Polynomial complexity despite the fermionic sign

R. Rossi1, N. Prokof’ev2,3, B. Svistunov2,3,4,5,1, K. Van Houcke1 and F. Werner5

1Laboratoire de Physique Statistique, Ecole Normale Supérieure, UPMC, Université Paris Diderot, CNRS
24 rue Lhomond, 75005 Paris, France
2Department of Physics, University of Massachusetts - Amherst, MA 01003, USA
3National Research Center “Kurchatov Institute” - 123182 Moscow, Russia
4Wilczek Quantum Center, Zhejiang University of Technology - Hangzhou 310014, China
5Laboratoire Kastler Brossel, Ecole Normale Supérieure, UPMC, CNRS, Collège de France
24 rue Lhomond, 75005 Paris, France

received 29 March 2017; accepted in final form 10 May 2017
published online 30 May 2017

PACS 02.70.Ss – Quantum Monte Carlo methods
PACS 71.10.Fd – Lattice fermion models (Hubbard model, etc.)

Abstract – It is commonly believed that in unbiased quantum Monte Carlo approaches to fermionic many-body problems, the infamous sign problem generically implies prohibitively large computational times for obtaining thermodynamic-limit quantities. We point out that for convergent Feynman diagrammatic series evaluated with a recently introduced Monte Carlo algorithm (see Rossi R., arXiv:1612.05184), the computational time increases only polynomially with the inverse error on thermodynamic-limit quantities.

The notion of fermion sign problem (FSP) was originally formulated in the context of auxiliary-field, path-integral and diffusion quantum Monte Carlo (QMC) methods [1–4]. There, it was observed that the computational time required for calculating properties of the fermionic system to a given accuracy scales exponentially with the system volume. Later, the notion of FSP was implicitly extended to an arbitrary QMC approach dealing with interacting fermions, referring to the stochastic sampling of a non-sign-definite quantity with a near cancellation between positive and negative contributions. Sign-free Monte Carlo (MC) algorithms were emerging only as exceptions confirming the rule: In each such case, the absence of FSP was due to some special property of the simulated model (see, e.g., [5–11] and references therein). Nowadays the FSP is generally perceived as one of the most important unsolved problems in the field of numerical studies of interacting fermionic systems in dimensions $d > 1$ (see footnote 1).

The main message of this letter follows from a simple observation. Suppose some quantity $Q$ is computed from a limit $Q = \lim_{n \to \infty} Q_n$, with an exponentially fast convergence, $|Q - Q_n| \sim e^{-\#n}$ (where $\#$ denotes some positive constant). Then in order to compute $Q$ up to an error $|Q - Q_n| = \epsilon$ it is sufficient to take $n \sim \ln \epsilon^{-1}$. Hence, even if the computational time increases exponentially as a function of $n, t \sim e^{\#n}$, the increase of $t$ as a function of $\epsilon^{-1}$ is only polynomial, $\ln t \sim \ln \epsilon^{-1}$.

This observation applies to the simulation of interacting fermions by the algorithm introduced in ref. [12], denoted hereafter by the acronym CDet, for connected determinant diagrammatic MC. This algorithm works directly in the thermodynamic limit since it evaluates the series of connected Feynman diagrams. It exploits two advantages of the fermionic sign: First, for fermions on a lattice at finite temperature, the series has a finite radius of convergence, so that the convergence as a function of diagram order $n$ is exponential; second, a factorial number of connected Feynman diagrams can be evaluated in exponential time using determinants (and a recursive formula).

In general, sign-alteration of observables simulated by MC methods is neither sufficient nor necessary to state that the problem is intractable. One should rather focus on what we will call the “computational complexity problem” (CCP) instead of the FSP. The key question is the one that is most relevant practically2: How easily

---

1Frustrated spin and frustrated bosonic (with restricted on-site Hilbert space) lattice models can be mapped to a system of interacting fermions and thus are part of the present discussion.

