse. We use the Moore-Penrose pseudoinverse of **B** matrix for the inverse transformation,

$$\mathbf{B}^{+}\delta q = \delta x \tag{3.6}$$

It's worth noting that not every change in internal coordinates is realizable. Consider, for example, three atoms forming a triangular structure, characterized by three bond lengths and three bond angles. You cannot change any bond angle in isolation, because any change that leads to the sum of three bond angles other than 180° is not physically realizable. To convert these nonphysical internal coordinates, we project them onto a realizable space spanned by **B** matrix with minimum error distance,

$$\delta\tilde{q} = \hat{\mathbf{P}}\delta q \tag{3.7}$$

where

$$\hat{\mathbf{P}} = \mathbf{B}\mathbf{B}^{+} \tag{3.8}$$

3.3.4 Transformation between Internal and Cartesian Coordinates

Here we use x, g_x, H_x and q, g_q, H_q to denote the energy, gradient, and Hessian in Cartesian and internal coordinates, respectively. Given the Wilson **B** matrix,

$$g_x = B^T g_q \tag{3.9}$$

$$H_x = B^T H_q B + K \tag{3.10}$$

where K is

$$K_{jk} = \sum_i [g_q]_i b'_{ijk} \tag{3.11}$$

where b'_{ijk} is the derivative of the elements in the **B** matrix,

$$b'_{ijk} \equiv \frac{\partial^2 q_i}{\partial x_j \partial x_k} = \frac{\partial b_{ij}}{\partial x_k} \tag{3.12}$$

Similarly, the energy derivatives in internal coordinates can be computed through,

$$g_q = (B^T)^{+} g_x \tag{3.13}$$

$$H_q = (B^T)^{+}(H_x - K)B^{+} \tag{3.14}$$

Equation 3.13 is essential because most quantum chemistry software compute energy and its derivatives in Cartesian coordinates.

Eqs. 3.5 and 3.6 are only valid for interconverting infinitesimal changes in Cartesian and intenral coordinates. During the process of a geometry optimization in internal coordinates, larger changes are made. Specifically, starting from a realizable molecular structure with Cartesian coordinates $\mathbf{x}^{(k)}$ and its corresponding internal coordinates $\mathbf{q}(\mathbf{x})^{(\mathbf{k})}$, after a non-infinitesimal step in internal coordinates $\mathbf{s}$, the new internal coordinates would be

$$\mathbf{q}^{(k+1)} = \mathbf{q}(\mathbf{x})^{(k)} + \mathbf{s} \tag{3.15}$$

This new configuration in internal coordinates space will rarely be located on the $3N-6$ manifold realizable by Cartesian coordinates. Specifically, the Cartesian representation,

$$x = x^{(k)} + B^{+}s^{(k)} \tag{3.16}$$

will not be the exact counterpart of the target internal coordinates. To maintain a consistent transformation between the two system, we choose x_{k+1} as the closest point on the $3N-6$ manifold to the target internal value, q_{k+1},

$$\mathbf{x}^{(k+1)} \equiv \arg \underbrace{\min}_{mathbfx} |\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(k+1)}|^2 \tag{3.17}$$

More details about the manifold projection strategy can be found in 2.

3.3.5 Select key internal coordinates in optimization

For most chemical reactions, the active reaction sites can be characterized by a few key internal coordinates. These coordinates are usually related to the bond formation/breaking and bond angle opening/closing. Motivated by this idea, a

reduced dimensional potential energy surface can be generated to describe the energy changing during the reaction. Given a system consisting of M_{int} internal coordinates, the crucial information related to the reaction mechanism is located in the K-dimensional key internal coordinate space. The PES around the reaction site is chiefly dependent on the changes of these key internal coordinates.

It is best for researchers to specify the key coordinates based on the chemical process(es) of interest. However, if no user input is provided, it is also possible to automatically select the key internal coordinates. A set of protocols are conducted to select proper key internal coordinates based on the difference between the reactant and product structure. Initially, a set of internal coordinates is specified as the union of the internal coordinates from the reactant, product, and the TS guess structure.

A given coordinate is selected as a key internal coordinate if:

- An inter-atomic distance changes more than half the sum of the composing covalent radii
- An angle changes by at least 30°

No intuitive and reliable criterion can be easily generalized to describe the behaviors of dihedrals, so they are not included in the auto-selection scheme. Note, however, that in some reactions (e.g., nucleophilic elimination reactions for cyclic hydrocarbons) it is clearly appropriate to include a dihedral among the key coordinates.

3.3.6 Construct delocalized reduced internal coordinates

We developed a geometry optimization algorithm based on reduced internal coordinates. Our strategy is similar to the idea proposed to Baker.[43, 44] The geometry of a chemical system can be specified by $3N-6$ independent coordinates. To effectively represent the reaction process without introducing extra redundancy, we treat the key internal coordinates and the non-key internal coordinates separately. After selected the K key internal coordinates, the non-key coordinates are constructed as a linear combination of all the other redundant internal coordinates.

$$v^{(j)} = \sum_{i=1}^{M_{int}} v_i^{(j)} q_i \tag{3.18}$$

The **V** matrix to transform reduced delocalized internal coordinates to redundant internal coordinates. is denoted as

$$V = \begin{bmatrix} v_1^{(1)} & v_1^{(2)} & \dots & v_1^{(3N_{atoms}-6)} \\ v_2^{(1)} & v_2^{(2)} & \dots & v_2^{(3N_{atoms}-6)} \\ \vdots & \vdots & & \vdots \\ v_{M_{int}}^{(1)} & v_{M_{int}}^{(2)} & \dots & v_{M_{int}}^{(3N_{atoms}-6)} \end{bmatrix} \tag{3.19}$$

To construct the **V** matrix, $3N_{atoms}-6$ non-zero singular vectors are selected from the Wilson **B** matrix through singular value decomposition. These vectors are denoted as

$$\mathbf{a^{(i)}} = \begin{bmatrix} a_1^{(i)} & a_2^{(i)} & \dots & a_{M_{int}}^{(i)} \end{bmatrix} \quad i = 1, 2, \dots, 3N_{atoms} - 6 \tag{3.20}$$

These vectors spans the same space as Baker's delocalized internal coordinates since the singular vectors are eigenvectors of matrix $\mathbf{BB^T}$.

$$\mathbf{B}_{M_{int}\times 3N} = \mathbf{U}_{M_{int}\times M_{int}} \cdot \mathbf{\Sigma}_{M_{int}\times 3N} \cdot \mathbf{V^*}_{3N\times 3N} \tag{3.21}$$

$$\mathbf{BB^T} = \mathbf{U}_{M_{int}\times M_{int}} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^T_{M_{int}\times M_{int}} \tag{3.22}$$

To separate the changes in key and non-key internal coordinates, we impose a small changes on each key internal coordinate successively without changing other internal coordinates. This operation usually results in an unphysical structure. We then project the unrealizable structure onto the physical space through,

$$\mathbf{c}^{(j)} = \mathbf{P}\hat{\mathbf{e}} = \mathbf{BB}^{+}\hat{\mathbf{e}}^{(j)} \tag{3.23}$$

where $\hat{\mathbf{e}}^{(j)}$ is a unit vector with 1 in the j^{th} position but 0 anywhere else. The vectors $\mathbf{c}^{(j)}$ are not orthonormal. To form a well-behaved basis, we orthogonalize them through,

$$\mathbf{CC}^T = \mathbf{V\Lambda V}^T \tag{3.24}$$

where $\mathbf{C} = [\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(k)}]$. We pick the eigenvectors $\mathbf{v}_i$ from $\mathbf{V}$ with non-zero eigenvalues λ_i.

$$\mathbf{V}_{key} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_k \end{bmatrix} \tag{3.25}$$

where k is the number of independent reduced coordinates. Normally, k should be equal to the number of key internal coordinates selected. If not, it indicates that

there is redundancy in the key internal space. We then should reduced the dimensionality of key-internal space by only including the independent eigenvectors.

To construct the full $\mathbf{V}$ space for the non-key internal coordinates, we need to project out the key-internal space,

$$\mathbf{d}^{(j)} \equiv (\mathbf{I} - \mathbf{P}_{key})\mathbf{a}^{(j)} \tag{3.26}$$

$$= \mathbf{a}^{(j)} - \mathbf{V}_{key}\mathbf{V}_{key}^{T}\mathbf{a}^{(j)} \tag{3.27}$$

where $\mathbf{P}_{key}$ is the projection operator onto the key-coordinate space. After projecting out the key internal space, the leftover vectors $\mathbf{D} = [\mathbf{d}^{(1)}, \mathbf{b}^{(2)}, \ldots, \mathbf{d}^{(3N-6-k)}]$ are not orthonormal either. The same procedures are conducted for the non-key space to generate a orthonormal basis:

$$\mathbf{D}\mathbf{D}^{T} = \mathbf{V}'\mathbf{\Lambda}\mathbf{V}'^{T} \tag{3.28}$$

There are $3N - 6 - k$ eigenvectors with non-zero eigenvalues in $\mathbf{V}'$.

$$V_{nonkey} = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \ldots & \mathbf{v}'_{3N-6-k} \end{bmatrix} \tag{3.29}$$

Combining the key- and non-key spaces,

$$\mathbf{V} = \begin{bmatrix} V_{key} & V_{nonkey} \end{bmatrix} \tag{3.30}$$

we obtain the complete $\mathbf{V}$ matrix for transforming redundant internal coordinates to the delocalized internal space.$\mathbf{V}$ matrix needs to be constructed at each iteration of the optimization process. The choice of the basis for non-key internal space

is almost arbitrary, so the $\mathbf{V}$ determined may vary dramatically between each iteration. To keep the minimal variance between each $\mathbf{V}$ matrix, a rotation is applied for maximum overlap,

$$\mathbf{V}_{new} = \mathbf{Q}\mathbf{V}_{old} \tag{3.31}$$

$$\mathbf{Q} = \mathbf{V}_{new}\mathbf{V}_{old}^{T} \tag{3.32}$$

$$= \mathbf{U}\mathbf{\Sigma}\mathbf{W}^{T} \tag{3.33}$$

The optimal orthonormal rotation matrix is

$$\mathbf{Q}^{(max)} = \mathbf{U}\mathbf{W}^{T} \tag{3.34}$$

and the maximally aligned new basis is

$$\mathbf{V}_{new} = \mathbf{Q}^{(max)}\mathbf{V}_{old} \tag{3.35}$$

During the optimization process, we map the molecular structure from Cartesian coordinates to redundant internal coordinates, then to key+non-key delocalized internal coordinates. With the $\mathbf{V}$ matrix, the conversion of gradient, Hessian,

and optimization step is straightforward,

$$\mathbf{g}_v = \mathbf{V}^T \mathbf{g}_q \tag{3.36}$$

$$\mathbf{g}_q = \mathbf{V}\mathbf{g}_v \tag{3.37}$$

$$\mathbf{H}_v = \mathbf{V}^T \mathbf{H}_q \mathbf{V} \tag{3.38}$$

$$\mathbf{H}_q = \mathbf{V}\mathbf{H}_v \mathbf{V}^T \tag{3.39}$$

$$\Delta\mathbf{v} = \mathbf{V}^T \Delta\mathbf{q} \tag{3.40}$$

$$\Delta\mathbf{q} = \mathbf{V}\Delta\mathbf{v} \tag{3.41}$$

$$\tag{3.42}$$

3.3.7 The secant condition in reduced coordinates

In the *GOpt* algorithm, the Hessian matrix of energy is updated through Quasi-Newton methods where the value is computed based on the difference of gradients between iterations. Since the main optimization is performed in $\mathbf{V}$ space, the secant condition for $\mathbf{H}_v$ is selected to be

$$\mathbf{H}_v^{old}\delta\mathbf{v} \approx \delta\mathbf{g}_v - (\mathbf{V}^{old})^T\Big((\mathbf{B}^{old})^T\Big)^+\Big((\mathbf{B}^{old})^T\delta\mathbf{V}\mathbf{g}_v^{old} + (\delta\mathbf{B})^T\mathbf{g}_q^{old}\Big) \tag{3.43}$$

3.3.8 Quasi-Newton Updates

In *GOpt*, the Hessian matrix is updated by quasi-Newton methods. These methods approximate the Hessian with the gradient changes from the previous iterations. The four methods we introduce here only require the gradient difference between the latest two structures.

$$\mathbf{s}_v^{(k)} = \mathbf{v}^{(k+1)} - \mathbf{v}^{(k)} \tag{3.44}$$

and the secant condition

$$\mathbf{y}^{(k)} = (\mathbf{g}_v^{(k+1)} - \mathbf{g}_v^{(k)}) - (\mathbf{V}^{(k)})^T \Big((\mathbf{B}^{(k)})^T \Big)^+ \Big((\mathbf{B}^{(k)})^T (\mathbf{V}^{(k+1)} - \mathbf{V}^{(k)}) \mathbf{g}_v^{(k)} + (\mathbf{B}^{(k+1)} - \mathbf{B}^{(k)})^T \mathbf{g}_q^{(k)} \Big) \tag{3.45}$$

The four major methods we consider in *GOpt* are the symmetric-rank-one update(SR1)

$$\mathbf{H_v}^{(k+1)} = \begin{cases} \mathbf{H}_v^{(k)} & \frac{\left\| (\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}) \cdot \mathbf{s}_v^{(k)} \right\|}{\left\| \mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)} \right\| \cdot \left\| \mathbf{s}_v^{(k)} \right\|}) \leq 1e^{-9} \\ \mathbf{H}_v^{(k)} + \frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}) \cdot \mathbf{s}_v^{(k)}} & Otherwise \end{cases} \tag{3.46}$$

the Powell-symmetric-Broyden update (PSB)

$$\begin{aligned} \mathbf{H_v}^{(k+1)} = \mathbf{H}_v^{(k)} + & \frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})(\mathbf{s}_v^{(k)})^T + \mathbf{s}_v^{(k)} (\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^T \mathbf{s}_v^{(k)}} \\ & - \left(\frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T (\mathbf{s}_v^{(k)})}{(\mathbf{s}_v^{(k)})^T \mathbf{s}_v^{(k)}} \right) \mathbf{s}_v^{(k)} (\mathbf{s}_v^{(k)})^T \end{aligned} \tag{3.47}$$

the Broyden-Fletcher-Goldfarb-Shanno update (BFGS)

$$\mathbf{H_v}^{(k+1)} = \mathbf{H}_v^{(k)} + \frac{\mathbf{y}_v^{(k)} (\mathbf{y}_v^{(k)})^T}{(\mathbf{y}_v^{(k)})^T s_v^{(k)}} - \frac{(\mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})(\mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^t \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}} \tag{3.48}$$

and Bofill's 1994 update (Bofill), which is a mixed method of the SR1 and PSB updates

$$\mathbf{H}_{Bofill}^{(k+1)} = (1-\psi)\mathbf{H}_{SR1}^{(k+1)} + \psi\mathbf{H}_{PSB}^{(k+1} \tag{3.49}$$

$$\psi = 1 - \frac{|\mathbf{s}_v^{(k)} \cdot \left(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}\right)|^2}{|\mathbf{s}_v^{(k)}|^2|\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2} \tag{3.50}$$

$$= \frac{|\mathbf{s}_v^{(k)} \times \left(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}\right)|^2}{|\mathbf{s}_v^{(k)}|^2|\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2} \tag{3.51}$$

where ψ is the square of the sine value of the angle between the step, $\mathbf{s}^{(k)}$, and the difference between the gradient difference $\mathbf{y}_v^{(k)}$ and the change in gradient that accompanies the step $\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}$. The original form of SR1 will encounter numerical problem when the $\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}$ is close to 0. We avoided this problem by imposing a value check before the update takes place. The BFGS method circumvents this misbehavior by ensuring the Hessian update preserves positive-definiteness. SR1 and PSB do not preserve the positive semi-definite during the Hessian update process. In the TS optimization, it is important to maintain one negative eigenvalue during the process. This makes SR1, PEB, and Bofill the ideal candidates for TS optimization, leaving BFGS a superior choice for minimization.

