HU-EP-13/05
SFB/CPP-13-13
DESY 13-022

Quasi-Monte Carlo methods for lattice systems: a first look

K. Jansen[1,2], H. Leovey[3], A. Nube[1,4], A. Griewank[3], and M. Müller-Preussker[4]

[1] *NIC, DESY Zeuthen, Platanenallee 6, D-15738 Zeuthen, Germany*
[2] *Department of Physics, University of Cyprus, P.O.Box 20537, 1678 Nicosia, Cyprus*
[3] *Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin*
[4] *Institut für Physik, Humboldt-Universität zu Berlin, Newtonstr. 15, D-12489 Berlin*

February 27, 2013

Abstract

We investigate the applicability of Quasi-Monte Carlo methods to Euclidean lattice systems for quantum mechanics in order to improve the asymptotic error behavior of observables for such theories. In most cases the error of an observable calculated by averaging over random observations generated from an ordinary Markov chain Monte Carlo simulation behaves like $N^{-1/2}$, where N is the number of observations. By means of Quasi-Monte Carlo methods it is possible to improve this behavior for certain problems up to N^{-1}. We adapted and applied this approach to simple systems like the quantum harmonic and anharmonic oscillator and verified an improved error scaling.

1 Introduction

Markov chain-Monte Carlo (Mc-MC) techniques are commonly the method of choice for the numerical evaluation of partition functions in statistical physics or path integrals in Euclidean time for models in high energy physics. The reason is that they are based on importance sampling and hence select the integration points automatically according to the corresponding weight in the integrand. Many algorithms have been developed to implement a Mc-MC, starting from the Metropolis algorithm, heatbath and over-relaxation to cluster and hybrid Monte Carlo algorithms, see e.g. refs. [1, 2]. In this way, simulations of demanding 4-dimensional quantum field theories became possible and, in fact, were carried out very successfully to e.g. compute the low-lying hadron spectrum [3] or deriving bounds for the Higgs boson mass [4].

The drawback of Mc-MC is that it estimates the desired quantity stochastically and hence the results are affected by a statistical error which sometimes needs very long and computer time extensive samplings. Quantitatively, this sampling error behaves as $1/\sqrt{N}$ for a (thermalized) sample size of N. This

error scaling behaviour is often a real stumbling block in such Mc-MC simulations. If we consider lattice quantum chromodynamics as a typical system for Mc-MC calculations in high energy physics, then due to this error scaling and the very high computational demand of these simulations, it is often impossible to significantly decrease the error to the targeted precision. It would therefore be very desirable to have Monte Carlo methods available that possibly show a better error scaling.

Such methods in fact exist in form of Quasi-Monte Carlo (QMC) techniques [5],[6], where it is known that the error scaling can be improved to an $1/N$ behaviour in the optimal case. However, although QMC methods have been analyzed theoretically very thoroughly and comprehensively, in practice they have found only limited application and, to the best of our knowledge, were never tested in models that are relevant for high energy physics.

In this paper we therefore want to perform a very first step towards the very challenging goal of applying QMC methods to generic field theories by looking at the non-trivial case of the anharmonic quantum mechanical oscillator discretized on a finite Euclidean time lattice and evaluated in the corresponding path integral formulation [7]. For the case of the anharmonic oscillator the system is not Gaussian anymore and a successful application of QMC methods would be a first non-trivial test. Of course, even if such a test is successful, there is a long way to address eventually 4-dimensional quantum field theories, but a proof of concept would certainly open the promising road to attack field theories in the future.

We will start our discussion with a description of the harmonic oscillator in its time-discretized path integral formulation. Here the problem is fully Gaussian and the application of QMC methods should lead to the optimal error scaling of $1/N$, an expectation that we will see to be fulfilled. Nevertheless, the harmonic oscillator example can serve well to explain how QMC methods work and how this expected optimal error scaling behaviour is realized.

We will then proceed to look at the anharmonic oscillator and we will demonstrate that also in this case QMC leads to an improved error scaling, although we will not be able to reach the optimal one. We consider this, nevertheless to be a very promising, non-trivial result which bears the potential that also other models in quantum mechanics, e.g. the topological quantum mechanical action of ref. [8] and even field theories can be evaluated by QMC methods. For a first account of our studies, we refer to the proceedings contribution of ref. [9].

2 Quantum Mechanical Harmonic and Anharmonic Oscillator

In this section we will discuss the basic steps for the quantization of the theory in the path integral approach and the discretization on a time lattice. The first step is the construction of the Lagrangian (resp. the action) of the corresponding classical mechanical system for a given path $x(t)$ of a particle with mass M_0. For a numerically stable evaluation of the path integral it is essential to pass on

to Euclidean time. In this case the Lagrangian L and the action S is given by:

$$L(x,t) = \frac{M_0}{2}\left(\frac{dx}{dt}\right)^2 + V(x) \tag{1}$$

$$S(x) = \int_0^T L(x,t)\, dt. \tag{2}$$

Depending on the scenario (harmonic or anharmonic oscillator) the potential $V(x)$ consists of two parts

$$V(x) = \underbrace{\frac{\mu^2}{2}x^2}_{\text{harmonic part}} + \underbrace{\lambda\, x^4}_{\text{anharmonic part}}, \tag{3}$$

such that the parameter λ controls the anharmonic part of the theory. It should also be mentioned that in the anharmonic case the parameter μ^2 can take on negative values, leading then to a double well potential.

The next step is to discretize time into equidistant time slices with a spacing of a. The path is then only defined on the time slices:

$$t \to t_i = (i-1)\cdot a \quad i = 1 \ldots d \tag{4}$$

$$x(t) \to x_i = x(t_i)\,. \tag{5}$$

On the lattice the derivative with respect to the time appearing in (1) (first term) will be replaced by the forward finite difference $\nabla x_i = \frac{1}{a}(x_{i+1} - x_i)$. The choice of the lattice derivative is not unique and requires special care, particularly if one considers more complicated models like lattice QCD. But in [7] it was shown that the lattice derivative chosen here permits a well defined continuum limit. Putting all the ingredients together, we can write down the lattice action for the (an)harmonic oscillator

$$S^{\text{latt}}(x) = a\sum_{i=1}^{d}\left(\frac{M_0}{2}\left(\nabla x_i\right)^2 + V(x_i)\right)\,. \tag{6}$$

For the path a cyclic boundary condition $x_{d+1} = x_1$ can be assumed. In the following the superscript "latt" will be dropped, as we will only refer to the lattice action from now on. The expectation value of an observable O of the quantized theory expressed in terms of the path integral reads as follows:

$$\langle O(x)\rangle \;=\; \frac{\int_{\mathbb{R}^d} O(x) e^{-S(x)} dx_1 ... dx_d}{\int_{\mathbb{R}^d} e^{-S(x)} dx_1 ... dx_d}\,. \tag{7}$$

This expression is suitable for a numerical evaluation of certain quantities of the underlying theory. Up to now only Monte Carlo methods are known to give reliable results for dimensions $d \gg 10$. One type of such methods, often used in physics, is the Markov chain-Monte Carlo approach mostly applying the weight $\propto e^{-S(x)}$ for sampling paths $\{x_i\}$ (so-called "importance sampling"). In the next sections, we will provide a summary of the mathematical results for the QMC methods and particularly recapitulate in a rather mathematical language the strict error scaling bounds for this method. The reader more interested directly in the results may move to section 7 directly.

