

# Introduction aux méthodes de Monte Carlo quantique

Julien Toulouse  
Laboratoire de Chimie Théorique  
Université Pierre et Marie Curie et CNRS, 75005 Paris, France  
julien.toulouse@upmc.fr

26 janvier 2015

## Table des matières

<b>1</b>	<b>Monte Carlo variationnel</b>	<b>2</b>
1.1	Idée de base . . . . .	2
1.2	Estimation de l'incertitude statistique . . . . .	3
1.3	Coût de calcul . . . . .	5
1.4	Procédure d'échantillonnage . . . . .	5
<b>2</b>	<b>Monte Carlo diffusionnel</b>	<b>9</b>
2.1	Idée de base . . . . .	9
2.2	Mise en œuvre stochastique . . . . .	11
2.3	Approximation des nœuds fixés . . . . .	13
	<b>Références</b>	<b>14</b>

Ce document est une introduction aux deux principales variantes de méthodes de Monte Carlo quantique (QMC) pour les calculs de structure électronique : Monte Carlo variationnel (VMC) et Monte Carlo diffusionnel (DMC). Pour plus de détails sur ces méthodes, consulter par exemple les références [1–6]. Pour des revues sur les applications des méthodes QMC en chimie et en physique de la matière condensée, consulter par exemple les références [7, 8].

## 1 Monte Carlo variationnel

### 1.1 Idée de base

L'idée de la méthode VMC est simplement de calculer les intégrales multidimensionnelles apparaissant en mécanique quantique par une technique d'intégration numérique Monte Carlo. La quantité de premier intérêt est l'énergie variationnelle associée à un hamiltonien  $\hat{H}$  et une fonction d'onde  $\Psi$  qui peut s'écrire

$$E_v = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int d\mathbf{R} \Psi(\mathbf{R})^2 E_L(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R})^2} = \int d\mathbf{R} \rho(\mathbf{R}) E_L(\mathbf{R}), \quad (1)$$

où  $\Psi$  est supposée réelle et  $E_L(\mathbf{R}) = (H\Psi(\mathbf{R}))/\Psi(\mathbf{R})$  est l'*énergie locale* dépendant des  $3N$  coordonnées  $\mathbf{R}$  des  $N$  électrons, et  $\rho(\mathbf{R}) = \Psi(\mathbf{R})^2 / \int d\mathbf{R} \Psi(\mathbf{R})^2$  est la densité de probabilité normalisée. L'énergie locale  $E_L(\mathbf{R})$  peut être considérée comme une variable aléatoire continue associée à la distribution de probabilité  $\rho(\mathbf{R})$ , et l'énergie variationnelle peut alors être estimée par la valeur moyenne de  $E_L(\mathbf{R})$  sur un échantillon de  $M$  points  $\mathbf{R}_k$  distribués suivant  $\rho(\mathbf{R})$ ,

$$E_v \approx \langle E_L \rangle_\rho = \frac{1}{M} \sum_{k=1}^M E_L(\mathbf{R}_k), \quad (2)$$

pour  $M$  assez grand. En pratique, les points  $\mathbf{R}_k$  sont obtenus par un algorithme d'échantillonnage de type Metropolis.

L'avantage de cette approche est qu'elle n'implique pas d'intégration analytique de la fonction d'onde, et donc n'impose pas de contraintes particulières dans le choix de la forme de la fonction d'onde. Les fonctions d'onde couramment utilisées en QMC sont de type Jastrow-Slater,

$$\Psi(\mathbf{R}) = J(\mathbf{R})\Phi(\mathbf{R}), \quad (3)$$

où  $J(\mathbf{R})$  est un facteur de Jastrow et  $\Phi(\mathbf{R})$  est un déterminant de Slater ou une petite combinaison linéaire de déterminants de Slater<sup>1</sup>. Le facteur de Jastrow est généralement de la forme  $J(\mathbf{R}) = e^{\sum_{i<j} f(r_{ij})}$ , c'est-à-dire qu'il dépend explicitement des distances interparticules  $r_{ij}$ , ce qui lui permet de décrire efficacement la corrélation électronique dite « dynamique ».

En pratique, la méthode VMC a deux types d'erreurs :

- une *erreur systématique*, due à l'utilisation d'une fonction d'onde approchée (comme dans toute méthode de chimie quantique),
- une *incertitude statistique*, due à l'échantillonnage de taille  $M$  finie (et spécifique aux méthodes Monte Carlo).

Bien sûr, l'erreur systématique n'est en général pas connue, puisque sa détermination requière la connaissance de la solution exacte. L'incertitude statistique peut au contraire s'estimer simplement par des raisonnements statistiques. Pour cela, réexaminons de façon plus précise la signification de l'équation (2). La moyenne de l'énergie locale  $\langle E_L \rangle$  sur un échantillon fini est

---

1. En QMC, il est pratique d'utiliser des fonctions d'onde pour lesquelles les valeurs des coordonnées de spin ont été fixées, et on peut donc ne considérer que les coordonnées d'espace  $\mathbf{R}$ .

elle-même une variable aléatoire, prenant différentes valeurs sur différents échantillons. Le théorème central limite établit que, si  $E_L(\mathbf{R}_k)$  sont des variables aléatoires *indépendantes* (c'est-à-dire non corrélées) et *identiquement distribuées*, d'espérance mathématique  $E[E_L]$  et de variance  $V[E_L] = E[(E_L - E_v)^2]$  finies, alors, dans la limite d'un échantillon de taille infinie ( $M \rightarrow \infty$ ), la distribution de probabilité de la variable aléatoire  $\langle E_L \rangle$  converge (dans le sens mathématique de la convergence en distribution) vers une loi normale de valeur moyenne  $E[E_L]$  et de variance  $V[E_L]/M$ ,

$$E[\langle E_L \rangle] = E[E_L] = E_v, \quad (4a)$$

$$V[\langle E_L \rangle] = \frac{V[E_L]}{M}. \quad (4b)$$

Ceci signifie que  $\langle E_L \rangle$  est un *estimateur* de  $E_v$  avec une incertitude statistique pouvant être définie par l'écart type de sa loi normale

$$\sigma[\langle E_L \rangle] = \sqrt{V[\langle E_L \rangle]} = \sqrt{\frac{V[E_L]}{M}}. \quad (5)$$

La signification de cet écart type est que la valeur recherchée  $E_v$  se trouve à 68,3% de chance dans l'intervalle  $[\langle E_L \rangle - \sigma, \langle E_L \rangle + \sigma]$ , à 95,5% de chance dans l'intervalle  $[\langle E_L \rangle - 2\sigma, \langle E_L \rangle + 2\sigma]$ , et à 99,7% de chance dans l'intervalle  $[\langle E_L \rangle - 3\sigma, \langle E_L \rangle + 3\sigma]$ . Attention, si la variance  $V[E_L]$  est infinie mais l'espérance  $E[E_L]$  est finie, alors la loi des grands nombres assure la convergence de  $\langle E_L \rangle$  vers  $E[E_L]$  mais avec une incertitude statistique plus difficile à estimer et qui décroît plus lentement que  $1/\sqrt{M}$ .

Il est important de noter que l'incertitude statistique décroît en  $1/\sqrt{M}$  *indépendamment de la dimensionnalité du problème*. Ceci doit être comparé aux méthodes d'intégration numériques déterministes (sur grille) dont la convergence de l'erreur d'intégration se détériore avec la dimensionnalité  $d$  de l'espace. Par exemple, la méthode de Simpson converge en  $1/M^{(4/d)}$  de sorte que, en principe, dès  $d > 8$  les méthodes Monte Carlo sont plus efficaces.

L'incertitude statistique est d'autant plus faible que la variance de l'énergie locale  $V[E_L]$  est faible. Dans la limite où  $\Psi$  est une fonction propre exacte de  $\hat{H}$ , l'énergie locale  $E_L$  devient l'énergie exacte, indépendante de  $\mathbf{R}$ , et donc sa variance  $V[E_L]$  et l'incertitude statistique sur  $\langle E_L \rangle$  s'annulent. Ceci est connu sous le nom de propriété de *zéro variance*. Puisque l'erreur systématique (ou biais) sur l'énergie variationnelle  $\Delta E = E_v - E_0$  (où  $E_0$  est l'énergie exacte) s'annule également dans cette limite, il y a également une propriété de zéro biais.

