A Multilevel Method for the Global Optimization of Potential Energy Functions

Nadine Pastuszka

Born 7th June 1989 in Brühl, Germany

29th April 2014

Master's Thesis Mathematics

Advisor: Prof. Dr. Michael Griebel

Institute for Numerical Simulation

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Contents

1 Introduction

Global optimization is of great relevance in many different areas of current research. Many problems are based on complex processes and their description often requires numerous parameters. The goal of global optimization is the determination of the best possible value for each of them. Examples can be found in the field of economics and finance which aims the maximization of the profit and sales. In statistics and data mining, the task is to find the optimal parameters in order to minimize the difference between modeled and empirical data.

In this thesis the area of materials science is considered dealing with the classification, the synthesis and the determination of properties of materials. Usually, their structure on the nanoscale, meaning all interactions between the corresponding atoms or molecules, is necessary in order to gain a better understanding of their effective macroscopic properties.

In this context, real experiments often lead to numerous problems such as a great amount of work, high expenses or even the fact that their conduct is impossible. These issues can be avoided by the use of computer simulations and on top of that they enable detailed predictions about the behavior of all corresponding particles. For the simulation of processes on the atomic level, there is the need of a mathematical model that describes the properties and interactions of atoms.

At that point, the laws of quantum mechanics have to be used according to which the quantum states evolution of time is described by the Schrödinger equation. The solution of this partial differential equation is a wave function that contains all measurable quantities of the underlying physical system, including the potential energy. The Schrödinger equation is so complex that it is impossible to be solved analytically, except for a few simple cases and a small number of particles. The main problem in this context is its high dimension. However, often thousands or millions of particles have to be taken into account in applications and consequently a number of approximations have to be applied in order to be able to compute an approximated solution. The determination of such approximations nevertheless remains to be a difficult task and involves a great computational effort.

All interactions within a physical system depend on the force acting on each particle. Its computation is based on a potential energy function, which can be determined via the potential energy information from the approximated solution of the Schrödinger equation.

The force acting on each particle is obtained by the partial derivative of the potential energy function with respect to the corresponding atom coordinates. Thus, every minimum of the potential energy function represents a stable configuration in which all of the atoms are in a certain equilibrium between attractive and repulsive forces. If it is possible, the configuration of a physical system always transitions into a state with a lower energy value, a minimum of the potential energy function therefore represents a preferred configuration state. Its stability depends on the properties of the corresponding minimum and its energy value. The global minimum of the potential energy function, the so-called *ground state*, is the most stable atom configuration since there is no other state with a lower energy value.

Applications often include many-body systems with a large number of atoms. All upcoming interactions for each possible atom configuration have to be taken into account and this

number grows rapidly with an increasing number of interacting particles. This is called the *curse of dimension*. Since additionally each approximated solution of the Schrödinger equation requires high computational effort, a classical discretization of potential energy functions is impossible.

The problem can be tackled by the use of a dimension-wise decomposition that represents the original function by its lower-order contributions which can be interpreted as the interaction of the different particles with one another. Examples for this lower-order contributions in the context of the potential energy are given by single particle energy contributions or pair- and triple-interactions. In many of the applications it suffices to take into account up to at most quadruple-interactions in order to approximate the underlying potential energy functions well by a dimension-wise decomposition. Nevertheless, it can be a very complex issue to describe all relevant interactions in a physical system and still, a very complex structured potential energy function has to be assumed.

Due to the complex structure of potential energy functions, their global optimization is a very demanding task. However, the knowledge of the ground state is of great relevance for experimental and theoretical investigations in science and efficient methods in order to solve this global optimization problem are of great interest.

Problem Definition

Potential energy functions play an important role in the context of the effective macroscopic properties of a physical system such as molecules or clusters. In the context of optimization, it must be emphasized that smoothness can be presumed for potential energy functions but neither linearity nor convexity. In the case of strong convex functions it holds that any local minimum is also the global one and given a computed minimum there is a possibility to decide whether it is a solution or not. Furthermore, there are several convergence criteria allowing statements about the maximal number of optimization steps needed until the global minimum is reached. On the other hand, the situation is completely different with non-convex functions for which multiple local minima have to be assumed. In general there are no precise statements about the actual number of these local minima, except for simple structured functions such as polynomials.

These issues apply in particular for potential energy functions, since they are non-convex and very complex. Without knowing the actual number of local minima, it is assumed to grow exponential with the number of atoms in the underlying physical system. Considering systems with a large number of particles this is an enormous number and function or derivative evaluations require high computational costs. Thus, there is no possibility of searching for all local minima in order to find the global one which leads to the lowest energy value. This yields the main problem of finding an optimization method which finds the smallest minimum out of the numerous local ones. There are no exact statements when or even whether a method convergences towards the global minimum and without further assumptions, there is no possibility of verifying a computed solution.

Thus, it is of great interest to find methods that search efficiently for the minimal function value and can distinguish as best as possible between local and global minima in order to solve the resulting high dimensional, non-linear, non-convex global optimization problem.

Solution Approaches

Methods for global optimization can be based on many different principles. Concerning the minimization of potential energy functions algorithms which are motivated by biological or physical processes and include stochastic elements have been successful.

In this thesis another approach from the field of subspace correction methods will be considered. It takes advantage of the potential energy functions special construction, based on the dimension-wise decomposition in its lower-order contributions. The main idea is a partition of the original problem into lower-dimensional subproblems via space decomposition. These subproblems are solved iteratively, while ensuring that the intermediate result in each iteration step does not get worse. The key aspect is to perform a correction of the current solution candidate in each subspace in order to get an approximated solution for the original problem.

Concerning the minimization of potential energy functions, the subspaces will be chosen such that the different contributions of the dimension-wise decomposition are considered separately in each iteration, which means that only a specific atom subset is optimized in each iteration step. In the area of materials science many-body systems with a large number of atoms are of great interest and thus, the main advantage using the subspace correction approach is the opportunity to reduce the dimension of the optimization problem enormously. One of the central issues of this thesis is the choice of subspaces and optimization procedures in order to get an efficient global optimization algorithm.

Applying the subspace correction approach, the interactions of specific atom sets are optimized iteratively, but it is impossible to change all positions at once in one iteration step. Considering the case that the optimal configuration would be reached by multiplying all coordinates of the starting configuration with an appropriate scaling factor. Since the optimization only affects subsets of atom coordinates, the subspace correction method is unable to use such information. Examples from the field of materials science are given by crystal structures which are characterized by a periodic atom configuration, a so-called crystal lattice. They include lattice vibrations which are wave-like perturbations that can be expressed by an appropriate expression such as a sinus function. In order to improve the efficiency of the global optimization of their potential energy functions, a multilevel method based on two different coordinate transformations will be developed. Applied to some initial atom configuration, the transformations are able to change all atom positions dependent on the coefficients of a certain orthogonal basis expansion, as employed in spectral methods. Motivated by the lattice vibrations we will choose the Fourier basis.

In general, such expansions are defined by an infinite sum which has to be cut after a certain number of terms. Initially, it is not known how many terms are necessary in order to get a good approximation of the transformation function. The approach will consist of a multilevel method in which different restrictions of the number of terms in the orthogonal basis expansion represent different levels. Furthermore, a sparse grid approach for the reduction of the number of coefficients to be optimized will be presented in order to improve the efficiency.

For the minimization of potential energy functions the optimization problem changes in the following sense. Instead of searching for the optimal atom configuration directly, the coefficients of the respective transformation functions are optimized. Applied to an initial configuration, the optimal transformation function then gives the optimal atom positions.

The final multilevel optimization method will be a combination of the transformation approach and the direct optimization of the atoms coordinates.

All algorithms were implemented in MATLAB and we used the NLopt nonlinear-optimization package from Steven G. Johnson, see [Johnson, 2013].

Own Contributions

In the following, there is an overview of the contributions established by this work.

- Application of the subspace correction for optimization to potential energy functions.
- Development of a new multilevel algorithm based on a coordinate transformation including an orthogonal basis expansion.
- Implementation of the subspace correction as optimization method.
- Implementation of a new multilevel optimization method.

Structure of the Thesis

This thesis is structured as follows.

Chapter 2 describes a general approach for dimension-wise decompositions of high dimensional functions. Furthermore, the ANOVA decomposition and the high dimension model reduction are described as special cases which both can be applied to potential energy functions.

In **Chapter 3** the development of potential energy functions is described based on the approximated solution of the Schrödinger equation, the Born-Oppenheimer approximation and the dimension-wise decomposition. The ground state is explained as the most preferred configuration state which is given by the global minimum of the potential energy function. Furthermore, some examples for potentials are presented.

Chapter 4 gives a description of high-dimensional, non-linear, non-convex global optimization and the resulting problems. Some known worst complexity results connected to this topic are presented, including results for dimension-wise decomposed functions. The subspace correction method for global optimization is motivated and described. Its definition is based on a decomposition of the original problem into lower-dimensional subproblems via space decomposition. Regarding the global optimization of high-dimensional functions, it is a decisive advantage that only the lower-dimensional subproblems have to be solved in order to compute an approximated solution for the original problem. Afterwards, the application of the subspace correction method on the optimization of potential energy functions is presented. In order to improve the efficiency of the optimization a new multilevel optimization method for global optimization is presented, based on a coordinate transformation.

Chapter 5 contains the numerical results using the subspace correction optimization for the example of Lennard-Jones and Gupta clusters. Furthermore, the coordinate transformation based optimization is tested for a bond potential and Lennard-Jones clusters including periodic boundary conditions.

Chapter 6 provides a summary of this thesis and suggestions for further works.

2 Representation of High Dimensional Functions

Many problems in current research are based on complex processes, thus we often need a lot of parameters in order to describe them accurately enough by a function.

We take advantage of the fact that for many problems the parameters affect each other more or less strongly and thus the individual terms contribute differently to the total function. Under this assumption, we can introduce weights which represent the relative importance for each parameter set or we can omit those terms that only make a small contribution to the total function.

In this master thesis we consider the problem of global minimization of functions which describe the potential energy of a physical system, dependent on the spatial coordinates of the corresponding atoms. Due to the fact that systems of practical relevance often contain a large number of atoms, we have to assume high dimensional functions of possibly more than thousand variables. The approach to reduce the complexity of those functions is a dimension-wise decomposition that represents the original function by a sum of its lower-order contributions, which can be interpreted as the interaction of the different parameters with one another.

In the following sections, we describe the derivation of a general dimension-wise decomposition, its minimality and we define the notion of effective dimensions in this context. The ANOVA decomposition and the high dimensional model reduction (HDMR) will be presented as special cases. We follow the works of [F. Y. Kuo and Wozniakowski, 2010] and [Holtz, 2008].

2.1 General Dimension-wise Decomposition

Let $\Omega^d \subseteq \mathbb{R}^d$ be a domain and f be a d-dimensional function in the linear space of all real functions $\mathcal{F}$ defined on D. We will use the notation $f(\mathbf{x}) = f(x_1, \ldots, x_d)$ and we define the independence of $f \in \mathcal{F}$ on the variable x_j for some $j \in \{1, \ldots, d\}$ by the condition

$$\mathbf{x}_{\{1,\ldots,d\}\setminus\{j\}} = \mathbf{y}_{\{1,\ldots,d\}\setminus\{j\}} \implies f(\mathbf{x}) = f(\mathbf{y}) \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in D. \tag{2.1}$$

Furthermore, we denote the restriction of $\mathbf{x}$ onto a subset of its indices $\mathbf{u} \subseteq \{1, \ldots, d\}$ by $\mathbf{x}_\mathbf{u}$, that is $\mathbf{x}_\mathbf{u} = (x_j)_{j \in \mathbf{u}}$.

Next, we introduce the commuting projection P_j which, applied on $f \in \mathcal{F}$, leads to a function that is no longer dependent on the j-th variable x_j while all other variables remain unchanged. The vanishing of the variable x_j is achieved by a certain method, for example by setting x_j to a constant value or by integrating out x_j. The set of commuting projections $\{P_j\}_{j=1}^d$ for all variables $x_1, \ldots, x_d$ thus fulfills

$$P_j(f) = f \quad \text{if } f \text{ is independent of } x_j, \text{ and this also holds for } P_j(f).$$

We may have to take into account certain restrictions on the domain Ω^d or the function space $\mathcal{F}$, according to the choice of the P_j. For example, if we choose P_j as a method that integrates out x_j, we have to assume integrability of f on Ω^d.

We define the projection for a set $\mathbf{u} \subseteq \{1, \ldots, d\}$ by $P_{\mathbf{u}} = \prod_{j \in \mathbf{u}} P_j$. Moreover, we set $P_\emptyset := I$ and define the composition of two projections $P_{\mathbf{u}}$ and $P_{\mathbf{v}}$ by the projection induced by their union, that is $P_{\mathbf{u}} P_{\mathbf{v}} = P_{\mathbf{u} \cup \mathbf{v}}$ for all $\mathbf{u}, \mathbf{v} \subseteq \{1, \ldots, d\}$.

In the context of the function space $\mathcal{F}$ we will use the notation $\mathcal{F}_{\mathbf{u}}$ for the restriction of functions $f \in \mathcal{F}$ to the subset $\mathbf{u} \subseteq \{1, \ldots, d\}$. Thus, the $f_{\mathbf{u}}$ are only dependent on the subset $\mathbf{x}_{\mathbf{u}}$ of all variables $\mathbf{x} = (x_1, \ldots, x_d)$ of f and the remaining variables have fixed values.

In the following theorem, taken from [F. Y. Kuo and Wozniakowski, 2010], we present a decomposition formula for f in terms of different $f_{\mathbf{u}}$'s and equivalent representations of the functions $f_{\mathbf{u}}$.

Theorem 2.1. *Let $\{P_j\}_{j=1}^d$ be commuting projections on $\mathcal{F}$ satisfying (2.1). For every function $f \in \mathcal{F}$ and each subset $\mathbf{u} \subseteq \{1, \ldots, d\}$, we define*

$$f_{\mathbf{u}} := \left(\prod_{j \in \mathbf{u}} (I - P_j) \right) P_{\{1,\ldots,d\} \setminus \mathbf{u}}(f). \tag{2.2}$$

This leads to a decomposition of f given by

$$f = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} f_{\mathbf{u}}, \tag{2.3}$$

where $f_{\mathbf{u}}$ depends only on variables with indices in $\mathbf{u}$.

a) Define the functions $g_{\mathbf{u}}$ recursively with respect to the cardinality of $\mathbf{u} \subseteq \{1, \ldots, d\}$ by

$$g_\emptyset := P_{\{1,\ldots,d\}}(f), \qquad g_{\mathbf{u}} := P_{\{1,\ldots,d\} \setminus \mathbf{u}}(f) - \sum_{\mathbf{v} \subsetneq \mathbf{u}} g_{\mathbf{v}}.$$

Then $f_{\mathbf{u}} = g_{\mathbf{u}}$ for all $\mathbf{u} \subseteq \{1, \ldots, d\}$.

b) Define the functions $h_{\mathbf{u}}$ for $\mathbf{u} \subseteq \{1, \ldots, d\}$ by

$$h_{\mathbf{u}} := \sum_{\mathbf{v} \subsetneq \mathbf{u}} (-1)^{|\mathbf{u}| - |\mathbf{v}|} P_{\{1,\ldots,d\} \setminus \mathbf{v}}(f).$$

Then $f_{\mathbf{u}} = h_{\mathbf{u}}$ for all $\mathbf{u} \subseteq \{1, \ldots, d\}$.

c) Assume that $f = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} w_{\mathbf{u}}$, where $w_{\mathbf{u}}$ depends only on the variables with indices in the set $\mathbf{u}$ and satisfies the "annihilating" property

$$P_j(w_{\mathbf{u}}) = 0 \quad \text{for all} \quad j \in \mathbf{u}. \tag{2.4}$$

Then $f_{\mathbf{u}} = w_{\mathbf{u}}$ for all $\mathbf{u} \subseteq \{1, \ldots, d\}$. In other words (2.3) is the unique decomposition of this form with the annihilating property.

For the proof, see [F. Y. Kuo and Wozniakowski, 2010].

This theorem enables us to determine the individual terms of the decomposition (2.3) recursively or directly. In the next section we will describe the meaning of the preference to lower-order terms in this context.

2.1.1 Minimality

The dimension-wise decomposition (2.3) with terms defined as in theorem 2.1 has the property that we never include unnecessary subsets $\mathbf{u}$ and corresponding terms $f_{\mathbf{u}} \neq 0$ into the sum. With unnecessary, we mean those terms which include variables that are canceled out by the variables of other terms. For example, for $f_{\{1\}} = x_1$, $f_{\{2\}} = -x_2$ and $f_{\{1,2\}} = -(x_1 - x_2)$, we obtain $f_{\{1\}} + f_{\{2\}} + f_{\{1,2\}} = x_1 - x_2 - (x_1 - x_2) = 0$ and thus, we can omit these terms. In other words, if there exists a way to express a higher-order term in terms of lower-order terms, they are left out. This way, we ensure giving preference to lower-order terms, independent of the choice of the projection $\{P_j\}_{j=1}^d$. We state this minimal property in the following theorem from [F. Y. Kuo and Wozniakowski, 2010].

Theorem 2.2. *Let f be a function from $\mathcal{F}$, $\mathbf{z} \subseteq \mathbf{u} \subseteq \{1, \ldots, d\}$ with $\mathbf{z}$ given. Assume that there exists a decomposition*

$$f = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} t_{\mathbf{u}} \qquad \text{and} \qquad t_{\mathbf{u}} = 0 \text{ for all } \mathbf{u} \text{ containing } \mathbf{z}.$$

Then the decomposition from theorem 2.1 also fulfills

$$f_{\mathbf{u}} = 0 \text{ for all } \mathbf{u} \text{ containing } \mathbf{z}.$$

See [F. Y. Kuo and Wozniakowski, 2010] for the proof.

This minimality property motivates the definition of a number which gives us the maximal number of terms required for an exact representation or at least an approximation of the function f. We will give the different definitions of such numbers by explaining the effective dimension of f.

2.1.2 Effective Dimensions

For many functions it holds that the higher-order terms make much less contribution to the total function than the lower-order terms. In order to solve problems based on high-dimensional functions, it can be of great advantage if we can leave out several higher-order terms and thus reduce the dimension of f.

There exist different notions of the dimension of a d-dimensional function f connected with the decomposition (2.3).

Cutoff Dimension/ Truncation Dimension

The cutoff dimension describes the smallest number $k \leq d$ for which a function f of d variables has a decomposition

$$f = \sum_{\mathbf{u} \subseteq \{1,\ldots,k\}} f_{\mathbf{u}}.$$

If we allow a predefined error in this truncated decomposition that is defined in a certain norm, we denote k as the truncation dimension [Holtz, 2008].

Order/Superposition Dimension

The order of f is defined as the smallest number $q \leq d$ for which the decomposition can be written as a sum of functions which depend on at most q of all d variables, formally

$$f = \sum_{\substack{\mathbf{u} \subseteq \{1,\dots,d\} \\ |\mathbf{u}| \leq q}} f_{\mathbf{u}}. \tag{2.5}$$

As in the definition of the truncation dimension, including a predefined error of the representation (2.5), q is defined as the superposition dimension [Holtz, 2008].

In the next section, we are going to describe the ANOVA decomposition as a special case of the decomposition (2.3).

2.2 The ANOVA Decomposition

The ANOVA decomposition describes the representation of a square-integrable function defined on the d-dimensional domain Ω^d. In particular, we can decompose $f \in \mathcal{L}^2(\Omega^d)$ into a sum of 2^d functions $f_{\mathbf{u}}$ defined on $|\mathbf{u}|$-dimensional subspaces $\Omega^{|\mathbf{u}|}$, which are induced by a subset $\mathbf{u} \subseteq \{1, \dots, d\}$.

The so-called ANOVA terms $f_{\mathbf{u}} : \Omega^{|\mathbf{u}|} \to \mathbb{R}$ are only dependent on the subset $\mathbf{x}_{\mathbf{u}}$ of variables of f and can be interpreted as the contribution of each parameter set given by $\mathbf{u}$. According to the notation of the previous section, we define $\mathcal{F}$ as the function space $\mathcal{L}^2(\Omega^d)$, containing all d-dimensional square-integrable, real functions. The set of projections $\{P_j\}_{j=1}^d$ is characterized by an integration with respect to a product measure. The ANOVA decomposition of f is then given by (2.3).

The subspaces are defined by the $|\mathbf{u}|$-dimensional product set of connected sets $\Omega_j \subseteq \mathbb{R}$, that is

$$\Omega^{|\mathbf{u}|} := \bigotimes_{j \in \mathbf{u}} \Omega_j \subseteq \mathbb{R}^{|\mathbf{u}|}.$$

Furthermore, we define

$$d\mu(\mathbf{x}) = \prod_{j=1}^{d} d\mu_j(x_j)$$

as the d-dimensional product measure defined on Borel subsets of Ω^d with $\mu_j, \;\; j = 1, \dots, d$, describing probability measures on Borel subsets $\mathcal{B}(\Omega_j)$ of Ω_j.

We further consider $\mathcal{F}$ as the Hilbert space $V^{(d)}$ of all functions $f : \Omega^d \to \mathbb{R}$ equipped with the $\mathcal{L}^2$-inner product

$$(f, g) := \int_{\Omega^d} f(\mathbf{x}) g(\mathbf{x}) d\mu(\mathbf{x}),$$

two functions f and g are therefore called orthogonal if $(f, g) = 0$.

With this, we can define the projection $P_{\mathbf{u}} : V^{(d)} \to V^{(d-|\mathbf{u}|)}$ of a square-integrable function f, which is induced by the measure μ in the following way

$$P_{\mathbf{u}} f(\mathbf{x}_{\mathbf{u}}) := \int_{\Omega^{|\mathbf{u}|}} f(\mathbf{x}) d\mu_{\mathbf{u}}(\mathbf{x})$$

with $\mathbf{u} \subseteq \{1, \dots, d\}$ given, $\mathbf{x_u} \in \mathbb{R}^{|\mathbf{u}|}$ as the vector containing $\{x_j : j \in \mathbf{u}\}$ and $d\mu_{\mathbf{u}}(\mathbf{x}) := \prod_{j \in \mathbf{u}} d\mu_j(x_j)$.

For a fixed measure μ we can represent $f \in V^{(d)}$ by the decomposition (2.3), and the corresponding ANOVA terms can be defined according to theorem 2.1 recursively by *a)*, directly by *b)* or they can be characterized by the annihilating property *c)*.

Furthermore, the ANOVA decomposition has the property that it is orthogonal in the sense that

$$(f_{\mathbf{u}}, f_{\mathbf{v}}) = \int_{\Omega^d} f_{\mathbf{u}}(\mathbf{x_u}) f_{\mathbf{v}}(\mathbf{x_v}) d\mu(\mathbf{x}) = 0 \quad \text{for all } \mathbf{u} \neq \mathbf{v}.$$

For the proof and more details see for example [Oettershagen, 2011] and the references therein.

The Classical ANOVA Decomposition

The classical ANOVA decomposition is given by the Lebesgue measure $\mu = \lambda$ and $d\mu(\mathbf{x}) = d\mathbf{x}$ or a measure that is absolutely continuous with respect to the Lebesgue measure, denoted by $\mu_i = \varphi_i \lambda$ for $i = 1, \dots, d$ and by $\mu_{\mathbf{u}} = \varphi_{\mathbf{u}} \lambda$ for a subset $\mathbf{u} \subseteq \{1, \dots, d\}$. We obtain the projection for the Lebesgue measure

$$P_{\mathbf{u}} f(\mathbf{x_u}) = \int_{\Omega^{|\mathbf{u}|}} f(\mathbf{x})\, d\mathbf{x_u}$$

and for the absolutely continuous measure

$$P_{\mathbf{u}} f(\mathbf{x_u}) = \int_{\Omega^{|\mathbf{u}|}} f(\mathbf{x}) \varphi^{\mathbf{u}}(\mathbf{x})\, d\mathbf{x_u}.$$

Using the transformation theorem, we can compute each ANOVA term $f_{\mathbf{u}}$ by an integral over the unit cube $[0,1]^d$. For a more detailed explanation, see for example [Oettershagen, 2011].

The ANOVA terms $f_{\mathbf{u}}$ for the Lebesgue measure with the domain $\Omega^d = [0,1]^d$ have the following properties from theorem 2.1

a)

$$f_{\emptyset} = \int_{[0,1]^d} f(\mathbf{x})\, d\mathbf{x}, \quad \text{and} \quad f_{\mathbf{u}}(\mathbf{x_u}) = \int_{[0,1]^{d-|\mathbf{u}|}} f(\mathbf{x})\, d\mathbf{x}_{\{1,\dots,d\}\setminus \mathbf{u}} - \sum_{\mathbf{v} \subsetneq \mathbf{u}} f_{\mathbf{v}}(\mathbf{x_v}),$$

b)

$$f_{\mathbf{u}} = \sum_{\mathbf{v} \subsetneq \mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} \int_{[0,1]^{d-|\mathbf{v}|}} f(\mathbf{x})\, d\mathbf{x}_{\{1,\dots,d\}\setminus \mathbf{v}},$$

c)

$$\int_0^1 f_{\mathbf{u}}(\mathbf{x_u})\, dx_j = 0 \quad \text{for all} \quad j \in \mathbf{u},$$

see [F. Y. Kuo and Wozniakowski, 2010].

The Anchored ANOVA Decomposition

The anchored ANOVA decomposition is defined by the choice of the Dirac measure $\mu = \delta_{\mathbf{a}}$ at a fixed point $\mathbf{a} \in \Omega^d$ for μ. The projection is defined by

$$\begin{aligned} P_{\mathbf{u}} f(\mathbf{x}_{\mathbf{u}}) &= \int_{\Omega^{|\mathbf{u}|}} f(\mathbf{x}) \, d\delta_{\mathbf{u}}^{\mathbf{a}}(\mathbf{x}) \\ &= f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}}, \end{aligned}$$

with $f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}}$ denoting the evaluation at the point

$$\mathbf{x} = \begin{cases} x_j & \text{, for } j \notin \mathbf{u}; \\ a_j & \text{, for } j \in \mathbf{u}. \end{cases}$$

For this, we have to assume that for all $\mathbf{x} \in \Omega^d$ it holds that $f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}} \in \Omega^d$.

Thus, we can describe the corresponding ANOVA terms by the following equations

a)

$$f_{\emptyset} = f(\mathbf{a}), \quad \text{and} \quad f_{\mathbf{u}}(\mathbf{x}_{\mathbf{u}}) = f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}} - \sum_{\mathbf{v} \subsetneq \mathbf{u}} f_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}),$$

b)

$$f_{\mathbf{u}} = \sum_{\mathbf{v} \subsetneq \mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|} \cdot f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}},$$

c)

$$f(\mathbf{x})_{|\mathbf{x}_{\mathbf{u}}=\mathbf{a}_{\mathbf{u}}} = 0 \quad \text{whenever } x_j = a_j \quad \text{for all} \quad j \in \mathbf{u},$$

see [F. Y. Kuo and Wozniakowski, 2010].

