



HAL
open science

New transition state optimization and reaction path finding algorithm with reduced internal coordinates

Xiaotian Yang

► **To cite this version:**

Xiaotian Yang. New transition state optimization and reaction path finding algorithm with reduced internal coordinates. Theoretical and/or physical chemistry. Sorbonne Université; McMaster university (Hamilton, Canada), 2021. English. NNT : 2021SORUS481 . tel-03824918

HAL Id: tel-03824918

<https://theses.hal.science/tel-03824918>

Submitted on 21 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**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 © Copyright by Xiaotian YANG December 3,
2020

Sorbonne Université © Copyright by Xiaotian YANG
December 3, 2020

McMaster University

Doctor of Philosophy (2020)

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: xiii, 171

Sorbonne Université

Doctor of Philosophy (2020)

Paris, France (Laboratoire de Chimie Théorique)

TITRE: New Transition State Optimization and Reaction Path Finding Algorithm
with Reduced internal coordinates

AUTEURE: Xiaotian YANG (Sorbonne Université)

SUPERVISEUSE: Dr. Julia CONTRERAS-GARCÍA

NOMBRE DE PAGES: xiii, 171

Abstract

Geometry optimization is a fundamental step in 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 reaction and product, which are local minima on the potential energy surface, 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 reliable algorithm and high accuracy. One can follow the gradient descent direction to pursuit the local minimum until convergence is reached. But for transition state, Either up-hill or down-hill direction allowed in the process makes the determination more challenging.

Motivated by the difficulty of the obstacle, 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 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 are introduced to effectively represent the molecular structures, and 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 problem. In Chapter 4, A novel algorithm to generating reaction path based on redundant internal coordinates is illustrated with real chemical reactions. Last but no least,

in Chapter 6, a systematic exploration between different methods available in the optimization is conducted. A well-performed combination of optimization methods is drawn for generic optimization purpose.

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

Acknowledgements

Ph.D. is long journey, so is life. The day I finish my defence gonna mark an end for this 6-year experience, but also gonna initiate a new chapter. "life's like a box of chocolates. You never know what you're gonna get." Hereby, please allow me to thank Dr. Ayers to guide me in the last 6 years to become a better person, then a better researcher. Also great appreciation to Dr. Contreras-garcia. She is my supervisor, mentor, and also close friend. What I learnt the most is not only the professional insights, but also the optimistic and enthusiastic attitude towards life.

I also like to thank my committee members, lab-mates, parents, friends, etc. Too many details I couldn't disclose in my thesis. So, please let me recap all my gratitude. Thank you all for the support.

Contents

Abstract	iii
Acknowledgements	v
Declaration of Authorship	xiii
1 Introduction	1
1.1 Introduction	1
1.2 The Potential Energy Surface	4
1.2.1 The Born-Oppenheimer Approximation	4
1.2.2 Numerical Calculations on Potential Energy Surfaces	9
1.3 Coordinate System	10
1.3.1 Cartesian Coordinates	10
1.3.2 Internal Coordinates	11
1.3.3 Transformation between Cartesian coordinates and redundant internal coordinates	14
1.4 Numerical Methods for Optimization	17
1.4.1 Quasi-Newton method	19
1.5 Iteration strategies	22
1.5.1 Line Search method	23
1.5.2 Trust-Region Methods	25

1.6	Transition State Optimization	28
1.6.1	Single-Ended Method	29
1.6.2	Double-Ended Methods	32
1.7	Summary	36
1.8	References	39
2	Generating Initial Guesses for Transition States with Redundant Internal Coordinates and Robust Dihedrals	52
2.1	Abstract	52
2.2	Introduction	53
2.3	Methodology	55
2.3.1	Normal Redundant Internal Coordinates	55
2.3.2	Robust Redundant Internal Coordinates	57
2.3.3	Mapping between Internal coordinates and Cartesian coor- dinates	61
2.4	Results and Discussion	65
2.4.1	Testing Protocol	65
2.4.2	Overview of result	65
2.4.3	Result and Discussion	66
2.5	Conclusion	69
2.6	Reference	73
3	A Robust Algorithm for geometry optimization and transition state search with Reduced Internal Coordinates	77
3.1	Abstract	77
3.2	Introduction	78

3.3	Methodology	81
3.3.1	Overview	81
3.3.2	Selection of redundant internal coordinates	83
3.3.3	Coordinate transformations	84
3.3.4	Transformation between Reduced Internal and Cartesian	86
3.3.5	Select key internal coordinates in optimization	87
3.3.6	Construct delocalized reduced internal coordinates	88
3.3.7	The secant condition in reduced coordinates	92
3.3.8	Quasi-Newton Updates	93
3.3.9	Hessian Finite Differences Update	95
3.3.10	Hessian Modification	96
3.3.11	Step Size Control	98
3.3.12	Trust Radius Determination	100
3.3.13	Convergence Criterion	105
3.3.14	Summary of the Algorithm	105
3.4	Results and Discussion	107
3.4.1	Testing Protocol	107
3.4.2	<i>GOpt</i> default methods	107
3.4.3	Comparison with <i>Berny</i> Algorithm	107
3.4.4	Results and Discussion	108
3.5	Summary	110
3.6	Reference	111
4	Bisect hyperplane optimization algorithm for reaction path finding	117

4.1	Abstract	117
4.2	Introduction	118
4.3	Methodology	121
4.3.1	Coordinates system	121
4.3.2	Method overview	124
4.4	Examples and Cases	131
4.4.1	Muller-Brown Potential	131
4.4.2	Chemical Reactions	134
4.5	Conclusion	137
4.6	Reference	140
5	Systematic Assessment on performance and robustness of <i>GOpt</i> algorithm	146
5.1	Abstract	146
5.2	Introduction	147
5.3	Testing protocol	148
5.3.1	A Systematic Method for Generating Initial Guess	148
5.3.2	Criterion on Test assessment	151
5.3.3	Results and Comparison	151
5.4	Summary	155
5.5	Reference	158
6	Systematic Assessment on components on the optimization algorithm	159
6.1	Abstract	159
6.2	Introduction	160

6.3	Testing Protocol	161
6.3.1	Secant Condition	161
6.3.2	Quasi-Newton Update	163
6.3.3	Trust Radius Update	164
6.3.4	Energy-Based Trust Radius Update	165
6.3.5	Testing Methods	167
6.4	Results and Discussion	167
6.5	Reference	171

List of Figures

2.1	Illustration of internal coordinates consist of 4 atoms	62
4.1	Reaction path finding on Muller-Brown surface with Bisection method	131
4.2	Path points generated by Bisection algorithm	135
4.3	Energy curve of reaction HCN -> CNH along the reaction coordinates	135
4.4	HSNO -> HONS isomerization mechanism	138
4.5	HSNO -> HONS Energy vs Reaction process	138
5.1	Convergence rate for random Cartesian perturbation	157
5.2	Convergence rate for random key internal perturbation	157

List of Tables

2.1	Number of iterations and negative eigenvalues for generated guess structures	71
2.2	Average performance of conventional dihedrals and robust dihedrals	72
2.3	Optimization iteration needed after removing ill-defined conventional dihedrals	72
3.1	Number of gradient evaluation needed for transition state optimization	109
4.1	Analytical and interpolated stationary points from the reaction paths	133
4.2	Transition state from interpolation and analytic Computation for HCN -> CNH reaction	136
4.3	Transition state from interpolation and analytic Computation for HSNO -> HONS reaction	137
5.1	Test results from <i>GOpt</i> and <i>Berny</i> algorithm when applied random perturbation in Cartesian coordinates	153
5.2	Test results from <i>GOpt</i> and <i>Berny</i> algorithm when applied random perturbation in key internal coordinates	154
6.1	Test results for different secant conditions	170
6.2	Test results for different quasi-Newton update methods	170
6.3	Test results for different trust-radius update methods	170

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 connecting 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 a these algorithms are affected by many factors, among them the choice of 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 Schrödinger equation,

$$\mathcal{H}|\Psi\rangle = \mathcal{E}|\Psi\rangle \quad (1.1)$$

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

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

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} ,

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

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

$$\mathcal{H}_{elec} \Psi_{elec} = \mathcal{E}_{elec} \Psi_{elec} \quad (1.4)$$

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

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

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

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

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

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 Schrödinger 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 nuclei 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 which 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 1st 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.

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:

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

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,

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

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, $\mathbf{H}(\mathbf{x})$,

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

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]

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

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, analytic computation of the Hessian is significantly more expensive than 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 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 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 position 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 not 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 non-redundant, 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 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 \mathbf{B} matrix, which is the Jacobian of the transformation from Cartesian to internal coordinates, with elements,^[32]

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \quad (1.13)$$

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

$$\delta \mathbf{q} = \mathbf{B} \cdot \delta \mathbf{x} \quad (1.14)$$

For most molecules with more than a few atoms, the number of redundant internal coordinates is far greater than $3M$, so \mathbf{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

$$\mathbf{B}^+ \cdot \delta \mathbf{q} = \delta \mathbf{x} \quad (1.15)$$

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:

$$\mathbf{g}_x = \mathbf{B}^T \mathbf{g}_q \quad (1.16)$$

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

$$\mathbf{H}_x = \mathbf{B}^T \mathbf{H}_q \mathbf{B} + \mathbf{K} \quad (1.18)$$

$$\mathbf{H}_q = (\mathbf{B}^T)^+ (\mathbf{H}_x - \mathbf{K}) \mathbf{B}^+ \quad (1.19)$$

where \mathbf{K} is the matrix including the second derivatives of the internal coordinates with respect to Cartesian coordinates,

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

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

$$\mathbf{s}_0 = \mathbf{q}_{target} - \mathbf{q}_0 \quad (1.21)$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{B}^+ \mathbf{s}_0 \quad (1.22)$$

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

$$\mathbf{s}_1 = \Delta \mathbf{q} = \mathbf{q}_{target} - \mathbf{q}_1 \quad (1.24)$$

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

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

$$\mathbf{s}_k = \mathbf{q}_{target} - \mathbf{q}_k \quad (1.26)$$

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 aim 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} ,

$$\underbrace{\min}_x \left(\mathbf{q}_{target} - \mathbf{q}(\mathbf{x}) \right)^T \mathbf{W} \left(\mathbf{q}_{target} - \mathbf{q}(\mathbf{x}) \right) \quad (1.27)$$

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 nearby energy contour can be estimated through Taylor expansion

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

Due to the limitation of computation power, high order derivatives of energy versus coordinates are not regularly available. This expansion is normally approximated with only the first and second-order derivatives. The same Taylor expansion can also be applied to the gradient at \mathbf{x}_0 .

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}_0 + \mathbf{H}_0(\mathbf{x} - \mathbf{x}_0) + \dots \quad (1.29)$$

Truncating any higher order derivative after Hessian matrix, Eqn.1.29 becomes

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

where $\Delta\mathbf{x}$ is the step defined as $\mathbf{x} - \mathbf{x}_0$ under given coordinate system.

Points of interest like reaction, product, or transition state are all stationary points on the potential energy surface with zero gradient, $\mathbf{g}(\mathbf{x}) = 0$.

$$\mathbf{0} = \mathbf{g}_x = \mathbf{g}_0 + \mathbf{H}_0\Delta x \quad (1.31)$$

Reform the above equation, one can obtain the formula for optimization step

$$\mathbf{H}_0\Delta\mathbf{x} = -\mathbf{g}_0 \quad (1.32)$$

$$\Delta\mathbf{x} = -\mathbf{H}_0^{-1}\mathbf{g}_0 \quad (1.33)$$

The step $\Delta\mathbf{x}$ in Eqn.1.33 is 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 potential energy is exact as the quadratic approximation where all the higher-order derivatives are zero, the gradient of $\mathbf{g}(x_1)$ will be exact zeros and the optimization is finished. But this is not the case in most scenarios. When the quadratic model is not exact, one needs to repeat the procedures until convergence is reached. This iterative process of solving the linear equation is called *Newton-Raphson* method.

Newton-Raphson method is based on the hypothesis that the initial point is on a contour which can be approximated by a paraboloid. It's an efficient way to achieve optimization goal when the high order derivative is small. It's the fundamental basis for other more sophisticated numerical optimization algorithm.[33–36] However, if the step $\Delta\mathbf{x}$ is too big, the surface does not follow the quadratic approximation in the range. The high order derivatives omitted in 1.33 become crucial. The Newton step from 1.33 is no longer reliable. Certain constraints

need to be applied to ensure the step does not reach out to the region where the quadratic approximation is inaccurate. One of the 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 optimization $\mathbf{x}_i = \mathbf{x}_{i-1} + \alpha\Delta\mathbf{x}$, where α is the parameter to be optimized.

1.4.1 Quasi-Newton method

Newton’s method is accurate and efficient when the quadratic approximation is accurate on the potential energy surface. It also requires the analytic computation of the Hessian matrix. Unlike energy and gradient evaluation, computation of the Hessian matrix is a time-consuming process. It’s computationally prohibitive for a large molecule to take Newton’s step at every optimization iteration. Motivated by the limitation, quasi-Newton methods are brought forward as a more efficient substitute. The biggest improvement of the quasi-Newton method is it uses the gradient(s) calculated from previous successive iterations(s) to approximate the Hessian matrix of the new configuration. Because gradient calculation is at the comparable cost as the energy calculation, the quasi-Newton method can successfully replace the tedious Hessian evaluation in the optimization process.

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 a pure identity matrix with a scaling factor but it may lead to slow convergence

or even failures due to the lack of information regarding the potential surface curvature.[44]

After each iteration, a new step \mathbf{s}_n is taken place. 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+1} through a standard computation. The updated Hessian needs to ensure the gradient changes match the step taken under the quadratic approximation for local potential.

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

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

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

Secant condition \mathbf{y} is the key to many different quasi-newton update schemes. Various methods adapted \mathbf{y} for a more flexible version.

Broyden-Fletcher-Goldfarb-Shanno (BFGS)

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

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

One of the great features of the BFGS method is it preserves the positive-definite nature of the original Hessian matrix, given it is positive-definite as well. All the positive eigenvalues ensure the update step is always towards the energy decreasing direction. This feature makes BFGS the ideal choice for minimization, such as

locating reaction and product, but not effective for transition state where uphill steps may be taken during the process.

Symmetric-Rank-1 (SR1)

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

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

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 problem would be raised due to division over 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 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]

$$\begin{aligned} \mathbf{H}^{new} = \mathbf{H}^{old} + & \frac{(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x})\Delta\mathbf{x}^T + \Delta\mathbf{x}(\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x})^T}{\|\Delta\mathbf{x}\|^2} \\ & - \left(\frac{\Delta\mathbf{x} \cdot (\mathbf{y} - \mathbf{H}^{old}\Delta\mathbf{x})}{\|\Delta\mathbf{x}\|^4} \right) \Delta\mathbf{x}\mathbf{x}^T \end{aligned} \quad (1.38)$$

Bofill

Both RS1 and PSB are proper candidates for geometry optimization. Bofill proposed a mixed method combining SR1 and PSB methods.[52]

$$\begin{aligned} \mathbf{H}^{Bofill} &= \phi \mathbf{H}^{SR1} + (1 - \phi) \mathbf{H}^{PSB} \\ \phi &= \frac{\|(\mathbf{y} - \mathbf{H}^{old} \Delta \mathbf{x})^T \Delta \mathbf{x}\|^2}{\|\mathbf{y} - \mathbf{H}^{old} \Delta \mathbf{x}\|^2 \|\Delta \mathbf{x}\|^2} \end{aligned} \quad (1.39)$$

The Bofill method takes a linear combination of the two methods. Combining the advantages from each to form a great candidate for transition state optimization.

The Hessian matrix for transition state needs to have exact one negative eigenvalue. It’s corresponding eigenvector is tangent to the reaction direction at given structure. Knowing the information about bond-breaking and forming can also improve the performance and effectiveness of the optimization process. In Chapter 3, we introduce a new optimization algorithm to exploit the chemical intuition from the reaction. From the difference between the reactant and product, the algorithm can recognize the key internal coordinates as a reduced representation of the reaction. These reduced coordinates guides the Hessian matrix to have the correct eigenvalue information.

1.5 Iteration strategies

In the optimization process, One starts from the initial structure \mathbf{x}_0 to iterate towards the points of interest on the energy surface. The iterative process is terminated when a result is reached within the preset threshold. With gradient \mathbf{g}_i ,

Hessian \mathbf{H}_i of current point \mathbf{x}_i and previous points $\mathbf{x}_{i-1}, \mathbf{x}_{i-2} \dots \mathbf{x}_0$, a proper step need to be determined.

