

## 2 Simulation moléculaire

### 2.1 Description

La simulation moléculaire [Tuck10] consiste à modéliser l'évolution de la matière à l'échelle des atomes. L'objectif est de comprendre comment des phénomènes visibles et mesurables à l'échelle macroscopique (le changement de conformation d'une protéine, une réaction chimique, un changement de phase, ...) émergent à partir de modèles à l'échelle microscopique. Ce "microscope numérique" a pris une importance considérable dans de nombreux domaines, en particulier en biologie, chimie, physique et sciences des matériaux. L'objectif de ce projet est d'explorer sur un modèle jouet les difficultés posées par ces simulations, notamment du fait de l'importance jouée par des événements rares.



▷ **Modélisation.** L'ingrédient essentiel de modélisation à l'échelle microscopique est une fonction potentielle, qui à une configuration du système (la position des noyaux des atomes) associe une énergie. En principe, cette fonction doit être obtenue par un calcul quantique, en simulant la structure électronique de l'état fondamental du système, pour une position des noyaux donnée [SJ96]. Bien souvent, on a recours à une description approchée, basée sur un champ de force empirique [Tuck10].

Une fois la fonction potentielle choisie, l'évolution du système est modélisée par une dynamique qui dépend de l'environnement thermodynamique du système : dynamique hamiltonienne pour un système isolé, ou bien dynamique de Langevin pour un système à température donnée, par exemple.

On se propose de se placer pour simplifier les choses en dimension 1 d'espace, et de considérer la dynamique de Langevin sur-amortie :

$$dX(t) = -V'(X(t))dt + \sqrt{2\epsilon}dW(t) \quad (1)$$

où  $X(t) \in \mathbb{R}$  est la position,  $V : \mathbb{R} \rightarrow \mathbb{R}$  est la fonction potentielle mentionnée ci-dessus,  $(W(t))_{t \geq 0}$  est un mouvement brownien, et  $\epsilon > 0$  est un paramètre proportionnel à la température du système et qui jouera un rôle fondamental dans la suite. En pratique, on simulera une discrétisation de cette dynamique, pour un pas de temps  $h > 0$  :

$$X_{n+1} = X_n - V'(X_n)h + \sqrt{2\epsilon h}G_n \quad (2)$$

où  $(G_n)_{n \geq 0}$  est une suite de gaussiennes ( $X_n$  est une approximation de  $X(nh)$ , d'autant meilleure que  $h$  est petit). On considérera le potentiel double-puits :

$$V(x) = \frac{x^4}{4} - \frac{x^2}{2} + \frac{x^3}{3}.$$

On notera  $x_1 < x_2$  les deux minima locaux et  $z \in (x_1, x_2)$  le maximum local de la fonction  $V$ . Dans la limite  $\epsilon \rightarrow 0$ , la dynamique (1) passe l'essentiel de son temps aux voisinages de  $x_1$  ou de  $x_2$ , avec de très rares transitions de  $x_1$  à  $x_2$ . On appelle ces voisinages des *états métastables* du système, qui en pratique correspondent à des conformations macroscopiques du système (penser à une conformation d'une protéine, à des réactifs et des produits, ou à différentes phases d'un matériau). Etre capable de simuler efficacement des transitions entre états métastables et le temps moyen pour observer de telles transitions est un enjeu majeur en simulation moléculaire.

## 2.2 Objectifs

### ▷ Trajectoires réactives.

Pour commencer, on s'intéresse aux trajectoires réactives, c'est-à-dire les portions de trajectoires qui réalisent une transition entre  $x_1$  et  $x_2$ . On considère pour cela une condition initiale  $X_0 = x_1 + 0.1$  et on s'intéresse à l'évènement :

$$R = \{\tau_2 < \tau_1\}$$

où  $\tau_1 = \inf\{n, X_n < x_1\}$  et  $\tau_2 = \inf\{n, X_n > x_2\}$ .

1. Simuler par une méthode de Monte Carlo naïve l'évènement  $R$ , en commençant par des valeurs de  $\epsilon$  assez grande. On pourra tracer les trajectoires associées, et la distribution de leur longueur par exemple. Que se passe-t-il quand  $\epsilon$  tend vers 0 ? Tester également l'influence du pas de temps  $h > 0$ .

Quand la température  $\epsilon$  tend vers 0, il faut adopter une méthode de simulation d'évènements rares pour pouvoir simuler l'évènement  $R$ . On se propose d'utiliser l'algorithme Adaptive Multilevel Splitting (AMS) suivant [CG07, BGGLR16].

### Initialisation.

On simule  $M$  trajectoires

$$(X^{m,0})_{1 \leq m \leq M}$$

où pour tout  $m$ ,  $X^{m,0} = (X_n^{m,0})_{0 \leq n \leq \tau_1^{m,0} \wedge \tau_2^{m,0}}$  est une réalisation de (2) (avec la condition initiale  $X_0^{m,0} = x_1 + 0.1$ ) jusqu'à l'indice de temps  $\tau_1^{m,0} \wedge \tau_2^{m,0}$ .