It should be noted that for the anchored ANOVA decomposition, we only have to evaluate the function at the point $\mathbf{a}$ which is much cheaper than computing an integral as in the classical ANOVA decomposition. However, we have to take into account the fact that the approximation quality of the anchored ANOVA decomposition can be strongly dependent on the choice of the anchor $\mathbf{a}$, see for example [Zhang et al., 2011].

2.3 High Dimensional Model Reduction

In molecular chemistry, the decomposition (2.3) is often described as high dimensional model representation (HDMR), see for example [Hayes et al., 2006].

In this context, we describe the properties of a molecule by the function $f : \mathbb{R}^{3d} \to \mathbb{R}$, that means we consider a molecule of d atoms. The subsets $\mathbf{u} \subseteq \{1, \ldots, 3d\}$ can be interpreted as molecular subcomponents of the atoms with indices $i \in \mathbf{u}$ and the terms $f_{\mathbf{u}}$ are their contributions to the total function f.

The so-called cut-HDMR is similar to the anchored decomposition where we choose a reference point $\mathbf{a}$ and evaluate the different contributions at $\mathbf{a}$. The choice of the anchor or reference point and the resulting error estimates for the approximations are for example discussed in [Wang, 2008].

For molecules, it turns out that the contribution of higher-order terms of the decomposition (2.3) decreases rapidly such that we can assume a superposition dimension $q \ll d$.

We give a more detailed explanation of the determination of potential energy functions and the resulting dimension-wise decomposition in the next chapter.

3 Physical Background

In the field of materials science we are interested in the classification, the synthesis and the determination of properties of materials. In order to gain a better understanding of their effective macroscopic properties, we have to know their structure on the nanoscale. This means that we need information about all interactions between the corresponding atoms or molecules. With the help of simulations, we can avoid problems arising in the conduct of real experiments, different approaches can be tested without the need for synthesis of the material itself and thus expenses and a great amount of work in laboratories can be saved. Furthermore, we can observe the resulting intermediate steps more precisely than in a real experiment.

In a physical system all corresponding particles influence each other in a certain way. This is the reason why we need to study all their properties and interactions to be able to carry out computer simulations that describe their dynamic processes. In this context, the positions of the particles to one another and their behavior is of great importance.

In the following, we will pay closer attention to clusters which consist of a certain number of unbonded atoms and molecules consisting of two or more atoms linked by chemical bonds. Atoms are comprised of a positively charged nucleus that is surrounded by a system of negatively charged electrons. Electrons do not move around the nucleus along a fixed orbit and thus we cannot determine their exact location. This special electron behavior can be explained by the so-called *wave-particle dualism*, which implies that electrons show both particle-like and wave-like properties. Hence, we describe the nucleus surrounding system of electrons via so-called orbitals, which are wave functions that indicate the probability of presence for each electron. For simulations on the atomic level, we need to employ the laws of quantum mechanics which characterize electrons by wave functions in order to identify the probability that an electron is located at a particular location.

The underlying mathematical model is given by the Schrödinger equation from which we can derive all properties of the considered physical system. The issue at hand is that this differential equation is impossible to solve analytically in its original formulation, except for a few simple cases. Even the computation of numerical solutions is restricted to some simple physical systems with a small number of particles, far away from an interesting size.

The approach to solve this problem is the formulation of different approximations of the Schrödinger equation which are briefly presented in the next section.

The task in this thesis is to find the ground state and thus, the most stable configuration of physical systems such as molecules or clusters. Therefore, we consider the simplified problem definition of finding the configuration with the lowest potential energy value.

In this chapter we motivate the derivation of such energy functions based on the solution of an approximated formulation of the Schrödinger equation. Thereafter, we explain the importance of finding the most stable atom-configuration that corresponds to the global minimum of the underlying potential energy function and we present some examples for potential energy functions. We will mainly follow the work of [Griebel et al., 2007], where more detailed

information can be found.

3.1 The Schrödinger Equation

The Schrödinger equation is a partial differential equation that identifies the state of electrons and nuclei and has a complex-valued wave function as a solution. For a system of N nuclei and K electrons, we have the time-dependent state function

$$\Psi(\mathbf{R}_1, \dots, \mathbf{R}_N, \mathbf{r}_1, \dots, \mathbf{r}_K, t) = \Psi(\mathbf{R}, \mathbf{r}, t),$$

with $\mathbf{R}_i \in \mathbb{R}^3$, $i = 1, \dots N$, the spatial coordinates of the nuclei and $\mathbf{r}_j \in \mathbb{R}^3$, $j = 1, \dots K$, those of the electrons, t is the time. This means that the dimension of the space of configurations of all particles is of dimension $3(N+K)$. The probability density for finding a given configuration in a certain domain at a particular time is then obtained via the square of the absolute value of the wave function $\Psi^*\Psi$. Furthermore, we assume a normalized state function Ψ, that is $\int \Psi(\mathbf{R}, \mathbf{r}, t)^* \Psi(\mathbf{R}, \mathbf{r}, t)\ d\mathbf{R}\, d\mathbf{r} = 1$.

Without the consideration of spin and relativistic effects together with the assumption of no external influences on the system, its Hamilton operator is defined by

$$\mathcal{H} = \underbrace{-\frac{\hbar^2}{2\,m}\Delta}_{\text{kinetic energies}} + \underbrace{V(\mathbf{R}, \mathbf{r}, t)}_{\text{potential energies}},$$

with m as the particle mass and $\hbar = \frac{h}{2\pi}$, where h denoting Planck's constant. Both terms of the sum are operators with $-\frac{\hbar^2}{2\,m}\Delta$ describing the kinetic energy of the nuclei and electrons and $V(\mathbf{R}, \mathbf{r}, t)$ describing the potential energy between all interactions of nuclei and electrons, that means all nucleus-nucleus, electron-electron and nucleus-electron interactions.

With this, the time-dependent Schrödinger equation is given by

$$\mathrm{i}\hbar\ \frac{\partial \Psi(\mathbf{R}, \mathbf{r}, t)}{\partial t} = \mathcal{H}\Psi(\mathbf{R}, \mathbf{r}, t) \tag{3.1}$$

with i the imaginary unit and the state function Ψ as the solution.

Since the resulting problem is complex and of high dimension, it is impossible to solve it analytically for most cases. The exceptions are restricted to a few simple cases such as the solution of the Schrödinger equation for the hydrogen atom. This is the reason why we have to use approximations in order to reduce the complexity of this high dimensional partial differential equation.

The first approximation is the assumption that the Hamilton operator is not explicitly time-dependent. We use the following separation ansatz for the wave function

$$\Psi(\mathbf{R}, \mathbf{r}, t) = f(t)\ \cdot \tilde{\Psi}(\mathbf{R}, \mathbf{r}).$$

After plugging this into the time-dependent Schrödinger equation (3.1) and separation we obtain the following two equations

$$\mathrm{i}\hbar \frac{1}{f(t)}\ \frac{df(t)}{dt} = E \tag{3.2}$$

$$\tilde{\mathcal{H}}\tilde{\Psi}(\mathbf{R}, \mathbf{r}) = E\tilde{\Psi}(\mathbf{R}, \mathbf{r}). \tag{3.3}$$

Now, only equation (3.2) is dependent on the time t and we can state its general solution that is given by

$$f(t) = ce^{-iEt/\hbar}.$$

Equation (3.3) is an eigenvalue problem for $\tilde{\mathcal{H}}$ which is called the time-independent or stationary Schrödinger equation. We identify the wave function $\tilde{\Psi}$ as the eigenfunction with the corresponding eigenvalue E that describes the state of the underlying system.

The stationary Schrödinger equation is still very complex and this is why we have to use further methods in order to simplify the problem.

3.1.1 Born-Oppenheimer Approximation

The next simplification is based on the idea that we only take into account one state of the wave function $\tilde{\Psi}$. From a physical point of view, the most interesting eigenvalue is the smallest one, because then the solution describes the state with the smallest energy value, the so-called ground state which is the atoms most preferred one. Furthermore, this is also the most important eigenvalue with respect to optimization. In the following, we will denote the wave function in the ground state by ϕ_0 and the smallest eigenvalue with E_0.

We will make use of the Born-Oppenheimer approximation, other methods are given for example by the Car-Parinello method or the Ehrenfest molecular dynamics, see [Griebel et al., 2007].

The Born-Oppenheimer approximation exploits the awareness of the much smaller electron mass in contrast to the mass of the nuclei. Nevertheless, the electrons are subjected to comparable attractive and repulsive forces with the consequence that they react more strongly on external factors. We assume that the electrons conform to the nuclei positions in a negligible time and are therefore in the quantum mechanical ground state at any time, dependent on the configuration of the nuclei.

This is the motivation to split the Schrödinger equation into two equations, which then describe the state of electrons and nuclei separately. Due to the dependency of the electrons on the positions of the nuclei, these two equations are coupled.

Assuming fixed nuclear configurations, the electron-dependent part is then described by an effective potential given by the electronic Schrödinger equation. The nuclei are considered as classical particles and their interactions are computed by Newton's equations from classical mechanics. After several computations, we obtain the following equations

$$m_k \ddot{\mathbf{R}}_k(t) = -\Delta_{\mathbf{R}_k} \min_{\phi_0} \Big\{ \int \phi_0{}^*(\mathbf{R}(t), \mathbf{r}) \mathcal{H}_e(\mathbf{R}(t), \mathbf{r}) \phi_0(\mathbf{R}(t), \mathbf{r}) \; d\mathbf{r} \Big\} \tag{3.4}$$

$$\tilde{\mathcal{H}}_e \, \phi_0(\mathbf{R}(t), \mathbf{r}) = E_0\big(\mathbf{R}(t)\big) \, \phi_0(\mathbf{R}(t), \mathbf{r}), \tag{3.5}$$

with $\tilde{\mathcal{H}}_e$ describing the electronic Hamilton operator.

3.1.2 The Determination of Potential Energy Hypersurfaces

With (3.4) we can define the potential function $U^{BO} : \mathbb{R}^{3N} \to \mathbb{R}$, dependent on the positions of the nuclei and based on the Born-Oppenheimer approximation and restricted to the ground state as follows

$$U^{BO}\big(\mathbf{R}(t)\big) := \min_{\phi_0} \Big\{ \int \phi_0{}^*(\mathbf{R}(t), \mathbf{r}) \mathcal{H}_e(\mathbf{R}(t), \mathbf{r}) \phi_0(\mathbf{R}(t), \mathbf{r}) \; d\mathbf{r} \Big\}.$$

In order to compute this potential energy function, we need to solve the time-independent Schrödinger equation for its ground state. Note that (3.5) is still a complex problem because it is dependent on the number of electrons and thus of dimension $3K$. We still need to consider further simplifications in order to find approximated solutions.

For a further reduction of the problem size, several methods have been developed such as the density functional theory, the self consistent field method as a Hartree-Fock method and Post-Hartree-Fock methods, such as configuration interaction methods, Møller-Plesset methods and coupled cluster methods. Detailed explanations of these methods can be found for example in [Parr and Yang, 1989, Szabo and Ostlund, 2012, Møller and Plesset, 1934]. By using such methods, we have the possibility to solve (3.5) approximately but for increasing size of molecules these approximations fail and we have to consider further simplifications.

In order to determine an analytic potential function, many different approaches have been used, for example the so-called *ab initio methods* that use approximated solutions of (3.5) or the *fitting* of given analytic potentials to experimental or quantum mechanical results. For further information and other methods see [Griebel et al., 2008] and the references therein.

In the following, we present a method based on the dimension-wise decomposition as described in chapter 2. The main idea is to reduce the high number of variables of the potential function using the decomposition (2.3). Thus, we decompose a d-dimensional function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^d$, into the following sum

$$f = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} f_{\mathbf{u}}.$$

The subcomponents $f_{\mathbf{u}}$ are chosen hierarchically and are each dependent on the number of variables induced by the subset $\mathbf{u}$. If it is possible to achieve a good approximation of the complete energy function U in a few terms of low-dimensional parts, we can for example perform *ab initio* calculations just for these low dimensional parts and transfer the results to U.

Since the aim is to express the complete potential as a sum of different functions as fragments, we need to define them in more detail. We consider that the whole energy function can be interpreted as a sum of functions which is characterized by the contributions of all atoms acting independently, then all pair, triple- and quadruple-interactions and so on.

With $\mathbf{R} = (\mathbf{R}_1, \dots, \mathbf{R}_N)$, $\mathbf{R}_i \in \mathbb{R}^3$, describing the nucleus coordinates of all atoms we can formally write the dimension-wise decomposition the following way

$$U(\mathbf{R}) \approx V_0 + \sum_{i}^{N} V_i(\mathbf{R}_i) + \sum_{i,j}^{N} V_{i,j}(\mathbf{R}_i, \mathbf{R}_j) + \sum_{i,j,k}^{N} V_{i,j,k}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k) + \cdots + V_{1,2,\dots,N}(\mathbf{R}). \quad (3.6)$$

According to [Griebel et al., 2008], we can define the subcomponents by

$$\begin{aligned} V_0 &= 0 \\ V_i(\mathbf{R}_i) &= U_i(\mathbf{R}_i) - V_0 \\ V_{i,j}(\mathbf{R}_i, \mathbf{R}_j) &= U_{i,j}(\mathbf{R}_i, \mathbf{R}_j) - V_i(\mathbf{R}_i) - V_j(\mathbf{R}_j) - V_0 \\ &\vdots \end{aligned}$$

and in the case of the cut-HDMR, as in [Hayes et al., 2006] ,we choose a reference point $\bar{\mathbf{R}} \in \mathbb{R}^{3N}$ and define

$$\begin{aligned} V_0 &= U(\bar{\mathbf{R}}) \\ V_i(\mathbf{R}_i) &= U(\bar{\mathbf{R}}^i, \mathbf{R}_i) - V_0 \\ V_{i,j}(\mathbf{R}_i, \mathbf{R}_j) &= U((\bar{\mathbf{R}}^{ij}, \mathbf{R}_i, \mathbf{R}_j) - V_i(\mathbf{R}_i) - V_j(\mathbf{R}_j) - V_0 \\ &\vdots \end{aligned}$$

Here, $\bar{\mathbf{R}}^i$ is the notation for $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_N)$ with all variables set to their reference values, except for $\mathbf{R}_i$ and fixed by $\bar{\mathbf{R}}$. The notation $\bar{\mathbf{R}}^{ij}$ describes all variables set to their reference values, except for $\mathbf{R}_i$ and $\mathbf{R}_j$, and so on.

In both cases, the subcomponents describe the contributions of the atoms on which coordinates they depend. Therefore V_0 is a constant describing the contribution of each atom without considering interactions. The functions $V_i(\mathbf{R}_i)$ are each identified by the contribution of the i-th atom for $i = 1, \ldots, N$, the $V_{i,j}(\mathbf{R}_i, \mathbf{R}_j)$ describe all pair-interaction contributions, $V_{i,j,k}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k)$ all contributions of triple-interactions and analogously for the remaining terms of the sum. In this expansion we have 2^N different terms.

The crucial point in those decompositions is the fact that in most physical and chemical applications the terms of low order are the significant contributions for the resulting energy. So, the hope is that we only have to take into account q terms of the sum in (3.6), with $q \ll N$, and still have a good approximation for the many-body potential. According to chapter 2, we search for the superposition dimension q, that is

$$U \approx \sum_{\substack{\mathbf{u} \subseteq \{1,\ldots,N\} \\ |\mathbf{u}| \leq q}} V_{\mathbf{u}}.$$

Clearly, we can only take advantage of this decomposition if the higher-order terms only make a small contribution to the total energy. If it is the last term which is the most important one, the decomposition would be useless. We can motivate this approach by the good results in their application in molecular dynamics by assuming two-, three, four- or many-body potentials.

3.2 The Ground State

The goal of this thesis is the determination of the most stable configuration of a physical system described by a potential energy function. The corresponding particles always prefer those states which result in a low energy value. Whenever it is possible, there is a transition into a state with a lower energy value than the current one.

We obtain the force acting on each particle by the partial derivative of the potential energy function with respect to its spatial coordinates. Thus, each minimum of the potential energy function represents a state in which all atoms are in a certain equilibrium between attractive and repulsive forces. Since there is no other configuration with a lower energy value near a minimum, it represents a preferred and stable configuration state. The stability is dependent

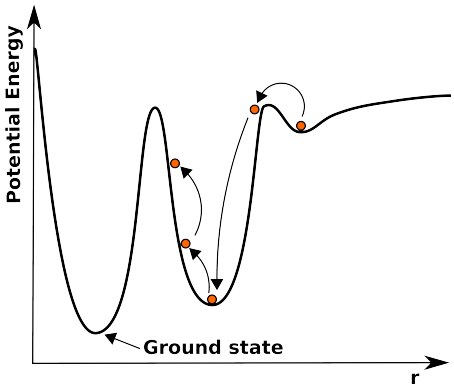


Figure 3.1: Energy barriers in a potential energy function, r describes the distance between two atoms.

on the properties of the corresponding minimum and the resulting energy value. The so-called *ground state* is the most stable atom configuration. It is given by the global minimum of the potential energy function because there exists no other state with a lower energy value. It should be noted that there may exist more that one ground state, those are called degenerate ground states. In this case, it is sufficient to find one on them.

A one-dimensional example is given by figure 3.1. The transition from a local minimum with low energy barriers into a smaller energy configuration requires only a small effort. The situation is different for the transition from a state with high energy barriers. In order to achieve another state with smaller energy value, the physical system previously has to accept higher energy values. This transition often requires external influences. An example for this given by the reduction of the temperature, the consequence being that the energy barriers become much lower.

Since the ground state is the most stable configuration, its knowledge is of great relevance for experimental and theoretical investigations in science. Furthermore, most of the substances, as they are found in nature, are in the ground state.

In order to determine the ground state, the task is to find the coordinates that give the minimal potential energy. Concerning the potential energy function U, we cannot make any further assumptions than smoothness, especially neither linearity nor convexity. The issue at hand is that we often have to consider a large number of particles and the corresponding potential energy function becomes high-dimensional and complex structured. Thus, the resulting nonlinear, non-convex global optimization problem is a very demanding task. A more detailed problem description will be presented in section 4.4.

3.3 Examples for Potentials

In this section, we will consider atomic clusters or molecules of N atoms in the three dimensional space. Potential energy functions can depend on the Cartesian coordinates of the corresponding atoms or on internal coordinates such as bond lengths, bond angles and torsions. In the following, we describe the potential energy by a function $U : \mathbb{R}^{3N} \to \mathbb{R}$, that

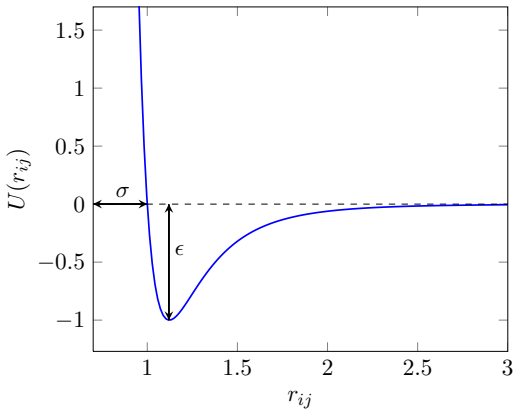


Figure 3.2: Lennard-Jones Potential with $\epsilon = \sigma = 1$.

gives the potential energy dependent on the coordinates of the N atomic nuclei $\mathbf{R}_i \in \mathbb{R}^3$, $i = 1, \ldots, N$.

For the rest of this chapter we will use the notation $r_{ij} := ||\mathbf{R}_j - \mathbf{R}_i||$ with $||\cdot||$ describing the Eucledian norm in order to describe the distance between two atoms.

Pair Potentials

The first example is a cluster that contains only two-body interactions that are described by pair potentials $U_{ij}^{pair} : \mathbb{R} \to \mathbb{R}$ for each atom-pair (i, j). The resulting energy function for the cluster is given by

$$U(\mathbf{R}_1, \ldots \mathbf{R}_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} U_{ij}^{pair}(r_{ij}). \tag{3.7}$$

Considering the ANOVA-decomposition (3.6) we truncate the sum after $q = 2$ terms and thus obtain a strong simplification. In the following, we present common examples for U_{ij}^{pair}.

The first one is the Coulomb Potential

$$U_{ij}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

with ϵ_0 describing the dielectric constant and q_i and q_j are the atomic charges.

The Lennard-Jones-$(12, 6)$-Potential is defined by

$$U^{LJ}(r_{ij}) = U_{ij}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{3.8}$$

where the parameter ϵ_{ij} describes the depth of the potential and σ_{ij} describes the distance with energy zero. The resulting function graph for the parameters $\epsilon := \epsilon_{ij} = 1$ and $\sigma := \sigma_{ij} = 1$ is shown in figure 3.2. The function achieves its minimal value at the minimizing atom-configuration which is called ground state. For the Lennard-Jones-$(12, 6)$-Potential it

is known that its ground state corresponds to the distance $r_{ij} = 2^{1/6} \cdot \sigma$, for lower distances the two particles show a repulsive behavior and for larger ones they attract each other. Furthermore, we can conclude that for large values of ϵ we obtain more stable bonds than for lower values, because the depth of the potential provides information about how strongly the particles repulse and attract each other.

The Morse Potential [Doye and Wales, 1997] is given by

$$U(r_{ij}) = \epsilon \cdot e^{-\delta_0 \cdot (r_{ij} - r_0)} \left(e^{\delta_0 \cdot (r_{ij} - r_0)} - 2 \right), \tag{3.9}$$

where ϵ describes the bonds dissociation energy, r_0 is the equilibrium bond length and δ_0 is a constant that controls the attractive and repulsive range of the potential.

The van der Waals Potential is described by the following potential function

$$U_{ij}(r_{ij}) = -a \left(\frac{1}{r_{ij}} \right)^6$$

with a defined as a constant dependent on the interaction between the particles.

Although the presented potentials are only simple pair interactions, it turned out that they work well for the description of the potential energy of noble gases. Nevertheless we need to consider some more complicated potentials in order to represent other substances.

Potentials for Linear Molecules

In this section, we will discuss the potential functions of molecules which consist of particles with already known bond relations. We restrict in the following to molecules with linear internal structure, that is a bond angle of 180° and only intramolecular interactions. Intramolecular interactions are those between the atoms in one single molecule, contrary to the intermolecular ones between atoms of different molecules. They are usually much stronger than the intermolecular interactions. The most frequent used potentials in order to describe these interactions are the bond potential, the angle potential and the torsion potential, see figure 3.3.

The bond potential is defined through a fixed set of atom pairs which are bonded. The resulting function that describes these interaction is commonly a harmonic potential, such as

$$U_b(r_{ij}) = \frac{1}{2} k_b (r_{ij} - r_0)^2,$$

where r_{ij} is the bond length between the i-th and j'th atom, r_0 is the equilibrium distance and k_b is the spring constant.

Next, we include the interactions between three atoms i, j, k out of which two are bonded. We assume that there is a bond between atom i and j and a bond between j and k; atoms i and k are not bonded. We define the angle between those three atoms as θ_{ijk} and one common used angle potential form given by

$$U_a(\theta_{ijk}) = -k_\theta \big(cos(\theta - \theta_0) - 1 \big).$$

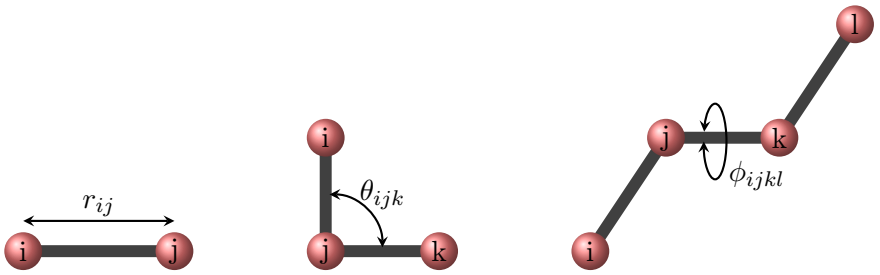


Figure 3.3: Illustration of a bond (left), angle (middle) and torsion potential (right).

If θ is close to the equilibrium value θ_0, we can also use the approximation

$$U_a(\theta_{ijk}) \approx \frac{1}{2}\, k_\theta (\theta - \theta_0)^2.$$

The commonly used torsion potential, between four atoms which are successively bonded, is given by

$$U_t(\phi_{ijkl}) = \sum_{n=1}^{3} k_{\phi_n} \big(\cos(n\phi - \delta) + 1 \big),$$

where ϕ is the angle between the planes spanned by the atoms i, j, k and j, k, l.

In order to obtain the potential function for the whole molecule consisting of N atoms with coordinates $\mathbf{R}_i$, $i = 1, \dots N$, we now sum up the above described potentials plus a Lennard-Jones Potential for all non-bonded interactions and obtain

$$U(\mathbf{R}_1, \dots, \mathbf{R}_N) = \underbrace{\sum_{(i,j)} U_b(r_{ij})}_{\text{bonds}} + \underbrace{\sum_{(i,j,k)} U_a(\theta_{ijk})}_{\text{angles}} + \underbrace{\sum_{(i,j,k,l)} U_a(\phi_{ijkl})}_{\text{torsions}} + \underbrace{\sum_{(i,j)} U^{LJ}(r_{ij})}_{\text{nonbonded interactions}}$$

Applications for this potential functions are alkenes and alkanes such as butane, for a more detailed explanation see [Griebel et al., 2007].

Gupta Clusters for Zinc and Cadmium

In the following, we present a potential function that consists of pair and many-body contributions an can be used for the representation of metals like zinc and cadmium. The Gupta cluster [Doye, 2003] is defined by

$$U(\mathbf{R}_1, \dots, \mathbf{R}_N) = \underbrace{\sum_{i+1}^{N} \sum_{j=i+1}^{N} \phi(r_{ij})}_{\text{pair potential}} + \underbrace{\sum_{i=1}^{N} F(\bar{\rho}_i)}_{\text{many-body contribution}}, \tag{3.10}$$

with $F(\bar{\rho})$ describing a many-body embedding function dependent on the following sum of an "atomic density" function

$$\bar{\rho}_i = \sum_{\substack{j=1 \\ i \neq j}}^{N} \rho(r_{ij}).$$

The remaining functions are defined by

$$\phi(r) = 2Ae^{-p(r/r_0-1)}$$

$$F(\bar{\rho}) = -\xi\sqrt{\bar{\rho}}$$

$$\rho(r) = e^{-2q(r/r_0-1)}.$$

Since these functions are non unique, we define transformations for ϕ and F that applied to (3.10) results in the same energy values. These transformations are defined as follows

$$\phi'(r) = \phi(r) + 2g\rho(r) \text{ and}$$

$$F'(\bar{\rho}) = F(\bar{\rho}) - g\bar{\rho}$$

with

$$g = \left.\frac{dF}{d\bar{\rho}}\right|_{\bar{\rho}=\bar{\rho}_{\text{xtal}}}.$$

The transformation function $F'(\bar{\rho})$ has a minimum at $\bar{\rho} = \bar{\rho}_{\text{xtal}}$ which is a fixed value that represents the value of the atomic density function $\bar{\rho}$ in the equilibrium crystal.

