Some complements to the lecture

"Quantum Monte Carlo"

Michel Caffarel

Lab. Chimie et Physique Quantiques, CNRS-Université de Toulouse, France.

Contact (do not hesitate): caffarel@irsamc.ups-tlse.fr

CONTENTS

I. SOME REVIEW PAPERS

DMC and VMC for continuous space [1], [2], [3], [4]

For matrices:

[5]

II. THE VARIATIONAL MONTE CARLO (VMC) METHOD

In VMC, the probability density associated with Ψ_T is sampled

$$\pi(\mathbf{x}) = \frac{\Psi_T^2(\mathbf{x})}{\int d\mathbf{x}\Psi_T^2(\mathbf{x})^2}.$$

The properties are computed as probabilistic averages over sampled configurations.

In the case of the energy, the variational energy E_v is obtained as

$$E_v = \frac{\langle\Psi_T|H|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \frac{\int d\mathbf{x}|\Psi_T|^2\frac{H\Psi_T}{\Psi_T}}{\int d\mathbf{x}|\Psi_T|^2}$$

that is

$$\boxed{E_v = \int d\mathbf{x}\pi(\mathbf{x})E_L(\mathbf{x})}$$

and

$$E_{var} = \langle E_L\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E_L(\mathbf{x}_i)$$

Other properties can be computed in a similar way

$$\frac{\langle\Psi_T|O|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} = \int d\mathbf{x}O(\mathbf{x})\pi(\mathbf{x}) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{K}O(\mathbf{x}_i)$$

How to sample Ψ^2?

• Use of the drifted brownian motion.

$$\boxed{\frac{\partial f(\mathbf{x},t)}{\partial t} = \frac{1}{2}\nabla^2 f(\mathbf{x},t) - \nabla[\mathbf{b}(\mathbf{x}) f(\mathbf{x},t)]}$$

The stationary distribution is given by the condition

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = 0$$

which gives

$$f(\mathbf{x}) = \frac{\Psi_T^2}{\int d\mathbf{x} \Psi_T^2}$$

Proof. $\frac{1}{2}\nabla^2 f - \nabla[\mathbf{b}f] = \frac{1}{2}\nabla^2\Psi_T^2 - \nabla\frac{\nabla\Psi_T}{\Psi_T}\Psi_T^2 = \frac{1}{2}\nabla^2\Psi_T^2 - \frac{1}{2}\nabla\nabla\Psi_T^2 = 0$

The averages can be computed over the set of configurations generated by the drifted brownian motion.

Nice but practical problem is the residual **short-time error** due to the finite time-step τ.

- Use of the Metropolis algorithm

Very important algorithm. It belong to **the Top 10 list of the most employed numerical algorith used in science and technology**

A. The Metropolis algorithm

The aim is to compute

$$\boxed{\mathrm{I} = \int_{\mathcal{E}} d\mathbf{x} \Pi(\mathbf{x}) F(\mathbf{x})}$$

where $\mathbf{x} \in \mathcal{E}$= configuration space (continuous or discrete).

$\Pi(\mathbf{x})$ probability density, that is

$$\Pi(\mathbf{x}) \geq 0 \quad \text{and} \quad \int_{\mathcal{E}} d\mathbf{x}\ \Pi(\mathbf{x}) = 1$$

The Metropolis algorithm generates step by step configurations $\mathbf{x}^i$ in configuration space distributed according to $\Pi(\mathbf{x})$.

We then have

$$\mathrm{I} = \lim_{P \to \infty} \frac{1}{P} \sum_{i=1}^{P} F(\mathbf{x}^i)$$

In practice, a finite number of configurations are generated and we have

$$\mathrm{I} = \frac{1}{\mathrm{P}} \sum_{\mathrm{i}=1}^{\mathrm{P}} \mathrm{F}(\mathrm{x}^{\mathrm{i}}) + \frac{\mathrm{c}}{\sqrt{\mathrm{P}}} \quad \text{for P large enough}$$

The fundamental quantity of the algorithm is the **trial transition probability density** denoted here as $P(\mathbf{x} \to \mathbf{y})$. The algorithm is as follows.

METROPOLIS ALGORITHM

At each Monte Carlo step a new state $\mathbf{x}_{i+1}$ is generated from the current state $\mathbf{x}_i$ by a two-step procedure:

1) Draw a "trial" state denoted as $\mathbf{x}^T$ using the trial transition probability density $P(\mathbf{x} \to \mathbf{y})$

2) Accept the trial state as the new state ($\mathbf{x}_{i+1} = \mathbf{x}^T$) or reject it ($\mathbf{x}_{i+1} = \mathbf{x}_i$) with probability $q(\mathbf{x}_i, \mathbf{x}_T)$ $(0 \leq q \leq 1)$ given by

$$q = Min\big[1, \frac{\pi(\mathbf{x}^T) P(\mathbf{x}^T \to \mathbf{x}_i)}{\pi(\mathbf{x}_i) P(\mathbf{x}_i \to \mathbf{x}^T)}\big] \tag{1}$$

• $P(\mathbf{x} \to \mathbf{y})$ **must be easy to sample.** In practice, we (almost) always use a product of one-dimensional uniform or gaussian probability densisities. A universal choice inspired by the drifted brownian motion defined above is

$$P(\mathbf{x} \to \mathbf{y}) = \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}-\mathbf{x}_0-\mathbf{b}(\mathbf{x}_0)\tau)^2}{2\tau}}$$

where the drift is

$$\mathbf{b}(\mathbf{x}) = \frac{1}{2}\frac{\nabla\Pi}{\Pi}$$

• $P(\mathbf{x} \to \mathbf{y})$ **must be ergodic** ("go everywhere")

• The Metropolis algorithm converges to π **independently** on the choice of the trial transition probability and/or the initial conditions $\mathbf{x}_0$. Such quantities only determines the rate of convergence of the Markov chain towards π.

For a derivation of the Metropolis algorithm in the discrete case, see appendix E

B. The trial wavefunction

In QMC a great freedom in choosing the functional form of the trial wavefunction

The standard choice: The multi-determinant Slater-Jastrow wavefunction.

$$\Psi_T = e^{J(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \sum_{k=1}^{N_{det}} c_k Det_k(\{\Phi_i^\alpha\}) Det_k(\{\Phi_i^\beta\}), \tag{2}$$

where $\{\Phi_i^\sigma\}(\sigma = \alpha, \beta)$ is a set of molecular orbitals and e^J is the Jastrow factor. The role of the Jastrow factor is to impose the exact behavior of the wavefunction in the $[r_{ij} \to 0]$-limit (electron-electron cusp condition) and, also, to incorporate some two-body (electron-electron and electron-nucleus) and three-body (electron-electron-nucleus) correlations (to describe the best as possible the shape of the Coulomb hole[6]). Many different forms for the Jastrow factor have been introduced. Typically,

$$J = \sum_{i<j} u(r_{ij}) + \sum_i \sum_\alpha v(r_{i\alpha}) + \sum_{i<j} \sum_\alpha w(r_{ij}, r_{i\alpha}, r_{j\alpha})$$

where $r_{ij} = |r_i - R_\alpha|$, and $r_{i\alpha} = |r_i - R_\alpha|$. Various forms for the functions u,v, and w have been tested. For example, the minimal Padé form for u

$$u(r_{ij}) = \frac{a r_{ij}}{1 + b r_{ij}}.$$

But many other forms, see appendix I

C. The trial wavefunction optimization

Aim: To find the "best" parameters of the trial wavefunction

- Minimization of the variational energy

$$E(\mathbf{p}) = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

where $\mathbf{p}$ denotes the set of parameters of $\Psi_T(\mathbf{x}, \mathbf{p})$

- Minimization of the variance of the Hamiltonian

$$\sigma^2(\mathbf{p}) = \frac{\langle \Psi_T | [H - E(\mathbf{p})]^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

Motivations:

- Reduce the statistical fluctuations (remember the zero-variance property)

- Reduce the fixed-node error

III. THE DIFFUSION MONTE CARLO (DMC) METHOD

Very similar to the method presented above. It differs only in the way the local energy term in the diffusion equation is taken into account.

$$\boxed{\frac{\partial f(\mathbf{x},t)}{\partial t} = \frac{1}{2}\nabla^2 f(\mathbf{x},t) - \nabla[\mathbf{b}(\mathbf{x})f(\mathbf{x},t)] - (E_L(\mathbf{x}) - E_T)f(\mathbf{x},t)}$$

The equation of evolution of the local energy part is given by

$$\frac{\partial f(\mathbf{x},t)}{\partial t} = -(E_L(\mathbf{x}) - E_T)f(\mathbf{x},t)$$

whose solution is

$$f(\mathbf{x},t) = f(\mathbf{x},t=0)e^{-t(E_L(\mathbf{x})-E_T)}$$

Instead of considering $e^{-t(E_L(\mathbf{x})-E_T)}$ as a weight for the drifted brownian trajectories, we simulate this term as a **birth-death process or branching process.**
In the branching process the variation of density is reproduced by **killing or duplicating a certain number of times each walker** at position $\mathbf{x}$ proportionally to $e^{-t(E_L(\mathbf{x})-E_T)}$
The stationary density is now

$$\boxed{\pi_{DMC} = \Psi_T \Phi_0} \tag{3}$$

when E_T has ben taken equal to E_0.

