MICHEL CAFFAREL

Talking Across Fields: A Physicist’s Presentation of some Mathematical Aspects of Quantum Monte Carlo Methods


<http://afst.cedram.org/item?id=AFST_2015_6_24_4_949_0>

© Université Paul Sabatier, Toulouse, 2015, tous droits réservés.

L’accès aux articles de la revue « Annales de la faculté des sciences de Toulouse Mathématiques » (http://afst.cedram.org/), implique l’accord avec les conditions générales d’utilisation (http://afst.cedram.org/legal/). Toute reproduction en tout ou partie de cet article sous quelque forme que ce soit pour tout usage autre que l’utilisation à fin strictement personnelle du copiste est constitutive d’une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.
Talking Across Fields: 
A Physicist’s Presentation 
of some Mathematical Aspects 
of Quantum Monte Carlo Methods

MICHEL CAFFAREL(1)

Résumé. — Dans cet article nous discutons quelques aspects mathématiques des méthodes Monte Carlo quantique du point de vue du physicien. Une liste (non-exhaustive) de techniques probabilistes utilisées et développées en physique et de problèmes ouverts est présentée. Afin d’illustrer l’approche des physiciens, nous décrivons en détail une des variantes des méthodes Monte Carlo quantique basée sur la formule de Feynman-Kac. Le problème de l’efficacité numérique au cœur des applications physiques où l’espace de configuration est en général de très grande dimension est présenté, ainsi que la solution adoptée. Finalement, nous explicitons les contraintes spécifiques associées à la simulation des systèmes de fermions et présentons le fameux “problème du signe” considéré comme un des problèmes les plus importants à résoudre en physique numérique.

Abstract. — This paper discusses some mathematical aspects related to the use of probabilistic techniques in quantum Monte Carlo (QMC) methods from a physicist’s point of view. A selected list of problems and techniques employed in computational physics and of interest to the applied probability community is presented. One of the variants of QMC approaches based on the Feynman-Kac formula is described in some detail. The problem of numerical efficiency at the heart of physical applications defined in (very) high-dimensional space is discussed and the commonly used solution through importance sampling is presented. Finally, the specific constraints related to fermion systems in QMC are presented and the celebrated “fermion sign problem” (considered as one of the most important open problem in computational physics) is discussed.

(1) Lab. de Chimie et Physique Quantiques, CNRS-Université de Toulouse, France.
1. Introduction

Quantum Monte Carlo (QMC) is a broad family of numerical approaches widely used in computational physics and aiming at solving the Schrödinger equation using stochastic techniques. Ultimately, physical quantities are written as expectation values of random variables taking values in a state space $\mathbb{R}^D$ and admitting a density $\pi(x)$ with respect to Lebesgue measure $dx$. A critical feature of applications considered in physics is the high-dimensional nature of the state space (say, $D$ greater at least than a few thousands and most often much more). Tight constraints are thus imposed on what kind of probabilistic techniques can be efficiently used in practice. In Monte Carlo simulations of classical systems where $\pi$ is typically the Boltzmann distribution or some variation of it, the dimension is already large, $D = dN$ (or $2dN$ when velocities are considered), where $N$ is the number of particles and $d$ the dimension of the space in which particles live (usually $d = 3$). When the quantum character is considered, the space-time formalism of quantum physics based on path-integrals [31], [32] can be used to express once again the physical properties as expectation values of random variables. However, in order to get a stochastic interpretation of such path-integrals (leading to Feynman-Kac-type formulae), a so-called Wick’s rotation must be first introduced, a transformation in the complex plane that substitutes the physical real time $t$ into a mathematical imaginary time $-it$. As far as we are concerned with stationary solutions (that is, independent on time), quantum averages of interest are insensitive to this rotation. By using a path-integral formalism, the state space is now made of all possible time-paths for the particles and thus becomes infinite-dimensional (the interested reader may find a mathematical presentation of all these aspects close to the physicists’view in Glimm and Jaffe [40]). In actual simulations, paths are discretized using a finite time-step and quantum expectation values are then expressed back as ordinary finite-dimensional integrals in $\mathbb{R}^D$ where the new “quantum” dimension $D$ is much increased. Denoting $P$ the number of time intervals along the paths (stochastic trajectories), $D$ is given as the product of the $dN$ dimensions of the classical state space times the “time” integer $P$. Exact quantum expectation values are formally obtained for infinitely long trajectories ($P = +\infty$) and infinitely small time-steps. In practice, when $P$ is chosen large enough but finite and the time-step small enough, the bias on expectation values can be made smaller than the statistical error.

To summarize, physical properties of a $N$-particle system can always be written as an expectation value $\mathbb{E}_\pi[f(X)]$ where $X$ is a random variable distributed according to $\pi(x)dx$ and $f$ a real-valued function defined over $\mathbb{R}^D$. At the classical level, $D$ is finite and is proportional to the number
of particles. At the quantum level, $D$ is formally infinite but can be made finite at the price of a small and controlled error.

An important aspect to consider in quantum simulations is the bosonic or fermionic nature of the particles. For bosons, QMC algorithms are particularly well suited (see, e.g. ref [23]) and are nowadays considered as state-of-the-art approaches. Unfortunately, in the important case of fermions (including all applications involving electrons, such as, among others, the domains of material sciences, nano-sciences and chemistry) the situation is different. Mathematically, dealing with fermions imposes to restrict the set of all possible paths to a subset of “fermionic paths” compatible with the Pauli exclusion principle (that is, leading to a wavefunction with the suitable antisymmetry properties under the exchange of fermions). Despite much effort and many proposals, it has not been possible so far to devise a probabilistic algorithm allowing to compute fermionic expectation values, that is both stable (bounded variance) and exact (unbiased estimators). This problem -known under the name of “sign problem” is of uttermost practical importance and is viewed as one of the most important problems to be solved in computational many-body physics [63, 51, 71, 66]. To contribute to its solution is an exciting challenge both for the physical and mathematical communities.