The resulting Gupta potential in its so-called effective pair format is given by

$$\phi_{\text{eff}} = 2Ae^{-p(r/r_0-1)} - \frac{\xi}{\sqrt{\bar{\rho}_{\text{xtal}}}}e^{-2q(r/r_0-1)},$$

$$F_{\text{eff}} = -\xi\sqrt{\bar{\rho}}\left(1 - \frac{1}{2}\sqrt{\frac{\bar{\rho}}{\bar{\rho_{\text{xtal}}}}}\right).$$

All the parameters are dependent on the metal to be represented, the choice for zinc is given in chapter 5.

Periodic Boundary Conditions

In this section, we go into more detail on periodic boundary conditions which are used for the simulation of crystals. We follow the work of [Griebel and Hamaekers, 2005].

For this purpose, we replicate a simulation cell of finite volume with N particles infinitely in each direction of the three-dimensional space and each atom in the simulation cell now interacts not only with all atoms except for itself in this cell, but also with all atoms in all replicated periodic cells. Since this is an infinite number of particles, we use a cut-off radius which defines the neighborhood of each particle in which the interaction results in an energy contribution unequal to zero. All other interactions that are outside this radius have not to be taken into account.

In order to describe the shape of the simulation cell, we define the 3×3 transformation matrix

$$h = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3],$$

which consists of the three linear independent basis vectors that span the cell.

With this, we can express each coordinate vector $\mathbf{R}_i$ of a particle i by the product of the transformation matrix and the scaled coordinate vector of the particle, that is

$$\mathbf{R}_i = h\mathbf{s}_i, \qquad i = 1, \ldots, N.$$

We assume

$$\mathbf{s}_i \in [0,1)^3$$

in order to prevent an overlap between the different cell boundaries. We obtain the vector coordinates of the particles in the periodic cells via translation vectors $\mathbf{S} \in \mathbb{Z}^3$ and define the translated image of particle i by

$$\mathbf{R}_{i\mathbf{S}} := \mathbf{R}_i + h\mathbf{S}.$$

In order to describe the interactions of the particles, we need to adapt the corresponding potential to the periodic boundary conditions. Therefore, we have to use a periodic potential which is explicitly dependent on h in most of the cases and gives the potential energy per unit cell.

At first, we define the periodic potential for pair potentials, such as those presented in section 3.3. Including periodic boundary condition the potential function is given by

$$U_{pair}^{\text{periodic}}(\mathbf{R}_1, \ldots, \mathbf{R}_N, h) = \sum_{i=1}^{N} \left(\sum_{\substack{j=1 \\ i<j}}^{N} U_{ij}(r_{ij}) + \sum_{j=1}^{N} \sum_{\mathbf{S}\in\mathcal{S}} U_{ij}(r_{ij_\mathbf{S}}) \right), \tag{3.11}$$

with $\mathcal{S} := \mathbb{Z}^3 \backslash \binom{0}{0}{0}$ and $r_{ij_\mathbf{S}}$ describing the distance between the i-th particle and the periodic replication of particle j obtained by applying the translation vector $\mathbf{S}$ on $\mathbf{R}_j$, that is

$$r_{ij_\mathbf{S}} = ||\mathbf{R}_{j\mathbf{S}} - \mathbf{R}_i|| = ||\mathbf{R}_j + h\mathbf{S} - \mathbf{R}_i||.$$

To ensure that the sum over the translated images of each particle is finite, we define the cut-off radius $r_c > 0$ such that

$$u_{ij}(r) = 0, \qquad \text{for all } r > r_c.$$

An two-dimensional example for $N = 3$ atoms is illustrated in figure 3.4. The original simulation cell is framed by blue lines, the replicated ones by black lines. For each replicated cell, the transformation matrix is written down in the corresponding cell. All atoms which are surrounded by the red circle interact with atom i since they are inside the radius r_c, this is symbolized by the connecting line between atoms i and j and between atoms i and k.

Lennard-Jones Potential for Periodic Boundary Conditions

Based on the presented definitions for periodic potential in order to derive the periodic Lennard-Jones potential, the cut-off radius ensures that the sum in (3.11) is finite. In order to avoid that the potential becomes discontinuous, we multiply a smoothing function $S(r)$ defined as follows

$$S(r) = \begin{cases} 1, & : \ r \le r_l, \\ 1 - (r - r_l)^2(3r_c - r_l - 2r)/(r_c - r_l)^3, & : \ r_l < r < r_c, \\ 0, & : \ r \ge r_c. \end{cases}$$

Multiplying S with the Lennard-Jones Potential yields the function

$$U_{ij}^{LJ}(r_{ij}) = 4\epsilon_{ij} \cdot S(r_{ij}) \cdot \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right)$$

which decreases continuously to zero between r_l and r_c.

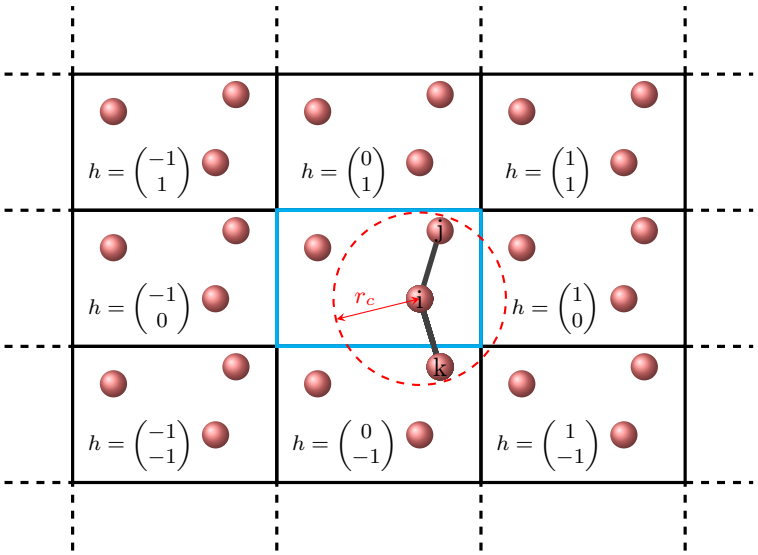


Figure 3.4: Example of Periodic boundaries in two dimensions for $N = 3$ atoms.

Crystal lattice

Crystals have a periodic structure of the corresponding atoms. In order to describe a crystal structure unambiguously, we have to specify the length, shape, and the smallest set of repeating elements, which is called the unit cell.

The so-called *Bravais lattice* [Chatterjee, 2008] describes a classification of different lattice systems of the unit cell based on discrete translation operations. In the three-dimensional space, each lattice point can be identified by

$$\mathbf{B} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

with integers n_i, $i = 1, 2, 3$ and $\mathbf{a}_i \in \mathbb{R}^3$ describe the vectors that span the lattice.

Based on their point group, each lattice is classified with one of the seven crystal systems. The different types of crystal systems, dependent on the parameters of the simulation cell are described in table 3.1. Here $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are the side lengths of the cell and α, β and γ the corresponding angles. Considering the unit cell, there are four different ways of adding lattice points in order to create a new pattern, while the shape of the cell remains unchanged. The corresponding lattices are defined as follows.

- Primitive: The lattice points are on each corner of the cell.
- Body centered: The lattice points are on each corner and at the center of the cell.
- Face centered: The lattice points are on each corner and at the center of each face of the cell.
- Base centered: The lattice points are on each corner and at the center of two faces of the cell.

All resulting 14 *Bravais lattices* are listed in table 3.2, including some examples.

Crystal class or systems	Relation between **a**, **b** and **c**	Relation between α, β and γ
Triclinic	$\mathbf{a} \neq \mathbf{b} \neq \mathbf{c}$	$\alpha \neq \beta \neq \gamma$
Monoclinic	$\mathbf{a} \neq \mathbf{b} \neq \mathbf{c}$	$\alpha = \gamma = 90^\circ \neq \beta$
Orthorhombic	$\mathbf{a} \neq \mathbf{b} \neq \mathbf{c}$	$\alpha = \beta = \gamma = 90^\circ$
Rhombohedral	$\mathbf{a} = \mathbf{b} = \mathbf{c}$	$\alpha = \beta = \gamma \neq 90^\circ$
Tetragonal	$\mathbf{a} = \mathbf{b} \neq \mathbf{c}$	$\alpha = \beta = \gamma = 90^\circ$
Hexagonal	$\mathbf{a} = \mathbf{b} \neq \mathbf{c}$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Cubic	$\mathbf{a} = \mathbf{b} = \mathbf{c}$	$\alpha = \beta = \gamma = 90^\circ$

Table 3.1: Unit cells of the seven crystal systems [Chatterjee, 2008].

The goal of this thesis is the determination of the lowest energy value, i.e. the most stable atom configuration of a potential energy function. In the following chapter, we describe the global optimization of high-dimensional, non-linear and non-convex functions and methods in order to minimize the potential energy of a physical system efficiently. Thereby, we take advantage of the special construction of potential energy functions based on the decomposition in its lower-order contributions.

Crystal system	Bravais lattice	Examples
Triclinic	Primitive	Copper sulfate, sodium bisulfate
Monoclinic	Primitive	Malachite, azunite
	Base centered	Gypsum, mica
Orthorhombic	Primitive	Topaz, aragonite
	Base centered	Chlorine, iodine
	Body centered	Thorium, tetrafluoride
	Face centered	Sulfur, plutonium
Rhombohedral	Primitive	Quartz, tourmaline, antimony, arsenic
Tetragonal	Primitive	Rutile, cassiterite
	Body centered	Tin, indium
Hexagonal	Primitive	Zinc, titanum, magnesium, cadmium
Cubic	Primitive	Polonium, sodium, chlorate
	Body centered	Iron, chromium, sodium, potassium, tungsten, vanadium
	Face centered	Copper, silver, gold, diamond, sodium chloride

Table 3.2: The 14 *Bravais lattices* and examples [Chatterjee, 2008].

4 Global Optimization

The determination of an optimal state is of great relevance in many different areas of current interest. Assuming that the considered problem can be described by a function dependent on a set of different parameters, the goal of global optimization is to find the best possible value for each of them. The issue at hand is that we often have to deal with complex processes and thus, the resulting functions depend on many variables. In chapter 2, we already presented a dimension-wise decomposition that enables the reduction of the complexity of high-dimensional functions.

Nevertheless, global optimization is a very challenging task. In most of the applications, we have to assume non-linear, non-convex and high-dimensional functions and thus, an unknown number of many local minima out of which we have to find the smallest one. Furthermore, due to the high dimension, function and derivative computations yield high computational cost. In order to solve this problem, we are interested in efficient algorithms, that find the smallest function value using only a small number of optimization steps.

In this chapter, we give a short overview of definitions in the context of the global optimization of non-convex, high dimensional functions and the resulting problems. Furthermore, some known worst case error bounds are described, in particular this includes those which are based on dimension-wise decomposed functions. Besides, we present the subspace correction method for linear systems and the extension to nonlinear optimization problems.

Thereafter, we give a more detailed description of the optimization of potential energy functions and give a short overview of methods that already have been successful. The application of the subspace correction as optimization method is presented and a new multilevel algorithm based on a coordinate transformation is developed.

4.1 Problem Definition

Let $f : X^d \to Y$ be a d-dimensional objective function that describes the underlying optimization problem. The aim of global optimization is to find either the best possible variables $\mathbf{x}^* := (x_1^*, x_2^*, \ldots, x_d^*) \in X^d$ or the best possible function value $f(\mathbf{x}^*) \in Y$. In this thesis, we restrict to the latter case which means that the aim is to find a configuration $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_d)$ such that the resulting function value $f(\mathbf{x})$ is as close as possible to the global minimum value $f(\mathbf{x}^*)$. Furthermore, we only consider problems in which the domain X^d of f can be chosen arbitrary, dependent on the specific problem definition, but with the range and codomain Y being subsets of the set of real numbers, that is $Y \subseteq \mathbb{R}$. If we allow an arbitrary range and codomain $Y \subseteq \mathbb{R}^d$, the resulting problem is called multi-objective optimization. It is different topic that will not be further discussed.

It should be noted that we will describe every problem in the context of minimization, even though it is used for maximization. We can reformulate every maximization problem with objective function f in minimizing $-f$.

Without further assumptions, we obtain the following global optimization problem

$$\boxed{\text{Find} \qquad f(\mathbf{x}^*) = \min_{x \in X^d} f(\mathbf{x}).}$$

Dependent on the underlying problem definition, we may have to consider additional constraints. In order to restrict the global optimization problem to a subdomain of X^N, we can make use of bound constraints. Each element then is bounded such that $b_i \leq x_i \leq B_i$ for $i = 1, \ldots, d$, with b_i the lower and B_i the upper bound for x_i. Another possibility is the definition of equality and inequality constraints of the input variables or via auxiliary functions.

There exist many different optimization algorithms which are based on different ideas and heuristics. Basically, we can distinguish between deterministic and stochastic algorithms that include either only function evaluations or additionally, information of derivatives. This information includes the determination of gradients and optionally also Hessians. The choice which one is the best depends on the properties of the underlying objective function. Furthermore, we have to consider how much information is given about f and the computational cost required for its determination.

Similar to the choice of the algorithm, the complexity of the optimization problem also depends on the properties of the objective function. For instance, it can be a very challenging task to find the optimal state of a function which is discontinuous or not differentiable. On top of that, we have to take into account that if the dimension d is large, function evaluations can become very expensive as well as the computation of derivatives.

For convex functions it holds that each local minimum is also the global one, nevertheless its optimization can be a demanding task for high dimensions. For non-convex functions, we often have to assume a very large number of local minima out of which we have to find the global one. The optimization algorithm thus has to distinguish between local and global optima, which is the main difference between global and local optimization methods.

In figure 4.1, some different properties of f in the case of a one-dimensional objective function are illustrated. It is a great challenge to detect the global optimum due to the small area in which the function and derivative values indicate the existence of this global minimum. Another difficulty occurs in areas in which neither the function value nor the gradient of the objective function changes its value significantly. Hence, the optimization algorithm has to choose the search direction almost randomly. Furthermore, there may be multiple local minima which may not all be detected. Especially due to high barriers between these local minima, there is the risk of getting stuck in some of them without finding the smallest one. For a more detailed overview of global optimization problems, see [Weise, 2009].

Another important problem in the context of global optimization is that we cannot guarantee whether we found the exact solution or not. In a finite number of steps, it is impossible to verify that the computed minimum $\tilde{\mathbf{x}} \in X^d$ is the global one, see [Addis et al., 2004]. Thus, we have to be satisfied with a good local minimum which can be computed in a certain amount of time and satisfies

$$f(\tilde{\mathbf{x}}) \leq f(\mathbf{x}^*) + \epsilon,$$

for some $\epsilon > 0$. In many applications such an approximation can already be sufficient.

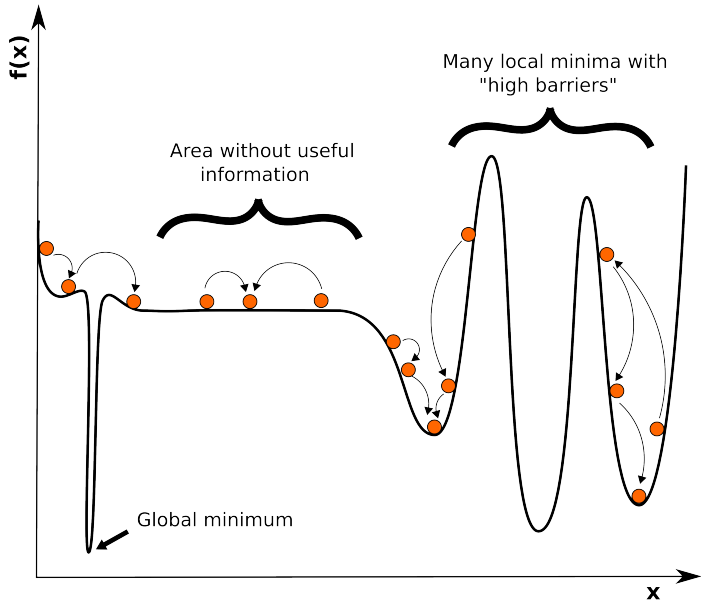


Figure 4.1: Different properties of an objective function with respect to minimization.

In the following, we will explain in more detail some worst-case error estimates for global optimization problems.

4.2 Worst Case Error for Optimization

In this section, we will present some results in the context of worst-case analysis for two problems which are similar to global optimization. We will describe the relation between the worst case error of global optimization and L_∞-approximation, using only function evaluations and give a brief discussion on the resulting problems in this context. It should be noted that global optimization is a non-linear problem, whereas the problem of L_∞-approximation is a linear one.

We have to take into account that there are different error bounds for different function classes and algorithms. We cannot assume a numerical method that gives good results for all possible functions and thus, we have to consider function classes which enables us to give error estimates but still contain functions of our interest. The results are taken from [Novak and Woźniakowski, 2010], [Novak, 1988] and [Novak and Ritter, 1996].

Let $f \in F$ be a d-dimensional function defined on a nonempty set X. We define a functional that describes the problem to be solved with respect to f by

$$S : F \to M$$

with M being a given set. Examples for the functional S are given by the determination of an approximation, an integral or the global minimum of the function f. Since many problems cannot be solved exactly or in an acceptable amount of time, we have to consider approximations to the solution of the problem S.

We denote the set of all admissible approximation methods by A. Thereby we distinguish between the following two subsets of approximation that both use n function values $f(x_i)$, $i = 1, \ldots, n$.

- $A_n^{\text{ad}} \subset A$ contains all approximations to S using adaptively chosen function values.
- $A_n \subset A$ contains all approximations to S using non-adaptively chosen function values.

In this context, we assume that we can evaluate the underlying function exactly, that is without the consideration of any rounding or measurement errors. Nevertheless, we have to take into account that the use of approximation methods causes an error. The aim is to quantify this error in order to make statements about the quality of the computed solution.

Considering the error function $\Delta : M \times F \to \mathbb{R}_0^+$ we define the error of an approximation $\tilde{S} \in A$ to the problem S by

$$\Delta(\tilde{S}(f), f).$$

For example for the case of M being a metric space with the metric d, we can write

$$\Delta(\tilde{S}(f), f) = d(\tilde{S}(f), S(f)).$$

In order to give general bounds we need to consider the worst case error. Assuming a previously fixed method $\tilde{S}$, we thus have to choose the function f for which the error $\Delta(\tilde{S}(f), f)$ has the greatest value and define the worst case error of $\tilde{S}$ by

$$\Delta_{max}(\tilde{S}) := \sup_{f \in F} \Delta(\tilde{S}(f), f).$$

Next, we consider the best of all admissible approximations, instead of a fixed method $\tilde{S}$. By distinguishing between methods using adaptively and non-adaptively chosen function values, we obtain the following two error definitions

$$e_n^{\text{ad}}(F, S) = \inf_{\tilde{S} \in A_n^{\text{ad}}} \Delta_{max}(\tilde{S}) \quad \text{and} \quad e_n(F, S) = \inf_{\tilde{S} \in A_n} \Delta_{max}(\tilde{S}).$$

When turning to the problem of global optimization, we assume that F is the class of all real valued, bounded, d-dimensional functions on the set M. At this point, we repeat the fact that it does not matter if we consider the global maximum or minimum since we can maximize $-f$ in order to obtain the minimal value of f.

The problem definition is now dependent on whether we search for the minimizing elements $\mathbf{x} = (x_1, \ldots, x_d)$ or only for the minimal function value $f(\mathbf{x})$. We will denote the former problem by *Opt* and the latter by *Opt**. The definitions of these two problems are given by finding a solution such that

- *Opt* $:\Delta^{Opt}(\tilde{S}(f), f) := |\tilde{S}(f) - \inf f|$ is sufficiently small, and
- *Opt** $:\Delta^{Opt^*}(\tilde{S}(f), f) := |f(\tilde{S}(f)) - \inf f|$ is sufficiently small.

For these errors, we can give an upper and a lower bound using the error estimates for another problem, namely the L_∞-approximation *App*. The definition of *App* is to find the best approximation $\tilde{S}(f)$ of a function f, i.e. finding a solution such that

- *App*: $\Delta^{App}(\tilde{S}(f), f) := ||\tilde{S}(f) - f||_\infty$ is sufficiently small.

Let $B(X)$ be the Banach space of all bounded real-valued functions on a given set X. With this and the previous definitions, we can formulate the following propositions, taken from [Novak, 1988].

Proposition 4.1. *Let $F \subset B(X)$ be convex and symmetric, and let $n \in \mathbb{N}$. Then*

$$e_{n+1}(F, App) \leq e_n^{ad}(F, Opt) \leq e_n(F, Opt) \leq 2 \cdot e_n(F, App).$$

Proposition 4.2. *Let $F \subset B(X)$ be convex and symmetric. Then*

$$\frac{1}{2} e_n(F, App) \leq e_n^{ad}(F, Opt^*) \leq e_n(F, Opt^*) \leq e_n(F, App).$$

In order to describe the worst case error of the problem *App*, we need the definition of the n-width $d_n(F)$ of $F \subset B(X)$ according to Kolmogorov, that is

$$d_n(F) = \inf_V \sup_{f \in F} \inf_{v \in V} ||f - v||_\infty,$$

where $V \subseteq B(X)$ with $\dim(V) = n$. Thus, we have the following error estimate.

Proposition 4.3. *Let $F \subset B(X)$. Then the following inequality holds*

$$e_n(F, App) \leq (1 + n) \cdot d_n(F).$$

The constant $(1 + n)$ is optimal in the following sense: For each n there is a set F with $e_n(F, App) = (1 + n) \cdot d_n(F)$.

For the proofs and a more detailed description of special cases, we refer to the works of [Novak and Woźniakowski, 2010, Novak, 1988] and the references therein.

As a result we can conclude that there exists a global optimization algorithm for *Opt* and *Opt**, based on n function values, that derives the same error rate as the best approximation algorithm for *App*. Furthermore, there exists an adaptive method which has an error bound that is always as good as the best non-adaptive one.

Next, we consider the special case of the optimization of Hölder continuous functions, defined on

$$[a, b] := \prod_{i=1}^{d} [a_i, b_i] \subset \mathbb{R}^d.$$

The following results are taken from [Novak and Ritter, 1996], the following theorem gives a lower error bound for the optimization for this function class. In the following $D^r f$, $r = (r_1, \ldots, r_d)$ is defined as the partial derivative of order $|r| := \sum_{i=1}^{d} r_i$.

Theorem 4.1. *Let F be a Hölder class of the form*

$$F = \{f : [a, b] \to \mathbb{R} \mid |D^r f(x) - D^r f(y)| \leq ||x - y||^\alpha \text{ for all } r \text{ with } |r| \leq k\},$$

where $0 < \alpha \leq 1$ and $k \in \mathbb{N}_0$. Then there is a constant $c > 0$ such that

$$e_n(F, Opt) > c \cdot n^{-\frac{(k+\alpha)}{d}} \tag{4.1}$$

for any $n \in \mathbb{N}$. This lower bound is sharp and can be achieved with a passive algorithm using function values from a regular grid.

It is easy to see that the error (4.1) becomes great for fixed k and α and a large dimension d. The bound is based on an algorithm which only uses non-adaptively chosen function evaluations, but even by including adaptive methods, gradient information (of order less or equal k) or by using stochastic methods we cannot achieve better results. Thus, the global optimization of Hölder functions turns out to be a difficult task.

Next, we present two suggestions in order to improve the error bound (4.1). One approach in this context is the use of an average instead of the worst case error. For this purpose, we have to derive an error estimate that holds for many $f \in F$, instead of all. Since in practice, adaptive or stochastic algorithms often give much better results in global optimization, the hope is to be able to show this with the help of average case analysis. The issue at hand is that such studies are very difficult and this topic is stated as an open problem for arbitrary dimension in [Novak, 1988]. We refer to [Novak and Ritter, 1996] and the references therein for some results in the one dimensional case.

Another possibility in order to derive smaller error estimates is the restriction of the function class F. With this, we can make use of the properties of the contained functions, as for example functions that have decompositions as described in chapter 2. We recall the decomposition from theorem 2.1, given by

$$f = \sum_{\mathbf{u} \subseteq \{1,\dots,d\}} f_{\mathbf{u}}.$$

The challenge now would be to derive connections between the function spaces of the $f_{\mathbf{u}}$ in and the function space of f.

If we assume a given decomposition of the function $f : \mathbb{R}^d \to \mathbb{R}$ and additionally, that f has order $q \ll d$ (as defined in section 2.1.2), that is

$$f = \sum_{\substack{\mathbf{u} \subseteq \{1,\dots,d\} \\ |\mathbf{u}| \leq q}} f_{\mathbf{u}}.$$

For example, we can assume that f has order $q = 2$ and thus can be rewritten by

$$f(x) = \sum_{i,j=1}^{d} f_{i,j}(x_i, x_j).$$

If we furthermore claim that $f_{i,j} \in C^2([a,b])$ for all pairs (i,j), it follows that

$$f \in C^2([a,b])$$

and we can conclude that $D^r f$ exists for all r, such that $r_i \leq 1$ for every i.

The motivation of this approach is the fact that we can give an error bound for the space $W_\infty^{\mathbf{1}}([a,b])$, defined as follows

$$W_\infty^{\mathbf{1}}([a,b]) = \{f : [a,b] \to \mathbb{R} \big| \ ||D^r f||_\infty \leq 1 \text{ for all } r \text{ with } r_i \leq 1\}.$$

There exist methods $\tilde{S}$ for $F = W_\infty^{\mathbf{1}}([a,b])$ with the error

$$\Delta_{max}(\tilde{S}) = \sup_{f \in F} \Delta(\tilde{S}(f), f) \leq c \cdot n^{-1} (\log n)^{2(d-1)},$$

see [Novak and Ritter, 1996] and the references therein.

In this context, it is an interesting question whether we can also generalize this to functions with decompositions in which each term $f_{\mathbf{u}}$ is s times continuously differentiable, that is $f_{\mathbf{u}} \in C^s$ for all $\mathbf{u}$, $f \in C^s$ and thus $D^r f$, with $r_i \leq \lfloor s/q \rfloor$, exist.

In the following, we present a global optimization method that may take advantage of dimension-wise decomposed function, the method of subspace correction.

