# On Tractability of Path Integration

Columbia University Computer Science Department Report CUCS-022-96

Grzegorz W. Wasilkowski and Henryk Woźniakowski *

September 1995

**Abstract**

Many applications require approximate values of path integrals. A typical approach is to approximate the path integral by a high dimensional integral and apply a Monte Carlo (randomized) algorithm. However, Monte Carlo algorithm requires roughly $\varepsilon^{-2}$ integrand evaluations to provide an $\varepsilon$-approximation. Moreover, the error bound of $\varepsilon$ is guaranteed only in a stochastic sense.

Do we really need to use randomized algorithms for path integrals? Perhaps, we can find a *deterministic* algorithm that is more effective even in the worst case setting. To answer this question, we study the *worst case complexity* of path integration, which, roughly speaking, is defined as the minimal number of the integrand evaluations needed to compute an approximation with error at most $\varepsilon$. We consider path integration with respect to a Gaussian measure, and for various classes of integrands.

Tractability of path integration means that the complexity depends polynomially on $1/\varepsilon$. We show that for the class of $r$ times Frechet differentiable integrands, tractability of path integration holds iff the covariance operator of the Gaussian measure has finite rank. Hence, if the Gaussian measure is supported on an infinite dimensional space then path integration is intractable. In this case, there exists no effective deterministic algorithm, and the use of randomized algorithms is justified. In fact, for this class of integrands, the classical Monte Carlo algorithm is (almost) optimal and the complexity in the randomized setting is proportional to $\varepsilon^{-2}$.

On the other hand, for a particular class of entire integrands, the worst case complexity of path integration is at most of order $\varepsilon^{-p}$ with $p$ depending on the Gaussian

*The first author was partially supported by the the National Science Foundation under Grant CCR-9420543, and the second by the National Science Foundation and the Air Force Office of Scientific Research. This work was partly done while the second author was visiting the Australian National University.

measure. Hence, path integration is now tractable. Furthermore, for any Gaussian measure, the exponent $p$ is less than or equal to 2. For the Wiener measure, $p = 2/3$. For this class, we provide effective deterministic algorithms which solve the path integration problem with (worst case) cost that is usually much less than the (randomized) cost of the classical Monte Carlo algorithm.

# I. Introduction

Approximate computation of integrals is undoubtedly one of the most important problems of computational mathematics. In many cases, integrals involve functions of finitely many variables $d$. Not surprisingly, the univariate case $d = 1$ is best understood and has a rich and well-developed theory. Elements of the classical theory of univariate integration can be found in almost all numerical analysis textbooks. The study of the complexity of continuous problems has started from the pioneering work of Sard and Nikolskij on univariate integration, see [1, 2].

The multivariate case, with $d$ finite and greater than one, is much harder and is a subject of very active research. For large $d$, a typical approach is to use Monte Carlo (randomized) algorithms. However, for some classes of integrands, deterministic algorithms can be also very effective. An example is provided by integrands with bounded mixed derivatives, for which algorithms based on low discrepancy points can be used. The state of the art can be found in [3]. Complexity of multivariate integration in various settings is also an active research area. An account of recent progress can be found in [4, 5, 6, 7, 8, 9, 10, 11].

In this paper we consider the case $d = +\infty$. That is, we deal with integrals of functions of infinitely many variables. This is usually called the *path integration* problem. The name is derived from the most typical case, in which we integrate over continuous functions (paths) with respect to the Wiener measure, see [12]. Sometimes, instead of path integration, the name *functional* integration is used, see [13, 14]. The latter stresses that we integrate over a class of functions.

One may suspect that the path integration problem is merely of theoretical interest. However, the opposite is true. Path integrals occur in many applied fields, including quantum physics and chemistry, differential equations, and financial mathematics, as well as average case complexity. Here are a few examples. In the forties, R. P. Feynman introduced path integration in quantum physics, see [12]. The work of Feynman initiated a very fruitful stream of research in quantum physics and chemistry which continues to be active, see e.g., [13, 15, 16, 17, 18, 19, 20, 21, 22, 23]. A rigorous mathematical foundation for Feynman path integration can be found in [24]. In the fifties, M. Kac observed that the approach of Feynman can be used for the solution of parabolic differential equations, and established what today is called the Feynman-Kac formula, see [25]. In fact, solutions of many differential and operator equations can be expressed as path integrals, see [14]. Also many problems in financial mathematics are expressed as generalized Feynman-Kac formulas, and hence their solution may be reduced to computing path integrals, see [26, 27, 28]. Finally, in average case complexity, we need to estimate the average error of an algorithm; this error is, once more, given as a path integral. A more complete list of applications is given in the introduction of [14].

A typical approach to computing path integrals is to switch to a multivariate integral and apply a Monte Carlo (randomized) algorithm. That is, the infinite dimensional integral is approximated by a $d$ dimensional integral, where $d$ may be large (or even huge). Then the classical Monte Carlo algorithm can be used since its speed of convergence, although not great, does not depend on $d$. This approach usually requires on the order of $\varepsilon^{-2}$ integrand evaluations to obtain the expected error at most $\varepsilon$, see Section 2 for more details.

Due to this relatively high cost of Monte Carlo and only stochastic error assurance, one would like to know weather there is an effective *deterministic* algorithm which approximates path integrals with a small (deterministic) error. Obviously, the existence of such a deterministic algorithm depends on the probability measure $\mu$ occurring in the path integral as well as on the class $F$ of integrands. Hence, for a given measure $\mu$ and a given class $F$, we wish to find the *worst case complexity* of path integration. Roughly speaking, the worst case complexity is proportional to the minimal number of integrand evaluations needed to compute an approximation with worst case error at most $\varepsilon$.

We are mainly interested in how the complexity depends on $\varepsilon$. If the complexity is of order $\varepsilon^{-p}$ with $p < 2$ then we beat the bound $\varepsilon^{-2}$ of the classical Monte Carlo algorithm. For $p = 2$, the bounds are of the same order. However, even for $p > 2$, we may prefer to use a deterministic algorithm since its error is guaranteed to be at most $\varepsilon$, whereas for the classical Monte Carlo algorithm we only know that its expected error is at most $\varepsilon$. Moreover, with deterministic algorithms, we do not have to cope with the problem of generating random numbers or functions.

This discussion motivates the concept of tractability of path integration. Namely, we say that the path integration problem is *tractable* if the worst case complexity depends polynomially on $1/\varepsilon$. In other settings such as the randomized or average case settings, tractability of path integration is obvious in classes for which the $L_2$ norms of integrands are uniformly bounded. However, for problems that are not path integration, the study of tractability in other settings is an interesting subject, see [29, 30].

Tractability of path integration depends on the probability measure $\mu$ and the class $F$ of integrands. We now comment on the assumptions regarding $\mu$ and $F$.

We begin with the measure $\mu$. In most applications of path integration, the classical Wiener measure (Brownian motion) is used. The Wiener measure is an example of a Gaussian measure which is appropriate for many applications. Its role can be hardly overestimated. It would be tempting to study path integration only for the Wiener measure. However, we prefer to be more general and to study path integration for arbitrary Gaussian measures, and to illustrate the results for the Wiener measure as a primary example. In this way we will better understand the influence of the Gaussian measure on the complexity of path integration. Of course, it would also be interesting to study path integration for a non-Gaussian measure, although no such application is known to us.

We now turn to the class $F$ of integrands. Here, the situation is far more complex since there is no class of integrands which plays a dominant role corresponding to the Wiener measure. Even for the multivariate case, there is no class which is singled out. On contrary, many different classes seem to be relevant and their choice depends on the particular application. Usually these classes are characterized by some global smoothness properties of the integrands.

For path integration, we follow the multivariate approach and we analyze classes defined by global smoothness. First we consider the class of integrands that are $r$ times continuously Frechet differentiable. We prove that in this case, tractability of path integration holds iff the covariance operator of the Gaussian measure has finite rank. Hence, the problem of path integration is *intractable* if the Gaussian measure is supported on an infinite dimensional space. Then it is reasonable to switch to the randomized setting. It turns out that the classical Monte Carlo algorithm is (almost) optimal and the complexity in the randomized setting is proportional to $\varepsilon^{-2}$.

Next we analyze a specific class of entire functions. For this class, the path integration problem is tractable, and the worst case complexity is of order $\varepsilon^{-p}$ with $p$ depending, in particular, on the Gaussian measure used. For the Wiener measure, we have $p = 2/3$, which means that we need substantially fewer integrand evaluations than for the classical Monte Carlo algorithm even though we guarantee that the worst case error is at most $\varepsilon$. We stress, however, that to get this result we assume that integrands are entire functions and that we can use derivatives as permissible information. It is well known that the classical Monte Carlo algorithm requires no smoothness of the integrands; it is enough to assume that they are square integrable.

