CHAPTER 1

Introduction

A true challenge in simulating structural evolution in materials is the long time and large length scales that one needs to capture while accurately including important (atomic) details. As an example, polymers in melts, blends and solutions include structural features that span the whole range from nanometers, microns, millimeters to even larger sizes, which, depending on the property/phenomena of interest, may be important for particular material properties. The corresponding time scales of the dynamic process relevant for these different material properties vary from femtoseconds to milliseconds or even seconds or hours. This wide range of length and time scales is responsible for the variability of physical properties, making (rational) design of structures for intricate applications possible. Unfortunately there is no model or simulation algorithm that spans the whole range of length and time scales. For this reason, multiscale modeling - bridging and linking different computational methods that each provides a realistic description on well-defined time and length scales - is gaining importance for predicting and designing structure resulting in specific material properties.

The leading thread in this thesis is a stochastic Quasi-Newton (S-QN) model that contains both mathematical (sampling) and physical (diffusion) elements. In particular, we start from a conventional Langevin equation for particles diffusion, that uses a separation in time and length scales of different components of the system, for instance a polymer in solvent, to represent the fastest motions of the system (solvent) by a random fluctuation force. By averaging over the fastest motions, only the slower modes of the system are explicitly represented and the functional collective modes would hypothetically come within simulation reach. However, the resulting description in terms of coarse grained variables can also be categorized in terms of slow or fast modes, but now on a coarse-grained scale. Again, the fastest modes restrict the time step for stable time integration, resulting in substantial computational effort for capturing slow, collective motions on this scale. Since the sampling performance is always dictated by the smallest time or length scale in system descriptions, we seek to improve the simulation performance by automatically updating all modes roughly equally fast. Our method, described in this thesis, automatically accelerates the slow modes of the system. Since fast modes are simultaneously slowed down, we effectively integrate the Langevin equation with a much larger time step. The mode-scaling is caused by the proposed mobility and gives rise to much more efficient sampling of the energy landscape.

In Chapter 2 the new method is introduced: an alternative to the standard Langevin equations for Brownian dynamics. The focus is on introducing the new stochastic quasi Newton technique, which could both be seen as a quasi-Newton method with noise and Langevin equations with alternative mobility. We compare its performance to that of the standard technique and provide a detailed analysis of its thermodynamic and computational properties. We conclude that, even for lower dimensional systems, the new method is an order of magnitude more efficient in crossing energy barriers than the standard approach.

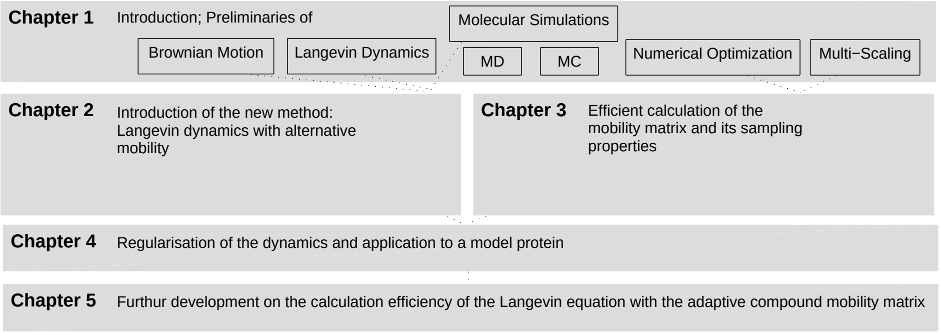
In Chapter 3, a new update scheme is introduced for updating the mobility matrix. The bottle-neck of our proposed method in Chapter 2 is the calculation of the mobility and its factorized form. Especially for the latter, an efficient update scheme in factorized form is needed for high-dimensional systems. Without such extension, practical application to molecular systems will be limited; with the extension, a new field is opened. Here, we focus on the details of such an algorithm and a detailed analysis of the performance and, in addition, discuss general guidelines for the choice of the free parameters in our methodology. We have chosen to focus on the physical context, i.e. critical slowing down phenomena and the automated separation of time scales in molecular simulation (multiplicity of timescales), and the conceptual relation to existing general methodology (Fourier acceleration techniques). Our results here are for a rather "academic" high-dimensional system but nevertheless relevant within this context. In particular, the system contains the critical slowing down features, represents a "stepping stone" for the simulation of more realistic systems and is very well suited for a quantitatively analysis of performance. The results are decisive in that the method is indeed capable of incorporating a multiplicity of time steps in an automated way.

Chapter 4 deals with an application of the new method to a minimal model of a protein. In order to avoid numerical problems, we consider the problem of ill-conditioning. This problem is solved by refining the update scheme for the mobility such that regularization occurs in the ill-conditioned cases. We analyze the effect of regularization and apply the new scheme to the minimal model of the chosen protein in a stepwise manner, by including an increasing number of terms of the total coarse-grained energy expression. A detailed analysis of the collective dynamics and sampling behavior for the considered coarse-grained protein is provided.

Although the proposed factorized update (FSU) in Chapter 3 makes the compute expensive factorization (Cholesky decomposition) redundant, it is useful to develop a limited memory implementation, much in the spirit of existing limited-memory optimization methods like L-BFGS. The proposed L-FSU method is not only more efficient in term of memory storage, since no full matrices need to be stored, but also computationally. Although the limited memory update for the new mobility was already introduced in Chapter 3, the incorporation within our S-QN method (especially the spurious drift term) is not straightforward. Chapter 5 provides detail how to integrate L-FSU and S-QN.

Although the proposed factorized update (FSU) in Chapter 3 makes the compute expensive factorization (Cholesky decomposition) redundant, it is useful to develop a limited memory implementation, much in the spirit of existing limited-memory optimization methods like L-BFGS. The proposed L-FSU method is not only more efficient in term of memory storage, since no full matrices need to be stored, but also computationally. Although the limited memory update for the new mobility was already introduced in Chapter 3, the incorporation within our S-QN method (especially the spurious drift term) is not straightforward. Chapter 5 provides detail how to integrate L-FSU and S-QN. Since our method is based on the Langevin equation, this introduction starts with the history of Brownian motion with its forthcoming Langevin equation. The Langevin equation consist of a deterministic part and stochastic/noise part, where the deterministic part only (NewtonŠs second law) gives the common approach in performing molecular simulations: molecular dynamics (MD). In addition, a purely stochastic approach is also available for performing molecular simulations: Monte Carlo (MC). Both simulation methods are shortly covered here. The new proposed mobility in the Langevin equation is a space dependent mobility. Since it is a space dependent mobility, it needs to be updated at every time step. The update scheme for the mobility is very much related to the update scheme in mathematical minimization for the quasi newton method. This method is briefly described in this chapter of preliminaries.