2We closely follow ideas expressed by D. Ceperley at the Meeting of the Simons Collaboration on the Many-Electron Problem, New York, November 19–20, 2015.
can one indefinitely increase the accuracy of the computed thermodynamic-limit answer? This leads to the following definition of the CCP that can be applied to any numerical scheme. Let $Q$ be the intensive quantity of interest in the thermodynamic limit. A numerical scheme is said to have CCP if the computational time $t$ required to obtain $Q$ with an error $\epsilon$ diverges faster than any polynomial function of $\epsilon^{-1} \rightarrow \infty$. The CCP is considered to be solved if

$$ t(\epsilon) = O(\epsilon^{-\alpha}). $$

(1)

Note that we consider unbiased methods, i.e., $\epsilon \rightarrow 0$ is the difference between computed value and exact value.

In what follows, we show in some detail that CDet solves the CCP, at least at finite temperature and small enough interaction. In parallel, we also discuss the conventional diagrammatic Monte Carlo approach (hereafter denoted by DiagMC) in which the sum over diagram topologies is done stochastically [12–14]. Assuming that we are inside the radius of convergence, the convergence is exponential, with the sum over all possible connected diagram topologies connecting the internal and external vertices.

Let us denote the contribution of a diagram of topology $T$ for fixed internal variables $X$ by $D(T, X)$. Let $a_n$ be the sum of all Feynman diagrams of order $n$:

$$ a_n = \int dX \sum_{T \in S_n} D(T, X), $$

(5)

with $S_n$ the set of all diagram topologies at order $n$. This can be rewritten in the form of eq. (2):

$$ a_n = \langle A_n \rangle_{P_n}, $$

(6)

with the unnormalised distribution to be sampled chosen to be

$$ P_n(T, X) = |D(T, X)| \quad \text{(DiagMC)}, $$

$$ P_n(X) = \sum_{T \in S_n} |D(T, X)| \quad \text{(CDet)} $$

(7)

and

$$ A_n = \left\{ \begin{array}{l}
|z_n \cdot \text{sign}[D(T, X)]| \quad \text{(DiagMC)}, \\
|z_n \cdot \text{sign} \left[ \sum_{T \in S_n} D(T, X) \right] | \quad \text{(CDet)}
\end{array} \right. $$

(8)

with the normalisation factors

$$ z_n = \left\{ \begin{array}{l}
\int dX \sum_{T \in S_n} |D(T, X)| \quad \text{(DiagMC)}, \\
\int dX \sum_{T \in S_n} D(T, X) \quad \text{(CDet)}
\end{array} \right. $$

(9)

So, in DiagMC the diagrams are sampled according to the distribution $|D(T, X)|$, while in CDet diagrams are grouped together via determinants and in the MC part of the algorithm one samples $X$ according to the distribution $\left| \sum_{T \in S_n} D(T, X) \right|$. In what follows we neglect the statistical error on the normalisation factors $z_n$ since they are obtained by sampling a sign-positive quantity.

Here we consider fermions on a lattice at finite temperature, so that the radius of convergence of the diagrammatic series is finite [12–14]. Assuming that we are inside the radius of convergence, the convergence is exponential,

$$ |a_n| = O(R^{-n}) $$

(10)

with $R > 1$ a constant. Here and in what follows we omit multiplicative constants and power laws which do not affect the dominant scaling behavior.
The number of diagrams scales factorially with the order \( n \). For CDet, however, one takes into account cancellations between different diagram topologies. More specifically, we expect that for fermions on a lattice at finite temperature,
\[
\begin{align*}
z_n &\sim n! R_D^n \quad \text{(DiagMC)}, \\
z_n &\sim R_C^n \quad \text{(CDet)}
\end{align*}
\]
with \( R_D \) and \( R_C \) positive constants.

Let us discuss the behavior of the average sign, \( \langle \text{sign} \rangle := \langle \text{sign } A_n \rangle p_n = a_n/z_n \), as a function of the order \( n \). For DiagMC, we see that \( \langle \text{sign} \rangle \) tends to zero factorially, a manifestation of the near-compensation between different diagrams. For CDet, \( \langle \text{sign} \rangle \) tends to zero exponentially in the generic case where \( R_C < R \).