## 1.2 Estimation de l'incertitude statistique

En pratique, la densité de probabilité  $\rho(\mathbf{R})$  est échantillonnée par l'algorithme de Metropolis-Hastings qui fournit une suite de points  $\mathbf{R}_k$  correctement distribués suivant  $\rho(\mathbf{R})$  mais *séquentiellement corrélés* (c'est-à-dire non indépendants). Ceci vient du fait que chaque point est généré à partir du précédent, et on peut alors définir un *temps d'autocorrélation* donnant une idée du nombre de points minimum moyen séparant deux points non corrélés. Cette corrélation séquentielle doit être prise en compte dans l'application du théorème central limite pour évaluer l'incertitude statistique. Ceci est réalisé par la technique de *l'échantillonnage par blocs*, qui est détaillée ici.

Considérons une suite de  $M$  réalisations  $X_i$  (séquentiellement corrélées) d'une variable aléatoire d'espérance  $E[X]$  et de variance  $V[X]$ . Par exemple,  $X$  peut être l'énergie locale  $E_L$ . On divise cette suite en  $M_b$  blocs consécutifs de  $M_s$  pas chacun. La *moyenne sur un bloc*  $\bar{X}$  est

$$\bar{X} = \frac{1}{M_s} \sum_{i=1}^{M_s} X_i. \quad (6)$$

L'espérance mathématique de  $\bar{X}$  est aussi l'espérance de  $X$ , c'est-à-dire  $E[\bar{X}] = E[X]$ , mais sa variance n'est pas simplement  $V[X]/M_s$  puisque les variables  $X_i$  ne sont pas indépendantes. On peut définir à présent la *moyenne globale*  $\langle \bar{X} \rangle$  sur l'ensemble de l'échantillon par la moyenne sur tous les blocs des moyennes sur chaque bloc  $\bar{X}_k$

$$\langle \bar{X} \rangle = \frac{1}{M_b} \sum_{k=1}^{M_b} \bar{X}_k. \quad (7)$$

La moyenne globale  $\langle \bar{X} \rangle$  est une autre variable aléatoire de même espérance que  $X$ , c'est-à-dire  $E[\langle \bar{X} \rangle] = E[\bar{X}] = E[X]$ . Si la longueur des blocs est grande devant le temps d'autocorrélation alors les moyennes par bloc  $\bar{X}_k$  peuvent être considérées comme indépendantes, et la variance de la moyenne globale est simplement

$$V[\langle \bar{X} \rangle] = \frac{V[\bar{X}]}{M_b}, \quad (8)$$

ce qui conduit à l'incertitude statistique sur  $\langle \bar{X} \rangle$

$$\sigma[\langle \bar{X} \rangle] = \sqrt{V[\langle \bar{X} \rangle]} = \sqrt{\frac{V[\bar{X}]}{M_b}}. \quad (9)$$

En pratique, l'incertitude statistique sur un échantillon fini est calculée par

$$\sigma[\langle \bar{X} \rangle] \approx \sqrt{\frac{1}{M_b - 1} \left( \frac{1}{M_b} \sum_{k=1}^{M_b} \bar{X}_k^2 - \left( \frac{1}{M_b} \sum_{k=1}^{M_b} \bar{X}_k \right)^2 \right)}, \quad (10)$$

où le terme  $M_b - 1$  qui apparaît à la place de  $M_b$  est nécessaire pour avoir un estimateur non biaisé de l'écart type sur l'échantillon.

Pour finir, examinons la variance  $V[\bar{X}]$ . Puisque les variables  $X_i$  ne sont pas indépendantes, le développement de  $V[\bar{X}]$  fait intervenir les covariances des variables

$$V[\bar{X}] = \frac{1}{M_s^2} \sum_{i,j} \text{Cov}[X_i, X_j] = \frac{V[X]}{M_s} + \frac{2}{M_s^2} \sum_{i<j} \text{Cov}[X_i, X_j] = T_c \frac{V[X]}{M_s}, \quad (11)$$

définissant le temps d'autocorrélation de  $X$

$$T_c = 1 + \frac{2}{V[X]M_s} \sum_{i<j} \text{Cov}[X_i, X_j]. \quad (12)$$

Le temps d'autocorrélation est égal à 1 en absence de corrélation entre les variables, c'est-à-dire  $\text{Cov}[X_i, X_j] = 0$  pour  $i \neq j$ , mais peut être grand en présence de corrélation séquentielle. Il est instructif d'exprimer l'incertitude statistique en fonction de  $T_c$

$$\sigma[\langle \bar{X} \rangle] = \sqrt{T_c \frac{V[X]}{M_s M_b}} = \sqrt{T_c \frac{V[X]}{M}}, \quad (13)$$

où  $M = M_s M_b$  est la taille totale de l'échantillon. L'expression (13) permet d'interpréter  $T_c$  comme un facteur donnant le nombre de points effectivement indépendants dans l'échantillon,  $M_{\text{eff}} = M/T_c$ . En pratique, il est utile de calculer le temps d'autocorrélation par  $T_c = M_s V[\bar{X}]/V[X]$  et de vérifier si la longueur des blocs est assez grande pour une estimation correcte de l'incertitude statistique, par exemple  $M_s > 100 T_c$ . Si  $M_s$  n'est pas largement supérieur à  $T_c$ , alors l'incertitude statistique  $\sigma[\langle \bar{X} \rangle]$  et le temps d'autocorrélation  $T_c$  seront sous-estimés.

### 1.3 Coût de calcul

Le temps de calcul nécessaire pour atteindre une incertitude statistique donnée  $\sigma[\langle \bar{X} \rangle]$  est

$$t = t_s M = t_s \frac{T_c V[X]}{\sigma[\langle \bar{X} \rangle]^2} \quad (14)$$

où  $t_s$  est le temps de calcul par itération. La dépendance en  $1/\sigma[\langle \bar{X} \rangle]^2$  implique que diminuer l'incertitude statistique par un facteur 10 nécessite d'augmenter le temps de calcul par un facteur 100. Cette dépendance quadratique provient directement du théorème central limite et semble incontournable. On peut cependant jouer avec les trois autres paramètres :

- $T_c$  dépend de l'algorithme d'échantillonnage et de la variable aléatoire  $X$ . Pour des algorithmes performants comme celui d'Umrigar [3,9], le temps d'autocorrélation de l'énergie locale est proche de 1 et peu d'amélioration semble à présent possible ;
- $t_s$  est habituellement dominé par le coût d'évaluation de  $X$ . Pour l'énergie locale, le coût d'évaluation dépend de la forme de la fonction d'onde ;
- $V[X]$  dépend du choix de la variable aléatoire  $X$  avec sa distribution de probabilité associée, la seule contrainte étant que la valeur moyenne  $E[X]$  doit être égale à la valeur de l'observable que nous voulons calculer (sinon, il s'agit d'un estimateur biaisé). Le choix d'une bonne distribution de probabilité est habituellement appelé *échantillonnage suivant l'importance*. Même avec une distribution de probabilité fixée, nous avons souvent la liberté d'utiliser un estimateur  $X$  de plus faible variance possible. Ceci a été exploité pour construire des estimateurs améliorés pour diverses observables. Il y a parfois un compromis à trouver entre un faible temps de calcul par itération  $t_s$  et une faible variance  $V[X]$ .

### 1.4 Procédure d'échantillonnage

La densité de probabilité,  $\rho(\mathbf{R}) = \Psi(\mathbf{R})^2 / \int d\mathbf{R} \Psi(\mathbf{R})^2$ , est généralement compliquée et ne peut pas être échantillonnée par une méthode directe telle que la méthode de la transformée inverse. On utilise donc une méthode plus générale : l'algorithme de Metropolis-Hastings (ou Metropolis généralisé), utilisant un processus stochastique et plus particulièrement une chaîne de Markov.

