Sorbonne Université

Ecole Doctorale 388 - Chimie Physique et Chimie Analytique de Paris-Centre

*Laboratoire de Chimie Théorique - équipe « Théorie de la Structure Electronique »*

# Développement d'une méthode Monte Carlo quantique à fragments pour de grands systèmes

Par Antoine BIENVENU

Thèse de doctorat de Chimie Physique

Dirigée par Roland Assaraf et Julien Toulouse

Présentée et soutenue publiquement le 04/07/2022

Devant un jury composé de :

VUILLEMIER Rodolphe, Professeur, Président du Jury

CAFFAREL Michel, Directeur de Recherche, Rapporteur
HOLZMANN Markus, Directeur de Recherche, Rapporteur

EHRLACHER Virginie, Maîtresse de Conférences, Examinatrice
GRIGORI Laura, Directrice de Recherche, Examinatrice
JOURDAIN Benjamin, Professeur, Examinateur

ASSARAF Roland, Chargé de Recherches, Encadrant
TOULOUSE Julien, Maître de Conférences, Directeur de Thèse

# Table des matières

# Résumé

Les méthodes de Monte Carlo sont extrêmement utilisées dans de nombreux domaines, dont la chimie quantique et la physique statistique. En effet, leur capacité à explorer un espace configurationnel sans connaissances préalables les rend très pratiques par rapport à de nombreuses méthodes déterministes. Un des principaux facteurs limitants de leur usage, cependant, est le surcoût causé par la croissance des fluctuations statistiques sur le calcul de propriétés extensives avec la taille du système. En effet, pour un système à $N$ particules, on doit multiplier le scaling [1] habituel en $\mathcal{O}\left(N^{2-3}\right)$ par $\mathcal{O}\left(N\right)$ pour le calcul de l'énergie, et jusqu'à $\mathcal{O}\left(N^3\right)$ pour des quantités plus complexes, comme des fonctions de réponse (polarisabilité) ou la matrice hessienne de l'énergie par rapport à des paramètres variationnels. Nous nous sommes tout particulièrement intéressés à la méthode de Monte Carlo Variationnelle (VMC). Bien qu'étant parmi les plus simples méthodes de Monte Carlo quantique, celle-ci requiert en général l'exploration d'une densité de probabilité $|\psi_T|^2$ et d'une quantité d'intérêt, l'énergie locale, hautement complexes par rapport aux modèles usuels de physique statistique.

Dans cette thèse, nous construisons dans un premier temps une méthode générale pour réduire le scaling des fluctuations pour des quantités d'intérêt extensives locales sans introduire de biais. Cette méthode, que nous appellons Monte Carlo Partitionnelle (PMC), se base sur le principe de "Diviser pour régner", et met en jeu trois idées-clés : la partition de notre système en fragments ; des sous-échantillonnages sur ces fragments à faible coût (avec une méthode de réduction matricielle) ; ainsi qu'un estimateur amélioré se servant de ces sous-échantillonnages. L'intérêt principal de cet estimateur est qu'il est non biaisé, et que dans la limite où les fragments sont indépendants, il est de variance nulle. Nous démontrons de plus que l'estimateur amélioré a une variance systématiquement plus faible que l'estimateur classique. Nous présentons l'application de cette méthode générale à la méthode VMC sur des modèles de Hubbard. On observe des gains importants en efficacité, qui augmentent avec la taille du système. En effet, même avec une fonction d'onde métallique, avec une longueur de corrélation infinie, on gagne de l'ordre de $\mathcal{O}\left(N\right)$ en efficacité pour le calcul de l'énergie variationnelle.

Dans un second temps, nous étendons la méthode PMC de manière théorique au calcul de quantités d'intérêt extensives plus complexes s'exprimant sous forme de dérivées d'une énergie. Ces quantités, qui entrent en jeu dans l'optimisation de la fonction d'onde par rapport à des paramètres, ou dans le calcul de fonctions de réponse, peuvent se mettre sous forme de covariances généralisées (cumulants). Nous commençons en construisant un estimateur amélioré pour le calcul de covariances qui partage la propriété de zéro-variance dans la limite de fragments indépendants. Cet estimateur est ensuite généralisé au calcul de cumulants de tout ordre. Nous présentons ensuite comment ces estimateurs de cumulants peuvent être appliqués au calcul de dérivées de l'énergie variationnelle par rapport aux paramètres de la fonction d'onde.

1. Loi d'échelle de la complexité du calcul

# Introduction

La chimie quantique est un champ de recherche vaste, aux applications variées. Le plus souvent, ces applications mettent en jeu la résolution d'un problème de structure électronique. Ces problèmes sont très souvent ramenés au calcul de la solution de plus basse énergie d'une équation de Schrödinger, aussi connu sous le nom de calcul *ab initio*. Pour ce faire, on dispose de nombreuses méthodes numériques qui, à mesure que s'accroît la capacité de calcul, permettent de s'intéresser à des systèmes de plus en plus grands. Dans ces conditions, le scaling des méthodes desquelles on se sert est un facteur limitant de la taille des systèmes auquels on peut les appliquer.

On dispose pour les calculs *ab initio* de trois grandes familles de méthodes. Dans la première, on résout directement et de manière approchée l'équation de Schrödinger en construisant une fonction d'onde polyélectronique. Parmi ces méthodes, qu'on appellera ci-dessous "méthodes déterministes à fonction d'onde", les plus précises sont également de loin les plus coûteuses, et choisir une méthode de cette famille revient à faire un compromis entre scaling du temps de calcul, et précision. La deuxième famille de méthodes est issue de la théorie de la fonctionnelle de la densité (DFT), et cherche à représenter la répartition des électrons sous la forme d'une densité électronique, plus facile à optimiser. Cependant, un des termes de la fonctionnelle est inconnu et doit être remplacé par une approximation. Ces méthodes déterministes à densité électronique ont donc un faible coût calculatoire, mais il n'y a pas de manière systématique d'augmenter la précision du résultat. Enfin, les méthodes de Monte Carlo quantique emploient une exploration stochastique pour explorer la densité de probabilité issue de la fonction d'onde ; elles pourraient présenter une précision comparable aux méthodes déterministes à fonction d'onde pour un coût plus faible, mais les erreurs statistiques intrinsèques à l'approche Monte Carlo augmentent significativement le coût calculatoire.

Nous nous sommes intéressés tout particulièrement à la méthode de Monte Carlo quantique la plus simple, la méthode de Monte Carlo variationnelle. Dans cette thèse, nous allons vous présenter une nouvelle méthode dérivée de cette dernière, fondée sur le principe de « diviser pour régner », et en se basant sur un principe de zéro-variance. Là où des méthodes précédentes employant l'un ou l'autre de ces principes ont permis des gains intéressants, de manière générale, celles n'utilisant que le premier se fondent sur un équilibre entre un biais systématique et coût de calcul, tandis que des méthodes n'employant que le second réduisaient de manière général la longueur de corrélation ou la variance des estimateurs statistiques, permettant ainsi certes des gains, mais qui ne croissent pas avec la taille du système. La méthode que nous présentons ici se fonde sur un compromis entre la variance et le coût de calcul qui permet un gain croissant avec la taille du système sur lequel on travaille.

Dans une première partie, nous commencerons par fournir une présentation plus poussée de l'approche Monte Carlo et de son intérêt par rapport à des méthodes d'exploration systématique dans des systèmes de forte dimensionalité, ainsi que des différentes méthodes de calcul *ab initio*. Notre seconde partie développera notre nouvelle méthode, en partant de ses aspects les plus théoriques, pour ensuite se tourner vers les astuces d'implémentation qui lui donnent son intérêt, pour enfin fournir les résultats obtenus par cette méthode sur un système modèle simple et peu favorable, le modèle de Hubbard avec une fonction d'onde métallique. Enfin, notre troisième partie explorera comment on peut adapter et étendre notre méthode pour améliorer le calcul de quantités plus complexes, telle la polarisabilité, qui se mettent sous la forme de dérivées ou de covariances.

En annexe, vous pourrez trouver un résumé des notations employées et introduites au cours de cette thèse ; quelques rappels et démonstrations de statistiques mettant en jeu les moments et cumulants, ainsi que les estimateurs classiques de ceux-ci ; quelques démonstrations supplémentaires que nous n'avons pas senti le besoin de développer dans le coeur de la thèse ; ainsi que les textes des deux articles qui sont en cours de publication.

# Première partie

# Contextualisation

# Chapitre 1

# Un bref aperçu de l'approche Monte Carlo

Dans cette partie, nous allons chercher à remettre dans leur contexte les méthodes de Monte Carlo quantique, d'abord au sein de la famille plus vaste des méthodes employant l'approche Monte Carlo, et ensuite au sein des méthodes employées dans les calculs ab initio.

Les méthodes Monte Carlo (MC) forment une vaste famille de méthodes de calcul et de simulation numérique caractérisées par l'emploi de variables aléatoires. Si les premières idées à ce sujet datent du XVIIIe siècle, avec l'aiguille de Buffon et les travaux de Laplace, elles ne furent vraiment employées et théorisées qu'à partir des années 1930-1940, avec l'article de Metropolis et Ulam [1]. Cette famille de méthodes s'est ensuite rapidement développée grâce à ses applications dans de nombreux domaines – notamment en physique, chimie, ingénierie et finance – pour résoudre des problèmes d'optimisation ou d'intégration.

(Les notations utilisées sont rappelées dans l'annexe A.)

## 1.1 Optimisation en Monte Carlo

Les problèmes d'optimisation abordés à l'aide de méthodes Monte Carlo peuvent généralement être ramenés à un niveau ou à un autre à la recherche d'extrema d'une fonction dans un espace configurationnel. Soit donc un espace configurationnel $\Omega$ de dimension $d$, qu'on assimilera à un domaine de $\mathbb{R}^d$. L'emploi d'une méthode simple, où on utilise une grille de points comportant $M$ points dans chaque dimension, a bien évidemment un coût qui se comporte en $\mathcal{O}\left(M^d\right)$. De plus, pour peu que l'espace configurationnel soit suffisamment large, on reste bien souvent éloigné du point optimal.

Or, pour un système physique à $N$ particules, le nombre de degré de libertés, est donc de dimensions à explorer, est donné par $d = \mathcal{O}(N)$. On se retrouve alors avec un scaling exponentiel du coût par rapport au nombre de particules. Un tel scaling est généralement impraticable.

Dans ces conditions, on peut décider d'employer une méthode locale, qui utilise les dérivées partielles successives de notre fonction. L'exemple typique est la méthode d'optimisation de Newton, qui requiert le calcul de l'inverse de la matrice hessienne. Cela résulte en un coût de calcul en $\mathcal{O}\left(d^3\right)$. Cependant, si on est garanti avec ces méthodes d'arriver à un extremum, la dynamique d'exploration est en général attirée et piégée par l'extremum local le plus proche, et non l'extremum global. Combiner cette méthode avec une recherche par grille peut mener à l'extremum global avec plus de précision, au coût d'un scaling bien plus élevé ; ainsi, prendre une grille de $m$ points dans chaque direction nous donne un scaling en $\mathcal{O}\left(d^3 m^d\right)$.

Il faut donc une méthode qui garde un scaling raisonnable, qui puisse être attirée mais non piégée par des extrema. En introduisant un élément de marche aléatoire dans une méthode locale, on transforme celle-ci en méthode d'optimisation Monte Carlo à exploration stochastique (voir [2], chap. 5, pp 157-173). On peut ensuite améliorer ces méthodes soit en jouant sur l'amplitude du pas de la promenade au cours du temps, ce qui donne une méthode dite de "Simulated Annealing" (ou, en français, "Recuit simulé" ; voir [2], pp 163-169 en particulier), soit en ajoutant une étape d'acceptation ou de rejet.

Ces méthodes sont notamment employées pour la construction et l'optimisation de modèles complexes. Des méthodes avancées d'optimisation Monte Carlo peuvent ainsi servir dans les méthodes d'apprentissage informatique (deep learning).

## 1.2 Intégration en Monte Carlo

Le calcul de la valeur d'une intégrale d'une fonction, qu'elle soit propre ou impropre, est un problème complexe qui ne se prête que rarement à une résolution analytique. Dans le cas général, il est nécessaire d'employer des méthodes numériques.

Soit $\Omega$ un domaine de $\mathbb{R}^d$, d'hypervolume $V = |\Omega| > 0$ qu'on supposera dans un premier temps fini ; et soit $f \in \mathcal{L}^1(\Omega, \mathbb{R})$ une application définie sur $\Omega$ et intégrable sur celui- ci. Si l'on cherche à calculer de manière numérique et déterministe l'intégrale $I = \int_\Omega f$, on se heurte bien évidemment à des problèmes de dimensionnalité. En effet, une méthode de type Riemann montre assez vite ses limites, pour les mêmes raisons que l'optimisation ; mais même des méthodes plus élaborées – comme la méthode de Simpson, qui requiert une application quatre fois différentiable – donnent une erreur d'intégration au mieux en $\mathcal{O}\left(M^{-4/d}\right)$ pour $M$ points d'intégration. Cela requiert dans tous les cas un nombre de points qui augmente de manière exponentielle avec la dimensionnalité $d$ de l'espace d'intégration, et indirectement, le nombre de degrés du liberté du système pour lequel on intègre.

Cependant, tant que l'hypervolume $V$ est fini, il est possible de générer des configurations (une suite $(\mathcal{R}_M) \in \Omega^{\mathbb{N}}$ de configurations indépendantes) de manière aléatoire à partir de la distribution uniforme sur $\Omega$. On peut les utiliser pour calculer la valeur de l'intégrale. En effet, puisque $f$ est intégrable, la variable aléatoire $f(\mathcal{R})$ admet une espérance mathématique. On peut alors appliquer la loi faible des grands nombres, qui nous donne le résultat suivant :

$$\forall \varepsilon > 0,\ \lim_{M\to\infty} P\left(\left(\left(\frac{V}{M}\sum_{i=1}^{M} f[\mathcal{R}(i)]\right) - I\right) \geq \varepsilon\right) = 0\ ; \tag{1.1}$$

duquel on tire,

$$\lim_{M\to\infty}\left(\frac{1}{M}\sum_{i=1}^{M} f[\mathcal{R}_i]\right) = \bar{f} = \frac{\int_\Omega f}{\int_\Omega 1} = \frac{I}{V}\ . \tag{1.2}$$

L'équation (1.1) est en fait un résultat plus fort que l'équation (1.2), puisqu'elle exprime que la suite des valeurs moyennes ne peut converger vers d'autre valeur que celle de l'espérance mathématique de la variable aléatoire $f(\mathcal{R})$. Il s'agit d'une propriété connue en statistiques sous le nom de convergence en probabilité.

Si l'on suppose en plus que $f$ est de carré intégrable – c'est à dire que $f \in \mathcal{L}^2(\Omega, \mathbb{R})$, ou, de manière équivalente, $\int_\Omega f^2$ est finie (et vaut $V\overline{f^2}$)– on peut employer des propriétés plus fortes. La première est bien évidemment la loi forte des grands nombres, qui nous dit que la suite de nos moyennes partielles converge presque sûrement vers $\bar{f}$ :

$$P\left(\lim_{M\to\infty}\left(\frac{V}{M}\sum_{i=1}^{M} f[\mathcal{R}(i)]\right) = I\right) = 1\ . \tag{1.3}$$

Une description intéressante de cette équation est que l'on dit que l'estimateur $\frac{V}{M}\sum_{i=1}^{M} f[\mathcal{R}(i)]$ est fortement convergent. Il s'agit d'une terminologie vers laquelle on retournera. La deuxième propriété d'intérêt, et la plus importante, est le théorème central limite. Celui-ci nous fournit la convergence de la loi de probabilité à laquelle est soumise la valeur moyenne vers une gaussienne, dont l'écart-type (et

donc l'ordre de grandeur de l'erreur statistique) pour la $M$-ième moyenne partielle est donné par :

$$\varepsilon = V\sqrt{\frac{\overline{f^2} - \overline{f}^2}{M-1}} \ . \tag{1.4}$$

On se retrouve avec un scaling en $\varepsilon = \mathcal{O}\left(M^{-1/2}\right)$, indépendamment de la dimensionnalité. Cela montre donc que les méthodes Monte Carlo sont comparativement plus efficaces lorsque l'on a un grand nombre de degrés de liberté sur lesquels intégrer, ce qui est souvent le cas dans des systèmes physiques et/ou chimiques (voir [2], chapitre 3, pp 79-107). En pratique, bien sûr, la variance dépend de la dimensionalité du système. Cependant, de manière générale, cette dépendance est linéaire ou polynomiale, ce qui mène à un scaling bien inférieur.

Quand on passe sur un domaine d'hypervolume infini, le problème change quelque peu, puisqu'on ne peut plus générer un échantillon de configurations à partir d'une loi uniforme. Cependant, pour peu qu'on puisse générer des configurations à partir d'une loi de probabilité $\rho$ définie sur $\Omega$ ou sur $\mathbb{R}^d$ tout entier, alors on peut la faire apparaître dans l'intégrale :

$$\int_\Omega f = \int_\Omega \frac{f}{\rho}\rho \ . \tag{1.5}$$

On peut alors voir le cas fini comme correspondant à $\rho = 1$. En pratique, cependant, on est passé de calculer la moyenne de $f$ à la moyenne de $f/\rho$. On doit donc étendre les notations précédentes en assimilant $V$ à $\int_\Omega \rho$, $\bar{f}$ à la moyenne de $f/\rho$ sur $\rho$ dans $\Omega$, et $\overline{f^2}$ à la moyenne de $(f/\rho)^2$ sur $\rho$ dans $\Omega$ si celle-ci existe.

Dans certains cas, $\rho$ est imposée par le système. De manière générale, cependant, on peut choisir notre densité $\rho$ pour éviter que $\overline{f^2}$ soit infinie, afin que la convergence soit gouvernée par le théorème central limite. Les méthodes qui portent sur le choix d'une densité $\rho$ optimale sont connues sous le nom de méthodes d'*importance sampling* ("échantillonnage préférentiel" en français, voir [2], pp 90-107). Bien que la densité $\left|f - \bar{f}\right|$ soit optimale du point de vue de la réduction de la variance, elle ne l'est pas toujours d'un point de vue de la facilité à générer de variables aléatoires.

La construction de celles-ci se fait généralement en combinant d'une manière ou d'une autre deux méthodes : la transformation et la réjection. Les méthodes sont présentées ci-dessous sur un exemple à une dimension.

La méthode de génération par transformation consiste à appliquer une bijection à une ou plusieurs variables aléatoires uniformes pour générer des variables soumises à d'autres distributions. En effet, un effort important a été porté dans la construction d'algorithmes informatiques simples capables de créer des variables pseudo-aléatoires de distribution uniforme, et qui sont fonctionnellement indépendantes les unes des autres.

Prenons en exemple la distribution de Cauchy, définie par $p(x) = 1/(1+x^2)$, et supposons qu'on veuille générer une variable aléatoire $x$ à partir de celle-ci en n'ayant accès qu'à une variable aléatoire de distribution uniforme $\zeta \in ]0,1[$. Alors, on peut prendre $x = \tan(\pi\zeta - \frac{\pi}{2})$. D'une manière plus générale, si la densité de probabilité à une dimension $\rho_{1D}$ admet une primitive $G$, alors celle-ci est bijective, et on peut générer à partir de $\rho_{1D}$ en utilisant $x = G^{-1}(\zeta)$.

D'autres méthodes de transformation existent, bien sûr. Pour construire des gaussiennes, par exemple, on peut se servir de l'algorithme de Box-Muller, qui transforme une paire de variables aléatoires uniformes en une paire de variables gaussiennes ; ou bien on peut se servir du théorème central limite et sommer un nombre important de variables aléatoires uniformes successives (par exemple 12). On voit cependant donc que ces méthodes ne peuvent permettre que de générer des coordonnées pour des distributions simples, dont les primitives sont connues. Si on travaille dans un espace multidimensionnel, la plupart des cas, se servir exclusivement de transformation requiert que l'on puisse factoriser cette densité de probabilité en un produit de distributions à faible dimensionnalité à partir desquelles on sait générer des variables aléatoires.

La méthode par rejet consiste à majorer la densité $\rho$ à partir de laquelle on cherche à générer des variables aléatoires par une densité $\rho'$ à un facteur multiplicatif $m$ près. Alors, si on génère une configuration $\mathcal{R}$ à partir de $\rho'$, et que celle-ci est acceptée avec une probabilité $\rho(\mathcal{R})/(m\rho'(\mathcal{R}))$, cela revient à générer à partir de $\rho$. Il est cependant évident qu'en moyenne la probabilité d'acceptation, et donc l'efficacité de la méthode, est en $1/m$. Ainsi, si cela nous permet l'accès à des distributions dont la forme est plus complexe que celles auxquelles la transformation nous donne accès, pour des distributions complexes dans un espace à forte dimensionnalité, l'efficacité de cette méthode décroît exponentiellement avec la taille du système. .

Certains algorithmes utilisent une description grossière de $\rho'$ qu'ils affinent au cours de la simulation à l'aide des valeurs de $\rho$ générées au cours de celle-ci afin que l'efficacité de $m$ augmente au cours de la simulation. C'est d'ailleurs l'idée principale derrière l'algorithme ARS (pour les distributions log-concaves) (voir [2], chapitre 2, pp 56-61).

## 1.3 Intégration en Monte Carlo par chaînes de Markov

Jusqu'ici, on s'est principalement intéressé à des méthodes pour générer une famille de variables aléatoires identiques et indépendantes à partir d'une densité de probabilité arbitrairement choisie. Cependant, lorsque la dimensionnalité du système avec lequel on travaille augmente, les méthodes par transformation pure deviennent inutilisables à moins qu'il ne soit possible de factoriser la densité recherchée, et les méthodes par transformation et réjection deviennent de moins en moins efficaces.

Qui plus est, il est très fréquent qu'en physique et en chimie, une partie importante de la forme employée par la densité de probabilité soit imposée par le problème. En physique statistique, on doit travailler avec la distribution de Gibbs ; et en physique et chimie quantiques, le carré du module de la fonction d'onde – $\rho = |\psi|^2 = \psi\psi^*$. Ces densités ne se factorisent pas aisément en dehors de systèmes très simples ou constitués de sous-systèmes indépendents.

C'est pour cela qu'on va s'intéresser à la génération de familles de variables aléatoires $(\mathcal{R}_m)_{m\in\mathbb{N}^*} \in \Omega^{\mathbb{N}^*}$ au moyen de processus stochastiques. Ces processus sont caractérisés par la dépendance de la densité de probabilité à partir de laquelle la $(m+1)$-ième variable $(\mathcal{R}_{m+1})$ envers les valeurs des $m$ variables précédentes. En particulier, les chaînes de Markov (Pour plus de détails, voir [3]) sont définies par la propriété ci-dessous (en utilisant la notation traditionnelle $P(A|B)$ pour la probabilité conditionnelle de $A$ sachant $B$) :

$$P\left(\mathcal{R}_{m+1}|\mathcal{R}_1, ..., \mathcal{R}_m\right) = P\left(\mathcal{R}_{m+1}|\mathcal{R}_m\right) \ . \tag{1.6}$$

En d'autres termes, la densité de probabilité d'une nouvelle variable ne dépend que de la valeur précédente. L'exemple typique de chaîne de Markov est bien évidemment une promenade aléatoire : si $(\vec{u}_m)_{m\in\mathbb{N}^*}$ est une suite de vecteurs aléatoires, identiques et indépendants de $\mathbb{R}^3$, alors il est évident que $(\vec{v}_m = \sum_m^{k=1} \vec{u}_k)_{m\in\mathbb{N}^*}$ est une chaîne de Markov, définie par la relation $\vec{v}_{m+1} = \vec{v}_m + \vec{u}_{m+1}$.

De manière générale, une chaîne de Markov est générée à l'aide d'un algorithme de génération de déplacements qui, dans un univers discret, prend la forme d'une matrice réelle positive stochastique par colonnes $\boldsymbol{T} \in \mathcal{M}_\Omega([0,1])$ telle que l'élément de matrice $\boldsymbol{T}_{\mathcal{R}',\mathcal{R}}$ corresponde à la probabilité conditionnelle, ou probabilité de transition, $P(\mathcal{R}'|\mathcal{R})$. Afin de retenir l'aspect matriciel et l'aspect de probabilité conditionnel, on utilisera la notation $\boldsymbol{T}(\mathcal{R}'|\mathcal{R})$. Bien qu'on utilise des notations matricielles discrètes, il va sans dire que les applications et propriétés sont les mêmes – mutatis mutandis – dans un univers continu.

On dira qu'une densité de probabilité $\rho \in [0,1]^\Omega$ ($\mathcal{L}_1(\Omega, \mathbb{R}^+)$ dans le cas continu) est une densité stationnaire de la matrice de transition, si pour toute variable $\mathcal{R}$ générée par celle-ci, les flux entrants et sortants de cette densité s'équilibrent :

$$\forall \mathcal{R} \in \Omega, \sum_{\mathcal{R}'\in\Omega} \boldsymbol{T}(\mathcal{R}|\mathcal{R}')\rho(\mathcal{R}') = \sum_{\mathcal{R}'\in\Omega} \boldsymbol{T}(\mathcal{R}'|\mathcal{R})\rho(\mathcal{R}) = \rho(\mathcal{R}) \ . \tag{1.7}$$

On peut montrer que toute distribution initiale converge vers la distribution stationnaire si et seulement si les deux conditions suivantes sont remplies : d'une part, l'algorithme de transition est ergodique, autrement dit s'il existe un moyen de voyager de n'importe quelle configuration de densité non nulle à n'importe quelle autre - ce qui, en termes mathématiques, revient à imposer qu'il existe un entier $n \in \mathbb{N}$ tel que $\boldsymbol{T}^n$ soit strictement positive ; et d'autre part que $\rho$ est la seule densité stationaire de la matrice de transition, et toute autre densité vecteur propre de celle-ci a une valeur propre plus petite que 1. Comme toute matrice est trigonalisable dans $\mathbb{C}$, cela se ramène à l'expression suivante :

$$\forall \sigma \in \mathbb{C}^{\Omega}, \, (\exists \lambda \in \mathbb{C}, \, \boldsymbol{T}\sigma = \lambda\sigma) \Rightarrow [\sigma \in \mathbb{C}\rho \text{ ou } |\lambda| < 1] \ . \tag{1.8}$$

Afin de construire les algorithmes Monte Carlo, on préfère souvent imposer une condition plus stricte sur les flux : la condition d'équilibre détaillé, qui impose que le flux entre toute paire de configurations soit nul :

$$\forall (\mathcal{R}, \mathcal{R}') \in \Omega^2, \ \boldsymbol{T}(\mathcal{R}'|\mathcal{R})\rho(\mathcal{R}) = \boldsymbol{T}(\mathcal{R}|\mathcal{R}')\rho(\mathcal{R}') \ . \tag{1.9}$$

Cela permet de se servir de la chaîne de Markov comme d'une suite de configurations identiques et indépendentes pour les propriétés de convergence, ce qui inclut la loi forte des grands nombres et le théorème central limite. Qui plus est, la condition d'équilibre détaillé est très utile lorsqu'il s'agit de construire aisément des algorithmes dont on connaît la densité stationnaire. En effet, elle nous permet d'appliquer le théorème de Rosenbluth, selon lequel l'écart quadratique moyen entre la densité cible $\rho$, et la densité à partir de laquelle on échantillonne vraiment $(\boldsymbol{T}^n\mathcal{R}_0)$ converge de manière monotone vers 0. (pour une démonstration, voir [4], pp 35-38).

La méthode Monte Carlo à chaînes de Markov (MCMC) par excellence est l'algorithme de Metropolis-Hastings. Développé en 1949 par Metropolis et Ulam [1] et étendu en 1953 [5], il est généralisé en 1970 par Wilfred K. Hastings [6] et est constitue de nos jours une des méthodes de simulation numérique les plus employées, toutes familles confondues.

Dans le cadre de cette méthode, la valeur de la matrice de transition est réécrite sous la forme d'un produit entre les termes d'une matrice de proposition $\boldsymbol{P}$, stochastique par colonnes, et d'une matrice d'acceptation $\boldsymbol{A}$ :

$$\begin{aligned} \forall (\mathcal{R} \neq \mathcal{R}') \in \Omega^2, \boldsymbol{T}(\mathcal{R}'|\mathcal{R}) &= \boldsymbol{P}(\mathcal{R}'|\mathcal{R})\boldsymbol{A}(\mathcal{R}'|\mathcal{R}) \\ \forall \mathcal{R} \in \Omega, \boldsymbol{T}(\mathcal{R}|\mathcal{R}) &= \boldsymbol{P}(\mathcal{R}|\mathcal{R}) + \sum_{\substack{\mathcal{R}' \in \Omega \\ \mathcal{R}' \neq \mathcal{R}}} \boldsymbol{P}(\mathcal{R}'|\mathcal{R})(1 - \boldsymbol{A}(\mathcal{R}'|\mathcal{R})) \ . \end{aligned} \tag{1.10}$$

Ces équations montrent bien le procédé sous-jacent à l'algorithme : d'abord on propose un déplacement, puis on lui donne une chance d'être accepté. S'il est accepté, le déplacement a lieu, sinon la configuration est répétée. La probabilité d'acceptation est obtenue grâce à la condition d'équilibre détaillé :

$$\forall (\mathcal{R}', \mathcal{R}) \in \Omega^2, \boldsymbol{A}(\mathcal{R}'|\mathcal{R}) = \min\left(1, \frac{\boldsymbol{P}(\mathcal{R}|\mathcal{R}')\rho(\mathcal{R}')}{\boldsymbol{P}(\mathcal{R}'|\mathcal{R})\rho(\mathcal{R})}\right) \ . \tag{1.11}$$

Cette expression traduit qu'il est facile de se déplacer vers des régions de plus haute densité de probabilité, et difficile vers celle de plus basse densité de probabilité. L'ergodicité dépend ensuite exclusivement du choix de la matrice de proposition.

Cet algorithme dans sa version initiale était donné pour une matrice de proposition symétrique, la généralisation à une matrice de proposition asymétrique provenant des travaux de Wilfred Hastings. Le choix d'une matrice de transition ou de proposition biaisée peut permettre d'augmenter le taux d'acceptation des déplacements, ce qui conduit de manière générale à un algorithme plus efficace en réduisant le temps de corrélation. [1]

1. Choisir de reproduire le gradient logarithmique de $\sqrt{\rho}$ plutôt qu'un déplacement à taux de proposition constant mène à de bien meilleurs taux d'acceptation. Il en est de même pour un déplacement de type dérive-diffusion comme en Monte Carlo diffusionnelle. Il faut cependant faire attention dans ces cas aux points où la densité de probabilité est non dérivable, comme le centre d'une orbitale de type Slater en $\exp(-r)$.

L'algorithme de Metropolis-Hastings est ainsi capable d'échantillonner n'importe quelle densité de probabilité calculable, même non normalisée. Il forme ainsi un outil incroyablement versatile, qui a pu être adapté au-delà des problèmes d'intégration.

En effet, pour des problèmes d'optimisation il suffit de mémoriser la position du plus bas minimum visité lors d'une exploration. La flexibilité énorme du choix de matrice de proposition permet d'intégrer des déplacements parfois surprenants. Ainsi, pour des systèmes tels l'ensemble grand canonique, où volume et nombre de particules sont des paramètres variables, et les hyperparamètres de simulation sont la pression, température, et potentiel chimique, il est possible de proposer des mouvements changeant ces paramètres variables, comme par exemple la création et destruction de particules ou augmentation ou réduction du volume. Cela s'applique également à l'optimisation de modèles, où on peut intégrer des mouvements d'ajout ou de suppression de variables, voire même de passage d'un type de modèle à un autre.

L'algorithme de Metropolis-Hastings a également facilité le développement des méthodes de Monte Carlo quantique (QMC), bien que la généralisation de Hastings de l'algorithme de Metropolis soit arrivée en 1970, soit cinq à six années après les premiers calculs QMC par Kalos [7] et par McMillan [8].

## Récapitulatif

Nous avons rappelé ici comment les méthodes directes d'exploration de propriétés, tant en optimisation qu'en intégration, peinaient lorsque la dimensionnalité de l'espace ou variété à explorer augmente. Or, pour des systèmes physiques, la dimensionalité de l'espace de configuration croît linéairement avec le nombre de particules, ce qui se traduit par un coût en pratique exponentiel pour ces méthodes directes. En comparaison, l'erreur intrinsèque aux méthodes Monte Carlo est beaucoup plus faible et ne croît pas avec la dimensionalité de l'espace à explorer.

Bien sûr, il n'est pas toujours possible de choisir la densité de probabilité à partir de laquelle on doit générer des configurations, c'est d'ailleurs généralement le cas en mécanique quantique. Dans ces conditions, il est possible de se servir de méthodes de Monte Carlo à chaîne de Markov pour échantillonner à partir d'une densité de probabilité complexe, sans pour autant avoir à utiliser un algorithme complexe et avec beaucoup de pertes pour générer des variables aléatoires à partir de cette densité.

### Bibliographie

[1] Nicholas Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247) :335–341, 1949.

[2] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2 edition, 2004.

[3] Julien Toulouse, Roland Assaraf, and Cyrus J. Umrigar. Chapter fifteen - introduction to the variational and diffusion monte carlo methods. In Philip E. Hoggan and Telhat Ozdogan, editors, *Electron Correlation in Molecules – ab initio Beyond Gaussian Quantum Chemistry*, volume 73 of *Advances in Quantum Chemistry*, pages 285–314. Academic Press, 2016.

[4] James E. Gubernatis, Naoki Kawashima, and Philipp Werner. *Quantum Monte Carlo Methods, Algorithms for Lattice Models*. Cambridge University Press, 2 edition, 2016.

[5] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6) :1087–1092, 1953.

[6] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.

[7] M. H. Kalos. Monte carlo calculations of the ground state of three- and four-body nuclei. *Physical Review*, 128 :1791–1795, 1962.

[8] W. L. McMillan. Ground state of liquid $He^4$. *Physical Review*, 138 :A442–A451, 1965.

# Chapitre 2

# Calculs ab-initio et QMC

Les calculs ab-initio forment le type de calcul le plus répandu dans le monde de la simulation quantique, et pour une bonne raison. En effet, un calcul ab-initio cherche à construire une représentation d'un système à partir d'une paramétrisation la plus simple possible. En pratique, cette paramétrisation consiste à définir le hamiltonien de ce système. À partir de ce hamiltonien, une représentation est construite, soit au moyen d'une fonction d'onde, soit d'une densité électronique, de laquelle on peut ensuite extraire les propriétés recherchées - typiquement, l'énergie de l'état fondamental.

Il existe trois grandes familles de méthodes pour les calculs ab-initio que nous présenterons tour à tour : les méthodes de fonction d'onde déterministes intensives en intégrales, les méthodes à densité électronique, et enfin les méthodes de Monte Carlo Quantique (QMC).

Pour un rappel des notations introduites à ce chapitre, voir l'annexe A.

## 2.1 Equation de Schrödinger et problème ab-initio

Soit un système $\mathcal{S}$ contenant $N$ particules dans un espace physique $\omega$, de positions généralisées $\vec{r}_1, ..., \vec{r}_N$. $\omega$ est un espace physique quelconque : il peut s'agir d'un domaine de $\mathbb{R}^3$, de $[[1, L]]^2 \times \{\uparrow, \downarrow\}$... On peut condenser la notation en posant $\Omega = \omega^N$ l'espace configurationnel, et $\mathcal{R} = (\vec{r}_i)_{i \in [[1,N]]} \in \Omega$ une configuration du système. Alors, il existe une fonction $\phi \in \mathcal{F}(\mathbb{R} \times \Omega, \mathbb{C})$ dite fonction d'onde qui remplit les trois caractéristiques suivantes : premièrement, elle est de carré intégrable à tout instant,

$$\forall t \in \mathbb{R},\ (\mathcal{R} \mapsto \phi(t, \mathcal{R})) \in \mathcal{L}^2(\Omega, \mathbb{C})\ ; \tag{2.1}$$

la densité de probabilité de trouver les particules à un instant, dans une configuration donnée, est le carré de son module,

$$\forall (t, \mathcal{R}) \in (\mathbb{R} \times \Omega),\ \rho(t, \mathcal{R}) = |\phi(t, \mathcal{R})|^2\ ; \tag{2.2}$$

et elle répond à l'équation de Schrödinger dite dynamique [1] :

$$\forall (t, \mathcal{R}) \in (\mathbb{R} \times \Omega),\ \boldsymbol{i}\hbar \left(\frac{\partial \phi}{\partial t}\right)(t, \mathcal{R}) = (\hat{H}\phi)(t, \mathcal{R})\ , \tag{2.3}$$

où $\hbar = h/2\pi$ est la constante de Planck réduite, et $\hat{H}$ est l'opérateur hamiltonien, qui décrit l'énergie du système. En pratique, on pose $\phi(t, \mathcal{R}) = \psi(\mathcal{R}) \exp(-\boldsymbol{i}Et/\hbar)$ et on se ramène à l'équation de Schrödinger dite statique, une équation aux valeurs propres :

$$\hat{H}\psi = E\psi\ . \tag{2.4}$$

L'opérateur hamiltonien décrit complètement l'état énergétique du système et peut être séparé en parties correspondant à l'énergie cinétique et à chacun des potentiels, sous la forme $\hat{H} = \hat{T} + V(\mathcal{R})$. Dans cette expression, $\hat{T}$ correspond à l'énergie cinétique et $V$ à l'énergie potentielle. Ainsi, pour le cas d'un proton (1) et d'un électron (2) dans le vide, l'hamiltonien prend la forme suivante :

$$\hat{H} = -\frac{\hbar^2}{2m_1}\vec{\nabla}_1^2 - \frac{\hbar^2}{2m_2}\vec{\nabla}_2^2 - \frac{e^2}{4\pi\varepsilon_0||\vec{r}_1 - \vec{r}_2||}\ . \tag{2.5}$$

Cela donne lieu à une équation de Schrödinger statique correspondante très compliquée, et, sauf à employer des approximations drastiques, comme la théorie de Hückel, l'équation de Schrödinger est impossible à résoudre analytiquement dès que l'on travaille avec plus de deux particules. Même l'emploi de l'approximation de Born-Oppenheimer – fixer les noyaux pour se ramener à un système électronique - ne suffit généralement pas à simplifier suffisamment le système. On se retrouve alors avec trois pistes : partir d'une approximation de champ moyen et corriger ensuite, avec les méthodes déterministes à fonction d'onde ; essayer de réécrire le problème d'une manière complètement différente, avec la théorie de la fonctionnelle de la densité ; ou bien utiliser une méthode stochastique à fonction d'onde, c'est-à-dire de Monte Carlo Quantique.

## 2.2 Méthodes déterministes à fonction d'onde

La méthode la plus intuitive consiste à chercher à construire une fonction d'onde approximative que l'on optimise. Une fois ramené à un système électronique au moyen de l'approximation de Born Oppenheimer, on peut d'abord commencer par restreindre l'espace de recherche. On définit alors l'espace $\mathcal{L}^2_a(\Omega,\mathbb{C})$ comme l'espace des fonctions de carré intégrable antisymétriques ci-dessous :

$$\mathcal{L}^2_a(\Omega,\mathbb{C}) = \left\{\psi \in \mathcal{L}^2(\Omega,\mathbb{C}) / \forall \mathcal{R} \in \Omega, (N \geq i > j \geq 1), \psi(...,\vec{r}_i,...,\vec{r}_j,...) = -\psi(...,\vec{r}_j,...,\vec{r}_i,...)\right\} . \tag{2.6}$$

Lorsqu'on travaille dans un espace $\Omega$ borné (par exemple, avec une particule dans une boite) on doit de plus ajouter des conditions sur $\psi$ à la frontière de $\Omega$, $\partial\Omega$, dans la définition de $\mathcal{L}^2_a$, comme par exemple les conditions de Dirichlet ($\forall \mathcal{R} \in \partial\Omega$, $\psi(\mathcal{R}) = 0$). Une fois ceci fait, on réalise deux approximations.

La première consiste à chercher à factoriser notre fonction en un produit de Hartree. Dans notre cas antisymétrique, cela revient à écrire notre fonction d'onde polyélectronique $\psi_S$ sous la forme d'un déterminant de Slater de fonctions d'onde monoélectroniques $\chi_i$ :

$$\psi_S(\mathcal{R}) \approx \frac{1}{\sqrt{N!}} \det\left((\chi_i(\vec{r}_j))_{(i,j)\in[[1,N]]^2}\right) . \tag{2.7}$$

La seconde est l'approximation des orbitales moléculaires, ou approximation LCAO, qui revient à projeter les fonctions d'onde monoélectroniques (ou orbitales moléculaires) $\chi_i$ sur une famille libre d'orbitales atomiques $\varphi_k$ :

$$\chi_i \approx \sum_k c_{ik}\varphi_k . \tag{2.8}$$

Quand cette famille libre est une base complète de $\mathcal{L}^2(\omega,\mathbb{C})$, alors cette dernière approximation devient exacte. Sinon, cette famille génère un espace, qu'on notera $\mathcal{L}^2_v(\omega,\mathbb{C})$ et qu'on appellera espace variationnel monoélectronique, auquel on restreint les fonctions d'onde monoélectroniques $\chi_i$.

Le terme "variationnel" provient du principe variationnel, qui précise que la fonctionnelle énergie variationnelle $E_v$ a pour minimum absolu la fonction d'onde de l'état fondamental ; et qu'il suffit donc de minimiser la valeur de cette fonctionnelle pour approcher la fonction d'onde de l'état fondamental. La fonctionnelle $E_v$ est définie par :

$$E_v : \psi \in \mathcal{L}^2_a(\Omega,\mathbb{C}) \mapsto \frac{\left\langle \psi \middle| \hat{H} \middle| \psi \right\rangle}{\langle\psi|\psi\rangle} = \frac{\int_\Omega \psi^*(\hat{H}\psi)}{\int_\Omega |\psi|^2} . \tag{2.9}$$

Le principe variationnel se démontre par décomposition de $\psi$ sur la base orthonormale des fonctions propres $(\psi_i)$ du hamiltonien – autrement dit, la base des états du système – sous la forme $\psi = \sum_i c_i\psi_i$, en supposant que celle-ci est discrète. Si on insère cette décomposition dans l'expression de notre fonctionnelle, on obtient :

$$E_v\left(\sum_i c_i\psi_i\right) = \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} = E_0 + \frac{\sum_{i>0} |c_i|^2 (E_i - E_0)}{\sum_i |c_i|^2} , \tag{2.10}$$

où $i = 0$ correspond à l'état fondamental, et $E_i$ à l'énergie de l'état $\psi_i$. Il est donc évident que $c_0 \neq 0$ et $(c_i)_{i>0} = 0$, autrement dit $\psi = \lambda\psi_0$, $\lambda \in \mathbb{C}^*$, minimise l'énergie variationnelle.

La forme la plus simple du principe variationnel se met ainsi sous la forme suivante :

$$E_0 = \min_{\substack{\psi \in \mathcal{L}^2_a(\Omega,\mathbb{C}) \\ ||\psi||=1}} \left\langle \psi \left| \hat{H} \right| \psi \right\rangle \ . \tag{2.11}$$

En pratique, les deux approximations qu'on a présentées reviennent à travailler dans l'espace variationnel :

$$E_{HF} = \min_{\substack{\chi_i \in \mathcal{L}^2_v(\omega,\mathbb{C}) \\ \langle\chi_i|\chi_j\rangle = \delta_{ij}}} \left\langle \psi_S \left| \hat{H} \right| \psi_S \right\rangle \geq E_0 \ . \tag{2.12}$$

L'emploi de multiplicateurs de Lagrange conduit à l'écriture du Lagrangien suivant dont on cherche les valeurs extrémales :

$$\mathcal{L}[(\chi_i)_{i\in[[1,N]]}] = \left\langle \psi \left| \hat{H} \right| \psi \right\rangle - \sum_{i=1}^{N} \varepsilon_i \langle \chi_i|\chi_i\rangle \ ; \tag{2.13}$$

On peut décider de dériver cette équation par rapport aux $\chi_i^*$ pour obtenir $N$ équations de Schrödinger avec des hamiltoniens monoélectroniques de la forme suivante :

$$(\hat{t} + v_{ext} + v_{HF}[\psi])\chi_i = \varepsilon_i \chi_i \ . \tag{2.14}$$

Dans cette équation, $\hat{t}$ est l'opérateur énergie cinétique (continu dans le cas d'un espace réel continu), $v_{ext}$ correspond au potentiel extérieur, et le potentiel $v_{HF}$ une sorte de potentiel moyen auquel est soumis l'électron de la fonction d'onde monoélectronique $\chi_i$. C'est de là que vient le nom de théorie du champ moyen pour cette approche. La résolution approximative de ce système d'équations aux valeurs propres non linéaires de manière itérative est connue sous le nom de méthode du champ auto-cohérent (self-consistent field ou SCF), ou tout simplement méthode Hartree-Fock.

Pour une base de l'espace variationnel comportant $q$ fonctions, la méthode du champ autocohérent coûte en général de l'ordre de $\mathcal{O}\left(q^3\right)$ à $\mathcal{O}\left(q^4\right)$. De plus, le calcul des termes des équations de Fock requiert le calcul d'intégrales à un coût faible. En pratique, ces nécessités font que les bases utilisées se composent de fonctions gaussiennes et de combinaisons linéaires fixées de fonctions gaussiennes, dites "gaussiennes contractées". Ainsi, la base STO-3G consiste d'orbitales de Slater (STO voulant dire Slater-Type Orbital) décrites sous forme d'une somme de trois gaussiennes.

On se retrouve donc avec une méthode qui est limitée par sa base et par les approximations qu'elle emploie. En effet, même dans la limite d'une base infinie, Hartree-Fock demeure strictement au-dessus de la valeur expérimentale de l'énergie. Cependant, il existe de nombreuses méthodes pour aller au-delà, qui utilisent Hartree-Fock comme point de départ. Soit on cherche à améliorer le type de fonction d'onde, soit on cherche à améliorer le traitement avec des méthodes dites perturbatives.

La première approche consiste à utiliser une fonction prenant la forme d'une combinaison linéaire de déterminants de Slater, pour exploiter tout l'espace variationnel antisymétrique $\mathcal{L}^2_v(\Omega,\mathbb{C}) = \bigwedge^N \mathcal{L}^2_v(\omega,\mathbb{C})$. Cette famille de méthodes est connue sous le nom de méthode d'interaction configurationnelle (ou CI). Dans une base contenant $q$ fonctions, cela donne lieu à $\binom{N}{q}$ déterminants. On en tire un coût global en $\mathcal{O}\left(q!\right)$, mais on obtient la valeur optimale de l'énergie variationnelle qu'on peut obtenir avec la base ; et à base complète on retrouve l'énergie exacte de l'état fondamental. On peut choisir d'utiliser une méthode tronquée (par opposition à l'interaction configurationnelle complète, ou Full CI), mais si on réduit le scaling du coût, on perd la cohérence en taille (size-consistency) : on n'obtient pas les mêmes résultats pour la réunion de deux systèmes indépendants que pour les deux séparés, contrairement à la Full CI ou à la méthode Hartree-Fock (si on n'impose pas de conditions liées aux spins).

Les méthodes perturbatives ont un coût, et une réduction du biais, intermédiaires entre les méthodes Hartree-Fock et CI, mais on perd le principe variationnel. Dans cette famille, les méthodes de Møller-Plesset sont issues de l'interprétation de l'énergie Hartree-Fock comme les deux premiers termes d'un développement par la théorie de la perturbation de Rayleigh-Schrödinger, et ont un coût plus faible ;

et on a d'autre part les méthodes de clusters couplés, utilisant un ansatz exponentiel tronqué de la somme des opérateurs qui font sauter un électron d'une orbitale à une autre, qui sont plus coûteuses mais reprennent en partie l'idée de l'interaction configurationnelle. On peut aussi citer dans cette famille les méthodes utilisant la théorie perturbative de Dyson sur la fonction de Green.

Pour plus de détail, voir [2] et [3].

## 2.3 Théorie de la fonctionnelle de la densité

La Théorie de la fonctionnelle de la densité, ou DFT, repose sur l'idée que, d'un point de vue physique, puisque les électrons sont interchangeables, la densité électronique $\rho_e$ définie ci-dessous donne lieu à une description suffisante du système pour la plupart des observables ; et qu'on peut donc tenter d'écrire l'énergie comme une fonctionnelle de la densité électronique :

$$\rho_e(\vec{r}) = N \int_{\omega^{N-1}} \rho(\vec{r}, \vec{r}_2, ..., \vec{r}_N) d\vec{r}_2 ... d\vec{r}_N \ . \tag{2.15}$$

Cette idée a été mise en oeuvre au moyen du théorème de Hohenberg et Kohn, démontré en 1964 [4]. En effet, celui-ci précise que le potentiel extérieur peut théoriquement être obtenu à une constante près à partir de la densité électronique de l'état fondamental. Le potentiel extérieur nous permet alors de reconstruire l'opérateur Hamiltonien, et d'arriver à la fonction d'onde. La densité électronique permet ainsi de reconstruire la fonction d'onde de laquelle elle est issue ; les deux descriptions sont alors équivalentes.

Il faut bien se rendre compte, cependant, que l'on n'a fait que déplacer la complexité. En effet, dans l'approche de fonction d'onde, on travaille avec une fonction d'onde complexe (à $3N$ dimensions) qui est donc difficile à optimiser, mais qui dispose d'une fonctionnelle simple, l'énergie variationnelle. En comparaison, la densité électronique est une quantité bien plus simple et intuitive, étant une fonction positive, intégrable et à seulement 3 dimensions... mais la fonctionnelle est bien plus complexe.

De manière générale, le théorème de Hohenberg et Kohn démontre (mais ne donne pas) l'existence d'une fonctionnelle universelle $F[\rho_e]$ telle que la fonctionnelle énergie se mette sous la forme $E[\rho_e] = F[\rho_e] + \int_\omega \rho_e v_{ext}$. Les travaux de Kohn et Sham en 1965 [5] proposent de séparer l'énergie cinétique sans interaction de la fonctionnelle universelle, laissant l'énergie potentielle interélectronique sous la forme d'une fonctionnelle d'énergie dite de Hartree, et laissant une fonctionnelle dite d'échange-corrélation $E_{xc}[\rho_e]$ pour corriger les erreurs. Cela laisse le problème sous la forme suivante :

$$E_0 = \min_{\substack{\rho_e \in \mathcal{L}^1(\omega, \mathbb{R}^+) \\ \int_\omega \rho_e = N}} \left[ \min_{\psi \to \rho_e} \left\langle \psi \left| \hat{T} \right| \psi \right\rangle + \int_\omega \rho_e(\vec{r}) v_{ext}(\vec{r}) d\vec{r} + \iint_{\omega^2} \frac{\rho_e(\vec{r}) \rho_e(\vec{r}')}{2||\vec{r} - \vec{r}'||} d\vec{r} d\vec{r}' + E_{xc}[\rho_e] \right] \ . \tag{2.16}$$

Dans cette équation, le premier terme correspond à la fonctionnelle d'énergie cinétique sans interactions, le second à la fonctionnelle d'énergie potentielle extérieure, le troisième à la fonctionnelle de Hartree, et le dernier à la fonctionnelle d'échange-corrélation. Ces trois premières fonctionnelles sont connues en pratique et peuvent être exprimées sous forme intégrale. On peut alors tirer leur valeur en se servant des équations de Kohn-Sham pour passer d'une densité électronique à des fonctions d'onde monoélectronique. Cependant, aucune expression analytique exacte utilisable n'est connue pour la fonctionnelle d'échange-corrélation.

C'est pourquoi on est obligé de se reposer sur des approximations pour la valeur de la fonctionnelle d'échange-corrélation. De plus, le calcul d'intégrales mis en jeu lors du développement sur les équations de Kohn-Sham limite à l'utilisation d'un espace variationnel comme pour la méthode du champ auto-cohérent, avec un coût global en $\mathcal{O}\left(q^3\right)$ pour des approximations standard pour une base à $q$ fonctions de l'espace variationnel. De plus, contrairement aux méthodes à fonction d'onde, où on peut atteindre en théorie le résultat exact avec une base complète avec soit la méthode Full CI, soit une méthode perturbative dont le développement aurait été poussé à l'infini, il n'existe pas de méthodes systématiques pour améliorer le résultat en DFT ; surtout qu'un certain nombre de fonctionnelles approchées largement employées sont en réalité semi-empiriques.

Pour une introduction plus poussée aux méthodes DFT, voir [6].

## 2.4 Méthode de Monte Carlo Variationnelle

La méthode de Monte Carlo Variationnelle, ou VMC, repose sur la combinaison de la propriété (2.2) et de l'expression de la fonctionnelle énergie variationnelle (2.9). En effet, la première de ces deux propriétés nous permet d'envisager $\rho = |\psi|^2$ comme une densité de probabilité (si $\psi$ est normalisée) et donc d'assimiler à une variable aléatoire toute fonction de $\mathbb{C}^\Omega$. La seconde nous permet de réinterpréter la fonctionnelle énergie variationnelle comme une moyenne d'une fonction énergie locale $E_l$ ... et donc comme l'espérance mathématique $\boldsymbol{E}_\rho$ de $E_l$ assimilée à une variable aléatoire :

$$E_v(\psi) = \int_\Omega \psi^*(\hat{H}\psi) = \int_\Omega \rho \frac{\hat{H}\psi}{\psi} \equiv \int_\Omega \rho E_l \equiv \boldsymbol{E}_\rho(E_l) \ . \tag{2.17}$$

En utilisant un algorithme de Monte Carlo à chaîne de Markov, on se retrouve donc avec une méthode assez simple d'un point de vue théorique, et qui présente, par rapport aux méthodes précédentes des avantages incontestables.

Pour commencer, la forme que peut prendre la fonction d'onde est beaucoup plus flexible, n'étant plus contrainte par la nécessité de calculer de vastes nombres d'intégrales. Non seulement les fonctions de base de l'espace variationnel monoélectronique ne sont plus restreintes à des gaussiennes ou gaussiennes contractées, mais il devient possible de multiplier le (ou les) déterminant de Slater par un terme supplémentaire symétrique exponentiel, ou facteur de Jastrow, afin d'inclure de façon compacte des effets de corrélation du système dans la fonction d'onde :

$$\psi_{JS}(\mathcal{R}) = e^{\sum\limits_{i\neq j} J(\vec{r}_i,\vec{r}_j)} \det\left((\chi_i(\vec{r}_j))_{i,j}\right) \ , \tag{2.18}$$

où $J : \omega^2 \to \mathbb{R}$ est une fonction symétrique. Un exemple de propriétés physiques du système que l'on peut chercher à intégrer dans le facteur de Jastrow est le comportement là où l'énergie potentielle diverge. En effet, Kato a prouvé en 1957 [7] que, dans un état propre, l'énergie cinétique devait compenser ces divergences de l'énergie potentielle, là où la distance entre deux particules tend vers zéro. Cela résulte en des conditions de pointe ("cusp") dites de Kato, qui ne peuvent être satisfaites par un déterminant de Slater pour une paire d'électrons, mais qui peuvent l'être par l'adjonction d'un facteur de Jastrow à la fonction d'onde.

De plus, comme seul entrent en jeu le calcul de la valeur locale de la fonction d'onde, et de l'énergie locale, le coût calculatoire reste de manière générale de l'ordre de $\mathcal{O}\left(N^3\right)$. On voit donc qu'il est facile d'améliorer les résultats pour un surcoût assez faible.

Cependant, la méthode VMC n'est pas une panacée. Le facteur limitant de son usage est provient de l'incertitude numérique d'origine statistique que ne présentent ni les méthodes Hartree-Fock ou post-Hartree- Fock, ni les méthodes DFT, ces deux familles étant toutes deux déterministes. Or, il s'agit d'une incertitude qui s'accroît à mesure que le système augmente en taille. La solution classique consiste à étendre l'échantillon. Cependant, cela se traduit par une augmentation du coût assez significative ; même dans les cas où la convergence est régie par le théorème central limite, le coût réel de la méthode est multiplié par un facteur qui suit le scaling de la variance de la quantité que l'on cherche à calculer. Pour l'énergie, cela représente $\mathcal{O}(N)$, pour ses dérivées $\mathcal{O}\left(N^2\right)$, et pour la matrice hessienne $\mathcal{O}\left(N^3\right)$. Et, lorsqu'on n'a pas accès au théorème central limite (voir les exemples de [8]), la convergence est plus lente encore, et l'impact sur le scaling plus important encore.

L'emploi pour de grands systèmes de la méthode VMC est donc clairement limité par le scaling des fluctuations statistiques de celle-ci. On peut alors à chercher à contourner ce problème, en évitant le calcul de dérivées de l'énergie en échantillonnant indirectement à partir d'une distribution que l'on sait correspondre à un état propre sans la connaître ; ou à le mitiger, en améliorant l'estimation de l'énergie variationnelle.

## 2.5 Méthode de Monte Carlo Diffusionnelle

L'idée derrière la méthode de Monte Carlo Diffusionnelle (ou DMC) consiste à créer un algorithme Monte Carlo à chaîne de Markov qui soit capable d'utiliser l'état fondamental sans même le connaître. Au lieu d'utiliser la fonctionnelle énergie variationnelle, on utilise l'expression suivante de l'énergie exacte $E_0$ de l'état fondamental :

$$E_0 = \frac{\left\langle \psi_0 \middle| \hat{H} \middle| \psi \right\rangle}{\langle \psi_0 | \psi \rangle} = \frac{\int_\Omega \psi_0^* \psi E_l}{\int_\Omega \psi_0^* \psi} \, . \tag{2.19}$$

On rappelle que $\psi_0$ est la fonction d'onde de l'état fondamental, et $\psi$ une fonction d'onde quelconque de recouvrement non nul avec $\psi_0$. On peut, de manière similaire, chercher à tirer d'autres propriétés de l'état fondamental analytique sans connaître l'expression de sa fonction d'onde.