1.5.1 Line Search method

In linear search method, a direction vector \mathbf{p}_i is chosen. The algorithm optimize along the direction for a point with a lower energy or smaller gradient magnitude. The direction vector simplifies a multidimensional optimization problem into a dimension optimization problem. To find the proper answer, one need to solve

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

The exact minimization may take too expensive and unnecessary as too many trial steps is required. Instead, a point with approximate minimum value is more preferred. With the new point \mathbf{x}_{i+1} , a new \mathbf{p}_{i+1} is selected, and the algorithm is repeated.[\[44\]](#)

Steepest Descent

When choosing a optimization direction, the easiest choice the 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,

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

To minimize the value of ΔU is equivalent to

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

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

Linear search along the steepest descent direction is an effective way lowering the energy of the system. One can minimize the energy of the system without computing the Hessian matrix. When taking the step, the direction of the move is perpendicular to the contour of the energy surface. It's effective when the contour of the objective function is well scaled in each directions. However, when the contour is off regular shape, steepest descent can take many extra iterations before converging.[53]

Newton direction

Besides steepest descent direction, Newton's direction is another important one for line search method. The direction is derived from Eqn.1.33 with local quadratic approximation. For minimization, linear search method requires a positive-definite Hessian matrix to fulfill the descent requirement. When adapting Newton step, the scalar factor α is not applied when the step generate a structure with energy reduction.

Newton step leads to a fast and more robust convergence. Utilizing the extra information from Hessian matrix, the step taken is impartial from the curvature of each direction. The limitation to Newton step is the computation cost for Hessian matrix in each iteration. Without explicitly compute the Hessian matrix, one

can either using finite-difference to estimate the Hessian or using quasi-Newton method in 1.4.1 to approximate the Hessian matrix according to the information from previous structure.

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 at near region. The newton step from 1.33 is calculated premising the quadratic energy model. The model is accurate when the step is close to current structure \mathbf{x}_i . If the step is too long, the higher order derivatives omitted in the quadratic model become non-negligible.[44, 54] Then the approximated energy from the local model will no longer be accurate. To ensure a proper step size, it is sensible to limit the maximum length of the step at each iteration,

$$\|\Delta x\| \leq \tau \quad (1.43)$$

where τ is the trust radius set according to the accuracy of current model. When a step surpass the trust radius, the step is deemed risky and need to be rescaled. When the step from Eqn.1.33 is larger than τ , a constrained optimization is needed,

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

where $\tilde{\lambda}$ is the Lagrange multiplier. When minimizing the objective function, $\tilde{\lambda}$ is set to be positive. For maximizing problem, $\tilde{\lambda}$ is set to be negative. Expanding Hessian matrix with spectral theorem,

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

The solution for optimization step is rewritten as[52, 55]:

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

where λ is determined at the constraint that

$$\|\Delta \mathbf{x}(\tilde{\lambda})\| = \tau \quad (1.47)$$

$\text{sgn}(\lambda_i)$ denotes the sign of each eigenvalue. This shift ensures the negative eigenvalue remains negative while scale down the relative step length along the eigenvector direction. 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

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

When the two energy is close, the quadratic model is considered accurate, so τ is increased in the next iteration. If the energy is poorly matched, the radius is decreased. If the model generates moderate result, the radius is keep unchanged.

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

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

λ_p and λ_n are selected so that $\lambda_1 - \lambda_n < 0$ and $\lambda_i + \lambda_p > 0$.

Another popular method is 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

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

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

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

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 are a proper candidate for minimization tasks, while the second smallest eigenvalue is more suitable for transition state optimization.

To derive a more generic solution to 1.49, one can solve two separate generalized eigenvalue problems of negative-curvature and positive-curvature respectively for the Lagrange multiplier.

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

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

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 geometry optimization: minimization, and saddle point optimization. Maximization problem can be convert to a minimization problem by multiplying the objective function with -1. When conducting minimization, the procedures are relatively straightforward. One can take a step towards the energy descent direction until a convergence is reached. The saddle point optimization is more complicated. As a stationary point on the PES, the structure is the maximum in some directions but minimum in the other directions. 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 gradient when taking a good step towards the transition state. Without extra information

about the uphill landscape, saddle optimization remains an obscure problem.

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 trying to automate the searching procedure by exploiting the information from the reaction, product or both.

1.6.1 Single-Ended Method

One common method is to start the geometry optimization from one end of the reaction, normally the reactant or product, but it is not necessary. The path to the transition state requires the initial structure 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.

Coordinate driving method premises the reaction can be characterized by one coordinate. The algorithm takes major steps towards the direction of the selected coordinates. At each iteration, a constrained optimization is performed to minimize all the other coordinates. The selected coordinate is believed to carry the reaction from the reactant to product through the transition state. Coordinate driving method creates a line path consist of each optimized point. The point with the highest energy is regarded as the best transition state guess. This is an effective algorithm when a reaction can be simply characterized by one internal coordinate. But when the path is curved where the driving coordinate changes

during the reaction process, the coordinate driving method is not effective and may fail due to the lack of flexibility.[60–63]

To extend the applicability of coordinate driving, one can select more proper internal coordinate as the driven coordinate or compound current coordinates to form a new coordinate representation. For example, in the hydrogen transfer reaction where the hydrogen atom need to travel from on end to the other such as isomerization from HCN to CNH. The bond angle \angle_{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 to drive energy uphill is the direction-of-least-ascent.[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 simple mechanism. When the molecules are large, the vibration modes increases, Following the lest-ascent direction normally ends up reaching a conformational changes rather than a reaction product. For large molecule, a safer choice is to select the eigenvector of the Hessian matrix that corresponds to the reaction path of interest. [66, 67]

Besides, dimer method is also an applicable method to drive the energy uphill to the desired transition state.[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 dimers along the axis. The direction of next step is determined by rotating dimers to align with the lowest curvature direction. Then a displacement is taken place for the midpoint along the transition state. Without

calculating the Hessian matrix exactly, the dimers method provides a versatile method for finding the best reaction direction. The efficiency of the method is mainly determined by the rotation of the dimers for the lowest eigenvalue mode.

Gradient-extremal path is an alternative useful 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 is locally defined path where each point on the path $\mathbf{x}(s)$ follows

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

where s is the arc length in the path, $\mathbf{g}(s)$ and $\mathbf{H}(s)$ is the gradient and Hessian at $x(s)$. 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. Gradient-extremal path is designed to pass through stationary point such as minimum, transition state, or high order saddle point. 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. Also, in each step, an analytical calculation for Hessian matrix is required. A direct implementation of the gradient-extremal path is computational-heavy.

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, In RGF path, though starting from

different direction, different trajectories intersect at the stationary points on the surface.

All these methods have some intrinsic drawbacks as the algorithm is searching for a one-dimensional parameterized path for transition state. The performance is good when the molecule is small. But when the system is large, these methods are no longer reliable because the excessive dimensionality of the objective energy surface. Multi-dimensional surface walking is a more 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 keep all the other coordinates minimized.

These multi-dimensional surface walk approaches are more robust as all the important coordinates are properly included. But these methods are also computationally costly. Properly selecting the reduce internal coordinates is crucial to the success of the algorithm.

1.6.2 Double-Ended Methods

Double-ended methods, unlike its single-ended counterparts searching from one end, exploiting the information from both reactant and product for a more comprehensive description of the reaction and the transition state. In the double-ended methods, the reaction path is initially represented by a series points interpolated by the structure difference between the reactant and the product. During the optimization process, each point are optimized from their initial guess structure to the desired equilibrium state on the reaction path.

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

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

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

$$\mathbf{g} = \mathbf{g}^{spring} + \mathbf{g}^{PES} \quad (1.56)$$

where $\mathbf{g}^{spring} = \frac{dV^{spring}}{ds}$ and \mathbf{g}^{PES} is the surface potential gradient. 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

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

$$\mathbf{g}_{\parallel}^{spring} = \tau \tau^T \mathbf{g}^{spring} \quad (1.58)$$

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

where τ is the unit vector tangent to the reaction path. The $\mathbf{g}_{\parallel}^{spring}$ component in \mathbf{g}^{NEB} is solely for displace equal distance between points while the \mathbf{g}_{\perp}^{PES} is used for drag the point downhill to the optimal position in the perpendicular space. This separation decouples the interference of spring force from the relaxation process, prohibiting the cutting-corner path when the reaction path is curved.

For some reactions, the energy of the system changes rapidly without enough restoring force acting on it. This situation causes a kinky reaction path, 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.

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

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

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

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

In 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

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

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

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

Following the Eqn.1.63 the force gradually evolves the initial guess towards the desired MEP. Impossible to optimize infinite points one the path, the practical

implementation is to present the string path with discrete states and connecting through all the states by a interpolation curve, normally a cubic spline. The object of the optimization is to minimize the $\sum_i \|\mathbf{g}_\perp^{(i)}\|$ at each state along the whole string path. One can adopt various numerical optimization methods such as steepest-decent or Quasi-Newton method to update the state.

Unlike NEB method, there's no direct spring energy between neighbouring points, Less coupled states make string method converge more smoothly. Without the nudged force, the displacement of each state on the string is need to be redistributed to ensure adequate coverage near the transition state region. Also, extra "kinking" force is required to straighten out the path.

In growing string method (GSM), the number of states representing the path increases over iterations.[97–100] Starting from both ends, the states grows systematically until both part meet and form a complete reaction path. A parameterized density function and a indicator function are defined to keep track of the newly added points and the spacing. When the growing string step into new interior, the density function and indicator function are evolved adaptively to ensure a uniformly distribution. 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 method 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.

1.7 Summary

Effectively modelling chemical reactions is one of the most fundamental and important problem in theoretical chemistry. There are many aspects of the problem need to be carefully handled to achieve a satisfactory result including accurately determination of stable structures on the potential energy surface, properly generation of transition state guess, effectively location of transition state, and the detailed description of reaction path.

In this chapter, we reviewed recent works of popular methods for geometry optimization. The difficulty of geometry optimization lies in many different aspects of the problem. A good initial guess structure is important for geometry optimization. It's relatively easy to generate a good starting point for minimization such as the reaction, product, or the intermediate. It's more challenging and obscure to find a good guess for transition state. Different types of geometry optimization methods are introduced in chapter.1.6. In chapter.2, a new efficient and unambiguous method is introduced. The generated structure is a 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 optimization process is based on the Taylor expansion of the local potential approximation. When high order derivatives after Hessian matrix are omitted, the update step is denoted as Newton step or quasi-Newton step, depends on whether the Hessian is analytic or approximated. Various quasi-Newton update schemes are introduced in chapter.1.4. Transition state optimization requires the Hessian matrix to have one exact negative eigenvalue. We presents a new hybrid method

incorporating the reaction information. The entries associated with reaction space are explicitly evaluated through finite difference while the remaining elements are updated through quasi-Newton method. This new approach combined the advantage of accurate Hessian information with fast update method. In Chapter.3, a systematic testing proves the promising potential for effectively converge transition state optimization.

In Chapter.5, a comprehensive testing is conducted to measure the performance of our new approach. Different initial guess structure are generated by a certain amount 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 are better converging optimization task even the initial guess is poor.

Compared with locating transition state, finding a reaction path on the potential. Depends on where the reaction path was initiated, the methods are categorised into single-ended and double-ended methods. Chapter1.6. In Chapter.5, we introduced a new approach to generate the path point sequentially by a bisection optimization algorithm. The optimization technique it utilizes is the same technique as the one in Chapter.3 except the key coordinates is selected as the reaction reaction vector. Unlike the transition state optimization looking for a saddle point, at each iteration of the path point, a local minimum in the hyper-plane 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-García, 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 coordinates can be calculated directly from the Cartesian coordinates:

$$q_{AB}^{bond} = \|\mathbf{R}_{AB}\| \quad (2.1)$$

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

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

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

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

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 \mathbf{B} matrix, with elements:[12]

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \quad (2.5)$$

Owing to the redundancy of the internal coordinates, the \mathbf{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:

$$\delta \mathbf{q} = \mathbf{B} \cdot \delta \mathbf{x} \quad (2.6)$$

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

$$\delta \mathbf{x} = \mathbf{B}^+ \cdot \delta \mathbf{q} \quad (2.7)$$

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,

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

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 reevaluating 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:

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

2. Hydrogen bonds are designated between hydrogen atoms covalently bonded to one atom with strong electronegativity, $\mathbf{X} = \text{N, O, F, P, S, Cl}$, and located in the peripheral area of another strong electronegative atom, $\mathbf{Y} = \text{N, O, F, P, S, Cl}$. 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° .

$$R_{HY} \leq 0.9 * (r_H^{vdW} + r_Y^{vdW}) \quad (2.10)$$

$$\angle_{XHY} > 90^\circ \quad (2.11)$$

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 specifying 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.

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

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 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.

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

Bond Angles

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

Conventional Dihedral Angles

Dihedral $\angle_{\alpha\beta\gamma\delta}$ 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 defined from the cross products of the

$$\hat{R}_{\beta\alpha} = \frac{R_\alpha - R_\beta}{\|R_\alpha - R_\beta\|} \quad (2.14)$$

where $\hat{R}_{\beta\alpha}$ denotes the unit vector of bond linking atom β and α . The normal vector \hat{n} is defined as,

$$\hat{n}_{\alpha\beta\gamma} = \frac{\hat{R}_{\beta\gamma} \times \hat{R}_{\beta\alpha}}{\|\hat{R}_{\beta\gamma} \times \hat{R}_{\beta\alpha}\|} \quad (2.15)$$

The dihedral angle is the angle between these two normal vectors,

$$\angle_{\alpha\beta\gamma\delta} = \cos^{-1}(\hat{n}_{\alpha\beta\gamma} \cdot \hat{n}_{\beta\gamma\delta}) \quad (2.16)$$

Including all possible dihedrals in the system would lead to the explosion of internal coordinates. To reduce the unnecessary redundancy, we restrict the dihedrals to be the one including non-auxiliary bond. More specifically, 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 $\alpha\beta\gamma*$ and added to the internal coordinates set. 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 center 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 $\angle_{\alpha\beta\gamma}$, $\angle_{\alpha\beta\delta}$, and $\angle_{\gamma\beta\delta}$ 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 of the angle between the $\alpha\beta$ and $\gamma\delta$ bonds and the volume of the parallelepiped enclosed by $\alpha\beta\gamma\delta$.

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

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

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 test the performance of the robust dihedral indicators, the comprehensive test results are demonstrated and discussed in the next chapter.

2.3.3 Mapping between Internal coordinates and Cartesian coordinates

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 coordinates is as close as possible to the specified redundant internal coordinates.

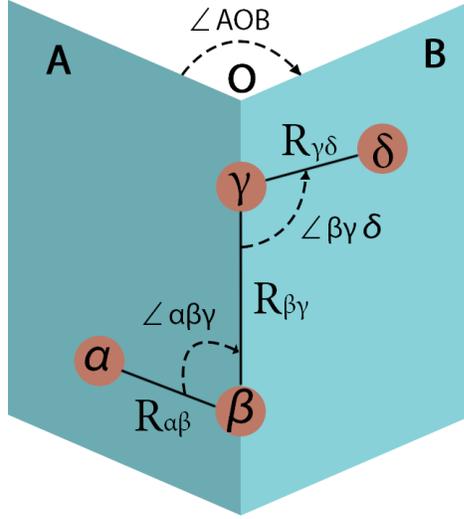


FIGURE 2.1: Illustration of internal coordinates consist of 4 atoms

Based on the protocol we proposed in the previous section, the amount of selected robust internal coordinates should be way higher than $3N - 6$. Randomly selecting $3N - 6$ internal coordinates is very unlikely to represent a physically realizable structure. The feasible structure of certain molecule is located on a $3N - 6$ Manifold inside a M_{int} -dimensional space.