3 Direct Monte Carlo and quasi–Monte Carlo methods

In many practical applications one is interested in calculating quotients of the form (7) where the action $S(.)$ and the observables $O(.)$ are usually smooth functions in high dimensions. In some special situations where one would like to deal with integrands of moderately high dimensions, one possible approach is to consider estimators $\hat{I}_1$,$\hat{I}_2$ for the integrals I_1,I_2 in the numerator and in the denominator of (7) separately, and then take $\hat{I}_1/\hat{I}_2$ as an estimation of $\langle O(x)\rangle$. Another possible approach one can consider is given by the so called weighted uniform sampling (WUS) estimator, analyzed in [10]. In the latter case, one take a joint estimator for the total quantity $\langle O(x)\rangle$, using a single direct sampling method. We will show some characteristics of the WUS estimator in section 7, and we will refer from now on to the latter two approaches as *direct* sampling methods for estimating (7). In many interesting examples, we encounter the case were the action $S(.)$ and the observable $O(.)$ lead to integrals I_1,I_2 of Gaussian type. Then the integrals I_1,I_2 can be written in the form

$$I_i \;=\; \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \int_{\mathbb{R}^d} g_i(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\top C^{-1}\mathbf{x}} d\mathbf{x}, \quad \mathbf{x} = (x_1,\ldots,x_d),\; i = 1,2 \quad ,$$

where C denotes the covariance matrix of the Gaussian density function. A transformation to the unit cube in $\mathbb{R}^d$ can be applied such that the corresponding integrals take the form

$$I = \int_{[0,1]^d} g(A\mathbf{\Phi}^{-1}(\boldsymbol{z}))d\boldsymbol{z} = \int_{[0,1]^d} f(\boldsymbol{z})d\boldsymbol{z} = I_{[0,1]^d}(f), \quad \boldsymbol{z} = (z_1,\ldots,z_d)\,. \tag{8}$$

Here $AA^\top = C$ is some symmetric factorization of the covariance matrix, and $\mathbf{\Phi}^{-1}(\boldsymbol{z}) := (\Phi^{-1}(z_1),\ldots,\Phi^{-1}(z_d))^\top$, where $\Phi^{-1}(\cdot)$ represents the inverse of the normal cumulative distribution function $\Phi(\cdot)$.
In the classical direct Monte–Carlo (MC) approach one tries to estimate (8) by generating samples pseudo-randomly. One starts with a finite sequence of independent identically distributed (i.i.d.) samples $P_N = \{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N\}$, where the points $\boldsymbol{z}_j,\; 1 \le j \le N$, have been generated from the uniform distribution in $[0,1]^d$. Then, the quadrature rule is fixed by taking the average of the function evaluations for f

$$Q_N := \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{z}_j),$$

as an approximation of the desired integral $\int_{[0,1]^d} f(\boldsymbol{z})\; d\boldsymbol{z}$. The resulting estimator $\hat{Q}_N$ is unbiased. The integration error can be approximated via the central limit theorem, given that f belongs to $L_2([0,1]^d)$. The variance of the estimator $\hat{Q}_N$ is given by

$$\frac{\sigma^2}{N} = \frac{1}{N}\left(\int_{[0,1]^d} f^2(\boldsymbol{z})\; d\boldsymbol{z} - \left(\int_{[0,1]^d} f(\boldsymbol{z})\; d\boldsymbol{z}\right)^2\right).$$

As measured by its standard deviation from zero the integration error associated with the MC approach is then of order $O(N^{-\frac{1}{2}})$. The quality of the MC samples

relies on the selected pseudo–random number generators of uniform samples, here we use the *Mersenne Twister* generator from Matsumoto and Nishimura (see [11]). MC is in general a very reliable tool in high–dimensional integration, but the order of convergence is in fact rather poor.

In contrast, quasi–Monte Carlo (QMC) methods generates deterministically point sets that are more regularly distributed than the pseudo–random points from MC (see [5], [12], [13], [6]). Typical examples of QMC are shifted lattice rules and low–discrepancy sequences. In order to give a short introduction to the subject, we define now the classical notion of discrepancy of a finite sequence of points P_N in $[0,1)^d$. Given $P_N = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$ a set of points in $[0,1)^d$, and a nonempty family $\mathbb{I}$ of Lebesgue-measurable sets in $[0,1)^d$, we define the classical discrepancy function by

$$D(\mathbb{I}; P_N) := \sup_{B \in \mathbb{I}} \left| \frac{\sum_{i=1}^{N} c_B(\boldsymbol{z}_i)}{N} - \lambda_d(B) \right|,$$

where c_B is the characteristic function of B. This allows us to define the so called *star discrepancy*

Definition 3.1 *We define the star discrepancy $D^\star(P_N)$ of the point set P_N by $D^\star(P_N) := D(\mathbb{I}; P_N)$, where $\mathbb{I}$ is the family of all sub-intervals of the form $\prod_{i=1}^{d}[0, u_i)$, with $u_i \geq 0,\ 1 \leq i \leq d$.*

The *star discrepancy* can be considered as a measure of the worst difference between the uniform distribution and the sampled distribution in $[0,1)^d$ attributed to the point set P_N. The usual way to analyze QMC as a deterministic method is by choosing a class of integrand functions F, and a measure of discrepancy $D(P_N)$ for the point sets P_N. Then, the deterministic integration error is usually given in the form

$$|Q_N - \int_{[0,1]^d} f(\boldsymbol{z})\, d\boldsymbol{z}| \ \leq\ D(P_N)V(f),$$

where $V(f)$ measures a particular variation of the function $f \in F$. A classical particular error bound in this form is the famous Koksma–Hlawka inequality, where $D(P_N)$ is taken to be the *star discrepancy* of the point set P_N, and $V(f)$ is the variation in the sense of Hardy and Krause of f. In the context of QMC, a sequence of points in $[0,1)^d$ is called a low–discrepancy sequence if $D^\star(P_N) = O(N^{-1}(\log(N))^d)$ for all truncations of the sequence to its first N terms.

3.1 Quasi–Monte Carlo errors and complexity

For error analysis of QMC methods, there are certain reproducing kernel Hilbert spaces $\mathbb{F}_d$ of functions $f : [0,1]^d \to \mathbb{R}$ that are particularly useful (see [14]). Let us denote now with $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the inner product and norm in $\mathbb{F}_d$. A reproducing kernel is a function $K : [0,1]^d \times [0,1]^d \to \mathbb{R}$ satisfying the properties

1. $K(\cdot, \boldsymbol{y}) \in \mathbb{F}_d$ for each $\boldsymbol{y} \in [0,1]^d$
2. $\langle f, K(\cdot, \boldsymbol{y}) \rangle = f(\boldsymbol{y})$ for each $\boldsymbol{y} \in [0,1]^d$ and $f \in \mathbb{F}_d$

If the integral

$$I(f) = \int_{[0,1]^d} f(\boldsymbol{z}) d\boldsymbol{z}$$

is a continuous functional on the space $\mathbb{F}_d$, then the worst case quadrature error

$$e_N(\mathbb{F}_d) := \sup_{f \in \mathbb{F}_d, \|f\| \leq 1} |I(f) - Q_N(f)|$$

for point sets $P_N = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}$ and quasi-Monte Carlo algorithms for the space $\mathbb{F}_d$ can be given by

$$e_N(\mathbb{F}_d) = \sup_{\|f\| \leq 1} |\langle f, h_N \rangle| = \|h_N\|$$

for some $h_N \in \mathbb{F}_d$ due to Riesz' representation theorem. In this case, the *representer* h_N of the quadrature error is given explicitly in terms of the kernel by

$$h_N(\boldsymbol{z}) = \int_{[0,1]^d} K(\boldsymbol{z}, \boldsymbol{y}) d\boldsymbol{y} - \frac{1}{N} \sum_{i=1}^{N} K(\boldsymbol{z}, \boldsymbol{z}_i), \quad \forall \boldsymbol{z} \in [0,1]^d.$$

Tensor product reproducing kernel Hilbert spaces are of particular interest, since the multivariate kernel result as the product of the underlying univariate kernels. In QMC error analysis, the weighted (anchored) tensor product Sobolev space introduced in [15] is often considered

$$\mathbb{F}_d = \bigotimes_{i=1}^{d} W_2^1([0,1]),$$

also denoted with $\mathbb{F}_d = \mathcal{W}_{2,\text{mix}}^{(1,\ldots,1)}([0,1]^d)$. The weighted norm $\|f\|_\gamma^2 = \langle f, f \rangle_\gamma$ results from the inner product

$$\langle f, g \rangle_\gamma = \sum_{u \subseteq \{1,\ldots,d\}} \prod_{j \in u} \gamma_j^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial \boldsymbol{z}_u} f(\boldsymbol{z}_u, \boldsymbol{1}) \frac{\partial^{|u|}}{\partial \boldsymbol{z}_u} g(\boldsymbol{z}_u, \boldsymbol{1}) d\boldsymbol{z}_u, \tag{9}$$

where for $u \subseteq \{1, \ldots, d\}$ we denote by $|u|$ its cardinality, and $(\boldsymbol{z}_u, \boldsymbol{1})$ denotes the vector containing the coordinates of $\boldsymbol{z}$ with indices in u, and the other coordinates set equal to 1.