The classes of integrands studied in this paper are characterized by global smoothness properties. In a forthcoming paper, see [31], we consider a different class of integrands. This class is related to the Feynman-Kac formula. More precisely, this is the class of potential and initial conditions functions which define the heat equation. Although these functions do not need to be very smooth, we prove tractability of path integration, and in many cases, the worst case complexity is substantially smaller than $\varepsilon^{-2}$.

# II. Formulation of the problem

In this section we formulate the path integration problem and explain a typical computational approach to approximating path integrals. We also define the worst case complexity and tractability of path integration.

Let $X$ be a separable Banach space. The norm in $X$ is denoted by $\|\cdot\|_X$. An example of $X$ is provided by the space $X = C([0,1])$ of continuous scalar functions defined on $[0,1]$ with the sup norm, $\|x\|_X = \sup_{t\in[0,1]} |x(t)|$.

We assume that $X$ is equipped with a zero mean Gaussian measure $\mu$, see, e.g., [32]. An example of $\mu$ is provided by the Wiener measure $\mu = w$ for which $X = C([0,1])$ and

$$\int_{C[0,1]} x(t_1)x(t_2)\, w(dx) \;=\; \min\{t_1, t_2\}.$$

It is known that $X$ can be embedded in the Hilbert space $L_2([0,1])$. It is an obvious corollary of the Banach (sometimes called the Banach-Mazur or Banach-Alaoglu) theorem which states that $X$ is isometrically isomorphic to a subspace of $C([0,1])$ which, in turn, can be treated as a subspace of $L_2([0,1])$. This means that there exists a one-to-one linear continuous mapping $\mathrm{Im} : X \to L_2([0,1])$. We denote the inner product of $L_2([0,1])$ by $\langle\cdot,\cdot\rangle$. Then the measure $\nu = \mu\,\mathrm{Im}^{-1}$ is also a zero mean Gaussian measure on $L_2([0,1])$. Let $C_\nu : L_2([0,1]) \to L_2([0,1])$ be the covariance operator of $\nu$. The operator $C_\nu$ is self adjoint, nonnegative definite and has finite trace. That is, there exists an orthonormal system $\{\eta_i\}_i$ of $L_2([0,1])$, $\langle\eta_i,\eta_j\rangle = \delta_{i,j}$, for which

$$C_\nu\,\eta_i \;=\; \lambda_i\,\eta_i, \quad \lambda_1 \,\ge\, \lambda_2 \,\ge\, \cdots \,\ge\, 0 \quad \text{and} \quad \sum_{i=1}^{\infty} \lambda_i \,<\, +\infty. \tag{1}$$

Observe that if all $\lambda_i = 0$ then the path integration problem becomes trivial. Indeed, $\mu$ is then an atomic measure at zero and $S(f) = f(0)$. This, of course, can be solved exactly by using one function value. To omit this trivial case, we assume that at least one eigenvalue is positive,

$$\lambda_1 \,>\, 0.$$

Without loss of generality, we may assume that $\eta_i \in \mathrm{Im}(X)$. Indeed, the measure $\nu$ is concentrated on $\mathrm{Im}(X)$, $\nu(\mathrm{Im}(X)) = 1$, and $\eta_i \in \overline{\mathrm{Im}(X)}$, where the closure of $\mathrm{Im}(X)$ is taken in the norm of $L_2([0,1])$. Hence, we can approximate $\eta_i$ with an arbitrarily small error by elements of $\mathrm{Im}(X)$. To avoid this cumbersome approximation of $\eta_i$, we assume, for simplicity, that $\eta_i$ belong to $\mathrm{Im}(X)$.

For the Wiener measure $\mu = w$, we have

$$\mathrm{Im}(x) = x, \qquad \eta_i = \sqrt{2}\,\sin\left(\frac{2i-1}{2}\pi x\right), \qquad \lambda_i = \frac{4}{\pi^2(2i-1)^2}.$$

Let $F$ be a class of (Borel) measurable real functions defined on $X$. An example of such $F$ studied in this paper is the class $F = F^r$ of $r$ times Frechet differentiable functions for which $\|f^{(i)}\| = \sup_{x\in X}\|f^{(i)}(x)\| \leq 1$ for $i = 0, 1, \ldots, r$. Here, $f^{(i)}(x)$ is an $i$-linear form from $X^i$ to $\mathbb{R}$, and its norm is defined as $\|f^{(i)}(x)\| = \sup_{\|x_j\|_X \leq 1} |f^{(i)}(x)x_1x_2\cdots x_i|$.

The *path integration* problem is defined as approximating integrals of $f$ from $F$. That is, we want to approximate the expectation of $f$ with respect to the Gaussian measure $\mu$,

$$S(f) = \int_X f(x)\,\mu(dx), \quad \forall\, f \in F. \tag{2}$$

Since $X$ is usually infinite dimensional, the integrand $f$ in (2) depends on infinitely many variables. That is why the path integration problem can be viewed as an integration of functions of infinitely many variables.

We now illustrate the path integrals problem for finite and infinite dimensional spaces $X$. Assume first that $X = \mathbb{R}^d$ for some finite $d$, and let $\mu$ be the standard Gaussian measure with $\lambda_i$ as the eigenvalues of its covariance operator. Then (2) becomes

$$S(f) = \frac{1}{(2\pi)^{d/2}}\frac{1}{\sqrt{\lambda_1\lambda_2\cdots\lambda_d}}\int_{\mathbb{R}^d} f(t_1,t_2,\ldots,t_d)\,\exp\left(-t_1^2/(2\lambda_1) - \cdots - t_d^2/(2\lambda_d)\right)\,d\vec{t},$$

where $\vec{t} = [t_1, t_2, \ldots, t_d] \in \mathbb{R}^d$.

Hence, for finite dimensional spaces $X$, the path integration problem reduces to finite dimensional integration with respect to a Gaussian measure.

Assume now that $X$ is of infinite dimension. We now show how $S(f)$ can be approximated by finite dimensional integrals. Let $x = \mathrm{Im}^{-1}(\mathrm{Im}\,x)$. Note that $\mathrm{Im}^{-1}$ is well defined on the set $\mathrm{Im}(X)$ which is of a full $\nu$-measure. Hence, $\mathrm{Im}^{-1}(\mathrm{Im}\,x)$ is defined almost everywhere. By changing variables $y = \mathrm{Im}(x)$ we may rewrite (2) as

$$S(f) = \int_{L_2([0,1])} f\,\left(\mathrm{Im}^{-1}(y)\right)\,\nu(dy).$$

The elements $y$ from $L_2([0,1])$ can be approximated by

$$P_d y = \sum_{i=1}^{d} \langle y, \eta_i\rangle\,\eta_i.$$

Then the integral $S(f)$ is approximated by $S_d(f_d)$, where $f_d : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$f_d(\vec{t}) = f\left(\mathrm{Im}^{-1}\left(t_1\eta_1 + t_2\eta_2 + \cdots + t_d\eta_d\right)\right),$$

for $\vec{t} = [t_1, t_2, \ldots, t_d] \in \mathbb{R}^d$, and

$$S_d(f_d) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\lambda_1\lambda_2\cdots\lambda_d}} \int_{\mathbb{R}^d} f_d(\vec{t}) \exp\left(-t_1^2/(2\lambda_1) - \cdots - t_d^2/(2\lambda_d)\right) d\vec{t}. \qquad (3)$$

Observe that $S_d$ is a finite dimensional integral, as for the case of a finite dimensional space $X$. However, unlike the latter case, the eigenvalues $\lambda_i$ tend to zero, and $\lambda_i \le a/i$ with $a = \sum_{i=1}^{\infty} \lambda_i < +\infty$. Hence, there is a decreasing dependence on the successive variables $t_i$ in (3).