4.3 Subspace Correction Methods

Subspace correction methods are based on the divide and conquer principle and allows the definition of a unified framework of the ideas of well-known iterative methods such as Jacobi, Gauss-Seidel, domain decomposition or multigrid methods.

The main idea is a partition of the original problem into lower-dimensional subproblems, based on space decomposition. These subproblems are solved iteratively, while ensuring that the intermediate result in each iteration step does not get worse. The key aspect is to perform a correction of the current solution candidate in each subspace in order to get an approximated solution for the original problem. Thus, we have the possibility to reduce the dimension of the original problem significantly which is a main advantage of these types of methods.

First, we will give an overview on subspace correction methods for finding an approximated solution of symmetric positive definite problems for linear systems. With this, we will motivate the extension on non-linear optimization problems. We will mainly follow the work of [Xu, 1992].

4.3.1 Linear Systems and Space Decomposition

Subspace correction methods are used to find an approximation of the solution $\mathbf{u}$ of a linear equation

$$\mathbf{A}\mathbf{u} = \mathbf{f}, \tag{4.2}$$

with $\mathbf{A}$ describing a symmetric positive definite (SPD) linear operator on a finite-dimensional linear vector space $\mathcal{V}$ and $\mathbf{f}, \mathbf{u} \in \mathcal{V}$ with $\mathbf{f}$ given.

The basic idea in order to find a solution of (4.2) is the following iterative method. In each iteration step, the latest approximated solution $\mathbf{u}^{\text{old}}$ of the problem is updated to a better or equal approximation $\mathbf{u}^{\text{new}}$ using the computation of a residuum. This procedure usually contains the following three steps.

1. $\mathbf{r}^{old} = \mathbf{f} - \mathbf{A}\mathbf{u}^{old}$;
2. $\hat{\mathbf{e}} = \mathbf{B}\mathbf{r}^{old}$ with $\mathbf{B} \approx \mathbf{A}^{-1}$ (approximation of: solve $\mathbf{A}\mathbf{e} = \mathbf{r}^{old}$);
3. $\mathbf{u}^{new} = \mathbf{u}^{old} + \hat{\mathbf{e}}$ (update).

Based on these iteration steps, it is of great importance how we choose the approximate inverse of $\mathbf{A}$, because with $\mathbf{B}$ we can compute the key aspect of subspace correction methods,

the correction $\hat{\mathbf{e}}$. Next, we will specify the subspaces. The idea is to turn the original high-dimensional problem space into $J \geq 1$ smaller subspaces $\mathcal{V}_i \subset \mathcal{V}$ of lower dimension and solve an appropriate problem iteratively on these subspaces.

We consider a partition of the d-dimensional space $\mathcal{V}$ into J subspaces $\mathcal{V}_i \subset \mathcal{V}$, $i = 1, \ldots, J$, of lower dimension such that

$$\mathcal{V} = \sum_{i=1}^{J} \mathcal{V}_i. \tag{4.3}$$

Consequently, for each $\mathbf{v} \in \mathcal{V}$, there exist $\mathbf{v}_i \in \mathcal{V}_i$, $i = 1, \ldots, J$, with $v = \sum_{i=1}^{J} \mathbf{v}_i$.

Let $(\cdot, \cdot)_{\mathbf{A}}$ with $\mathbf{A}$ SPD be an inner product on $\mathcal{V}$ given by

$$(\mathbf{u}, \mathbf{v})_{\mathbf{A}} := (\mathbf{A}\mathbf{u}, \mathbf{v}) \quad \text{for } \mathbf{u}, \mathbf{v} \in \mathcal{V}.$$

Defining the restriction of $\mathbf{A}$ on $\mathcal{V}_i$ by the operator $\mathbf{A}_i : \mathcal{V}_i \to \mathcal{V}_i$ which is also SPD by

$$(\mathbf{A}_i \mathbf{u}_i, \mathbf{v}_i) = (\mathbf{A}\mathbf{u}_i, \mathbf{v}_i), \qquad \mathbf{u}_i, \mathbf{v}_i \in \mathcal{V}_i,$$

we can compute all steps in the subspace correction iteration for each subspace $\mathcal{V}_i$. Furthermore, we introduce the orthogonal projections $\mathbf{Q}_i, \mathbf{P}_i : \mathcal{V} \to \mathcal{V}_i$, defined by

$$\begin{aligned} (\mathbf{Q}_i \mathbf{u}, \mathbf{v}_i) &= (\mathbf{u}, \mathbf{v}_i) \ , \quad \text{and} \\ (\mathbf{P}_i \mathbf{u}, \mathbf{v}_i)_{\mathbf{A}} &= (\mathbf{u}, \mathbf{v}_i)_{\mathbf{A}} \quad \text{for} \quad \mathbf{u} \in \mathcal{V}, \mathbf{v}_i \in \mathcal{V}_i. \end{aligned}$$

According to these definitions we can conclude that the identity

$$\mathbf{A}_i \mathbf{P}_i = \mathbf{Q}_i \mathbf{A} \tag{4.4}$$

holds.

From (4.4) we can conclude that for u being a solution of (4.2), its restriction on each subspace $\mathcal{V}_i$ is given by

$$\mathbf{A}_i \mathbf{u}_i = \mathbf{f}_i \tag{4.5}$$

with $\mathbf{u}_i = \mathbf{P}_i \mathbf{u}$ and $\mathbf{f}_i = \mathbf{Q}_i \mathbf{f}$.

Since our aim is a correction of the current solution in each iteration step, we consider an approximated solution for each subspace problem and define the SPD operator $\mathbf{R}_i : \mathcal{V}_i \to \mathcal{V}_i$ with $\mathbf{R}_i \approx \mathbf{A}_i^{-1}$ and the computed solution can be expressed by $\mathbf{u}_i = \mathbf{R}_i \mathbf{f}_i$.

In the next section, we present an additive and multiplicative approach for the subspace correction method.

4.3.2 Parallel and Successive Method

In the following, we present the additive approach, called parallel subspace correction (PSC). Using this type of method, each problem (4.5) is solved separately on the corresponding subspace $\mathcal{V}_i$, similar to the well-known Jacobi method. We use the three subspace correction iteration steps, restricted on the $\mathcal{V}_i$, and define $\mathbf{e}$ such that the exact solution of (4.2) is given by $\mathbf{u} = \mathbf{u}^{old} + \mathbf{e}$.

Consequently, we have to solve

$$\mathbf{A}_i \mathbf{e}_i = \mathbf{Q}_i \mathbf{r}^{old}$$

on each subspace $\mathcal{V}_i$. The steps 2. and 3. of the subspace correction iteration become

$$\hat{\mathbf{e}}_i = \mathbf{R}_i \mathbf{Q}_i \mathbf{r}^{old} \qquad \text{and}$$

$$\mathbf{u}^{new} = \mathbf{u}^{old} + \sum_{i=1}^{J} \hat{\mathbf{e}}_i.$$

Equivalently, with $\mathbf{B} := \sum_{i=1}^{J} \mathbf{R}_i \mathbf{Q}_i$ we can write

$$\mathbf{u}^{new} = \mathbf{u}^{old} + \mathbf{B}(\mathbf{f} - \mathbf{A}\mathbf{u}^{old}). \tag{4.6}$$

Since this additive approach solves the lower dimensional subproblems separately we can implement and run the algorithm in parallel. Thus, we can reduce the computing time for each iteration step.

In the multiplicative approach, called successive subspace correction, we use the best solution approximation in each iteration step. This means that we correct successively in each particular subspace and use therefore the information of each previous step, in principle it is similar to the Gauss-Seidel method. In each iteration we obtain for a certain subspace $\mathcal{V}_i$ and $\mathbf{u}^{old}$ denoting the starting or the most updated approximated solution

$$\mathbf{u}^{new} = \mathbf{u}^{old} + \mathbf{R}_i \mathbf{Q}_i (\mathbf{f} - \mathbf{A}\mathbf{u}^{old}). \tag{4.7}$$

4.3.3 Subspace Correction for Optimization

The use of subspace correction methods for optimization problems is motivated by the fact that the problem (4.2) is equivalent to the minimization of the $\mathbf{A}$-norm error of the difference between the approximated solution $\mathbf{u}$ and the exact solution $\mathbf{u}^* = \mathbf{A}^{-1}\mathbf{f}$, see [Xu, 2001]. The $\mathbf{A}$-norm is given by

$$||\mathbf{u}||_{\mathbf{A}} = \sqrt{(\mathbf{v}, \mathbf{v})_{\mathbf{A}}}.$$

Note, that this defines a norm since $\mathbf{A}$ is SPD. We consider

$$||\mathbf{u} - \mathbf{u}^*||_{\mathbf{A}}^2 \overset{\mathbf{A}\ \text{SPD}}{=} \mathbf{u}^T \mathbf{A} \mathbf{u} - 2\mathbf{u}^T \mathbf{f} + \mathbf{f}^T \mathbf{A}^{-1} \mathbf{f}$$

which is equivalent to minimizing the quadratic functional

$$F(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T \mathbf{A} \mathbf{u} - \mathbf{u}^T \mathbf{f} = \frac{1}{2}(\mathbf{A}\mathbf{u}, \mathbf{u}) - (\mathbf{f}, \mathbf{u}).$$

According to [Tai and Xu, 2002] the additive and multiplicative subspace correction algorithms are defined by assuming the update steps (4.6) and (4.7) to be understood as an optimization problem, respectively. The resulting optimization algorithms are described in algorithm 1 and algorithm 2.

There are many strategies for the choice of the relaxation parameters α_i, $i = 1, \ldots, J$, in the additive subspace correction method for optimization. For example, we can choose the α_i,

Algorithm 1: Parallel Subspace Correction for Optimization

Given $\mathbf{u}^0 \in \mathcal{V}$
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{u}^k$
  for $i = 1, 2, \ldots, J$ **do**
    $\hat{\mathbf{e}}_i = \underset{\mathbf{e} \in \mathcal{V}_i}{\operatorname{argmin}} \; F(\mathbf{u}^k + \mathbf{e})$
  end
  $\mathbf{v} := \mathbf{v} + \sum_{i=1}^{J} \alpha_i \hat{\mathbf{e}}_i$
  $\mathbf{u}^{k+1} = \mathbf{v}$
end

Algorithm 2: Successive Subspace Correction for Optimization

Given $\mathbf{u}^0 \in \mathcal{V}$
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{u}^k$
  for $i = 1, 2, \ldots, J$ **do**
    $\hat{\mathbf{e}} = \underset{\mathbf{e} \in \mathcal{V}_i}{\operatorname{argmin}} \; F(\mathbf{u}^k + \mathbf{e})$
    $\mathbf{v} := \mathbf{v} + \hat{\mathbf{e}}$
  end
  $\mathbf{u}^{k+1} = \mathbf{v}$
end

$i = 1, \ldots, J$, as proposed in [Frommer and Renaut, 1999], the method is called the selection strategy.

In theis strategy we choose only this correction that results the smallest value of $\tilde{J}$, i.e. we have to find the index t such that

$$F(\mathbf{v} + \alpha_t \hat{\mathbf{e}}_t) = \min \{F(\mathbf{v} + \alpha_1 \hat{\mathbf{e}}_1), \ldots, F(\mathbf{v} + \alpha_J \hat{\mathbf{e}}_J)\}. \tag{4.8}$$

The correction is then defined by setting $\alpha_t = 1$ and $\alpha_i = 0$ for all $i \neq t$.

There exist many other possible methods of choosing the relaxation parameters α_i for the parallel subspace correction optimization.

For instance, we can modify the selection strategy (4.8) by choosing the index t not out of optimizations in all subspaces but out of a random subset of them. This way we can reduce the number of optimizations in each iteration step what can become crucial if we have a large number of subspaces. Another method is the optimal strategy proposed by [Frommer and Renaut, 1999]. It is defined by choosing the best possible relaxation parameters α_i, which means that we have to solve the following J-dimensional optimization problem

$$(\alpha_1, \ldots, \alpha_J) = \underset{(\alpha_1, \ldots, \alpha_J) \in \mathbb{R}^J}{\operatorname{argmin}} F(\mathbf{v} + \sum_{i=1}^{J} \alpha_i \hat{\mathbf{e}}_i).$$

It should be noted that we have to perform an additional optimization in each iteration step by using this strategy. The dimension of this optimization problem is J, which means that it depends on the number of subspaces and we need to consider wether this strategy is useful or not.

We concentrate on the different subspace correction methods and their applicability more than on the different relaxation strategies, on this account we will restrict to the selection strategy and refer to [Frommer and Renaut, 1999, Tai and Xu, 2002, Sharapov, 1997] and the references therein for possible alternatives and the convergence analysis under certain conditions for the objective function.

Although the parallel and successive subspace correction for optimization are motivated by the linear problem (4.2), we can transfer them to non-linear functionals F.

The results and convergence analysis for F being a non-linear convex functional are presented in [Tai and Xu, 2002].

In [Richtárik and Takáč, 2012] a randomized block-coordinate descent method for the minimization of the sum of a smooth convex and a simple non-smooth block-separable convex function and its complexity results is described.

Propositions regarding the convergence of subspace correction methods for optimization turn out to be much more difficult for the non-convex case, an example is given by an eigenvalue problem. This problem has been discussed by [Chan and Sharapov, 2002], who presented a multilevel method and its qualitative convergence for solving partial differential equation based eigenvalue problems. The method is a modified multiplicative subspace correction in order to minimize the Rayleigh quotient. Nevertheless, the conclusion of a uniformly optimal convergence for the non-linear, non-convex case is still not possible .

In this thesis, we concentrate on the minimization of potential energy functions for which we cannot make any convexity assumptions. Thus, we will not further discuss the specific

convergence analysis of subspace correction methods for this special case. In the following, we describe the global optimization of potential energy functions.

4.4 Global Optimization of Potential Energy Functions

We already described the relationship between potential energy functions and the ground state in chapter 3.

In the following, we will go into more detail in the global optimization of potential energy functions. Since they are high-dimensional and complex structured, we have to assume a very large number of local minima and expensive function and derivative evaluations. Thus, we are interested in efficient methods for this non-linear, non-convex global optimization problem.

The potential energy of a physical system is dependent on the spatial coordinates of each corresponding atom. Considering a system of N atoms with coordinates $\mathbf{x}_i \in \mathbb{R}^3$, $i = 1, \ldots, N$, its potential energy function is described by $U : \mathbb{R}^{3N} \to \mathbb{R}$. This function is based on the complex interactions of all corresponding particles. In chapter 3, we already described some simplifications in order to get an analytic potential energy function of the form

$$U(\mathbf{x}_1, \ldots, \mathbf{x}_N) \approx \sum_{\substack{\mathbf{u} \subseteq \{1,\ldots,N\} \\ |\mathbf{u}| \leq q}} V_{\mathbf{u}}(\mathbf{x}_1, \ldots, \mathbf{x}_{|\mathbf{u}|}) \tag{4.9}$$

with $q \ll N$. The different $V_{\mathbf{u}}$ represent the original function by its lower-order contributions, which can be interpreted as the interaction of the different particles with each other. Examples for this lower-order contributions in the context of the potential energy are given by single particle contributions ($|\mathbf{u}| = 1$) or pair- and triple-interactions ($|\mathbf{u}| = 2$ and $|\mathbf{u}| = 3$). Nevertheless, the potential energy function (4.9) remains high-dimensional and complex structured.

We use the fact that the potential energy value of a physical system does not change by translating the coordinates of all atoms, as long as all distances and angles remain the same. Thus, we can restrict to a bound-constrained minimization problem which means that we only allow configurations in the subset

$$M = [a, b]^{3N} \subset \mathbb{R}^{3N},$$

with $a, b \in \mathbb{R}$ being the lower and upper bound for the domain of U, respectively. Furthermore, we assume that U is sufficiently smooth. The resulting problem definition is given by

$$\boxed{\begin{array}{ll} \text{Find} & U(\mathbf{x}^*) = \min\limits_{x \in \mathbb{R}^{3N}} U(\mathbf{x}), \\ \text{s.t.} & x \in M. \end{array}} \tag{4.10}$$

The issue at hand is that we have to assume a large number of local minima, due to the complex structure of U. More precisely, the number of local minima of potential functions often increases exponentially with the number of particles. This applies even for potential with only simple pair interactions. For the Lennard-Jones cluster [Hoare, 1979] made the assumption that for $N \leq 4$ atoms there is one local minimum and that this number increases in $\mathcal{O}(1.03^{N^2})$. He established this by interpolating his computed results up to nine atoms.

Thus, it is not a bound for the number of local minima, but gives an impression of how the growth behaves dependent on the number of particles. According to this assumption, for $N = 10$ atoms there are only 19 local minima, but for $N = 15$ atoms, we already have to assume at least 773 local minima out of which we have to find the smallest one.

The optimization of clusters has been studied by many authors by the use of different optimization approaches. Examples for the optimization of only pair-potentials are given by the Lennard-Jones Cluster, which was studied in [Wales and Doye, 1997] and the Morse Cluster, described in [Doye and Wales, 1997]. A short overview of frequently used global optimization algorithms in the context of the minimization of potential energy functions is given in the next section.

4.4.1 Overview

In this section, we present an overview of some algorithms that turned out to be very successful for the global minimization of potential energy functions.

Algorithms which are motivated by the processes in nature are called evolutionary or genetic algorithms. They combine mechanisms like mutation, crossover, natural selection and survival of the fittest with the context of optimization. In each optimization step the possible minimizing variables are given in a set of elements that represent the individuals in a population or the chromosomes in a gene pool. One main advantage using these meta-heuristic types of algorithms is that we do not have to make many assumptions on f. For the application on potential energy functions see for example [Deaven and Ho, 1995, Moloi and Ali, 2005].

Another popular type of methods in order to minimize energy functions are so-called Graduated-Non-Convexity (GNC) methods which were originally developed for the area of visual reconstruction, see for example [Blake and Zisserman, 1987]. They are based on the fact that for each convex function f, we have to find any local minimum because it is also the global one. The idea is to create a series of functions which represent a stepwise convex relaxation on the original function f. This convex relaxation is realized by the use of appropriate transformations of the original function and usually continues until we have a function with a significant smaller number of local minima. The optimization procedure now works as follows. In each step, we minimize a transformation of the original function, beginning with the function which has the fewest number of local minima. The starting point of each minimization is the minimizer of the previous step. The algorithm is successful if it converges after a certain number of iteration steps towards the global minimum of the original function f. The main problem here is the choice of the convexifying function series. If we do not know much about the structure properties of f, the use of specific, function dependent transformations turns out to be a very difficult task. But also the use of general transformations can be problematical, because there is the possibility of loosing important properties of the original objective function. In the following, we propose different approaches for function transformations that can be applied for the minimization of potential energy functions.

In the Diffusion Equation Method (DEM), we consider the objective function as the solution of the diffusion equation. Its application in the context of the optimization of Lennard-Jones clusters is presented in [Kostrowicki et al., 1991]. The convex relaxation is obtained by solving a non-linear, isotropic diffusion equation. Since its solution can be rewritten into the convolution with a Gaussian function with a width dependent on the time t, the convex

relaxation can be understood as a spatial averaging over an increasing domain with progressing time.

The method of Simulated Annealing tries to imitate the annealing process of liquid metals and has its origin in the field of combinatorial optimization, see [Kirkpatrick et al., 1983]. For the context of the minimization of potential energy functions, we refer to [Xue, 1994]. The method of Simulated Annealing is motivated by the fact that for high temperatures the particles can move freely and are able to break down energy barriers. If the temperature is decreased they loose this property and the metal becomes solid. By decreasing the temperature slowly, the particles have enough time to reorder in a stable configuration which is called a crystal. But if the annealing is done too quickly, the resulting end-configuration only represents a metastable state, for example a polycrystalline. The main idea of the method is the decision wether to stay in the current best minimum or to update the input by random values in an appropriate neighborhood. The Simulated Annealing algorithm allows intermediate results with a higher function value up to a certain value, which are usually determined via the Metropolis criterion. By using this type of method, we thus have the possibility to overcome energy barriers in order to find the minimum with the smallest energy value out of all local ones.

Another method, which has turned out to be very successful in the context of the minimization of potential energy functions, is the basin hopping algorithm based on Monte Carlo sampling. An example for its usage for the optimization of Lennard-Jones clusters is given by [Wales and Doye, 1997]. The transformation assigns each set of variables an energy value, which is obtained by applying a minimization with this configuration as a starting point. Thus, the transformed potential energy function has different basins of attraction. Each of them represents all configurations that lead to the same local minimum. For the function exploration, a Monte Carlo simulation with a fixed number of randomly generated configurations at a fixed temperature is used. In each step, the current configuration is modified by random numbers in a certain range and the decision whether to except this new optimal atom coordinates or not is made via the Metropolis criterion. The main task is to find the best constant temperature and a sufficient number of steps which often have to be determined by different test runs.

4.4.2 Application of Subspace Correction

The basic idea of subspace correction methods is the partition of the original problem into lower-dimensional subproblems, based on space decomposition. The general method for nonlinear optimization problems was already described in section 4.3.

Next, we apply the subspace correction on the optimization of potential energy functions. We can exploit their special construction, based on the dimension-wise decomposition in its lower-order contributions, in order to define the subspaces. Furthermore, the main advantage is that we can reduce the dimension of the optimization problem significantly, since we have to solve only a lower-dimensional subproblem in each iteration step.

We choose the subspaces such that the different contributions of the dimension-wise decomposition are considered separately in each iteration. This means that a specific atom subset is optimized in each iteration step. The subspaces $\mathcal{V}_m$, $m = 1 \ldots, J$, are defined by

$$\mathcal{V}_m = \left\{ \mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in M \;\middle|\; x_{(k)} = 0 \text{ if } k \notin I_m \right\} \tag{4.11}$$

with I_m describing an index set that contains all indices of the coordinates of those atoms, that correspond to the m-th subspace.

Furthermore, we modify the potential energy function such that it just depends on the coordinates of this atom subset, while all other coordinates are fixed. Therefore, it suffices to take into account only those interactions, which involve the considered subset of atoms.

By using this choice of subspaces, we ensure that each solution of the respective subproblems is a feasible solution for the original problem, i. e. the minimization of a potential energy function.

One-Atom Optimization

For the case of optimizing only one atom in each iteration step, the index set I_m in (4.11) contains only the index of the m-th atom. That is

$$\mathcal{V}_m^{\text{1-atom}} = \Big\{ \mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in M \;\Big|\; \mathbf{x}_i = 0 \;\text{ if }\; i \neq m \Big\},$$

for $m = 1, \ldots, N$. Furthermore, we define the restricted function $U_m : \mathbb{R}^3 \to \mathbb{R}$, which is only dependent on the coordinates of the m-th atom $\mathbf{x}_m \in \mathbb{R}^3$.

Consequently, we obtain N different subspaces and we have to solve N three-dimensional optimization problems. With this, the idea is to optimize U_m for each atom position $\mathbf{x}_m$, $i = 1, \ldots, N$, and repeat this until convergence or until a given maximal number of iterations is reached via the parallel or successive subspace correction for optimization. The resulting algorithm for the parallel method including the selection strategy is described in algorithm 3 and the one for the successive method in algorithm 4.

Algorithm 3: Parallel Subspace Correction for One-Atom Optimization

Given $\mathbf{x}^0 \in M$
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{x}^k$
  for $m = 1, 2, \ldots, N$ **do**

$$\hat{\mathbf{e}}_i = \operatorname*{argmin}_{\mathbf{e} \in \mathcal{V}_m^{\text{1-atom}}} U_m(\mathbf{x}^k + \mathbf{e})$$

  end
  Find t such that

$$U_m(\mathbf{v} + \alpha_t \hat{\mathbf{e}}_t) = \min\left(U_m(\mathbf{v} + \alpha_1 \hat{\mathbf{e}}_1), \ldots, U_m(\mathbf{v} + \alpha_N \hat{\mathbf{e}}_N)\right)$$

  and set

$$\alpha_t = 1 \text{ and } \alpha_m = 0 \text{ for all } m \neq t \quad \% \text{ Selection Strategy}$$

  $\mathbf{v} := \mathbf{v} + \sum_{m=1}^{N} \alpha_m \hat{\mathbf{e}}_m$
  $\mathbf{u}^{k+1} = \mathbf{v}$
end

In figure 4.2 and 4.3, respectively, the basic ideas of the parallel and successive method for the one-atom optimization are shown for an example of $N = 4$ atoms. Figure 4.2 shows two

Algorithm 4: Successive Subspace Correction for One-Atom Optimization

Given $\mathbf{x}^0 \in M$
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{x}^k$
  for $m = 1, 2, \ldots, N$ **do**
    $\hat{\mathbf{e}} = \underset{\mathbf{e} \in \mathcal{V}_m^{\text{1-atom}}}{\operatorname{argmin}} \; U_m(\mathbf{x}^k + \mathbf{e})$
    $\mathbf{v} := \mathbf{v} + \hat{\mathbf{e}}$
  end
  $\mathbf{x}^{k+1} = \mathbf{v}$
end

iteration steps of the parallel subspace correction algorithm. All atom positions are optimized separately by using only the information of the iteration step before. After the optimization of all atom coordinates, we update these positions and repeat this in the next iteration step until some stopping criterion is reached. In figure 4.3, there are the particular steps of one iteration of the successive subspace correction algorithm. The coordinates of each atom are optimized successively, including the most updated positions of all the other atoms.

Two- and Three-Atom Optimization

The one-atom optimization can be extended to a two- or three-atom optimization. We modify the previously described procedures slightly by considering two or three atoms at once instead of one. It should be noted that in these cases, we have more different subspaces than in the one-atom optimization. The number of subspaces for the two-atom optimization is

$$J_{\text{pair}} = \frac{N(N-1)}{2}$$

and the number for the three-atom optimization is given by

$$J_{\text{triple}} = \frac{N(N-1)(N-2)}{6}.$$

Therefore, the index set I_m in (4.11) contains the indices of the coordinates of the considered atom pair (i, j) and the atom triple (i, j, k), respectively. That is

$$\mathcal{V}_m^{\text{2-atom}} = \Big\{ \mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in M \;\Big|\; \mathbf{x}_i = 0 \text{ and } \mathbf{x}_j = 0 \;\text{ if }\; (i,j) \neq m \Big\}$$

for $m = 1, \ldots, J_{\text{pair}}$ and

$$\mathcal{V}_m^{\text{3-atom}} = \Big\{ \mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in M \;\Big|\; \mathbf{x}_i = 0 \text{ and } \mathbf{x}_j = 0 \text{ and } \mathbf{x}_k = 0 \;\text{ if }\; (i,j,k) \neq m \Big\}$$

for $m = 1, \ldots, J_{\text{triple}}$.