In this case the reproducing kernel is given by

$$K_{d,\gamma}(\boldsymbol{z}, \boldsymbol{y}) = \prod_{j=1}^{d} (1 + \gamma_j [1 - \max(z_j, y_j)]) \quad \text{for } \boldsymbol{z}, \boldsymbol{y} \in [0,1]^d.$$

The weighted tensor product Sobolev space allow for explicit QMC constructions deriving error estimates of the form

$$e_N(\mathbb{F}_d) \leq C(\delta) N^{-1+\delta} \quad \text{for } \delta \in (0, \tfrac{1}{2}], \tag{10}$$

where the constant $C(\delta)$ is independent on the dimension d, if the sequence of weights (γ_j) satisfies the condition (see [16])

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \,.$$

Traditional unweighted function spaces considered for integration suffer from the curse of dimensionality. Their weighted variants describe a setting where the variables or group of variables may vary in importance, corresponding to an anisotropic setting. Many integration problems in practice start with an isotropic setting but can be modified to an anisotropic setting using a proper transformation. This gives a partial explanation of why some very high-dimensional spaces become tractable for QMC.
Explicit QMC constructions satisfying (10) are *shifted lattice rules* for weighted spaces. The rate (10) can be also obtained for Niederreiter and Sobol' sequences (see [17]). The idea of "weighting" the norm of the spaces to obtain tractable results can be applied in fact to more general function spaces than smooth function spaces of tensor product form, and many integration examples can be found in [12]. In our numerical experiments, we used so far QMC algorithms based on a particular type of low–discrepancy sequences. Numerical experiments with shifted lattice rules will be carried out in the near future, following new techniques for fixing adequate weights introduced in [18].

4 Low–discrepancy (t, d)-sequences

The most well known type of low–discrepancy sequences are the so called (t, d)-sequences. To introduce how (t, m, d)-nets and (t, d)-sequences are defined, we consider first *elementary intervals* in a integer base $b \geq 2$. Let E be any subinterval of $[0, 1)^d$ of the form $E = \prod_{i=1}^{d} [a_i b^{-c_i}, (a_i + 1)b^{-c_i})$ with $a_i, c_i \in \mathbb{N}, c_i \geq 0$, $0 \leq a_i < b^{-c_i}$ for $1 \leq i \leq d$. An interval of this form is called an elementary interval in base b.

Definition 4.1 *Let $0 \leq t \leq m$ be integers. A (t, m, d)-net in base b is a point set P_N of $N = b^m$ points in $[0, 1)^d$ such that every elementary interval E in base b with $\lambda_d(E) = \frac{b^t}{b^m}$ contains exactly b^t points.*

Definition 4.2 *Let $t \geq 0$ be an integer. A sequence $\mathbf{z}_1, \mathbf{z}_2, \ldots$ of points in $[0, 1)^d$ is a (t, d)-sequence in base b if for all integers $k \geq 0$ and $m > t$, the point set consisting of $N = b^m$ points $\mathbf{z}_i$ with $kb^m \leq i < (k + 1)b^m$, is a (t, m, d)-net in base b.*

The parameter t is called the *quality parameter* of the (t, d)–sequences. In [19], theorem 4.17, it is shown that (t, d)-sequences are in fact low–discrepancy sequences. We reproduce this result in the following

Theorem 4.3 *The star-discrepancy $D^\star$ of the first N terms P_N of a (t, d)-sequence in base b, satisfies*

$$ND^\star(P_N) \leq C(d, b)b^t(log(N))^d + O(b^t(log(N))^{d-1}),$$

where the implied constants depend only on b and d. If either $d = 2$ or $b = 2$, $d = 3, 4$, we have

$$C(d, b) = \frac{1}{d}\left(\frac{b-1}{2log(b)}\right)^d,$$

and otherwise

$$C(d, b) = \frac{1}{d!}\frac{b-1}{2\lfloor b/2 \rfloor}\left(\frac{\lfloor b/2 \rfloor}{log(b)}\right)^d.$$

Explicit constructions of (t, d)-sequences are available. Examples are the generalized Faure, Sobol', Niederreiter and Niederreiter–Xing sequences. All these examples fall into the category of constructions called *digital sequences*. We refer to [13] for further reading on this topic.

5 Randomized QMC

There are some advantages in retaining the probabilistic properties of the sampling. There are practical hybrid methods permitting us to combine the good features of MC and QMC. Randomization is an important tool for QMC if we are interested for a practical error estimate of our sample quadrature Q_N to the desired integral. One goal is to randomize the deterministic point set P_N generated by QMC in a way that the estimator $\hat{Q}_N$ preserves unbiasedness. Another important goal is to preserve the better equidistribution properties of the deterministic construction.

The simplest form of randomization applied to *digital sequences* seems to be the technique called *digital b–ary shifting*. In this case, we add a random shift $\Delta \in [0,1)^d$ to each point of the deterministic set $P_N = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_N\}$ using operations over the selected ring $\mathbb{F}_b$. The application of this randomization preserves in particular the t value of any projection of the point set (see [5] and references therein). The resulting estimator is unbiased.
The second randomization method we consider is the one introduced by Owen ([20]) in 1995. He considered (t, m, d)-nets and (t, d)-sequences in base b and applied a randomization procedure based on permutations of the digits of the values of the coordinates of points in these nets and sequences. This can be interpreted as a random scrambling of the points of the given sequence in such a way that the net structure remains unaffected. We do not discuss here in detail Owen's randomization procedure, or from now on called *Owen's scrambling*. The main results of this randomization procedure can be stated in the following

Proposition 5.1 ***(Equidistribution)***
A randomized (t, m, d)-net in base b using Owen's scrambling is again a (t, m, d)-net in base b with probability 1. A randomized (t, d)-sequence in base b using Owen's scrambling is again a (t, d)-sequence in base b with probability 1.

Proposition 5.2 ***(Uniformity)***
Let $\tilde{\boldsymbol{z}}_i$ be the randomized version of a point $\boldsymbol{z}_i$ originally belonging to a (t, m, d)-net in base b or a (t, d)-sequence in base b, using Owen's scrambling. Then $\tilde{\boldsymbol{z}}_i$ has the uniform distribution in $[0,1)^d$, that is, for any Lebesgue measurable set $G \subseteq [0,1)^d$, $P(\tilde{\boldsymbol{z}}_i \in G) = \lambda_d(G)$, with λ_d the d-dimensional Lebesgue measure.

The last two propositions state that after *Owen's scrambling* of *digital sequences* we retain unaffected the low discrepancy properties of the constructions, and that after this randomization procedure we obtain random samples uniformly distributed in $[0,1)^s$.

