

RANDOMIZING THE TRAPEZOIDAL RULE GIVES THE OPTIMAL RMSE RATE IN GAUSSIAN SOBOLEV SPACES

TAKASHI GODA, YOSHIHITO KAZASHI, AND YUYA SUZUKI

ABSTRACT. Randomized quadratures for integrating functions in Sobolev spaces of order $\alpha \geq 1$, where the integrability condition is with respect to the Gaussian measure, are considered. In this function space, the optimal rate for the worst-case root-mean-squared error (RMSE) is established. Here, optimality is for a general class of quadratures, in which adaptive non-linear algorithms with a possibly varying number of function evaluations are also allowed. The optimal rate is given by showing matching bounds. First, a lower bound on the worst-case RMSE of $O(n^{-\alpha-1/2})$ is proven, where n denotes an upper bound on the expected number of function evaluations. It turns out that a suitably randomized trapezoidal rule attains this rate, up to a logarithmic factor. A practical error estimator for this trapezoidal rule is also presented. Numerical results support our theory.

1. INTRODUCTION

We study numerical integration of functions with respect to the standard Gaussian measure. Given a function $f : \mathbb{R} \rightarrow \mathbb{R}$, our problem is to approximate the integral

$$I(f) = \int_{\mathbb{R}} f(x)\rho(x) dx \quad \text{with} \quad \rho(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2},$$

based on pointwise function evaluations.

Of particular interest in this paper are randomized algorithms, which typically combine a fixed (deterministic) quadrature rule and the Monte Carlo sampling; see for example [13, Section 12.7] and references therein as well as [19, 6, 23]. One benefit of these algorithms is that one may construct online error estimators to assess their accuracy, which is often difficult with deterministic algorithms. Another is that they are robustly efficient for various smoothness of the integrand. Even in the absence of smoothness of the integrand, which is typically required for deterministic quadratures, they may work well at least as accurate as the plain vanilla Monte Carlo method; in the presence of smoothness, they exploit it and may achieve accuracy superior to the Monte Carlo methods.

How accurate can a randomized algorithm be? To address this question, we fix a function class and a figure of merit. We then consider a general class of algorithms and derive a lower bound for the error, which quantifies the best possible accuracy using this class of algorithms in the sense specified below.

Date: July 21, 2023.

2010 Mathematics Subject Classification. Primary 65C05, 65D30, 65D32, 68W20.

Key words and phrases. trapezoidal rule, Gaussian Sobolev space, randomized setting, root-mean-squared error, lower bound.

This work was supported by JSPS KAKENHI Grant Number 20K03744 (T.G.) and by NTNU project grant 81617985 (Y.S.).

It turns out that this best possible rate is, up to a logarithmic factor, attained by a randomized trapezoidal rule, which we propose in this paper. This result is a random counterpart of our recent result [9], in which we showed that the trapezoidal rule is, up to a logarithmic factor, optimal to approximate $I(f)$ in a suitable sense.

Following [9], throughout this paper, we assume that f is in a Sobolev space of integer order $\alpha \geq 1$, denoted by \mathcal{H}_α . More precisely, with the normed vector space $L_\rho^2(\mathbb{R})$ defined by

$$\|f\|_{L_\rho^2}^2 := \int_{\mathbb{R}} |f(x)|^2 \rho(x) dx < \infty,$$

for an integer $\alpha \geq 1$, \mathcal{H}_α is defined by

$$\mathcal{H}_\alpha := \left\{ f \in L_\rho^2(\mathbb{R}) \mid \|f\|_\alpha := \left(\sum_{\tau=0}^{\alpha} \|f^{(\tau)}\|_{L_\rho^2}^2 \right)^{1/2} < \infty \right\},$$

where $f^{(\tau)}$ denotes the τ -th weak derivative of f .

For this class of functions, we quantify the efficiency of the algorithms in terms of the worst-case root-mean-squared error (RMSE), where the mean is taken on the underlying probability space where the randomization is considered. To define this, we first introduce a class of deterministic algorithms; realizations of the randomized algorithm we consider take their values in the following class of deterministic algorithms.

In the deterministic worst-case setting which we build upon, we approximate $I(f)$ by a (deterministic) quadrature rule of the form

$$(1) \quad A_m^{\det}(f) = \varphi_{m(f)}(x_1, x_2, \dots, x_{m(f)}, f(x_1), f(x_2), \dots, f(x_{m(f)})),$$

with nodes $x_i \in \mathbb{R}$, $i = 1, \dots, m(f)$, where $\varphi_{m(f)} : \mathbb{R}^{2m(f)} \rightarrow \mathbb{R}$ is a (linear or non-linear) mapping. Here, each node x_i can be chosen sequentially by using the information already obtained through x_1, \dots, x_{i-1} and $f(x_1), \dots, f(x_{i-1})$, for $i \geq 2$, i.e., with x_1 pre-selected independently of f , $x_i = \psi_i(x_1, \dots, x_{i-1}, f(x_1), \dots, f(x_{i-1}))$ for some (not necessarily linear) function ψ_i , $i \geq 2$. This sequential choice of nodes is continued until a stopping criterion is satisfied, and the number of quadrature nodes m is not determined a priori. More precisely, the sequential choice of nodes is terminated as soon as

$$(2) \quad \text{ter}_i(x_1, \dots, x_i, f(x_1), \dots, f(x_i)) = 1,$$

where $\text{ter}_i : \mathbb{R}^{2i} \rightarrow \{0, 1\}$ is a Boolean function for each i . Then the positive integer $m(f)$, representing the total number of function evaluations, is given by

$$m(f) = \min \{i \in \mathbb{N} \mid \text{ter}_i(x_1, \dots, x_i, f(x_1), \dots, f(x_i)) = 1\}.$$