Last but not least in this chapter, an example of a system with a slow and fast mode is given and shows illustratively that integration with larger time steps for slow modes causes faster decrease towards (energy) minimum.


Chapter 1
Introduction; Preliminaries of
Molecular Simulations
Brownian Motion
Langevin Dynamics
MD
MC
Numerical Optimization
Multi−Scaling
Chapter 2
Introduction of the new method: Langevin dynamics with alternative mobility
Chapter 3
Efficient calculation of the mobility matrix and its sampling properties
Chapter 4
Regularisation of the dynamics and application to a model protein
Chapter 5
Furthur development on the calculation efficiency of the Langevin equation with the adaptive compound mobility matrix

.

Figure 1.1: *Schematic drawing outline thesis*

1.1. Brownian Motion

Brownian motion is named after the Scottish botanist Robert Brown (1773 - 1858), who observed pollen grains with ’rapid oscillatory motion’ in water. However, it was not Brown who discovered this type of motion. In [1], the author cleverly point out that anyone looking at water through a microscope is able to see small things moving around. Brown himself also mentions precursors in his papers [2, 3]. Brown’s original contribution was that the motion is present in organic as well as inorganic matter (by that time one thought that the particles were alive), and that the motions have their origin in the particles themselves and not in the surrounding fluid. Unfortunately he couldn't work out what was causing the motion.

In 1905 Albert Einstein (1879- 1955) explained, in addition to his work on the special theory of relativity and the photoelectric effect, the Brownian motion. Einstein predicted that the random motions of molecules in a liquid impacting on larger suspended particles would result in irregular, random motions of the particles: the Brownian motion. He published his theory on Brownian motion in 1905 in a paper titled: 'On the movement of small particles in stationary liquids required by the molecular-kinetic theory of heat' [4]. This work was extended in his doctoral dissertation − also published in 1905 − that shows how to calculate Avogadro's number and hence the sizes of molecules [5].

Einstein derived the diffusion equation

$$\frac{\partial p}{\partial t} = D\nabla^2 p, \tag{1.1}$$

where $p = p(x,t)$ is the probability that a brownian particle is at x at time t and D is the so called diffusion constant. The solution (with initial condition $p(x,0) = \delta(x)$, the delta distribution) is given as

$$p(x,t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(-\frac{|x|^2}{4Dt}\right), \tag{1.2}$$

where $|\cdot|$ is the Euclidean norm for the multidimensional case . By considering a system of brownian particles in equilibrium between osmotic pressure and friction force which obeys Stokes' law, Einstein derived an expression for the diffusion coefficient. The expression is given in terms of temperature T, viscosity η and radius a of the Brownian particles

$$D = \frac{kT}{6\pi\eta a}, \tag{1.3}$$

where $k = k_B$ is the Boltzmann constant. By observations and equation(1.2) one can determine D and from (1.3) the Boltzmann constant is obtained.

Einstein argued that the velocity of a brownian particle can not be measured due to the rapid change of direction and magnitude of the particles in such a short time scale. He introduced the mean squared displacement $\overline{\Delta_x^2}$ of the displacement Δ_x in the x-direction in a certain time interval τ, of the Brownian particle as the observable quantity in Brownian motion. From the solution of the diffusion equation, the mean squared displacement could derived and is given by

$$\overline{\Delta_x^2} = \frac{RT}{N_{\text{av}}} \frac{1}{3\pi\eta a}\tau, \tag{1.4}$$

where R is the gas constant and N_{av} is the number of Avogadro.
In 1906, Marian Smoluchowski (1972 - 1917) derived independently from Einstein the same form for the mean squared displacement of a Brownian particle [6]. Smoluchowski's used a dynamical approach for describing the Brownian motion and obtained the mean squared displacement which differs from Einsteins by a factor $\frac{64}{27}$.
Paul Langevin (1872-1946), was well aware of the findings of Einstein and Smoluchowski regarding Brownian motion. Also using a dynamical approach Langevin was able to derive exact the same expression as Einstein for the mean squared displacement [7]. Langevin described the motion of free particles of mass m in direction x with two force terms: an according to Stokes' formula a viscous resistance term and an independent fluctuation force term X

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t} = -6\pi\eta a\frac{\mathrm{d}x}{\mathrm{d}t} + X. \tag{1.5}$$

Multiplying equation(1.5) by x Langevin obtained

$$\frac{m}{2}\frac{\mathrm{d}^2x^2}{\mathrm{d}t} - mv^2 = -3\pi\eta a\frac{\mathrm{d}x^2}{\mathrm{d}t} + Xx. \tag{1.6}$$

Considering a large number of identical particles of mass m, making use of $m\overline{v^2} = RT/N_{\mathrm{av}}$ and averaging over equation(1.6) gives

$$\frac{m}{2}\frac{\mathrm{dz}}{\mathrm{d}t} + 3\pi\eta az = \frac{RT}{N_{av}}, \tag{1.7}$$

where $z = \frac{\overline{\mathrm{d}x^2}}{\mathrm{d}t}$. Notice that the average of Xx is zero due to the irregularity of X. Langevin found a solution for z which is in the regime of a constant rate of agitation given as

$$\frac{\overline{\mathrm{d}x^2}}{\mathrm{d}t} = \frac{RT}{N_{av}}\frac{1}{3\pi\mu a}. \tag{1.8}$$

In a time interval τ holds

$$\overline{x^2} - \overline{x_0^2} = \frac{RT}{N_{av}}\frac{1}{3\pi\mu a}\tau, \tag{1.9}$$

which is exact the same as Einsteins equation(1.4).
The validity of the expression for the displacement has been verified experimentally by Jean Perrin (1870-1942) and in 1908 he indeed confirmed that Einsteins expression for the displacement was exact [8].

Langevin's equation, a differential equation with a random force term X, was the first stochastic differential equation. Stochastic differential equations are widely used these days. For example in financial modeling (the famous Black-Scholes formula), neuroscience and off course as stochastic dynamics in molecular physics.
Throughout this thesis the following form of the Langevin equation will be considered

$$\boxed{\mathrm{d}\mathbf{x} = [-M(\mathbf{x})\nabla\Phi(\mathbf{x}) + k_B T\nabla \cdot M(\mathbf{x})]\,\mathrm{d}t + \sqrt{2k_B T}L(\mathbf{x})\mathrm{d}W(t)}\,. \tag{1.10}$$

Below we show how this form of the Langevin equation is obtained and also discuss some basic properties of the stochastic differential equation.