The energy can be computed as

$$E_0 = \frac{\int \Phi_0 H \Psi_T}{\int \Phi_0 \Psi_T} = \frac{\int \Phi_0 \Psi_T \frac{H\Psi_T}{\Psi_T}}{\int \Phi_0 \Psi_T}$$

and then

$$\boxed{E_0 = \int d\mathbf{x} \pi_{DMC}(\mathbf{x}) E_L(\mathbf{x})} \tag{4}$$

Appendix A: Uniform distribution and random generator

1. Uniform distribution

The uniform distribution over $(0,1)$ is given by the density

$$P(x) = \begin{cases} 1, & \text{if } x \in (0,1), \\ 0, & \text{if } x \notin (0,1). \end{cases} \tag{A1}$$

In practice, the uniform distribution is realized using a Random Number Generator (RNG). Most generators are based on the use of a deterministic algorithm "mimicking" randomness as best as possible (pseudo-random generators). A common one is the simple linear congruential generator

$$x_{n+1} = (ax_n + c) \bmod m \tag{A2}$$

where x_0 is defined as the "seed" of the generator. Note that once the seed has been chosen, the entire series of "random" numbers can be reproduced. A vast literature is devoted to the problem of producing randomness as pure as possible (minimization of correlations between pseudo-random numbers). A popular good quality-RNG has been proposed by L'Ecuyer[7]

2. L'Ecuyer pseudo-random generator

The L'Ecuyer pseudo-random generator is a combined multiple recursive generator

$$z_n = (x_n - y_n) \bmod m_1$$

where x_n and y_n are

$x_n = (a_1 x_{n-1} + a_2 x_{n-2} + a_3 x_{n-3}) \bmod m_1$

$y_n = (b_1 y_{n-1} + b_2 y_{n-2} + b_3 y_{n-3}) \bmod m_2$

with coefficients $a_1 = 0, a_2 = 63308, a_3 = -183326, b_1 = 86098, b_2 = 0, b_3 = -539608$, and moduli $m_1 = 2^{31} - 1 = 2147483647$ and $m_2 = 2145483479$.

The period is approximately 2^{185} (about 10^{56}).

Appendix B: Gausian distribution and random generator

1. One-dimensional gaussian distribution over $(-\infty, +\infty)$

As a consequence of the central-limit theorem, the gaussian distribution is ubiquitous in real applications. The one-dimensional version is defined as

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{B1}$$

where μ is the mean of the distribution

$$\mu = \langle x \rangle = \int_{-\infty}^{+\infty} dx\ x\ P(x) \tag{B2}$$

and σ^2 its variance

$$\sigma^2 = \langle (x-\mu)^2 \rangle = \int_{-\infty}^{+\infty} dx\ (x-\mu)^2\ P(x) \tag{B3}$$

When $\mu = 0$ and $\sigma^2 = 1$, the distribution is known as the normal distribution.

A simple and widely employed gaussian random generator based on the use of a uniform random generator is the **Box-Muller algorithm** given by

$$\begin{cases} x = \sqrt{-2\ln u_1}\cos(2\pi u_2) \\ y = \sqrt{-2\ln u_1}\sin(2\pi u_2) \end{cases} \tag{B4}$$

where u_1, u_2 are two uniform random numbers over (0,1). The two values x and y are independent and gaussian distributed.

2. Generalization to arbitrary dimension d

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d detC}} \exp\left[-\frac{1}{2}\sum_{i,j}(\mathbf{x}-\mu)_i C^{-1}_{ij}(\mathbf{x}-\mu)_j\right] \tag{B5}$$

where μ is the mean vector

$$\mu_i = \langle x_i \rangle \quad i = 1, d \tag{B6}$$

and C the $d \times d$ covariant matrix given by

$$\boxed{C_{ij} = \langle (\mathbf{x}-\mu)_i (\mathbf{x}-\mu)_j \rangle} \tag{B7}$$

To sample this d-dimensional distribution, diagonalize the covariant matrix, factorize the probability distribution into a product of d one-dimensional gaussian distribution using the eigensolutions of C, and then sample the 1d distributions independently.

Appendix C: Stochastic process

1. General stochastic process

Stochastic process $X(t)$ = Series of random variables indexed by a time t.

The fundamental quantities are the *n-time probability distributions*. In the continuous case, it is written as

$$P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) \tag{C1}$$

with $0 \leq t_1 \leq t_2 \leq ... \leq t_n$, $\mathbf{x}_i$ denoting the state, or configuration, of the system at time t_i [typically, $\mathbf{x} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$, N number of particles]. The interpretation of the probability distribution density is as follows.

$$P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) d\mathbf{x}_1 d\mathbf{x}_2 ... d\mathbf{x}_n \tag{C2}$$

is the probability of finding the system between $\mathbf{x}_1 + d\mathbf{x}_1$ at time t_1, $\mathbf{x}_2 + d\mathbf{x}_2$ at time t_2, etc with

$$\int d\mathbf{x}_1 d\mathbf{x}_2 ... d\mathbf{x}_n P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) = 1 \tag{C3}$$

By integrating the n-time distribution over all states at k first times, we can generate $(n-k)$-time probability distribution densities

$$\boxed{P_{n-k}(\mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) = \int d\mathbf{x}_1 ... d\mathbf{x}_k P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n)} \tag{C4}$$

Let us now define the *conditional probability densities* as follows

$$\boxed{P_{k|(n-k)}(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) = \frac{P_n(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k; \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n)}{P_k(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k)}} \tag{C5}$$

With this definition

$$P_{k|(n-k)}(\mathbf{x}_1, t_1; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; ...; \mathbf{x}_n, t_n) d\mathbf{x}_{k+1} d\mathbf{x}_{k+2} ... d\mathbf{x}_n \tag{C6}$$

is the probability of finding the system between $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1}, ...,$\mathbf{x}_n + d\mathbf{x}_n$ at time t_n *knowing* that the system was at $\mathbf{x}_1$ at time t_1, $\mathbf{x}_2$ at time t_2,...,$\mathbf{x}_k$ at time t_k.

Stochastic process are now classified according to the nature of their n-time probability distributions.

2. Fully decorrelated process: The case of the branching process

Fully decorrelated process are the simplest stochastic process we can think of. They describe a time series of independent random variables. The probability of being between $\mathbf{x}_{k+1}$ and $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1}, knowing that we are at $\mathbf{x}_k$ at time t_k, is independent on $\mathbf{x}_k$ (and, then, on all previous states). In terms of conditional probability densities it is written as (for all possible k)

$$\boxed{P_{k|1}(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; ...; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) = P_1(\mathbf{x}_{k+1}, t_{k+1})} \tag{C7}$$

where $P_1(\mathbf{x}, t)$ is the probability distribution at time t, namely

$$P_1(\mathbf{x}, t) = \int d\mathbf{x}_2 ... d\mathbf{x}_n P_n(\mathbf{x}, t; \mathbf{x}_2, t_2; ...; \mathbf{x}_n, t_n) \tag{C8}$$

Using Eqs.(C5) and (C7) the n-time probability distribution can be written as

$$\boxed{P_n(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2, ...) = \prod_k P_1(\mathbf{x}_k, t_k)} \tag{C9}$$

Because of their simplicity and lack of time correlations such process are usually not very useful for modelizing physical situations. As a simple example, we could use them for describing the dynamics of a Brownian particle (pollen grain in water) when observation times t_k are separated by long time intervals (say, several minutes or more). Another more interesting exemple is the so-called branching or birth-death process as it is defined in DMC simulations.