3.3.9 Hessian Finite Differences Update

GOpt is an efficient algorithm as it can effectively describe the PES changes within the key/reduced-coordinate space. It is particularly important to maintain the accuracy of the elements of the approximate Hessian corresponding to the key coordinates during the whole optimization process. To achieve this, we update the first k rows/columns (corresponding to the key reduced coordinates) with finite

difference when needed. Denote the perturbation in the r^{th} key reduced coordinate as $\delta\mathbf{v} = \epsilon\mathbf{e}_r$, where $\mathbf{e}_r$ is the unit vector with all 0's except a '1' at the r^{th} position. The finite-difference formula for the r^{th} row/column of the Hessian matrix is similar to Eqn. 3.43,

$$\mathbf{H}\mathbf{e}_r = \frac{d\mathbf{g}_v}{d\epsilon} - \mathbf{V}^T(\mathbf{B}^T)^+\left(\mathbf{B}^T\frac{d\mathbf{V}}{d\epsilon}\mathbf{g}_v + \left(\frac{d\mathbf{B}}{d\epsilon}\right)^T\mathbf{g}_q\right) \tag{3.52}$$

The r^{th} row/columns of the Hessian matrix are approximated by

$$\frac{d\mathbf{g}(\mathbf{v}+\epsilon\mathbf{e_r})}{d\epsilon} = \frac{\mathbf{g}(\mathbf{v}+\epsilon\mathbf{e_r}) - \mathbf{g}(\mathbf{v})}{\epsilon} \tag{3.53}$$

The default value for ϵ we use in *GOpt* is 0.001.

Finite difference evaluation is a time-consuming step, requiring an additional energy and derivative evaluation. It is inefficient and unnecessary to update the Hessian at each iteration of the process. With the proper choice of the Quasi-Newton method, the key-coordinate portion of the Hessian matrix is often accurate. Therefore, the finite difference method is only invoked when the following criteria are met:

$$|\mathbf{g}_v^{(r)}| > \omega\frac{|\mathbf{g}_v|}{\sqrt{3N_{atoms}-6}} \tag{3.54}$$

$$|\mathbf{H}_v^{(k)}\mathbf{e}_r - \mathbf{H}_v^{(k-1)}\mathbf{e}_r| > \nu|\mathbf{H}_v^{(k-1)}\mathbf{e}_r| \tag{3.55}$$

The user parameters ω and ν are set, by default, to be 1.0. Criterion 3.54 checks the norm of the gradient in a specific direction compared with the overall gradient. There is no reason to ensure that the Hessian elements corresponding to an element of the gradient that is nearly zero are accurate, since in that case, the optimization

along that key coordinate is not hindering the convergence efficiency. Criterion 3.55 compares the discrepancy between Hessian matrix updated by the quasi-Newton methods. When the difference is small, it indicates that the update for that row of the Hessian is possibly accurate enough without the need for a finite-difference update.

3.3.10 Hessian Modification

A transition state is a saddle point on the PES. This requires the structure to have exactly one negative eigenvalue in the Hessian matrix. The corresponding eigenvector indicates the direction along which the energy is ascending.

It is important for the Hessian matrix to preserve exactly one negative eigenvalue with its eigenvector related to the chemical reaction. But it's not always the case when the molecular structure under optimization is far from the TS or the approximated Hessian is inaccurate. To ensure a well-behaved Hessian matrix, we modify the matrix to force the right structure.

First, we do not want the negative eigenvalue to be associated with molecular coordinates that are not related to the chemical reaction of interest. To prevent this, we ensure the non-key block of the Hessian matrix is positive semi-definite. If not, we replace the negative eigenvalue in the subblock with zero and reconstruct the block.

Second, one exact sufficiently negative eigenvalue in the key-coordinate space is needed. We diagonalize key subblock and check its eigenvalue(s). If the smallest eigenvalue is larger than the threshold λ_n (by default, -0.005 a.u.), we set the value to that threshold. If there are more than one negative eigenvalues, we reset all but the smallest negative eigenvalues to 0 and reconstruct the key-reduced block.

Third, after examine the two subblocks separately, we then do a complete eigenvalue check on the entire Hessian matrix.

$$\mathbf{H}_v^{(k)} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{3.56}$$

where $\mathbf{\Lambda}$ is the diagonal matrix with i^{th} eigenvalues λ_i. We list the eigenvalues in ascending order, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{3N-6}$. If there's only one negative eigenvalue less than the threshold λ_n, and all the other positive eigenvalue are greater than threshold λ_p, no further action is needed. This is most frequent situation.

$$\lambda_1 \leq \lambda_n \tag{3.57}$$

$$\lambda_i \geq \lambda_p \quad i = 2, 3, \ldots, 3N-6 \tag{3.58}$$

However, sometimes the Hessian matrix does not satisfied this criterion. If there is one negative eigenvalues, but it does not meet the requirements, we modify the eigenvalues,

$$\lambda_1 = \min(\lambda_1, \lambda_n) \tag{3.59}$$

$$\lambda_i = \max(\lambda_i, \lambda_p) \quad i = 2, 3, \ldots, 3N-6 \tag{3.60}$$

If there are multiple negative eigenvalues, we pick the one whose corresponding eigenvector has the most overlap with the key-reduced space. For each eigenvector with a negative eigenvalue, we sums up its components in the key-reduced space,

$$p_i = \sum_{r=1}^{R} |\chi_{i;r}|^2 \tag{3.61}$$

$\chi_{i:r}$ is the r^{th} element of the i^{th} eigenvector. We then retain the negative eigenvalue of the eigenvector with the largest p_i value, replacing all the other eigenvalues to $\max(\lambda_p, \lambda_i)$.

If no negative eigenvalue is presented, then p_i is computed for each eigenvector. Among all the eigenvectors with $p_i \geq 0.5$, we pick the one with the smallest eigenvalue as the candidate, modifying its eigenvalue to λ_n. The other positive eigenvalues are set to $\max(\lambda_p, \lambda_i)$.

3.3.11 Step Size Control

Given a non-stationary structure with its Hessian $\mathbf{H}_v$ and gradient $\mathbf{g}_v$, we can locate the TS through by taking the Newton step on the potential energy surface. Starting from the structure $\mathbf{v}_k$, the TS, which has zero gradients, is expected to be at $\mathbf{v}_{k+1}$ such that,

$$\mathbf{g}_v^{(k)} + \mathbf{H}_v^{(k)}(\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}) = 0 \tag{3.62}$$

the step to obtain the TS is then,

$$\mathbf{s}_v^{(k)} = \mathbf{v}^{(k+1)} - \mathbf{v}^{(k)} = -\left(\mathbf{H}_v^{(k)}\right)^{-1}\mathbf{g}_v^{(k)} \tag{3.63}$$

Expanding 3.63 with its eigenvalues and eigenvectors,

$$\mathbf{s}_v^{(k)} = -\sum_{i=1}^{3N-6}\left(\frac{\chi_i\chi_i^T\mathbf{g}_v^{(k)}}{\lambda_i}\right) = -\sum_{i=1}^{3N-6}\left(\frac{\chi_i^T\mathbf{g}_v^{(k)}}{\lambda_i}\right)\chi_i \tag{3.64}$$

The step, $\mathbf{s}_v^{(k)}$ would lead to the exact TS structure if the objective function is quadratic and the Hessian matrix were exact. However, the PES is not quadratic,

and our Hessian matrix is usually approximate. Therefore the above-computed step is only reliable when $v^{(k+1)}$ is close enough to $v^{(k)}$ for the (approximate) quadratic model to be accurate.. To ensure the $\mathbf{s}_v^{(k)}$ does not step out the valid region, a spherical region defined by radius τ is introduced.[45, 46] When the calculated stepsize from 3.63 is larger than the trust radius τ, it needs to be re-scaled to fit in the trust region.

Trust-region image method (TRIM)

To scale down oversized optimization step, TRIM modifies the eigenvalues by an undetermined variable $\tilde{\lambda} \geq 0$,

$$\mathbf{s}_v^{(k)}(\tilde{\lambda}) = -\sum_{\lambda_i<0}\left(\frac{\chi_1^T\mathbf{g}_v^{(k)}}{\lambda_1-\tilde{\lambda}}\right)\chi_1 - \sum_{\lambda_i\geq 0}\left(\frac{\chi_i^T\mathbf{g}_v^{(k)}}{\lambda_i+\tilde{\lambda}}\right)\chi_i \tag{3.65}$$

The value of $\tilde{\lambda}$ is computed by solving the nonlinear equation,

$$|\mathbf{s}_v^{(k)}(\tilde{\lambda})| = \tau \tag{3.66}$$

3.3.12 Trust Radius Determination

The value of τ should be neither too large to violate the validity of 3.63, nor too small to hinder the optimization convergence. The step-size should also scaled correctly as the number of atoms increases. To make sure the step is under proper

range, we define the trust radius as,

$$\tau_{init} = 0.35\sqrt{N_{atoms}}\text{a.u.} \tag{3.67}$$

$$\tau_{min} = 0.1\sqrt{N_{atoms}}\text{a.u.} \tag{3.68}$$

$$\tau_{max} = \sqrt{N_{atoms}}\text{a.u.} \tag{3.69}$$

$$\tag{3.70}$$

where τ_{init} is the initial trust radius, which is used in the first optimization step.

In the *GOpt* algorithm, if the magnitude of the gradient decreases after taking a step, the optimization step is accepted and the trust radius range is to be updated using either the "energy-based criterion" or the "gradient-based criterion". If the gradient increases instead, the step length is shortened by a factor of 4, $\tau_{new} = \frac{\tau_{old}}{4}$, and the step is recomputed. If the step is too small, $\tau_{new} < \frac{\tau_{min}}{10}$, then the step is taken and the trust radius is reset to its minimum value, $\tau_{new} = \tau_{min}$. This criterion is implemented to make sure when the structure is far from the ideal TS, the guess structure can make a step in the uphill direction from a near-minimum area, even if this step increases the magnitude of the gradient.

Energy-based update

The energy-based update compares the actual energy difference between the new geometry and the previous geometry to the estimated energy from quadratic approximation. The method uses the ratio of these two values to assess the accuracy

of the local quadratic approximation,

$$\Delta m^{(k)} = \mathbf{g}_v^{(k)} \cdot \mathbf{s}_v^{(k)} + \frac{1}{2}(\mathbf{s}_v^{(k)})^T \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)} \tag{3.71}$$

$$\Delta U^{(k)} = U(\mathbf{x}^{(k+1)}) - U(\mathbf{x}^{(k)}) \tag{3.72}$$

where Δm is the approximated energy calculated in v-space and ΔU is the actual energy difference between the latest two structures. When

$$\frac{2}{3} < \frac{\Delta m^{(k)}}{\Delta U^{(k)}} < \frac{3}{2}, \tag{3.73}$$

$$\text{then} \quad \tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max}) \tag{3.74}$$

It indicates the quadratic approximation of local potential energy is very accurate because the actual approximated energy is close to the real one. We then double the trust radius for the next step. If

$$\frac{1}{3} < \frac{\Delta m^{(k)}}{\Delta U^{(k)}} < 3, \tag{3.75}$$

$$\text{then} \quad \tau_{new} = \max(\tau_{old}, \tau_{min}) \tag{3.76}$$

It shows the quadratic model is moderately accurate. It is safer to keep the trust radius unchanged. Otherwise, we view the energy approximation as inaccurate and the trust radius is reduced for the next optimization step:

$$\tau_{new} = \max(\tfrac{1}{4}\tau_{old}, \tau_{min}) \tag{3.77}$$

Gradient-based update

The energy-based method is effective for minimization but when locating a transition state, it's more intuitive and appropriate to use a trust-radius update based on the accuracy of the predicted gradient.The predicted gradient is calculated as

$$\mathbf{g}_{v;predict}^{(k+1)} = \mathbf{g}_v^{(k)} + \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k+1)} \tag{3.78}$$

There are two indicators used in gradient-based method,

- The agreement in magnitude of gradient between the predicted and actual one, measured by the ratio.

$$\rho = \frac{|\mathbf{g}_{v;predict}^{(k+1)}| - |\mathbf{g}_v^{(k)}|}{|\mathbf{g}_v^{(k+1)}| - |\mathbf{g}_v^{(k)}|} \tag{3.79}$$

- The agreement in direction of gradient between the predicted and actual ones, measured by the cosine of the angle between them

$$\cos(\theta) = \frac{(\mathbf{g}_{v;predict}^{(k+1)} - \mathbf{g}_v^{(k)}) \cdot (\mathbf{g}_v^{(k+1)} - \mathbf{g}_v^{(k)})}{|\mathbf{g}_{v;predict}^{(k+1)} - \mathbf{g}_v^{(k)}| \cdot |\mathbf{g}_v^{(k+1)} - \mathbf{g}_v^{(k)}|} \tag{3.80}$$

The angle aligned by the gradient difference is sensitive measurement to the dimension of the system. As the dimension get larger, the chance of two vectors being aligned in the same direction decreases. For example, if one generates a large number of random vectors in d dimensions, 10% of the them will fulfill Eqn.3.81

and 40% will satisfy Eqn. 3.82

$$\cos(\theta) \geq p_{10}(d) \approx \sqrt{\frac{1.6424}{d} + \frac{1.11}{d^2}} \tag{3.81}$$

$$\cos(\theta) \geq p_{40}(d) \approx \sqrt{\frac{0.064175}{d} + \frac{0.0946}{d^2}} \tag{3.82}$$

The approximated expression are derived by least-square fitting to a much more complicated analytical expression.