#### Processus stochastique

Un *processus stochastique* représente l'évolution – le plus souvent « temporelle » – d'une variable aléatoire. Il est décrit par une trajectoire de points consécutifs  $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M$  avec une loi de probabilité associée  $P(\mathbf{R}_M, \dots, \mathbf{R}_2, \mathbf{R}_1)$ . L'idée d'une évolution temporelle peut être mise en évidence en décomposant la probabilité de la trajectoire complète en produits de la probabilité conditionnelle pour chaque point sachant que les points précédents ont déjà été réalisés. Par exemple, pour  $M = 3$ , la probabilité de la trajectoire est

$$P(\mathbf{R}_3, \mathbf{R}_2, \mathbf{R}_1) = P(\mathbf{R}_3|\mathbf{R}_2, \mathbf{R}_1)P(\mathbf{R}_2|\mathbf{R}_1)P(\mathbf{R}_1). \quad (15)$$

#### Chaîne de Markov

Une *chaîne de Markov* est un processus stochastique pour lequel la probabilité conditionnelle de transition à un nouveau point  $\mathbf{R}_k$  ne dépend que du point précédent  $\mathbf{R}_{k-1}$

$$P(\mathbf{R}_k|\mathbf{R}_{k-1}, \dots, \mathbf{R}_1) = P(\mathbf{R}_k|\mathbf{R}_{k-1}), \quad (16)$$

c'est-à-dire que le processus « oublie » la façon dont il est arrivé au point  $\mathbf{R}_{k-1}$ . La probabilité d'une trajectoire s'écrit donc simplement, par exemple pour  $M = 3$ ,

$$P(\mathbf{R}_3, \mathbf{R}_2, \mathbf{R}_1) = P(\mathbf{R}_3|\mathbf{R}_2)P(\mathbf{R}_2|\mathbf{R}_1)P(\mathbf{R}_1), \quad (17)$$

et  $P(\mathbf{R}_f|\mathbf{R}_i)$  est appelé probabilité de transition d'un point  $\mathbf{R}_i$  à un point  $\mathbf{R}_f$ . Notons que, en général, cette probabilité de transition peut dépendre du temps (mesuré par l'indice  $k$ ). Nous considérerons ici uniquement le cas d'un chaîne de Markov stationnaire pour laquelle la probabilité de transition est indépendante du temps.

Dans la suite, nous utilisons des notations correspondant au cas d'états  $\mathbf{R}_k$  d'un espace continu (« intégrales » au lieu de « sommes »), mais nous ignorons les différences mathématiques parfois subtiles entre le cas continu et discret, et utilisons souvent le vocabulaire du cas discret (par exemple, « matrice »). La probabilité de transition  $P(\mathbf{R}_f|\mathbf{R}_i)$  pourra donc être aussi appelée matrice de transition et a les deux propriétés suivantes :

$$P(\mathbf{R}_f|\mathbf{R}_i) \geq 0 \quad (\text{non négativité}), \quad (18a)$$

$$\int d\mathbf{R}_f P(\mathbf{R}_f|\mathbf{R}_i) = 1 \quad (\text{normalisation en ligne}), \quad (18b)$$

la deuxième propriété exprimant le fait que la probabilité qu'un point  $\mathbf{R}_i$  se trouve à quelque part à l'étape suivante doit être 1. On dit qu'il s'agit une *matrice stochastique*. On peut montrer que toutes ses valeurs propres sont comprises entre 0 et 1, et qu'il y a au moins une valeur propre égale à 1. L'idée pour échantillonner la distribution de probabilité  $\rho(\mathbf{R})$  souhaitée est de construire une chaîne de Markov qui converge vers  $\rho(\mathbf{R})$ . Une condition nécessaire est que la distribution  $\rho(\mathbf{R})$  soit un vecteur propre (à droite) de  $P(\mathbf{R}_f|\mathbf{R}_i)$  avec la valeur propre 1

$$\int d\mathbf{R}_i P(\mathbf{R}_f|\mathbf{R}_i)\rho(\mathbf{R}_i) = \rho(\mathbf{R}_f) = \int d\mathbf{R}_i P(\mathbf{R}_i|\mathbf{R}_f)\rho(\mathbf{R}_f) \quad \forall \mathbf{R}_f, \quad (19)$$

où la deuxième égalité provient simplement de la condition de normalisation (18b). L'équation (19) est une *condition de stationnarité* pour  $\rho(\mathbf{R})$ . Elle signifie que si nous partons de la distribution souhaitée  $\rho(\mathbf{R})$  alors nous continuerons à échantillonner la même distribution en appliquant la chaîne de Markov. Cependant, nous avons besoin de plus que ça. Nous voulons que toute distribution initiale  $\rho_{\text{ini}}(\mathbf{R})$  évolue vers la distribution stationnaire désirée  $\rho(\mathbf{R})$  par applications répétées de la matrice de transition

$$\lim_{M \rightarrow \infty} \int d\mathbf{R}_1 P^M(\mathbf{R}|\mathbf{R}_1)\rho_{\text{ini}}(\mathbf{R}_1) = \lim_{M \rightarrow \infty} \int d\mathbf{R}_1 d\mathbf{R}_2 \dots d\mathbf{R}_M P(\mathbf{R}|\mathbf{R}_M)P(\mathbf{R}_M|\mathbf{R}_{M-1}) \dots P(\mathbf{R}_2|\mathbf{R}_1)\rho_{\text{ini}}(\mathbf{R}_1) = \rho(\mathbf{R}), \quad (20)$$

c'est-à-dire  $\rho(\mathbf{R})$  doit être le vecteur propre dominant de  $P$  (le vecteur propre unique de plus grande valeur propre). Si la condition de stationnarité (19) est satisfaite alors ce sera toujours le cas sauf si  $P$  a plusieurs vecteurs propres de valeur propre 1. On peut montrer que la matrice  $P$  a un seul vecteur propre de valeur propre 1 si  $P$  est une matrice régulière (ou primitive) c'est-à-dire s'il existe un entier  $k \geq 1$  tel que tous les éléments de la matrice  $P^k$  soient strictement positifs,  $P^k(\mathbf{R}_i|\mathbf{R}_j) > 0$ ,  $\forall \mathbf{R}_i, \mathbf{R}_j$ . Cela signifie que, en un certain nombre d'itération  $k$ , toute transition entre n'importe quelle paire d'états  $\mathbf{R}_i$  et  $\mathbf{R}_j$  doit être possible. Ceci assure que tout l'espace des états est bien visité, et que la chaîne de Markov converge vers la distribution stationnaire unique  $\rho(\mathbf{R})$ . La chaîne de Markov est alors dite *ergodique*.

Au lieu d'imposer la condition de stationnarité (19), il est souvent plus facile d'imposer la condition plus forte d'*équilibre local*

$$P(\mathbf{R}_f|\mathbf{R}_i)\rho(\mathbf{R}_i) = P(\mathbf{R}_i|\mathbf{R}_f)\rho(\mathbf{R}_f), \quad (21)$$

qui force le flux de probabilité entre deux états  $\mathbf{R}_i$  et  $\mathbf{R}_f$  d'être le même dans les deux sens. C'est une condition suffisante (mais pas nécessaire) pour que  $\rho(\mathbf{R})$  soit une distribution stationnaire. Une chaîne de Markov qui satisfait la condition (21) est qualifiée de réversible.

Concrètement, une chaîne de Markov est réalisée par une *marche aléatoire*. On part d'un point initial  $\mathbf{R}_1$  (ou marcheur) – c'est-à-dire une distribution  $\delta(\mathbf{R} - \mathbf{R}_1)$  – on calcule le second point  $\mathbf{R}_2$  par tirage dans la distribution de probabilité  $P(\mathbf{R}_2|\mathbf{R}_1)$ , puis le troisième point  $\mathbf{R}_3$  par tirage dans la distribution de probabilité  $P(\mathbf{R}_3|\mathbf{R}_2)$ , et ainsi de suite. Après une phase de convergence (appelé *équilibre*) prenant un certain nombre d'itérations  $M_{\text{eq}}$ , la marche aléatoire échantillonne la distribution stationnaire  $\rho(\mathbf{R})$  dans le sens que  $\rho(\mathbf{R}) \propto \sum_{k>M_{\text{eq}}} \delta(\mathbf{R} - \mathbf{R}_k)$  et les valeurs moyennes des estimateurs des observables souhaitées sont calculées. La vitesse de convergence vers la distribution stationnaire  $\rho(\mathbf{R})$  et le temps d'autocorrélation pour les estimateurs d'observables sont déterminés par la deuxième plus grande valeur propre de la matrice  $P$ . La marche aléatoire doit être suffisamment longue pour avoir plusieurs transitions entre les voisinages de toutes paires d'états ayant une contribution non négligeable à la valeur moyenne. Si les transitions entre deux parties du système sont trop improbables, par exemple entre deux atomes dissociés, alors le risque est que la marche aléatoire reste bloquée sur une partie du système et semble convergée, mais que la véritable distribution stationnaire ne soit pas encore atteinte.