Branching process.

We describe now the so-called "branching" or "birth-death" process as it is defined in QMC. It will be used in the Diffusion Monte Carlo (DMC) algorithm presented below. Note that it is actually a very particular case of more general branching process introduced in mathematics.

Let us consider a weight $w \geq 0$ (we will see that this weight will depend on electronic configuration). The branching process is defined as

$$\boxed{X = E(w + U)} \tag{C10}$$

where U is the uniform random variable over (0,1) and E the integer part. X takes on integer values. The probability of having n is denoted as

$$P_n = P(X = n) \tag{C11}$$

Now, it is clear that for a given w, only two values of n with non-zero probability are possible: n_c and n_c+1 where $n_c \equiv E(w)$. Now, we have

$$P_{n_c+1} = 1 - (n_c + 1 - w) \tag{C12}$$

$$P_{n_c} = n_c + 1 - w \tag{C13}$$

Of course, as it should be, $P_{n_c+1} + P_{n_c+1} = 1$. Let us compute the mean

$$\bar{n} = n_c(n_c + 1 - w) + (n_c + 1)(1 - (n_c + 1 - w)) = w \tag{C14}$$

We thus have

$$\boxed{\langle X \rangle = w} \tag{C15}$$

3. General Markov process

These are the key process used in the vast majority of stochastic simulations. The probability of being between $\mathbf{x}_{k+1}$ and $\mathbf{x}_{k+1} + d\mathbf{x}_{k+1}$ at time t_{k+1} is now dependent on the previous configurations $\mathbf{x}_k$ but not on the oldest ones $\mathbf{x}_{l<k}$. It is common to say (in a loosely way) that for a Markov process, the future (at time t_{k+1}) depends on the present (time t_k) but not on the past (times $t_{l<k}$). More precisely, **the Markov hypothesis** is written as

$$\boxed{P_{k|1}(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2, ..., \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) = P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1})} \tag{C16}$$

The fundamental quantity $P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1})$ characterizing the Markov process is called the transition kernel or transition probability density. In what follows we shall use the convenient notation

$$P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = P_{1|1}(\mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}) \tag{C17}$$

It is easy to check that the n-time probability density can now be written as

$$\boxed{P_n(\mathbf{x}_1, t_1; ...; \mathbf{x}_n, t_n) = P_1(\mathbf{x}_1, t_1) \prod_{k=1}^{n-1} P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}).} \tag{C18}$$

From Eqs.(C4) and (C5) we have

$$\int d\mathbf{x}_{k+1} P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = 1 \tag{C19}$$

In practice, most of the Markov process used in simulations are invariant under a time shift, they are said to be *homogeneous*. In that case

$$P(\mathbf{x}_k, t_k \to \mathbf{x}_{k+1}, t_{k+1}) = P(\mathbf{x}_k \to \mathbf{x}_{k+1}, t_{k+1} - t_k) \tag{C20}$$

For simplicity, the time interval will be denoted as t and the transition probability as $P(\mathbf{x} \to \mathbf{y}, t)$. Because of the time-shift invariance, the one-body density $P_1(\mathbf{x})$ is now independent on time. Let us derive the equation obeyed by $P_1(\mathbf{x})$. We have

$$P(\mathbf{x} \to \mathbf{y}, t) = \frac{P_2(\mathbf{x}; \mathbf{y}, t)}{P_1(\mathbf{x})} \tag{C21}$$

Mutiplying the equation by $P_1(\mathbf{x})$ and integrating over $\mathbf{x}$ we get

$$\int d\mathbf{x} P_1(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \int d\mathbf{x} P_2(\mathbf{x}; \mathbf{y}, t) = P_1(\mathbf{y}). \tag{C22}$$

Following a popular tradition, we shall denote, here and in what follows, **the stationary distribution density** as π

$$\pi(\mathbf{x}) = P_1(\mathbf{x}) \tag{C23}$$

The equation obeyed by π is thus $\boxed{\int d\mathbf{x} \pi(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \pi(\mathbf{y})}$

Starting from the distribution $\pi(\mathbf{x})$ and applying the transition kernel to all $\mathbf{x}$ leads to configurations $\mathbf{y}$ also distributed according to π. It clearly illustrates the interpretation of π as the stationary distribution of the stochastic process.

Let us now adopt an alternative point of view. As already mentioned, the transition probability density characterizes the Markov process. Considered as the kernel of a linear operator, the properties of its eigensolutions can be studied. A first remark is that the transition probability is in general not symmetric, $P(\mathbf{x} \to \mathbf{y}, t) \neq P(\mathbf{y} \to \mathbf{x}, t)$. As a consequence, it is necessary to distinguish between left- and right-eigenvectors and, in addition, the eigenvalues are not necessarily real. However, because $P(\mathbf{x} \to \mathbf{y}, t) \geq 0$ and $\int d\mathbf{y} P(\mathbf{x} \to \mathbf{y}, t) = 1$ it can be shown that the modulus of all eigenvalues ≤ 1 and that the left-eigenstate associated with the maximal eigenvalue $\lambda = 1$ is positive everywhere (Krein-Rutman theorem, a generalization of the Perron-Frobenius theorem to operators [8) The integral equation

$$\int d\mathbf{x} \pi(\mathbf{x}) P(\mathbf{x} \to \mathbf{y}, t) = \pi(\mathbf{y}) \tag{C24}$$

is thus recovered where $\pi(\mathbf{x}) \geq 0$ is the maximal eigenvector of the transition kernel which defines the stationary distribution of the stochastic process.

In the preceding section we have derived an integral equation allowing to compute the stationary density π when the transition kernel is known. Let us now consider the problem of the computation of the kernel itself. The fundamental equation for $P(\mathbf{x} \to \mathbf{y}, t)$ is a simple consequence of the Markov hypothesis. It is obtained by observing that if we introduce an arbitrary intermediate time $u \in (0, t)$ and consider the probability of going from $\mathbf{x}$ to $\mathbf{y}$ in a time t we must have

$$\boxed{P(\mathbf{x} \to \mathbf{y}, t) = \int d\mathbf{z} P(\mathbf{x} \to \mathbf{z}, u) P(\mathbf{z} \to \mathbf{y}, t - u)} \tag{C25}$$

It is known under the name of **Chapman-Kolmogorov equation**. A much more interesting form is its local form relating time and space derivatives.

Let us derive such an equation in the one-dimensional case. The generalization to an arbitrary dimension is elementary. The following derivation follows closely that of [9] Let $h(x)$ be an arbitrary smooth function and consider the time derivative of the transition probability. We can write

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \int dy h(y) \lim_{\Delta t \to 0} \frac{P(x \to y, t + \Delta t) - P(x \to y, t)}{\Delta t}$$

Applying the Chapman-Kolmogorov equation we have

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big[\int dy h(y) \int dz P(x \to z, t) P(z \to y, \Delta\tau) - \int dy h(y) P(x \to y, t) \Big]$$

Changing the name of the dummy variable y into z in the last integral of the RHS and using $\int dy P(z \to y, \Delta t) = 1$ then

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big[\int dz P(x \to z, t) \int dy P(z \to y, \Delta\tau) [h(y) - h(z)] \Big]$$

Now, we introduce a Taylor expansion of $h(y)$ around z:

$$h(y) = h(z) + \sum_{n=1}^{\infty} h^{(n)}(z) \frac{(y - z)^n}{n!}$$

and defining the "jump moments"

$$D^{(n)}(z) = \frac{1}{n!} \lim_{\Delta t \to 0} \int dy (y - z)^n P(z \to y, \Delta\tau)$$

we get

$$\int dy h(y) \frac{\partial P(x \to y, t)}{\partial t} = \int dz P(x \to z, t) \sum_{n=1}^{\infty} D^{(n)}(z) h^{(n)}(z)$$