As a result, the variance on individual measurements behaves as
\[
\sigma_{A_n} = z_n \sqrt{1 - \langle \text{sign} \rangle^2 p_n} \sim z_n, \tag{13}
\]
which, together with eq. (4), gives for the statistical error bar on the \( n \)-th order contribution \( a_n \)
\[
\epsilon_{\text{stat}}(n) \sim z_n \sqrt{\frac{2 \tau_{\text{auto}}(n) + 1}{N_{MC}(n)}}. \tag{14}
\]
Here \( \tau_{\text{auto}}(n) \) is expected to increase at most polynomially with \( n \), which we have checked numerically for CDet; we will neglect this \( n \)-dependence of \( \tau_{\text{auto}} \) since it will not affect the final scalings. An appropriate dependence \( N_{MC}(n) \) of the number of MC steps on order will be specified below.

Note that the relative statistical error given by the absolute value of \( \epsilon_{\text{stat}}(n)/a_n \propto z_n/a_n = 1/\langle \text{sign} \rangle p_n \) diverges factorially for DiagMC and exponentially for CDet. This can be viewed as a sign problem for diagrammatic Monte Carlo methods, limiting the order that can be reached. On the other hand, the exponential convergence (10), which is only possible thanks to the fermionic sign, implies that reaching very high orders is not necessary, as we now quantify.

The assumption of exponential convergence implies that the systematic error due to the finite diagram-order cut-off \( N \),
\[
\epsilon_{\text{sys}}(N) = \sum_{n=N+1}^{\infty} a_n, \tag{15}
\]
decreases exponentially,
\[
\epsilon_{\text{sys}}(N) \sim \frac{O(R^{-N})}{N}. \tag{16}
\]
To achieve a final error \( \sim \epsilon \), it is then natural to work in a regime where systematic and statistical errors are on the

same order. We thus choose \( N \) so that \( R^{-N} \sim \epsilon \), and we take a computational time \( t \) such that the total statistical error is \( \epsilon_{\text{stat}} \sim \epsilon \). Neglecting correlations between different orders, we have \( \epsilon_{\text{stat}} \approx \sum_{n=0}^{N} c_{\text{stat}}(n) \), which leads us to choose \( N_{MC}(n) \) such that \( \epsilon_{\text{stat}}(n) \) is \( n \)-independent. Equation (14) together with eqs. (11), (12) then yield
\[
\begin{align*}
t_n &\sim \left\{ \begin{array}{ll} 
\frac{1}{c^2 R_D^n} & \text{(DiagMC)}, \\
\frac{1}{c^2} \left( \frac{3}{R_C} \right)^n & \text{(CDet)},
\end{array} \right. \tag{17}
\end{align*}
\]
where the factor 3 for CDet comes from the fact that the computational time per MC step is \( \sim 3^n \), because of the recursive formula that needs to be evaluated in order to eliminate disconnected diagrams [12]. As a result, for DiagMC most time is spent sampling the highest order, while for CDet this is the case only for \( R_C < \sqrt{3} \). Finally, we get
\[
\begin{align*}
t(\epsilon) &\sim \epsilon^{-\# \ln(\ln^{-1})} \quad \text{(DiagMC)}, \tag{18} \\
t(\epsilon) &\sim \epsilon^{-\alpha} \quad \text{(CDet)}. \tag{19}
\end{align*}
\]
Hence polynomial scaling is nearly reached with DiagMC, and is achieved with CDet. The exponent for CDet is given by
\[
\alpha = 2 + \frac{\ln(3/R_C^2)}{\ln R}. \tag{20}
\]
if \( R_C < \sqrt{3} \).

The case \( R_C > \sqrt{3} \) is particularly instructive. Here, most time is spent sampling low diagram orders, and one has \( \alpha = 2 \), which is the best scaling one can achieve in any Monte Carlo computation. We thus conclude that fermionic sign—all by itself—does not necessarily leads to any qualitative effect on the scaling of computational time with \( \epsilon \).