The physical properties being ultimately expressed as expectation values of random variables, virtually all probabilistic techniques and algorithms employed in physics have also been considered in applied probability theory. Making the bridge between both communities is therefore important. In this spirit, we now present a list of probabilistic techniques and problems addressed in many-body physics. In each case, some references from the physics literature are given. Although this list is by no way exhaustive and strongly reflects the author’s interests, we nevertheless hope that it will be of some help to the interested mathematically minded reader.

Metropolis-Hastings algorithm in QMC. As in most domains of natural sciences, the Metropolis-Hastings (MH) algorithm [54],[44] is widely used in computational physics and to give a fair account of its various applications is just impossible. Here, we restrict ourselves to mention three important uses. In variational Monte Carlo (VMC) it is employed to compute expectation values with respect to the quantum-mechanical probability density associated with an approximate trial wavefunction $[\pi(x) = \Psi^2(x)$, assuming normalized wavefunctions] as invariant distribution [24]. In applications VMC is mostly used to explore the respective quality of approximate wavefunctions with different physical contents and thus to get insight into the nature of the quantum state studied. Another application is the calculation of ex-
act ground-state properties expressed as path-integrals, the density sampled being now the weight of each (Brownian) path, see for example the Reptation Monte Carlo (RMC) approach [14]. A third important application is the computation of quantum statistical averages at finite-temperature. The inverse temperature ($\beta = \frac{1}{k_B T}$, $k_B$ Boltzmann constant) playing now the role of the imaginary time of ground-state applications ($\beta = it$), thermodynamic averages can also be written as a sum of over paths. Such approaches are usually referred to as Path-Integral Monte Carlo (PIMC) methods [23]. Let us emphasize that in each case (VMC, RMC, PIMC) the probability to be sampled is known explicitly, thus making Metropolis a natural method.

**Use of diffusion processes.** Although in principle the Metropolis-Hastings algorithm can always be employed to sample the quantum path density, the use of continuous diffusion processes (in particle coordinates) may be preferred for a number of practical and theoretical reasons. This can done by introducing a (forward) Fokker-Planck (FP) operator built from the original Hamiltonian. In most implementations the FP operator consists of a diagonal and constant diffusion matrix and a drift vector expressed as the logarithmic derivative of some good-quality approximation of the density. The sample paths of the process are built using an explicit Euler scheme for the associated Stochastic Differential Equation (SDE). Note also that in practical implementations a Metropolis rule is added to get an unbiased stationary density despite time-discretization. Important examples of QMC approaches where a continuous diffusion processes is used are the popular Diffusion Monte Carlo (DMC) methods [41, 2, 59] (a birth-death process is also introduced, see below).

**Langevin equation.** In order to get a better sampling of the configuration space alternative stochastic dynamics can be introduced. One such example is the recent proposal [61] of extending the diffusion process defined in the space of particle positions to the classical phase space including both particle positions and momenta. In this case the equation governing the stochastic evolution of the system is the Langevin equation with inertia derived from the Newtonian classical mechanics. It can be shown that the detailed balance condition can be imposed to the dynamics and thus a Metropolis algorithm can be used.

**Branching or birth-death processes.** To enhance the convergence of estimators a branching process may be introduced. Its role is to kill sample paths spending too much time in low-probability regions and to duplicate trajectories in high-probability regions. To keep the population size reasonable, a population control step must be introduced. The price to pay is then the introduction of correlations between paths (interacting particles model).
Non-trivial mathematical aspects are related to the way of imposing this control, usually via a nonlinear feedback term \([59]\). Widely employed Diffusion Monte Carlo methods introduce such a branching process \([41, 2, 59]\).

**Feynman-Kac approaches.** From a general perspective, all variants of QMC approaches can actually be rewritten using Feynman-Kac-type formulae expressed under various forms. However, in practice most of QMC methods do not explicitly mention such a relationship. Counterexamples are \([30, 60, 18]\). An example of QMC method using explicitly Feynman-Kac formalism is described below in some detail.

**Reconfiguration process.** To allow a better mathematical control of birth-death-type events, a stochastic reconfiguration process may also be introduced (a reconfiguration step is defined here as choosing \(N\) objects among \(N\) objects according to some probabilities). In this process the population size is kept fixed and the finite population bias is removed through the introduction of a global Feynman-Kac weight for the population. Such methods can be viewed as a combination of DMC and Feynman-Kac approaches and are known under the name of Stochastic Reconfiguration Monte Carlo (SRMC) methods \([45, 64, 4]\).

**Correlating close stochastic processes.** A great variety of important physical properties are expressed as a (very) small difference of two (very) close expectation values

\[
\delta_\epsilon = \mathbb{E}_{\pi_\epsilon}[f_\epsilon(X)] - \mathbb{E}_\pi[f(X)]
\]

where \(\epsilon\) is a small parameter connecting the distributions \(\pi\) and \(\pi_\epsilon\), and the functions \(f\) and \(f_\epsilon\). Certainly, by far the most important example is the case where \(\mathbb{E}[f(X)]\) is the total energy of the system. To be able to compute efficiently such small differences is an important issue of computational physics (see, discussion in ref \([10]\)). Most of the approaches proposed so far are based on the natural idea of correlating the two close stochastic processes involved in the difference (see, e.g. refs \([33, 10, 46, 5]\)). However, much work remains to be done and this important issue must still be considered as an open problem in the general case.

**Improved estimators.** As in any Monte Carlo approach it is desirable to build improved estimators (or control variates) with (much) lower variances. A number of proposals developed within the context of QMC approaches have been presented \([6-9, 11]\). Among them, let us point out the Zero-Variance Monte Carlo approach described in \([6]\) which turns out to be of very broad applicability (see, applications in the applied probability community \([29, 55, 57, 37, 56]\)).
Bayesian inference. Bayesian approaches and max-entropy methods have been developed and adapted for modeling the noisy data of QMC, see e.g. refs [42], [20].

Stochastic optimization. An important step in most QMC applications is to optimize the parameters of an analytically known trial wavefunction chosen to be a good approximation of the unknown eigenfunction. In general it is done by minimizing the variational energy (Rayleigh quotient of operator $H$) computed over a random sample. Such an optimization step involving a large number of parameters whose majority are non-linear and a noisy objective function to minimize is not easy to perform in the general case. A number of techniques have been proposed and applied to make this optimization step as efficient as possible within the framework of QMC approaches [67, 43, 65, 68].

