I. (IMAGINARY-)TIME DEPENDENT VARIATIONAL MONTE CARLO (IN CONTINUOUS SPACE)

Markus Holzmann

LPMMC, CNRS and UGA, Grenoble,

markus.holzmann@grenoble.cnrs.fr

(Dated: July 8, 2020)

My first set of lecture notes prepared for the VAQUM20.

A. Time dependent Schrödinger equation

One of the major challenges in computational physics is to provide accurate solutions to the many-body Schrödinger equation

$$i\partial_t \Psi(\mathbf{R}, t) = H\Psi(\mathbf{R}, t) \tag{1}$$

for a generic system described by the non-relativistic Hamiltonian

$$H = -\frac{\nabla^2}{2m} + V(\mathbf{R}) + V_{ext}(\mathbf{R}), \quad \nabla^2 \equiv \sum_{i=1}^{N} \nabla_i^2 \tag{2}$$

of interacting particles

$$V(\mathbf{R}) = \sum_{i<j} v(r_{ij}) \tag{3}$$

within some external potential

$$V_{ext}(\mathbf{R}) = \sum_{i=1}^{N} v_{ext}(\mathbf{r}_i) \tag{4}$$

Whereas for non-interacting particles, solutions of the many-body problem can be directly obtained from the corresponding single particle Schroödinger equation, difficulties generally arise from the presence of a non-vanishing interparticle interaction introducing correlations between the N particles, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots \mathbf{r}_N)$. In the following I will therefore mainly focus on homogeneous systems with vanishing or spatially periodic (crystal) external potential, imposing periodic boundary conditions on the wave function, aiming to describe extended matter in the macroscopic (thermodynamic) limit.

To be concrete, I list some examples of physical systems realized in nature which we aim to describe.

- *Jellium.* An important model for condensed matter is the homogeneous electron gas where electrons interact with each other by the bare Coulomb interaction, $v(r) = e^2/r$, but the interacting with the positive ions $v_{ext}(r)$ is reduced to a homogeneous positive charge background such that charge neutrality of the whole system is guaranteed.

- *Electronic structure in solids.* Here the interaction of the electrons with the nuclei in solids enters in the electronic (Born-Oppenheimer) Schrödinger equation via an external periodic lattice potential.

- *Electronic structure in atoms and molecules.* In contrast to the above examples which aim to describem macroscopic systems (thermodynamic limit), here one adresses localized systems with fixed number of electrons and fixed nuclear position in space (quantum chemistry).

- *Ultracold atomic gases.* Phases and properties of atoms interacting with a phenomenological two-body interaction between two (neutral) atoms are adressed. The interaction is frequently given by a single (or a few) well defined quantities of scattering theory, e.g. the s-wave scattering length, and can be mimicked by any short-range potential (hard-core, gaussian, etc) given the same asymtotic scattering properties for two atoms.

- *Helium liquids and solids.* The phenomenological interaction between two helium atoms is very well known, to first order given by a hard-core or Lennard-Jones interaction, but more precise interactions are known combining knowlegd from quantum chemistry and experiment. Helium liquids and solids are the prototypes of quantum liquids and solids.

1. Exact Diagonalization (ED)

General solutions of the time-dependent Schrödinger equation can be found by expanding in the eigenstates of the Hamitonian. In order to determine them exactly, we can expand the full wave function in terms of some single-particle orthonomal basis set $\phi_n(\mathbf{r})$

$$\Psi(\mathbf{R}) = \sum_{n_1,\ldots,n_N} c_{n_1,n_2,\ldots,n_N} \phi_{n_1}(\mathbf{r}_1)\phi_{n_2}(\mathbf{r}_2)\cdots\phi_{n_N}(\mathbf{r}_N) \tag{5}$$

truncate the basis to include only states with $n_i \leq M$, use the orthonormality between different basis states, and determine the energy eigenfunctions from diagonalizing the Hamiltonian matrix

$$H_{nm} = \int d\mathbf{R}\phi_{n_1}^*(\mathbf{r}_1)\phi_{n_2}^*(\mathbf{r}_2)\cdots\phi_{n_N}^*(\mathbf{r}_N)H\phi_{m_1}(\mathbf{r}_1)\phi_{m_2}(\mathbf{r}_2)\cdots\phi_{m_N}(\mathbf{r}_N) \tag{6}$$

It is clear that building the matrix including only few states becomes rapidly unaffordable already for very few number of particles. Does this mean that precise solutions of the many-body problem are intrinsically out of reach?

In order to understand the problem let us consider a homogeneous gas of few particles inside a (periodic) box. As basis states, we may use plane wave states, $\phi_n(x) \sim e^{ik_nx}$, or, equivalently, simply discretize space, $\phi_n(x) \sim \delta_{x,n\Delta}$ leading to a lattice model. Increasing the number of plane wave states corresponds to a more dense discretization of space. For single particle systems, we then fill up uniformly the space in our box. However, in the many-body problem, we uniformly fill up a box in a much higher dimensions, since one configuration corresponds to the N-particle vector. Using an expansion in single particle states, all particles can reach all positions independently. Even within a very crude discretization, we immediatly include states where all or almost all particles sit in the same corner of the box. Our physical expectation tells us that the probability to find such situations to be exponentially small, so that the amplitude of this state should be neglegible in the exact wave function. Even worse, the more states we include into our diagonalization, the more we include such physically irrelevant states. Increasing the single particle basis set for ED, we thus end up with most of our states sitting in irrelevant corner regions of the many-body Hilbert space.

Although inefficient for large systems, exact diagonalization is of course essential to provide exact and unbiased bechmark results for systems with few degress of freedom.

2. Variational approach – ground state

Trying to address the limit of large number of particles, we clearly should (a) avoid any explicit calculation of the Hamiltonian in terms of a matrix and (b) look for basis states suited to describe the relevant degrees of freedom of the problem. Both of these issues are practically calling for a variational approach. (Of course, (a) can be adressed using iterative solver where, instead of the full matrix only the state vector is stored, but even storing vectors will soon request too much memory rescources.)