For example, if we define the functions ter_i by $\text{ter}_i \equiv 0$ if $i < m$ and $\text{ter}_m \equiv 1$ for some fixed $m \in \mathbb{N}$, then the number of quadrature nodes is equal to m independently of the integrand f . Another example is the case where ter_i is a function of an error estimator

$$\text{Estimator}(x_1, \dots, x_i, f(x_1), \dots, f(x_i))$$

and a user-specified error tolerance TOL; ter_i is defined so that it returns 1 if the estimated error is below TOL, and returns 0 otherwise. In this case, the number of quadrature nodes in general depends on f .

In this paper, following the standard terminology in the context of information-based complexity [16, Sections 4.1.1 & 4.2.1] and [24, 15], we say that this type

of sequential node selection is *adaptive* and that (1) is an *adaptive algorithm*. If the number $m(f)$ does not vary with f and each node x_i is chosen without the information $f(x_1), \dots, f(x_m)$, we say that (1) is a *non-adaptive algorithm*. In this context, the adaptive stopping rule and the adaptive node selection are usually considered separately, and each type of adaptivity improves an algorithm in different ways; see [24, 15].

Our interest in this paper is in the randomized setting. We follow [16, Definition 4.35] for the class of algorithms we consider:

Definition 1.1. *A randomized quadrature rule A is a pair of a probability space (Ω, Σ, μ) and a family $(A^\omega)_{\omega \in \Omega}$ of mappings such that the following holds:*

- (1) *For each fixed $\omega \in \Omega$, the mapping $A^\omega : \mathcal{H}_\alpha \rightarrow \mathbb{R}$ is a deterministic quadrature rule of the form (1).*
- (2) *Let $M(f, \omega)$ be the number of nodes used in A^ω for $f \in \mathcal{H}_\alpha$ and ω . Then the function $\omega \mapsto M(f, \omega)$ is measurable for each fixed f .*

Finally, for a randomized quadrature rule A , its *cardinality* is defined as the supremum of the expected number of nodes over $f \in \mathcal{H}_\alpha$:

$$\#(A) := \sup_{f \in \mathcal{H}_\alpha} \int_{\Omega} M(f, \omega) \, d\mu(\omega).$$

We quantify the efficiency of the algorithm by the worst-case error. For a deterministic algorithm A_m^{\det} as in (1), the (deterministic) worst-case error in \mathcal{H}_α is given by

$$(3) \quad e^{\text{wor}}(A_m^{\det}, \mathcal{H}_\alpha) := \sup_{\substack{f \in \mathcal{H}_\alpha \\ \|f\|_\alpha \leq 1}} |I(f) - A_m^{\det}(f)|.$$

For the randomized setting, we consider the worst-case RMSE

$$e^{\text{rmse}}(A, \mathcal{H}_\alpha) := \sup_{\substack{f \in \mathcal{H}_\alpha \\ \|f\|_\alpha \leq 1}} \left(\int_{\Omega} (I(f) - A^\omega(f))^2 \, d\mu(\omega) \right)^{1/2}.$$

For A_m^{\det} we have $e^{\text{rmse}}(A_m^{\det}, \mathcal{H}_\alpha) = e^{\text{wor}}(A_m^{\det}, \mathcal{H}_\alpha)$.

The main contribution of this paper is two-fold.

- (1) We establish a lower bound for the worst-case RMSE $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$ for any integer $\alpha \geq 1$. More precisely, we prove that for any $n \geq 1$ and any randomized quadrature rule A with cardinality $\#(A) \leq n$, $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$ is bounded from below by $c_\alpha n^{-\alpha-1/2}$ with a constant $c_\alpha > 0$ depending only on α ; see Section 2.
- (2) We propose an algorithm that achieves this rate up to a logarithmic factor; hence, our algorithm is optimal, up to a logarithmic factor, in the sense of the worst-case RMSE. More precisely, we develop a suitably truncated randomized trapezoidal rule with the number of nodes $M(f, \omega)$ being randomly chosen independently of $f \in \mathcal{H}_\alpha$. This quadrature rule is shown to be an unbiased estimator of $I(f)$, which allows for practical error estimation.

As briefly mentioned earlier, the second contribution above is motivated by our recent paper [9], in which we considered a suitably truncated trapezoidal rule and established the optimality in terms of the worst-case error. There, building upon the work of Nuyens and Suzuki [18], we showed that for any $\alpha \geq 1$ a suitably

truncated trapezoidal rule achieves the optimal rate $\mathcal{O}(n^{-\alpha})$ up to a logarithmic factor.

As already mentioned, we randomize this trapezoidal rule and establish the optimality in the sense of the worst-case RMSE. Without randomizing the same algorithm is sub-optimal; it will achieve only $\mathcal{O}(n^{-\alpha})$ in the sense of the worst-case RMSE, while the lower bound we establish in this paper is $n^{-\alpha-1/2}$ up to a constant factor. We also note that the worst-case deterministic error for the standard, deterministic, Gauss–Hermite quadrature in \mathcal{H}_α is bounded from below and above by $n^{-\alpha/2}$ up to some constant factors (note also that, from $e^{\text{rmse}}(A_m^{\text{det}}, \mathcal{H}_\alpha) = e^{\text{wor}}(A_m^{\text{det}}, \mathcal{H}_\alpha)$ for deterministic algorithms, these matching bounds of the rate $n^{-\alpha/2}$ also apply to the worst-case RMSE for the standard deterministic Gauss–Hermite quadrature). This rate is merely half of the best possible rate; see [8] together with [9].