The above scalings are not purely academic considerations, as we illustrate with an example for CDet. We analyse the computation, reported in ref. [12], of the pressure of Fermi-Hubbard model in two dimensions. The diagrammatic scheme is a bare series with bare tadpoles taken into account through a shift of the chemical potential. The Hubbard parameters are: interaction \( U = 2 \), chemical potential \( \mu = 0.55978 \) and inverse temperature \( \beta = 8 \) (with hopping = 1); this corresponds to a density \( n = 0.875600(2) \). Figure 1(a) shows that \( |a_n| \) approaches an exponential behavior \( R^{-n} \) with \( R = 2.5(1) \), while fig. 1(b) shows that \( z_n \) approaches \( R^{-n} \) with \( R_C = 0.75(3) \) (see footnote 4). We can make three important observations. First, the exponent \( \alpha = 3.8(2) \) is not too large. Second, we clearly reach the asymptotic regime where eqs. (12), (16), and therefore also eq. (19), are valid. Third, \( |a_n| \) at low \( n \) is \( \sim 100 \) times larger than the extrapolation to low orders of the large-order behavior shown by the straight line in

\footnote{4}{For the density and the kinetic energy, we find the same value of \( R_C \) within our error bars.}
Carlo, for the pressure of the Fermi-Hubbard model (at
iteration space of connected determinant diagrammatic Monte
$U_{\mu} = 2$, $\beta = 8$, $n \simeq 0.875$). The lines are linear fits to the data at
large $n$.

These three observations explain why it was possible to obtain a $10^{-9}$ relative accuracy for the pressure in ref. [12]. Interestingly, the second and third observations hold independently of $U$ (with $\mu(U) = \mu_0 + Un_0/2$ as in [12]). In contrast, $\alpha$ diverges when $U$ tends to the critical value $U_c = 2R(U = 2) \simeq 5.1$ such that $R(U_c) = 1$. Divergent-series summation methods may allow to approach this point and even to go beyond it; we leave this for future study.

To avoid possible confusion, let us remark that we do not claim any connection between our results and the computational complexity theory of computer science. In this theory, a “problem instance” is defined by $N$ parameters, and the $P$ complexity class is defined by polynomial scaling of computational time with respect to $N$ for $N \to \infty$ (in the worst case with respect to all possible instances).

In the spin-glass problem discussed in [18], an instance is a disorder realisation, and $N$ is the number of random couplings (which happens to coincide with the system volume). In contrast, here we consider problems defined by a fixed (usually small) number of model-parameters (e.g., $T$, $U$ and $\mu$ for the Hubbard model in the thermodynamic limit).

We turn to “traditional” QMC, by which we mean here path-integral or auxiliary-field QMC. We consider the computation of an intensive quantity at finite temperature. Generically the FSP leads to an exponential scaling with spatial volume and inverse temperature of the average sign, and hence, due to eq. (4), of the statistical error (see, e.g., [1,2,4,10]):

$$\epsilon_{\text{stat}}(L) \sim \frac{e^{#2L^d}}{\sqrt{T}},$$

where $L$ is the linear system size.

Beside the statistical error, we also need to take into account the systematic error $\epsilon_{\text{sys}}(L)$ coming from the finite size $L$. The total error $\epsilon$ entering the CCP is $\epsilon \sim \epsilon_{\text{stat}}(L) + \epsilon_{\text{sys}}(L)$. We assume that finite-size corrections decrease exponentially,

$$\epsilon_{\text{sys}}(L) \sim e^{-\#L},$$

which is expected generically (away from second-order phase transitions and at finite temperature). For a given computational time $t$, the optimal strategy is to choose $L$ so that $\epsilon_{\text{sys}} \sim \epsilon_{\text{stat}}$, which yields

$$t(\epsilon) \sim e^{#23L^d}.$$ 

So for $d > 1$ the scaling of $t$ with $\epsilon^{-1}$ is quasi-polynomial and there is a CCP. In one dimension there is no CCP, which is another illustration of the simple observation presented in the introduction.