How one would expect a typical many-body wave function to look like? In a classical gas, a typical configuration of particles is given by the Boltzmann distribution, $\sim \exp[-\beta V(\mathbf{R})]$. In the classical limit, we would therefore expect the dominante amplitudes in the density matrix (and therefore also of a typical wave function) to have a similar functional form

$$\Psi_T(\mathbf{R}) \sim \exp\left[-\sum_{i<j} u(r_{ij})\right] \tag{7}$$

However, after all, we are adressing quantum systems! Although we may accept that the system ressembles a classical gas on large scales, details certainly don't match, so setting $u(r)$ proportional to the classical potential energy $v(r)$ might not be so a good idea. Instead, we can get the a-priori best form of $u(r)$ for the ground state by minimizing the variational energy

$$E_T = \frac{\int d\mathbf{R}\Psi_T^*(\mathbf{R})H\Psi_T(\mathbf{R})}{\int d\mathbf{R}|\Psi_T(\mathbf{R})|^2} \tag{8}$$

with respect to $u(r)$, e.g. by expressing $u(r) = \sum_n^M \alpha_n\phi_n(r)$ in a set of (1D) basis functions $\phi_n(r)$ and minimizing the corresponding function with respect to the parameters α_n, $n = 1,\ldots,M$. Later, we will present more concrete motivations for this ansatz and discuss how to improve it systematically.

At the optimal parameters, $\widetilde{\alpha}$, the variational energy is stationary, $\partial_{\alpha_n} E_T|_{\widetilde{\alpha}} = 0$, and the trial energy $\widetilde{E}_T$ provides the best upper bound for the ground state energy we can obtain. Since the energy plays a central role in physics, the trial energy gives us a unique criteria to decide which is the best wave function. However, up to what extend does this "best" wave function also represent well ground state properties other than the energy? There is in general no guarantee, as one can easily construct counterexamples, e.g. double well potential or Bose-Fermi mapping in 1D, where the ground state energy can be arbitrarily closely approached with a trial wave function having the wrong symmetry.

To get a bit more insight, let us expand the trial wave function around the minimum,

$$\Psi_T(\mathbf{R}|\alpha) = \widetilde{\Psi}_T(\mathbf{R}) + \sum_n \delta\alpha_n \partial_{\alpha_n} \widetilde{\Psi}_T(\mathbf{R}) + \ldots \tag{9}$$

with $\widetilde{\Psi}_T(\mathbf{R}) \equiv \Psi_T(\mathbf{R}|\widetilde{\alpha})$ and $\delta\alpha_n \equiv \alpha_n - \widetilde{\alpha}_n$ we get

$$E_T(\delta\alpha) = \widetilde{E}_T + \sum_n \left[\frac{\int d\mathbf{R}\widetilde{\Psi}_T^*(\mathbf{R})(H - \widetilde{E}_T)\partial_{\alpha_n}\widetilde{\Psi}_T(\mathbf{R})}{\int d\mathbf{R}|\widetilde{\Psi}_*(\mathbf{R})|^2} \delta\alpha_n + cc \right] \tag{10}$$

From the stationary condition of the trial energy at the optimal variational parameters $\widetilde{\alpha}$, we know that the first order terms in $\delta\alpha_n$ and $\delta\alpha_n^*$ must vanish separately (equivalently, we can split $\delta\alpha_n = \delta\alpha_n' + i\delta\alpha_n''$ into its real and imaginary part and the resulting expressions for real and imaginary part must vanish independently). We therefore must have

$$\int d\mathbf{R} \left[\partial_{\alpha_n} \widetilde{\Psi}_T^*(\mathbf{R}) \right] H\widetilde{\Psi}_T(\mathbf{R}) = \widetilde{E}_T \int d\mathbf{R} \left[\partial_{\alpha_n}^* \widetilde{\Psi}_T(\mathbf{R}) \right] \widetilde{\Psi}_T(\mathbf{R}) \tag{11}$$

for all components α_n independently.

Let us consider the vector space spanned by the functions $\{\widetilde{\Psi}_T(\mathbf{R}), \partial_{\alpha_1}\widetilde{\Psi}_T(\mathbf{R}), \ldots \partial_{\alpha_M}\widetilde{\Psi}_T(\mathbf{R})\}$, the above equation tells us that $\widetilde{\Psi}_T(\mathbf{R})$ can be considered as the solution to the Schrödinger equation

$$H\widetilde{\Psi}_T(\mathbf{R}) = \widetilde{E}_T\widetilde{\Psi}_T(\mathbf{R}) + r(\mathbf{R}) \tag{12}$$

within this vector space, e.g.

$$\int d\mathbf{R} \left[\partial_{\alpha_n}\widetilde{\Psi}_T^*(\mathbf{R}) \right] r(\mathbf{R}) = 0 \tag{13}$$

for all n. We see that the variational approach corresponds to determining the lowest eigenvalue obtained within the basis spanned by the variational state and its derivatives with respect to the variational paramters. We may therefore expect that observables O with $O(\mathbf{R})\widetilde{\Psi}_T(\mathbf{R})$ having major overlap inside this vector space will be accurately represented. On the otherhand, we should certainly not trust any properties of the trial wave function, where the corresponding observable is out of reach, e.g. by symmetry, even if the trial energy may be exact.

3. Variational approach – dynamics

Let us now turn to adress the time evolution of a quantum-many-body wave function, assuming that its initial state at $t = 0$ is well described by a variational wave function $\Psi(\mathbf{R}|\alpha(0))$, the time-dependence can always be expressed through the time-dependence of the variational parameters, $\Psi(\mathbf{R}, t) = \Psi_{\alpha(t)}(\mathbf{R}) \equiv \Psi(\mathbf{R}|\alpha(t))$, if we enlarge the set of "variational parameters" to include a global time-dependent phase and normalization term.

Plugging our time-dependent variational wave function into the time-dependent Schrödinger equation, we obtain

$$i\partial_t \Psi_{\alpha(t)}(\mathbf{R}) = H\Psi_{\alpha(t)}(\mathbf{R}) + r(\mathbf{R}, t) \tag{14}$$

where the reminder $r(\mathbf{R}, t)$ vanishes if our variational wave function is exact. In order to determine the a-priori best time evolution of our parameters, it is natural to extend our previous insights and consider the vector space which can be reached from our variational state within infinitesimal small time which naturally lies within the space spanned by the partial derivatives with respect to the variational parameters.

$$\Psi_{\alpha(t+\delta t)}(\mathbf{R}) = \Psi_{\alpha(t)}(\mathbf{R}) + \sum_n (\alpha_n(t + \delta t) - \alpha_n(t))\partial_{\alpha_n}\Psi_{\alpha(t)}(\mathbf{R}) \tag{15}$$

