

Kapitel 4

Geometrieoptimierung

4.1 Einleitung

Die Geometrie, die ein System bei tiefen Temperaturen im Gleichgewicht einnimmt, ist durch ein Minimum auf der Grundzustandsenergiefläche (1.2) ($n = 0$) charakterisiert. Im allgemeinen Fall finden sich mehrere lokale Energieminima, deren Gesamtenergien teilweise nur wenig voneinander abweichen. Dann kann das System mit endlicher Wahrscheinlichkeit in jeder dieser Strukturen, Isomere genannt, auftreten. Es sollte an dieser Stelle betont werden, dass im Gleichgewicht nicht die *innere Energie* minimiert wird, sondern je nach Randbedingung die *freie Energie*, *Enthalpie* oder *freie Enthalpie*. Gleichgewichtsgeometrien werden vom System erst angenommen, wenn die innere Energie, z.B. durch Energieabgabe an die Umgebung, soweit abgesunken ist, dass sich die verbleibende Energie nur noch in Vibrationen der Atomkerne um ihre Gleichgewichtslagen äußern kann. Die Gleichgewichtsgeometrie entspricht dann der hypothetischen, starren Struktur des Systems am absoluten Nullpunkt. Hypothetisch deshalb, weil auch am absoluten Temperaturnullpunkt infolge der Heisenbergschen Unschärferelation immer noch die Nullpunktsschwingungen im System auftreten.

Die Gleichgewichtsgeometrie ist durch das Verschwinden des Gradienten an die Potentialhyperfläche, d.h. aller Kräfte auf die Atomkerne, charakterisiert. Zu diesem Zweck muss ein Ausdruck für die Ableitung der Gesamtenergie nach den Kernkoordinaten gefunden werden. Da im Programmpaket *StoBe* sowohl die Elektronendichte als auch die Austauschkorrelationsenergiedichte mithilfe von analytischen Anpassungen als Linearkombinationen von Gaußfunktionen dargestellt werden [70–72], ist eine solche Ableitung für jede Kombination von Austauschkorrelationsfunktionalen analytisch möglich. Darüberhinaus stellt sich heraus, dass die Ableitungen der Orbitalkoeffizienten sowie der zwei Typen von Anpasskoeffizienten nach den Kernpositionen wegfallen, so dass ein störungstheoretischer Ansatz nicht notwendig ist. Dieser Umstand ist eine direkte Folge des Hellmann-Feynman-Theorems. Es bleiben dann nur noch die Pulay-Korrekturen, die infolge der Abhängigkeit des Basissatzes von den Atomkernpositionen auftreten. Die in *StoBe* verwendete Energieableitung ist in [73] hergeleitet.

Zum Finden der Gleichgewichtsgeometrien gibt es eine Reihe von unterschiedlichen Verfahren, deren Grundschema aber meistens nach folgendem Muster abläuft:

1. Bestimmung der elektronischen Struktur durch selbstkonsistente Lösung der KS-Gleichungen für eine vorgegebene Geometrie,
2. Analytische Berechnung des Energiegradienten $\mathbf{g} = \frac{\partial E}{\partial \mathbf{R}}$ und eventuell der Hesse-Matrix $\mathbf{H} = \frac{\partial^2 E}{\partial \mathbf{R}^2}$, gegebenenfalls unter Berücksichtigung von Randbedingungen (z.B. Fixierung von Atomen, starre Atomgruppen)¹,

¹Der Vektor \mathbf{R} ist hier als ein $3N$ -dimensionaler Positionsvektor zu verstehen, der die Geometrie der N Atomkerne definiert.

3. Extrapolation der Geometrie zur Minimierung des Gradienten, falls er oberhalb einer bestimmten Schranke ist, und Rückkehr zu Punkt 1.

Die Methoden unterscheiden sich darin, wie die Extrapolation in 3 durchgeführt wird und welche Größen in 2 dazu berechnet werden müssen.

4.2 Newton-Raphson-Methode

Der in dieser Arbeit verwendete Algorithmus ist eine Abwandlung der Newton-Raphson-Methode, bei der anhand der ersten und zweiten Ableitung \mathbf{g}_k und \mathbf{H}_k ² der Energiefunktion diese um die aktuelle Geometrie \mathbf{R}_k (die aktuelle Geometrie) bis zur zweiten Ordnung entwickelt wird, wobei k den Index des Extrapolationsschrittes angibt:

$$E(\mathbf{R}) = E(\mathbf{R}_k) + \mathbf{g}_k \cdot (\mathbf{R} - \mathbf{R}_k) + \frac{1}{2} {}^t (\mathbf{R} - \mathbf{R}_k) \mathbf{H}_k (\mathbf{R} - \mathbf{R}_k). \quad (4.1)$$

Die Entwicklung (4.1) hat einen stationären Punkt ($\nabla E(\mathbf{R}) = 0$) bei der Geometrie

$$\mathbf{R}_{k+1} = \mathbf{R}_k - \mathbf{H}^{-1} \mathbf{g}_k, \quad (4.2)$$

für die im nächsten Schritt der Gradient \mathbf{g}_{k+1} und die Hesse-Matrix \mathbf{H}_{k+1} berechnet wird. (Ist (1.2) bereits eine quadratische Funktion in einem n -dimensionalen Raum reichen n Schritte zur Konvergenz, da die Hesse-Matrix nach n Schritten exakt definiert ist.)