1.2. Classical Langevin Equation and Stochastic Differential Equation

The most simple Langevin equation is the description of the Brownian motion

$$m\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\gamma v(t) + X(t), \tag{1.11}$$

where γ/m is the friction constant. Equation(1.11) can be written in a more general form of a stochastic differential equation (SDE)

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(x,t) + b(x,t)\zeta(t), \tag{1.12}$$

where x is the variable of interest, $a(x,t)$ and $b(x,t)$ are known functions and the fluctuation force $\zeta(t)$ is assumed to be a gaussian process with idealized mathematical description, i.e. for $t \neq t'$

$$\langle\zeta(t)\zeta(t')\rangle = \delta(t-t') \quad \text{and} \quad \langle\zeta(t)\rangle = 0. \tag{1.13}$$

Due to the idealization, the variable of interest x becomes a Markov process and therefore its transition probability is described by the Fokker-Planck equation (FPE)

$$\partial_t p(x,t|x_0,t_0) = -\partial_x[a(x,t)p(x,t|x_0,t_0)] + \frac{1}{2}\partial_x^2[b(x,t)^2 p(x,t|x_0,t_0)]. \tag{1.14}$$

By assuming the Brownian motion taken place in a medium in thermal equilibrium, we have

$$p_s(x) = \mathcal{N}\exp\left(-\frac{1}{2}\frac{mx^2}{kT}\right), \tag{1.15}$$

since the variable of interest x corresponds with the velocity v and therefore the stationary solution of the FPE must coincide with the Maxwell - Boltzmann distribution. Making use of $a(x,t) = -\frac{\gamma}{m}v = a$, one can obain $b(x,t) = b$ from

$$\frac{2a}{b^2} = p_s \partial_x p_s. \tag{1.16}$$

$$\Longrightarrow$$

$$p_s = C \exp\left(\int \frac{2a}{b^2}\,\mathrm{d}x\right). \tag{1.17}$$

Calculating the integral

$$\int \frac{2a}{b^2}\,\mathrm{d}x = -\int \frac{2\gamma v}{m}\frac{1}{b^2}\,\mathrm{d}v = -\frac{\gamma v^2}{mb^2}, \tag{1.18}$$

and setting it equal to $-\frac{1}{2}\frac{mv^2}{kT}$ yields

$$b^2 = \frac{2\gamma kT}{m^2}, \tag{1.19}$$

which is the relation between the systematic (frictional) force $a(v,t) = -\frac{\gamma}{m}v(t)$ and the random force $b(v,t)\zeta(t) = \frac{X(t)}{m}$: the so called *fluctuation dissipation theorem,*

$$\frac{2\gamma kT}{m^2}\langle\zeta(t)\zeta(t')\rangle = \frac{1}{m^2}\langle X(t)X(t')\rangle \tag{1.20}$$

$$\Longrightarrow$$

$$\langle X(t)X(t')\rangle = 2\gamma kT\delta(t-t'). \tag{1.21}$$

Similar derivation and results hold for Brownian motion applied in a potential field Φ, where the Langevin equation becomes

$$m\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\gamma v(t) - \partial_x\Phi(x) + X(t), \tag{1.22}$$

and where the considered stationary solution for the transition probability equals

$$p_s(x,v) = \mathcal{N}\exp\left(-\frac{1}{kT}\{\tfrac{1}{2}mv^2 + \Phi(x)\}\right). \tag{1.23}$$

In this thesis we will consider particles in over-damped limit. Which means that the Brownian behavior of particles interacting through a potential is in the high friction limit. Under this assumption one finds the positional Langevin equation ($\frac{\mathrm{d}v}{\mathrm{d}t} = 0$)

$$\gamma v(t) = -\partial_x\Phi(x) + X(t). \tag{1.24}$$

The corresponding integral equation of (1.12) is given as

$$x(t) - x(0) = \int_0^t a[x(s), s]\,\mathrm{d}s + \int_0^t b[x(s), s]\zeta(s)\,\mathrm{d}s. \tag{1.25}$$

Assuming $\int_0^t \zeta(s)\,\mathrm{d}s = W(t)$, one can write

$$\int_0^t b[x(s), s]\zeta(s)\,\mathrm{d}s = \int_0^t b[x(s), s]\,\mathrm{d}W(s), \tag{1.26}$$

which is a stochastic integral with respect to a sample function $W(s)$. More details on stochastic integrals can be found in [9]. Here, we only consider the integrated form of the positional Langevin equation

$$\mathrm{d}x = -\gamma^{-1}\partial_x\Phi(x)\,\mathrm{d}t + \sqrt{2\gamma^{-1}kT}\,\mathrm{d}W(t). \tag{1.27}$$

From the assumption $\int_0^t \zeta(s)\,\mathrm{d}s = W(t)$, one finds that $W(t)$ is a markov process with its corresponding FPE given by (1.14), with $a(x,t) = 0$ and $b(x,t) = 1$. By solving the Fokker Planck equation, with the delta distribution as its initial condition, $W(t)$ is determined as a Gaussian process with $\langle W(t)\rangle = W(t_0) = w_0$ and $\langle [W(t) - w_0]^2\rangle = t - t_0$, which will be called the Wiener process. More precisely the corresponding FPE is of the same form as Einsteins diffusion equation (1.1), therefore this one dimensional Wiener process is also often called Brownian motion. Notice that due to the irregularity of the sample paths (brownian motion), $W(t)$ is not differentiable which contradicts that the integral of $\zeta(t)$ is $W(t)$. Nevertheless, equation (1.27) can be used and interpreted consistently.

The positional Langevin equation in the multi-dimensional form with space dependent $\gamma = \gamma(\mathbf{x})$ is derived in the same way and given as

$$\mathrm{d}\mathbf{x} = [-M(\mathbf{x})\nabla\Phi(\mathbf{x}) + k_BT\nabla\cdot M(\mathbf{x})]\,\mathrm{d}t + \sqrt{2k_BT}L(\mathbf{x})\mathrm{d}W(t), \tag{1.28}$$

where $M(\mathbf{x}) = \gamma(\mathbf{x})^{-1}$ is the mobility, $M(\mathbf{x}) = L(\mathbf{x})L^T(\mathbf{x})$ and $W(t)$ becomes the multivariate Wiener process with $\langle \mathrm{d}W_i(t)\mathrm{d}W_j(t)\rangle = \delta_{ij}\mathrm{d}t$. More details of the derivation is found in appendix A.