The basic results about the variance of the randomized QMC estimator $\hat{Q}_N$ after applying *Owen's scrambling* to (t, m, d)-nets in base b (or of (t, d)-sequences in base b) can be found in [21]. We summarize these results in the following

Theorem 5.3 *Let $\tilde{\boldsymbol{z}}_i$, $1 \le i \le N$, be the points of a scrambled (t, m, d)-net in base b, and let f be a function on $[0,1)^d$ with integral I and variance $\sigma^2 = \int (f-I)^2 d\boldsymbol{z} < \infty$. Let $\hat{Q}_N = N^{-1} \sum_{i=1}^N f(\tilde{\boldsymbol{z}}_i)$, where $N = b^m$. Then for the variance $V(\hat{Q}_N)$ of the randomized QMC estimator it holds*

$$V(\hat{Q}_N) = o(1/N), \quad \text{as } N \to \infty, \quad \text{and} \quad V(\hat{Q}_N) \le \frac{b^t}{N} \left(\frac{b+1}{b-1} \right)^d \sigma^2.$$

For $t = 0$ we have

$$V(\hat{Q}_N) \le \frac{1}{N} \left(\frac{b}{b-1} \right)^{d-1} \sigma^2.$$

The above theorem says that the variance of a QMC estimator $\hat{Q}_N$ using scrambled $(0, m, d)$–nets is always smaller than a small multiple of the variance of the corresponding MC estimator. The bound of the theorem above can be improved (see for example theorem 13.9 in [13]). If the integrand at hand is smooth enough, using *Owen's scrambling* it can be shown that one can obtain an improved asymptotic error estimate of order $O(N^{-\frac{3}{2}-\frac{1}{d}+\delta})$, for any $\delta > 0$, see [22]. Improved scrambling techniques have been developed in [23],[24].

6 Effective dimensions and sensitivity indices

In many practical applications, one encounters functions for which the total variance is concentrated in a small part of its ANOVA terms. The notion of *effective dimension* of a function was first introduced in [25] to describe the contribution of a group of variables to the total variance.

6.1 ANOVA Decomposition

Using the *ANOVA (Analysis of Variance) decomposition* we decompose a function into a sum of simpler functions, see [26]. Let $D = \{1, \ldots, d\}$. For any subset $\mathbf{i} \subseteq D$, let $|\mathbf{i}|$ denote its cardinality and $(D - \mathbf{i})$ be its complementary set in D. Let $\boldsymbol{z}_\mathbf{i} = (z_j : \ j \in \mathbf{i})$ be the $|\mathbf{i}|-$dimensional vector containing the coordinates of $\boldsymbol{z}$ with indices in $\mathbf{i}$. Now assume that f is a square integrable function. Then we can write f as the sum of 2^d ANOVA terms:

$$f(\boldsymbol{z}) = \sum_{\mathbf{i} \subseteq D} f^\mathbf{i}(\boldsymbol{z}) \ ,$$

where the ANOVA terms $f^\mathbf{i}(\mathbf{x})$ are defined recursively by

$$f^\mathbf{i}(\boldsymbol{z}) = \int_{[0,1]^{d-|\mathbf{i}|}} f(\boldsymbol{z}_\mathbf{i}, \boldsymbol{z}_{D-\mathbf{i}}) d\boldsymbol{z}_{D-\mathbf{i}} \ - \sum_{\mathbf{j} \subsetneq \mathbf{i}} f^\mathbf{j}(\boldsymbol{z}) \ ,$$

and $f^\varnothing = I(f)$. The sum of the right–hand side is over strict subsets $\mathbf{j} \neq \mathbf{i}$, and we use the convention $\int_{[0,1]^0} f(\boldsymbol{z}) d\boldsymbol{z}_\varnothing = f(\boldsymbol{z})$. The ANOVA terms enjoy the following interesting properties:

1. $\int_0^1 f^\mathbf{i}(\boldsymbol{z}) dz_j = 0$ for $j \in \mathbf{i}$.

2. The decomposition is orthogonal, in that $\int_{[0,1]^d} f^{\mathbf{i}}(\boldsymbol{z}) f^{\mathbf{j}}(\boldsymbol{z}) d\boldsymbol{z} = 0$ whenever $\mathbf{i} \neq \mathbf{j}$.

3. Let $\sigma^2(f) = \int_{[0,1]^d} f(\boldsymbol{z})^2 \, d\boldsymbol{z} - (I(f))^2$ be the variance of f, then we have:

$$\sigma^2(f) \;=\; \sum_{\mathbf{i} \subseteq D} \sigma_{\mathbf{i}}^2(f), \qquad \text{where} \qquad \sigma_{\mathbf{i}}^2(f) = \int_{[0,1]^d} f^{\mathbf{i}}(\boldsymbol{z})^2 \, d\boldsymbol{z}$$

for $|\mathbf{i}| > 0$ is the variance of $f^{\mathbf{i}}$ and $\sigma_{\varnothing}^2(f) = 0$.

Definition 6.1

1. *f is said to have effective dimension in the superposition sense d_s with proportion p, for $0 < p < 1$, if d_s is the smallest integer that satisfies*

$$\sum_{|\mathbf{i}| \leq d_s} \sigma_{\mathbf{i}}^2(f) \geq p\sigma^2(f).$$

2. *f is said to have effective dimension in the truncation sense d_t with proportion p, for $0 < p < 1$, if d_t is the smallest integer that satisfies*

$$\sum_{\mathbf{i} \subseteq \{1,\ldots,d_t\}} \sigma_{\mathbf{i}}^2(f) \geq p\sigma^2(f).$$

One can estimate the effective dimension in truncation sense based on the algorithm proposed by Wang and Fang [27]. They show that the following equality holds

$$\int_{[0,1]^{2d-|u|}} f(\boldsymbol{z}) f(\boldsymbol{z}_u, \mathbf{y}_{D-u}) d\boldsymbol{z} d\mathbf{y}_{D-u} = \sum_{\mathbf{i} \subseteq u} \sigma_{\mathbf{i}}^2(f) + f^{\varnothing}.$$

Thus, for estimating the effective dimension in truncation sense, we need to estimate the following tree type of integrals

$$\int_{[0,1]^d} f(\boldsymbol{z}) d\boldsymbol{z}, \quad \int_{[0,1]^d} f^2(\boldsymbol{z}) d\boldsymbol{z}, \quad \int_{[0,1]^{2d-|u|}} f(\boldsymbol{z}) f(\boldsymbol{z}_u, \mathbf{y}_{D-u}) d\boldsymbol{z} d\mathbf{y}_{D-u}, \quad (11)$$

for $u = \{1, \ldots, l\}$, $l = 1, 2, \ldots$, using MC or QMC, until the proportion of variance defining the effective dimension is reached. In many applications, the proportion value is usually taken as $p = 0.99$. For problems exhibiting low *effective dimensions* with an effective part of the function belonging to the weighted tensor product Sobolev space, one can expect that QMC methods outperform MC for high–dimensional integration. If the truncation effective dimension is small, than few variables are important for sampling. If the superposition effective dimension is small, say $d_s = 2$ or maybe $d = 3$, than some QMC sequences are also expected to outperform MC, because they can exhibit more equidistributed low dimensional projections than MC.
The ordering of the variables of the integrand is important for achieving a reduction of the effective dimension in the truncation sense d_t, and usually affects the performance of QMC in practice. Sensitivity indices usually help to order the variables in a convenient way for integration.