For a random structure represented in internal coordinates, it's highly possible there isn't a corresponding Cartesian counterpart. In order to map every point from the M dimension space onto the $3N - 6$ 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,

$$\mathbf{x}(\mathbf{q}^{(target)}) = \underset{\mathbf{x}}{\operatorname{arg\,min}} (\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)})^T \mathbf{W} (\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)}) \quad (2.19)$$

where \mathbf{W} is a weight matrix with only diagonal elements. It is used to demonstrate the importance of each internal coordinates when conducting the manifold mapping. Initially, \mathbf{W} is the identity matrix. Under different circumstances, the value can vary according to one's desire. If some coordinates are selected as frozen ones, a large value will be assigned, When conventional dihedral is included in the system, its weight is adjusted to $\sin^2 \angle_{\alpha\beta\gamma} \sin^2 \angle_{\beta\gamma\delta}$ to reduce the impact of the possible collinear situation. The objective cost function of transformation adapted square function for bonds and robust indicators while utilizing cosine function as a discrepancy measurement for angles,

$$cost_{bond} = \left(q(x) - q^{(target)} \right)^2 \quad (2.20)$$

$$cost_{angle} = \left(\cos(\theta(x)) - \cos(\theta^{(target)}) \right)^2 \quad (2.21)$$

As for conventional dihedral,

$$cost_{dihed} = \left(\cos(\phi(x)) - \cos(\phi^{(target)}) \right)^2 + \left(\sin(\phi(x)) - \sin(\phi^{(target)}) \right)^2 \quad (2.22)$$

The complete objective function to be minimized is defined as,

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

With the assistance of redundant internal coordinates and the manifold projection method, We adopted the double-ended scheme to generate a reasonable initial

guess structure. After the reactant and product structures have been expressed in terms of redundant internal coordinates, it is important to generate a sensible guess for the transition-state structure. To achieve this, we make a line segment that interpolates from the reactant to the product, and (in the absence of further information) choose the midpoint of this segment as a guess for the transition-state structure.

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

In general, the transition-state-guess in redundant coordinates will not be realizable, so the manifold projection method is used to find corresponding atomic positions. We observe that this structure is, in most cases, an excellent initial guess for transition-state optimization.

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

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

where p is the fractional variable determining the interpolation ratio. When $p = 0$, the guess structure is the same as the reactant, or $p = 1$ for 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.

2.4 Results and Discussion

2.4.1 Testing Protocol

A set of 32 chemical reactions involving various mechanisms is constructed as the test set. It includes proton or hydrogen transfer reactions, pericyclic reactions, Diels-Alder reactions, intramolecular reactions, *Sn2* reaction, free radical, etc. The detailed reaction information is presented in the appendix.

All the calculations are conducted at HF level with 6-31+G basis, except two reactions using 6-31+G(d,p). The reactant and product structures used in the double-end method are obtained from an intrinsic reaction coordinate(IRC) process of the known TS. The transition state optimization update is performed by *GOpt*, a geometry optimization package developed by us along with *Gaussian 16* for computing Energy and gradient at each optimization iteration.

2.4.2 Overview of result

When constructing the guess conformation, the union of the coordinates from the reactant and product are used. Robust dihedral and conventional dihedral are deployed separately to form the respective initial guess structure, as well as in the following optimization iterations. The optimization process is handled by *GOpt* towards the same target TS. All other meta parameters are set to be the same. The purpose is to evaluate the efficiency, robustness, and consistency of the two different sets of dipoles.

2.4.3 Result and Discussion

First of all, we generalize the basic trends derived from the table 2.1. Later we analyze the performance on specific cases to draw out a more comprehensive understanding of the new method.

Negative eigenvalues of the hessian matrix indicate the number of directions along which the potential surface are maxima. For a TS structure, only one negative eigenvalue is preferred because it is only the maximum along the reaction path while the minima among the rest. The capability of generating the right amount of eigenvalues is crucial to the TS optimization, especially when the initial guess structure is far from the ideal position. From the 2.1, the average number of negative eigenvalues are 0.97 and 1.03 respectively, within the same error range of ± 0.3 . Both are qualified for generating a plausible initial guess with *GOpt* double-end method. Based on the test data, there isn't a noticeable difference between the two methods.

The average convergence iterations needed for robust dihedral, 7.94 steps, is at the same level as the conventional dihedral's 7.15 steps for the successful cases. No substantial difference is observed from the average result directly. However, the convergence rate is quite different. All the reactions converges to the expected TS with robust dihedrals while 5 out of 32 reactions failed instantly due to the ill-defined conventional dihedral. The remedy for this problem is possible. Thanks to the extra redundancy in the redundant internal coordinates, all the failed system are still well-specified after the failed conventional dihedrals being removed. The results in Tab.2.3 shows the number of iterations needed after the alteration. If these results are also included to represents conventional dihedral's performance, the total average steps of conventional dihedrals increase slightly to 7.47.

Though the average result looks very close between the two methods, their performance varies depending on different types of reactions and their initial geometries. Among all the reactions successfully converged, conventional dihedral outperforms robust dihedral in 9 reactions with less iterations, reaches a draw in 13 reactions, and only behaves worse in 5 cases. It is a sensible result. Unlike the conventional dihedral, which is represented by the cosine angle between the two planes, robust dihedral implemented two different indicators rather than one. For a molecule with M dihedral planes, $2M$ robust dihedral indicators are included in the system. Though more information is preserved, the extra redundancy also leads to slower convergence. As the number of internal coordinates increases, the dimension of the space the molecule resides increases, resulting in a harder minimization problem when projecting the internal structure back to its Cartesian counterpart through manifold projection. When a nonphysical configuration is obtained, compromise in coordinates change are inevitably made between all redundant internal coordinates. All these factors contribute to the slow convergence of the highly redundant system.

Convention dihedral is a commonly used internal coordinates in many optimization occasions. It's not a suitable candidate for dealing with collinear systems by definition,

$$\cos(\phi) = \frac{\mathbf{n}_A \cdot \mathbf{n}_B}{|\mathbf{n}_A||\mathbf{n}_B|} \quad (2.27)$$

where \mathbf{n}_A and \mathbf{n}_B are the normal vectors of the two planes. If \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 are constituting vectors of the two planes, the dihedral angle can be reformed as,

$$\cos(\phi) = \frac{(\mathbf{b}_1 \times \mathbf{b}_2) \cdot (\mathbf{b}_2 \times \mathbf{b}_3)}{|\mathbf{b}_1 \times \mathbf{b}_2||\mathbf{b}_2 \times \mathbf{b}_3|} \quad (2.28)$$

when any two adjacent \mathbf{b}_i vectors are near collinear, the norm of the cross product is near 0. This design leads to an error-prone internal representation of a molecule. Moreover, the transformation matrix between internal and Cartesian coordinates is approximated by the Wilson B matrix, which is defined as $B_{ij} = \frac{\partial x}{\partial q}$. This approximation is established when both changes are minuscule. However, at the near collinear situation, tiny changes in Cartesian may result in drastic changes in internal, undermining the stability and validity of the transformation.

Viewed from the Tab.2.1, multiple reactions failed without further alteration due to the collapse of the conventional dihedrals. In these reactions, at least three atoms are collinear. Fortunately, the TS structure resides in the same line space defined by the collinear atoms. The information reserved by the redundant internal coordinates is normally more than enough to specify the system. With only bonds, angles, and other dihedrals, the system is still well defined. By simply removing the defective dihedrals, these guess structures can converge to the desired TS without difficulty. After reducing the redundancy, most of the failed reactions performed even better compared to their robust counterpart. The drawback of removing collinear dihedrals is that the out-of-line movement is also eliminated from the optimization process. Without the constraints from certain dihedrals, taking a step out of the collinear structure becomes a harder task. Fortunately, all the test reactions, that converge after the tweak, all have their TS achieved without extra moves for peripheral atoms.

2.5 Conclusion

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 connection 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.

In this chapter, we 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 dihedral. The robustness and consistency are the key advantages of the new representation.

Among all 32 test reactions, robust dihedral compete all the tasks with 7.94 number of iterations and 100% successful rate while conventional dihedral with 7.15 steps and 0.84% rate. Though averagely, robust dihedral indicators converge slightly slower, It exhibits great potential and robustness tackling collinear structures where conventional dihedral usually performs poorly or even fail.

The main drawback of this new indicators is their slightly slower convergence in optimization process. In redundant internal coordinates, the total amount of specified dihedrals are normally more than the degree of freedom. Using two

indicators for dihedrals brings extra redundancy to the system. In the test, for the cases both methods succeed to converge without modification, conventional dihedral excels in 33% of the total tests regarding to optimization iteration, while only been outperformed for 19% of the tests.

Though slightly slower compared to the conventional dihedral, the new dihedral indicators excels in consistency and robustness. These properties serve these new indicators to be a promising candidate for solving chemical reaction involving collinear structures.

TABLE 2.1: Number of iterations and negative eigenvalues for generated guess structures

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	$C_5H_6 + C_2H_4$	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	$N_2C_2 + C_2H_4$	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	71	36
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. Gonzalez-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 Dalton 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 pentatomic tetrahedral molecules. *Physical review* 38(11) (1931), 1969.

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 is proposed in this chapter. One of the salient advantage of this method is the separation between key internal coordinates, The total $3N - 6$ reduced internal coordinates space is divided into the key-space corresponding the bond-breaking and bond-forming process, and the non-key space with leftover coordinates. Quasi-Newton update schemes are deployed each iteration to ensure an efficient optimization process. When the elements of Hessian matrix associated with the reaction activate site are inaccurate, finite difference is applied to correct the Hessian matrix. Redundant internal coordinates are constructed with bonds, angles, and the robust dihedral indicators.

A trust radius scheme is adopted to constrain the gradient information as well as a Hessian modification scheme to adjust the Hessian eigenvalues in a proper form. A set of 32 reactions consist of various reaction types are used to compare the performance between our algorithm and the popular *Berny* algorithm. All the energy and gradient evaluation is computed in Gaussian[1]. Compared with *Berny* algorithm, our new approach exhibits more robust and consist performance. The new algorithm and all the advanced features will be accessible in our coming open-source Python package *GOpt*.

3.2 Introduction

Geometry Optimization is a vital procedures in many quantum chemistry researches regarding reaction mechanism and molecular reactivity. 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) with reactant and product as the local minima, and TS as the first order saddle point. It's relatively easy to locate the reactant and product as one may follow the gradient downhill to the minimum. However, finding the TS is a more challenging task. For a first order saddle point, it's the maximum along only one dimension while minimum in all the other perpendicular ones. Many sophisticated optimization methods have been developed in the last several decades. Three major types of methods are devised to explore the transition state. The first one is a straightforward optimization method 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 initial guess and the researcher’s chemical intuition. The other two methods can be divided into two categories: the single-ended method and the double-ended method according to the starting point structure.

The most popular optimization method is the *Berny* algorithm.[2] It starts from a provided initial guess structure. With the calculated Hessian matrix at the first step, the algorithm displaces the guess along the direction with the negative eigenvalue. As an efficient method, Berny algorithm has made a great success and was regarded as one of the enhanced and effective algorithm. It is the default option in Gaussian series software for geometry optimization. Berny algorithm is sensitive to the initial structure. A guess conformation without exact one negative eigenvalue is not welcome. It also requires the researcher to be keen on possible active sites when involving complicated mechanism.

Single-ended methods normally start from the reactant or the product, one of the valley on the potential energy surface. An ascending direction is selected to drive the energy go 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 utilizing the eigenvector information. Gradient-extremal-method[12–14] and reduced-gradient-following[15–17] are also effective methods in generating paths passing through stationary points. Though these paths are normally not the lowest-energy-path but they sheds light on discovering more possible 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 elaborated methods, such as nudged elastic band (NEB) method,[18–26]

string method (SM),[\[27–32\]](#) and growing string method (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. An minimum-energy path is obtained on the PES as the initial guess for TS optimization. It’s a relative expensive approach since multiple energy and gradient evaluations are needed. Besides, QST approximated paths are often very different from the real reaction paths.

Though continual progress has been made for various computational methods, exploring TS is still an unpleasant experience. Optimization Failures still frequently occur. These inconsistency may attribute to (1) computation failure from the underlying quantum chemistry software, (2) unphysical structure obtained during the optimization process, (3) Inaccurate Hessian matrix resulting poor convergence.

Many factors may 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 step control scheme.[\[39–41\]](#) Motivated by the problems of existing methods, we herein propose a robust algorithm with minimal efficiency loss, the *GOpt* algorithm.

In *GOpt* algorithm, we adapted the redundant internal coordinates from the work of Pulay. 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 to be 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.[\[42, 43\]](#)

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, are also called the reduced coordinates. 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 nonreduced internal basis, a reduced-internal transformation V matrix 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 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 only 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 become 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 describly 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, inter-fragments bonds, long distance bonds for linear chains, and auxiliary bonds. All these bonds are measure in atomic unit. Bond angles are measured between any two non-auxiliary bonds. The value is measure in cosine function rather than in direct angle for more robust performance at near collinear structure. One of the major improvement is the selection of dihedrals compared to normal redundant internal schemes. In **GOpt**, for every non-auxiliary bond $\beta\gamma$, the atoms α is selected as the atom bonded to β with the most bonds connection to other atoms. Then all possible $\alpha\beta\gamma*$ are included. Symmetrically, the dihedral defined by $*\beta\gamma\delta$ are also added to the system where δ is set to be the atom with the most bonded atom. Normal conventional dihedral may fail frequently when the system is in the collinear structure. When three atoms in the dihedral are located near one line, the plane defined by these atoms is very inconsistent. A small change in the Cartesian coordinates of these atoms may result in 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

$$\hat{R}_{\beta\alpha} = \frac{R_\alpha - R_\beta}{|R_\alpha - R_\beta|} \quad (3.3)$$

is the unit vector along the direction $\beta\alpha$ 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 coordinate in geometry optimization process. When a small perturbation is imposed on one coordinates, the corresponding changes in the transformed coordinates will be small as well, guarantees the consistency of the coordinates transformation.

3.3.3 Coordinate transformations

The Cartesian coordinate is the direct and straightforward representation yet non-efficient nor intuitive for chemistry while the internal coordinate is ideal for optimization but clumsy when computing energy and it's derivatives. That's why a functional and robust transformation scheme is indispensable in geometric optimization process. Because the transformation from Cartesian to internal is not linear, the best approximation for the transformation is by Wilson **B** matrix. It is the Jacobian matrix of the transformation from the $3N_{atoms}$ Cartesian coordinates to the N_{int} internal coordinates. The entries of the Jacobian matrix is defined as,

$$b_{ij} = \frac{\partial q_i}{\partial x_j} \quad (3.4)$$

These elements can be obtained from trigonometric functions. Usually, the **B** matrix is rectangular because the internal coordinates are normally way more

than $3N$. Since only $3N - 6(5)$ degree of freedom is needed to fully specify it's structure, the \mathbf{B} matrix is also singular with only $3N - 6(5)$ nonzero singular values. Their corresponding vectors are called the delocalized internal coordinates. By rearranging eqn. 3.4, we obtained the the matrix format,

$$\mathbf{B}\delta x = \delta q \quad (3.5)$$

This equation expresses the change of internal coordinates when a small change occurred in Cartesian coordinates. Because \mathbf{B} matrix is rectangular, there isn't a unique inverse. We adapted Moore-Penrose pseudoinverse of \mathbf{B} matrix for the inverse transformation,

$$\mathbf{B}^+\delta q = \delta x \quad (3.6)$$

It's worth noting that not every changes in internal coordinates is realizable. For example, three atoms forming a triangular structure. Any change leads to the sum of three angles other than 60° is not physically allowable. To converted these nonphysical internal coordinates, we project them onto realizable space spanned by \mathbf{B} matrix with minimum error distance,

$$\delta\tilde{q} = \hat{\mathbf{P}}\delta q \quad (3.7)$$

where

$$\hat{\mathbf{P}} = \mathbf{B}\mathbf{B}^+ \quad (3.8)$$

3.3.4 Transformation between Reduced Internal and Cartesian

In the numerical optimization process, most software compute the energy, energy gradient, and its Hessian matrix in 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. If known the Willson \mathbf{B} matrix,

$$g_x = B^T g_q \quad (3.9)$$

$$H_x = B^T H_q B + K \quad (3.10)$$

K is calculated by

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

where b'_{ijk} is defined as,

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

It's the derivative of elements in \mathbf{B} matrix. Inversely, the energy derivatives in internal coordinates can be computed through,

$$g_q = (B^T)^+ g_x \quad (3.13)$$

$$H_q = (B^T)^+ (H_x - K) B^+ \quad (3.14)$$

Equation 3.13 is essential because most quantum chemistry software compute energy and its derivatives in Cartesian coordinates. Due to the nonlinear transformation between cartesian and internal coordinates, the transformation is only valid when the changes in each system is infinitesimal. Points in cartesian space

are of $3N$ dimensions while in internal space of N_{int} dimension which is usually way bigger than $3N$. Converting from cartesian X to internal is straightforward trigonometric application. The inverse is not always clearly defined. If one starts with \mathbf{x}_k and its corresponding internal coordinates $\mathbf{q}(\mathbf{x})_k$, after a non-infinitesimal optimization step s , new ideal internal coordinates \mathbf{q}_{k+1} would be

$$\mathbf{q}_{k+1} = \mathbf{q}(\mathbf{x})_k + s \quad (3.15)$$

This new configuration in internal coordinates space will rarely located on the $3N - 6$ manifold representable by Cartesian coordinates. That is

$$x = x_k + B^+ s_k \quad (3.16)$$

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

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

The detailed implementation 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 bond forming, breaking, and angle swing. Motivated by this idea, a reduced-dimensionality

potential energy surfaces can be generated to describe the energy changing during the reaction. Given a system consisting of n internal coordinates, the crucial information related to the reaction mechanism is located in the few key internal coordinates space. The PES around the reaction site will be also changes depends on the changes of these key internal coordinates. To effectively identify the key internal coordinates, the most accurate way would be getting it directly from the researcher. The users could specify the key internal coordinates involved in the reaction based on their chemical intuition as the input for the program.

