
Samenvatting

In dit proefschrift heb ik de Langevin vergelijking beschouwd met een alternatieve mobiliteits matrix. Waar men veelal de mobiliteits matrix in de Langevin vergelijking als constant betracht (inverse van de frictie coëfficiënt), beschouw ik de mobiliteits matrix als een plaatsafhankelijke matrix die in elke simulatie stap vernieuwd moet worden. De Langevin vergelijking is een stochastische differentiaalvergelijking die een beschrijving geeft van de tijdsevolutie van een set veranderende macroscopische variabelen. Deze set van variabelen veranderen relatief langzaam ten op zichte van andere (microscopische) variabelen. Een van de eerste beschrijvingen die gebruik maakt van de Langevin vergelijking is de beschrijving van Brownse beweging: de beweging van (macroscopisch grote) pollen in water door botsingen met de snel bewegende water moleculen.

In een moleculair systeem kunnen de posities (coördinaten) van de grote deeltjes (moleculen/atomen) gezien worden als een set macroscopische variabelen en de posities van de kleine deeltjes kunnen gezien worden als de microscopische variabelen. Dit komt overeen met de relatief snellere bewegingen van de kleinere deeltjes ten opzichte van de bewegingen van de grotere deeltjes. De onderverdeling in variabelen maakt de Langevin vergelijking uiterst geschikt voor de beschrijving van deze "coarse-grained" modellen.

In een simulatie, gebruikmakend van de klassieke moleculaire dynamica (MD), worden de atomen voorgesteld als deeltjes en de bewegingen van deze deeltjes volgen uit de bewegingsvergelijkingen van Newton. Een complex systeem heeft echter vele verschillende lengte- en tijdschalen waardoor simulaties gebaseerd op fundamentele bewegingsvergelijkingen heel veel reken/simulatie-tijd vergen voor het in kaart brengen van functionele en relevante bewegingen. Dit probleem wordt "critical slowing down" genoemd. Om dit probleem te overbruggen worden (behalve coarse-

grained modellen) multi-schaal simulatie technieken toegepast, die verschillende tijdstap groottes toestaan en dus mogelijk maken om de relevante (langzame) bewegingen te accelereren.

Het doel van dit proefschrift is een stochastische quasi Newton (S-QN) methode te ontwikkelen, waarbij door multi-schaling de relevante bewegingen effectief met een grotere tijdstap worden genomen. Door langzame bewegingen met grotere tijdstappen te integreren wordt "critical slowing down" voorkomen. Het detecteren van de langzaamste modes (deze corresponderen met de relevante bewegingen) in elke tijdstap en de bewegingen in deze modes accelereren is een mogelijkheid. Echter, het detecteren van de verschillende modes in een systeem in elke tijdstap is veel te kostbaar. Vandaar de ontwikkeling van een methode die automatisch de mobiliteits matrix in de Langevin vergelijking aanpast, zodanig dat in de resulterende bewegingen de langzame modes automatisch met een grotere tijdstap worden geïntegreerd.

In hoofdstuk 2 wordt de nieuwe mobiliteits matrix gepresenteerd: de inverse Hessiaan van de energie potentiaal. Door deze keuze voor de mobiliteits matrix te nemen lijkt de deterministische term van de Langevin vergelijking op de verplaatsing in de Newton minimalisatie methode. We verwachten, analoog aan het prestatieverschil tussen de Newton methode en de steepest descent methode, een betere convergentie naar het energie minimum vergeleken met een constante mobiliteits matrix. De stochastische term (ruis-term) in de Langevin vergelijking, geconstrueerd met de gefactoriseerde mobiliteit, is essentieel voor correcte thermodynamische distributies. Aan de hand van eenvoudige 1 en 2 dimensionale systemen wordt aangetoond dat de alternatieve mobiliteit thermodynamisch consistent is: de energie potentiaal heeft een Boltzmann verdeling. De keuze van de inverse Hessiaan draagt ook bij aan de snellere verplaatsing over energie barrières. Deze laatste is kwantitatief gemeten aan de hand van de gemiddelde eerste passage tijd waarbij gekeken wordt naar de gemiddelde benodigde tijd om van onder aan het energielandschap tot aan de top van het energielandschap te komen. Gebruikmakend van de inverse Hessiaan is deze gemiddelde tijd een orde kleiner dan de gemiddelde tijd nodig met een constante mobiliteit. Omdat het uitrekenen van de analytische inverse Hessiaan erg veel rekentijd vergt, wordt de inverse Hessiaan benaderd. Deze benadering is afgeleid van een bestaand update schema in de quasi Newton methode: het DFP-update schema.