1.3. Molecular simulation preliminaries

There are two main approaches for performing molecular simulations: a deterministic approach and a stochastic approach. The deterministic approach, called Molecular

Dynamics (MD), simulates the time evolution of the molecular system and gives the actual trajectory of the system. The stochastic approach, Monte Carlo (MC), explores the energy surface by randomly accessing different configurations of the molecular system. The different configurations allows the calculation of the system's thermodynamic properties. In the remainder of this section the most common molecular dynamic and Monte Carlo methods are given.

1.3.1. Molecular dynamics

In MD the classical equations of motion are solved in time for a set of atoms. Newton's equation of motion can be discretized by Taylor expansion

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\Delta t^2}{2m_i}\mathbf{f}_i(t) + \frac{\Delta t^3}{3!}\dddot{\mathbf{r}}_i(t) + \mathcal{O}(\Delta t^4), \tag{1.29}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \mathbf{f}_i(t)\frac{\Delta t}{m_i} + \frac{\Delta t^2}{2}\ddot{\mathbf{v}}_i(t) + \frac{\Delta t^3}{3!}\dddot{\mathbf{v}}_i(t) + \mathcal{O}(\Delta t^4), \tag{1.30}$$

where $\mathbf{r}_i$ and $\mathbf{v}_i$ is the position and velocity respectively of particle i and $\mathbf{f}_i$ the force on particle i. The most simple scheme is the Euler scheme which calculates the positions using the first three terms of expansion (1.29) ($\mathcal{O}(\Delta t^2)$) and calculates the velocities using the first two terms of expansion (1.30) ($\mathcal{O}(\Delta t)$). Not only is the Euler algorithm an unstable scheme, it is neither time-reversible nor phase-space preserving, which are important properties in physical systems.

1.3.1a. Verlet

It is possible to solve Newton's equation of motion without explicitly calculating the velocity. Consider the Taylor expansion at $t - \Delta t$

$$\mathbf{r}_i(t-\Delta t) = \mathbf{r}_i(t) - \mathbf{v}_i(t)\Delta t + \frac{\Delta t^2}{2m_i}\mathbf{f}_i(t) - \frac{\Delta t^3}{3!}\dddot{\mathbf{r}}_i(t) + \mathcal{O}(\Delta t^4). \tag{1.31}$$

The Verlet scheme [10, 11] for updating the positions is obtained by adding equations (1.29) and (1.31), which gives after simple algebraic operations

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\Delta t^2}{m_i}\mathbf{f}_i(t) + \mathcal{O}(\Delta t^4), \tag{1.32}$$

and analogously by substracting the velocity is obtained

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2). \tag{1.33}$$

Although the velocity is not necessary needed for determining the positions in the Verlet scheme, they are often needed to calculate physical quantities like the kinetic energy. However the update scheme for the velocity in the basic Verlet algorithm needs the positions at $t + \Delta t$ to calculate the velocity at time t. This deficiency can be solved by using the Leap Frog scheme, where the velocities are updated at half time steps and 'leap' ahead the positions [12].

1.3.1b. Velocity Verlet

An better implementation of the same basic algorithm is the so-called velocity Verlet scheme, where positions, velocities and accelerations at time $t+\Delta t$ are obtained from the same quantities at time t in the following way:

$$\begin{aligned} \mathbf{r}_i(t+\Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\Delta t^2}{2m_i}\mathbf{f}_i(t) && (1.34) \\ \mathbf{v}_i(t+\Delta t) &= \mathbf{v}_i(t+\Delta t/2) + \frac{\Delta t}{2m_i}\mathbf{f}_i(t+\Delta t), && (1.35) \end{aligned}$$

where $\mathbf{v}_i(t + \Delta t/2)$ is given by

$$\mathbf{v}_i(t+\Delta t/2) = \mathbf{v}_i(t) + \frac{\Delta t}{2m_i}\mathbf{f}_i(t). \tag{1.36}$$

The velocity Verlet scheme is very stable and is a widely used MD algorithm [13].

1.3.2. Monte Carlo methods

The Monte Carlo method is the most popular method used to generate a certain assemble. Assume that we want to calculate the integral

$$I = \int_a^b f(x)\,\mathrm{d}x, \tag{1.37}$$

which can be rewritten as

$$I = (b-a)\langle f(x)\rangle \approx (b-a)\frac{1}{L}\sum_{i=1}^{L} f(x_i), \tag{1.38}$$

where L is the number of random (hence, the name Monte Carlo) points x_i taken between (a, b). The variance of $\langle f(x)\rangle$ equals

$$\sigma^2 = \frac{1}{L}[\langle f(x)^2\rangle - \langle f(x)\rangle^2]. \tag{1.39}$$

Thus for sufficient large L the discrete approximation of the integral is quite accurate. Clearly, the method works best for constant f. Consider $a = 0$ and $b = 1$ and introduce the weight function $w(x)$

$$I = \int_0^1 w(x)\frac{f(x)}{w(x)}\,\mathrm{d}x = \int_0^1 \frac{f(x(y))}{w(x(y))}\,\mathrm{d}y, \tag{1.40}$$

where $w(x)$ satisfies

$$\int w(x)\,\mathrm{d}x = 1 \quad y(x) = \int_0^x w(x')\,\mathrm{d}x',$$

such that $\mathrm{d}y = w(x)\mathrm{d}x$, $y(0) = 0$ and $y(1) = 1$ hold. For making use of Monte Carlo sampling for $\frac{f(x(y))}{w(x(y))}$ with random sampling in y from the distribution $w(x)$, $w(x)$ must be chosen such that the fraction is nearly constant

$$I = \frac{1}{L}\sum_{i=1}^{L}\frac{f(x(y_i))}{w(x(y_i))}. \tag{1.41}$$

Since the inversion of $y(x)$ is almost always difficult to obtain, one needs another method to approach the integral I.

1.3.3. Importance Sampling

Consider a canonical ensemble where the corresponding integral is denoted by

$$\langle A\rangle_{NVT} = \int p(\mathbf{x})A(\mathbf{x})\,\mathrm{d}\mathbf{x} = \langle\frac{A(\mathbf{x})p(\mathbf{x})}{w(\mathbf{x})}\rangle_w, \tag{1.42}$$

where $\langle \quad \rangle_w$ represents an average over all samples of $\mathbf{x}$ from the distribution $w(\mathbf{x})$, see equation (1.40) where $f(x)$ is replaced by $A(\mathbf{x})p(\mathbf{x})$. If $w(\mathbf{x}) = p(\mathbf{x})$ then $\langle A\rangle_{NVT} = \langle A\rangle_w$. The problem is how to take the samples such that at the end, $\mathbf{x}$ is distributed according to the chosen distribution. Before introducing the method found by Metropolis *et al* [14], consider first the following.