The restricted potential energy functions are given by $U_{i,k} : \mathbb{R}^6 \to \mathbb{R}$ for the two-atom optimization and by $U_{i,k,l} : \mathbb{R}^9 \to \mathbb{R}$ for the three-atom optimization. The resulting optimization

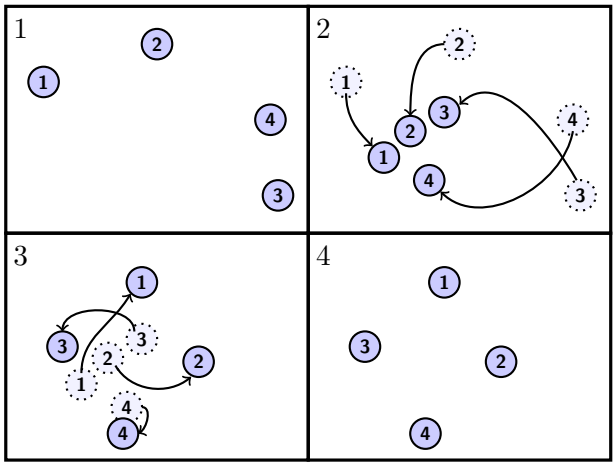


Figure 4.2: Example of the parallel subspace correction optimization steps for $N = 4$ atoms.

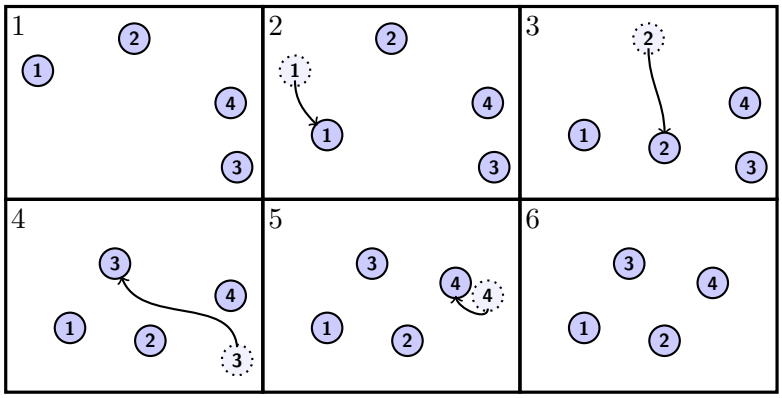


Figure 4.3: Example of the successive subspace correction optimization steps for $N = 4$ atoms.

problem is then six and nine-dimensional and we have to go through all pairs and all triples, respectively.

It may be of relevance in which order we run through these subspaces. In this thesis, we consider two different types of orders. The first is a random choice of pairs or triples and the second is based on the idea that all of the atoms are optimized once before we consider one of them a second time. In the following, we will call this the *improved order of atoms*. An example for $N = 4$ different atoms is given in the following table.

m	(i, j)
1	$(1, 2)$
2	$(3, 4)$
3	$(2, 3)$
4	$(1, 4)$
5	$(1, 3)$
6	$(2, 4)$

The same idea can be transferred to the three-atom optimization and the corresponding order of the subspaces $\mathcal{V}_m$, $m = 1 \ldots, J_{\text{triple}}$. Especially for the parallel subspace correction method this is essential, since the subspaces have to be chosen non-overlapping in each iteration.

Based on these three optimization methods, it is possible to combine them in order to obtain a multilevel method. The different three levels are defined by the one-, two- and three-atom optimization. However, it has to be determined how many times we repeat one level and we have to decide from which level we start.

The choice of these different optimization procedures gives us the possibility to change the positions of one, two or three atoms at once in each subspace-iteration. However, there is no possibility of changing all positions at once in one iteration step. Considering the case that the optimal configuration would be reached by multiplying all coordinates of the starting configuration with an appropriate scaling factor, the subspace correction method is unable to use such information since the optimization only affects subsets of atom coordinates. In order to improve the convergence results for such and similar cases, it would be advantageous to change the coordinates of all atoms in one single step.

In the next section we will describe a method which enables the realization of this idea.

4.4.3 Multilevel Coordinate Transformation

In this section, we describe a coordinate transformation which enables us the changing of the positions of all atom coordinates by changing three values of this transformation function. Assuming that there exists a function $T : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$, we obtain the optimal configuration $\mathbf{x}^*$ by applying t on the initial configuration $\mathbf{x}^{(0)}$, i.e.

$$T\big(\mathbf{x}^{(0)}\big) = \mathbf{x}^*.$$

Now, instead of optimizing the coordinates directly, we search for the optimal transformation function T. For the general case, this is impossible since we do not know any properties

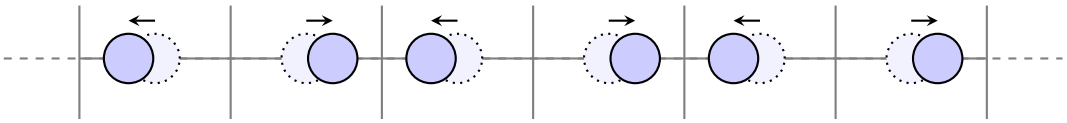


Figure 4.4: Lattice vibrations for a one-dimensional lattice model with periodic boundaries.

of this transformation function. Regarding to the method of subspace correction, we therefore are satisfied with a coordinate transformation that can compute a better solution candidate in the iteration steps that the direct optimization of the atom coordinates. If this is possible, we can improve the efficiency of the resulting optimization algorithm.

Preliminary to the definition of the coordinate transformation, we give a brief motivation for its choice. The following example is taken from [Srivastava, 2011]. Considering a potential energy function for a crystal structure, we have to take into account periodic boundary conditions, i.e. a periodic structure of the corresponding atoms. The configurations in the equilibrium (ground state) can be described by the *Bravais lattices* according to their symmetries. This was already described in section 3.3.

If the position of one atom changes, this affects all other atoms which are close enough such that they interact. After that, each atom tries to move again to the position where the equilibrium of the forces acting on them is reached which again causes motions for all atoms. Hence, there exist certain dynamic which is called lattice vibration.

A simplified example is given in figure 4.4 which shows a one-dimensional lattice with periodic boundaries of $N = 6$ identical atoms. We assume that the optimal configuration is reached if each atom is at center of its cell. Furthermore, each atom only interacts with its direct neighbors and the force can be described as they are used for elastic springs.

By applying Newton's second law of motion the equation of motion of the h-th atom is given by

$$m\ddot{x}_h = C(x_{h+1} - x_h) + C(x_{h-1} - x_h)$$

with M describing the mass of the atom and C a spring constant. Now, considering a one-dimensional lattice of N atoms, this leads to a system of N coupled differential equations and the solutions due to the periodic grid structure, we assume oscillations in the solutions x_h. The so-called normal mode solution is given by

$$x_h = x_0 \cdot \exp(i(kh - \omega t)).$$

This expression results in a wave with wavevector k including the frequency ω and the amplitude x_0. This is the motivation for a transformation which is based on a function expansion including the Fourier basis as it is used in spectral methods.

Spectral methods [Gottlieb et al., 1977] in general describe the numerical treatment of partial differential equations (PDE's). The solution function that numerically solves the PDE is written in a linear combination of its spectrum, which is defined by a set of certain orthogonal basis functions. The key aspect is that instead of searching for the solution function, we search for the coefficients of its linear combination in order to get a solution. By applying the PDE on this function we get a system of conditions for the coefficients and a truncation leads to the finite problem of finding all coefficients to get an approximation of the exact solution.

The main difference between spectral methods and other numerical methods is the global approach of finding a solution. Each coefficient of the linear combination depends on the information of the entire domain, in contrast to for example finite element, finite difference or finite volumes methods which just depend on the information of their neighborhood. Therefore, we need to consider those orthogonal basis functions in the linear combination that are not only nonzero on a small subdomain, but on the entire domain of the function.

Regarding the problem (4.10), instead of searching for the optimal configuration $\mathbf{x}^*$ we search for the coefficients of the expansion of an appropriate coordinate transformation.

Let I be a set of multi-indices, a function $g : \mathbb{R}^3 \to \mathbb{R}$ with sufficient conditions can be expanded with the help of suitable basis functions, i.e.

$$g(\mathbf{x}) = \sum_{\mathbf{k} \in I} \tilde{c}_{\mathbf{k}} \cdot \varphi_{\mathbf{k}}(\mathbf{x})$$

where $\varphi_{\mathbf{k}} : \mathbb{R}^3 \to \mathbb{R}$ describes an appropriate spectral basis function. Examples are given by the Fourier basis and Chebyshev polynomials with the index sets $I = \mathbb{Z}^3$ and $I = \mathbb{N}^3$, respectively. A sum over a truncated index set I_K, defined by

$$I_K = \{\mathbf{k} \in I \mid |\mathbf{k}|_\infty := \max_{j \in \{1,2,3\}} |k_j| \leq K\} \tag{4.12}$$

leads to a finite expansion.

Each multi-index $\mathbf{k} \in I_K$ has a unique coefficient vector $\tilde{\mathbf{c}} \in \mathbb{R}^{|I_K|}$ with entries $\tilde{c}_{\mathbf{k}}$. Thus, the finite expansion or rather the resulting approximated function is uniquely determined by the coefficient vector and we obtain

$$g_{\tilde{\mathbf{c}}}(\mathbf{x}) = \sum_{\mathbf{k} \in I_K} \tilde{c}_{\mathbf{k}} \varphi_{\mathbf{k}}(\mathbf{x}).$$

Next, we consider three auxiliary functions $g_{\mathbf{c}^1}, g_{\mathbf{c}^2}, g_{\mathbf{c}^3}$ which are uniquely defined by three coefficient vectors $\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3 \in \mathbb{R}^{|I_K|}$.

With regard to the coordinate vector of each atom $\mathbf{x}_i \in \mathbb{R}^3$, we introduce the following two coordinate transformations based on those functions. The additive transformation is given by

$$t^{\text{add}}_{\mathbf{x}_i}(\mathbf{c}) = \Big(x_{i1} + g_{\mathbf{c}^1}(\mathbf{x}_i), x_{i2} + g_{\mathbf{c}^2}(\mathbf{x}_i), x_{i3} + g_{\mathbf{c}^3}(\mathbf{x}_i)\Big) \tag{4.13}$$

and the multiplicative transformation is defined by

$$t^{\text{mult}}_{\mathbf{x}_i}(\mathbf{c}) = \Big(x_{i1} \cdot \big(1 + g_{\mathbf{c}^1}(\mathbf{x}_i)\big), x_{i2} \cdot \big(1 + g_{\mathbf{c}^2}(\mathbf{x}_i)\big), x_{i3} \cdot \big(1 + g_{\mathbf{c}^3}(\mathbf{x}_i)\big)\Big) \tag{4.14}$$

with $\mathbf{c} = \{\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3\}$ the collection of the designated coefficient vectors. Note, that the identity is obtained by choosing $g_{\mathbf{c}^j} \equiv 0$ for $j = 1, 2, 3$.

The transformation $t_{\mathbf{x}} : \mathbb{R}^{|I_K|} \to \mathbb{R}^{3N}$ for the complete configuration vector $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{3N}$ is defined by

$$t_{\mathbf{x}}(\mathbf{c}) = \big(t_{\mathbf{x}_1}(\mathbf{c}) \ldots, t_{\mathbf{x}_N}(\mathbf{c})\big) \tag{4.15}$$

where $t_{\mathbf{x}_i}$, $i = 1, \ldots, N$, describes either the additive transformation (4.13) or the multiplicative transformation (4.14) for each $\mathbf{x}_i$.

With this, the optimization of potential energy functions changes in the following sense. Instead of searching for the optimal atom configuration directly, we search for the optimal coefficients of the transformation function $t_{\mathbf{x}}$. Applied to an initial configuration $\mathbf{x}$, the optimal transformation function then gives the optimized atom positions. Thus, the task is the global optimization of the composite function $U \circ t_{\mathbf{x}} : \mathbb{R}^{3|I_K|} \to \mathbb{R}$ and the optimization problem (4.10) changes to

$$\boxed{\begin{array}{ll} \text{Find} & U(\mathbf{x}^*) = \min\limits_{\mathbf{c}} U\big(t_{\mathbf{x}}(\mathbf{c})\big) \\ & \text{s.t. } t_{\mathbf{x}}(\mathbf{c}) \in M \end{array}} \tag{4.16}$$

with a given starting configuration $\mathbf{x}$.

The dimension of this optimization problem is no longer dependent on the number of atoms in the underlying physical system but on the number of coefficients in the function expansions $g_{\mathbf{c}^1}, g_{\mathbf{c}^2}$ and $g_{\mathbf{c}^3}$. Thus, this method can also be applied for the optimization of physical systems which are dependent on a high number of particles.

Since initially, it is not known how many terms are necessary in order to get a good approximation of the transformation function, we define a multilevel method in which each number for K in the index set (4.12) represents a different level. For the multiplicative transformation, we define additionally the level $L = 0$ by setting $g_{\mathbf{c}^j} = c^j$ with constant $c^j \in \mathbb{R}$, for $j = 1, 2, 3$. Thereby, we can eliminate scaling factors in the atom configuration.

Multilevel Optimization via Subspace Correction

For the optimization of the coefficients we use the subspace correction method. The subspaces are chosen such that we optimize only a subset of all coefficients for all of the three function expansions. An example is given by the choice of optimizing one coefficient $c_{\mathbf{k}^j}$ for each function $g_{\mathbf{c}^j}$, $j = 1, 2, 3$ in each iteration step. This yields a three-dimensional optimization problem for each subspace. Consequently, the subspaces for each Level L are given by

$$\mathcal{V}_m^L = \Big\{ \mathbf{c} \in M_c^L \;\Big|\; c_{\mathbf{k}}^j = 0 \;\text{ for } j = 1, 2, 3 \;\text{ if }\; \mathbf{k} \notin I_m \Big\}$$

with $I_m \subset I_K$ describing the set of multi-indices which is considered for the optimization in the m-th subspace. In the case of optimizing three coefficients at once, we have N_L subspaces in each level L.

As motivated for the optimization of the atom coordinates themselves, we only perform a correction in each subspace after which the candidate solution for the configuration is updated. Thus, we restrict the space of coefficients to the subset

$$M_c^L = [a_c, b_c]^{3N_L} \subset \mathbb{R}^{3N_L}$$

with $a_c, b_c \in \mathbb{R}$, N_L being the number of coefficients in each g_j, $j = 1, 2, 3$, for the level L.

This can also be extended for a larger number of coefficients that are optimized in one iteration step.

Due to the definition of the transformation (4.15), each subset of coefficients affects all coordinates of $\mathbf{x}$. In the context of the optimization of potential energy functions, we can change all coordinates of $\mathbf{x}$ and thus, change the configuration of the entire physical system by using only one subspace correction iteration step. The definition of subspace correction

methods ensures that in each step the solution candidate does not get worse. Thus, each multilevel optimization step enables the possibility of changing all atom coordinates in order to get a new (better) starting point for the next level.

The final multilevel algorithm is a combination of the previously described multilevel method including the transformation $t_{\mathbf{x}}$ and the direct optimization of the coordinates of each atom $\mathbf{x}_i$, $i = 1, \ldots, N$. Thus, we have the possibility to change the positions of all coordinates for different levels and thereafter, we can do an accurate optimization of each atom, each atom pair or each atom triple.

The resulting algorithm for the parallel subspace correction method is given by algorithm 5 and the successive one is given by algorithm 6. In this context, we can also choose the algorithms such that a parallel subspace correction is used for the multilevel method and the successive method for the direct atom coordinate optimization or vice versa.

Sparse Grid Approach

Next, we present a sparse grid approach in order to reduce the number of coefficients to be optimized in order to improve the efficiency of the algorithm.

The sparse grid method in general is used to compute numerical discretizations for high dimensional problems. Let d be the dimension of our problem and K be the number of grid points of every single coordinate direction. The resulting degrees of freedom for a full grid are in $\mathcal{O}(K^d)$, which can be reduced to $\mathcal{O}(K(\log K)^{d-1}))$ by the use of sparse grids, see for example [Gerstner and Griebel, 2010] for a more detailed description.

Applied to the transformation function $t_{\mathbf{x}}$ and the expansion g, we choose the norm

$$|\mathbf{k}|_1 := \sum_{j=1}^{3} |k_j|,$$

instead of the $|\cdot|_\infty$-norm. The expansion with a reduced number of coefficients is then given by

$$g(\mathbf{x}) \approx \sum_{\substack{\mathbf{k} \in I_K \\ |\mathbf{k}|_1 \leq K}} c_{\mathbf{k}} \cdot \varphi_{\mathbf{k}}(\mathbf{x})$$

with

$$I_K = \{\mathbf{k} \in I \mid |\mathbf{k}|_1 \leq K\}$$

The $|\cdot|_1$-norm has turned out to achieve good results in the context of expansions for polynomials. Thus, if we assume that the expansion can approximate the expansion including the $|\cdot|_1$-norm, we can increase the efficiency of the resulting algorithm.

4.4.4 The Fourier Basis

Since we are interested in a global approach, we consider basis functions that are not only nonzero on a small subdomain, but on the entire considered domain of the underlying function.

Consider a regulated 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$, i.e.

$$f(x + 2\pi) = f(x) \qquad \text{for all } x \in \mathbb{R}.$$

Algorithm 5: Multilevel Method (based on the Parallel Subspace Correction)

Given $\mathbf{x} \in M$, $\mathbf{c}^{(0)} = 0$

% multilevel method
for *each level* L **do**
  for $l = 0, 1, 2, \ldots$ **do**
    $\mathbf{v} := \mathbf{c}^{(l)}$
    for $m = 1, 2, \ldots, N_L$ **do**

$$\hat{\mathbf{e}}_i = \operatorname*{argmin}_{\mathbf{e} \in \mathcal{V}_m^L} U\big(t_{\mathbf{x}}(\mathbf{c} + \mathbf{e})\big)$$

    end
    Find t such that

$$U\big(t_{\mathbf{x}}(\mathbf{v} + \alpha_t \hat{\mathbf{e}}_t)\big) = \min\big(U\big(t_{\mathbf{x}}(\mathbf{v} + \alpha_1 \hat{\mathbf{e}}_1)\big), \ldots, U\big(t_{\mathbf{x}}(\mathbf{v} + \alpha_{N_L} \hat{\mathbf{e}}_{N_L})\big)\big)$$

    and set

$$\alpha_t = 1 \text{ and } \alpha_m = 0 \text{ for all } m \neq t \quad \% \text{ Selection Strategy}$$

    $\mathbf{v} := \mathbf{v} + \sum_{m=1}^{N_L} \alpha_m \hat{\mathbf{e}}_m$
    $\mathbf{c}^{l+1} = \mathbf{v}$
    $\mathbf{x} = t_{\mathbf{x}}(\mathbf{v})$
  end
end

% direct optimization of atom coordinates
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{x}^k$
  for $m = 1, 2, \ldots, N$ **do**

$$\hat{\mathbf{e}}_i = \operatorname*{argmin}_{\mathbf{e} \in \mathcal{V}_m^{1\text{-atom}}} U_m(\mathbf{x}^k + \mathbf{e})$$

  end
  Find t such that

$$U_m(\mathbf{v} + \alpha_t \hat{\mathbf{e}}_t) = \min\left(U_m(\mathbf{v} + \alpha_1 \hat{\mathbf{e}}_1), \ldots, U_m(\mathbf{v} + \alpha_N \hat{\mathbf{e}}_N)\right)$$

  and set

$$\alpha_t = 1 \text{ and } \alpha_m = 0 \text{ for all } m \neq t \quad \% \text{ Selection Strategy}$$

  $\mathbf{v} := \mathbf{v} + \sum_{m=1}^{N} \alpha_m \hat{\mathbf{e}}_m$
  $\mathbf{u}^{k+1} = \mathbf{v}$
end

Algorithm 6: Multilevel Method (based on the Successive Subspace Correction)

Given $\mathbf{x} \in M$, $\mathbf{c}^{(0)} = 0$

% multilevel method
for *each level* L **do**
  for $l = 0, 1, 2, \ldots$ **do**
    $\mathbf{v} := \mathbf{c}^{(l)}$
    for $m = 1, 2, \ldots, N_L$ **do**
      $\hat{\mathbf{e}} = \underset{\mathbf{e} \in \mathcal{V}_m^L}{\operatorname{argmin}} \, U\big(t_{\mathbf{x}}(\mathbf{c} + \mathbf{e})\big)$
      $\mathbf{v} := \mathbf{v} + \hat{\mathbf{e}}$
    end
  end
  $\mathbf{c}^{l+1} = \mathbf{v}$
  $\mathbf{x} = t_{\mathbf{x}}(\mathbf{v})$
end

% direct optimization of atom coordinates
for $k = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} := \mathbf{x}^k$
  for $m = 1, 2, \ldots, N$ **do**
    $\hat{\mathbf{e}} = \underset{\mathbf{e} \in \mathcal{V}_m^{\text{1-atom}}}{\operatorname{argmin}} \, U_m(\mathbf{x}^k + \mathbf{e})$
    $\mathbf{v} := \mathbf{v} + \hat{\mathbf{e}}$
  end
  $\mathbf{x}^{k+1} = \mathbf{v}$
end

The Fourier basis elements are defined by

$$e_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$$

and are orthonormal with respect to the L^2-hermitian inner product on $[0, 2\pi]$, because

$$\begin{aligned}\langle e_m(x), e_n(x)\rangle_{L^2} &= \int_0^{2\pi} \overline{\frac{1}{\sqrt{2\pi}} e^{imx}} \frac{1}{\sqrt{2\pi}} e^{inx} dx \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{i(n-m)x} dx \\ &= \begin{cases} 1 & \text{if } m = n; \\ 0 & \text{if } m \neq n. \end{cases}\end{aligned}$$

The Fourier series of f is defined by

$$Sf(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx} \tag{4.17}$$

with coefficients

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ikt} dt \in \mathbb{C},$$

see for example [Königsberger, 2003] for further details.

Next, we extend the Fourier series (4.17) to a three dimensional 2π-periodic function $f : \mathbb{R}^3 \to \mathbb{C}$, that is 2π-periodic in each variable, and obtain basically the same expression. We can write

$$\begin{aligned}Sf(\mathbf{x}) = Sf(x_1, x_2, x_3) &= \sum_{k_1,k_2,k_3=-\infty}^{\infty} c_{k_1} e^{ikx_1} \cdot c_{k_2} e^{ik_2x_2} \cdot c_{k_3} e^{ik_3x_3} \\ &= \sum_{k_1,k_2,k_3=-\infty}^{\infty} c_{\mathbf{k}}\, e^{i(k_1x_1+k_2x_2+k_3x_3)} \\ &= \sum_{\mathbf{k}\in\mathbb{Z}^3} c_{\mathbf{k}}\, e^{i\mathbf{kx}}. \end{aligned} \tag{4.18}$$

Since we take this function representation for the function g in the transformation function (4.15) in the context of real valued potential energy functions, we can assume that g is also real-valued. For that reason we can conclude that for the coefficients in (4.18) it holds that

$$\overline{c_{\mathbf{k}}} = c_{-\mathbf{k}}, \tag{4.19}$$

with $\overline{c_{\mathbf{k}}}$ denoting the complex conjugate of $c_{\mathbf{k}}$ and $-\mathbf{k} = (-k_1, -k_2, -k_3)$.

Furthermore we will apply the following identities that follow from Euler's formula [Königsberger, 2003]

$$\sin(x) = \frac{1}{2\mathrm{i}} \left(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}\right) \qquad \text{and} \qquad \cos(x) = \frac{1}{2} \left(\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}\right) \tag{4.20}$$

and set

$$c_{\mathbf{k}} = \frac{1}{2}(a_{\mathbf{k}} - ib_{\mathbf{k}}).$$

We can split the elements of $\mathbb{Z}^3$ into two ones $\mathcal{Z}^+$ and $\mathcal{Z}_0^-$, which are each identified by containing all elements of $\mathbb{Z}^3$ with opposite signs and just the latter one contains the zero such that

$$\mathbb{Z}^3 = \mathcal{Z}^+ \cup \mathcal{Z}_0^- \quad \text{with} \quad \mathcal{Z}^+ \cap \mathcal{Z}_0^- = \emptyset.$$

It follows for the function $g_{\mathbf{c}}$ in the transformation (4.15) that

$$\begin{aligned} g_{\mathbf{c}}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^3} c_{\mathbf{k}}\, e^{i\mathbf{k}\mathbf{x}} \quad &= c_0 + \sum_{\mathbf{k}\in\mathcal{Z}^+} c_{\mathbf{k}} e^{i\mathbf{k}\mathbf{x}} + c_{-\mathbf{k}} e^{-i\mathbf{k}\mathbf{x}} \\ &\overset{(4.19)}{=} c_0 + \sum_{\mathbf{k}\in\mathcal{Z}^+} \frac{1}{2}(a_{\mathbf{k}} - ib_{\mathbf{k}})e^{i\mathbf{k}\mathbf{x}} + \frac{1}{2}(a_{\mathbf{k}} + ib_{\mathbf{k}})e^{-i\mathbf{k}\mathbf{x}} \\ &\overset{(4.20)}{=} c_0 + \sum_{\mathbf{k}\in\mathcal{Z}^+} \Big(a_{\mathbf{k}} \cos(\mathbf{k}\mathbf{x}) + b_{\mathbf{k}} \sin(\mathbf{k}\mathbf{x})\Big). \end{aligned} \tag{4.21}$$

This is finally the Fourier series based transformation function.

Furthermore, one should note that the choice of a 2π-periodic function was arbitrary, because we can rescale and replace the expressions in (4.21) by

$$g_{\mathbf{c}}(\mathbf{x}) = c_0 + \sum_{\mathbf{k}\in\mathcal{Z}^+} \left(a_{\mathbf{k}} \cos\Big(\frac{2\pi}{P}\,\mathbf{k}\mathbf{x}\Big) + b_{\mathbf{k}} \sin\Big(\frac{2\pi}{P}\mathbf{k}\mathbf{x}\Big)\right),$$

with P denoting the periodicity of the underlying function.

5 Numerical Results

In this chapter we present the results of the subspace correction optimization and the multilevel optimization method described in section 4.4. Therefore, we need different test examples in order to evaluate their applicability.

We test the subspace correction by the application on the Lennard-Jones potential which is only dependent on pair-interactions. Furthermore, we consider the optimization of the Gupta cluster for zinc which is described by a many-body contribution, additional to a pair potential. The multilevel method is tested for a bond potential and the Lennard-Jones potential with periodic boundary conditions by assuming perturbed starting configurations.