Pour ce faire, supposons que l'on dispose d'un opérateur $\hat{P}$ qui commute avec le Hamiltonien. Alors il partage les vecteurs propres de l'hamiltonien, et on peut écrire ses valeurs propres $\lambda_i$. Alors, si $|\lambda_0| = 1$ et $\forall i > 0$, $|\lambda_i| < 1$, alors on a :

$$\forall \psi \in \mathcal{L}^2(\Omega, \mathbb{C}), \lim_{n \to \infty} \hat{P}^n \psi \in \mathbb{C}\psi_0 \, . \tag{2.20}$$

On peut alors appeler cet opérateur un projecteur, bien qu'il n'en constitue pas un au sens de l'algèbre linéaire. C'est ce qui permet le calcul des propriétés de l'état fondamental. Celui le plus communément employé provient d'une résolution en temps imaginaire de l'équation de Schrödinger dynamique. En effet, si on développe la fonction d'onde à l'instant zéro sur les états propres $\psi_i$ en $\phi(0) = \sum_i c_i \psi_i$, alors on a par l'équation de Schrödinger dynamique :

$$\phi(t) = \sum_i c_i e^{-\boldsymbol{i} t E_i} \psi_i \, , \tag{2.21}$$

avec bien sûr $E_i$ l'énergie associé à l'état de fonction d'onde $\psi_i$. En passant en temps imaginaire (en posant $t = -\boldsymbol{i}\tau$), cela nous donne :

$$\phi(-\boldsymbol{i}\tau) e^{\tau E_0} = \sum_i c_i e^{-\tau(E_i - E_0)} \psi_i \, . \tag{2.22}$$

En pratique, comme les opérateurs d'énergie cinétique et potentielle ne commutent pas, on doit discrétiser dans le temps avec un pas $\delta\tau$ et réaliser une décomposition de Trotter, ce qui résulte bien évidemment en une erreur de discrétisation. De plus, on n'a pas accès à $E_0$ et on doit donc utiliser la valeur d'essai $E_T$, ce qui nous donne le projecteur exponentiel :

$$\hat{P}_{\delta\tau} = e^{\delta\tau(E_T \hat{1} - \hat{H})} \, . \tag{2.23}$$

Pour construire notre matrice de transition, on peut donc partir de l'expression de la fonction de Green (ou propagateur en temps imaginaire) :

$$G_{\delta\tau}(\mathcal{R}'|\mathcal{R}) = \left\langle \mathcal{R}' \middle| \hat{P}_{\delta\tau} \middle| \mathcal{R} \right\rangle \, . \tag{2.24}$$

Comme il s'agit d'un opérateur destiné à faire évoluer $\psi$ vers $\psi_0$, on doit donc le modifier pour obtenir la matrice de transition qui fait converger $|\psi|^2$ vers la "densité mixte" $\psi^* \psi_0$. Cette matrice est aussi connue sous le nom de fonction de Green d'échantillonnage préférentiel (en anglais, "importance sampling Green function") :

$$\tilde{G}_{\delta\tau}(\mathcal{R}'|\mathcal{R}) = \frac{\psi^*(\mathcal{R}')}{\psi^*(\mathcal{R})} G_{\delta\tau}(\mathcal{R}'|\mathcal{R}) \, . \tag{2.25}$$

La décomposition de cette matrice de transition en se servant de la formule de Trotter nous donne un terme de diffusion provenant de l'opérateur énergie cinétique $\hat{T}$ – d'où le nom de Monte Carlo Diffusionnelle – un terme de dérive (issu du gradient du potentiel $V$), et un terme de pondération.

Comme la pondération est multiplicative, la méthode DMC doit alors employer une population de marcheurs – là où la VMC peut le plus souvent se permettre d'en employer un seul – accompagnée de processus de naissance et de mort de marcheurs, afin d'éviter le risque de se faire piéger par quelques configurations de poids important. Cette population est régulée en jouant sur la valeur de l'approximation de l'énergie de l'état fondamental $E_T$ de manière dynamique.

La faiblesse majeure de la méthode DMC provient du fait qu'on travaille de manière générale avec des systèmes électroniques, et donc fermioniques [1]. En effet, les électrons sont soumis à la statistique de Fermi-Dirac. Or, on sait que l'énergie fondamental d'un système de fermions est systématiquement plus haut en énergie que l'état fondamental d'un système de bosons équivalent. Comme la distribution de départ est généralement une distribution de Dirac, la fonction d'onde d'essai "effective" est toujours entâchée d'un peu de fonction d'onde fondamentale bosonique. La méthode DMC – et, de manière plus générale, les autres méthodes de Monte Carlo projectionnelles – font converger tout système vers son état fondamental bosonique.

Pour combattre ce phénomène, un certain nombre d'approches ont été tentées. On peut par exemple partir du principe que l'état fondamental fermionique est plus bas en énergie que le plus bas état excité bosonique, et constitue donc l'état, hors le fondamental bosonique, le plus durable ; et que, par ailleurs, le coefficient initial de l'état fondamental bosonique pour une fonction d'onde de départ antisymétrique est évidemment faible. Cela revient donc à extraire le comportement fermionique du comportement transitoire... mais on doit alors composer bruit qui augmente exponentiellement avec la longueur de simulation et la taille du système par rapport au signal, plutôt que de décroître exponentiellement comme on aurait pu l'espérer en DMC. On peut gagner en précision en plaçant une population à l'état initial sur les zones de fonction d'onde positive, et une autre sur les zones de fonction d'onde négative. La différence des répartitions de ces populations converge vers l'état fondamental fermionique, mais l'amplitude décroît exponentiellement.

Une autre approche provient de l'approximation des surfaces nodales fixées (Fixed Node), introduite par Anderson [9] dans les années 70, et consiste à imposer que la fonction d'onde projetée comporte les mêmes surfaces nodales (hypersurfaces où la fonction d'onde est nulle) que la fonction d'onde de départ, et donc à fixer des zones de signe. Du point de vue de l'hamiltonien, cela revient à placer des barrières de potentiel infini là où se situent les surfaces nodales de la fonction d'onde de départ. L'énergie vers laquelle on converge est alors variationnelle. Pour peu que ces barrières soient bien placées, on peut obtenir la densité mixte recherchée.

Enfin, la méthode de Monte Carlo Fermionique de Kalos [10] permet de gagner en précision en employant des populations de marcheurs signés capables de s'annuler les uns les autres. Cette méthode a été récemment améliorée par Hutcheon dans un récent article [11]. En effet, l'introduction de mouvements d'échange (permutation de particules dans un marcheur) dans la fonction de Green, combinée à la définition dynamique dans l'étape diffusionnelle de surfaces nodales que les marcheurs ne peuvent traverser, permet, sans employer de fonction d'onde initiale, à la DMC de reproduire de manière quasi-exacte le comportement réel des électrons. Il ne s'agit cependant en ce moment que d'une preuve de concept sur quelques atomes.

Les problèmes de variance exponentiellement croissante avec la taille du système en DMC sont connus sous le nom de problème du signe. On rencontre un autre problème du signe, dit dynamique, lorsqu'on travaille avec les états excités d'un système bosonique. Des méthodes ont également été créées pour gérer ce problème, comme dans le cas de l'algorithme inchworm [12].

Pour un peu plus de bases mathématiques en DMC, voir [13], pp 298-312. On voit cependant que toutes ces méthodes dérivées de la Monte Carlo Diffusionnelle cherchent à faire des compromis entre plusieurs formes d'erreur systématique et le temps de calcul.

1. le cas où on apparie les électrons en paires de spin opposé, au comportement bosonique, est exceptionnel

## 2.6 Echantillonnages et estimateurs améliorés

Pour comprendre comment améliorer les méthodes de VMC, nous effectuerons d'abord quelques rappels sur les notions d'estimateur, puis nous présenterons comment on peut s'en servir pour réduire les fluctuations statistiques par diverses méthodes.

### 2.6.1 Rappels sur les estimateurs

Soit $\Omega$ un univers muni de la loi de probabilité $\rho$. On appellera variable aléatoire de $\Omega$ les élements de $\mathbb{C}^{\Omega}$. Soit $\bar{\Omega}$ l'ensemble dont les éléments sont tous les $n$-tuplets d'issues de $\Omega$, quelque soit $n$, et $\kappa$ une variable d'intérêt constante. On appellera alors estimateur toute variable aléatoire de $\bar{\Omega}$ en ce qu'elle sert à calculer la valeur de $\kappa$.

Soit $k$ un estimateur de $\kappa$. On peut définir le biais, la variance et l'erreur quadratique moyenne de $k$ :

$$\begin{aligned} \text{Biais}(k) &= \boldsymbol{E}\left(k-\kappa\right) \ ; \\ \text{Var}\left(k\right) &= \boldsymbol{E}\left(\left|k-\boldsymbol{E}\left(k\right)\right|^2\right) \ ; \\ \text{MSE}(k) &= \boldsymbol{E}\left(\left|k-\kappa\right|^2\right) = \text{Var}\left(k\right) + \left|\text{Biais}(k)\right|^2 \ . \end{aligned} \tag{2.26}$$

Un estimateur est dit non biaisé si il a pour biais 0, et zéro-variant s'il a pour variance 0.

Une suite d'estimateurs $(k_n)_{n\in\mathbb{N}}$ est dite convergente si elle converge en probabilité :

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} P(|k_n - \kappa| > \varepsilon) = 0 \ ; \tag{2.27}$$

et elle est dite fortement convergente si elle converge presque sûrement :

$$P\left(\lim_{n\to\infty} k_n = \kappa\right) = 1 \ . \tag{2.28}$$

Pour finir, on définira la notion d'expression d'estimateur, comme une application qui à la variable $X$ associe un estimateur $k(X)$ de la valeur de la fonctionnelle $\kappa(X)$. Ainsi, les expressions classiques d'estimateurs de l'espérance mathématique sont les éléments de la suite des moyennes partielles :

$$m_n : X \mapsto \frac{1}{n}\sum_{i=1}^{n} X_i \ . \tag{2.29}$$

Il est facile de vérifier que les estimateurs qui en résultent sont non biaisés, et que leur variance est $\text{Var}\left(X\right)/n$. De même, les expressions classiques d'estimateurs de la covariance $\text{Cov}\left(\cdot,\cdot\right)$ sont :

$$c_n : (X,Y) \mapsto m_n(XY) - m_n(X)m_n(Y) \ . \tag{2.30}$$

On vérifie facilement que ces estimateurs ont pour biais $-\text{Cov}\left(X,Y\right)/n$. Pour plus de résultats, nous vous invitons à consulter l'annexe B.

Par la suite, nous travaillerons sur des estimateurs construits à partir de variables aléatoires quelconques, pour manipuler les expressions qu'ils induisent ; et lorsqu'on devra travailler avec des suites d'estimateurs ou d'expressions, on emploiera des notations ne précisant pas le rang de la suite.

### 2.6.2 Méthodes de réduction de la variance

La première méthode de réduction de la variance consiste bien évidemment en l'échantillonnage préférentiel, abordé avec l'équation (1.5). En pratique, cependant, l'échantillonnage préférentiel se

repose sur une autre expression. Si on pose $\rho' = f\rho$, alors on a :

$$\boldsymbol{E}_{\rho}(X) = \frac{\int_{\Omega} X\rho}{\int_{\Omega} \rho} = \frac{\frac{\int_{\Omega} X\rho'/f}{\int_{\Omega} \rho'}}{\frac{\int_{\Omega} \rho'/f}{\int_{\Omega} \rho'}} = \frac{\boldsymbol{E}_{\rho'}(X/f)}{\boldsymbol{E}_{\rho'}(1/f)} \ . \tag{2.31}$$

On peut bien évidemment débattre du choix de $f$. Cependant, il faut prendre en compte qu'il n'est généralement pas plus facile d'échantilloner à partir de $\rho'$, mais que l'on rend le calcul de l'incertitude sur le résultat final significativement plus difficile. Pour obtenir un gain significatif, on doit donc avoir une bonne connaissance de la forme de la fonction d'onde et de la quantité que l'on échantillonne. Cependant, cela peut permettre de se ramener d'un système de variance infinie à un système de variance finie pour peu que $f$ diverge là ou $X$ le fait également.

D'autre part, il est possible de chercher à réduire la variance en changeant d'estimateur. Ainsi, Assaraf et Chevreau [14] ont développé récemment une méthode originale mettant en jeu deux répliques indépendantes d'un système, et une méthode de construction dynamique de clusters en utilisant les relations entre ceux-ci sur la base d'un mouvement d'échange. À partir de cela, ils ont pu construire un estimateur de covariances – et donc de dérivées – non biaisé et dont la variance a un scaling de $\mathcal{O}(N)$, ce qui représente un gain de l'ordre de $\mathcal{O}(N)$.

Une méthode plus simple pour changer l'estimateur nous provient de l'algèbre linéaire. En effet, ajouter à une variable aléatoire $X$ une variable aléatoire $O$ d'espérance nulle ne change pas son espérance mathématique. On peut donc développer la variance d'une combinaison linéaire :

$$\mathrm{Var}(X + \lambda \cdot O) = \mathrm{Var}(X) + 2\lambda \cdot \mathrm{Cov}(X, O) + \lambda^2 \cdot \mathrm{Var}(O) \ . \tag{2.32}$$

Il s'agit d'une fonction quadratique du paramètre $\lambda$, qui admet donc un minimum, et on a :

$$\min_{\lambda} \mathrm{Var}(X + \lambda \cdot O) = \mathrm{Var}(X)\left[1 - \frac{\mathrm{Cov}(X, O)^2}{\mathrm{Var}(O)\,\mathrm{Var}(X)}\right] \ . \tag{2.33}$$

On voit donc qu'on peut introduire des variables d'espérance nulle, ou "variables de contrôle", pour modifier (et réduire) la variance à laquelle est soumis un estimateur sans introduire de biais ; et que la réduction de la variance est d'autant plus important que la variable de contrôle est corrélée. En pratique, le gain n'est vraiment significatif que si la variable de contrôle est fortement corrélée avec $X$.

On peut fort aisément interpréter la méthode des variables de contrôle comme une projection orthogonale, dans l'espace des variables aléatoires et pour le produit scalaire covariance, par rapport à l'espace orthogonal aux variables de contrôles. En effet, si on note $\lambda_0 = -\mathrm{Cov}(X, O)/\mathrm{Var}(O)$, on peut réécrire l'équation (2.33) comme :

$$\mathrm{Var}(X) = \lambda_0^2 \mathrm{Var}(O) + \min_{\lambda} \mathrm{Var}(X + \lambda \cdot O) = \lambda_0^2 \mathrm{Var}(O) + \mathrm{Var}(X + \lambda_0 \cdot O) \ . \tag{2.34}$$

Il s'agit clairement d'une équation de Pythagore, puisque $\mathrm{Cov}(X + \lambda_0 \cdot O, O) = 0$. On peut donc chercher la meilleure direction possible pour projeter. Dans la plupart des cas, le choix de la direction de projection ne permet de ne gagner qu'un préfacteur sur la variance, mais cela n'a pas empêché d'être développées des méthodes à variables de contrôle pour une large variété de variables d'intérêt, notamment les densités électroniques [15] [16], densités de paire [17] [18], forces [19] [20]...

Mes travaux s'inscrivent dans la continuité des méthodes à variable de contrôle. Cependant, le choix de la direction de projection – lié à ce qu'on appellera le principe de zéro-variance – nous permettra d'obtenir des résultats permettant un gain en scaling plutôt qu'en préfacteur.

## Récapitulatif

Après avoir commencé par rappeler la forme du problème *ab initio*, nous avons présenté brièvement les deux grandes familles de méthodes déterministes de calcul *ab initio* : méthodes à fonction d'onde (Hartree-Fock et post-Hartree-Fock) d'une part, et méthodes à densité électronique d'autre part, avec leurs avantages et inconvénients : équilibrage permanent coût/biais avec un coût qui augmente vite dans un cas, biais inconnu à coût faible dans l'autre.

Nous avons ensuite développé les méthodes de Monte Carlo variationnelle et diffusionnelle. Si la seconde est bien plus puissante que la première, elle souffre cependant de problèmes exponentiels provenant du problème du signe fermionique. La méthode VMC, elle, allie les avantages des méthodes de densité électronique – un faible coût, en théorie – avec ceux des méthodes déterministes à fonction d'onde – la capacité à réduire le biais à une valeur arbitrairement basse pour peu qu'on y mette l'effort. Elle souffre cependant d'un problème de la variance qui augmente significativement son coût pratique. Qui plus est, bien qu'il existe des méthodes permettant de réduire la variance, nulle ne permettait, jusqu'ici, de gagner sur le scaling.

### Bibliographie


[1] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28 :1049–1070, 1926.

[2] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory*. Dover, New York, 1996.

[3] T. Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. Wiley, Chichester, 2002.

[4] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136 :B 864, 1964.

[5] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140 :A1133, 1965.

[6] Julien Toulouse. Review of approximations for the exchange-correlation energy in density-functional theory, 2021.

[7] Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2) :151–177, 1957.

[8] J. R. Trail. Heavy-tailed random error in quantum monte carlo. *Physical Review E*, 77 :016703, Jan 2008.

[9] J. B. Anderson. Quantum chemistry by random walk. H $^2P$, $H_3^+$ $D_{3h}$ $^1A_1'$, $H_2$ $^3\Sigma_u^+$, $H_4$ $^1\Sigma_g^+$, Be $^1S$. *Journal of Chemical Physics*, 65 :4121, 1976.

[10] M. H. Kalos and Francesco Pederiva. Exact monte carlo method for continuum fermion systems. *Phys. Rev. Lett.*, 85 :3547–3551, Oct 2000.

[11] Michael Hutcheon. Stochastic nodal surfaces in quantum monte carlo calculations. *Physical Review E*, 102 :042105, 2020.

[12] Zhenning Cai, Jianfeng Lu, and Siyao Yang. Inchworm monte carlo method for open quantum systems. *Communications on Pure and Applied Mathematics*, 73(11) :2430–2472, 2020.

[13] Julien Toulouse, Roland Assaraf, and Cyrus J. Umrigar. Chapter fifteen - introduction to the variational and diffusion monte carlo methods. In Philip E. Hoggan and Telhat Ozdogan, editors, *Electron Correlation in Molecules – ab initio Beyond Gaussian Quantum Chemistry*, volume 73 of *Advances in Quantum Chemistry*, pages 285–314. Academic Press, 2016.

[14] Roland Assaraf and Hilaire Chevreau. en preparation.

[15] Roland Assaraf and Michel Caffarel. Zero-variance principle for monte carlo algorithms. *Physical Review Letters*, 83 :4682–4685, 1999.

[16] R. Assaraf, M. Caffarel, and A. Scemama. Improved Monte Carlo estimators for the one-body density. *Physical Review E*, 75 :035701(R), 2007.

[17] J. Toulouse, R. Assaraf, and C. J. Umrigar. Zero-variance zero-bias quantum Monte Carlo estimators of the spherically and system-averaged pair density. *Journal of Chemical Physics*, 126 :244112, 2007.

[18] Daniel Borgis, Roland Assaraf, Benjamin Rotenberg, and Rodolphe Vuilleumier. Computation of pair distribution functions and three-dimensional densities with a reduced variance principle. *Molecular Physics*, 111(22-23) :3486–3492, 2013.

[19] R. Assaraf and M. Caffarel. Computing forces with quantum Monte Carlo. *Journal of Chemical Physics*, 113 :4028, 2000.

[20] R. Assaraf and M. Caffarel. Zero-variance zero-bias principle for observables in quantum Monte Carlo : Application to forces. *Journal of Chemical Physics*, 119 :10536, 2003.

Deuxième partie

# Développement de la méthode de Monte Carlo Partitionnelle

# Chapitre 3

# Construction de l'estimateur de Monte Carlo Partitionnelle

Dans ce chapitre, nous commencerons par présenter le principe de zéro-variance. Nous développerons ensuite un formalisme pour l'emploi de partitions de notre système en fragments, et d'espérances conditionnelles sur ces fragments. Dans un troisième temps, nous construirons un estimateur théorique nouveau, de variance nulle quand la fonction d'onde est un état propre du système ou quand le système peut être scindé en sous-systèmes indépendants. Dans un quatrième temps, nous montrerons comment cet estimateur théorique est transformé en un estimateur pratique de l'espérance mathématique, l'estimateur de Monte Carlo Partitionnelle (ou PMC) que nous avons développé. Enfin, nous conclurons ce chapitre en calculant la variance de l'estimateur PMC, afin de démontrer qu'il bénéficie bien d'une variance significativement plus faible que l'estimateur traditionnel. Les notations utilisées ici sont développées en annexe A.

## 3.1 Principe de zéro-variance

Dans sa forme la plus simple, le principe de zéro-variance est une forme faible de continuité pour la variance. Lorsque celui-ci s'applique, si un estimateur est zéro-variant dans un cas limite, alors sa variance s'amenuise à mesure que l'on se rapproche de ce cas limite. L'exemple typique est l'estimateur classique de l'espérance mathématique de l'énergie locale, utilisé en VMC pour le calcul de l'énergie variationnelle. En effet, dans cet exemple le cas limite est celui où notre fonction d'onde est état propre du hamiltonien, auquel cas l'énergie locale est constante et égale à l'énergie de l'état. Si on développe $\boldsymbol{E}\left(E_l^2\right)$ en terme d'intégrales, on trouve :

$$\boldsymbol{E}\left(E_l^2\right) = \int_\Omega \left(\frac{\hat{H}\psi}{\psi}\right)^2 \psi\psi^* = \int_\Omega (\hat{H}\psi)(\hat{H}\psi)^* = \left\langle \hat{H}\psi | \hat{H}\psi \right\rangle = \left\langle \psi \left| \hat{H}^2 \right| \psi \right\rangle \ . \tag{3.1}$$

Si on développe $\psi = \sum_i c_i\psi_i$, où $\psi_i$ est l'état propre d'énergie $E_i$, on arrive donc à :

$$\mathrm{Var}\left(E_l\right) = \sum_i |c_i|^2 E_i^2 - \left(\sum_i |c_i|^2 E_i\right)^2 \ . \tag{3.2}$$

Cela revient à la variance d'une variable aléatoire qui prend la valeur $E_i$ avec la loi discrète de probabilité $|c_i|^2$ et une probabilité zéro en dehors du spectre de l'hamiltonien. Si on fait tendre vers 0 tous les $c_i$ sauf un, la variance converge vers zéro.

On va chercher à mettre en place un principe de zéro-variance afin de gérer des systèmes de grande taille, dans lesquels les extrémités d'un système intéragissent assez peu. Pour cela, on va s'intéresser à la limite dans laquelle on peut découper un système en sous-systèmes indépendants, la limite de séparabilité.

## 3.2 Notion de partition du système, et opérateurs conditionnels

En mathématiques, on appelle partition d'un ensemble $E$ toute famille $(F_i)_{i\in[[1,p]]} \in \mathcal{P}(E)^p$ de sous-ensembles disjoints de $E$ (où $\mathcal{P}(E)$ est l'ensemble des parties de $E$) dont la réunion redonne le système tout entier, autrement dit que $\bigcup_{i=1}^{p} F_i = E$. Chaque élément de $E$ appartient alors à un et un seul sous-ensemble $F_i$ :

$$\forall x \in E,\ \exists! i \in [[1,p]],\ x \in F_i\ . \tag{3.3}$$

En pratique, on se doit de considérer notre système $\mathcal{S}$ comme à la fois un espace géographique $\omega$ et un certain nombre $N$ de particules numérotées évoluant dans celui-ci ; on notera ainsi $\mathcal{S} = (\omega, [[1,N]])$. On dira alors que $\mathcal{S}_i = (\omega_i, J_i)$, est un sous-système de $\mathcal{S}$ si $\omega_i \subset \omega$ et $J_i = \{j_1, \dots, j_n\} \in \mathcal{P}([[1,N]])$. Cela nous permet de définir la réunion et l'intersection de sous-systèmes par $\mathcal{S}_i \cup \mathcal{S}_j = (\omega_i \cup \omega_j, J_i \cup J_j)$ et $\mathcal{S}_i \cap \mathcal{S}_j = (\omega_i \cap \omega_j, J_i \cap J_j)$.

On dira, par ailleurs, que la famille de sous-systèmes $(\mathcal{S}_1, ..., \mathcal{S}_p)$ forme une partition de $\mathcal{S}$ si, pour toute configuration $\mathcal{R} \in \Omega = \omega^N$ du système, chaque particule appartient à un et un seul sous-système :

$$\forall \mathcal{R} \in \Omega,\ \forall j \in [[1,N]],\ \exists! i \in [[1,p]],\ j \in \mathcal{S}_i\ ; \tag{3.4}$$

en utilisant la notation $j \in \mathcal{S}_i$ pour représenter $j \in J_i$. En pratique, on peut distinguer les partitions électroniques, où on peut poser $\mathcal{S}_i = (\omega, J_i)$ où $J_i \in \mathcal{P}([[1,N]])$ est une partie statique de l'ensemble des particules ; et les partitions géographiques, où on construit une partition de $\omega$ en sous-secteurs disjoints $\omega_i$, qui induisent pour toute configuration une partition de l'ensemble des particules liée à leur position. Cela revient à écrire les sous-systèmes sous la forme $\mathcal{S}_i = (\omega_i, \{k | \vec{r}_k \in \omega_i\})$.

On peut alors définir pour tout sous-système $\mathcal{S}_i$ son environnement $\mathcal{S}_{\bar{\imath}}$, tel que $(\mathcal{S}_i, \mathcal{S}_{\bar{\imath}})$ réalisent une partition de $\mathcal{S}$. On définit également la sous-configuration associée $\mathcal{R}_i = (\vec{r}_j)_{j\in\mathcal{S}_i}$. On peut alors réordonner les électrons pour écrire que $\mathcal{R} = (\mathcal{R}_1, ..., \mathcal{R}_p) = (\mathcal{R}_i, \mathcal{R}_{\bar{\imath}})$.

Cela nous permet de définir l'espace conditionnel $\Omega_{\bar{\imath}}(\mathcal{R})$ comme l'ensemble des configurations partageant le même environnement de $\mathcal{S}_i$ :

$$\Omega_{\bar{\imath}}(\mathcal{R}) = \{\mathcal{R}' \in \Omega / \mathcal{R}_{\bar{\imath}} = \mathcal{R}'_{\bar{\imath}}\}\ . \tag{3.5}$$

On peut alors définir l'espérance conditionnelle d'une variable aléatoire $X$ pour la densité de probabilité $\rho$ et connaissant l'environnement, comme la valeur moyenne de $X$ sur le sous-système, à environnement figé :

$$\boldsymbol{E}\left(X|\bar{\imath}\right)(\mathcal{R}) = \frac{\displaystyle\int_{\Omega_{\bar{\imath}}(\mathcal{R})} X\rho}{\displaystyle\int_{\Omega_{\bar{\imath}}(\mathcal{R})} \rho}\ . \tag{3.6}$$

Il s'agit visiblement d'une variable aléatoire de $\Omega$. On peut ainsi considérer qu'il s'agit de l'action d'un opérateur $\mathcal{E}_i$ sur la variable aléatoire $X$ :

$$\begin{array}{rccc} \mathcal{E}_i : & \mathbb{C}^\Omega & \longrightarrow & \mathbb{C}^\Omega \\ & X & \mapsto & \boldsymbol{E}\left(X|\bar{\imath}\right) \end{array}\ . \tag{3.7}$$

Cherchons maintenant à déterminer les propriétés de cet opérateur. Cet opérateur est bien évidemment linéaire ; il s'agit donc d'un endomorphisme de l'espace des variables aléatoires $\mathbb{C}^\Omega$. De plus, comme toute variable aléatoire résultante $\mathcal{E}_i X$ est constante sur les espaces conditionnels de $\mathcal{S}_i$ (les $\Omega_{\bar{\imath}}(\mathcal{R})$), on a $\mathcal{E}_i\mathcal{E}_i = \mathcal{E}_i$. Il s'agit donc d'un projecteur sur l'espace des variables constantes sur les espaces conditionnels de $\mathcal{S}_i$.

De plus, si $\mathcal{S}_k = \mathcal{S}_i \cup \mathcal{S}_j$, alors on a $\mathcal{E}_k = \mathcal{E}_i\mathcal{E}_j$. On en tire donc que les projecteurs $\mathcal{E}_i$ commutent deux à deux. De plus, comme moyenner sur un sous-système, puis sur l'environnement, revient à moyenner sur le système tout entier, le projecteur $\mathcal{E}_{\bar{\imath}}\mathcal{E}_i$ projète sur le sous-espace vectoriel des variables aléatoires constantes. Comme celui-ci est de dimension 1, on peut l'assimiler à $\mathbb{C}$, et donc assimiler ce projecteur à la forme linéaire espérance mathématique ; ce qui signifie que $\mathcal{E}_{\bar{\imath}}\mathcal{E}_i = \boldsymbol{E}$.

On peut donc définir le projecteur complémentaire $\Delta_i = \hat{1} - \mathcal{E}_i$ qui à $X$ associe $X - \boldsymbol{E}(X|\bar{\imath})$. Si $\mathcal{E}_i$ moyenne sur chacun des espaces conditionnels de $\mathcal{S}_i$, alors $\Delta_i$ peut s'interpréter par l'action de centrer la variable aléatoire sur chacun d'entre eux. On a bien évidemment $\Delta_i\mathcal{E}_i = \mathcal{E}_i\Delta_i = 0$.

Une propriété intéressante de ces deux familles d'opérateurs, qu'on appellera opérateurs conditionnels, apparaît si l'on s'intéresse à la covariance $\mathrm{Cov}(\cdot,\cdot)$, définie par :

$$\begin{array}{rccc} \mathrm{Cov}(\cdot,\cdot): & (\mathbb{C}^\Omega)^2 & \longrightarrow & \mathbb{C} \\ & (X,Y) & \mapsto & \boldsymbol{E}(XY) - \boldsymbol{E}(X)\,\boldsymbol{E}(Y) \end{array} . \tag{3.8}$$

En effet, la covariance est un produit scalaire sur le sous-espace vectoriel des variables aléatoires centrées. Alors, pour deux variables aléatoires $X$ et $Y$, on a :

$$\begin{aligned} \mathrm{Cov}(\Delta_i X, \mathcal{E}_i Y) &= \boldsymbol{E}(\Delta_i X\, \mathcal{E}_i Y) - \boldsymbol{E}(\Delta_i X)\,\boldsymbol{E}(\mathcal{E}_i Y) \\ &= \mathcal{E}_{\bar{\imath}}\mathcal{E}_i(\Delta_i X\, \mathcal{E}_i Y) - \mathcal{E}_{\bar{\imath}}\mathcal{E}_i\Delta_i X\, \mathcal{E}_{\bar{\imath}}\mathcal{E}_i\mathcal{E}_i Y \\ &= \mathcal{E}_{\bar{\imath}}(\mathcal{E}_i\Delta_i X\, \mathcal{E}_i Y) - 0\boldsymbol{E}(Y) \\ &= 0\ . \end{aligned} \tag{3.9}$$

On voit donc que $\mathcal{E}_i$ et $\Delta_i$ sont des projecteurs orthogonaux pour le produit scalaire covariance, et on a par simple développement de $X = \mathcal{E}_i X + \Delta_i X$ :

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(\mathcal{E}_i X, \mathcal{E}_i Y) + \mathrm{Cov}(\Delta_i X, \Delta_i Y)\ . \tag{3.10}$$

## 3.3 Construction de l'estimateur théorique

Intéressons nous maintenant à la limite de séparabilité. Il s'agit d'un cas limite caractérisé par le fait qu'on puisse écrire notre système comme une réunion de sous-systèmes indépendants entre eux. On parle ici d'une double indépendance. D'une part, il faut que nos sous-systèmes soient physiquement indépendents, c'est-à-dire qu'ils n'intéragissent pas entre eux ; et d'autre part qu'ils soient statistiquement indépendants. Cette première condition revient à imposer que le hamiltonien puisse être décomposé en une somme de hamiltoniens sur les sous-systèmes, commutant entre eux ($\hat{H} = \sum_i \hat{h}_i$), et donc que toute observable extensive s'écrive comme une somme de variables aléatoires sur chacun des sous-systèmes ; tandis que la seconde impose que la densité de probabilité $\rho = |\psi_T|^2$ se mette sous la forme d'un produit de densités sur chacun des sous-systèmes.

Dans ces conditions, la limite de séparabilité peut exprimée comme l'existence d'une partition de notre système $\mathcal{S}$ en $p \geq 2$ sous-systèmes indépendants $(\mathcal{S}_i)_{1\leq i\leq p}$. En supposant que la variable aléatoire $X$ qui nous intéresse est extensive (comme c'est le cas des termes énergétiques), on peut alors l'écrire comme une somme de variables aléatoires indépendantes $X_i$ provenant de chacun des sous-systèmes, $X = \sum_{i=1}^p X_i$. Pour peu que l'on connaisse l'espérance mathématique de chacune de ces variables aléatoires, on peut alors reconstruire l'espérance de $X$ à partir de celles des $X_i$.

Si on cherche à développer l'espérance conditionnelle $\mathcal{E}_i X$ sur les $X_j$, la condition d'indépendance des systèmes entre eux nous donne que $\mathcal{E}_i X_i = \boldsymbol{E}(X_i)$ et $\mathcal{E}_i X_j = X_j$ sinon. On en tire alors le résultat suivant :

$$\begin{aligned} \mathcal{E}_i X &= \sum_{j=1}^{p} \mathcal{E}_i X_j \\ &= \boldsymbol{E}(X_i) + \sum_{\substack{j=1 \\ j\neq i}}^{p} X_j \\ &= \boldsymbol{E}(X_i) + (X - X_i)\ . \end{aligned} \tag{3.11}$$

On peut alors réorganiser cette dernière expression pour en tirer l'expression de $\boldsymbol{E}(X_i)$ :

$$\boldsymbol{E}(X_i) = X_i - \Delta_i X\ . \tag{3.12}$$

En d'autre mots, $\boldsymbol{E}(X_i)$, c'est $X_i$ à laquelle on retire les variations de $X$ sur l'espace conditionnel. On peut alors sommer cette expression pour retrouver l'expression de l'espérance mathématique de $X$ :

$$\boldsymbol{E}(X) = X - \sum_{i=1}^{p} \Delta_i X \ . \tag{3.13}$$

Il s'agit clairement d'une structure en "variable d'intérêt plus variable de contrôle", ou plutôt, somme de variables de contrôle, où chacune retranche à $X$ ses variations dans l'espace conditionnel lié à un de ses sous-systèmes. On peut donc s'en servir pour construire une variable aléatoire $\tilde{X}_{\text{th}} = X - \sum_{i=1}^{p} \Delta_i X$, qu'on appellera la variable $X$ corrigée théorique, et dont l'estimateur traditionnel de l'espérance mathématique est l'estimateur PMC théorique, non biaisé dans le cas général et zéro-variant et dans la limite de séparabilité, et dans la limite d'un état propre :

$$\overline{\tilde{X}_{\text{th}}} = \overline{X} - \sum_{i=1}^{p} \overline{\Delta_i X} \ ; \tag{3.14}$$

où la barre est une notation pour la moyenne sur l'échantillon.

## 3.4 Construction de l'estimateur pratique

On se rend rapidement compte que, pour calculer la valeur de l'estimateur que l'on vient de développer, il est nécessaire d'estimer les valeurs des $\Delta_i X$, et donc des espérances conditionnelles, à chaque configuration de la dynamique Monte Carlo. Pour cela, la méthode PMC introduit de courtes sous-dynamiques latérales (side-walks) sur chacun des sous-systèmes de la partition.

Ces sous-dynamiques latérales consistent de dynamiques Métropolis-Hastings basées sur une restriction de la matrice de transition globale au sous-système, de manière à ce que la densité stationnaire soit la restriction de $|\psi|^2$ à l'espace conditionnel $\Omega_{\bar{i}}(\mathcal{R})$ sur lequel on travaille. En pratique, cela revient à utiliser le même algorithme de génération de mouvements, mais à rejeter quand on sort du sous-système, ou, si notre système global avait déjà des conditions périodiques aux bords, à utiliser celles-ci aux bords du sous-système.

Si on note $\mathcal{R}^K$ la $K$-ième configuration de la dynamique principale (et $\mathcal{R}_i^K$ et $\mathcal{R}_{\bar{i}}^K$ ses sous-configurations, avec $\mathcal{R}^K \equiv (\mathcal{R}_i^K, \mathcal{R}_{\bar{i}}^K)$), et $\mathcal{R}_i^{Kk}$ la $k$-ième sous- configuration de la $K$-ième sous-dynamique dans le sous-système $\mathcal{S}_i$, alors l'expression qu'on utilise pour l'approximation de $\Delta_i X(\mathcal{R}^K)$ est :

$$\Delta_i X(\mathcal{R}^K) \approx \frac{1}{m} \sum_{k=1}^{m} X(\mathcal{R}_i^K, \mathcal{R}_{\bar{i}}^K) - X(\mathcal{R}_i^{Kk}, \mathcal{R}_{\bar{i}}^K) \ . \tag{3.15}$$

Il s'agit visiblement de l'expression de l'estimateur classique de l'espérance de $X(\mathcal{R}^K) - X$ sur l'espace conditionnel $\Omega_{\bar{i}}(\mathcal{R}^K)$. On peut choisir d'utiliser une expression améliorée, dans le cadre de ce qu'on appellera la méthode Monte Carlo partitionnelle multi-échelles (MS-PMC), qu'on présente à la section 4.1 ; il s'agira alors d'une expression récursive, en utilisant un estimateur PMC ou MS-PMC de l'espérance conditionnelle.

Par ailleurs, comme $\sum_i \Delta_i X$ est une variable de contrôle, on peut introduire un paramètre $c$ comme préfacteur de la variable de contrôle dans l'équation (3.14), ce qui nous donne comme expression globale de l'estimateur PMC de la variable $X$ :

$$X_{\text{PMC}} = \frac{1}{M} \sum_{K=1}^{M} X(\mathcal{R}^K) - \frac{c}{mM} \sum_{K=1}^{M} \sum_{i=1}^{p} \sum_{k=1}^{m} \left[ X(\mathcal{R}^K) - X(\mathcal{R}_i^{Kk}, \mathcal{R}_{\bar{i}}^K) \right] \ . \tag{3.16}$$

En particulier, pour l'énergie, on peut séparer énergie cinétique et énergie potentielle, voire même énergie potentielle externe et énergie potentielle électron-électron. En restant sur une simple partition

en $E_l = T_l + V_l$, et en utilisant des coefficients séparés $c_t$ et $c_v$ pour $\sum_i \Delta_i T_l$ et $\sum_i \Delta_i V_l$, on arrive à :

$$\begin{aligned} E_{\text{PMC}} = \frac{1}{M}\sum_{K=1}^{M} E_l(\mathcal{R}^K) - \frac{c_t}{mM}\sum_{K=1}^{M}\sum_{i=1}^{p}\sum_{k=1}^{m}\left[T_l(\mathcal{R}^K) - T_l(\mathcal{R}_i^{Kk}, \mathcal{R}_{\bar{i}}^K)\right] \\ - \frac{c_v}{mM}\sum_{K=1}^{M}\sum_{i=1}^{p}\sum_{k=1}^{m}\left[V_l(\mathcal{R}^K) - V_l(\mathcal{R}_i^{Kk}, \mathcal{R}_{\bar{i}}^K)\right] \ . \end{aligned} \tag{3.17}$$

Par la suite, on notera $\bar{X}_i(\mathcal{R}^K) = (1/m)\sum_{k=1}^{m} X(\mathcal{R}_i^{Kk}, \mathcal{R}_{\bar{i}}^K)$ la valeur moyenne de $X$ sur la $K$-ième sous-dynamique sur le sous-système $\mathcal{S}_i$, et on notera $\tilde{X}_{\text{pr}}$ la variable corrigée pratique, définie ci-dessous :

$$\tilde{X}_{\text{pr}} = X - \sum_{i=1}^{p}(X - \bar{X}_i) = (1-p)X + \sum_{i=1}^{p}\bar{X}_i \ . \tag{3.18}$$

On peut la relier à $\tilde{X}_{\text{th}}$ en se servant de la relation suivante :

$$\tilde{X}_{\text{th}} = \lim_{m\to\infty} \tilde{X}_{\text{pr}} \ . \tag{3.19}$$

On va maintenant chercher à démontrer que cette variable aléatoire est bel et bien de variance moindre que $X$.

## 3.5 Calcul de Var($\tilde{X}$)

Pour commencer, on notera $\text{Cov}\,(X, Y|\bar{\imath}) = \mathcal{E}_i(XY) - \mathcal{E}_i X \mathcal{E}_i Y$ la covariance conditionnelle ou covariance interne, et $\text{Var}\,(X|\bar{\imath}) = \text{Cov}\,(X, X|\bar{\imath})$ la variance conditionnelle ou variance interne, liées au sous-système $\mathcal{S}_i$. Alors, on va démontrer une expression qu'on nommera "partition des covariances" vraie pour tout sous-système $\mathcal{S}_i$ de $\mathcal{S}$ et toute paire de variables aléatoires $X$ et $Y$ :

$$\begin{aligned} \text{Cov}\,(X,Y) &= \boldsymbol{E}\,(XY) - \boldsymbol{E}\,(X)\,\boldsymbol{E}\,(Y) = \boldsymbol{E}\,(\mathcal{E}_i(XY)) - \boldsymbol{E}\,(\mathcal{E}_i X)\,\boldsymbol{E}\,(\mathcal{E}_i Y) \\ &= \boldsymbol{E}\,(\mathcal{E}_i(XY)) - \boldsymbol{E}\,(\mathcal{E}_i X\,\mathcal{E}_i Y) + \boldsymbol{E}\,(\mathcal{E}_i X\,\mathcal{E}_i Y) - \boldsymbol{E}\,(\mathcal{E}_i X)\,\boldsymbol{E}\,(\mathcal{E}_i Y) \\ &= \boldsymbol{E}\,(\text{Cov}\,(X, Y|\bar{\imath})) + \text{Cov}\,(\mathcal{E}_i X, \mathcal{E}_i Y) \ . \end{aligned} \tag{3.20}$$

On peut aisément identifier cette expression à l'équation (3.10) pour en tirer
$\text{Cov}\,(\Delta_i X, \Delta_i Y) = \boldsymbol{E}\,(\text{Cov}\,(X, Y|\bar{\imath}))$.

D'autre part, intéressons-nous à un développement en covariances sur chacun des sous-systèmes. Dans le cadre de l'hypothèse de séparabilité, pour une variable aléatoire extensive $X$, on a par indépendence des sous-systèmes :

$$\text{Var}\,(X_i) = \text{Var}\,(X|\bar{\imath}) \equiv V_i \ . \tag{3.21}$$

Autrement dit, la variance interne correspond à la variance de la variable aléatoire du sous-système correspondant dans la limite de séparabilité. Cependant, dans le cas général, $V_i = \text{Var}\,(X|\bar{\imath})$ n'est pas une constante mais une variable aléatoire. Le premier ordre du développement en covariances prend donc la forme suivante :

$$\text{Var}\,(X) = \sum_{i=1}^{p}\boldsymbol{E}\,(V_i) + \ldots \ . \tag{3.22}$$

Le second ordre de ce développement revient à ne considérer que des corrélations indépendantes des sous-systèmes de la partition deux à deux, par exemple pour prendre en compte des effets de bord de corrélations à courte portée. Si on pose $\mathcal{S}_k$ la réunion des sous-systèmes $\mathcal{S}_i$ et $\mathcal{S}_j$, on va noter $V_{ij} = \text{Var}\left(X|\bar{k}\right)$ la variance conditionnelle sur la réunion. Cela revient à écrire :

$$\text{Var}\,(X) = \sum_{i=1}^{p}\boldsymbol{E}\,(V_i) + \sum_{i=1}^{p}\sum_{j=i+1}^{p}\boldsymbol{E}\,(V_{ij} - V_i - V_j) + \ldots \ . \tag{3.23}$$

Il s'agit d'un simple développement en sommes téléscopiques que l'on peut rapprocher de la formule de Poincaré en combinatoire. Ainsi, pour une partition en trois sous-systèmes, le développement complet donne :

$$\begin{aligned}\mathrm{Var}\,(X) = {} & \boldsymbol{E}\,(V_1) + \boldsymbol{E}\,(V_2) + \boldsymbol{E}\,(V_3) + \boldsymbol{E}\,(V_{12} - V_1 - V_2) + \boldsymbol{E}\,(V_{13} - V_1 - V_3) + \boldsymbol{E}\,(V_{23} - V_2 - V_3) \\ & + \boldsymbol{E}\,(\mathrm{Var}\,(X) - V_{12} - V_{23} - V_{13} + V_1 + V_2 + V_3)\ .\end{aligned} \tag{3.24}$$

Maintenant que l'on dispose de ces équations, commençons par réécrire l'expression de $\tilde{X}_{\mathrm{th}}$ défini plus haut. On a :

$$\tilde{X}_{\mathrm{th}} = X - \sum_{i=1}^{p} \Delta_i X = (1-p)X + \sum_{i=1}^{p} \mathcal{E}_i X\ . \tag{3.25}$$

Cela se développe en l'équation suivante :

$$\begin{aligned}\mathrm{Var}(\tilde{X}_{\mathrm{th}}) = {} & (1-p)^2\,\mathrm{Var}\,(X) + 2(1-p)\sum_{i=1}^{p}\mathrm{Cov}\,(X, \mathcal{E}_i X) \\ & + \sum_{i=1}^{p}\mathrm{Var}\,(\mathcal{E}_i X) + 2\sum_{i=1}^{p}\sum_{j=i+1}^{p}\mathrm{Cov}\,(\mathcal{E}_i X, \mathcal{E}_j X)\ .\end{aligned} \tag{3.26}$$

La partition des variances avec le sous-système $\mathcal{S}_i$ nous donne de manière transparente l'expression suivante :

$$\mathrm{Var}\,(\mathcal{E}_i X) = \mathrm{Cov}\,(X, \mathcal{E}_i X) = \mathrm{Var}\,(X) - \boldsymbol{E}\,(V_i)\ . \tag{3.27}$$

Enfin, pour le dernier terme de l'équation (3.26), on utilise la composition des variances sur le système réunion $\mathcal{S}_k = \mathcal{S}_i \cup \mathcal{S}_j$ :

$$\begin{aligned}\mathrm{Cov}\,(\mathcal{E}_i X, \mathcal{E}_j X) &= \mathrm{Var}\,(\mathcal{E}_k X) + \boldsymbol{E}\left(\mathrm{Cov}\left(\mathcal{E}_i X, \mathcal{E}_j X | \bar{k}\right)\right) \\ &= (\mathrm{Var}\,(X) - \boldsymbol{E}\,(V_{ij})) + \boldsymbol{E}\left(\mathcal{E}_k(\mathcal{E}_i X\, \mathcal{E}_j X) - (\mathcal{E}_k X)^2\right) \\ &= (\mathrm{Var}\,(X) - \boldsymbol{E}\,(V_{ij})) + \boldsymbol{E}\left(\mathcal{E}_j\mathcal{E}_i(\mathcal{E}_i X\, \mathcal{E}_j X) - (\mathcal{E}_k X)^2\right)\ . \\ &= (\mathrm{Var}\,(X) - \boldsymbol{E}\,(V_{ij})) + \boldsymbol{E}\left(\mathcal{E}_j(\mathcal{E}_i X\, \mathcal{E}_k X) - (\mathcal{E}_k X)^2\right) \\ &= \mathrm{Var}\,(X) - \boldsymbol{E}\,(V_{ij})\ .\end{aligned} \tag{3.28}$$

En resommant dans l'équation (3.26), on en tire donc l'expression suivante :

$$\mathrm{Var}(\tilde{X}_{\mathrm{th}}) = \mathrm{Var}\,(X) - \sum_{i=1}^{p}\boldsymbol{E}\,(V_i) - 2\sum_{i=1}^{p}\sum_{j=i+1}^{p}\boldsymbol{E}\,(V_{ij} - V_i - V_j)\ . \tag{3.29}$$

Intéressons-nous maintenant à l'estimateur pratique $\tilde{X}_{\mathrm{pr}}$ défini à l'équation (3.18), qui met en jeu les moyennes de sous-dynamiques $\bar{X}_i$ plutôt que les espérances conditionnelles $\mathcal{E}_i X$. On a bien évidemment $\mathcal{E}_i \bar{X}_i = \mathcal{E}_i X$. Si on cherche à le développer comme on l'a fait pour $\tilde{X}_{\mathrm{th}}$ à l'équation (3.26), on trouve :

$$\begin{aligned}\mathrm{Var}(\tilde{X}_{\mathrm{pr}}) &= (1-p)^2\,\mathrm{Var}\,(X) + 2(1-p)\sum_{i=1}^{p}\mathrm{Cov}\left(X, \bar{X}_i\right) + \sum_{i=1}^{p}\mathrm{Var}\left(\bar{X}_i\right) + 2\sum_{i=1}^{p}\sum_{j=i+1}^{p}\mathrm{Cov}\left(\bar{X}_i, \bar{X}_j\right) \\ &= \mathrm{Var}(\tilde{X}_{\mathrm{th}}) \\ &\quad + 2(1-p)\sum_{i=1}^{p}\left[\mathrm{Cov}\left(X, \bar{X}_i\right) - \mathrm{Cov}\,(X, \mathcal{E}_i X)\right] \\ &\quad + \sum_{i=1}^{p}\left[\mathrm{Var}\left(\bar{X}_i\right) - \mathrm{Var}\,(\mathcal{E}_i X)\right] \\ &\quad + 2\sum_{j=i+1}^{p}\left[\mathrm{Cov}\left(\bar{X}_i, \bar{X}_j\right) - \mathrm{Cov}\,(\mathcal{E}_i X, \mathcal{E}_j X)\right]\ .\end{aligned} \tag{3.30}$$

Traitons alors cette expression terme à terme.

Le troisième terme est le plus facile. En effet, on peut utiliser la partition des covariances pour faire apparaître la variance interne de $\bar{X}_i$ dans le sous-système $\mathcal{S}_i$, ce qui prend la forme de la variance d'un estimateur classique de l'espérance, qu'on munit d'un temps d'autocorrélation $\tau_i$ :

$$\mathrm{Var}\left(\bar{X}_i\right) - \mathrm{Var}\left(\mathcal{E}_i X\right) = \boldsymbol{E}\left(\mathrm{Var}\left(\bar{X}_i|\bar{i}\right)\right) = \boldsymbol{E}\left(\frac{V_i\tau_i}{m}\right) . \tag{3.31}$$

On fait de même apparaître une fonction d'autocorrélation dans le second terme :

$$\mathrm{Cov}\left(X, \bar{X}_i\right) - \mathrm{Cov}\left(X, \mathcal{E}_i X\right) = \boldsymbol{E}\left(\mathrm{Cov}\left(X, \bar{X}_i|\bar{i}\right)\right) = \boldsymbol{E}\left(V_i\frac{\tau_i - 1}{2m}\right) . \tag{3.32}$$

Pour le dernier terme de l'équation (3.30), on réemploie la partition des covariances utilisée à l'étape précédente, ce qui nous donne :

$$\mathrm{Cov}\left(\bar{X}_i, \bar{X}_j\right) - \mathrm{Cov}\left(\mathcal{E}_i X, \mathcal{E}_j X\right) = \boldsymbol{E}\left(\mathrm{Cov}\left(\bar{X}_i, \bar{X}_j|\bar{k}\right)\right) . \tag{3.33}$$

Il nous reste donc à évaluer cette covariance entre deux sous-dynamiques. Pour cela, on se propose d'utiliser les fonctions d'autocorrélation pour obtenir une approximation de $\bar{X}_i$ en le projetant sur $(\mathcal{E}_i X, \Delta_i X)$. Cela nous donne :

$$\bar{X}_i \underset{m\to\infty}{=} \mathcal{E}_i X + \frac{\tau_i - 1}{2m}\Delta_i X . \tag{3.34}$$

On peut alors utiliser cette expression pour développer notre covariance. L'emploi de la partition des variances sur chacun des termes nous donne alors :

$$\boldsymbol{E}\left(\mathrm{Cov}\left(\bar{X}_i, \bar{X}_j|\bar{k}\right)\right) = \boldsymbol{E}\left(\frac{\tau_i - 1}{2m}(V_{ij} - V_j)\right) + \boldsymbol{E}\left(\frac{\tau_j - 1}{2m}(V_{ij} - V_i)\right) - \boldsymbol{E}\left(\frac{(\tau_i - 1)(\tau_j - 1)}{4m^2}(V_{ij} - V_i - V_j)\right) . \tag{3.35}$$

En réintégrant tous les termes ainsi obtenus dans l'équation (3.30), on arrive au résultat suivant :

$$\mathrm{Var}(\tilde{X}_{\mathrm{pr}}) = \mathrm{Var}\left(X\right) - \sum_{i=1}^{p}\boldsymbol{E}\left(V_i(1 - \frac{\tau_i}{m})\right) - 2\sum_{i=1}^{p}\sum_{j=i+1}^{p}\boldsymbol{E}\left((V_{ij} - V_i - V_j)(1 - \frac{\tau_i - 1}{2m})(1 - \frac{\tau_j - 1}{2m})\right) . \tag{3.36}$$

On voit donc que l'on retrouve bien pour la variable corrigée pratique la variance de la variable corrigée théorique dans la limite de sous-dynamiques de longueur infinie, ce qui est bien ce à quoi on s'attendait. De plus, dès lors que le deuxième ordre de l'approximation de séparabilité est exact, on arrive à un estimateur zéro-variant. C'est le cas lorsque les systèmes ne sont corrélés que deux à deux, par exemple pour des effets de bords sur des sous-systèmes de grande taille par rapport à la longueur de corrélation entre particules (sur une partition géographique). Cela revient ainsi à compenser les effets de bords.

De plus, comme les $\tau_i$ sont une fonction croissante convergente de $m$ majorée par $m$, on est garanti d'avoir un gain en variance systématique, indépendemment de l'optimisation du paramètre de la variable de contrôle.

## Récapitulatif

Dans ce chapitre, nous avons commencé par exprimer le principe de zéro-variance, en le rapprochant de la propriété de continuité, et en prenant pour exemple la limite des états propres pour l'estimateur classique de l'énergie variationnelle. Ensuite, nous avons défini la notion de partition du système, ainsi que les projecteurs orthogonaux d'espérance conditionnelle $\mathcal{E}_i$ et $\Delta_i$.

Grâce à ces opérateurs, nous avons pû construire un estimateur théorique non biaisé qui soit zéro-variant dans la limite de sous-systèmes indépendents, puis créer à partir de celui-ci un estimateur pratique non biaisé, zéro-variant dans la limite de sous-systèmes indépendents avec des sous-échantillonnages infinis. On peut interpréter cet estimateur pratique comme la moyenne de la variable aléatoire munie d'une somme de termes correctifs liés à chacun des sous-systèmes.

Enfin, nous avons calculé la variance de cet estimateur pratique. Non seulement avons-nous réussi à prouver que celui-ci réduisait la variance de manière systématique, mais en plus que la variance se comportait en $A + B/m$. Maintenant, l'étape suivante consiste à implémenter cet estimateur pour pouvoir le tester.

# Chapitre 4

# Implémentation

Dans ce chapitre, nous chercherons à couvrir les détails qui séparent la théorie de la méthode de Monte Carlo Partitionnelle de son implémentation. Après une brève exploration de l'implémentation récursive de notre méthode – que nous appellerons Monte Carlo Partitionnelle Multi-Échelle – nous présenterons les formalismes de réduction matricielle développés par Filippi, Assaraf et Moroni [1] qui nous servent à générer les sous-dynamiques latérales à un coût faible. Après quoi, nous présenterons le modèle discret sur lequel nous avons travaillé, ainsi que les méthodes de calcul de l'énergie cinétique – et des différences de celle-ci – que nous avons développées sur celui-ci. Nous finirons en discutant de l'intérêt d'intégrer les sous-dynamiques à la dynamique principale. Pour les notations employées à ce chapitre, voir l'annexe A.

## 4.1 Monte Carlo Partitionnelle Multi-Échelle

Supposons que l'on dispose de deux partitions de notre système $\mathcal{S}$, $(\mathcal{S}_i)_{i\in[[1,p]]}$ et $(\mathcal{S}_j)_{j\in[[1,q]]}$ telles que tout élément de $(\mathcal{S}_i)_i$ admet une partition en éléments de $(\mathcal{S}_j)_j$. Alors, lorsqu'on évalue $\Delta_i X$, on peut alors envisager de l'évaluer au moyen de la méthode de Monte Carlo Partitionnelle plutôt qu'une simple sous-dynamique latérale. Cela revient à utiliser l'estimateur suivant, en utilisant et étendant les notations de l'équation (3.15) et en utilisant $\mathfrak{m}$ et $\mathfrak{K}$ pour la sous-dynamique de la sous- dynamique :

$$\Delta_i X(\mathcal{R}^K) \approx \frac{1}{m}\sum_{k=1}^{m}\left[X(\mathcal{R}_i^K,\mathcal{R}_{\bar{i}}^K) - X(\mathcal{R}_i^{Kk},\mathcal{R}_{\bar{i}}^K) - \frac{1}{\mathfrak{m}}\sum_{\mathcal{S}_j\subset\mathcal{S}_i}\sum_{\mathfrak{K}=1}^{\mathfrak{m}} X(\mathcal{R}_{ij}^{Kk},\mathcal{R}_{i\bar{j}}^{Kk},\mathcal{R}_{\bar{i}}^K) - X(\mathcal{R}_{ij}^{Kk\mathfrak{K}},\mathcal{R}_{i\bar{j}}^{Kk},\mathcal{R}_{\bar{i}}^K)\right]. \tag{4.1}$$

Il s'agit visiblement de la deuxième étape dans une application récursive de la Monte Carlo Partionnelle, que l'on appellera Monte Carlo Partionnelle Multi-Échelle (en anglais, Multiscale Partition Monte Carlo, ou MS-PMC). La MS-PMC requiert l'existence d'une famille de partitions munie d'une relation d'ordre absolue, que l'on appellera partition hiérarchisée, telle que pour chaque paire de partitions de la famille, il existe une partition de chaque élément de la partition supérieure en éléments de la partition inférieure. Alors, on peut disposer d'une variable de contrôle supplémentaire pour chacune des partitions de la famille hiérarchisée, que l'on peut bien évidemment munir d'un coefficient indépendent. Ne retenir que la variable de contrôle correspondant à la partition la plus haute revient alors à exploiter une simulation MS-PMC comme une simple simulation PMC.

D'un point de vue intuitif, si l'on considère que la méthode VMC revient à explorer une trajectoire dans l'espace configurationnel et la méthode PMC à introduire des branchements d'une manière comparable à un goupillon, alors la MS-PMC revient à introduire des branchements supplémentaires de manière fractale tant que faire se peut.

## 4.2 Formalisme de réduction matricielle

Le formalisme de Filippi, Assaraf et Moroni a été créé pour évaluer efficacement les valeurs de fonctions d'onde à plusieurs déterminants. Cependant, nous présenterons ici une version réduite, dont l'utilisation se justifie par la nécessité de minimiser le coût de calcul des sous-dynamiques afin de tirer un gain calculatoire de celles-ci. Soit $\psi_{\mathrm{S}}$ une fonction d'onde à un seul déterminant de Slater :

$$\psi_{\mathrm{S}}(\mathcal{R}) = \det\left((\chi_i(\vec{r}_j))_{(i,j)\in[[1,N]]^2}\right) . \tag{4.2}$$

Dans le cadre de l'approximation des orbitales moléculaires, on développe les $\chi_i$ dans la base des $N_b$ orbitales atomiques $\varphi_k$, ce qui nous donne $\chi_i = \sum_{k=1}^{N_b} c_{ik}\varphi_k$, et nous permet d'écrire :

$$\begin{aligned}\psi_{\mathrm{S}}(\mathcal{R}) &= \det\left(\left(\sum_{k=1}^{N_b} c_{ik}\varphi_k(\vec{r}_j)\right)_{(i,j)\in[[1,N]]^2}\right) \\ &= \det\left((c_{ik})_{(i,k)\in[[1,N]]\times[[1,N_b]]}(\varphi_k(\vec{r}_j))_{(k,j)\in[[1,N_b]]\times[[1,N]]}\right) .\end{aligned} \tag{4.3}$$

Cela revient à traduire l'approximation des orbitales moléculaires comme un produit matriciel. On définit alors les trois matrices suivantes :

$$\begin{aligned}\boldsymbol{A}(\mathcal{R}) &= (\chi_i(\vec{r}_j))_{(i,j)\in[[1,N]]^2} \\ \boldsymbol{C} &= (c_{ik})_{(i,k)\in[[1,N]]\times[[1,N_b]]} \\ \boldsymbol{X}(\mathcal{R}) &= (\varphi_k(\vec{r}_j))_{(k,j)\in[[1,N_b]]\times[[1,N]]} .\end{aligned}$$

On appelle $\boldsymbol{X}$ la matrice des orbitales atomiques, $\boldsymbol{C}$ la matrice LCAO, qui code les coefficients des orbitales moléculaires, et $\boldsymbol{A}$ la matrice de Slater. $\boldsymbol{C}$ et $\boldsymbol{X}$ sont de manière générale des matrices rectangulaires. L'approximation des orbitales moléculaires revient alors à poser $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{X}$. Maintenant, supposons que l'on cherche à calculer $\psi_{\mathrm{S}}(\mathcal{R})$, en connaissant la valeur, non nulle, dans une configuration de référence $\mathcal{R}_0$. Alors, on a :

$$\psi_{\mathrm{S}}(\mathcal{R}) = \psi_{\mathrm{S}}(\mathcal{R}_0)\det(\boldsymbol{A}(\mathcal{R}_0)^{-1}\boldsymbol{C}\boldsymbol{X}(\mathcal{R})) \equiv \psi_{\mathrm{S}}(\mathcal{R}_0)\det(\boldsymbol{D}(\mathcal{R}_0)\boldsymbol{X}(\mathcal{R})) . \tag{4.4}$$

La matrice $\boldsymbol{D} = \boldsymbol{A}^{-1}\boldsymbol{C}$ est appelée la matrice dérivée. On peut l'interpréter comme la matrice des coefficients de la dérivée logarithmique de $\psi_{\mathrm{S}}$ par rapport aux éléments de $\boldsymbol{X}$ :

$$d_{jk} = \frac{\partial \log\det(\boldsymbol{C}\boldsymbol{X})}{\partial x_{kj}} . \tag{4.5}$$

On a en particulier l'expression assez évidente $\boldsymbol{D}(\mathcal{R}_0)\boldsymbol{X}(\mathcal{R}_0) = \boldsymbol{I}_N$, la matrice identité.