Consider a large number of particles moving independently through the space. Define $\rho_n(X)$ as the density of particles at X after n-movements. Then the net number of particles moving from X to Y in the next step equals

$$\begin{aligned}\Delta N &= \rho_n(X)\pi(X \to Y) - \rho_n(Y)\pi(Y \to X)\\ &= \rho_n(Y)\pi(X \to Y)\left[\frac{\rho_n(X)}{\rho_n(Y)} - \frac{\pi(Y \to X)}{\pi(X \to Y)}\right],\end{aligned} \tag{1.43}$$

where $\pi(X \to Y)$ is the probability that the particle will move from X to Y. If there are lots of particle in X then $\Delta N(X) > 0$, hence particles are moving from X to Y (towards the equilibrium state). The expression in the square brackets in (1.43) also shows that equilibrium is obtained if

$$\frac{\rho_n(X)}{\rho_n(Y)} = \frac{\pi(Y \to X)}{\pi(X \to Y)} = \frac{\rho_e(X)}{\rho_e(Y)}, \tag{1.44}$$

where $\rho_e(X)$ is the density of particles at X in equilibrium state. From here we go back to the problem how to take samples such that the samples form the distribution $w(X)$. We need to couple $\rho_e(X)$ and $w(X)$. The transition probability $\pi(X \to Y)$ can be rewritten as

$$\pi(X \to Y) = \alpha(X \to Y)\kappa(X \to Y), \tag{1.45}$$

where $\alpha(X \to Y)$ is probability of making a trial step from $X \to Y$ and $\kappa(X \to Y)$ is the probability of accepting this step. In equilibrium the following holds

$$\frac{\rho_e(X)}{\rho_e(Y)} = \frac{\pi(Y \to X)}{\pi(X \to Y)} = \frac{\alpha(Y \to X)\kappa(Y \to X)}{\alpha(X \to Y)\kappa(X \to Y)} = \frac{\kappa(Y \to X)}{\kappa(X \to Y)}, \tag{1.46}$$

where the latter equality holds if we assume that the trial step probability from X to Y is the same as the trial step probability back (from Y to X), i.e. $\alpha(X \to Y) = \alpha(Y \to X)$. Rest us to define the acceptance probability, which gives us the chance to include the distribution function $w(X)$. Metroplis *et al* introduced the following asymmetric scheme by defining the acceptance probability in the following way

$$\left.\begin{aligned} \kappa(Y \to X) &= 1 \\ \kappa(X \to Y) &= \frac{w(Y)}{w(X)} \end{aligned}\right\} \quad \text{if } w(X) > w(Y), \tag{1.47}$$

and in the other case

$$\left.\begin{aligned} \kappa(X \to Y) &= 1 \\ \kappa(Y \to X) &= \frac{w(X)}{w(Y)} \end{aligned}\right\} \quad \text{if } w(X) < w(Y). \tag{1.48}$$

Hence, in both the cases the equilibrium state satisfies

$$\frac{\rho_e(X)}{\rho_e(Y)} = \frac{w(X)}{w(Y)}. \tag{1.49}$$

The transition probability becomes

$$\pi(X \to Y) = \begin{cases} \alpha(X \to Y) & w(Y) \geq w(X) \qquad (1.50a) \\ \\ \alpha(X \to Y)\frac{w(Y)}{w(X)} & w(Y) < w(X) \qquad (1.50b) \end{cases}$$

$$\pi(X \to X) = 1 - \sum_{X \neq Y} \pi(X \to Y). \tag{1.51}$$

This gives the following algorithm for drawing samples from distribution $w(\mathbf{x})$

Algorithm 1.3.1 *Importance Sampling Monte Carlo*

```
1. initiate $\mathbf{x}_n$ at random
2. change $\mathbf{x}_n$ with a random $\Delta\mathbf{x}$
3. calculate $\frac{w(\mathbf{x}_n+\Delta\mathbf{x})}{w(\mathbf{x}_n)}$
4. draw a random number r between $(0, 1)$
    • if $r < \min(1, \frac{w(\mathbf{x}_n+\Delta\mathbf{x})}{w(\mathbf{x}_n)})$
      then $\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta\mathbf{x}$
    • else $\mathbf{x}_{n+1} = \mathbf{x}_n$
5. repeat from step 2 for $n = n + 1$
```

1.3.3a. Force-Bias Monte Carlo

It is not necessary that $\alpha(X \to Y) = \alpha(Y \to X)$. If $\alpha(X \to Y) \neq \alpha(Y \to X)$, then the acceptance probability needs to be modified without violating $\frac{\rho_e(X)}{\rho_e(Y)} = \frac{w(X)}{w(Y)}$. The acceptance probability becomes

$$\left.\begin{aligned} \kappa(Y \to X) &= 1 \\ \kappa(X \to Y) &= \frac{\alpha(Y \to X)}{\alpha(X \to Y)}\frac{w(Y)}{w(X)} \end{aligned}\right\} \quad \text{if } \alpha(X \to Y)w(X) > \alpha(Y \to X)w(Y), \tag{1.52}$$

and in the other case

$$\left.\begin{aligned} \kappa(X \to Y) &= 1 \\ \kappa(Y \to X) &= \frac{\alpha(X \to Y)}{\alpha(Y \to X)}\frac{w(X)}{w(Y)} \end{aligned}\right\} \quad \text{if } \alpha(X \to Y)w(X) < \alpha(Y \to X)w(Y). \tag{1.53}$$

The transition probability becomes

$$\pi(X \to Y) = \begin{cases} \alpha(X \to Y) & \alpha(X \to Y)w(X) \leq \alpha(Y \to X)w(Y) \quad (1.54a) \\ \alpha(X \to Y)\frac{\alpha(Y \to X)}{\alpha(X \to Y)}\frac{w(Y)}{w(X)} & \alpha(X \to Y)w(X) > \alpha(Y \to X)w(Y) \quad (1.54b) \end{cases}$$

$$\pi(X \to X) = 1 - \sum_{X \neq Y} \pi(X \to Y). \tag{1.55}$$

Step 4 in algorithm 1.3.1 should now be modified by changing

$$\min(1, \frac{w(\mathbf{x}_n + \Delta\mathbf{x})}{w(\mathbf{x}_n)}) \quad \text{into} \quad \min(1, \frac{\alpha(\mathbf{x}_n + \Delta\mathbf{x} \to \mathbf{x}_n)w(\mathbf{x}_n + \Delta\mathbf{x})}{\alpha(\mathbf{x}_n \to \mathbf{x}_n + \Delta\mathbf{x})w(\mathbf{x}_n)}). \tag{1.56}$$