5.1 Subspace Correction Optimization

In this section, we consider the direct optimization of the coordinates of each atom $\mathbf{x}_i \in \mathbb{R}^3$, $i = 1, \ldots, N$, for clusters and the corresponding potential energy function $U : \mathbb{R}^{3N} \to \mathbb{R}$ via the subspace correction optimization. The resulting non-linear, non-convex optimization problem is given by

$$\boxed{\begin{array}{ll} \text{Find} & U(\mathbf{x}^*) = \min\limits_{x \in \mathbb{R}^{3N}} U(\mathbf{x}), \\ \text{s.t.} & x \in M. \end{array}}$$

The choice of the set M is motivated by the fact that the energy for each atom pair tends to zero for distances that are sufficiently large. Thus, we assume that the optimal configuration is inside a domain in which each atom has at most some distance $b \in \mathbb{R}^+$ to all other atoms, see figure 5.1. For convenience we choose M such that

$$M = [-b, b]^{3N}.$$

The idea of subspace correction methods is to correct the current solution candidate in each iteration. We place particular emphasis on the repetition of the iterations instead of finding

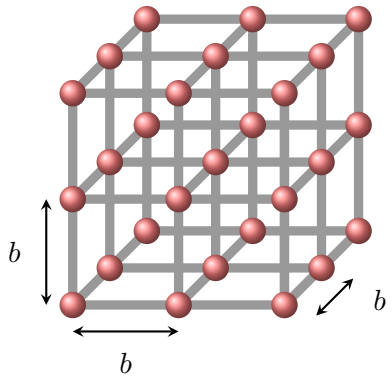


Figure 5.1: Configuration of $N = 27$ atoms, in which each atom has distance b to all other atoms.

the best possible minimal value for a subproblem in each iteration step. Instead of running a local optimization until convergence, we apply only a fixed number of global optimization steps. Thereby, we try to avoid the problem of getting stuck in local minima.

For the optimization in each iteration, we use the Multi-Level-Single-Linkage (MLSL) optimization algorithm from the NLopt nonlinear-optimization package from Steven G. Johnson, see [Johnson, 2013]. It describes a global optimization algorithm which uses a multistart technique in order to find the smallest function value. This means that the objective function is optimized via certain number of local optimizations from a set of different starting points. The number of starting points, the so-called population, can be varied and are chosen randomly but using a single linkage clustering. This method classifies regions in which we already run a local search and no smaller minimum is expected in order to prevent repeated searches. We use both the derivative-free and the derivative-based implementation. For a more detailed information of this algorithm, see [Kan and Timmer, 1987a, Kan and Timmer, 1987b]. In the context of the optimization via subspace correction, we choose the fixed number of 200 function evaluations for each global optimization with a population of 50.

We already described in chapter 4 that it is a demanding task to find any stopping criteria since we cannot make any statements about whether or when a global optimization algorithm converges towards the global minimum. Hence, the risk is that we stop the algorithm before the global minimum is found. For the subspace correction optimization, we choose to stop the algorithm if the energy value of the original function does not change 20 times up to the sixth decimal place or the maximal number of 2000 iterations is reached.

Regarding the starting configuration, a frequently used approach in the context of potential energy functions is based on the best configuration for the same function including one atom less. Given the positions of the first $N-1$ atoms, we choose the position for the N-th atom randomly. However, since we are interested in testing the algorithm with respect to its applicability, we test the method without using any knowledge about good starting configurations. Therefore, they are chosen randomly and we test $T = 5$ different starting points $\mathbf{x}^{(0)^j} \in M$, $j = 1, \ldots, T$.

In order to compare our results, we use the so far best known global minima of Lennard-Jones and Gupta clusters, collected in the Cambridge Cluster database (CCD) [Wales et al., 2014] on the internet.

We test the following methods described in section 4.4.2.

- **One-Atom Parallel Optimization (1-PSC)**. We optimize the three-dimensional coordinates of each atom i, $i = 1, \ldots, N$, in one iteration step by using the parallel subspace correction method for optimization with the selection strategy.

- **Two-Atom Parallel Optimization (2-PSC)**. We optimize the three-dimensional coordinates of each pair of atoms (i, j), $i, j = 1, \ldots, N$, in one iteration step by using the parallel subspace correction method for optimization with the selection strategy and the improved order of atoms.

- **One-Atom Successive Optimization (1-SSC)**. We optimize the three-dimensional coordinates of each atom i, $i = 1, \ldots, N$, in one iteration step by using the successive subspace correction method for optimization.

- **Two/Three-Atom Successive Optimization (2a/3a-SSC)**. We optimize the six- and nine-dimensional coordinates of each pair (i,j), $i,j = 1,\dots,N$, and triple (i,j,k), $i,j,k = 1,\dots,N$, of atoms in one iteration step, respectively, by using the successive subspace correction method for optimization.

- **Two/Three-Atom Successive Optimization with improved order of atoms (2b/3b-SSC)**. We optimize the six- and nine-dimensional coordinates of each pair (i,j), $i,j = 1,\dots,N$, and triple (i,j,k), $i,j,k = 1,\dots,N$, of atoms in one iteration step, respectively, by using the successive subspace correction method for optimization and an improved order of atoms.

In order to evaluate the results, we consider the number of successes, the mean error and the development of the function value dependent on the number of function evaluations for each method. Thereby, a success is denoted by an optimization which leads to the best known energy value known from the CCD. The mean error represents a measure for the average distance between the computed solution and the reference value. Let $U(\tilde{\mathbf{x}}^j)$ be the computed solution based on the starting configuration $\mathbf{x}^{(0)^j}$, the mean error is defined by

$$e_{mean} := \sum_{j=1}^{T} \frac{|U(\tilde{\mathbf{x}}^j) - U(\mathbf{x}^*)|}{T}$$

with $U(\mathbf{x}^*)$ describing the minimal energy value from the CCD and T as the number of different starting points.

5.1.1 Lennard-Jones Cluster

In order to analyze the applicability of the subspace correction optimization on potential energy functions, we consider the Lennard-Jones-$(12,6)$-Potential, which was defined in section 3.3. We recall that this potential is described by

$$U^{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] = 4\left[\left(\frac{1}{r_{ij}}\right)^{12} - \left(\frac{1}{r_{ij}}\right)^{6}\right]$$

with $r_{ij} = ||\mathbf{x}_j - \mathbf{x}_j||$ describing the distance between the i-th and j-th atom.

Hence, the Lennard-Jones Cluster for N atoms with coordinates $\mathbf{x}_1,\dots,\mathbf{x}_N \in \mathbb{R}^3$ is given by

$$U(\mathbf{x}) = U(\mathbf{x}_1,\dots\mathbf{x}_N) = \sum_{i=1}^{N}\sum_{j=i+1}^{N} U^{LJ}(r_{ij}).$$

Furthermore, we choose equal parameters $\sigma_{ij} := \sigma$ and $\epsilon_{ij} := \epsilon$ for all $i,j = 1,\dots,N$, such that

$$\sigma_{ij} = \epsilon_{ij} = 1.$$

For the Lennard-Jones cluster, the choice of M is motivated by the fact that each $U^{LJ}(r_{ij})$ achieves its minimum at $r_{min} = 2^{1/6}\cdot\sigma = 2^{1/6}$. Since we search for the global minimum of the

	Derivative-free		Gradient-based	
Method	#successes	e_{mean}	#successes	e_{mean}
1-PSC	4/5	0.2896	1/5	1.2873
2-PSC	1/5	4.7675	2/5	0.8687
1-SSC	4/5	0.3816	3/5	0.5791
2a-SSC	4/5	0.9301	4/5	0.8203
2b-SSC	3/5	1.0380	5/5	0.0000

Table 5.1: Results for the optimization of a Lennard-Jones cluster with $N = 15$ atoms via different methods.

full cluster, we assume the optimal configuration to be inside a domain in which each atom has at most the distance r_{min} to all other atoms, i.e.

$$b = 2^{1/6} \cdot \sqrt[3]{N-1} \quad \text{and} \quad M = [-b, b]^{3N}.$$

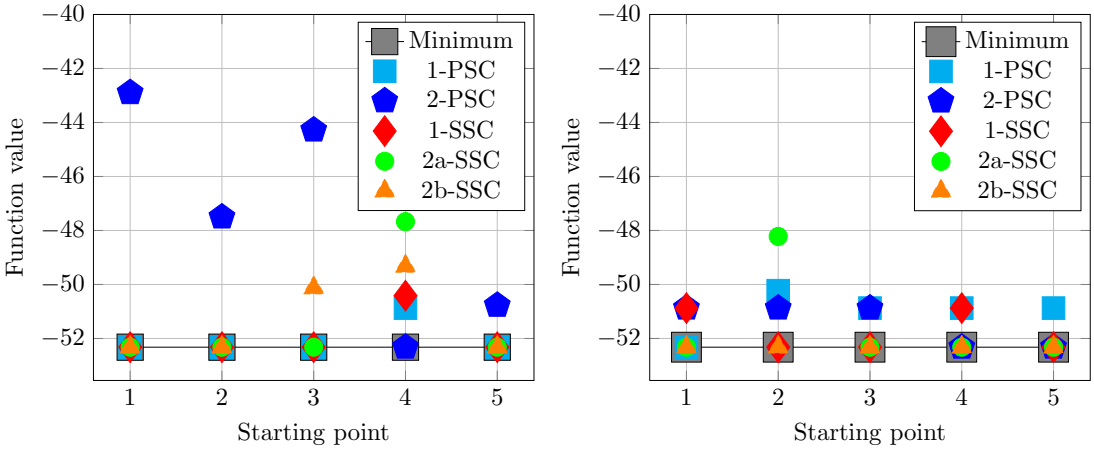
We begin with the different subspace correction methods for a Lennard cluster with $N = 15$ atoms in order to compare the parallel and the successive types of methods. The best known minimum listed in the CCD yields a potential energy value of -52.322627. In table 5.1, we listed the number of successes and the mean errors of all tested methods for five different random starting points. We denote the methods using gradient information in the MLSL algorithm by gradient-based and those doing not by derivative-free. For comparison, we also run an optimization using only the derivative-free and gradient-based MLSL algorithm directly for the original function $U(\mathbf{x}) = U(\mathbf{x}_1, \ldots, \mathbf{x}_{15})$ for each starting point. The best result out of all optimizations only leads to a potential energy value of -37.2251 after $2 \cdot 10^6$ function evaluations. This clearly shows the advantage of the subspace correction method compared to the optimization only using the MLSL algorithm.

We point out that the best number of successes and mean errors was achieved by the 2b-SSC method, which was the only one that found the best known minimum in all of the five cases. Furthermore, except for the derivative-free 1-PSC, the parallel methods lead to worse results than the successive methods for the same subspaces. A more detailed description of this results is given by figure 5.2, where each value of the computed minima is shown.

Next, we analyze the development of the function value dependent on the number of function evaluations of the different methods. In figures 5.3-5.7 we see the results starting from the different random configurations for the derivative-free and the gradient-based types of methods. For almost all different starting points, the function value of the PSC methods decrease slower than for the SSC methods. Together with the fact that they have a less number of successes, we conclude that the successive subspace correction method is more suitable for the optimization of potential energy functions.

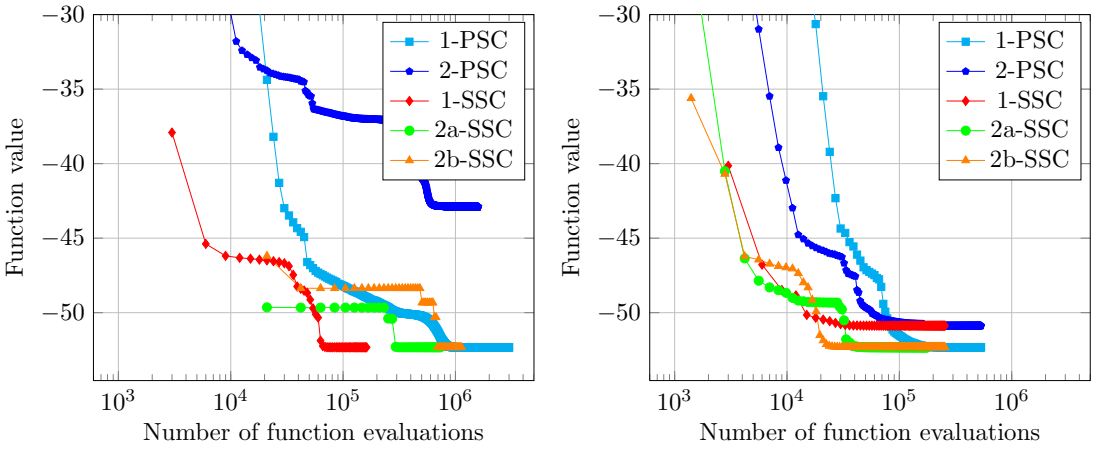
Regarding the derivative-free SSC methods it is remarkable that the one-atom optimization achieves in most of all cases by far the best results. The situation is different for the gradient-based types of algorithms. Here, the two-atom optimization with improved order of atoms becomes significantly more efficient than the same method without using gradient information.

The reason for these much better results of the gradient-based algorithms may be due to the fact that we have to perform six-dimensional optimizations during the two-atom subspace correction. Using the derivative-free method, we have too less information about the function



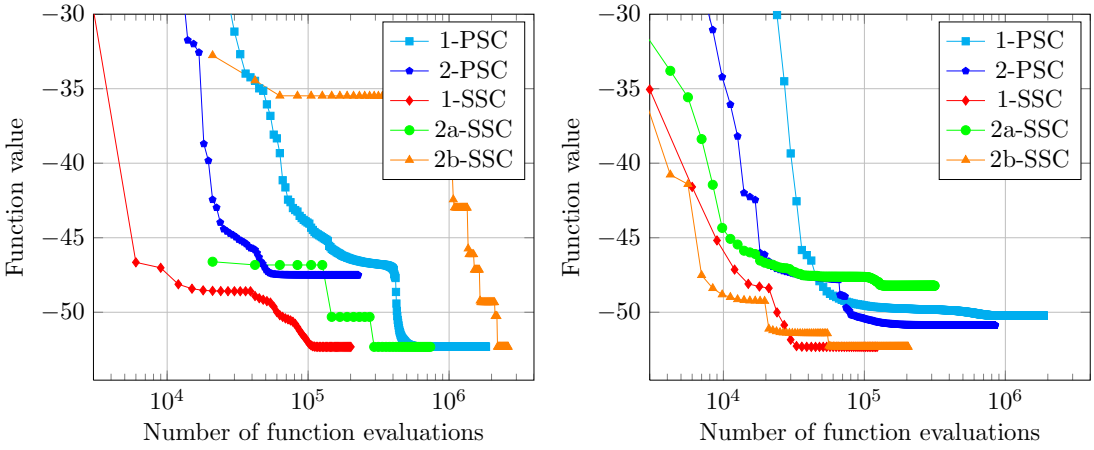
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.2: Found minima for Lennard-Jones cluster ($N = 15$).



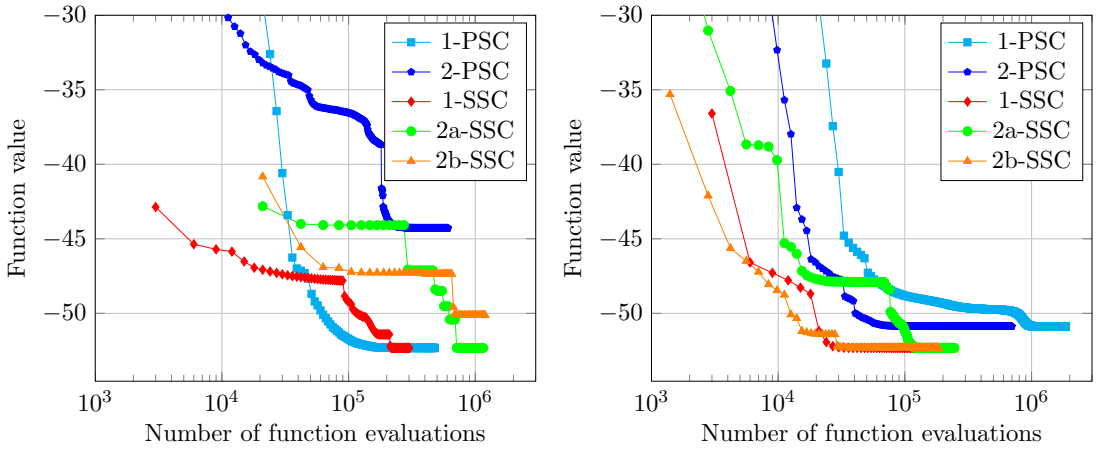
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.3: Development of the function value of starting point 1.



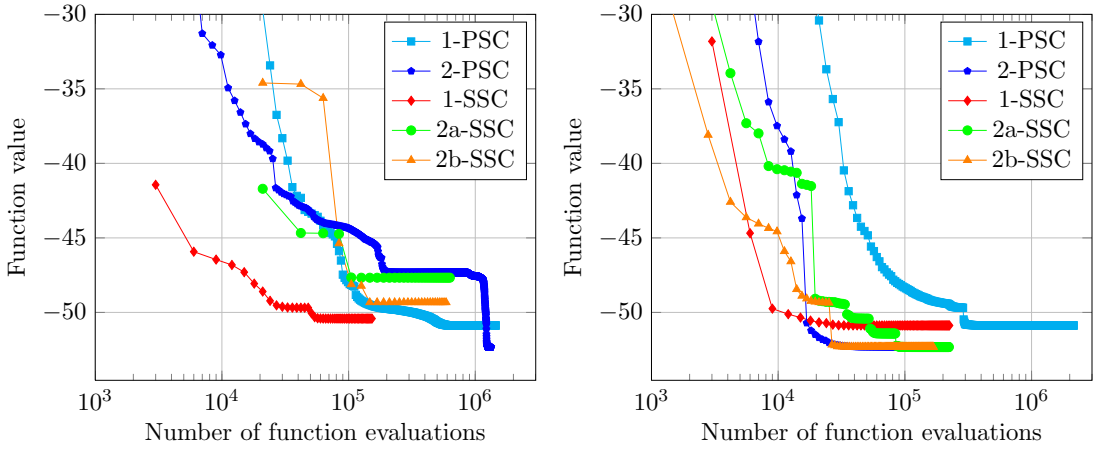
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.4: Development of the function value of starting point 2.



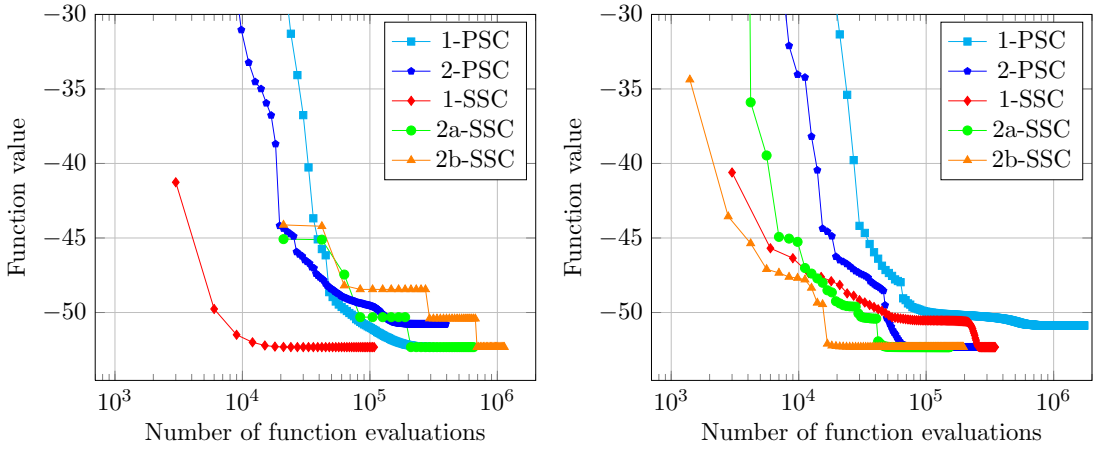
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.5: Development of the function value of starting point 3.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.6: Development of the function value of starting point 4.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.7: Development of the function value of starting point 5.

Method	derivative-free		gradient-based	
	#successes	e_{mean}	#successes	e_{mean}
1-SSC	1/5	0.5736	0/5	2.5907
2a-SSC	2/5	0.2653	4/5	0.4328
2b-SSC	1/5	0.7431	2/5	0.8172
3a-SSC	0/5	11.2647	4/5	0.0832
3b-SSC	0/5	11.5120	5/5	0.0000

Table 5.2: Results for the optimization of a Lennard-Jones cluster with $N = 25$ atoms via different methods.

and thus the optimization in each iteration step cannot yield lower function values.

In order to test whether we can improve the efficiency of the subspace correction optimization via the three-atom optimization, we consider the Lennard-Jones cluster consisting of $N = 25$ atoms. Here, we restrict to the successive method.

Again, we first evaluate the number of successes and the mean error for each type of method, see table 5.2. All of the derivative-free algorithms have a low number of successes, nevertheless we can conclude by the mean error that they achieved minima with a low distance to the reference value. Both of the derivative-free three-atom optimization methods have large mean errors. Note, that we have to perform a nine-dimensional optimization in each iteration step which can be very challenging task. Completely different it behaves for the gradient-based types of methods. Here, the three-atom method including the improved order of atoms was the only one which found the best known minimum for all of the five random starting configurations.

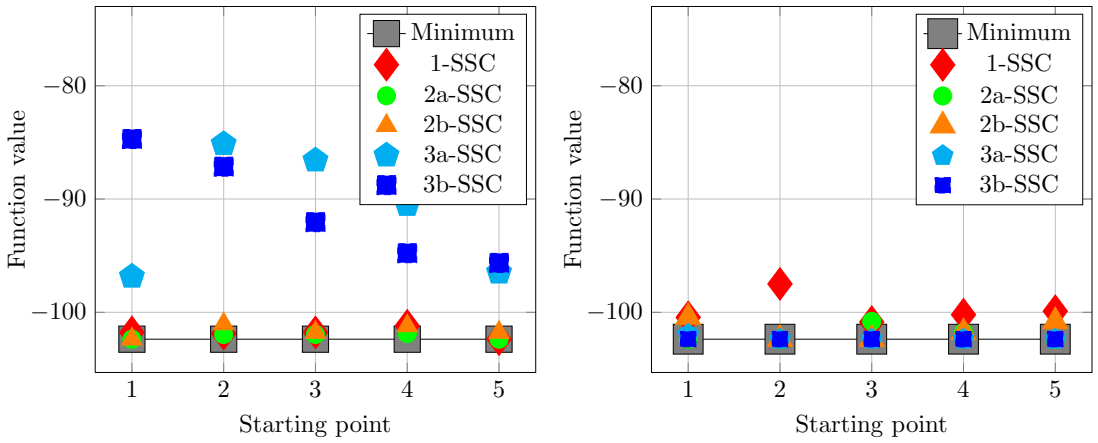
The specific results for each method is shown in figure 5.8 and the development of the function value dependent on the number of function evaluations is given by figures 5.9-5.13. For the derivative-free methods the results are comparable to those considering $N = 15$ atoms in the Lennard-Jones cluster. Regarding the gradient-based algorithms, the efficiency of the two- and three-atom based methods could be improved significantly by using gradient information.

Thus, we can conclude that the subspace correction optimization can be improved by considering a larger subset of atoms in the optimization for each subspace, even for a larger number of atoms. Additionally, we emphasize that the improved order of atoms can further improve the results. Using the subspace correction method, we found all best known minima for the Lennard-Jones cluster from the CCD up to 150 atoms.

Since the Lennard-Jones cluster consists only of pair-potentials, we consider another cluster which includes also many-body contributions.

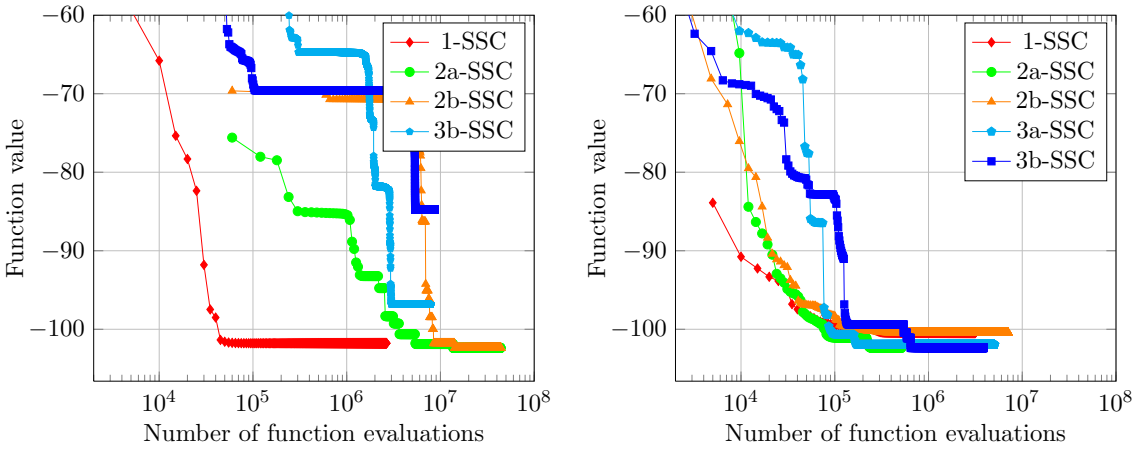
5.1.2 Gupta Cluster

The Gupta Cluster for zinc was defined in section 3.3 and is given by



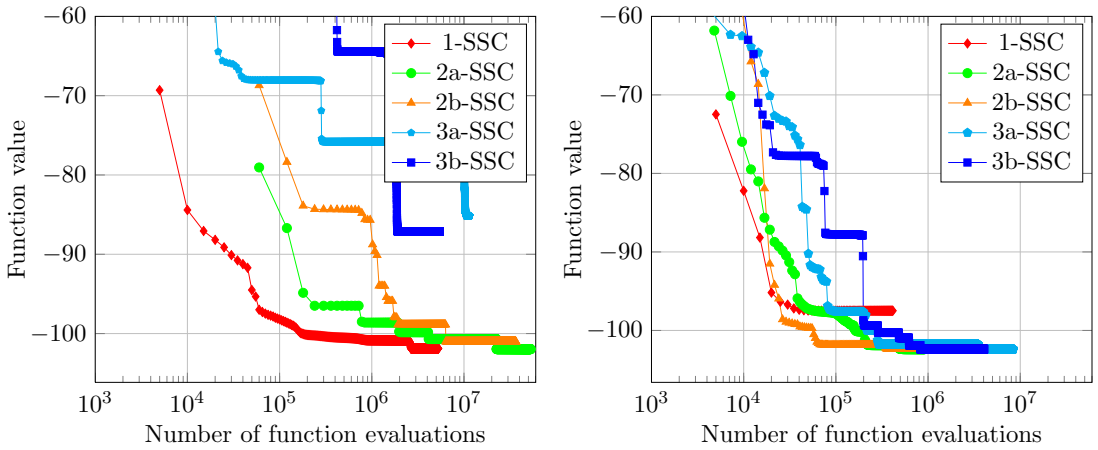
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.8: Found minima for Lennard-Jones cluster ($N = 25$).