6.2 Derivative based sensitivities

As pointed out by Sobol' and Kucherenko in [28], very often derivative based measures of sensitivities can successfully be used for detecting non essential variables. Small values of first order derivatives of a function implies small values of one–dimensional total Sobol' sensitivity indices. Let $\sigma^2_{\mathbf{i}}(f)$ denote the partial variance corresponding to the ANOVA term $f^{\mathbf{i}}$. Define

$$\sigma^2_{\{j\}}(f)^{\text{tot}} = \sum_{\mathbf{i}\subset D: j\in\mathbf{i}} \sigma^2_{\mathbf{i}}(f),$$

then it is shown in [26] and [28] that

$$\sigma^2_{\{j\}}(f)^{\text{tot}} = \frac{1}{2}\int_{[0,1]^d}\int_0^1 [f(\boldsymbol{z}) - f(z_1,\cdots,z_{j-1},z'_j,z_{j+1},\cdots,z_n)]^2 d\boldsymbol{z}dz'_j,$$

from which one obtain the following two results:

1. if $c < |\frac{\partial f}{\partial z_j}| < C$, then

$$\frac{c^2}{12} \leq \sigma^2_{\{j\}}(f)^{\text{tot}} \leq \frac{C^2}{12}\,,$$

2. and if $\frac{\partial f}{\partial z_j} \in L_2([0,1]^d)$, then

$$\sigma^2_{\{j\}}(f)^{\text{tot}} \leq \frac{1}{\pi^2}\int_{[0,1]^d}\left(\frac{\partial f}{\partial z_j}(\mathbf{x})\right)^2 d\boldsymbol{z}. \tag{12}$$

As a consequence of the bounds stated above, the total variance corresponding to non–essential variables of a function can be bounded using first order derivatives information. In a wide variety of problems in practice, the gradient of a scalar function can be efficiently computed trough algorithmic differentiation (see [29]), at a cost at most 4 times of that for evaluating the original function. Thus, a cheap method for estimating derivative based sensitivities, and an upper bound on the effective dimension in the truncation sense (as stated in the following simple Proposition), may be available using algorithmic differentiation. The variance is, clearly, invariant to a permutation of the variables. This allowed us to consider the following

Definition 6.2 *Given an bijection (permutation) $\pi : \{1,\ldots,d\} \to \{1,\ldots,d\}$, f is said to have π–effective dimension in the truncation sense d_t with proportion p, for $0 < p < 1$, if d_t is the smallest integer that satisfies*

$$\sum_{\mathbf{i}\subset\{\pi^{-1}(1),\ldots,\pi^{-1}(d_t)\}} \sigma^2_{\mathbf{i}}(f) \geq p\sigma^2(f).$$

Proposition 6.3 *Let $f \in L_2([0,1]^d)$ such that $\frac{\partial f}{\partial z_j} \in L_2([0,1]^d)$ $\forall$ $1 \leq j \leq n$. Consider the derivative based sensitivities*

$$v_i := \frac{1}{\pi^2}\int_{[0,1]^d}\left(\frac{\partial f}{\partial z_j}(\boldsymbol{z})\right)^2 d\boldsymbol{z},$$

and consider any permutation $\pi^* : \{1, \ldots, d\} \to \{1, \ldots, d\}$ *such that*

$$v_{(\pi^*)^{-1}(k)} \geq v_{(\pi^*)^{-1}(k+1)}, \quad \forall\, 1 \leq k \leq d-1,$$

(a non-increasing ordering of the sensitivities v_i*'s, resulting in what is called by the authors a "Diff–decay–ordering"* π^**).*
Let $0 < p < 1$ *be a fixed proportion parameter. If there exists an integer* m *such that*

$$\sum_{j=m+1}^{d} v_{(\pi^*)^{-1}(j)} \leq (1-p)\sigma^2(f) \tag{13}$$

then, it follows that the π^**–effective dimension in the truncation sense with proportion* p *is at most* $d - m$*.*

Proof.

Let d_t denote the π^* –effective dimension in the truncation sense with proportion p. Consider m satisfying (13) and define $T_m = \{\mathbf{i} : \mathbf{i} \subset \{(\pi^*)^{-1}(1), ..., (\pi^*)^{-1}(m)\}\}$ and $f_{T_m} = \sum_{\mathbf{i} \in T_m} f^{\mathbf{i}}(\boldsymbol{z})$. It follows from the L_2 orthogonality of ANOVA decomposition and (12) that

$$\sigma^2(f) - \sigma^2(f_{T_m}) = \sigma^2(f - f_{T_m}) = \sum_{\{\mathbf{i}:\mathbf{i}\subset D \wedge \mathbf{i} \notin T_m\}} \sigma^2_{\mathbf{i}}(f) \leq \sum_{j=m+1}^{d} \sum_{\{\mathbf{i}\subset D:(\pi^*)^{-1}(j)\in\mathbf{i}\}} \sigma^2_{\mathbf{i}}(f)$$

$$= \sum_{j=m+1}^{d} \sigma^2_{\{(\pi^*)^{-1}(j)\}}(f)^{\mathrm{tot}} \leq \sum_{j=m+1}^{d} v_{(\pi^*)^{-1}(j)} \leq (1-p)\sigma^2(f).$$

It follows $\sigma^2(f_{T_m}) \geq p\sigma^2(f)$ and thus $d_t \leq m$, what was required to be proved.

□

7 Weighted uniform sampling

In this section we will discuss the method we used to approximate observables as they are defined in equation (7). Before the weighed uniform sampling (WUS) method can be applied to this expression, it is necessary to perform a transformation of the variables x_i to the d-dimensional unit cube, $[0,1]^d$. In the cases we will consider in section 8, this transformation will always be of the form

$$x_i = \sum_j A_{ij} \Phi^{-1}(z_j)\ , \tag{14}$$

with A being a positive definite matrix and Φ^{-1} the inverse of the PDF of the standard normal distribution. After the transformation equation (7) reads:

$$\langle O \rangle = \frac{\int_{[0,1]^d} O(A\Phi^{-1}(\boldsymbol{z}))W(\boldsymbol{z})dz_1 \ldots dz_d}{\int_{[0,1]^d} W(\boldsymbol{z})dz_1 \ldots dz_d} \tag{15}$$

$$W(\boldsymbol{z}) = \exp\left[-S(A\Phi^{-1}(\boldsymbol{z})) + \frac{1}{2}\sum_i (\Phi^{-1}(z_i))^2\right]\ .$$

Now, in the WUS method points $\boldsymbol{z}_j,\ 1 \leq j \leq N$, are generated from a uniform distribution in $[0,1]^d$. Using these points, a quotient of integrals of the form

$$R := \frac{\int_{[0,1]^d} f_1(\boldsymbol{z}) d\boldsymbol{z}}{\int_{[0,1]^d} f_2(\boldsymbol{z}) d\boldsymbol{z}}$$

can then be obtained by taking the estimator

$$\hat{R}_N := \frac{\sum_{j=1}^N f_1(\boldsymbol{z}_j)}{\sum_{j=1}^N f_2(\boldsymbol{z}_j)} , \tag{16}$$

where the functions f_i could be of very general, in particular non-Gaussian nature. For our example these functions can be read off from equation (15): $f_1 = O(A\Phi^{-1}(\boldsymbol{z}))W(\boldsymbol{z})$ and $f_2 = W(\boldsymbol{z})$. For the case that $W(\boldsymbol{z})$ is really a function of $\boldsymbol{z}$ (and not just a constant), this way of evaluating integrals over certain weight functions W is known as reweighting technique in field theory or statistical physics. A crucial element of the WUS (reweighting) method is that the sampling points have a large enough overlap with the weight functions f_i considered. The estimator in 16 was analyzed in [10] and applications have been investigated for example in [30] and [31]. The bias and the root mean square error (RMSE) of this estimator satisfy

$$\begin{aligned} Bias(\hat{R}_N) &= \frac{R\, var(f_2)}{N} - \frac{cov(f_1, f_2)}{N} + O(N^{-\frac{3}{2}}) \\ RMSE(\hat{R}_N) &= \frac{\sqrt{var(f_1) + R^2 var(f_2) - 2R\, cov(f_1, f_2)}}{\sqrt{N}} + O(N^{-\frac{3}{4}}) . \end{aligned}$$

The bias of the estimator is asymptotically negligible compared with the RMSE. One clear disadvantage of WUS against Mc-MC or Importance Sampling for problems with large regions of relative low values of the integrands is that with WUS we sample over the entire unit cube $[0,1]^d$ uniformly, thus the method is dependent on how we transformed the problem to the unit cube. In contrast, Mc-MC and Importance Sampling based techniques for models in high-energy or statistical physics usually focus on characteristic or important regions of the integrands aiming to sample directly from the underlying distribution of the problem, using in this way only the most relevant sample points.