Our idea for randomization comes from a classical work of Bakhvalov [2], who introduced the method of choosing the number of quadrature nodes randomly. His method has been explored further quite recently in the context of quasi-Monte Carlo integration over the high-dimensional unit cube [7, 11]. Another related work is by Wu [25], who considered a randomized trapezoidal rule similar to ours. Compared to our methods, Wu used a fixed number of nodes; the function space considered there was the fractional Sobolev spaces of order less than 2 over a finite interval, for which the author showed that the convergence rate in L^p ($p \in [2, \infty)$), where the p -integrability is with respect to the underlying probability measure, improves the standard convergence rate by 1/2. In contrast, our results show that, for the function classes we consider, we are able to exploit smoothness more than 2 and improve the L^2 -convergence rate by 1/2. See also Remark 3.10 for a related discussion.

We start by showing a general lower bound for the worst-case RMSE in Section 2. The lower bound established in this section shows the best possible rate. In Section 3, we show that this rate, up to a logarithmic factor, is achieved by a randomized trapezoidal rule. Numerical results in Section 4 support our theory.

2. A GENERAL LOWER BOUND

In this section, we prove the following general lower bound on $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$ that holds for any randomized quadrature rule A with its cardinality $\#(A)$ at most n .

Theorem 2.1 (General lower bound on the worst-case RMSE). *Let integers $\alpha \geq 1$ and $n \geq 1$ be given. For any randomized quadrature rule A , possibly adaptive or nonlinear, as in Definition 1.1 with $\#(A) \leq n$, the inequality*

$$e^{\text{rmse}}(A, \mathcal{H}_\alpha) \geq \frac{c_\alpha}{n^{\alpha+1/2}}$$

holds, where the constant $c_\alpha > 0$ depends on α but is independent of n and A .

To show this result, we first construct fooling functions, which are crucial for our proof. For $n \in \mathbb{N}$ given, consider $10n$ intervals

$$(4) \quad \iota_j := (\xi_{j-1}, \xi_j) := ((j-1)/n, j/n), \quad j = -5n+1, \dots, 5n.$$

By construction, ι_j , $j = -5n+1, \dots, 5n$ are disjoint. For each j , we construct a function $f_j \in \mathcal{H}_\alpha$ supported on ι_j that integrates to a large value, still having a unit Sobolev norm.

Lemma 2.2 (Fooling function and its integral). *Let integers $\alpha \geq 1$ and $n \geq 1$ be given. Then, for each $j = -5n + 1, \dots, 5n$, there exists a function $f_j \in \mathcal{H}_\alpha$ such that $f_j \geq 0$, $\text{supp}(f_j) = \iota_j$, $\|f_j\|_\alpha = 1$, and*

$$I(f_j) \geq \frac{\eta_\alpha}{n^{\alpha+1/2}},$$

where $\eta_\alpha > 0$ is a constant depending on α but independent of n , j and f_j . Here, $\text{supp}(f_j)$ denotes the set-theoretical support $\text{supp}(f_j) = \{x \in \mathbb{R} \mid f_j \neq 0\}$.

Proof. Define $h_j : \mathbb{R} \rightarrow \mathbb{R}$ such that $h_j \geq 0$ and $\text{supp}(h_j) = \iota_j$ by

$$h_j(x) = \begin{cases} \left(\frac{x - \xi_{j-1}}{\xi_j - \xi_{j-1}} \right)^\alpha \left(1 - \frac{x - \xi_{j-1}}{\xi_j - \xi_{j-1}} \right)^\alpha & \text{if } x \in \iota_j, \\ 0 & \text{otherwise.} \end{cases}$$

As we showed in [9, Proof of Lemma 3.1], we have $h_j \in \mathcal{H}_\alpha$. Moreover, with

$$S_{\alpha,\tau} := \sum_{\ell_1, \ell_2=0}^{\alpha} (-1)^{\ell_1+\ell_2} \binom{\alpha}{\ell_1} \binom{\alpha}{\ell_2} \frac{(\alpha + \ell_1)!}{(\alpha + \ell_1 - \tau)!} \frac{(\alpha + \ell_2)!}{(\alpha + \ell_2 - \tau)!},$$

we have

$$\begin{aligned} \|h_j\|_\alpha^2 &= \sum_{\tau=0}^{\alpha} \int_{\iota_j} |h_j^{(\tau)}(x)|^2 \rho(x) dx \leq \sum_{\tau=0}^{\alpha} \frac{e^{-\min(\xi_{j-1}^2, \xi_j^2)/2}}{\sqrt{2\pi}} \int_{\iota_j} |h_j^{(\tau)}(x)|^2 dx \\ &\leq \sum_{\tau=0}^{\alpha} \frac{S_{\alpha,\tau}}{\sqrt{2\pi}(\xi_j - \xi_{j-1})^{2\tau-1}} = \sum_{\tau=0}^{\alpha} \frac{n^{2\tau-1} S_{\alpha,\tau}}{\sqrt{2\pi}} \leq n^{2\alpha-1} \sum_{\tau=0}^{\alpha} \frac{S_{\alpha,\tau}}{\sqrt{2\pi}}, \end{aligned}$$

and

$$\begin{aligned} I(h_j) &= \int_{\iota_j} h_j(x) \rho(x) dx \geq \frac{e^{-\max(\xi_{j-1}^2, \xi_j^2)/2}}{\sqrt{2\pi}} \int_{\iota_j} h_j(x) dx \\ &\geq \frac{e^{-25/2}}{\sqrt{2\pi}} (\xi_j - \xi_{j-1}) \int_0^1 x^\alpha (1-x)^\alpha dx = \frac{(\alpha!)^2}{n(2\alpha+1)! \sqrt{2\pi} e^{25}}. \end{aligned}$$