Integrating by parts n times we get

$$\int dz h(z) \Big[\frac{\partial P(x \to z, t)}{\partial t} - \sum_{n=1}^{\infty} \big(- \frac{\partial}{\partial z} \big)^n [D^{(n)}(z) P(x \to z, t)] \Big] = 0$$

and finally this integral being valid for any h the equation for the transition probability can be written as

$$\frac{\partial P(x \to y, t)}{\partial t} = \sum_{n=1}^{\infty} \big(- \frac{\partial}{\partial y} \big)^n [D^{(n)}(y) P(x \to y, t)] \tag{C26}$$

In its general d-dimensional version it writes

$$\boxed{\frac{\partial P(\mathbf{x} \to \mathbf{y}, t)}{\partial t} = \sum_{n=1}^{\infty} (-1)^n \sum_{j_1 \ldots j_n} \frac{\partial^n}{\partial y_{j_1} \cdots \partial y_{j_n}} \Big[D^{(n)}_{j_1, \ldots, j_n}(\mathbf{y}) P(\mathbf{x} \to \mathbf{y}, t) \Big] .} \tag{C27}$$

This equation is known under the name of **Kramers-Moyal expansion** (of the master equation). Here, the "jump moments" are defined as

$$\boxed{D^{(n)}_{j_1, \ldots, j_m}(\mathbf{y}) = \frac{1}{n!} \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big\langle \prod_{\mu=1}^{n} [Y_{j_\mu}(t + \Delta t) - Y_{j_\mu}(t)] \Big\rangle \Big|_{Y_k(t) = y_k} .} \tag{C28}$$

This equation is known under the name of Kramers-Moyal expansion (of the master equation). Let us now discuss the Markovian process at the heart of QMC approaches presented below.

4. Markovian process at work in QMC

- *Free diffusion or brownian process.*

The free diffusion process is invariant by space translation and thus, $D^{(1)} = 0$. It is defined by a constant diagonal diffusion matrix $D^{(2)}_{ij} = \frac{1}{2}$ and $D^{(n>2)} = 0$

In one dimension the Kramers-Moyal expansion is written as

$$\boxed{\frac{\partial P(x \to y, t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial y^2} P(x \to y, t)} \tag{C29}$$

with initial condition, $P(x \to y, t = 0) = \delta(x - y)$ This equation is known under the name of **free diffusion (or heat) equation**. By using a Fourier transform the gaussian solution of this equation is easily obtained. We have

$$p(x \to y, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} \tag{C30}$$

In d dimensions the solution is a product of independent one-dimensional gaussian distributions for each coordinate

$$p(\mathbf{x} \to \mathbf{y}, t) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y_i - x_i)^2}{2t}} = \frac{1}{\sqrt{2\pi t}^d} e^{-\frac{(\mathbf{y}-\mathbf{x})^2}{2t}} \tag{C31}$$

Using the gaussian transition probability density, brownian trajectories can be generated step-by-step. From Eq.(C31) it is seen that the quantities $\frac{(y_i - x_i)}{\sqrt{t}}$ are independent and normally distributed. $\mathbf{y}$ can thus be obtained from $\mathbf{x}$ by drawing a gaussian number for each coordinate

$$\frac{(y_i - x_i)}{\sqrt{t}} = \eta_i \quad i = 1, d \tag{C32}$$

where η is a normal random vector. The previous expression can be rewritten as

$$\boxed{y_i = x_i + \sqrt{t}\eta_i \quad i = 1, d} \tag{C33}$$

This last equation is the simplest example of a discretized form of the so-called **Stochastic Differential Equation (SDE)** associated with a diffusion process.

• *Drifted diffusion or drifted Brownian motion.* As we shall see later, QMC methods are based on a more general version of the free Brownian motion where a drift part is introduced to enhance the Monte Carlo convergence (importance sampling). In this case, both $D^{(1)}$ and $D^{(2)}$ are non-vanishing. The first jump moment is known as the drift vector

$$\mathbf{b}(\mathbf{x}) = \mathbf{D^{(1)}}(\mathbf{x}) \tag{C34}$$

In this case, the equation of evolution (KM expansion) is known as the **Fokker-Planck equation**. It is written as

$$\boxed{\frac{\partial P(\mathbf{x} \to \mathbf{y}, t)}{\partial t} = \frac{1}{2}\nabla_y^2 P(\mathbf{x} \to \mathbf{y}, t) - \nabla_y[\mathbf{b}(\mathbf{y}) P(\mathbf{x} \to \mathbf{y}, t)]} \tag{C35}$$

In the case of a constant drift vector $\mathbf{b}$ this equation can still be solved using a Fourier transform, we get

$$P(\mathbf{x} \to \mathbf{y}, t) = \frac{1}{\sqrt{2\pi t}^d} e^{-\frac{(\mathbf{y}-\mathbf{x}-\mathbf{b}\ t)^2}{2t}} \tag{C36}$$

Stochastic trajectories are generated using the discretized SDE

$$y_i = x_i + b_i(x_1, ..., x_d)t + \sqrt{t}\eta_i \;\; i = 1 \;\; to \;\; d \tag{C37}$$

In the case of a general drift $\mathbf{b}(\mathbf{x})$, no analytical solution exists. However, it is still possible to generate trajectories by using a small enough time-step τ instead of an arbitrary time t as above. For that, we need to introduce a short-time approximation of the transition probability. When the time-step is sufficiently small, the variation of position is small and at leading order the drift vector can be considered as constant. The transition probability density is thus approximated as

$$\boxed{P(\mathbf{x} \to \mathbf{y}, \tau) = \frac{1}{\sqrt{2\pi t}^d} \exp - \frac{(\mathbf{y} - \mathbf{x} - \mathbf{b}(\mathbf{x})\tau)^2}{2\tau}} \tag{C38}$$

This qualitative statement can be made more rigorous by looking at the small time-step limit of the exact solution of the Fokker-Planck equation, Eq.(C35). Having a short-time gaussian expression for the transition probability, stochastic trajectories can be generated according to

$$\boxed{y_i = x_i + b_i(\mathbf{x})\tau + \sqrt{\tau}\eta_i \quad i = 1, d} \tag{C39}$$

Note that the equations for each component are now coupled through the drift vector.

The stationary density π of the process can be obtained by solving $\frac{\partial P(\mathbf{x}\to\mathbf{y},t)}{\partial t} = 0$ that is

$$\frac{1}{2}\nabla^2\pi - \nabla(\mathbf{b}\pi) = 0$$

It is easily seen that this equality is fulfilled when

$$\boxed{\mathbf{b}(\mathbf{x}) = \frac{1}{2}\frac{\nabla\pi(\mathbf{x})}{\pi(\mathbf{x})}} \tag{C40}$$

Markov process with drift can thus be used to sample a given distribution $\pi(\mathbf{x})$ (for example, the Boltzmann distribution $\pi(\mathbf{x}) = \frac{e^{-\beta E(\mathbf{x})}}{Z}$). For that, we choose a drift vector according

to Eq.(C40) (here, $\mathbf{b} = -\frac{\beta}{2}\nabla E(\mathbf{x})$) and we generate trajectories using the stochastic differential equation, Eq.(C37). Note that with such a scheme a (small) bias on the stationary distribution related to the use of a small but finite time-step is present. In contrast, it is not the case with the Metropolis algorithm presented in the next section.

• *Other Markov process.* There exist a great variety of Markovian process. Let us just say a few words about two important examples.

i) *The Lévy flight: A generalization of the browian motion allowing large moves*

Probability distribution:

$$f(x;\mu,c) = \sqrt{\frac{c}{2\pi}} \quad \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

where $x > \mu$, μ = location parameter, and c =scale parameter.

"Heavy-tailed" probability distribution (large values of x have non-negligible probability to occur). Note that $< x^2 >= \infty$ (mean), $< x^2 >= \infty$ (variance)!!

Kramers-Moyal equation derived above

$$\frac{\partial P(x \to y,t)}{\partial t} = \sum_{n=1}^{\infty} \big(-\frac{\partial}{\partial y}\big)^n [D^{(n)}(y) P(x \to y,t)]$$

becomes here

$$\frac{\partial P(x \to y,t)}{\partial t} = -(-\frac{\partial^\alpha}{\partial y^\alpha})[D^{(2)}(y) P(x \to y,t)] - \frac{\partial}{\partial y}[D^{(1)} P(x \to y,t)]$$

with fractional derivative ($0 < \alpha \leq 2$).