The trial probability function α is included in the acceptance criteria of the algorithm. Taking a closer look to the definition of the trial probability, one can conclude that incorporating α in the criteria, gives us an opportunity to bias moves to preferred directions. Back to the canonical ensemble, with a system of N particles, we have

$$w(\mathbf{x}) = \mathcal{N} \exp[-\beta\Phi(\mathbf{x})], \tag{1.57}$$

where $\beta = \frac{1}{kT}$, $\Phi(\mathbf{x})$ is the potential energy and $\mathbf{x}$ is a vector of length $3N$. It is obvious that when the trial probability function is defined by $\alpha(\mathbf{x} + \Delta\mathbf{x} \to \mathbf{x}) = w(\mathbf{x})$ and $\alpha(\mathbf{x} \to \mathbf{x} + \Delta\mathbf{x}) = w(\mathbf{x} + \Delta\mathbf{x})$ that then every move will be accepted. Unfortunately it is not possible to do so, since this will require pre-knowledge of the whole distribution. For the canonical ensemble, we can approach the trial distribution α by expanding $w(\mathbf{x} + \Delta\mathbf{x})$ around $\mathbf{x}$, which means for one particle move

$$\alpha(x_i \to x_i + \Delta x_i) = c \exp[-\beta(\nabla_{x_i}\Phi(\mathbf{x}) \cdot \Delta x_i)], \tag{1.58}$$

where c is the normalization constant. Since $\Phi(\mathbf{x})$ is the potential energy, $-\nabla_{x_i}\Phi(\mathbf{x})$ is the force $f_i(\mathbf{x})$ on particle i in state $\mathbf{x}$. This shows that displacements in the direction of the force are more often accepted. From here we will use the abbreviation α_{on} for trial probability from the old state to the new state $\alpha(x_i \to x_i + \Delta x_i)$ and vice versa α_{no} for the trial probability $\alpha(x_i + \Delta x_i \to x_i)$. For $w(\mathbf{x})$ we will do the same, i.e. $w(\mathbf{x}) = w_o$ and $w(\mathbf{x} + \Delta x) = w_n$. The displacement is written as $\Delta x_i = x_i^n - x_i^o = \Delta x_i^{on}$ and the force on particle i in the old state is written as f_i^o and similarly in the new state as f_i^n. The acceptance ratio becomes

$$\begin{aligned}\frac{\alpha_{no} w_n}{\alpha_{on} w_o} &= \frac{c_n}{c_o} \frac{\exp[\beta f_i^n (x_i^o - x_i^n)]}{\exp[\beta f_i^o (x_i^n - x_i^o)]} \frac{\exp[-\beta \Phi(\mathbf{x} + \Delta \mathbf{x})]}{\exp[-\beta \Phi(\mathbf{x})]} \\ &= \frac{c_n}{c_o} \exp[-\beta((f_i^o + f_i^n) \Delta x_i^{on} + \Delta \Phi)], \end{aligned} \tag{1.59}$$

where $\Delta\Phi$ is the change in energy. This scheme is called the force-bias scheme by Pangali, Rao and Berne [15]. The displacement Δx_i^{on} here is a random displacement. In the next subsection we relate this displacement with Brownian motion.

1.3.3b. Smart Monte Carlo

P.J. Rossky, J.D. Doll and H.L. Friedman [16] used the displacement of the brownian dynamics from D.L. Ermak [17] as the trial displacement Δx_i^{on} of particle i from the old state to the new state

$$\Delta x_i^{on} = \beta A f_i^o + \delta x_i, \tag{1.60}$$

where δx_i is a Gaussian random variable, with zero mean and variance $2A$. The trial distribution function equals

$$\alpha^{on} = (4A\pi)^{-\frac{3}{2}} \exp[-(\Delta x_i^{no} - \beta A f_i^o)^2 \frac{1}{4A}], \tag{1.61}$$

and the acceptance ratio equals

$$\frac{\alpha_{no} w_n}{\alpha_{on} w_o} = \exp[-\beta(\Delta\Phi + \frac{(f_i^n + f_i^o)}{2} \Delta x_i^{no} + \frac{\beta A}{4}((f_i^n - f_i^o)^2 + 2 f_i^o (f_i^n - f_i^o)))]. \tag{1.62}$$

Notice the similarities between the force-bias and smart Monte Carlo method.

1.3.4. Simulated Annealing

Simulated annealing was found by modeling a physical phenomenon, and is a generalization of a Monte Carlo method for global optimization. In a physical annealing process, a melt is slowly cooled down. The initially high temperature of the melt corresponds to a disordered system and as cooling proceeds, the system becomes more ordered. Due to the slowly cooling down, the system is in thermodynamic equilibrium any time, and at the end (where the temperature approaches zero) the lowest energy state is obtained. If the initial temperature is too low or cooling is done too

fast, the system may become quenched, forming defects or freezing out in metastable states (i.e. trapped in a local minimum energy state).
At the initial high temperature Monte Carlo sampling is applied and after sufficient time the temperature is decremented and again a Monte Carlo sampling is applied. The entire process is repeated until a frozen state is achieved at T=0, corresponding with the lowest energy state and thus the global minimum of the energy potential Φ.

Simulated annealing is also applicable to combinatorial problem [18, 19]. The energy potential corresponds to the objective function and each state of the thermodynamic system correspond to the current solution of the combinatorial problem. The difficulty is finding an analogy in the combinatorial problem for the temperature. For more details we refer to [20].

1.4. Mathematical preliminaries on Unconstrained Optimization

Using the potential energy function to find the geometry of a molecule (or an assemblage of molecules) which corresponds to a minimum potential energy, is a very common goal in molecular simulations. The minimization of the potential energy function (i.e., geometry optimization) involves finding the minima of the potential field Φ. The minima correspond to stable states of the system; any movement away from a minimum gives a higher energy. Since finding a minimum is a typical problem in mathematics, it is not surprising that fundamental mathematical methods has been used in the search for these stable states.

1.4.1. Introduction in numerical mathematics

In unconstrained optimization an objective function $f(\mathbf{x})$, $f : \mathbb{R}^n \to \mathbb{R}$, has to be minimized with respect to $\mathbf{x}$. Mathematically, this can be written as

$$\min_{\mathbf{x}} f(\mathbf{x}). \tag{1.63}$$

Since $n \gg 1$ for large systems, one often needs a numerical approach to solve equation (1.63). There are two common approaches in numerical minimization: the line search strategy and the trust region strategy. Trust region methods first choose a step size and approximate the objective function within a certain region, then a direction is chosen for the displacement in this trust region. Line search methods first choose a direction and then a step size is chosen. Our method is partially based on the this method. Below we describe the most popular line search method: the quasi Newton (QN) method.