Diffusion processes in a functional space. The vast majority of diffusion processes introduced in QMC are defined in the classical space of particle positions (position representation of quantum mechanics). Alternative representations more adapted to the quantum nature of particles may also be considered. Along this line, a number of works have introduced diffusion processes defined in a space of determinants. See e.g. the Auxiliary Field QMC (AFQMC) method of Zhang et al. [70, 72, 1] and the recent FCIQMC method of Alavi and collaborators [16], [28].

etc.

Of course, entering into the details of these various items is out of the scope of the present account. Instead, we have chosen to focus on the implementation of one of the variants of the QMC approaches based on the Feynman-Kac formula. This well-known formula in the probabilistic community will allow us to illustrate two of the most fundamental issues of QMC in realistic applications, namely the need for efficiency in high-dimensions and the challenging problem of imposing the fermionic nature of particles within a stochastic framework, this latter problem being known as the “fermion sign problem”. Before doing this we shall briefly present in Section II the mathematical notations and properties used here. In section III the standard Feynman-Kac formula is derived in the way it is usually done in physics. In section IV, how to generalize this formula to make it efficient in high dimensions is discussed. Finally, in section V the fermion sign problem and the standard approximate solution employed in practical applications (fixed-node approximation) are presented.
Quantum systems are described by a self-adjoint Hamiltonian operator, $H$ defined in the Hilbert space, $L^2(\mathbb{R}^{dN}, dx)$. Most Hamiltonians are written as a sum of a kinetic operator (responsible for the delocalization of particles) and a potential operator describing the interaction between particles. To make things explicit, we shall consider the following standard form for $H$ (Schrödinger-type Hamiltonian)

$$
H = -\frac{1}{2} \nabla_x^2 + V(x)
$$

(2.1)

where $x = (r_1, r_2, ..., r_N)$ denotes the position of each of the $N$ particles ($x \in \mathbb{R}^{dN}$). The Laplacian operator is written as

$$
\nabla_x^2 = \sum_{i=1}^{N} \sum_{l=1}^{d} \frac{\partial^2}{\partial x_i^l^2}
$$

(2.2)

where $(x_1^1, x_1^2, ..., x_1^d)$ in the $d$-uplet giving the position $r_i$ of each particle $i$.

In QMC we are concerned with bound states (not scattering states, see note [73]) that is, eigensolutions $\Psi$ of the stationary Schrödinger equation verifying

$$
\int dx |\Psi|^2 = 1
$$

(2.3)

In such a case the spectrum of $H$ consists entirely of isolated eigenvalues [47] and the eigensolutions can be labeled by an integer $i$ (traditionally, $i = 0$ for the ground-state). The stationary Schrödinger equation to be solved is finally written as

$$
H\Psi_i(x) = E_i \Psi_i(x)
$$

(2.4)

with $E_0 \leq E_1 \leq \cdots$. In QMC methods discussed here we are essentially interested in evaluating $(E_0, \Psi_0)$ and eventually the first low-lying eigensolutions, not the entire spectrum of the Hamiltonian.

In most applications $H^* = H$ (with respect to $L^2(\mathbb{R}^{dN}, dx)$). The reality of $H$ results from the fact that $V$ is real-valued (an important exception not considered here is in the presence of a magnetic field). $H$ can thus be restricted to acting on real valued functions in $L^2(\mathbb{R}^{dN}, dx)$ and thus no complex conjugate asterisk on eigenstate will be introduced in the formulae to follow.
Let us denote $K(x, y, t)$ the kernel of the linear operator $e^{-tH}$ defined through
\[
(e^{-tH}\Psi)(x) = \int dy K(x, y, t)\Psi(y)
\]
where $\Psi$ is some test function.

In the physics literature, $K$ is usually referred to as the imaginary-time propagator (probability amplitude of evolving from $x$ to $y$ in a time interval $t$). Using the spectral decomposition of the operator $e^{-tH}$ we have
\[
K(x, y, t) = \sum_i \Psi_i(x)\Psi_i(y)e^{-tE_i}
\]
where $\sum$ is a short-hand notation for representing the sum and integral over the discrete and continuous parts of the eigenspectrum, respectively (the series is supposed to converge). Using this relation, it is easy to show that the kernel $K$ obeys the Chapman-Kolmogorov [58] equation
\[
K(x, y, t) = \int dz K(x, z, t-u)K(z, y, u) \quad u \in (0, t)
\]

3. The Feynman-Kac formula

In physics the Feynman-Kac is generally derived in two steps as follows. First, the Chapman-Kolmogov equation, Eq.(2.7), is iterated by breaking the total time $t$ into small pieces. Let $\tau$ be a small but finite time-step such that $t = P\tau$, $P$ integer. We can write
\[
K(x_0, x_t, t) = \int d\mathbf{x}_1...d\mathbf{x}_{P-1} \prod_{k=1}^P K(x_{k-1}, x_k, \tau)
\]
where the initial point at $t = 0$ is denoted as $x_0$ and the final point at time $t$, $x_t$. Here, $\mathbf{x}_P$ is identified to $x_t$.