### Algorithme de Metropolis-Hastings

Dans l'algorithme de Metropolis-Hastings, on réalise une chaîne de Markov par la marche aléatoire suivante. Partant d'un point  $\mathbf{R}_i$ , le nouveau point  $\mathbf{R}_f$  est déterminé en deux étapes :

1. un point temporaire  $\mathbf{R}'_f$  est proposé avec la probabilité  $P_{\text{prop}}(\mathbf{R}'_f|\mathbf{R}_i)$ ,
2. le point  $\mathbf{R}'_f$  est accepté (c'est-à-dire  $\mathbf{R}_f = \mathbf{R}'_f$ ) avec la probabilité  $P_{\text{acc}}(\mathbf{R}'_f|\mathbf{R}_i)$ , ou rejeté (c'est-à-dire  $\mathbf{R}_f = \mathbf{R}_i$ ) avec la probabilité  $P_{\text{rej}}(\mathbf{R}'_f|\mathbf{R}_i) = 1 - P_{\text{acc}}(\mathbf{R}'_f|\mathbf{R}_i)$

La probabilité de transition correspondante peut s'écrire

$$P(\mathbf{R}_f|\mathbf{R}_i) = \begin{cases} P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i)P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) & \text{si } \mathbf{R}_f \neq \mathbf{R}_i \\ 1 - \int d\mathbf{R}'_f P_{\text{acc}}(\mathbf{R}'_f|\mathbf{R}_i)P_{\text{prop}}(\mathbf{R}'_f|\mathbf{R}_i) & \text{si } \mathbf{R}_f = \mathbf{R}_i \end{cases} \quad (22)$$

ou, en une seule expression,

$$P(\mathbf{R}_f|\mathbf{R}_i) = P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i)P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) + \left[ 1 - \int d\mathbf{R}'_f P_{\text{acc}}(\mathbf{R}'_f|\mathbf{R}_i)P_{\text{prop}}(\mathbf{R}'_f|\mathbf{R}_i) \right] \delta(\mathbf{R}_i - \mathbf{R}_f). \quad (23)$$

La probabilité de proposition  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  est une matrice stochastique, c'est-à-dire  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) \geq 0$  et  $\int d\mathbf{R}_f P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) = 1$ , ce qui assure que  $P(\mathbf{R}_f|\mathbf{R}_i)$  respecte la condition de non-négativité (18a). Le deuxième terme de l'équation (23) avec la fonction delta assure que  $P(\mathbf{R}_f|\mathbf{R}_i)$  respecte la condition de normalisation (18b). La probabilité d'acceptation est choisie de façon à satisfaire la condition d'équilibre local (21), pour  $\mathbf{R}_f \neq \mathbf{R}_i$ ,

$$\frac{P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i)}{P_{\text{acc}}(\mathbf{R}_i|\mathbf{R}_f)} = \frac{P_{\text{prop}}(\mathbf{R}_i|\mathbf{R}_f)\rho(\mathbf{R}_f)}{P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)\rho(\mathbf{R}_i)}. \quad (24)$$

Plusieurs choix sont possibles. Celui préconisé par Metropolis *et al.* maximise la probabilité d'acceptation

$$P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ 1, \frac{P_{\text{prop}}(\mathbf{R}_i|\mathbf{R}_f)\rho(\mathbf{R}_f)}{P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)\rho(\mathbf{R}_i)} \right\}. \quad (25)$$

Notons que la probabilité d'acceptation n'est pas une matrice stochastique. Une fois la probabilité d'acceptation fixée, il ne reste qu'à choisir la probabilité de proposition  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$ . Notons enfin que, puisque seul le rapport  $\rho(\mathbf{R}_f)/\rho(\mathbf{R}_i)$  intervient dans l'équation (25), il n'est pas nécessaire de calculer la constante de normalisation de la densité de probabilité  $\rho(\mathbf{R})$ .

## Choix de la probabilité de proposition

On a beaucoup de liberté dans le choix de la probabilité de proposition  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$ . Les seules contraintes sont que  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  doit être une matrice stochastique ergodique et qu'il doit être possible d'échantillonner efficacement  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  avec une méthode d'échantillonnage directe. La probabilité de proposition détermine la taille moyenne des déplacements proposés  $\mathbf{R}_i \rightarrow \mathbf{R}_f$  et le taux d'acceptation moyen de ces déplacements. Afin de limiter la corrélation séquentielle, nous devons faire des déplacements les plus grands possibles mais avec un grand taux d'acceptation. En pratique, il y a un compromis à trouver entre la taille moyenne des déplacements proposés et le taux d'acceptation moyen.

Le choix le plus simple pour  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  est une distribution uniforme à l'intérieur d'un petit cube  $\Omega(\mathbf{R}_i)$  centré en  $\mathbf{R}_i$  et de côté  $\Delta$  et une distribution nulle à l'extérieur

$$P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) = \begin{cases} \frac{1}{\Delta^{3N}} & \text{si } \mathbf{R}_f \in \Omega(\mathbf{R}_i) \\ 0 & \text{sinon} . \end{cases} \quad (26)$$

En pratique, un déplacement suivant l'équation (26) est réalisé en faisant

$$\mathbf{R}_f = \mathbf{R}_i + \frac{\Delta}{2} \boldsymbol{\chi}, \quad (27)$$

où  $\boldsymbol{\chi}$  représente un vecteur de  $3N$  nombres aléatoires tirés de la loi uniforme entre  $-1$  et  $1$ . La taille du cube  $\Delta$  peut être ajusté pour minimiser le temps d'autocorrélation, mais celui-ci reste grand et l'échantillonnage est inefficace.

Des choix plus intelligents utilisent l'information sur la distribution  $\Psi(\mathbf{R})^2$  pour guider l'échantillonnage. Un choix pour  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  qui conduirait à des grands déplacements et avec une probabilité d'acceptation égale à 1 serait  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) = \rho(\mathbf{R}_f)$ , indépendamment de  $\mathbf{R}_i$ , mais nous serions alors revenu au problème initial de l'échantillonnage de la distribution compliquée  $\rho(\mathbf{R})$ . Un bon choix pour  $P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$  est la fonction de Green de l'équation de Fokker-Planck dans l'approximation des temps courts

$$P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) = \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\mathbf{R}_f - \mathbf{R}_i - \mathbf{v}(\mathbf{R}_i)\tau)^2}{2\tau}}, \quad (28)$$

où  $\mathbf{v}(\mathbf{R}_i) = \nabla\Psi(\mathbf{R}_i)/\Psi(\mathbf{R}_i)$  est appelé *vitesse de dérive* de la fonction d'onde et  $\tau$  est le pas de temps qui peut être ajusté pour obtenir un temps d'autocorrélation minimal. En pratique, un déplacement suivant l'équation (28) est réalisé en faisant

$$\mathbf{R}_f = \mathbf{R}_i + \mathbf{v}(\mathbf{R}_i)\tau + \boldsymbol{\eta}, \quad (29)$$

où  $\boldsymbol{\eta}$  représente un vecteur de  $3N$  nombres aléatoires tirés de la loi normale de moyenne 0 et d'écart-type  $\sqrt{\tau}$ . Le terme  $\boldsymbol{\eta}$  décrit un processus de diffusion isotrope gaussien (ou processus de Wiener). Le terme  $\mathbf{v}(\mathbf{R}_i)\tau$  est un terme de dérive qui a tendance à déplacer la marche aléatoire vers les régions où la valeur absolue de la fonction d'onde  $|\Psi(\mathbf{R})|$  est élevée. Le pas de temps optimal  $\tau$  est d'autant plus faible que  $\mathbf{v}(\mathbf{R})$  est grand. En pratique, pour un système atomique ou moléculaire, le pas de temps optimal est limité par les régions de l'espace physique où le gradient de la fonction d'onde est le plus grand, c'est-à-dire les régions de cœur. Il est plus efficace de faire des déplacements de tailles différentes pour les électrons de cœur et de valence. Ceci est fait par exemple dans l'algorithme VMC d'Umrigar [3,9] où les déplacements des électrons sont réalisés en coordonnées sphériques et la taille des déplacements radiaux dépend de la distance au noyau le plus proche. Cet algorithme permet d'atteindre le plus souvent un temps d'autocorrélation de l'énergie locale proche de 1.