For a function $f$ that satisfies the Banach-Lebesgue theorem, (it is enough to assume that $f$ is continuous and $|f(\mathrm{Im}^{-1}(P_d y))| \le g(\mathrm{Im}^{-1}(y))$, $\forall\, d$, for some function $g$ for which $S(g)$ is finite), we have

$$S(f) = \lim_{d\to\infty} S_d(f_d).$$

This suggests that to approximate $S(f)$ it is enough to choose a sufficiently large $d$ and approximate a finite dimensional integral $S_d(f_d)$. The choice of $d$ depends on the smoothness of the elements of $F$. For example, assume that $F = F_{\mathrm{Lip}}$ is the class of Lipschitz functions,

$$F_{\mathrm{Lip}} = \{f : X \to \mathbb{R},\colon\ |f(x_1) - f(x_2)| \le K\|\mathrm{Im}(x_1) - \mathrm{Im}(x_2)\|_{L_2([0,1])},\ \forall\, x_1, x_2 \in X\},$$

for some positive constant $K$. Then for $f \in F_{\mathrm{Lip}}$ we have

$$|S(f) - S_d(f_d)| \le K\left(\int_{L_2([0,1])} \sum_{i=d+1}^{\infty} \langle y, \eta_i\rangle^2\, \nu(dy)\right)^{1/2} = K\left(\sum_{i=d+1}^{\infty} \lambda_i\right)^{1/2}.$$

Hence, to guarantee that the error $|S(f) - S_d(f_d)| \le \varepsilon$, $\forall f \in F_{\mathrm{Lip}}$, it is enough to define $d$ as the smallest integer for which

$$\sum_{i=d+1}^{\infty} \lambda_i \le \varepsilon^2/K^2.$$

For $\lambda_i = \Theta(i^{-k})$ with $k > 1$, we get

$$d = \Theta\left((K/\varepsilon)^{2/(k-1)}\right) \quad \text{as } \varepsilon \to 0^+.$$

For the Wiener measure we have $k = 2$ and

$$d \,=\, \frac{1}{\pi^2}\left(\frac{K}{\varepsilon}\right)^2 (1 \,+\, o(1)) \quad \text{as } \varepsilon \to 0^+.$$

Hence, for infinite dimensional spaces $X$, we can approximate path integration by $d$ dimensional integrals with respect to a Gaussian measure, where $d = d(\varepsilon)$ goes to infinity as the error tolerance $\varepsilon$ goes to zero. How fast $d(\varepsilon)$ goes to infinity depends on the decay of the eigenvalues of $\lambda_i$.

In either case of $X$, we see that path integrals may be approximated by $d$ dimensional integrals, where $d$ is typically (very) large. For $f \in L_2(X,\mu)$, the high dimensional integration is usually done by the classical Monte Carlo algorithm applied to the function $f_d$,

$$S_d(f_d) \,\sim\, \mathrm{MC}_n(f_d;\vec{u}\,) \,=\, \frac{1}{n}\sum_{i=1}^{n} f_d(u_i),$$

where $\vec{u} \,=\, [u_1, u_2, \ldots, u_n] \,\in\, \mathbb{R}^{n\,d}$ and $u_i$ are independent random points of $\mathbb{R}^d$ which are distributed according to the Gaussian measure of zero mean and variances $\lambda_1, \lambda_2, \ldots, \lambda_d$. It is well known that

$$\mathrm{E}\,(S_d(f_d) - \mathrm{MC}_n(f_d;\vec{u}\,))^2 \,=\, \frac{1}{n}\left(\int_{L_2([0,1])} h^2(P_d y)\,\nu(dy) \,-\, \left(\int_{L_2([0,1])} h(P_d y)\,\nu(dy)\right)^2\right),$$

where $h \,=\, f\circ \mathrm{Im}^{-1}$ and E stands for the expectation with respect to the random selection of the points $u_i$.

Note that

$$\mathrm{E}\,(S(f) - \mathrm{MC}_n(f_d;\vec{u}\,))^2 \,=\, |S(f) - S_d(f_d)|^2 \,+\, \mathrm{E}\,(S_d(f_d) - \mathrm{MC}_n(f_d;\vec{u}\,))^2$$

and

$$\begin{aligned}\int_{L_2([0,1])} h^2(P_d y)\,\nu(dy) \;&=\; \int_{L_2([0,1])} (h(P_d y) - h(y) + h(y))^2\,\nu(dy)\\ &\le\; 2\int_{L_2([0,1])} (h(y) - h(P_d y))^2\,\nu(dy) \,+\, 2\int_{L_2([0,1])} h^2(y)\,\nu(dy).\end{aligned}$$

Obviously, $\int_{L_2([0,1])} h^2(y)\,\nu(dy) = \int_X f^2(x)\,\mu(dx)$. Hence, for the class $F_{\mathrm{Lip}}$ of Lipschitz functions we have

$$\mathrm{E}\,(S(f) - \mathrm{MC}_n(f_d;\vec{u}\,))^2 \,\le\, K^2\,\left(1 + 2n^{-1}\right)\sum_{i=d+1}^{\infty}\lambda_i \,+\, \frac{2}{n}\int_X f^2(x)\,\mu(dx).$$

To guarantee that the randomized error of the classical Monte Carlo algorithm is at most $\varepsilon$, we choose $n$ of order $\varepsilon^{-2}$ and $d$ such that $\sum_{i=d+1}^{\infty} \lambda_i$ is of order $(\varepsilon/K)^2$. For $\lambda_i = \Theta(i^{-k})$, the cost of the classical Monte Carlo algorithm with randomized error at most $\varepsilon$ is proportional to the cost of computing $\varepsilon^{-2}$ values of functions of $d = \Theta((K/\varepsilon)^{2/(k-1)})$ variables.

The goal of this paper is to investigate whether path integration can be solved by *deterministic* algorithms in the *worst case* setting. More precisely, we are interested in the *worst case complexity* $\mathrm{comp}(\varepsilon, F)$ of path integration. This is defined as the minimal cost among all deterministic algorithms which compute an approximation whose error is at most $\varepsilon$ for all $f \in F$. In what follows, we assume that the cost of one integrand evaluation is $\mathbf{c}$, and the cost of one arithmetic operation or comparison of real numbers is unity. Of course, $\mathbf{c} \geq 1$, and in many cases $\mathbf{c}$ is much larger than unity. The precise definition of $\mathrm{comp}(\varepsilon, F)$ can be found, e.g., in [8]. Here we only mention that in our case $\mathrm{comp}(\varepsilon, F)$ can be (roughly) defined as the minimal number of integrand evaluations needed to compute an approximation whose error is at most $\varepsilon$ for all $f \in F$.

It is usually difficult to find $\mathrm{comp}(\varepsilon, F)$. That is why we settle for some characteristics of $\mathrm{comp}(\varepsilon, F)$. We say that the path integration problem is *tractable* in the worst case setting iff there exist two nonnegative numbers $K$ and $p$ such that

$$\mathrm{comp}(\varepsilon, F) \leq K \, \mathbf{c} \, \varepsilon^{-p}, \qquad \forall \, \varepsilon \in (0,1). \tag{4}$$

The smallest (or rather infimum of) $p$ for which (4) holds is called the *exponent* of the path integration problem,

$$p(F) = \inf \left\{ p \, : \, \limsup_{\varepsilon \to 0+} \varepsilon^p \, \mathrm{comp}(\varepsilon, F) < +\infty \right\}.$$

# III. Finite regularity

In this section we study tractability of path integration for the class $F = F^r$ of $r$ times continuously Frechet differentiable functions where $r$ is a nonnegative integer,

$$F^r = \left\{ f : X \to \mathbb{R} \, : \, f^{(r)} \text{ is continuous and } \|f^{(k)}(x)\| \leq 1, \; \forall \, x \in X, \; k = 0, 1, \dots, r \right\}.$$

As we shall see, tractability of path integration depends on the eigenvalues $\lambda_i$ of the correlation operator $C_\nu$ in (1).

**Theorem 1** **(i)** *If $r = 0$ or all the eigenvalues $\lambda_i$ are positive, i.e., $\lambda_i > 0, \forall \, i \geq 1$, then the path integration problem is intractable.*

**(ii)** *If $r \geq 1$ and only $k$ eigenvalues $\lambda_i$ are positive, i.e., $\lambda_k > 0$ and $\lambda_{k+1} = 0$, then the path integration problem is tractable with exponent $k/r$, i.e.,*

$$\mathrm{comp}(\varepsilon, F^r) = \Theta\left(\mathbf{c}\,\varepsilon^{-k/r}\right).$$

The assumption that all eigenvalues $\lambda_i$ are positive is natural since, otherwise, the measure $\mu$ is concentrated on a finite dimensional subspace of $X$ which contradicts the essence of the path integration problem. Hence, Theorem 1 provides a negative result about tractability of path integration. It indicates that the class $F^r$ of finite smoothness is too large to permit tractability of path integration in the worst case setting. To get tractability in the worst case setting, we need to shrink the class $F^r$. This can be done in different ways. One of them is to consider a class of entire functions, i.e., functions with infinite smoothness $r = +\infty$, and this is the subject of Section 4. Another one will be reported in a forthcoming paper [31].

For completeness, we also consider the case where only $k$ eigenvalues are positive. Then, as we shall see, the path integration problem becomes a $k$ dimensional weighted integration problem and is tractable with exponent $k/r$. Note, however, that if $k$ is large relative to $r$ then the exponent is large.