We then may determine the time dependence of the parameters by projecting the time-dependent Schrödinger equation into this linearized parameter space, e.g. we require

$$\int d\mathbf{R} \left[\partial_{\alpha_n} \Psi^*_{\alpha(t)}(\mathbf{R}) \right] r(\mathbf{R}, t) = 0 \tag{16}$$

as a necessary (but not sufficient) condition that the variational wave function satisfies the Schrödinger equation. We then get the following equations

$$i \int d\mathbf{R} \Psi^*_{\alpha(t)}(\mathbf{R}) \partial_t \Psi_{\alpha(t)}(\mathbf{R}) = \int d\mathbf{R} \Psi^*_{\alpha(t)}(\mathbf{R}) H \Psi_{\alpha(t)}(\mathbf{R}) \tag{17}$$

$$i \int d\mathbf{R} \left[\partial_{\alpha_n} \Psi^*_{\alpha(t)}(\mathbf{R}) \right] \partial_t \Psi_{\alpha(t)}(\mathbf{R}) = \int d\mathbf{R} \left[\partial_{\alpha_n} \Psi^*_{\alpha(t)}(\mathbf{R}) \right] H \Psi_{\alpha(t)}(\mathbf{R}) \tag{18}$$

The first equation can be eliminated by a global phase transformation

$$\Psi_{\alpha(t)}(\mathbf{R}) \rightarrow \Psi_{\alpha(t)}(\mathbf{R}) \exp\left[-i \int_0^t dt' E_{\alpha(t')} \right], \quad E_\alpha = \frac{\int d\mathbf{R} \Psi^*_\alpha(\mathbf{R}) H \Psi_\alpha(\mathbf{R})}{\int d\mathbf{R} |\Psi_\alpha(\mathbf{R})|^2} \tag{19}$$

Please note that the linearized parameter space varies and follows the time evolution of the parameters in the wave function

$$i \sum_m \int d\mathbf{R} \left[\partial_{\alpha_n} \Psi^*_{\alpha(t)}(\mathbf{R}) \right] \left[\partial_{\alpha_m} \Psi_{\alpha(t)}(\mathbf{R}) \right] \partial_t \alpha_m(t) = \int d\mathbf{R} \left[\partial_{\alpha_n} \Psi^*_{\alpha(t)}(\mathbf{R}) \right] H \Psi_{\alpha(t)}(\mathbf{R}) \tag{20}$$

or, using a bra-ket notation, $\langle \mathbf{R} | \alpha(t) \rangle = \Psi_{\alpha(t)}(\mathbf{R})$, $\langle \mathbf{R} | n(t) \rangle = \partial_{\alpha_n} \Psi_{\alpha(t)} \Psi(\mathbf{R})$ we have

$$i \sum_m \langle n(t) | m(t) \rangle \dot{\alpha}_m = \langle n(t) | H | \alpha(t) \rangle \tag{21}$$

For orthonomal basis states $\langle n(t) | m(t) \rangle = \delta_{nm}$, e.g. chosing a variational parameters which linearly couple to time-independent orthonomal basis states, we would recover the usual matrix expression for the time dependent Schrödinger equation. However, variational wave functions typically lead to non-vanishing overlaps $\langle n(t) | m(t) \rangle$ between different tangent vectors.

The time-dependent variational equations can also be derived from a variational principle of the action functional [1]

$$S = \int dt \frac{\langle \Psi(t) | i\partial_t - H | \Psi(t) \rangle}{\langle \Psi(t) | \Psi(t) \rangle} \tag{22}$$

Requesting the action to be stationary with respect to variations in $\delta \Psi(t) = \sum_n \delta \alpha_n \partial_{\alpha_n} \Psi_{\alpha(t)}$, we reobtain the previous equation of motion.

4. Variational Monte Carlo

Closed form analytical results with correlated variational wave function are in general out of scope, cluster expansions and diagrammatic summation methods have been developped to approximately calculate ground state energies and time-dependent correlation functions [2, 3]. Unbiased results can be obtained by variational Monte Carlo (VMC) calculations, numerically sampling the probability distribution

$$p(\mathbf{R}) \sim |\Psi_\alpha(\mathbf{R})|^2 \tag{23}$$

by Markov chain Monte Carlo methods. Expectation values of any operator A can be accessed by calculating the mean value of the associated expression $[\Psi(\mathbf{R})]^{-1}[A\Psi(\mathbf{R})]$. The relevant observables in the present context are the local energy

$$E_L^\alpha(\mathbf{R}) = \frac{H \Psi_\alpha(\mathbf{R})}{\Psi_\alpha(\mathbf{R})} \tag{24}$$

and

$$B_n^\alpha(\mathbf{R}) = \frac{\partial_{\alpha_n}\Psi_\alpha(\mathbf{R})}{\Psi_\alpha(\mathbf{R})} = \partial_{\alpha_n}\log\Psi_\alpha(\mathbf{R}) \tag{25}$$

which measures the change of the wave function with respect to the variational parameters.

Denoting

$$\langle\ldots\rangle_\alpha = \int d\mathbf{R}\, p(\mathbf{R})\ldots = \frac{\int d\mathbf{R}[\Psi(\mathbf{R})|^2\ldots}{\int d\mathbf{R}|\Psi(\mathbf{R})|^2} \tag{26}$$

we see that we can compute the energy via

$$E_\alpha = \langle E_L^\alpha\rangle_\alpha \tag{27}$$

as well as the correlation functions determining the dynamical evolution within our variational space

$$i\sum_m \langle [B_n^\alpha]^* B_m^\alpha\rangle_\alpha \dot{\alpha}_m = \langle [B_n^\alpha]^* E_L^\alpha\rangle_\alpha \tag{28}$$

Variational Monte Carlo calculations are very powerful, since they are (relatively) simple: We can apply the same method (and code) to quite different physical systems (different interactions, gas, liquids, solids, ...) and quantities (ground/excited states, dynamics, ...) within major changes involved... The main difficulty is to obtain reliable error estimations!

Here, I will not discuss the determinantion of the stochastic error due to the Monte Carlo sampling, which can be found in many text books, but attention is needed when the wave function contains non-analytical points or vanishes, since then estimators may diverge and error estimations typically relies on the existence of the variance of the quantity under interest.

In general, the variational ansatz introduces a systematic bias for the physical quantities of interest, e.g. the ground state energy. Estimating, reducing, and controlling this bias is a major challenge. Therefore, in the following I will discuss different forms of variational wave functions in continuous space, and possible ways to improve them and estimate their quality.

B.   Many-body trial wave functions

Above, I have given a very heuristic motivation for the use of a pair-product "Jastrow" wave function

$$\Psi \sim \exp\left[-\sum_{i<j} u(r_{ij})\right] = \prod_{i<j}\phi(r_{ij}), \quad \phi(r) = e^{-u(r)} \tag{29}$$

The Jastrow wave function is physically sound, the structure of pair correlations also arises from perturbative considerations, e.g. Bogoliubov approximation of a weakly interacting gas. Is there a way to justify this functional form from some simple considerations which may also be useful to guide us to obtain more elaborate wave functions? In the following I will focus on bosonic wave functions, afterwards I will comment on the extension for fermionic systems.