Supposons qu'on ne travaille alors que dans le sous-secteur $\mathcal{S}_i$ comportant $n$ électrons, dans l'espace conditionnel $\Omega_{\bar{i}}(\mathcal{R}_0)$. Alors, les colonnes de $\boldsymbol{A}^{-1}(\mathcal{R}_0)\boldsymbol{A}(\mathcal{R})$ correspondant aux $N-n$ électrons de $\mathcal{S}_{\bar{i}}$ seront égales à celle de l'identité. On peut donc introduire une matrice $\boldsymbol{P}_i$ creuse de taille $N\times n$, dont les lignes sont nulles pour les électrons de $\mathcal{S}_{\bar{i}}$, et dont la restriction aux électrons de $\mathcal{S}_i$ donne $\boldsymbol{I}_n$. Alors on a :

$$\frac{\psi_{\mathrm{S}}(\mathcal{R})}{\psi_{\mathrm{S}}(\mathcal{R}_0)} = \det({}^t\boldsymbol{P}_i\boldsymbol{D}(\mathcal{R}_0)\boldsymbol{X}(\mathcal{R})\boldsymbol{P}_i) . \tag{4.6}$$

Le coût de calcul du déterminant à proprement parler n'est alors que de $\mathcal{O}\left(n^3\right)$, mais le coût de l'établissement de la matrice à l'intérieur est de $\mathcal{O}\left(n^2N_b\right) \approx \mathcal{O}\left(n^2N\right)$, le coût d'un mouvement à $n$ electrons par la méthode de Sherman et Morrison. Pour réduire le coût plus avant, on va alors chercher à restreindre les orbitales atomiques auxquelles les électrons de $\mathcal{S}_i$ ont accès (ce qui est bien plus évident pour une partition géographique qu'une partition électronique de $\mathcal{S}$).

On introduit alors une matrice creuse $\boldsymbol{Q}_i$ qui projette de la même manière sur les $n_b$ orbitales atomiques qu'on a choisies :

$$\frac{\psi_{\mathrm{S}}(\mathcal{R})}{\psi_{\mathrm{S}}(\mathcal{R}_0)} \approx \det({}^t\boldsymbol{P}_i\boldsymbol{D}(\mathcal{R}_0)\boldsymbol{Q}_i\ {}^t\boldsymbol{Q}_i\boldsymbol{X}(\mathcal{R})\boldsymbol{P}_i) \equiv \det(\boldsymbol{C}_2(\mathcal{R}_0,\mathcal{S}_i)\boldsymbol{X}_2(\mathcal{R},\mathcal{S}_i)) . \tag{4.7}$$

Il s'agit d'une approximation qui n'a aucun impact en terme de biais, et qui est d'ailleurs exacte si les orbitales non retenues sont nulles sur l'espace géographique $\omega_i$ dans lequel évoluent les électrons du sous-système. Si ce n'est pas le cas, alors l'erreur relative sur la fonction d'onde est au plus donnée par l'expression suivante :

$$\left|\frac{\delta\psi}{\psi}\right| \leq \sum_{j=1}^{n}\sum_{k=1}^{N_b-n_b} \max_{\vec{r}\in\omega_i} |d_{jk}\varphi_k(\vec{r})| \ . \tag{4.8}$$

On peut minimiser cette erreur avec un choix judicieux des orbitales. Dans un espace continu, on peut par exemple décider d'introduire un *cutoff* sur les valeurs des orbitales atomiques.

On voit par ailleurs que l'équation (4.7) fait réapparaître une structure en $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{X}$. On appelle les matrices $\boldsymbol{C}_2 =^t \boldsymbol{P}_i\boldsymbol{D}(\mathcal{R}_0)\boldsymbol{Q}_i$ et $\boldsymbol{X}_2 =^t \boldsymbol{Q}_i\boldsymbol{X}(\mathcal{R})\boldsymbol{P}_i$ (ainsi que $\boldsymbol{A}_2 = \boldsymbol{C}_2\boldsymbol{X}_2$) des matrices réduites, car ayant une moindre dimension. On voit clairement que l'environnement est factorisé dans la matrice LCAO réduite. Ainsi, on se ramène à un coût de calcul par configuration en $\mathcal{O}\left(n^2 n_b\right)$, ce qui se ramène le plus souvent à un coût en $\mathcal{O}\left(n^3\right)$. L'indice 2 est lié à la notion de niveau que l'on présente à la section 4.1.

## 4.3 Modèle de Hubbard

Le modèle duquel on s'est servi est le modèle de Hubbard, qui consiste en un graphe quelconque de sites dont l'ensemble est $\omega_s$, chacun capable de contenir au plus un électron de chaque spin. Cela correspond à un espace à une particule (ou espace géographique) $\omega = \omega_s \times \{\uparrow, \downarrow\}$. Il s'agit d'un système discret, où chaque électron est capable de sauter d'un site à un site adjacent, et interagit avec l'électron sur le même site. En utilisant la notation $\hat{a}_{\vec{\imath}\sigma}$ pour l'opérateur annihilation du spin-site $(\vec{\imath}, \sigma)$ et $\hat{a}^\dagger_{\vec{\imath}\sigma}$ pour l'opérateur création correspondant, le hamiltonien du modèle de Hubbard est [2] :

$$\hat{H} = -\sum_{\sigma\in\{\uparrow,\downarrow\}}\sum_{\vec{\imath}\neq\vec{\jmath}\in\omega_s} t_{\vec{\imath}\vec{\jmath}}\hat{a}^\dagger_{\vec{\imath}\sigma}\hat{a}_{\vec{\jmath}\sigma} + v\sum_{\vec{\imath}\in\omega_s}\hat{n}_{\vec{\imath}\uparrow}\hat{n}_{\vec{\imath}\downarrow} \ , \tag{4.9}$$

où $v \geq 0$ est un paramètre, les $t_{\vec{\imath}\vec{\jmath}}$ sont les éléments de la matrice d'adjacence $\boldsymbol{T}$ qui vaut 1 si deux sites sont adjacents et 0 sinon, et les $\hat{n}_{\vec{\imath}\sigma} = \hat{a}^\dagger_{\vec{\imath}\sigma}\hat{a}_{\vec{\imath}\sigma}$ sont les opérateurs nombre d'occupation des spin-sites.

Travailler avec un système discret nous permet d'utiliser la base orthonormale des fonctions de Dirac liées à chacun des sites comme famille génératrice de l'espace variationnel monoélectronique, transformant l'approximation des orbitales moléculaires en un développement exact. La matrice des orbitales atomiques est alors creuse, et la complexité est déportée vers la matrice LCAO.

En particulier, on travaille avec une grille régulière bidimensionnelle carrée de $L \times L$ sites, munie de conditions de bord périodiques, contenant $N \approx L^2$ électrons également répartis entre les deux spins, dont on peut visualiser une représentation à la figure 4.1.

Dans ce cas, on a $\omega_s = (\mathbb{Z}/L\mathbb{Z})^2$ ($[[0, L-1]]^2$ muni de coordonnées modulo $L$), et on peut poser les positions sous la forme $\vec{r}_i = (x_i, y_i, \sigma_i) \in \omega$. La fonction d'onde que l'on emploie est alors un déterminant d'ondes planes dont les nombres d'ondes sont choisis pour minimiser l'énergie cinétique. Il s'agit alors de la meilleure fonction d'onde à un seul déterminant dans le cas $v = 0$ :

$$\psi_{\mathrm{S}}(\mathcal{R}) = \det\left(\left(\delta_{\sigma_i\sigma_j}\exp\left(\boldsymbol{i}(x_i k_{xj} + y_i k_{yj})\right)\right)_{(i,j)\in[[1,N]]^2}\right) \ . \tag{4.10}$$

Il s'agit d'un vecteur propre de l'opérateur énergie cinétique monoélectronique, et l'énergie cinétique globale est la somme de celles des ondes planes. Les nombres d'onde $\vec{k}_j = (k_{xj}, k_{yj}, \sigma_j) \in (\frac{2\pi}{L}\mathbb{Z}/L\mathbb{Z})^2 \times \{\uparrow, \downarrow\}$ ont pour énergie cinétique orbitalaire $T_j = -2\cos(k_{xj}) - 2\cos(k_{yj})$ (voir démonstration en annexe C.1). On choisit les $N$ orbitales aux plus basses valeurs de $T_j$.

Cette fonction d'onde présente plusieurs avantages. Non seulement il s'agit d'un vecteur propre de l'opérateur énergie cinétique global, mais elle se résume pour la méthode Hartree-Fock à réaliser une somme sur l'espace réciproque. On a donc accès à une valeur de référence de l'énergie variationnelle pour vérifier le comportement de notre algorithme. De plus, il s'agit d'une fonction d'onde qui possède

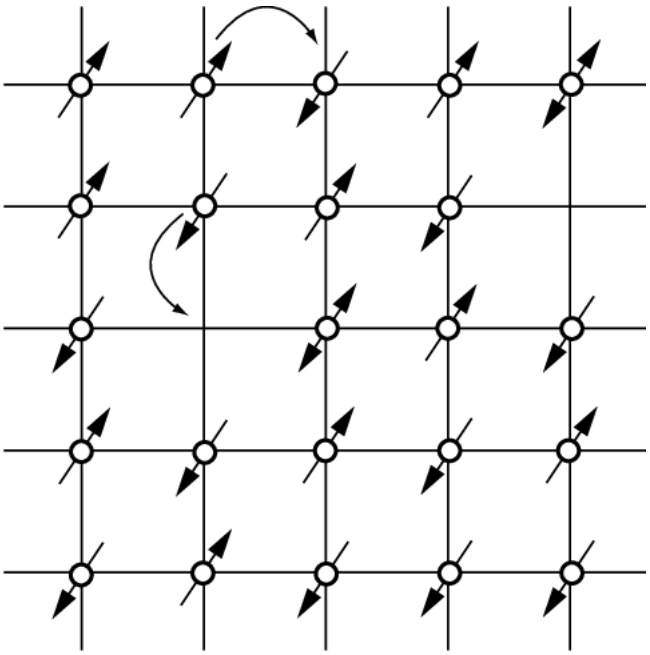

FIGURE 4.1 – Le modèle de Hubbard sous forme de grille 2D.

une longueur de corrélation infinie car les orbitales moléculaires sont complètement délocalisées ; et lorsque $L$ tend vers l'infini, on se retrouve dans un système métallique. Cela revient donc à se placer dans un des cas les moins favorables à l'application de la PMC, loin de la limite de séparabilité.

On a aussi considéré employer une fonction de Jastrow-Slater :

$$\psi_{\mathrm{JS}}(\mathcal{R}) = e^{J(\mathcal{R})}\psi_{\mathrm{S}}(\mathcal{R}) \,, \tag{4.11}$$

où on écrit $J$ sous la forme d'un produit matriciel ${}^t\boldsymbol{U}{}^t\boldsymbol{X}(\mathcal{R})\boldsymbol{J}\boldsymbol{X}(\mathcal{R})\boldsymbol{U}$, où $\boldsymbol{U}$ est le vecteur colonne à $2N$ électrons, ne comportant que des 1, et $\boldsymbol{J}$ est une matrice de diagonale nulle, qui commute avec toute translation de la grille des sites. Cela revient à écrire $J$ comme une somme de produits deux à deux de nombre d'occupations de sites en fonction de leurs positions relatives :

$$J(\mathcal{R}) = j_{0,0,\uparrow\downarrow}\sum_{\vec{i}\in\omega_s} n_{\vec{i}\uparrow}n_{\vec{i}\downarrow} + \frac{j_{0,1,\uparrow\uparrow}}{2}\sum_{\substack{(\vec{i},\vec{j})\in\omega_s^2\\ \boldsymbol{T}_{\vec{i}\vec{j}}=1}} \left[n_{\vec{i}\uparrow}n_{\vec{j}\uparrow} + n_{\vec{i}\downarrow}n_{\vec{j}\downarrow}\right] + j_{0,1,\uparrow\downarrow}\sum_{\substack{(\vec{i},\vec{j})\in\omega_s^2\\ \boldsymbol{T}_{\vec{i}\vec{j}}=1}} n_{\vec{i}\uparrow}n_{\vec{j}\downarrow} + \ldots \tag{4.12}$$

La forme de sous-système la plus évidente, et celle qu'on a choisie, consiste à utiliser pour sous-systèmes des sous-grilles carrées de $l \times l$ sites adjacents, avec $l$ un diviseur entier de $L$. On les munit pour les besoins de l'échantillonage de conditions périodiques aux bords. Les mouvements proposés sont tous des mouvements à un seul électron, et on extrait une configuration tous les $N$ (ou $n$ dans les sous-systèmes) mouvements.

## 4.4 Calcul exact des différences d'énergie cinétique locale

### 4.4.1 Energie cinétique locale pour un déterminant de Slater

Prenons donc une configuration $\mathcal{R}$ quelconque de notre modèle de Hubbard telle que $\psi_{\mathrm{S}}(\mathcal{R}) \neq 0$. Alors l'énergie cinétique locale s'écrit :

$$T_l(\mathcal{R}) = \frac{\hat{T}\psi_{\mathrm{S}}}{\psi_{\mathrm{S}}}(\mathcal{R}) = \frac{-1}{\psi_{\mathrm{S}}(\mathcal{R})}\left[\sum_{\sigma\in\{\uparrow,\downarrow\}}\sum_{\vec{i}\neq\vec{j}\in\omega_s} t_{\vec{i}\vec{j}}\hat{a}^\dagger_{\vec{i}\sigma}\hat{a}_{\vec{j}\sigma}\psi_{\mathrm{S}}\right](\mathcal{R}) \,. \tag{4.13}$$

En pratique, on peut beaucoup simplifier cette expression. En effet, pour chaque site, il existe exactement quatre sites adjacents, et on peut donc se ramener à l'application de 4 translations $\hat{t}_{\vec{\delta}}$ correspondant aux quatre vecteurs $\uparrow\equiv(0,-1)$, $\downarrow\equiv(0,1)$, $\leftarrow\equiv(-1,0)$, et $\rightarrow\equiv(1,0)$. On pose $\omega_d=\{\downarrow,\uparrow,\leftarrow,\rightarrow\}$ pour faciliter la notation, et on éclate plus avant les opérateurs de translation sur les spin-sites en posant $\hat{t}_{\vec{\imath}\sigma\vec{\delta}}=\hat{a}^{\dagger}_{\vec{\imath}+\vec{\delta}\,\sigma}\hat{a}_{\vec{\imath}\sigma}$ :

$$T_l(\mathcal{R})=\frac{-1}{\psi_{\mathrm{S}}(\mathcal{R})}\left[\sum_{(\vec{\imath}\sigma)\in\omega}\sum_{\vec{\delta}\in\omega_d}\hat{t}_{\vec{\imath}\sigma\vec{\delta}}\psi_{\mathrm{S}}\right](\mathcal{R})\,. \tag{4.14}$$

De même, on peut se restreindre aux sites sur lesquels existent des électrons :

$$T_l(\mathcal{R})=-\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\frac{\hat{t}_{\vec{r}_i\vec{\delta}}\psi_{\mathrm{S}}}{\psi_{\mathrm{S}}}\right](\mathcal{R})\,. \tag{4.15}$$

Passons maintenant au formalisme matriciel. Cela nous donne :

$$\begin{aligned}T_l(\mathcal{R})&=-\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\det(\boldsymbol{A}^{-1}(\mathcal{R})\boldsymbol{t}_{\vec{r}_i\vec{\delta}}\boldsymbol{C}\boldsymbol{X}(\mathcal{R}))\right]\\&=-\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\det(\boldsymbol{D}(\mathcal{R})\boldsymbol{t}_{\vec{r}_i\vec{\delta}}\boldsymbol{X}(\mathcal{R}))\right]\,.\end{aligned} \tag{4.16}$$

En pratique, il s'agit d'un déplacement à un seul électron, l'électron $e_i$, donc on peut simplifier par simple application de la formule de Sherman et Morrison. On a alors :

$$T_l(\mathcal{R})=-\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}d_{i,\vec{r}_i+\vec{\delta}}(\mathcal{R})\right]\,. \tag{4.17}$$

On adoptera la notation condensée $d_{i\vec{\delta}}$ pour l'élement de matrice $d_{i,\vec{r}_i+\vec{\delta}}(\mathcal{R})$. Cela nous donne donc l'expression simplifiée :

$$T_l(\mathcal{R})=-\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}d_{i\vec{\delta}}\,. \tag{4.18}$$

Cherchons maintenant à écrire cela sous forme matricielle. Comme il s'agit d'une somme, nous allons chercher à mettre cette expression sous la forme d'une trace. Il nous faut donc une première étape qui transforme $i$ en $\vec{r}_i$, une qui transforme $\vec{r}_i$ en une somme des $\vec{r}_i+\vec{\delta}$, et enfin la matrice $\boldsymbol{D}$. Cela nous donne :

$$T_l(\mathcal{R})=-\mathrm{Tr}(\boldsymbol{D}\boldsymbol{T}\boldsymbol{X})\,. \tag{4.19}$$

Ce résultat est général et s'applique pour un système de Hubbard quelconque, pas seulement notre grille bidimensionnelle. Qui plus est, il se généralise assez bien au cas d'un système continu ; il suffit de remplacer la matrice d'adjacence $\boldsymbol{T}$ par le laplacien à $N$ électrons.

### 4.4.2 Développement pour un sous-système

Supposons que l'on dispose de deux configurations $\mathcal{R}$ et $\mathcal{R}'$ ne différant que par un sous-système $\mathcal{S}_i$. On adoptera les notations rapides suivantes : $\boldsymbol{X}$ pour $\boldsymbol{X}(\mathcal{R})$, $\boldsymbol{X}'$ pour $\boldsymbol{X}(\mathcal{R}')$, et ainsi de suite pour $\boldsymbol{D}$ et $\boldsymbol{A}$. On notera $\boldsymbol{X}_2={}^t\boldsymbol{Q}_i\boldsymbol{X}'\boldsymbol{P}_i$, $\boldsymbol{C}_2={}^t\boldsymbol{P}_i\boldsymbol{D}\boldsymbol{Q}_i$, et $\boldsymbol{D}_2={}^t\boldsymbol{P}_i\boldsymbol{D}'\boldsymbol{Q}_i$ les matrices réduites pour $\boldsymbol{X}'$, $\boldsymbol{D}$ et $\boldsymbol{D}'$ pour le sous-système $\mathcal{S}_i$, et de même $\boldsymbol{A}_2=\boldsymbol{C}_2\boldsymbol{X}_2$.

On va chercher à calculer la valeur de $T_l(\mathcal{R}')=\mathrm{Tr}(\boldsymbol{D}'\boldsymbol{T}\boldsymbol{X}')$ connaissant celle de $T_l(\mathcal{R})$. Pour cela, on va se servir de l'égalité suivante :

$$\boldsymbol{D}'=(\boldsymbol{C}\boldsymbol{X}')^{-1}\boldsymbol{C}=(\boldsymbol{C}\boldsymbol{X}')^{-1}(\boldsymbol{C}\boldsymbol{X})(\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{C}=((\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{C}\boldsymbol{X}')^{-1}\boldsymbol{D}=(\boldsymbol{D}\boldsymbol{X}')^{-1}\boldsymbol{D}\,. \tag{4.20}$$

Cherchons alors à développer $\boldsymbol{DX'}$ par blocs :

$$\boldsymbol{DX'} = \begin{bmatrix} \boldsymbol{I}_{\mathcal{S}_{\bar{i}}} & \boldsymbol{N} \\ 0 & \boldsymbol{A}_2 \end{bmatrix} , \tag{4.21}$$

où $\boldsymbol{N} = {}^t\boldsymbol{P}_{\bar{i}}\boldsymbol{DX'}\boldsymbol{P}_i$ est la partie de $\boldsymbol{DX'}$ qui provient du mouvement des électrons de $\mathcal{S}_i$ mais n'entre pas en jeu dans le déterminant. On va alors prendre l'inverse de cette matrice par blocs :

$$(\boldsymbol{DX'})^{-1} = \begin{bmatrix} \boldsymbol{I}_{\mathcal{S}_{\bar{i}}} & -\boldsymbol{N}\boldsymbol{A}_2^{-1} \\ 0 & \boldsymbol{A}_2^{-1} \end{bmatrix} . \tag{4.22}$$

Si l'on repart de l'équation (4.19), on obtient l'équation suivante :

$$\begin{aligned} T_l(\mathcal{R}') &= -\operatorname{Tr}(\boldsymbol{D'TX'}) = -\operatorname{Tr}((\boldsymbol{DX'})^{-1}\boldsymbol{DTX'}) \\ &= -\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\sum_{j=1}^{N}((\boldsymbol{DX'})^{-1})_{ij}d_{j,\vec{r}_i'+\vec{\delta}} . \end{aligned} \tag{4.23}$$

Décomposons $i$ et $j$ entre environnement et sous-système :

$$\begin{aligned} T_l(\mathcal{R}') &= -\sum_{i=1}^{n}\sum_{\vec{\delta}\in\omega_d}\sum_{j=1}^{n}((\boldsymbol{DX'})^{-1})_{ij}d_{j,\vec{r}_i'+\vec{\delta}} - \sum_{i=n+1}^{N}\sum_{\vec{\delta}\in\omega_d}\sum_{j=1}^{N}((\boldsymbol{DX})^{-1})_{ij}d_{j,\vec{r}_i'+\vec{\delta}} \\ &= -\sum_{i=1}^{n}\sum_{\vec{\delta}\in\omega_d}(\boldsymbol{A}_2^{-1}\boldsymbol{C}_2)_{i,\vec{r}_i'+\vec{\delta}} - \sum_{i=n+1}^{N}\sum_{\vec{\delta}\in\omega_d}d_{i,\vec{r}_i+\vec{\delta}} - \sum_{i=n+1}^{N}\sum_{\vec{\delta}\in\omega_d}\sum_{j=1}^{n}((\boldsymbol{DX})^{-1})_{ij}d_{j,\vec{r}_i'+\vec{\delta}} . \end{aligned} \tag{4.24}$$

On retranche alors $T_l(\mathcal{R})$ des deux côtés pour simplifier l'expression :

$$T_l(\mathcal{R}') - T_l(\mathcal{R}) = -\sum_{i=1}^{n}\sum_{\vec{\delta}\in\omega_d}(d'_{i\vec{\delta}} - d_{i\vec{\delta}}) + \sum_{i=n+1}^{N}\sum_{\vec{\delta}\in\omega_d}\sum_{j=1}^{n}\sum_{k=1}^{n}d_{i,\vec{r}_k'}(\boldsymbol{A}_2^{-1})_{kj}d_{j,\vec{r}_i+\vec{\delta}} . \tag{4.25}$$

La somme sur les électrons du sous-système a un coût de calcul en $\mathcal{O}\left(n^3\right)$, et celle sur les électrons de l'environnement un coût de calcul en $\mathcal{O}\left(Nn^2\right)$. On peut cependant effectuer en amont de la sous-dynamique les sommes sur $i$ et $\vec{\delta}$ de sorte à se ramener à un coût par configuration en $\mathcal{O}\left(n^2\right)$ pour le second terme.

### 4.4.3 Extension pour une fonction d'onde de Jastrow-Slater

Pour une fonction d'onde de Jastrow-Slater, on doit repartir de l'équation (4.15) en replaçant $\psi_{\mathrm{S}}$ par $\psi_{\mathrm{JS}}$ :

$$T_l(\mathcal{R}) = -\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\frac{\hat{t}_{\vec{r}_i\vec{\delta}}\psi_{\mathrm{JS}}}{\psi_{\mathrm{JS}}}\right](\mathcal{R}) . \tag{4.26}$$

Cette expression se factorise aisément en reprenant l'expression de $\psi_{\mathrm{JS}}$ :

$$T_l(\mathcal{R}) = -\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\frac{\hat{t}_{\vec{r}_i\vec{\delta}}e^J}{e^J}d_{i\vec{\delta}}\right](\mathcal{R}) . \tag{4.27}$$

Pour le calcul de cette expression, on construit la matrice $\boldsymbol{K}$ telle que :

$$(\boldsymbol{K})_{i\vec{j}} = \sum_{\substack{k=1 \\ k\neq i}}^{N}(\boldsymbol{J})_{\vec{r}_k\vec{j}} . \tag{4.28}$$

On peut alors réécrire l'expression de l'énergie cinétique sous la forme suivante :

$$T_l(\mathcal{R}) = -\left[\sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d}\exp(2((\boldsymbol{K})_{i\vec{r}_i+\vec{\delta}}-(\boldsymbol{K})_{i\vec{r}_i}))d_{i\vec{\delta}}\right](\mathcal{R}) \equiv \sum_{i=1}^{N}\sum_{\vec{\delta}\in\omega_d} z_{i\vec{\delta}}d_{i\vec{\delta}}\,. \tag{4.29}$$

Si on suppose que la portée du Jastrow est finie ($\exists l_j < L/2$, $\max(|x_1-x_2|,|y_1-y_2|) > l_j \Rightarrow (\boldsymbol{J})_{\vec{r}_1\vec{r}_2} = 0$), alors on peut faire une division entre environnement proche, qui interagit avec le système au travers du Jastrow, et environnement lointain qui ne le fait pas. On les notera respectivement $\mathcal{S}_{\bar{i}j}$ et $\mathcal{S}_{\bar{i}\bar{j}}$.

Si l'on reprend et étend les notations de la sous-section précédente, alors on a :

$$\begin{aligned}
T_l(\mathcal{R}') - T_l(\mathcal{R}) =& \sum_{\substack{i\in\mathcal{S}\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) \\
=& \sum_{\substack{i\in\mathcal{S}_i\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) + \sum_{\substack{i\in\mathcal{S}_{\bar{i}j}\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) + \sum_{\substack{i\in\mathcal{S}_{\bar{i}\bar{j}}\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) \\
=& \sum_{\substack{i\in\mathcal{S}_i\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) + \sum_{\substack{i\in\mathcal{S}_{\bar{i}j}\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}} - z_{i\vec{\delta}})d_{i\vec{\delta}} + \sum_{\substack{i\in\mathcal{S}_{\bar{i}j}\\ \vec{\delta}\in\omega_d}} z'_{i\vec{\delta}}(d'_{i\vec{\delta}} - d_{i\vec{\delta}}) + \sum_{\substack{i\in\mathcal{S}_{\bar{i}\bar{j}}\\ \vec{\delta}\in\omega_d}} z_{i\vec{\delta}}(d'_{i\vec{\delta}} - d_{i\vec{\delta}})\,.
\end{aligned} \tag{4.30}$$

Le premier de ces quatre termes correspond à l'énergie cinétique liée au sous-système, le second à la modification de l'énergie cinétique des électrons proches suite au Jastrow, le troisième à celle des électrons proches liée au déterminant de Slater, et le dernier à la correction liée au déterminant des systèmes lointains. On peut alors décomposer de cette manière :

$$\begin{aligned}
T_l(\mathcal{R}') - T_l(\mathcal{R}) =& \sum_{\substack{i\in\mathcal{S}_i\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}}d'_{i\vec{\delta}} - z_{i\vec{\delta}}d_{i\vec{\delta}}) + \sum_{\substack{i\in\mathcal{S}_{\bar{i}j}\\ \vec{\delta}\in\omega_d}} (z'_{i\vec{\delta}} - z_{i\vec{\delta}})d_{i\vec{\delta}} \\
&+ \sum_{\substack{i\in\mathcal{S}_{\bar{i}j}\\ \vec{\delta}\in\omega_d}} \sum_{(j,k)\in\mathcal{S}_i^2} d_{i,\vec{r}'_k}(\boldsymbol{A}_2^{-1})_{kj} d_{j,\vec{r}_i+\vec{\delta}} z'_{i\vec{\delta}} \\
&+ \sum_{(j,k)\in\mathcal{S}_i^2} (\boldsymbol{A}_2^{-1})_{kj} \sum_{\substack{i\in\mathcal{S}_{\bar{i}\bar{j}}\\ \vec{\delta}\in\omega_d}} d_{j,\vec{r}_i+\vec{\delta}} z_{i\vec{\delta}} d_{i,\vec{r}'_k}\,.
\end{aligned} \tag{4.31}$$

Le premier terme est en $\mathcal{O}\left(n^3\right)$, le second en $\mathcal{O}\left(n^{3/2}l_j\right)$, le troisième en $\mathcal{O}\left(n^{5/2}l_j\right)$, et le troisième en $\mathcal{O}\left(n^2\right)$. On arrive bel et bien à quelque chose environ de l'ordre de $\mathcal{O}\left(n^3\right)$, mais le troisième terme est celui des termes environnementaux dont le calcul est le plus coûteux.

## 4.5 Amélioration de l'échantillonnage par intégration des sous-dynamiques

Il est facile de se rendre compte qu'échantillonner à l'aide de la méthode présentée ci-dessus pour les sous-systèmes est bien moins coûteuse qu'échantillonner sur le système global. Ainsi, si on divise un système à $N$ électrons en $p$ sous-systèmes comportant $n$ particules chacun, réaliser une nouvelle configuration coûte $\mathcal{O}\left(N^3\right)$ pour le système global et $\mathcal{O}\left(pn^3\right)$ pour les sous-systèmes. On peut donc envisager un gain en $\mathcal{O}\left(p^2\right)$. Ainsi, dans un système muni d'une longueur de corrélation finie, on peut imaginer que la taille optimale des sous-systèmes converge vers un nombre fini, permettant un gain global en $\mathcal{O}\left(N^2\right)$.

On a donc tout intérêt à se servir le plus possible des sous-dynamiques. Outre la réduction de la variance, nous avons envisagé de réduire également la longueur de corrélation entre configurations successives en intégrant les sous-dynamiques à la dynamique principale. En effet, bien que cela impose des modifications à l'estimateur pratique que l'on emploie, le nouvel estimateur que l'on peut construire pour représenter les calculs effectués n'en est pas biaisé pour autant, pour peu que l'on maintienne l'ergodicité de l'échantillonnage global. Cet estimateur est même zéro-variant dans la limite de séparabilité. La succession des sous-dynamiques peut alors être comparée à un échantillonneur de Gibbs (voir [3], pp 337-424, pour une étude en profondeur de l'échantilloneur de Gibbs).

Comment s'assurer alors de l'ergodicité ? En effet, si la forme d'échantillonneur de Gibbs nous permet de s'assurer de la représentation de l'interaction entre sous-systèmes, lorsque nos sous-dynamiques sont réalisées sur des sous-systèmes réalisant une partition, on ne peut pas prendre en compte les possibilités d'échange entre sous-systèmes. Pour ce faire, on a deux choix. Soit conserver une dynamique principale, coûteuse, soit abandonner la partition – et par là-même l'estimateur zéro-variant dans la limite de séparabilité par construction – pour permettre l'échange entre sous-systèmes. Cet échange prend la forme, dans un découpage de la liste des électrons induisant une partition de l'espace, d'un recouvrement sur les sous-listes entrant en jeu, permettant un échange d'espace ; et pour un découpage dans l'espace en secteurs induisant une partition de la liste des électrons, d'un recouvrement sur les secteurs mis en jeu, permettant un échange d'électrons.

L'algorithme sur le modèle de Hubbard avec lequel j'ai travaillé a pour temps de corrélation 1, ce qui ne permet pas l'emploi de ces idées pour réduire le temps de corrélation ; et nous ne présenterons pas de résultats à ce sujet dans le chapitre suivant. Cependant, elle a été testée en pratique pour des séparations coeur-valence dans le travail effectué avec Jonas Feldt et Roland Assaraf, ce qui a mené à l'article présenté dans l'annexe D.

## Récapitulatif

Dans ce chapitre, nous avons élaboré deux variantes possibles à la méthode de Monte Carlo Partitionnelle : une variante qui utilise une partition hiérarchisée dans une approche multi-échelle, et une qui cherche à intégrer les sous-dynamiques à la dynamique principale afin de réduire le temps de corrélation.

Nous nous sommes également intéressés aux astuces qui rendent efficace la méthode PMC : d'une part, le calcul accéléré dans les sous-dynamiques grâce à la factorisation matricielle, qui permet pour un sous-système à $n$ électrons de se ramener à un coût par configuration des sous-dynamiques de $\mathcal{O}\left(n^3\right)$ ; et d'autre part, le calcul de la valeur exacte de l'énergie cinétique dans le modèle de Hubbard pour une fonction d'onde de type Jastrow-Slater.

Nous avons également décrit le système modèle en question, ainsi que montré que pour une partition en $p$ sous-systèmes, on peut s'attendre à un gain réel se comportant au mieux en $\mathcal{O}\left(p^2\right)$.

Maintenant, ils nous faut passer à la pratique, et fournir des résultats.

### Bibliographie

[1] C. Filippi, R. Assaraf, and S. Moroni. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *Journal of Chemical Physics*, 144 :194105, 2016.

[2] M Cyrot. The Hubbard hamiltonian. *Physica B+ C*, 91 :141–150, 1977.

[3] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2 edition, 2004.

# Chapitre 5

# Résultats

Dans ce chapitre, nous présenterons les résultats obtenus sur l'application de la méthode PMC à des systèmes de Hubbard. Dans un premier temps, nous vérifierons que la méthode PMC, ainsi que son implémentation récursive, n'introduisent pas de biais. Ensuite, nous explorerons le comportement de la méthode PMC en fonction de la variation de trois paramètres clé. Enfin, nous conclurons par la présentation des gains issus de la méthode MS-PMC. Pour les notations employées à ce chapitre, voir l'annexe A.

## 5.1 Convergence

Dans cette section, notre objectif est d'observer si la méthode PMC et son implémentation récursive (MS-PMC) mettent en jeu des estimateurs biaisés en pratique (et en dépit de la théorie sous-jacente). Pour ce faire, nous utiliserons une unique simulation et comparerons les résultats obtenus sur celle-ci en utilisant tout ou partie de l'information. En effet, comme on l'a cité à la section 4.1, la méthode MS-PMC utilise une variable de contrôle par partition, liée aux sous-dynamiques sur cette partition. On peut donc élaguer notre arbre des trajectoires, pour n'exploiter que la dynamique principale si on veut les résultats en VMC, et n'exploiter que la dynamique principale et les sous-dynamiques qui s'y rattachent directement si on veut les résultats en PMC.

Le système auquel on s'intéresse est un système de Hubbard carré de taille $L = 27$, à moitié rempli et muni de sa fonction d'onde Hartree-Fock (pour plus de détails voir section 4.3). On le munit d'une partition $\mathcal{P}_1$ en sous-systèmes carrés de taille $l = 9$, et d'une partition $\mathcal{P}_2$ en sous-systèmes de taille $\mathfrak{l} = 3$. La simulation que nous avons effectuée consistait d'une dynamique principale de longueur $M = 5000$, avec des sous-dynamiques sur $\mathcal{P}_1$ de longueur $m = 64$, et sur $\mathcal{P}_2$ de longueur $\mathfrak{m} = 16$.

Notre première sous-section se concentre sur la comparaison entre l'estimateur VMC et l'estimateur pratique PMC ; la seconde s'intéresse à l'estimateur pratique MS-PMC.

### 5.1.1 Comparaison VMC/PMC

Dans le cadre de cette étude de convergence, nous avons décidé d'optimiser le coefficient de la variable de contrôle $c$ pour chacun des points de la simulation de manière à représenter à chaque point le résultat final qui aurait été obtenu si la simulation avait été arrêtée à ce point. La formule des deux estimateurs réels que nous comparons est alors donnée ci-dessous, pour une dynamique principale de longueur partielle $x$ :

$$E_{\mathrm{VMC}}(x) = \frac{1}{x}\sum_{K=1}^{x} E_l(\mathcal{R}^K) \; ;$$

$$E_{\mathrm{PMC}}(x) = E_{\mathrm{VMC}}(x) + \frac{c(x)}{mx}\sum_{K=1}^{x}\sum_{\mathcal{S}_i\in\mathcal{P}_1}\sum_{k=1}^{m}\left[E_l(\mathcal{R}_i^{Kk}, \mathcal{R}_{\tilde{\imath}}^K) - E_l(\mathcal{R}^K)\right] \equiv \frac{1}{x}\sum_{K=1}^{x} E_l(\mathcal{R}^K) + c(x)\Delta^1 E(K) \; ;$$

où nous avons introduit la notation $\Delta^1 E(K)$ pour représenter la variable de contrôle PMC $\tilde{E}_{\mathrm{pr}} - E_l$.

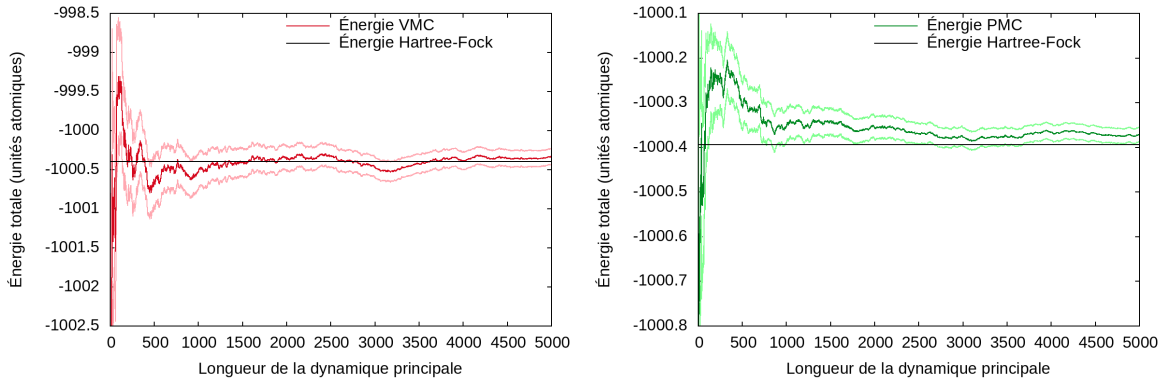


FIGURE 5.1 – Convergence de l'énergie en fonction de la longueur de la dynamique principale, en VMC (gauche) et PMC (droite). Les courbes plus claires correspondent aux barres d'erreur.

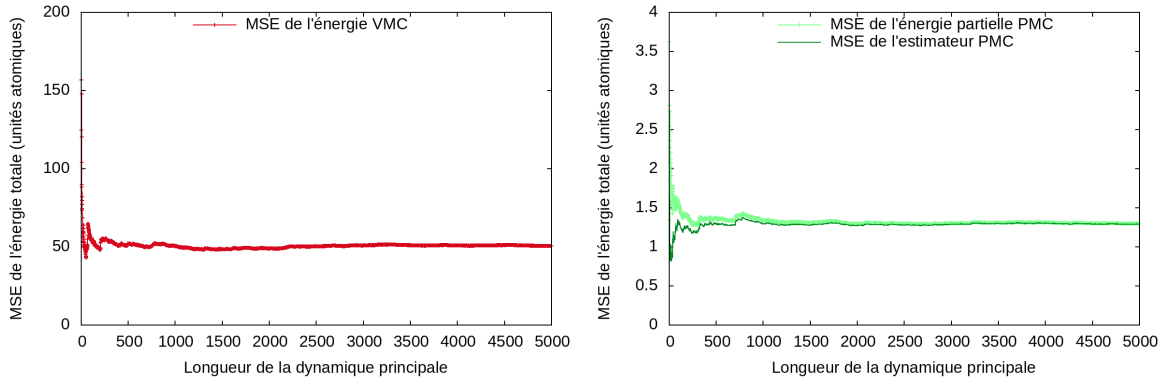


FIGURE 5.2 – Erreurs quadratiques de l'énergie en fonction de la longueur de la dynamique principale, en VMC (gauche) et PMC (droite).

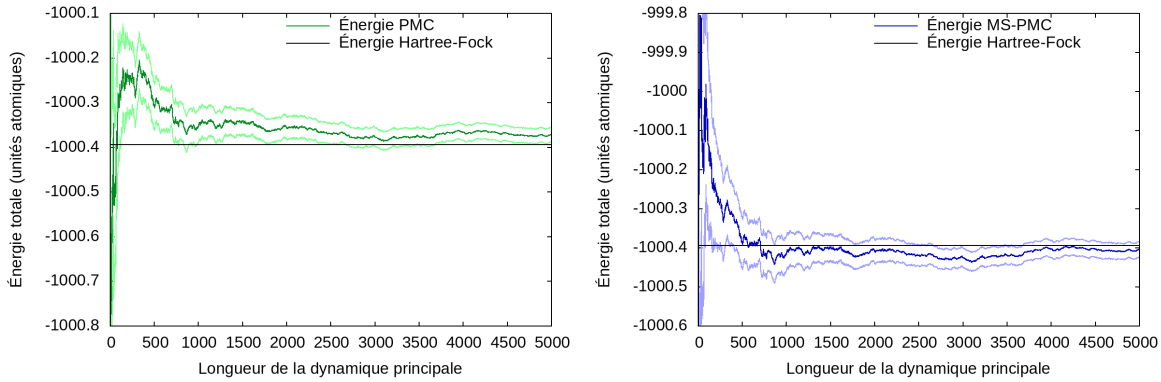


FIGURE 5.3 – Convergence de l'énergie en fonction de la longueur de la dynamique principale, en PMC (gauche) et MS-PMC (droite). Les courbes plus claires correspondent aux barres d'erreur.

Dans la figure 5.1, nous présentons l'évolution de la valeur prise par ces variables, munie de sa barre d'erreur à mesure que la dynamique principale progresse, en VMC (a) et en PMC (b). Dans chacun des deux graphes, on a placé l'énergie variationnelle obtenue par Hartree-Fock comme valeur de référence. La différence d'échelle entre les deux graphes – d'un facteur 5 – est significative et montre bien à quel point la convergence est accélérée en PMC.

Nous disposons d'un certain nombre de moyens d'estimer la vitesse de convergence ainsi que l'existence ou non d'un biais. Premièrement, on peut faire appel au nombre moyen, au cours de la dynamique, de déviations standard entre la valeur observée et l'énergie variationnelle :

$$\overline{(n_{\delta E})_{\mathrm{VMC}}} = \frac{1}{M}\sum_{x=2}^{M}\frac{|E_v - E_{\mathrm{VMC}}(x)|}{\sigma(x)} \; ;$$
$$\overline{(n_{\delta E})_{\mathrm{PMC}}} = \frac{1}{M}\sum_{x=2}^{M}\frac{|E_v - E_{\mathrm{PMC}}(x)|}{\tilde{\sigma}(x)} \; ;$$

où $\sigma(x)$ fait référence à l'écart-type de la simulation VMC interrompue à cet instant, obtenu à partir de la variance de $E_l$, et $\tilde{\sigma}(x)$ à l'écart type de la simulation PMC interrompue à cet instant, obtenu à partir de la variance minimisée de $\tilde{E}_{\mathrm{pr}}$. La valeur pour l'estimateur VMC prend la valeur de 0.62 sur les 500 premiers points et de 0.39 au bout des 5000 ; l'estimateur PMC de 1.77 et 1.34 respectivement. Si biais il y a, cela suggère donc que celui-ci décroît plus vite que $\mathcal{O}\left(1/\sqrt{M}\right)$.

Une autre quantité qu'on peut chercher à employer est l'erreur quadratique moyenne (ou MSE, pour *Mean Square Error*). Si cette quantité est facile à construire pour l'estimateur VMC ($\mathrm{MSE}(E)_{\mathrm{VMC}}$), on peut l'obtenir de deux manières différentes suivant la manière dont on décide d'interpréter l'énergie PMC. La première des deux interprétations revient à considérer que l'énergie PMC se comporte comme une pseudo-valeur moyenne d'une énergie $E'(K) = KE_{\mathrm{PMC}}(K) - (K-1)E_{\mathrm{PMC}}(K-1)$, et donc de calculer l'erreur quadratique moyenne de l'énergie PMC ($\mathrm{MSE}(E)_{\mathrm{PMC}}$) comme l'erreur quadratique moyenne de $E'$. Cette interprétation revient donc à considérer le coefficient $c$ comme étant variable. L'autre interprétation revient à considérer que sur la dynamique de longueur $x$, le coefficient $c$ n'est pas variable, mais constant. On peut alors utiliser la variance minimisée pour calculer l'erreur quadratique moyenne $\mathrm{MSE}(E_{\mathrm{PMC}})$). L'expression de ces trois quantités est fournie ci-dessous, et représentée dans la figure 5.2.

$$\mathrm{MSE}(E)_{\mathrm{VMC}}(x) = \frac{1}{x}\sum_{K=1}^{x}\left(E_l(\mathcal{R}^K) - E_{\mathrm{HF}}\right)^2 \; ;$$
$$\mathrm{MSE}(E)_{\mathrm{PMC}}(x) = \frac{1}{x}\sum_{K=1}^{x}\left(KE_{\mathrm{PMC}}(K) - (K-1)E_{\mathrm{PMC}}(K-1) - E_{\mathrm{HF}}\right)^2 \; ;$$
$$\mathrm{MSE}(E_{\mathrm{PMC}})(x) = \frac{1}{x}\sum_{K=1}^{x}\left[E(\mathcal{R}^K) + \frac{c(x)}{m}\sum_{\mathcal{S}_i\in\mathcal{P}_1}\sum_{k=1}^{m}\left[E_l(\mathcal{R}_i^{Kk},\mathcal{R}_{\bar{i}}^K) - E_l(\mathcal{R}^K)\right]\right]^2 - E_{\mathrm{PMC}}(x)^2$$
$$+ \left(E_{\mathrm{PMC}}(x) - E_{\mathrm{HF}}\right)^2 .$$

On observe que là où la VMC converge vers une valeur de l'erreur quadratique moyenne de l'ordre de 50, celle vers laquelle convergent les deux fonctions d'erreur quadratique moyenne en PMC est de l'ordre de 1,3. De plus, les deux fonctions d'erreur quadratique semblent converger vers une même valeur, ce qui suggère que le coefficient $c(x)$ est assez stable et qu'on n'a pas d'effets de surcompensation d'un pas à l'autre. On peut ainsi envisager que la variance a été réduite d'un facteur total entre 30 et 40.

### 5.1.2 Comparison PMC/MS-PMC

Rajoutons maintenant la partition inférieure et la variable de contrôle qui y est rattachée. L'expression de l'estimateur MS-PMC est donnée ci-dessous (en faisant disparaître dans la notation, mais

sans l'oublier, l'environnement $\mathcal{R}_{\tilde{i}}^{K}$ ) :

$$E_{\text{MS-PMC}}(x) = E_{\text{VMC}}(x) + c_1(x)\Delta^1 E(K) + \sum_{\substack{K\in[[1,x]]\\ k\in[[1,m]]}} \sum_{\mathcal{S}_i\in\mathcal{P}_1} \frac{c_2(x)}{xm\mathfrak{m}} \sum_{\substack{\mathcal{S}_j\in\mathcal{P}_2\mathfrak{K}=1\\ \mathcal{S}_j\subset\mathcal{S}_i}}^{\mathfrak{m}} \left[E_l(\mathcal{R}_{ij}^{Kk\mathfrak{K}}, \mathcal{R}_{i\tilde{j}}^{Kk}) - E_l(\mathcal{R}_i^{Kk})\right]$$

$$\equiv \frac{1}{x}\sum_{K=1}^{x} E_l(\mathcal{R}^K) + c_1(x)\Delta^1 E(K) + c_2(x)\Delta^2 E(K)\ .$$

Nous avons représenté son évolution à côté de celle de l'estimateur PMC à la figure 5.3. On voit, à l'échelle et à la courbe, que la convergence en MS-PMC est du même ordre de grandeur (et de vitesse) qu'en PMC. Le calcul de $\overline{(n_{\delta E})_{\text{MS-PMC}}}$ nous donne 0.91 déviation standard d'écart en moyenne sur les 500 premiers points, et 0.67 déviation standard sur la dynamique au complet. Quand on reconstruit les erreurs quadratiques moyennes en MS-PMC, on arrive aux expressions suivantes :

$$\text{MSE}(E)_{\text{MS-PMC}}(x) = \frac{1}{x}\sum_{K=1}^{x}\left(KE_{\text{MS-PMC}}(K) - (K-1)E_{\text{MS-PMC}}(K-1) - E_{\text{HF}}\right)^2 ;$$

$$\text{MSE}(E_{\text{MS-PMC}})(x) = \frac{1}{x}\sum_{K=1}^{x}\left[E_l(\mathcal{R}^K) + c_1(x)\Delta^1 E(K) + c_2(x)\Delta^2 E(K)\right]^2 - E_{\text{MS-PMC}}(x)^2 + \left(E_{\text{MS-PMC}}(x) - E_{\text{HF}}\right)^2 ;$$

où nous utilisons la notation $\Delta^2 E(K)$ pour représenter la variable de contrôle supplémentaire introduite en MS-PMC, et correspondant à la seconde partition $\mathcal{P}_2$.

La figure 5.4 représente côte-à-côte ces quantités pour l'estimateur PMC et pour l'estimateur MS-PMC.

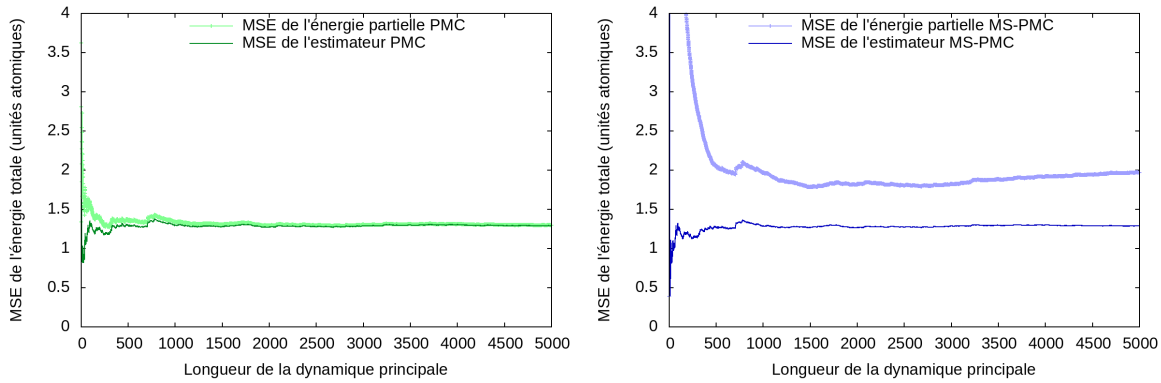


FIGURE 5.4 – Erreurs quadratiques de l'énergie en fonction de la longueur de la dynamique principale, en PMC (gauche) et MS-PMC (droite).

On voit que l'erreur quadratique moyenne observée est du même ordre pour l'énergie MS-PMC traitée en temps qu'estimateur global, mais si on la traîte comme la moyenne d'une variable aléatoire, on se retrouve avec une erreur quadratique moyenne qui converge vers quelque chose proche de 2. Cela s'interprète par des coefficients des variables de contrôle significativement plus instables, avec des effets de surcompensation de l'énergie moyenne.

## 5.2 Optimisation de la méthode PMC

Dans cette partie, nous explorerons dans un premier temps le comportement de la méthode PMC en fonction des trois hyperparamètres clé de la simulation : la taille du système $L$, la taille des sous-systèmes $l$, et la longueur des sous-dynamiques $m$, que nous traiterons un par un dans l'ordre inverse. Ce

faisant, nous définirons trois fonctions de gain qui nous permettront de caractériser les comportements observés.

Dans cette section, sauf mention contraire les simulations ont lieu avec une dynamique principale de longueur $M = 500$.

### 5.2.1 Variance en fonction de la longueur des sous-dynamiques

Nous avons démontré à la section 3.5 une relation explicite de dépendence de la variance de la variable corrigée $\tilde{X}_{\text{pr}}$ en fonction de la longueur des sous-dynamiques $m$, dont nous rappellons la forme ci-dessous :

$$\text{Var}\left(\tilde{X}_{\text{pr}}\right) \underset{\infty}{=} V_\infty + \frac{\delta V}{m} + \mathcal{O}\left(\frac{1}{m^2}\right) .$$

C'est, avant toute autre chose, ce que nous avons cherché à vérifier. Pour cela, nous avons pris un système de taille $L = 20$, muni de sous-systèmes de taille $l = 5$, et calculé la variance de l'estimateur amélioré pour des valeurs de $m$ variant de 1 à 1024, dont nous avons représenté l'évolution à la figure 5.5.

On observe bel et bien une forme qui ressemble à une fonction inverse. Pour confirmer, nous avons effectué une régression linéaire en fonction de $1/m$, qui doit nous donner une droite. Cette régression est représentée à la figure 5.6.

On obtient bel et bien une droite, ce qui confirme la forme de l'expression démontrée à la section 3.5.

### 5.2.2 Fonctions de gain

Pour aller plus avant, on doit maintenant construire des fonctions de gain, qui nous permettront de caractériser le comportement de la méthode PMC. Comme l'idée consiste à faire diminuer la variance pour un faible coût calculatoire, on va pour commencer introduire le gain en variance $G_v$ et le facteur d'augmentation du temps de calcul $G_t$ définis par :

$$\begin{aligned} G_v &= \frac{\text{Var}\left(E_l\right)}{\text{Var}\left(\tilde{E}_{\text{pr}}\right)} , \\ G_t &= \frac{\tau_{\text{PMC}}}{\tau_{\text{VMC}}} \end{aligned} \quad , \tag{5.1}$$

où $\tau_{\text{PMC}}$ représente le temps de calcul pour une dynamique de longueur $M = 500$ en PMC et $\tau_{\text{VMC}}$ le temps pour un calcul équivalent en VMC. Enfin, on introduit ce qu'on appellera le gain en efficacité calculatoire, ou "gain réel", $G_r$, défini comme le ratio des coûts calculatoires (qui se simplifie dans notre cas où la longueur de corrélation sur la dynamique principale est de 1) :

$$G_r = \frac{\sigma^2_{\text{VMC}} \tau_{\text{VMC}}}{\sigma^2_{\text{PMC}} \tau_{\text{PMC}}} \equiv \frac{\text{Var}\left(E_l\right) \tau_{\text{VMC}}}{\text{Var}\left(\tilde{E}_{\text{pr}}\right) \tau_{\text{PMC}}} = \frac{G_v}{G_t} . \tag{5.2}$$

Il est alors évident que $G_r$ est la quantité que l'on va chercher à maximiser.

### 5.2.3 Fonctions de gain et longueur des sous-dynamiques

Reprenons maintenant les simulations que l'on a effectuées pour la figure 5.5, et extrayons-en les valeurs prises par les fonctions de gain. Les valeurs qu'on obtient sont données dans le tableau 5.1, et représentées en échelles log-log dans la figure 5.7.

On observe, très clairement, trois régimes, qu'on triera par $m$ croissant.

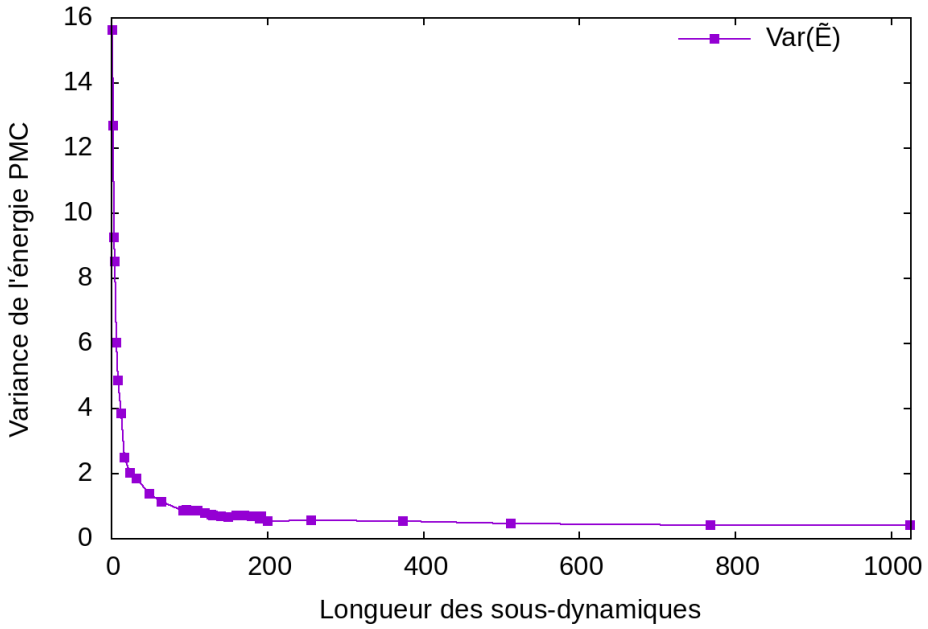


FIGURE 5.5 – Evolution de la variance de l'estimateur amélioré PMC de l'énergie totale en fonction de la longueur des sous-dynamiques.

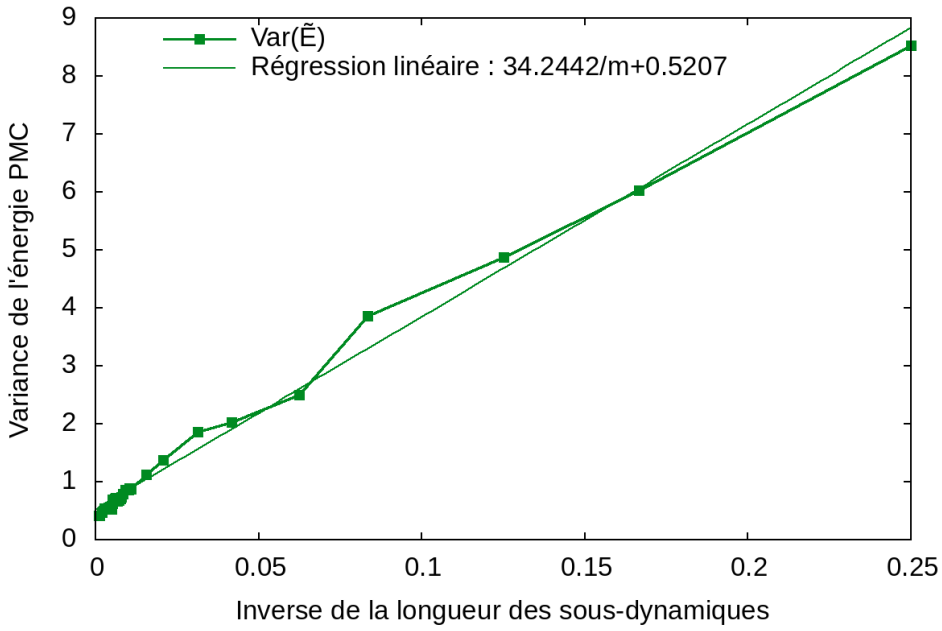


FIGURE 5.6 – Régression linéaire de la variance de l'estimateur amélioré PMC de l'énergie totale par rapport à l'inverse de la longueur des sous-dynamiques.