For gradient-based trust-radius update, after a step is taken, if

$$\frac{4}{5} < \rho < \frac{5}{4} \tag{3.83}$$

$$\text{and} \quad p_{10}(3N-6) < \cos(\theta) \tag{3.84}$$

$$\text{then} \quad \tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max}) \tag{3.85}$$

then the approximate Hessian is accurate and we double the trust radius. If

$$\frac{1}{5} < \rho < 6 \tag{3.86}$$

$$\text{and} \quad p_{40}(3N-6) < cos(\theta) \tag{3.87}$$

$$\text{then} \quad \tau_{new} = \max(\tau_{old}, \tau_{min}) \tag{3.88}$$

Otherwise, we deem the approximate gradient from Eq. (3.78) inaccurate, so we halve the current trust radius,

$$\tau_{new} = \max(\tfrac{1}{2}\tau_{old}, \tau_{min}) \tag{3.89}$$

Recapitulation

These procedures for updating the trust radius are rather complicated, so we summarize:

1. If $g^{(k+1)} < g^{(k)}$, accept the step and update the trust radius with (a) the energy-based method or (b) the gradient-based method.

2. Otherwise, change the current trust radius to $\tau_{new} = \frac{1}{4}\tau_{old}$. If $\tau_{new} \geq \frac{1}{10}\tau_{min}$ recompute a new step with the shorter length to go back to step 1. Otherwise, set $\tau_{new} = \tau_{min}$ and take the step anyway.

3.3.13 Convergence Criterion

In *GOpt*, we use a similar criterion to the one proposed by Baker and Chan. We regard the optimization has achieved convergence if the largest component of the Cartesian gradient is less than 3.0×10^{-4} a.u. If the optimization doesn't converge after 100 iterations, it is then considered as a failed trial.

3.3.14 Summary of the Algorithm

Here, we put together all the components of the entire *GOpt* algorithm for geometry optimization

1. Start from the Cartesian coordinates; this set of coordinates is the input for *GOpt*.

2. Form a complete set of redundant internal coordinates to describe the system with preset protocols. Select the user-specified coordinates as the key internal

coordinates. If no key coordinates are specified, select the coordinates from the structural difference between reactant and product.

3. Construct the Wilson **B** matrix based on the selected redundant internal coordinates; construct reduced **V** matrix according to given key internal coordinates

4. Invoke external quantum chemistry software to compute the energy, energy gradient, and (possibly) Hessian.

5. Transform gradient and Hessian from Cartesian space to redundant internal space with the **B** matrix and to reduced internal space with **V** matrix.

6. Check if finite-differences are needed to update the rows/columns corresponding to the key reduced internal space. If criterion 3.54 and 3.55 are met, external software is invoked to conduct extra calculations.

7. Modify the hessian to ensure there is only one, sufficiently negative, eigenvalue.

8. Compute the optimization step in V-space within the trust radius.

9. Transform the step from V-space to internal coordinates. Then use manifold-projection to convert the target internal coordinates to the Cartesian coordinates of a molecular structure. Compute the energy and gradient of the new structure with external software.

10. If the magnitude of gradient decreases, accept the step. Otherwise, decrease the trust radius, and go back to step 8 to recompute a shortened step. If the

trust radius is less than $\tau_{min}/10$, accept the step anyway and then reset the trust radius to τ_{min}.

11. Construct **B** matrix and **V** matrix for the new structure. Align the new **V** with the previous one.

12. Update the Hessian matrix **H** with one of the quasi-Newton methods.

13. Check if the new structure meets the convergence criterion, if yes, return the latest structure as the final TS result. if not, go back to 5 to start next iteration of the optimization.

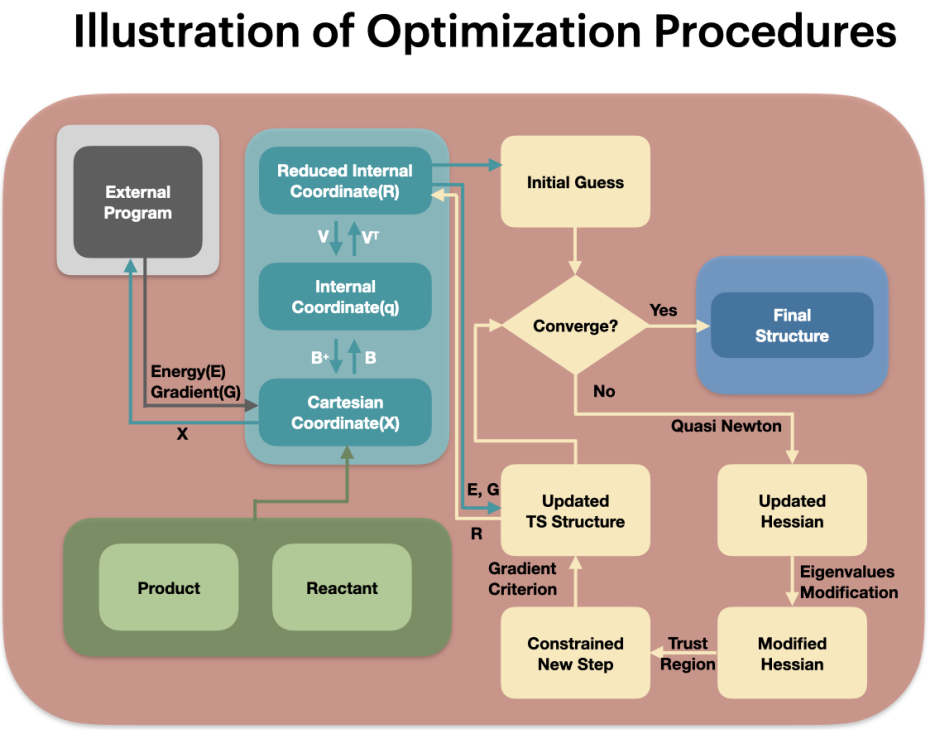


FIGURE 3.1: Illustration of optimization Procedures

3.4 Results and Discussion

3.4.1 Testing Protocol

To test the general performance of the algorithm, we use *GOpt* to optimize the TS for 32 randomly selected reactions of different types. Details of the reactions are list in Table 3.1. All energy calculations are performed by *Gaussian 16* with HF/6-31+G(d,p). Only the initial Hessian is computed analytically with *Gaussian*.

3.4.2 *GOpt* default methods

GOpt provides many different options for geometry optimization. Different combinations may perform differently for different reaction types. In this study, we selected methods we've found to have good overall performance. The details of the optimization comparison between different methods are elaborated upon in Chapter 5. By default, *GOpt* uses the Bofill quasi-Newton update, TRIM for step-size control, and gradient-based trust radius updates.

3.4.3 Comparison with *Berny* Algorithm

The *Berny* Algorithm is one of the most popular geometry optimization algorithm, implemented in many renowned software including *Gaussian*. To compare the performance, both algorithms start from the same initial guess and try to optimize the structure to the TS of interest.

3.4.4 Results and Discussion

The data presented in the 3.1 show the performance difference between the *GOpt* and *Berny* algorithms. Overall, *GOpt* needs averagely 7.38 steps to reach convergence, compared with 10.8 steps by *Berny's*. Also, among all the random sampled reactions, *GOpt* has converged all the guess structure to the desired transition state while *Berny* failed in 1 case. 78.1% of the reactions converge with the same, or fewer, gradient evaluation with *GOpt*. The failed case for *Berny* is reaction 32. During the optimization, the guess structure was displaced to an unrealistic structure where the SCF energy calculation cannot converge, resulting in an SCF error. In test. 6, 11 and 29, the performance of *GOpt* is substantially better than *Berny*. Using reduced internal coordinates and proper step control methods help eliminate excessive energy oscillation.

All these results suggest that *GOpt* is a promising algorithm for geometry optimization with robustness, efficiency, and versatility.

TABLE 3.1: Number of gradient evaluation needed for transition state optimization

index	Reaction	Num. of iterations	
		GOpt algorithm	*Berny* Algorithm
1	$C_4H_6 + C_2H_4$	6	6
2	$C_5H_6 + C_2H4$	4	5
3	$C_4H_4Si + C_2H_4$	4	6
4	$C_6H_8O + C_2H_4$	5	6
5	$C_4H_5N + C_2H_4$	5	6
6	C_4H_6	4	5
7	C_6H_8	3	6
8	C_8H_8	4	5
9	$C_{12}H_{18}$	6	20
10	$N_2O + C_2H_4$	5	7
11	$N_3 + C_2H_4$	11	30
12	$N2C_2 + C_2H_4$	6	7
13	$ONC + C_2H_4$	11	9
14	$N_2C_H + C_2H_4$	5	10
15	$HF + C_2H_4$	10	6
16	$C_2H_4 + H_2$	10	14
17	$HCN + H_2$	6	8
18	$HNC + H_2$	7	8
19	$C_2H_6 + SiH_2$	5	12
20	$HONS$	5	8
21	$HNCS$	6	7
22	$C_3H_4O_2$	7	14
23	C_6H_8	8	10
24	$CH_3F + Cl^-$	10	7
25	$CH_3Cl + F^-$	8	6
26	$CH_3F + F^-$	11	7
27	$CH_3OH + F^-$	6	17
28	$CH_3OH + \cdot OOH$	11	13
29	$CH_3OH + \cdot CH_3$	11	54
30	$HF + \cdot CH_3$	7	6
31	$N_2O + \cdot H$	13	10
32	$H_2O + \cdot CH_3$	16	failed
Ave. iterations		7.38	9.37
Converge Rate		100%	93.8%

The number of gradient evaluation needed to achieve convergence from the same TS guess for *GOpt* algorithm and *berny* algorithm.

3.5 Summary

Here, we present a new algorithm, *GOpt*, for geometry optimization. It uses the reduced internal coordinates with reaction-related key internals, which ensure that the Hessian matrix is accurate for the most chemically-important coordinates and to ensure that the TS optimization approaches the direction we desire. *GOpt* algorithm only needs to evaluate the Hessian matrix once (in the first iteration); in subsequent steps, the Hessian was updated using a Quasi-Newton update and finite-differences for key coordinates. In a future paper, we will show that even an approximate initial Hessian (e.g., from a minimal basis-set calculation) often suffices to initialize the procedure.

The optimization process is conducted in reduced internal space while the major quantum chemistry properties, like energy, gradient, are computed in Cartesian coordinates. To effectively convert geometries between different representations, we introduced the robust dihedral descriptors and the manifold-projection method. These methods provide us a failsafe way to interconvert between redundant internal coordinates and Cartesian coordinates.

Due to these improvements, our transition-state algorithm performs significantly better than those in traditional quantum chemistry software. Specifically, the incidence of convergence failure is reduced, and the number of gradient evaluations required to converge from a reasonably good initial guess is reduced.

3.6 Reference

[1] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, et al. *Gaussian 16*. 2016.

[2] H. B. Schlegel. Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry* 3(2) (1982), 214–218.

[3] J. E. Straub. Reaction rates and transition pathways. *Eastern Hemisphere Distribution* (2001), 199.

[4] G. Crippen and H. Scheraga. Minimization of polypeptide energy: XI. The method of gentlest ascent. *Archives of biochemistry and biophysics* 144(2) (1971), 462–466.

[5] D. Poppinger. On the calculation of transition states. *Chemical Physics Letters* 35(4) (1975), 550–554.

[6] C. J. Cerjan and W. H. Miller. On finding transition states. *The Journal of chemical physics* 75(6) (1981), 2800–2806.

[7] J. Simons, P. Joergensen, H. Taylor, and J. Ozment. Walking on potential energy surfaces. *The Journal of Physical Chemistry* 87(15) (1983), 2745–2753.

[8] G. Henkelman and H. Jónsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of chemical physics* 111(15) (1999), 7010–7022.

[9] J. Kästner and P. Sherwood. Superlinearly converging dimer method for transition state search. *The Journal of chemical physics* 128(1) (2008), 014106.

[10] C. Shang and Z.-P. Liu. Constrained Broyden minimization combined with the dimer method for locating transition state of complex reactions. *Journal of Chemical Theory and Computation* 6(4) (2010), 1136–1144.

[11] A. Heyden, A. T. Bell, and F. J. Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *The Journal of chemical physics* 123(22) (2005), 224101.

[12] J. Pancir. Calculation of the least energy path on the energy hypersurface. *Collection of Czechoslovak Chemical Communications* 40(4) (1975), 1112–1118.

[13] M. Basilevsky and A. Shamov. The local definition of the optimum ascent path on a multi-dimensional potential energy surface and its practical application for the location of saddle points. *Chemical Physics* 60(3) (1981), 347–358.

[14] D. K. Hoffman, R. S. Nord, and K. Ruedenberg. Gradient extremals. *Theoretica chimica acta* 69(4) (1986), 265–279.

[15] W. Quapp, M. Hirsch, O. Imig, and D. Heidrich. Searching for saddle points of potential energy surfaces by following a reduced gradient. *Journal of computational chemistry* 19(9) (1998), 1087–1100.

[16] M. Hirsch and W. Quapp. Improved RGF method to find saddle points. *Journal of computational chemistry* 23(9) (2002), 887–894.

[17] R. Crehuet, J. M. Bofill, and J. M. Anglada. A new look at the reduced-gradient-following path. *Theoretical Chemistry Accounts* 107(3) (2002), 130–139.

[18] G. Henkelman, B. P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics* 113(22) (2000), 9901–9904.

[19] G. Henkelman and H. Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of chemical physics* 113(22) (2000), 9978–9985.

[20] S. A. Trygubenko and D. J. Wales. A doubly nudged elastic band method for finding transition states. *The Journal of chemical physics* 120(5) (2004), 2082–2094.

[21] J.-W. Chu, B. L. Trout, and B. R. Brooks. A super-linear minimization scheme for the nudged elastic band method. *The Journal of chemical physics* 119(24) (2003), 12708–12717.

[22] P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman, and E. Kaxiras. Adaptive nudged elastic band approach for transition state calculation. *The Journal of chemical physics* 117(10) (2002), 4651–4658.

[23] I. F. Galván and M. J. Field. Improving the efficiency of the NEB reaction path finding algorithm. *Journal of computational chemistry* 29(1) (2008), 139–143.

[24] D. Sheppard, R. Terrell, and G. Henkelman. Optimization methods for finding minimum energy paths. *The Journal of chemical physics* 128(13) (2008), 134106.

[25] D. R. Alfonso and K. D. Jordan. A flexible nudged elastic band program for optimization of minimum energy pathways using ab initio electronic

structure methods. *Journal of computational chemistry* 24(8) (2003), 990–996.

[26] N. González-Garcia, J. Pu, À. González-Lafont, J. M. Lluch, and D. G. Truhlar. Searching for saddle points by using the nudged elastic band method: an implementation for gas-phase systems. *Journal of chemical theory and computation* 2(4) (2006), 895–904.

[27] E. Weinan, W. Ren, and E. Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *Journal of Chemical Physics* 126(16) (2007), 164103.

[28] A. Samanta and E. Weinan. Optimization-based string method for finding minimum energy path. *Communications in Computational Physics* 14(2) (2013), 265–275.

[29] P. Y. Ayala and H. B. Schlegel. A combined method for determining reaction paths, minima, and transition state geometries. *The Journal of chemical physics* 107(2) (1997), 375–384.

[30] S. K. Burger and W. Yang. Sequential quadratic programming method for determining the minimum energy path. *The Journal of chemical physics* 127(16) (2007), 164107.

[31] S. K. Burger and W. Yang. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *The Journal of chemical physics* 124(5) (2006), 054109.