If no user input is provided, a sets of protocol is conducted to select proper key internal coordinates based on the difference between the reactant and product structure. Initially, a union set of internal coordinates are generated from the internal coordinates of the reactant, product, and the TS guess structure.

The coordinates would be selected as the key internal coordinates 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.

3.3.6 Construct delocalized reduced internal coordinates

We developed a geometry optimization algorithm based on the reduced internal coordinates. It is similar to the idea proposed to Baker. 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 \mathbf{K} key internal coordinates, the non-key coordinates are constructed through a linear combination of all the other redundant internal coordinates.

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

The \mathbf{V} matrix to transform reduced delocalized internal coordinates to redundant internal coordinates. is denoted as

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

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

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

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

$$\mathbf{B}_{N_{int} \times 3N} = \mathbf{U}_{N_{int} \times N_{int}} \cdot \mathbf{\Sigma}_{N_{int} \times 3N} \cdot \mathbf{V}_{3N \times 3N}^* \quad (3.21)$$

$$\mathbf{B}\mathbf{B}^T = \mathbf{U}_{N_{int} \times N_{int}} \cdot \mathbf{\Lambda} \cdot \mathbf{U}_{N_{int} \times N_{int}}^T \quad (3.22)$$

To separate the changes in key internal coordinates space and non-internal coordinates space, we impose a small changes each key internal coordinates successively

without changing others. This operation usually results in an unrealizable structure. We then project the unrealizable structure to the realizable space through,

$$\mathbf{b}^{(j)} = \mathbf{P}\hat{\mathbf{e}} = \mathbf{B}\mathbf{B}^+\hat{\mathbf{e}}^{(j)} \quad (3.23)$$

where $\hat{\mathbf{e}}^{(j)}$ is a unit vector with 1 in the j^{th} position but 0s anywhere else. These $\mathbf{b}^{(j)}$ delocalized vectors correspond to the changes when the key internal coordinates changes the redundant internal coordinates space.

The vectors $\mathbf{b}^{(j)}$ are not orthonormal, we orthogonalize it through,

$$\mathbf{B}\mathbf{B}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \quad (3.24)$$

where $\mathbf{B} = [\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(k)}]$. We pick the eigenvectors \mathbf{v}_i from \mathbf{U} with non-zero eigenvalues λ_i .

$$\mathbf{V}_{key} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_k \end{bmatrix} \quad (3.25)$$

where k is the number of independent reduced coordinates. Normally, k should be equal to the number of key internal coordinates. If not, it indicates that there is redundancy in the key internal space. We then reduced the dimensionality of key-internal space by only including the independent eigenvectors.

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

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

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

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

$$\mathbf{D}\mathbf{D}^T = \mathbf{V}'\mathbf{\Lambda}\mathbf{V}'^T \quad (3.28)$$

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

$$V_{nonkey} = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \dots & \mathbf{v}'_{3N-6-k} \end{bmatrix} \quad (3.29)$$

Combining the key reduced space and non-key reduced space,

$$\mathbf{V} = \begin{bmatrix} V_{key} \\ V_{nonkey} \end{bmatrix} \quad (3.30)$$

we obtained the complete \mathbf{V} matrix for transforming redundant internal coordinates to delocalized reduced internal space. \mathbf{V} matrix need to be construct in 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,

$$\mathbf{V}_{new} = \mathbf{Q}\mathbf{V}_{old} \quad (3.31)$$

$$\mathbf{Q} = \mathbf{V}_{new}\mathbf{V}_{old}^T \quad (3.32)$$

$$= \mathbf{U}\mathbf{\Sigma}\mathbf{W}^T \quad (3.33)$$

The optimal orthonormal rotation matrix is

$$\mathbf{Q}^{(min)} = \mathbf{U}\mathbf{W}^T \quad (3.34)$$

and the maximally aligned new basis is

$$\mathbf{V}_{new} = \mathbf{Q}^{(min)}\mathbf{V}_{old} \quad (3.35)$$

During the optimization process, we map the molecular structure from Cartesian coordinates to redundant internal coordinates, then to reduced internal coordinates. With the \mathbf{V} matrix, the conversion of gradient, Hessian, and optimization step is straightforward,

$$\mathbf{g}_v = \mathbf{V}^T \mathbf{g}_q \quad (3.36)$$

$$\mathbf{g}_q = \mathbf{V} \mathbf{g}_v \quad (3.37)$$

$$\mathbf{H}_v = \mathbf{V}^T \mathbf{H}_q \mathbf{V} \quad (3.38)$$

$$\mathbf{H}_q = \mathbf{V} \mathbf{H}_v \mathbf{V}^T \quad (3.39)$$

$$\Delta \mathbf{v} = \mathbf{V}^T \Delta \mathbf{q} \quad (3.40)$$

$$\Delta \mathbf{q} = \mathbf{V} \Delta \mathbf{v} \quad (3.41)$$

$$(3.42)$$

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 updated based on the difference of gradient

between iterations. Due to the main optimization is conducted in reduced coordinates space, the secant condition for \mathbf{H}_v is selected to be

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

3.3.8 Quasi-Newton Updates

In *GOpt* optimization process, the Hessian matrix is updated by quasi-Newton methods. These methods approximate the Hessian with the properties changes from the near iteration points. The four methods we introduce here are using the last step

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

and the secant condition

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

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

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

the Powell-symmetric-Broyden update (PSB)

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

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

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

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

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

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

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

ψ is the square of the sine value of the angle between the step, $s(k+1)$, and the error in the v Hessian's approximation to the change in gradient that accompanies the step. The form of the SR1 update is designed to avoid numerical problems when $\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}$ is close to 0. BFGS method avoids this kind misbehavior by making the Hessian update to be positive definite. SR1 and PSB does not preserve the positive semi-definite during the Hessian update process. In the TS optimization, it is crucial to maintain one negative eigenvalue during the process. This makes SR1, PEB, and Bofill to be the idea candidate for TS optimization, leave BFGS as the good choice for minimization.

3.3.9 Hessian Finite Differences Update

GOpt is an efficient algorithm as it can effectively describe PES changes within the key-reduced space. It is important to keep the Hessian as an accurate approximation during the whole optimization process. We update the first R rows/columns for the key reduced coordinates with finite difference when needed. Assume the perturbation in the r^{th} key reduced coordinate is $\delta\mathbf{v} = \epsilon\mathbf{e}_r$, where \mathbf{e}_r is the unit vector with all 0's except a '1' at the r^{th} position. The update formula for the r^{th} row/column of the Hessian matrix is similar to Eqn. 3.43,

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

The r^{th} row/columns of the Hessian matrix is approximate by

$$\frac{d\mathbf{f}(\mathbf{v} + \epsilon\mathbf{e}_r)}{d\epsilon} = \frac{\mathbf{f}(\mathbf{v} + \epsilon\mathbf{e}_r) - \mathbf{f}(\mathbf{v})}{\epsilon} \quad (3.53)$$

The ϵ we used in the *GOpt* is set to 0.001.

Finite difference method is a time-consuming step, requiring an additional energy and derivative evaluation by quantum chemistry software. It is inefficient and unnecessary to update hessian at each iteration of the process. With the proper choice of Quasi-Newton method, the Hessian matrix especially the first R rows/columns, corresponding to the key reduced space, are often accurate during the optimization. Finite difference method is only invoked when the following

criteria are met:

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

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

The user parameter ω and ν are tentatively selected as 1.0.

Criterion 3.54 is checking the norm of the gradient in a specific direction compared with overall gradient in case the optimization along that direction is hindering the convergence efficiency. Criterion 3.55 compares the hessian update difference by the quasi-Newton method. When the difference is small, it reflects the update for that row of the Hessian is highly possible accurate enough without the need for finite-difference update.

3.3.10 Hessian Modification

Transition state is the saddle point on the PES. This requires the structure to have exactly one negative eigenvalue in the Hessian matrix. The corresponding eigenvector features the direction along which the energy is maximum on the reaction path.

It is important for Hessian to preserve exact 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. We circumvent this problem by modifying the Hessian matrix to conform with the ideal structure.

First, we do not want the negative eigenvalue happen in the region non-related

to the chemical reaction. To do so, we ensure the non-key reduced block of the Hessian matrix to be positive semi-definite. If not, we replace the negative eigenvalue in the subblock to zero and reconstruct the block.

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

Third, after examine the two subblock separately, we then do a complete diagnose on the entire Hessian matrix.

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

where $\mathbf{\Lambda}$ is the diagonal matrix with i^{th} eigenvalues at λ_i . We list the eigenvalues in ascending order, $\lambda_1 \leq \lambda_2 \dots \lambda_{3N-6}$. If there's only one negative eigenvalue less than the threshold λ_n , with all the other positive eigenvalue greater than threshold λ_p . This is the most observed situation.

$$\lambda_1 \leq \lambda_n \quad (3.57)$$

$$\lambda_i \geq \lambda_p \quad i = 2, 3, \dots, 3N - 6 \quad (3.58)$$

However, sometimes the Hessian matrix do not satisfied the criterion. If there is one negative eigenvalues, but not meet the requirements, we modify them by,

$$\lambda_1 = \min(\lambda_1, \lambda_n) \quad (3.59)$$

$$\lambda_i = \max(\lambda_i, \lambda_p) \quad i = 2, 3, \dots, 3N - 6 \quad (3.60)$$

If there are multiple negative eigenvalues, we would 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,

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

$\chi_{i;r}$ 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, the 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 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,

$$\mathbf{g}_{v;k} + \mathbf{H}_{v;k}(\mathbf{v}_{k+1} - \mathbf{v}_k) = 0 \quad (3.62)$$

the step to obtain the TS is then,

$$\mathbf{s}_{v;k} = \mathbf{v}_{k+1} - \mathbf{v}_k = -\mathbf{H}_{v;k}^{-1}\mathbf{g}_{v;k} \quad (3.63)$$

Expanding 3.63 with Spectral theorem

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

The step, $\mathbf{s}_{v;k}$ would lead to the exact TS structure When the objective function is quadratic and the Hessian matrix is analytical. However, the PES is way more complicated. The above step is only valid if the v_k and v_{k+1} lands within the PES quadratic approximation where the high order corrections are negligible. To ensure during the iteration, the $\mathbf{s}_{v;k}$ would not step out the valid region, a spherical region defined by radius τ is introduced.[44, 45] When the calculated stepsize from 3.63 is larger than the trust radius τ , it need to be re-scaled to fit in the trust region.

Trust-region image potential (TRIM)

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

$$\mathbf{s}_v^k(\tilde{\lambda}) = \left(\frac{\chi_1^T \mathbf{g}_{v;k}}{\lambda_1 - \tilde{\lambda}} \right) \chi_1 + \sum_{i=2}^{3N-6} \left(\frac{\chi_i^T \mathbf{g}_{v;k}}{\lambda_i + \tilde{\lambda}} \right) \chi_i \quad (3.65)$$

until

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

3.3.12 Trust Radius Determination

In each step of the optimization, the (quasi-)Newton step is computed in the nonredundant internal coordinate space, $s_v = -H^{-1}g$. The step, s_v would lead to the exact TS structure if one were very close to the solution or if the objective function were quadratic. Sometimes these conditions are nearly satisfied, but when the new coordinates, v_{k+1} , are far from the previous coordinates v_k , the quadratic approximation fails because higher order corrections are nonnegligible. To ensure the step, s_v , does not exceed the region where the quasi-Newton Hessian is reliable, a spherical region defined by the trust radius τ is introduced. When the calculated step-size is larger than the trust radius τ , the step is reduced by the trust-radius image method to fit in the trust region. We implemented separate energy-based (for minimization) and gradient-based (for transition states) methods for determining appropriate trust radii during the optimization processes.

The value of τ should be neither too large to violate the validity of 3.63 validity, nor too small to hinder the optimization convergence. The step should also be impartial towards small or large system. To make sure the step is under proper range, we define the trust radius as,

$$\tau_{init} = 0.35\sqrt{N_{atoms}a.u.} \quad (3.67)$$

$$\tau_{min} = 0.1\sqrt{N_{atoms}a.u.} \quad (3.68)$$

$$\tau_{max} = \sqrt{N_{atoms}a.u.} \quad (3.69)$$

$$(3.70)$$

where τ_{init} is the initial stepsize of the first optimization step.

In the algorithm, in the *GOpt* algorithm, gradient decreasing is a preferred

result. If the magnitude of the gradient after taking a step decreases, the optimization step is accepted and the trust radius range is to be updated with 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 recompute the new step. If the step is too small, $\tau_{new} < \tau_{min}/10$, then the step would be set to $\tau_{new} = \tau_{min}$ and taken no matter the magnitude of the gradient change. This criterion is implemented to make sure when the structure is far from the ideal TS, the guess structure can make a step towards the uphill direction from a near minimum area. Two types of trust-radius updating approaches are included in *GOpt*

Energy-based update

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

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

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

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

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

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

It indicates the quadratic approximation of local potential energy is vary accurate because the actual approximated energy is close to the real one. We then double the trust radius with enough confidence.If

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

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

It shows the quadratic model is moderate accurate. It is safer to keep the trust radius unchanged. In other situations, we regards the energy approximation inaccurate for such a big step. So a reduced step is expected for the next optimization iteration

$$\tau_{new} = \max\left(\frac{1}{4}\tau_{old}, \tau_{min}\right) \quad (3.77)$$

Gradient-based update

The energy-based method is an outstanding method in minimization. But when refer to locating transition state, it's more intuitive and appropriate to use gradient-based updating scheme. This method, as it's name suggests, uses the difference between actual and approximated gradient to evaluate the accuracy of local potential surface model. The predicted gradient is calculated as

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

There are two indicators used in gradient-based method,

- The change of magnitude of gradient between the predicted and actual ones, measured by the ratio.

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

- The change in direction of gradient between the predicted and actual ones, measured by the cosine value

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

The angle aligned by the gradient difference is a sensitive measurement to the dimension of the system. As the dimension get larger, the chance of two vectors aligned in the same direction decreases. It's more easily for two random vectors to be aligned in low dimension space. 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

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

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

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

For gradient-based method, after a step is taken, if

$$\frac{4}{5} < \rho < \frac{5}{4} \quad (3.83)$$

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

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

This shows the approximation from the model is accurate and we double the trust radius; if

$$\frac{1}{5} < \rho < 6 \quad (3.86)$$

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

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

Otherwise, we deem the 3.78 inaccurate, so we halve the current trust radius,

$$\tau_{new} = \max\left(\frac{1}{2}\tau_{old}, \tau_{min}\right) \quad (3.89)$$

recap

The trust radius update method:

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 $\tau_{new} = \frac{1}{4}\tau_{old}$. 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 as having 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 in 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 algorithm, the *GOpt*, for geometry optimization

1. Collect Cartesian coordinates of the target system as the input for *GOpt* algorithm
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 user input is given, select the coordinates related to reaction based on the difference between reactant and product.
3. Construct Wilson **B** matrix based on selected redundant internal coordinates, and reduced **V** matrix according to given key internal coordinates
4. Invoke external quantum chemistry software to compute the energy, energy gradient (and Hessian for the initial guess)
5. Transform gradient and Hessian from Cartesian space to redundant internal space with **B** matrix and to reduced internal space with **V** matrix.