TABLE 5.1 – Gain de variance $G_v$, facteur d'augmentation du temps de calcul $G_t$, et gain réel $G_r$ pour le calcul de l'énergie en PMC par rapport au VMC, pour différentes longueurs des sous-dynamiques $m$. La valeur en rouge est la valeur optimale de $m$ obtenue par régressions linéaires.

| $m$ | 1 | 2 | 3 | 4 | 6 | 8 | 12 | 16 | 24 | 32 | 48 | 64 | 68 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $G_v$ | 1.758 | 2.165 | 2.965 | 3.220 | 4.556 | 5.634 | 7.117 | 11.00 | 13.54 | 14.77 | 19.95 | 24.33 | 27.62 |
| $G_t$ | 1.026 | 1.031 | 1.039 | 1.048 | 1.061 | 1.074 | 1.103 | 1.125 | 1.179 | 1.233 | 1.328 | 1.428 | 1.456 |
| $G_r$ | 1.714 | 2.099 | 2.855 | 3.074 | 4.296 | 5.244 | 6.450 | 9.773 | 11.49 | 11.97 | 15.02 | 17.04 | 18.96 |
| $m$ | 72 | 76 | 80 | 84 | 88 | 92 | 96 | 100 | 104 | 108 | 110 | 112 | 120 |
| $G_v$ | 26.27 | 27.84 | 26.21 | 27.24 | 30.68 | 31.42 | 31.11 | 31.83 | 30.97 | 34.01 | 31.81 | 31.47 | 34.56 |
| $G_t$ | 1.577 | 1.614 | 1.603 | 1.551 | 1.540 | 1.569 | 1.601 | 1.628 | 1.766 | 1.805 | 1.766 | 1.786 | 1.842 |
| $G_r$ | 16.65 | 17.24 | 16.34 | 17.57 | 19.92 | 20.03 | 19.42 | 19.56 | 17.53 | 18.85 | 18.02 | 17.62 | 18.76 |
| $m$ | 128 | 130 | 140 | 150 | 160 | 170 | 180 | 190 | 192 | 200 | 256 | 384 | 512 |
| $G_v$ | 37.56 | 38.62 | 40.08 | 41.24 | 37.72 | 38.64 | 39.25 | 43.66 | 39.67 | 51.49 | 48.88 | 52.98 | 58.24 |
| $G_t$ | 1.974 | 1.993 | 1.974 | 2.100 | 2.225 | 2.189 | 2.238 | 2.302 | 2.265 | 2.315 | 2.671 | 3.486 | 4.097 |
| $G_r$ | 19.03 | 19.38 | 20.30 | 19.64 | 16.95 | 17.65 | 17.53 | 18.97 | 17.52 | 22.25 | 18.30 | 15.20 | 14.21 |
| $m$ | 768 | 1000 | | | | | | | | | | | |
| $G_v$ | 66.42 | 66.82 | | | | | | | | | | | |
| $G_t$ | 5.974 | 7.859 | | | | | | | | | | | |
| $G_r$ | 11.12 | 8.804 | | | | | | | | | | | |

Dans le premier, l'augmentation du temps de calcul demeure essentiellement négligeable, et la diminution de variance est l'aspect dominant de la variation du gain réel. Cela s'interprète par le fait que le coût des sous-dynamiques demeure faible, mais l'augmentation de l'espace configurationnel exploré est importante.

Dans le second, le gain en variance commence à converger vers sa valeur maximale, tandis que l'augmentation du temps de calcul commence à monter. Cela résulte en un gain réel qui plafonne, avec comme on le voit un plateau autour de la valeur optimale de la longueur des sous-dynamiques. Les fluctuations entre simulations sur le temps de calcul et la variance de notre variable finale deviennent plus visibles : on observe ainsi une incertitude relative de 10-15% sur le gain réel.

Dans le troisième, la diminution de variance a essentiellement atteint sa valeur maximale, ou presque, mais l'augmentation du temps de calcul est importante, et est l'aspect dominant de la variation du gain réel. Cela s'interprète par le fait qu'on a essentiellement exploré tout l'espace configurationnel des sous-secteurs au point qu'on n'arrive plus à tirer de gain supplémentaire, tandis que les sous-dynamiques conservent un coût proportionnel à la longueur de celles-ci.

Ces trois régimes sont très généraux en PMC. Certes, il n'est pas toujours possible de les observer tous les trois ; ainsi, pour de trop petits systèmes ($L = 4$ ou $L = 6$), le coût relatif des sous-dynamiques est important, ce qui nous place dès le début dans le second régime. D'autre part, pour des sous-systèmes complètement décorrélés, on n'arrive jamais au troisième régime ; de même, pour des grands systèmes, où la valeur optimale de $m$ est suffisamment élevée, le troisième régime est hors de portée calculatoire.

D'autre part, la régression que l'on a réalisée à la figure 5.6, combiné avec une régression linéaire du temps de calcul de la forme $\tau_0 + m\delta\tau$, nous permet de déterminer une valeur empirique de la longueur optimale des sous-dynamiques. La valeur optimale ainsi obtenue, $m = 92$, a été portée en rouge dans le tableau, et se situe en plein dans le plateau du second régime. Nous avons ainsi pu concevoir une méthode simple pour obtenir les paramètres des deux régressions linéaires, et ainsi la longueur optimale des sous-dynamiques, en se servant d'une courte dynamique – 10 à 50 configuration suffisent – munie de longues sous-dynamiques avant l'exécution de la dynamique principale.

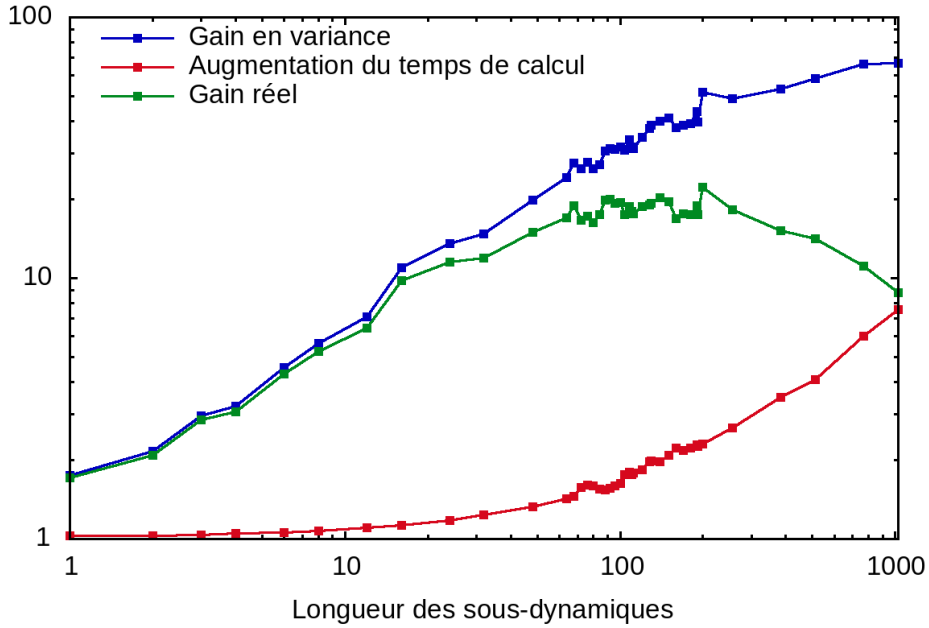


FIGURE 5.7 – Évolution des trois fonctions de gain en fonction de la longueur des sous-dynamiques.

### 5.2.4 Taille des sous-systèmes

Maintenant que nous avons pu caractériser le comportement à partition fixée des fonctions de gain en fonction de la longueur des sous-dynamiques, intéressons nous à l'impact de la taille des sous-systèmes. Pour ce faire, nous avons réalisé un spectre similaire de simulations pour un système de taille $L = 20$ avec des sous-systèmes de taille $l = 2$, $l = 4$, $l = 5$, et $l = 10$, avec des sous-dynamiques de longueur $m = 1$ à $m = 1024$. La figure 5.8 contient les valeurs prises par le gain en variance $G_v$, à gauche, et le surcoût en temps, $G_t$, à droite, pour chacune des partitions, en fonction de la longueur des sous-dynamiques.

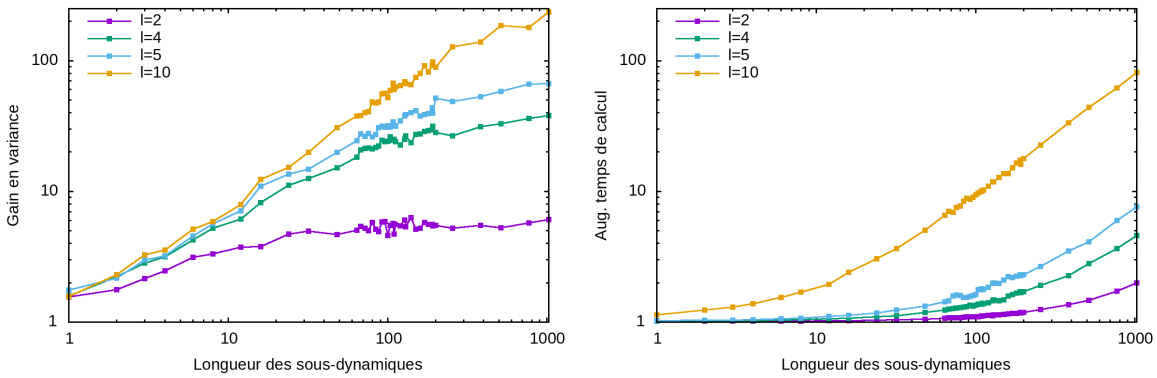


FIGURE 5.8 – Évolution du gain en variance (à gauche) et de l'augmentation du temps de calcul (à droite) dans le calcul de l'énergie en PMC par rapport au VMC, en fonction de la longueur des sous-dynamiques, pour différentes partitions en sous-systèmes.

Il est clair que, à mesure que $m$ croît, $G_v$ augmente jusqu'à un plafond plus élevé lorsque $l$ est plus grand, mais que $G_t$ augmente également significativement plus vite. On doit donc équilibrer, encore une fois, l'augmentation du coût de calcul avec la réduction de la variance. Pour confirmer cela, nous avons sélectionné pour chaque partition la simulation qui nous avait donné le meilleur gain réel. Les valeurs que prennent alors les fonctions de gain et la longueur des sous-dynamiques ont été placées dans le tableau 5.2 et dans la figure 5.9, pour lequel nous avons pris en abcisse le facteur d'échelle

$\ln(l)/\ln(L)$ afin de pouvoir visualiser les symétries.

TABLE 5.2 – Valeurs prises par les fonctions de gain dans le calcul de l'énergie en PMC par rapport au VMC, pour la longueur des sous-dynamiques ayant donné la plus grande valeur du gain réel, en fonction de la taille des sous-systèmes. La meilleure valeur prise pour le gain en variance est également représentée.

| $l$ | 2 | 4 | 5 | 10 |
|---|---|---|---|---|
| $G_v(m_{\text{opt}})$ | 6.275 | 26.23 | 51.49 | 67.32 |
| $G_t(m_{\text{opt}})$ | 1.138 | 1.368 | 2.315 | 9.946 |
| $G_r(m_{\text{opt}})$ | 5.515 | 19.18 | 22.25 | 6.769 |
| $m_{\text{opt}}$ | 140 | 104 | 200 | 108 |
| $\max G_v$ | 6.275 | 38.07 | 66.82 | 236.5 |

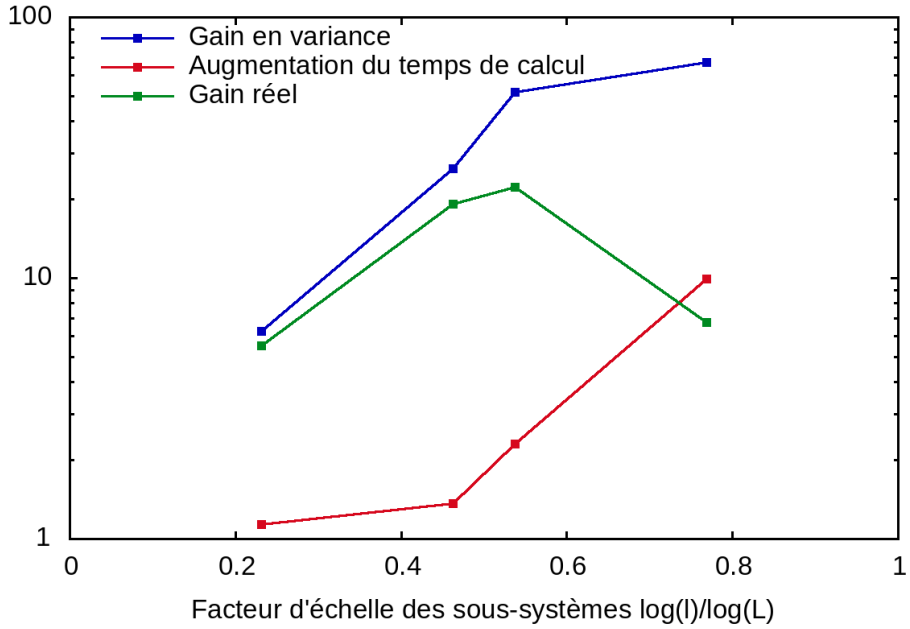


FIGURE 5.9 – Évolution des fonctions de gain en fonction du facteur d'échelle des sous-systèmes $\ln(l)/\ln(L)$, pour la longueur optimale des sous-dynamiques.

Il est évident que pour $l$ petit, on a de très faibles valeurs de $G_t$, et le gain en variance gouverne la valeur du gain réel. Cela signifie que la variance résiduelle est le facteur limitant du gain réel pour de faibles valeurs de $l$. A contrario, pour $l$ grand, on se retrouve avec de très grandes valeurs de $G_t$. Le coût, trop élevé, des configurations dans les sous-systèmes est alors le facteur limitant du gain réel. Dans ce cas, la meilleur valeur est obtenue pour $l = 5$.

Nous avons réalisé ce travail pour différentes valeurs de $L$ afin de faire apparaître soit une loi d'échelle, soit une valeur optimale de $l$ pour de grands systèmes. Les valeurs qu'on a utilisées sont $L = 12$, $L = 20$ et $L = 30$ ; et nous avons représenté les valeurs du gain réel $G_r$ optimisé pour la longueur des sous-dynamiques en fonction du facteur d'échelle $\ln(l)/\ln(L)$ dans la figure 5.10.

On observe un maximum autour d'un facteur d'échelle de 0.5, soit donc pour $l = \sqrt{L}$, avec de meilleures tolérances pour $l$ plus grand.

### 5.2.5 Scaling en fonction de la taille du système

Suite aux tendances que nous avons pu déduire des deux sous-sections précédentes, nous avons réalisé des simulations VMC et PMC pour des systèmes de taille $L = 4$ à $L = 56$, munis de sous-

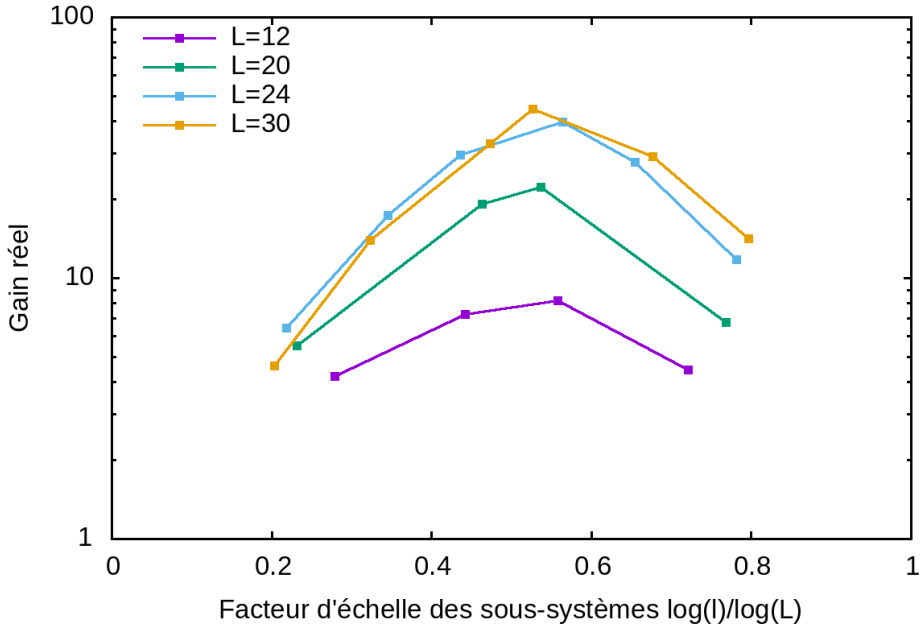


FIGURE 5.10 – Évolution du gain réel en fonction du facteur d'échelle des sous-systèmes $\ln(l)/\ln(L)$, pour la longueur optimale des sous-dynamiques, et pour différentes tailles de systèmes.

systèmes de taille $l = \left\lceil \sqrt{L} \right\rceil$, et avec une longueur de sous-dynamiques optimisée. Les valeurs obtenues pour le temps de calcul, la variance de $E_l$ et $\tilde{E}_{\text{pr}}$, et le gain réel, ont été portés dans le tableau 5.3. La figure 5.11 ci-dessous reproduit, quant à elle, le scaling observé pour le gain réel, et la figure 5.12 les courbes correspondant au tableau tout entier.

On observe une variance qui croit lentement et de manière moins que linéaire avec la taille du système, mais un gain réel qui se comporte de manière linéaire par rapport à la taille du système, approximativement comme $N/20$. Ce n'est pas surprenant. En effet, si on reprend les considérations théoriques mises en jeu à la section 4.5, on doit trouver un gain en $\mathcal{O}\left(p^2\right)$. Or, ici $p = (L/l)^2$. Comme $p$ se comporte en $\mathcal{O}\left(\sqrt{N}\right)$, il en découle alors un gain global en $\mathcal{O}\left(N\right)$. Des simulations réalisées sur des systèmes modèles de chaînes d'hydrogène métalliques nous permettent d'observer un comportement similaire pour le gain réel, comme on peut le voir aux résultats présentés par l'article fourni à l'annexe E.

## 5.3 Résultats en MS-PMC

Les résultats et la loi d'échelle qu'on a pu observer en PMC simple laissent supposer que les tailles optimales de sous-systèmes dans les partitions de MS-PMC doivent également se conformer à des lois d'échelle. Pour cela, nous avons envisagé trois stratégies différentes pour ces lois d'échelle.

La première de ces stratégies consiste à considérer que la MS-PMC revient à utiliser la PMC sur chacun des sous-systèmes, et qu'il faut donc utiliser $l \approx \sqrt{l}$, et de manière générale pour la partition $\mathcal{P}_k$ on a $l_k \approx L\hat{}(1/2^k)$. Cette stratégie consiste à utiliser une structure en racines successives par le bas, et sera donc notée en abrégé RSI (Racines Successives Inférieures). Cela s'interprête comme une stratégie qui utilise les sous-dynamiques du troisième niveau pour accélérer le calcul des espérances conditionnelles, et non pour réduire la variance résiduelle.

La deuxième stratégie consiste au contraire à considérer que la PMC est une stratégie de reconstruction, et que c'est le nombre de sous-systèmes qui doit être pris en racines successives. Ainsi, pour

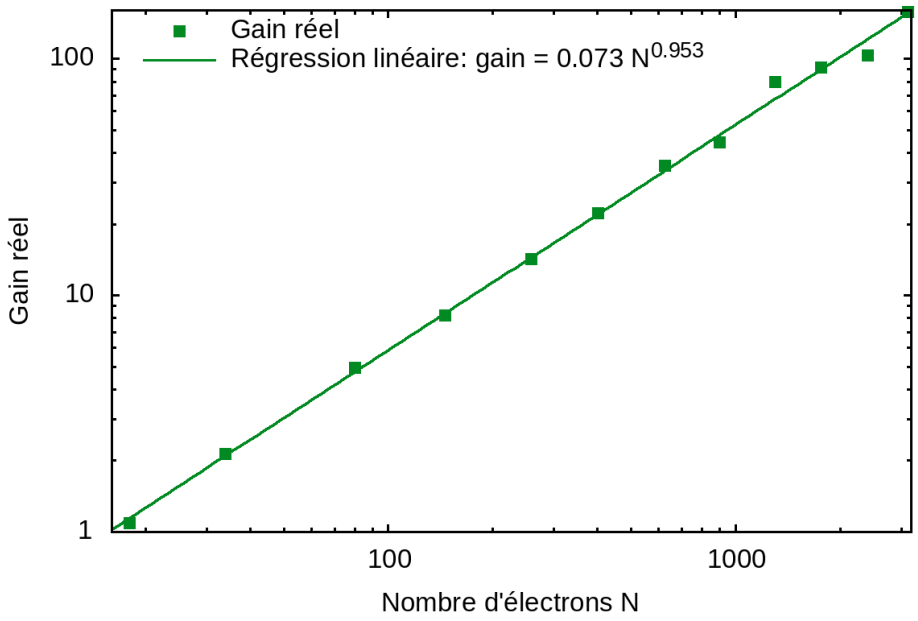


FIGURE 5.11 – Scaling du gain réel optimal pour le calcul de l'énergie en PMC par rapport au VMC, avec la fonction d'onde d'essai HF.

une partition hiérarchisée mettant en jeu $K$ partitions, la taille $l_k$ des sous-systèmes est de l'ordre de : $l_k \approx L\hat{}(1 - 2^k/2^{1+K})$. Cette stratégie en racines successives par le haut sera notée en abrégé RSS (Racines Successives Supérieures). Cela s'interprête comme une stratégie d'introduction de partitions intermédiaires pour regrouper les données et réduire la variance résiduelle, mais sans réduire le coût global des sous-dynamiques.

La dernière stratégie cherche à espacer de manière géométrique les valeurs des $l_k$ entre 1 et $L$. Ainsi, pour une simulation à $q-1$ partitions, on aura pour la partition $\mathcal{P}_k$ des sous-systèmes de taille $l_k \approx L\hat{}(1 - k/q)$. Cette stratégie d'espacement géométrique sera notée en abrégé GS (Geometrical Scaling). Cette stratégie cherche alors à jouer sur les deux tableaux, c'est à dire à avoir une variance résiduelle plus petite tout en ayant un coût effectif de réduction de la variance moins élevé.

On s'est restreint, lors de l'application de ces stratégies, à l'emploi d'au plus deux partitions. En effet, les limitations en termes de coût de calcul rendent les systèmes où une troisième partition aurait été envisageable – c'est-à-dire $L = 256$ en RSI et RSS et $L = 81$ en GS – hors de notre portée. Les résultats obtenus ont été portés dans le tableau 5.4.

Ce qu'on observe est que la stratégie RSI semble peu performante, à la fois pour de faibles et grandes tailles de $L$. Les stratégies RSS et GS semblent toutes deux plus performantes, permettant un gain d'un facteur 2 supplémentaire pour des systèmes de grande taille, et de ces deux stratégies GS semble la plus efficace.

TABLE 5.3 – Comparaison des valeurs prises par le temps de calcul et la variance de l'estimateur employés en VMC et PMC à taille des sous-systèmes et longueur des sous-dynamiques optimaux, ainsi que le gain réel optimal correspondant, avec la fonction d'onde d'essai HF.

| $L$ | $N$ | $\tau_{\text{VMC}}$ | $\sigma^2_{\text{VMC}}$ | $\tau_{\text{PMC}}$ | $\sigma^2_{\text{PMC}}$ | $G_r$ |
|---|---|---|---|---|---|---|
| 4 | 18 | $6.252{\cdot}10^{-2}$ | 1.05 | $1.845{\cdot}10^{-1}$ | 0.329 | 1.08 |
| 6 | 34 | $2.277{\cdot}10^{-1}$ | 2.31 | $5.835{\cdot}10^{-1}$ | 0.424 | 2.13 |
| 9 | 80 | $2.131{\cdot}10^{0}$ | 6.00 | $4.895{\cdot}10^{0}$ | 0.527 | 4.96 |
| 12 | 146 | $1.174{\cdot}10^{1}$ | 10.2 | $2.247{\cdot}10^{1}$ | 0.648 | 8.21 |
| 16 | 258 | $7.128{\cdot}10^{1}$ | 17.2 | $1.218{\cdot}10^{2}$ | 0.707 | 14.3 |
| 20 | 402 | $2.658{\cdot}10^{2}$ | 27.5 | $6.153{\cdot}10^{2}$ | 0.533 | 22.2 |
| 25 | 624 | $1.306{\cdot}10^{3}$ | 46.0 | $2.066{\cdot}10^{3}$ | 0.826 | 35.2 |
| 30 | 898 | $5.191{\cdot}10^{3}$ | 54.2 | $8.145{\cdot}10^{3}$ | 0.778 | 44.5 |
| 36 | 1298 | $2.133{\cdot}10^{4}$ | 85.1 | $2.010{\cdot}10^{4}$ | 1.131 | 79.9 |
| 42 | 1762 | $5.986{\cdot}10^{4}$ | 112 | $7.055{\cdot}10^{4}$ | 1.036 | 91.5 |
| 49 | 2400 | $1.536{\cdot}10^{5}$ | 160 | $1.947{\cdot}10^{5}$ | 1.227 | 103 |
| 56 | 3138 | $3.740{\cdot}10^{5}$ | 229 | $4.108{\cdot}10^{5}$ | 1.325 | 158 |

TABLE 5.4 – Résultats obtenus en MS-PMC.

| $L$ | $l_2$ | $l_3$ | Stratégies | $\max G_r$ | $\max_{\text{PMC}} G_r$ |
|---|---|---|---|---|---|
| 8 | 4 | 2 | GS | 3.39 | |
| 12 | 4 | 2 | RSI | 8.05 | 7.85 |
| 12 | 6 | 2 | GS | 8.32 | |
| 12 | 6 | 3 | RSS | 8.42 | |
| 16 | 4 | 2 | RSI | 15.0 | 14.3 |
| 16 | 8 | 2 | | 12.9 | |
| 16 | 8 | 4 | RSS | | |
| 27 | 9 | 3 | GS | 45.8 | |
| 36 | 6 | 3 | RSI | 88.2 | 75.4 |
| 36 | 12 | 4 | GS | 155 | |
| 36 | 12 | 6 | RSS | 140 | |

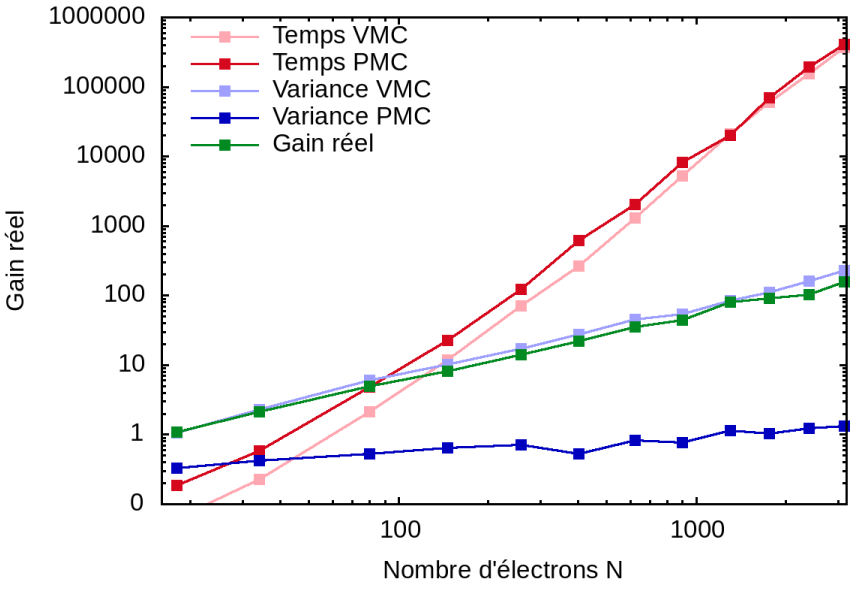


FIGURE 5.12 – Scaling des performances comparatives en VMC et PMC pour la fonction d'onde d'essai HF.

## Récapitulatif

Après avoir vérifié que la méthode PMC et son implémentation multi-échelle étaient non biaisées, ou que si biais il y avait, celui-ci décroissait plus vite que $\mathcal{O}\left(1/\sqrt{N}\right)$, nous nous sommes penché sur le comportement de la méthode PMC en fonction de ses deux paramètres, taille des sous-systèmes et longueur des sous-dynamiques. Dans chacun des deux, nous avons mis en évidence un phénomène d'équilibrage entre réduction de la variance et augmentation du temps de calcul.

Qui plus est, nous avons pu effectuer des régressions linéaires sur la variance et le temps de calcul en fonction de la longueur des sous-dynamiques $m$, ce qui nous permet de calculer une valeur optimale de $m$ suffisamment bonne, et mettre en évidence une loi d'échelle pour la taille optimale des sous-systèmes dans le modèle de Hubbard métallique. Cela s'est traduit par un gain réel en $\mathcal{O}(N)$, soit bel et bien une diminution du scaling des fluctuations statistiques à un faible surcoût en temps.

Sur la méthode MS-PMC, on observe un faible gain supplémentaire croissant avec la taille du système. Cependant, les limitations sur les tailles des systèmes exploîtables avec la capacité de calcul disponible n'ont pas permis de dégager de grandes tendances.

Cependant, jusqu'ici tous nos résultats obtenus ont été pour un simple calcul de l'énergie variationnelle. Si l'on veut obtenir davantage de résultats exploitables, il va falloir pouvoir optimiser la fonction d'onde, ce qui requiert le calcul de dérivées de l'énergie variationnelle. C'est donc à cela qu'on va s'attacher dans la prochaine partie.

Troisième partie

# Extension de la méthode aux cumulants d'ordre supérieur

# Chapitre 6

# Construction théorique d'estimateurs de $\mathrm{Cov}\,(X, Y)$

Nous avons vu à la partie précédente que la méthode de Monte Carlo Partitionnelle nous permettait de calculer de manière efficace la valeur de l'énergie variationnelle. Cependant, la plupart des quantités d'intérêt que l'on peut chercher à calculer sont plus difficiles à calculer. En effet, la plupart de ces quantités d'intérêt se mettent sous la forme de dérivées logarithmiques ; dans un système de physique statistique, la quantité à dériver est la fonction de partition ; et pour un système de chimie quantique, il s'agit de la densité $\langle\psi|\exp(-t\hat{H})|\psi\rangle$. En particulier, l'énergie variationnelle se met sous la forme suivante :

$$E_v[\psi] = \frac{\left\langle \psi \middle| \hat{H} \middle| \psi \right\rangle}{\langle\psi|\psi\rangle} = -\frac{\partial}{\partial t} \ln \left[ \left\langle \psi \middle| e^{-t\hat{H}} \middle| \psi \right\rangle \right] \Big|_{t=0} . \tag{6.1}$$

Pour de nombreuses quantités d'interêt – fonctions de réponse, dérivées de l'énergie variationnelle – on doit cependant travailler avec des dérivées logarithmiques secondes et troisièmes de cette densité, dont l'expression met en jeu des cumulants, des formes multilinéaires, symétriques, homogènes et additives pour deux multiplets indépendents de variables aléatoires, qui généralisent la notion de covariance.

Dans cette partie, nous allons donc nous intéresser à étendre la méthode de Monte Carlo Partitionnelle à ces variables d'intérêt. Nous nous intéresserons dans un premier chapitre au développement d'une expression d'estimateurs améliorés de la covariance ; pour ensuite chercher à construire une expression générale nous permettant de construire ces estimateurs pour tout cumulant. Enfin, notre dernier chapitre s'intéressera à l'implémentation de ces estimateurs, ainsi qu'à la manière dont on peut reconstruire nos dérivées à partir des cumulants qu'on aura calculés. Nous ferons fortement référence tout au long de cette partie à l'annexe B, qui contient de nombreuses propriétés sur les cumulants, ainsi que les expressions classiques d'estimateurs de ceux-ci. Pour les notations introduites dans cette partie, voir l'annexe A.

Au cours de ce chapitre, nous commencerons par rappeler les propriétés des estimateurs classiques de la covariance de deux variables aléatoires $X$ et $Y$. Nous construirons ensuite de deux manières différentes un même estimateur amélioré de la covariance de ces variables, que vous pourrez retrouver à l'équation (6.24). Notre quatrième section s'intéressera aux propriétés de cet estimateur et interprètera son expression ; et nous finirons en développant brièvement la variance de cet estimateur pour chercher à caractériser la réduction en variance à laquelle on peut s'attendre.

## 6.1 Propriétés de l'estimateur classique de la covariance

L'expression classique des estimateurs de la covariance de deux variables aléatoires $X$ et $Y$ est donné par l'expression suivante :

$$\bar{C}(X,Y) = \frac{1}{M}\sum_{K=1}^{M} X(\mathcal{R}^K)Y(\mathcal{R}^K) - \left(\frac{1}{M}\sum_{K=1}^{M} X(\mathcal{R}^K)\right)\left(\frac{1}{M}\sum_{K=1}^{M} Y(\mathcal{R}^K)\right) = \overline{XY} - \overline{X}\,\overline{Y} . \tag{6.2}$$

Il s'agit d'estimateurs biaisés, qui ont, pour des issues successives indépendentes, pour espérance $(1 - 1/M)\text{Cov}(X, Y)$, et dont la variance est donnée par l'expression suivante (démontrée dans l'annexe B.2.2) :

$$\text{Var}\left(\bar{C}(X,Y)\right) = \frac{M-1}{M^2}(\text{Var}(X)\,\text{Var}(Y) + \text{Cov}(X,Y)^2) + \frac{(M-1)^2}{M^3}\gamma_4(X,X,Y,Y)\,. \tag{6.3}$$

Dans cette équation, $\gamma_4$ est un cumulant quaternaire. La variance, la covariance et $\gamma_4$ étant des cumulants, ceux-ci sont extensifs, et leur valeur se comporte donc de manière proportionnelle au nombre d'électrons $N$. La variance de cet estimateur se comporte ainsi en $\mathcal{O}\left(N^2\right)$. On peut montrer, de manière générale, que pour le cumulant d'ordre $q$, la variance des estimateurs classiques de celui-ci se comporte en $\mathcal{O}\left(N^q\right)$. On voit donc que le bruit augmente significativement plus vite que le signal pour les dérivées d'ordre supérieur...

## 6.2 Construction par développement d'une covariance

À la section 3.5, on a cherché à développer la variance de la variable aléatoire améliorée pratique $\tilde{X}_{\text{pr}}$, avec l'équation (3.30) :

$$\text{Var}\left(\tilde{X}_{\text{pr}}\right) = (1-p)^2\text{Var}(X) + 2(1-p)\sum_{i=1}^{p}\text{Cov}\left(X,\bar{X}_i\right) + \sum_{i=1}^{p}\sum_{j=1}^{p}\text{Cov}\left(\bar{X}_i,\bar{X}_j\right)\ ; \tag{6.4}$$

où $\bar{X}_i$ est la valeur moyenne de $X$ prise sur une sous-dynamique dans le sous-système $\mathcal{S}_i$. Cependant, comme le dernier terme met en jeu des covariances entre des moyennes sur des sous-dynamiques différentes, il est évident qu'il n'est pas facile de se servir de Var $(\tilde{X}_{\text{pr}})$ pour construire un estimateur de la variance de $X$.

Cependant, si on développe de la même manière Cov $(\tilde{X}_{\text{pr}}, Y)$, avec $Y$ une variable aléatoire quelconque, cette même décomposition nous donne :

$$\text{Cov}\left(\tilde{X}_{\text{pr}},Y\right) = (1-p)\text{Cov}(X,Y) + \sum_{i=1}^{p}\text{Cov}\left(\bar{X}_i,Y\right)\ . \tag{6.5}$$

On peut développer cette expression en nous servant de la partition des covariances :

$$\begin{aligned}
\text{Cov}\left(\tilde{X}_{\text{pr}},Y\right) &= (1-p)\text{Cov}(X,Y) + \sum_{i=1}^{p}\left[\text{Cov}\left(\mathcal{E}_i\bar{X}_i,\mathcal{E}_iY\right) + \boldsymbol{E}\left(\text{Cov}\left(\bar{X}_i,Y|\bar{\imath}\right)\right)\right] \\
&= (1-p)\text{Cov}(X,Y) + \sum_{i=1}^{p}\left[\text{Cov}\left(\mathcal{E}_iX,\mathcal{E}_iY\right) + \boldsymbol{E}\left(\text{Cov}\left(\bar{X}_i,Y|\bar{\imath}\right)\right)\right] \\
&= (1-p)\text{Cov}(X,Y) + \sum_{i=1}^{p}\left[\text{Cov}(X,Y) - \boldsymbol{E}\left(\text{Cov}\left(X,Y|\bar{\imath}\right)\right) + \boldsymbol{E}\left(\text{Cov}\left(\bar{X}_i,Y|\bar{\imath}\right)\right)\right] \\
&= \text{Cov}(X,Y) - \boldsymbol{E}\left(\sum_{i=1}^{p}\text{Cov}\left(X,Y|\bar{\imath}\right)\right) + \boldsymbol{E}\left(\sum_{i=1}^{p}\text{Cov}\left(\bar{X}_i,Y|\bar{\imath}\right)\right)\ .
\end{aligned} \tag{6.6}$$

Pour ce dernier terme, on devra se resservir de la projection de l'équation (3.34) :

$$\bar{X}_i \underset{m\to\infty}{=} \mathcal{E}_iX + \frac{\tau_i - 1}{2m}\Delta_iX\ ;$$

où $\tau_i$ est le facteur d'autocorrélation et $m$ la longueur des sous-dynamiques. Si on l'introduit dans l'expression de Cov $(\tilde{X}_{\text{pr}}, Y)$, cela nous donne :

$$\text{Cov}\left(\tilde{X}_{\text{pr}},Y\right) = \text{Cov}(X,Y) - \sum_{i=1}^{p}\boldsymbol{E}\left(\text{Cov}\left(X,Y|\bar{\imath}\right)\left(1 - \frac{\tau_i-1}{2m}\right)\right)\ . \tag{6.7}$$

On peut donc envisager de partir de l'estimateur classique de la covariance de $\tilde{X}_{\text{pr}}$ et de $Y$ pour construire un estimateur de la covariance de $X$ et de $Y$. Il nous faudra cependant le compléter avec une somme d'estimateurs de $\boldsymbol{E}\left(\text{Cov}\left(X, Y|\bar{\imath}\right)\right)$. Pour cela, on choisit l'estimateur $\overline{\bar{C}(X, Y|\bar{\imath})}$, la moyenne partielle des estimateurs classiques de la covariance conditionnelle de $X$ et de $Y$.

On arrive ainsi à l'expression suivante pour l'estimateur $\tilde{C}(X, Y)$ :

$$\tilde{C}(X, Y) = \bar{C}(\tilde{X}_{\text{pr}}, Y) + \sum_{i=1}^{p} \overline{\bar{C}(X, Y|\bar{\imath})} \; ; \tag{6.8}$$

où $\bar{C}(X, Y|\bar{\imath})$ correspond à l'estimateur classique de la covariance, calculé sur une sous-dynamique. On voit que cette expression n'est pas symétrique par permutation de $X$ et $Y$. Nous construirons à la section suivante une expression symétrique par permutation.

## 6.3 Construction par séparabilité

Dans cette section, nous allons chercher à construire une expression améliorée de l'estimateur de la covariance en repartant de l'hypothèse de séparabilité. Nous allons donc chercher à construire une quantité $Q$ d'espérance nulle telle que, dans la limite de séparabilité, on ait (avec $C$ un estimateur quelconque non biaisé de la covariance) :

$$\text{Cov}\left(X, Y\right) = Q + C \; .$$

Nous allons, de manière similaire au raisonnement employé à la section 3.3, nous placer dans l'hypothèse de séparabilité et poser $X_{\bar{\imath}} = X - X_i$ et $Y_{\bar{\imath}} = Y - Y_i$ ; et développer dans un premier temps l'expression de $-\sum_i \Delta_i XY$, d'abord avec $X_i$ et $Y_i$, puis en les faisant disparaître de l'expression. Pour ce faire, partons de l'expression de $\mathcal{E}_i X$. Dans l'hypothèse de séparabilité, celle-ci est :

$$\mathcal{E}_i X = \mathcal{E}_i X_{\bar{\imath}} + \mathcal{E}_i X_i = X_{\bar{\imath}} + \boldsymbol{E}\left(X_i\right) \; . \tag{6.9}$$

On peut alors développer de la même manière $\mathcal{E}_i XY$. Cela nous donne :

$$\begin{aligned} \mathcal{E}_i XY &= \mathcal{E}_i((X_i + X_{\bar{\imath}})(Y_i + Y_{\bar{\imath}})) \\ &= \mathcal{E}_i(X_i(Y_i + Y_{\bar{\imath}})) + X_{\bar{\imath}}\mathcal{E}_i(Y_i + Y_{\bar{\imath}}) \\ &= \boldsymbol{E}\left(X_i Y_i\right) + Y_{\bar{\imath}}\boldsymbol{E}\left(X_i\right) + X_{\bar{\imath}}\boldsymbol{E}\left(Y_i\right) + X_{\bar{\imath}}Y_{\bar{\imath}} \; . \end{aligned} \tag{6.10}$$

On en tire celle de $\Delta_i XY$ :

$$\begin{aligned} \Delta_i XY &= XY - \mathcal{E}_i XY \\ &= (X_i Y_i + X_i Y_{\bar{\imath}} + Y_i X_{\bar{\imath}} + X_{\bar{\imath}} Y_{\bar{\imath}}) - (\boldsymbol{E}\left(X_i Y_i\right) + Y_{\bar{\imath}}\boldsymbol{E}\left(X_i\right) + X_{\bar{\imath}}\boldsymbol{E}\left(Y_i\right) + X_{\bar{\imath}}Y_{\bar{\imath}}) \\ &= \Delta_i(X_i Y_i) + Y_{\bar{\imath}}\Delta_i X_i + X_{\bar{\imath}}\Delta_i Y_i \\ &= \Delta_i(X_i Y_i) + Y_{\bar{\imath}}\Delta_i X + X_{\bar{\imath}}\Delta_i Y \; . \end{aligned} \tag{6.11}$$

Et on en déduit l'expression de $-\sum_i \Delta_i XY$, au signe près :

$$\begin{aligned} \sum_{i=1}^{p} \Delta_i XY &= \sum_{i=1}^{p} \Delta_i(X_i Y_i) + Y_{\bar{\imath}}\Delta_i X + X_{\bar{\imath}}\Delta_i Y \\ &= \sum_{i=1}^{p} \Delta_i(X_i Y_i) + \sum_{i=1}^{p} Y_{\bar{\imath}}\Delta_i X + \sum_{i=1}^{p} X_{\bar{\imath}}\Delta_i Y \\ &= \sum_{i=1}^{p} \Delta_i(X_i Y_i) + Y\sum_{i=1}^{p} \Delta_i X - \sum_{i=1}^{p} Y_i\Delta_i X_i + X\sum_{i=1}^{p} \Delta_i Y - \sum_{i=1}^{p} X_i\Delta_i Y_i \\ &= \sum_{i=1}^{p} \left[\Delta_i(X_i Y_i) - Y_i\Delta_i X_i - X_i\Delta_i Y_i\right] + Y\left(X - \boldsymbol{E}\left(X\right)\right) + X\left(Y - \boldsymbol{E}\left(Y\right)\right) \; . \end{aligned} \tag{6.12}$$

Pour faire disparaître $X_i$ et $Y_i$, on va employer les covariances conditionnelles $C_i = \mathrm{Cov}\,(X, Y|\bar{\imath})$, qui sont constantes dans l'hypothèse de séparabilité et valent :

$$C_i = \mathcal{E}_i XY - \mathcal{E}_i X \mathcal{E}_i Y = \mathrm{Cov}\,(X_i, Y_i)\ . \tag{6.13}$$

Si on cherche à faire apparaître $C_i$ dans la somme de l'équation (6.12), on arrive à :

$$\begin{aligned}\Delta_i(X_iY_i) - Y_i\Delta_i X_i - X_i\Delta_i Y_i &= X_iY_i - \mathcal{E}_i X_iY_i - Y_iX_i + Y_i\mathcal{E}_i X_i - Y_iX_i + X_i\mathcal{E}_iY_i \\ &= -(\mathcal{E}_iX_iY_i - \mathcal{E}_iX_i\mathcal{E}_iY_i) - (\mathcal{E}_iX_i - X_i)(\mathcal{E}_iY_i - Y_i) \\ &= -C_i - \Delta_iX_i\Delta_iY_i\ .\end{aligned} \tag{6.14}$$

De plus, remarquer que $\Delta_i X_i = \Delta_i X$ nous permet de faire disparaître toute référence aux $X_i$ et $Y_i$. On peut alors réinjecter cette expression dans l'équation (6.12), ce qui nous donne, dans la limite de séparabilité :

$$-\sum_{i=1}^{p}\Delta_i XY = \sum_{i=1}^{p}\left[C_i + \Delta_i X\Delta_i Y\right] - Y(X - \boldsymbol{E}\,(X)) - X(Y - \boldsymbol{E}\,(Y))\ . \tag{6.15}$$

On sait que, dans la limite de séparabilité, $\sum_i C_i = \mathrm{Cov}\,(X, Y)$. On va donc chercher à altérer l'équation (6.15) pour obtenir l'expression de $2Q$. Or, on a :

$$\mathcal{E}_i(\Delta_i X\Delta_i Y) = \boldsymbol{E}\,((X - \boldsymbol{E}\,(X|\bar{\imath}))(Y - \boldsymbol{E}\,(Y|\bar{\imath}))|\bar{\imath}) = \mathrm{Cov}\,(X, Y|\bar{\imath}) = C_i\ . \tag{6.16}$$

On peut donc décomposer $\Delta_i X\Delta_i Y$, et soustraire des deux côtés $\Delta_i(\Delta_i X\Delta_i Y)$. Cela nous donne :

$$-\sum_{i=1}^{p}\left[\Delta_i XY + \Delta_i(\Delta_i X\Delta_i Y)\right] = 2\sum_{i=1}^{p}C_i - Y(X - \boldsymbol{E}\,(X)) - X(Y - \boldsymbol{E}\,(Y)) = 2Q\ . \tag{6.17}$$

Décomposer $(\Delta_i XY + \Delta_i(\Delta_i X\Delta_i Y))$ nous donne :

$$\begin{aligned}\Delta_i XY + \Delta_i(\Delta_i X\Delta_i Y) &= XY - \mathcal{E}_i XY + [\Delta_i X\Delta_i Y - \mathcal{E}_i(\Delta_i X\Delta_i Y)] \\ &= XY - \mathcal{E}_i XY + [(X - \mathcal{E}_i X)(Y - \mathcal{E}_i Y) - \mathcal{E}_i(X - \mathcal{E}_i X)(Y - \mathcal{E}_i Y)] \\ &= XY - \mathcal{E}_i XY + [XY - X\mathcal{E}_i Y - Y\mathcal{E}_i X + \mathcal{E}_i X\mathcal{E}_i Y - \mathcal{E}_i XY + \mathcal{E}_i X\mathcal{E}_i Y] \\ &= 2XY - 2\mathcal{E}_i XY + 2\mathcal{E}_i X\mathcal{E}_i Y - X\mathcal{E}_i Y - Y\mathcal{E}_i X \\ &= 2XY - 2C_i - X\mathcal{E}_i Y - Y\mathcal{E}_i X \\ &= -2C_i\ .\end{aligned} \tag{6.18}$$

Réinsérons cette expression dans l'équation (6.17) :

$$2\sum_{i=1}^{p}C_i - \sum_{i=1}^{p}\left[X\Delta_i Y + Y\Delta_i X\right] = 2Q\ . \tag{6.19}$$

On arrive alors aux expressions suivantes de $Q$ et $C$ :

$$\begin{aligned}Q &= \sum_{i=1}^{p}\mathrm{Cov}\,(X, Y|\bar{\imath}) - \sum_{i=1}^{p}\frac{X\Delta_i Y + Y\Delta_i X}{2}\ ; \\ C &= \frac{X(Y - \boldsymbol{E}\,(Y)) + Y(X - \boldsymbol{E}\,(X))}{2}\ .\end{aligned} \tag{6.20}$$

On vérifie aisément que $Q$ est par construction bien d'espérance nulle, et que $C$ est bien par construction un estimateur non biaisé de $\mathrm{Cov}\,(X, Y)$. On peut donc prendre comme estimateur amélioré théorique $\tilde{C}_{\mathrm{th}} = C + Q$, c'est-à-dire :

$$\begin{aligned}\tilde{C}_{\mathrm{th}} &= \sum_{i=1}^{p}\mathrm{Cov}\,(X, Y|\bar{\imath}) + \frac{X}{2}\left[Y - \sum_{i=1}^{p}\Delta_i Y - \boldsymbol{E}\,(Y)\right] + \frac{Y}{2}\left[X - \sum_{i=1}^{p}\Delta_i X - \boldsymbol{E}\,(X)\right] \\ &= \sum_{i=1}^{p}\mathrm{Cov}\,(X, Y|\bar{\imath}) + \frac{X}{2}\left[\tilde{Y}_{\mathrm{th}} - \boldsymbol{E}\,(Y)\right] + \frac{Y}{2}\left[\tilde{X}_{\mathrm{th}} - \boldsymbol{E}\,(X)\right]\ .\end{aligned} \tag{6.21}$$

On reconnaît dans cette équation l'expression de la covariance de $\tilde{X}_{\text{th}}$ et $Y$, et vice versa.

Si on veut passer d'un estimateur théorique à un estimateur pratique, on doit alors remplacer les variables améliorées théoriques pratiques par des variables améliorées pratiques et utiliser l'estimateur classique de la covariance $\bar{C}$ pour les covariances internes, ce qui nous donne :

$$\tilde{C} = \sum_{i=1}^{p} \bar{C}(X, Y|\bar{\imath}) + \frac{X}{2}\left[\tilde{Y}_{\text{pr}} - \boldsymbol{E}\left(Y\right)\right] + \frac{Y}{2}\left[\tilde{X}_{\text{pr}} - \boldsymbol{E}\left(X\right)\right] \ . \tag{6.22}$$

On peut remarquer que ce $\tilde{C}$ ne se sert que d'une configuration de la dynamique principale. Pour un estimateur les utilisant toutes, on arrive à l'estimateur amélioré pratique $\tilde{C}_{\text{pr}}$ :

$$\tilde{C}_{\text{pr}} = \sum_{i=1}^{p} \overline{\bar{C}(X, Y|\bar{\imath})} + \frac{1}{2}\left[\bar{C}(\tilde{X}, Y) + \bar{C}(X, \tilde{Y})\right] \ . \tag{6.23}$$

## 6.4 Partition des covariances

On a vu à la section 6.3 comment on est arrivé à l'estimateur pratique $\tilde{C}_{\text{pr}}$ suivant :

$$\tilde{C}_{\text{pr}} = \sum_{i=1}^{p} \overline{\bar{C}(X, Y|\bar{\imath})} + \frac{\bar{C}(\tilde{X}_{\text{pr}}, Y) + \bar{C}(X, \tilde{Y}_{\text{pr}})}{2} \ ; \tag{6.24}$$

où $\tilde{X}_{\text{pr}}$ est la variable améliorée pratique définie à l'équation 3.18, et $\bar{C}(X, Y|\bar{\imath})$ est l'estimateur classique de la covariance, calculé sur une sous-dynamique. On reconnaît dans cet estimateur une version symétrique par permutation de $X$ et $Y$ de l'estimateur obtenu à la section 6.2, à l'équation (6.8). Il s'agit d'un estimateur non biaisé à échantillonnages et sous-échantillonnages infinis, et zéro-variant dans la limite de séparabilité avec des sous-échantillonnages infinis.

On peut interpréter l'estimateur $\tilde{C}_{\text{pr}}$ de façon assez aisée, en le comparant à la partition de la covariance. On peut alors voir le premier terme comme la somme des covariances sur les sous- systèmes, et le second terme comme la covariance externe, entre les sous-systèmes, et qu'on notera $C_e$. Cela revient à écrire que :

$$\text{Cov}\left(X, Y\right) = C_e + \sum_{i=1}^{p} \boldsymbol{E}\left(C_i\right) \ . \tag{6.25}$$

On pourra évaluer $C_e$ en pratique en utilisant n'importe quel estimateur de la forme $a\bar{C}(X, \tilde{Y}_{\text{pr}}) + b\bar{C}(Y, \tilde{X}_{\text{pr}})$ avec $a + b = 1$.

Cherchons maintenant à calculer le biais de $\tilde{C}_{\text{pr}}$. En effet, si $\tilde{C}_{\text{th}}$ est non-biaisé par construction, ce n'est pas nécessairement le cas pour $\tilde{C}_{\text{pr}}$. On a, en supposant un facteur d'autocorrélation de $\tau_i$ dans les sous-dynamiques, et des points de la dynamique principale indépendents deux à deux :

$$\begin{aligned}
\boldsymbol{E}\left(\tilde{C}_{\text{pr}}\right) &= \boldsymbol{E}\left(\sum_{i=1}^{p} \overline{\bar{C}(X, Y|\bar{\imath})}\right) + \boldsymbol{E}\left(\frac{\bar{C}(\tilde{X}_{\text{pr}}, Y) + \bar{C}(X, \tilde{Y}_{\text{pr}})}{2}\right) \\
&= \sum_{i=1}^{p} \boldsymbol{E}\left(\bar{C}(X, Y|\bar{\imath})\right) + \frac{M-1}{2M}\left[\text{Cov}\left(\tilde{X}_{\text{pr}}, Y\right) + \text{Cov}\left(X, \tilde{Y}_{\text{pr}}\right)\right] \\
&= \frac{m - \tau_i}{m} \sum_i \boldsymbol{E}\left(C_i\right) + \frac{M-1}{M}\left[C_e + \sum_{i=1}^{p} \boldsymbol{E}\left(C_i \frac{\tau_i - 1}{2m}\right)\right] \ .
\end{aligned} \tag{6.26}$$

On trouve deux termes dominants au biais, à $M$ et $m$ finis : un pour la covariance externe, en $C_e/M$, et un pour les covariances internes, en $\sum_i \boldsymbol{E}\left(C_i(1 + \tau_i)\right)/2m$. On voit que la covariance externe compense en partie le biais de sous-échantillonnage fini pour les sous-systèmes.

## 6.5 Développement de la variance de l'estimateur de la covariance

À la section 3.5, nous avons réussi à démontrer que les variables améliorées théorique, $\tilde{X}_{\text{th}}$, et pratique, $\tilde{X}_{\text{pr}}$, avaient une variance systématiquement moindre que la variable aléatoire d'origine $X$. Dans cette partie, nous effectuerons un bref développement de la variance de $\tilde{C}_{\text{pr}}$ afin d'essayer de caractériser dans le cas général la réduction en variance de l'estimateur. On supposera $X$ et $Y$ centrées, puisque notre estimateur est invariant par translation.

Le développement de la variance nous donne alors :

$$\text{Var}(\tilde{C}) = \text{Var}\left(\frac{\bar{C}(X,\tilde{Y}) + \bar{C}(Y,\tilde{X})}{2}\right) + \text{Cov}\left(\bar{C}(X,\tilde{Y}) + \bar{C}(Y,\tilde{X}), \overline{\sum_{i=1}^{p} \bar{C}_i}\right) + \text{Var}\left(\overline{\sum_{i=1}^{p} \bar{C}_i}\right) \ . \quad (6.27)$$

Le premier terme de l'équation (6.27) nous est donné par l'expression de la variance (et covariance) d'estimateurs standard de la covariance (voir Annexe B.2.2) :

$$M\text{Var}\left(\bar{C}(X,\tilde{Y}_{\text{pr}})\right) = \text{Var}\,(X)\,\text{Var}(\tilde{Y}) + \text{Cov}(\tilde{Y},X)^2 + \gamma_4(X,X,\tilde{Y},\tilde{Y}) + \mathcal{O}\left(\frac{1}{M}\right) \ . \quad (6.28)$$

Le gain en variance sur $Y$ se reporte sur le premier terme ; le second est égal au carré de la covariance externe $C_e$, qui est en général bien plus faible que la covariance totale ; et le troisième terme est un cumulant quaternaire, au scaling linéaire.

Le second terme de l'équation (6.27) nous donne de manière similaire une covariance :

$$M\text{Cov}\left(\bar{C}(X,\tilde{Y}), \overline{\sum_{i=1}^{p} \bar{C}_i}\right) \approx \text{Cov}\left(X\tilde{Y}, \sum_{i=1}^{p} \bar{C}_i\right) = \text{Cov}\left(XY, \sum_{i=1}^{p} \bar{C}_i\right) - \text{Var}\left(\sum_{i=1}^{p} \bar{C}_i\right) \ . \quad (6.29)$$

Cette covariance s'écrit sous la forme d'un cumulant ternaire (voir annexe B.5.1), au scaling linéaire ; et on peut s'attendre à ce qu'il soit négatif, car il s'agit de la covariance de $C_e$ et des $C_i$.

Quant au dernier terme, il nous est donné par la variance d'une moyenne :

$$M\text{Var}\left(\overline{\sum_{i=1}^{p} \bar{C}_i}\right) = \text{Var}\left(\sum_{i=1}^{p} \bar{C}_i\right) \ . \quad (6.30)$$

On voit alors apparaître un facteur en $1/m$ dans les variances et covariances de chacune des covariances internes. On peut donc s'attendre à ce que ce terme se comporte avec un gain de l'ordre de $m$.

Pour peu que le ratio de la covariance externe à la covariance globale soit du même ordre de grandeur que celui de la variance de $\tilde{X}_{\text{pr}}$ à celle de $X$, on peut alors s'attendre à observer un gain en variance similaire sur les covariances à celui observé pour les variables $X$ et $Y$.

## Récapitulatif

Dans ce chapitre, nous avons montré comment on pouvait se servir de la variable améliorée PMC, qu'on a définie à la partie précédente, pour construire un estimateur amélioré de la covariance de deux variables aléatoires $X$ et $Y$ extensives. Qui plus est, nous sommes arrivés à l'expression de cet estimateur par le moyen de deux approches complètement différente. Par ailleurs, nous avons pu noter que le biais de cet estimateur, par construction, avait tendance à être significativement moindre que l'estimateur classique, et qu'une analyse rapide de la variance de cet estimateur suggère une réduction de la variance au moins comparable à celle observée pour $X$ ou $Y$.

Cependant, les covariances ne suffisent pas à calculer des dérivées secondes, que l'on peut chercher à obtenir. On va donc s'intéresser par la suite au cumulant ternaire.

# Chapitre 7

# Estimateurs zéro-variants de cumulants d'ordre supérieur

Dans ce chapitre, nous nous intéresserons à la construction d'estimateurs pour des quantités extensives de complexité supérieure. La condition d'extensivité nous mène à nous servir de cumulants (pour plus d'informations sur ceux-ci, voir l'annexe B), et nous nous intéresserons à construire une expression générale d'estimateurs de cumulants zéro-variants dans la limite de séparabilité, et non-biaisés pour un échantillonage et sous-échantillonage infinis. Afin de simplifier les expressions, nous utiliserons exclusivement des cumulants symétriques.