It follows that, with $f_j := h_j / \|h_j\|_\alpha$, we have $f_j \in \mathcal{H}_\alpha$, $f_j \geq 0$, $\text{supp}(f_j) = \iota_j$, $\|f_j\|_\alpha = 1$ and

$$I(f_j) = \frac{I(h_j)}{\|h_j\|_\alpha} \geq \frac{(\alpha!)^2}{n^{\alpha+1/2} (2\alpha+1)! \sqrt{2\pi} e^{25}} \left(\sum_{\tau=0}^{\alpha} \frac{S_{\alpha,\tau}}{\sqrt{2\pi}} \right)^{-1/2} =: \frac{\eta_\alpha}{n^{\alpha+1/2}},$$

for $j = -5n + 1, \dots, 5n$. This completes the proof. \square

We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. Let f_j , $j = -5n + 1, \dots, 5n$, be as in Lemma 2.2. By construction, we have $\|f_j\|_\alpha = 1$. Hence, the definition of $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$ implies

$$\begin{aligned} &e^{\text{rmse}}(A, \mathcal{H}_\alpha) \\ &\geq \max_{f \in \{\pm f_j \mid j = -5n+1, \dots, 5n\}} \left(\int_{\Omega} (I(f) - A^\omega(f))^2 d\mu(\omega) \right)^{1/2} \\ &\geq \left(\frac{1}{10n} \sum_{j=-5n+1}^{5n} \int_{\Omega} \frac{(I(f_j) - A^\omega(f_j))^2 + (I(-f_j) - A^\omega(-f_j))^2}{2} d\mu(\omega) \right)^{1/2} \end{aligned}$$

$$(5) = \left(\frac{1}{10n} \int_{\Omega} \sum_{j=-5n+1}^{5n} \frac{(I(f_j) - A^{\omega}(f_j))^2 + (I(-f_j) - A^{\omega}(-f_j))^2}{2} d\mu(\omega) \right)^{1/2}.$$

We will show that at least n out of $10n$ functions f_j , $j = -5n + 1, \dots, 5n$, satisfy $(I(f_j) - A^{\omega}(f_j))^2 + (I(-f_j) - A^{\omega}(-f_j))^2 \geq 2\eta_{\alpha}^2/n^{2\alpha+1}$ on a set of positive measure $E \subset \Omega$. Here, the choice of these n functions may depend on $\omega \in E$.

Define $E \in \Sigma$, where Σ is the σ -algebra as in Definition 1.1, by

$$E := \left\{ \omega \in \Omega \mid \frac{1}{10n} \sum_{j=-5n+1}^{5n} M(f_j, \omega) < 2n \right\}.$$

Moreover, for $\omega \in E$ denote by F_0^{ω} the subset of f_j 's whose corresponding quadrature $A^{\omega}(f_j)$ uses at most $4n$ nodes: with

$$\text{Ind}_0^{\omega} := \{j \in \{-5n+1, \dots, 5n\} \mid M(f_j, \omega) \leq 4n\},$$

let

$$F_0^{\omega} := \{f_j \in \{f_{-5n+1}, \dots, f_{5n}\} \mid j \in \text{Ind}_0^{\omega}\}.$$

For $i = 1, 2, \dots, 4n$, we recursively define

$$\text{Ind}_i^{\omega} := \{j \in \text{Ind}_{i-1}^{\omega} \mid i \leq M(f_j, \omega), x_i^{(j)} \notin \iota_j\} \cup \{j \in \text{Ind}_{i-1}^{\omega} \mid M(f_j, \omega) < i\},$$

where $x_i^{(j)} = x_i^{(j)}(\omega)$ denotes the i -th node of $A^{\omega}(f_j)$ and ι_j is defined in (4), and

$$F_i^{\omega} := \{f_j \in F_{i-1}^{\omega} \mid j \in \text{Ind}_i^{\omega}\}.$$

F_i^{ω} is constructed so that if $i \leq M(f_j, \omega)$, i.e., when the i -th node exists, then $f_j \in F_i^{\omega}$ implies $f_j(x_i^{(j)}) = 0$, and if $M(f_j, \omega) < i$, i.e., when the i -th node does not exist, then we keep $f_j \in F_{i-1}^{\omega}$ in F_i^{ω} . We will show $|F_{4n}^{\omega}| \geq n$ for $\omega \in E$ and that F_{4n}^{ω} is the sought set that gives the aforementioned lower bound.