6. Check if finite-difference is need to update the rows/columns corresponding to the key reduced internal space. If criterion 3.54 and 3.55 are met, the external software are invoked to conduct extra calculations.
7. Modify the hessian to ensure there is only one negative enough eigenvalues.
8. Compute the optimization step in V-space within the trust radius.
9. Express the step from reduced space to internal space. Then Using manifold-projection convert the target internal coordinates to Cartesian coordinates. 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 step side depends on the trust radius update method, go back to step 8 to recompute a shorted new step. If the trust radius is less than $\tau_{min}/10$, accept the step anyway.
11. Construct \mathbf{B} matrix and \mathbf{V} matrix for the new structure. Align the new \mathbf{V} with the previous one.
12. Update the Hessian matrix \mathbf{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.

3.4 Results and Discussion

3.4.1 Testing Protocol

To test the general performance of algorithm, we use *GOpt* to optimize the TS for 30 randomly selected reactions of different types. Details of the reactions are list in the appendix. All calculations were conducted by *Gaussian 16* with HF/6-31+G(d,p). The initial Hessian is computed analytically with *Gaussian* and updated through Quasi-Newton method.

3.4.2 *GOpt* default methods

GOpt introduces many different methods and choices for geometry optimization. Different combination may perform differently among different reaction types. We select the methods with the best performance in most general reactions. The details of the optimization comparison between different methods are listed in chapter 5.

By default, *GOpt* uses Bofill as the quasi-Newton update scheme, TRIM as the step control method, and gradient-based method for updating trust radius during the iterations.

3.4.3 Comparison with *Berny* Algorithm

Berny Algorithm is one of the most popular geometry optimization algorithm implemented in many renowned software including *Gaussian*. To compare the performance, both algorithms starts from the same initial guess and trying to

optimize the structure to the TS of interest. The result is presented in the next section.

3.4.4 Results and Discussion

The data presented in the 3.1 show the performance difference between *GOpt* and *Berny* algorithm. Viewing from the general result, *GOpt* needs averagely 7.38 steps to reach convergence, compared with 10.8 steps by *Berny*. Also, among all the random sampled reactions, *GOpt* has converged all the guess structure to the desired transition state while *Berny* failed in 1 test cases. 78.1% of the reactions converges with the same or less gradient evaluation with *GOpt*. In test 32, the guess structure was converged to a nonphysical structure where the SCF energy calculation cannot converge, result in an SCF error. In test. 6, 11 and 29, the performance of *GOpt* is substantially better. Using reduced internal step along with 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	
		<i>GOpt</i> algorithm	<i>Berny</i> Algorithm
1	$C_4H_6 + C_2H_4$	6	6
2	$C_5H_6 + C_2H_4$	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	$N_2C_2 + C_2H_4$	6	7
13	$ONC + C_2H_4$	11	9
14	$N_2CH + C_2H_4$	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 internal as the effective representation of the system. Treating the key internal coordinates separately allows the reaction to progress towards the direction we desires. *GOpt* also only need to evaluate the Hessian matrix once for the first iteration. In the following optimization process, the Quasi-Newton update and key space finite-differences method, *GOpt* can accurately approximate the potential energy surface, leading to a fast and robust convergence.

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 representation, we introduced the robust dihedral descriptors and manifold-project method. These methods provide us a fail-safe way to convert the non-linear transformation between redundant internal coordinates and Cartesian coordinates.

With these improvements, the transition-state algorithm performs significantly better than those in traditional quantum chemistry software, e.g., Gaussian. Specifically, the frequency of convergence failure is reduced by a factor of two or more, at comparable computational cost.

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-García, 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] 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.
- [43] 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.
- [44] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.
- [45] R. Fletcher. *Practical methods of optimization* john wiley & sons. *New York* 80 (1987), 4.

Chapter 4

Bisect hyperplane optimization algorithm for reaction path finding

4.1 Abstract

A novel double-ended bisection algorithm is proposed to locate reaction path. The algorithm is based on the robust reduced internal coordinates introduced in chapter 3. 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 hyperplane perpendicular to the reaction vector. Coordinated with quasi-Newton update methods and trust radius schemes, the optimization is efficient and robust. With the generated path, the interpolated chemical property is in great agreement with the result from analytical quantum chemistry calculation. Real chemical reaction models are performed and discussed.

4.2 Introduction

Chemical reactions represent the core field of many chemical researches. With the fast development of the electronic structure theory, it is more feasible to fully describe a reaction process with theoretical tools. The first step towards an effective model of a chemical reaction is to identify the correct reaction path connecting the reactant and product. This path is closely linked to the transition states and intermediates, providing detailed information regarding the reaction mechanism.

Different methods have been devised and developed in the past several decades. Based on the beginning structure(s) of the initial paths, these methods can be categorized into the single-ended methods, and double-ended methods. The first concept is going uphill on the potential energy surface from a single stable structure, either the reactant or the product, to search the possible transition state. However, taking a random step towards any direction is an uphill move. Extra information is needed to locate the desired transition. The most common one is coordinate driving method.[1] This method relies on the researcher to pick one coordinate associated with the reaction as the dominant direction. When optimizing along the path, it is important to keep all other coordinates stay in the minimum. But this method may not work well if the reaction involves multiple main coordinates during the reaction process.[2-5] The alternative idea is to follow the direction along the eigenvector of Hessian matrix with the smallest eigenvalue.[6-9] This is equivalent to step towards the direction with the least energy ascent on the potential energy surface. One more 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 for the latter methods. Firstly, the costly evaluation of Hessian matrix is required; Secondly, kinky paths are hardly avoided when following Hessian related methods. For gradient extremal method, the paths generated are normally different from the minimum energy path.

Rather than looking for the TS from one side of the reaction, some other 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 group methods in the double-ended family. For NEB method, One or more extra virtual spring potentials are appended to the original potential expression of the system. 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 the equal spacing. In 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 through a spline path. During the optimization, each point follows the steepest gradient descent in hyperplane of reaction path. The chain-of-state methods like NEB and SM can effectively locate the reaction path from the reactant to the product while bypassing the calculation of the exact transition state. With the complete pathway, one can effectively conclude the reaction mechanism and the kinetic properties. The main drawback of these chain-of-state methods is the demanding calculation requirement in the optimization process for each state. High performance parallel computing is commonly implemented to circumvent the computation power limitation. 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]

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 algorithm starts with two stable structures on the potential energy surface. By connecting the two end points on selected coordinates space, a interpolated mid structure is generated as an initial guess. Constrained minimization is performed in the hyperplane perpendicular to the reaction vector. This procedure can effectively locate the valley on the potential energy surface between the reactant and product. Then an iterative bisection process is performed to complete the path with points in between. Unlike the SM and NEB methods where the points are even spaced, in the bisection algorithm, all the points are discretized depends on the reaction and user’s desire. This provide the advantage to increase the resolution at near the transition state structure. It also eliminates the kinky pathway on the surface as the each points is optimized in its own hyperplane without disturbance from other states. The remaining chapter elaborates the implementation details. Two reaction examples are presented to illustrate the advantages of this methods. Discussion and future improvement are augmented at the end of this paper.

4.3 Methodology

4.3.1 Coordinates system

To effectively describe a chemical reaction with meaningful coordinates, we normally select internal coordinates system, such as chemical bonds, angles, and dihedrals between planes, to represent the system during the reaction, denoted as $q_{N_{int}}$. There are various schemes to select different sets of internal coordinates. In this chapter, we stick to the same internal coordinates selection procedures as the one described in Chapter 2.

Redundant internal coordinates

Here, we briefly recap the coordinates selection idea for the redundant internal coordinates. When given a structure, five types of inter-atomic bonds, including covalent bond, hydrogen bond, inter-fragments bonds, long distance bonds as well as auxiliary bonds, are added depends the types of atoms and the distance between them. We include the angles forming by each pair of bonds, excluding auxiliary bonds, connected to the same atom. For dihedrals, we replace the conventional one with our robust dihedral indicator which performs more robust and consistent. 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 any non-selected atom bonded to the other atom.

The transformation from Cartesian to redundant internal coordinates are implemented by \mathbf{B} matrix

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

while the inverse is done by an iterative manifold projection method

$$\mathbf{x}(\mathbf{q}^{(target)}) = \underset{\mathbf{x}}{\operatorname{arg\,min}} (\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)})^T \mathbf{W} (\mathbf{q}(\mathbf{x}) - \mathbf{q}^{(target)}) \quad (4.2)$$

where \mathbf{x} is the Cartesian representation; $\mathbf{q}(\mathbf{x})$ is the set of internal coordinates corresponds to \mathbf{x} ; $\mathbf{q}^{(target)}$ is the desired internal set but may not be a physical structure. The forward and back transformation is not symmetric, because the dimension in redundant internal coordinates are normally way higher than Cartesian. For a set of Cartesian coordinates, there is always a corresponding internal representation but not vice versa.

Reduced internal coordinates

Though the redundancy of internal system grants us more connectivity information, it slows down the optimization process, hindering convergence efficiency. To fully specify a system, only $3N - 6$ degree of freedom is needed. In this bisection algorithm, we adapted the similar idea as the one elaborated in chapter 3 with minor modification to fit our path-finding theme.

Rather than separate the reduced internal coordinates into key and non-key partition, here we split the reduced internal space into the reaction space and the non-reaction space.

To generate the proper reduced space, we start with choosing $3N - 6$ non-zero singular vectors from \mathbf{B} matrix, denotes as

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

To separate the reaction path space from the non-reaction space, we compute the coordinate displacement $\delta\mathbf{q}_{path}$ between the reaction and product

$$\mathbf{q}_{path} = \mathbf{q}_{start} - \mathbf{q}_{end} \quad (4.4)$$

We project the reaction path indicator into the realizable space

$$\mathbf{q}'_{path} = \mathbf{B}\mathbf{B}^+ \mathbf{q}_{path} \quad (4.5)$$

The \mathbf{V}_{react} is consist solely of reaction path vector

$$\mathbf{V}_{react} = \frac{\mathbf{q}'_{path}}{|\mathbf{q}'_{path}|} \quad (4.6)$$

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

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

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

where \mathbf{P}_{react} is the projection operator of the key internal space. In the leftover \mathbf{D} space, we pick the $3N - 5$ orthonormal as the basis,

$$\mathbf{D}\mathbf{D}^T = \mathbf{V}'\mathbf{\Lambda}\mathbf{V}'^T \quad (4.9)$$

There are $3N - 5$ eigenvectors with non-zero eigenvalues in the \mathbf{V}' .

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

$$V_{nonreact} = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \dots & \mathbf{v}'_{3N-5} \end{bmatrix} \quad (4.10)$$

4.3.2 Method overview

Especially for complicated multi-step reactions, it can be very difficult to locate transition states, or even to propose reasonable reaction intermediates. In such cases, it is best to determine the chemical reaction path directly. This is also useful for detailed studies of reactions, where having an atomistic description of the reaction pathway is useful.

Unfortunately, most existing methods for finding the minimum energy path between reactant and product are prone to failure and unsuitable for parallel programming, which is especially problematic given the cost of reaction-path-finding. To remedy this, we developed a bisection method. Like other 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.

The first step in the bisection method is to take the reactant and product,

denoted by $q_{reactant}$ and $q_{product}$ as their redundant internal coordinate representation. The reaction path vector is initialised to the vector between the two structures, $q_{path} = q_{product} - q_{reactant}$. Minimizing starting from the midpoint between the structures, $q_{guess} = (q_{product} + q_{reactant})/2$ on the hyperplane perpendicular to the reaction path vector is guaranteed to find a point on a minimum energy path. The manifold projection algorithm is used to find a molecular structure corresponding to q_{guess} . Since the constrained minimization and the manifold projection algorithm are robust, this method converges.

Note that this algorithm works for any two structures. Denoting the Cartesian structures of two points as x_{start} and x_{end} , one can uniquely define the redundant internal coordinates, denoted as q_{start} and q_{end} respectively. The reaction path vector is $q_{path} = q_{end} - q_{start}$ and the initial guess state structure is a linear combination of each end $q_{init} = (q_{start} + q_{end})/2$, where the manifold projection method is used to locate the closest structure $q(x)_{init}$ 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 each set of sequential points, with a maximum of 2^{n-1} simultaneous constrained optimizations at each step. 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.

Generate initial state on the path

The reaction path starts from two end structure, normally the reactant and the product, but our bisection method do not restrict that. For simplicity, we denote the Cartesian representation of the two ends as \mathbf{x}_{start} and \mathbf{x}_{end} . Following the protocol in 4.3.1, redundant internal coordinates are selected for both structures, denoted as \mathbf{q}_{start} and \mathbf{q}_{end} respectively. The initial guess state structure is a linear combination of each end $\mathbf{q}_{init} = (\mathbf{q}_{start} + \mathbf{q}_{end})/2$. However, the guess state in internal coordinates space may not be a physical one in Cartesian space. An iterative manifold projection method is used to locate the closest structure $\mathbf{q}(\mathbf{x})_{init}$ on $3N$ Cartesian space corresponding to \mathbf{q}_{init} .

Construct non-react space

Redundancy in the internal coordinates is of great help when specifying the connectivity of structure. On the other hand, it also hinders the optimization convergence efficiency. For a system with N atoms, only $3N - 6$ degree of freedom is needed. From the redundant internal coordinates representation of the two end points, we obtained the reaction path vector through $\mathbf{q}_{path} = \mathbf{q}_{end} - \mathbf{q}_{start}$. The actual unit direction vector $\hat{\mathbf{q}}_{path}$ is the realizable unit counterpart of \mathbf{q}_{path} :

$$\hat{\mathbf{q}}_{path} = \frac{\mathbf{B}\mathbf{B}^+\mathbf{q}_{path}}{|\mathbf{B}\mathbf{B}^+\mathbf{q}_{path}|} \quad (4.11)$$

The unit direction vector solemnly consist of the \mathbf{V}_{react} space. Then the hyper-plane perpendicular to the reaction space is the whole $3N - 6$ space without the reaction direction. To project out the reaction path space, one can easily follow Eqn.4.10 to construct the non-react space $\mathbf{V}_{nonreact}$.

Normally the energy, gradient, and Hessian (or approximate one) are computed in Cartesian coordinates. The conversion between cartesian, redundant internal, and reduced internal is the same as described in 3.36 expect the \mathbf{V} is replaced with the new $\mathbf{V}_{nonreact}$.

Optimization process

After projecting out the reaction path vector, the reaction path finding task is downgraded to a constrained minimization problem within the leftover subspace.

Energy, gradient, and Hessian are normally computed in Cartesian coordinates. To conversion between different coordinate system are conducted through

$$\tilde{\mathbf{g}}_v = \tilde{\mathbf{V}}^T \mathbf{g}_q \quad (4.12)$$

$$\mathbf{g}_q = \tilde{\mathbf{V}} \tilde{\mathbf{g}}_v \quad (4.13)$$

$$\tilde{\mathbf{H}}_v = \tilde{\mathbf{V}}^T \mathbf{H}_q \tilde{\mathbf{V}} \quad (4.14)$$

$$\mathbf{H}_q = \tilde{\mathbf{V}} \tilde{\mathbf{H}}_v \tilde{\mathbf{V}}^T \quad (4.15)$$

$$\Delta \tilde{\mathbf{v}} = \tilde{\mathbf{V}}^T \Delta \mathbf{q} \quad (4.16)$$

$$\Delta \mathbf{q} = \tilde{\mathbf{V}} \Delta \tilde{\mathbf{v}} \quad (4.17)$$

$$(4.18)$$

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

$$\tilde{\mathbf{s}} = -\tilde{\mathbf{H}}^{-1}\tilde{\mathbf{g}} \quad (4.19)$$

Minimization is easy to reach as one can follow the gradient descend direction. To ensure a proper direction is taken for the optimization step, the hessian matrix is required to have all positive eigenvalues. If any eigenvalues are negative, an increase in energy yet decrease in gradient structure is obtained. This step will push the structure towards the transition state or even higher order saddle point. When encounter Hessian matrix with negative eigenvalue, a Hessian shift function is called to alter the negative or small positive value to the preset threshold λ_p .

$$\lambda_i = \begin{cases} \lambda_p & \text{if } \lambda_i < \lambda_p \\ \lambda_i & \text{otherwise} \end{cases} \quad (4.20)$$

Newton method is exact and swift if the contour is quadratic. However if the potential energy surface is usually elusive and complicated, a simple Newton step is incapable to locate the minimum. The Hessian evaluation in the Newton methods is time-consuming yet unnecessary, especially for minimization. To circumvent the problem, the quasi-Newton method is adapted. Rather than exact hessian, quasi-Newton approximate the Hessian matrix based on the difference between previous step(s). It can be expressed in general,

$$\mathbf{H}^{new} = \mathbf{H}^{old} + \Delta\mathbf{H} \quad (4.21)$$

so that the new step is

$$\mathbf{s}^{new} = -(\mathbf{H}^{new})^{-1}\mathbf{g} \quad (4.22)$$

There are many different ways of updating Hessian matrix like SR1, Bofill family, Broden family and so on. Among those methods, BFGS is one of the most successful. The general update scheme for BFGS is expressed as:

$$\mathbf{H}^{new} = \mathbf{H}^{old} + \frac{\Delta\mathbf{g}(\Delta\mathbf{g})^T}{(\Delta\mathbf{g})^T\mathbf{s}} - \frac{(\mathbf{H}\mathbf{s})(\mathbf{H}\mathbf{s})^T}{\mathbf{s}^T\mathbf{H}\mathbf{s}} \quad (4.23)$$

One of the greatest feature for BFGS method is the output matrix is always positive-definite, make it an ideal candidate for quasi-Newton method in minimization.