8 Numerical experiments

We consider for our numerical tests the *quantum mechanical harmonic and anharmonic oscillator* in the *path integral approach* as described in section 2. For definiteness we repeat here the expression for the action of the system:

$$S(x) = \frac{a}{2} \sum_{i=1}^{d} \left(\frac{M_0}{a^2} (x_{i+1} - x_i)^2 + \mu^2 x_i^2 + 2\lambda x_i^4 \right) . \tag{17}$$

We investigate the two observable functions

$$O_1(x) = \frac{1}{d} \sum_{i=1}^{d} x_i^2 ,\ O_2(x) = \frac{1}{d} \sum_{i=1}^{d} x_i^4 ,$$

using the notation $\langle X^2\rangle$,$\langle X^4\rangle$ for $\langle O_1(x)\rangle$,$\langle O_2(x)\rangle$ in our tests. In addition, we will look at the ground state energy E_0 which, by virtue of the virial theorem, is related to O_1 and O_2 by $E_0 = \mu^2 O_1 + 3\lambda O_2 + \frac{\mu^4}{16}$.

8.1 Harmonic Oscillator

For the harmonic oscillator we can apply immediately the direct sampling approach described in sections 3 and 7 for calculating estimates of observables O by setting

$$f_1 = O(A\Phi^{-1}(\boldsymbol{z}))\ , \quad f_2 = 1$$

in (16). The matrix A is a square root of C, the covariance matrix of the variables x_i, appearing in the action if it is expressed as a bi-linear form: $S = \frac{1}{2}x^T C^{-1}x$, written explicitly

$$C^{-1}_{ij} = \frac{2M_0}{a}\left[u\delta_{ij} - \frac{1}{2}\left(\delta_{i+1\ j} + \delta_{i\ j+1}\right)\right],\ u = 1 + \frac{a^2\mu^2}{2M_0}\ . \tag{18}$$

Different factorizations, namely Cholesky and PCA (principle component analysis) have been tried out. The PCA based factorization turned out to perform better in our tests, which is the reason why we will only show results for this method. The PCA factorization can be explicitly obtained for circulant Toeplitz matrices and the matrix–vector products can be efficiently computed by means of the fast Fourier transform. Given that the covariance matrix C is circulant Toeplitz, we have that $C = G\Lambda G^T$, with $G := Re(F) + Im(F)$,

$$(F)_{kl} = \frac{1}{\sqrt{d}}e^{-\frac{2\pi i}{d}kl} \tag{19}$$

being the Fourier matrix and Λ the diagonal matrix of positive eigenvalues (Lemma 4 in [32]). Thus $A = G\Lambda^{\frac{1}{2}}$ is a factorization of C, and in this case one can follow a recipe for generating normals with randomized QMC based on the discrete Fourier transform and using fast Fourier transform (FFT) techniques as described in [32]:

1. Generate a randomized QMC point $\tilde{\boldsymbol{z}}$.
2. Compute $\tilde{\mathbf{y}} = \Phi^{-1}(\tilde{\boldsymbol{z}})$.
3. Compute $\tilde{\mathbf{w}} = (\sqrt{\beta_1}\tilde{y}_{\pi^{-1}(1)}, \ldots, \sqrt{\beta_d}\tilde{y}_{\pi^{-1}(d)})$, where

$$\beta_j = \left[\frac{2M_0}{a}\left(u - \cos(2\pi j/d)\right)\right]^{-1},\ 1 \le j \le d \tag{20}$$

   are the eigenvalues in the diagonal matrix Λ, and $\pi(.)$ is a fixed permutation of the variables.
4. Compute $\tilde{\mathbf{v}} = \mathrm{FFT}(\tilde{\mathbf{w}})$.
5. Take $\tilde{\mathbf{x}} = Re(\tilde{\mathbf{v}}) + Im(\tilde{\mathbf{v}})$ as the resulting point sample.

We reorder the sampling points in step 3 in order to be able to apply standard FFT. It is (strongly) recommended to fix first the permutation $\pi(.)$ such that $(\beta_{\pi(j)})_{j=1}^{d}$ are in non-increasing order, and this permutation was taken in our experiments. If this permutation of variables does not lead to satisfactory results, the analysis described in 6.2 can be carried out to investigate if a possible different permutation leads to more effective dimension reduction and better results.
In the ordinary Mc-MC approximation, we used the Mersenne Twister[11] pseudo random number generator. We note in passing that the Mc-MC samples were generated in exactly the same way as described above for QMC with the only difference that in step one Mc-MC points were generated according to the Gaussian measure of the harmonic oscillator. This corresponds to a heatbath algorithm, where all variables x_i are updated at the same time and a reweighting procedure in the anharmonic case. For the QMC tests, we use randomly scrambled Sobol' sequences using the technique proposed by J. Matoušek[23]. For a description of how to implement Sobol sequences, we refer to [33] and [34]. Moreover, one can find a three-pages-note with a short and simple description on the website `http://web.maths.unsw.edu.au/~fkuo/sobol/index.html`. The error of $\langle X^2 \rangle$ was obtained by scrambling 10 times the QMC sequence and making 10 runs of an Mc-MC simulation (with different seeds). This procedure is repeated 30 times in both cases to obtain the error of the error. From the results shown in figure 1, we can see a scaling of the errors for both methods that agrees perfectly with the expected behavior, namely $N^{-0.5}$ for Mc-MC and N^{-1} for QMC, for large N. Although this example is trivial, it was our first successful application of the QMC approach in a physical lattice model and motivated us to pass on to more complicated models.

8.2 Anharmonic Oscillator

The WUS (reweighting) approach was also used for the problem of the anharmonic oscillator to estimate $\langle X^4 \rangle$, $\langle X^2 \rangle$ and the ground state energy of the system (E_0). With the anharmonic term in action, the probability distribution function (PDF) of the variables x_i is of non-Gaussian nature and hence becomes very complicated. This makes it very hard to generate the samples directly from the PDF of the anharmonic oscillator. Instead of this, the anharmonic term and a fraction of the harmonic term is treated as part of the weight functions f_1 and f_2 in (16), leaving the sampling procedure of the x_i basically unchanged as compared to the harmonic oscillator. For the weight functions in 16 we have chosen

$$f_1(\boldsymbol{z}) = O(\mathbf{x}) f_2(\boldsymbol{z}) \;, \quad f_2(\boldsymbol{z}) = e^{-\sum_i \left(a \left(\frac{\mu^2 - \mu_{sim}^2}{2} \right) x_i^2 + a\lambda x_i^4 \right)} \;, \qquad (21)$$
$$\mathbf{x} = A\Phi^{-1}(\boldsymbol{z}) \;.$$

For the evaluation of $\mathbf{x}$, not the original parameter μ^2 in equation (18) was taken but a shifted parameter μ_{sim}^2 which is exactly compensated by the weight function f_2. This procedure is necessary, because of $C = A^T A$ being positive definite only if $\mu_{sim}^2 > 0$, whereas $\mu^2 < 0$ in our test cases. Thus, we have to shift the value of μ in the harmonic piece of the action to guarantee positivity. Besides the requirement of positivity, one is free in the choice of the parameter μ_{sim}. In fact, it turned out that one can take advantage of this freedom to