1.   Local energy method

Projection methods based on propagating any starting wave function Ψ_T in imaginary time

$$\Psi_\tau \sim e^{-\tau H}\Psi_T \tag{30}$$

will approach the true ground state exponentially fast increasing τ. Stochastic implementations of this idea are behing quantum Monte Carlo alogorithm, e.g. Diffusion and Green's function Monte Carlo as well as Variational Path Integral and Reptation Monte Carlo provide systematic stochastic approaches to sample the true ground state. In order to use this idea to improve our variational wave function, we write the projection in differential form

$$\partial_\tau\Psi(\tau) = -H\Psi(\tau) \tag{31}$$

which corresponds to Schrödiger's equation with imaginary time $t = i\tau$. We have

$$-\partial_\tau \log \Psi_\tau(\mathbf{R}) = E_L(\mathbf{R}) \tag{32}$$

in the position representation used in most of VMC calculations. Wee see that, in order to improve our starting wave function, we should include expressions in $\log \Psi$ of the same functional form as our local energy.

Let us first note that we expect the logarithm of a many-body wave function to be extensive, $U = -\log \Psi \sim N$, in the sense that

$$\langle U \rangle \sim \langle E_L \rangle \sim N \tag{33}$$

in close similarity with a classical partition function of an extended system in statistical physics. Further, since the projection works for any initial wave function, we can apply the above procedure iteratively.

Explicitly, let us start with the simplest possible bosonic wave function, Ψ_1 constant, the ideal gas ground state where the local energy is simply the potential energy, $E_L^{(1)}(\mathbf{R}) = V(\mathbf{R})$ and we obtain the Jastrow form

$$U_2 = -\log \Psi_2 = \sum_{i<j} u(r_{ij}) \tag{34}$$

but we use our variational freedom to replace $\tau v(r)$ with a general function $u(r)$ which we can parametrize and optimize.

For the next step, we will need the local energy of a Jastrow wave function

$$E_L^{(2)}(\mathbf{R}) = \frac{[\nabla U_2]^2}{2m} + \frac{\nabla^2 U_2}{2m} + V \tag{35}$$

Writing out ∇U_2 and $\nabla^2 U_2$,

$$\nabla_i \sum_{j<k} u(r_{jk}) = \sum_{j \neq i} \nabla_i u(r_{ij}) \tag{36}$$

$$\sum_i \nabla_i^2 \sum_{j<k} u(r_{jk}) = \sum_{i<j} \nabla_i^2 u(r_{ij}) \tag{37}$$

we see that the last two terms of the local energy, involving $\nabla^2 U_2$ and V, are of the same functional form as U_2. Only the term involving $[\nabla U]^2$ gives rise to a new functional form which we write as

$$U_3 = \lambda \sum_i \mathbf{w}_i \cdot \mathbf{w}_i, \quad \mathbf{w}_i = \sum_j \mathbf{r}_{ij} w(r_{ij}) \tag{38}$$

Notice that this is a 3-body correlation, but not of the most general form with evaluation cost $\sim N^3$, but one which takes the norm of a vector functions $\mathbf{w}_i$ so that the total cost remains $\sim DN^2$ having the same scaling with N as the 2-body Jastrow correlation. However, since this is a non-linear form in $\mathbf{w}_i$, we may generalize the form to $U_3 = \sum_a \lambda_a \sum_i \mathbf{w}_i^{(a)} \mathbf{w}_i^{(a)}$ allowing for several different functions $w^{(a)}(r)$ to be optimized.

Working out the local energy, the terms involving the gradient of U_3 will create 4-body terms [10], but their form is strongly constraint to only two possible forms

$$U_4 = \sum_i \mathbf{v}_i \cdot \underline{T}_i^{(2)} \cdot \mathbf{v}_i \tag{39}$$

containing a tensor $\underline{T}_i^{(2)} = \sum_j \mathbf{r}_{ij} \otimes \mathbf{r}_{ij} t(r_{ij})$ bracked by two vectors $\mathbf{v}_i$. The second possibility is

$$U_4 = \sum_{i,j} \mathbf{v}_i \cdot \underline{T}_{ij}^{(2)} \cdot \mathbf{v}_j \tag{40}$$

with a tensor $\underline{T}_{ij}^{(2)} = \mathbf{r}_{ij} \otimes \mathbf{r}_{ij} t(r_{ij})$. Here, all tensors, and vectors generally involve each time new functions $t(r)$ and $v(r)$.

Instead of going on, let us try to construct the general structure we may obtain. New functional terms in the local energy can only arise through the gradient which always act on a functional forms stemming from a pair correlation. Therefore, we can only create vectors $\mathbf{v}_i$ and arbitrary higher order tensors of the form $\underline{T}_i^{(n)} = \sum_j \mathbf{r}_{ij} \otimes \mathbf{r}_{ij} \cdots \otimes \mathbf{r}_{ij} t(r_{ij})$

and $\underline{T}_{ij}^{(n)} = \mathbf{r}_{ij} \otimes \mathbf{r}_{ij} \cdots \otimes \mathbf{r}_{ij} t(r_{ij})$ Since $\log \Psi$ must be a scalar, we can simply include all possible scalars we can construct from terms up to a certain order.

Although going beyond 4th order is possible, the number of possible combinations rapidely grows, and, since all these expressions are highly non-linear, we may in principle have to include linear superpositions as indicated already for U_3 before. A more efficient way to construct these higher order terms indirectly, is by the so-called "iterative backflow renormalization".

There, we use that a vector potential $\mathbf{v}_i = \sum_{j \neq i} \mathbf{r}_{ij} v(r_{ij})$ transforms exactly like the coordinate $\mathbf{r}_i$ of a particle, but contains information on the actual state $\mathbf{R}$ via the two-body correlation term. It is then straighforward to construct vector potential including higher order correlations by using a superposition of $\mathbf{r}_i$ and $\mathbf{v}_i$ in the construction (called "backflow coordinates" for historical reasons), e.g.

$$\mathbf{v}_i^{(2)} = \sum_{j \neq i} \mathbf{q}_{ij}^{(1)} v_2(q_{ij}^{(1)}), \quad \mathbf{q}_i^{(1)} = \mathbf{r}_i + \mathbf{v}_i \tag{41}$$

which is easily iterated to higher order. All derivatives needed for calculating the local energy can be worked out by use of the chain rule. Importantly, they can be calculated efficiently, and the computational cost does not increase strongly going to higher order iterations (the scaling with the number of particles remains the same independent of the order). The correspondning formulas are given in the appendix of Ref. [11]. We can construct iterated vector and tensor potentials and construct scalars in the same way as described above.