Eqn.4.19 is only valid when the potential energy surface fulfill the quadratic approximation. This estimation is appropriate if the optimization step taken is within a certain trust radius τ . [35, 36]

$$\tau_{init} = 0.35\sqrt{N_{atoms}a.u.} \quad (4.24)$$

$$\tau_{min} = 0.1\sqrt{N_{atoms}a.u.} \quad (4.25)$$

$$\tau_{max} = \sqrt{N_{atoms}a.u.} \quad (4.26)$$

$$(4.27)$$

where τ_{init} is the initial stepsize of the first optimization step. The trust radius is updated every time a new step is taken. The approximated energy change by quadratic estimation is

$$\Delta\mathbf{E} = \mathbf{g} \cdot \mathbf{s} + \frac{1}{2}(\mathbf{s})^T\mathbf{H}\mathbf{s} \quad (4.28)$$

while the actual energy difference is

$$\Delta U = E^{new} - E^{old} \quad (4.29)$$

We use the ratio of the estimated energy difference against the real one as the indicator guiding the trust radius update. When the ratio is between $[\frac{2}{3}, \frac{3}{2}]$, we regards the quadratic approximation as an accurate estimate. So it's safe to expand τ . If the ratio falls in the range between $\frac{1}{3}$ and 3, then the quadratic model is considered moderate, so we keep the trust radius unchanged. However, if the ratio doesn't land in the above ranges, the quadratic model is not accurate for the same size step we just took. It's sensible at this situation reduce the trust radius. The detailed scheme can be generalized as follow

$$\tau_{new} = \begin{cases} \min(\max(2\tau_{old}, \tau_{min}), \tau_{max}) & \frac{2}{3} < \frac{\Delta E}{\Delta U} < \frac{3}{2} \\ \max(\tau_{old}, \tau_{min}) & \frac{1}{3} < \frac{\Delta E}{\Delta U} < 3 \\ \max(\frac{1}{4}\tau_{old}, \tau_{min}) & \text{otherwise} \end{cases} \quad (4.30)$$

Convergence Criteria

Since the reaction path point optimization is not as rigid as minimization or transition state optimization, here we adapted the similar strategy proposed by Baker and Chan yet with a slight loosen standard. We consider the optimization converged when the largest component of Cartesian gradient corresponding to the non-react $\mathbf{V}_{nonreact}$ space is less than 1×10^{-3} . If the process doesn't reach convergence in 100 steps, a *Converge Failure* error would be raised.

4.4 Examples and Cases

4.4.1 Muller-Brown Potential

Muller-Brown potential is a two-variable parametric potential.[37] It is an ideal testing ground for the algorithm. The minimum energy path connects the reactant and product through a intermediate and two saddle points. Also, the path is deviated from the linear interpolation of the two minimum points where a non-convex minimization is needed.

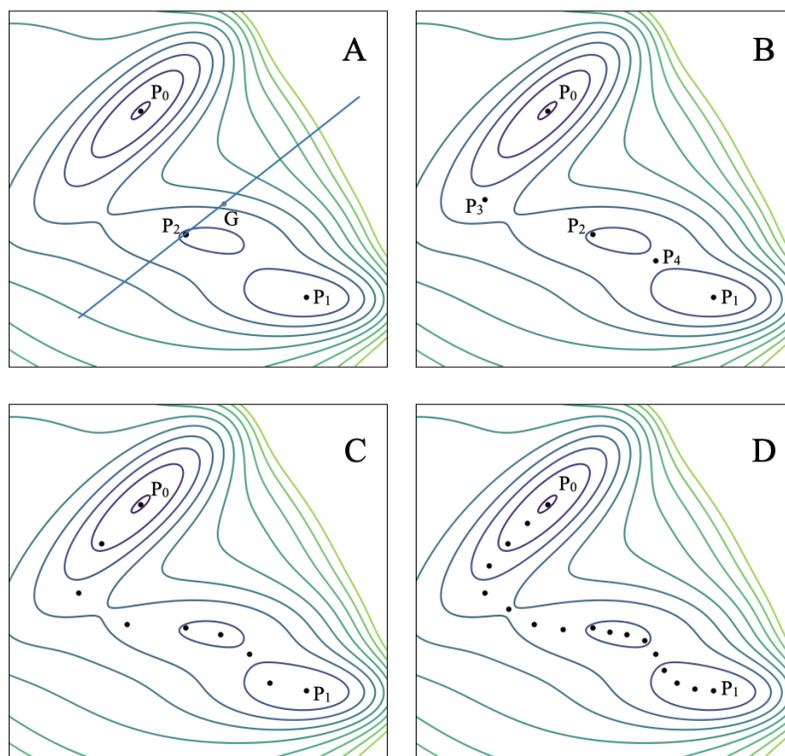


FIGURE 4.1: Reaction path finding on Muller-Brown surface with Bisection method

To better illustrate the process, we implemented the algorithm on the Muller-Brown two-parametric potential. The whole process is presented in the Fig. 4.1. Clearly we can see there are three minima along with two transition states. To obtain the starting points, two individual unconstrained optimizations are performed with two initial guesses. For brevity, We denote these two starting points P_0 and P_1 respectively. To find the first point on the path, a guess point is generated by linear interpolation between P_0 and P_1 in selected coordinates. With the initial guess G , and the reaction space spanned by direction vector connecting P_0 and P_1 , a constrained minimization is performed in the hyperplane perpendicular to the direction vector. On the two dimensional potential surface, it is located at the minimum on the bisector line P_2 (Fig. 4.1A). After the first middle point P_2 anchored on the potential surface, new points are generated with P_2 as the starting point (Fig. 4.1B). This process is continued until enough points are generated to represent the reaction path. Unlike other popular reaction path methods, the Bisection hyperplane optimization does not keep the equidistant points. This grants the algorithm the ability to put denser points at the area where the path curves, providing a more detailed description among potential twist region.

The whole process is set to finish if certain number of points have been generated indicated by the users, or other preset criterions are met such the maximum distance between nearby points are less than certain threshold measured in given coordinates. In Fig. 4.1C & D, we generate 9 points and 17 points respectively to describe the reaction path. There are three critical points on the path, the intermediate and two saddle points. Because the Muller-Brown surface is analytical, we can calculate the exact coordinates and potential of them. Comparatively, the

same information is obtained through the reaction paths from bisection optimization method.

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

The data in Tab.4.1 illustrates the performance for different reaction path. The intermediate and saddle points are interpolated by a cubic spline going through all path points. The path in Fig. 4.1B is not shown in the table as a monotonic curve is produced without enough data support. In the two sample reaction paths shown in the table, 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 between is ± 0.2 for Path C and ± 0.03 for D. The whole trend of the reaction path is quickly defined by the first several round of points evaluation. On the two-dimensional Muller-Brown potential, after projecting out the one direction vector, the hyperplane of the middle point is a perpendicular line. The minimization process brings the initial guess to the valley of the intersection. This guarantees a path-point in that local region a reaction path must go through. The procedures is also independent from potential gradient outside the hyperplane, preventing forming a kinky pathway.

From the data in Fig.4.1 and Tab.4.1, Bisection optimization algorithm exhibits great efficiency and accuracy as the method aims at describing the big picture of the reaction path before elaborating the details in the local region. In the real implementation, the generated path provides a proper guess for TS transition optimization. Because every point is obtained through a constrained minimization, no real Hessian matrix evaluation is required in the whole process.

4.4.2 Chemical Reactions

1. HCN Isomerization

The degree of freedom of a system is $3N - 6$ where N is the number of atoms in the system. To test the bisection optimization method in real chemical reaction, the HCN \rightarrow HNC isomerization is selected as the first multi-dimensional test example. This isomerization reaction has been studied thoroughly both experimentally and theoretically.[38–40] As a molecule of three atoms, the total degree of freedom needed to fully specify the structure is 3. It functions as an idea reaction to testify the performance of bisection method on multi-dimensional potential surface.

The two starting structure for taking the bisection optimization are optimized with HF/6-31+G method. All the reaction path points optimization is also determined in the gas-phase with the same HF/6-31+G method.

If using the bond length of CH, CN and the angle between HCN as the three independent variables, the bond length between C and N are of little change during the whole process. To better illustrate the reaction path, we depict the CN bond length as a constant to majorly emphasize the H transfer trajectory. The number

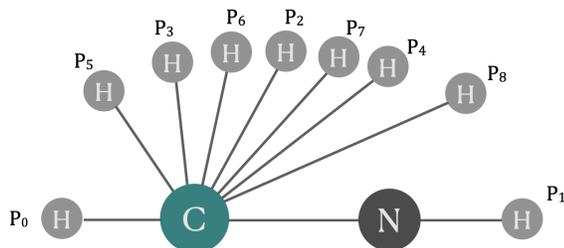


FIGURE 4.2: Path points generated by Bisection algorithm

P_0 and P_1 are stable structures for reactant and product. The following indices indicate the sequence of generation

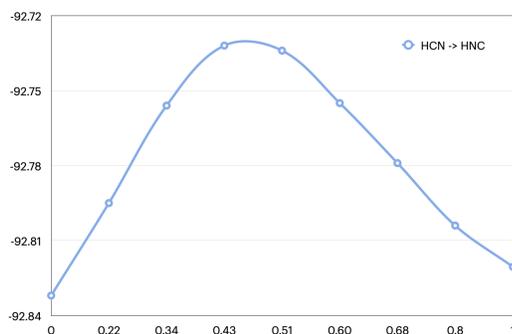


FIGURE 4.3: Energy curve of reaction HCN -> CNH along the reaction coordinates

in subscript represent the sequence being generated by the bisection algorithm if performed no-parallelly.

As shown in Fig.4.2, the reaction path conforms with many experimental results. Parameterizing the reaction path by the arc length between each adjacent path points, an interpolated continuous reaction path is generated as shown in Fig.4.3. To the interest of most researchers, the transition state of the isomerization is obtained by interpolating the maximum on the spline curve. The results from

TABLE 4.2: Transition state from interpolation and analytic Computation for HCN \rightarrow CNH reaction

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

interpolation and quantum calculation is in Tab.4.2. The result from the reaction path curve highly agrees with the exact calculation reference. Though only 7 points are produced by the procedures, they suffice to provide a detailed description along the reaction, especially among the area near transition state. It is worthy noting that in the actual result from the bisection optimization, the bond length between C and N is not constant. The actual bond is slight stretched from the initial 2.19 a.u. to maximal 2.26 a.u. near the transition state.

2. HSNO \rightarrow HONS Isomerization

The second reaction is HSNO \rightarrow HONS isomerization.[41–43] It may look alike to the previous example, but the real reaction path behind is far more complicated. Many computational researches have been conducted to disclose the reaction mechanism. The simple H atom migration is involved with multiple available route, intermediates, and transition states. The details is shown in Fig.4.4.

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

From the structures of the points from the bisection optimization method, the

TABLE 4.3: Transition state from interpolation and analytic Computation for HSNO \rightarrow HONS reaction

	Interpolation	Quantum Computation	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%

path we obtained is the same as the reaction mechanism 1 in Fig.4.4A. The path successfully goes through two transition states, and the connecting intermediate. With the path, the energy of the structure of great interests are shown in the Tab.4.3. Compared with the reference energy from quantum computation, the interpolated the results are accurate to the third decimal places for transition states, and the second decimal places for Intermediate. These points provide a quantitatively description of the whole reaction process. Also from the the graph, there's a tiny peak around $x = 0.13$, which corresponds to the hydrogen rotation barrier. Guided by the neighbouring points, the current path leads the structure to the valley side of that rotation barrier.

4.5 Conclusion

The bisection optimization method introduced an efficient and accurate way to generate a minimum energy pathway connecting the reactant and product. It is designed to be kinky-free and flexible. No extra information is need about the reaction except for two ends structures. No limitation is imposed on what starting and ending points are along the reaction path.

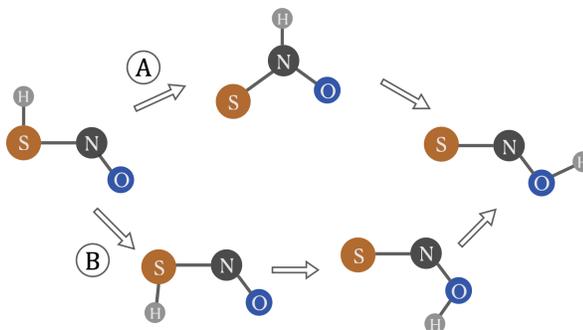


FIGURE 4.4: HSNO \rightarrow HONS isomerization mechanism

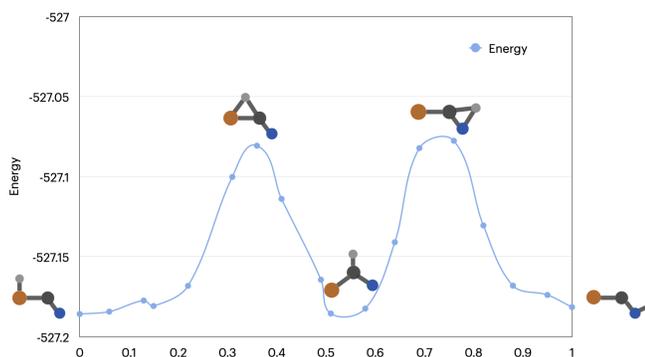


FIGURE 4.5: HSNO \rightarrow HONS Energy vs Reaction process

In bisection optimization methods, the reaction space is divided into the tangent path-direction vector and its hyperplane. This simplifies the process into a local constrained optimization problem. The generated energy path provides quantitatively detailed characterization of the reaction process. Based on the required accuracy, more points can be added to the system and located at the region where better description is needed.

Three examples of different difficulty levels illustrate the effectiveness. The first example is a tricky two-variable parametrized potential surface where the energy

is not convex when locating the minimum energy path. This obstacle is overcome by *GOpt* hessian modification scheme. The numerical result is highly in agree with the analytical one. The second example is the isomerization of HCN. This is a simple 3-dimensional real reaction. However, the reaction path for this one is relatively easy and straight forward. The pathway with 9 points can clearly characterize the migration of H from one side to the other. The interpolated the TS energy is every close to calculated one. The last example is the isomerization of HSNO. This reaction involves two possible reaction mechanism, multiple saddle points and intermediates. From the pathway generated by *GOpt*, it takes one of the mechanism, goes through all the TS and intermediate on that path. The interpolated energy of those key structures are as accurate as calculated with errors less than 0.0007%.

Based on the promising performance of the bisection optimization algorithm, we hope it along with *GOpt* can become a helpful and powerful tools for more chemical researches and application.

4.6 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] J. E. Dennis Jr. *RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations*. 1983.
- [36] R. Fletcher. *Practical methods of optimization* john wiley & sons. *New York* 80 (1987), 4.
- [37] 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.
- [38] K. Ishida, K. Morokuma, and A. Komornicki. The intrinsic reaction coordinate. An abinitio calculation for $\text{HNC} \rightarrow \text{HCN}$ and $\text{H} + \text{CH}_4 \rightarrow \text{CH}_3 + \text{H}$. *The Journal of Chemical Physics* 66(5) (1977), 2153–2156.
- [39] 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.

- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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 optimization algorithm *GOpt*. The test set consist of 32 reactions from various types. A random perturbation is applied to the known transition state to generate 10 random initial structures. Both *GOpt* algorithm and the benchmark *Berny* algorithm method are deployed to find the desired transition state on the potential energy surface. The increments of perturbation range from 0.05, 0.1, 0.2, to 0.4 atomic unit. In general, *GOpt* performs more efficient when the perturbation is small, and is marginally slower when the perturbation increased to 0.4 a.u.. But *GOpt* possesses higher convergence rate throughout all test scenarios, marking it a promising optimization candidate.