**Remark 1** Intractability of path integration in the worst case setting can be broken by switching to the randomized setting. Indeed, for the class $F^r$, the Monte Carlo algorithm applied to $S_d(f_d)$, as discussed in Introduction, yields an approximation whose expected error is at most $\varepsilon$ and cost equals $(\mathbf{c}+1)\,\varepsilon^{-2}$. The Monte Carlo algorithm is almost optimal. Indeed, it can be proven that the complexity of the path integration problem for the class $F^r$ in the randomized setting is

$$\mathrm{comp}^{\mathrm{ran}}(\varepsilon, F^r) = \Theta(\mathbf{c}\,\varepsilon^{-2}), \quad \forall\, r \geq 0,$$

assuming that the eigenvalues $\lambda_i$ of (1) do not go to zero too fast, i.e., $\lambda_i = \Omega(i^{-k})$ for some $k > 1$.

Without any assumption on the eigenvalues $\lambda_i$, one can prove that the complexity $\mathrm{comp}^{\mathrm{ran}}(\varepsilon, F^r)$ goes to infinity faster than $\varepsilon^{-2+\delta}$ for any positive $\delta$.

For the definition of the randomized setting see, e.g., [8]. The proof of the lower bound on the complexity uses results of [33] and [4] for the finite dimensional case, and the proof of Theorem 1.

*Proof of Theorem 1:* We first prove (i). Suppose on the contrary that we have tractability, i.e.,

$$\mathrm{comp}(\varepsilon, F^r) \leq K\,\mathbf{c}\,\varepsilon^{-p} \tag{5}$$

for some nonnegative $K$ and $p$.

For $r = 0$ let $d = 1$, and for $r \geq 1$ let $d$ be an integer such that $d > r\,p$. For $D = [0,1]^d$ define the class $C^{r,d}(D)$ of functions $g : \mathbb{R}^d \to \mathbb{R}$ which are $r$ times continuously differentiable, whose support is contained in the set $D$, and for which $\|g^{(i)}(\vec{t}\,)\| \leq 1$ for all $\vec{t} \in \mathbb{R}^d$ and $i = 0, 1, \ldots, r$. Here, the norm $\|g^{(i)}(\vec{t}\,)\|$ is defined as in Section 2 with the 2-norm of $\vec{t}$, i.e., $\|\vec{t}\,\|^2 = \sum_{j=1}^d t_j^2$.

For $g \in C^{r,d}(D)$, $Ry = [\langle y, \eta_1\rangle, \langle y, \eta_2\rangle, \ldots, \langle y, \eta_d\rangle]$ for $y \in L_2([0,1])$, and $P = R \circ \mathrm{Im} : X \to \mathbb{R}^d$ define

$$f(x) = g\,(P\,x), \quad \forall\, x \in X.$$

It is easy to check that

$$f^{(i)}(x)\, x_1 x_2 \cdots x_i = g^{(i)}(Px)\,(Px_1)^T (Px_2)^T \cdots (Px_i)^T, \quad \forall\, x_1, x_2, \ldots x_i \in X.$$

This yields

$$|f^{(i)}(x)\, x_1 \cdots x_i| \leq \|Px_1\| \cdots \|Px_i\| \leq \|\mathrm{Im}(x_1)\| \cdots \|\mathrm{Im}(x_i)\| \leq \beta^i\, \|x_1\|_X \cdots \|x_i\|_X,$$

where $\beta = \|\mathrm{Im}\|$. Hence

$$\|f^{(i)}(x)\| \leq \beta^r, \quad \forall\, x \in X,\ \forall\, i \leq r.$$

This means that $\beta^{-r} f$ belongs to $F^r$. Since $\eta_i \in \mathrm{Im}(X)$, we have $P(X) = \mathbb{R}^d$ and it is easy to check that

$$S(f) = \int_D g(\vec{t}\,)\, \rho_d(\vec{t}\,)\, d\,\vec{t}, \tag{6}$$

where the weight $\rho_d$ is given by

$$\rho_d(\vec{t}\,) = \frac{1}{(2\pi)^{d/2}\sqrt{\lambda_1 \lambda_2 \cdots \lambda_d}} \exp\left(-t_1^2/(2\lambda_1) - t_2^2/(2\lambda_2) - \cdots - t_d^2/(2\lambda_d)\right).$$

Observe that $\rho_d$ is well defined since the $\lambda_i$ are positive for $1 \leq i \leq d$. Indeed, if $r = 0$ then $d = 1$ and $\lambda_1 > 0$, and if $r \geq 1$ then all eigenvalues $\lambda_i$ are positive.

The essence of (6) is that the $(\varepsilon/\beta^r)$-complexity of path integration cannot be smaller than the $\varepsilon$-complexity of $d$ dimensional weighted integration in the class $C^{r,d}(D)$. Since (5) holds, this implies that the latter complexity is also $O(\mathbf{c}\,\varepsilon^{-p})$. We now show that this is not true.

Let $m_n = m_n(C^{r,d}(D))$ denote the minimal error of algorithms using $n$ function values for the weighted integration problem (6) in the class $C^{r,d}(D)$. It is known, see e.g., [8] p. 58, that

$$m_n = \inf_{\vec{t_1}, \vec{t_2}, \ldots, \vec{t_n} \in D} \ \sup_{g \in C^{r,d}(D),\, g(\vec{t_i}) = 0} \int_D g(\vec{t}\,)\, \rho_d(\vec{t}\,)\, d\,\vec{t}.$$

Clearly, the above supremum will not increase if we additionally constrain $g$ by adding $g \geq 0$. There exists a positive number $\alpha = \alpha(d, \lambda_i)$ such that $\rho_d(\vec{t}\,) \geq \alpha > 0$, $\forall \vec{t} \in [0,1]^d$. Hence

$$m_n \geq \alpha \inf_{\vec{t}_1, \vec{t}_2, \ldots, \vec{t}_n \in D} \; \sup_{g \in C^{r,d}(D),\, g \geq 0,\, g(\vec{t}_i)=0} \int_D g(\vec{t}\,)\, d\vec{t}.$$

The right-hand side is known to be $\Theta(n^{-r/d})$, see [4], Hence, $m_n = \Omega(n^{-r/d})$.

If $r = 0$, then $m_n$ is bounded uniformly from below in $n$ by a positive number. This means that for small $\varepsilon$, the complexity is infinite, and we have intractability of path integration.

If $r \geq 1$, then to guarantee $m_n \leq \varepsilon$ we have to take $n = \Omega(\varepsilon^{-d/r})$. This also means that the complexity is $\Omega(\mathbf{c}\, \varepsilon^{-d/r})$. Since $d/r > p$ this is a contradiction, which completes the proof of (i).

We now prove (ii). We will be using the notation and results from the proof of part (i). Since we have only $k$ positive eigenvalues,

$$S(f) = \int_{\mathbb{R}^k} g(\vec{t}\,)\, \rho_k(\vec{t}\,)\, d\vec{t},$$

where

$$g(\vec{t}\,) = f\left(\mathrm{Im}^{-1}\left(t_1\eta_1 + t_2\eta_2 + \cdots + t_k\eta_k\right)\right).$$

As in the proof of (i), we conclude that there exists a positive number $\gamma = \gamma(k)$ such that $\gamma\, g \in C^{r,k}(\mathbb{R}^k)$.

From (i) we also conclude that

$$\mathrm{comp}(\varepsilon, F^r) = \Omega(\mathbf{c}\, \varepsilon^{-k/r}).$$

Thus, we need a matching upper bound. First we change variables $u_i = t_i/\sqrt{\lambda_i}$ to get

$$S(f) = \int_{\mathbb{R}^k} h(\vec{u}\,)\, w_k(\vec{u}\,)\, d\vec{u},$$

where $h(\vec{u}\,) = g\left(\sqrt{\lambda_1}\, u_1, \ldots, \sqrt{\lambda_k}\, u_k\right)$ and

$$w_k(\vec{u}\,) = (2\pi)^{-k/2} \exp\left(-\|\vec{u}\,\|^2/2\right), \quad \|\vec{u}\,\|^2 = \sum_{i=1}^k u_i^2.$$

There exists a positive number $M = M(k, r, \{\lambda_i\})$ depending on $k$, $r$, and the eigenvalues, such that $M^{-1}\, h \in C^{r,k}(\mathbb{R}^k)$.