[32] E. F. Koslover and D. J. Wales. Comparison of double-ended transition state search methods. *The Journal of chemical physics* 127(13) (2007), 134102.

[33] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of chemical physics* 120(17) (2004), 7877–7886.

[34] A. Goodrow, A. T. Bell, and M. Head-Gordon. Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *The Journal of chemical physics* 129(17) (2008), 174109.

[35] A. Goodrow, A. T. Bell, and M. Head-Gordon. Transition state-finding strategies for use with the growing string method. *The Journal of chemical physics* 130(24) (2009), 244108.

[36] A. Goodrow, A. T. Bell, and M. Head-Gordon. A strategy for obtaining a more accurate transition state estimate using the growing string method. *Chemical Physics Letters* 484(4-6) (2010), 392–398.

[37] T. A. Halgren and W. N. Lipscomb. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chemical Physics Letters* 49(2) (1977), 225–232.

[38] C. Peng and H. Bernhard Schlegel. Combining synchronous transit and quasi-newton methods to find transition states. *Israel Journal of Chemistry* 33(4) (1993), 449–454.

[39] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* 17(1) (1996), 49–56.

[40] V. Bakken and T. Helgaker. The efficient optimization of molecular geometries using redundant internal coordinates. *The Journal of chemical physics* 117(20) (2002), 9160–9174.

[41] P. Pulay and G. Fogarasi. Geometry optimization in redundant internal coordinates. *The Journal of chemical physics* 96(4) (1992), 2856–2860.

[42] P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. *Journal of the American Chemical Society* 101(10) (1979), 2550–2560.

[43] J. Baker, A. Kessi, and B. Delley. The generation and use of delocalized internal coordinates in geometry optimization. *The Journal of chemical physics* 105(1) (1996), 192–212.

[44] J. Baker, D. Kinghorn, and P. Pulay. Geometry optimization in delocalized internal coordinates: An efficient quadratically scaling algorithm for large molecules. *The Journal of chemical physics* 110(11) (1999), 4986–4991.

[45] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.

[46] R. Fletcher. Practical methods of optimization john wiley & sons. *New York* 80 (1987), 4.

Chapter 4

Bisecting hyperplane optimization for reaction path finding

4.1 Abstract

A new double-ended algorithm is proposed to locate the reaction path. The algorithm is based on the robust reduced internal coordinates introduced in chapter2. The reaction vector is determined by the two end structure and updated over the optimization process. Each state of the reaction path is optimized to the minimum structure in the bisecting hyperplane perpendicular to the reaction vector between two structures. The resulting paths agree with reaction paths obtained from other techniques, but this method is explicitly parallelizable and avoids numerical ill-conditioning and kinking problems associated with other techniques.

4.2 Introduction

Especially for complicated multi-step reactions, it can be very difficult to locate transition states, or even to propose reasonable reactive intermediates. In such

cases, it is best to determine the chemical reaction path directly, rather than attempt to guess stable structures along the reaction path. Having a full reaction pathway is also useful for detailed mechanistic studies of reactions, where having an atomistic description of the reaction pathway is useful. By definition, the reactive pathway will pass through reactive intermediates and transition states. Though several choices of reaction pathways are present in the literature, the most common is the minimum energy pathway, which is normally coincident with the intrinsic reaction coordinate obtained as the steepest-descent path from the transition state(s). The minimum energy path forms a leading line about which reactive trajectories cluster (in the limit of zero temperature).

The main two strategies for finding reaction pathways are the single-ended and double-ended schemes. Single-ended methods ascend uphill towards a transition state from a single stable structure, either from the reactant or the product. However, taking a random step towards any direction is an uphill move, so extra information is needed to locate the desired transition. The most common strategy is the coordinate driving method, which relies on the researcher to pick a coordinate associated with the reaction; the energy is minimized with respect to the other coordinates.[1]. Coordinate-driving is challenging when it is difficult to identify a single key coordinate.[2–5] An alternative idea is to follow the direction along the eigenvector of Hessian matrix with the smallest eigenvalue.[6–9] This is equivalent to choosing the direction with the least energy ascent on the potential energy surface, but it often leads towards transition states associated with conformational change, rather than chemical bond formation/breaking. Another popular method is to follow the gradient extremal path to the transition state.[10–12] This method

is designed to pass through stationary points on the potential energy surface. Different pathways from gradient extremal intersect at transition states. There are two common drawbacks to the latter methods. First, the costly evaluation of the Hessian matrix is required; Second, paths that weave around the potential energy surface are often obtained when using gradient extremals. For the gradient extremal method, the paths generated are normally different from the minimum energy paths.

Rather than looking for the paths from one side of the reaction, some methods try to locate the path by a set of discrete points connecting the reactant to the product, namely the chain-of-states methods. The nudged elastic band (NEB) method and the string method (SM) are the two main types of methods in the double-ended family. For the NEB method, one or more extra virtual spring potentials are appended to the original potential energy expression for the atomic nuclei. The gradient of the spring potential and surface potential are used to adjust the states in the direction along the reaction path and the perpendicular hyperplane respectively to maintain equal spacing between states in the chain. In the string method, the reaction pathway is described as a string connecting the reactant and product. When implemented, the string is represented by multiple discrete points connected by a spline path. During the optimization, each point follows the steepest gradient descent in the hyperplane of the reaction path. Chain-of-state methods like NEB and SM can effectively locate the reaction path from the reactant to the product while bypassing the direct calculation of the exact transition state. The main drawback of these chain-of-state methods is the demanding computational requirements. High-performance parallel computing is commonly used, but the coupling between states in the NEB method makes this less than

totally straightforward. Beside NEB and SM, other chain-of-states methods are also actively developed such as conjugate peak refinement,[13] replica path,[14–17] line-integral,[18–22] and different derived string methods including zero temperature string methods,[23–26] finite temperature string methods,[27] quadratic string method,[28] and growing string methods.[29–34]

4.3 Overview

Here we presented a new bisection algorithm for locating a reaction pathway to help identify the transition state between the reactant and the product. The motivation for this strategy is to construct a method for finding the minimum energy pathway that is easily parallelized and robust. Like other two-ended methods, the bisection method works by defining the reaction path as a sequence of points. Unlike other techniques, every step in the bisection method is a simple, robust, local minimization that can be performed simultaneously with other minimizations.

The first step in the bisection method is to take the reactant and product, denoted by $\mathbf{q}_{reactant}$ and $\mathbf{q}_{product}$ in redundant internal coordinates. The reaction path vector is initialised to the vector between the two structures, $\mathbf{q}_{path} = \mathbf{q}_{product} - \mathbf{q}_{reactant}$. Minimizing starting from the midpoint between the structures, $\mathbf{q}_{guess} = (\mathbf{q}_{product} + \mathbf{q}_{reactant})/2$ on the hyperplane perpendicular to the reaction path-vector always locates a point on a minimum energy path. The manifold projection algorithm is used to find a molecular structure corresponding to $\mathbf{q}_{guess}$. Since the constrained minimization and the manifold projection algorithm are robust, this method always converges.

Note that this algorithm works for any two structures. Denoting the Cartesian structures of two points as $\mathbf{x}_{start}$ and $mathbfx_{end}$, one can uniquely define the redundant internal coordinates, denoted as $\mathbf{q}_{start}$ and $\mathbf{q}_{end}$ respectively. The reaction path vector is $\mathbf{q}_{path} = \mathbf{q}_{end} - \mathbf{q}_{start}$ and the initial guess structure is a linear combination of each end $\mathbf{q}_{init} = (\mathbf{q}_{start} + \mathbf{q}_{end})/2$, where the manifold projection method is used to locate the closest structure $\mathbf{q}(\mathbf{x})_{init}$ in Cartesian space.

Therefore, after the first point between the reactant and product is located, one can bisect between this point and the reactant and product, setting up two parallel constrained minimizations. This procedure can be repeated for any number of sequential points, and every time computational resources become available, one can choose any interval between two previous completed calculations and bisect it. Typically one would pick an interval based on chemical interest (e.g., proximity to a stationary point on the potential surface) or mathematical necessity (e.g., a region of the reaction path where the curvature is high or the distance between previous structures is large). If at the end of this procedure, a continuous reaction path is obtained, this is *guaranteed* to be a minimum energy reaction path. If a continuous reaction path is not located (which can happen when the reactant and product structures are very different and the topology of the potential energy surface is very complicated, one nonetheless knows that every continuous segment of the curve is a minimum-energy pathway. Then, by adding additional structures to extend these curves, a full minimum energy pathway can be constructed, to whatever precision is desired (by bisecting to the degree desired). This method, therefore, eliminates the non-robustness (convergence failures) and kinked-pathway problems (due to local minimum tracks) that are associated with the competing elastic-band and string methods.

4.4 Methodology

4.4.1 Coordinate system

To effectively describe a chemical reaction with meaningful coordinates, we normally select redundant internal coordinates, such as chemical bonds, angles, and dihedrals between planes, to represent the system during the reaction, denoted $\{q_1, q_2, \ldots, q_{M_{int}}\}$. There are various methods to select internal coordinates but here we use the protocol described in Chapter 2.

Specifically, we consider five types of inter-atomic bonds, including covalent bonds, hydrogen bonds, inter-fragment bonds, long-distance bonds, and auxiliary bonds; these are added based on the types of atoms involved and the distance between them. We include the angles formed by each pair of bonds that are connected to the same atom, excluding auxiliary bonds. For dihedrals, we replace the conventional definition with our robust dihedral indicator. To limit the number of dihedrals in the system, we only include the dihedral consisting of $\alpha\beta\gamma*$ and $*, \beta\gamma\delta$ where α and δ represent the most connected atoms bonded to β and γ respectively, while * represents a non-selected atom.

The transformation from an infinitesimal change in Cartesian coordinates to an infinitesimal change in redundant internal coordinates is defined by the $\mathbf{B}$ matrix

$$\delta\mathbf{q}(\mathbf{x}) = \mathbf{B}\delta\mathbf{x} \tag{4.1}$$

while the inverse is performed by iterative manifold projection method

$$\mathbf{x}(\mathbf{q}^{(target)}) = \arg\underbrace{\min}_{\mathbf{x}}(\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)})^T\mathbf{W}(\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)}) \tag{4.2}$$

where $\mathbf{x}$ is the Cartesian coordinates; $\mathbf{q}(\mathbf{x})$ is the set of internal coordinates corresponding to $\mathbf{x}$; $\mathbf{q}^{(target)}$ is the desired internal set, which may not be a physically realizable molecular structure. The forward and back transformation is not symmetric, because the number of redundant internal coordinates is far higher than the number of Cartesian coordinates. Therefore, for a set of Cartesian coordinates, corresponding internal coordinates can always be explicitly computed, but not *vice versa*.

Delocalized internal coordinates

To fully specify the structure of a (nonlinear) N-atom molecule, only $3N-6$ degrees of freedom are needed. In this bisection algorithm, we adopted a similar idea as the one elaborated in chapter 2 to generate initial guesses for points on the reaction pathway, then we refine these guesses.

The delocalized internal coordinates are found by choosing $\mathbf{3N-6}$ non-zero singular vectors from the $\mathbf{B}$ matrix, denoted as

$$\mathbf{a^{(i)}} = \begin{bmatrix} a_1^{(i)} & a_2^{(i)} & \dots & a_{N_{int}}^{(i)} \end{bmatrix} \quad i = 1, 2, \dots, 3N_{atoms} - 6 \tag{4.3}$$

To separate the reaction-path coordinate from the non-reaction coordinates, we compute the coordinate displacement $\mathbf{q_{path}}$ between the reaction and product

$$\mathbf{q}_{path} = \mathbf{q}_{start} - \mathbf{q}_{end} \tag{4.4}$$

We then project the reaction path indicator into the realizable space

$$\mathbf{q}'_{path} = \mathbf{B}\mathbf{B}^{+}\mathbf{q}_{path} \tag{4.5}$$

The $\mathbf{V}_{react}$ consists solely of the reaction path vector

$$\mathbf{V}_{react} = \frac{\mathbf{q}'_{path}}{|\mathbf{q}'_{path}|} \tag{4.6}$$

To construct the full non-react space without the reaction path vector, we need to project out the path vector first

$$\mathbf{d}^{(j)} \equiv (\mathbf{I} - \mathbf{P}_{react})\mathbf{a}^{(j)} \tag{4.7}$$

$$= \mathbf{a}^{(j)} - \mathbf{V}_{react}\mathbf{V}_{react}^{T}\mathbf{a}^{(j)} \tag{4.8}$$

where $\mathbf{P}_{react}$ is the projection operator for the reaction-path vector. In the leftover $\mathbf{D}$ space, we pick the $3N-7$ non-singular orthonormal eigenvectors as the basis,

$$\mathbf{D}\mathbf{D}^{T} = \mathbf{V}'\mathbf{\Lambda}\mathbf{V}'^{T} \tag{4.9}$$

These orthonormal vectors are a basis for the hyperplane perpendicular to the reaction vector, denoted as

$$V_{nonreact} = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \dots & \mathbf{v}'_{3N-7} \end{bmatrix} \tag{4.10}$$

Optimization Process

After projecting out the reaction path vector, finding a point on the minimum-energy path reduces to a constrained minimization problem within the leftover subspace. The energy, gradient, and Hessian are normally computed in Cartesian

coordinates. Conversion between the different coordinate system are defined as:

$$\tilde{\mathbf{g}}_v = \tilde{\mathbf{V}}^T \mathbf{g}_q \tag{4.11}$$

$$\mathbf{g}_q = \tilde{\mathbf{V}} \tilde{\mathbf{g}}_v \tag{4.12}$$

$$\tilde{\mathbf{H}}_v = \tilde{\mathbf{V}}^T \mathbf{H}_q \tilde{\mathbf{V}} \tag{4.13}$$

$$\mathbf{H}_q = \tilde{\mathbf{V}} \tilde{\mathbf{H}}_v \tilde{\mathbf{V}}^T \tag{4.14}$$

$$\Delta\tilde{\mathbf{v}} = \tilde{\mathbf{V}}^T \Delta\mathbf{q} \tag{4.15}$$

$$\Delta\mathbf{q} = \tilde{\mathbf{V}} \Delta\tilde{\mathbf{v}} \tag{4.16}$$

$$\tag{4.17}$$

Here, we use $\tilde{X}$ to denoting all quantities in the $\mathbf{V}_{nonreact}$ space. The Newton step truncated at the second order derivative is

$$\tilde{\mathbf{s}} = -\tilde{\mathbf{H}}^{-1} \tilde{\mathbf{g}} \tag{4.18}$$

One could minimize the hyperplane by simply following the (constrained) gradient descent direction. It's significantly more efficient to use the Hessian matrix, but to ensure a proper direction is taken for the optimization step, the hessian matrix is required to have all positive eigenvalues. Therefore, whenever the Hessian matrix has a negative eigenvalue, a Hessian shift function is called to alter the negative or small positive value to the preset threshold λ_p.

$$\lambda_i = \begin{cases} \lambda_p & \text{if } \lambda_i < \lambda_p \\ \lambda_i & \text{otherwise} \end{cases} \tag{4.19}$$

The BFGS quasi-Newton method is used to update the Hessian. This ensures that given an initial positive-definite Hessian, the updated Hessian preserves this structure. The energy-based trust-radius update presented 3 is used to ensure global convergence. We consider the calculation to have converged when the largest component of the Cartesian gradient in the non-reacting $\mathbf{V}_{nonreact}$ space is less than 1×10^{-3}.