The second step consists in introducing a short-time approximation for the propagator $K(x_{k-1}, x_k, \tau)$. Using the Baker-Campbell-Hausdorff formula [13] allowing to split the exponential operator of a sum of two general operators into a product of two exponentials
\[
e^{-\tau(A+B)} = e^{-\tau A}e^{-\tau B} + O(\tau^2)
\]
the short-time propagator can be written
\[
K(x_{k-1}, x_k, \tau) = K_0(x_{k-1}, x_k, \tau)e^{-\tau V(x_k)} + O(\tau^2)
\]
where $K_0$ is the free diffusion kernel of the kinetic (Laplacian) operator

$$K_0(x_{k-1}, x_k, \tau) = \left( \frac{1}{\sqrt{2\pi T}} \right)^d \int dN e^{-\frac{(x_k-x_{k-1})^2}{2\tau}}$$

(3.4)

Finally, the exact kernel $K(x_0, x_t, t)$ can be written in the limit $\tau$ goes to zero and $P$ to infinity with $t = P\tau$ finite as

$$K(x_0, x_t, t) = \lim_{P \to +\infty} \int dx_1 ... dx_{P-1} \prod_{k=1}^P K_0(x_{k-1}, x_k, \tau) e^{-\tau \sum_{k=1}^P V(x_k)}$$

(3.5)

which is formally written as

$$K(x_0, x_t, t) = K_0(x_0, x_t, t) \mathbb{E}_{X(0)=x_0, X(t)=x_t} [e^{-\int_0^t ds V[X(s)]}]$$

(3.6)

where the expectation value is defined over the set of the Brownian trajectories $X(s)$ starting at $x_0$ and ending at $x_t$ at time $t$. More rigorously, for any bounded test function $\phi(x)$ we can write

$$\int \phi(x_t) K(x_0, x_t, t) dx_t = \mathbb{E}_{X(0)=x_0} [\phi(X_t) K_0(x_0, X_t, t) e^{-\int_0^t ds V[X(s)]}]$$

(3.7)

where $t \mapsto X_t$ is a standard Brownian motion. This equation is the Feynman-Kac (FK) formula for the kernel.

As a consequence of Eq.(2.6) the long-time behavior of the kernel reads

$$K(x_0, x_t, t) \sim e^{-tE_0} \Psi_0(x_0) \Psi_0(x_t) \quad \text{at large } t$$

(3.8)

up to some exponentially small corrections $O[e^{-t(E_1-E_0)}]$. From this relation, expressions for the exact ground-state wavefunction and energy can be derived. In the case of the exact wavefunction the value of $\Psi_0$ at point $x$ can be obtained as (see, note [74])

$$\frac{\Psi_0(x)}{\int dx \Psi_0(x)} = \lim_{t \to +\infty} \frac{\int dy K(x, y, t)}{\int dx dy K(x, y, t)}$$

(3.9)

which leads to the expression

$$\frac{\Psi_0(x)}{\int dx \Psi_0(x)} = \lim_{t \to +\infty} \frac{\mathbb{E}_{X(0)=x} [e^{-\int_0^t ds V[X(s)]}]}{\int dx \mathbb{E}_{X(0)=x} [e^{-\int_0^t ds V[X(s)]}]}.$$  

(3.10)

In the case of the ground-state energy, $E_0$ may be obtained as

$$E_0 = - \lim_{t \to +\infty} \frac{1}{t} \ln \left[ \int dx \mathbb{E}_{X(0)=x} [e^{-\int_0^t ds V[X(s)]}] \right].$$

(3.11)
From a computational point of view FK-type formulae are particularly attractive since they provide a simple and constructive way of obtaining the lowest eigensolutions of $H$ in any dimension just by constructing $dN$-dimensional Brownian trajectories, by computing the exponential of the time-integral of the potential function along each trajectory (a simple one-dimensional integral) and, finally, by averaging contributions coming from a large enough number of trajectories. Furthermore, calculations on different trajectories being independent, it may lead to massively parallel simulations (for example, one trajectory built on each compute core).

However, at this point it should be clear that such a scheme cannot work in practice since there is no hope to get converged estimators in high-dimension using a naive uniform sampling of the configuration space. To get realistic estimators with reasonably small variances requires some sort of “clever” sampling taking into account the magnitude of the potential function (regions of low $V$ must be preferentially sampled). A practical solution to this problem is presented in the next section.

4. A computationally realistic Feynman-Kac formula through importance sampling

To allow a realistic sampling of the configuration space, an approximation $\Psi_T$ (called the trial wavefunction) of the unknown exact solution $\Psi_0$ is introduced. A well-known property of Schrödinger operators is that the ground-state wavefunction has a constant sign (say, positive) and vanishes only at infinity [Krein-Rutman theorem, generalization of the Perron-Frobenius theorem to operators [50]. Note that the positivity of the ground-state wavefunction can be directly seen from Eq.(3.10). As a consequence, $\Psi_T$ is chosen here strictly positive at any finite distance. In the following section we shall see that this condition must be released for fermionic systems and that some specific treatment will be required.

Let us now define a generalized (similarity-transformed) kernel as follows

$$\tilde{K}(x, y, t) \equiv \frac{\Psi_T(y)}{\Psi_T(x)} K(x, y, t)$$

(4.1)

It is easy to check that this new kernel also verifies the Chapman-Kolmogorov relation, Eq.(2.7), and thus a FK-type representation for it can be built as in Eq.(3.1). The second step consists in getting a short-time approximation of the generalized kernel, similarly to Eq.(3.3). Recalling that $K$ was the kernel of the operator $e^{-tH}$

$$(e^{-tH}\Psi)(x) = \int dy K(x, y, t) \Psi(y)$$

- 958 –
and using definition, Eq.(4.1), $\tilde{K}$ is easily found to be the kernel of the exponential operator $e^{tL}$

$$(e^{tL}\Psi)(x) = \int dy \tilde{K}(x, y, t)\Psi(y)$$  (4.2)

where $L$ is given through the following similarity transformation

$$L = -\frac{1}{\Psi_T}H\Psi_T.$$  (4.3)

This transformation is the unitary transformation of the ambient Hilbert space $L^2(\mathbb{R}^{dN})$ obtained by multiplication with respect to the function $\Psi_T$. Note that the spectrum of $H$ is left unmodified.