2. Path-integral method

Backflow coordinates and the iterative renormalization occur rather naturally by considering the projection in the path integral formulation – in contrast to the derivative version which has led us to the local energy method above.

In the variational path integral we improve our initial trial wave function, by imaginary time projection applying the propagator $G(\mathbf{R}, \mathbf{R}'; \tau) = \langle \mathbf{R} | \exp[-\tau H] | \mathbf{R}' \rangle$

$$\Psi_\tau(\mathbf{R}) = \int d\mathbf{R}' G(\mathbf{R}, \mathbf{R}'; \tau) \Psi_T(\mathbf{R}') = \sum_n \langle \mathbf{R} | E_n \rangle e^{-\tau E_n} \langle E_n | \Psi_T \rangle \tag{42}$$

which suppresses the overlap of any trial wave function with excited energy eigenstates of energy E_n by $\exp[-\tau[E_n - E_0]$, so that Ψ_τ is guaranteed to provide lower energy bounds than Ψ_T.

For short (imaginary) time, the form of the propagator is explicitly known

$$G(\mathbf{R}, \mathbf{R}'; \tau \to 0) \sim e^{-m(\mathbf{R} - \mathbf{R}')^2/2\tau} e^{-\tau V(\mathbf{R})} \tag{43}$$

Let us know apply it to a Jastrow trial function, $\Psi_T = \exp[-U]$,

$$\Psi_\tau(\mathbf{R}) \sim \int d\mathbf{R}' e^{-m(\mathbf{R} - \mathbf{R}')^2/2\tau - \tau V(\mathbf{R}) - U(\mathbf{R}')} \tag{44}$$

(The stochastic integral of this one-step application of the propagator is also called shadow wave function). Since we expect a convergent integral due to gaussian part of the propagator, we may suppose a sharp peak around some point $\mathbf{Q} \approx \mathbf{R}$, the precise point will be determined later, and expand the Jastrow correlation

$$U(\mathbf{R}') \approx U(\mathbf{Q}) + (\mathbf{R}' - \mathbf{Q}) \nabla U(\mathbf{Q}) \tag{45}$$

which gives

$$\Psi_\tau(\mathbf{R}) \approx \int d\mathbf{R}' e^{-\lambda(\mathbf{R}' - \mathbf{R} + \nabla U/2\lambda)^2} e^{-\tau V(\mathbf{R}) - U(\mathbf{Q}) + [\nabla U]^2/4\lambda} \tag{46}$$

Since we have assumed the integral over $\mathbf{R}'$ to be sharply peaked for around $\mathbf{Q}$, this point should coincide with the center of the gaussian distribution. We thus obtain

$$\mathbf{Q} = \mathbf{R} - \nabla U(\mathbf{Q})/2\lambda^2 \tag{47}$$

Writing out $\mathbf{Q}$ explicitly, we get the "backflow" form

$$\mathbf{q}_i = \mathbf{r}_i - \sum_{j \neq i} \mathbf{r}_{ij} u(r_{ij}) \tag{48}$$

where I have used $\mathbf{Q} \approx \mathbf{R}$ for simplification.

The approximate integration of the shadow wave function has led us to an explicit form

$$\Psi_\tau(\mathbf{R}) \approx e^{-\tau V(\mathbf{R}) - U(\mathbf{Q}) + [\nabla U]^2/4\lambda} \tag{49}$$

which we can identify with the first iteration of backflow improved potentials. Subsequent applications of the propagator then leads to iterated backflow form [12]. The backflow iterated backflow wave functions can be regarded as a deep neural network, though not of the canonic form of a restricted Boltzmann machine used on lattice Hamiltonian [13], but much closer to recent continuous space versions applied to describe electrons in light atoms and small molecules [14, 15].

3. Fermions and inhomogeneous systems

Above we have basically shown how to systematically improve trial wave functions for homogeneous Bose systems, but the heuristic derivation can be extended also to Fermions and inhomogeneous systems.

In the presence of an external potential, it is natural to include a single body Jastrow function, e.g. $\exp[-\sum_i u_1(\mathbf{r})] = \prod_i \phi(\mathbf{r}_i)$, in our lowest order trial wave functions. Its gradient naturally couples to the gradient term of a two particle Jastrow correlation, e.g. $\sum_i [\nabla_i \phi(\mathbf{r}_i)] \nabla_i \sum_j u_2(r_{ij})$.

Fermionic functions can be build out of antisymmetrizing a general asymmtric wave function where each fermion occupies a different orbital, $\sum_P (-1)^{|P|} \prod_i \phi_i(\mathbf{r}_{P(i)})$, leading to a Slater determinant, or a Slater-Jastrow wave function when multiplied by a symmetric Jastrow correlation,

$$\Psi_{SJ} \sim D(\mathbf{R}) e^{-U(\mathbf{R})}, \quad D(\mathbf{R}) = \det_{ni} \phi_n(\mathbf{r}_i) \tag{50}$$

where $U(P\mathbf{R}) = U(\mathbf{R})$ for any permutation by construction, e.g. $U = \sum_{i<j} u(r_{ij})$, and $D(\mathbf{R})$ assures the antisymmetry. Then the coupling between single particle and two particle correlations give rise $\sum_i [\nabla \log D] \nabla_i U(\mathbf{R})$ which, to leading order can be written either as a change of coordinates to backflow-coordinates

$$\Psi_{bf} \sim \det_{ni} \phi_n(\mathbf{q}_i) e^{-U(\mathbf{R})}, \quad \mathbf{q}_i = \mathbf{r}_i + \sum_j \mathbf{r}_{ij} \eta(r_{ij}) \tag{51}$$

in the case of (quasi-) unform systems, or, in strongly inhomogeneous situations, as an orbital backflow

$$\Psi_{obf} \sim \det_{ni} \widetilde{\phi}_n(\mathbf{r}_i; \mathbf{R}) e^{-U(\mathbf{R})}, \quad \widetilde{\phi}_n(\mathbf{r}_i; \mathbf{R}) = \phi_n(\mathbf{r}_i) + [\nabla_i \phi_n(\mathbf{r}_i)] \sum_j \mathbf{r}_{ij} \eta(r_{ij}) \tag{52}$$

Using similar arguments as before, we then obtain the iterative backflow wave functions for fermions.