4.5 Examples and Cases

4.5.1 Muller-Brown Potential

The Muller-Brown potential is a two-variable parametric potential that is commonly used to test reaction-pathfinding methods.[35] The minimum energy path connects the reactant and product through an intermediate and two saddle points. Also, the path deviates strongly from linear interpolation between the reactant and product structures, which is challenging for some approaches.

The path-optimization process for the bisecting hyperplane method is presented in Fig. 4.1. The initial guess structure, G is halfway between the reactant, P_0, and product, P_1 structures, and when it is optimized in the hyperplane perpendicular to the vector $\mathbf{q}_{P_1} - \mathbf{q}_{P_0}$ gives the structure P_2 (Fig. 4.1A). After the first middle point P_2 anchors the reaction pathway, new points are generated halfway between P_0 and P_2 and halfway between P_1 and P_2, and optimized on the corresponding bisecting hyperplanes (Fig. 4.1B). This process is continued until enough points are generated to represent the reaction path. Unlike other popular reaction path methods, bisecting hyperplane optimization does not keep even-spaced points. This grants the algorithm the ability to put more points in regions where the

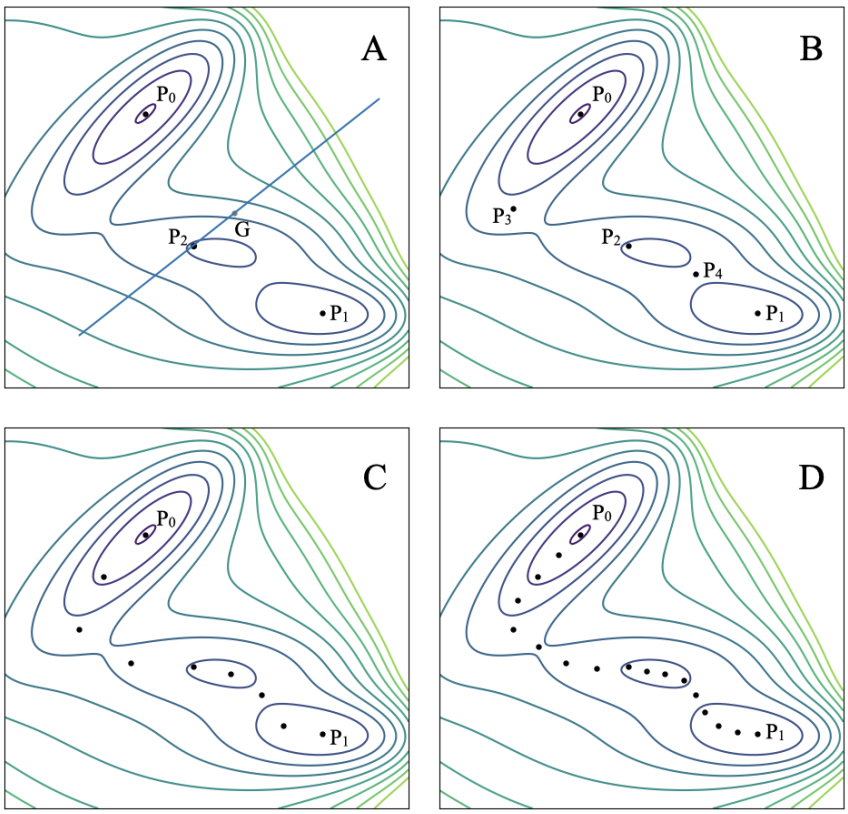


FIGURE 4.1: Reaction path finding on Muller-Brown surface with Bisection method

curvature of the path is large, where the gradient is small (near stationary points), etc..

The process converges when a user-specified number of points have been generated, or other preset criteria (e.g. the maximum distance between consecutive points on the reaction path is sufficiently small) In Fig. 4.1C & D, we generate 9 points and 17 points respectively to describe the reaction path. Additional points could be considered, if desired.

The data in Tab.4.1 illustrates the accuracy of the reaction paths in Fig. 4.1C & D. The intermediate and saddle points are interpolated by a cubic spline going

TABLE 4.1: Analytical and interpolated stationary points from the reaction paths

	Intermediate	Saddle Point1	Saddle Point1
	Coordinates & Potential		
Analytic	(-0.050, 0.467) -80.768	(-0.822, 0.624) -40.665	(0.212, 0.293) -72.249
path 4.1C	(-0.040, 0.473) -80.714	(-0.845, 0.602) -40.421	(0.222, 0.298) -72.218
path 4.1D	(-0.050, 0.464) -80.761	(-0.814, 0.632) -40.633	(0.220, 0.298) -72.230

through all path points. (The path in Fig. 4.1B is not shown in the table because there is not enough data to locate all the stationary points.) The two saddle points and the intermediate are successfully identified and located. The absolute error in coordinates is ±0.02 and ±0.01 for C and D respectively. The energy difference is ±0.2 for Path C and ±0.03 for D. Structures with this level of accuracy suffice as initial guesses for subsequent geometry optimization for the reactive intermediate(s) and/or transition state(s).

4.5.2 Chemical Reactions

1. HCN Isomerization

To test the bisecting hyperplane method for real chemical reactions, we consider first the HCN → HNC isomerization. This isomerization reaction has been studied thoroughly both experimentally and theoretically and, as a 3-atom molecule has just 3 degrees of freedom.[36–38] All calculations, including the initial reactant and product structures, were determined using the HF/6-31+G method.

Considering the CH and CN bond lengths and the ∠HCN as the three internal coordinates, one observes that the CN bond length does not change very much during the reaction process. The reaction path, therefore, can be illustrated by neglecting the change in this coordinate, as depicted in Figure 4.2. It is worth noting that in the actual result from the bisecting hyperplane optimization, the bond length between is C and N is not constant. The CN bond slightly stretches from 2.19 a.u. in the reactant to 2.26 a.u. near the transition state.

TABLE 4.2: Transition state from interpolation and analytic Computation for HCN -> CNH reaction

	Interpolation	Analytic Computation
Energy	-92.72995 a.u.	-92.72972 a.u.
CH bond HCN angle	2.2801 a.u. 71.44°	2.2862 a.u. 71.83°

Parameterizing the interpolated reaction path by its arc length allows one to generate the continuous reaction coordinate diagram in Fig.4.3. The transition state of the isomerization can be obtained by finding the maximum on the spline. As seen in Table4.2, the energy and molecular structure from the interpolated reaction path agrees closely with the reference obtained by normal transition-state optimization methods, even though only 7 points were used to define the spline.

2. HSNO ← HONS Isomerization

The isomerization from HSNO to HONS may seem similar to the previous example, but it is far more complicated.[39–41] There are several competing reaction mechanisms, with multiple reactive intermediates and transition states, as shown in Fig.4.4.

With such a complicated chemical reaction, more points along the reaction path are needed. Starting from the stable structures of HSON and HNOS, a reaction path consisting of 17 points (including two starting points) is generated. The energy change along the chemical reaction process is shown in Fig.4.5.

Based on the structures we observe, the bisecting hyperplane optimization method locates reaction mechanism A in Fig.4.4. The path successfully goes through two transition states, together with the connecting intermediate, and the interpolated structures and their energies are very accurate (see Table 4.3), providing quantitative accuracy for the whole reaction path. One can even obtain additional insight: the tiny peak around $x = 0.13$, which one might at first presume to be a computational artifact, corresponds to the hydrogen-atom-rotational barrier.

TABLE 4.3: Energies from interpolation of the reaction path and direct optimization of key structures in the HSNO ← HONS reaction

	Interpolation	Direct Optimization	Relative error
TS1 Energy	-527.080177 a.u.	-527.080348 a.u.	0.000032%
TS2 Energy	-527.068519 a.u.	-527.068274 a.u.	0.000046%
Intermediate Energy	-527.192585 a.u.	-527.189130 a.u.	0.00066%

4.6 Conclusion

The bisecting hyperplane optimization method is an efficient, accurate, and robust way to generate a minimum energy pathway connecting the reactant and product. By construction, the method provides a kink-free pathway, wherein several points on the reaction path can be optimized simultaneously. The only required input

is the two end structures. No limitation is imposed on what starting and ending points are selected; they do not even have to be stable reactant/product geometries.

The basic idea of the algorithm is to divide the space of internal coordinates into a path-direction vector (defined by the points before and after the current point on the reaction path) and its orthogonal hyperplane. Optimization in the orthogonal hyperplane is performed directly as a minimization in a reduced-dimensionality space, and yields a point that is guaranteed to be on a minimum energy pathway. Additional points can be strategically allocated in regions where a better description of the potential energy surface is needed.

Three different examples were used to illustrate the effectiveness of the algorithm. The first is the venerable Muller-Brown model, which is difficult because of its curved reaction pathway. In this reaction, the potential energy surface is nonconvex, and the *Gopt* Hessian-matrix modification scheme, which forces the Hessian matrix to be positive definite, was helpful. The second and third examples were chemical reactions of varying complexity, the isomerization of HCN, and of HSNO. Minimum energy pathways without difficulty in both cases, and energies and molecular structures of the reactive intermediate and transition states were accurately located. This suggests that the bisecting hyperplane optimization method should be considered as an alternative to other, alternative, two-ended approaches to reaction path optimization like the nudged elastic band and string families of methods.

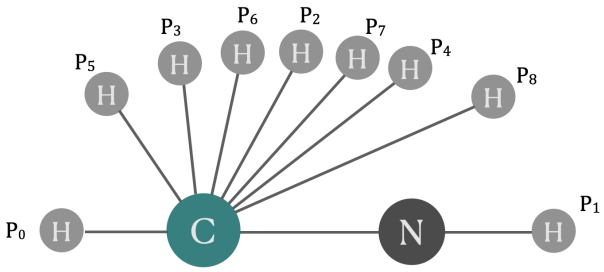


FIGURE 4.2: Path points generated by Bisection algorithm for HCN -> CNH reaction

P_0 and P_1 are stable structures for reactant and product. The labels indicate the sequence in which points were generated by the serial bisecting hyperplane algorithm.

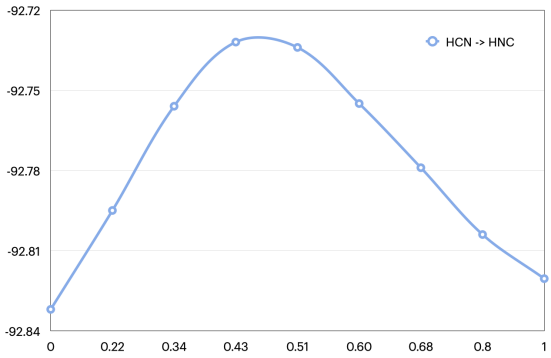


FIGURE 4.3: Energy curve of reaction HCN -> CNH along the reaction coordinates

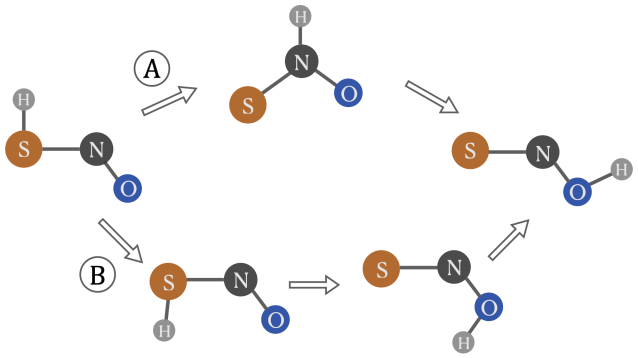


FIGURE 4.4: HSNO ← HONS isomerization mechanism

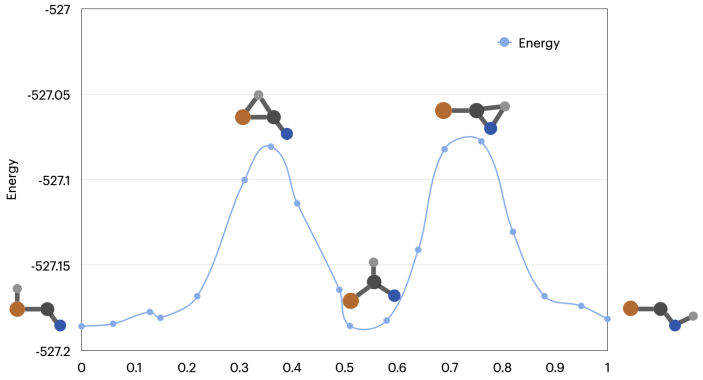


FIGURE 4.5: HSNO ← HONS Energy vs Reaction process

4.7 Reference

[1] J. E. Straub. Reaction rates and transition pathways. *Eastern Hemisphere Distribution* (2001), 199.

[2] U. Burkert and N. L. Allinger. Pitfalls in the use of the torsion angle driving method for the calculation of conformational interconversions. *Journal of computational chemistry* 3(1) (1982), 40–46.

[3] I. H. Williams and G. M. Maggiora. Use and abuse of the distinguished-coordinate method for transition-state structure searching. *Journal of Molecular Structure: THEOCHEM* 89(3-4) (1982), 365–378.

[4] P. Scharfenberg. Theoretical analysis of constrained minimum energy paths. *Chemical Physics Letters* 79(1) (1981), 115–117.

[5] M. J. Rothman and L. L. Lohr Jr. Analysis of an energy minimization method for locating transition states on potential energy hypersurfaces. *Chemical Physics Letters* 70(2) (1980), 405–409.

[6] G. Crippen and H. Scheraga. Minimization of polypeptide energy: XI. The method of gentlest ascent. *Archives of biochemistry and biophysics* 144(2) (1971), 462–466.

[7] D. Poppinger. On the calculation of transition states. *Chemical Physics Letters* 35(4) (1975), 550–554.

[8] C. J. Cerjan and W. H. Miller. On finding transition states. *The Journal of chemical physics* 75(6) (1981), 2800–2806.

[9] J. Simons, P. Joergensen, H. Taylor, and J. Ozment. Walking on potential energy surfaces. *The Journal of Physical Chemistry* 87(15) (1983), 2745–2753.

[10] J. Pancir. Calculation of the least energy path on the energy hypersurface. *Collection of Czechoslovak Chemical Communications* 40(4) (1975), 1112–1118.

[11] M. Basilevsky and A. Shamov. The local definition of the optimum ascent path on a multi-dimensional potential energy surface and its practical application for the location of saddle points. *Chemical Physics* 60(3) (1981), 347–358.

[12] D. K. Hoffman, R. S. Nord, and K. Ruedenberg. Gradient extremals. *Theoretica chimica acta* 69(4) (1986), 265–279.