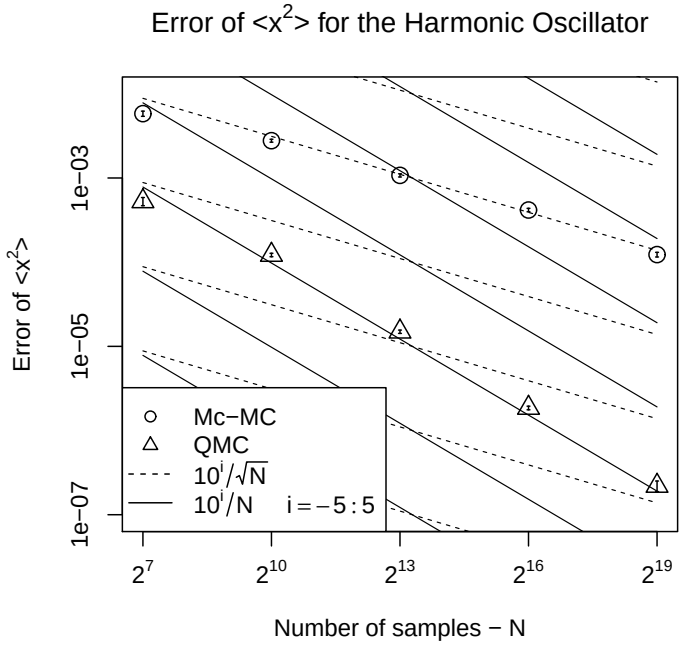

Figure 1: The error of $\langle X^2 \rangle$ in dependence on the number of samples N. The parameters here were chosen as $\lambda = 0$ (harmonic oscillator), $d = 51$, $M_0 = 0.5$ and $\mu^2 = 2.0$. The error of the error was obtained by repeating the numerical experiment 30 times, see also the text.

tune μ_{sim} to a value that reduces the fluctuations of the weights. This leads to observable averages with less variances and therefore smaller errors. Further, it is important to note that the PCA factorization during the generation of the Gaussian samples is essential for an efficient reduction of the effective dimension (see [25]) of the problem. For the parameters listed below, we estimated the effective dimensions of the functions (21) to be close to 20 (for a 99% variance concentration), for estimating the integrals described in (11) with the dimension of the original system up to $d = 1000$. Thus, we observe a drastic reduction of dimensionality.

On the other hand we found that the effective dimension depends very strongly on the parameter $T = da$, i.e. the physical time extent of the system. For small T-values, say $T < 0.2$, the effective dimension is reduced sufficiently good, such that the QMC approach leads to an $1/N$ error scaling. The situation changes for $T = 1.5$, where the effective dimension is significantly increased and, correspondingly, the error behaves like $1/N^\alpha$ with $\alpha \approx 0.75$ only, see below. Tests with values of $T \geq 5$ indicate that the simulations become more and more difficult in the sense that one needs more and more samples to achieve the same accuracy of an observable as compared to estimates at $T = 1.5$. Thus, in such situations the overlap of the sampling points with the functions f_i in 16 seem to

	O	α	$\log C$	χ^2/dof
$d = 100$	X^2	-0.763(8)	2.0(1)	7.9 / 6
	X^4	-0.758(8)	4.0(1)	13.2 / 6
	E_0	-0.737(9)	4.0(1)	8.3 / 6
$d = 1000$	X^2	-0.758(14)	2.0(2)	5.0 / 4
	X^4	-0.755(14)	4.0(2)	5.7 / 4
	E_0	-0.737(13)	4.0(2)	4.0 / 4

Table 1: Shown are the results for fit parameters of the error scaling for the observables considered, i.e. X^2 , X^4 and E_0, where the used fit function takes the form error $= CN^\alpha$, see equation 22. We also provide the χ^2 values as well as the number of degrees of freedom in the fit (dof).

be too small to reduce the fluctuations sufficiently. However, we are presently exploring a more general approach for selecting an optimal matrix C (resp. A in (14)) in the sampling procedure to improve the situation for larger values of T.

For our numerical experiments, the parameters were set to $M_0 = 0.5$, $\lambda = 1.0$, $\mu^2 = -16$. In the two tests the lattice spacing a was adjusted such, that T was kept fixed. The optimal value of μ^2_{sim} generally depends on all physical parameters of the system and in particular on a. Thus, we have to adjust also μ_{sim} when the lattice spacing a is changed. In particular, we set $a = 0.015$ and $\mu^2_{sim} = 0.176$ for $d = 100$, whereas for $d = 1000$ $a = 0.0015$ and $\mu^2_{sim} = 0.2$ was chosen. The error analysis of $\langle X^2 \rangle$ and $\langle X^4 \rangle$ was carried through in the same way as described for the harmonic oscillator test case discussed in the last subsection, 8.1. We show in figure 2 the error of $\langle X^2 \rangle$ and E_0 as a function of the number of samples. In addition, we represent by the dashed line in figure 2 a fit to the data for the computed errors using the formula

$$\log(\text{ Error } (\langle O \rangle)) = \log C + \alpha \log N \; ; \quad O = \{x^2, x^4, E_0\} \; , \tag{22}$$

with C and α left as free parameters. From this analysis we can obtain a quantitative determination of the exponent of the error scaling. The results for the fit parameters are listed in table 1.

As can be inferred from table 1 in the case of the anharmonic oscillator the error scaling exponent is not the optimal one of $\alpha = 1$ but assumes a value of $\alpha \approx 0.76$. However, this constitutes still a much improved error scaling compared to a Mc-MC simulation with a corresponding large gain in the number of required samples to reach a desired accuracy. Moreover, the value of α is consistent for all observables considered here and independent from the dimension of the problem employed, a finding which is clearly encouraging.

We finally mention that the resulting estimates of the ground state energy matches in at least two significant digits with the theoretical value, $E_0 = 3.863$, calculated in [35], namely $\hat{E}_0 = 3.857 \pm 0.004$ for $d = 100$ and $\hat{E}_0 = 3.862 \pm 0.004$ for $d = 1000$.

9 Concluding Remarks

In this article we have performed a first application of the QMC method to Euclidean lattice models. The goal was to see, whether the QMC algorithm provides also in the case of non-Gaussian systems an improved error scaling behaviour with respect to Markov-chain Monte Carlo methods. As a prototype system, we have considered the quantum mechanical oscillator discretized on a Euclidean time lattice, both in its harmonic (Gaussian) form as well as adding a non-Gaussian quartic term (anharmonic oscillator). For the harmonic oscillator we found a large-N (N being the number of sample points) error behavior as expected, i.e. for QMC ($\sim 1/N$) and for Mc-MC ($\sim 1/\sqrt{N}$).

The main result of our investigation is that also for the anharmonic oscillator, which is a non-Gaussian problem, the QMC approach leads to a significant improvement of the error scaling N^α with $\alpha \approx -0.76$, see table 1 for the exact values of α for different observables and different physical situations.

Further, we found that the choice of the lattice spacing a does not seem to have any effect on the error scaling behaviour and the accessible range of $T = 1.5$ values gives already estimates of the ground state energy, compatible (within errors) with the theoretical prediction (valid in the limit $T \to \infty$ and $a \to 0$). For the case that the improved error scaling and the mild dependence on the lattice spacing a found here will also be present in more elaborate models, the QMC has the potential to become very valuable in the future. On the other hand we observed that the applicability of the WUS (reweighting) approach seems to be limited by the physical time extent $T = da$ of the system. Stable results could only be found for values $T \leq 1.5$.

It is clear that the here considered quantum mechanical systems are rather simple models and still a long way has to be gone, if generic quantum field theories, especially gauge theories are to be studied. Nevertheless, it is very reassuring that we find an improved error scaling behaviour in the case of a quartic potential and hence a non-Gaussian system. This promising result is certainly a strong motivation for studying further QMC methods in lattice field theories and statistical mechanics.