Apart from these asymptotic scalings, there are also practical advantages of diagrammatic methods over traditional QMC. For traditional QMC, the condition for getting close to the thermodynamic limit is typically $L$ much larger than the correlation length. Equation (21) then yields a computational time $t \propto e^{#23L^d}$ which is often prohibitive, meaning that one cannot get close to the thermodynamic limit, and that one cannot even reach the asymptotic scaling regimes (22), (23). The situation is very different in diagrammatic expansions, where as shown by the above example, the asymptotic regime is accessible, and, moreover, the lowest orders typically set the scale while higher-order contributions are merely corrections.

In conclusion, unbiased numerical methods for solving quantum many-fermion problems should be evaluated on the basis of their scaling of computational time with respect to the final error bar on thermodynamic-limit quantities. The presence of a fluctuating sign does not suffice to say that a problem is intractable by Monte Carlo. Nothing prevents in principle a polynomial scaling of the CPU time vs. the inverse error bar. We demonstrated that such polynomial complexity is indeed achieved by the recently proposed CDet method when inside the radius of convergence of the Feynman diagrammatic series. Since this method offers the possibility to calculate properties of many-fermion systems in polynomial time, it is fair to
say that the sign problem has become irrelevant here and a numerical solution to the many-fermion problem is available, at least in some region of parameter space.

***

We acknowledge support from ERC Grant Thermodynamix (NP and FW), the Simons Foundation’s Many Electron Collaboration, the National Science Foundation under the grant PHY-1314735, and the MURI Program New Quantum Phases of Matter from AFOSR (NP and BS). Some of us are members of Paris Sciences et Lettres, Sorbonne Universités (RR, KVH and FW) and Sorbonne Paris-Cité (RR and KVH).

REFERENCES

[1] Loh E. Y., Gubernatis J. E., Scalettar R. T., White S. R., Scalapino D. J. and Sugar R. L., Phys. Rev. B, 41 (1990) 9301.
[2] Ceperley D. M., Path integral Monte Carlo methods for fermions, in Monte Carlo and Molecular Dynamics of Condensed Matter Systems, edited by Binder K. and Ciccotti G., Vol. 49 (Editrice Compositori, Bologna, Italy) 1996.
[3] Ceperley D. M., Quantum Monte Carlo Methods for Fermions, in Proceedings of the Les Houches Summer School, Session 56, Strongly Interacting Fermions and High Tc Superconductivity, edited by Doucet B. and Zinn-Justin J. (Elsevier) 1995.
[4] Assaad F. F. and Evertz H. G., World-line and Determinantal Quantum Monte Carlo Methods for Spins, Phonons and Electrons, in Computational Many-Particle Physics, edited by Feibik H., Schneider R. and Weiße A., Lect. Notes Phys., Vol. 739 (Springer) 2008, pp. 277-356.
[5] Chandrasekharan S. and Wiese U.-J., Phys. Rev. Lett., 83 (1999) 3116.
[6] Berg E., Metlitski M. A. and Sachdev S., Science, 338 (2012) 1606.
[7] Wei Z. C., Wu C., Yi Li, Zhang S. and Xiang T., Phys. Rev. Lett., 116 (2016) 256001.
[8] Alet F., Damle K., and Pujari S., Phys. Rev. Lett., 117 (2016) 197203.
[9] Honecker A., Wessel S., Kerkdyk R., Pruschke T., Mila F. and Normand B., Phys. Rev. B, 93 (2016) 054408.
[10] Iazzi M., Soluyanov A. A. and Troyer M., Phys. Rev. B, 93 (2016) 115102.
[11] Capponi S., J. Phys.: Condens. Matter, 29 (2017) 043002.
[12] Rossi R., arXiv:1612.05184.
[13] Van Houcke K., Kozik E., Prokof’ev N. and Svistunov B., Phys. Procedia, 6 (2010) 95.
[14] Benfatto G., Giuliani A. and Mastropietro V., Ann. Henri Poincaré, 7 (2006) 809.
[15] Abdesselam A. and Rivaasseau V., Lett. Math. Phys., 44 (1998) 77.
[16] Magnen J., private communication.
[17] Zinn-Justin J., Phys. Rep., 70 (1981) 109.
[18] Troyer M. and Wiese U.-J., Phys. Rev. Lett., 94 (2005) 170201.