## Déplacer les électrons tous ensemble ou un par un ?

Dans tout ce qui précède nous avons supposé que, dans le cas d'un système à plusieurs électrons, les électrons sont tous déplacés ensemble puis que ce déplacement est accepté ou rejeté en bloc. En fait, il est aussi possible de déplacer les électrons un par un, c'est-à-dire déplacer le premier électron, accepter ou rejeter ce déplacement, puis déplacer le deuxième électron, accepter ou rejeter ce déplacement, etc. Dans ce cas, la probabilité de transition des  $N$  électrons se décompose formellement suivant

$$P(\mathbf{R}_f|\mathbf{R}_i) = P(\mathbf{r}_{1,f}\mathbf{r}_{2,f}\dots\mathbf{r}_{N,f}|\mathbf{r}_{1,f}\mathbf{r}_{2,f}\dots\mathbf{r}_{N,i}) \times \dots \times P(\mathbf{r}_{1,f}\mathbf{r}_{2,f}\dots\mathbf{r}_{N,i}|\mathbf{r}_{1,f}\mathbf{r}_{2,i}\dots\mathbf{r}_{N,i}) \times P(\mathbf{r}_{1,f}\mathbf{r}_{2,i}\dots\mathbf{r}_{N,i}|\mathbf{r}_{1,i}\mathbf{r}_{2,i}\dots\mathbf{r}_{N,i}), \quad (30)$$

où chaque probabilité de transition à un électron (sachant que les autres électrons sont fixes) consiste en une probabilité de proposition et une probabilité d'acceptation comme précédemment. Si chaque probabilité de transition à un électron satisfait la condition stationnarité (19), alors la probabilité de transition globale la satisfait aussi.

Déplacer les  $N$  électrons un par un demande plus de temps de calcul que déplacer les électrons en une seule fois, puisque la fonction d'onde doit être recalculée à chaque déplacement pour calculer la probabilité d'acceptation. Le temps de calcul n'augmente pas d'un facteur  $N$  comme on pourrait naïvement le penser mais typiquement d'un facteur 2 si la valeur de la fonction d'onde est recalculée intelligemment après le déplacement d'un seul électron. Par exemple, pour un déterminant de Slater, nous pouvons utiliser la formule de Sherman-Morrison pour recalculer rapidement la valeur du déterminant quand une seule ligne ou colonne a été changée. Malgré l'augmentation du temps de calcul, il a été plusieurs fois montré dans la littérature que, pour les systèmes avec beaucoup d'électrons, déplacer les électrons un par un conduisait à un algorithme plus efficace : le taux d'acceptation moyen est plus grand, les points  $\mathbf{R}_k$  sont moins séquentiellement corrélés et donc le temps d'autocorrélation est plus petit (d'un facteur plus grand que celui nécessaire pour compenser le plus grand coût de calcul par itération).

## 2 Monte Carlo diffusionnel

### 2.1 Idée de base

Alors que la méthode VMC est limitée par l'utilisation d'une fonction d'onde approchée  $\Psi$ , l'idée de la méthode DMC est d'obtenir la fonction d'onde exacte  $\Psi_0$  de l'état fondamental du système. Si nous disposons de cette fonction d'onde exacte  $\Psi_0$ , alors l'énergie exacte  $E_0$  associée peut être exprimée par projection, en utilisant la fonction d'onde d'essai  $\Psi$ ,

$$E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi \rangle}{\langle \Psi_0 | \Psi \rangle} = \frac{\int d\mathbf{R} \Psi_0(\mathbf{R}) \Psi(\mathbf{R}) E_L(\mathbf{R})}{\int d\mathbf{R} \Psi_0(\mathbf{R}) \Psi(\mathbf{R})}, \quad (31)$$

puisque  $\Psi_0$  est fonction propre de l'hamiltonien  $\hat{H}$ . L'avantage de la formule de projection (31) est qu'elle ne nécessite pas le calcul de l'action de  $\hat{H}$  sur  $\Psi_0$ . L'intégrale de l'équation (31) met en jeu l'énergie locale de la fonction d'onde essai,  $E_L(\mathbf{R}) = (H\Psi(\mathbf{R}))/\Psi(\mathbf{R})$ , et peut être estimée de manière similaire qu'en VMC en calculant la valeur moyenne de  $E_L(\mathbf{R})$  sur un échantillon de points  $\mathbf{R}_k$  représentant la distribution mixte  $\Psi_0(\mathbf{R})\Psi(\mathbf{R})/\int d\mathbf{R} \Psi_0(\mathbf{R})\Psi(\mathbf{R})$ .

Mais comment obtenir la fonction d'onde exacte  $\Psi_0$  ? Considérons l'action de l'opérateur d'évolution en temps imaginaire ( $t \rightarrow -it$ ) sur une fonction d'onde quelconque comme la fonction d'onde d'essai  $\Psi$

$$|\Psi(t)\rangle = e^{-(\hat{H}-E_T)t}|\Psi\rangle, \quad (32)$$

où  $E_T$  est un décalage d'énergie pour l'instant arbitraire. Utilisant la décomposition spectrale de l'opérateur d'évolution (écrite avec les états propres  $\Psi_i$  et les énergies propres  $E_i$  de  $\hat{H}$ ), nous

voyons que la limite dans un temps d'évolution infini est dominée par l'état  $\Psi_0$  de plus basse énergie ayant un recouvrement non nul avec  $\Psi$

$$\lim_{t \rightarrow \infty} |\Psi(t)\rangle = \lim_{t \rightarrow \infty} \sum_i e^{-(E_i - E_T)t} |\Psi_i\rangle \langle \Psi_i | \Psi \rangle = \lim_{t \rightarrow \infty} e^{-(E_0 - E_T)t} |\Psi_0\rangle \langle \Psi_0 | \Psi \rangle, \quad (33)$$

puisque tous les autres états d'énergies  $E_i > E_0$  donnent des exponentielles décroissant plus rapidement. L'exponentielle  $e^{-(E_0 - E_T)t}$  peut être éliminée en ajustant  $E_T$  à  $E_0$ , et on obtient alors que  $\Psi(t)$  devient proportionnel à  $\Psi_0$

$$\lim_{t \rightarrow \infty} |\Psi(t)\rangle \propto |\Psi_0\rangle. \quad (34)$$

En représentation de position, l'équation (32) s'écrit

$$\Psi(\mathbf{R}_f, t) = \int d\mathbf{R}_i G(\mathbf{R}_f | \mathbf{R}_i; t) \Psi(\mathbf{R}_i), \quad (35)$$

où  $G(\mathbf{R}_f | \mathbf{R}_i; t) = \langle \mathbf{R}_f | e^{-(\hat{H} - E_T)t} | \mathbf{R}_i \rangle$  est appelé *fonction de Green* (ou propagateur de  $\mathbf{R}_i$  à  $\mathbf{R}_f$ ). Multipliant et/ou divisant par  $\Psi(\mathbf{R}_f)$  et  $\Psi(\mathbf{R}_i)$ , nous obtenons l'équation d'évolution de la distribution mixte  $f(\mathbf{R}, t) = \Psi(\mathbf{R}, t) \Psi(\mathbf{R})$

$$f(\mathbf{R}_f, t) = \int d\mathbf{R}_i \tilde{G}(\mathbf{R}_f | \mathbf{R}_i; t) \Psi(\mathbf{R}_i)^2, \quad (36)$$

où  $\tilde{G}(\mathbf{R}_f | \mathbf{R}_i; t)$  est la fonction de Green transformée, dite avec *échantillonnage suivant l'importance*,

$$\tilde{G}(\mathbf{R}_f | \mathbf{R}_i; t) = \Psi(\mathbf{R}_f) G(\mathbf{R}_f | \mathbf{R}_i; t) \frac{1}{\Psi(\mathbf{R}_i)}. \quad (37)$$

Dans la limite d'un temps infini, la distribution mixte devient proportionnelle à la distribution stationnaire souhaitée :  $f(\mathbf{R}) = \lim_{t \rightarrow \infty} f(\mathbf{R}, t) \propto \Psi_0(\mathbf{R}) \Psi(\mathbf{R})$ .