[13] S. Fischer and M. Karplus. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical physics letters* 194(3) (1992), 252–261.

[14] R. Czerminski and R. Elber. Reaction path study of conformational transitions and helix formation in a tetrapeptide. *Proceedings of the National Academy of Sciences* 86(18) (1989), 6963–6967.

[15] R. Czerminski and R. Elber. Reaction path study of conformational transitions in flexible systems: applications to peptides. *The Journal of chemical physics* 92(9) (1990), 5580–5601.

[16] H. L. Woodcock, M. Hodošček, P. Sherwood, Y. S. Lee, H. F. Schaefer Iii, and B. R. Brooks. Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to

prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theoretical Chemistry Accounts* 109(3) (2003), 140–148.

[17] J. B. Brokaw, K. R. Haas, and J.-W. Chu. Reaction path optimization with holonomic constraints and kinetic energy potentials. *Journal of chemical theory and computation* 5(8) (2009), 2050–2061.

[18] R. Elber and M. Karplus. A method for determining reaction paths in large molecules: Application to myoglobin. *Chemical Physics Letters* 139(5) (1987), 375–380.

[19] R. Czerminski and R. Elber. Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *International Journal of Quantum Chemistry* 38(S24) (1990), 167–185.

[20] A. Ulitsky and R. Elber. A new technique to calculate steepest descent paths in flexible polyatomic systems. *The Journal of chemical physics* 92(2) (1990), 1510–1511.

[21] C. Choi and R. Elber. Reaction path study of helix formation in tetrapeptides: Effect of side chains. *The Journal of chemical physics* 94(1) (1991), 751–760.

[22] W. Nowak, R. Czerminski, and R. Elber. Reaction path study of ligand diffusion in proteins: application of the self penalty walk (SPW) method to calculate reaction coordinates for the motion of CO through leghemoglobin. *Journal of the American Chemical Society* 113(15) (1991), 5627–5637.

[23] W. Ren et al. Higher order string method for finding minimum energy paths. *Communications in Mathematical Sciences* 1(2) (2003), 377–384.

[24] E. Weinan, W. Ren, and E. Vanden-Eijnden. String method for the study of rare events. *Physical Review B* 66(5) (2002), 052301.

[25] E. Weinan, W. Ren, and E. Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *Journal of Chemical Physics* 126(16) (2007), 164103.

[26] M. Cameron, R. V. Kohn, and E. Vanden-Eijnden. The string method as a dynamical system. *Journal of nonlinear science* 21(2) (2011), 193–230.

[27] E. Weinan, W. Ren, and E. Vanden-Eijnden. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* 109(14) (2005), 6688–6693.

[28] S. K. Burger and W. Yang. Quadratic string method for determining the minimum-energy path based on multiobjective optimization. *The Journal of chemical physics* 124(5) (2006), 054109.

[29] B. Peters, A. Heyden, A. T. Bell, and A. Chakraborty. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The Journal of chemical physics* 120(17) (2004), 7877–7886.

[30] W. Quapp. A growing string method for the reaction pathway defined by a Newton trajectory. *The Journal of chemical physics* 122(17) (2005), 174106.

[31] A. Goodrow, A. T. Bell, and M. Head-Gordon. Development and application of a hybrid method involving interpolation and ab initio calculations for the determination of transition states. *The Journal of chemical physics* 129(17) (2008), 174109.

[32] A. Goodrow, A. T. Bell, and M. Head-Gordon. Transition state-finding strategies for use with the growing string method. *The Journal of chemical physics* 130(24) (2009), 244108.

[33] W. Quapp. The Growing String Method for Flows of Newton Trajectories by a Second-order Method. *Journal of Theoretical and Computational Chemistry* 8(01) (2009), 101–117.

[34] A. Goodrow, A. T. Bell, and M. Head-Gordon. A strategy for obtaining a more accurate transition state estimate using the growing string method. *Chemical Physics Letters* 484(4-6) (2010), 392–398.

[35] K. Muller and L. D. Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta* 53(1) (1979), 75–93.

[36] K. Ishida, K. Morokuma, and A. Komornicki. The intrinsic reaction coordinate. An abinitio calculation for HNC→ HCN and H-+ CH4→ CH4+ H-. *The Journal of Chemical Physics* 66(5) (1977), 2153–2156.

[37] H. Tang, S. Jang, M. Zhao, and S. A. Rice. On the classical theory of the rate of isomerization of HCN. *The Journal of chemical physics* 101(10) (1994), 8737–8746.

[38] T. L. Nguyen, J. H. Baraban, B. Ruscic, and J. F. Stanton. On the HCN–HNC energy difference. *The Journal of Physical Chemistry A* 119(44) (2015), 10929–10934.

[39] M. Nonella, J. Huber, and T.-K. HA. Photolytic preparation and isomerization of HNSO, HOSN, HSNO, and HONS in an argon matrix: an experimental and theoretical study. *Journal of physical chemistry (1952)* 91(20) (1987), 5203–5209.

[40] L. V. Ivanova, B. J. Anton, and Q. K. Timerghazin. On the possible biological relevance of HSNO isomers: a computational investigation. *Physical Chemistry Chemical Physics* 16(18) (2014), 8476–8486.

[41] C.-H. Lai, E. Y. Li, and P.-T. Chou. Isomerization reactions of RSNO (R= H, C n H 2n+ 1 n<4). *Theoretical Chemistry Accounts* 117(1) (2007), 145–152.

Chapter 5

Systematic Assessment on performance and robustness of *GOpt* algorithm

5.1 Abstract

A comprehensive and systematic test is conducted to assess the overall effectiveness and robustness of the newly elaborated *GOpt* optimization algorithm for transition states. The test set consists of 32 reactions of various types. A random distortion is applied to the known transition state to generate 10 random initial structures, generating distorted geometries from 0.05 a.u. to 0.4 a.u. away from the actual transition state. Then the *GOpt* algorithm and the benchmark *Berny* algorithm are tested for their ability to converge to the original transition state. For small distortions, the *GOpt* is more efficient than *Berny*. For larger distortions, *Gopt* is marginally slower. In all cases, *Gopt* converges more frequently, suggesting that it is among the best algorithms for optimizing transition states.

5.2 Introduction

The *Gopt* algorithm for optimizing chemical structure is introduced in chapter 3. *Gopt* is a quasi-Newton method, where the Hessian update includes finite-difference updates for key chemically-relevant coordinates, as well as suitable eigenvalue shifts to ensure convergence to a structure with the appropriate number of negative eigenvalues. A gradient-based trust-radius method is used to control the optimization step. The algorithm shows promising efficiency and robustness in preliminary tests. In these tests, as summarized in Table.3.1, *GOpt* converges faster and more robustly than *Berny*. However, these tests used a good initial guess from interpolation between the reactant and the product. However, a good initial guess is not always available, especially when the reaction mechanism is complicated, with multiple intermediates and TS sharing the same vicinity of the potential energy surface.

In this chapter, a more systematic and thorough assessment is conducted to explore how well *Gopt* works for poor initial guesses. As before, we use program parameters (related to the transformation between redundant internal and Cartesian coordinates, quasi-Newton update, secant conditions, and trust-radius protocols) that were optimized separately, see Chapter.6. (It is simply not practical to optimize this many parameters for a large dataset like that considered here.)

5.3 Testing protocol

A database of chemical reactions for testing

To effectively test a computational algorithm, it is important to construct a broad and relevant database. Our database includes diverse reaction types focussing on organic (Diels-Alder cycloadditions,[1] electrocyclizations,[2, 3] Huisgen cycloadditions,[4], and S_N2 nucleophilic substitutions, and radical additions, together with main-group inorganic radical addition and proton-transfer reactions.[5] In a few cases methyl groups were added to reagents to ensure that there were competing low-energy barriers (corresponding to conformational rearrangements). All quantum chemistry computations are conducted in *Gaussian 16* with HF/6-31+G.[6] The exact Hessian is only computed at the very first step; subsequently the Hessian is updated using finite differences and/or quasi-Newton Hessian updates.

5.3.1 Systematic Methods to Generating Initial Guesses of Varying Quality

To systematically generate a set of random initial guesses of varying quality, we randomly distort the true transition state. Specifically, we start with a random vector, $\mathbf{a}$, in Cartesian coordinates. The random vector is projected using the Wilson $\mathbf{B}$ matrix and its pseudo-inverse $\mathbf{B}^+$,[7] and then normalized

$$\hat{\mathbf{u}} = \frac{\mathbf{B}^+\mathbf{B}\mathbf{a}}{|\mathbf{B}^+\mathbf{B}\mathbf{a}|} \tag{5.1}$$

This produces a random perturbation excluding molecular translation and rotation. A set of random structures are generated by adding the transition state

geometry with a scaled perturbation.

$$\mathbf{x}_{guess} = \mathbf{x}_{t.s.} + \sqrt{3N} \cdot \epsilon \hat{\mathbf{u}} \tag{5.2}$$

Here N is the number of atoms in the molecule and $3N$ is the dimensionality of the Cartesian-coordinate vector. The factor ϵ contols the average amount of perturbation on each atom in the molecule. We choose to generate 10 random initial guess for each ϵ. We chose $\epsilon \in \{0.05, 0.1, 0.2, 0.3\}$. We did not consider larger values of ϵ because already when $\epsilon = 0.3$, in many cases the initial guesses were so unreasonable (e.g., near-collisions between atoms) that it was impossible to converge the Hartree-Fock calculations.

When guessing a TS, one can usually accurately predict the positions of various "spectator" atoms, and only has significant uncertainty about the values of the key internal coordinates. To test this sort of error, we generated a random vector in internal coordinates, $\mathbf{a}$, with nonzero entries only for the key internal coordinates. We then project out the redundancy due to internal coordinates,

$$\mathbf{v} = \mathbf{B}\mathbf{B}^{+}\mathbf{a}_{int} \tag{5.3}$$

The generated perturbation is applied to the target transition state structure. The transformation from the internal coordinates to Cartesian coordinates is carried out by the manifold projection method introduced as Eqn.2.19,

$$\underbrace{\min}_{\mathbf{x}_\kappa} |\mathbf{q}(\mathbf{x}_\kappa) - (\mathbf{q}_{t.s.} + \cdot\kappa\hat{\mathbf{v}})|^2 \tag{5.4}$$

where $\mathbf{q}_{t.s}$ is the redundant internal coordinates of the exact transition state and

κ is the factor used to scale the perturbation of the equilibrium structure. The distance of the perturbed Cartesian coordinates from the original transition state structure is computed in Cartesian coordinates.

$$||\mathbf{x}_\kappa - \mathbf{x}_{t.s.}|| = \epsilon \cdot \sqrt{N_{key}} \tag{5.5}$$

Where N_{key} is the number of key internal coordinates that are perturbed. In order to remove the irrelevant rotation and translation, Kabsch's Algorithm was used to align the two structures.[8] Again, We generate 10 random initial guesses for each choice of $\epsilon \in \{0.1, 0.2, 0.3, 0.4\}$. We did not consider $\epsilon > 0.4$ because such structures are extremely distorted.

5.3.2 Results and Comparison

We compare the performance of *GOpt* with the default *Berny* algorithm in *Gaussian* based on the convergence rate and the number of gradient evaluations performed. For large values of ϵ, the initial structure is very poor. Large initial steps and aggressive step-size updates can lead to severe overshooting of the transition state. Therefore, we set the initial stepsize in *GOpt* to $0.15 * \sqrt{N}$, and reduced the rate at which the trust radius is increased from 2.0 to 1.5. These adjustments slow the convergence, but decrease the rate at which the optimization fails. Even though some of the initial structures are very poor, we retained the default condition for the maximum number of steps: the optimization is considered a failure if it doesn't converge within 100 steps. Though the step limit is 100, extra gradient evaluations may occur due to the finite-difference hessian update and the recalculation of rejected steps.

To compare the performance of *GOpt* and *Berny*, we count the number of gradient evaluations required. (The gradient evaluation is the most expensive step in the geometry optimization, especially for large molecules.)

Table 5.1 shows the performance of *GOpt* and *Berny* when the positions of all atoms are perturbed. For the small displacements ($\epsilon = 0.05$ and 0.1), *GOpt* converges with fewer gradient evaluations than *Berny*. As the displacement gets larger, the quality of the test structures deteriorates and the number of gradient evaluations required dramatically increases. The number of gradient evaluations required by *Gopt* and *Berny* seem comparable, but this result is biased by the fact that *Gopt* converges from difficult initial structures (which tend to require additional gradient evaluations) where *Berny* fails. While the computational cost of *Gopt* and *Berny* are similar, *Gopt* converges much more robustly in all cases. Indeed, the failure rate for *Gopt* is 20-30% that of *Berny*. This is not so critical for good initial guesses (where *Berny* still converges for about 90% of structures), but is especially impressive for the poor initial guesses. For example, *Berny* converges for less than half the initial structures with $\epsilon = 0.3$, but *Gopt* converges for more than 80% of those structures.

The performance plot in Figure 5.3 shows the percentage of reactions that have converged after a specified number of gradient evaluations have occurred. For transition-state guess structures that converge in 20 or fewer gradient evaluations, *GOpt* and *Gauss* perform very similarly. For poor guess structures, significantly more gradient iterations are required, and *Gopt* performs significantly better than *Berny*. We do not consider what happens beyond the 100^{texth} evaluation of the gradient because by this point, the calculation has already failed or, alternatively, the optimization is converging towards an undesired transition state. The salient

features of *GOpt* are its efficiency (which is comparable to *Berny*) and its robustness (even for very poor initial guesses, *Gopt* often converges). The good performance of *Gopt* can be attributed to the emphasis it places on the key internal coordinates, which ensure that the single negative eigenvalue of the Hessian is sensibly imposed. Moreover, when the initial structure has no negative Hessian eigenvalues, *GOpt* is able to identify a sensible direction in which to take an uphill step, and thereby navigate the potential energy surface towards the right structure.

TABLE 5.1: Test results from *GOpt* and *Berny* algorithm for random perturbations of all atoms' Cartesian coordinates of specified magnitude ϵ. The number of gradient evaluations is averaged over the number of initial structures for which convergence to the desired transition state was attained by a given method. I.e., for $\epsilon = 0.3$, *Gopt* requires on average 41.7 gradient evaluations for the 80% of optimizations it converges, while *Berny* requires on average 41.0 gradient evaluations for the 43% of optimizations it converges.

Methods / ϵ(Bohr)	*GOpt* Algorithm	*Berny* Algorithm
Average Gradient Evalution		
0.05	7.5	9.7
0.10	13.7	17.2
0.20	28.2	29.0
0.30	41.7	41.0
Convergence Rate		
0.05	0.98	0.94
0.10	0.97	0.89
0.20	0.93	0.74
0.30	0.80	0.43

FIGURE 5.1: The percentage of transition-state optimizations converged by *GOpt* and *Berny* for a given number of gradient evaluations, where the initial guess structures are constructed by randomly perturbing the Cartesian coordinates of the exact transition-state structure.