An intense activity aout the modelization of the paths followed by animals or humans when searching for food, hunting, (or even searching for lost keys on the beach...) has been developed. See, for example, the influential work by H. Eugene Stanley and collaborators of 1999 ("Optimizing the success of random searches"[10]).

ii) *The Poisson process: A simple example of discrete Markov process*

Poisson process of intensity λ ($\lambda > 0$. Equation of evolution of discrete variable X

$$\frac{p(X=n)(t+\Delta t) - p(X=n)(t)}{\Delta t} = p(X=n-1)(t) - p(X=n)(t)$$

when Δt goes to zero, the probability distribution is given by

$$\mathbb{P}(X=n,t) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}, \;\; n \;\; \text{integer}$$

5. Stochastic process with memory effects (beyond Markov ones)

. Being almost never used in realistic simulations, they will not discussed here.

Appendix D: Derivation of G_0 by Fourier transform

The equation obeying $G_0(\mathbf{x}, \mathbf{x}', t)$ is

$$\frac{\partial G_0(\mathbf{x}, \mathbf{x}', t)}{\partial t} = \frac{1}{2}\nabla^2_{\mathbf{x}} G_0(\mathbf{x}, \mathbf{x}', t)$$

with $G_0(\mathbf{x}, \mathbf{x}', 0) = \delta(\mathbf{x} - \mathbf{x}')$.

Let use the Fourier representation of G_0

$$G_0(\mathbf{x}, \mathbf{x}', t) = \int d\mathbf{k} e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')} \tilde{G}(\mathbf{k}, t)$$

Injecting into the equation, we have

$$\frac{\partial \tilde{G}(\mathbf{k}, t)}{\partial t} = -\frac{\mathbf{k}^2}{2}\tilde{G}(\mathbf{k}, t)$$

with

$$\tilde{G}(\mathbf{k}, 0) = 1$$

The solution reads

$$\tilde{G}(\mathbf{k}, t) = e^{-\frac{\mathbf{k}^2}{2}t}$$

and

$$G_0(\mathbf{x}, \mathbf{x}', t) = \int d\mathbf{k} e^{\frac{t}{2}\left[\mathbf{k} - i\frac{(\mathbf{x}-\mathbf{x}')}{t}\right]^2} e^{-\frac{1}{2}\frac{(\mathbf{x}-\mathbf{x}')^2}{t}}$$

$$\boxed{G_0(\mathbf{x}, \mathbf{x}', t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(\mathbf{x}-\mathbf{x}')^2}{2t}}}$$

Appendix E: Derivation of the Metropolis algorithm, convergence properties, and statistical error

1. Derivation in the discrete case

- Def. 1 Probability distribution $\pi_i \geq 0$ i=1,N and $\sum_i \pi_i = 1$
- Def. 2 transition probability (or stochastic matrix) $P_{i\to j}$:

  i. $P_{i\to j} \geq 0$

  ii. $\sum_{j=1}^{N} P_{i\to j} = 1$ (independent on i)
- Def. 3 *Ergodic* transition probability

  $\forall i_0 \ \ \forall i$ there exist a non-zero probability that after a *finite* number of steps starting from i_0 we end at i.

- Def. 4 Stationary (or invariant) distribution π:

$$\sum_i \pi_i P_{i\to j} = \pi_j$$

Metropolis algorithm

Let $P^T_{i\to j}$ being a trial ergodic transition probability, then $P_{i\to j}$ defined as follows

$$\begin{cases} P_{i\to j} = P^T_{i\to j}\mathrm{Min}(1, R_{ij}) & \mathrm{j} \neq \mathrm{i} \\ P_{i\to i} = P^T_{i\to i} + \sum_{k\neq i} P^T_{i\to k}(1 - \mathrm{Min}(1, R_{ik})) & \mathrm{j} = \mathrm{i} \\ \text{with } \ R_{ij} = \frac{\pi_j P^T_{j\to i}}{\pi_i P^T_{i\to j}} \end{cases}$$

is an ergodic transition probability admitting π_i as stationary distribution.

Proof:

1. $P_{i\to j}$ is a transition probability

   - $P_{i\to j} \geq 0$ obvious
   - $\sum_{j=1}^{N} P_{i\to j} = \sum_{j\neq i} P_{i\to j} + P_{i\to i}$
     $= \sum_{j\neq i} P^T_{i\to j} + P^T_{i\to i}$
     $= 1$

2. Stationary distribution

   We have to show: $\sum_i \pi_i P_{i\to j} = \pi_j$

   For that we first show that $\{P_{i\to j}; \pi_i\}$ obeys detailed balance

   $$\pi_i P_{i\to j} = \pi_j P_{j\to i} \quad \forall(i,j)$$

   Proof:

   - i=j obvious
   - i≠j: the ratio of the two sides of the previous equality is

     $$\frac{\pi_j P_{j\to i}}{\pi_i P_{i\to j}} = \frac{R_{ij}\mathrm{Min}(1, R_{ji})}{\mathrm{Min}(1, R_{ij})}.$$

     Remarking that $R_{ij} = 1/R_{ji}$ and distinguising between the two cases corresponding to $R_{ij} \geq 1$ and $R_{ij} < 1$, we easily verify that this ratio is equal to 1.

   Finally, using the detailed balance relation we get

   $$\sum_i \pi_i P_{i\to j} = \sum_i \pi_j P_{j\to i} = \pi_j$$

   thus, π_i is the stationary distrbution.

2. Convergence of the Metropolis algorithm

Let us precise the way the distribution converges to the stationary one.

Let $f^{(k)}$ be a distribution, that is a set of N positive real numbers. The application of the stochastoc matrix to this distribution is written as

$$f_i^{(k+1)} = \sum_j f_j^{(k)} P_{j\to i} \equiv P f_i^{(k)}$$

We have the following property

$$\lim_{n\to\infty} f_i^{(n)} \sim P^n f_i^{(0)} = \pi_i \qquad \forall f^{(0)}$$

The different steps of the proof are as follows.

- Let us associate to $P_{i\to j}$ a *symmetric* real matrix defined as follows

$$M_{ij} = \sqrt{\pi_i} P_{i\to j} \frac{1}{\sqrt{\pi_j}}$$

  Let us insist that the stochastic matrix is in general *not symmetric*.

- It is easy to check that $\sqrt{\pi}$ is eigenstate of M with eigenvalue 1

$$\sum_j M_{ij}\sqrt{\pi_j} = \sqrt{\pi_i}$$

- We also see that

$$P^n f^{(0)} = \sqrt{\pi} M^n \frac{f^{(0)}}{\sqrt{\pi}}$$

- Let us now use the spectral decomposition of M. For large n, M^n becomes the projector in the eigenspace associated with the largest eigenvalue. Due to its particular structure, it can be shown that M has eigenvalues λ_i such that $0 \leq |\lambda_i| \leq 1$ and in the case where π does not vanish, the associated eigenspace is not degenerate. As a consequence

$$P^n f^{(0)} = c\ \pi$$

  where c is the overlap between the initial distribution $f^{(0)}/\sqrt{\pi}$ and the eigenstate $\sqrt{\pi}$ of matrix M.

3. The statistical error

The Metropolis algorithm is a simple and efficient algorithm for generating states distributed according to an arbitrary density. However, the price to pay for such a simplicity is the fact that the successive states produced are correlated. Accordingly, some care is needed when estimating the statistical error associated with the arithmetic averages computed. First of all, it is important to check that we are not in the transient regime associated

with the initial configuration used. Second we have to estimate the correlation time of the Markov chain.