Dans un premier temps, nous construirons l'expression pour les cumulants d'ordre 3 et 4. Une fois ceux-ci construits, on démontrera par récurrence l'expression générale que l'on peut deviner à partir de ces deux estimateurs. Enfin, nous finirons par traduire en terme de fonctions génératrices cette expression générale.

## 7.1 Démonstration rapide à l'ordre 3

Partons, en premier lieu, de la formule du cumulant d'ordre 3, $\gamma_3$. Celle-ci est donnée par :

$$\gamma_3(X) = \gamma_3(X,X,X) = \boldsymbol{E}\left(X^3\right) - 3\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right) + 2\boldsymbol{E}\left(X\right)^3 \ . \tag{7.1}$$

L'estimateur classique du cumulant d'ordre 3, ou "bare estimator", s'écrit de la manière suivante :

$$\bar{\gamma_3}(X) = \bar{\gamma_3}(X,X,X) = \overline{X^3} - 3\overline{X}\,\overline{X^2} + 2\overline{X}^3 \ . \tag{7.2}$$

En pratique, on utilisera plutôt l'expression suivante, qui correspond plus à un estimateur naïf et est absolument non biaisée, pour notre construction :

$$\bar{\gamma_3}(X) = \overline{X^3} - 3\boldsymbol{E}\left(X\right)\overline{X^2} + 2\boldsymbol{E}\left(X\right)^2\overline{X} \ . \tag{7.3}$$

Par analogie avec l'expression de l'estimateur amélioré de la covariance, on va alors chercher à faire apparaître la moyenne des cumulants internes $\overline{\gamma_3(X|\bar{\imath})} = \overline{\mathcal{E}_i X^3} - 3\overline{\mathcal{E}_i X}\,\overline{\mathcal{E}_i X^2} + 2\overline{(\mathcal{E}_i X)^3}$, en rajoutant des deux côtés de l'équation des variables de contrôle de la forme $\sum_i \overline{\mathcal{E}_i A \Delta_i B}$ dans l'équation (7.3). En effet, on peut aisément vérifier que celles-ci sont bien d'espérance nulle ; on a montré que les projecteurs $\mathcal{E}_i$ et $\Delta_i$ étaient orthogonaux à l'équation (3.9), et on a $\boldsymbol{E} = \mathcal{E}_{\bar{\imath}}\mathcal{E}_i$. Commençons par ajouter des deux côtés $-\sum_i \overline{\Delta_i X^3}$ :

$$\begin{aligned}\bar{\gamma_3}(X) - \sum_{i=1}^{p}\overline{\Delta_i X^3} &= \overline{X^3} - 3\boldsymbol{E}\left(X\right)\overline{X^2} + 2\boldsymbol{E}\left(X\right)^2\bar{X} + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X^3} - X^3\right] \\ &= (1-p)\overline{X^3} - 3\boldsymbol{E}\left(X\right)\overline{X^2} + 2\boldsymbol{E}\left(X\right)^2\bar{X} + \sum_{i=1}^{p}\overline{\mathcal{E}_i X^3} \ .\end{aligned} \tag{7.4}$$

On poursuit en ajoutant des deux côtés $\sum_i \overline{\mathcal{E}_i X \Delta_i X^2}$ :

$$\bar{\gamma_3}(X) + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X\, \Delta_i X^2} - \overline{\Delta_i X^3}\right] = (1-p)\overline{X^3} + \sum_{i=1}^{p} \overline{X^2 \mathcal{E}_i X} - 3\boldsymbol{E}(X)\,\overline{X^2} + 2\boldsymbol{E}(X)^2\,\bar{X} + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X^3} - \overline{\mathcal{E}_i X^2 \mathcal{E}_i X}\right] . \tag{7.5}$$

On reconnaît dans les deux premiers termes l'expression de $\overline{X^2 \tilde{X}_{\text{th}}}$ :

$$\bar{\gamma_3}(X) + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X\, \Delta_i X^2} - \overline{\Delta_i X^3}\right] = \overline{X^2 \tilde{X}_{\text{th}}} - 3\boldsymbol{E}(X)\,\overline{X^2} + 2\boldsymbol{E}(X)^2\,\bar{X} + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X^3} - \overline{\mathcal{E}_i X^2 \mathcal{E}_i X}\right] . \tag{7.6}$$

Par analogie avec la covariance externe et les covariances internes, on va alors faire apparaître les termes de cumulants internes $\sum_i \overline{\gamma_3(X|\bar{\imath})}$ et un cumulant externe $\bar{\gamma_3}(X, X, \tilde{X}_{\text{th}})$ dans le membre de droite de l'équation (7.6), et ensuite compléter le membre de droite pour maintenir l'égalité. Or, si on construit un estimateur de $\bar{\gamma_3}(X, X, \tilde{X}_{\text{th}})$ de la même manière que l'équation (7.3), on arrive à :

$$\bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) = \overline{X^2 \tilde{X}_{\text{th}}} - 2\boldsymbol{E}(X)\,\overline{X \tilde{X}_{\text{th}}} - \boldsymbol{E}(X)\,\overline{X^2} + 2\boldsymbol{E}(X)^2\,\bar{X} . \tag{7.7}$$

Cela nous donne donc :

$$\bar{\gamma_3}(X) + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X\, \Delta_i X^2} - \overline{\Delta_i X^3}\right] = \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) - 2\boldsymbol{E}(X)\,\overline{X^2} + 2\boldsymbol{E}(X)\,\overline{X \tilde{X}_{\text{th}}} + \sum_{i=1}^{p} \left[\overline{\gamma_3(X|\bar{\imath})} - \overline{2(\mathcal{E}_i X)^3} + \overline{2\mathcal{E}_i X^2 \mathcal{E}_i X}\right] . \tag{7.8}$$

En réorganisant et en faisant apparaître la variance interne, on arrive à :

$$\bar{\gamma_3}(X) + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X\, \Delta_i X^2} - \overline{\Delta_i X^3}\right] = \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) + 2\boldsymbol{E}(X)\,\overline{(X \tilde{X}_{\text{th}} - X^2)} + \sum_{i=1}^{p} \overline{\gamma_3(X|\bar{\imath})} + 2\sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X \mathrm{Var}(X|\bar{\imath})}\right] . \tag{7.9}$$

On reconnait que $\tilde{X}_{\text{th}} - X = \sum_i \mathcal{E}_i X - X$, et en se servant de la démonstration au chapitre 6.2, on a $\boldsymbol{E}(X\tilde{X}_{\text{th}}) = \boldsymbol{E}(X^2) - \sum_i \boldsymbol{E}(\mathrm{Var}(X|\bar{\imath}))$. On peut alors rajouter des deux côtés la variable de contrôle $\sum_i \overline{2\boldsymbol{E}(X)\,\Delta_i X^2}$. Cela nous donne :

$$\begin{aligned}
\tilde{\bar{\gamma_3}}(X) &= \bar{\gamma_3}(X) + \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X\, \Delta_i X^2} - \overline{\Delta_i X^3} + 2\boldsymbol{E}(X)\,\overline{\Delta_i X^2}\right] \\
&= \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) + 2\boldsymbol{E}(X) \sum_{i=1}^{p} \left[(\overline{X \mathcal{E}_i X} - \overline{X^2}) + (\overline{X^2} - \overline{\mathcal{E}_i X^2})\right] + \sum_{i=1}^{p} \overline{\gamma_3(X|\bar{\imath})} + 2\sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X \mathrm{Var}(X|\bar{\imath})}\right] \\
&= \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) - 2\boldsymbol{E}(X) \sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X^2} - \overline{X \mathcal{E}_i X}\right] + \sum_{i=1}^{p} \gamma_3(X|\bar{\imath}) + 2\sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X \mathrm{Var}(X|\bar{\imath})}\right] \\
&= \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) + 2\sum_{i=1}^{p} \left[\overline{\mathcal{E}_i X \mathrm{Var}(X|\bar{\imath})} - \overline{\boldsymbol{E}(X) \mathrm{Var}(X|\bar{\imath})}\right] + \sum_{i=1}^{p} \overline{\gamma_3(X|\bar{\imath})}
\end{aligned} \tag{7.10}$$

$$\tilde{\bar{\gamma_3}}(X) = \bar{\gamma_3}(X, X, \tilde{X}_{\text{th}}) + 2\mathrm{Cov}\left(X, \sum_{i=1}^{p} \mathrm{Var}(X|\bar{\imath})\right) + \sum_{i=1}^{p} \overline{\gamma_3(X|\bar{\imath})} . \tag{7.11}$$

Si on prend les termes de l'équation (7.11) dans la limite de séparabilité, on peut remarquer que $\tilde{X}_{\text{th}}$, $\text{Var}\,(X|\bar{\imath})$ et $\gamma_3(X|\bar{\imath})$ sont tous des constantes ; les deux premiers termes sont alors nuls et le dernier constant, ce qui nous donne un estimateur théorique non biaisé, et zéro-variant dans la limite de séparabilité. On pourra donc en déduire un estimateur pratique zéro-variant dans la limite de séparabilité à sous-échantillonnages infinis, et non biaisé à échantillonnage et sous-échantillonnages infinis.

On peut résumer la construction de cette section sous la forme suivante, plus générale :

$$\gamma_3(X) = \gamma_3(X, X, \tilde{X}_{\text{th}}) + \sum_{i=1}^{p} 2\gamma_2\left(X, \gamma_2(X|\bar{\imath})\right) + \sum_{i=1}^{p} \gamma_1(\gamma_3(X|\bar{\imath}))\ . \tag{7.12}$$

## 7.2 Démonstration rapide à l'ordre 4

Repartons ici encore de l'expression exacte du cumulant quaternaire. On peut montrer aisément, en se servant des développements en cumulants des moments, qu'il s'agit de :

$$\gamma_4(X) = \gamma_4(X, X, X, X) = \boldsymbol{E}\left(X^4\right) - 4\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^3\right) - 3\boldsymbol{E}\left(X^2\right)^2 + 12\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right)^2 - 6\boldsymbol{E}\left(X\right)^4\ . \tag{7.13}$$

Cela nous permet de construire assez facilement un estimateur naif, non biaisé, du cumulant quaternaire :

$$\bar{\gamma_4}(X) = \bar{\gamma_4}(X, X, X, X) = \overline{X^4} - 4\boldsymbol{E}\left(X\right)\overline{X^3} - 3\boldsymbol{E}\left(X^2\right)\overline{X^2} + 12\boldsymbol{E}\left(X\right)^2\overline{X^2} - 6\boldsymbol{E}\left(X\right)^3\bar{X}\ . \tag{7.14}$$

En intégrant à notre expression des variables de contrôle en $\sum_i \overline{\mathcal{E}_i A \Delta_i B}$, on va chercher à trouver une expression simple comme la précédente. On va commencer par $-\sum_i \overline{\Delta_i X^4}$ :

$$\bar{\gamma_4}(X) - \sum_{i=1}^{p}\overline{\Delta_i X^4} = (1-p)\overline{X^4} - 4\boldsymbol{E}\left(X\right)\overline{X^3} - 3\boldsymbol{E}\left(X^2\right)\overline{X^2} + 12\boldsymbol{E}\left(X\right)^2\overline{X^2} - 6\boldsymbol{E}\left(X\right)^3\bar{X} + \sum_{i=1}^{p}\overline{\mathcal{E}_i X^4}\ . \tag{7.15}$$

Et poursuivre avec $\sum_i \overline{\mathcal{E}_i X \Delta_i X^3}$ pour faire apparaître $X^3\tilde{X}_{\text{th}}$ :

$$\begin{aligned}\bar{\gamma_4}(X) + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X \Delta_i X^3} - \overline{\Delta_i X^4}\right] = \overline{X^3\tilde{X}_{\text{th}}} - 4\boldsymbol{E}\left(X\right)\overline{X^3} - 3\boldsymbol{E}\left(X^2\right)\overline{X^2} + 12\boldsymbol{E}\left(X\right)^2\overline{X^2} - 6\boldsymbol{E}\left(X\right)^3\bar{X} \\ + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X^4} - \overline{\mathcal{E}_i X^3 \mathcal{E}_i X}\right]\ .\end{aligned} \tag{7.16}$$

On peut maintenant transformer l'estimateur naïf de $\gamma_4$ pour construire un estimateur naif de $\gamma_4(X, X, X, \tilde{X}_{\text{th}})$ :

$$\begin{aligned}\bar{\gamma_4}(X, X, X, \tilde{X}_{\text{th}}) = \overline{X^3\tilde{X}_{\text{th}}} - \boldsymbol{E}\left(X\right)\overline{X^3} - 3\boldsymbol{E}\left(X\right)\overline{X^2\tilde{X}_{\text{th}}} - 3\boldsymbol{E}\left(X^2\right)\overline{X\tilde{X}_{\text{th}}} \\ + 6\boldsymbol{E}\left(X\right)^2\overline{X^2} + 6\boldsymbol{E}\left(X^2\right)\overline{X\tilde{X}_{\text{th}}} - 6\boldsymbol{E}\left(X\right)^3\bar{X}\ .\end{aligned} \tag{7.17}$$

Si on introduit cette expression dans l'estimateur en cours de construction de $\gamma_4$, on en tire :

$$\begin{aligned}\bar{\gamma_4}(X) + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X \Delta_i X^3} - \overline{\Delta_i X^4}\right] = \bar{\gamma_4}(X, X, X, \tilde{X}_{\text{th}}) + 3\overline{\left[2\boldsymbol{E}\left(X\right)^2 X - \boldsymbol{E}\left(X\right)X^2 - \boldsymbol{E}\left(X^2\right)X\right](X - \tilde{X}_{\text{th}})} \\ + \sum_{i=1}^{p}\overline{\mathcal{E}_i X^4} - \overline{\mathcal{E}_i X^3 \mathcal{E}_i X}\ .\end{aligned} \tag{7.18}$$

On fait ensuite apparaître les moyennes des cumulants quaternaires internes $\overline{\gamma_4(X|\bar{\imath})}$ dans l'équation globale :

$$\bar{\gamma_4}(X) + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X \Delta_i X^3} - \overline{\Delta_i X^4}\right] = \bar{\gamma_4}(X,X,X,\tilde{X}_{\text{th}}) + 3\overline{\left[2\boldsymbol{E}\left(X\right)^2 X - \boldsymbol{E}\left(X\right)X^2 - \boldsymbol{E}\left(X^2\right)X\right](X-\tilde{X}_{\text{th}})} + \sum_{i=1}^{p}\left[\overline{\gamma_4(X|\bar{\imath})} + 3\overline{\mathcal{E}_i X^3 \mathcal{E}_i X} + 3\overline{(\mathcal{E}_i X^2)^2} - 12\overline{\mathcal{E}_i X^2 (\mathcal{E}_i X)^2} + 6\overline{(\mathcal{E}_i X)^4}\right] . \tag{7.19}$$

On peut alors faire apparaître la variance interne $\mathrm{Var}\left(X|\bar{\imath}\right)$ et le cumulant ternaire interne $\gamma_3(X|\bar{\imath})$ dans la somme :

$$\bar{\gamma_4}(X) + \sum_{i=1}^{p}\left[\overline{\mathcal{E}_i X \Delta_i X^3} - \overline{\Delta_i X^4}\right] = \bar{\gamma_4}(X,X,X,\tilde{X}_{\text{th}}) + 3\overline{\left[2\boldsymbol{E}\left(X\right)^2 X - \boldsymbol{E}\left(X\right)X^2 - \boldsymbol{E}\left(X^2\right)X\right](X-\tilde{X}_{\text{th}})} + \sum_{i=1}^{p}\overline{\gamma_4(X|\bar{\imath})} + 3\overline{\mathcal{E}_i X \gamma_3(X|\bar{\imath})} + 3\overline{\mathcal{E}_i X^2 \mathrm{Var}\left(X|\bar{\imath}\right)} . \tag{7.20}$$

Dans la poursuite d'une expression similaire à l'équation (7.12), on va alors chercher à faire apparaître $\gamma_3(X,X,\mathrm{Var}\left(X|\bar{\imath}\right))$ et $\mathrm{Cov}\left(X,\gamma_3(X|\bar{\imath})\right)$. On va donc devoir décomposer ceux-ci ainsi que leurs estimateurs. Pour $\gamma_3(X,X,\mathrm{Var}\left(X|\bar{\imath}\right))$, cela nous donne :

$$\begin{aligned}
\gamma_3(X,X,\mathrm{Var}\left(X|\bar{\imath}\right)) &= \boldsymbol{E}\left(X^2\mathrm{Var}\left(X|\bar{\imath}\right)\right) - \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(\mathrm{Var}\left(X|\bar{\imath}\right)\right) \\
&\quad - 2\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X\mathrm{Var}\left(X|\bar{\imath}\right)\right) + 2\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(\mathrm{Var}\left(X|\bar{\imath}\right)\right) \\
&= \boldsymbol{E}\left(\mathcal{E}_i X^2\mathrm{Var}\left(X|\bar{\imath}\right)\right) - \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(\mathrm{Var}\left(X|\bar{\imath}\right)\right) \\
&\quad - 2\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(\mathcal{E}_i X\mathrm{Var}\left(X|\bar{\imath}\right)\right) + 2\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(\mathrm{Var}\left(X|\bar{\imath}\right)\right) ; \qquad (7.21a) \\
\bar{\gamma_3}(X,X,\mathrm{Var}\left(X|\bar{\imath}\right)) &= \overline{\mathcal{E}_i X^2\mathrm{Var}\left(X|\bar{\imath}\right)} - \boldsymbol{E}\left(X^2\right)\overline{\mathrm{Var}\left(X|\bar{\imath}\right)} - 2\boldsymbol{E}\left(X\right)\overline{\mathcal{E}_i X\mathrm{Var}\left(X|\bar{\imath}\right)} + 2\boldsymbol{E}\left(X\right)^2\overline{\mathrm{Var}\left(X|\bar{\imath}\right)} \\
&= \overline{\mathcal{E}_i X^2\mathrm{Var}\left(X|\bar{\imath}\right)} - \boldsymbol{E}\left(X^2\right)\left[\overline{\mathcal{E}_i X^2} - \overline{X\mathcal{E}_i X}\right] - 2\boldsymbol{E}\left(X\right)\left[\overline{\mathcal{E}_i X^2\mathcal{E}_i X} - \overline{(\mathcal{E}_i X)^3}\right] \\
&\quad + 2\boldsymbol{E}\left(X\right)^2\left[\overline{\mathcal{E}_i X^2} - \overline{X\mathcal{E}_i X}\right] . \qquad (7.21b)
\end{aligned}$$

On voit apparaître $\overline{\mathcal{E}_i X^2 \mathcal{E}_i X}$, sur lequel on peut faire disparaître l'opérateur $\mathcal{E}_i$ qui nous convient le plus à l'aide de variables de contrôle, et $\overline{(\mathcal{E}_i X)^3}$, que l'on ne peut pas aisément gérer. Maintenant, développons $\mathrm{Cov}\left(X,\gamma_3(X|\bar{\imath})\right)$ :

$$\begin{aligned}
\mathrm{Cov}\left(X,\gamma_3(X|\bar{\imath})\right) &= \boldsymbol{E}\left(X\gamma_3(X|\bar{\imath})\right) - \boldsymbol{E}\left(X\right)\boldsymbol{E}\left(\gamma_3(X|\bar{\imath})\right) \\
&= \boldsymbol{E}\left(\mathcal{E}_i X\gamma_3(X|\bar{\imath})\right) - \boldsymbol{E}\left(X\right)\boldsymbol{E}\left(\mathcal{E}_i X^3 - 3\mathcal{E}_i X^2\mathcal{E}_i X + 2(\mathcal{E}_i X)^3\right) \qquad (7.22) \\
\bar{C}(X,\gamma_3(X|\bar{\imath})) &= \overline{\mathcal{E}_i X\gamma_3(X|\bar{\imath})} - \boldsymbol{E}\left(X\right)\left[\overline{\mathcal{E}_i X^3} - 3\overline{\mathcal{E}_i X^2\mathcal{E}_i X} + 2\overline{(\mathcal{E}_i X)^3}\right] .
\end{aligned}$$

On peut sommer les deux expressions pour simplifier :

$$\begin{aligned}
\bar{\gamma_3}(X,X,\mathrm{Var}\left(X|\bar{\imath}\right)) + \bar{C}(X,\gamma_3(X|\bar{\imath})) &= \overline{\mathcal{E}_i X^2\mathrm{Var}\left(X|\bar{\imath}\right)} + \overline{\mathcal{E}_i X\gamma_3(X|\bar{\imath})} - \boldsymbol{E}\left(X^2\right)\left[\overline{\mathcal{E}_i X^2} - \overline{X\mathcal{E}_i X}\right] \\
&\quad + \boldsymbol{E}\left(X\right)\left[\overline{\mathcal{E}_i X^2\mathcal{E}_i X} - \overline{\mathcal{E}_i X^3}\right] + 2\boldsymbol{E}\left(X\right)^2\left[\overline{\mathcal{E}_i X^2} - \overline{X\mathcal{E}_i X}\right] \\
&= \overline{\mathcal{E}_i X^2\mathrm{Var}\left(X|\bar{\imath}\right)} + \left(2\boldsymbol{E}\left(X\right)^2 - \boldsymbol{E}\left(X^2\right)\right)\left[\overline{X^2} - \overline{\Delta_i X^2} - \overline{X\mathcal{E}_i X}\right] \\
&\quad + \overline{\mathcal{E}_i X\gamma_3(X|\bar{\imath})} + \boldsymbol{E}\left(X\right)\left[\overline{X^2\mathcal{E}_i X} + \overline{\Delta_i X^3} - \overline{\mathcal{E}_i X\Delta_i X^2} - X^3\right] .
\end{aligned} \tag{7.23}$$

Si on réinsère cela dans notre équation (7.20), on arrive à :

$$\begin{aligned}\bar{\gamma_4}(X) + \sum_{i=1}^{p} \mathcal{E}_i X \Delta_i X^3 - \Delta_i X^4 = &\sum_{i=1}^{p} \left[3\bar{\gamma_3}(X, X, \mathrm{Var}\,(X|\bar{\imath})) + 3\bar{C}(X, \gamma_3(X|\bar{\imath})) + \overline{\gamma_4(X|\bar{\imath})}\right] \\ &- 3\sum_{i=1}^{p} \left[\boldsymbol{E}\,(X)\,\overline{\Delta_i X^3} + \boldsymbol{E}\left(X^2\right)\overline{\Delta_i X^2} - 2\boldsymbol{E}\,(X)^2\,\overline{\Delta_i X^2} - \boldsymbol{E}\,(X)\,\overline{\mathcal{E}_i X \Delta_i X^2}\right] \\ &+ \bar{\gamma_4}(X, X, X, \tilde{X}_{\mathrm{th}})\ . \end{aligned} \tag{7.24}$$

Les termes de la seconde ligne sont évidemment des variables de contrôle de la forme $\sum_i \overline{\mathcal{E}_i A \Delta_i B}$, que l'on peut changer de côté pour arriver à l'estimateur amélioré suivant :

$$\tilde{\gamma_4}(X) = \bar{\gamma_4}(\tilde{X}_{\mathrm{th}}, X, X, X) + \sum_{i=1}^{p} \left[3\bar{\gamma_3}(X, X, \mathrm{Var}\,(X|\bar{\imath})) + 3\mathrm{Cov}\,(X, \gamma_3(X|\bar{\imath})) + \overline{\gamma_4(X|\bar{\imath})}\right]\ . \tag{7.25}$$

Pour les mêmes raisons que l'estimateur (7.11), il s'agit d'un estimateur non-biaisé, zéro-variant dans la limite de séparabilité. On pourra donc en déduire des estimateurs pratiques, non biaisés dans la limite d'échantillonnages et sous-échantillonnages infinis, et zéro-variant dans la limite de sous-systèmes indépendants avec des sous-échantillonnages infinis.

On peut résumer la construction de cette section sous la forme suivante, plus générale :

$$\gamma_4(X) = \gamma_4(X, X, X, \tilde{X}) + 3\sum_{i=1}^{p} \gamma_3(X, X, \gamma_2(X|\bar{\imath})) + 3\sum_{i=1}^{p} \gamma_2(X, \gamma_3(X|\bar{\imath})) + \sum_{i=1}^{p} \gamma_1(\gamma_4(X|\bar{\imath}))\ . \tag{7.26}$$

On peut remarquer que les formes des estimateurs qu'on a obtenus suggère l'existence d'une expression générale de forme binômiale. C'est ce que l'on va s'attacher à démontrer à la section suivante.

## 7.3 Généralisation

Si l'on reprend les expressions (7.12) et (7.26), on peut remarquer que $\tilde{X}_{\mathrm{th}}$ n'apparaît de manière directe que dans le premier terme, le cumulant d'ordre supérieur. De plus, on peut remarquer que les démonstrations qu'on a employées sont tout aussi valides si on avait pris un seul sous-système $\mathcal{S}_i$ et non une partition du système, seules les propriétés de l'estimateur final vis-à-vis de la zéro-variance étant affectées. On peut alors poser que :

$$\begin{aligned} &\forall \mathcal{S}_i \subset \mathcal{S}, \\ &\gamma_3(X) = \gamma_3(X, X, \gamma_1(X|\bar{\imath})) + 2\gamma_2(X, \gamma_2(X|\bar{\imath})) + \gamma_1(\gamma_3(X|\bar{\imath})) \\ &\gamma_4(X) = \gamma_4(X, X, X, \gamma_1(X|\bar{\imath})) + 3\gamma_3(X, X, \gamma_2(X|\bar{\imath})) + 3\gamma_2(X, \gamma_3(X|\bar{\imath})) + \gamma_1(\gamma_4(X|\bar{\imath}))\ . \end{aligned} \tag{7.27}$$

Cela nous donne donc une forme bien plus simple à démontrer. La propriété que l'on devine, et que l'on va alors chercher à démontrer, est la suivante :

$$\forall q \in \mathbb{N},\ \gamma_{1+q}(X) = \sum_{j=0}^{q} \frac{q!}{j!(q-j)!} \gamma_{1+j}(X \ldots X, \gamma_{1+q-j}(X|\bar{\imath}))\ . \tag{7.28}$$

Pour cela, on va avoir besoin de plusieurs résultats. Pour commencer, on va devoir utiliser la décomposition des moments sur les cumulants (qui est démontrée à l'annexe C.2). Pour ce, on définit la fonction $f$ sur les suites presque nulles (nulles à partir d'un certain rang) de $\mathbb{N}$ par :

$$f : (x \in (\mathbb{N})^{\mathbb{N}^*}) \mapsto \sum_{k=1}^{\infty} k x_k\ . \tag{7.29}$$

Cette fonction nous permet de définir, pour tout entier $q$ strictement positif l'ensemble $I_q$ comme la préimage de $\{q\}$ par $f$ – autrement posé, $I_q = \{x \in \mathbb{N}^{\mathbb{N}*}, f(x) = q\}$. Alors, le développement du moment d'ordre $q$ sur les cumulants est donné par :

$$\boldsymbol{E}\left(X^q\right) = \sum_{x\in I_q} \frac{q!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \,. \tag{7.30}$$

Cette expression, que nous démontrons à l'annexe C.2, nous permet d'obtenir un résultat intéressant sur la décomposition du moment mixte d'ordre $q+1$ $\boldsymbol{E}\left(X^qY\right)$ sur les cumulants mixtes $\gamma_1(Y)$, $\gamma_2(X,Y)$, et de manière générale $\gamma_j(X\dots X,Y)$. En effet, si on cherche à asymétriser l'expression d'ordre $q+1$, on peut asymétriser chacun des termes de la suite en raisonnant en termes probabilistes. Si on a $x \in I_{q+1}$, alors la probabilité que $Y$ remplace un $X$ dans un cumulant d'ordre $k$ est $kx_k/(q+1)$. Introduire cette brisure de symétrie dans l'équation (7.30) nous donne alors :

$$\boldsymbol{E}\left(X^qY\right) = \sum_{x\in I_{q+1}} \frac{(q+1)!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \left[\sum_{k=1}^{\infty} \frac{kx_k\gamma_k(Y,X\dots X)}{(q+1)\gamma_k(X)}\right] \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \,. \tag{7.31}$$

On va ensuite réorganiser cette équation pour faire apparaître les préfacteurs des cumulants mixtes $\gamma_j(Y,X\dots X)$ :

$$\begin{aligned}
\boldsymbol{E}\left(X^qY\right) &= \sum_{x\in I_{q+1}} \frac{(q+1)!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \left[\sum_{k=1}^{\infty} \frac{kx_k\gamma_k(Y,X\dots X)}{(q+1)\gamma_k(X)}\right] \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \\
&= \sum_{j=1}^{q+1} \sum_{\substack{x\in I_{q+1}\\ x_j>0}} \frac{(q+1)!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \left[\frac{jx_j\gamma_j(Y,X\dots X)}{(q+1)\gamma_j(X)}\right] \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \,;
\end{aligned} \tag{7.32}$$

$$\begin{aligned}
\boldsymbol{E}\left(X^qY\right) &= \sum_{j=1}^{q+1} \frac{\gamma_j(Y,X\dots X)}{\gamma_j(X)} \sum_{\substack{x\in I_{q+1}\\ x_j>0}} \frac{q!jx_j}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \\
&= \sum_{j=1}^{q+1} \gamma_j(Y,X\dots X) \sum_{x\in I_{q+1-j}} \frac{q!}{(j-1)!\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \\
&= \sum_{j=1}^{q+1} \frac{q!}{(j-1)!(q+1-j)!} \gamma_j(Y,X\dots X) \sum_{x\in I_{q+1-j}} \frac{(q+1-j)!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]} \prod_{k=1}^{\infty} \gamma_k(X)^{x_k} \,.
\end{aligned} \tag{7.33}$$

On peut reconnaître $\boldsymbol{E}\left(X^{q+1-j}\right)$ dans la dernière expression, donc on peut recondenser (en changeant l'ordre des indices) :

$$\boldsymbol{E}\left(X^qY\right) = \sum_{j=0}^{q} \frac{q!}{j!(q-j)!} \boldsymbol{E}\left(X^{q-j}\right) \gamma_{1-j}(Y,X\dots X) \,. \tag{7.34}$$

Le résultat suivant dont on a besoin concerne le moment d'ordre $q+1$. L'estimateur classique de celui-ci est bien évidemment :

$$\bar{m}_{q+1}(X) = \overline{X^{q+1}} \,. \tag{7.35}$$

On peut y rajouter des variables de contrôle en $\overline{\mathcal{E}_i A \Delta_i B}$ pour construire l'estimateur suivant :

$$\begin{aligned}
\tilde{m}_{q+1}(X) &= \bar{m}_{q+1}(X) - \overline{\Delta_i X^{q+1}} + \overline{\varepsilon_i X \Delta_i X^q} \\
&= \overline{\mathcal{E}_i X^{q+1}} + \overline{X^q \mathcal{E}_i X} - \overline{\mathcal{E}_i X \mathcal{E}_i X^q} \\
&= \overline{X^q \mathcal{E}_i X} + \overline{\mathrm{Cov}\left(X^q, X|\bar{\imath}\right)} \,.
\end{aligned} \tag{7.36}$$

Il s'agit bien évidemment d'un estimateur non biaisé (que l'on aurait également pû tirer à l'aide de l'estimateur de la covariance de $X$ et $X^q$).

On va maintenant démontrer la propriété (7.28) par un raisonnement par récurrence forte. Soit, $\forall q \in \mathbb{N}^*$, $\mathcal{P}(q)$ la proposition suivante :

$$\gamma_{q+1}(X) = \sum_{j=0}^{q} \frac{q!}{j!(n-j)!} \gamma_{j+1}\left(X \dots X, \sum_{i=1}^{p} \gamma_{1+q-j}(X|\bar{\imath})\right) .$$

**Initialisation** : On a démontré plus haut que la propriété était vraie aux rang 1 (il s'agit de la partition des covariances), 2 et 3 (7.27).

**Hérédité** : Soit $q \in \mathbb{N}^*$ tel que $\forall k < q$, $\mathcal{P}(k)$ vraie. On a alors :

$$\boldsymbol{E}\left(X^{q+1}\right) = \boldsymbol{E}\left(X^q \mathcal{E}_i X\right) + \boldsymbol{E}\left(\mathrm{Cov}\left(X, X^q|\bar{\imath}\right)\right) . \tag{7.37}$$

On peut alors introduire l'équation (7.34), en posant dans un premier temps $Y = X$ et dans un second temps $Y = \mathcal{E}_i X$. Cela nous donne :

$$\sum_{j=0}^{q} \frac{q!}{j!(q-j)!} \boldsymbol{E}\left(X^{q-j}\right) \gamma_{1+j}(X) = \sum_{j=0}^{q} \frac{q!}{j!(q-j)!} \boldsymbol{E}\left(X^{q-j}\right) \gamma_{1+j}(\mathcal{E}_i X, X \dots X) + \boldsymbol{E}\left(\mathrm{Cov}\left(X, X^q|\bar{\imath}\right)\right) . \tag{7.38}$$

Dans nos sommes, les termes de rang 0 se compensent car $\gamma_1(X) = \boldsymbol{E}(X) = \gamma_1(\mathcal{E}_i X)$, et le terme de rang $q$ est bien évidemment celui qu'on cherche à garder. Cela nous donne alors :

$$\gamma_{1+q}(X) = \gamma_{1+q}(X \dots X, \mathcal{E}_i X) + \sum_{j=1}^{q-1} \frac{q!}{j!(q-j)!} \boldsymbol{E}\left(X^{q-j}\right) \left[\gamma_{1+j}(X \dots X, \mathcal{E}_i X) - \gamma_{1+j}(X)\right] + \boldsymbol{E}\left(\mathrm{Cov}\left(X, X^q|\bar{\imath}\right)\right) . \tag{7.39}$$

On peut maintenant faire intervenir l'hypothèse de récurrence forte pour obtenir l'expression des termes restants.

$$\begin{aligned}\gamma_{1+q}(X) = {} & \gamma_{1+q}(X \dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\mathcal{E}_i X^{1+q}\right) - \boldsymbol{E}\left(\mathcal{E}_i X \mathcal{E}_i X^q\right) \\ & + \sum_{j=1}^{q-1} \frac{q!}{j!(q-j)!} \boldsymbol{E}\left(X^{q-j}\right) \left[\gamma_{1+j}(X \dots X, \mathcal{E}_i X) - \sum_{k=0}^{j} \frac{j!}{k!(j-k)!} \gamma_{1+j-k}(X \dots X, \gamma_{1+k}(X|\bar{\imath}))\right] .\end{aligned} \tag{7.40}$$

On peut bien évidemment remarquer que le rang 0 de la somme intérieure se compense avec le terme $\gamma_{1+j}(X \dots X, \mathcal{E}_i X)$ et condenser l'expression, ce qui nous donne :

$$\begin{aligned}\gamma_{1+q}(X) = {} & \gamma_{1+q}(X \dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\mathcal{E}_i X^{1+q}\right) - \boldsymbol{E}\left(\mathcal{E}_i X \mathcal{E}_i X^q\right) \\ & - \sum_{j=1}^{q-1} \sum_{k=1}^{j} \frac{q! \boldsymbol{E}\left(X^{q-j}\right)}{k!(j-k)!(q-j)!} \gamma_{1+j-k}(X \dots X, \gamma_{1+k}(X|\bar{\imath})) .\end{aligned} \tag{7.41}$$

Si on intervertit les sommes avec $j' = j - k$, cela devient :

$$\begin{aligned}\gamma_{1+q}(X) = {} & \gamma_{1+q}(X \dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\mathcal{E}_i X^{1+q}\right) - \boldsymbol{E}\left(\mathcal{E}_i X \mathcal{E}_i X^q\right) \\ & - \sum_{k=1}^{q-1} \sum_{j'=0}^{q-k-1} \frac{q! \boldsymbol{E}\left(X^{q-k-j'}\right)}{k! j'! (q-j'-k)!} \gamma_{1+j'}(X \dots X, \gamma_{1+k}(X|\bar{\imath})) .\end{aligned} \tag{7.42}$$

L'étape suivante revient à développer $\mathcal{E}_i X^{n+1}$ sur ses cumulants, à l'aide de l'équation (7.34) :

$$\gamma_{1+q}(X) = \gamma_{1+q}(X\dots X, \mathcal{E}_i X) - \boldsymbol{E}\left(\mathcal{E}_i X \mathcal{E}_i X^q\right) - \sum_{k=1}^{q-1}\sum_{j'=0}^{q-k-1} \frac{q!\boldsymbol{E}\left(X^{q-k-j'}\right)}{k!j'!(q-k-j')!}\gamma_{1+j'}(X\dots X, \gamma_{1+k}(X|\bar{\imath})) + \boldsymbol{E}\left(\sum_{j=0}^{q}\frac{q!\mathcal{E}_i X^{q-j}}{j!(q-j!)}\gamma_{1+j}(X|\bar{\imath})\right) . \tag{7.43}$$

Le terme de rang zéro de cette somme se compense avec $\boldsymbol{E}\left(\mathcal{E}_i X \mathcal{E}_i X^q\right)$, et on peut alors rassembler les sommes (en remplaçant $j$ par $k$ et faisant sortir le terme de rang $q$) :

$$\gamma_{1+q}(X) = \gamma_{1+q}(X\dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\gamma_{1+q}(X|\bar{\imath})\right) + \sum_{k=1}^{q-1}\frac{q!}{k!(q-k!)}\left[\boldsymbol{E}\left(\mathcal{E}_i X^{q-k}\gamma_{1+k}(X|\bar{\imath})\right) - \sum_{j'=0}^{q-k-1}\frac{(q-k)!\boldsymbol{E}\left(X^{q-k-j'}\right)}{j'!(q-k-j')!}\gamma_{1+j'}(X\dots X, \gamma_{1+k}(X|\bar{\imath}))\right] . \tag{7.44}$$

Comme on a $\boldsymbol{E}\left(\mathcal{E}_i X^{q-k}\gamma_{1+k}(X|\bar{\imath})\right) = \boldsymbol{E}\left(X^{q-k}\gamma_{1+k}(X|\bar{\imath})\right)$, on peut réemployer l'équation (7.34) pour le décomposer pour obtenir des sommes qui se compensent :

$$\begin{aligned}
\gamma_{1+q}(X) &= \gamma_{1+q}(X\dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\gamma_{1+q}(X|\bar{\imath})\right) \\
&\quad + \sum_{k=1}^{q-1}\frac{q!}{k!(q-k!)}\left[\sum_{j'=0}^{q-k}\frac{(q-k)!\boldsymbol{E}\left(X^{q-k-j'}\right)}{j'!(q-k-j')!}\gamma_{1+j'}(X\dots X, \gamma_{1+k}(X|\bar{\imath}))\right] \\
&\quad - \sum_{k=1}^{q-1}\frac{q!}{k!(q-k!)}\left[\sum_{j'=0}^{q-k-1}\frac{(q-k)!\boldsymbol{E}\left(X^{q-k-j'}\right)}{j'!(q-k-j')!}\gamma_{1+j'}(X\dots X, \gamma_{1+k}(X|\bar{\imath}))\right] \\
&= \gamma_{1+q}(X\dots X, \mathcal{E}_i X) + \boldsymbol{E}\left(\gamma_{1+q}(X|\bar{\imath})\right) + \sum_{k=1}^{q-1}\frac{q!}{k!(q-k!)}\gamma_{1+q-k}(X\dots X, \gamma_{1+k}(X|\bar{\imath})) \\
&= \sum_{k=0}^{q}\frac{q!}{k!(q-k!)}\gamma_{1+q-k}(X\dots X, \gamma_{1+k}(X|\bar{\imath})) .
\end{aligned} \tag{7.45}$$

C'est là $\mathcal{P}(q)$. On a donc $(\forall k \in [[1, q-1]],\ \mathcal{P}(k)$ vraie$) \Longrightarrow \mathcal{P}(q)$ vraie.

**Conclusion :** On a donc, par récurrence forte, $\mathcal{P}(n)$ vraie pour tout ordre $n$ entier non nul, c'est-à-dire l'équation (7.28) :

$$\forall q \in \mathbb{N},\ \gamma_{q+1}(X) = \sum_{j=0}^{q}\frac{q!}{j!(n-j)!}\gamma_{j+1}\left(X\dots X, \sum_{i=1}^{p}\gamma_{1+q-j}(X|\bar{\imath})\right) .$$

Pour construire un estimateur général théorique non biaisé, zéro-variant dans la limite de séparabilité, on peut remarquer que :

$$\begin{aligned}
\forall q \in \mathbb{N},\ \gamma_{q+1}(X) &= \gamma_{q+1}(X) + \sum_{i=1}^{p}\left[\gamma_{q+1}(X) - \gamma_{q+1}(X)\right] \\
&= \gamma_{q+1}(X) + \sum_{i=1}^{p}\left[\sum_{j=0}^{q}\frac{q!}{j!(n-j)!}\gamma_{j+1}\left(X\dots X, \sum_{i=1}^{p}\gamma_{1+q-j}(X|\bar{\imath})\right) - \gamma_{q+1}(X)\right]
\end{aligned} \tag{7.46}$$

On n'a alors qu'à remplacer les cumulants principaux par des estimateurs classiques non biaisés, pour obtenir :

$$\begin{aligned}\tilde{\gamma}_{1+q,\mathrm{th}}(X) &= \bar{\gamma}_{1+q}(X) + \sum_{i=1}^{p}\left[\sum_{j=0}^{q}\frac{q!}{j!(q-j)!}\bar{\gamma}_{1+q-k}(X\dots X,\gamma_{1+k}(X|\bar{\imath})) - \bar{\gamma}_{1+q}(X)\right] \\ &= \bar{\gamma}_{1+q}(X\dots X,\tilde{X}_{\mathrm{th}}) + \sum_{j=0}^{q}\frac{q!}{j!(q-j)!}\bar{\gamma}_{1+q-k}\left(X\dots X,\sum_{i=1}^{p}\gamma_{1+k}(X|\bar{\imath})\right)\ .\end{aligned} \tag{7.47}$$

Par ailleurs, la forme binômiale de la propriété (7.28) suggère une propriété sous-jacente sur des séries entières. On va alors chercher à la section suivante à ramener cette propriété à l'égalité de deux développements en série de Taylor.

## 7.4 Passage à l'exponentielle et fonctions génératrices

On rappelle qu'en statistiques, la fonction génératrice des moments (resp. des cumulants) de $X$ est la fonctions $G_m(X)$ (resp $G_c(X)$) de $\mathbb{C}$, holomorphe sur son espace de définition, construite telle que son développement en série de Taylor en l'origine $z=0$ donne les moments successifs de $X$ (resp. les cumulants successifs de $X$) :

$$\begin{aligned}\left.\frac{d^nG_m(X)}{dz^n}\right|_{z=0} &= m_n(X) = \boldsymbol{E}\left(X^n\right)\ ; \\ \left.\frac{d^nG_c(X)}{dz^n}\right|_{z=0} &= \gamma_n(X)\ .\end{aligned} \tag{7.48}$$

Le développement en série de Taylor de l'application $z\mapsto \boldsymbol{E}\left(e^{zX}\right) = \int_\Omega \exp(zX(\mathcal{R}))\rho(\mathcal{R})d\mathcal{R}$ nous montre que celle-ci satisfait bien la condition qui définit $G_m$. Donc, partout où l'intégrale converge, on a $G_m(X) = \boldsymbol{E}\left(e^{tX}\right)$.[1] La fonction génératrice des cumulants est elle définie par $G_c(X) = \log(G_m(X))$. Des démonstrations plus avancées sont données dans l'annexe B.1.

Intéressons-nous maintenant à la propriété (7.28). On peut reconnaître en celle-ci une formule binômiale, qui rappelle beaucoup la formule de Leibnitz.

Or, on a pu observer (voir annexe B.5) que la covariance d'un estimateur classique de cumulant d'ordre $q$ et d'une variable semble prendre la forme d'un cumulant d'ordre $q+1$. On peut exprimer cette propriété sous la forme suivante. Si $\mathrm{Cov}\,(X,\cdot) : Y\mapsto \mathrm{Cov}\,(X,Y)$, alors :

$$\gamma_{n+1}(X,...) = \mathrm{Cov}\,(X,\cdot)\,(\gamma_n(...))\ . \tag{7.49}$$

On peut interpréter cela comme une équation différentielle sur $G_c$ :

$$\frac{\partial}{\partial z}G_c(X) = \mathrm{Cov}\,(X,\cdot)\,(G_c(X))\ . \tag{7.50}$$

Cela nous permet d'exprimer $G_c$ de la manière suivante :

$$G_c(X) = \exp(\mathrm{Cov}\,(X,\cdot))(1)\ . \tag{7.51}$$

Dans ces conditions, l'équation (7.28) semble pouvoir se mettre sous la forme suivante :

$$\exp(\mathrm{Cov}\,(X,\cdot))(X) = \exp(\mathrm{Cov}\,(\mathcal{E}_iX,\cdot) + \mathrm{Cov}\,(X,\cdot))(\mathcal{E}_iX)\ . \tag{7.52}$$

D'autre part, si on réutilise l'expression (7.30) pour reconstruire une expression conditionnelle des moments, on arrive à :

$$\tilde{m}_q = \sum_{j=1}^{q}\frac{q!}{j!(q-j)!}X^{q-j}\gamma_j(X|\bar{\imath})\ . \tag{7.53}$$

1. La fonction génératrice des moments n'est pas nécessairement définie partout. Par exemple, si $\rho(x) = e^{-x}$ et $X = x$ sur $\mathbb{R}^+$, notre application est seulement définie sur $]-\infty,1[+i\mathbb{R}$. Cependant, il est suffisant qu'elle soit définie sur un voisinage de 0.

Regrouper en une seule série de Taylor donne l'expression suivante :

$$G_m(X)(z) - 1 = \boldsymbol{E}\left(\sum_{k=0}^{\infty}\frac{z^k X^k}{k!}\sum_{n=1}^{\infty}\frac{z^n \gamma_n(X|\bar{\imath})}{(n-1)!(n+k)}\right) . \tag{7.54}$$

Le facteur en $1/(n+k)$ suggère que l'on travaille avec la mauvaise somme, on passe donc à la dérivée, car on a reconnu l'expression de la dérivée de $G_c(X|\bar{\imath})$ :

$$\frac{dG_m(X)}{dz} = \boldsymbol{E}\left(\frac{dG_c(X|\bar{\imath})}{dz}e^{zX}\right) . \tag{7.55}$$

Cela se ramène à l'expression suivante en se ramenant sur $G_c$ :

$$\frac{dG_c(X)}{dz} = \frac{\boldsymbol{E}\left(\frac{dG_c(X|\bar{\imath})}{dz}e^{zX}\right)}{\boldsymbol{E}\left(e^{zX}\right)} . \tag{7.56}$$

On peut vérifier que l'expression 7.56 nous permet bien de retrouver l'expression générale des cumulants trouvée à la section précédente. La comparaison avec l'expression générale de la dérivée première de la fonction génératrice des cumulants nous donne :

$$\boldsymbol{E}\left(Xe^{zX}\right) = \boldsymbol{E}\left(\frac{dG_c(X|\bar{\imath})}{dz}e^{zX}\right) . \tag{7.57}$$

Ou, autrement posé :

$$\boldsymbol{E}\left(Xe^{zX}\right) = \boldsymbol{E}\left(\boldsymbol{E}\left(Xe^{zX}|\bar{\imath}\right)\frac{e^{zX}}{\boldsymbol{E}\left(e^{zX}|\bar{\imath}\right)}\right) . \tag{7.58}$$

Ces expressions pourront bien sûr être ultérieurement recomposées comme on a fait à la fin de la section 7.3 pour reconstruire des estimateurs théoriques zéro-variants dans la limite de séparabilité.

Les deux équations précédentes peuvent rappeler diverses expressions, parmi lesquelles les fonctions de partitions d'ensembles canoniques, l'échantillonage par importance, ainsi que certaines manières de poser le problème du signe fermionique. On peut donc proposer comme perspectives l'application de ces équations à ces divers domaines.

## Récapitulatif

Dans ce chapitre, nous avons construit une formule générale permettant le calcul d'estimateurs améliorés de cumulants d'une variable aléatoire $X$ extensive, qui sont non biaisés dans la limite d'échantillonnages et sous-échantillonnages infinis et zéro-variants dans la limite de sous-systèmes indépendents avec des sous-échantillonnages infinis. Qui plus est, nous avons réussi à ramener cette expression à une propriété sur les fonctions génératrices, ce qui peut potentiellement trouver une application dans le cadre du problème du signe (mentionné section 2.5).

Maintenant, il nous reste à montrer comment, en pratique, ces estimateurs pourront nous servir à calculer les valeurs de dérivées.

# Chapitre 8

# Implémentation et applications

Dans ce chapitre, nous allons nous intéresser aux détails de l'implémentation des estimateurs qu'on a construit aux chapitres précédents ainsi qu'à l'usage que l'on peut en faire. Ainsi, dans la première section, nous nous appuierons sur les résultats de l'annexe B.5 pour élaborer un algorithme rapide de construction pratique de cumulants. Dans la deuxième section, nous rappellerons les liens analytiques qui relient les cumulants et le calcul de dérivées. Enfin, dans la troisième section, nous développerons en pratique, en se resservant du formalisme matriciel qu'on a présenté au chapitre 4, la manière dont on peut, sur le modèle de Hubbard, obtenir les quantités nécessaires aux cumulants qu'on aura retenues à la fin de la section précédente.

## 8.1 Algorithme de construction pratique de cumulants

Pour une variable extensive quelconque, on peut envisager plusieurs difficultés pour des systèmes de grande taille. En effet, si notre variable a une valeur moyenne en $\mathcal{O}(N)$, ses fluctuations liées à la variance sont d'ordre $\mathcal{O}\left(\sqrt{N}\right)$, et les cumulants successifs sont d'autant plus faibles. On ne peut donc pas se contenter de passer d'une variable extensive à une variable intensive (en divisant par $N$). Pour une raison similaire, on ne peut pas non plus se contenter de travailler en stockant des moyennes, car cela risque de multiplier les erreurs numériques successives liées au processeur.

C'est pourquoi la première étape, si on dispose d'un vecteur de variables aléatoires $\vec{X}$ comportant les variables aléatoires $X_1, X_2, ...X_q$, consiste à pseudo-centrer ces variables, c'est à dire à retrancher à $\vec{X}$ un vecteur $\vec{X_0}$ raisonnablement proche de $\boldsymbol{E}(\vec{X})$, de manière à faire disparaître le terme dominant en $\mathcal{O}(N)$. On peut ensuite décider (ou non) de diviser le vecteur par $\sqrt{N}$.

Une fois qu'on dispose de ce vecteur $\vec{X'}$, on peut ensuite décider d'empiler les moments de ce vecteur. Si on souhaite travailler en pratique avec des valeurs de cumulants allant jusqu'à l'ordre $K$, on doit se servir de structures de données ayant jusqu'à $2K$ dimensions. On peut cependant gagner significativement sur la complexité en mémoire si on utilise les symétries des moments et cumulants en se servant d'une structure de données invariante par permutation.

Une fois la dynamique principale terminée, on peut passer nos piles à des valeurs moyennes, pour ensuite centrer nos moments. Passer des moments aux cumulants peut ensuite être réalisé aux moyen des méthodes développées dans l'annexe B.4. On peut alors passer des cumulants ainsi construits aux estimateurs améliorés "composites", et se servir des expressions de l'annexe B.5 pour en calculer les variances.

Bien sûr, comme l'estimateur comparable à $\bar{X}$ est $\bar{V}$, on se servira non de la variance totale donnée par les résultats de l'annexe, mais $M$ fois celle-ci.

## 8.2 Application pratique au calcul de dérivées

Soit $\alpha$ un paramètre de fonction d'onde que l'on cherche à optimiser. On va chercher à calculer la dérivée de l'énergie variationnelle par rapport à ce paramètre, $\partial E_v/\partial\alpha$. Alors, on a (voir, par exemple, [1] et [2]) :

$$\begin{aligned}\frac{\partial E_v}{\partial\alpha} &= \frac{\partial}{\partial\alpha}\left[\frac{\int_\Omega \rho(\mathcal{R})E_l(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\right] \\ &= \frac{1}{\left(\int_\Omega \rho(\mathcal{R})d\mathcal{R}\right)^2}\left[\frac{\partial\int_\Omega \rho(\mathcal{R})E_l(\mathcal{R})d\mathcal{R}}{\partial\alpha}\int_\Omega \rho(\mathcal{R})d\mathcal{R} - \frac{\partial\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\partial\alpha}\int_\Omega \rho(\mathcal{R})E_l(\mathcal{R})d\mathcal{R}\right] .\end{aligned} \tag{8.1}$$

On peut simplifier cette expression en :

$$\frac{\partial E_v}{\partial\alpha} = \frac{1}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\left[\frac{\partial\int_\Omega \rho(\mathcal{R})E_l(\mathcal{R})d\mathcal{R}}{\partial\alpha} - E_v\frac{\partial\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\partial\alpha}\right] . \tag{8.2}$$

Si on suppose qu'on peut faire les interversions de limites pour échanger dérivée partielle et intégrale, alors cela donne :

$$\frac{\partial E_v}{\partial\alpha} = \frac{1}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\int_\Omega\left[\rho(\mathcal{R})\frac{\partial E_l}{\partial\alpha}(\mathcal{R}) + (E_l(\mathcal{R}) - E_v)\frac{\partial\rho}{\partial\alpha}(\mathcal{R})\right]d\mathcal{R} . \tag{8.3}$$

Si on passe de $\rho$ à $\psi$, cela nous donne :

$$\frac{\partial E_v}{\partial\alpha} = \frac{1}{\int_\Omega |\psi(\mathcal{R})|^2\, d\mathcal{R}}\left[\int_\Omega \frac{\partial E_l}{\partial\alpha}(\mathcal{R})\,|\psi(\mathcal{R})|^2\, d\mathcal{R} + \int_\Omega\left[2\Re(\frac{1}{\psi}\frac{\partial\psi}{\partial\alpha})(E_l(\mathcal{R}) - E_v)\,|\psi(\mathcal{R})|^2\right]d\mathcal{R}\right] . \tag{8.4}$$

En repassant sur des espérances mathématiques, on arrive à :

$$\begin{aligned}\frac{\partial E_v}{\partial\alpha} &= \boldsymbol{E}\left(\frac{\partial E_l}{\partial\alpha}\right) + 2\boldsymbol{E}\left(\frac{\partial\ln\psi}{\partial\alpha}(E_l - E_v)\right) \\ &= \boldsymbol{E}\left(\frac{\partial E_l}{\partial\alpha}\right) + 2\boldsymbol{E}\left(\frac{\partial\ln\psi}{\partial\alpha}E_l\right) - 2E_v\boldsymbol{E}\left(\frac{\partial\ln\psi}{\partial\alpha}\right) \\ &= \boldsymbol{E}\left(\frac{\partial E_l}{\partial\alpha}\right) + 2\,\mathrm{Cov}\left(\frac{\partial\ln\psi}{\partial\alpha}, E_l\right) .\end{aligned} \tag{8.5}$$

On peut montrer que $\boldsymbol{E}\left(\frac{\partial E_l}{\partial\alpha}\right) = 0$ car l'opérateur Hamiltonien est hermitien, et s'en servir comme d'une variable de contrôle.

Ceci fait, on va chercher à calculer de la même manière la dérivée seconde de l'énergie variationnelle

(et élément de la matrice hessienne de celle-ci) $\partial^2 E_v/\partial\alpha\partial\beta$ :

$$\begin{aligned}
\frac{\partial^2 E_v}{\partial\alpha\partial\beta} &= \frac{\partial}{\partial\beta}\frac{\partial}{\partial\alpha}\left[\frac{\int_\Omega \rho(\mathcal{R})E_l(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\right] \\
&= \frac{\partial}{\partial\beta}\left[\frac{\frac{\partial}{\partial\alpha}\int_\Omega E_l(\mathcal{R})\rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}} - E_v\frac{\frac{\partial}{\partial\alpha}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\right] \\
&= \frac{\frac{\partial^2}{\partial\alpha\partial\beta}\int_\Omega E_l(\mathcal{R})\rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}} - \frac{\frac{\partial}{\partial\alpha}\int_\Omega E_l(\mathcal{R})\rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\frac{\frac{\partial}{\partial\beta}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}} - E_v\frac{\frac{\partial^2}{\partial\alpha\partial\beta}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}} \\
&+ 2E_v\frac{\frac{\partial}{\partial\alpha}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\frac{\frac{\partial}{\partial\beta}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}} - \frac{\frac{\partial}{\partial\beta}\int_\Omega E_l(\mathcal{R})\rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\frac{\frac{\partial}{\partial\alpha}\int_\Omega \rho(\mathcal{R})d\mathcal{R}}{\int_\Omega \rho(\mathcal{R})d\mathcal{R}}\,.
\end{aligned} \tag{8.6}$$

On a ainsi cinq termes à évaluer individuellement. On admet qu'on peut faire toutes les interversions de limites.

Le premier terme revient à calculer $\partial^2(\psi^*\psi E_l)/\partial\alpha\partial\beta$, ce qui nous donne :

$$\frac{\partial\,|\psi|^2 E_l}{\partial\alpha\partial\beta} = |\psi|^2\frac{\partial^2 E_l}{\partial\alpha\partial\beta} + 2E_l\Re\left(\psi^*\frac{\partial^2\psi}{\partial\alpha\partial\beta}\right) + 2\Re\left(\frac{\partial\psi^*}{\partial\alpha}\frac{\partial\psi}{\partial\beta}\right)E_l + 2\psi\frac{\partial E_l}{\partial\alpha}\Re\left(\frac{\partial\psi}{\partial\beta}\right) + 2\psi\frac{\partial E_l}{\partial\beta}\Re\left(\frac{\partial\psi}{\partial\alpha}\right)\,. \tag{8.7}$$

Le second, dernier, et quatrième termes se déduisent aisément de ceux qu'on a déjà trouvés pour la dérivée première, et le troisième terme revient à calculer $\partial^2(\psi^*\psi)/\partial\alpha\partial\beta$, ce qui nous donne :

$$\frac{\partial\,|\psi|^2}{\partial\alpha\partial\beta} = 2\Re\left(\psi^*\frac{\partial^2\psi}{\partial\alpha\partial\beta}\right) + 2\Re\left(\frac{\partial\psi^*}{\partial\alpha}\frac{\partial\psi}{\partial\beta}\right)\,. \tag{8.8}$$

En se ramenant aux espérances mathématiques, on arrive au résultat final suivant :

$$\frac{\partial^2 E_v}{\partial\alpha\partial\beta} = 2\gamma_3(E_l, \frac{\partial\ln\psi}{\partial\alpha}, \frac{\partial\ln\psi}{\partial\beta}) + 2\mathrm{Cov}\left(E_l, \frac{\partial^2\ln\psi}{\partial\alpha\partial\beta}\right) + \boldsymbol{E}\left(\frac{\partial E_l}{\partial\alpha}\frac{\partial\ln\psi}{\partial\beta}\right) + \boldsymbol{E}\left(\frac{\partial E_l}{\partial\beta}\frac{\partial\ln\psi}{\partial\alpha}\right)\,. \tag{8.9}$$

Cela nous permet de remarquer, en se servant des expressions des variances définies dans l'annexe B, que la variance d'une dérivée première se comporte le plus souvent en $\mathcal{O}\left(N^2\right)$ et d'une dérivée seconde en $\mathcal{O}\left(N^3\right)$, ce qui rend une optimisation de fonction d'onde par une méthode d'optimisation de type Newton assez chère.