To explicit the operator $L$, we shall first define a trial Hamiltonian $H_T$ defined as the Schrödinger Hamiltonian admitting $\Psi_T$ as ground-state wave-function. The trial Hamiltonian is easily built and reads

$$H_T = H - [E_L(x) - E_T]$$  (4.4)

where $E_T$ is some arbitrary reference energy and $E_L(x)$ a multiplicative operator known in QMC under the name of “local energy”

$$E_L(x) = \frac{H\Psi_T(x)}{\Psi_T(x)}.$$  (4.5)

Finally, $\Psi_T$ obeys the following Schrödinger equation

$$H_T\Psi_T(x) = E_T\Psi_T(x).$$  (4.6)

Using the trial Hamiltonian the operator $L$ can be decomposed as

$$L = -\frac{1}{\Psi_T}(H_T - E_T)\Psi_T - E_L(x).$$  (4.7)

Simple algebra shows that the first operator, $-\frac{1}{\Psi_T}(H_T - E_T)\Psi_T$, in the right-hand-side of this equation is a Fokker-Planck operator written as

$$L_T = \frac{1}{2}\nabla_x^2 + b(x)\cdot\nabla_x$$  (4.8)

where the drift vector $b$ is given by

$$b(x) = \frac{\nabla_x\Psi_T}{\Psi_T}.$$  (4.9)

Approximations for the solution of the Fokker-Planck equation at short time are well-known. They are simple generalizations of the Gaussian free
diffusion kernel, Eq.(3.4) in presence of a drift term. At the lowest order we have
\[ \tilde{K}_T(x_{k-1}, x_k, \tau) \simeq \left( \frac{1}{\sqrt{2\pi \tau}} \right)^d N e^{-\frac{(x_k-x_{k-1}-b(x_{k-1})\tau)^2}{2\tau}} \] (4.10)

Using as previously the Baker-Campbell-Hausdorff formula, Eq.(3.3) generalizes as follows
\[ \tilde{K}(x_{k-1}, x_k, \tau) = \tilde{K}_T(x_{k-1}, x_k, \tau)e^{-\tau E_L(x_k)} + O(\tau^2) \] (4.11)

A generalized version of the Feynman-Kac formula now follows by iterating the Chapman-Kolmogorov equation and using the previous short-time expression
\[ \tilde{K}(x_0, x_t, t) = \lim_{P \to +\infty} \int dx_1...dx_{P-1} \prod_{k=1}^P \tilde{K}_T(x_{k-1}, x_k, \tau)e^{-\tau \sum_{k=1}^P E_L(x_k)} \] (4.12)

which is formally written as
\[ \tilde{K}(x_0, x_t, t) = \tilde{K}_T(x_0, x_t, t)\tilde{E}X(0)=x_0, X(t)=[e^{-\int_0^t ds E_L[X(s)]}] \] (4.13)

where the new expectation value \( \tilde{E} \) is defined now over the sample paths of the continuous diffusion process described by the FP operator. The trajectories can be generated using the associated stochastic differential equation (SDE)
\[ dX(t) = b[X(t)]dt + dW_t \] (4.14)

where \( W_t \) is the \( dN \)-dimensional Wiener process. In practice, some discretized version of the SDE is employed, for example
\[ X_{k+1} = X_k + b(X_k)\tau + \sqrt{\tau} \eta, \] (4.15)

where \( \eta \) is a random vector whose each independent component has a normal distribution. It should be noted that formula (4.15) is just a practical way of realizing the Gaussian distribution given in Eq.(4.10). In QMC simulations it is common usage to call ”walker” the representative point \( X_k \) of each trajectory. In contrast with some other fields (e.g. use of particle methods for the simulation of continuous systems of fluid mechanics) the term “particle” is not employed here to avoid confusion with the physical particles.

A new estimator for the wavefunction using the FK formula can be obtained from (recall that \( \Psi_T(x) \) is a known computable function)
\begin{align}
\frac{\Psi_T(x) \Psi_0(x)}{\int dx \Psi_T(x) \Psi_0(x)} &= \lim_{t \to +\infty} \frac{\int dy \Psi_T(x) K(x, y, t) \Psi_T(y)}{\int dx \int dy \Psi_T(x) K(x, y, t) \Psi_T(y)} \\
&= \Psi_T^2(x) \lim_{t \to +\infty} \frac{\mathbb{E}_{X(0) = x} [e^{-\int_0^t ds E_L[X(s)]]}}{\int dx \Psi_T^2(x) \mathbb{E}_{X(0) = x} [e^{-\int_0^t ds E_L[X(s)]]}} \quad (4.16)
\end{align}

and \( E_0 \) can be computed as

\begin{align}
E_0 &= -\lim_{t \to +\infty} \frac{1}{t} \ln \left[ \int dx \Psi_T^2(x) \mathbb{E}_{X(0) = x} [e^{-\int_0^t ds E_L[X(s)]]} \right]. \quad (4.17)
\end{align}

The free diffusion process introduced in the preceding section is a special case of the more general process considered here. All formulas can be recovered by taking \( \Psi_T = 1 \), \( b = 0 \), and \( E_L = V \). The introduction of a trial wavefunction \( \Psi_T \) has two important practical consequences. Instead of the uniform sampling resulting from Brownian paths, the configuration space is now sampled through drifted Brownian paths realizing a stationary (invariant) distribution close to the exact quantum-mechanical one. The invariant distribution \( \pi \) (verifying \( \int \pi \tilde{K}_T = \pi \), that is \( L^\ast \pi = 0 \), where \( L^\ast \) is the adjoint of operator \( L \)) is given as (normalized distribution)

\begin{equation}
\pi(x) = \Psi_T^2. \quad (4.18)
\end{equation}

In practice, because of the drift term walkers are pushed into regions where the trial wavefunction is large. Second, the fluctuations of the exponential Feynman-Kac weight are considerably reduced. The variations are no longer determined by the “bare” potential \( V \) but by the new “renormalized” (screened) potential, \( E_L \). The magnitude of the fluctuations is directly related to the quality of the trial function, the better it is the smaller the fluctuations of the local energy are. In the limit of an exact trial wavefunction the statistical fluctuations entirely vanish (zero-variance property, see ref [35]).

A number of applications using the approach presented (or some equivalent version of it) have shown that in practice “good” enough trial wavefunctions can be built to make this scheme work with the accuracy needed, even in very high dimensions.

5. The sign problem

In the preceding sections the FK formulas have been elaborated without taking care of the eventual fermionic nature of particles. As it is, this algorithm can be directly employed for quantum systems with no Fermi constraints (bosonic systems, quantum oscillators, ensemble of distinguishable
particles, etc.). In such cases, the ground-state eigenfunction is nodeless (say, positive) and may be obtained from Eq.(4.16). Unfortunately, for fermionic systems such an eigenstate is physically forbidden by the Pauli exclusion principle [69].