First, we show that E has a positive measure. By the definition of the cardinality of a randomized quadrature rule, it holds that

$$\begin{aligned} \int_{\Omega} \frac{1}{10n} \sum_{j=-5n+1}^{5n} M(f_j, \omega) d\mu(\omega) &\leq \sup_{j=-5n+1, \dots, 5n} \int_{\Omega} M(f_j, \omega) d\mu(\omega) \\ &\leq \sup_{f \in \mathcal{H}_{\alpha}} \int_{\Omega} M(f, \omega) d\mu(\omega) \leq n. \end{aligned}$$

With $E^c := \Omega \setminus E$, it follows from Markov's inequality that

$$\mu(E) = 1 - \mu(E^c) \geq 1 - \frac{1}{2n} \int_{\Omega} \frac{1}{10n} \sum_{j=-5n+1}^{5n} M(f_j, \omega) d\mu(\omega) \geq \frac{1}{2}.$$

For $\omega \in E$ it turns out $|F_0^{\omega}| \geq 5n$, i.e., for at least half of the $10n$ functions f_j , $j = -5n + 1, \dots, 5n$, the integral estimate $A^{\omega}(f_j)$ is computed by no more than $4n$ nodes. Indeed, from Markov's inequality for the uniform measure ν on $(\{-5n+1, \dots, 5n\}, 2^{\{-5n+1, \dots, 5n\}})$ and the corresponding expectation \mathbb{E}_{ν} , together with the definition of E we see that the number of f_j 's whose corresponding nodes exceed $4n$ is less than half:

$$\frac{|\{j \in \{-5n+1, \dots, 5n\} \mid M(f_j, \omega) > 4n\}|}{10n} = \nu(\{j \mid M(f_j, \omega) > 4n\})$$

$$\leq \frac{1}{4n} \mathbb{E}_\nu[M(f, \omega)] = \frac{1}{4n} \cdot \frac{1}{10n} \sum_{j=-5n+1}^{5n} M(f_j, \omega) < \frac{1}{2}.$$

Next, we will show $|F_i^\omega| \geq |F_{i-1}^\omega| - 1$ for $i = 1, \dots, 4n$. For this, we will show that, given $\omega \in E$, the first i quadrature nodes of $A^\omega(f_j)$ for $f_j \in F_i^\omega$ are uniquely determined, independently of $f_j \in F_i^\omega$.

To see this, first fix $f_j \in F_i^\omega \subset F_{i-1}^\omega$ arbitrarily. By construction of F_i^ω , the interval ι_j does not contain any of the first $i-1$ nodes $x_1^{(j)}, \dots, x_{i-1}^{(j)}$ of $A^\omega(f_j)$. Although the next node $x_i^{(j)}$ is (possibly adaptively) determined by the set of $x_1^{(j)}, \dots, x_{i-1}^{(j)}$ and $f_j(x_1^{(j)}), \dots, f_j(x_{i-1}^{(j)})$, these function values are all equal to 0. Now recall that the first node $x_1^{(j)} = x_1$ of $A^\omega(f_j)$ does not depend on f_j , since initially we do not have any information on f_j . Therefore, by the recursive definition of F_i^ω , we see that for $f_j \in F_i^\omega$ the first i quadrature nodes of $A^\omega(f_j)$ are all equal for $f_j \in F_i^\omega$, independently of the corresponding index j . Indeed, for $f_j, f_{j'} \in F_i^\omega$, we have

$$\begin{aligned} x_\ell^{(j)} &= \psi_\ell(x_1^{(j)}, x_2^{(j)}, \dots, x_{\ell-1}^{(j)}, f_j(x_1^{(j)}), \dots, f_j(x_{\ell-1}^{(j)})) \\ &= \psi_\ell(x_1, x_2^{(j)}, \dots, x_{\ell-1}^{(j)}, 0, \dots, 0) \\ &= \psi_\ell(x_1, x_2^{(j')}, \dots, x_{\ell-1}^{(j')}, 0, \dots, 0) = x_\ell^{(j')} \quad \text{for } \ell = 2, \dots, \min\{i, M(f_j, \omega)\}, \end{aligned}$$

and if $\min\{i, M(f_j, \omega)\} = M(f_j, \omega)$ then for $\ell = M(f_j, \omega) + 1, \dots, i$ the point $x_\ell^{(j)}$ does not exist. Here, note that if $M(f_j, \omega) < i$ then $M(f_j, \omega) = M(f_{j'}, \omega)$, and if $i \leq M(f_j, \omega)$ then $i \leq M(f_{j'}, \omega)$. Indeed, for $\ell = 1, \dots, i$, the Boolean function (2) satisfies

$$\begin{aligned} &\text{ter}_\ell(x_1^{(j')}, x_2^{(j')}, \dots, x_\ell^{(j')}, f_{j'}(x_1^{(j')}), \dots, f_{j'}(x_\ell^{(j')})) \\ &= \text{ter}_\ell(x_1^{(j)}, x_2^{(j)}, \dots, x_\ell^{(j)}, 0, \dots, 0) \\ &= \text{ter}_\ell(x_1^{(j)}, x_2^{(j)}, \dots, x_\ell^{(j)}, f_j(x_1^{(j)}), \dots, f_j(x_\ell^{(j)})). \end{aligned}$$

Now, for $f_j \in F_{i-1}^\omega$, we have $f_j \notin F_i^\omega$ if and only if $x_i^{(j)} \in \iota_j$. Since the i -th node $x_i^{(j)}$ can be contained in at most one interval $\iota_{\bar{j}} = ((\bar{j}-1)/n, \bar{j}/n)$ among the supports $(\iota_k)_k$ corresponding to $(f_k)_k \subset F_{i-1}^\omega$, the above argument leads to $|\{f_j \in F_{i-1}^\omega \mid f_j \notin F_i^\omega\}| \leq 1$, and thus

$$|F_i^\omega| \geq |F_{i-1}^\omega| - 1.$$

Therefore, $|F_{4n}^\omega| \geq |F_0^\omega| - 4n \geq n$ holds.