Let $f(x)$ a quantity whose expectation value is to be computed, $I(f) = \int dx\pi(x)f(x)$ A unbiased estimator of the expectation value is the arithmetic sum

$$\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \tag{E1}$$

where n is a finite number of configurations drawn with the Metropolis algorithm. Note that $\bar{f}_n$ is a random variable and that its value depends on the series of random numbers used to generate the successive states of the sum. Unbiased means here that if the finite sum is computed an infinite number of times with different random realizations, then

$$\langle \bar{f}_n \rangle = \frac{1}{n}\sum_{i=1}^{n} \langle f(x_i) \rangle = I(f) \tag{E2}$$

Due to the central limit theorem valid for Markov process, we know that for sufficiently large n the distribution of the random variable $\bar{f}_n$ becomes gaussian

$$P(\bar{f}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(\bar{f}_n - \langle f_n \rangle)^2}{2\sigma_n^2}}$$

where

$$\sigma_n^2 = \langle \bar{f}_n{}^2 \rangle - \langle \bar{f}_n \rangle^2 \tag{E3}$$

Now, a practical way to compute the error bar is to realize a certain number of independent calculations of $\bar{f}_n$ and to estimate the variance of the distribution $P(\bar{f})$. Let N_b the number of independent calculations, we denote $\bar{f}_n{}^k \quad k = 1, N_b$, the values obtained for each calculation. Unbiased estimates of the mean and variance are

$$\langle \bar{f}_n \rangle = \frac{1}{N_b}\sum_{k=1}^{N_b} \bar{f}_n{}^k$$

and

$$\sigma_n^2 = \frac{1}{N_b - 1}\sum_{k=1}^{N_b} \left(\bar{f}_n{}^k - \langle \bar{f}_n \rangle\right)^2$$

An estimate of the statistical error δf on the estimate of $I(f)$ is then $\delta f = \frac{\sqrt{\sigma_n^2}}{\sqrt{N_b}}$, that is

$$\delta f = \frac{1}{\sqrt{N_b(N_b - 1)}} \sqrt{\sum_{k=1}^{N_b} \left(\bar{f}_n{}^k - \langle \bar{f}_n \rangle\right)^2} \tag{E4}$$

In practical calculations, the N_b calculations are never fully independent and some correlation are introduced. Such correlations can be explicited as follows. By inserting (E2) into (E3) we get

$$\sigma^2 = \frac{1}{n}[c_0 + 2\sum_{i=1}^{n-1}(1 - \frac{i}{n})c_i]$$

where

$$c_i = \langle f_k f_{k+i} \rangle - \langle f_k \rangle \langle f_{k+i} \rangle$$

(time translation implies independence on k). Calculation of the c_i can be performed by estimating the various correlators from the N_b realizations. Formula (E4) can be easily generalized using such correlators. For a discussion of such aspects, see for example[11].

Appendix F: Derivation of the importance-sampled diffusion equation

$$\frac{\partial \Psi(\mathbf{x},t)}{\partial t} = \frac{1}{2}\nabla^2\Psi(\mathbf{x},t) - V(\mathbf{x})\Psi(\mathbf{x},t) \tag{F1}$$

Let $\Psi_T(\mathbf{x})$ the (time-independent) trial wavefunction and introduce the mixed density f as follows

$$f(\mathbf{x},t) \equiv \Psi_T(\mathbf{x})\Psi(\mathbf{x},t)$$

Let us multiply each side of the equation by Ψ_T we have

$$\frac{\partial f}{\partial t} = \frac{1}{2}\Psi_T\nabla^2\frac{f}{\Psi_T} - Vf$$

Let us compute $\nabla^2\frac{f}{\Psi_T}$.

$$\nabla\frac{f}{\Psi_T} = \frac{\nabla f}{\Psi_T} - \frac{f\nabla\Psi_T}{\Psi_T^2}$$

It is convenient to use the drift vector

$$\mathbf{b} = \frac{f\nabla\Psi_T}{\Psi_T}$$

So we have $\nabla\frac{f}{\Psi_T} = \frac{1}{\Psi_T}\big[\nabla f - f\mathbf{b}\big]$ Then

$$\nabla^2\frac{f}{\Psi_T} = -\frac{\nabla\Psi_T}{\Psi_T^2}\big[\nabla f - f\mathbf{b}\big] + \frac{1}{\Psi_T}\big[\nabla^2 f - \nabla f\mathbf{b} - f\nabla\mathbf{b}\big]$$

and

$$\Psi_T\nabla^2\frac{f}{\Psi_T} = -\mathbf{b}\big[\nabla f - f\mathbf{b}\big] + \nabla^2 f - \nabla f\mathbf{b} - f\nabla\mathbf{b} = f\mathbf{b}^2 + \nabla^2 f - 2\mathbf{b}\nabla - f\nabla\mathbf{b}$$

Now, using

$$\mathbf{b}\nabla f = \nabla(\mathbf{b}f) - \nabla\mathbf{b}f$$

we finally get

$$\Psi_T\nabla^2\frac{f}{\Psi_T} = \nabla^2 f - 2\nabla(\mathbf{b}f) + (\nabla\mathbf{b} + \mathbf{b}^2)f$$

The time equation becomes

$$\frac{\partial f}{\partial t} = \frac{1}{2}\nabla^2 f - \nabla(\mathbf{b}f) + \Big[\frac{1}{2}(\nabla\mathbf{b} + \mathbf{b}^2)\Big] - Vf \tag{F2}$$

Now, the local energy is given as

$$E_L = \frac{-\frac{1}{2}\nabla^2}{\Psi_T} + V = -\frac{1}{2}(\nabla \mathbf{b} + \mathbf{b}^2) + V$$

So we get the final equation of evolution for f

$$\frac{\partial f}{\partial t} = \frac{1}{2}\nabla^2 f - \nabla(\mathbf{b}f) - E_L f$$

Appendix G: Short-time approximation for the drifted brownian

We want to show that

$$G = \langle \mathbf{x}|e^{\tau\frac{1}{2}\nabla^2.-\tau\nabla[\mathbf{b}.]}|\mathbf{x}_0 \sim_{\tau\to 0} \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}-\mathbf{x}_0-\mathbf{b}(\mathbf{x}_0)\tau)^2}{2\tau}}$$

With an error $O[\tau^2]$, we can "break" the exponential

$$e^{\tau\frac{1}{2}\nabla^2.-\tau\nabla[\mathbf{b}.]} \sim e^{\tau\frac{1}{2}\nabla^2.} e^{-\tau\nabla[\mathbf{b}.]}$$

Then, we introduce the resolution of the identity between the exponentials

$$\begin{aligned} G &= \int d\mathbf{x}_1 \langle \mathbf{x}|e^{\tau\frac{1}{2}\nabla^2.|\mathbf{x}_1\rangle}\langle \mathbf{x}_1|e^{-\tau\nabla[\mathbf{b}.]}|\mathbf{x}_0\rangle \\ &= \int d\mathbf{x}_1 \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}-\mathbf{x}_1)^2}{2\tau}} \langle \mathbf{x}_1|e^{-\tau\nabla[\mathbf{b}.]}|\mathbf{x}_0\rangle \end{aligned} \tag{G1}$$

Let us denote

$$G_b = \langle \mathbf{x}|e^{-t\nabla[\mathbf{b}.]}|\mathbf{x}_0\rangle$$

$G_b(t)$ obeys the equation

$$\frac{\partial G_b}{\partial t} = -\nabla[\mathbf{b}G_b]$$

with

$$G_b(0) = \delta(\mathbf{x}-\mathbf{x}_0)$$

At small time, it can be shown that it is possible to consider the drift constant [$= \mathbf{b}(\mathbf{x}_0) = \mathbf{b}_0$] on the time interval, the varying part of the drift entering at a subleading order in τ.