For fermions the functional space of wave functions is divided into two orthogonal spaces

$$L^2(\mathbb{R}^{dN}) = B \oplus F$$

(5.1)

where $F$ is the vector space of “fermionic” wavefunctions defined as follows:

$$\Psi \in F \text{ if and only if } \Psi[\sigma(x)] = \text{sgn}(\sigma)\Psi(x)$$

(5.2)

where $\sigma$ ranges in some permutation subgroup of the symmetric group $S_N$ leaving invariant some 2-subsets partition of $\{1, ..., N\}$ (corresponding to “spin up” or “spin down” electrons). In particular, all totally skew-symmetric functions are in this case. $B$, the vector space of “bosonic” wavefunctions, is then simply the orthogonal of $F$. In particular, all totally symmetric functions are in $B$.

The Pauli principle can then be summarized by saying that the “fermionic” eigensolutions of $H$ physically admissible are those obtained by restricting the Hamiltonian to the vector space $F$. In particular, the totally symmetric nodeless lowest eigenstate of $H$ is forbidden for fermions (the so-called “bosonic” ground-state).

Note that in contrast with standard presentations of the Pauli exclusion principle, no spin coordinates have been introduced here. Actually, at the non-relativistic level such coordinates are not needed, see e.g. [69], [53]. However, they are of common use since within a spin-space representation the Pauli exclusion principle is particularly simple to express. The eigenstates are written as a combination of space and spin functions and only those that are totally antisymmetric under the exchange of space-spin coordinates of any pair of particles are physically allowed. In a spin-free (space-only) formalism as used here, the spatial wavefunctions $\Psi(x)$ just need to be antisymmetric under permutations within two subsets of particles that can be formally associated with spin “up” and “down” particles.

Because the Schrödinger Hamiltonian is spin-independent and the diffusion processes introduced are defined in a pure space representation, the use of spin coordinates is not adapted and is thus avoided in QMC.

Finally, the problem to solve in QMC is to design an efficient algorithm allowing to converge to the ground-state fermionic eigenfunction (lowest eigenstate of $H$ restricted to vector space $F$). Unfortunately, up to now it
has not been possible to define a computationally tractable (polynomial) algorithm implementing exactly such a property for a general fermionic system ("sign problem"). This problem is addressed in the two following sections.

5.1. Exact Feynman-Kac formula for fermions

To go further it is convenient to distinguish the fermionic and bosonic parts of the spectral decomposition of the kernel, eq (2.6).

\[ K(x, y, t) = \sum_i \Psi_i^B(x)\Psi_i^B(y)e^{-tE_i^B} + \sum_i \Psi_i^F(x)\Psi_i^F(y)e^{-tE_i^F} \]  

(5.3)

where the superscripts F (Fermion) and B (Boson) refer to eigensolutions that belong to vector space \( F \) or \( B \), respectively.

The objective now is to evaluate \( E_0^F \) and \( \Psi_0^F \). As in the preceding section [see, eq.(4.1)] we introduce a similarity-transformed kernel \( \tilde{K} \) using a positive and totally symmetric wavefunction \( \Psi_G \) (bosonic-type wave function).

\[ \tilde{K}(x, y, t) = \frac{\Psi_G(y)}{\Psi_G(x)} K(x, y, t). \]  

(5.4)

Here, \( \Psi_G \) will be referred to as the “guiding” wavefunction. Note that, in contrast with the preceding section 4 where the Feynman-Kac formula was derived without considering any type of symmetry, the positive bosonic guiding function is now distinguished from the fermionic trial wavefunction \( \Psi_T^F \) that will be chosen to be a good approximation of the fermionic ground-state, \( \Psi_0^F \).

By filtering out the bosonic components of the generalized kernel, the unknown fermionic ground-state \( \Psi_0^F \) can be extracted from the large-time limit

\[ \frac{\Psi_T^F(x)\Psi_0^F(x)}{\int dx\Psi_T^F(x)\Psi_0^F(x)} = \lim_{t \to +\infty} \frac{\int dy\Psi_T^F(x)K(x, y, t)\Psi_T^F(y)}{\int dx\int dy\Psi_T^F(x)K(x, y, t)\Psi_T^F(y)}. \]  

(5.5)

Using the Feynman-Kac formula, \( \Psi_0^F \) can be obtained as (up to the known and computable trial wavefunction \( \Psi_T^F \))

\[ \frac{\Psi_T^F(x)\Psi_0^F(x)}{\int dx\Psi_T^F(x)\Psi_0^F(x)} = \]  

\[ \lim_{t \to +\infty} \frac{\tilde{E}_{\Psi_0^F(x)} = \int [\rho(x)\rho(t)]w[x(t)]w[x(t)]e^{-\int_0^t dsE_{\Psi_0^F}^G[x(s)]}}{\int dx\Psi_T^F(x)\tilde{E}_{\Psi_0^F(x)} = \int [\rho(x)\rho(t)]w[x(t)]w[x(t)]e^{-\int_0^t dsE_{\Psi_0^F}^G[x(s)]}} \]  

(5.6)
where $\sigma$ is the sign of the fermionic trial wavefunction

$$\sigma(x) = \text{sgn} [\Psi^F_T(x)], \quad (5.7)$$

$w$ some positive weight

$$w(x) = \frac{|\Psi^F_T(x)|}{\Psi_G(x)}, \quad (5.8)$$

and $E^G_L$ the local energy associated with $\Psi_G$

$$E^G_L(x) = \frac{H \Psi_G(x)}{\Psi_G(x)}. \quad (5.9)$$

As usual [see, Eqs.(3.11),(4.17)] the ground-state energy can be computed from the long-time behavior of the logarithm of the denominator of the right-hand-side of eq.(5.6).