Note that our problem can be expressed as

$$S(f) \;=\; \int_{\|\vec{u}\|\,\leq\,2} h(\vec{u}\,)\, w_k(\vec{u}\,)\, d\,\vec{u} \;+\; \sum_{i=1}^{\infty} \int_{2^i\,\leq\,\|\vec{u}\|\,\leq\,2^{i+1}} h(\vec{u}\,)\, w_k(\vec{u}\,)\, d\,\vec{u}. \tag{7}$$

Without loss of generality assume that $n$ is a power of two. We will approximate the successive terms in (7), $i \;=\; 0, 1, \ldots, -1 + \log_2 n$, using $n/2^{i+1}$ points. We choose these points $\vec{t_{i,j}}$ to minimize the error of approximating the function $h$ in the $L_2$-sense over the domain

$$D_i \;=\; \left\{\vec{t} \in \mathbb{R}^k \,:\, \|\vec{t}\| \;\leq\; 2^{i+1}\right\}.$$

Let $m_j(D_i)$ denote the error of such an approximation when $j$ function values are used. It is easy to check that

$$m_j(D_i) \;=\; O\left(2^{i(r+k/2)}\, m_j(D_0)\right), \quad \forall\, i.$$

Here and below the factors in the $O$ notation may depend on $k$ and $r$. It is known, see e.g., [4], that

$$m_j(D_0) \;=\; \Theta(j^{-r/k}).$$

There is a linear algorithm, $A_i(h) \;=\; \sum_{j=1}^{n/2^i} h(\vec{t_{i,j}})h^*_{i,j}$ for some functions $h^*_{i,j}$, whose error is $m_j(D_0)$. We approximate the $i$th term of (7) by

$$\int_{D_i \setminus D_{i-1}} A_i(h)(\vec{u}\,)\, w_k(\vec{u}\,)\, d\,\vec{u}, \qquad \text{where} \quad D_{-1} = \emptyset.$$

Clearly, its error is bounded by

$$\|h \;-\; A_i(h)\|_{L_2(D_i)}\, \|w_k\|_{L_2(D_i \setminus D_{i-1})}.$$

Observe that

$$\|w_k\|_{L_2(D_i \setminus D_{i-1})} \;=\; O\left(\exp(-2^{2i-1})\right).$$

Hence, the total error $e_n$ of this approximation is bounded by

$$e_n \;=\; O\left(m_{n/2}(D_0) \;+\; \sum_{i=1}^{-1+\log_2 n} \exp(-2^{2i-1})\, m_{n/2^{i+1}}(D_i) \;+\; \int_{\|\vec{u}\|\,\geq\,n} \exp(-\|\vec{u}\,\|^2/2)\, d\,\vec{u}\right).$$

Since the last term is $O(\exp(-n))$ we finally have

$$e_n \;=\; O\left(n^{-r/k}\left(1 \;+\; \sum_{i=1}^{-1+\log_2 n} \exp(-2^{2i-1})\, 2^{i(r+k/2+r/k)}\right)\right).$$

Since the series $\sum_{i=1}^{\infty} \exp(-2^{2i-1})\, 2^{i(r+k/2+r/k)}$ is convergent, we conclude that $e_n \;=\; O(n^{-r/k})$. Setting $n \;=\; O(\varepsilon^{-k/r})$ and keeping in mind that our algorithm is linear (so that its cost is proportional to $\mathbf{c}\, n$) we conclude that the complexity is bounded by $O(\mathbf{c}\, \varepsilon^{k/r})$. This completes the proof of (ii). $\square$

# IV. Entire $F$

In this section, we demonstrate tractability of path integration for a certain class $F$ of entire functions defined on an infinite dimensional space $X$. We do this assuming additionally that we can compute the derivatives of integrands at zero.

First, we need to analyze the case of entire functions defined on the finite dimensional space $\mathbb{R}^d$, and then we extend the analysis to the space $X$. In what follows, the spaces of entire functions will depend on a sequence of positive numbers $\beta_k$ such that

$$\max_k \lambda_k \beta_k \;<\; 1 \qquad \text{and} \qquad \max_k \beta_k \;<\; \infty. \tag{8}$$

Without loss of generality, we assume that

$$\lambda_k \beta_k \;\geq\; \lambda_{k+1}\beta_{k+1}, \qquad \forall\, k.$$

Since $\lambda_k$ are summable, so are $\lambda_k\beta_k$ and, in particular, they converge to zero. As we shall see, the complexity of the problem depends on how fast they decay. More precisely, it depends on the *sum-exponent* $p_{\lambda\beta}$ of the sequence $\{\lambda_k\beta_k\}_k$ defined as follows:

$$p_{\lambda\beta} \;=\; \inf\left\{ p \;:\; \sum_{k=1}^{\infty} (\lambda_k\beta_k)^p < \infty \right\}. \tag{9}$$

Of course, we always have $p_{\lambda\beta} \leq 1$.

## 0.1 Finite dimensional case

In this subsection we consider $d$ dimensional integration for a class of entire functions defined on $\mathbb{R}^d$.

Let $\mathbb{N}_+^d$ denote the set of multi-indices $\vec{i} = [i_1, i_2, \ldots, i_d]$ with nonnegative integers $i_k$. By $|\vec{i}\,|$ we mean $\sum_{k=1}^d i_k$. Consider the Hilbert space $H_d$ of entire functions $f : \mathbb{R}^d \to \mathbb{R}$ with the inner product

$$\langle f, g\rangle_{H_d} \;=\; \sum_{\vec{i}\in\mathbb{N}_+^d} \frac{f^{(\vec{i}\,)}(0)\, g^{(\vec{i}\,)}(0)}{\prod_{k=1}^d i_k!\beta_k^{i_k}}, \qquad \text{where} \quad f^{(\vec{i}\,)}(0) \;=\; \frac{\partial^{|\vec{i}\,|} f(0)}{\partial x_1^{i_1} \cdots \partial x_d^{i_d}}.$$

Define the function

$$R_d(x,t) \;=\; \exp\left(\sum_{k=1}^d x_k t_k \beta_k\right), \quad x,t \in \mathbb{R}^d.$$

Note that

$$R_d^{(\vec{i})}(\cdot,t)\Big|_{x=0} \;=\; \prod_{k=1}^d t_k^{i_k} \beta_k^{i_k},$$

and therefore

$$\langle f, R_d(\cdot,t)\rangle_{H_d} \;=\; \sum_{\vec{i}\in\mathbb{N}_+^d} f^{(\vec{i})}(0) \prod_{k=1}^d \frac{t_k^{i_k}}{i_k!} \;=\; f(t).$$

Note that the last series is absolutely convergent since it is bounded by

$$\left(\sum_{\vec{i}\in\mathbb{N}_+^d} \frac{f^{(\vec{i})}(0)^2}{\prod_{k=1}^d i_k!\,\beta_k^{i_k}}\right)^{1/2} \left(\sum_{\vec{i}\in\mathbb{N}_+^d} \prod_{k=1}^{\infty} \frac{t_k^{2i_k}\beta_k^{i_k}}{i_k!}\right)^{1/2} \;=\; \|f\|_{H_d} \exp\left(\sum_{k=1}^d t_k^2 \beta_k/2\right).$$

This verifies that $H_d$ is a space of entire functions and that $R_d$ is its reproducing kernel. (Basic information on reproducing kernel Hilbert spaces can be found, e.g., in [34].) Let $F_d$ be the unit ball of $H_d$,

$$F_d \;=\; \{f \in H_d : \; \|f\|_{H_d} \le 1\}.$$

As in the previous section, let $S_d$ be the following linear functional from $H_d$ to $\mathbb{R}$:

$$S_d(f) \;=\; \frac{1}{\prod_{k=1}^d \sqrt{2\pi\lambda_k}} \int_{\mathbb{R}^d} \exp\left(-\sum_{k=1}^d t_k^2/(2\lambda_k)\right) f(t)\,dt.$$

We approximate $S_d$ by algorithms that use $f^{(\vec{j})}(0)$ as information about $f$. For a given finite subset $M \subset \mathbb{N}_+^d$, we compute $f^{(2\vec{i})}(0)$ for $\vec{i} \in M$ and define the algorithm

$$A_{d,M}(f) \;=\; \sum_{\vec{i}\in M} f^{(2\vec{i})}(0) \prod_{k=1}^d \frac{\lambda_k^{i_k}(2i_k-1)!!}{(2i_k)!}. \tag{10}$$

We prove that $A_{d,M}$ is *optimal* in the class of algorithms that use information $f^{(2\vec{i})}(0)$ for $\vec{i} \in M$. Here, optimality is understood in the sense of minimizing the worst case error. The worst case error of the algorithm $\phi(f) = \phi(f^{(2\vec{i})}(0) : \vec{i} \in M)$ in the unit ball $F_d$ of $H_d$ is defined as

$$e(\phi) \;=\; \sup_{f\in F_d} |S_d(f) - \phi(f)|.$$

**Theorem 2** *The algorithm $A_{d,M}$ is optimal and we have*

$$e^2(A_{d,M}) \;=\; \sum_{\vec{i}\notin M} \prod_{k=1}^{d} \frac{(\lambda_k\beta_k)^{2i_k}((2i_k-1)!!)^2}{(2i_k)!}. \tag{11}$$

*For any $\alpha \in (0,2)$,*

$$e^2(A_{d,M}) \;\le\; C_{d,\alpha} \max_{\vec{i}\notin M} \prod_{k=1}^{d} (\lambda_k\beta_k)^{i_k\alpha} \tag{12}$$

*with*

$$C_{d,\alpha} = \prod_{k=1}^{d} \left(1-(\lambda_k\beta_k)^{2-\alpha}\right)^{-1/2}.$$

*Proof:* Take an arbitrary algorithm $\phi$ that uses information $f^{(2\vec{i})}(0)$ for $\vec{i} \in M$. Due to Smolyak's theorem, see e.g., [8] p. 76, we know that the worst case error is minimized by a linear algorithm. Therefore we may assume that $\phi$ is linear, $\phi(f) = \sum_{\vec{i}\in M} a_{\vec{i}} f^{(2\vec{i})}(0)$ for some weights $a_{\vec{i}}$. Observe that for $a_{\vec{i}} = a^*_{\vec{i}}$ with $a^*_{\vec{i}} = \prod_{k=1}^{d} \lambda_k^{i_k}(2i_k-1)!!/(2i_k)!$ we have $\phi = A_{d,M}$.