However, I want to note that the above arguments essentially rely on the smoothness of the potentials being expanded. In the case of Fermions, typical values of the logarithm of the determinant are extensive in the number of particles, but we know that the value of the determinant itself must vanish at the Fermion nodes due to antisymmetry. Therefore, the landscape of variations of the determinant cannot considered as smooth, and derivation cannot be taken too serious (Similar for strongly inhomogeneous orbitals which may vanish in some regions of space). Still, the resulting forms of the wave functions have been shown to be very efficient for lowering the wave functions, so the wave functions do lead to noticeable improvements also in the case of fermions and inhomogeneous systems.

C. Optimization

Without having detailed a-priori information, the trial energy provides a unique critierum to decide on the best wave function: the lower, the better. We thus aim to minimize the trial energy

$$E_\alpha = \langle E_L^\alpha \rangle_\alpha \tag{53}$$

with respect to the variational parameters α. Since we cannot calculate E_α exactly, we have to minimize an estimator affected by a stochastic random error.

Several strategies have been used:

- Simply ignore the stochastic error and use your favorite optimizer for non-linear optimization problems to minimize the estimator of the trial energy.

- use stochastic reconfiguration, solve the imaginary time Schrödinger equation in the variational space (see real time below but with $t = i\tau$.

Much can and should be added, but I will not enter the subject of opimization here.

Note that denoting $r(\mathbf{R}) = (H - E_\alpha)\Psi_\alpha(\mathbf{R})$, $|r(\mathbf{R})|^2$ is a measure of how far the projected Schrödinger equation solves the full Schrödinger equation. Thus, we can also minimize $\sigma^2 = \int d\mathbf{R}|r(\mathbf{R})|^2/\int d\mathbf{R}|\Psi_\alpha(\mathbf{R})|^2 = \langle H^2 \rangle_\alpha - E_\alpha^2$, the variance of the energy with respect to the trial wave function parameters. For vanishing variance of the energy, we have an exact eigenstate. Energy vs variance extrapolations can be used to estimate the true ground state energy, particularly important for Fermions to quantify the error of the fixed-node approximation.

D. Time-dependent Variational Monte Carlo method

Let us now turn to search to the best variational approximation for the time evolution of a quantum state $\Psi_{\alpha(t)}(\mathbf{R}) = \langle \mathbf{R}|\alpha(t)\rangle$, Using as before the notation $\partial_{\alpha_n}\Psi_{\alpha(t)}\Psi(\mathbf{R}) = \langle \mathbf{R}|n(t)\rangle$ we have a first order differential equation for the evolution of our variational parameter $\alpha(t)$ in time

$$i\sum_m \langle n(t)|m(t)\rangle\dot{\alpha}_m(t) = \langle n(t)|H|\alpha(t)\rangle \tag{54}$$

or

$$i\sum_m S_{nm}\dot{\alpha}_m(t) = F_n(t) \tag{55}$$

Given a wave function expressed by parameters $\alpha_n(t)$ at some time t, we can sample $|\Psi_{\alpha(t)}(\mathbf{R})|^2$ by variational Monte Carlo and calculate estimators for S_{nm} and F_n

$$S_{nm} = \left\langle [\partial_{\alpha_n}\log\Psi_\alpha]^* [\partial_{\alpha_m}\log\Psi_\alpha]\right\rangle_{\alpha(t)} \tag{56}$$

$$F_n = \left\langle [\partial_{\alpha_n}\log\Psi_\alpha]^* E_L^\alpha \right\rangle_{\alpha(t)} \tag{57}$$

and use these estimators in the equation of motion which can then be integrated over a small time step. After each time step, we update the variational parameters and perform a new Monte Carlo sampling with this time-propagated wave function. Most of time-dependent variational Monte Carlo calculations have been performed on lattice models [7, 13], but they can equally be extended to continuous space [8].

It is immediately clear, that time-dependent calculations are computationally much more demanding than their static, ground state analog, simply since a) we have to perform VMC calculations at (almost) all time steps, and b) we have to parametrize real and imaginary part of the logarithm of the wave function, introducing at least twice as much variational parameters, but likely much more (typically $10^2 - 10^5$ are currently used). Still the basic implementation of time-dependent variational Monte Carlo is rather straightforward. In particular, setting $t = i\tau$, we obtain an imaginary time evolution which should drive us to the best ground state within the variational reach; in the context of the ground state optimization methods, this imaginary time projection is called *stochastic reconfiguration* [19].

However, in a basic implementation, time-dependent Variaional Monte Carlo poses also new problems and questions. One of the most important ones are:

- long time stability: what is the influence of stochastic noise affecting the correlation functions and the time discretization in solving the ODE?

- large number of parameters: singular matrix, impossibility of storing the matrix elements

- control of the overall systematic bias

These questions can be adressed to some extend theoretically, but, ultimatively, must also be checked by numerical experiments. Let us now go into some of the details.

1. Integrating the ODE

Let us first assume that all matrix elements S_{nm} as well as the "forces" F_n can be exactly calculated. We then have to solve an ODE of the type $\dot{y}(t) = F(y(t))$, a classical problem. There are many way to discretize this ODE for

short times and solve the associated finite difference equations. From classical molecular dynamics simulations, it is well known that different discretization schemes may strongly affect long time stability, in particular they may lead to drifts in conserved quantities, e.g. the mean energy.

I will not enter into details, but just mention that the naive first order discretization, $y(t + dt) = y(t) + dt F(y(t))$, although accurate at small times, does not perform well at large times, and higher order methods should be prefered. In [8] adaptive 4th order Runge-Kutta was used, the stability of different schemes in time-dependent VMC have been discussed in the recent Refs [20, 21].

One further complication arises that for solving any discretized time evolution, we typically have to calculate the inverse of the matrix S_{nm}, which may get problematic when the matrix gets singular, e.g. one of the eigenvalues of S_{nm} approaches zero. Then any error will potentially propagate into the time evolution, and, in particular, time-discretization errors will blow up if the ratio of largest to smallest eigenvalues becomes too large. In lattice models, the eigenvalues of the matrix are bounded, so that might not pose a problem in principle (or can be circumvented), but this is more delecate in the continuum. I will come back later to this point, which for us is intrinsically connected with the stochastic nature of VMC calculations.