5.2 Introduction

A novel algorithm for optimizing chemical structure is introduced in chapter 3. In the proposed algorithm, quasi-Newton methods are adopted along with trust-radius method to control the optimization step. Elaborated Hessian modification method are implemented to locate the uphill direction for saddle point. The algorithm exhibits promising efficiency and robustness at converging various geometry optimization tasks in a straightforward tests.

In this chapter, a more systematic and thorough assessment is conducted to explore the range of effectiveness. In 3.1, *GOpt* reaches less iteration cycles and better convergence rate compared with Berny algorithm.[1] Its great performance is partially accredited to the good initial guess from the interpolation methods, and partially to the sophisticated algorithm leading the structure towards the proper TS. When the reaction mechanism is relatively straightforward, the guess structure is not too far away from the transition state, the optimization may only need to take a few iterations to step towards the gradient decreasing direction. However, a good initial guess is not always available. When the reaction mechanism is complicated, multiple intermediates and TS involved, the optimization algorithm may need to take a step towards a energy increasing or even gradient increasing direction. Though *GOpt* has shown promising results in the general tests, it is still not clear to what limit, *GOpt* can robustly handle the non-ideal initial guess.

In the default setting of *GOpt*, many meta-parameter are included to initialize the optimization algorithm. To accustom to different optimization cases, we implemented 2 cost functions for coordinate transformation, 4 different quasi-Newton update schemes, 3 types of secant conditions, and 2 kinds of trust-radius updating

protocols. Among all the choices compatible, the optimal combinations and the premised condition are unknown.

Most often, when new approaches established, people normally test the new idea against a small testset.

In this chapter, a systematic assessment is designed to investigate the performance of *GOpt* under different parameter, and for different initial guess structure.

5.3 Testing protocol

A database of chemical reactions for testing

To effectively test a computational algorithm, it is crucial to construct a broad and relevant database. Most of the reactions included were taken from the test sets that we used for testing transition-state optimizer. The database involves various reaction types such as Diels Alder,[2] electrocyclic,[3, 4] Huisgen cycloaddition,[5] Addition, Proton transfer, S_N2 substitution, free radical reactions.[6] Some reaction were constructed by replacing or adding extra functional groups. This procedure produced several sterially-hindered reactions that requires the optimization algorithm to correctly identify the true transition state from other low-energy barrier corresponding to conformation changes. All quantum chemistry computation are conducted in *Gaussian*[7] 16 with HF/6-31+G. The exact Hessian is computed on the very first step.

5.3.1 A Systematic Method for Generating Initial Guess

As a optimization algorithm, it's important to perform consistent and robust with initial structures of different quality. Our goal here is generate a set of random

initial guesses by imposing a random perturbation of selected scale to the the true transition state. We start with a random vector, \mathbf{a} , in Cartesian coordinates. The random vector is projected onto the Wilson \mathbf{B} matrix and it's pseudo-inverse,[8] and then normalized

$$\hat{\mathbf{u}} = \frac{\mathbf{B}^+ \mathbf{B} \mathbf{a}}{|\mathbf{B}^+ \mathbf{B} \mathbf{a}|} \quad (5.1)$$

This projection produces a random perturbation of the internal $3N-6$ degree of freedom without redundant translation and rotation. With the unit perturbation, a set of random structures are generated by adding the transition state geometry with a scaled perturbation.

$$\mathbf{x}_{guess} = \mathbf{x}_{t.s.} + \sqrt{3N} \cdot \epsilon \hat{\mathbf{u}} \quad (5.2)$$

The factor ϵ is to adjust the average amount of perturbation on each atom in the molecule. To average out extreme cases, we choose to generate 10 random initial guess for each ϵ . We also gradually increase the value of ϵ until the test result deteriorated dramatically. The ϵ is selected to be 0.05, 0.1, 0.2, and 0.3.

When the ϵ equals 0.3, it is observed that in many test cases, the initial guesses failed instantly from quantum chemistry calculation due to illegitimate structure. Large factor permits each atom in the system to deform away from their equilibrium position, often leading to collision of near atoms or the distance between them become unrealistic.

Here we propose another testing protocol involving only the key internal coordinates. During the chemical reaction, key internal coordinates are selected as the representation of the chemical process. The ability to stably converge a deformed structure at near key-coordinates area is a strong indicator to the robustness of the

testing algorithm. We started with a set of random generated variables \mathbf{a} matches the size of internal coordinates. The vector \mathbf{a} has non-zero entries only at the key internal space. We then project out the redundancy due to overly-defined internal coordinates, and then normalized

$$\mathbf{v} = \mathbf{B}\mathbf{B}^+ \mathbf{a}_{int} \quad (5.3)$$

The generated perturbation are then applied to the target transition state structure. The transformation from the internal coordinates to Cartesian coordinates are carried out by the manifold projection method introduced as Eqn.2.19,

$$\underbrace{\min}_{x_\kappa} |\mathbf{q}(\mathbf{x}_\kappa) - (\mathbf{q}_{t.s.} + \cdot \kappa \hat{\mathbf{v}})|^2 \quad (5.4)$$

where $\mathbf{q}_{t.s.}$ is the redundant internal coordinates of the transition state and κ is the factor to scale the perturbation applied on the equilibrium structure. The distance of the perturbed Cartesian coordinates from the original transition state structure is computed in Cartesian coordinates.

$$\|\mathbf{x}_\kappa - \mathbf{x}_{t.s.}\| = \epsilon \cdot \sqrt{N_{key}} \quad (5.5)$$

Where N_{key} is the number of key internal coordinates that are perturbed. In order to remove the irrelevant rotation and translation as well as minimizing distance calculated, we adapted Kabsch's Algorithm to align the two structures.[9] Again, We generate 10 random initial guesses for each choice of ϵ . As the changes from key internal coordinates are more accurate and tangential to transition state, we choose a slightly larger set of ϵ , 0.1, 0.2, 0.3, and 0.4.

5.3.2 Criterion on Test assessment

We assess the performance of *GOpt* with default *Berny* algorithm in *Gaussian* by the successful convergence rate and the gradient evaluation performed. As the factor ϵ get larger, the potential energy surface become more delicate and complex than area near the transition state. A big initial step and aggressive step-size update scheme may lead to a overshoot step among certain optimization direction. Therefore, we tune down the initial stepsize in *GOpt* to $0.15 * \sqrt{N}$ where N is the number of atoms in the system, and the step-size update factor from 2 to 1.5 in the gradient-base approach. The adjustment of these hyper-parameters though may result in a slightly slower convergence yet a less error-prone convergence process.

The upper limit of steps taken for *GOpt* is set to 100, if the optimization process couldn't converge in 100 steps, we deem it fails to reach the transition state. Though the step limit is set to 100, some extra gradient evaluation maybe need during *GOpt* optimization due to finite-difference hessian update or reject step recalculation.

5.3.3 Results and Comparison

Here, we assess the performance of *GOpt* in comparasion with the default *Berny* algorithm in *Gaussian*. In the geometry optimization process, the major time-consumption step is determined by the gradient evaluation. The number of gradient evaluation needed for the optimization process is an important indicator to the efficiency. We also consider the convergence rate under different perturbation factor. Though slow convergence is not preferred, we think the robustness of an optimization algorithm is more critical.

In Table.5.1, we compared the performance of *GOpt* with *Berny* algorithm on test cases where all atoms are perturbed. The perturbation are measure as distance between test structure and the real transition state structure in Cartesian coordinates space. For small displacement when $\epsilon = 0.05$ or 0.1 , *GOpt* converges with less gradient evaluation than *Berny*'s. As the displacement get larger, the quality of the test structure deteriorates, resulting the average evaluation needed to reach the transition state increase dramatically. When $\epsilon = 0.2$, the evaluation needed by *GOpt* and *Berny* is close. And When ϵ reaches 0.3 , the *Berny*'s is the slightly efficient one by tiny margin.

While the evaluation of gradient required to reach saddle points are close for *GOpt* and *Berny*. The convergence rate trend is completely different. When the ϵ is 0.05 . both methods perform excellent, more than 94% reactions converge without problem. However, as the displacement increases, the result from *Berny* deteriorate dramatically from 0.94 to 0.44 when ϵ increase from 0.05 to 0.3 . On the contrary, the result from *GOpt* is very stable. The advantage of the robustness is inconspicuous at the beginning when ϵ is small. But as the displacement increases, *GOpt* exhibits great competence converging initial guesses that are distorted and skewed. At the most extreme case where $\epsilon=0.3$, the performance of *GOpt* is close to two times better than the result from *Berny*, revealing that *GOpt* can be a helpful and promising tools in locating transition state.

Fig.5.1 shows the trend of convergence rate increase as the number of gradient evaluation increase. When the number of evaluation is less the 20 , the curve for *GOpt* and *Gauss* is almost overlapped. This correspond to the similar performance of the two algorithms when the initial guess structure from the perturbation is in good quality. As the guess structures quality deteriorate, more steps are needed

to handle as well as more error-prone structure is formed, leading a less favorable convergence result. We stop the graph at the 100th evaluation of the gradient as at this point, the structure still running optimizations either have already failed or are optimizing towards a wrong transition state. The robustness of *GOpt* is the key feature of our algorithm. This mainly accredit to the proper predefined key-internal coordinates. With a clear information of where the negative eigenvalue would be located. Certain amount of perturbation would, though distorting the structure, not affect the steps the algorithm takes towards the negative eigenvalue. Moreover, *GOpt* are granted protocols to seek the proper direction to step uphill when the current structure is with no negative eigenvalues or multiple negative ones, lending *GOpt* the capability navigating towards the right structure.

TABLE 5.1: Test results from *GOpt* and *Berny* algorithm when applied random perturbation in Cartesian coordinates

Methods $\epsilon(\text{Bohr})$	<i>GOpt</i> Algorithm	<i>Berny</i> Alogorithm
Average Gradient Evaluation		
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	0.95	0.75
0.30	0.81	0.44

In Tab.5.2, a more constructive random guess structures are used. With a more reasonable starting point, the number of gradient evaluation shown in the Tab.5.2 are substantially smaller than the one needed in full-random test-cases. Viewing

TABLE 5.2: Test results from *GOpt* and *Berny* algorithm when applied random perturbation in key internal coordinates

Methods	<i>GOpt</i> Algorithm	<i>Berny</i> Alogorithm
$\epsilon(\text{Bohr})$		
Average Gradient Evaluation		
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	0.91	0.85
0.4	0.87	0.81

from the average gradient evaluation needed, Tab.5.2 share the similar trend as Tab.5.1. *GOpt* possesses efficiency advantage over *Gauss* in small perturbation but the latter quickly catches up in tests when the perturbation is larger. Also in Tab.5.2, the maximum displacement used is $\epsilon = 0.4$, provide a clearer view of the trend. There are two factors contribute to this effect. One is the relative conservative step control scheme implemented in *GOpt*. Because the initial structure are far from the real transition state, we encourage the algorithm take more small steps in case slipping into unexpected energy saddle area. The second is *GOpt* is a Hessian-free algorithm with any analytic evaluation of Hessian matrix in the optimization process except the starting structure. This required the algorithm to take extra gradient evaluation for the sake of hessian finite difference approximation.

As for convergence rate, the data from Tab.5.2 are also Superior than the corresponding ones from Tab.5.1. The difference between the two methods still exists but way subtle than Tab.5.1. This is because the changes in key-internal space

normally allow the guess structure retained the proper eigenvectors with negative eigenvalues. This is a crucial improve when optimizing with *Berny* algorithm. Though the leading margin is not as obvious as Tab.5.1, *GOpt* still outperformed *Berny* in every tested ϵ by 3 - 6 percentage, showing great consistency and robustness in geometry optimization task.

Fig.5.2 exhibit a different trend of convergence rate to number of gradient evaluation than Fig.5.1. Due to the higher quality guess structure in each test cases, both algorithms have 70% of their optimization tests converged in less than 20 rounds. The major divergence of the two lines takes place where *number of gradient evaluation* = 15. This is majorly because when the guess structures are good, both algorithm converge fast and swift, bring the structure close to the real transition state. But if the optimization takes more than 15 steps, there's a chance the initial guess is not of high quality. This leaves more challenges to the optimization algorithm. Equipped with more tools and step control protocol, *GOpt* has the ability to handle these problems more effectively. This proves the *GOpt* is a promising algorithm for geometry optimization with both efficiency and robustness.

5.4 Summary

The goals of this chapter is to construct a protocol for testing the performance of different geometry optimization methods. Here we use out newly developed *GOpt* and the the popular methods *Berny* algorithm from the *Gaussian* program as the candidates.

There are many key features of this testing protocol, the first is a relative large and broad database of 32 chemical reactions involving various reaction types.

All of the reactions provided involved in chapter 2. The Second, we designed a method to systematic generating initial guesses with different level of perturbation. This granted us the ability to construct testing cases of various quality. Two kinds of perturbation methods are adapted here. (1) Generate a complete random perturbation in every coordinates of every atom; (2) Generate a selected random perturbation aiming certain reaction coordinates space.

The last but not least, we assess the two methods compared through this protocol, the *Berny* algorithm from *Gaussian* and our developed algorithm *GOpt*. From the result, for both kinds of perturbation, *GOpt* outperformed *Berny* in convergence rate and robustness. When the structure is relatively close to the real transition state, *GOpt* is also the more efficient choice. Yet when the quality of initial guess deteriorated, the *Berny* start to catup in efficiency. From the test result, we can confidently conclude that *GOpt* is the more suitable candidate for transition state optimization. We are also looking forward to more improvement in the *GOpt* future performance.

FIGURE 5.1: Convergence rate for random Cartesian perturbation

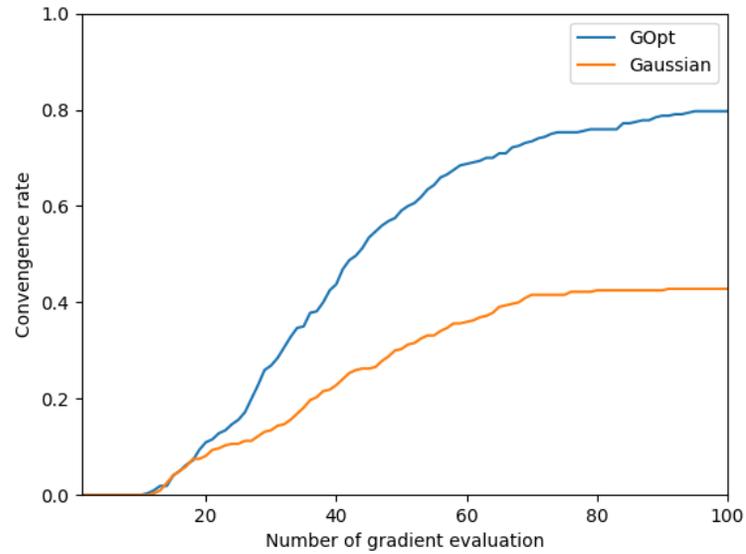
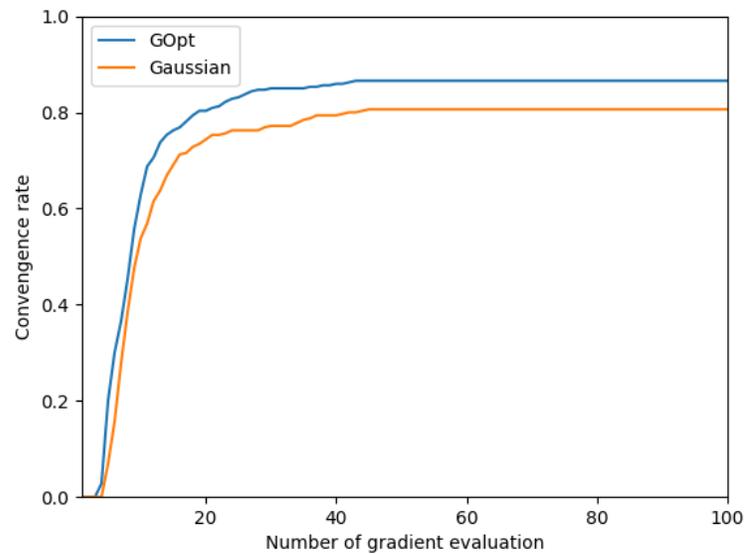


FIGURE 5.2: Convergence rate for random key internal perturbation



5.5 Reference

- [1] H. B. Schlegel. Optimization of equilibrium geometries and transition structures. *Journal of Computational Chemistry* 3(2) (1982), 214–218.
- [2] F. Fringuelli and A. Taticchi. *The Diels-Alder reaction: selected practical methods*. John Wiley & Sons, 2002.
- [3] R. B. Woodward and R. Hoffmann. Stereochemistry of electrocyclic reactions. *Journal of the American Chemical Society* 87(2) (1965), 395–397.
- [4] H. Longuet-Higgins and E. Abrahamson. The electronic mechanism of electrocyclic reactions. *Journal of the American Chemical Society* 87(9) (1965), 2045–2046.
- [5] 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.
- [6] J. Clayden. *Organic chemistry*. 2001.
- [7] 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.
- [8] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1980.
- [9] 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

Systematic Assessment on components on the optimization algorithm

6.1 Abstract

GOpt is a robust and efficient optimization algorithm. Different methods and schemes are included in the software package for versatility and flexibility. A test set consisting of 20 reactions of various sorts are selected to illustrate the general performance of each candidates. Different methods from the same category are arranged in the same group. Starting from the same initial guess structure, different methods are deployed to test the optimal combination for general optimization tasks. The result are assessed from two aspects, efficiency (number of gradient evaluations) and the robustness (the convergence rate). Viewed from the average performance, BFGS and energy-based update is the best default choice for minimization tasks while Bofill with gradient-based update is the optimal choice for transition state optimization.