On appelle  $\max_{0 \leq n \leq \tau_1^{m,0} \wedge \tau_2^{m,0}} X_n^{m,0}$  le *niveau* atteint par la trajectoire  $m$  et on note

$$K^0 = \min_{1 \leq m \leq M} \max_{0 \leq n \leq \tau_1^{m,0} \wedge \tau_2^{m,0}} X_n^{m,0}$$

le minimum des niveaux atteints par les trajectoires. Finalement, on note

$$I^0 = \left\{ m, \max_{0 \leq n \leq \tau_1^{m,0} \wedge \tau_2^{m,0}} X_n^{m,0} = K^0 \right\}$$

les indices des trajectoires qui atteignent ce minimum des niveaux.

### Itérations.

On itère sur  $q \geq 0$  les étapes suivantes, tant que  $K^q \leq x_2$  :

#### (i) Branchement.

Pour tous les indices de trajectoire  $m \in I^q$ , on effectue les étapes suivantes :

- On choisit un indice  $p$  d'une trajectoire parent choisie uniformément au hasard parmi  $\{1, \dots, M\} \setminus I^q$  (ce sont donc des trajectoires de niveau supérieur à  $K^q$ ) ;
- On recopie  $X_n^{m,q+1} = X_n^{p,q}$  pour tout  $n \leq \inf\{n, X_n^{p,q} > K^q\}$
- On termine ensuite la simulation de  $X_n^{m,q+1}$  jusqu'à  $n = \tau_1^{m,q+1} \wedge \tau_2^{m,q+1}$ , en suivant la dynamique (2).

Cette étape a donc permis de remplacer toutes les trajectoires de niveau  $K^q$  par des trajectoires de niveaux supérieurs à  $K^q$ . Les autres trajectoires ne sont pas modifiées : pour tout  $m \in \{1, \dots, M\} \setminus I^q$ ,  $X_n^{m,q+1} = X_n^{m,q}$ .

#### (ii) Calcul du niveau.

On note

$$K^{q+1} = \min_{1 \leq m \leq M} \max_{0 \leq n \leq \tau_1^{m,q+1} \wedge \tau_2^{m,q+1}} X_n^{m,q+1}$$

le minimum des niveaux atteint par les nouvelles trajectoires. Finalement, on note

$$I^{q+1} = \left\{ m, \max_{0 \leq n \leq \tau_1^{m,q+1} \wedge \tau_2^{m,q+1}} X_n^{m,q+1} = K^{q+1} \right\}$$

les indices des trajectoires qui atteignent ce minimum des niveaux.

On incrémente finalement  $q$  puis on passe à l'itération suivante.

Un estimateur de la probabilité de l'évènement  $R$  est alors donné par

$$\hat{p} = \left(1 - \frac{K^0}{M}\right) \dots \left(1 - \frac{K^Q}{M}\right)$$

où  $Q = \inf\{q \geq 0, K^q > x_2\}$  est le nombre d'itérations effectuées. On peut montrer que c'est un estimateur non biaisé (cf. [BGGLR16]), et on peut donc

obtenir une estimation de  $\mathbb{P}(R)$  en moyennant  $\hat{p}$  sur plusieurs réalisations indépendantes de l'algorithme.

2. Simuler par la méthode AMS l'évènement  $R$ . On commencera par des valeurs de  $\epsilon$  qui permettent de comparer avec le calcul Monte Carlo naïf de la question précédente, puis on diminuera la valeur de  $\epsilon$ . A nouveau, on pourra tracer les trajectoires associées, et la distribution de leur longueur, et on testera également l'influence du pas de temps  $h > 0$ .

▷ **Temps de transition.**

On s'intéresse désormais au temps moyen  $T_{12}$  d'une transition de  $x_1$  à  $x_2$ . Plus précisément, on considère une longue trajectoire  $(X_t)_{t \geq 0}$  de (1) et on s'intéresse à

$$T_{12} = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K (\tau_2^k - \tau_1^k)$$

où, pour  $k \geq 1$ ,

$$\tau_1^k = \inf\{t > \tau_2^{k-1}, X_t < x_1\} \text{ et } \tau_2^k = \inf\{n > \tau_1^k, X_t > x_2\}$$

sont les temps de visite successifs de  $x_1$  et  $x_2$  (avec la convention  $\tau_2^0 = -\infty$ ).

3. Estimer par une méthode de Monte Carlo naïve le temps  $T_{12}$ , en commençant par des valeurs de  $\epsilon$  assez grandes. Tracer des trajectoires de transitions  $(X_n)_{\tau_1^k \leq n \leq \tau_2^k}$ . Que se passe-t-il quand  $\epsilon$  tend vers 0 ?