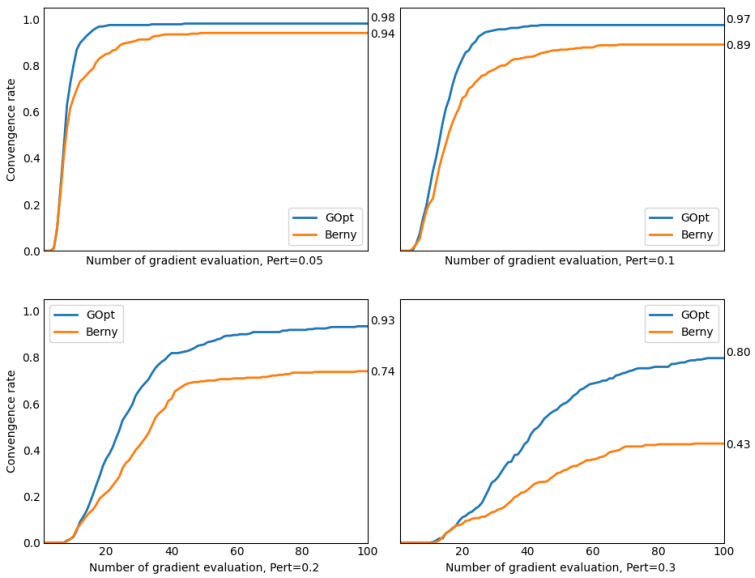
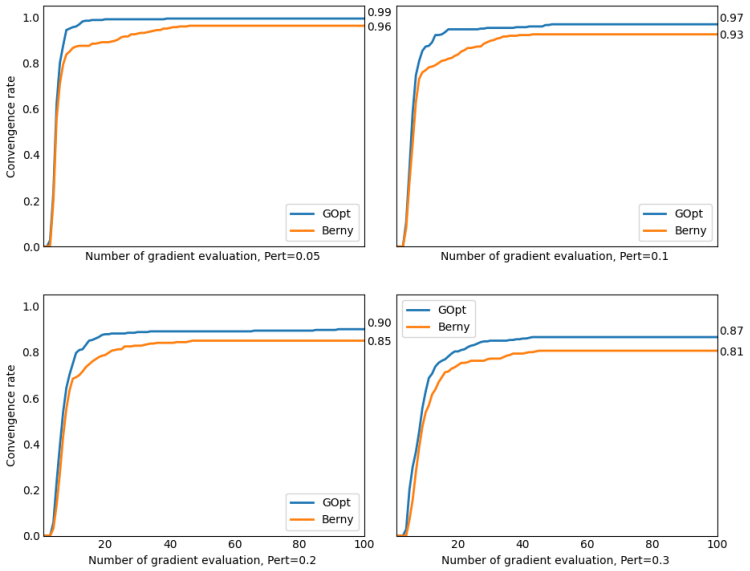


TABLE 5.2: Test results from *GOpt* and *Berny* algorithm for a random perturbation of key internal coordinates. The number of gradient evaluations is averaged over the number of initial structures for which convergence to the desired transition state was attained by a given method. I.e., for $\epsilon = 0.3$, *Gopt* requires on average 9.2 gradient evaluations for the 90% of optimizations it converges, while *Berny* requires on average 8.6 gradient evaluations for the 85% of optimizatoins it converges.

Methods / ϵ(Bohr)	*GOpt* Algorithm	*Berny* Algorithm
Average Gradient Evalution		
0.1	4.7	6.6
0.2	6.4	7.9
0.3	9.2	8.6
0.4	10.4	10.0
Convergence Rate		
0.1	0.99	0.96
0.2	0.97	0.93
0.3	0.90	0.85
0.4	0.87	0.81

FIGURE 5.2: The percentage of transition-state optimizations converged by *GOpt* and *Berny* for a given number of gradient evaluations, where the initial guess structures are constructed by randomly perturbing the key internal coordinates of the exact transition-state structure



To assess how *Gopt* and *Berny* perform for more chemically reasonable initial guess structures, we use initial structures generated by random displacements of the key internal coordinates. As seen in Table 5.2, these initial guess structures are much better, and convergence is usually attained with fewer than 10 gradient evaluations. Similar trends persist: when it converges, *Gopt* requires slightly fewer gradient evaluations than *Berny* for good initial guesses ($\epsilon = 0.1$ and 0.2), and slightly more for poorer initial guesses ($\epsilon = 0.3$ and 0.4)). However, in all cases, *Gopt* converges more initial guess structures than *Berny*. The computational cost of *Gopt* can be reduced by allowing larger initial step-sizes and expanding the trust radius more aggressively, but then slightly fewer calculations converge.

The performance plot in Fig.5.2 shows that for both *Gopt* and *Berny*, 70%

of the optimizations converge to the targeted transition state with 20 or fewer gradient evaluations. Directly comparing the results for perturbations of all atomic positions vs. only the key internal coordinates, as in Figure 5.4, indicates that for calculations that converge with 15 or fewer gradient evaluations, the algorithms are comparable. For calculations that converge more slowly, *Gopt* is clearly preferable.

5.4 Summary

We have constructed a systematic testing protocol for transition-state optimization and used it to test our recently proposed *Gopt* algorithm and compare its performance to the *Berny* algorithm that is used in the *Gaussian16* software package, and prevalent elsewhere in the literature. Our protocol is based on a database of 32 organic and main-group inorganic reactions, with known transition states at the HF/6-31+G level. Initial guesses for these transition states are generated by either randomly displacing all the atoms in the structure (Cartesian perturbations) or the key internal coordinates (internal perturbations) by a specified amount, ϵ. As expected, when the initial structure is less chemical reasonable (Cartesian perturbations) and further from the transition state (larger values of ϵ), transition-state optimization is more likely to fail and, if it converges, more evaluations of the gradient are required. While *Gopt* and *Berny* require similar numbers of gradient evaluations, *Gopt* is much more robust, converging more than 80% of the time even for chemically unreasonable ($\epsilon = 0.3$; Cartesian displacements) or deformed chemically reasonable ($\epsilon = 0.4$; key internal displacements) molecular structures. From these results, it is difficult to conceive of a situation where the *Berny* algorithm would be deemed preferable to *Gopt*, but *Berny* is certainly a capable

transition-state optimizer when an accurate, chemically sensible, transition-state guess structure is available.

FIGURE 5.3: Convergence rate for random Cartesian perturbation

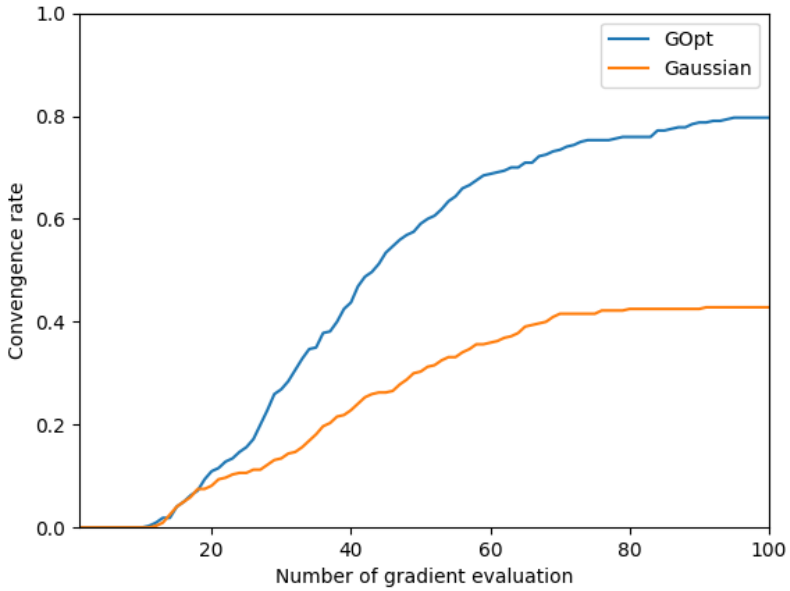
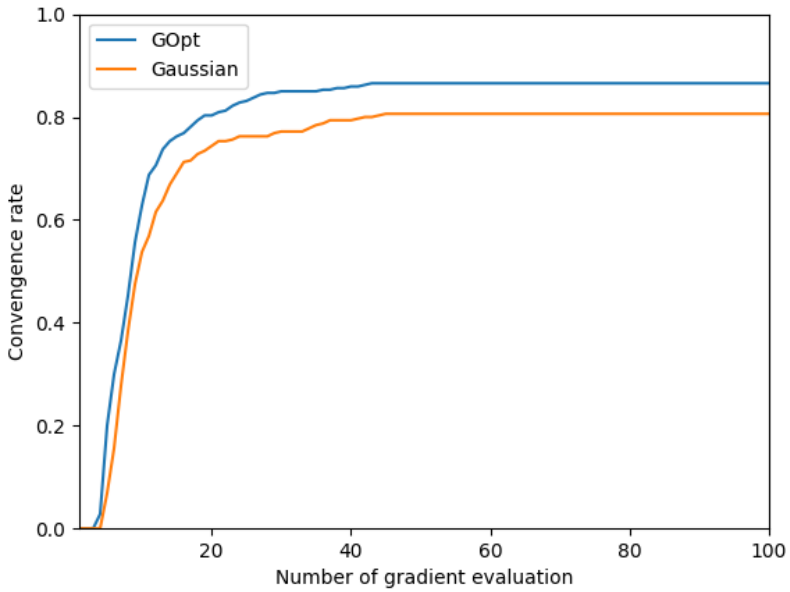


FIGURE 5.4: Convergence rate for random key internal perturbation



5.5 Reference

[1] F. Fringuelli and A. Taticchi. *The Diels-Alder reaction: selected practical methods*. John Wiley & Sons, 2002.

[2] R. B. Woodward and R. Hoffmann. Stereochemistry of electrocyclic reactions. *Journal of the American Chemical Society* 87(2) (1965), 395–397.

[3] H. Longuet-Higgins and E. Abrahamson. The electronic mechanism of electrocyclic reactions. *Journal of the American Chemical Society* 87(9) (1965), 2045–2046.

[4] R. Huisgen, G. Mloston, and E. Langhals. The first two-step 1, 3-dipolar cycloadditions: non-stereospecificity. *Journal of the American Chemical Society* 108(20) (1986), 6401–6402.

[5] J. Clayden. *Organic chemistry*. 2001.

[6] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, G. Petersson, H. Nakatsuji, et al. *Gaussian 16*. 2016.

[7] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1980.

[8] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 34(5) (1978), 827–828.

Chapter 6

Summary and Future Work

6.1 Abstract

In the preceding chapters, we have established that *GOpt* is a robust and efficient optimization package. Here we review the key ideas and features of *Gopt* and discuss a few key program options which were included in the package for versatility and flexibility. Some preliminary tests of these options are presented, using a small set of 20 reactions. Specifically, we test different secant conditions, different quasi-Newton methods, and different trust-radius updates. The BFGS quasi-Newton method with the energy-based trust-radius update is the best default choice for optimizing to a minimum on the potential energy surface, while the Bofill quasi-Newton method with the gradient-based trust-radius update is the optimal choice for transition state optimization.

6.2 Introduction

In this thesis, we have presented the key features of the *Gopt* software package, which is conceived as a general-purpose package for molecular structure optimization and reaction-path finding. The first key component of *Gopt* is the use of redundant internal coordinates. While redundant internal coordinates are preferable for geometry optimization to Cartesian coordinates or nonredundant (e.g., Z-matrix) internal coordinates, there are issues associated with internal coordinates in general (e.g., ill-defined dihedral angles if one of the bonds rotating is nearly collinear with the central bond about which it rotates), and redundant internal coordinates in particular (the difficulty of mapping physically unrealizable redundant internal coordinates to a realizable molecular structure in Cartesian coordinates). Chapter 2 shows how to overcome these obstacles by defining a robust set of redundant internal coordinates (including interfragment bonds, special interatomic distances for long linear chains, and robust dihedral coordinates that do not fail for torsions around collinear bonds). In addition, a manifold projection strategy, wherein a specified set of redundant internal coordinates is mapped to the closest realizable molecular structure, solves the issue of nonphysical redundant internal coordinates.

Using the developments in chapter 2, we noted that a good initial guess for transition-state optimizations is often obtained by averaging the reactant and product structures in redundant internal coordinates. Where this initial guess is inadequate, it can be refined by constrained minimization in the hyperplane that is orthogonal to the reaction-path-vector, $\mathbf{q}_{\text{path}} = \mathbf{q}_{\text{product}} - \mathbf{q}_{\text{reactant}}$. The resulting structure is guaranteed to be on a minimum-energy pathway on the molecular

potential energy surface. If this structure is still a poor guess for the transition state, the procedure can be iterated by adding, and optimizing, additional points between the structures of the reactant and the transition-state guess, and between the transition-state guess and the product. By iteratively optimizing structures on the bisecting hyperplane between previously determined structures along the reaction path, a minimum energy path that connects reactants and products can be attained. This bisecting hyperplane method is tested in Chapter 4, and observed to overcome many of the problematic issues associated with other two-ended approaches for constructing chemical reaction pathways.

While *Gopt* includes utilities for redundant internal coordinates and methods for transition-state guessing and reaction-path finding, the primary use-case of *Gopt* is geometry optimization. *Gopt* is designed to be user-friendly, including both sensible default parameters but also the flexibility to easily adjust these defaults and interface *Gopt* to other software packages.

The primary innovation in *Gopt* is the identification, either automatically or by the user, of certain *key* coordinates, which typically correspond to the bonds that fracture/form and the angles that open/close during the reaction. By identifying these coordinates, ensuring that the Hessian matrix for these coordinates is accurate (using finite-differences where necessary), and ensuring that the negative curvature direction features these coordinates (using explicit shifts of Hessian eigenvalues where necessary), *Gopt* converges with comparable computational cost to the commonly used *Berny* algorithm that is included as the default in *Gaussian16* and commonly mimicked elsewhere. However, *Gopt* is significantly more robust than the *Berny* method. For example, in about half the cases where the

Berny algorithm fails to converge, *Gopt* successfully locates the targeted transition state. The details of the *Gopt* algorithm are described in Chapter 3, with systematic tests in Chapter 4 establishing its low computational cost and high robustness, at least compared to the *Berny* algorithm.

The remainder of this chapter presents some of our preliminary data on selecting appropriate default parameters for the *Gopt* program. Specifically, we will explore different quasi-Newton updates, trust-radius update methods, and secant conditions.

6.3 Testing Protocol

20 chemical reactions with 6 different mechanisms were explored using redundant internal coodinates. The initial guess structure was generated using the protocol in chapter 2. Specifically, we averaged the reduced internal coordinate representation of the reactant and the product,

$$\mathbf{x}^{init} = \arg\underbrace{\min}_{\mathbf{x}} \left\| \mathbf{q}(\mathbf{x}) - \left[\tfrac{1}{2}\mathbf{q}^{(reactant)} + \tfrac{1}{2}\mathbf{q}^{(product)} \right] \right\| \tag{6.1}$$

Where $\mathbf{x}$ and $\mathbf{q}$ denote the system in Cartesian coordinates and in reduced internal coordinates, respectively. All calculations were performed using *Gaussain16* at the HF/6-31+G level, with the initial Hessian computed exactly. Calculations that did not converge to the targetted transition state within 100 iterations are considered to be convergence failures.