Since $H_d$ is a reproducing kernel Hilbert space and both $S_d$ and $\phi$ are continuous linear functionals, the worst case error of $\phi$ is equal to the average case error of $\phi$ for a certain space and a certain measure, see e.g., [8] p. 304 and [35]. More precisely, there exists a separable Banach space $B_d$ of functions defined on $\mathbb{R}^d$ such that $H_d$ is a dense subset of $B_d$. The space $B_d$ is equipped with a zero mean Gaussian measure $\nu_d$ whose covariance function is the reproducing kernel $R_d$,

$$R_d(x,t) \;=\; \int_{B_d} f(x)f(t)\,\nu_d(df).$$

Then

$$e^2(\phi) \;=\; \int_{B_d} |S_d(f)-\phi(f)|^2 \;\nu_d(df).$$

Since for every $x = [x_1,\dots,x_d] \in \mathbb{R}^d$ and every $f \in H_d$,

$$f(x) = \sum_{\vec{i}\in\mathbb{N}_+^d} f^{(\vec{i})}(0) \prod_{k=1}^{d} \frac{x_k^{i_k}}{i_k!},$$

we have

$$S_d(f) = \sum_{\vec{i}\in\mathbb{N}_+^d} f^{(2\vec{i})}(0) \prod_{k=1}^{d} \frac{\lambda_k^{i_k}(2i_k-1)!!}{(2i_k)!}.$$

It is easy to check that

$$\int_{B_d} f^{(\vec{i})}(0)\, f^{(\vec{j})}(0)\, \nu_d(df) \;=\; R_d^{(\vec{i},\vec{j})}(0,0) \;=\; \delta_{\vec{i},\vec{j}}\, \prod_{k=1}^{d} i_k!\beta_k^{i_k},$$

where $\delta_{\vec{i},\vec{j}}$ is the Kronecker delta. From this, we conclude that

$$e^2(\phi) \;=\; \sum_{\vec{i}\in M} (a_{\vec{i}} - a^*_{\vec{i}})^2 \prod_{k=1}^{d} (2i_k)!\beta_k^{2i_k} \;+\; \sum_{\vec{i}\notin M} \prod_{k=1}^{d} \frac{(\lambda_k\beta_k)^{2i_k}((2i_k-1)!!)^2}{(2i_k)!}.$$

Hence, the worst case error is minimized iff $a_{\vec{i}} = a^*_{\vec{i}}$. That is, $\phi = A_{d,M}$ and the error of $A_{d,M}$ is given by (11).

We now estimate $e^2(A_{d,M})$ as follows

$$e^2(A_{d,M}) \;\le\; \left(\max_{\vec{i}\notin M} \prod_{k=1}^{d} (\lambda_k\beta_k)^{i_k\alpha}\right) \left(\sum_{\vec{i}\in \mathbb{N}_+^d} \prod_{k=1}^{d} \frac{(\lambda_k\beta_k)^{i_k(2-\alpha)}((2i_k-1)!!)^2}{(2i_k)!}\right).$$

We now compute the last sum. Note that it has the same form as (11) for $M=\emptyset$ and $\widetilde{\lambda}_k = \lambda_k^{1-\alpha/2}$, $\widetilde{\beta}_k = \beta_k^{1-\alpha/2}$. This corresponds to the square of the error of the zero algorithm for approximating

$$\widetilde{S}_d(f) \;=\; \frac{1}{\prod_{k=1}^{d} \sqrt{2\pi\widetilde{\lambda}_k}} \int_{\mathbb{R}^d} e^{-\sum_{k=1}^{d} t_k^2/(2\widetilde{\lambda}_k)}\, f(t)\, dt,$$

where the functions $f$ are now from a reproducing kernel Hilbert space with kernel $\widetilde{R}_d(x,t) = \exp\left(\sum_{k=1}^{d} x_k t_k \widetilde{\beta}_k\right)$. That is, denoting this sum by $\gamma$, we have

$$\begin{aligned}
\gamma \;&=\; \int_{\widetilde{B}_d} |\widetilde{S}_d(f)|^2\, \widetilde{\nu}_d(df) \\
&=\; \int_{\mathbb{R}^d} \frac{\exp\left(-\sum_{k=1}^{d} x_k^2/(2\widetilde{\lambda}_k)\right)}{\prod_{k=1}^{d}\sqrt{2\pi\widetilde{\lambda}_k}} \int_{\mathbb{R}^d} \frac{\exp\left(-\sum_{k=1}^{d} t_k^2/(2\widetilde{\lambda}_k)\right)}{\prod_{k=1}^{d}\sqrt{2\pi\widetilde{\lambda}_k}} \widetilde{R}_d(x,t)\, dt\, dx.
\end{aligned}$$

It is easy to verify that the right-hand side equals $\prod_{k=1}^{d}(1-(\lambda_k\beta_k)^{2-\alpha})^{-1/2}$. This completes the proof. $\square$

We now choose a subset $M$ such that the error of the algorithm $A_{d,M}$ is at most $\varepsilon$. For $\alpha \in (0, 2-p_{\lambda\beta})$, let

$$M_{d,\alpha}(\varepsilon) \;=\; \left\{\vec{i} \in \mathbb{N}_+^d \;:\; \prod_{k=1}^{d} (\lambda_k\beta_k)^{i_k} > (\varepsilon^2/C_{d,\alpha})^{1/\alpha}\right\}. \tag{13}$$

Note that the sequence $\{C_{d,\alpha}\}_d$ is increasing and

$$C_\alpha \;=\; \lim_{d\to\infty} C_{d,\alpha} \;=\; \prod_{k=1}^{\infty} \left(1-(\lambda_k\beta_k)^{2-\alpha}\right)^{-1/2} \tag{14}$$

exists. It is finite since $\alpha < 2 - p_{\lambda\beta}$ implies that $\sum_{k=1}^{\infty}(\lambda_k\beta_k)^{(2-\alpha)} < +\infty$. Denote

$$A_{d,\alpha}(\varepsilon) \;=\; A_{d,M_{d,\alpha}(\varepsilon)}. \tag{15}$$

The next theorem presents the error and cost bounds of the algorithm $A_{d,\alpha}(\varepsilon)$.

**Theorem 3** *For every $d$, $\varepsilon > 0$, and $\alpha \in (0, 2-p_{\lambda\beta})$, the algorithm $A_{d,\alpha}(\varepsilon)$ has error at most $\varepsilon$, and its cost is at most $(\mathbf{c}+2)\, n_{d,\alpha}(\varepsilon)$. Here, $n_{d,\alpha}(\varepsilon)$ is the cardinality of the set $M_{d,\alpha}(\varepsilon)$ and denotes how many derivative evaluations of $f$ are used, and*

$$n_{d,\alpha}(\varepsilon) \;\le\; K_\alpha\, \varepsilon^{-2(2-\alpha)/\alpha}, \tag{16}$$

*where*

$$K_\alpha \;=\; \prod_{i=1}^{\infty}(1-(\lambda_i\beta_i)^{2-\alpha})^{-1} \;\sup_{\varepsilon\in(0,1)} \varepsilon^{2(2-\alpha)/\alpha}\left(2+\frac{\ln C_\alpha/\varepsilon^2}{\alpha\,\ln 1/(\lambda_1\beta_1)}\right) \;<\; +\infty.$$

*Proof:* The bound $\varepsilon$ on the error of the algorithm follows directly from (12) of Theorem 2.

We now prove the bound on $n_{d,\alpha}(\varepsilon)$. Note that $\alpha < 2-p_{\lambda\beta}$ implies that $K_\alpha$ is finite. Let $p = 2(2-\alpha)/\alpha$.