To get an efficient sampling of the configuration space, the nodeless wavefunction $\Psi_G$ that determines the sampled distribution ($\pi = \Psi^2_G$) must be chosen close to the modulus of the fermionic trial wavefunction $\Psi^F_T$, for example

$$\Psi_G(x) = \sqrt{(\Psi^F_T(x))^2 + \epsilon(x)} \quad (5.10)$$

where $\epsilon(x)$ is some “small” strictly positive function decaying sufficiently rapidly at large distances. Note that with such a choice $w(x)$ is a positive function close to 1. From formula (5.6) we can see that the fermion ground-state is expressed as the expectation value of the (weighted) time auto-correlation function of the fermionic sign, $\sigma[\mathbf{X}(0)]\sigma[\mathbf{X}(t)]$, at very large time.

Let us now decompose the expectation values into their positive and negative contributions. We can write

$$\Psi^2_G(x) \lim_{t \to +\infty} \frac{\tilde{E}_{\mathbf{X}(0)=x,\sigma_0\sigma_t=1}[F_t] - \tilde{E}_{\mathbf{X}(0)=x,\sigma_0\sigma_t=-1}[F_t]}{\int d\mathbf{x} \Psi^2_G(\mathbf{x}) \Psi^F_T(\mathbf{x})} =$$

$$\Psi^2_G(x) \lim_{t \to +\infty} \frac{\tilde{E}_{\mathbf{X}(0)=x,\sigma_0\sigma_t=1}[F_t] - \tilde{E}_{\mathbf{X}(0)=x,\sigma_0\sigma_t=-1}[F_t]}{\int d\mathbf{x} \Psi^2_G(\mathbf{x})} \quad (5.11)$$

where the positive functional $F_t$ is defined as

$$F_t[\mathbf{X}(s)] = w[\mathbf{X}(0)]w[\mathbf{X}(t)]e^{-\int_0^t ds E^G_L[\mathbf{X}(s)]} \quad (5.12)$$

and the new expectation values, $\tilde{E}_{\mathbf{X}(0)=x,\sigma_0\sigma_t=\pm 1}$ are introduced by making a distinction between the set of all trajectories verifying $\sigma_0\sigma_t = 1$ and those
verifying $\sigma_0 \sigma_t = -1$. Both expectation values can be obtained from eq.(4.16) by applying the kernel to the positive and negative part of $\Psi_T$

$$\Psi_T^{\pm} = \text{Max}(0, \pm \Psi_T^F)$$

(5.13)

instead of $\Psi_T^F$ itself. $\Psi_T^{\pm}$ being positive, the long-time behavior of the expectation values of the denominator of (5.11) is governed by the bosonic ground-state

$$\mathbb{E}_{\mathbf{x}(0)=\mathbf{x}, \sigma_0 \sigma_t = \pm 1}[F_t] \sim e^{-tE^B_0} \quad \text{at large } t,$$

(5.14)

with the same behavior for the expectations value of the denominator. In contrast, the fermionic expectation value we are interested in behaves as

$$\mathbb{E}_{\mathbf{x}(0)=\mathbf{x}}[\sigma_0 \sigma_t F_t] \sim e^{-tE^F_0} \quad \text{at large } t.$$  

(5.15)

The fermionic wavefunction is thus obtained as the ratio of the difference of two bosonic-type expectation values, each of them increasing in time as $e^{-tE^B_0}$ (by convention, the energy of bound-states for a coulombic potential is negative), that is exponentially large with respect to the fermionic contribution by a factor of $e^{t(E^F_0 - E^B_0)}$ (recall that $E^F_0 > E^B_0$). In practice, this situation is particularly unfavorable since the statistical error related to the expectation value of the square (variance) of the weighted sign will also increase in a bosonic way. Finally, the signal-to-noise ratio (SNR) of fermionic QMC calculations is found to vanish exponentially at large times

$$\text{SNR} \sim e^{-t\Delta} \quad \text{at large } t$$

(5.16)

where $\Delta = E^F_0 - E^B_0 > 0$ is the so-called Fermi-Bose gap. For systems having a small enough gap, converged fermionic expectation values can be obtained. It is the case, for example, for very small molecules [25] or for sufficiently uniform systems such as the electron gas [27]. However, in the general case the gap is large and increases polynomially with the number of fermions. In this situation to get converged QMC calculations becomes exponentially difficult and impossible in practice, this is the fermion sign-problem.

5.2. Approximate Feynman-Kac formula for fermions: The Fixed-Node approximation

In the exact fermionic QMC algorithm just presented, the stochastic trajectories are able to reach any point in the configuration space (ergodic property). It is true because the invariant distribution, $\pi = \Psi^2_G$, is integrable and strictly positive at finite distances [48]. In practice, $\Psi_G$ being chosen
close to $|\Psi_T|$ [see, eq.(5.10)] the trajectories remain trapped some time in a
given nodal domain (around some maximum of $|\Psi_T|$) and then, from time
to time, leave the domain for a neighboring one, and so on. The fixed-node
approach consists in forcing the trajectories to stay in the nodal domain
where they started from. This is easily realized by imposing the positive
guiding function $\Psi_G$ to vanish at the boundaries of the nodal cells, for
example by taking $\epsilon = 0$ in (5.10)

$$\Psi_G = |\Psi_T|. \quad (5.17)$$

With such a choice the drift vector guiding the walkers diverges at the nodes
and the nodal boundaries play now the role of infinitely repulsive barriers;
the stochastic trajectories are trapped forever within their nodal domain.

Under such circonstances, we have

$$\sigma[\mathbf{X}(0)]\sigma[\mathbf{X}(t)] = 1 \quad \forall t \quad \text{and} \quad w = 1 \quad (5.18)$$

for each trapped trajectory and the Feynman-Kac formula, eq.(5.6), becomes

$$\frac{\Psi_T(x)\Psi_0^{FN}(x)}{\int dx \Psi_T(x)\Psi_0^{FN}(x)} = \Psi_T^2(x) \lim_{t \to +\infty} \frac{\hat{E}_{\mathbf{X}(0)=x}[e^{-\int_0^t ds E_L[\mathbf{X}(s)]]}}{\int dx \Psi_T^2(x)\hat{E}_{\mathbf{X}(0)=x}[e^{-\int_0^t ds E_L[\mathbf{X}(s)]}]} \quad (5.19)$$