In hoofdstuk 3 wordt een gefactoriseerd secant update (FSU) geconstrueerd. Deze update is een gefactoriseerde equivalent van de DFP-update, waardoor er geen expliciete factorisatie van de mobiliteit meer nodig is, hetgeen de complexiteit reduceert van $O(n^3)$ naar $O(n^2)$. De FSU is ook zodanig geconstrueerd dat het een basis vormt

voor de constructie van een gelimiteerd geheugen versie (LFSU), waarbij geen matrices worden opgeslagen en geen matrix vermenigvuldigingen plaatsvinden. In plaats van n^2 matrix elementen opslaan worden er $3m$ vectoren van lengte n opgeslagen, waarbij m de afkap parameter of de historie diepte is. Voor LFSU wordt de complexiteit nog verder gereduceerd naar $O(mn)$. Aan de hand van een systeem met deeltjes waarbij harmonische interactie tussen de buren plaatsvindt, wordt getoond dat het nieuwe update schema naar de inverse Hessiaan convergeert. Ook geldt dat FSU en LFSU automatisch bijdragen aan het accelereren van de verschillende tijd en lengte schalen, met als resultaat een betere sampling prestatie vergeleken met de conventionele Langevin dynamica (constante mobiliteits matrix).

In hoofdstuk 4 wordt de Langevin dynamica met de door FSU geconstrueerde mobiliteit toegepast op een model eiwit. Voor de stabiliteit van de simulatie wordt het FSU schema aangepast zodanig dat er regularisatie plaatsvindt voor slecht geconditioneerde systemen: systemen met grote conditie getallen (grootste eigenwaarde van de mobiliteit matrix gedeeld door de kleinste eigenwaarde). Het model eiwit is gebaseerd op 3 soorten 'groffe' deeltjes die ofwel een hydrofobe ofwel een hydrofiele ofwel een neutrale peptide representeren. Met dit eiwit model is het mogelijk om ketenafhankelijke eigenschappen van een eiwit te bepalen. Met de S-QN methode worden de langzame modes geaccelereerd, hetgeen de verplaatsing over energie barrières versneld en een groter gebied van het energie landschap wordt bewandeld. Het gevolg is dat verschillende lokale minima zijn gevonden, waaronder ook het globale minimum. Gedetailleerde analyse van de SQ-N methode laat zien dat automatische multi-schaling bijdraagt aan het eerst equilibreren van de bindings lengtes, bindings hoeken en torsie hoeken. Nadat deze bindings eigenschappen zijn gestabiliseerd wordt er een plotselinge daling in het totale energie potentiaal waargenomen, hetgeen correspondeert met de niet bindings-interacties die een rol gaan spelen. De plotselinge daling in de totale energie potentiaal komt overeen met de daling in de overlap functie. Deze functie geeft aan in hoeverre de configuratie van het eiwit overeenkomt met de configuratie van een referentie eiwit. Als referentie eiwit wordt vaak de natieve conformatie genomen, waarbij een totale overlap correspondeert met functie waarde 0 en geen overlap met functie waarde 1. De evolutie van de overlap functies van de partiële structuren(sub-domeinen) laat zien dat deze domeinen alvorens de plotselinge daling al zijn gevormd en collectief bewegen. De collectieve bewegingen zijn afwezig in de conventionele Langevin dynamica zodat minder minima worden gevonden vanwege de kritieke vertraging.

Het voorgestelde FSU schema draagt bij aan de automatische multi-schaling in de

Langevin dynamica en een efficiënte berekening van de ruis-term. Ook is de FSU zodanig opgezet dat een gelimiteerde variant voor het berekenen van de mobiliteit mogelijk is. Echter de Langevin vergelijking met een plaatsafhankelijke mobiliteit bezit een correctie term voor de ruis, waarbij de divergentie van de mobiliteit berekend moet worden. Deze term heeft geen bijdrage in de conventionele Langevin vergelijking (divergentie van constante mobiliteit levert geen bijdrage). Omdat de divergentie term uitrekenen computationeel erg duur is, wordt dit vermeden door een predictie en correctie schema, waarbij wel de inverse van de mobiliteit moet worden uitgerekend. In hoofdstuk 5 wordt op de computationele aspecten van het uitrekenen van de Langevin vergelijking met plaatsafhankelijke mobiliteit ingegaan. Voor de inverse van de mobiliteit wordt een gelimiteerde geheugen schema geconstrueerd, waardoor de gehele Langevin vergelijking in een gelimiteerde geheugen vorm geschreven kan worden. Standaard computationele methoden zijn van $O(n^3)$, terwijl FSU van $O(n^2)$ is en LFSU van $O(mn)$ en minder opslagcapaciteit nodig heeft. Tezamen met de automatische multi-schaling eigenschap, vormt (L)FSU een krachtige methode voor het uitvoeren van moleculaire simulaties.