Nirgends im Newton-Raphson-Verfahren wird aber vorausgesetzt, dass die Hesse-Matrix \mathbf{H} positiv-definit ist. Daher konvergiert das Verfahren streng genommen nicht unbedingt in Richtung eines lokalen Minimums (auch wenn dies in der Praxis der wahrscheinlichste Fall ist), sondern zum nächsten stationären Punkt auf der Energiehyperfläche, so dass dieses Verfahren auch zur Lokalisation von Übergangszuständen geeignet ist. Die Identifizierung des stationären Punktes als Minimum, Maximum oder Sattelpunkt kann dann mithilfe einer Vibrationsanalyse (siehe Kapitel 5) erfolgen, mit der sich die topologische Umgebung des stationären Zustandes sondieren lässt. Darüberhinaus lässt sich aus ihr auch entnehmen, wie die Geometrie weiter optimiert werden muss, um zum endgültigen Minimum (keine imaginären Frequenzen) oder zum endgültigen Übergangszustand (eine imaginäre Frequenz) zu gelangen.

4.3 Pseudo-Newton-Raphson-Methode

Die Berechnung der Hesse-Matrix stellt sich jedoch bei der Verwendung von lokalisierten Basisfunktionen als extrem aufwändig heraus [75]. Daher ist in *StoBe* die sogenannte Pseudo-Newton-Raphson-Methode implementiert, die keine analytische Berechnung der Hesse-Matrix erfordert. Stattdessen wird die Hesse-Matrix \mathbf{H}_k bei jedem Schritt k so aktualisiert, dass sämtliche bereits berechneten Gradienten $\mathbf{g}_{l \leq k}$ kongruent zu ihr sind. Das heißt, dass die aktualisierte Hesse-Matrix \mathbf{H}_{k+1} der Bedingung

$$\mathbf{g}_l - \mathbf{g}_{l-1} = \mathbf{H}_{k+1} (\mathbf{R}_l - \mathbf{R}_{l-1}) \quad (4.3)$$

für $l = 1, \dots, k$ gehorcht. Die neue Geometrie \mathbf{R}_{k+1} wird dann aus der Gleichung (4.2) mit $\mathbf{H} = \mathbf{H}_{k+1}$ ermittelt. Auf diese Weise ergeben sich nur genäherte Hesse-Matrizen, die aber, wie sich herausstellt, für den Zweck der Geometrieoptimierung ausreichen. Die Bedingung (4.3) definiert die mathematische Vorschrift, anhand der die Hesse-Matrix (oder ihre Inverse) bei jedem Schritt aktualisiert wird, nicht exakt. Es gibt daher verschiedene Typen von Pseudo-Newton-Raphson-Methoden, z.B. Broyden-Fletcher-Goldfarb-Shanno (BFGS), Davidon-Fletcher-Powell (DFP) und Murtaugh-Sargent (MS) [76], die sich in der Art und Weise unterscheiden, wie die Hesse-Matrix aktualisiert wird.

²Eine Ableitung der analytischen zweiten (und dritten) Energieableitung in Abhängigkeit der Kernpositionen findet sich in [74].

Es hat sich herausgestellt, dass alle Algorithmen bei der Suche nach lokalen Minima auf der Energiehyperfläche ein gutes Konvergenzverhalten zeigen. Übergangszustände ließen sich jedoch nur mit dem Algorithmus von Murtaugh und Sargent finden, der im Folgenden exemplarisch vorgestellt wird.

Beim Murtaugh-Sargent-Algorithmus wird anstelle der Hesse-Matrix \mathbf{H} ihre Inverse \mathbf{H}^{-1} aktualisiert, so dass die neue Matrix \mathbf{H}_{k+1}^{-1} der Bedingung

$$\mathbf{R}_l - \mathbf{R}_{l-1} = \mathbf{H}_{k+1}^{-1} (\mathbf{g}_l - \mathbf{g}_{l-1}) \quad (4.4)$$

für $l = 1, \dots, k$ gehorcht. Dadurch erspart man sich die Invertierung der Matrix in (4.2). Die aktualisierte Inverse \mathbf{H}_{k+1}^{-1} ergibt sich aus \mathbf{H}_k^{-1} anhand der Rekursionsformel

$$\mathbf{H}_{k+1}^{-1} = \mathbf{H}_k^{-1} + \frac{(\mathbf{R}_k - \mathbf{R}_{k-1} - \mathbf{H}_k^{-1} (\mathbf{g}_k - \mathbf{g}_{k-1})) \cdot {}^t (\mathbf{R}_k - \mathbf{R}_{k-1} - \mathbf{H}_k^{-1} (\mathbf{g}_k - \mathbf{g}_{k-1}))}{{}^t (\mathbf{R}_k - \mathbf{R}_{k-1} - \mathbf{H}_k^{-1} (\mathbf{g}_k - \mathbf{g}_{k-1})) (\mathbf{g}_k - \mathbf{g}_{k-1})}, \quad (4.5)$$

die (4.4) erfüllt.