For $d=1$, we have

$$n_{1,\alpha}(\varepsilon) \;\le\; 1+\lceil(\ln C_\alpha/\varepsilon^2)/(\alpha\ln 1/(\lambda_1\beta_1))\rceil \;\le\; K_{1,\alpha}\,\varepsilon^{-p},$$

where

$$K_{1,\alpha} \;=\; \sup_{\varepsilon\in(0,1)} \varepsilon^p\left(2+\frac{\ln C_\alpha/\varepsilon^2}{\alpha\,\ln 1/(\lambda_1\beta_1)}\right).$$

Hence (16) holds for $d=1$ since $K_{1,\alpha}\le K_\alpha$.

By induction, suppose that $n_{d-1,\alpha}(\varepsilon) \le K_{d-1,\alpha}\varepsilon^{-p}$. Then

$$n_{d,\alpha}(\varepsilon) \;\le\; \sum_{i=1}^{\infty} n_{d-1,\alpha}\left(\varepsilon/(\lambda_d\beta_d)^{\alpha i/2}\right) \;\le\; K_{d-1,\alpha}\,\varepsilon^{-p}\sum_{i=0}^{\infty}(\lambda_d\beta_d)^{\alpha i p/2} \;=\; K_{d,\alpha}\,\varepsilon^{-p},$$

where

$$
\begin{aligned}
K_{d,\alpha} &= K_{d-1,\alpha}\,(1-(\lambda_d\beta_d)^{2-\alpha})^{-1} \;=\; K_{1,\alpha}\prod_{i=1}^{d}(1-(\lambda_i\beta_i)^{2-\alpha})^{-1} \\
&\le K_{1,\alpha}\prod_{i=1}^{\infty}(1-(\lambda_i\beta_i)^{2-\alpha})^{-1} \;=\; K_\alpha.
\end{aligned}
$$

This proves (16). The algorithm $A_{d,M}$ is linear and its weights $a^*_{\vec{i}}$ can be precomputed. Hence, its cost is equal to $n_{d,\alpha}(\varepsilon)$ derivative evaluations, $n_{d,\alpha}(\varepsilon)$ multiplications, and $n_{d,\alpha}(\varepsilon)-1$ additions. Hence, it is bounded by $(\mathbf{c}+2)\,n_{d,\alpha}(\varepsilon)$, as claimed. This completes the proof. □

Theorem 3 presents an upper bound on the worst case complexity of multivariate integration $S_d$ in the unit ball $F_d$. Namely

$$
\mathrm{comp}(\varepsilon) \;\le\; (\mathbf{c}+2)\,K_\alpha\,\varepsilon^{-p^*},
$$

where $p^* = 2(2-\alpha)/\alpha$. Since $\alpha$ can be arbitrarily close to $\alpha^* = 2 - p_{\lambda\beta}$, we may have

$$
p^* \;\simeq\; \frac{2p_{\lambda\beta}}{2-p_{\lambda\beta}}.
$$

Since $p_{\lambda\beta} \le 1$, the exponent $p^*$ of the multivariate problem $S_d$ is always bounded by two. Hence, it is no larger than the exponent of $1/\varepsilon$ in the cost function of the classical Monte Carlo method. It can be even smaller than 2. For instance, for $\lambda_k\beta_k = \Theta(k^{-r})$ with $r > 1$, we have

$$
p^* \;\simeq\; \frac{2}{2r-1}. \tag{17}
$$

Observe that $\lambda_k\beta_k = \Theta(k^{-r})$ holds for $\beta_k \simeq 1$ and for the $(r-2)$-fold Wiener measure $\mu$. For the classical Wiener measure $\mu$, we have $r = 2$ and $p^* \simeq 2/3$.

## 0.2 Infinite dimensional case

We now consider the infinite dimensional case, $d = +\infty$. To that end, let $\mathbb{N}_+^\infty$ denote the set of infinite multi-indices $\vec{i} = [i_1, i_2, \ldots]$ with nonnegative integers $i_k$ for which $|\vec{i}\,| = \sum_{k=1}^{\infty} i_k$ is finite. This means that any $\vec{i} \in \mathbb{N}_+^\infty$ has almost all coefficients equal zero.

We define the Hilbert space $H_\infty$ of entire functions as a limiting case of spaces $H_d$. That is, $H_\infty$ is a space of entire functions $f : X \to \mathbb{R}$ with inner product

$$
\langle f, g\rangle_{H_\infty} \;=\; \sum_{\vec{i}\in\mathbb{N}_+^\infty} \frac{\left(f^{(\vec{i})}(0)\,\prod_{k=1}^{\infty}(\mathrm{Im}^{-1}\eta_k)^{i_k}\right)\,\left(g^{(\vec{i})}(0)\,\prod_{k=1}^{\infty}(\mathrm{Im}^{-1}\eta_k)^{i_k}\right)}{\prod_{k=1}^{\infty} i_k!\,\beta_k^{i_k}},
$$

where, as before, $\eta_k$ are the eigenelements of the covariance operator $C_\nu$.

This is a reproducing kernel Hilbert space whose reproducing kernel is

$$R_\infty(x,t) \;=\; \exp\left(\sum_{k=1}^{\infty} \langle \mathrm{Im}(x), \eta_k\rangle \; \langle \mathrm{Im}(t), \eta_k\rangle \; \beta_k\right), \qquad x,t \,\in X.$$

Consider the class $F$ as the unit ball of $H_\infty$, i.e.,

$$F \;=\; \{f \in H_\infty : \; \|f\|_{H_\infty} \leq 1\,\}\,.$$

This class is a limiting case of the unit balls $F_d$ in the spaces $H_d$, and all the results from the previous section apply. More precisely, for a given finite subset $M$ of $\mathbb{N}_+^\infty$, let

$$A_M(f) \;=\; \sum_{\vec{i}\in M} \left(f^{(2\vec{i})}(0)\,\prod_{k=1}^{\infty}(\mathrm{Im}^{-1}\eta_k)^{2i_k}\right)\,\prod_{k=1}^{\infty}\frac{\lambda_k^{i_k}(2i_k-1)!!}{(2i_k)!}. \tag{18}$$

For $\alpha \in (0, 2-p_{\lambda\beta})$ and

$$C_\alpha \;=\; \prod_{k=1}^{\infty}\left(1-(\lambda_k\beta_k)^{2-\alpha}\right)^{-1/2},$$

see (14), let

$$M_\alpha(\varepsilon) \;=\; \left\{\vec{i} \in \mathbb{N}_+^\infty \;:\; \prod_{k=1}^{\infty}(\lambda_k\beta_k)^{i_k} > \left(\varepsilon^2/C_\alpha\right)^{1/\alpha}\right\}. \tag{19}$$

We have the following theorem.

**Theorem 4** **(i)** *The algorithm $A_M$ is optimal in the class of algorithms that use the same information, and its worst case error is given by*

$$e^2(A_M) \;=\; \sum_{\vec{i}\in\mathbb{N}_+^\infty\setminus M}\;\prod_{k=1}^{\infty}\frac{(\lambda_k\beta_k)^{2i_k}((2i_k-1)!!)^2}{(2i_k)!}.$$

**(ii)** *For every $\varepsilon$ and $\alpha \in (0,2-p_{\lambda\beta})$, the algorithm $A_{M_\alpha(\varepsilon)}$ has error at most $\varepsilon$, and its cost is bounded by*

$$(\mathbf{c}+2)\,K_\alpha\,\varepsilon^{-2(2-\alpha)/\alpha}.$$

**(iii)** *The path integration problem is tractable and its exponent $p(F)$ is bounded by*

$$p(F) \;\leq\; \frac{2p_{\lambda\beta}}{2-p_{\lambda\beta}}.$$

*For $\lambda_i\beta_i = \Theta\left(i^{-r}\right)$ with some $r > 1$, we have*

$$p(F) \;=\; \frac{2p_{\lambda\beta}}{2-p_{\lambda\beta}} \;=\; \frac{2}{2r-1}$$

*if the information is restricted to function and derivative values at zero.*

*Proof:* The proof of (i) and (ii) follows from the results of the previous section. Indeed, given $d$ and $f \in F$, let

$$f_d(\vec{t}\,) \;=\; f\left(\mathrm{Im}^{-1}(t_1\eta_1+\cdots+t_d\eta_d)\right), \qquad \text{with} \quad \vec{t} = [t_1,\ldots,t_d] \;\in\; \mathbb{R}^d.$$

Then $f_d^{(\vec{i}\,)}(0) \;=\; f^{(|\vec{i}\,|)}(0)(\mathrm{Im}^{-1}\eta_1)^{i_1}\ldots(\mathrm{Im}^{-1}\eta_d)^{i_d}$, and $\|f_d\|_{H_d} \;\le\; \|f\|_{H_\infty}$. Hence, $f_d$ belongs to the unit ball of the space $H_d$.