Acknowledgment

The authors wish to express their gratitude to Alan Genz (Washington State University) and Frances Kuo (University of New South Wales, Sydney) for inspiring comments and conversations, which helped to develop the work in this article. Frances Kuo collaborated with us during her visit to the Humboldt-University Berlin in 2011. A.N., K.J. and M.M.-P. acknowledge financial support by the DFG-funded corroborative research center SFB/TR9. K. J. was supported in part by the Cyprus Research Promotion Foundation under contract ΠΡΟΣΕΛΚΥΣΗ/ΕΜΠΕΙΡΟΣ/0311/16.

References

[1] C. Gattringer and C. B. Lang. Quantum chromodynamics on the lattice. *Lect.Notes Phys.*, 788:1–211, 2010.

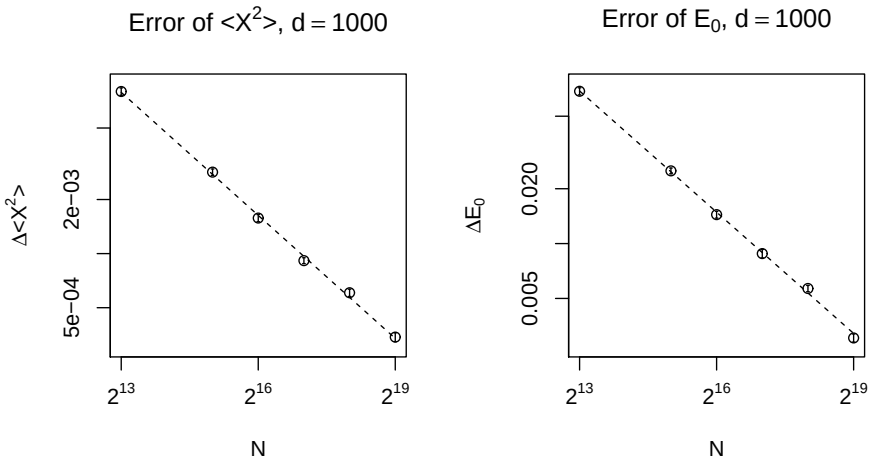

Figure 2: We show the error of the observables $\langle X^2 \rangle$ and E_0 as a function of the number of samples in a double logarithmic graph. The error of the error was obtained by performing 30 repetitions of the experiment with parameters chosen as $\lambda = 1.0$, $\mu^2 = -16$, $a = 0.0015$ and $d = 1000$. For the sample generation randomly scrambled Sobol' (RQMC) was used with $2^{13}, 2^{15}, 2^{16}, 2^{17}, 2^{18}$ and 2^{19} points. The dashed line shows the fit to the data points using a fit function $\log(\Delta\langle O \rangle) \sim \log(C) + \alpha \log(N)$. The fitted exponents are $\alpha = -0.758(14)$ for $\langle X^2 \rangle$ and $\alpha = -0.737(13)$ for E_0, see also table 1.

[2] I. Montvay and G. Münster. *Quantum fields on a lattice*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1994.

[3] Z. Fodor and C. Hölbling. Light Hadron Masses from Lattice QCD. *Rev.Mod.Phys.*, 84:449, 2012.

[4] J. Bulava, P. Gerhold, K. Jansen, J. Kallarackal, B. Knippschild, et al. Higgs-Yukawa model in chirally-invariant lattice field theory. 2012.

[5] P. L'Ecuyer and C. Lemieux. Recent advances in randomized quasi-monte carlo methods. In Moshe Dror, Pierre L'Ecuyer, and Ferenc Szidarovszky, editors, *Modeling Uncertainty*, volume 46 of *International Series in Operations Research & Management Science*, pages 419–474. Springer US, 2005.

[6] F. Kuo, Ch. Schwab, and I. Sloan. Quasi-monte carlo methods for high-dimensional integration: the standard (weighted hilbert space) setting and beyond. *ANZIAM Journal*, 53(0), 2012.

[7] M. Creutz and B. A. Freedman. A statistical approach to quantum mechanics. *Ann. Phys.*, 132:427–462, 1981.

[8] W. Bietenholz, U. Gerber, M. Pepe, and U.-J. Wiese. Topological Lattice Actions. *JHEP*, 1012:020, 2010.

[9] K. Jansen, H. Leovey, A. Nube, A. Griewank, and M. Mueller-Preussker. A first look at quasi-Monte Carlo for lattice field theory problems. 2012.

[10] M. J. D. Powell and J. Swann. Weighted uniform sampling – a monte carlo technique for reducing variance. *J.Inst.Maths Applics*, 2:228–236, 1966.

[11] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, January 1998.

[12] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Volume II: Standard information for functionals*, volume 12 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2010.

[13] J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, New York, NY, USA, 2010.

[14] F. J. Hickernell. A generalized discrepancy and quadrature error bound. *Math. Comp*, 67:299–322, 1998.

[15] I. H. Sloan and H. Woźniakowski. When are quasi-monte carlo algorithms efficient for high dimensional integrals? *J. Complexity*, 14:1–33, 1997.

[16] F. Y. Kuo. Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces. *J. Complexity*, 19(3):301–320, 2003. Numerical integration and its complexity (Oberwolfach, 2001).

[17] X. Wang. Strong tractability of multivariate integration using quasi-monte carlo algorithms. *Math. Comp.*, 72(242):823–838, April 2003.

[18] A. Griewank, L. Lehmann, H. Leovey, and M. Zilberman. Automatic evaluations of cross-derivatives. *to appear in Math. Comp.*, 2012.

[19] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[20] A. B. Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In H. Niederreiter and P. J.-S. Shiue, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106 of *Lecture Notes in Statistics*, pages 299–317. Springer-Verlag, 1995.

[21] A. B. Owen. Monte carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910, October 1997.

[22] A. B. Owen. Local antithetic sampling with scrambled nets. *Ann. Statist.*, 36:2319–2343, 2008.

[23] J. Matoušek. On the L_2-discrepancy for anchored boxes. *Journal of Complexity*, 14:527–556, 1998.

[24] S. Tezuka and H. Faure. I-binomial scrambling of digital nets and sequences. *Journal of Complexity*, 19(6):744 – 757, 2003.

[25] R. E. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997.

[26] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulation*, 55(1-3):271–280, 2001. The Second IMACS Seminar on Monte Carlo Methods (Varna, 1999).

[27] X. Wang and K.-T. Fang. The effective dimension and quasi-Monte Carlo integration. *J. Complexity*, 19(2):101–124, 2003.

[28] I.M. Sobol' and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math. Comput. Simul.*, 79(10):3009–3017, 2009.

[29] A. Griewank and A. Walther. *Evaluating derivatives*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2008. Principles and techniques of algorithmic differentiation.

[30] J. Spanier and E. H. Maize. Quasi-random methods for estimating integrals using relatively small samples. *SIAM Rev.*, 36(1):18–44, March 1994.

[31] R. E. Caflisch and B. Moskowitz. Modified monte carlo methods using quasi-random sequences. In *Lecture Notes in Statistics 106*, pages 1–16. Springer-Verlag, 1995.

[32] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. *Journal of Computational Physics*, 230(10):3668 – 3694, 2011.

[33] S. Joe and F. Y. Kuo. Remark on algorithm 659: Implementing sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.*, 29(1):49–57, March 2003.

[34] S. Joe and F. Y. Kuo. Constructing sobol sequences with better two-dimensional projections. *SIAM J. Sci. Comput.*, 30(5):2635–2654, August 2008.

[35] R. Blankenbecler, T. A. DeGrand, and R. L. Sugar. Moment Method for Eigenvalues and Expectation Values. *Phys.Rev.*, D21:1055, 1980.