On remarque également qu'on a besoin de trois familles de quantités pour pouvoir obtenir ces dérivées premières et secondes : la famille des dérivées premières logarithmiques de la fonction d'onde, $(\partial\ln\psi/\partial\alpha)_\alpha$) ; la famille des dérivées secondes logarithmiques de la fonction d'onde, $(\partial^2\ln\psi/\partial\alpha\partial\beta)_{\alpha\beta}$) ; et la famille des dérivées de l'énergie locale, $(\partial E_l/\partial\alpha)_\alpha$.

## 8.3 Calcul des termes des dérivées

À la fin de la section précédente, nous avons retiré trois familles de quantités nécessaires au calcul des dérivées de l'énergie variationnelle. Supposons maintenant que l'on travaille sur le modèle de Hubbard, avec les notations et formalismes introduits au chapitre 4 (rappel en annexe A), avec une fonction d'onde d'essai de type Jastrow-Slater définie de la manière suivante :

$$\psi_{JS}(\mathcal{R}) = \det(\boldsymbol{C}\boldsymbol{X}(\mathcal{R}))e^{J(\mathcal{R})}\,; \tag{8.10}$$

avec, on le rappelle, $J(\mathcal{R}) = {}^t\boldsymbol{U}^t\boldsymbol{X}(\mathcal{R})\boldsymbol{J}\boldsymbol{X}(\mathcal{R})\boldsymbol{U}$. On a alors deux familles de paramètres à optimiser : les $c_{ik}$, éléments de la matrice LCAO $\boldsymbol{C}$, et les $j_i$, élements de la matrice de Jastrow $\boldsymbol{J}$.

Si on cherche à calculer les dérivées premières logarithmiques de la fonction d'onde, on trouve :

$$\frac{\partial \ln \psi_{JS}}{\partial j_i}(\mathcal{R}) = \frac{\partial \ln \exp J}{\partial j_i}(\mathcal{R}) = {}^t\boldsymbol{U}^t\boldsymbol{X}(\mathcal{R})\frac{\partial \boldsymbol{J}}{\partial j_i}\boldsymbol{X}(\mathcal{R})\boldsymbol{U} \; ; \tag{8.11a}$$

$$\begin{aligned}\frac{\partial \ln \psi_{JS}}{\partial c_{ik}}(\mathcal{R}) &= \frac{\partial \ln \det(\boldsymbol{C}\boldsymbol{X})}{\partial c_{ik}}(\mathcal{R}) = \left(\boldsymbol{X}(\mathcal{R})\boldsymbol{A}^{-1}(\mathcal{R})\right)_{ki} \\ &= \mathrm{Tr}\left(\boldsymbol{E}_{ik}\boldsymbol{X}\boldsymbol{A}^{-1}\right)(\mathcal{R}) \; ;\end{aligned} \tag{8.11b}$$

où $\boldsymbol{E}_{ik}$ est la matrice de la base classique telle que $(\boldsymbol{E}_{ik})_{jl} = \delta_{ij}\delta_{kl}$.

On voit ainsi que ces deux familles de paramètres avec lesquelles on travaille mettent en jeu deux structures de données complètement différentes. D'une part, pour les éléments de la matrice LCAO, on trouve la matrice $\boldsymbol{X}\boldsymbol{A}^{-1}$, qui peut être comprise comme étant la dérivée logarithmique de $\det \boldsymbol{A}$ par rapport à $\boldsymbol{C}$ comme $\boldsymbol{D}$ l'est par rapport à $\boldsymbol{X}$ ; et d'autre part, chacun des termes entrant en jeu dans le calcul du facteur de Jastrow. En effet, on a, dans notre cas précis :

$$J(\mathcal{R}) = \sum_i j_i \frac{\partial J}{\partial j_i}(\mathcal{R}) = \sum_i j_i \frac{\partial \ln \psi}{\partial j_i}(\mathcal{R}) \; . \tag{8.12}$$

Bien sûr, chacun de ces termes prend une forme similaire à l'énergie potentielle, consistant de sommes de produits de nombres d'occupations, invariant par permutation entre les axes de l'espace, et par réflection ou translation du système tout entier.

Passons maintenant aux dérivées logarithmiques secondes de la fonction d'onde. Il est évident que, par construction, si une des dérivations est par rapport à un des $j_i$, la dérivée seconde est nulle. Reste donc les dérivées secondes par rapport aux valeurs de la matrice $\boldsymbol{C}$. Pour cela, on trouve, en se servant de l'expression de la dérivée de $\boldsymbol{A}^{-1}$ :

$$\begin{aligned}\frac{\partial^2 \ln \psi_{JS}}{\partial c_{ik}\partial c_{i'k'}} &= -\mathrm{Tr}\left(\boldsymbol{E}_{ik}\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{E}_{i'k'}\boldsymbol{X}\boldsymbol{A}^{-1}\right) \\ &= -(\boldsymbol{X}\boldsymbol{A}^{-1})_{ki'}(\boldsymbol{X}\boldsymbol{A}^{-1})_{k'i} \\ &= -\frac{\partial \ln \psi_{JS}}{\partial c_{i'k}}\frac{\partial \ln \psi_{JS}}{\partial c_{ik'}} \; .\end{aligned} \tag{8.13}$$

Il s'agit donc de quelque chose qui est très pratique à calculer.

Il nous reste à calculer les termes de dérivées de l'énergie locale. Si l'énergie potentielle est bien évidemment indépendente de la valeur prise par la fonction d'onde, ce n'est pas le cas de l'énergie cinétique. Il nous suffit alors de dériver la somme des $z_{j\vec{\delta}}d_{j\vec{\delta}}$, ce qui nous donne :

$$\begin{aligned}\frac{\partial E_l}{\partial c_{ik}} &= \frac{\partial}{\partial c_{ik}}\sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}}d_{j\vec{\delta}} = \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}}\frac{\partial d_{j\vec{\delta}}}{\partial c_{ik}} \\ &= \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}}\frac{\partial}{\partial c_{ik}}\left[\mathrm{Tr}(\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{T}_{\vec{\delta}}\boldsymbol{X}\boldsymbol{E}_{jj})\right] \\ &= \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}}\left[\mathrm{Tr}(\boldsymbol{A}^{-1}\boldsymbol{E}_{ik}\boldsymbol{T}_{\vec{\delta}}\boldsymbol{X}\boldsymbol{E}_{jj}) - \mathrm{Tr}(\boldsymbol{A}^{-1}\boldsymbol{E}_{ik}\boldsymbol{X}\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{T}_{\vec{\delta}}\boldsymbol{X}\boldsymbol{E}_{jj})\right] \\ &= \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}}(a^{-1})_{ji}((\boldsymbol{I}_\omega - \boldsymbol{X}\boldsymbol{D})\boldsymbol{T}_{\vec{\delta}}\boldsymbol{X})_{kj} \; ;\end{aligned} \tag{8.14a}$$

$$\begin{aligned}\frac{\partial E_l}{\partial j_i} &= \frac{\partial}{\partial j_i}\sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}} d_{j\vec{\delta}} = \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} \frac{\partial z_{j\vec{\delta}}}{\partial j_i} d_{j\vec{\delta}} \\ &= \sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}} d_{j\vec{\delta}} \frac{\partial 2\left((\boldsymbol{K})_{j,\vec{r}_j+\vec{\delta}} - (\boldsymbol{K})_{j,\vec{r}_j}\right)}{\partial j_i} \\ &= 2\sum_{j=1}^{N}\sum_{\vec{\delta}\in\omega_\delta} z_{j\vec{\delta}} d_{j\vec{\delta}} \sum_{\substack{k=1 \\ k\neq j}}^{N} \left[\left(\frac{\partial \boldsymbol{J}}{\partial j_i}\right)_{\vec{r}_k,\vec{r}_j+\vec{\delta}} - \left(\frac{\partial \boldsymbol{J}}{\partial j_i}\right)_{\vec{r}_k,\vec{r}_j}\right] . \end{aligned} \tag{8.14b}$$

Pour calculer les cumulants de ces quantités, afin d'arriver à la valeur et à la variance des dérivées respectives, on se réfèrera aux formules qui ont été développées dans l'annexe B.

## Récapitulatif

Nous avons vu dans ce chapitre comment les méthodes et formalismes que nous avions développés jusqu'ici pouvaient nous permettre de calculer les valeurs des dérivées d'observables. En particulier, nous avons développé les expressions pratiques que l'on pouvait employer pour les dérivées de l'énergie variationnelle par rapport aux paramètres d'une fonction d'onde de type Jastrow-Slater que nous avions mentionnée au chapitre 4.

L'étape suivante, d'un point de vue du développement en vue de l'implémentation, serait de transformer nos expressions globales en expressions restreintes afin de pouvoir calculer des différences exactes de ces dérivées dans nos sous-dynamiques, comme nous l'avons fait à la section 4.4.

L'application pratique à laquelle ce travail doit nous mener, cependant, est évidente. En effet, l'optimisation de fonction d'onde, que ce soit au moyen d'une méthode de Newton [3] ou de la méthode linéaire développée par Toulouse et Umrigar (voir [1], [2] et [4]), requiert le gradient de l'énergie locale, et dans le cas de la méthode de Newton, elle requiert également le calcul de la matrice Hessienne par rapport à chacun des paramètres que l'on cherche à optimiser. On peut par ailleurs se servir de la matrice Hessienne pour le calcul d'énergies d'excitation, en se servant de la théorie de la réponse linéaire [5].

### Bibliographie

[1] Julien Toulouse and C. J. Umrigar. Optimization of quantum Monte Carlo wave functions by energy minimization. *Journal of Chemical Physics*, 126(8) :084102, 2007.

[2] Julien Toulouse and C. J. Umrigar. Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. *Journal of Chemical Physics*, 128 :174101, 2008.

[3] C. J. Umrigar and Claudia Filippi. Energy and variance optimization of many-body wave functions. *Physical Review Letters*, 94(15) :150201, Apr 2005.

[4] C. J. Umrigar, Julien Toulouse, Claudia Filippi, S. Sorella, and R. G. Hennig. Alleviation of the fermion-sign problem by optimization of many-body wave functions. *Physical Review Letters*, 98 :110201, Mar 2007.

[5] Bastien Mussard, Emanuele Coccia, Roland Assaraf, Matthew Otten, Cyrus J. Umrigar, and Julien Toulouse. Chapter fourteen - time-dependent linear-response variational monte carlo. In Philip E. Hoggan, editor, *Novel Electronic Structure Theory : General Innovations and Strongly Correlated Systems*, volume 76 of *Advances in Quantum Chemistry*, pages 255 – 270. Academic Press, 2018.

# Conclusion

Au cours de cette thèse, nous avons présenté une nouvelle méthode de Monte Carlo quantique, dérivée de la méthode de Monte Carlo Variationnelle (VMC). La méthode de Monte Carlo Partitionnelle (PMC) exploite une partition du système en une collection de fragments, pour réaliser à chaque point de la dynamique Metropolis principale des sous-dynamiques latérales à faible coût sur chacun des fragments. Les sous- dynamiques sont ensuite réintégrées au calcul au moyen d'un estimateur amélioré de variance nulle lorsque les sous-dynamiques sont de longueur infinie et les fragments indépendents entre eux. Cet estimateur nous permet donc de réduire de manière très significative la variance, en se servant du principe de « diviser pour régner ». La méthode PMC met ainsi en jeu un compromis entre augmentation du temps de calcul et réduction de la variance. Nous avons pu montrer (à la section 4.5) que pour de grands systèmes isolants à $N$ électrons, on pouvait s'attendre à un gain allant jusqu'à $\mathcal{O}\left(N^2\right)$. Dans des conditions significativement plus défavorables – des systèmes métalliques, à longueur de corrélation infinie – nous avons pu obtenir un gain en $\mathcal{O}\left(N\right)$.

De plus, nous avons également construit des estimateurs non biaisés dans la limite d'échantillonnages et sous-échantillonnages infinis, et zéro-variants dans la limite de fragments indépendents avec des sous-échantillonnages infinis, de cumulants, afin de pouvoir mettre en œuvre le calcul de dérivées et dérivées secondes de l'énergie. Cela nous ouvre la porte pour étendre les gains de la méthode PMC à l'optimisation de la fonction d'onde d'essai, ainsi qu'aux calculs de diverses propriétés.

On peut imaginer un champ d'applications varié à la méthode PMC. Si utiliser notre méthode sur des systèmes discrets, comme le modèle de Hubbard, est facile grâce aux symmétries sous-jacentes, les travaux présentés (et développés davantage) dans l'article fourni à l'annexe D appliquent la méthode PMC à une partition cœur-valence dans des systèmes atomiques et moléculaires. Qui plus est, un raisonnement en fragments permet aussi de s'attaquer à une variété de systèmes chimiques. En effet, dans des molécules de grande taille, l'interaction entre des parties distantes de cette molécule est généralement très faible. On peut ainsi imaginer un découpage arbitraire de l'espace dans lequel cette molécule est contenue, ou un découpage en fragments lié à la géométrie de la molécule. Enfin, on peut envisager une application aux cristaux. En effet, lorsqu'on emploie une méthode traitant de manière explicite les interactions coulombiennes, on se restreint en général à travailler sur une maille de simulation, munie de conditions périodiques aux limites ; et reproduire le comportement du cristal requiert de faire croître cette maille de simulation pour extrapoler un comportement à l'infini. Ces méthodes sont alors généralement mal adaptées à cause du fort coût de l'augmentation en taille. Cependant, la structure périodique des cristaux nous permet d'envisager de construire aisément des fragments composés d'une ou plusieurs mailles élémentaires, nous permettant d'envisager l'emploi de la méthode PMC afin de réduire le coût de l'augmentation en taille du système.

Bien entendu, cela ne fait pas de la méthode PMC une panacée d'office. En effet, la méthode PMC est dérivée de la VMC, qui est la plus simple des méthodes de Monte Carlo quantique ; et elle tire de cette méthode le besoin absolu de la connaissance de la densité de probabilité sur laquelle on travaille. L'adaptation des principes sous-jacents à la méthode PMC et des idées qui ont mené à son élaboration à la méthode de Monte Carlo Diffusionnelle (DMC) semble alors fort compromis. Par ailleurs, bien que nous ayons présenté une méthode théorique pour étendre la méthode PMC aux dérivées successives de l'énergie, l'application pratique de celle-ci reste en projet. On peut au mieux s'attendre à une répercution du gain sur la variance de l'énergie à la variance des dérivées de celle-ci. Or, dans le meilleur des

cas, on gagnerait $\mathcal{O}\left(N^2\right)$... là où la variance des dérivées secondes de l'énergie se comporte en $\mathcal{O}\left(N^3\right)$. On ne peut donc pas voir en la méthode PMC une résolution complète du problème de la variance.

Pour pallier à ces difficultés et développer plus avant la méthode PMC, nous avons plusieurs pistes. Si l'application récursive, multi-échelle de la méthode PMC présente un gain limité, il est également possible de réduire la variance davantage en envisageant un développement en fragments, inspiré de la formule de Poincaré en combinatoire (voir, par exemple, [1]). Qui plus est, bien que la méthode DMC semble significativement plus difficile d'accès, il n'est pas impensable de changer d'espace, et de chercher à travailler dans l'espace des trajectoires plutôt que celui des configurations. Ce changement d'espace est à l'origine de la méthode de Monte Carlo Reptationnelle (voir [2]), une cousine moins performante de la méthode DMC, à laquelle nous envisageons d'étendre les principes clé de la méthode PMC.

## Bibliographie

[1] F. Zahariev and M. S. Gordon. Combined quantum monte carlo – effective fragment molecular orbital method : fragmentation across covalent bonds. *Physical Chemistry Chemical Physics*, 23 :14308–14314, 2021.

[2] Stefano Baroni and Saverio Moroni. Reptation quantum monte carlo : A method for unbiased ground-state averages and imaginary-time correlations. *Physical Review Letters*, 82(24) :4745, 1999.

# Quatrième partie

# Annexes

# Annexe A

# Notations et définitions

## Notations générales classiques

- $\Omega$ est un domaine de $\mathbb{R}^d$. S'il est fini, il est d'hypervolume $V = |\Omega|$.
- $\mathcal{F}(\Omega, \mathbb{C}) \equiv \mathbb{C}^\Omega$ est l'ensemble des fonctions à valeurs complexes définies sur $\Omega$.
- $\mathcal{L}^1(\Omega, \mathbb{R})$ est l'ensemble des fonctions à valeurs réelles, intégrables sur $\Omega$. $\mathcal{L}^2(\Omega, \mathbb{C})$ est l'ensemble des fonctions à valeurs complexes dont le carré du module est intégrable.
- $\mathcal{L}_\mathbb{K}(E)$ est l'ensemble des applications linéaires (ou endomorphismes) de $E$.
- $\boldsymbol{E}$ est l'espérance mathématique, forme linéaire de l'ensemble des variables aléatoires. En particulier, si $\rho$ est une densité de probabilité, on a :

$$\boldsymbol{E}_\rho = X \mapsto \frac{\displaystyle\int_\Omega X\rho}{\displaystyle\int_\Omega \rho}\ .$$

## Chapitre 1

Les notations introduites au chapitre 1 ne sont pas conservées par la suite.

- $\varepsilon$ est une erreur ou un écart-type.
- $f \in \mathcal{L}^1(\Omega, \mathbb{R})$ est une fonction intégrable sur $\Omega$, dont on cherche à calculer l'intégrale $I = \displaystyle\int_\Omega f$.
- $\mathcal{R}$ est un élément de $\Omega$. La suite $(\mathcal{R}_m)_{m\in\mathbb{N}^*} \in \Omega^{\mathbb{N}^*}$ est une suite de variables aléatoires de $\Omega$. Dans la section 1.2, les variables sont identiques et indépendantes. Dans la section 1.3 elles sont générées de manière stochastique.
- $\bar{f}$ est la valeur moyenne de $f$ et vaut $\frac{I}{V}$. $\overline{f^2}$ est de même la moyenne de $f^2$ si $f$ est de carré intégrable.
- $\rho \in \mathcal{L}^1(\Omega)$ est une densité de probabilité.
- $\xi$ est une variable aléatoire générée à partir de la distribution uniforme de $]0, 1[$.
- $\boldsymbol{T}$ est une matrice de transition, de terme général $\boldsymbol{T}(\mathcal{R}|\mathcal{R}')$. $\boldsymbol{P}$ est une matrice de proposition, et $\boldsymbol{A}$ une matrice d'acceptation.

## Chapitre 2

Les notations introduites au chapitre 2 sont :

- $\mathcal{S}$ est un système physique à $N$ particules dans un domaine $\omega$ de $\mathbb{R}^3$.
- Les particules ont pour position $\vec{r}_1, \ldots, \vec{r}_N$ et spins $\sigma_1, \ldots, \sigma_n$.
- L'espace configurationnel est $\Omega = \omega^N$. Les configurations sont les $\mathcal{R} = (\vec{r}_i)_{i\in[[1,N]]} \in \Omega$.
- $\hat{H}$ est l'opérateur Hamiltonien ; $\hat{T}$ l'opérateur d'énergie cinétique et $V$ l'énergie potentielle.
- $\phi$ est une fonction d'onde dépendente du temps ; $\psi \in \mathcal{L}^2(\Omega, \mathbb{C})$ est une fonction d'onde statique.
- $\chi$ est une orbitale moléculaire, fonction d'onde monoélectronique. $\varphi$ est une orbitale atomique, fonction de base monoatomique.

- $E_v$ est la fonctionnelle énergie variationnelle.
- $\psi_i$ est l'état propre du système d'énergie $E_i$ ; le fondamental correspond à $i = 0$.
- $\omega_v$ est l'espace variationnel monoélectronique, et $\Omega_v$ l'espace variationnel polyélectronique.
- $\rho = |\psi|^2$ est la densité de probabilité, $\rho_e$ est la densité électronique.
- $E_l = (\hat{H}\psi)/\psi$ est l'énergie locale.
- $J(\vec{r}_1, \vec{r}_2)$ est une fonction symétrique, terme du facteur de Jastrow.

## Chapitre 3

Les notations introduites au chapitre 3 sont :
- $\mathcal{S} = (\omega, [[1, N]])$ est un système, composé d'un espace physique $\omega$ (par exemple $\{\uparrow, \downarrow\} \times (\mathbb{Z}/(L\mathbb{Z}))^2$) et d'une liste d'indices de particules.
- $\mathcal{S}_i = (\omega_i, J_i)$ est un sous-système de $\mathcal{S}$ si $\omega_i \subset \omega$ et $J_i \subset [[1, N]]$.
- Si $\mathcal{S}_i$ et $\mathcal{S}_j$ sont deux sous-systèmes, alors $\mathcal{S}_i \cup \mathcal{S}_j = (\omega_i \cup \omega_j, J_i \cup J_j)$ et $\mathcal{S}_i \cap \mathcal{S}_j = (\omega_i \cap \omega_j, J_i \cap J_j)$. On notera $\emptyset$ le sous- système $(\emptyset, \emptyset)$.
- $(\mathcal{S}_i)_{i\in[[1,p]]} \in \mathcal{P}(\mathcal{S})$ est une partition de $\mathcal{S} \iff \bigcup_{i=1}^{p} \mathcal{S}_i = \mathcal{S}$ et $i \neq j \iff \mathcal{S}_i \cap \mathcal{S}_j = \emptyset$.
- $\mathcal{S}_{\bar{\imath}} = \bigcup_{i\neq j} \mathcal{S}_j = \mathcal{S}/\mathcal{S}_i$ est l'environnement de $\mathcal{S}_i$.
- $\mathcal{R}_i$ est la sous-famille de $\mathcal{R}$ ne comportant que les positions de $\mathcal{S}_i$.
- $\Omega_{\bar{\imath}}(\mathcal{R}) = \{\mathcal{R}' \in \Omega/\mathcal{R}_{\bar{\imath}} = \mathcal{R}'_{\bar{\imath}}\}$, $\mathcal{R}_i|\mathcal{R}$ la complétion de $\mathcal{R}_i$ en un élement de $\Omega_{\bar{\imath}}(\mathcal{R})$.
- Les partitions hiérarchisées sont une famille de partitions, telle que tout élément d'une partition antérieure admet une partition en élements de la partition ultérieure.
- Si $X$ est une variable aléatoire extensive et $(\mathcal{S}_i)$ est une partition en sous-systèmes indépendents, $X = \sum_{i=1}^{p} X_i$.
- $1_A$ est la fonction caractéristique du sous-ensemble $A$ et vaut 1 pour un élément de $A$ et 0 sinon.
- $\boldsymbol{E}(X|A) = \boldsymbol{E}_{\rho 1_A}(X)$ est l'espérance conditionnelle de $X$ sachant $A$.
- $\boldsymbol{E}(X|\bar{\imath}) = \mathcal{R} \mapsto \boldsymbol{E}_{\rho 1_{\Omega_{\bar{\imath}}(\mathcal{R})}}(X)$ est l'espérance conditionnelle de $X$ à environnement $\mathcal{S}_{\bar{\imath}}$ figé.
- $\varepsilon_i = X \mapsto \boldsymbol{E}(X|\bar{\imath})$ est l'endomorphisme projecteur qui moyenne sur $\mathcal{S}_i$. On a $\varepsilon_i\varepsilon_j = \varepsilon_{i\cup j}$.
- En assimilant les variables aléatoires constantes à $\mathbb{C}$, on a $\varepsilon_i\varepsilon_{\bar{\imath}} = \boldsymbol{E}_\rho$.
- $\Delta_i = \mathrm{Id}_{\mathcal{L}_{\mathbb{C}}(\mathbb{C}^\Omega)} - \varepsilon_i$ est le projecteur complémentaire de $\varepsilon_i$, et ne donne que des variables de contrôle.
- $V_i = \mathrm{Var}(X|\bar{\imath}) = \varepsilon_i(X^2) - (\varepsilon_i X)^2$ est la variance conditionnelle de $X$ à environnement $\mathcal{S}_{\bar{\imath}}$ figé.
- $C_i = \mathrm{Cov}(X, Y|\bar{\imath}) = \varepsilon_i(XY) - \varepsilon_i X \varepsilon_i Y$ est la covariance conditionnelle de $X$ et $Y$ à environnement $\mathcal{S}_{\bar{\imath}}$ figé.
- $\mathcal{R}^K$ est la $K$-ième configuration de la dynamique principale, qui a pour longueur $M$.
- $\mathcal{R}_i^{Kk}$ est la $k$-ième configuration de la sous-dynamique liée au sous-système $\mathcal{S}_i$ et qui a pour environnement $\mathcal{R}_{\bar{\imath}}^K$, l'environnement du sous-système $\mathcal{S}_i$ dans $\mathcal{R}^K$. La sous- dynamique a pour longueur $m$.
- $\tilde{X}_{\mathrm{th}}$ est la variable corrigée théorique $X - \sum_i \Delta_i X$ ; $\tilde{X}_{\mathrm{pr}}$ est son équivalent pratique. L'estimateur PMC est la moyenne de la variable corrigée pratique.
- $V_{ij} = \mathrm{Var}(X|\overline{i \cup j})$.
- $\bar{X}_i$ est une moyenne sur une sous-dynamique sur le secteur $\mathcal{S}_i$.
- $\tau_i$ est le temps d'autocorrélation sur une sous-dynamique sur le secteur $i$. Il dépend de l'environnement.

## Chapitre 4

Les notations introduites au chapitre 4 sont :
- $n$ est le nombre d'électrons dans un sous-système. Il est bien évidemment variable, mais est de l'ordre de $N/p$.
- $\mathfrak{K}$, $\mathfrak{m}$ et $\mathfrak{n}$ sont les équivalents respectifs de $k$, $m$ et $n$ pour une sous-dynamique sur une seconde partition.

- $\mathcal{R}_{ij}^{Kk\mathfrak{K}}$ est l'équivalent de $\mathcal{R}_i^{Kk}$ pour la sous-dynamique sur la seconde partition, pour le sous-système $\mathcal{S}_j$ du sous-système $\mathcal{S}_i$.
- $\psi_S$ est une fonction d'onde de type déterminant de Slater.
- $\chi$ et $\varphi$ sont respectivement les orbitales moléculaires et atomiques introduites au chapitre 2 (voir A).
- $\boldsymbol{A} = (\chi_i(\vec{r}_j))_{ij}$ est la matrice de Slater.
- $\boldsymbol{C} = (c_{ik})_{ik}$ est la matrice LCAO.
- $\boldsymbol{X} = (\varphi_k(\vec{r}_j))_{kj}$ est la matrice des orbitales atomiques.
- $\boldsymbol{D} = (d_{ij})_{ij} = \boldsymbol{A}^{-1}\boldsymbol{C}$ est la matrice dérivée, et on a $\boldsymbol{D}(\mathcal{R})\boldsymbol{X}(\mathcal{R}) = \boldsymbol{I}_N$.
- $\boldsymbol{P}_i$ et $\boldsymbol{Q}_i$ sont deux matrices rectangulaires de projection sur les aspects respectivement électroniques et géographiques du secteur $\mathcal{S}_i$.
- $\boldsymbol{T} = (t_{\vec{\imath}\vec{\jmath}})_{\vec{\imath}\vec{\jmath}}$ est la matrice d'adjacence d'un système discret. Elle est remplacée dans le cas continu par l'opérateur laplacien.
- $\hat{a}_{\vec{\imath}\sigma}$ est l'opérateur annihilation lié au spinsite $\vec{\imath}\sigma$. $\hat{a}^\dagger_{\vec{\imath}\sigma}$ est l'opérateur création qui y est lié.
  On a $\hat{a}_{\vec{\imath}\sigma}\hat{a}^\dagger_{\vec{\imath}'\sigma'} + \hat{a}^\dagger_{\vec{\imath}'\sigma'}\hat{a}_{\vec{\imath}\sigma} = \delta_{ii'}\delta\sigma\sigma'$.
- $\hat{n}_{\vec{\imath}\sigma} = \hat{a}^\dagger_{\vec{\imath}\sigma}\hat{a}_{\vec{\imath}\sigma}$ est l'opérateur nombre d'occupation du spinsite $\vec{\imath}\sigma$.
- $L$ est la taille de la grille du modèle de Hubbard. Les sous-systèmes de la première partition ont pour taille $l$, ceux de la seconde partition $\mathfrak{l}$.
- $\vec{k} = (k_x, k_y)$ est un vecteur d'onde dont $k_x$ et $k_y$ sont les nombres d'onde.
- $J(\mathcal{R})$ est l'exposant du facteur de Jastrow. $\boldsymbol{J}$ est la matrice correspondante.
- $\omega_d$ est l'ensemble des translations monoélectroniques correspondant à un saut d'un site à un autre adjacent.

# Chapitre 5

Les notations introduites au chapitre 5 sont :
- $G_v$ est le gain en variance, la variance VMC divisée par la variance PMC.
- $G_t$ est l'augmentation du temps de calcul. Le temps CPU PMC divisé par le temps CPU VMC.
- Le coût calculatoire est défini par le temps de calcul, multiplié par le carré de l'erreur statistique.
- $G_r$ est le gain réel. Le coût calculatoire VMC divisé par le coût calculatoire PMC. Assimilé au ratio $G_v/G_t$ car on travaille avec un temps de corrélation de 1.
- RSI est la stratégie qui consiste à prendre la racine de la taille du secteur de la $k$-ième partition pour avoir la taille du secteur de la $(k+1)$-ième partition.
- GS est la stratégie qui consiste à répartir géométriquement les tailles des secteurs des partitions entre 1 et $L$.
- RSS est la stratégie qui consiste à prendre la racine de $L$ fois la taille du secteur de la $k$-ième partition pour avoir la taille du secteur de la $(k+1)$-ième partition.

# Chapitres 6 et 7

Les notations employées au chapitres 6 et 7 sont :
- $X$ et $Y$ sont deux variables aléatoires correspondant à des observables extensives, que l'on peut décomposer sur les sous-systèmes.
- $\gamma_q$ est le cumulant d'ordre $q$, une forme $q$-linéaire extensive.
- $G_m$ est la fonction génératrice des moments, ou fonction caractéristique ; $G_c$ est la fonction génératrice des cumulants, ou deuxième fonction caractéristique.

# Annexe B

# Cumulants

Dans cette annexe, nous allons dans un premier temps introduire les notions de moments et cumulants desquelles on se sert aux chapitres 6 et 7, les définir à partir de leur fonction génératrice (et inversement), pour finir par retrouver le théorème central limite pour une variable aléatoire $X$ dont les moments sont tous définis et finis.

Dans un second temps, nous nous intéresserons aux estimateurs standard de la variance et de la covariance. En particulier, nous évaluerons leur biais et leur variance. Nous finirons par étendre les résultats obtenus pour la covariance d'estimateurs standard de la covariance de quatre variables aléatoires différentes.

La troisième section de cette annexe s'intéressera de manière similaire aux cumulants ternaires et aux covariances généralisées ternaires, afin d'évaluer leurs biais, variance et covariances.

La quatrième section s'intéresse aux cumulants d'ordre 4 à 6, afin de proposer des méthodes d'obtention d'estimateurs "standard" non biaisés de ceux-ci à l'aide des expressions et méthodes déployées aux sections précédentes.

Enfin, nous finirons en exploitant les méthodes et raisonnements présentés aux parties précédentes pour proposer des méthodes pratiques d'obtention de variances d'estimateurs composites.

## B.1 Moments et cumulants

### B.1.1 Moments et fonction génératrice des moments

Soit $\Omega$ un univers d'événements probabilistes $\mathcal{R}$ muni d'une loi de probabilité $\rho$, et $X \in \mathbb{C}^{\Omega}$ une variable aléatoire complexe. Alors, pour tout $n$ entier strictement positif, on définit le moment d'ordre $n$ de $X$, par, si celle-ci existe, l'espérance mathématique de $X^n$ :

$$\forall n \in \mathbb{N}^*,\ m_n(X) = \int_{\Omega} X(\mathcal{R})^n \rho(\mathcal{R}) d\mathcal{R} = \boldsymbol{E}_{\rho}(X^n)\ . \tag{B.1}$$

Si on suppose que ces moments existent pour tout $n$, alors on peut les interpréter comme dérivant d'une fonction génératrice complexe $G_m$. Cette fonction serait définie sur un voisinage de 0 dans $\mathbb{C}$ et holomorphe sur son espace de définition, de manière à ce que les moments soient les dérivées successives de $G_m$ en 0. On peut alors la reconstruire par recomposition de la série de Taylor de $G_m$ et en se servant de la propriété de forme linéaire de l'espérance mathématique :

$$\begin{aligned} G_m(X)(z) &= \sum_{n=0}^{\infty} \frac{m_n(X) z^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{\boldsymbol{E}\left(X^n\right) z^n}{n!} \\ &= \boldsymbol{E}\left( \sum_{n=0}^{\infty} \frac{X^n z^n}{n!} \right) \\ &= \boldsymbol{E}\left(e^{zX}\right)\ . \end{aligned} \tag{B.2}$$

$G_m$ est par construction entière et holomorphe sur son espace de définition. En particulier, si $X$ est bornée, ses moments sont majorés par une suite géométrique, donc $G_m$ est définie sur $\mathbb{C}$ tout entier. À cause de ces propriétés, $G_m$ est également connue sous le nom de fonction caractéristique de $X$.

Par ailleurs, supposons qu'on dispose de deux variables aléatoires $X$ et $Y$ dont les moments sont finis et définis. Alors, si $X$ et $Y$ sont indépendantes, alors on a pour tous entiers positifs $n$ et $m$, $\boldsymbol{E}\left(X^m Y^n\right) = \boldsymbol{E}\left(X^m\right)\boldsymbol{E}\left(Y^n\right)$, et donc en particulier $m_n(XY) = m_n(X)m_n(Y)$. Le moment est ainsi une propriété multiplicative pour des variables aléatoires indépendantes. Il s'agit d'une propriété intéressante, mais peu exploitable dans le cas général.

### B.1.2 Cumulants

Afin se rapprocher de la propriété d'extensivité, qui nous est plus utile, nous allons chercher à construire des quantités d'intérêt qui soient non multiplicatives mais bien additives pour des variables aléatoires indépendantes. Pour ce faire, nous alons chercher à commencer par développer la série de Taylor de $G_m(X+Y)$. Cela nous donne :

$$\begin{aligned}\forall n \in \mathbb{N},\ \left.\frac{d^n G_m(X+Y)}{dz^n}\right|_0 &= \boldsymbol{E}\left((X+Y)^n\right) \\ &= \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \boldsymbol{E}\left(X^i\right)\boldsymbol{E}\left(Y^{n-i}\right) \\ &= \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \left.\frac{d^i G_m(X)}{dz^i}\right|_0 \left.\frac{d^{n-i} G_m(Y)}{dz^{n-i}}\right|_0 .\end{aligned} \tag{B.3}$$

On voit que ce terme de la série de Taylor s'écrit comme une formule de Leibnitz de la dérivée $n$-ième d'un produit au point $z = 0$. Comme $G_m$ est une fonction entière, on peut généraliser cette expression en 0 pour tout point. On voit alors que si $X$ et $Y$ sont indépendantes, la fonction caractéristique génératrice des moments est multiplicative, $G_m(X+Y) = G_m(X)G_m(Y)$. Cela ressemble à l'exponentielle d'une fonction qui serait elle additive. On peut alors définir la fonction génératrice $G_c = \ln G_m$, qui, elle, est additive. $G_c$ est également connue sous le nom de deuxième fonction caractéristique de $X$.

Cela nous permet alors de définir, pour tout entier naturel non nul $n$, le cumulant symétrique d'ordre $n$, $\gamma_n$, comme la valeur de la dérivée $n$-ième de $G_c$ en 0 :

$$\forall n \in \mathbb{N}^*,\ \gamma_n(X) = \left.\frac{d^n G_c(X)}{dz^n}\right|_0 . \tag{B.4}$$

Le développement limité aux deux premiers ordres nous permet de remarquer qu'on connaît déjà les deux premiers cumulants, qui sont respectivement l'espérance mathématique (cumulant d'ordre 1) et la variance (cumulant d'ordre 2). Dans ce conditions, il n'existe que deux familles de distributions n'ayant qu'un nombre fini de cumulants non nuls : les distributions de Dirac, qui définissent des variables aléatoires constantes, décrites par leur seule valeur moyenne ; et les distributions normales, qui définissent des variables aléatoires décrites par valeur moyenne et variance.

Une propriété intéressante des cumulants est leur homogénéité. En effet, si on s'intéresse à la variable aléatoire $\alpha X$, avec $\alpha \in \mathbb{C}$ un complexe quelconque, alors dans un voisinage de 0, on peut écrire la fonction génératrice des moments de $\alpha X$ comme :

$$G_m(\alpha X)(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \boldsymbol{E}\left((\alpha X)^n\right) = \sum_{n=0}^{\infty} \frac{(\alpha z)^n}{n!} \boldsymbol{E}\left(X^n\right) = G_m(X)(\alpha z) . \tag{B.5}$$

On a donc $G_c(\alpha X)(z) = G_c(X)(\alpha z)$. Cela nous permet d'en déduire l'homogénéité :

$$\forall n \in \mathbb{N}^*,\ \gamma_n(\alpha X) = \left.\frac{d^t G_c(X)(\alpha z)}{dz^n}\right|_0 = \alpha^n \left.\frac{d^t G_c(X)}{dz^n}\right|_0 = \alpha^n \gamma_n(X) . \tag{B.6}$$

De plus, on peut remarquer que, les cumulants d'ordre $n > 1$ sont tous invariants par translation. En effet, une translation sur la variable aléatoire $X$ revient à y ajouter une variable aléatoire constante $X_0$. Comme $X_0$ est constante, elle est indépendente de toute autre variable aléatoire, et les cumulants de $X + X_0$ sont donc obtenus en ajoutant les cumulants deux à deux. $X_0$ ayant pour seul cumulant non nul le premier, de valeur $X_0$, les autres cumulants sont alors invariants par translation.

### B.1.3 Moyennage et convergence

Définissons pour commencer une suite de répliques identiques et indépendantes de notre variable aléatoire $(X_i)_{i\in\mathbb{N}^*}$, et la fonction moyenne partielle $\bar{X}$ qui à $m$ associe la moyenne des $m$ premières valeurs de $X$ :

$$\bar{X} : m \in \mathbb{N}^* \mapsto \frac{1}{m}\sum_{i=1}^{m} X_i . \tag{B.7}$$

Alors par identicité, indépendance des $X_i$, et homogénéité des cumulants, on obtient la propriété suivante :

$$\begin{aligned}\forall (n,m) \in (\mathbb{N}^*)^2, \gamma_n(\bar{X}(m)) &= \gamma_n\left(\frac{1}{m}\sum_{i=1}^{m} X_i\right) \\ &= \frac{1}{m^n}\sum_{i=1}^{m}\gamma_n(X_i) = m^{1-n}\gamma_n(X)\ .\end{aligned} \tag{B.8}$$

À partir de cette expression, on peut retrouver en deux lignes le théorème central limite. Il faut cependant remarquer que le fait que la fonction caractéristique soit définie sur un voisinage de zéro est une hypothèse très forte, vu qu'elle requiert que tous les moments soit finis.

### B.1.4 Écriture des moments en fonction des cumulants

Si l'on souhaite se déplacer entre les moments et les cumulants, on doit de manière générale se servir d'équivalences obtenues au moyen d'identifications de séries de Taylor. Le calcul des cumulants à partir des moments est malaisé, car on doit développer le logarithme d'une somme. À contrario, obtenir les moments à partir des cumulants est bien plus aisé, car il suffit de développer en série de Taylor $G_c$ dans un voisinage de 0 puis de développer l'exponentielle de ce développement.

Ce développement nous donne l'équation (7.30), qui est démontrée à l'annexe C, et que nous reproduisons ci-dessous :

$$\forall q \in \mathbb{N}^*,\ \boldsymbol{E}\left(X^q\right) = \sum_{x\in I_q} \frac{q!}{\prod_{k=1}^{\infty}\left[(k!)^{x_k}x_k!\right]}\prod_{k=1}^{\infty}\gamma_k(X)^{x_k}\ ; \tag{B.9}$$

où $I_q$ est l'ensemble des suites presque nulles de $\mathbb{N}^{\mathbb{N}^*}$ de somme $q$.

On peut interpréter cette équation en se servant d'un ensemble de $q$ particules indiscernables. En effet, dans notre somme, l'indice de sommation correspond à toutes les partitions possibles de notre ensemble en sous-ensembles ; le préfacteur au nombre de manières différentes de les remplir, et le produit des cumulants correspond à chacun des ensembles des partitions. L'aspect de dénombrement dans les expressions des cumulants est alors évident.

Nous avons ci-dessous développé pour la variable aléatoire $X$, d'espérance mathématique $\mu$ et de variance $\sigma^2$, et dont les moments sont tous finis, l'équation (B.9) pour des valeurs de $q$ allant de 1 à 6 :

$$\boldsymbol{E}\left(X\right) = \gamma_1 \equiv \mu\ ; \tag{B.10a}$$
$$\boldsymbol{E}\left(X^2\right) = \mu^2 + \gamma_2 \equiv \mu^2 + \sigma^2\ ; \tag{B.10b}$$
$$\boldsymbol{E}\left(X^3\right) = \mu^3 + 3\mu\sigma^2 + \gamma_3\ ; \tag{B.10c}$$
$$\boldsymbol{E}\left(X^4\right) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 + 4\mu\gamma_3 + \gamma_4\ ; \tag{B.10d}$$
$$\boldsymbol{E}\left(X^5\right) = \mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4 + 10\mu^2\gamma_3 + 10\sigma^2\gamma_3 + 5\mu\gamma_4 + \gamma_5\ ; \tag{B.10e}$$
$$\boldsymbol{E}\left(X^6\right) = \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6 + 20\mu^3\gamma_3 + 60\mu\sigma^2\gamma_3 + 10\gamma_3^2 + 15\mu^2\gamma_4 + 15\sigma^2\gamma_4 + 6\mu\gamma_5 + \gamma_6\ . \tag{B.10f}$$

### B.1.5 Cumulants asymétriques et covariances généralisées

Le plus souvent, on cherche à se servir d'expressions non symétriques des cumulants pour caractériser les interactions entre variables non indépendantes. L'exemple typique est celui de la covariance, qui est le cumulant asymétrique d'ordre 2. On peut ainsi voir le concept de cumulants non symétriques comme des covariances généralisées (ce qu'ils sont généralement appelés).

Dans ce cas, si on dispose de $n$ variables aléatoires $X_1, X_2, \ldots, X_n$, on peut écrire le cumulant d'ordre $n$ comme :

$$\gamma_n(X_1,\ldots,X_n) = \left.\frac{\partial^n \ln\left(\boldsymbol{E}\left(e^{\sum_{i=1}^{n} z_i X_i}\right)\right)}{\partial z_1 \partial z_2 \ldots \partial z_n}\right|_0\ . \tag{B.11}$$

Cette définition nous permet de définir un moment à plusieurs variables... dont on connait déjà l'expression :

$$\boldsymbol{E}\left(\prod_{i=1}^{n} X_i\right) = m_n(X_1,\ldots,X_n) = \left.\frac{\partial^n \boldsymbol{E}\left(e^{\sum_{i=1}^{n} z_i X_i}\right)}{\partial z_1 \partial z_2 \ldots \partial z_n}\right|_0 . \tag{B.12}$$

On peut alors adapter la démonstration de l'équation (B.9) pour cette expression. Le résultat qu'on obtiendra ne sera pas développé ci-dessous, mais se comprend aisément en reprennant l'interprétation que l'on a donnée pour cette équation, mais en remplaçant l'ensemble de $q$ particules indiscernables, liées à $q$ variables aléatoires toutes égales à $X$, et donc indiscernables, par un ensemble de $n$ particules discernables liées aux variables aléatoires $X_1,\ldots,X_n$.

De manière générale, on peut alors définir une procédure de désymétrisation pour les expressions des cumulants. Pour ce faire, il faut suivre trois contraintes :

- Maintien de la symétrie : les moments et cumulants sont invariants par permutation de variables aléatoires, donc il faut que l'expression désymétrisée le soit aussi (par exemple, $\gamma_3(X,Y,Z) = \gamma_3(X,Z,Y)$).
- Retour à l'expression symétrique : si on remplace toutes les variables par $X$ dans l'expression désymétrisée, on doit retrouver l'expression symétrique. ($\gamma_2(X,X) = \gamma_2(X)$, $\gamma_3(X,X,X) = \gamma_3(X)$, et ainsi de suite)
- Maintien du degré dans chacun des termes : dans chaque terme de degré $n$, il faut que chacune des variables aléatoires apparaisse exactement autant de fois qu'elle le fait dans les arguments du moment ou du cumulant que l'on évalue.

Ces trois contraintes suffisent à construire une expression désymétrisée unique du cumulant asymétrique.

## B.2 Estimateurs standard de cumulants d'ordre 2

Soient $X$ et $Y$ deux variables aléatoires dont les moments sont tous finis, d'espérances mathématiques respectives $\mu_x$ et $\mu_y$. On notera $\mathrm{Var}\,(X) = \sigma_x^2$, $\mathrm{Var}\,(Y) = \sigma_y^2$, et $\mathrm{Cov}\,(X,Y) = \sigma_x\sigma_y c_{xy}$ ; $c_{xy}$ est le coefficient de corrélation entre $X$ et $Y$.

L'expression du moment centré d'ordre 4 de $X$ en fonction des cumulants se tire facilement de celle des moments donnée à l'équation (B.10d) en fixant $\mu_x = 0$ :

$$\boldsymbol{E}\left((X-\mu_x)^4\right) = 3\sigma_x^4 + \gamma_4(X) . \tag{B.13}$$

Cette expression se désymétrise facilement (en introduisant deux variables aléatoires supplémentaires $Z$ et $T$) :

$$\boldsymbol{E}\left((X-\mu_x)(Y-\mu_y)(Z-\mu_z)(T-\mu_t)\right) = \sigma_x\sigma_y\sigma_z\sigma_t(c_{xy}c_{zt} + c_{xz}c_{yt} + c_{xt}c_{yz}) + \gamma_4(X,Y,Z,T) . \tag{B.14}$$

Cela nous permet de construire les expressions pour 3 $X$ et 1 $Y$, et pour 2 $X$ et 2 $Y$ :

$$\boldsymbol{E}\left((X-\mu_x)^3(Y-\mu_y)\right) = 3c_{xy}\sigma_x^3\sigma_y + \gamma_4(X,X,X,Y) ; \tag{B.15a}$$

$$\boldsymbol{E}\left((X-\mu_x)^2(Y-\mu_y)^2\right) = (1+2c_{xy}^2)\sigma_x^2\sigma_y^2 + \gamma_4(X,X,Y,Y) . \tag{B.15b}$$

### B.2.1 Estimateur standard de la variance

On rappelle que l'expression initiale de la variance est donnée par celle du moment centré d'ordre 2 :

$$\mathrm{Var}\,(X) = \boldsymbol{E}\left((X - \boldsymbol{E}\,(X))^2\right) = \boldsymbol{E}\left(X^2\right) - \boldsymbol{E}\,(X)^2 . \tag{B.16}$$

Celle-ci coincide avec celle du cumulant d'ordre 2. Pour une expérience statistique finie de longueur $m$, bien sûr, on utilise l'estimateur standard du moment centré d'ordre 2, dont l'expression est donnée ci-dessous :

$$\bar{V} = \overline{(X-\bar{X})^2} = \overline{X^2} - \bar{X}^2 . \tag{B.17}$$

On sait, par homogénéité des cumulants et décomposition des moments sur les cumulants, que $\bar{X}^2$ a pour espérance $\mu_x^2 + \sigma_x^2/m$. $\bar{V}$ a donc pour espérance mathématique $(m-1)\sigma_x^2/m$, ce qui en fait un estimateur non biaisé à un facteur multiplicatif près. Il est, du fait de son expression, invariant par translation. On peut ainsi se contenter de travailler dans l'espace des variables aléatoires centrées et fixer $\mu_x = 0$.

Si l'on cherche à obtenir l'expression de la variance de $\bar{V}$, il nous faut bien évidemment nous intéresser à l'espérance mathématique de $\bar{V}^2$. Si on met au carré l'expression de $\bar{V}$, on obtient :

$$\boldsymbol{E}\left(\bar{V}^2\right) = \boldsymbol{E}\left(\overline{X^2}^2\right) - 2\boldsymbol{E}\left(\bar{X}^2\overline{X^2}\right) + \boldsymbol{E}\left(\bar{X}^4\right) . \tag{B.18}$$

Le dernier terme s'obtient facilement en combinant les équations (B.13) et (B.8) :

$$\boldsymbol{E}\left(\bar{X}^4\right) = \frac{3\sigma_x^4}{m^2} + \frac{\gamma_4}{m^3} \,. \tag{B.19}$$

Le premier terme s'obtient à l'aide du changement de variable $Y = (X - \mu_x)^2$. L'espérance de $Y$ est alors bien évidemment $\sigma_x^2$. On peut alors réinterpréter l'équation (B.13) comme donnant l'espérance de $Y^2$, et en tirer le résultat à l'aide de l'équation (B.8) :

$$\boldsymbol{E}\left(\overline{X^2}^2\right) = \boldsymbol{E}\left(\bar{Y}^2\right) = (\sigma_x^2)^2 + \frac{2\sigma_x^4 + \gamma_4}{m} \,. \tag{B.20}$$

Pour le terme central, cependant, on ne pourra se contenter d'utiliser ces méthodes. On va alors devoir décomposer la somme implicite aux valeurs moyennes et réorganiser ses termes :

$$\begin{aligned}
\boldsymbol{E}\left(\overline{X^2}\bar{X}^2\right) &= \frac{1}{m^3}\sum_{i,j,k}^{m} \boldsymbol{E}\left(X_i^2 X_j X_k\right) \\
&= \frac{1}{m^2}\left[\boldsymbol{E}\left(X^4\right) + (m-1)\boldsymbol{E}\left(X^2\right)^2 + 2(m-1)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^3\right) + (m-1)(m-2)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right)^2\right] \\
&= \frac{\boldsymbol{E}\left(X^4\right)}{m^2} + (m-1)\frac{\boldsymbol{E}\left(X^2\right)^2}{m^2} \\
&= \frac{3\sigma_x^4 + \gamma_4}{m^2} + \frac{(m-1)\sigma_x^4}{m^2} \,.
\end{aligned} \tag{B.21}$$

En intégrant ces espérances à l'équation (B.18), on arrive à l'expression suivante pour $\boldsymbol{E}\left(\bar{V}^2\right)$ :

$$\boldsymbol{E}\left(\bar{V}^2\right) = \frac{m^2-1}{m^2}\sigma_x^4 + \frac{(m-1)^2}{m^3}\gamma_4 \,. \tag{B.22}$$

On en déduit alors l'expression de la variance et de l'erreur quadratique moyenne de $\bar{V}$ :

$$\mathrm{Var}\left(\bar{V}(X)\right) = \frac{2(m-1)}{m^2}\sigma_x^4 + \frac{(m-1)^2\gamma_4}{m^3} \underset{\infty}{=} \frac{2\sigma_x^4 + \gamma_4}{m} + \mathcal{O}\left(\frac{1}{m^2}\right) \,; \tag{B.23a}$$

$$\mathrm{MSE}(\bar{V}(X)) = \mathrm{Var}\left(\bar{V}(X)\right) + \left(\frac{\sigma_x^2}{m}\right)^2 \underset{\infty}{=} \frac{2\sigma_x^4 + \gamma_4}{m} + \mathcal{O}\left(\frac{1}{m^2}\right) \,. \tag{B.23b}$$

On peut remarquer que, bien que l'on ne puisse pas appliquer la loi des grands nombres à $\bar{V}(X)$, celle-ci étant échantillonnée que sur une seule simulation, sa variance a un comportement semblable en fonction de la longueur de la simulation à la variance d'une valeur moyenne. On peut donc assimiler de manière quelque peu abusive à une pseudo-moyenne de valeurs de la variance, et ainsi considérer $2\sigma_x^4 + \gamma_4(X)$ comme la variance intrinsèque de l'estimateur de la variance.

### B.2.2 Variance de l'estimateur standard de la covariance

La désymétrisation de l'estimateur standard de la variance nous permet d'obtenir de manière très directe l'estimateur standard de la covariance :

$$\bar{C}(X,Y) = \overline{XY} - \bar{X}\bar{Y} \,. \tag{B.24}$$

L'espérance mathématique se retrouve de la même manière que celle de l'estimateur standard de la variance, et vaut $\boldsymbol{E}\left(\bar{C}(X,Y)\right) = \frac{m-1}{m}\mathrm{Cov}\left(X,Y\right)$. Pour calculer la variance de l'estimateur standard de la covariance, on va, comme pour la variance, devoir en évaluer l'espérance mathématique du carré. On a :

$$\boldsymbol{E}\left(\bar{C}(X,Y)^2\right) = \boldsymbol{E}\left(\overline{XY}^2\right) + \boldsymbol{E}\left(\bar{X}^2\bar{Y}^2\right) - 2\boldsymbol{E}\left(\overline{XY}\bar{X}\bar{Y}\right) \,. \tag{B.25}$$

Si on réutilise les mêmes raisonnements qu'à la partie précédente, mais en remplaçant l'équation (B.13) par l'équation (B.15b), on obtient facilement les valeurs des deux premiers termes :

$$\boldsymbol{E}\left(\overline{XY}^2\right) = (c_{xy}\sigma_x\sigma_y)^2 + \frac{1+c_{xy}^2}{m}\sigma_x^2\sigma_y^2 + \frac{\gamma_4(X,X,Y,Y)}{m} \,; \tag{B.26a}$$

$$\boldsymbol{E}\left(\bar{X}^2\bar{Y}^2\right) = \frac{1+2c_{xy}^2}{m^2}\sigma_x^2\sigma_y^2 + \frac{\gamma_4(X,X,Y,Y)}{m^3} \,. \tag{B.26b}$$

Il ne nous reste alors qu'à décomposer et réorganiser le dernier terme, de la même manière que pour la variance :

$$\begin{aligned}\boldsymbol{E}\left(\overline{XY}\bar{X}\bar{Y}\right) &= \frac{1}{m^3}\sum_{i,j,k}^{m}\boldsymbol{E}\left(X_iY_iX_jY_k\right)\\ &= \frac{1}{m^2}\left[\boldsymbol{E}\left(X^2Y^2\right)+(m-1)\boldsymbol{E}\left(XY\right)^2\right]\\ &= \frac{\sigma_x^2\sigma_y^2(1+2c_{xy}^2)+\gamma_4(X,X,Y,Y)}{m^2}+\frac{(m-1)\sigma_x^2\sigma_y^2c_{xy}}{m^2}\ .\end{aligned} \tag{B.27}$$

On peut ainsi reconstruire la valeur de l'expression de $\boldsymbol{E}\left(\bar{C}^2\right)$ :

$$\boldsymbol{E}\left(\bar{C}^2\right) = \frac{c_{xy}^2(m-1)^2+(1+c_{xy}^2)(m-1)}{m^2}\sigma_x^2\sigma_y^2+\frac{(m-1)^2}{m^3}\gamma_4(X,X,Y,Y)\ ; \tag{B.28}$$

et on en tire l'expression de la variance et l'erreur quadratique moyenne de $\bar{C}(X,Y)$ :

$$\mathrm{Var}\left(\bar{C}(X,Y)\right) = \frac{(1+c_{xy}^2)(m-1)}{m^2}\sigma_x^2\sigma_y^2+\frac{(m-1)^2\gamma_4(X,X,Y,Y)}{m^3} \underset{\infty}{=} \frac{\sigma_x^2\sigma_y^2(1+c_{xy}^2)+\gamma_4}{m}+\mathcal{O}\left(\frac{1}{m^2}\right)\ ; \tag{B.29a}$$

$$\mathrm{MSE}(\bar{C}(X,Y)) = \mathrm{Var}\left(\bar{C}(X,Y)\right)+\left(c_{xy}\frac{\sigma_x\sigma_yc_{xy}}{m}\right)^2 \underset{\infty}{=} \frac{\sigma_x^2\sigma_y^2(1+c_{xy}^2)+\gamma_4}{m}+\mathcal{O}\left(\frac{1}{m^2}\right)\ . \tag{B.29b}$$

Par la suite, puisqu'un biais en $1/m$ n'affecte pas le comportement à grand $m$ de l'erreur quadratique moyenne, nous n'en reparlerons pas.

### B.2.3 Covariances d'estimateurs standard

On peut adapter le raisonnement qu'on a présenté et employé aux deux sous-sections précédentes pour des covariances de deux estimateurs standard quelconques de covariances. Cela nous donne le résultat général suivant :

$$\frac{m^3}{(m-1)}\mathrm{Cov}\left(\bar{C}(X,Y),\bar{C}(Z,T)\right) = m\mathrm{Cov}\left(X,Z\right)\mathrm{Cov}\left(Y,T\right)+m\mathrm{Cov}\left(Y,Z\right)\mathrm{Cov}\left(Z,T\right)+(m-1)\gamma_4(X,Y,Z,T)\ . \tag{B.30}$$

On peut aisément vérifier que cette expression nous redonne bien les résultats présentés aux sous-sections B.2.1 et B.2.2.