On note  $x_0 = x_1 + 0.1$ . Par la relation de Hill, on peut montrer (pour le processus en temps continu, cf. [BGL20]) que

$$T_{12} = \mathbb{E}(T_{101}) \left( \frac{1}{\mathbb{P}^{x_0}(\tau_2 < \tau_1)} - 1 \right) + \mathbb{E}(T_{102}) \quad (3)$$

où :

- $\mathbb{P}^{x_0}(\tau_2 < \tau_1)$  est la probabilité, partant de  $X_0 = x_0$  d'atteindre  $x_2$  avant  $x_1$  (qu'on peut estimer par AMS, cf. ci-dessus) ;
- $\mathbb{E}(T_{101})$  est le temps moyen, pour une trajectoire partant de  $x_1$ , d'atteindre  $x_0$  puis de revenir en  $x_1$  (qu'on peut estimer par Monte Carlo naïf) ;
- $\mathbb{E}(T_{102})$  est le temps moyen, pour une trajectoire partant de  $x_1$ , d'atteindre  $x_0$  puis d'atteindre  $x_2$  sans revenir en  $x_1$  (qu'on peut estimer par Monte Carlo naïf pour la partie de  $x_1$  à  $x_0$  puis par AMS pour la partie de  $x_0$  à  $x_2$ ).

3. Expliquer heuristiquement pourquoi la relation de Hill (3) est raisonnable. En utilisant la discrétisation (2), estimer le temps  $T_{12}$  en utilisant la relation de Hill, en combinant des calculs Monte Carlo naïfs et l’algorithme AMS. On pourra commencer par des valeurs de  $\epsilon$  assez grandes, pour comparer aux résultats obtenus à la question précédente par Monte Carlo. Dans la limite  $\epsilon \rightarrow 0$ , on pourra comparer la valeur obtenue à la formule d’Eyring-Kramers (cf. [BEGK04, HTB90, Kra40]) :

$$T_{12} \approx \frac{2\pi}{\sqrt{|V''(z)| V''(x_1)}} e^{(V(z)-V(x_1))/\epsilon}.$$

## 2.3 Extensions possibles

Voici quelques prolongements possibles à ce projet :

- On pourra s’intéresser à généraliser cette approche en dimension plus grande, par exemple pour le potentiel bi-dimensionnel du type de la figure 1, c’est-à-dire  $V(x, y) = \frac{x^4+y^4}{4} - \frac{x^2+y^2}{2} + \gamma \frac{(x-y)^2}{2}$  avec  $\gamma < 1/2$ .

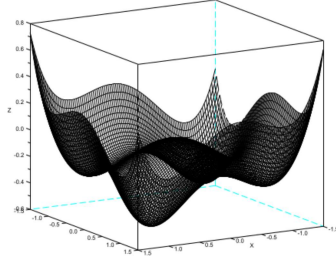


FIGURE 1 – Exemple de potentiel.

- On pourra chercher à comprendre les propriétés mathématiques de l’estimateur  $\hat{p}$  de l’algorithme AMS dans un cas simple, cf. [BLR15] ;
- On pourra également comprendre les preuves mathématiques permettant de démontrer la relation de Hill (3), cf. [BGL20].

## Références

- [BGL20] M. Baudel, A. Guyader, and T. Lelièvre. On the Hill relation and the mean reaction time for metastable processes, <https://arxiv.org/abs/2008.09790>.

- [BEGK04] A. Bovier, M. Eckhoff, V. Gaynard, and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4) :399–424, 2004
- [BLR15] C.E. Bréhier, T. Lelièvre and M. Rousset. Analysis of Adaptive Multilevel Splitting algorithms in an idealized case, *ESAIM P&S*, 19, 361 - 394, 2015.
- [CG07] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis, *Stochastic Analysis and Applications*, 25(2) :417–443, 2007.
- [BGGLR16] C.-E. Bréhier, M. Gazeau, L. Goudenège, T. Lelièvre, and M. Rousset. Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms. *Annals of Applied Probability*, 26(6), 3559-3601, 2016.
- [HTB90] P. Hanggi, P. Talkner, and M. Borkovec. Reaction-rate theory : fifty years after Kramers. *Rev. Modern Phys.*, 62(2) :251–341, 1990.
- [Kra40] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7, 1940.
- [OV05] E. Olivieri and M.E. Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications*. Chapitre 5. Cambridge University Press, Cambridge, 2005.
- [Tuck10] M. Tuckerman. *Statistical Mechanics : Theory and Molecular Simulation*. Oxford University Press, 2010.
- [SJ96] G.D. Smith and R.L. Jaffe. Quantum chemistry study of conformational energies and rotational energy barriers in n-alkanes. *The Journal of Physical Chemistry*, 100(48) :18718–18724, 1996.