Now, from $F_{4n}^\omega \subset F_0^\omega$, for $f_j \in F_{4n}^\omega$ no quadrature point of $A^\omega(f_j)$ is in the support of f_j . Therefore, $|F_{4n}^\omega| \geq n$ means that among $10n$ functions f_j , $j = -5n+1, \dots, 5n$, at least n of them have no quadrature node of $\{A^\omega(f_j)\}_{j=-5n+1, \dots, 5n}$ in their supports. Hence, for these (at least) n functions, the function values used in $A^\omega(f_j)$ are all 0, which also holds true for $-f_j$. Therefore, for these f_j 's, we obtain the equality $A^\omega(f_j) = A^\omega(0) = A^\omega(-f_j)$ and

$$\begin{aligned} &(I(f_j) - A^\omega(f_j))^2 + (I(-f_j) - A^\omega(-f_j))^2 \\ &= (I(f_j) - A^\omega(f_j))^2 + (-I(f_j) - A^\omega(f_j))^2 \\ &= 2(I(f_j))^2 + 2(A^\omega(f_j))^2 \geq 2(I(f_j))^2 \geq \frac{2\eta_\alpha^2}{n^{2\alpha+1}}. \end{aligned}$$

For the rest of the (at most $4n$) functions, $(I(f_j) - A^\omega(f_j))^2 + (I(-f_j) - A^\omega(-f_j))^2$ is trivially bounded below by 0.

Applying the results to (5), we obtain

$$e^{\text{rmse}}(A, \mathcal{H}_\alpha) \geq \left(\mu(E) \cdot \left(\frac{n}{10n} \cdot \frac{\eta_\alpha^2}{n^{2\alpha+1}} + \frac{9n}{10n} \cdot 0 \right) \right)^{1/2} \geq \frac{\eta_\alpha}{2\sqrt{5}n^{\alpha+1/2}},$$

which completes the proof of the theorem. \square

Remark 2.3. *The idea of proving a lower bound on the randomized error by the average-case error over bump functions with mutually disjoint supports goes back to Bakhvalov [1]. We also refer to [14, Chapter 2] and [17, Chapter 17] for more information on error analyses for the randomized setting.*

3. RANDOMIZED TRAPEZOIDAL RULE

Here we develop a randomized quadrature rule A that achieves the almost optimal rate of convergence in the sense of the worst-case RMSE $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$. This algorithm is based on the suitably truncated deterministic trapezoidal rule we considered in [9], which is given by

$$A_{n,T}^* = \frac{2T}{n} \sum_{j=0}^{n-1} f(\xi_j^*) \rho(\xi_j^*) \quad \text{with} \quad \xi_j^* = T \left(\frac{2j}{n} - 1 \right), j = 0, \dots, n-1.$$

Here, n denotes the number of nodes and $T > 0$ is a parameter that controls the cut-off of the integration domain from \mathbb{R} to $[-T, T]$. As shown in [9], $A_{n,T}^*$ turns out to achieve, up to a logarithmic factor, the optimal rate of convergence in the sense of the worst-case error in \mathcal{H}_α , for any integer $\alpha \geq 1$.

In this paper, we introduce its randomized counterpart that makes $e^{\text{rmse}}(A, \mathcal{H}_\alpha)$ small. In what follows, we denote the cumulative distribution function of the standard Gaussian distribution by Φ .

Algorithm 3.1. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $n \geq 4$ and $T > 0$ be given. To approximate the integral $I(f)$, with independent random variables $M_\star \sim U\{\lfloor n/2 \rfloor, \dots, n-2\}$ and $\delta \sim U(0, 1)$, we define the randomized trapezoidal rule $A_{n,T} = (A_{n,T}^{M_\star, \delta})_{M_\star, \delta}$ by*

$$(6) \quad A_{n,T}^{M_\star, \delta} = \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} f(\xi_j^*) \rho(\xi_j^*) + \Phi(-T) f(\xi_{\text{left}}^*) + (1 - \Phi(T)) f(\xi_{\text{right}}^*),$$

where we let

$$\xi_j^* := T \left(\frac{2(j + \delta)}{M_\star} - 1 \right), \quad j = 0, \dots, M_\star - 1,$$

and the end nodes ξ_{left}^* and ξ_{right}^* are independent of M_\star and sampled from the truncated normal distributions on $(-\infty, -T]$ and $[T, \infty)$, respectively.

Each node ξ_j in the first term of (6) is uniformly distributed over the interval $(T(2j/M_\star - 1), T(2(j+1)/M_\star - 1))$, $j = 0, \dots, M_\star - 1$. For ξ_{left}^* and ξ_{right}^* , we specify only their marginal distributions, truncated normals. Any random variables ξ_{left}^* and ξ_{right}^* having the marginals specified above that are independent of M_\star , in particular those that are not independent of ξ_j 's, allow the results in this paper to hold. For instance, one straightforward implementation is to use the inversion method, i.e., $\xi_{\text{left}}^* := \Phi^{-1}(\delta\Phi(-T))$, and $\xi_{\text{right}}^* := \Phi^{-1}((1 - \delta)\Phi(T) + \delta)$, where δ is as in Algorithm 3.1. As such, we omit ξ_{left}^* and ξ_{right}^* from the notation $A_{n,T}^{M_\star, \delta}$.