For a given $M$, let $d_M = \max\{k \,:\, i_k \neq 0 \text{ for some } \vec{i} \in M\}$. Then, for every $f \in F$, we have

$$A_M(f) \;=\; A_{d,M}(f_d), \qquad \forall\, d \ge d_M.$$

Since $S(f) = \lim_{d\to\infty} S_d(f_d)$, the first part of Theorem 4 follows from Theorem 2.

Let $n_\alpha(\varepsilon)$ be the cardinality of the set $M_\alpha(\varepsilon)$. Since $n_{d,\alpha}(\varepsilon) \le n_{d+1,\alpha}(\varepsilon)$ for every $d$, and $n_{d,\alpha}(\varepsilon) = n_\alpha(\varepsilon)$ for $d \ge d_{M_\alpha(\varepsilon)}$, the second part follows from Theorem 3.

We now prove (iii). Obviously, it is enough to show that $2/(2r-1)$ is a lower bound on $p(F)$ in the class of information restricted to function and derivative values at zero. Consider arbitrary information consisting of $n$ function and derivative values at 0. This corresponds to a subset $M$ of $\mathbb{N}_+^\infty$ having cardinality $n$. The minimal worst case error is given by $e^2(A_M)$, where the algorithm $A_M$ is given by (18). We have

$$e^2(A_M) \;=\; \sum_{\vec{i}\in\mathbb{N}_+^\infty}\prod_{k=1}^{\infty}\frac{(\lambda_k\beta_k)^{2i_k}((2i_k-1)!!)^2}{(2i_k)!} - \sum_{\vec{i}\in M}\prod_{k=1}^{\infty}\frac{(\lambda_k\beta_k)^{2i_k}((2i_k-1)!!)^2}{(2i_k)!}.$$

Observe that the first sum is greater than $1+\sum_{k=1}^{\infty}(\lambda_k\beta_k)^2/2$. Since the second sum has only $n$ elements, we have

$$e^2(A_M) \;\ge\; \frac{1}{2}\sum_{k=n}^{\infty}(\lambda_k\beta_k)^2 \;=\; \Theta\left(n^{-(2r-1)}\right).$$

To guarantee that the error is at most $\varepsilon$ we have to take $n = \Omega\left(\varepsilon^{-2/(2r-1)}\right)$. This completes the proof of Theorem 4. $\square$

## Acknowledgment

We wish to express our sincere thanks to S. Heinrich, E. Novak, K. Ritter, J. F. Traub, and A. G. Werschulz for useful comments.

We owe special thanks to W. Paciorek and B. Whaley for providing relevant literature on path integration in physics and chemistry.

## References

[1] S. M. Nikolskij, On the problem of approximation estimate by quadrature formulas (in Russian), *Usp. Mat. Nauk* 5, 165-177 (1950).

[2] A. Sard, Best approximate integration formulas; best approximation formulas, *Amer. J. Math.* 17, 80-91 (1949).

[3] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, *CBMS-NSF Reg. Conf. Series Appl. Math.*, 63, *SIAM*, Philadelphia, 1992.

[4] E. Novak, Deterministic and Stochastic Error Bounds in Numerical Analysis, *Lecture Notes in Mathematics*, 1349, Springer-Verlag, Berlin, 1988.

[5] S. Paskov, Average case complexity of multivariate integration for smooth functions, *J. Complexity* 9, 291-312 (1993).

[6] K. Ritter, G. W. Wasilkowski, and H. Woźniakowski, On multivariate integration for stochastic processes, in *Numerical Integration* (H. Braß, G. Hämmerlin,eds.), International Series of Numerical Mathematics, vol. 112, *Birkhäuser*, Basel, 331-347 (1993).

[7] K. Ritter, G. W. Wasilkowski, and H. Woźniakowski, Multivariate integration and approximation for random fields satisfying Sacks-Ylvisaker conditions, *Annals of Appl. Prob.* 5, 518-540 (1995).

[8] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, Information-based Complexity, *Academic Press*, New York, 1988.

[9] G. W. Wasilkowski, Integration and approximation of multivariate functions: average case complexity with isotropic Wiener measure, *Bull. Amer. Soc. (N.S.)* 28, 308-314 (1993). Full version *J. Approx. Theory* 77, 212-227 (1994).

[10] G. W. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems, *J. Complexity* 11, 1-56 (1995).

[11] H. Woźniakowski, Average case complexity of multivariate integration, *Bull. Amer. Math. Soc. (N.S)* 24, 185-194 (1991).

[12] R. P. Feynman and A. R. Hibbs, Quantum Mechanics and Path-Integrals, *McGraw-Hill*, New York, 1965.

[13] C. DeWitt-Morette, Editor, Special issue on functional integration, *J. of Math. Phys.* 36, no.5 (1995).

[14] A. D. Egorov, P. I. Sobolevsky and L. A. Yanovich, Functional Integrals: Approximate Evaluation and Applications, *Kluwer Academic Publishers*, Dordrecht, 1993.

[15] A. Das, Field Theory: A Path Integral Approach, Lecture Notes in Physics, 52, *World Scientific*, Singapore, 1993.

[16] C. DeWitt-Morette, A. Maheshwari and B. Nelson, Path integration in non-relativistic quantum mechanics, *Physics Reports (Review of Physics Letter)* 50, 255-372 (1979).

[17] C. Grosche and F. Steiner, How to solve path integrals in quantum mechanics, *DESY 94-180*, 1-39 (1994), and also *J. of Math. Phys.* 36, 2354-2385 (1995).

[18] D. C. Khandekar, S. V. Lawande and K. V. Bhagwat, Path-Integral Methods and their Applications, *World Scientific*, Singapore, 1993.

[19] H. Kleinert, Path Integrals in Quantum Mechanics, Statistics and Polymer Physics, *World Scientific*, Singapore, 1990.

[20] P. K. MacKeown, Evaluation of Feynman path integrals by Monte Carlo methods, *Amer. J. of Physics* 53, 880-885 (1985).

[21] N. Makri, Numerical path integral techniques for long time dynamics, *J. of Math. Phys.* 36, 2430-2457 (1995).

[22] L. S. Schulman, Techniques and Applications of Path Integration, *J. Wiley & Sons*, New York, 1981.

[23] F. W. Wiegel, Path Integral Methods in Physics and Polymer Physics, *World Scientific*, Singapore, 1986.

[24] S. A. Albeverio and R. J. Høegh-Krohn, Mathematical Theory of Feynman Path Integrals, Lecture Notes in Mathematics, 523, *Springer Verlag*, Berlin, 1976.

[25] M. Kac, On distributions of certain Wiener functionals, *Trans. Amer. Math. Soc.* 65, 1-13 (1949).

[26] D. Duffie, Dynamic Asset Pricing Theory, *Princeton University Press*, Princeton, New Jersey, 1992.

[27] J. Hull, Options, Futures, and Other Derivative Securities, *Prentice Hall*, Englewood Cliffs, N.J., second edition, 1993.

[28] R. Merton, Continuous-Time Finance, *Basil Blackwell*, Oxford, 1990.

[29] H. Woźniakowski, Tractability and strong tractability of multivariate tensor product problems, in Proceedings of 2nd Gauss Symposium, Muenchen, 1993, *J. Computing and Information* 4, 1-19 (1994).

[30] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, *J. Complexity* 10, 96-128 (1994).

[31] G. W. Wasilkowski and H. Woźniakowski, Worst case complexity of Feynman-Kac path integration, in preparation.

[32] N. N. Vakhania, V. I. Tarieladze, and S. A. Chobanyan, Probability Distributions on Banach Spaces, *Reidel, Dordrecht*, 1987.

[33] N. S. Bakhvalov, On approximate calculation of integrals (in Russian), *Vestnik MGV, Ser. Mat. Mekh. Astron. Fiz. Khim.* 4, 3-18 (1959).

[34] N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* 68, 337-404 (1950).

[35] G. Wahba, Spline Models for Observational Data, *CBMS-NSF Reg. Conf. Series Appl. Math.*, 59, *SIAM*, Philadelphia, 1990.

**Authors' Addresses:**

Grzegorz W. Wasilkowski, Department of Computer Science, University of Kentucky, Lexington, KY 40506, USA, email: greg@cs.engr.uky.edu

Henryk Woźniakowski, Department of Computer Science, Columbia University,
New York, NY 10027, USA, and Institute of Applied Mathematics, University of Warsaw,
ul. Banacha 2, 02-097 Warszawa, Poland, email: henryk@cs.columbia.edu