6.2 Introduction

Geometry optimization is a complicated process involving various fundamental components. Designed as a user-friendly software, we provide users with a default optimization configuration that fits the most-broad reaction types. We also offer users the flexibility to choose or implement some components of their own. When starting a optimization task. One can easily follow the template of our program and write their own code to work seamlessly with existing *GOpt* package.

In *GOpt* algorithm, the Hessian matrix during the optimization is approximated through a quasi-newton update. This estimation lower the computation cost of the program, making *GOpt* more affordable when computing larger system. It also broaden the compatibility of *GOpt* with other quantum chemistry software without analytic Hessian calculation. When conducting a quasi-newton update, the secant condition is the indispensable elements. Two coordinates transformation are implemented in *GOpt* when converting structure from Cartesian coordinates to Internal coordinates, then to Reduced Internal coordinate. After each optimization step is taken, the trust radius τ is required to be updated basic on the accuracy of the current quadratic approximation.

The choices of quasi-Newton update schemes, the trust-radius update methods, and the secant condition values , have great impact on the performance of *GOpt* algorithm. Some methods are designed in favor of minimization than transition than their counterparts. In the chapter, we are emphasizing the individual contribution of each component to the overall optimization convergence. The idea is to setup a set of test reactions from different reaction types for each candidates. All the optimization will start from the same initial guess structure and towards the

same final transition state. The performance is assessed by the convergence rate, the iteration steps taken, and the gradient evaluation during the process.

6.3 Testing Protocol

To test the performance of different quasi-Newton update schemes, the trust-radius update methods, and the secant condition values, a set of 20 test cases are picked 6 typical reaction types. To coordinate with new algorithm implemented in *GOpt*, reduced internal coordinates are selected. The initial guess structure are generated as a linear interpolation between the reduced internal coordinate representation of the reactant and the product.

$$\mathbf{x}^{init}(p) = \underset{x}{\operatorname{arg\,min}} \left\| \mathbf{q}(x) - \left[(1-p)\mathbf{q}^{(reactant)} + p\mathbf{q}^{(product)} \right] \right\| \quad (6.1)$$

Where \mathbf{x} and \mathbf{q} denote the system in Cartesian coordinates and in reduced internal coordinates, respectively. The value p is a fractional number ranging from 0 to 1. In this test, we set the value of $p = 0.5$.

The optimization of each test case starts from the initial guess structure generate from Eqn.6.1. The energy, gradient, and initial Hessian are computed through external quantum chemistry software. The program employed during this test case is *Gaussian16*. All calculation are performed at HF/6-31+G level.

6.3.1 Secant Condition

Analytic Hessian matrix evaluation is a time-consuming procedure. It is not computationally sensible to compute the exact hessian matrix at each iterations during

the optimization process and it's especially prohibitive when the system is large. One common way to substitute the Hessian matrix during the optimization is using Quasi-Newton update to approximate the Hessian matrix through the information from previous several points. To do so, the first property needs to be obtained is the secant condition \mathbf{y} . The optimization step is calculated in the reduced coordinates, so the secant condition in \mathbf{V} space is

$$\mathbf{H}_v \delta \mathbf{v} \approx \mathbf{g}_v(\mathbf{v} + \delta \mathbf{v}) - \mathbf{g}_v(\mathbf{v}) \quad (6.2)$$

Eqn.6.2 is derived from the original approximation in Cartesian coordinates

$$\mathbf{H}^{old} \delta \mathbf{x} = \mathbf{g}_x \quad (6.3)$$

Following the chain rule of derivatives for transformation matrix \mathbf{B} and \mathbf{V}

$$\mathbf{H}_v^{old} \delta \mathbf{v} \approx \delta \mathbf{g}_v - (\mathbf{V}^{old})^T ((\mathbf{B}^{old})^T)^+ ((\mathbf{B}^{old})^T \delta \mathbf{V} \mathbf{g}_v^{old} + (\delta \mathbf{B})^T \mathbf{g}_q^{old}) \quad (6.4)$$

Another secant condition is obtained by switching the variable

$$\mathbf{H}_v^{old} \delta \mathbf{v} \approx \delta \mathbf{g}_v + (\mathbf{V}^{old})^T (\delta \mathbf{B}^+)^T \mathbf{g}_x^{old} + (\delta \mathbf{V}^T) \mathbf{g}_q^{old} \quad (6.5)$$

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} .

$$\mathbf{H}^{old} \delta \mathbf{v} = \mathbf{H}^{new} - \delta \mathbf{v} \quad (6.6)$$

Eqn.6.4 and 6.5 do not maintain the symmetry. So the third choice is proposed as

$$H\delta v = \delta \mathbf{g}_v + \frac{1}{2} \left((\mathbf{V}^{old})^T (\delta \mathbf{B}^+)^T \mathbf{g}_x^{old} + (\mathbf{V}^{new})^T (\delta \mathbf{B}^+)^T \mathbf{g}_x^{new} + \frac{1}{2} (\delta \mathbf{V}^T) (\mathbf{g}_q^{old}) + \mathbf{g}_q^{new} \right) \quad (6.7)$$

6.3.2 Quasi-Newton Update

Quasi-Newton methods are used to update Hessian matrix approximation in each iteration. In the *GOpt* algorithm, four most widely-used quasi-Newton update are selected as the test objects: SR1,[1] PSB,[2] BFGS,[3–7] and Bofill.[8]

Symmetric-Rank-One update (SR1)

$$\mathbf{H}_v^{k+1} = \begin{cases} \mathbf{H}_v^{(k)} & \frac{\|(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}) \cdot \mathbf{s}_v^{(k)}\|}{\|\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}\| \cdot \|\mathbf{s}_v^{(k)}\|} \leq 1e^{-9} \\ \mathbf{H}_v^{(k)} + \frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}) \cdot \mathbf{s}_v^{(k)}} & \text{Otherwise} \end{cases} \quad (6.8)$$

the Powell-symmetric-Broyden update (PSB)

$$\mathbf{H}_v^{k+1} = \mathbf{H}_v^{(k)} + \frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})(\mathbf{s}_v^{(k)})^T + \mathbf{s}_v^{(k)}(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^T \mathbf{s}_v^{(k)}} - \left(\frac{(\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T (\mathbf{s}_v^{(k)})}{(\mathbf{s}_v^{(k)})^T \mathbf{s}_v^{(k)}} \right) \mathbf{s}_v^{(k)} (\mathbf{s}_v^{(k)})^T \quad (6.9)$$

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

$$\mathbf{H}_v^{k+1} = \mathbf{H}_v^{(k)} + \frac{\mathbf{y}_v^{(k)} (\mathbf{y}_v^{(k)})^T}{(\mathbf{y}_v^{(k)})^T \mathbf{s}_v^{(k)}} - \frac{(\mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}) (\mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)})^T}{(\mathbf{s}_v^{(k)})^T \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}} \quad (6.10)$$

Bofill's 1994 update (Bofill)

$$\mathbf{H}_{Bofill}^{(k+1)} = (1 - \psi)\mathbf{H}_{SR1}^{(k+1)} + \psi\mathbf{H}_{PSB}^{(k+1)} \quad (6.11)$$

$$\psi = 1 - \frac{|\mathbf{s}_v^{(k)} \cdot \mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}|^2}{|\mathbf{s}_v^{(k)}|^2 |\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}|^2} \quad (6.12)$$

$$= \frac{|\mathbf{s}_v^{(k)} \times \mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}|^2}{|\mathbf{s}_v^{(k)}|^2 |\mathbf{y}_v^{(k)} - \mathbf{H}_v^{(k)} \mathbf{s}_v^{(k)}|^2} \quad (6.13)$$

6.3.3 Trust Radius Update

At each iteration, the optimization step is calculated through (quasi-)Newton step $\mathbf{s} = \mathbf{H}^{-1}\mathbf{g}$. When the energy is exact quadratic, the step would lead to the exact stationary point of interest. However, if the higher orders of the energy approximation is non-negligible, the premise of the optimization is no longer valid. In *GOpt* algorithm, the trust radius are imposed to adjust the step-size in case the optimization step reaching area beyond the quadratic optimization. It is defined as a spherical region centered at current optimization point. Each step calculated is constrained by $\|s\| \leq \tau$. [9, 10] Based on the accurate of local energy approximation, the radius τ is updated accordingly. The value of τ should behave unbiased towards the size of system.

When a new optimization point is obtained, the trust radius is subjected to update based on the information difference between the current electronic property and the previous one. In *GOpt*, we implemented two general types of trust radius update schemes to cater minimization such as stable structure optimization and saddle point optimization such as transition state determination.

6.3.4 Energy-Based Trust Radius Update

In the energy-based method, the estimation of the accuracy of the quadratic approximation is measured by the energy difference. The predicted energy change is calculated

$$\Delta m_{k+1} = \mathbf{g}_k \cdot \mathbf{s}_k + \frac{1}{2} \mathbf{s}_k^T \mathbf{H}_k \mathbf{s}_k \quad (6.14)$$

while the real energy change is $\Delta U_{k+1} = U(\mathbf{x}_{k+1} - \mathbf{x}_k)$. The accuracy of the approximation is divided into three cases. If

$$\frac{2}{3} < \frac{\Delta m_k}{\Delta U_K} < \frac{3}{2} \quad (6.15)$$

Then, $\tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max})$. If

$$\frac{1}{3} < \frac{\Delta m_k}{\Delta U_K} < 3 \quad (6.16)$$

Then, $\tau_{new} = \max(\tau_{old}, \tau_{min})$. Any cases do not fall into the two situation will have $\tau_{new} = \min(\frac{1}{4}\tau_{old}, \tau_{min})$

Gradient-Based Trust Radius Update

When searching for the stationary points rather than minimization, the gradient is a better indicator to the process 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 vector are considered,

$$\rho = \frac{\|\mathbf{g}_{predict}^{k+1}\| - \|\mathbf{g}^k\|}{\|\mathbf{g}^{k+1}\| - \|\mathbf{g}^k\|} \quad (6.17)$$

$$\cos(\theta) = \frac{(\mathbf{g}_{predict}^{k+1} - \mathbf{g}^k) \cdot (\mathbf{g}^{k+1} - \mathbf{g}^k)}{\|\mathbf{g}_{predict}^{k+1} - \mathbf{g}^k\| \|\mathbf{g}^{k+1} - \mathbf{g}^k\|} \quad (6.18)$$

As dimension increases, the chance of two gradient vectors aligned in the same direction decreases. To normalize the affect of high-dimension system, two benchmark values are produced to characterize the accuracy. The first one is p_{10} representing the 10% of total random vector in the space satisfying the requirement

$$\cos(\theta) \geq p_{10}(d) \approx \sqrt{\frac{1.6424}{d} + \frac{1.11}{d^2}} \quad (6.19)$$

The second one is at 40 percentile stage,

$$\cos(\theta) \geq p_{40}(d) \approx \sqrt{\frac{0.064175}{d} + \frac{0.0946}{d^2}} \quad (6.20)$$

At a new structure,

$$\frac{4}{5} < \rho < \frac{5}{4} \quad (6.21)$$

$$p_{10}(3N - 6) < \cos(\theta) \quad (6.22)$$

then $\tau_{new} = \min(\max(2\tau_{old}, \tau_{min}), \tau_{max})$. If

$$\frac{1}{5} < \rho < 6 \quad (6.23)$$

$$p_{40}(3N - 6) < \cos(\theta) \quad (6.24)$$

then $\tau_{new} = \max(\tau_{old}, \tau_{min})$, Otherwise, $\tau_{new} = \min(\frac{1}{2}, \tau_{min})$.

6.3.5 Testing Methods

The same version of *GOpt* package is used in all the tests. The initial guess structure are selected from various reactions from different types. The initial reactant and product structures are from a normal IRC of a known transition state. Initial Hessian matrix of the initial guess structure is computed analytically. Each methods from each categories are tested independently. If any crush or non-convergence within 100 steps happened in the optimization process will be marked as failure. The average number of gradient evaluation is the main indicator to assess the performance of each method.

6.4 Results and Discussion

The Table.6.1 compared the performance between different secant conditions. The overall results is quite across each method. The first variant with straight chain-rule implementation of $\mathbf{H}_v \delta_V$ generate a slight over result in both steps and gradient evaluation needed by a margin of less 5%. The symmetric version of secant condition doesn't exhibit extra advantage over the general ones. The main contributor to the secant condition is the $\delta \mathbf{g}_v$. Confined by trust radius in each optimization step, the configuration change introduced is insignificant. The variation of the secant value resulted from $\delta \mathbf{B}$ and $\delta \mathbf{V}$ is normally in a smaller magnitude than $\delta \mathbf{g}_v$. In general, all three methods are qualified to compute a competent y vector.

Quasi-newton methods are more relevant to the performance of the optimizer. Table.6.2 shows the direct comparison between each methods. Viewed from a

comprehensive perspective, Bofill generates the most promising result, followed by PSB, SR1, and BFGS. It is not surprising that BFGS is the least efficient one. As mentioned in Chapter.1.4.1, BFGS is designed to maintain positive semi-definite of the original Hessian matrix. This property ensures optimization step generated from BFGS method towards a energy decreaseing direction, making salient method for minimization optimization but not transition state determination. Though, in *GOpt*, a Hessian modification is implemented to guarantee a proper negative eigenvalue in key-internal coordinates space, the extra step of correction slows down the overall efficiency. Taking Bofill as a benchmark, BFGS takes more than two times of gradient evaluation and 50% more steps to reach the convergence.

SR1, PSB, and Bofill are good candidates when updating Hessian matrix with negative eigenvalues. Relatively, Bofill and PSB have the better performance than SR1 with near 50% less gradient evaluation and 25% less iteration steps. This advantage is largely accredit to the rank-two level update. In rank-two update, the Hessian matrix is updated by two rank-one matrices. Rank-two updates produces a more accurate Hessian matrix approximation with slightly extra computation power. However, as the major time-consuming step in the geometry optimization is the quantum chemistry computing procedure. The extra cost for the rank-two hessian correction is actually insignificant.

The difference between the two trust-radius update method is not very conspicuous. The gradient-based method lead the result by a tiny margin. At most optimization task, the initial guess generated by *GOpt* is a good strcture not too far from the transition state. When the quadratic approximation is moderate, both energy-based and gradient-base update schemes fulfill the tasks. It is normally more dirable using gradient-based update when the initial structure are far from

the transition state. It's sensible to use energy-based method for minimization as any energy decrease direction is preferable while in transition state optimization, not only the change of energy, gradient change magnitude but also the direction of the change have a great impact on the optimization result.

The conclusion are drawn from the *GOpt* program calculation, but the efficiency of different methods is also applied to other quantum chemistry optimization package using the similar technique. Based on the results across different methods, we also setup the default choice for users with the general optimization purpose while still reserving customization to advanced user. For minimization task, the default combination is secant condition 1, BFGS method, and energy-based trust-radius update method while for transition state searching, Bofill method, and gradient-based trust-radius update are deployed with the same secant condition.

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.