for all point $x$ lying in the nodal domain considered. Here, $\Psi_0^{FN}$ denotes
the Fixed-Node (FN) ground-state eigenfunction obtained by imposing the
nodal boundaries to $\Psi_G$. Due to its very construction the fixed-node so-
lution has the same sign as the trial wavefunction ($\Psi_T(x)\Psi_0^{FN}(x) \geq 0$).
The fermionic problem defined over the entire configuration space $\mathbb{R}^{dN}$ is
thus recast in a sum of independent bosonic-type problems defined in each
nodal volume cut by the nodes of the approximate trial wavefunction. In-
stead of defining a unique Fokker-Planck operator with a non-divergent drift
vector over all space, a set of independent FP operators restricted to each
nodal cell domain is considered. Transposed into the original Hamiltonian
problem, it means that the Schrödinger equation is solved independently in
each nodal cell (mathematically, the $N$-body Schrödinger ground-state
is computed with additional Dirichlet boundary condition on the nodal set
$\mathcal{N}$ where $\Psi_T^F$ vanishes, $\mathcal{N} = \{ x \in \mathbb{R}^{dN} : \Psi_T^F(x) = 0 \}$. In the general case,
the zeroes of the trial wavefunction do not coincide with those of the un-
known fermionic eigensolution and we are thus left with a systematic bias,
the fixed-node error.

At this point, several important theoretical and practical aspects of the
fixed-node approximation must be discussed.
Mathematical foundation of the fixed-node approach. A mathematical analysis of the fixed-node approach and the justification of the statements given above can be found in Cancès et al. [22] and Rousset [60]. A convenient framework to analyze the fixed-node approach is to express it as a variational problem in the functional space of anti(skew)-symmetric functions with Dirichlet-type boundary conditions.

The tiling theorem. By solving the Schrödinger equation as a juxtaposition of independent problems, there is no reason why ground-state energies computed separately in each domain should be identical. The fixed-node energy is defined as the minimum of such energies. Unfortunately, in QMC calculations for non-trivial systems, the minimum found may depend on the initial conditions in the case where not all nodal domains are sampled, a situation that may arise since the number and localization of such domains in high dimension is in general not known. Hopefully, for fermionic ground-states Ceperley [26] has proved under physically reasonable conditions the existence of a tiling theorem for the exact ground-state: There is only one distinct kind of nodal regions. All others are related to it by permutational symmetry (with same energy). Unfortunately, in practice we need that Ψ_T satisfies the tiling property, not just the unknown ground-state. In actual simulations, it is generally assumed that Hartree-Fock or Kohn-Sham-type wavefunctions satisfy the tiling property. Results seem to validate such a statement. However, some (mathematical) work is needed to clarify this point.

Minimization of the fixed-node error. In a great variety of applications it has been found that -although small- the fixed-node error is still too large. To define an efficient and systematic approach to reduce it, is thus an important practical aspect. Up to now the best strategy involves two steps. A first one concerns the choice of the functional form for the trial wavefunction. To get accurate nodes the trial wavefunction must be built so as to incorporate as much as possible the main physical and mathematical features of the exact wavefunction. An intense activity has thus been developed to introduce and test various functional forms for Ψ_T taking into account important aspects of the wavefunction (e.g. refs [62, 34, 21, 12, 52, 3, 17, 15, 36, 39]). Once the trial wavefunction form has been chosen the next step consists in optimizing the many parameters (both linear and non-linear) of Ψ_T. It is usually done by minimizing the variational energy (or Rayleigh quotient, \(\int \Psi_T^* H \Psi_T \int \Psi_T^2\)) computed stochastically using a finite random sample. As already noticed in the introduction, a number of methods have been developed to make this difficult step as efficient as possible [67, 43, 65, 68]. Of course, it would be much more satisfactory to directly minimize the fixed-node error.
energy instead of miminizing the variational one. However, how to do that efficiently is still an open problem (see, for example, the proposal in [60]).

**Nodal properties.** Very little is known about the nodes of exact eigenfunctions (see, references cited in [26] and [19]). If exact nodes were known exact fermionic simulations would be possible and the sign problem would be solved. Unfortunately, to have a complete knowledge of the zeroes of a general antisymmetric wavefunction defined in \( \mathbb{R}^{dN} \) is an unsolved problem. In particular, we emphasize that the constraints resulting from the Pauli exclusion principle (the exact wavefunction must vanish when \( \mathbf{r}_i = \mathbf{r}_j \) for spin-like electrons) are not sufficient to determine the \((dN - 1)\)-dimensional nodal variety [49].

**Acknowledgments.** — This article is dedicated to Persi Diaconis as a souvenir of our weekly discussions in Toulouse that were so friendly and fruitful. I would like to thank Laurent Miclo for his careful reading of the manuscript and his useful suggestions. I also thank the referee for his/her invaluable comments which significantly contributed to improve the article. Finally, I acknowledge support by the Agence Nationale pour la Recherche (ANR) of our QMC project through Grant No ANR 2011 BS08 004 01.

**Bibliography**


Talking Across Fields


Talking Across Fields

Krein (M.G.) and Rutman (M.A.). — Linear operators leaving invariant a cone in a Banach space (Russian), Uspehi Mat. Nauk 3, No. 1, 23 p.3-95 (1948).


Scemama (A.), Lelièvre (T.), Stoltz (G.), Cancès (E.), and Caffarel (M.), An efficient sampling algorithm for Variational Monte Carlo J. Chem. Phys. 125, 114105 (2006).


In most applications to real systems, coulombic-type electrostatic potentials are considered. With such potentials the eigenstates of the Hamiltonian are square-integrable or not. In the first case the eigenstate describes a so-called bound state (the system remains localized in space) while in the second case we are in presence of a so-called scattering state (such states play a role, for example, in experiments where particle collision or diffusion are involved). For the systems considered in QMC applications, the first (lowest) eigenstates (including the ground-state) are bound states associated with isolated eigenvalues, while at higher energy a continuous eigenspectrum can be present (scattering states). In their standard setting, QMC approaches have been developed for bound states only.

Here, we suppose $\int dx \Psi_0(x) \neq 0$ which is true since $\Psi_0$ has a constant sign, see discussion in Sec.[5].