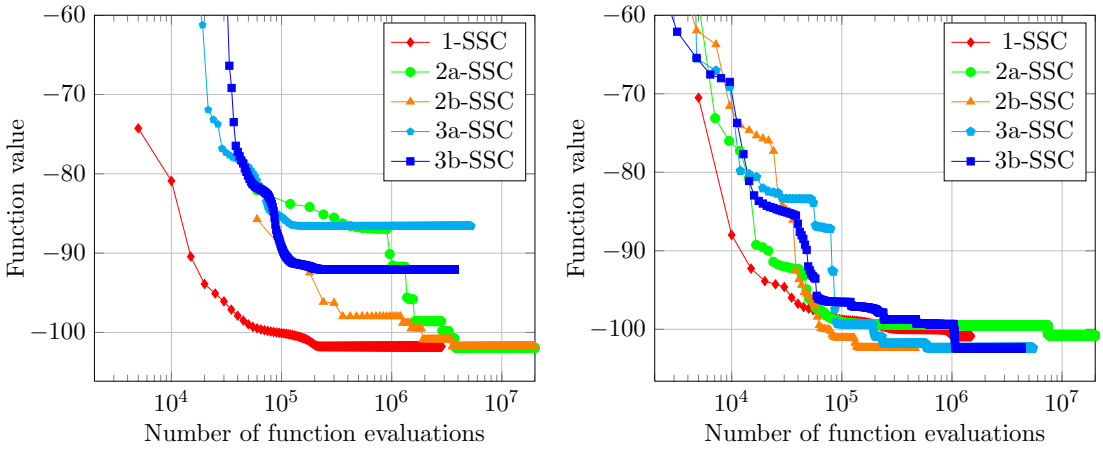
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.9: Development of the function value of starting point 1.



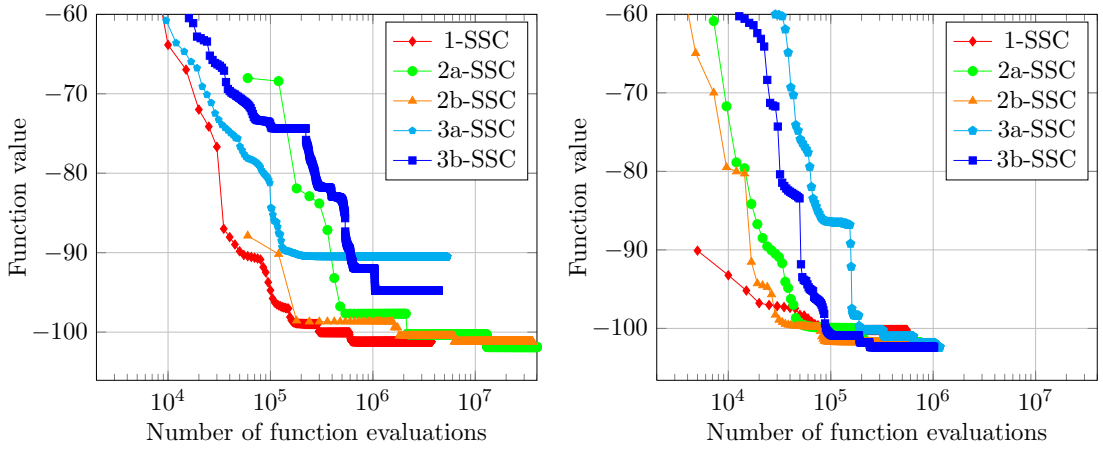
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.10: Development of the function value of starting point 2.



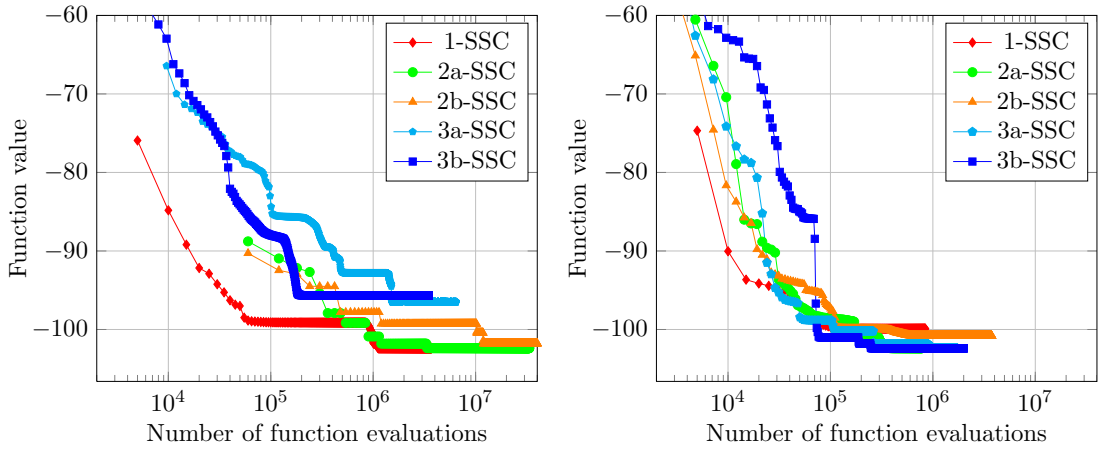
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.11: Development of the function value of starting point 3.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.12: Development of the function value of starting point 4.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.13: Development of the function value of starting point 5.

$$U(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \underbrace{\sum_{i+1}^{N}\sum_{j=i+1}^{N}\phi_{\text{eff}}(r_{ij})}_{\text{pair-potential}} + \underbrace{\sum_{i=1}^{N}F_{\text{eff}}(\bar{\rho}_i)}_{\text{many-body contribution}}$$

with

$$\phi_{\text{eff}} = 2Ae^{-p(r/r_0-1)} - \frac{\xi}{\sqrt{\bar{\rho}_{\text{xtal}}}}e^{-2q(r/r_0-1)},$$

$$F_{\text{eff}} = -\xi\sqrt{\bar{\rho}}\left(1-\frac{1}{2}\sqrt{\frac{\bar{\rho}}{\bar{\rho}_{\text{xtal}}}}\right)$$

$$\bar{\rho}_i = \sum_{\substack{j=1\\ i\neq j}}^{N}\rho(r_{ij}) \quad \text{and}$$

$$\rho(r) = e^{-2q(r/r_0-1)}.$$

The parameters are dependent on the metal represented by this potential energy function, the specific choices are taken from [Doye, 2003]. For zinc we choose the values given by

	p	q	A/eV	ξ/eV
Zn	9.689	4.602	0.1477	0.8900

with $\bar{\rho}_{xtal} = 8.638$ and we set $r_0 = 1$.

The function $F_{\text{eff}}(\bar{\rho})$ describes a many-body contribution to the total function U dependent on the sum of an "atomic density" function ρ. Based on the construction of F_{eff} and ϕ_{eff} and on the choice of parameters for zinc, the main energy contribution comes from the many-body term. Due to its more complex structure we expect the optimization to be more difficult than for pair-potentials.

The main characteristic concerning the lowest energy configurations of Gupta clusters is its disordered structure. This means that they are not in their energy minimum at the usual ordered form, for example given by icosahedral or decahedral forms. Hereby, the parameters for zinc lead to one of those clusters with the highest disorders. See [Doye, 2003] and the references therein for a more detailed description of those structures.

The choice of the bounding set M is motivated by the minimizing distances of the pair potential at

$$r_{min} = r_0\left(1+\frac{1}{p-2q}ln\left(\frac{Ap\sqrt{\bar{\rho}_{\text{xtal}}}}{\xi q}\right)\right).$$

We assume the optimal configuration to be inside a domain in which each atom has at most the distance $2r_{min}$ to all other atoms, i.e.

$$b = 1.0547\cdot\sqrt[3]{N-1} \quad \text{and} \quad M = [-b,b]^{3N}.$$

Indeed, the optimization of Gupta clusters proved to be more difficult than Lennard-Jones cluster. Although there are only two successes for all methods, the values of the mean errors are relatively small, see table 5.3. Thus, the optimizations lead to an minimal value that is close to the best known minimum. The detailed final results of all methods are shown in figure 5.14.

Method	derivative-free		gradient-based	
	#successes	e_{mean}	#successes	e_{mean}
1-SSC	0/5	0.0445	0/5	0.1001
2a-SSC	1/5	0.1077	0/5	0.1663
2b-SSC	1/5	0.0692	0/5	0.1976
3a-SSC	0/5	0.2932	0/5	0.1850
3b-SSC	0/5	0.2987	0/5	0.2636

Table 5.3: Results for the optimization of a Gupta cluster for zinc with $N = 10$ atoms via different methods.

The development of the function value dependent on the number of function evaluations for each randomly chosen starting point is given in figures 5.15-5.19. Here, it is not possible to give a clear statement which type of subspace correction method achieves the best results. Although, it is noticeable that at the beginning there is a faster decrease of the function value for the gradient-based methods than for the derivative-free ones, especially for the three-atom optimizations. However, only the two-atom methods without gradient information found the best known minimum. Note, that this results are based on randomly chosen starting points in order to test the quality of the subspace correction optimization.

The reference values for the global minima of Gupta clusters are given up to 125 atoms in the CCD. In order to test the applicability also for high dimensions, we tested the algorithm for $N = 50$ atoms starting from the best known configuration of $N - 1$ atoms. The resulting development of the function value for the 1-SSC and the 2b-SSC are shown in figure 5.20a. The 1-SSC clearly decreases faster and find the best known minimal value from the CCD as final result, the optimized end-configuration is shown in figure 5.20b.

Furthermore, we computed new minima of Gupta clusters with more atoms by using the 1-SSC and starting from the best known configuration of $N - 1$ atoms, given by the following table.

# atoms	computed minimum
126	-168.8215
127	-170.1066
128	-171.4166

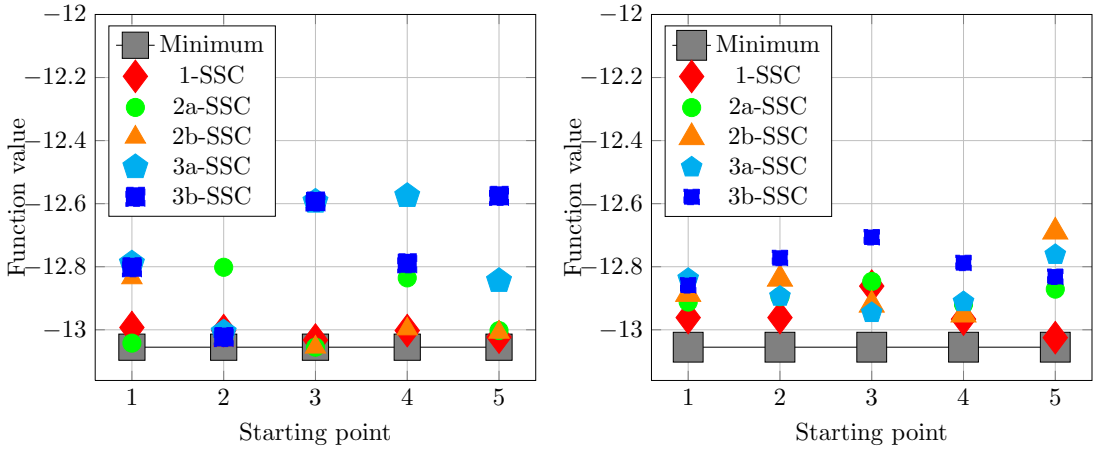
5.2 Multilevel Optimization

In this section, we consider the multilevel optimization of the coefficients of a function expansion by the use of the Fourier basis function. The resulting optimization problem is given by

$$\boxed{\begin{aligned} \text{Find} \quad & U(\mathbf{x}^*) = \min_{\mathbf{c}} U\big(t_{\mathbf{x}}(\mathbf{c})\big) \\ & \text{s.th. } t_{\mathbf{x}}(\mathbf{c}) \in M, \end{aligned}}$$

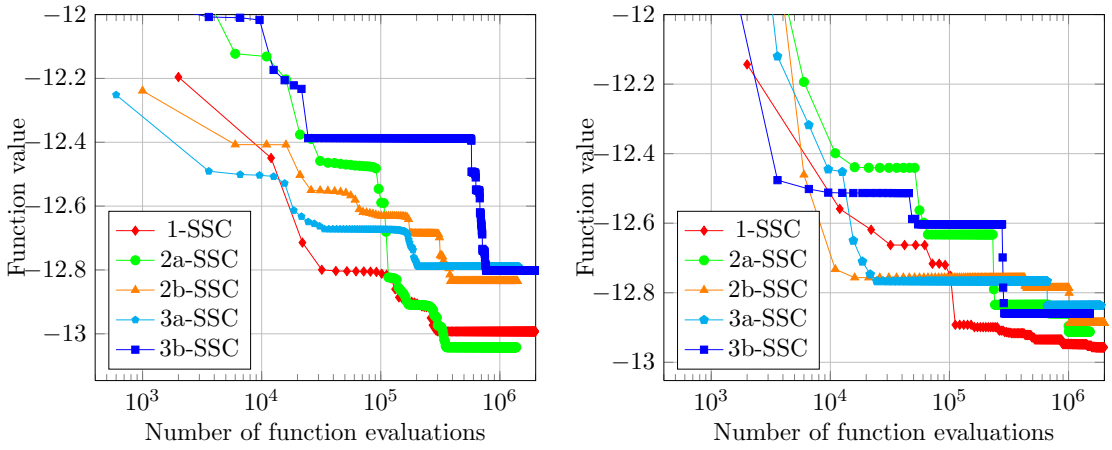
with $\mathbf{x}$ a given starting configuration.

The choice of the set M is the same as in the previous section. We test the following variants of the coordinate transformation optimization. Here, we restrict only to the successive



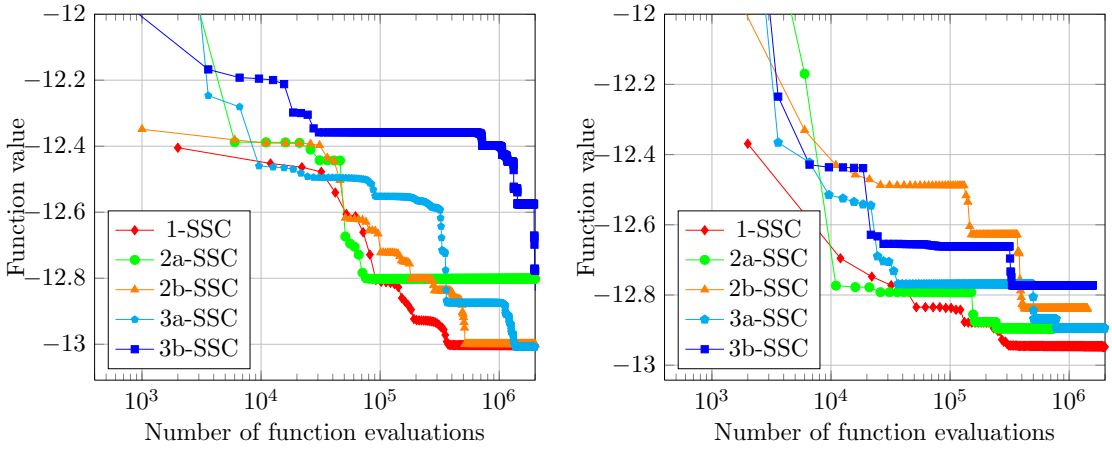
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.14: Found minima for Gupta cluster ($N = 10$).



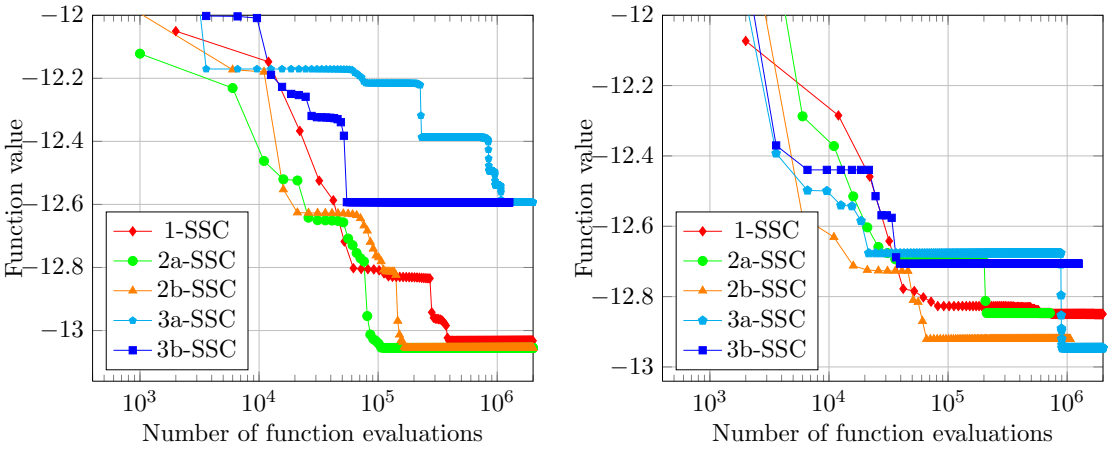
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.15: Development of the function value of starting point 1.



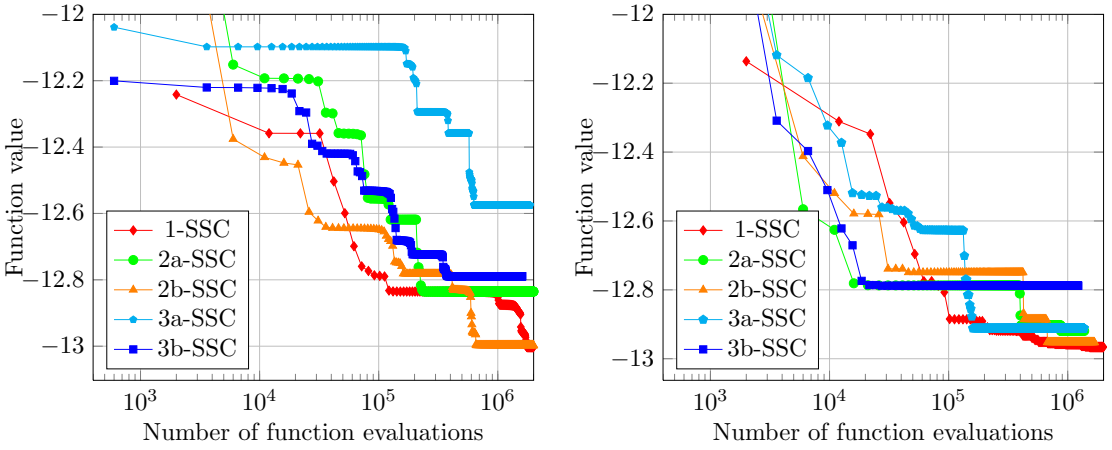
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.16: Development of the function value of starting point 2.



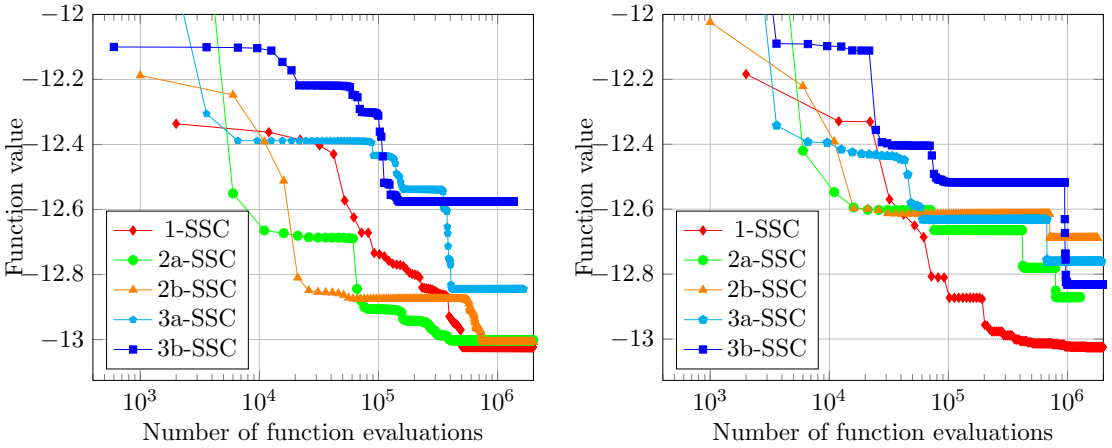
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.17: Development of the function value of starting point 3.



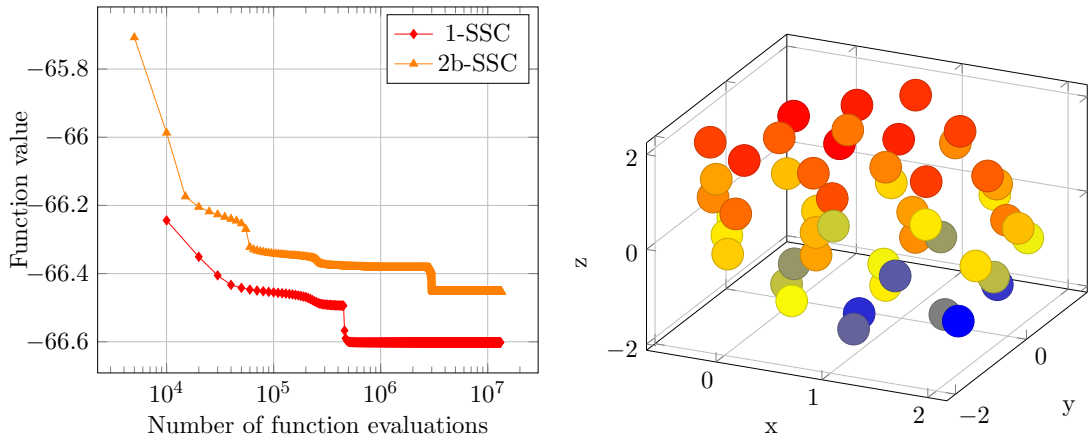
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.18: Development of the function value of starting point 4.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.19: Development of the function value of starting point 5.



(a) Development of the function value (derivative-free).

(b) Optimized configuration(1-SSC).

Figure 5.20: Gupta cluster optimization for $N = 50$ atoms.

methods and the 1-SSC for the derivative-free and the 2b-SSC for the gradient-based method since they turned out to be the most successful.

- **One Multilevel Full-Grid Optimization (1-FG$^+$/1-FG* $\mathbf{K = k}$).** We optimize the three-dimensional coordinates of each constant i, $i = 1, \ldots, N_c$, in one iteration step by using the successive subspace correction method for optimization. We use the full grid method with maximal level k with the additive (1-FG$^+$) and multiplicative (1-FG*) transformation, respectively. Thereafter, we use the 1-SSC.

- **One Multilevel Sparse-Grid Optimization (1-SG$^+$/1-SG* $\mathbf{K = k}$).** Same as (1-FG$^+$/1-FG* $\mathbf{K = k}$) with the sparse grid instead of the full grid method.

- **One-and-two Multilevel Full-Grid Optimization (1/2-FG$^+$/1/2-FG* $\mathbf{K = k}$).** We optimize the three-dimensional coordinates of each constant i, $i = 1, \ldots, N_c$, in one iteration step by using the successive subspace correction method for optimization. We use the full grid method with maximal level k with the additive (1/2-FG$^+$) and multiplicative (1/2-FG*) transformation, respectively. Thereafter, we use the 2b-SSC.

- **One-and-two Multilevel Sparse-Grid Optimization (1/2-SG$^+$/1/2-SG* $\mathbf{K = k}$).** Same as (1/2-FG$^+$/1/2-FG* $\mathbf{K = k}$) with the sparse grid instead of the full grid method.

For the subspace correction steps, we use the same settings as in section 5.1, i.e. only global optimization steps using the MLSL algorithm with 200 function evaluation and a population of 50. Since it is a difficult task to determine the interval $[-b_c, b_c]$ out of which the constants are chosen, we test different values for b_c. The motivation is again that we assume to correct the current solution candidate in each iteration step, instead of computing an accurate result. This is also the reason why we optimize each constant once in each level, then go to the next level and repeat these steps until a stopping criterion is reached. This procedure has turned out to be the most successful.

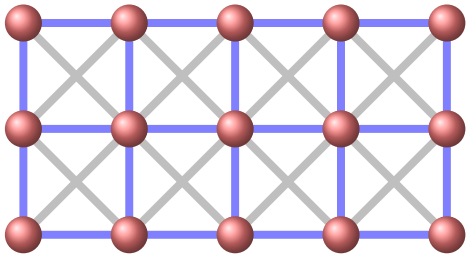


Figure 5.21: $N = 15$ atoms and the corresponding bonds in a uniform 3×5 grid.

5.2.1 Bond Potential

In order to test the multilevel coordinate transformation method, we use an example given by a system of atoms with a fixed bond structure as described in [Griebel et al., 2007] that can be chosen for elastic membranes.

We consider a uniform 3×5 grid of particles in which each atom only affects its direct neighbors provided that they are existent. This resulting bond structure is illustrated in figure 5.21.

We define its potential energy by the harmonic bond potential which we already mentioned in section 3.3. We distinguish between the potential for all diagonal bonds, given by

$$U_b^{diag}(r_{ij}) = \frac{1}{2}\, k_b(r_{ij} - \sqrt{2}\, r_0)^2$$

and the potential for all remaining bonds, described by

$$U_b(r_{ij}) = \frac{1}{2}\, k_b(r_{ij} - r_0)^2.$$

In two dimensions it is clear that the lowest energy value is achieved by the distance $\sqrt{2}r_0$ for all diagonal bonds and r_0 for all other ones, respectively. In three dimensions, we have to take into account that the system can overlap and in order to avoid this, we add the repulsive potential $\left(\frac{1}{r}\right)^{12}$ for all atom pairs. We obtain the following potential energy function

$$U(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \sum_{\substack{\text{bonds} \\ (i,j)}} U_b^{\text{tot}}(r_{ij}) + \sum_{\substack{i,j=1 \\ i<j}}^{N} \left(\frac{1}{r_{ij}}\right)^{12} \tag{5.1}$$

with the total bond term given by

$$U_b^{\text{tot}}(r_{ij}) = U_b(r_{ij}) + U_b^{diag}(r_{ij}).$$

In general, the parameter k_b has to be chosen differently for each bond, but for our purpose we use the fixed value $k_b = 300$. Furthermore, we set $r_0 = 2.2$.

Now, in order to test the applicability of the multilevel coordinate transformation, we use the aforementioned minimal configuration for the two-dimensional case with an additional perturbation as starting configuration. In the following, we denote the system of coordinates

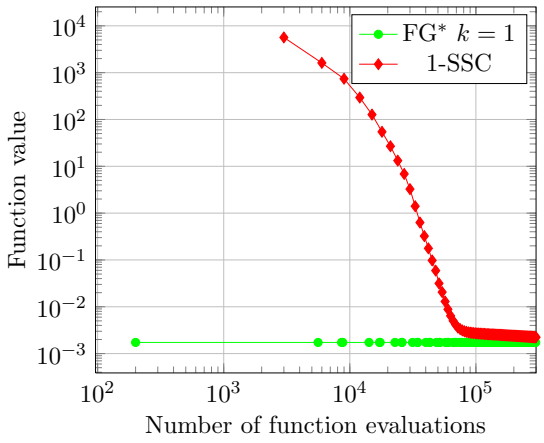


Figure 5.22: Development of the function value for $b_c = 1$.

of the grid of atoms including distances $\sqrt{2}r_0$ for all diagonal bonds and r_0 for all other ones by $\bar{\mathbf{x}} = (\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_{15})$.