6.3.1 Secant Condition

Because computing the Hessian exactly is time-consuming, it is preferable to use a quasi-Newton method, wherein the Hessian is updated after each iteration based on the computed gradient. Different quasi-Newton methods update the Hessian in different ways, using differing amounts of information from previous iterations. The key idea is that information about the Hessian can be deduced from the change in the gradient,

$$\mathbf{H}_v \delta\mathbf{v} \approx \mathbf{g}_v(\mathbf{v} + \delta\mathbf{v}) - \mathbf{g}_v(\mathbf{v}) \tag{6.2}$$

Here $\mathbf{v}$ is the molecular structure represented in the space of (nonredundant) delocalized internal coordinates. However, because this coordinate system *changes* as one moves along the manifold of physically realizable molecular structures, the secant condition is not unique. Different secant conditions are derived depending on how one treats the changes in $\mathbf{V}$ with molecular geometry.

One strategy is to start with the secant condition in Cartesian coordinates,

$$\mathbf{H}_x^{old} \delta\mathbf{x} = \delta\mathbf{g}_x \tag{6.3}$$

Using the chain rule of derivatives with transformation matrix $\mathbf{B}$ and $\mathbf{V}$, one obtains:

$$\mathbf{H}_v^{old} \delta\mathbf{v} \approx \delta\mathbf{g}_v - (\mathbf{V}^{old})^T((\mathbf{B}^{old})^T)^+\Big((\mathbf{B}^{old})^T \delta\mathbf{V}\mathbf{g}_v^{old} + (\delta\mathbf{B})^T\mathbf{g}_q^{old}\Big) \tag{6.4}$$

Another alternative is to directly evaluate the expression $\mathbf{H}^{old}\delta\mathbf{v}$,

$$\mathbf{H}_v^{old} \delta\mathbf{v} \approx \delta\mathbf{g}_v + (\mathbf{V}^{\mathbf{old}^T})(\delta\mathbf{B}^+)^T\mathbf{g}_x^{old} + (\delta\mathbf{V}^T)\mathbf{g}_q^{old} \tag{6.5}$$

When the local energy quadratic approximation is accurate, the gradient change is the same when taking a step $\delta\mathbf{v}$ at $\mathbf{H}^{old}$ as taking a step $-\delta\mathbf{v}$ at $\mathbf{H}^{new}$.

$$\mathbf{H}^{old}\cdot\delta\mathbf{v}=\mathbf{H}^{new}\cdot-\delta\mathbf{v} \tag{6.6}$$

Eqn.6.4 and 6.5 do not maintain the symmetry. However, we can enforce this condition by averaging the "old" and "new" versions of these equations. For example, corresponding to Eq. 6.5

$$\mathbf{H}\delta\mathbf{v}=\delta\mathbf{g}_v+\frac{1}{2}\Big((\mathbf{V}^{old})^T(\delta\mathbf{B}^+)^T\mathbf{g}_x^{old}+(\mathbf{V}^{new})^T(\delta\mathbf{B}^+)^T\mathbf{g}_x^{new}+(\delta\mathbf{V}^T)(\mathbf{g}_q^{old}+\mathbf{g}_q^{new})\Big) \tag{6.7}$$

6.3.2 Quasi-Newton Update

The second condition is used to update the Hessian matrix in quasi-Newton methods. Though we intend to add addition Hessian-update formulas to *Gopt*, for now we hae four choices, chosen due to their ubiquity in molecular optimization: simple-rank one (SR1),[1] Powell symmetric Broyden (PSB),[2] Broyden, Fletcher, Goldfarb, Shanno (BFGS),[3–7] and Bofill's 1994 update.[8] The SR1 update is slightly modified from its usual form to avoid divide-by-zero errors:

Symmetric-Rank-One update (SR1)

$$\mathbf{H_v}^{k+1}=\begin{cases}\mathbf{H}_v^{(k)} & \frac{\left\|(\mathbf{y}_v^{(k)}-\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})\cdot\mathbf{s}_v^{(k)}\right\|}{\left\|\mathbf{y}_v^{(k)}-\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}\right\|\cdot\left\|\mathbf{s}_v^{(k)}\right\|})\leq 1e^{-9}\\ \mathbf{H}_v^{(k)}+\frac{(\mathbf{y}_v^{(k)}-\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})(\mathbf{y}_v^{(k)}-\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})^T}{(\mathbf{y}_v^{(k)}-\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})\cdot\mathbf{s}_v^{(k)}} & Otherwise\end{cases} \tag{6.8}$$

Powell-symmetric-Broyden update (PSB)

$$\begin{aligned}\mathbf{H_v}^{k+1} = \mathbf{H}_v^{(k)} &+ \frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})(\mathbf{s}_v^{(k)})^T + \mathbf{s}_v^{(k)}(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^T\mathbf{s}_v^{(k)}} \\ &- \left(\frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})^T(\mathbf{s}_v^{(k)})}{(\mathbf{s}_v^{(k)})^T\mathbf{s}_v^{(k)}}\right)\mathbf{s}_v^{(k)}(\mathbf{s}_v^{(k)})^T\end{aligned} \tag{6.9}$$

Broyden-Fletcher-Goldfarb-Shanno update (BFGS)

$$\mathbf{H_v}^{k+1} = \mathbf{H}_v^{(k)} + \frac{\mathbf{y}_v^{(k)}(\mathbf{y}_v^{(k)})^T}{(\mathbf{y}_v^{(k)})^T s_v^{(k)}} - \frac{(\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})(\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^t\mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}} \tag{6.10}$$

Bofill's 1994 update (Bofill)

$$\mathbf{H}_{Bofill}^{(k+1)} = (1-\psi)\mathbf{H}_{SR1}^{(k+1)} + \psi\mathbf{H}_{PSB}^{(k+1} \tag{6.11}$$

$$\psi = 1 - \frac{|\mathbf{s}_v^{(k)} \cdot \mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2}{|\mathbf{s}_v^{(k)}|^2|\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2} \tag{6.12}$$

$$= \frac{|\mathbf{s}_v^{(k)} \times \mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2}{|\mathbf{s}_v^{(k)}|^2|\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)}\mathbf{s}_v^{(k)}|^2} \tag{6.13}$$

6.3.3 Trust Radius Update

At each iteration, a quadratic model for the potential energy surface is constructed using the quasi-Newton Hessian, with possible modifications to ensure that the eigenstructure of the Hessian is appropriate for converging to a saddle point (with exactly one negative-curvature direction associated with the key chemical coordinates) or to a minimum (with all positive-curvature directions). If the quadratic model were perfect, then the optimization step $\mathbf{s} = \mathbf{H}^{-1}\mathbf{g}$ would lead exactly to the stationary point. However, when the higher-order terms in the Taylor series expansion of the potential energy surface are significant, or when the quasi-Newton

approximate Hessian is inaccurate, the predicted optimization step may be inaccurate. As a sufficiently short optimization step (in the direction indicated by the gradient) is always beneficial, it seems sensible to optimize the quadratic model subject to the constraint that the solution lies within a trust radius where we believe the quadratic model will be accurate, $\|s\| \leq \tau$.[9, 10] Based on the accuracy of the quadratic model, the trust radius is adjusted after each step.

Specifically, the trust radius is updated based on the difference between the computed properties of the next iterate and the properties predicted for the next iterate by the quadratic model. In *GOpt*, we implemented two general types of trust radius update schemes: an energy-based method that is targeted primarily for structure minimization, and a gradient-based method that is targeted primarily towards transition states.

6.3.4 Energy-Based Trust Radius Update

In the energy-based method, accuracy of the quadratic approximation is measured by the predicted energy difference. The predicted energy change is calculated

$$\Delta m_{k+1} = \mathbf{g}_k \cdot \mathbf{s}_k + \frac{1}{2}\mathbf{s}_k^T \mathbf{H}_k \mathbf{s}_k \tag{6.14}$$

and compared to the real energy change is $\Delta U_{k+1} = U(\mathbf{x}_{k+1}) - U(\mathbf{x}_k)$. If the predicted energy change is accurate enough,

$$\frac{2}{3} < \frac{\Delta m_k}{\Delta U_K} < \frac{3}{2} \tag{6.15}$$

then the trust radius is doubled, $\tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max}))$, where τ_{min} and τ_{max} are minimum and maximum trust radii based on the characteristic length scale of oscillations on the potential energy surface. Similarly, if the accuracy of the quadratic model is moderate, the trust radius is retained. I.e., if

$$\frac{1}{3} < \frac{\Delta m_k}{\Delta U_K} < 3 \tag{6.16}$$

then, $\tau_{new} = \max(\tau_{old}, \tau_{min})$. Otherwise the quadratic model is inaccurate and the trust-radius is reduced by a factor of four, $\tau_{new} = \min(\frac{1}{4}\tau_{old}, \tau_{min})$

Gradient-Based Trust Radius Update

When searching for a stationary point rather than a minimum, the reducation of the gradient is a better indicator for the progress of the optimization. In the gradient-based scheme, the trust radius is updated based on the latest gradient and the one of previous step. When adjusting the trust radius, both the magnitude and the direction of the gradient vectors are considered,

$$\rho = \frac{\left\|\mathbf{g}_{predict}^{k+1}\right\| - \left\|\mathbf{g}^{k}\right\|}{\|\mathbf{g}^{k+1}\| - \|\mathbf{g}^{k}\|} \tag{6.17}$$

$$\cos(\theta) = \frac{(\mathbf{g}_{predict}^{k+1} - \mathbf{g}^{k}) \cdot (\mathbf{g}^{k+1} - \mathbf{g}^{k})}{\left\|\mathbf{g}_{predict}^{k+1} - \mathbf{g}^{k}\right\| \|\mathbf{g}^{k+1} - \mathbf{g}^{k}\|} \tag{6.18}$$

As dimension increases, the chance of two gradient vectors aligned in the same direction decreases. To treat high-dimensional systems appropriately, the function $p_x(d)$ is defined so that $x\%$ of d-dimensional vectors have a value of $\cos(\theta)$ greater

than $p_x(d)$. In the gradient-based trust-radius update, we use:

$$\cos(\theta) \geq p_{10}(d) \approx \sqrt{\frac{1.6424}{d} + \frac{1.11}{d^2}} \tag{6.19}$$

$$\cos(\theta) \geq p_{40}(d) \approx \sqrt{\frac{0.064175}{d} + \frac{0.0946}{d^2}} \tag{6.20}$$

The trust radius is then increased if the magnitude and the direction of the predicted gradient are accurate. I.e., if

$$\frac{4}{5} < \rho < \frac{5}{4} \tag{6.21}$$

$$p_{10}(3N-6) < \cos(\theta) \tag{6.22}$$

then $\tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max})$. Similarly, for moderately accurate predicted gradients, we retain the trust radius,

$$\frac{1}{5} < \rho < 6 \tag{6.23}$$

$$p_{40}(3N-6) < \cos(\theta) \tag{6.24}$$

then $\tau_{new} = \max(\tau_{old}, \tau_{min})$, Otherwise, the predicted gradient is inaccurate and we reduce the trust radius by a factor of two, $\tau_{new} = \min(\frac{1}{2}, \tau_{min})$.

6.4 Results and Discussion

Tables 6.1 - 6.3 compare the aforementioned program options. The results are nearly invariant with respect to the choice of secant condition (see Table 6.1),

which is reasonable since all the secant conditions are consistent with a quadratic model, and only differ at higher orders. We are already performing tests to see if when the initial guess structure is poorer, and therefore step sizes are larger, it may be possible to make meaningful distinctions between the quality of the various secant conditions.

By contrast, different quasi-Newton methods perform quite differently (see Table 6.2. It is not very surprising that BFGS does not perform well for saddle-point optimizations, nor is the significant advantage of the symmetric-rank-2 update (PSB) over a symmetric-rank-1 update (SR1) unexpected. The Bofill update (which combines PSB and SR1 in an advantageous way) works best of all. It would be interesting to see whether some of the more recent update formulas from Bofill's group might perform even better.

Finally, both trust radius updates perform well, though the gradient-based method is slightly better. This may be because we are testing with quite good initial guesses, so that the final structure already lies near, if not inside, the initial trust radius. We plan to revisit this result for a larger set of reactions, with more inaccurate initial guesses, to see if the results change.

These studies support the use of the first secant condition, the Bofill quasi-Newton method, and the gradient-based trust radius that we used, by default, elsewhere in this thesis. They also reinforce that not only is *Gopt* highly efficient (with comparable computational cost to state-of-the-art methods in the literature) and robust to poor initial guess structures (converging much more frequently than the best readily available molecular optimization packages), *Gopt* is also robust with respect to program options. That is, changing algorithmic parameters in *Gopt* rarely changes the performance much, and even when a remarkably poor choice is

made (e.g., using the BFGS quasi-Newton method for transition-state optimization), the method still converges without too many iterations. This helps us feel confident that the *Gopt* software should be widely used for molecular structure optimization and reaction-path finding.

TABLE 6.1: Test results for different secant conditions

	Secant 1 6.4	Secant 2 6.5	Secant 3 6.6
Converge Rate	100%	100%	100%
Steps	9.15	9.45	9.6
Gradient Eval.	10.9	11.35	11.4

The number of steps and gradient evaluations needed to achieve convergence from the same *GOpt* algorithm but different secant conditions.

TABLE 6.2: Test results for different quasi-Newton update methods

	BFGS	SR1	PSB	Bofill
Converge Rate	100%	100 %	100%	100%
Steps	15.35	12.65	9.9	9.15
Gradient Eval.	24.85	19.75	11.55	10.9

The number of steps and gradient evaluations needed with the same *GOpt* algorithm but different Quasi-Newton methods.

TABLE 6.3: Test results for different trust-radius update methods

	Energy-Based	Gradient-Based
Converge Rate	100%	100%
Steps	9.35	9.15
Gradient	11.65	10.9

The number of steps and gradient evaluations needed with the same *GOpt* algorithm but different trust-radius update schemes.

6.5 Reference

[1] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[2] M. J. Powell. A hybrid method for nonlinear equations. *Numerical methods for nonlinear algebraic equations* (1970).

[3] R. Fletcher. A new approach to variable metric algorithms. *The computer journal* 13(3) (1970), 317–322.

[4] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation* 24(109) (1970), 23–26.

[5] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics* 6(1) (1970), 76–90.

[6] D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation* 24(111) (1970), 647–656.

[7] D. F. Shanno and P. C. Kettler. Optimal conditioning of quasi-Newton methods. *Mathematics of Computation* 24(111) (1970), 657–664.

[8] J. M. Bofill. Updated Hessian matrix and the restricted step method for locating transition structures. *Journal of Computational Chemistry* 15(1) (1994), 1–11.

[9] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.

[10] R. Fletcher. Practical methods of optimization john wiley & sons. *New York* 80 (1987), 4.