## B.3 Estimateurs standard de cumulants d'ordre 3

Pour construire l'expression des moments centrés d'ordre 6, on introduit $\mu = 0$ dans l'équation (B.10f) :

$$\boldsymbol{E}\left((X-\mu_x)^6\right) = \gamma_6+15\gamma_4\sigma_x^2+10\gamma_3^2+15\sigma_x^6\ . \tag{B.31}$$

Cette expression se désymétrise aisément, en utilisant les six variables aléatoires $X$, $Y$, $Z$, $T$, $U$, $V$ qu'on considère centrées, et en acceptant de ne pas écrire toutes les permutations possibles :

$$\begin{aligned}\boldsymbol{E}\left(XYZTUV\right) = \gamma_6(X,Y,Z,T,U,V)+\gamma_3(X,Y,Z)\gamma_3(T,U,V)+...\\ +c_{xy}\sigma_x\sigma_y(\gamma_4(Z,T,U,V)+\sigma_z\sigma_t\sigma_u\sigma_vc_{zt}c_{uv})+...\end{aligned} \tag{B.32}$$

Cette expression nous donne, dans le cas particulier de $X$, $Y$, et $Z$ apparaissant deux fois chacune :

$$\begin{aligned}\boldsymbol{E}\left(X^2Y^2Z^2\right) = \sigma_x^2\sigma_y^2\sigma_z^2(1+2c_{xy}^2+2c_{xz}^2+2c_{yz}^2+8c_{xy}c_{xz}c_{yz})\\ +4\gamma_3(X,Y,Z)^2+2\gamma_3(X,X,Z)\gamma_3(Y,Y,Z)\\ +2\gamma_3(X,X,Y)\gamma_3(Y,Z,Z)+2\gamma_3(X,Y,Y)\gamma_3(X,Z,Z)\\ +\gamma_6(X,X,Y,Y,Z,Z)+\sigma_x^2\gamma_4(Y,Y,Z,Z)+\sigma_y^2\gamma_4(X,X,Z,Z)+\sigma_z^2\gamma_4(X,X,Y,Y)\\ +4\sigma_x\sigma_yc_{xy}\gamma_4(X,Y,Z,Z)+4\sigma_x\sigma_zc_{xz}\gamma_4(X,Y,Y,Z)+4\sigma_y\sigma_zc_{yz}\gamma_4(X,X,Y,Z)\ .\end{aligned} \tag{B.33}$$

### B.3.1 Estimateur standard du cumulant ternaire

On rappelle que dans le cas général, l'estimateur classique du cumulant ternaire symétrique est donnée par l'expression du moment centré d'ordre 3 :

$$\bar{\gamma_3} = \overline{X^3} - 3\overline{X}\,\overline{X^2} + 2\overline{X}^3 \ . \tag{B.34}$$

On peut réutiliser les raisonnements précédents pour déterminer les valeurs des deux termes extérieurs à l'aide de l'équation (B.10c) :

$$\begin{aligned} \boldsymbol{E}\left(\overline{X^3}\right) &= \boldsymbol{E}\left(X^3\right) = \mu_x^3 + 3\sigma_x^2\mu_x + \gamma_3(X) \ ; \\ \boldsymbol{E}\left(\bar{X}^3\right) &= \mu_x^3 + \frac{3}{m}\sigma_x^2\mu_x + \frac{1}{m^2}\gamma_3(X) \ . \end{aligned} \tag{B.35}$$

Et pour le terme central, on utilise le développement par décomposition :

$$\begin{aligned} \boldsymbol{E}\left(\overline{X^2}\bar{X}\right) &= \frac{1}{m^2}\sum_{i,j=1}^{m} \boldsymbol{E}\left(X_i^2 X_j\right) \\ &= \frac{m-1}{m}\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right) + \frac{1}{m}\boldsymbol{E}\left(X^3\right) \\ &= \mu_x^3 + \frac{m+2}{m}\sigma_x^2\mu_x + \frac{1}{m}\gamma_3(X) \ . \end{aligned} \tag{B.36}$$

On arrive donc à l'espérance suivante de l'estimateur standard du cumulant ternaire :

$$\boldsymbol{E}\left(\bar{\gamma_3}(X)\right) = (1-3+2)\mu_x^3 + \frac{3m-3(m+2)+6}{m}\sigma_x^2\mu_x + \frac{m^2-3m+2}{m^2}\gamma_3(X) = \frac{(m-1)(m-2)}{m^2}\gamma_3(X) \ . \tag{B.37}$$

Comme on s'y attendait, il s'agit bien d'un estimateur invariant par translation, et par la suite on travaillera dans l'espace des variables aléatoires centrées. Par ailleurs, on voit que là aussi, on a un estimateur non biaisé à un facteur multiplicatif près, et le biais est de l'ordre de $\mathcal{O}\left(1/m\right)$. Maintenant, si on cherche à s'intéresser à la variance de cet estimateur, on doit calculer la variance du carré de celui-ci :

$$\boldsymbol{E}\left(\bar{\gamma_3}^2\right) = \boldsymbol{E}\left(\overline{X^3}^2\right) - 6\boldsymbol{E}\left(\overline{X^3}\,\overline{X^2}\bar{X}\right) + 9\boldsymbol{E}\left(\overline{X^2}^2\bar{X}^2\right) + 4\boldsymbol{E}\left(\overline{X^3}\bar{X}^3\right) - 12\boldsymbol{E}\left(\overline{X^2}\bar{X}^4\right) + 4\boldsymbol{E}\left(\bar{X}^6\right) \ . \tag{B.38}$$

On développe les termes de ce développement en produits de moyennes de la même manière qu'à la section précédente, mais en se servant cette fois-ci de l'équation (B.31). Cela nous permet d'arriver aux expressions ci-dessous :

$$m\boldsymbol{E}\left(\overline{X^3}^2\right) = \boldsymbol{E}\left(X^6\right) + (m-1)\boldsymbol{E}\left(X^3\right)^2 \tag{B.39a}$$

$$= \gamma_6 + (m+9)\gamma_3^2 + 15\gamma_4\sigma_x^2 + 15\sigma_x^6 \ ; \tag{B.39b}$$

$$m^2\boldsymbol{E}\left(\overline{X^2}\,\overline{X}\,\overline{X^3}\right) = \boldsymbol{E}\left(X^6\right) + (m-1)\boldsymbol{E}\left(X^3\right)^2 + (m-1)\boldsymbol{E}\left(X^4\right)\boldsymbol{E}\left(X^2\right) \tag{B.39c}$$

$$= \gamma_6 + (m+9)\gamma_3^2 + (m+14)\gamma_4\sigma_x^2 + (3m+12)\sigma_x^6 \ ; \tag{B.39d}$$

$$m^3\boldsymbol{E}\left(\overline{X}^2\,\overline{X^2}^2\right) = \boldsymbol{E}\left(X^6\right) + 3(m-1)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X^4\right) + 2(m-1)\boldsymbol{E}\left(X^3\right)^2 + (m-1)(m-2)\boldsymbol{E}\left(X^2\right)^3 \tag{B.39e}$$

$$= \gamma_6 + (2m+8)\gamma_3^2 + (3m+12)\gamma_4\sigma_x^2 + (m^2+6m+8)\sigma_x^6 \ ; \tag{B.39f}$$

$$m^3\boldsymbol{E}\left(\overline{X}^3\,\overline{X^3}\right) = \boldsymbol{E}\left(X^6\right) + 3(m-1)\boldsymbol{E}\left(X^4\right)\boldsymbol{E}\left(X^2\right) + (m-1)\boldsymbol{E}\left(X^3\right)^2 \tag{B.39g}$$

$$= \gamma_6 + (m+9)\gamma_3^2 + (3m+12)\gamma_4\sigma_x^2 + (9m+6)\sigma_x^6 \ ; \tag{B.39h}$$

$$m^4\boldsymbol{E}\left(\overline{X}^4\,\overline{X^2}\right) = \boldsymbol{E}\left(X^6\right) + 4(m-1)\boldsymbol{E}\left(X^3\right)^2 + 7(m-1)\boldsymbol{E}\left(X^4\right)\boldsymbol{E}\left(X^2\right) + 3(m-1)(m-2)\boldsymbol{E}\left(X^2\right)^3 \tag{B.39i}$$

$$= \gamma_6 + (4m+6)\gamma_3^2 + (7m+8)\gamma_4\sigma_x^2 + (3m^2+12m)\sigma_x^6 \ ; \tag{B.39j}$$

$$m^5\boldsymbol{E}\left(\overline{X}^6\right) = \gamma_6 + 10m\gamma_3^2 + 15m\gamma_4\sigma_x^2 + 15m^2\sigma_x^6 \ . \tag{B.39k}$$

On peut réintroduire ces résultats dans l'expression globale, ce qui nous donne comme espérance du carré de $\bar{\gamma_3}$ :

$$\begin{aligned} \boldsymbol{E}\left(\bar{\gamma_3}(X)^2\right) = \ & \frac{(m-1)(m-2)^2(m+8)}{m^4}\gamma_3(X)^2 + \frac{(m-1)^2(m-2)^2}{m^5}\gamma_6(X) \\ & + 6\frac{(m-1)(m-2)}{m^3}\sigma_x^6 + 9\frac{(m-1)(m-2)^2}{m^4}\sigma_x^2\gamma_4(X) \ . \end{aligned} \tag{B.40}$$

On arrive ainsi à l'expression suivante de la variance de l'estimateur classique du cumulant ternaire :

$$\begin{aligned}\mathrm{Var}\left(\bar{\gamma_3}(X)\right) &= \frac{9(m-1)(m-2)^2}{m^4}\gamma_3^2 + \frac{(m-1)^2(m-2)^2}{m^5}\gamma_6 + 6\frac{(m-1)(m-2)}{m^3}\sigma_x^6 + 9\frac{(m-1)(m-2)^2}{m^4}\sigma_x^2\gamma_4 \\ &\underset{\infty}{=} \frac{1}{m}\left[9\gamma_3^2 + 6\sigma_x^6 + 9\sigma_x^2\gamma_4 + \gamma_6\right] + \mathcal{O}\left(\frac{1}{m^2}\right) .\end{aligned} \tag{B.41}$$

Afin de tester si on a ou non des relations de commutation entre opérateurs de cumulants (on ne sait jamais), on a également développé et calculé $\gamma_3(\bar{V}(X))$, ce qui nous donne :

$$\begin{aligned}\gamma_3(\bar{V}(X)) &= 8\frac{m-1}{m^3}\sigma_x^6 + 4\frac{(m-2)(m-1)}{m^4}\gamma_3^2 + 9\frac{(m-1)^2}{m^4}\sigma_x^2\gamma_4 + \frac{(m-1)^3}{m^5}\gamma_6 \\ &\underset{\infty}{=} \frac{1}{m^2}\left[8\sigma_x^6 + 4\gamma_3^2 + 9\sigma_x^2\gamma_4 + \gamma_6\right] + \mathcal{O}\left(\frac{1}{m^3}\right) .\end{aligned} \tag{B.42}$$

### B.3.2 Estimateur de la covariance étendue ternaire

On construit l'estimateur classique de la covariance étendue ternaire par désymétrisation de celui du cumulant ternaire symétrique :

$$\bar{\gamma_3}(X,Y,Z) = \overline{XYZ} - \overline{XY}\bar{Z} - \overline{XZ}\bar{Y} - \overline{YZ}\bar{X} + 2\bar{X}\bar{Y}\bar{Z} . \tag{B.43}$$

De même que l'estimateur duquel il est dérivé, il est soumis à un biais équivalent à un facteur multiplicatif de $\frac{(m-1)(m-2)}{m^2}$. Pour calculer la variance sur cet estimateur, on doit donc calculer $\boldsymbol{E}\left(\bar{\gamma_3}(X,Y,Z)^2\right)$.

Le développement du carré donne 15 termes, qui se ramènent à 7 par permutation. (le terme en $\boldsymbol{E}\left(\bar{X}^2\bar{Y}^2\bar{Z}^2\right)$ n'est pas précisé ici mais s'obtient de manière évidente à partir du terme du moment centré d'ordre 6 asymétrique)

$$m\boldsymbol{E}\left(\overline{XYZ}^2\right) = \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\boldsymbol{E}\left(XYZ\right)^2 \ ; \tag{B.44a}$$

$$\begin{aligned}m^3\boldsymbol{E}\left(\overline{X}^2\,\overline{YZ}^2\right) &= \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\left[2\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(Y^2Z^2\right) + 2\boldsymbol{E}\left(XYZ\right)^2\right] \\ &\quad + (m-1)(m-2)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right)^2 \ ;\end{aligned} \tag{B.44b}$$

$$\begin{aligned}m^3\boldsymbol{E}\left(\overline{X}\,\overline{YZ}\,\overline{Y}\,\overline{XZ}\right) &= \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\left[\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(XY^2Z\right) + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XYZ^2\right)\right] \\ &\quad + (m-1)\left[\boldsymbol{E}\left(XYZ\right)^2 + \boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(Y^2Z\right)\right] + (m-1)(m-2)\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(XZ\right) \ ;\end{aligned} \tag{B.44c}$$

$$m^2\boldsymbol{E}\left(\overline{XY}\,\overline{Z}\,\overline{XYZ}\right) = \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\left[\boldsymbol{E}\left(XYZ\right)^2 + \boldsymbol{E}\left(XYZ^2\right)\boldsymbol{E}\left(XY\right)\right] \ ; \tag{B.44d}$$

$$\begin{aligned}m^3\boldsymbol{E}\left(\overline{X}\,\overline{Y}\,\overline{Z}\,\overline{XYZ}\right) &= \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\left[\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XYZ^2\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(XY^2Z\right) + \boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X^2YZ\right)\right] \\ &\quad + (m-1)\boldsymbol{E}\left(XYZ\right)^2 \ ;\end{aligned} \tag{B.44e}$$

$$\begin{aligned}m^4\boldsymbol{E}\left(\overline{X}^2\,\overline{Y}\,\overline{Z}\,\overline{YZ}\right) &= \boldsymbol{E}\left(X^2Y^2Z^2\right) + (m-1)\left[\boldsymbol{E}\left(XYZ\right)^2 + \boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(YZ^2\right) + \boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(Y^2Z\right)\right] \\ &\quad + \left[2\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(Y^2Z^2\right) + 2\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(XY^2Z\right) + 2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XYZ^2\right)\right] \\ &\quad + (m-1)(m-2)\left[\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right)^2 + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(YZ\right)\right] .\end{aligned} \tag{B.44f}$$

Reconstruire l'espérance est ici un calcul fastidieux, qu'on a choisi d'accélérer en se servant de calcul matriciel. On retrouve l'expression observée à la sous-partie précédente en effectuant les substitutions suivantes (celle de $\gamma_6$ va de soi)

$$6\sigma_x^6 \Longleftrightarrow \sigma_x^2\sigma_y^2\sigma_z^2(1 + c_{xy}^2 + c_{xz}^2 + c_{yz}^2 + 2c_{xy}c_{xz}c_{yz}) \ ; \tag{B.45a}$$

$$9\gamma_3(X)^2 \Longleftrightarrow 3\gamma_3(X,Y,Z)^2 + 2\gamma_3(X,X,Z)\gamma_3(Y,Y,Z) + 2\gamma_3(X,X,Y)\gamma_3(Y,Z,Z) + 2\gamma_3(X,Y,Y)\gamma_3(X,Z,Z) \ ; \tag{B.45b}$$

$$\begin{aligned}9\gamma_4(X)\sigma_x^2 \Longleftrightarrow\ & \sigma_x^2\gamma_4(Y,Y,Z,Z) + \sigma_y^2\gamma_4(X,X,Z,Z) + \sigma_z^2\gamma_4(X,X,Y,Y) \\ & + 2\sigma_x\sigma_y c_{xy}\gamma_4(X,Y,Z,Z) + 2\sigma_x\sigma_z c_{xz}\gamma_4(X,Y,Y,Z) + 2\sigma_y\sigma_z c_{yz}\gamma_4(X,X,Y,Z) .\end{aligned} \tag{B.45c}$$

### B.3.3 Covariance de covariances étendues ternaires quelconques

On peut généraliser l'expression par désymétrisation. Dans ce cas, on part de deux cumulants ternaires mettant en jeu chacun deux triplets de variables aléatoires quelconques. Alors :

- Pour les termes mettant en jeu les cumulants ternaires, tous les termes sont dominants à part celui qui correspond aux deux triplets de départ.
- Pour les termes mettant en jeu les cumulants quaternaires, la covariance des termes dominants doit prendre une variable dans chacun des triplet de départ.
- Pour les termes ne mettant en jeu que des covariances, chacune des covariances des termes dominants doit prendre une variable dans chacun des triplets de départ.

## B.4 Construction d'estimateurs non biaisés de cumulants d'ordre 4 à 6

De manière général, construire des estimateurs non biaisés de $\mathrm{Var}\,(X)$ et $\gamma_3(X)$ n'est pas difficile, car ces quantités sont des moments centrés en plus d'être des cumulants. Cependant, ce n'est pas le cas pour les cumulants suivant. C'est pourquoi on va chercher à construire à partir des moments centrés des estimateurs non biaisés des cumulants d'ordre 4 à 6. Il faut cependant comprendre que ceux-ci seront soumis à de fortes variances de toute manière, car le terme dominant en terme de scaling pour le cumulant d'ordre $n$ est $\mathrm{Var}\,(X)^n$.

On partira du principe que l'on dispose des deux estimateurs suivants de biais nul :

$$\tilde{V}(X) = \frac{m}{m-1}[\bar{X^2} - \bar{X}^2]\,;$$
$$\tilde{\gamma_3}(X) = \frac{m^2}{(m-1)(m-2)}[\bar{X^3} + 2\bar{X}^3 - 3\bar{X}\,\bar{X^2}]\,.$$

Notre approche consistera dans cette partie à construire un système à $n$ équations à $n$ inconnues, où $n$ correspond au nombre de termes dans l'expression du moment centré du même ordre que le cumulant que nous cherchons, pour ensuite le résoudre, soit à la main, soit par inversion de matrice.

### B.4.1 Cumulant quaternaire

Pour le cumulant quaternaire, on doit construire un système à deux équations à deux inconnues. En d'autres termes, on doit trouver, outre l'expression du moment centré d'ordre 4, une équation d'ordre 4 centrée et invariante par translation. Nous choisirons pour ce faire l'équation (B.22).

Le moment centré d'ordre 4 a pour expression :

$$\bar{\mu_4}(X) = \bar{X^4} - 4\bar{X}\bar{X^3} + 6\bar{X^2}\bar{X^2} - 3\bar{X}^4\,. \tag{B.46}$$

On peut facilement en construire l'espérance en se servant des expressions construites à la partie B.2.1 et en notant que pour $X$ centrée, $\boldsymbol{E}\left(\bar{X}\overline{X^3}\right) = \boldsymbol{E}\left(X^4\right)/m$. On arrive alors au système à deux équations et deux inconnues suivant :

$$\begin{cases} \boldsymbol{E}\left(\bar{\mu_4}(X)\right) &= \frac{(m-1)(m^2-3m+3)}{m^3}\gamma_4(X) &+ 3\frac{(m-1)^2}{m^2}\sigma_x^4 \\ \boldsymbol{E}\left(\bar{V}^2(X)\right) &= \frac{(m-1)^2}{m^3}\gamma_4(X) &+ \frac{m^2-1}{m^2}\sigma_x^4 \end{cases}\,. \tag{B.47}$$

On peut inverser la deuxième ligne pour construire une expression de $\sigma_x^4$ :

$$\sigma_x^4 = \frac{m-1}{m+1}\boldsymbol{E}\left(\tilde{V}(X)^2\right) - \frac{m-1}{m(m+1)}\gamma_4(X)\,. \tag{B.48}$$

On peut ensuite insérer cette expression dans l'expression de $\bar{\mu_4}(X)$, pour construire l'estimateur suivant :

$$\boldsymbol{E}\left(\bar{\gamma_4}(X)\right) = \frac{(m-1)(m-2)(m-3)}{m^2(m+1)}\gamma_4(X) = \boldsymbol{E}\left(\bar{\mu_4}(X) - 3\frac{(m-1)^3}{m^2(m+1)}\tilde{V}^2(X)\right)\,. \tag{B.49}$$

C'est une expression non biaisée à un facteur multiplicatif près. L'expression non biaisée est :

$$\tilde{\gamma_4}(X) = \frac{m^2(m+1)}{(m-1)(m-2)(m-3)}(\bar{\mu_4}(X) - 3\frac{(m-1)^3}{m^2(m+1)}\tilde{V}^2(X))\,. \tag{B.50}$$

On peut remarquer que cela revient fondamentalement à utiliser l'estimateur classique $\bar{\gamma_4} = \bar{X^4} - 4\bar{X}\bar{X^3} - 3\bar{X^2}^2 + 12\bar{X^2}\bar{X^2} - 6\bar{X}^4$. En pratique, cependant, l'estimateur classique souffre d'un biais d'échantillonnage fini qui laisse un terme résiduel en $\frac{6}{m}\sigma_x^4$, qui dans un système physico-chimique a généralement un scaling plus important.

On peut généraliser cette méthode fort aisément au cas asymétrique, en partant de l'expression de la covariance d'estimateurs de la covariance fournie par l'équation B.30.

On peut également s'en servir pour construire un estimateur non biaisé de $\sigma_x^4$. En effet, il suffit de retrancher $\tilde{\gamma_4}/m$ à $\tilde{V}^2$ pour obtenir un estimateur non biaisé à un coefficient multiplicatif près.

### B.4.2 Cumulant quinaire

On peut remarquer que l'expression du moment d'ordre 5, centrée, donne un développement quasiment aussi simple que celle de l'ordre 4 :

$$\boldsymbol{E}\left((X-\mu_x)^5\right) = \gamma_5 + 10\sigma_x^2\gamma_3 \,. \tag{B.51}$$

On n'a donc à construire qu'un système à deux équations et deux inconnues, et le second terme nous indique que la deuxième équation consistera de l'expression de $\boldsymbol{E}\left(\bar{V}(X)\bar{\gamma_3}(X)\right)$. Après avoir déployé le raisonnement habituel par développement et décomposition, on arrive au système d'équations suivant :

$$\begin{cases} \boldsymbol{E}\left(\bar{\mu_5}(X)\right) &= \frac{(m-1)(m^3-4m^2+6m-4)}{m^4}\gamma_5(X) &+ 10\frac{(m-1)^2(m-2)}{m^3}\sigma_x^2\gamma_3 \\ \boldsymbol{E}\left(\bar{V}(X)\bar{\gamma_3}(X)\right) &= \frac{(m-1)^2(m-2)}{m^4}\gamma_5(X) &+ \frac{(m-1)(m-2)(m+5)}{m^3}\sigma_x^2\gamma_3 \end{cases} . \tag{B.52}$$

On retranche à la première ligne $10(m-1)/(m+5)$ fois la deuxième, ce qui nous donne pour estimateur non biaisé à un facteur multiplicatif près :

$$\boldsymbol{E}\left(\bar{\mu_5}(X) - 10\frac{m-1}{m+5}\bar{V}(X)\bar{\gamma_3}(X)\right) = \frac{(m-1)(m-2)(m-3)(m-4)}{m^3(m+5)}\gamma_5(X) \,. \tag{B.53}$$

Cela nous permet ainsi de faire disparaître un biais en $\frac{60}{m}\sigma_x^2\gamma_3(X)$.

Pour les variantes asymétriques, on peut utiliser la somme sur les permutations possibles des produits d'estimateurs classiques. Cependant, faire disparaître les termes parasites dans les estimateurs de produit se révèle plus difficile. En effet, on trouve :

$$\begin{aligned} m\mathrm{Cov}\left(\bar{\gamma_3}(X,Y,Z),\bar{C}(W,T)\right) = \frac{(m-1)(m-2)}{m^2}[&\sigma_{xw}\gamma_{yzt} + \sigma_{yw}\gamma_{xzt} + \sigma_{zw}\gamma_{xyt} + \sigma_{xt}\gamma_{yzw} + \sigma_{yt}\gamma_{xzw} \\ &+ \sigma_{zt}\gamma_{xyw}] + \frac{(m-1)^2(m-2)}{m^3}\gamma_5(X,Y,Z,T,W) \,. \end{aligned} \tag{B.54}$$

On ne peut de plus pas prélever le terme manquant dans le corps du produit comme on l'a fait pour le produit de covariances. On doit donc inverser une matrice 10×10 symétrique, dont le déterminant se comporte en $\frac{(m+5)(m-3)^5}{m^6}$.

On trouve que, une fois les covariances corrigées du terme lié au cumulant quinaire,

$$\begin{aligned} (m+5)(m-3)\mathrm{Cov}\left(W,T\right)\gamma_3(X,Y,Z) =&(m^2+3m-6)\bar{C}(W,T)\bar{\gamma_3}(X,Y,Z) + 4[\bar{C}(X,Y)\bar{\gamma_3}(Z,W,T) \\ &+ \bar{C}(X,Z)\bar{\gamma_3}(Y,W,T) + \bar{C}(Y,Z)\bar{\gamma_3}(X,W,T)] - (m+1)[\bar{C}(X,W)\bar{\gamma_3}(Y,Z,T) \\ &+ \bar{C}(Y,W)\bar{\gamma_3}(X,Z,T) + \bar{C}(X,T)\bar{\gamma_3}(Y,Z,W) + \bar{C}(Y,T)\bar{\gamma_3}(X,Z,W) \\ &+ \bar{C}(Z,T)\bar{\gamma_3}(X,Y,W) + \bar{C}(Z,W)\bar{\gamma_3}(X,Y,T)] \,. \end{aligned} \tag{B.55}$$

### B.4.3 Cumulant sénaire

Pour le cumulant sénaire, on a trois termes additionnels. On doit donc se servir d'un système à quatre équations et quatre inconnues, mettant en jeu les expressions de $\bar{V}^3$, ${\bar{\gamma_3}}^2$, et $\bar{V}\bar{\mu_4}$.

En utilisant les expressions employées à la sous-section B.3.1, on arrive au système d'équations suivant (écrit sous forme matricielle) :

$$\begin{bmatrix} \bar{V}^3 \\ {\bar{\gamma_3}}^2 \\ \bar{\gamma_4}\bar{V} \\ \bar{\mu_6} \end{bmatrix} = \begin{bmatrix} \frac{(m+3)(m^2-1)}{m^3} & 4\frac{(m-2)(m-1)}{m^4} & 3\frac{(m+3)(m-1)^2}{m^4} & \frac{(m-1)^3}{m^5} \\ 6\frac{(m-1)(m-2)}{m^3} & \frac{(m+8)(m-1)(m-2)^2}{m^4} & 9\frac{(m-1)(m-2)^2}{m^4} & \frac{(m-1)^2(m-2)^2}{m^5} \\ 0 & 6\frac{(m-2)(m-1)(m-3)}{m^3(m+1)} & \frac{(m+7)(m-1)(m-2)(m-3)}{m^3(m+1)} & \frac{(m-1)^2(m-2)(m-3)}{m^4(m+1)} \\ 15\frac{(m-1)^3}{m^3} & 10\frac{(m-1)^2(m-2)^2}{m^4} & 15\frac{(m-1)^5+(m-1)^2}{m^5} & \frac{(m-1)^6+(m-1)}{m^6} \end{bmatrix} \begin{bmatrix} \sigma^6 \\ \gamma_3^2 \\ \gamma_4\sigma^2 \\ \gamma_6 \end{bmatrix} . \tag{B.56}$$

Il s'agit bien évidemment d'une forme inconfortable à utiliser. On peut donc introduire les facteurs de normalisation pour revenir aux produits d'estimateurs non biaisés.

$$\begin{bmatrix} \tilde{V}^3 \\ \tilde{\gamma_3}^2 \\ \tilde{\gamma_4}\tilde{V} \\ \bar{\mu_6} \end{bmatrix} = \begin{bmatrix} \frac{(m+3)(m+1)}{(m-1)^2} & 4\frac{(m-2)}{m(m-1)^2} & 3\frac{(m+3)}{m(m-1)} & \frac{1}{m^2} \\ 6\frac{m}{(m-1)(m-2)} & 1+\frac{9}{m-1} & \frac{9}{m-1} & \frac{1}{m} \\ 0 & \frac{6}{(m-1)} & 1+\frac{8}{m-1} & \frac{1}{m} \\ 15\frac{(m-1)^3}{m^3} & 10\frac{(m-1)^2(m-2)^2}{m^4} & 15\frac{(m-1)^5+(m-1)^2}{m^5} & \frac{(m-1)^6+(m-1)}{m^6} \end{bmatrix} \begin{bmatrix} \sigma^6 \\ \gamma_3^2 \\ \gamma_4\sigma^2 \\ \gamma_6 \end{bmatrix} . \tag{B.57}$$

Pour calculer la matrice inverse, on a tout d'abord besoin du déterminant de celle-ci. L'expression de celui-ci donne, après quelques calculs :

$$\Delta = \frac{(m-3)(m-4)(m-5)(m+1)}{m(m-1)^3} . \tag{B.58}$$

On peut tirer de la formule de Sarrus l'expression de la comatrice, ce qui nous donne :

$$\begin{aligned}
\Delta\gamma_6 &= \frac{m(m^2+15m-4)(m+1)^2}{(m-1)^4(m-2)}\bar{\mu_6} - 15\frac{(m-3)(m+4)}{m^2}\tilde{\gamma_4}\tilde{V} \\
&\quad - 10\frac{(m+1)(m-2)}{m^3(m-1)}\tilde{\gamma_3}^2 - 15\frac{m^3+8m^2-17m+12}{m^2(m-2)}\tilde{V}^3 ; && \text{(B.59a)} \\
\Delta\sigma^2\gamma_4 &= -\frac{(m+1)(m+4)}{(m-1)^2(m-2)}\bar{\mu_6} + \frac{(m-3)(m^4-10m^2+45m-60)}{m^3(m-1)^2}\tilde{V}\tilde{\gamma_4} \\
&\quad + 2\frac{(m-2)(m+1)((2m^3+20-5m-5m^2)}{m^4(m-1)^2}\tilde{\gamma_3}^2 + 3\frac{60+5m^4-23m^3-85m+55m^2}{m^3(m-1)(m-2)}\tilde{V}^3 ; && \text{(B.59b)} \\
\Delta\gamma_3^2 &= -\frac{(m+1)(m^2-m+4)}{(m-1)^3(m-2)}\bar{\mu_6} + 3\frac{(m-3)(2m^3+20-5m-5m^2)}{m^3(m-1)^2}\tilde{V}\tilde{\gamma_4} \\
&\quad + \frac{(m-2)(m+1)(m^4-8m^3+25m^2-10m-40}{m^4(m-1)^2}\tilde{\gamma_3}^2 + 3\frac{-60+55m+3m^4-9m^3-5m^2}{m^3(m-1)(m-2)}\tilde{V}^3 ; && \text{(B.59c)} \\
\Delta\sigma^6 &= 2\frac{(m+1)}{(m-1)^3}\bar{\mu_6} - 3\frac{(m-2)(m-3)(m^2-5m+10}{m^3(m-1)^2}\tilde{V}\tilde{\gamma_4} \\
&\quad - 2\frac{(m-2)(m+1)(3m^2-15m+20)}{m^4(m-1)^2}\tilde{\gamma_3}^2 + \frac{m^4-15m^3+62m^2-120m+90}{m^3(m-1)}\tilde{V}^3 . && \text{(B.59d)}
\end{aligned}$$

Cette expression nous permet ainsi d'obtenir des estimateurs non biaisés du cumulant sénaire $\gamma_6$, ainsi que des produits de cumulants $\sigma^6$, $\sigma^2\gamma_4$ et$\gamma_3^2$. Dans le cas asymétrique, pour le calcul du seul cumulant sénaire, on devra bien évidemment employer des regroupements. Pour des calculs plus complexes, comme l'estimation de la pseudo-variance ou pseudo-covariance de cumulants ternaires, on devra inverser une matrice $41 \times 41$.

Dans le cas pratique, cependant, on laissera l'ordinateur inverser la matrice lui-même, pour se simplifier la vie.

## B.5 Variance d'estimateurs améliorés

Dans cette section, on va s'intéresser au calcul des estimateurs améliorés que l'on a construit aux chapitres 6 et 7. Il va sans dire que les expressions qu'on développe ici sont généralisables à d'autres estimateurs composites, ie construits par une somme d'estimateurs standard sans êtres eux-même des estimateurs standard.

### B.5.1 Estimateur composite de la variance

L'estimateur amélioré de la variance que l'on a pu déduire à la section 6.2 prend l'expression suivante :

$$\tilde{V}(X) = \bar{C}(X, \tilde{X}_{\text{pr}}) + \overline{\sum_{i=1}^{p} \bar{V}(X|\bar{\imath})} ; \tag{B.60}$$

où $\tilde{X}_{\text{pr}}$ est l'estimateur amélioré de $X$ qu'on a défini à la section 3.4, et $\bar{V}(X|\bar{\imath})$ est l'estimateur standard de la variance de $X$ dans l'espace conditionnel qu'on a défini à la section 3.2. Toutes les notations peuvent être retrouvées à l'annexe A.

On s'est déjà intéressés aux propriétés de zéro-variance et de biais de cet estimateur au chapitre 6. Pour simplifier davantage notre expression, on posera $V_i = \bar{V}(X|\bar{\imath})$. Bien sûr, nous disposons déjà de l'expression

générale de la variance du premier terme, et de celle du second terme. Nous reste alors à construire l'expression générale de la covariance d'une valeur moyenne et d'un estimateur classique de covariance... autrement dit, quelque chose qui prend la forme $\mathrm{Cov}\left(\bar{C}(X,Y),\bar{Z}\right)$.

Pour ce faire, on réutilise les expressions par développement de produit :

$$\begin{aligned} m\boldsymbol{E}\left(\overline{XY}\bar{Z}\right) &= (m-1)\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(Z\right)+\boldsymbol{E}\left(XYZ\right) \\ &= m\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(Z\right)+\mathrm{Cov}\left(XY,Z\right)\ ; \end{aligned} \tag{B.61a}$$

$$\begin{aligned} m^2\boldsymbol{E}\left(\bar{X}\bar{Y}\bar{Z}\right) &= (m-1)(m-2)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right)+(m-1)[\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(Z\right)+\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(Y\right) \\ &\quad +\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X\right)]+\boldsymbol{E}\left(XYZ\right) \\ &= (m^2)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right)+m[\boldsymbol{E}\left(Z\right)\mathrm{Cov}\left(X,Y\right)+\boldsymbol{E}\left(Y\right)\mathrm{Cov}\left(X,Z\right)+\boldsymbol{E}\left(X\right)\mathrm{Cov}\left(Z,Y\right)] \\ &\quad +\gamma_3(X,Y,Z)\ . \end{aligned} \tag{B.61b}$$

Cela nous ramène à :

$$\begin{aligned} m\mathrm{Cov}\left(\bar{C}(XY),\bar{Z}\right) &= \mathrm{Cov}\left(XY,Z\right)-\mathrm{Cov}\left(X,Z\right)\boldsymbol{E}\left(Y\right)-\mathrm{Cov}\left(Y,Z\right)\boldsymbol{E}\left(X\right)-\frac{\gamma_3(X,Y,Z)}{m} \\ &= \frac{(m-1)}{m}\gamma_3(X,Y,Z)\ . \end{aligned} \tag{B.62}$$

On se sert ici de l'expression suivante, que l'on peut déduire de l'expression de la covariance généralisée ternaire, et qui nous servira en pratique à calculer $\gamma_3(X,\Delta X,\sum_i V_i)$ :

$$\gamma_3(X,Y,Z)=\mathrm{Cov}\left(XY,Z\right)-\boldsymbol{E}\left(Y\right)\mathrm{Cov}\left(X,Z\right)-\boldsymbol{E}\left(X\right)\mathrm{Cov}\left(Y,Z\right)\ .$$

Si on réunit nos expressions, on en arrive alors à :

$$\begin{aligned} \mathrm{Var}\left(\tilde{V}(X)\right) &= \mathrm{Var}\left(\bar{C}(X,\tilde{X}_{\mathrm{pr}})\right)+2\mathrm{Cov}\left(\bar{C}(X,\tilde{X}_{\mathrm{pr}}),\overline{\sum_{i=1}^{p}V_i}\right)+\frac{1}{m}\mathrm{Var}\left(\sum_{i=1}^{p}V_i\right) \\ &\underset{\infty}{=} \frac{1}{m}\left[\mathrm{Var}\left(X\right)\mathrm{Var}\left(\tilde{X}_{\mathrm{pr}}\right)+\mathrm{Cov}\left(X,\tilde{X}_{\mathrm{pr}}\right)^2+\gamma_4(X,X,\tilde{X}_{\mathrm{pr}},\tilde{X}_{\mathrm{pr}})+\mathrm{Var}\left(\sum_{i=1}^{p}V_i\right)\right. \\ &\quad \left.+2\gamma_3\left(X,\tilde{X}_{\mathrm{pr}},\sum_{i=1}^{p}V_i\right)\right]+\mathcal{O}\left(\frac{1}{m^2}\right)\ . \end{aligned} \tag{B.63}$$

### B.5.2 Estimateur composite de la covariance

Si on veut utiliser l'estimateur composite de la covariance défini à l'équation (6.24), que l'on reproduit ci-dessous :

$$\tilde{C}(X,Y)=\sum_{i=1}^{p}\overline{\bar{C}(X,Y|\bar{\imath})}+\frac{\bar{C}(\tilde{X}_{\mathrm{pr}},Y)+\bar{C}(X,\tilde{Y}_{\mathrm{pr}})}{2}\ ;$$

on peut se contenter d'adapter par désymétrisation l'expression qu'on a obtenue à la fin de la sous-section précédente. Cela nous donne l'équation suivante :

$$\begin{aligned} \mathrm{Var}\left(\tilde{C}(X,Y)\right) &\underset{\infty}{=} \frac{1}{m}\left[\gamma_3\left(X,\tilde{Y}_{\mathrm{pr}},\sum_{i=1}^{p}C_i\right)+\gamma_3\left(Y,\tilde{X}_{\mathrm{pr}},\sum_{i=1}^{p}C_i\right)+\mathrm{Var}\left(\sum_{i=1}^{p}C_i\right)\right] \\ &\quad +\frac{1}{4m}\left[\mathrm{Var}\left(X\right)\mathrm{Var}\left(\tilde{Y}_{\mathrm{pr}}\right)+\mathrm{Var}\left(Y\right)\mathrm{Var}\left(\tilde{X}_{\mathrm{pr}}\right)+\mathrm{Cov}\left(X,\tilde{Y}_{\mathrm{pr}}\right)^2+\mathrm{Cov}\left(Y,\tilde{X}_{\mathrm{pr}}\right)^2\right. \\ &\quad +2\mathrm{Cov}\left(X,\tilde{X}_{\mathrm{pr}}\right)\mathrm{Cov}\left(Y,\tilde{Y}_{\mathrm{pr}}\right)+2\mathrm{Cov}\left(X,Y\right)\mathrm{Cov}\left(\tilde{X}_{\mathrm{pr}},\tilde{Y}_{\mathrm{pr}}\right)+\gamma_4\left(X,X,\tilde{Y}_{\mathrm{pr}},\tilde{Y}_{\mathrm{pr}}\right) \\ &\quad \left.+\gamma_4\left(Y,Y,\tilde{X}_{\mathrm{pr}},\tilde{X}_{\mathrm{pr}}\right)+2\gamma_4\left(X,Y,\tilde{X}_{\mathrm{pr}},\tilde{Y}_{\mathrm{pr}}\right)\right]+\mathcal{O}\left(\frac{1}{m^2}\right)\ . \end{aligned} \tag{B.64}$$

Il s'agit bien évidemment d'une expression bien plus lourde, et un algorithme de calcul de cumulants dédié peut être préférable.

### B.5.3 Estimateur composite du cumulant ternaire

L'estimateur composite dont on se servira est issu des développements de la section 7.1, et correspond à une implémentation pratique de celui décrit à l'équation (7.12) :

$$\tilde{\gamma_3}(X)=\bar{\gamma_3}\left(X,X,\tilde{X}_{\mathrm{pr}}\right)+2\bar{C}\left(X,\sum_i\mathrm{Var}\left(X|\bar{\imath}\right)\right)+\overline{\sum_i\gamma_3(X|\bar{\imath})}\ . \tag{B.65}$$

On peut facilement utiliser les expressions précédemment définies pour établir les valeurs des variances de chacun des termes, ainsi que la covariance des deux derniers. Pour calculer la covariance des deux termes manquants, on va devoir utiliser les méthodes précédentes pour développer $\mathrm{Cov}\left(\bar{\gamma}_3(X,X,Y),\bar{Z}\right)$ et $\mathrm{Cov}\left(\bar{\gamma}_3(X,X,Y),\bar{C}(X,Z)\right)$.

Le développement de la première de ces deux covariances nous donne :

$$m\boldsymbol{E}\left(\overline{X^2Y}\bar{Z}\right) = (m-1)\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(X^2YZ\right)\ ; \tag{B.66a}$$

$$\begin{aligned} m^2\boldsymbol{E}\left(\overline{X^2}\bar{Y}\bar{Z}\right) = {} & (m-1)(m-2)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + (m-1)\left[\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(Y\right)\right. \\ & \left. + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right)\right] + \boldsymbol{E}\left(X^2YZ\right)\ ; \end{aligned} \tag{B.66b}$$

$$\begin{aligned} m^2\boldsymbol{E}\left(\overline{XY}\bar{X}\bar{Z}\right) = {} & (m-1)(m-2)\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right) + (m-1)\left[\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X\right)\right. \\ & \left. + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right)\right] + \boldsymbol{E}\left(X^2YZ\right)\ ; \end{aligned} \tag{B.66c}$$

$$\begin{aligned} m^3\boldsymbol{E}\left(\bar{X}^2\bar{Y}\bar{Z}\right) = {} & (m-1)(m-2)(m-3)\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + (m-1)(m-2)\left[\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right)\right. \\ & \left. + 2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right) + 2\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right) + \boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X\right)^2\right] \\ & + (m-1)\left[2\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(XYZ\right) + \boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^2Z\right) + \boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^2Y\right) + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right)\right. \\ & \left. + 2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right)\right] + \boldsymbol{E}\left(X^2YZ\right)\ . \end{aligned} \tag{B.66d}$$

Si on suppose les variables aléatoires centrées, cela nous donne :

$$\begin{aligned} \mathrm{Cov}\left(\bar{\gamma}_3(X,X,Y),\bar{Z}\right) &= \frac{m^2-3m+2}{m^3}\boldsymbol{E}\left(X^2YZ\right) - (m-1)\left[\frac{(2m-4)}{m^3}\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right) + \frac{(m-2)}{m^3}\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right)\right] \\ &= \frac{(m-1)(m-2)}{m^3}\gamma_4(X,X,Y,Z)\ . \end{aligned} \tag{B.67}$$

Effectuer le calcul avec tous les termes (variables aléatoires non centrées) permet de retrouver le même résultat. Quant à la seconde covariance, le développement nous donne :

$$m\boldsymbol{E}\left(\overline{X^2Y}\overline{XZ}\right) = (m-1)\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(XZ\right) + \boldsymbol{E}\left(X^3YZ\right)\ ; \tag{B.68a}$$

$$\begin{aligned} m^2\boldsymbol{E}\left(\overline{X^2Y}\bar{X}\bar{Z}\right) = {} & (m-1)(m-2)\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right) + (m-1)\left[\boldsymbol{E}\left(X^3Y\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(X^2YZ\right)\boldsymbol{E}\left(X\right)\right. \\ & \left. + \boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(XZ\right)\right] + \boldsymbol{E}\left(X^3YZ\right)\ ; \end{aligned} \tag{B.68b}$$

$$\begin{aligned} m^2\boldsymbol{E}\left(\overline{X^2}\bar{Y}\overline{XZ}\right) = {} & (m-1)(m-2)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(XZ\right) + (m-1)\left[\boldsymbol{E}\left(X^3Z\right)\boldsymbol{E}\left(Y\right) + \boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X^2\right)\right. \\ & \left. + \boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(XZ\right)\right] + \boldsymbol{E}\left(X^3YZ\right)\ ; \end{aligned} \tag{B.68c}$$

$$m^2\boldsymbol{E}\left(\overline{XY}\bar{X}\overline{XZ}\right) = (m-1)(m-2)\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(XZ\right) + (m-1)\left[\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(XZ\right) + \boldsymbol{E}\left(X^2YZ\right)\boldsymbol{E}\left(X\right)\right.$$
$$\left.+\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X^2Z\right)\right] + \boldsymbol{E}\left(X^3YZ\right)\ ; \tag{B.68d}$$

$$m^3\boldsymbol{E}\left(\overline{XY}\bar{X}^2\bar{Z}\right) = (m-1)(m-2)(m-3)\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(Z\right) + (m-1)(m-2)\left[2\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right)\right.$$
$$\left.+\boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X\right)^2 + 2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X\right) + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^2\right)\right]$$
$$+ (m-1)\left[2\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X^2Z\right) + \boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^3Y\right)\right]$$
$$+ (m-1)\left[2\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X^2Y\right) + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(XYZ\right)\right] + \boldsymbol{E}\left(X^3YZ\right)\ ; \tag{B.68e}$$

$$m^3\boldsymbol{E}\left(\overline{XZ}\bar{X}^2\bar{Y}\right) = (m-1)(m-2)(m-3)\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(Y\right) + (m-1)(m-2)\left[2\boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\right.$$
$$\left.+\boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X\right)^2 + 2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^2\right)\right]$$
$$+ (m-1)\left[2\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X^2Y\right) + \boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^3Z\right)\right]$$
$$+ (m-1)\left[2\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X^2Z\right) + \boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(XYZ\right)\right] + \boldsymbol{E}\left(X^3YZ\right)\ ; \tag{B.68f}$$

$$m^3\boldsymbol{E}\left(\overline{X^2}\bar{X}\bar{Y}\bar{Z}\right) = (m-1)(m-2)(m-3)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + (m-1)(m-2)\left[\boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\right.$$
$$+\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(X^3\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^2\right) + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^2\right)$$
$$\left.+\boldsymbol{E}\left(YZ\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2\right)\right] + (m-1)\left[\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2YZ\right) + \boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^3Z\right) + \boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^3Y\right)\right.$$
$$\left.+\boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X^2\right)\right] + (m-1)\left[\boldsymbol{E}\left(X^3\right)\boldsymbol{E}\left(YZ\right) + \boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X^2Z\right) + \boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X^2Y\right)\right]$$
$$+ \boldsymbol{E}\left(X^3YZ\right)\ ; \tag{B.68g}$$

$$m^4\boldsymbol{E}\left(\bar{X}^3\bar{Y}\bar{Z}\right) = (m-1)(m-2)(m-3)(m-4)\boldsymbol{E}\left(X\right)^3\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + (m-1)(m-2)(m-3)\left[\boldsymbol{E}\left(X\right)^3\boldsymbol{E}\left(YZ\right)\right.$$
$$\left.+3\boldsymbol{E}\left(XY\right)\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(Z\right) + 3\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(X\right)^2\boldsymbol{E}\left(Y\right) + 3\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right)\right]$$
$$+ (m-1)(m-2)\left[\boldsymbol{E}\left(X^3\right)\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(Z\right) + 3\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Z\right) + 3\boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(Y\right)\right.$$
$$+ 3\boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X\right)^2 + 3\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(YZ\right) + 6\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(XZ\right)\boldsymbol{E}\left(YZ\right)$$
$$+3\boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(XZ\right) + 3\boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^2\right)\boldsymbol{E}\left(XY\right)\left.\right] + (m-1)\left[3\boldsymbol{E}\left(X\right)\boldsymbol{E}\left(X^2YZ\right)\right.$$
$$+ \boldsymbol{E}\left(Y\right)\boldsymbol{E}\left(X^3Z\right) + \boldsymbol{E}\left(Z\right)\boldsymbol{E}\left(X^3Y\right) + \boldsymbol{E}\left(X^3\right)\boldsymbol{E}\left(YZ\right) + 3\boldsymbol{E}\left(X^2Y\right)\boldsymbol{E}\left(XZ\right)$$
$$3\boldsymbol{E}\left(X^2Z\right)\boldsymbol{E}\left(YZ\right) + 3\boldsymbol{E}\left(XYZ\right)\boldsymbol{E}\left(X^2\right)\left.\right] + \boldsymbol{E}\left(X^3YZ\right)\ . \tag{B.68h}$$

On peut centrer nos variables aléatoires, ce qui élimine un grand nombre de termes, nous permettant d'arriver au résultat final suivant :

$$\mathrm{Cov}\left(\bar{\gamma_3}(X,X,Y),\bar{C}(X,Z)\right) = \frac{(m-1)^2(m-2)}{m^4}\left[\gamma_5(X,X,X,Y,Z) + \gamma_3(X)\mathrm{Cov}\left(X,Z\right)\right.$$
$$\left.+\gamma_3(X,X,Z)\mathrm{Cov}\left(X,Y\right) + 2\gamma_3(X,Y,Z)\mathrm{Var}\left(X\right) + 2\gamma_3(X,X,Y)\mathrm{Cov}\left(X,Z\right)\right]\ . \tag{B.69}$$

Un résultat somme toute peu étonnant, car il s'agit d'une asymétrisation du résultat employé à la sous-section B.4.2.

# Annexe C

# Démonstrations mineures

## C.1 Energie cinétique orbitalaire des ondes planes

On va commencer par prendre une version simplifiée de la fonction d'onde de Slater, déterminant d'ondes planes, que l'on a introduit à l'équation (4.10), en ne sélectionnant qu'un unique système de spin :

$$\psi(\mathcal{R}) = \det\left(\left(e^{\boldsymbol{i}(k_{xj}x_i + k_{yj}y_i)}\right)_{i,j}\right) . \tag{C.1}$$

Appliquer les 4 translations sur l'électron $i$, de position $\vec{r}_i = (x_i, y_i)$ dans l'onde plane $j$ donne :

$$\begin{aligned}\sum_{\vec{j}} t_{\vec{r}_i\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{r}_i}e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)} &= e^{\boldsymbol{i}(k_{xj}(x_i+1)+k_{yj}y_i)} + e^{\boldsymbol{i}(k_{xj}(x_i-1)+k_{yj}y_i)} + e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}(y_i+1))} + e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}(y_i-1))} \\ &= e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)}\left(e^{\boldsymbol{i}k_{xj}} + e^{-\boldsymbol{i}k_{xj}} + e^{\boldsymbol{i}k_{yj}} + e^{-\boldsymbol{i}k_{yj}}\right) \\ &= e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)}(2\cos k_{xj} + 2\cos k_{yj}) .\end{aligned} \tag{C.2}$$

On voit donc que l'orbitale moléculaire de vecteur d'onde $\vec{k}_j = (k_{xj}, k_{yj})$ a pour énergie cinétique $-2\cos k_{xj} - 2\cos k_{yj}$.

Passons maintenant de l'orbitale moléculaire à la fonction d'onde. On a :

$$\begin{aligned}\sum_{\vec{i},\vec{j}} t_{\vec{i}\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{i}}\psi(\mathcal{R}) &= \sum_{i=1}^{N}\sum_{\vec{j}} t_{\vec{r}_i\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{r}_i}\psi\mathcal{R} \\ &= \sum_{i=1}^{N}\sum_{\vec{j}} t_{\vec{r}_i\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{r}_i}\ \det(\boldsymbol{A}(\mathcal{R})) .\end{aligned} \tag{C.3}$$

On voit bien qu'alors seule est modifiée la ligne $i$ de la matrice $\boldsymbol{A}$. On peut donc calculer ce déterminant par développement selon la ligne $i$. En notant $B_{ij}$ le $j$-ième cofacteur de la matrice $\boldsymbol{A}$ pour un développement selon la ligne $i$, on a :

$$\begin{aligned}\sum_{\vec{i},\vec{j}} t_{\vec{i}\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{i}}\psi(\mathcal{R}) &= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{ij}\sum_{\vec{j}} t_{\vec{r}_i\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{r}_i}e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} B_{ij}(2\cos k_{xj} + 2\cos k_{yj})e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)} .\end{aligned} \tag{C.4}$$

On peut intervertir les sommes. On obtient une quantité que l'on peut interpréter comme le résultat d'un développement selon la colonne $j$ d'un déterminant :

$$\begin{aligned}\sum_{\vec{i},\vec{j}} t_{\vec{i}\vec{j}}\hat{a}^\dagger_{\vec{j}}\hat{a}_{\vec{i}}\psi(\mathcal{R}) &= \sum_{j=1}^{N}(2\cos k_{xj} + 2\cos k_{yj})\sum_{i=1}^{N} B_{ij}e^{\boldsymbol{i}(k_{xj}x_i+k_{yj}y_i)} \\ &= \sum_{j=1}^{N}(2\cos k_{xj} + 2\cos k_{yj})\det(\boldsymbol{A}(\mathcal{R})) \\ &= \psi(\mathcal{R})\sum_{j=1}^{N}(2\cos k_{xj} + 2\cos k_{yj}) .\end{aligned} \tag{C.5}$$

## C.2 Décomposition des moments sur les cumulants

Le décomposition en série de Taylor de la fonction génératrice des cumulants sur un voisinage de $z = 0$ donne :

$$G_c(X)(z) = \sum_{n=1}^{\infty} \frac{\gamma_n(X) z^n}{n!} = \lim_{q \to \infty} \sum_{n=1}^{q} \frac{\gamma_n(X) z^n}{n!} \ . \tag{C.6}$$

On a donc accès à une suite de fonctions $(f_q(X))_{q \in \mathbb{N}^*}$ qui converge simplement vers $G_c(X)$, et définie par :

$$\forall q \in \mathbb{N}^*,\ f_q(X) : z \mapsto \sum_{n=1}^{q} \frac{\gamma_n(X) z^n}{n!} \ . \tag{C.7}$$

On sait que $G_q = \exp \circ G_c$. Alors, comme l'exponentielle est entière sur tout $\mathbb{C}$, on peut dire que la suite $(g_q(X))_{q \in \mathbb{N}^*}$, définie par $g_q(X) = \exp \circ f_q(X)$, converge simplement vers $G_q(X)$.

Développons alors $g_q$ en série de Taylor au voisinage de 0. On a :

$$g_q(X)(z) = \prod_{n=1}^{q} e^{\frac{\gamma_n(X) z^n}{n!}} \tag{C.8a}$$

$$= \prod_{n=1}^{q} \sum_{k_n=0}^{\infty} \frac{\gamma_n(X)^{k_n} z^{n k_n}}{n!^{k_n} k_n!} \tag{C.8b}$$

$$= \sum_{\vec{k} \in \mathbb{N}^q} z^{\sum_{n=1}^{q} n k_n} \prod_{n=1}^{q} \frac{\gamma_n(X)^{k_n}}{n!^{k_n} k_n!} \ . \tag{C.8c}$$

en prenant $\vec{k} = (k_n)_{n \in [\![1,m]\!]}$. Comme $f_q$ est le développement limité à l'ordre $q$ de $G_c$ au voisinage de 0, alors le développement limité à l'ordre $q$ de $G_c(X)$ et celui à l'ordre $q$ de $g_q(X)$ coincident. En particulier, si on se concentre sur le terme de rang $m$ de ce développement limité, alors on a, en posant $J_q = \{\vec{k} \in \mathbb{N}^q, \sum_n n k_n = q\}$ :

$$\frac{m_q}{q!} = \sum_{\vec{k} \in J_q} \prod_{n=1}^{q} \frac{\gamma_n(X)^{k_n}}{n!^{k_n} k_n!} \ . \tag{C.9}$$

Pour obtenir l'équation (7.30), il ne reste plus qu'à transformer ces produits en produits infinis, les vecteurs en suites presque finies, et à faire passer le facteur $q!$ de l'autre côté, ce qu'on peut faire en se servant de la convergence simple de la suite des $g_q(X)$.

# Annexe D

# Article 1 — Stochastic effective core potentials, improving efficiency using a spin-dependent core definition

Cet article a été définitivement accepté par *Physical Chemistry Chemical Physics* le 22 juin 2022, et est actuellement sous presse. Cet article est disponible sur arXiv à l'addresse https://arxiv.org/abs/ et sur HAL à l'addresse https://hal.archives-ouvertes.fr/hal-03665065

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

# Stochastic effective core potentials, improving efficiency using a spin-dependent core definition

Jonas Feldt,[a*] Antoine Bienvenu[a], and Roland Assaraf[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Numerically cheap single-core subsamplings have been used to build improved estimators for molecular properties in the variational Monte Carlo framework[1]. The resulting estimators depend only on the valence electron positions and can be thought of as an exact effective core potential for the total energy. We are proposing a spin-dependent core definition which enables exploiting these single-core subsamplings (or sidewalks) not only to decrease the variance of the estimators but also to restrict the main variational Monte Carlo dynamics to the valence region. This results mainly in a simplification of the algorithm and additionally in a gain in efficiency as illustrated on alkane chains and silicon clusters. An evaluation of the efficiency on transition metal systems is done using cobalt clusters, a gain of up to two orders of magnitude is achieved compared to a standard all-electron calculation.

## 1 Introduction

Quantum Monte Carlo (QMC) methods employ a stochastic approach to solve the Schrödinger equation. Freedom in the choice of the wave function allows to treat equally dynamic and static correlation which is exploited for the investigation of materials and excited states.[2] The $N^{3-4}$ scaling of the computational cost with the system size $N$ is very favourable compared to deterministic quantum chemistry methods. Among recent developments, the cost of multideterminant expansions and optimizing has been greatly reduced[3,4] to the point where the optimization of a geometry and all parameters of the wave function scale the same as the computation of the total energy[5].

One of the remaining challenges to establish QMC methods as highly accurate and cost effective methods is the steep scaling with the atomic number $Z$. The consequence is that empirical effective core potentials (ECPs) are widely used which introduce a bias that cannot be easily judged a priori. For example the Burkatzki-Filippi-Dolg potentials have been parametrized for Hartree-Fock and do not take into account the correlation energy.[6] It has been demonstrated that this bias is even larger for excited states properties.[7] Furthermore, the unfavourable scaling with the effective nuclear charge $Z_{\text{eff}}$ remains, $Z_{\text{eff}}^{6.5}$ for the forces in diffusion Monte Carlo (DMC).[8]

Developments addressing this challenge have been mostly focusing on improving the sampling and reducing the correlation factor. For instance in variational Monte Carlo (VMC) the dynamic can be carried out very efficiently in spherical coordinates.[9] In DMC different grids based on a spatial discretization can be used for core and valence electrons. This enables to perform small moves adapted to the core electrons and large moves adapted to the valence electrons. This results in an acceleration of about one order of magnitude for very heavy elements ($Z = 118$) and achieves a scaling of $Z^5$ for all-electron calculations .[10] Adapting the moves to the core and valence electrons can be done also without discretizing the space i.e. in the usual framework of the (overdamped) Langevin Dynamics (drift and diffusion process) performed in Diffusion Monte Carlo methods and Variational Monte Carlo. For that purpose two time steps are introduced an optimized (one small one for the core electrons and a large for the valence electrons)[11], in this last reference the correlation factor was reduced by factor 2–4 in Variational Monte Carlo (up to the Neon atom).

Recently, a core-subsampling approach (sidewalks on the core electrons) was introduced to reduce not only the correlation factor but also the fluctuations coming from the core region[1]. The cost of these sidewalks scales linearly with the system size $N$ if we use the locality of the information in the core region (e.g. the atomic orbitals are highly local). This cost is negligible compared to the $\mathcal{O}(N^3)$ cost for the main walk.

Such an algorithm is equivalent to an on-the-fly construction of an *exact* effective core potential, because the resulting estimator is independent of the positions of the core electrons. We are proposing to use these sidewalks or subsamplings not only to remove the fluctuations coming from the core region but also alleviate the main walk to treat only the valence region. This can be done without loss of ergodicity thanks to a spin-dependent definition of the core-valence separation[11] and the use of the final

[a] *Laboratoire de Chimie Théorique - UMR7616, Sorbonne Université & CNRS, 75005 Paris, France*
[*] E-mail: jfeldt.theochem@gmail.com

configuration of the sidewalk to advance the main walk. These updates improve to some degree the variance but mainly the ergodicity, the correlation factor within the main walk and the computational time.

Limiting the main walk to the valence electrons results in a simplified algorithm,also because there is no more codependency of the parameters of this method (time steps and size of the sidewalks) in the core and valence regions. In particular the optimal size of the sidewalks can be now obtained automatically for any cluster using a single atom calculation (eqn 4) avoiding human time consuming optimizations. Besides efficiency comparisons using alkanes chains and silicon clusters, we apply the method on cobalt clusters to evaluate its efficiency on a transition metal system.

## 2 Algorithm

With $X$ being a random variable, for example the potential or the local energy, we are partitioning the variance based on the variance decomposition theorem

$$V(X) = \mathbb{E}(V(X|\Omega)) + V(\mathbb{E}(X|\Omega)) \quad (1)$$

into a contribution from the core electrons $\mathbb{E}(V(X|\Omega))$ and the remaining part, which stems from the valence electrons.[1] We remind that $\mathbb{E}(X)$ stands for the expectation value of $X$ (interpreted as a statistical average in the Monte Carlo framework) and $V$ stands for the variance $V(X) \equiv \mathbb{E}(X^2) - \mathbb{E}(X)^2$ which measures the statistical fluctuations.

The condition $\Omega$ is a constraint in real space which allows the core electrons alone to move within their core region while the valence electrons are frozen. Performing an average under this constraint produces the conditional expectation value $\mathbb{E}(X|\Omega)$ and the conditional variance $V(X|\Omega) = \mathbb{E}(X^2|\Omega) - \mathbb{E}(X|\Omega)^2$. $\mathbb{E}(X|\Omega)$ is an improved estimator of $X$ because it has the same expectation value (thanks to the law of total expectation $\mathbb{E}(\mathbb{E}(X|\Omega)) = \mathbb{E}(X)$) and a reduced variance $V(\mathbb{E}(X|\Omega) < V(X)$. As a by-product, using the estimator $\mathbb{E}(X|\Omega)$ does not change the accuracy. Note that $\mathbb{E}(X|\Omega)$ does not depend on the core electrons positions since they are averaged, it depends only on the valence positions.