The drift being constant, we need to solve

$$\frac{\partial G_b}{\partial t} = -\mathbf{b}_0\nabla[G_b]$$

Taking the Fourier transform of the equation, we get

$$\frac{\partial \tilde{G}_b(\mathbf{k},t)}{\partial t} = i\mathbf{b}_0\mathbf{k}\tilde{G}_b(\mathbf{k},t)$$

that is

$$\tilde{G}_b(k,t) = e^{i\mathbf{b}_0\mathbf{k}t}$$

In real space, we get

$$G_b(t) = \delta[\mathbf{x}-t\mathbf{b}_0]$$

Finally, injecting into the previous form into Eq. G1 and integrating we get

$$G \sim \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}-\mathbf{x}_0-\mathbf{b}(\mathbf{x}_0)\tau)^2}{2\tau}}$$

Appendix H: Derivation of the path-integral formula for the mixed density $\phi_0(\mathbf{x})\Psi_T(\mathbf{x})$

$$G(\mathbf{x},\mathbf{x}_0,t) = \langle \mathbf{x}|e^{-tH}|\mathbf{x}_0\rangle$$

$$G(\mathbf{x},\mathbf{x}_0,t) = \lim_{N\to\infty \ \tau=\frac{t}{N}} \int d\mathbf{x}_1 ... d\mathbf{x}_{N-1} \prod_{i=0}^{N-1} G(\mathbf{x}_i,\mathbf{x}_{i+1},\tau)$$

Now, introducing the trial wave function Ψ_T we get

$$\frac{1}{\Psi_T(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0,t)\Psi_T(\mathbf{x}) = \lim_{N\to\infty} \int d\mathbf{x}_1 ... d\mathbf{x}_{N-1} \prod_{i=0}^{N-1} \left[\frac{\Psi_T(\mathbf{x}_{i+1})}{\Psi_T(\mathbf{x}_i)} G(\mathbf{x}_i,\mathbf{x}_{i+1},\tau)\right]$$

Using the definition of the exponential of an operator it is easy to show that

$$\frac{\Psi_T(\mathbf{x}_{i+1})}{\Psi_T(\mathbf{x}_i)} G(\mathbf{x}_i,\mathbf{x}_{i+1},t) = \frac{\Psi_T(\mathbf{x}_{i+1})}{\Psi_T(\mathbf{x}_i)} \langle \mathbf{x}_{i+1}|e^{-tH}|\mathbf{x}_i\rangle = \langle \mathbf{x}_{i+1}|e^{tL}|\mathbf{x}_i\rangle$$

where the operator L is defined as

$$L = -\Psi_T H \frac{1}{\Psi_T}$$

Now, see derivation in appendix G

$$\langle \mathbf{x}_{i+1}|e^{\tau L}|\mathbf{x}_i\rangle \sim_{\tau\to 0} = \langle \mathbf{x}_{i+1}|e^{\tau\frac{1}{2}\nabla^2 . - \tau\nabla[\mathbf{b}.]}|\mathbf{x}_i\rangle e^{-\tau E_L(\mathbf{x}_i)} = \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}_{i+1}-\mathbf{x}_i-\mathbf{b}(\mathbf{x}_i)\tau)^2}{2\tau}} e^{-\tau E_L(\mathbf{x}_i)}$$

We, then, write

$$\frac{1}{\Psi_T(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0,t)\Psi_T(\mathbf{x}) = \lim_{N\to\infty} \int d\mathbf{x}_1 ... d\mathbf{x}_{N-1} \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\tau}^d} e^{-\frac{(\mathbf{x}_{i+1}-\mathbf{x}_i-\mathbf{b}(\mathbf{x}_i)\tau)^2}{2\tau}} \prod_{i=0}^{N-1} e^{-\tau E_L(\mathbf{x}_i)}$$

At large t

$$\frac{1}{\Psi_T(\mathbf{x}_0)} G(\mathbf{x},\mathbf{x}_0,t)\Psi_T(\mathbf{x}) = \frac{\phi_0(\mathbf{x}_0)}{\Psi_T(\mathbf{x}_0)} \phi_0(\mathbf{x})\Psi_T(\mathbf{x}) e^{-tE_0} + O[e^{-t(E_1-E_0)}]$$

which finally gives, up to some unessential prefactor

$$\phi_0(\mathbf{x})\Psi_T(\mathbf{x}) \sim \sum_{\textbf{drifted}\text{brownian paths arriving at } \mathbf{x}} e^{-\int_0^{+\infty} ds \mathbf{E_L}[\mathbf{x(s)}]}$$

Appendix I: Trial wavefunctions for molecules

• *The multi-determinant Slater-Jastrow wavefunction.*. See main text.

• *Use of a backflow term.* In trial wavefunctions including backflow, the electron coordinate $\mathbf{r}_i$ is replaced by a quasi-particle (dressed) coordinate $\bar{\mathbf{r}}_i = \mathbf{r}_i + \sum_{j\neq i} \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j)$ and is introduced in Slater forms. Physically, this backflow displacement is supposed to reproduce the characteristic "flow pattern" where the quantum fluid is pushed out of the way in front of a moving particle and fills in the space behind it. For more details, see Ref.[12]

• *Resonating VB form and geminal forms.* Let Φ be the pairing function (geminal) which takes into account the correlation between two electrons with opposite spin. If the system is unpolarized and the state is a spin singlet, the antisymmetrized geminal product (AGP) wavefunction is

$$\mathbf{\Psi}_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \hat{A}[\Phi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_2^{\downarrow})\Phi(\mathbf{r}_3^{\uparrow}, \mathbf{r}_4^{\downarrow}) \cdots \Phi(\mathbf{r}_{N-1}^{\uparrow}, \mathbf{r}_N^{\downarrow})], \tag{I1}$$

where $\hat{A}$ is an operator that antisymmetrizes the product in the square brackets and the geminal is a singlet:

$$\Phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) \frac{1}{\sqrt{2}} \left(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle \right), \tag{I2}$$

implying that $\phi(\mathbf{r}, \mathbf{r}')$ is symmetric under a permutation of its variables. Given this conditions, one can prove that the spatial part of the $\mathbf{\Psi}_{AGP}$ can be written in a very compact form:

$$\Psi_{AGP}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = det(A_{ij}), \tag{I3}$$

where A_{ij} is a $\frac{N}{2} \times \frac{N}{2}$ matrix defined as:

$$A_{ij} = \phi(\mathbf{r}_i^{\uparrow}, \mathbf{r}_j^{\downarrow}). \tag{I4}$$

For more details, see Ref.[13]

• *Perturbatively selected Configuration Interaction expansion.* In quantum chemistry Configuration Interaction (CI) expansions are widely used. They allow a systematic improvement of the wavefunction through increase of the number of determinants and of the basis set used. In QMC the use of CI expansions is problematic due to the very large number

of determinants. Indeed, at each Monte Carlo iteration -and there can be as many as one billion of such elementary steps- the first and second derivatives (Laplacian) must be computed for the current electronic configuration. However, despite these drawbacks, CI expansions have nevertheless been recently employed in QMC. It is possible only because 1) the CI expansion is reduced by a suitable selection of the most important determinants[14,15] 2) efficient techniques have been developed to make the CI expansion computable in a reasonable time.[16–18]. Some applications can be found in Ref.[15],[19].

• *Valence Bond trial wavefunction.* The use of Valance Bond (VB) wavefunctions is very attractive in quantum chemistry. Indeed, VB forms give a simple and very appealing interpretation of the electronic structure in terms of Lewis pairs (bound pairs, lone pair, etc.). Unfortunately, from a technical point of view VB wavefunctions are made of non-orthogonal determinants, a point which dramatically increases the computational effort (passing from a standard N^3 law to a $N!$ law). A number of QMC works using VB wavefunctions have been presented, see Ref.[20–22]

• *Multi-Jastrow form* The so-called Multi-Jastrow is obtained by replacing the global Jastrow form into local Jastrows attached to one-particle molecular orbitals. Using such local forms allows to describe the electron-electron correlation in a more specific way (electron correlation is different into a 1s orbitals, 3d orbitals, polarizable lone pairs, etc.) See[23].

• etc. (any home-made approximate wavefunction can be easily used in QMC). Describe various type of wavefunctions that can be used

Appendix J: Optimization of the trial wavefunction

1. The problem

When accurate results are searched for, we need to reduce the two following errors:
(1) The statistical fluctuations related to the finite number of Monte Carlo steps
(2) The fixed-node bias related to the use of an approximate nodal hypersurface.