At first we consider a perturbation characterized by a scaling, more precisely we multiply each coordinate of $\bar{\mathbf{x}}$ with the scaling factor 2. We obtain the starting configuration

$$\mathbf{x}^{(0)} = 2 \cdot \bar{\mathbf{x}}. \tag{5.2}$$

As a consequence, all distances in the grid of atoms are doubled.

Since this is a multiplicative perturbation, we use the multiplicative transformation for the optimization. Based on its construction, the transformation optimization based on the multiplicative approach can use the initial information about the optimal configuration more efficiently than the subspace correction optimization, see figure 5.22.

In the second example, we consider the starting configuration, obtained by adding a certain sinus-term to all coordinates, i.e.

$$\mathbf{x}^{(0)} = x_{ij} + \frac{1}{10}\sin(\mathbf{k}\mathbf{x}_i), \qquad \text{for } i = 1, \ldots, 15, \tag{5.3}$$

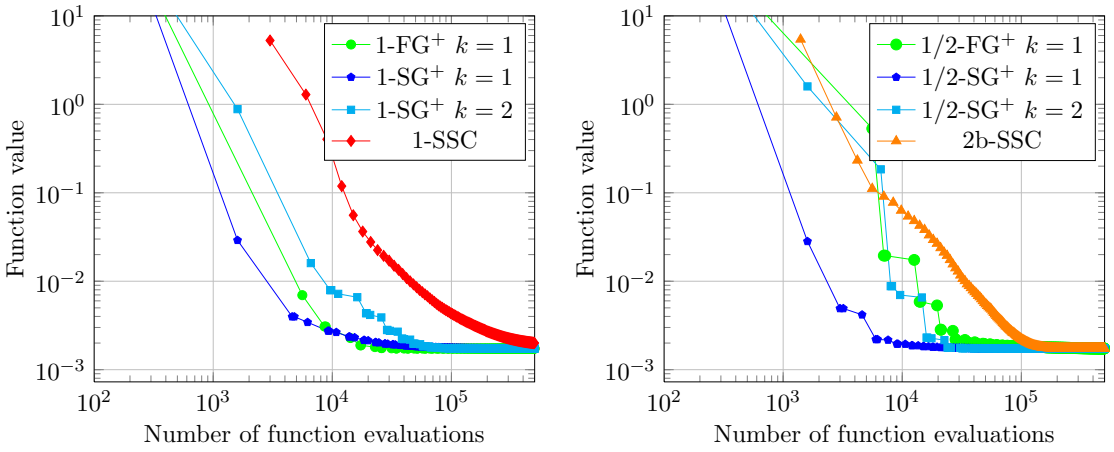
with $\mathbf{k} = (1, 0, 0)$ and $\mathbf{kx} = k_1x_1 + k_2x_2 + k_3x_3$.

We presume that the additive Fourier transformation leads to good results for this case. The development of the function value dependent on the number of function evaluations is shown for the derivative-free and gradient-based methods for different choices of b_c in figures 5.23-5.25. Again, we consider the function value dependent on the number of function evaluations. For comparison, we also run a subspace correction optimization and additionally plotted its respective best result. In each optimization, the multilevel method achieved better results and furthermore in most of all cases the sparse grid approach could further improve them.

Next, we consider an application from the field of materials science.

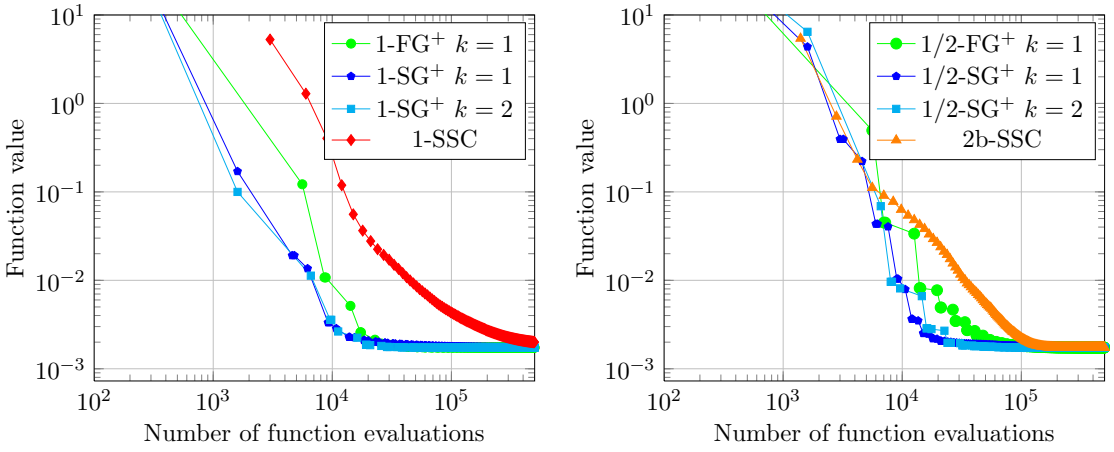
5.2.2 Lennard-Jones Potential with Periodic Boundary Conditions

In this section, we test the application of the presented global optimization methods in order to optimize the potential energy function of crystals. As already described in section 3.3 we



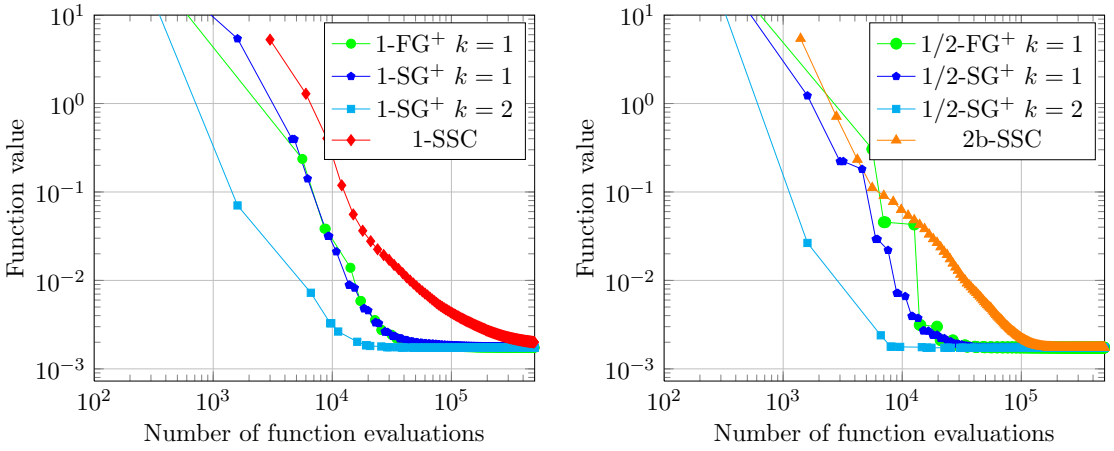
(a) Derivative-free. **(b)** Gradient-based.

Figure 5.23: Development of the function value for $b_c = 1$.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.24: Development of the function value for $b_c = 5$.



(a) Derivative-free. **(b)** Gradient-based.

Figure 5.25: Development of the function value for $b_c = 10$.

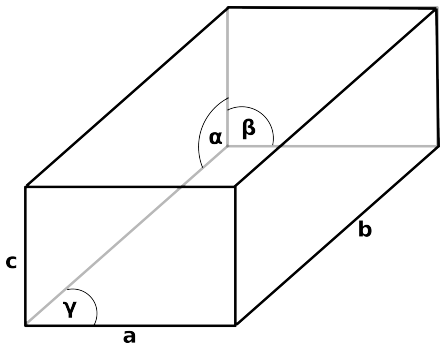


Figure 5.26: Simulation cell h.

have to include periodic boundary conditions. We consider the example of the Lennard-Jones potential with periodic boundary conditions. The resulting potential energy function is given by

$$U(\mathbf{x}_1, \ldots, \mathbf{x}_N, h) = \sum_{i=1}^{N} \left(\sum_{\substack{j=1 \\ i<j}}^{N} U^{LJ}(r_{ij}) + \sum_{j=1}^{N} \sum_{\mathbf{S} \in \mathcal{S}} U^{LJ}(r_{ij_\mathbf{S}}) \right),$$

with $\mathcal{S} := \mathbb{Z}^3 \backslash \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ and the transformation matrix h. Furthermore, we include the smoothing function $S(r)$ and set $r_l = 1.9\sigma$ and $r_c = 2.5\sigma$, see section 3.3.

The new aspect we have to take into account in the context of periodic boundary conditions is the explicit dependence of the potential energy function on the transformation matrix h. Thus, we have to optimize the shape of the cell as well as the atom coordinates.

We can express the transformation matrix $h = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ without loss of generality as a lower triangular matrix, that is

$$h = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

As a consequence, we have six variables more that have to be optimized. Since in this formulation it is difficult to choose the subspaces for the subspace correction optimization, we use the following equivalent notation

$$h = \begin{pmatrix} a & 0 & 0 \\ b\cos\gamma & b\sin\gamma & 0 \\ c\cos\beta & c\left(\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma}\right) & \frac{cW}{\sin\gamma} \end{pmatrix}$$

with

$$W = (1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\,\cos\beta\,\cos\gamma)^{1/2},$$

see [Gibson and Scheraga, 1995]. In this formulation, we can interpret a, b and c as the side lengths of the cell and α, β and γ as the angles of the resulting cell h, see figure 5.26. With this, we can use the subspace correction for optimizing the side lengths and angles separately.

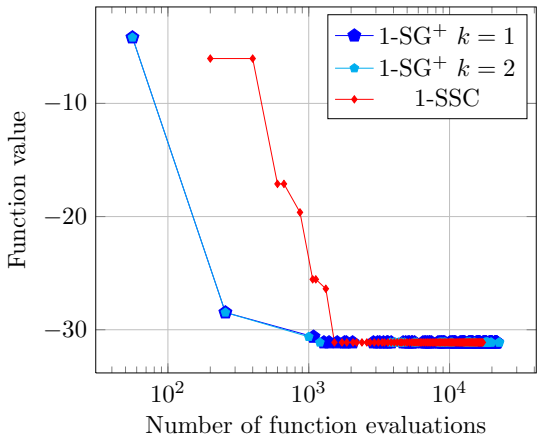


Figure 5.27: Development of the function value for $b_c = 1$.

In order to analyze the applicability of the multilevel method for the optimization of potential energy functions including periodic boundary conditions, we consider a perturbed configuration from one of the *Bravais lattices* which were described in section 3.3. This perturbation is expressed again by a sinus term, motivated by the lattice vibrations (see section 4.4.3). The starting configuration is defined by

$$\mathbf{x}_i^{start} = \mathbf{x}_i^{brl} + \frac{1}{10}\sin(\mathbf{k}\mathbf{x}_i^{obc}), \quad i = 1, \ldots, N$$

with $\mathbf{k} = (1, 0, 0)$ and $\mathbf{kx} = k_1x_1 + k_2x_2 + k_3x_3$.

In this example $\mathbf{x}^{obc}$ is a orthorhombic body centered starting configuration with $N = 2$ atoms forming the basis. We tested the methods 1-SG$^+$, $k = 1$ and 1-SG$^+$, $k = 2$ and for comparison the 1-SSC for the derivative-free case. The results are shown in figure 5.27.

Again, we can improve the efficiency of the optimization method by considering the multilevel coordinate transformation method. In the context of the optimization of potentials with periodic boundary conditions this is a very promising approach.

6 Summary and Future Prospects

Summary

In this thesis, we developed an efficient algorithm for the global optimization of potential energy functions by the use of subspace correction methods and a multilevel method based on coordinate transformations.

In the field of materials science potential energy functions are necessary for computer simulations in order to determine the properties of materials. Thereby, real experiments can be replaced and the great amount of work or the high expenses during their conduct can be avoided. Additionally, computer simulations allow the implementation of experiments which are impossible in reality.

Potential energy functions describe the interactions within a physical system dependent on the spatial coordinates of all corresponding atoms. They are based on the potential energy information obtained from approximated solutions of the high-dimensional Schrödinger equation. The most stable atom configuration, the ground state, is defined by the global minimum of the potential energy function. Within this state, all atoms are in a certain equilibrium between attractive and repulsive forces. The complex structure and the high dimension of potential energy functions yield a large number of local minima, as well as expensive function and derivative evaluations. These are the main reasons why their minimization is a very challenging task, however, the knowledge of the ground state is of great relevance for experimental and theoretical investigations in science.

High-dimensional models are omnipresent in many applications, but often it appears that the corresponding parameters affect each other more or less strongly. Since this also applies to potential energy functions, we presented a dimension-wise decomposition that represent the original function by a sum of its lower-order contributions. They can be interpreted as the interaction of the different parameters with each other. The consideration of a certain number of lower-dimensional terms can already give a good approximation for the underlying function. Therefore, we can reduce its complexity by considering one of the presented notions of dimension in the context of dimension-wise decompositions. Furthermore, we described the ANOVA decomposition and the high dimensional model reduction as special cases, which both can be applied to potential energy functions.

We described the development of approximations of potential energy based on multiple approximations of the high-dimensional Schrödinger equation and the dimension-wise decomposition. Moreover, some examples for potential energy functions and their specific applications were given.

We explained the challenges connected to non-linear, non-convex optimization problems and some known worst case error bounds for global optimization, based on different assumptions for the function space. These also included some results in the context of dimension-wise decomposed functions.

We presented a global optimization algorithm from the field of subspace correction. It can take advantage of the potential energy functions special construction, resulting from the

dimension-wise decomposition of its lower-order contributions. Preliminary, we described the application on linear systems by using a partition of the original problem into lower-dimensional subproblems, based on space decomposition. These subproblems are solved iteratively in order to get an approximated solution for the original problem, while ensuring that the solution does not get worse in each iteration step.

We explained the resulting iterative methods for the parallel and successive subspace correction method. On top of that, we motivated and formulated their application on non-linear optimization problems and the resulting algorithms.

Concerning the global minimization of potential energy functions, we chose the subspaces such that the different contributions of the dimension-wise decomposition are optimized separately in each iteration. Examples are given by the optimization of the three-dimensional coordinates of only one atom or the six-dimensional coordinates of one atom pair in one iteration step, while the remaining atoms positions are fixed. Since the potential functions of many-body systems include a large number of variables, the main advantage of the subspace correction approach is the enormous reduction of the dimension. In order to improve the efficiency in the case of optimizing two or more atoms in one iteration step, we additionally developed an improved order of running through the iterations, induced by the corresponding subspaces.

Since the subspace correction optimization method only affects subsets of atom coordinates, it is impossible to change all positions at once in a single iteration step. In order to enable a global approach, we developed a multilevel method based on a coordinate transformation.

The motivation for this transformation was given by lattice vibrations that describe the movements of crystal structured by a wave-like perturbation term. In order to transfer this to the problem of global optimization, we defined two different coordinate transformations, dependent on function expansion by the use of orthogonal basis functions. An example is given by the Fourier basis. Applied to some atom configuration, The transformations change all atom positions depending on the coefficients of the orthogonal basis expansion. Since such expansions are in general defined by an infinite sum, they have to be cut after a certain number of terms. We do not know initially how many terms have to be considered in order to get good results. Therefore, we developed a multilevel method in which each level represents a different number of terms in the orthogonal basis expansion.

Regarding to potential energy functions, the coefficients of the respective transformation were optimized instead of the atom coordinates themselves. The optimizations of the coefficients were performed via the subspace correction method and in order to further improve the results, we used a sparse grid approach which reduces the number of coefficients that have to be optimized.

The final multilevel algorithm was realized by a combination of optimizing the transformations coefficients on different levels and the subspace correction optimization of the coordinates themselves.

Based on these methods, we performed numerical experiments for the optimization of different potential energy functions.

First, we tested the applicability of the different subspace correction methods based on the direct optimization of a subset of atom coordinates. Therefore, we optimized the Lennard-Jones cluster which is described only by Lennard-Jones pair potentials starting from randomly chosen starting points. For this potential energy function we could show that the subspace

optimization algorithm can be improved by considering the optimization of subsets of two and three atoms instead of single atoms in each iteration step. In addition, we could further improve its efficiency by the use of the improved order of atoms.

Thus, we can conclude that the subspace optimization algorithm can take advantage of the special construction of a function based on the dimension-wise decomposition for this type of potential energy functions.

In order to test this type of algorithm for other potentials, we considered the Gupta cluster for zinc. The underlying potential function is more complex structured since it includes a many-body contribution additional to a pair-potential. Indeed, the optimization turned out to be more difficult for Gupta clusters than for Lennard-Jones clusters. A clear statement which type of method results in the best global minima was not possible. Nevertheless, we could find the best known global minima for this type of cluster by applying the subspace correction optimization for more than hundred atoms.

The multilevel method based on a coordinate transformation was tested by the application of a potential function that includes a harmonic bond potential for a fixed set of bonded atoms and a repulsive potential. Considering a good candidate for the configuration that gives the lowest energy value including a perturbation by a scaling factor or a sinus term, the multilevel approach always achieved better results than the subspace correction optimization. In addition, the application of the sparse grid approach could further improve its efficiency.

This could also be shown for the optimization of potential energy functions with periodic boundary conditions which are used for the description of crystals.

Consequently, we have developed new efficient methods in order to determine the global minimum of a potential energy function, which can take advantage of the special construction of these functions, based on the dimension-wise decomposition of its lower-order contributions.

Future Prospects

In this thesis, we developed an efficient algorithm for the global optimization of potential energy functions by the use of subspace correction methods and a multilevel method based on a coordinate transformation.

The fact that we can take advantage of the potential functions special construction based on the dimension-wise decomposition could be shown by the example of pair potentials. Since the Gupta cluster optimization turned out to be much more difficult, an interesting aspect for further work is the improvement of the subspace correction optimization procedure by modifying the optimization algorithm or the convergence criteria. Another approach in this context is the inclusion of more stochastic elements. It is motivated by the great success of algorithms allowing intermediate steps in which the current solution candidate gets worse up to a given value. The presented subspace correction algorithm thus can be modified in the following sense. If the current value does not change a specific number of times, we can randomly change a subset or all atoms coordinates in order to get a new starting point which leads to better results.

Furthermore, the consideration of other potentials and the corresponding choice of subspaces or transformations including orthogonal basis function are interesting approaches for following works. Especially, the application to more complicated potential functions in order to describe

crystals is of great relevance.

In the multilevel method and the underlying coordinate transformation, we can choose the basis functions differently. An example is given by the Chebyshev polynomials.

Finally, we would like to point out that the procedure based on subspace corrections can be transferred to other applications which can be described by dimension-wise decompositions. Thus, it is a general optimization approach not only for potential energy functions, but capable of solving problems of this kind given by multiple applications in the field of global optimization problems.

Bibliography

[Addis et al., 2004] Addis, B., Mosca, E., and Schoen, F. (2004). Global optimization using local searches. *Unpublished doctoral dissertation, Global Optimization Laboratory, University of Florence.*

[Blake and Zisserman, 1987] Blake, A. and Zisserman, A. (1987). *Visual reconstruction*, volume 2. MIT press Cambridge.

[Chan and Sharapov, 2002] Chan, T. F. and Sharapov, I. (2002). Subspace correction multilevel methods for elliptic eigenvalue problems. *Numerical linear algebra with applications*, 9(1):1–20.

[Chatterjee, 2008] Chatterjee, S. K. (2008). *Crystallography and the World of Symmetry*, volume 113. Springer.

[Deaven and Ho, 1995] Deaven, D. and Ho, K. (1995). Molecular geometry optimization with a genetic algorithm. *Physical Review Letters*, 75(2):288.

[Doye, 2003] Doye, J. P. (2003). Identifying structural patterns in disordered metal clusters. *Physical Review B*, 68(19):195418.

[Doye and Wales, 1997] Doye, J. P. and Wales, D. J. (1997). Structural consequences of the range of the interatomic potential a menagerie of clusters. *Journal of the Chemical Society, Faraday Transactions*, 93(24):4233–4243.

[F. Y. Kuo and Wozniakowski, 2010] F. Y. Kuo, I. H. Sloan, G. W. W. and Wozniakowski, H. (2010). On decompositions of multivariate functions. *Math. Comput.*, 79(270):953–966.

[Frommer and Renaut, 1999] Frommer, A. and Renaut, R. A. (1999). A unified approach to parallel space decomposition methods. *Journal of computational and applied mathematics*, 110(1):205–223.

[Gerstner and Griebel, 2010] Gerstner, T. and Griebel, M. (2010). Sparse grids. *Encyclopedia of Quantitative Finance.*

[Gibson and Scheraga, 1995] Gibson, K. and Scheraga, H. (1995). Crystal packing without symmetry constraints. 2. possible crystal packings of benzene obtained by energy minimization from multiple starts. *The Journal of Physical Chemistry*, 99(11):3765–3773.

[Gottlieb et al., 1977] Gottlieb, D., Orszag, S. A., and MA, C. H. I. (1977). *Numerical analysis of spectral methods*, volume 2. SIAM.

[Griebel and Hamaekers, 2005] Griebel, M. and Hamaekers, J. (2005). Molecular dynamics simulations of the mechanical properties of polyethylene-carbon nanotube composites. *Handbook of Theoretical and Computational Nanotechnology*, 9:409–454.

[Griebel et al., 2008] Griebel, M., Hamaekers, J., and Heber, F. (2008). BOSSANOVA: A bond order dissection approach for efficient electronic structure calculations. INS Preprint 0704, Institut für Numerische Simulation, Universität Bonn.

[Griebel et al., 2007] Griebel, M., Knapek, S., and Zumbusch, G. W. (2007). *Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications*, volume 5. Springer.

[Hayes et al., 2006] Hayes, M. Y., Li, B., and Rabitz, H. (2006). Estimation of molecular properties by high-dimensional model representation. *The Journal of Physical Chemistry A*, 110(1):264–272.

[Hoare, 1979] Hoare, M. (1979). Structure and dynamics of simple microclusters. *Adv. Chem. Phys*, 40:49–135.

[Holtz, 2008] Holtz, M. (2008). *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*. Dissertation, Institut für Numerische Simulation, Universität Bonn.

[Johnson, 2013] Johnson, S. G. (2013). *The NLopt nonlinear-optimization package*. `http://ab-initio.mit.edu/nlopt`.

[Kan and Timmer, 1987a] Kan, A. R. and Timmer, G. (1987a). Stochastic global optimization methods part i: Clustering methods. *Mathematical programming*, 39(1):27–56.

[Kan and Timmer, 1987b] Kan, A. R. and Timmer, G. (1987b). Stochastic global optimization methods part ii: Multi level methods. *Mathematical Programming*, 39(1):57–78.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Vecchi, M., et al. (1983). Optimization by simmulated annealing. *science*, 220(4598):671–680.

[Königsberger, 2003] Königsberger, K. (2003). *Analysis 1*, volume 1. Springer DE.

[Kostrowicki et al., 1991] Kostrowicki, J., Piela, L., Cherayil, B. J., and Scheraga, H. A. (1991). Performance of the diffusion equation method in searches for optimum structures of clusters of lennard-jones atoms. *The Journal of Physical Chemistry*, 95(10):4113–4119.

[Møller and Plesset, 1934] Møller, C. and Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618.

[Moloi and Ali, 2005] Moloi, N. and Ali, M. (2005). An iterative global optimization algorithm for potential energy minimization. *Computational Optimization and Applications*, 30(2):119–132.

[Novak, 1988] Novak, E. (1988). *Deterministic and stochastic error bounds in numerical analysis*, volume 1349. Springer-Verlag Berlin.

[Novak and Ritter, 1996] Novak, E. and Ritter, K. (1996). Global optimization using hyperbolic cross points. In *State of the Art in global Optimization*, pages 19–33. Springer.

[Novak and Woźniakowski, 2010] Novak, E. and Woźniakowski, H. (2010). *Tractability of Multivariate Problems: Standard information for functionals*, volume 12. European Mathematical Society.

[Oettershagen, 2011] Oettershagen, J. (2011). Reduktion der effektiven Dimension und ihre Anwendung auf hochdimensionale Probleme. Diplomarbeit, Institut für Numerische Simulation, Universität Bonn.

[Parr and Yang, 1989] Parr, R. G. and Yang, W. (1989). *Density-functional theory of atoms and molecules*, volume 16. Oxford university press.

[Richtárik and Takáč, 2012] Richtárik, P. and Takáč, M. (2012). Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, pages 1–38.

[Sharapov, 1997] Sharapov, I. A. (1997). *Multilevel subspace correction for large-scale optimization problems*. PhD thesis, Citeseer.

[Srivastava, 2011] Srivastava, J. (2011). *Elements of solid state physics*. PHI Learning Pvt. Ltd.

[Szabo and Ostlund, 2012] Szabo, A. and Ostlund, N. S. (2012). *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Dover Publications.

[Tai and Xu, 2002] Tai, X.-C. and Xu, J. (2002). Global and uniform convergence of subspace correction methods for some convex optimization problems. *Mathematics of Computation*, 71(237):105–124.

[Wales and Doye, 1997] Wales, D. J. and Doye, J. P. (1997). Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116.

[Wales et al., 2014] Wales, D. J., Doye, J. P. K., Dullweber, A., Hodges, M. P., Naumkin, F. Y., Calvo, F., Hernández-Rojas, J., and Middleton, T. F. (2014). *The Cambridge Cluster Database*. `http://www-wales.ch.cam.ac.uk/CCD.html`.

[Wang, 2008] Wang, X. (2008). On the approximation error in high dimensional model representation. In *Proceedings of the 40th Conference on Winter Simulation*, pages 453–462. Winter Simulation Conference.

[Weise, 2009] Weise, T. (2009). *Global Optimization Algorithms - Theory and Application*. Self-Published, second edition. Online available at http://www.it-weise.de/.

[Xu, 1992] Xu, J. (1992). Iterative methods by space decomposition and subspace correction. *SIAM review*, 34(4):581–613.

[Xu, 2001] Xu, J. (2001). The method of subspace corrections. *Journal of Computational and Applied Mathematics*, 128(1):335–362.

[Xue, 1994] Xue, G. (1994). Molecular conformation on the cm-5 by parallel two-level simulated annealing. *Journal of Global Optimization*, 4(2):187–208.

[Zhang et al., 2011] Zhang, Z., Choi, M., and Karniadakis, G. E. (2011). Anchor points matter in anova decomposition. In *Spectral and High Order Methods for Partial Differential Equations*, pages 347–355. Springer.