1.4.2. Quasi Newton Method

The line search method searches along a known direction **p** with unknown step length ϵ using

$$f(\mathbf{x}_k + \epsilon\mathbf{p}) = f_k + \epsilon\mathbf{p}^T\nabla f_k + \frac{1}{2}\epsilon^2\mathbf{p}^T\nabla^2 f(\mathbf{x}_k + \tau\mathbf{p})\mathbf{p}, \tag{1.64}$$

where $\tau \in (0, \epsilon)$ and $f_k = f(\mathbf{x}_k)$. The popular Newton direction is found by writing the object function as,

$$f(\mathbf{x}_k + \mathbf{p}) \approx f_k + \mathbf{p}^T\nabla f_k + \frac{1}{2}\mathbf{p}^T\nabla^2 f_k\mathbf{p} =: m_k(\mathbf{p}). \tag{1.65}$$

The approximation $f(\mathbf{x}_k + \mathbf{p}) \approx m_k(\mathbf{p})$ is very accurate for small $||\mathbf{p}||$. The unknown direction **p** is found by setting the derivative with respect to $\mathbf{p}_k$ of $m_k(\mathbf{p}_k)$ to zero. This gives the direction

$$\mathbf{p}_k = -\nabla^2 f_k^{-1}\nabla f_k, \tag{1.66}$$

which requires the existence of the inverse of $\nabla^2 f_k$. Therefore, $\nabla^2 f_k$ must be positive definite. If the inverse Hessian exists the direction (1.66) has the descending property, since for the second term of (1.64) holds

$$\nabla f_k^T\mathbf{p}_k = -\mathbf{p}_k^T\nabla^2 f_k\mathbf{p}_k \leq 0. \tag{1.67}$$

Not only the uncertainty of the existence of the inverse is a drawback in the Newton method but also it requires to calculate the Hessian $\nabla^2 f(\mathbf{x})$, which becomes a very expensive job for large systems. A method that avoids the calculation of the second derivatives, by using an iterative estimate of the Hessian, is called the Quasi Newton method.

In the next section the Quasi Newton method, BFGS (Broyden, Fletcher, Goldfarb and Shanno), will be described.

Using Taylor the following equation for a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ holds

$$f(\mathbf{x} + \mathbf{p}) = f(\mathbf{x}) + \nabla f(\mathbf{x} + \tau\mathbf{p}), \tag{1.68}$$

where $\tau \in (0,1)$. Moreover, if f is twice continuously differentiable then the following holds

$$\begin{aligned} \nabla f(\mathbf{x}+\mathbf{p}) &= \nabla f(\mathbf{x}) + \int_0^1 \nabla^2 f(\mathbf{x}+\tau\mathbf{p})\mathbf{p}\,\mathrm{d}\tau \\ &= \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\mathbf{p} + \int_0^1 [\nabla^2 f(\mathbf{x}+\tau\mathbf{p}) - \nabla^2 f(\mathbf{x})]\mathbf{p}\,\mathrm{d}\tau \end{aligned} \tag{1.69}$$

After setting $\mathbf{x} = \mathbf{x}_k$ and $\mathbf{p} = \mathbf{x}_{k+1} - \mathbf{x}_k$, one obtains

$$\nabla f_{k+1} = \nabla f_k + \nabla^2 f_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) + \mathcal{O}(||\mathbf{x}_{k+1} - \mathbf{x}_k||),$$

which gives after neglecting the (small) last term on the right hand side of the equation

$$\underbrace{\nabla f_{k+1} - \nabla f_k}_{\mathbf{y}_k} \approx \underbrace{\nabla^2 f_{k+1}}_{H_{k+1}} \underbrace{(\mathbf{x}_{k+1} - \mathbf{x}_k)}_{\mathbf{s}_k}. \tag{1.70}$$

The equation with H_{k+1} as the Hessian approximation,

$$H_{k+1}\mathbf{s}_k = \mathbf{y}_k, \tag{1.71}$$

is called the secant equation, which guarantees that the gradient of $m_{k+1}(\mathbf{p})$,

$$\nabla[m_{k+1}(\mathbf{p})] = \nabla[f_{k+1} + \nabla f_{k+1}^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T B_{k+1}\mathbf{p}], \tag{1.72}$$

matches the gradient of the objective function f at the latest two iterates $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$. The approximated Hessian needs to be positive definite, which holds if $\mathbf{s}_k$ and $\mathbf{y}_k$ satisfy the curvature condition

$$\mathbf{s}_k^T\mathbf{y}_k > 0. \tag{1.73}$$

When the object function f is strongly convex, the curvature condition is valid for any two points $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$. For nonconvex functions the curvature condition holds if the following conditions are set for the line search

$$f(\mathbf{x}_k + \epsilon_k\mathbf{p}_k) \leq f(\mathbf{x}_k) + c_1\epsilon_k\nabla f_k^T\mathbf{p}_k, \tag{1.74a}$$

$$\nabla f(\mathbf{x}_k + \epsilon_k\mathbf{p}_k)^T\mathbf{p}_k \geq c_2\nabla f_k^T\mathbf{p}_k, \tag{1.74b}$$

with $0 < c_1 < c_2 < 1$ and where ϵ_k is the step length in the direction $\mathbf{p}_k$. These conditions are called the Wolfe conditions which guarantees sufficient decrease in the object function. If the second condition holds then for condition (1.73) holds

$$\mathbf{s}_k^T\mathbf{y}_k \geq (c_2 - 1)\epsilon_k \nabla f_k^T \mathbf{p}_k, \tag{1.75}$$

where due to the descending direction of $\mathbf{p}_k$ the right hand side is positive and therefore, the curvature condition holds. Hence, the secant equation (1.71) has a solution. The secant equation gives n conditions. The requirement of positive definiteness of H_{k+1} gives also n conditions. But H_{k+1} has $\frac{n(n+1)}{2}$ degrees of freedom, so H_{k+1} is not uniquely determined. To determine H_{k+1} uniquely, an additional condition will be imposed; among all symmetric matrices, H is in some sense, closest to the current matrix H_k. This gives the following problem

$$\min_H \|H - H_k\|, \tag{1.76a}$$

subject to

$$H = H^T,\ H\mathbf{s}_k = \mathbf{y}_k,\ \text{and } \mathbf{v}^T H\mathbf{v} > 0 \text{ for all } \mathbf{v} \neq 0, \mathbf{v} \in \mathbb{R}^n. \tag{1.76b}$$