Both errors can be decreased by optimizing the parameters of the trial wavefunction. Different criteria can be used to define the "quality" of a trial wavefunction. The two most employed:

• Minimization of the variational energy

$$E(\mathbf{p}) = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

where $\mathbf{p}$ denotes the set of parameters of $\Psi_T(\mathbf{x}, \mathbf{p})$

• Minimization of the variance of the Hamiltonian

$$\sigma^2(\mathbf{p}) = \frac{\langle \Psi_T | [H - E(\mathbf{p})]^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

a. The correlated approach

The most natural idea to optimize the trial wavefunction is to minimize the total energy evaluated for a finite number of configurations N_c drawn in a preliminary Variational Monte Carlo step:

$$E(\mathbf{p}) \simeq \frac{1}{N_c} \sum_{i=1}^{N_c} E_L(\mathbf{x}_i)$$

In practice, such a idea is difficult to realize for two reasons:
(1) For a finite number of walkers $E(\mathbf{p})$ is not bounded from below and the minimizer can change parameters in a wird way so that to concentrate the wavefunction around one or a few points having a very low local energy.
(2) The stationary distribution, $\Psi_T^2(\mathbf{x}, \mathbf{p})$ depends on parameters $\mathbf{p}$, and thus new configurations must be redrawn at each change of parameters. The variational energy being calculated for a finite number of points, the energy curve $E(\mathbf{p})$ is then noisy and it is a tricky situation for the minimizer.

Practical solution:

(1) When a not too large number of configurations is used (a few thousand's) it is much preferable to minimize the variance since it is a quantity bounded from below ($\sigma^2 \geq 0$) for *any finite number* of configurations.

(2) To avoid the noisy character of $E(\mathbf{p})$ or $\sigma^2(\mathbf{p})$ a fixed set of configurations can be used and a correlated approach introduced?

$$\sigma^2(\mathbf{p}) = \frac{\frac{1}{N_c}\sum_{i=1}^{N_c} w_i (E_L - E)^2(\mathbf{x}_i, \mathbf{p})}{\frac{1}{N_c}\sum_{i=1}^{N_c} w_i}$$

where N_c number of configurations and $w_i = \frac{\Psi_T^2(\mathbf{x}_i,\mathbf{p})}{\Psi_T^2(\mathbf{x}_i,\mathbf{p}_0)}$. The configurations are drawn once for all according to $\Psi_T^2(\mathbf{x}_i, \mathbf{p}_0)$. In such conditions the energy curve is no longer noisy and standard minimizers (for example, quasi-Newton) can be employed.

Note that $\sigma^2(\mathbf{p})$ is a reasonable estimate of

$$\frac{\langle \Psi_T | (H-E)^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

only if the weights remain all close to one. It is thus important to quantify this aspect in some way, for example by introducing

$$\eta = \frac{1}{N_c} \frac{(\sum_i w_i)^2}{\sum_i {w_i}^2}$$

When η is close to one, the number of configurations playing a role is close to N_c and the estimation of the energy/variance is reasonable. In contrast, when only a few configurations contribute, η is close to zero and a new set of reference points must absolutely be drawn.

b. The linear method

The linear method has been recently introduced by Umrigar et al.[24] and is presently one of the most efficient approach to optimize a large number of parameters (both linear and non-linear).

The method is based on the minimization of the variational energy. Let us call N_p the number of parameters. The method consists in introducing a linear Taylor expansion around the current parameters p_0.

$$\Psi_T(\mathbf{x}, \mathbf{p}) = \Psi_T(\mathbf{x}, \mathbf{p}_0) + \sum_{i=1}^{N_p} (\mathbf{p} - \mathbf{p}_0)_i \Psi_i \qquad \text{(J1)}$$

where the functions Ψ_i are defined as

$$\Psi_i = \frac{\partial \Psi_T(\mathbf{x}, \mathbf{p}_0)}{\partial p_i}$$

Functions Ψ_i are now considered as a basis for the trial wavefunction and the energy is minimized in this basis set. Remarking that the Ψ_i are not orthogonal, the problem to solve is thus a generalized eigenvalue problem

$$H\Delta\mathbf{p} = ES\Delta\mathbf{p} \tag{J2}$$

where H and S are the Hamiltonian and overlap matrices, respectively.

$$H_{ij} = \langle\Psi_i|H|\Psi_j\rangle \text{ and } S_{ij} = \langle\Psi_i|\Psi_j\rangle$$

These quantities can be calculated in a VMC calculation using Ψ_0^2 as stationary distribution

$$H_{ij} = \langle\frac{\Psi_i}{\Psi_0}\frac{H\Psi_j}{\Psi_0}\rangle_{\Psi_0^2}$$

and

$$S_{ij} = \langle\frac{\Psi_i}{\Psi_0}\frac{\Psi_j}{\Psi_0}\rangle_{\Psi_0^2}$$

REFERENCES

[1]W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum monte carlo simulations of solids," Rev. Mod. Phys. **73**, 33–83 (2001).

[2]J. Kolorenc and L. Mitas, Rep. Prog. Phys. **74**, 026502 (2011).

[3]C. J. Umrigar, "Observations on variational and projector monte carlo methods," The Journal of Chemical Physics **143**, 164105 (2015), https://doi.org/10.1063/1.4933112.

[4]J. Toulouse, R. Assaraf, and C. J. Umrigar, "Introduction to the variational and diffusion Monte Carlo methods," Advances in Quantum Chemistry Electron Correlation in Molecules – ab initio Beyond Gaussian Quantum Chemistry, **73**, 285 (2016).

[5]M. Caffarel and R. Assaraf, "A pedagogical introduction to quantum monte carlo," Lecture Notes in Chemistry , 45 (2000).

[6]K. Schmidt and J. Moskowitz, "Correlated monte carlo wave functions for the atoms he through ne," J. Chem. Phys. **93**, 4172 (1990).

[7]P. L'Ecuyer, "Combined multiple recursive random number generators," Operations Research **44(5)**, 816–822 (1996).

[8]M. Krein and M. Rutman, "Linear operators leaving invariant a cone in a banach space," Uspehi Mat. Nauk **23**, 3–95 (1948).

[9]W. Coffey, Y. Kalmykov, and J. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry, and Electrical Engineering* (World Scientific, Singapore, 2004).

[10]G. M. Viswanathan, S. Buldyrev, S. Havlin, M. da Luz, E. Raposo, and H. Stanley, "Optimizing the success of random searches," Nature **401**, 911–914 (1999).

[11]U. Wolff, Comput. Phys. Commun. **156**, 143 (200).

[12]P. L. Rios, A. Ma, N. Drummond, M. Towler, and R. Needs, "Inhomogeneous backflow transformations in quantum monte carlo," Phys. Rev. E **74**, 066701 (2006).

[13]M. Casula, C. Attaccalite, and S. Sorella, "Correlated geminal wave function for molecules: An efficient resonating valence bond approach," J. Chem. Phys. **121**, 7110 (2004).

[14]E. Giner, A. Scemama, and M. Caffarel, "Using perturbatively selected configuration interaction in quantum monte carlo calculations," Can. J. Chem. **91**, 879 (2013).

[15]M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, "Using cipsi nodes in diffusion monte carlo," in *Recent Progress in Quantum Monte Carlo* (American Chemical Society (ACS), 2016) Chap. 2, pp. 15–46, http://pubs.acs.org/doi/pdf/10.1021/bk-2016-1234.ch002.

[16]B. K. Clark, M. A. Morales, J. McMinis, J. Kim, and G. E. Scuseria, "Computing the energy of a water molecule using multideterminants: A simple efficient algorithm," J. Chem. Phys. **135**, 244105 (2011).

[17]G. L. Weerasinghe, P. L. Ríos, and R. J. Needs, "Compression algorithm for multideterminant wave functions," Physical Review E **89** (2014), 10.1103/physreve.89.023304.

[18]A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, "Quantum monte carlo with very large multideterminant wavefunctions," Journal of Computational Chemistry **37**, 1866–1875 (2016).

[19]M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, J. Chem. Phys. **144**, 151103 (2016).

[20]A. Anderson and W. III, "Generalized valence bond wave functions in quantum monte carlo," J. Chem. Phys. **132**, 164110 (2010).

[21]F. Fracchia, C. Filippi, and C. Amovilli, "Size-extensive wave functions for quantum monte carlo: A linear scaling generalized valence bond approach," J. Chem. Theory Comput. **8**,

1943 (2012).

[22]B. Braida, J. Toulouse, M. Caffarel, and C. Umrigar, "Quantum monte carlo with jastrow valence-bond wave functions: application to bond breaking of some first-row diatomic molecules," J. Chem. Phys. **134**, 0184108 (2011).

[23]T. Bouabca, B. Braida, and M. Caffarel, J. Chem. Phys. **133**, 044111 (2010).

[24]C. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).