En pratique, une expression analytique de la fonction de Green n'est connue que dans la limite d'un temps d'évolution court,  $\tilde{G}(\mathbf{R}_f | \mathbf{R}_i; \tau)$ , où  $\tau$  est un petit pas de temps, et il est donc nécessaire d'itérer pour obtenir la distribution stationnaire

$$f(\mathbf{R}) = \lim_{M \rightarrow \infty} \int d\mathbf{R}_1 d\mathbf{R}_2 \dots d\mathbf{R}_M \tilde{G}(\mathbf{R} | \mathbf{R}_M; \tau) \tilde{G}(\mathbf{R}_M | \mathbf{R}_{M-1}; \tau) \dots \tilde{G}(\mathbf{R}_2 | \mathbf{R}_1; \tau) \Psi(\mathbf{R}_1)^2. \quad (38)$$

Une approximation aux temps courts pour la fonction de Green est obtenue en appliquant la formule de Trotter-Suzuki,  $e^{-(\hat{T} + \hat{V})\tau} = e^{-\hat{V}\tau/2} e^{-\hat{T}\tau} e^{-\hat{V}\tau/2} + O(\tau^3)$ , où  $\hat{T}$  et  $\hat{V}$  sont ici les opérateurs d'énergie cinétique et potentielle. En représentation de position, cette approximation conduit à l'expression suivante

$$G(\mathbf{R}_f | \mathbf{R}_i; \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\mathbf{R}_f - \mathbf{R}_i)^2}{2\tau}} e^{-\left(\frac{V(\mathbf{R}_f) + V(\mathbf{R}_i)}{2} - E_T\right)\tau}, \quad (39)$$

où  $V(\mathbf{R})$  est l'énergie potentielle. De manière similaire, supposant pour l'instant que la fonction d'onde d'essai est de même signe en  $\mathbf{R}_i$  et  $\mathbf{R}_f$ , c'est-à-dire  $\Psi(\mathbf{R}_f)/\Psi(\mathbf{R}_i) > 0$ , une approximation aux temps courts de la fonction Green avec échantillonnage suivant l'importance est

$$\tilde{G}(\mathbf{R}_f | \mathbf{R}_i; \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\mathbf{R}_f - \mathbf{R}_i - \mathbf{v}(\mathbf{R}_i)\tau)^2}{2\tau}} e^{-\left(\frac{E_L(\mathbf{R}_f) + E_L(\mathbf{R}_i)}{2} - E_T\right)\tau}, \quad (40)$$

où  $\mathbf{v}(\mathbf{R}) = \nabla \Psi(\mathbf{R}) / \Psi(\mathbf{R})$  est la vitesse de dérive de la fonction d'onde qui a été supposée constante entre  $\mathbf{R}_i$  et  $\mathbf{R}_f$ . Cette approximation aux temps courts entraîne une *erreur de pas de temps fini*, qui doit être en principe éliminée en extrapolant les résultats vers  $\tau = 0$ .

## 2.2 Mise en œuvre stochastique

La mise en œuvre stochastique de l'équation (38) est moins triviale que celle réalisée en VMC. En effet, la fonction de Green  $\tilde{G}(\mathbf{R}_f|\mathbf{R}_i; \tau)$  n'est généralement pas une matrice stochastique, puisqu'elle ne conserve pas la normalisation de la densité de probabilité :  $\int d\mathbf{R}_f \tilde{G}(\mathbf{R}_f|\mathbf{R}_i; \tau) \neq 1$ . Nous pouvons néanmoins la factoriser en une matrice stochastique  $P(\mathbf{R}_f|\mathbf{R}_i)$  et une matrice de poids  $W(\mathbf{R}_f|\mathbf{R}_i)$

$$\tilde{G}(\mathbf{R}_f|\mathbf{R}_i; \tau) = P(\mathbf{R}_f|\mathbf{R}_i)W(\mathbf{R}_f|\mathbf{R}_i), \quad (41)$$

où, dans l'approximation aux temps courts,  $P(\mathbf{R}_f|\mathbf{R}_i) = (2\pi\tau)^{-3N/2} e^{-(\mathbf{R}_f - \mathbf{R}_i - \mathbf{v}(\mathbf{R}_i)\tau)^2/2\tau}$  et  $W(\mathbf{R}_f|\mathbf{R}_i) = e^{-((E_L(\mathbf{R}_f) + E_L(\mathbf{R}_i))/2 - E_T)\tau}$ . Une marche aléatoire peut alors être réalisée en accumulant des poids. On pourrait penser procéder de la manière suivante. On part d'un marcheur à une position initiale  $\mathbf{R}_1$  avec un poids  $w_1 = 1$  – c'est-à-dire une distribution  $w_1\delta(\mathbf{R} - \mathbf{R}_1)$  – on calcule la position  $\mathbf{R}_2$  du marcheur à l'itération suivante par tirage dans la distribution de probabilité  $P(\mathbf{R}_2|\mathbf{R}_1)$  [suivant l'équation (29)] et on lui associe le poids  $w_2 = w_1 \times W(\mathbf{R}_2|\mathbf{R}_1)$ , puis la troisième position  $\mathbf{R}_3$  par tirage dans la distribution de probabilité  $P(\mathbf{R}_3|\mathbf{R}_2)$  et le poids  $w_3 = w_2 \times W(\mathbf{R}_3|\mathbf{R}_2)$ , et ainsi de suite. Après la phase d'équilibrage, la marche aléatoire échantillonne la distribution stationnaire  $f(\mathbf{R}) \propto \sum_{k > M_{\text{eq}}} w_k \delta(\mathbf{R} - \mathbf{R}_k)$ . En réalité, cette procédure est terriblement inefficace. En effet, le poids  $w_k$  a tendance à devenir soit de plus en plus petit soit de plus en plus grand au cours des itérations, de sorte que les valeurs moyennes sont dominées par un petit nombre de points de la chaîne de Markov correspondants aux poids les plus grands, et tous les autres points calculés ne contribuent pratiquement pas aux valeurs moyennes.

La solution est d'utiliser un ensemble de  $M_w$  marcheurs effectuant leurs marches aléatoires en parallèle, et de limiter la variation des poids avec un processus de mort/naissance des marcheurs. La variante la plus simple consiste à se débarrasser complètement des poids. Pour cela, à chaque itération chaque marcheur est remplacé par  $M_{\text{copies}}$  copies de lui-même pour l'itération suivante où  $M_{\text{copies}}$  est un entier égal à  $\text{int}[W(\mathbf{R}_f|\mathbf{R}_i)]$  ou à  $\text{int}[W(\mathbf{R}_f|\mathbf{R}_i)] + 1$  et de valeur moyenne  $W(\mathbf{R}_f|\mathbf{R}_i)$ . Mieux, les poids peuvent être conservés, et le processus de mort/naissance des marcheurs utilisé uniquement quand les poids dépassent un seuil inférieur ou supérieur, ce qui limite les fluctuations du nombre de marcheurs. La distribution stationnaire est alors représentée par une somme pondérée sur les itérations et les marcheurs,  $f(\mathbf{R}) \propto \sum_{k > M_{\text{eq}}} \sum_{\alpha=1}^{M_w} w_{\alpha,k} \delta(\mathbf{R} - \mathbf{R}_{\alpha,k})$ , où  $w_{\alpha,k}$  et  $\mathbf{R}_{\alpha,k}$  sont le poids et la position du marcheur  $\alpha$  à l'itération  $k$ . On a intérêt que ce que le terme  $W(\mathbf{R}_f|\mathbf{R}_i)$  fluctue le moins possible, puisque de grandes fluctuations entraînent beaucoup de duplications de marcheurs, donc des marcheurs plus corrélés, et donc un nombre de points statistiquement indépendants plus petit dans l'échantillon.

La valeur de  $E_T$  doit être ajustée au cours des itérations pour éviter l'explosion ou l'extinction de la population de marcheurs. Par exemple, un choix pour  $E_T$  à l'itération  $k+1$  est  $E_T(k+1) = E_0^{\text{est}}(k) - C \log(W_k/W_0)$  où  $E_0^{\text{est}}(k)$  est une estimation de  $E_0$  à l'itération  $k$ ,  $C$  est une constante,  $W_k$  est le poids total de la population de marcheurs et  $W_0$  est le poids total cible. À cause des fluctuations,  $E_T$  varie donc légèrement par rapport à  $E_0$  au cours des itérations, ce qui introduit un biais systématique sur les poids et donc sur la distribution stationnaire  $f(\mathbf{R})$ . En effet, l'ajustement de  $E_T$  rend  $f(\mathbf{R})$  trop petit dans les régions où  $E_L(\mathbf{R}) < E_0$  et trop grand dans les régions où  $E_L(\mathbf{R}) > E_0$ . Ceci est appelé *erreur de contrôle de population*. Cette erreur est généralement petite et décroît en augmentant le nombre de marcheurs  $M_w$ , jusqu'à disparaître dans la limite  $M_w \rightarrow \infty$ . Il est par ailleurs possible d'éliminer la quasi-totalité de cette erreur en introduisant des poids supplémentaires qui annulent le biais introduit par la variation de  $E_T$  sur un certain nombre d'itérations [6].