2. Stochastic determination of matrix, and inverse

In our case, the matrix S_{nm} and the forces F_n are not known exactly, however, we can calculate estimators within VMC

$$S_{nm}^{MC} = \frac{1}{N_{MC}} \sum_{\mathbf{R} \sim |\Psi_\alpha(\mathbf{R})|^2} [B_n^\alpha(\mathbf{R})]^* B_m^\alpha(\mathbf{R}) \tag{58}$$

$$F_n^{MC} = \frac{1}{N_{MC}} \sum [B_n^\alpha(\mathbf{R})]^* E_L^\alpha(\mathbf{R}) \tag{59}$$

where

$$B_n^\alpha(\mathbf{R}) = \partial_{\alpha_n} \log \Psi_\alpha(\mathbf{R}), \quad E_L^\alpha(\mathbf{R}) = \frac{H \Psi_\alpha(\mathbf{R})}{\Psi_\alpha(\mathbf{R})} \tag{60}$$

are local observables. Although S_{nm}^{MC} and F_n^{MC} approaches the exact values S_{nm} and F_n for infinite Monte Carlo samples ($N_{MC} \to \infty$), any finite Monte Carlo run introduces a stochastic error.

Actually, the local observables are not unique in the sense that we may replace them by different expressions which leads to the same correlation functions in average (or trivially related ones), but with a reduced variance such that the stochastic bias is reduced. A simple way to reduce slightly the bias is by actually considering only the fluctuating part of the local operators, e.g.

$$E_L^\alpha(\mathbf{R}) \to E_L^\alpha(\mathbf{R}) - E_\alpha^{MC} \tag{61}$$

where $E_\alpha^{MC} = N_{MC}^{-1} \sum E_L^\alpha(\mathbf{R})$ is the estimator of the trial energy for the given Monte Carlo sample, and similar for $B_n^\alpha(\mathbf{R})$. In the equation of motions, this substraction can be absorbed by global changes, e.g. in the normalization or global phase.

Of course, we will have to use sufficiently large Monte Carlo runs, N_{MC}, such that the stochastic noise is kept small enough. For given discrete time step, the stochatic noise must be small enough be represent a perturbation to the exact values. We will discuss some aspects how the noise enters into the dynamics later. Typically, one has to perform several calculations with different noise levels to judge experimentally the influence of the stochastistic bias.

Some care also have to be taken when the wave function can vanish, e.g. for fermions. Then the local operators in general diverge at the nodes. A sampling based on $|\Psi_\alpha(\mathbf{R})|^2$ leads to a vanishing weight at the nodes. Although mean value of one local observable is not affected, since the nodal region does not contribute to them, the nodal region may actually contribute to the average of correlation functions. This bias can in principle be adressed by regularizing the wave function at the nodes to get a finite probably to reach them in the random walk. In bosonic problems, as long as the time evolution results mainly in phase changes and does not introduce regions of vanishing amplitudes, this issue does not arise.

3. Handling of large and singular matrices

When adressing the time evolution of a many-body quantum systems, we are in general interested how excitation of a single or few modes relax in time being absorbed by the large degrees of freedom at disposition. For an accurate

description of such situations, we can expect to need a large variational freedom, such that our linearized (tangential) space includes the degrees of freedom involved in the relaxation. Thus, we expect far larger number of parameters than in typical ground state calculations and we immeditaly have to face the associated problems.

An immediate problem with large number of parameters is that we may not have enough memory at disposition to store the matrix S_{nm}. In this case, we have to use iterative solvers to obtain the inverse, e.g. we can start with some approximate vector $\dot{\alpha}_n^0$, and, instead of calculating the full matrix S_{nm} by VMC, we only calculate the vector $\sum_m S_{nm}\dot{\alpha}_n^0$ within a VMC run, or the corresponding vectors, e.g. needed in a conjugate gradient approach for the inverse. Each iteration of the iterative solver, needs a new VMC calculation. Reweighting methods can be used, but does not really need to a large speed up, since the sampling is usually fast compared to the evaluation of quantities like the local energy for a given configuration. Since changes in $\dot{\alpha}_n$ are small for small time steps, iterative solvers typically converge within a few iteration steps.

However, as already mentionned above, the matrix S_{nm} may become singular, and the inverse of the matrix is thus ill-defined. The intrinsic stochastic noise due to the Monte Carlo sampling will randomize the matrix, and the low eigenvalues and eigenvectors of the matrix will be dominated by noise. Pragmatically, we have to regularize the matrix, e.g. we can perform a singular value decomposition, or add a small positive diagonal matrix, such that the inverse is well defined.

In the context of a stochastic approach, the regularization does not pose a fundamental problem. After all, we have no access to the exact values, and, we can only hope to satisfy our differential equation

$$i \sum_m S_{nm}\dot{\alpha}_m - F_n = 0 \tag{62}$$

on *average*. Small eigenvalues of S_{nm} do not contribute significantly to the equation in general (even if all quantities are exactly known), but more importantly, they are likely to contribute less than the stochastic noise already present. In that case, it does not improve our calculation to solve an estimator of equation more accurately than the estimator itself. Regularization, therefore, does not introduce any further bias, of course, what we really should have a look at is, how stochastic noise (independent of its origin) affects our solution.

4. Stochastic noise: long-time bias

Let us now discuss more concretely possible bias introduced into the temporal evolution of our wave function. Therefore, we denote $\Psi_{\alpha(t)}(\mathbf{R})$ the actual outcome of a single realisation of a time-dependent Variational Monte Carlo run, starting with some initial conditions at $t = 0$.

Let us verify up to which extend $\Psi_{\alpha(t)}(\mathbf{R})$ can be considered as an accurate solution. Therefore, we simply plug this potential solution into the Schödinger equation

$$(i\partial_t - H)\Psi_{\alpha(t)}(\mathbf{R}) = r(\mathbf{R}, t) \tag{63}$$

If the reminder $r(\mathbf{R}, t)$ vanishes, we will have an exact solution. However, we don't expect this to happen in our problems, and we our stochastic and systematic (variational) bias will be at the origin.