However, in view of implementation, sampling from such truncated distributions by the inversion method can be numerically unstable and inefficient when T is large; see [4]. As pointed out therein, one needs to be careful when using the inversion method for the right tail of the distribution. Hence, it can be more convenient to sample them independently by other methods. We refer to [5, Chapter 9.1] and [3, 4] among many others for random number generation from the tail of normal distributions.

In Algorithm 3.1, the cut-off parameter $T > 0$ should be carefully chosen depending on n (and α). This point will be discussed later in this section. The symmetry of the standard Gaussian distribution around 0 implies $\Phi(-T) = 1 - \Phi(T)$ for any $T > 0$, i.e., the function values at ξ_{left}^* and ξ_{right}^* are weighted equally in $A_{n,T}^{M_*,\delta}$.

We emphasize that $A_{n,T}$ is a randomized non-adaptive linear quadrature rule; M_* , δ , ξ_{left}^* and ξ_{right}^* are the only ω -dependent components of $A_{n,T}$ and neither depends on the integrand f . Furthermore, we have $\#(A_{n,T}) \leq n$. This is because $M_* \sim U\{\lfloor n/2 \rfloor, \dots, n-2\}$, and moreover, for each realization of M_* and δ , the deterministic rule $A_{n,T}^{M_*,\delta}$ utilizes M_*+2 nodes: ξ_{left}^* , ξ_{right}^* , and ξ_j^* , $j = 0, \dots, M_*-1$. Hence, our result in Theorem 3.3 below shows that adaptivity and nonlinearity are not needed to achieve the optimality (up to a logarithmic factor) in the sense that is considered here (cf. the general lower bound in Theorem 2.1).

3.1. Unbiasedness and error estimation. Before proving an upper bound on $e^{\text{rmse}}(A_{n,T}, \mathcal{H}_\alpha)$, we show that our randomized quadrature rule is unbiased and allows for practical error estimation. Let us start by showing the unbiasedness of $A_{n,T}$.

Lemma 3.2 (Unbiasedness). *Let $\alpha \geq 1$, $n \geq 4$, and $T > 0$ be given. For any $f \in \mathcal{H}_\alpha$, we have*

$$\mathbb{E} \left[A_{n,T}^{M_*,\delta}(f) \right] := \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{m_*=\lfloor n/2 \rfloor}^{n-2} \int_0^1 A_{n,T}^{m_*,\delta}(f) \, d\delta = I(f).$$

Proof. Substituting (6) into $\mathbb{E} \left[A_{n,T}^{M_*,\delta}(f) \right]$, we have

$$\begin{aligned} \mathbb{E} \left[A_{n,T}^{M_*,\delta}(f) \right] &= \mathbb{E} \left[\frac{2T}{M_*} \sum_{j=0}^{M_*-1} f(\xi_j^*) \rho(\xi_j^*) + \Phi(-T) f(\xi_{\text{left}}^*) + (1 - \Phi(T)) f(\xi_{\text{right}}^*) \right] \\ &= \mathbb{E} \left[\frac{2T}{M_*} \sum_{j=0}^{M_*-1} f(\xi_j^*) \rho(\xi_j^*) \right] \\ &\quad + \Phi(-T) \mathbb{E} [f(\xi_{\text{left}}^*)] + (1 - \Phi(T)) \mathbb{E} [f(\xi_{\text{right}}^*)]. \end{aligned}$$

For the first term on the rightmost side above, we have

$$\begin{aligned} &\mathbb{E} \left[\frac{2T}{M_*} \sum_{j=0}^{M_*-1} f(\xi_j^*) \rho(\xi_j^*) \right] \\ &= \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{m_*=\lfloor n/2 \rfloor}^{n-2} \int_0^1 \left(\frac{2T}{m_*} \sum_{j=0}^{m_*-1} f(\xi_j^*) \rho(\xi_j^*) \right) \, d\delta \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \frac{2T}{m_\star} \sum_{j=0}^{m_\star-1} \int_0^1 f(\xi_j^*) \rho(\xi_j^*) d\delta \\
&= \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \sum_{j=0}^{m_\star-1} \int_{T(2j/m_\star-1)}^{T(2(j+1)/m_\star-1)} f(x) \rho(x) dx \\
&= \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \int_{-T}^T f(x) \rho(x) dx = \int_{-T}^T f(x) \rho(x) dx,
\end{aligned}$$

where the third equality follows from the change of variables $x = T(2(j+\delta)/m_\star - 1)$ for each $j = 0, \dots, m_\star - 1$. Regarding the second term, since ξ_{left}^* is independent of M_\star , we have

$$\Phi(-T) \mathbb{E}[f(\xi_{\text{left}}^*)] = \int_{-\infty}^{-T} f(x) \rho(x) dx.$$

In a similar way, for the third term it can be shown that

$$(1 - \Phi(T)) \mathbb{E}[f(\xi_{\text{right}}^*)] = \int_T^{\infty} f(x) \rho(x) dx.$$