Dans la limite d'un pas de temps infinitésimal, la matrice de transition  $P(\mathbf{R}_f|\mathbf{R}_i)$  a comme distribution stationnaire  $\Psi(\mathbf{R})^2$ , et le terme des poids  $W(\mathbf{R}_f|\mathbf{R}_i)$  permet de convertir cette distribution en la distribution mixte  $\Psi_0(\mathbf{R})\Psi(\mathbf{R})$ . On peut se débarrasser de l'erreur de pas de temps fini dans la matrice de transition  $P(\mathbf{R}_f|\mathbf{R}_i)$  en introduisant une étape d'acceptation/rejet

comme dans l'algorithme de Metropolis-Hastings. Pour cela, on redéfinit la matrice de transition comme  $P(\mathbf{R}_f|\mathbf{R}_i) = P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i)P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)$ , pour  $\mathbf{R}_i \neq \mathbf{R}_f$ , avec la probabilité de proposition

$$P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i) = \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\mathbf{R}_f - \mathbf{R}_i - \mathbf{v}(\mathbf{R}_i)\tau)^2}{2\tau}}, \quad (42)$$

et la probabilité d'acceptation

$$P_{\text{acc}}(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ 1, \frac{P_{\text{prop}}(\mathbf{R}_i|\mathbf{R}_f)\Psi(\mathbf{R}_f)^2}{P_{\text{prop}}(\mathbf{R}_f|\mathbf{R}_i)\Psi(\mathbf{R}_i)^2} \right\}. \quad (43)$$

Avec cette modification,  $P(\mathbf{R}_f|\mathbf{R}_i)$  a comme distribution stationnaire  $\Psi(\mathbf{R})^2$  même pour un pas de temps fini. Bien sûr, l'erreur de pas de temps fini persiste dans le terme  $W(\mathbf{R}_f|\mathbf{R}_i)$ . Comme certains déplacements sont rejetés,  $P(\mathbf{R}_f|\mathbf{R}_i)$  correspond à présent à un processus de diffusion et dérive avec un pas de temps effectif  $\tau_{\text{eff}} < \tau$ . Ce pas de temps effectif peut être estimé au cours du calcul à partir du taux d'acceptation moyen et il est cohérent de l'utiliser dans le terme  $W(\mathbf{R}_f|\mathbf{R}_i)$  à la place de  $\tau$ . On arrive donc à un algorithme très similaire à un algorithme VMC basé sur la même probabilité de proposition (42), avec seulement des poids en plus.

L'énergie est calculée par valeur moyenne de l'énergie locale sur la distribution  $f(\mathbf{R})/\int d\mathbf{R}f(\mathbf{R})$ . Pour  $M$  itérations (après la phase d'équilibration) et  $M_w$  marcheurs, on a

$$E_0 \approx \langle E_L \rangle_f = \frac{\sum_{k=1}^M \sum_{\alpha=1}^{M_w} w_{\alpha,k} E_L(\mathbf{R}_{\alpha,k})}{\sum_{k=1}^M \sum_{\alpha=1}^{M_w} w_{\alpha,k}}. \quad (44)$$

Tout comme en VMC, il existe un principe de zéro variance de l'énergie en DMC. Dans la limite où la fonction d'onde d'essai  $\Psi$  est une fonction propre exacte de l'hamiltonien, alors  $E_L(\mathbf{R})$  est indépendant de  $\mathbf{R}$ , les poids se réduisent à 1, et la variance sur  $\langle E_L \rangle_f$  s'annule. Notons qu'on peut utiliser la même moyenne pondérée pour une observable  $\hat{O}$  qui ne commute pas avec l'hamiltonien, mais dans ce cas la valeur moyenne sur la distribution mixte  $\langle O_L \rangle_f$  est un estimateur de  $\langle \Psi_0 | \hat{O} | \Psi \rangle / \langle \Psi_0 | \Psi \rangle$  qui n'est qu'une approximation à la valeur exacte  $\langle \Psi_0 | \hat{O} | \Psi_0 \rangle / \langle \Psi_0 | \Psi_0 \rangle$ .

Dans l'équation (40), nous avons supposé que la fonction d'onde d'essai  $\Psi(\mathbf{R})$  est toujours du même signe, c'est-à-dire qu'elle n'a pas de nœuds (points  $\mathbf{R}$  tels que  $\Psi(\mathbf{R}) = 0$ ). C'est le cas pour la fonction d'onde de l'état fondamental d'un système bosonique, ou pour quelques systèmes électroniques simples (deux électrons dans un état de spin singulet, comme dans l'état fondamental de l'atome He ou celui de la molécule  $\text{H}_2$ ). Dans ce cas, l'algorithme présenté ci-dessus permet d'obtenir l'énergie exacte du système, après élimination de l'erreur de pas de temps fini et de l'erreur de contrôle de population. Dans le cas où la fonction d'onde  $\Psi(\mathbf{R})$  à des nœuds, la fonction de Green  $\tilde{G}(\mathbf{R}_f|\mathbf{R}_i;t) = \Psi(\mathbf{R}_f)G(\mathbf{R}_f|\mathbf{R}_i;t)/\Psi(\mathbf{R}_i)$  peut être négative (si  $\Psi(\mathbf{R}_i)$  et  $\Psi(\mathbf{R}_f)$  sont de signes différents), ce qui est nécessaire pour convertir la distribution initiale positive  $\Psi(\mathbf{R})^2$  en la distribution mixte  $\Psi_0(\mathbf{R})\Psi(\mathbf{R})$  qui peut être négative car  $\Psi_0(\mathbf{R})$  n'a généralement pas les mêmes nœuds que  $\Psi(\mathbf{R})$ . On pourrait penser qu'il suffit d'inclure le signe de la distribution dans les poids, mais en réalité le problème est bien plus sérieux. En effet, si la fonction d'onde de l'état fondamental fermionique a des nœuds, alors il existe toujours au moins un état bosonique de plus basse énergie, et le véritable état fondamental de l'hamiltonien est un état bosonique dont la fonction d'onde  $\Psi_B(\mathbf{R})$  peut être choisie strictement positive. Si on appliquait la fonction de Green de manière exacte, la distribution mixte convergerait bien vers  $\Psi_0(\mathbf{R})\Psi(\mathbf{R})$  puisque la fonction d'onde d'essai antisymétrique (par rapport à l'échange de deux électrons) a un recouvrement nul avec tous les états bosoniques qui sont symétriques. Cependant, on applique en fait la fonction de Green par un échantillonnage fini dans l'espace des positions ce qui ne permet pas d'imposer l'antisymétrie. En partant d'une fonction d'onde  $\Psi$  antisymétrique, une petite composante sur  $\Psi_B$  peut donc se développer en appliquant la fonction

de Green et finit par dominer. La distribution tend vers  $\Psi_B(\mathbf{R})\Psi(\mathbf{R})$  et la formule de l'énergie de l'équation (44) ne donne que 0/0 (les poids positifs et négatifs se compensent) avec du bruit statistique. C'est une manifestation du célèbre *problème du signe fermionique*<sup>2</sup>. Pour éviter le problème du signe en DMC, on introduit l'*approximation des nœuds fixés*.