$\Omega_c^{(i)}$ being the constraint allowing only the core electrons of the atom $i$ to move, the improved estimator

$$\tilde{X} = X + \lambda \sum_i \left( \mathbb{E}(X|\Omega_c^{(i)}) - X \right) \quad (2)$$

eliminates the fluctuations of the core electrons completely when the core regions are independent[1] (while not modifying the expectation value since we add a zero-expectation value term). The coefficient $\lambda$ can be determined to minimize the variance [*]. The conditional expectation values $\mathbb{E}(X|\Omega_c^{(i)})$ are evaluated by sidewalks with a number of $M_s$ steps on the core electrons, using the ergodic theorem

$$\mathbb{E}(X|\Omega_c^i) = \lim_{M_s \to \infty} \frac{1}{M_s} \sum_{k=1}^{M_S} X^k \quad (3)$$

where $X^k$ is the value of $X$ for the $k^{th}$ step of the sidewalk.

One natural idea would be to use the sidewalks not only to lower the variance of the estimators but also to advance the main walk, moving the core electrons of the main walker to the last positions of the core sidewalk. With such updating scheme the sidewalk can be considered as a part of the main walk and as such can be called a "subwalk". In this context $\Omega_c^{(i)}$ would freeze beside the valence electrons the first $i-1$ cores in their new configuration and all remaining cores (excepted $i$) in their old configuration. This approach should make it possible to restrict the main walk to only the valence electrons. In practice the process would not be ergodic with the definition of the core region in Ref. 1. In this reference the core region was the largest nucleus-centred ball containing $n_c$ closest electrons (regardless their spin) to the nucleus. If there is no exchange between the valence and the core region the total spin of the core region is frozen. Hence, we will apply this idea with a different definition of the core and valence regions.

We introduce a spin-dependent core constraint, i.e. two core regions, one for the $\alpha$ electrons and one for the $\beta$ electrons. If we order the $\alpha$ electrons by their distance to the nucleus, the first $\alpha$ valence electron is on the nucleus-centered sphere which defines the boundary of the $\alpha$ core region (see Fig. 1). Of course the same definition applies to the $\beta$ core region.

Now we define equivalently the $\alpha$ (respectively $\beta$) valence region to be outside the (nucleus-centred) ball with the radius defined by the last $\alpha$ (respectively $\beta$) core electron (see Fig. 1). This core-valence separation definition is equivalent to the one introduced in reference[11].

Core and valence regions overlap because the space between the last core and the first valence electron can be explored by both core and valence electrons. With these definitions performing sidewalks in the core regions and a main walk restricted to the valence region should be now ergodic, because these regions have different radii which can evolve and exchange, ensuring for example the sampling of the spin in a given volume of the space. Note that with this new definition the core electrons have always the same indices (in particular we do not need anymore to reorder the electrons at the beginning of each sidewalk).

In summary, one iteration of the algorithm consists of the following steps:

1. Independent sidewalks for all cores.
2. Update of the full system after each core sidewalk.
3. Main walk only for the valence electrons .
4. Update of the full system with new valence positions.
5. Computation of the improved estimator $\tilde{X}$.

The update of the configuration after each core sidewalk is a small modification of the algorithm presented in Ref. 1 but it has

[*] Writing $\tilde{X} = X + \lambda C$, the variance can be expanded as $V(\tilde{X}) = V(X) + 2\lambda \ \mathrm{cov}(X,C) + \lambda^2 V(C)$ which is a quadratic function of $\lambda$ easy to minimize after computing the parameters (two variances and one covariance).

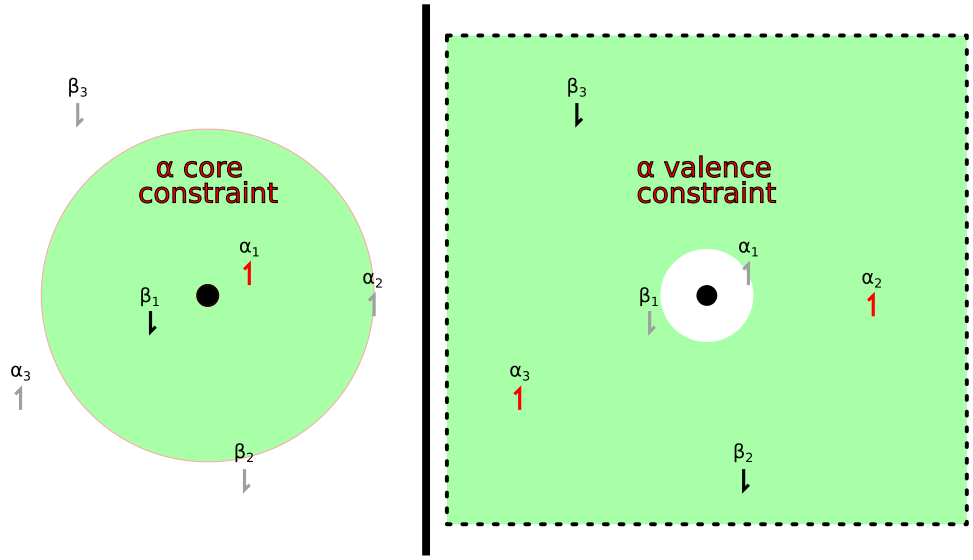


**Fig. 1** Representation of the $\alpha$ core and valence constraint for a nucleus surrounded by 3 $\alpha$ and 3 $\beta$ electrons. The electrons are labeled according to their distance to the nucleus. The constrained electrons are shown in red and can move freely within the green area, left for the core electrons and right for the two $\alpha$ valence electrons. The frozen electrons within the core sidewalk (left) and the valence main walk (right) are shown in gray. The valence space is infinite but here only shown within the dashed frame. Equivalent constraints apply to the $\beta$ electrons.

several consequences. The first is that we do not need to move the core electrons in the main walk as stated in step 3. This implies a further simplification of the dynamics which consisted in moving one electron at a time using two drift and diffusion processes with different time steps (a small one for the core region and a large one for the valence region) 1. Only one (large) time step for the valence region can be now used, avoiding the small time step move which was only efficient in the core region.

In this modified dynamics the treatment of core and valence electrons is now similar except that we do many moves per iteration for the core electrons and only one move for the valence electrons. With the new algorithm the estimator (eqn (2)) is formally the same except that $\Omega_c^{(i)}$ for two different indices $i$ represents different frozen configurations (i.e. the core electrons are not the same after one update). Note that if different core regions are independent the core updates do not modify this estimator and its variance, only the different definition of the core does.

The combination of all sidewalks (including updates) constitutes in itself an efficient (i.e. quickly decorrelating) move for all electrons. Hence the sidewalks are not only used to reduce the variance but also correlation.

## 3 Numerical Results

Simulations have been carried out for alkane chains and silicon clusters to compare the performance of the improved algorithm in this work which uses updates and a spin-dependent core definition with the previous algorithm[1] without updates and a purely spatial core definition. Due to the updates the estimator in eqn (2) is different from the previous estimator. The computational cost can be compared using the expression $\zeta = Vct$ with the variance $V$, the correlation factor $c$ and the computational time of a single step $t$. The gain in computational efficiency $G$ is defined as the inverse reduction of the computational cost. We carried out simulations for small model systems ($CH_4$ and $Si_1$) fitting the convergence of the variance with $M_s$ to estimate the reduction for very long sidewalks. This reduction in the limit of large $M_s$ is transferable to large systems where the optimal $M_s^*$ itself is increasingly large. The correlation factor is also transferable in this limit[1]. Therefore, we can estimate the gain in the computational efficiency (i.e. reduction of the cost) in the asymptotic limit of many atoms $G^\infty$ as 38 (alkanes) and 774 (Si clusters) compared to a regular all-electron VMC simulation. Compared to the previous algorithm it is an improvement by a factor of 3.7 (alkanes) and 3.9 (Si clusters).

We can further analyze the gain $G^\infty$ by separating according to the definition of the computational cost into contributions due to the variance $G_V^\infty$, the correlation factor $G_c^\infty$ and the computational time $G_t^\infty$. We begin with the estimator eqn (2) with a fixed value of $\lambda = 1$. The correlation factor was already small with the previous algorithm (1.8 for carbon and 1.5 for silicon), consequently we observe only an additional gain of 10% (alkanes) and 19% (Si clusters). This supports an improvement in ergodicity coming from the core updates but also, the new optimal time step for the valence dynamic is larger especially for the silicon clusters. However, we observe also a small loss in the variance by about 10% (alkanes) and 4% (silicons). To understand it, we carried out additional simulations without updates but with a spin-dependent core definition. Because we observe the same loss we attribute this loss to the new core definition. It turns out that contrary to the previous core definition the value $\lambda = 1$ is not optimal for the variance. This is a signature of a diminished core-valence separation coming from the new definition of these regions. Indeed $\lambda = 1$ is the optimal value to lower the variance when the valence and the core regions are independent, since in this limit eqn (2) cancels fully the fluctuations coming from the cores of the molecule. Here optimizing $\lambda$ recovers the loss in the vari-

ance and we even observe a small gain of about 26% (alkanes) and 11% (silicons). This reduced core separation should not be too surprising, since in rare cases a valence electron of one spin can be closer to the nucleus than a core electron of the opposite spin. Optimizing $\lambda$ however decreases the correlation factor by 8% (alkanes) or 21% (silicons) in comparison to Ref. 1. Nevertheless, this is compensated for by the gain in the variance. We will focus in the following on the estimator with optimal $\lambda$.

The appreciable effect on the overall gain is in the computational time $t$ which is improved by a factor of 2.9 (alkanes) and 3.4 (Si clusters). The cubic scaling with the system size (in the asymptotic limit of a large number of atoms) is the same for the new algorithm but the prefactor is reduced. On one hand, the computational time for the main walk is reduced because only the valence electrons need to be taken into account. On the other hand, an additional cost (scaling $\mathcal{O}(N^3)$) is added to the sidewalk which stems from the Sherman-Morrison update of the full configuration at the end of each sidewalk. Nevertheless, this is more efficient because we require less updates for more electrons: instead of one update for each single electron in the main walk we are carrying out only one update for all electrons of a core at once. The formulas for the update of $n_c$ electrons at once are shown in Appendix A. Also, we are exploiting the locality within the core subsystems for an efficient update and the $O(N^2)$ cost for the update of one core comes with a very low prefactor.

Next, we are looking at the results of simulations for alkanes up to 40 carbon atoms shown in Fig. 2 and for silicon clusters up to 24 atoms in Fig. 3 (red lines). This corresponds to about 350 electrons for the largest systems. The maximal gain for a given system is obtained by determining the optimal number of steps $M_s^*$ in the core sidewalk which is a balance between the reduction of the variance and the additional computational cost. A simple formula to determine $M_s^*$ is shown in eqn (4) in the computational details (see Sec. 4). For comparison we present also the results of the previous algorithm[1] (blue lines). It can be seen that the gain of the new algorithm is always at least as good as the previous gain. We observe similar results for alkane chains up to 20 carbon atoms and silicon clusters up to 5 atoms. For larger systems one can see an increasingly larger gain with the improved algorithm. For the largest systems studied here we gain a factor of 1.9 ($C_{40}H_{82}$) and 2.0 ($Si_{24}$).

The improvement for the alkanes comes mostly from the reduced computational time $t$ and additionally for more than 30 carbon atoms the updating scheme reduces the variance (in the optimal regime). This suggests that the convergence towards the asymptotic limit of many atoms is accelerated. For the silicon clusters the situation is different. Because of the large number of core electrons $n_c = 10$, the number of electrons in the main walk is strongly reduced (and consequently the computational time $t_0$, see eqn (4)). Therefore, the optimal sidewalk length $M_s^*$ is shorter which results in a reduced optimal gain in the variance multiplied by a factor of 0.35 (one atom) to 0.7 (24 atoms). With increasing system size this gain will approach its asymptotic value. The reduced optimal gain in the variance is counterbalanced by an even larger gain in the computational time $t$ of up to 4.1 times which leads to an overall improvement of a factor 2 for 24 atoms.

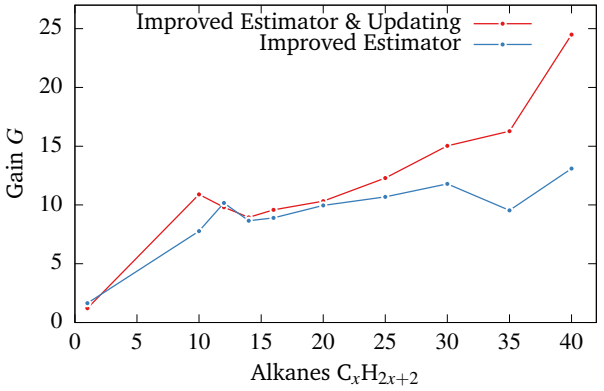


**Fig. 2** The gain $G$ for alkane chains comparing different algorithms.

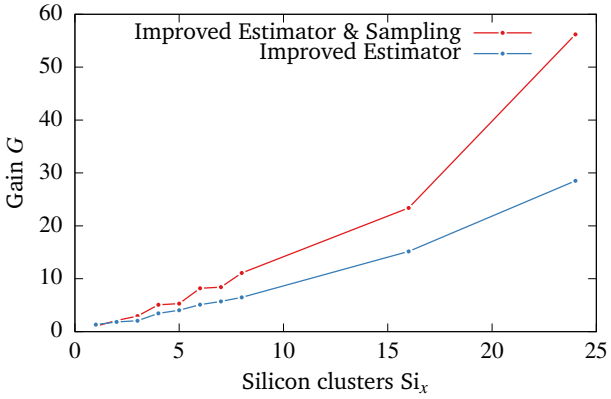


**Fig. 3** The gain $G$ for silicon clusters comparing different algorithms.

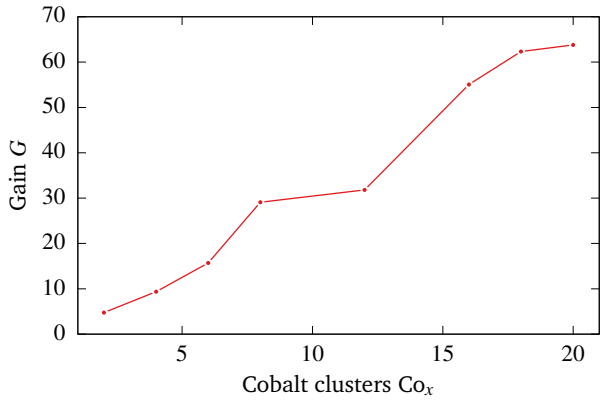


**Fig. 4** The gain $G$ for cobalt clusters comparing with a regular all-electron calculation.

Last, we are evaluating the efficiency of the new algorithm (compared to regular all-electron calculations) for transition metals on the example of clusters of cobalt atoms (hexagonal $P6_3/mmc$ space group[12]). In the asymptotic limit of large clusters the gain $G^\infty$ is given by about 2270 including a gain in the variance of about 130. The efficiency for clusters consisting of up to 20 cobalt atoms is shown in Fig. 4. The gain is still far from its asymptotic value. Nevertheless, we can obtain already an improvement of the efficiency by one order of magnitude for only four cobalt atoms and by a factor of 64 for the largest system $Co_{20}$. The efficiency for all three systems studied here is compared in Table 1 both for the asymptotic gain $G_\infty$ and for a gain for twenty atoms of either carbon, silicon or cobalt $G_{20}$. The gain in the asymptotic limit is increasing drastically with $Z$ by two orders of magnitudes going from $Z=6$ to 27. The gain for a medium-sized system of 20 atoms increases as well with $Z$ but at a slower rate (a factor 6.4 from $Z=6$ to $Z=27$).

**Table 1** The atomic number $Z$ as well as the number of core electrons $n_c$ for the alkanes, silicon and cobalt clusters and the gain in efficiency (compared to regular all-electron calculations) for twenty atoms $G_{20}$ and the asymptotic limit of many atoms $G_\infty$. The local energies $E_L$ and $\tilde{E}_L$ are given in atomic units with their statistical significance in brackets.

| | $CH_4$ | Si | Co |
|---|---|---|---|
| $Z$ | 6 | 14 | 27 |
| $n_c$ | 2 | 10 | 18 |
| $G_{20}$ | 10 | 40 | 64 |
| $G_\infty$ | 38 | 774 | 2270 |
| | $CH_4$ | Si | $Co_2$ |
| $E_L(\sigma)$ | -40.199(8) | -287.833(30) | -2750.908(46) |
| $\tilde{E}_L(\sigma)$ | -40.187(2) | -287.797(4) | -2750.822(31) |
| | $C_{20}H_{42}$ | $Si_{24}$ | $Co_{20}$ |
| $E_L(\sigma)$ | -781.011(37) | -6910.421(249) | -27507.3(1.8) |
| $\tilde{E}_L(\sigma)$ | -781.039(12) | -6910.470(33) | -27509.9(0.3) |

## 4 Computational Details

The main walk and the sidewalks are carried out with a drift and Brownian diffusion (overdamped Langevin process). Two time steps are used, $\tau_c$ and $\tau$, for the core and the valence region respectively. These two parameters and the number of iterations $M_s$ of any one-core sidewalk have to be optimized for a minimal cost $\zeta = Vct$ (we remind that $V$ is the variance, $c$ the correlation factor and $t$ the computational time for one iteration). These three parameters can be derived from a small number of simulations on very small systems (e.g. isolated atoms or $CH_4$ for alkanes). The time steps are directly transferable to larger systems, and the variance gain as a function of $M_s$ is also transferable[1]. For a given chemical element the reduction of the variance is a linear function of $1/M_s$ leading to $r_V = \tilde{V}/V = r_\infty + a/M_s$. These transferability considerations allow defining a simple formula for the optimal choice for $M_s$ (see appendix C of Ref. 1)

$$M_s^* = \sqrt{\frac{a}{r_\infty}\frac{t_0}{t_c}} \tag{4}$$

where $t_0$ is the computational time of one iteration excluding the sidewalks, i.e. the time to do a single walk on all the valence electrons and all the updates (it is scaling as $O(N^3)$). $t_c$ is the CPU time for all the core sidewalks (of the same element) with $M_s = 1$ (scaling as $O(N)$). This formula is valid with the assumption that the correlation factor in the main walk (when sampling the improved estimator) does not depend on $M_s$, which is obviously true for sufficiently large $M_s$. Without the updating scheme the correlation factor converges too slowly[1] to apply eqn (4) for small systems and tedious optimizations of $M_s$ have to be performed. We have checked that with the updating scheme this is no longer the case and eqn (4) can be applied to a small number of atoms (in all our calculations $M_s^* > 15$, see figures 5 and 6).

The optimal time steps are shown in Table 2. We found that

**Table 2** The optimal time steps for core and valence for alkane chains and silicon clusters

| | Alkanes | Silicon clusters |
|---|---|---|
| $\tau_c$ | 0.004 | 0.007 |
| $\tau$ | 0.8 | 1.8 |

the spin-dependent core definition has a negligible effect on the optimal value of $\tau_c$, the latter can be taken directly from Ref. 1. The optimal value of $\tau$ is larger, which is not surprising since it is now exclusively used for the valence electrons. The computational cost to move the valence electrons is below 10% of the total CPU time for silicon clusters even for $Si_{32}$. This is because most of the variance comes from the core region which has to be consequently sampled much more extensively. With these time steps we found an acceptance probability for the valence move of about 0.5 for silicon and 0.66 for carbon.

The value of $M_s^*$ for the alkanes are shown in Fig. 5 and for the silicon clusters in Fig. 6. One can observe that the linear scaling regime of $M_s^*$ is quickly reached for about 70 electrons for both alkanes and silicon clusters at which point a simple linear extrapolation can be used for even larger systems. Additionally, the simplicity of eqn (4) allows to determine the optimal value $M_s^*$ also for smaller systems where one cannot rely on the cubic scaling of $t_c$.

The wave function and the Jastrow factor have been taken from Ref. 1 and have been generated based on an SCF calcula-

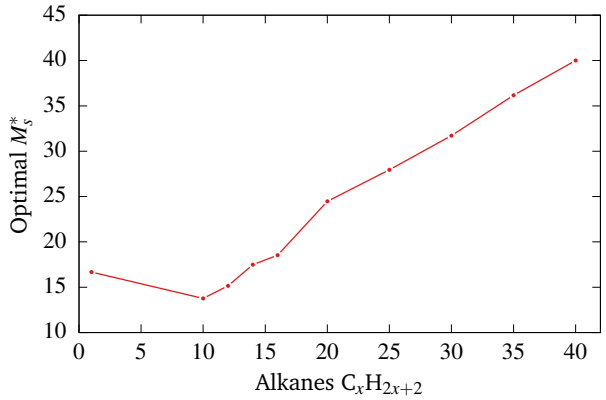


**Fig. 5** The optimal sidewalk length $M_s^*$ for alkane chains obtained from Eq. (4) .

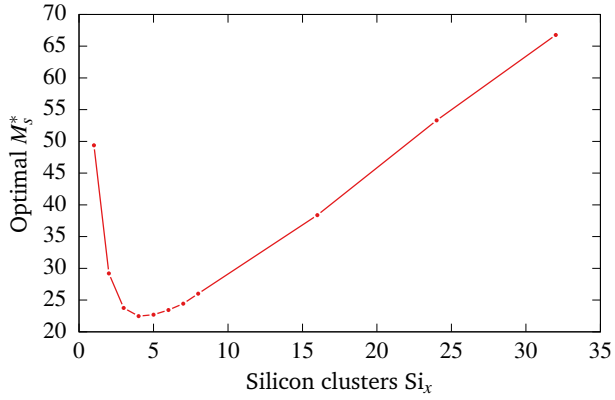


**Fig. 6** The optimal sidewalk length $M_s^*$ for silicon clusters obtained from Eq. (4).

tion carried out with Quantum Package.[13] The wave function for the cobalt clusters has been generated in the same manner. The very simple Jastrow factor ensures the electron-electron cusp condition. A Slater atomic orbital basis set[14] has been used with a TZP basis for the alkanes and SZ for the silicon and cobalt clusters which is expanded by a large sum of Gaussian functions for Quantum Package.

## 5 Conclusion

In this paper we are extending the stochastic ECP approach originally proposed to improve the estimator (removing fluctuations coming from core region using sidewalks) to the main dynamics itself: here the main walk moves only the valence electrons as we would expect in a complete ECP formalism, while the sidewalks focus exclusively on the core electrons. Key has been to update the system at the end of each core sidewalk: to maintain (and even improve) ergodicity we replaced the purely spatial core definition with a spin-dependent one. Compared to the previous algorithm the number of parameters (time steps) is reduced and the determination of the length of the sidewalk is extracted from calculations on a single atom which avoid a tedious optimization for large systems. The observed additional gain in efficiency (2–4) is growing with the system size as we observe an accelerated convergence towards the asymptotic limit of large systems. Tests include a transition metal (cobalt clusters). Large gains are observed with respect to traditional all-electron calculations (a factor 64 for 20 cobalt atoms) but are still small compared to the asymptotic gain (a factor 2270 for a very large cluster of cobalt atoms). This suggests that there is a large room of improvement, a keypoint of future developments would be to further improve the efficiency of the core sidewalks for the gain to converge much faster to the asymptotic limit. A natural idea would be for example to adapt this method to many shells. One of the most interesting perspective is the generalization to the diffusion Monte Carlo approach thanks to the efficient small time step dynamics in the core region and the following efficient updating of the full configuration of the electrons.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

J. F. acknowledges the Deutsche Forschungsgemeinschaft (DFG) for financial support (Grant FE 1898/1-1).

## A Updating the logarithmic gradient of the determinants

In Ref. 1 (appendix A) we described a method to update the determinants and its derivatives when a few electrons are moved, using a matrix $D$ representing the logarithmic gradient of the Slater matrix with respect to the atomic orbitals coefficients. Now we also have to update the matrix $D$ at the end of the sidewalk when a few (core) electrons have moved.

Let us first remind how to update the Slater determinant $\Phi$

when a few electrons are moved

$$\Phi = \det(XC) \tag{5}$$

where $X$ is the $N \times p$ matrix of atomic orbitals and C the $p \times N$ matrix of coefficients, $N$ and $p$ being respectively the number of electrons and molecular orbitals. The drift, the local energy and many other possible quantities depending on the electron positions involve logarithmic derivatives of $\Phi$ with respect to $X$.

$$\partial_\lambda \ln \Phi(X) = \mathrm{tr}(D\,\partial_\lambda X) \tag{6}$$

where

$$D \equiv C(XC)^{-1} \tag{7}$$

eqn (6) can be seen as the application of the chain rule involving $D$, the rectangular matrix representing the logarithmic gradient of $\Phi$ with respect to $X$. Note that eqn (6) must be valid for any parameter $\lambda$ and can thus be seen as the definition of $D$. If we are going to move only a few electrons within a core sidewalk, $X$ is modified in $X'$ which differs from $X$ by a few lines, $D(X)$ must be replaced by $D' = D(X')$ with efficient formulas.

First, we define the operator $P$ which applied on the left selects those lines, $PX'$ are the lines which may differ from the lines of $PX$. We also define the operator $Q^T$ which applied on the right of $PX$ or $PX'$ removes the zero columns (atomic orbitals which are zero because the electrons selected by $P$ are out of range). $P$ and $Q$ can be written in terms of rectangular matrices containing zeroes and ones[4]. $\bar{X}$ is the matrix of atomic orbitals within a subsystem and a submatrix of $X'$

$$\bar{X} \equiv PX'Q^T.$$

We obtain the changed Slater determinant $\Phi(X')$

$$\begin{aligned}\Phi(X') &= \det(XC)\det(PX'C(XC)^{-1}P^T)\\ &= \det(XC)\det(PX'Q^TQC(XC)^{-1}P^T)\\ &= \det(XC)\det(\bar{X}\bar{C})\end{aligned} \tag{8}$$

where $\bar{C}$ is a submatrix of $D = C(XC)^{-1}$.

Eqn (8) performs an update of the determinant of a product of two matrices, using the determinant of a reduced matrix. The last expression of $\Phi$ depends on $X'$ and we can write

$$X = P^TPX + (1 - P^TP)X' \tag{9}$$

where $P^TP$ represents the projector on the space spanned by the lines which have been modified. We note that the final expression should not depend on $PX$. Introducing

$$\begin{aligned}\bar{\alpha} &\equiv (\bar{X}\bar{C})^{-1}\\ \bar{D} &\equiv \bar{C}(\bar{X}\bar{C})^{-1}\end{aligned}$$

the logarithmic derivative of eqn (8) is

$$\begin{aligned}\partial_\lambda \ln\Phi(X') &= \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(\bar{D}\,\partial_\lambda \bar{X}) + \mathrm{tr}(\bar{\alpha}\bar{X}\,\partial_\lambda \bar{C})\\ &= \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(\bar{D}\,\partial_\lambda \bar{X})\\ &- \mathrm{tr}(\bar{\alpha}\bar{X}QD\,\partial_\lambda X\,DP^T)\end{aligned} \tag{10}$$

which should also be $\mathrm{tr}(D'\,\partial_\lambda X')$ so that $D'$ can be obtained by identification. First using the cyclic property of the trace we obtain

$$\begin{aligned}\partial_\lambda \ln\Phi(X') = \quad & \mathrm{tr}(D\,\partial_\lambda X) + \mathrm{tr}(Q^T\bar{D}P\,\partial_\lambda X') -\\ & \mathrm{tr}(DP^T\bar{\alpha}\bar{X}QD\,\partial_\lambda X).\end{aligned} \tag{11}$$

Since this expression should not depend on $P\,\partial_\lambda X$ we have

$$\begin{aligned}\partial_\lambda \ln\Phi(X') = \quad & \mathrm{tr}(D\,\partial_\lambda X') + \mathrm{tr}(Q^T\bar{D}P\,\partial_\lambda X') -\\ & \mathrm{tr}(DP^T\bar{\alpha}\bar{X}QD\,\partial_\lambda X')\end{aligned} \tag{12}$$

which can be also found using eqn (9). By identification we finally find the (Sherman Morrison) formula to update the logarithmic gradient

$$D' = D - DP^T\bar{\alpha}\bar{X}QD + Q^T\bar{D}P. \tag{13}$$

In the second term we select a few columns $(DP^T)$ and lines $(QD)$ of $D$ and build the product with the small rectangular subsystem matrix $\bar{\alpha}\bar{X}$. The cost for the update comes with a very small prefactor because we are exploiting the locality of the subsystem (thanks to the matrix $Q$ which selects a few orbitals). A bit more efficient formula can obtained as follows

$$\begin{aligned}QD' &= QD - \bar{D}\bar{X}QD + \bar{D}P\\ QD'P^T &= QDP^T - \bar{D}\bar{X}QDP^T + \bar{D}.\end{aligned}$$

Using that $\alpha\bar{X}QDP^T = P^TP$ we can decompose $D'$ as follows

$$D' = D(1 - P^TP) - DP^T\bar{\alpha}\bar{X}QD(1 - P^TP) + Q^T\bar{D}P \tag{14}$$

## B Updating a Jastrow-Slater Function

We define a Jastrow-Slater function

$$\Psi = J\Phi = e^U\Phi \tag{15}$$

where $\Phi$ is a Slater determinant and $J$ the Jastrow factor. We need to update $U$ when a few electrons move but also the spatial derivatives of $U$ which are involved in the drift. This is trivial since the new value of $U$ can be written as

$$U' = U + (U' - U) \tag{16}$$

If $U$ is a sum of pairwise interactions the only term to be computed is $U' - U$ which involves only a few pairs of electrons. The computational cost scales as $O(N)$ which can be reduced to $O(1)$ if the pairwise interactions are short range. This was the case for the silicon and cobalt clusters because we used a cutoff ( $r_{\mathrm{cutoff}} = 7.07$ au). The same discussion holds for the spatial deriva-

tives of $U$.

## References


1 J. Feldt and R. Assaraf, *J. Chem. Theory Comput.*, 2021, **17**, 1380–1389.

2 J. Feldt and C. Filippi, *Quantum Chemistry and Dynamics of Excited States*, John Wiley & Sons, Ltd, 2020, pp. 247–275.

3 B. K. Clark, M. A. Morales, J. McMinis, J. Kim and G. E. Scuseria, *J. Chem. Phys.*, 2011, **135**, 244105.

4 C. Filippi, R. Assaraf and S. Moroni, *J. Chem. Phys.*, 2016, **144**, 194105.

5 R. Assaraf, S. Moroni and C. Filippi, *J. Chem. Theory Comput.*, 2017, **13**, 5273–5281.

6 M. Burkatzki, C. Filippi and M. Dolg, *J. Chem. Phys.*, 2007, **126**, 234105.

7 A. Scemama, M. Caffarel, A. Benali, D. Jacquemin and P.-F. Loos, *Results Chem.*, 2019, **1**, 100002.

8 J. Tiihonen, R. C. Clay and J. T. Krogel, *J. Chem. Phys.*, 2021, **154**, 204111.

9 C. J. Umrigar, *Phys. Rev. Lett.*, 1993, **71**, 408–411.

10 K. Nakano, R. Maezono and S. Sorella, *Phys. Rev. B*, 2020, **101**, 155106.

11 D. Bressanini and P. J. Reynolds, *J. Chem. Phys.*, 1999, **111**, 6180–6189.

12 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. a. Persson, *APL Mater.*, 2013, **1**, 011002.

13 Y. Garniron, T. Applencourt, K. Gasperich, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J.-P. Malrieu, M. Véril, M. Caffarel, P.-F. Loos, E. Giner and A. Scemama, *J. Chem. Theory Comput.*, 2019, **15**, 3591–3609.

14 E. V. Lenthe and E. J. Baerends, *J. Comput. Chem.*, 2003, **24**, 1142–1156.

# Annexe E

# Article 2 — Systematic lowering of the scaling of Monte Carlo calculations by partitioning and subsampling

Cet article a été définitivement accepté par *Physical Review E* le 5 juillet 2022, et est actuellement sous presse. Cet article est disponible sur arXiv à l'addresse https://arxiv.org/abs/2205.00677 et sur HAL à l'addresse https://hal.archives-ouvertes.fr/hal-03656059.

# Systematic lowering of the scaling of Monte Carlo calculations by partitioning and subsampling

Antoine Bienvenu [1], Jonas Feldt [1], Julien Toulouse [1,2] and Roland Assaraf [1,*]
[1] *Laboratoire de Chimie Théorique, Sorbonne Université and CNRS, F-75005 Paris, France*
[2] *Institut Universitaire de France, F-75005 Paris, France*

(Received 9 February 2022; accepted 5 July 2022; published 1 August 2022)

We propose to compute physical properties by Monte Carlo calculations using conditional expectation values. The latter are obtained on top of the usual Monte Carlo sampling by partitioning the physical space in several subspaces or fragments, and subsampling each fragment (i.e., performing side walks) while freezing the environment. No bias is introduced and a zero-variance principle holds in the limit of separability, i.e., when the fragments are independent. In practice, the usual bottleneck of Monte Carlo calculations—the scaling of the statistical fluctuations as a function of the number of particles $N$—is relieved for extensive observables. We illustrate the method in variational Monte Carlo on the two-dimensional Hubbard model and on metallic hydrogen chains using Jastrow-Slater wave functions. A factor $\mathcal{O}(N)$ is gained in numerical efficiency.

DOI: 10.1103/PhysRevE.106.025301

## I. INTRODUCTION

Many domains of physics involve large dimensional integrals which can be computed efficiently with Monte Carlo methods, e.g., statistical physics [1], quantum physics applied to molecules and solids [2], or nuclear physics [3]. Monte Carlo methods reinterpret the energy or other properties as the expectation value of a random variable $O$ over a probability distribution $\pi$ on a configuration space $\Omega$,

$$\mathbb{E}(O) = \int_{x\in\Omega} O(x)\pi(x)dx. \tag{1}$$

Typically, the configuration $x$ corresponds to the $3N$ coordinates of the particles in physical space, but it can also correspond to the $N$ trajectories of the particles in the path-integral formulation of quantum mechanics. The probability distribution $\pi$ depends on the context. For example, in equilibrium statistical physics, $\pi$ is the Gibbs distribution. In variational Monte Carlo (VMC), $\pi = \Psi^2$ is the probability density of a wave function $\Psi$, and if $O = (H\Psi)/\Psi$ is the local energy for a given Hamiltonian $H$, then $\mathbb{E}(O)$ is the variational energy. Expectation values are computed using the ergodic theorem which states that the integral can be written as a time average, $\mathbb{E}(O) = \lim_{M\to\infty}(1/M)\sum_{i=1}^{M} O(x^i)$, where the sequence of $M$ configurations $(x^i)$ is built from a $\pi$-invariant ergodic stochastic process (usually a Markov chain). The sequence $(x^i)$ is called a sample of the distribution $\pi$.

A bottleneck of Monte Carlo methods comes from the statistical fluctuations which usually grow with the system size, as measured by the number of particles $N$. For a sample of sufficiently large size $M$, the statistical uncertainty $\sigma$ on the estimation of $\mathbb{E}(O)$ is

$$\sigma = \sqrt{\frac{V(O)c}{M}}, \tag{2}$$

where $V(O) = \mathbb{E}(O^2) - \mathbb{E}(O)^2$ is the variance of $O$ and $c > 1$ is a correlation factor which takes into account that the configurations are not fully independent. According to Eq. (2), reaching a given precision $\sigma$ requires a CPU time $t_M = Mt_1$ proportional to both the time $t_1$ of performing one step of the sampling and to the variance $V(O)$. The numerical efficiency of the method can then be measured by the asymptotically $M$-independent quantity

$$\sigma^2 t_M = V(O)ct_1, \tag{3}$$

which should be as small as possible for maximal efficiency. In the present paper, we will not be concerned about the correlation factor $c$ which sometimes diverges with $N$ (e.g., near criticality). A large corpus of work is devoted to reducing its scaling as a function of $N$, such as parallel tempering based methods (see, e.g., Refs. [4,5]). Equation (3) indicates a more crucial double penalty of Monte Carlo methods for large systems: Both $t_1$ and $V(O)$ grow with system size $N$. This double penalty is for example at the origin of the main bottleneck in computing the VMC energy of a fermionic system in real space [2,6]. Evaluating the wave function involves indeed calculating a Slater determinant of order $\mathcal{O}(N\times N)$ which costs $t_1 = \mathcal{O}(N^3)$ while the variance is typically extensive, $V(O) \propto N$, thus raising the scaling of the overall cost to $\mathcal{O}(N^4)$. This scaling is still larger than some deterministic methods such as the celebrated Kohn-Sham density-functional theory which scales as $\mathcal{O}(N^3)$ for a spatially delocalized (i.e., metallic) system [7].

The extensivity of the variance has a physical origin. A large system can in general be approximated by a collection of independent fragments. This ideal case corresponds to the separability limit where the random variable $O$ is the sum of independent variables $O_k$ on each fragment indexed by $k$, i.e., $O = \sum_k O_k$, and the variance is then $V(O) = \sum_k V(O_k) \propto N$. It is possible to reduce considerably the variance using an improved estimator $\tilde{O}$ built from the approximate solution of a partial differential equation [8–10]. But this type of improved

*assaraf@lct.jussieu.fr

 

estimator is still a sum of independent random variables in the separability limit, i.e., $\tilde{O} = \sum_k \tilde{O}_k$, and thus does not change the scaling with respect to $N$ but only reduces the prefactor [11].

To reduce the global computational scaling, a common and obvious strategy is to reduce the cost of the sampling. Some distributions $\pi$ can be sampled with a linear-scaling algorithm, i.e., $t_1 = \mathcal{O}(N)$, reducing the overall cost to an ideal scaling $\mathcal{O}(N^2)$. One can for example try to use the sparsity of the Slater matrix when localized Wannier functions are used [12]. But such sparsity is highly dependent on the physics of the system, and does not hold for a metallic system. In addition, this linear scaling is only theoretical because of a memory-access slowdown as $N$ increases. Another strategy consists in using a stable-versus-chaos stochastic dynamics [13], but finding such a stochastic dynamic is not straightforward [14].

Here, we propose to reduce the global computational scaling by using the locality of physical observables. The idea of using the locality of information to reduce the variance is not new: The strong locality in time of the Schrödinger equation (a first-order partial differential equation in time) has for example been exploited to remove the dynamical sign problem for bosonic systems [15]. Recently, a method was proposed [16] to exploit the low correlation between different core regions in a molecule, resulting in a reduced scaling as a function of the atomic charge $Z$. The present paper exploits the fact that in an extended physical system (including a metallic system) correlations between large fragments are small. We construct an improved estimator $\tilde{O}$ with a variance having a reduced scaling with respect to $N$, without changing the scaling of $t_1$, therefore achieving a reduction of the overall computational scaling. The present paper shares the same general philosophy as other fragment-based methods (see, e.g., Refs. [17–19]). However, while the latter methods are systematic techniques to find a good compromise between a smaller computational time and a larger systematic error, in the present method the reduction of the computational scaling is done without introducing any systematic error.

## II. THEORY

A configuration of particles is written as $x = (x_j)_{j\in J}$, where $x_j$ is the $j$th coordinate and $J$ is the list of coordinate indexes. For a given configuration $x = (x_j)_{j\in J}$, we define a partition of $J$ as $p$ disjoint sublists $J_k(x) \subset J$ such that $\bigcup_{k=1}^{p} J_k(x) = J$. We then define $p$ fragments as subsets $\Omega_k(x)$ of the configuration space $\Omega$ such that for all $x' \in \Omega_k(x)$, (i) $x'$ differ from $x$ only by the coordinates indexed by $J_k$, and (ii) $\Omega_k(x') = \Omega_k(x)$. In short, $\Omega_k$ can be seen as a parameter which specifies the positions of the frozen particles in the environment of a fragment. We then introduce the following improved estimator,

$$\tilde{O} \equiv O + \sum_{k=1}^{p} \lambda_k (\mathbb{E}(O|\Omega_k) - O), \tag{4}$$

where $\lambda_k$ are constants (or more generally functions of $\Omega_k$) and $\mathbb{E}(O|\Omega_k)$ is the conditional expectation value of the random variable $O$ with respect to $\Omega_k$, defined as the random variable obtained by partial averaging of $O$ over only configurations $x' \in \Omega_k$,

$$\mathbb{E}(O|\Omega_k) \equiv \frac{\int_{x'\in\Omega_k} O(x')\pi(x')dx'}{\int_{x'\in\Omega_k} \pi(x')dx'}. \tag{5}$$

The estimator $\tilde{O}$ in Eq. (4) is always not biased, i.e., $\mathbb{E}(\tilde{O}) = \mathbb{E}(O)$. Indeed $\mathbb{E}(O|\Omega_k) - O$ has a zero expectation value because of the well-known law of total expectation $\mathbb{E}[\mathbb{E}(O|\Omega_k)] = \mathbb{E}(O)$. This law can be proven starting from Eq. (1), i.e., $\mathbb{E}[\mathbb{E}(O|\Omega_k)] = \int \mathbb{E}(O|\Omega_k)\pi(x)dx$, and decomposing the integral over $x$ as an integral over the environment variable $\Omega_k$ and an integral over $x' \in \Omega_k$. Let us prove now that the estimator $\tilde{O}$ has a zero-variance property in the separability limit when we choose $\lambda_k = 1\ \forall k$. In this limit, $O$ is a sum of $p$ independent contributions on each fragment, $O = \sum_{k=1}^{p} O_k[(x_j)_{j\in J_k}]$. Independence implies that $\mathbb{E}(O_k|\Omega_k) = \mathbb{E}(O_k)$ and $\mathbb{E}(O_l|\Omega_k) = O_l$ if $l \neq k$, therefore $\mathbb{E}(O|\Omega_k) - O = \mathbb{E}(O_k) - O_k$ and

$$\tilde{O} = \sum_{k=1}^{p} \mathbb{E}(O_k) = \mathbb{E}(O). \tag{6}$$

In this limit $\tilde{O}$ is a constant, only one parent configuration $x$ is sufficient for sampling $\tilde{O}$, and the algorithm becomes equivalent to $p$ independent Monte Carlo simulations of the $p$ subsystems as we have to compute $p$ (conditional) expectation values $\mathbb{E}(O|\Omega_k)$.

We can sample $\mathbb{E}(O|\Omega_k)$ from the marginal distribution $\pi(\cdot|\Omega_k)$. This is done through a side walk which samples only $\Omega_k$, i.e., moving the coordinates indexed by $J_k$ in a given fragment while the other coordinates are frozen. From now on we will use the practical definition of the improved estimator

$$\tilde{O} \equiv O + \sum_{k=1}^{p} \frac{\lambda_k}{m_k} \sum_{i=1}^{m_k} \left(O_k^i - O\right), \tag{7}$$

where $O_k^i$ is the value of the random variable $O$ at the $i$th step of the $k$th side walk (moving only the coordinates indexed by $J_k$) of length $m_k$. A direct way to see that the estimator in Eq. (7) is not biased is to note that $\mathbb{E}(O_k^i - O) = 0$ as $O_k^i$ and $O$ share the same distribution $\pi$, since the side walk and the main walk both sample $\pi$. We expect this scheme that we call the partition Monte Carlo (PMC) method to reduce the variance with a low numerical cost because the $p$ subsamplings correspond to handling $p = \mathcal{O}(N)$ low-dimensional problems. The practical formula in Eq. (7) is equivalent to the theoretical definition in Eq. (4) in the limit $m_k \to \infty$ owing to the ergodic theorem. In practice, the parameters $\lambda_k$ and $m_k$ have to be adjusted to lower the variance of $\tilde{O}$ for a given CPU time. Also, for optimal efficiency, we can generalize the estimator $\tilde{O}$ in Eq. (7) using instead of $O_k^i - O$ the control variate $G_k^i - G_k$ provided it converges to the former in the separability limit. $G_k$ can be obtained from $O$ by neglecting terms outside of the fragment $k$, reducing the computational cost while retaining the unbiasedness and the zero-variance property in the separability limit. For example, when computing the variational energy of a molecule, i.e., $O = (H\Psi)/\Psi$, we

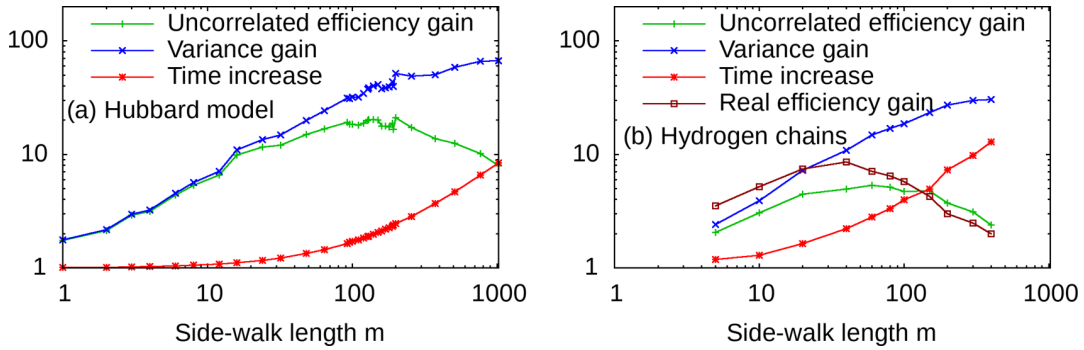


FIG. 1. Variance gain $V_{\mathrm{VMC}}/V_{\mathrm{PMC}}$, time increase $t_{\mathrm{PMC}}/t_{\mathrm{VMC}}$, and uncorrelated efficiency gain $(V_{\mathrm{VMC}}t_{\mathrm{VMC}})/(V_{\mathrm{PMC}}t_{\mathrm{PMC}})$ of the PMC method over the standard VMC method as a function of the side-walk length $m$ for (a) the 20×20 square Hubbard model (half filling) and (b) the $H_{320}$ metallic hydrogen chain. The real efficiency gain $(\sigma^2_{\mathrm{VMC}}t_{\mathrm{VMC}})/(\sigma^2_{\mathrm{PMC}}t_{\mathrm{PMC}})$ differs from the uncorrelated efficiency gain only for hydrogen chains.

take $G_k = (H_k\Psi)/\Psi$ where $H_k$ is the truncated Hamiltonian

$$H_k = \sum_{i=1}^{n_k}\left(-\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} + \sum_j \frac{1}{r_{ij}}\right), \tag{8}$$

where the index $i$ runs over the $n_k$ electrons in the fragment $k$. The first term is the kinetic-energy operator and the last two terms are the Coulomb interactions of the electrons of the fragment with the nuclei $A$ (charges $Z_A$) and electrons $j$ lying in a given neighborhood of the fragment.

Let us see now how the PMC method relieves the variance bottleneck. As an example, we consider VMC calculations using Jastrow-Slater wave functions

$$\Psi(x) = e^{J(x)}\Phi(x), \tag{9}$$

where $J(x)$ is any real symmetric function of the electron configuration $x$, and $\Phi(x) = \det(A)$ with the Slater matrix $A = XC$, where $X$ is a rectangular matrix of localized atomic orbitals (Kronecker functions in the case of a lattice model) and $C$ is the rectangular matrix of the orbital coefficients. For one fragment of the system we introduce now the matrix $P$ which selects the lines corresponding to the electrons of that fragment. For a side walk in that fragment, $X$ takes different values $X'$ such that only the lines $PX$ might differ from the lines $PX'$. The new determinant is [20,21]

$$\begin{aligned}\Phi(x') &= \det(X'C)\\ &= \det(A)\det(X'CA^{-1})\\ &= \det(A)\det(PX'Q^TQCA^{-1}P^T),\end{aligned} \tag{10}$$

where we have used the determinant lemma. We inserted the projector $Q^TQ$ where $Q^T$ selects on the right of $PX'$ only the few columns which may differ from zero for this fragment. These columns are very few because the atomic orbitals are localized. In conclusion, updating the determinant along the side walk is equivalent to multiplying it by a low-order effective Slater determinant

$$\Phi(x') = \det(A)\det(\bar{X}\bar{C}), \tag{11}$$

where $\bar{X} = PX'Q^T$ and $\bar{C} = QCA^{-1}P^T$. The matrix $\bar{C}$ represents effective orbitals for the fragment and is computed only once at each step of the usual main walk, at a total $\mathcal{O}(N^3)$ numerical cost for the $p$ subsystems. Once $\bar{C}$ has been built and stored, the side walk costs only $\mathcal{O}(n^3)$ where $n$ is the number of electrons in the fragment. The local energy of the subsystem involves a truncated Hamiltonian and can be computed with the same cost $O(n^3)$ [20,21]. The cost of subsampling $\mathcal{O}(N)$ fragments is thus $\mathcal{O}(N)$ for an extended system with a finite correlation length. This allows us to perform up to $\sum_k m_k = \mathcal{O}(N^3)$ total steps in the side walks without modifying the scaling of the main walk. Therefore, we can perform $m_k = \mathcal{O}(N^2)$ steps in each fragment and the improved estimator in Eq. (7) will have consequently a variance reduced by a factor up to $\mathcal{O}(N^2)$, which is achieved in the separability limit.

## III. RESULTS

We now illustrate the PMC method on the calculation of the ground-state energy of the two-dimensional (2D) Hubbard model and of metallic hydrogen chains.

The Hubbard systems that we employ consist in 2D square grids of $L\times L$ sites with periodic boundary conditions, filled to half capacity with $N \approx L^2$ electrons evenly distributed between the spins. Designating by $c^\dagger_{i\sigma}$ and $c_{i\sigma}$ the creation and annihilation operators of site $i$ with spin $\sigma \in \{\uparrow,\downarrow\}$, and by $n_{i\sigma} = c^\dagger_{i\sigma}c_{i\sigma}$ the corresponding number operators, the

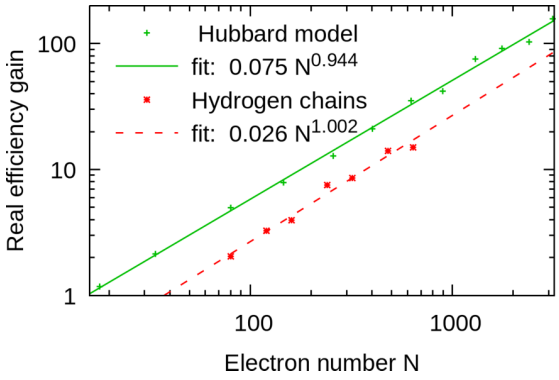


FIG. 2. Optimal real efficiency gain $(\sigma^2_{\mathrm{VMC}}t_{\mathrm{VMC}})/(\sigma^2_{\mathrm{PMC}}t_{\mathrm{PMC}})$ as a function of electron number $N$ for the Hubbard model and metallic hydrogen chains.

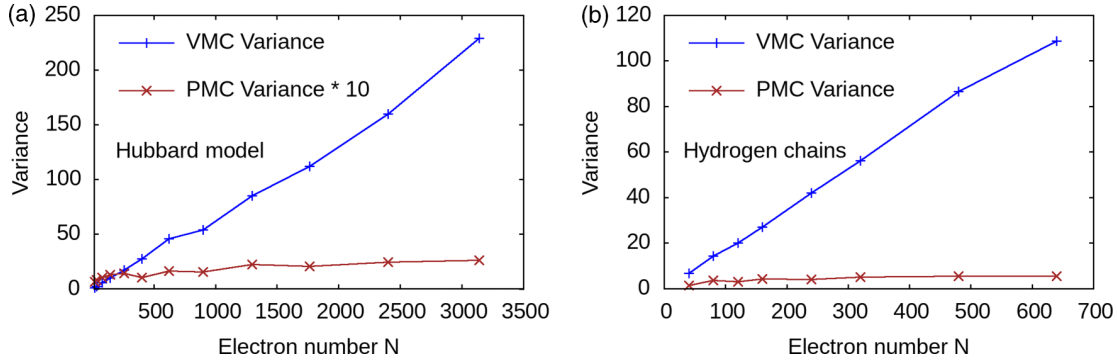


FIG. 3. Variance of the local energy in standard VMC and in PMC (for optimal $m$) as a function of electron number $N$ for (a) the Hubbard model (PMC variance multiplied by 10 on the plot) and (b) metallic hydrogen chains.

Hamiltonian takes the form [22]

$$H = -\sum_{i\neq j,\sigma} t_{ij} c^{\dagger}_{i\sigma} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow}, \tag{12}$$

where $t_{ij} = 1$ if $i$ and $j$ are adjacent, and $t_{ij} = 0$ otherwise, and $U = 1$ is the on-site interaction parameter. We have chosen the trial ground-state wave function to be a Slater determinant of plane waves without any Jastrow factor. We choose the subsystems as adjacent squares of $l \times l$ sites. The number of iterations of the main walk is kept constant at $M = 500$.

As an example of a simple system with a continuum configuration space, we consider metallic hydrogen chains with a regular interatomic distance of $1.4a_0$. The Hamiltonian is given by Eq. (8) except of course that there is no restriction in the sums for the full system. For the trial ground-state wave function, we use a simple Jastrow function [16] multiplied by the Hartree-Fock Slater determinant obtained from a basis made of the exact hydrogen 1s orbital on each atom. We choose the subsystems as consisting in $n$ adjacent hydrogen atoms.

The first parameter of the PMC method whose impact is to be explored is the side-walk length $m$ (chosen to be the same for all subsystems). Figure 1 reports the variance gain $V_{\mathrm{VMC}}/V_{\mathrm{PMC}}$, the CPU time increase $t_{\mathrm{PMC}}/t_{\mathrm{VMC}}$, and the uncorrelated efficiency gain $(V_{\mathrm{VMC}}t_{\mathrm{VMC}})/(V_{\mathrm{PMC}}t_{\mathrm{PMC}})$ (efficiency gain assuming a correlation factor $c = 1$) of the PMC method over the standard variational Monte Carlo (VMC) method. The efficiency gain is plotted as a function of the side-walk length $m$ for the 2D Hubbard model with total size $L = 20$ and subsystem size $l = 5$, and for hydrogen chains with $N = 320$ total atoms and $n = 12$ atoms in the subsystems. Two regimes are clearly visible. For small $m$, the variance gain increases linearly with $m$ while the CPU time is almost constant (the cost of a side-walk step is very small compared to that of a main-walk step). This leads to a linear increase of the uncorrelated efficiency gain. For large $m$, the variance gain saturates while the CPU time ratio increases linearly, driving the uncorrelated efficiency gain down. Between these two regimes, there is a plateau corresponding to optimal values of the side-walk length $m$. The saturation of the variance gain originates from the correlation between subsystems. Indeed, if the subsystems were independent, the variance would converge to zero as $m$ increases (zero-variance principle in the separability limit) and the variance gain to infinity.

One may ask the role of the correlation factor $c$ in Eq. (2). For the Hubbard model, $c$ has been found to be very close to 1, leading to a real efficiency gain almost identical to the uncorrelated efficiency gain. For the hydrogen chains $c \simeq 2.5$ for $m = 0$ (VMC) and $c$ is reduced for small $m$ (about 40% less for $H_{320}$ and $m \in [5, 40]$) before increasing slowly for larger values of $m$. This explains the difference between the uncorrelated efficiency gain and the real efficiency gain $(\sigma^2_{\mathrm{VMC}}t_{\mathrm{VMC}})/(\sigma^2_{\mathrm{PMC}}t_{\mathrm{PMC}})$ in Fig. 1. In particular, the optimal real efficiency gain is 40% higher than the optimal uncorrelated efficiency gain. We now consider systems of increasing sizes. For the Hubbard model, the optimal subsystem size has been found to be $l \approx \sqrt{L}$, and similarly for the metallic hydrogen chains we find $n \approx \sqrt{N/2}$. The fact that the optimal subsystem size does not saturate to a finite value as the system size increases is an indication of the nonseparability of the system. The optimal side-walk length $m$ also increases with system size since larger systems result in more decorrelated subsystems and cheaper side walks compared to the main walk. Figure 2 reports the real efficiency gain as a function of the electron number $N$ for the Hubbard model and the hydrogen chains up to $N$ of the order of $10^3$. Both metallic systems present a real efficiency gain scaling linearly with $N$, which hovers around $0.075N$ for the Hubbard model and $0.025N$ for the hydrogen chains. This real efficiency gain is almost entirely achieved by decreasing the variance of the local energy from $\mathcal{O}(N)$ to a behavior close to $\mathcal{O}(1)$, as shown in Fig. 3. Of course, we have checked that computing $\mathbb{E}(O)$ and $\mathbb{E}(\tilde{O})$ always gives the same answer within the error bars, in agreement with the unbiasedness of $\tilde{O}$.

## IV. CONCLUSIONS

We introduced a general and simple method to reduce the scaling of Monte Carlo calculations of extensive properties. It only requires to have an explicit formula [Eq. (1)] for the integral to be computed, and therefore can be used in any Markov chain Monte Carlo application. The method was illustrated on VMC calculations of metallic systems of $N$ particles, providing an efficiency gain of order $\mathcal{O}(N)$. The present idea can be applied in many contexts,

including fixed-node path-integral Monte Carlo approaches [23,24] since these schemes sample explicit probability distributions. Finally, the method can in principle be extended to derivatives of extensive properties to reduce the scaling for calculating response properties or optimizing variational wave functions.


[1] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*, 5th ed., Graduate Texts in Physics (Springer, Berlin, 2010).

[2] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

[3] J. Lynn, I. Tews, S. Gandolfi, and A. Lovato, Annu. Rev. Nucl. Part. Sci. **69**, 279 (2019).

[4] J. Goodman and A. D. Sokal, Phys. Rev. D **40**, 2035 (1989).

[5] J. Weare, Proc. Natl. Acad. Sci. USA **104**, 12657 (2007).

[6] J. Toulouse, R. Assaraf, and C. J. Umrigar, Adv. Quantum Chem. **73**, 285 (2016).

[7] S. Mohr, M. Eixarch, M. Amsler, M. J. Mantsinen, and L. Genovese, Nucl. Mater. Energy **15**, 64 (2018).

[8] R. Assaraf and M. Caffarel, Phys. Rev. Lett. **83**, 4682 (1999).

[9] A. Mira, R. Solgi, and D. Imparato, Stat. Comput. **23**, 653 (2013).

[10] D. Borgis, R. Assaraf, B. Rotenberg, and R. Vuilleumier, Mol. Phys. **111**, 3486 (2013).

[11] R. Assaraf and D. Domin, Phys. Rev. E **89**, 033304 (2014).

[12] A. J. Williamson, R. Q. Hood, and J. C. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).

[13] R. Assaraf, Phys. Rev. E **90**, 063317 (2014).

[14] R. Assaraf, B. Jourdain, T. Lelièvre, and R. Roux, Stoch. Partial Differ. Equ.: Anal. Comput. **6**, 125 (2017).

[15] G. Cohen, E. Gull, D. R. Reichman, and A. J. Millis, Phys. Rev. Lett. **115**, 266802 (2015).

[16] J. Feldt and R. Assaraf, J. Chem. Theory Comput. **17**, 1380 (2021).

[17] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[18] G. Knizia and G. K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012).

[19] F. Zahariev and M. S. Gordon, Phys. Chem. Chem. Phys. **23**, 14308 (2021).

[20] C. Filippi, R. Assaraf, and S. Moroni, J. Chem. Phys. **144**, 194105 (2016).

[21] R. Assaraf, S. Moroni, and C. Filippi, J. Chem. Theory Comput. **13**, 5273 (2017).

[22] M. Cyrot, Physica B+C **91**, 141 (1977).

[23] S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999).

[24] J. Shumway and M. Gilbert, J. Phys.: Conf. Ser. **35**, 190 (2006).