Different matrix norms can be used in (1.76a). A norm that allows easy solution of the minimization problem is the weighted Frobenius norm

$$\|A\|_W = \|W^{1/2}AW^{1/2}\|_F, \quad \text{with} \quad \|A\|_F = \Big(\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^2\Big)^{1/2}. \tag{1.77}$$

The weight matrix W is chosen such that the relation $W\mathbf{y}_k = \mathbf{s}_k$ is satisfied. The weight W is chosen as the inverse of the average Hessian defined by

$$[\overline{G}_k]^{-1} = \Big[\int_0^1 \nabla^2 f(\mathbf{x}_k + \tau\epsilon_k\mathbf{p}_k)\,\mathrm{d}\tau\Big]^{-1}, \tag{1.78}$$

having the property $\mathbf{y}_k = \overline{G}_k\epsilon_k\mathbf{p}_k = \overline{G}_k\mathbf{s}_k$, which follows from (1.69). Using this norm and weighting matrix, the unique solution of (1.76) is given by

$$H_{k+1} = (I - \frac{\mathbf{y}_k\mathbf{s}_k^T}{\mathbf{y}_k^T\mathbf{s}_k})H_k(I - \frac{\mathbf{s}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}) + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k}. \tag{1.79}$$

This formula is called the DFP(Davidson, Fletcher and Powell) updating formula [21]. Since

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \epsilon_k\mathbf{p}_k = \mathbf{x}_k - \epsilon_k H_k^{-1}\nabla f_k = \mathbf{x}_k - \epsilon_k B_k \nabla f_k,$$

it is useful to derive the updating formula for the inverse Hessian approximation B_k

$$B_{k+1} = B_k - \frac{B_k \mathbf{y}_k \mathbf{y}_k^T B_k}{\mathbf{y}_k^T B_k \mathbf{y}_k} + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \tag{1.80}$$

which is found by making use of the Sherman-Morrison-Woodbury formula.

Instead of imposing an additional condition on H_{k+1}, one can impose the same closeness condition on its inverse B_{k+1}, which gives the following problem

$$\min_B ||B - B_k|| \tag{1.81a}$$

subject to

$$B = B^T, \; B\mathbf{y}_k = \mathbf{s}_k, \text{ and } \mathbf{v}^T B \mathbf{v} > 0 \text{ for all } \mathbf{v} \neq 0, \mathbf{v} \in \mathbb{R}^n. \tag{1.81b}$$

Using again the weighted Frobenius norm with the average Hessian as the weight matrix satisfying $W\mathbf{s}_k = \mathbf{y}_k$, the unique solution is given by

$$B_{k+1} = (I - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}) B_k (I - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}, \tag{1.82}$$

which is called the BFGS updating formula. Again, using the Sherman-Morrison-Woodbury formula the inverse of B_{k+1} can be found

$$H_{k+1} = H_k - \frac{H_k \mathbf{s}_k \mathbf{s}_k^T H_k}{\mathbf{s}_k^T H_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}. \tag{1.83}$$

1.5. Multi-scaling; Slow and fast modes

Molecular dynamics is an useful technique for studying (thermo-) dynamics of physical systems like proteins and other biomolecules. However, using conventional MD it is nearly impossible to have access to all the time scales of motion with atomic resolution within reasonable time. The simulation performance of molecular systems, with various length and time scales, is dictated by the smallest scale in the model description. The small time step size limits the study of important conformational changes, which are determined by the slow modes of the system. For improving the simulation performance the basic idea is to separate the dynamics into fast and slow modes and integrate the slow motions with a larger time step. In Figure 1.2 the contours of a system of 2 coordinates ($\Phi(x_1, x_2)$), labeled x_1 and x_2 , are projected in a 2D plot.

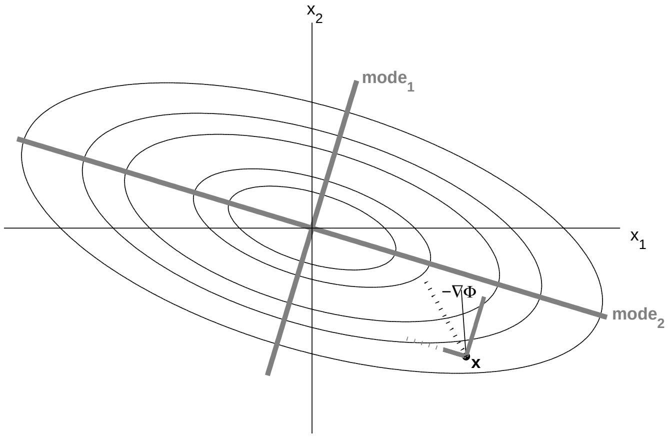


Figure 1.2: *Illustration of the different modes of a system.*

The minimum is located in the center and the normal mode directions are given by the gray lines, labeled mode_1 and mode_2 in Figure 1.2. Moving along a normal mode is related to the curvature of the potential along the mode direction. The normal modes and their frequencies are given by the eigenvectors and eigenvalues of the Hessian. The frequency is related to the curvature of the potential along the direction of the mode. Slow modes correspond to eigenvectors with small eigenvalues and fast modes correspond to eigenvectors with large eigenvalues. The normal mode directions are independent, which enables us to write any motion as a linear sum of the normal modes. In Figure 1.2 we have plotted the force $-\nabla\Phi$ at location **x**. The force is a compound motion of the slow (mode_2) and fast (mode_1) modes. Although the model is simplistic, it shows that the amplitude along the slow mode is smaller than the amplitude along the fast mode. By artificially increasing the amplitude along the slow mode, a larger motion toward the minimum is made, i.e. a faster decrease of the energy potential. In Figure 1.2 we amplified (grey dashed line along mode_2) the slow mode of the force at **x**, such that the amplitude for the slow mode becomes the same as the amplitude of the fast mode. The resulting total force, sum of the amplified slow mode and the original fast mode, is also given (black dashed line). Clearly amplifying the slow mode here results in a larger step toward the minimum.

As shown in the above example, partitioning the system into dynamics with fast modes and dynamics with slow modes allows for efficient propagation of the slow dynamics. Among popular approaches in separating modes of motion are *quasi-harmonic analysis* [22], *molecule optimal dynamic coordinates* [23] and *essential dynamics*[24].

In this thesis we propose a method which amplifies the slow modes of the system, which effectively means integration of the Langevin equation with a larger time step. The proposed method does not require any explicit partitioning of the space into dynamics with fast modes and dynamics with slow modes. Hence, the title of this thesis *Stochastic quasi Newton method for molecular simulations* could also be generalized into *Automated multi-scaling method with Langevin Dynamics for molecular simulations*.