Let us denote $\Phi(\mathbf{R}, t)$ the exact solution of the Schrödinger equation. We can then write down the formal solution of our inhomogeneous equation

$$\Psi_{\alpha(t)}(\mathbf{R}) = \Phi(\mathbf{R}, t) + \int d\mathbf{R}' \int dt' G(\mathbf{R} - \mathbf{R}', t - t') r(\mathbf{R}', t') \tag{64}$$

in terms of our exact Green's function

$$(i\partial_t - H)G(\mathbf{R}, t) = \delta(\mathbf{R})\delta(t) \tag{65}$$

To go further, we decompose our error $r(\mathbf{R}, t) = b(\mathbf{r}, t) + \xi(\mathbf{R}, t)$ into the systematic bias $b(\mathbf{R}, t)$ and the stochastic error $\xi(\mathbf{R}, t)$ with $\overline{\xi(\mathbf{R}, t)} = 0$, e.g. averaging over different realizations of our time-dependent Monte Carlo, the stochastic part must vanish. Since our random numbers are independent at different times, we further expect

$$\overline{\xi(\mathbf{R}, t)\xi(\mathbf{R}', t')} = \sigma^2(\mathbf{R} - \mathbf{R}')\delta(t - t') \tag{66}$$

Concerning the bias, $b(\mathbf{R}, t)$, we can introduce the projection operator $P(t)$ which project states into the (time-dependent) tangential space spanned by the best variational wave function $\Psi_{\alpha(t)}(\mathbf{R})$ and derivatives $\partial_{\alpha_n}\Psi_{alpha}(\mathbf{R})$ at time t. Together with the complement $Q(t) = 1 - P(t)$, we have $b(\mathbf{R}, t) = [Q(t) + P(t)]b(\mathbf{R}, t) = Q(t)b(\mathbf{R}, t)$, since

the variational equation projected the reminder out of the tangential space. This condition is satisfied and unbiased on average, so it holds for $b(\mathbf{R}, t)$.

Specifying on the expectation values of a local observable $A(\mathbf{R})$ at time t, we then get different terms for the overall error, averaged over different realizations,

$$\int d\mathbf{R} |\Psi_{\alpha(t)}(\mathbf{R}, t)|^2 A(\mathbf{R}) - \int d\mathbf{R} |\Phi(\mathbf{R}, t)|^2 A(\mathbf{R}) = \delta A_1 + \delta A_2^b + \delta A_2^s \tag{67}$$

where

$$\delta A_1 = \int dt' \int d\mathbf{R} d\mathbf{R}' \Phi^*(\mathbf{R}, t) A(\mathbf{R}) G(\mathbf{R} - \mathbf{R}', t - t') b(\mathbf{R}', t') + c.c. \tag{68}$$

is the potentially dominating systematic bias, basically reflecting all limitations of our variational trial wave function. Since we have $b(\mathbf{R}, t) = Q(\mathbf{R}, t) b(\mathbf{R}, t)$, this terms may be suppressed if modes out of our variational space to have significant overlap with the back-propagated final state $\Phi^*(\mathbf{R}) A(\mathbf{R})$. (Being interested only in a specific operator, we may adapt our variational approach to reduce further this overlap.)

$$\delta A_2^b = \int dt dt' \int d\mathbf{R} d\mathbf{R}' d\mathbf{R}'' A(\mathbf{R}) G(\mathbf{R} - \mathbf{R}', t - t') b(\mathbf{R}', t') G^*(\mathbf{R} - \mathbf{R}'', t - t'') b^*(\mathbf{R}'', t'') \tag{69}$$

is the corresponding second order term.

The second order term due to our stochastic errors writes

$$\begin{aligned} \delta A_2^s &= \int dt dt' \int d\mathbf{R} d\mathbf{R}' d\mathbf{R}'' A(\mathbf{R}) G(\mathbf{R} - \mathbf{R}', t - t') G^*(\mathbf{R} - \mathbf{R}'', t - t'') \overline{\xi(\mathbf{R}', t') \xi^*(\mathbf{R}'', t'')} \\ &\sim \int dt' \int d\mathbf{R} d\mathbf{R}' A(\mathbf{R}) |G(\mathbf{R} - \mathbf{R}', t - t')|^2 \sigma^2(\mathbf{R} - \mathbf{R}') \sim T \int d\mathbf{R} A(\mathbf{R}) \int d\mathbf{R}' \sigma^2(\mathbf{R}') \end{aligned} \tag{70}$$

This term would actually vanish if we calculate it with Ψ and Ψ^* from independent stochastic realizations, respectively. But, typically, it is calculated within the same time dependent VMC run, and then scales with the total time $\sim T$.

This estimations indicate that the stochastic noise can be controlled and simulations reaching typical (finite!) relaxation times should be feasable. Explicit computations experimentally confirm that the stochastic noise does not present the major source of uncertainty and the method succeeds to obtain numerically stable simulations over large times. However, similar to ground state variational calculations, the bias due to the variational wave function remains the main source of uncertainty.

---

[1] J.-P. Blaizot and G. Ripka, *Quantum Theory of Finite Systems*, MIT Press, Cambridge, Ma (1986).
[2] E. Feenberg, *Theory of Quantum Fluids*, Academic, New York (1969).
[3] A. Fabrocini, S. Fantoni, and E. Krotscheck (Eds), *Introduction to modern methods of quantum many-body theory and their applications*, World Scientific (2002), in particular chapters by M. Saarela and E. Krotscheck.
[4] W.L. McMillan, Phys. Rev. **138**, A442 (1965).
[5] D. M. Ceperley, Rev. Mod. Phys. **67**, 1601 (1995).
[6] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[7] G. Carleo, F. Becca, M. Schirò, and M. Fabrizio, Sci. Rep. **2**, 243 (2012).
[8] G. Carleo, L. Cevolani, L. Sanchez-Palencia, and M. Holzmann, Phys. Rev. X **7**, 031026 (2017).
[9] M. Holzmann, D. M. Ceperley, C. Pierleoni and K Esler, Phys. Rev. E **68**, 046707 (2003).
[10] M. Holzmann, B. Bernu, and D.M. Ceperley, Phys. Rev. B **74**, 104510 (2006).
[11] M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, Phys. Rev. B **91**, 115106 (2015).
[12] M. Ruggeri, S. Moroni, and M. Holzmann, Phys. Rev. Lett. **120**, 205302 (2018).
[13] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[14] D. Pfau, J. S. Spencer, A. G. de G. Matthews, and W. M. C. Foulkes, arXiv:1909.02487.
[15] J. Hermann, Z. Schätzle, and Frank Noé, arXiv:1909.08423.
[16] M. Holzmann and S. Moroni, Phys. Rev. B **99**, 085121 (2019).
[17] T. Comparin, R. Bombin, M. Holzmann, F. Mazzanti, J. Boronat, and S. Giorgini, Phys. Rev. A **99**, 043609 (2019).
[18] M. Holzmann and S. Moroni, Phys. Rev. Lett. **124**, 206404 (2020).
[19] F. Becca and S. Sorella, *Quantum Monte Carlo approaches for correlated systems*, Cambridge (2017).
[20] M. Schmitt and M. Heyl, arXiv:1912.08828 (2019).
[21] I. López-Gutiérrez and C.B. Mendl, arXiv:1912.08831 (2019).