### 2.3 Approximation des nœuds fixés

L'idée de cette approximation est de forcer la convergence vers une fonction d'onde approchant l'état fondamental fermionique en fixant les nœuds à ceux de la fonction d'onde d'essai  $\Psi(\mathbf{R})$ . Formellement, on peut définir un hamiltonien à nœuds fixés (FN),  $\hat{H}_{\text{FN}}$ , en ajoutant au vrai hamiltonien  $\hat{H}$  des barrières de potentiel infinies à la position des nœuds de  $\Psi(\mathbf{R})$ . La fonction d'onde de l'état fondamental de cet hamiltonien est appelée fonction d'onde FN,  $\Psi_{\text{FN}}$ , et son énergie associée est l'énergie FN,  $E_{\text{FN}}$ ,

$$\hat{H}_{\text{FN}}|\Psi_{\text{FN}}\rangle = E_{\text{FN}}|\Psi_{\text{FN}}\rangle. \quad (45)$$

Dans l'espace des positions  $\mathbf{R}$  à  $3N$  dimensions, les nœuds de  $\Psi(\mathbf{R})$  définissent des hypersurfaces de dimension  $3N - 1$ . L'espace des positions est alors partitionné en poches nodales de  $\Psi(\mathbf{R})$ , délimitées par les surfaces nodales, et dans lesquelles la fonction d'onde a un signe fixé. Dans chaque poche nodale, la fonction d'onde FN est la solution de l'équation de Schrödinger satisfaisant à la condition aux limites de s'annuler sur la surface nodale. La fonction de Green FN correspondant à l'hamiltonien  $\hat{H}_{\text{FN}}$  est

$$G_{\text{FN}}(\mathbf{R}_f|\mathbf{R}_i; t) = \langle \mathbf{R}_f | e^{-(\hat{H}_{\text{FN}} - E_T)t} | \mathbf{R}_i \rangle, \quad (46)$$

et ne permet les déplacements  $\mathbf{R}_i \rightarrow \mathbf{R}_f$  qu'à l'intérieur d'une poche nodale. La fonction de Green FN avec échantillonnage suivant l'importance,

$$\tilde{G}_{\text{FN}}(\mathbf{R}_f|\mathbf{R}_i; t) = \Psi(\mathbf{R}_f) G_{\text{FN}}(\mathbf{R}_f|\mathbf{R}_i; t) \frac{1}{\Psi(\mathbf{R}_i)}, \quad (47)$$

confine aussi les déplacements à l'intérieur d'une poche nodale, et est donc toujours positive ou nulle. Une approximation aux temps courts de  $\tilde{G}_{\text{FN}}(\mathbf{R}_f|\mathbf{R}_i; t)$  est alors encore donnée par l'équation (40). L'algorithme stochastique précédemment décrit peut donc être directement appliqué. Grâce à l'approximation FN, les poids restent toujours positifs, et la distribution mixte stationnaire  $f(\mathbf{R})$  est proportionnelle à  $\Psi_{\text{FN}}(\mathbf{R})\Psi(\mathbf{R})$ .

Puisque le gradient de la fonction d'onde d'essai  $\nabla\Psi(\mathbf{R})$  ne s'annule généralement pas à la position des nœuds, la vitesse de dérive  $\mathbf{v}(\mathbf{R}) = \nabla\Psi(\mathbf{R})/\Psi(\mathbf{R})$  diverge sur les nœuds. Pour des pas de temps grands, la fonction de Green de l'équation (40) conduit alors à des déplacements trop grands au voisinage des nœuds qui sont le plus souvent rejetés. Il est possible d'obtenir une meilleure approximation à la fonction de Green proche des nœuds en utilisant une simple modification de la vitesse de dérive décrivant mieux le comportement au voisinage des nœuds [6]. Pour une fonction d'onde  $\Psi(\mathbf{R})$  approchée, l'énergie locale  $E_L(\mathbf{R})$  et donc le terme de poids dans l'équation (40) divergent également sur les nœuds, et il est possible d'améliorer la fonction de Green en modifiant le terme de poids pour mieux décrire le comportement au voisinage des nœuds [6]. Notons enfin que, même s'il a été souvent proposé de rejeter les déplacements de marcheurs traversant les nœuds, en cohérence avec l'approximation FN, ceci n'est néanmoins pas nécessaire. En effet, la vitesse de dérive est perpendiculaire à la surface nodale et éloigne

2. On pourrait imposer l'antisymétrie en échantillonnant dans l'espace des déterminants de Slater au lieu de l'espace des positions. Néanmoins, même dans ce cas, le problème du signe est présent dans le sens où il n'est pas possible en principe de déterminer le signe des coefficients de la fonction d'onde exacte sur les déterminants de Slater, et dans la mise en œuvre la plus simple la moyenne de chaque coefficient tend vers 0.

les électrons de celle ci. Pour un pas de temps suffisamment petit, la probabilité de traverser un nœud est nulle. Pour un pas de temps grand, il est possible de traverser les nœuds mais ceci peut être considéré comme étant inclus dans l’erreur de pas de temps fini.

En appliquant le principe variationnel, il est facile de montrer que l’énergie FN est une borne supérieure à l’énergie exacte

$$E_{\text{FN}} = \frac{\langle \Psi_{\text{FN}} | \hat{H}_{\text{FN}} | \Psi_{\text{FN}} \rangle}{\langle \Psi_{\text{FN}} | \Psi_{\text{FN}} \rangle} = \frac{\langle \Psi_{\text{FN}} | \hat{H} | \Psi_{\text{FN}} \rangle}{\langle \Psi_{\text{FN}} | \Psi_{\text{FN}} \rangle} \geq E_0, \quad (48)$$

la deuxième égalité étant due au fait que les barrières de potentiel infinies dans  $\hat{H}_{\text{FN}}$  ne contribuent pas à la valeur moyenne puisque  $\Psi_{\text{FN}}$  est nulle sur la surface nodale. Puisque la fonction d’onde  $\Psi_{\text{FN}}$  est une fonction propre de  $\hat{H}_{\text{FN}}$ , l’énergie FN peut aussi s’exprimer par la formule de projection

$$E_{\text{FN}} = \frac{\langle \Psi_{\text{FN}} | \hat{H}_{\text{FN}} | \Psi \rangle}{\langle \Psi_{\text{FN}} | \Psi \rangle} = \frac{\langle \Psi_{\text{FN}} | \hat{H} | \Psi \rangle}{\langle \Psi_{\text{FN}} | \Psi \rangle}, \quad (49)$$

où l’hamiltonien  $\hat{H}_{\text{FN}}$  a été remplacé par  $\hat{H}$  pour la même raison que précédemment. En pratique, l’énergie FN est donc obtenue par la même formule (44).

La précision des résultats DMC avec l’approximation des nœuds fixés dépend donc de la qualité de la surface nodale de la fonction d’onde d’essai. Pour une fonction d’onde d’essai à un seul déterminant de Slater, l’erreur due à l’approximation FN peut souvent être significative sur les différences d’énergie. Afin de diminuer cette erreur, on peut utiliser plusieurs déterminants de Slater et optimiser les paramètres de la fonction d’onde (paramètres du facteur de Jastrow, coefficients des déterminants, coefficients des orbitales sur les fonctions de base, exposants des fonctions de base) en VMC (voir les références [10–12]). Cette procédure permet de s’approcher de la précision chimique (1 kcal/mol) en DMC pour les calculs de différences d’énergie comme les énergies d’atomisation de molécules [13].

## Remerciements

J’ai bénéficié de nombreuses discussions avec Cyrus Umrigar (Cornell University, USA) et Roland Assaraf (CNRS et UPMC, Paris) lors de l’écriture de ce document.

## Références

- [1] B. L. Hammond, J. W. A. Lester and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
- [2] M. P. Nightingale and C. J. Umrigar, in *Monte Carlo Methods in Chemistry*, edited by D. M. Ferguson, J. I. Siepmann and D. G. Truhlar (Wiley, NY, 1998), Advances in Chemical Physics Vol. 105, p. Chapter 4.
- [3] C. J. Umrigar, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by M. P. Nightingale and C. J. Umrigar (Kluwer, Dordrecht, 1999), NATO ASI Ser. C 525, p. 129.
- [4] W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- [5] P. J. Reynolds, D. M. Ceperley, B. J. Alder and W. A. Lester, *J. Chem. Phys.* **77**, 5593 (1982).
- [6] C. J. Umrigar, M. P. Nightingale and K. J. Runge, *J. Chem. Phys.* **99**, 2865 (1993).
- [7] B. M. Austin, D. Y. Zubarev and W. A. Lester, *Chem. Rev.* **112**, 263 (2012).
- [8] J. Kolorenč and L. Mitas, *Rep. Prog. Phys.* **74**, 026502 (2011).

- [9] C. J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993).
- [10] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella and R. G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).
- [11] J. Toulouse and C. J. Umrigar, J. Chem. Phys. **126**, 084102 (2007).
- [12] J. Toulouse and C. J. Umrigar, J. Chem. Phys. **128**, 174101 (2008).
- [13] F. R. Petruzielo, J. Toulouse and C. J. Umrigar, J. Chem. Phys. **136**, 124116 (2012).