Altogether we obtain

$$\begin{aligned}
\mathbb{E}[A_{n,T}^{M_\star, \delta}(f)] &= \int_{-T}^T f(x) \rho(x) dx + \int_{-\infty}^{-T} f(x) \rho(x) dx + \int_T^{\infty} f(x) \rho(x) dx \\
&= \int_{-\infty}^{\infty} f(x) \rho(x) dx = I(f),
\end{aligned}$$

which proves the statement. \square

In practice, we use the approximation

$$\overline{A_{n,T}(f)} := \frac{1}{r} \sum_{i=1}^r A_{n,T}^{M_\star^{(i)}, \delta^{(i)}}(f) \approx I(f),$$

with $r \in \mathbb{N}$ independent random variables

$$\begin{aligned}
&(M_\star^{(1)}, \delta^{(1)}, \xi_{\text{left}}^{*,(1)}, \xi_{\text{right}}^{*,(1)}), \dots, (M_\star^{(r)}, \delta^{(r)}, \xi_{\text{left}}^{*,(r)}, \xi_{\text{right}}^{*,(r)}) \\
&\in \{\lfloor n/2 \rfloor, \dots, n-2\} \times (0, 1) \times (-\infty, -T] \times [T, \infty),
\end{aligned}$$

which can be easily generated. Lemma 3.2 yields an error bound and a practical error estimator for this algorithm.

From Lemma 3.2, $A_{n,T}^{M_\star^{(i)}, \delta^{(i)}}(f)$, $i = 1, \dots, r$, are independent and unbiased estimators of $I(f)$. Thus $\overline{A_{n,T}(f)}$ is also an unbiased estimator of $I(f)$, and thus

$$(7) \quad \mathbb{E} \left[\left(\overline{A_{n,T}(f)} - I(f) \right)^2 \right] = \frac{1}{r} \mathbb{E} \left[\left(A_{n,T}^{M_\star, \delta}(f) - I(f) \right)^2 \right],$$

and

$$(8) \quad \mathbb{E} \left[\left(\overline{A_{n,T}(f)} - I(f) \right)^2 \right] = \mathbb{E} \left[\left(\overline{A_{n,T}(f)} - \mathbb{E} \left[\overline{A_{n,T}(f)} \right] \right)^2 \right].$$

As we shall show in the proof of Theorem 3.3, with a suitable $T > 0$, for any $f \in \mathcal{H}_\alpha$ the mean-squared error can be bounded as $\mathbb{E}[(A_{n,T}^{M_\star, \delta}(f) - I(f))^2] \leq$

$C_{\alpha,\lambda}^2 \|f\|_\alpha^2 \frac{(\ln n)^{\alpha+1/2}}{n^{2\alpha+1}}$, where $C_{\alpha,\lambda} > 0$ is a constant depending on α and λ but independent of f , and thus (7) implies the error bound

$$\mathbb{E} \left[\left(\overline{A_{n,T}(f)} - I(f) \right)^2 \right] \leq C_{\alpha,\lambda}^2 \|f\|_\alpha^2 \frac{1}{r} \frac{(\ln n)^{\alpha+1/2}}{n^{2\alpha+1}}.$$

From (8), the sample variance

$$(9) \quad \frac{1}{r(r-1)} \sum_{i=1}^r \left(A_{n,T}^{M_\star^{(i)}, \delta^{(i)}}(f) - \overline{A_{n,T}(f)} \right)^2$$

can be used as an online, unbiased mean-squared error estimator.

3.2. A bound on the root-mean-square error. As the second main result of this paper, we show an upper bound for the worst-case RMSE $e^{\text{rmse}}(A_{n,T}, \mathcal{H}_\alpha)$:

Theorem 3.3 (Upper bound on the worst-case RMSE). *Let $\alpha \geq 1$ and $n \geq 4$ be given. Fix $\lambda \in (1/2, 1)$ arbitrarily. Then, with*

$$(10) \quad T = \sqrt{\frac{2\alpha+1}{1-\lambda} \ln(n)},$$

$A_{n,T}$ as in Algorithm 3.1 satisfies

$$e^{\text{rmse}}(A_{n,T}, \mathcal{H}_\alpha) \leq C_{\alpha,\lambda} \frac{(\ln n)^{\alpha/2+1/4}}{n^{\alpha+1/2}},$$

where $C_{\alpha,\lambda} > 0$ is a constant depending only on α and λ .

This rate is optimal up to a logarithmic factor in the sense of the worst-case RMSE; indeed, Theorem 2.1 shows that the optimal rate is $\mathcal{O}(n^{-\alpha-1/2})$. To prove this theorem, first we need some preparations.

Remark 3.4. *In Theorem 3.3, the parameter $\lambda \in (1/2, 1)$ does not affect the convergence rate, including the logarithmic factor. However, it does change the constant $C_{\alpha,\lambda}$. This change is caused by changes of (i) the choice of the cut-off parameter T and (ii) the constant $C_{\alpha,\lambda}^*$, both of which depend on λ ; see (16) and (17) below. For increasing λ , T increases but $C_{\alpha,\lambda}^*$ decreases. It is not very straightforward to find the optimal λ minimizing the constant, and this is out of the scope of the present paper.*

3.2.1. Auxiliary results. Because our algorithm is based on the trapezoidal rule yet the integrand is not assumed to be periodic, error analysis requires extra work.

First, we introduce the following notations for an α -times weakly differentiable function $F: \mathbb{R} \rightarrow \mathbb{R}$. Let $F^{(\tau)}$ denote its τ -th weak derivative for $\tau = 0, \dots, \alpha$. Define

$$\|F\|_\alpha^* := \sup_{\substack{I \subset \mathbb{R} \\ |I| < \infty}} \|F\|_{\alpha,I}^*$$

with

$$\|F\|_{\alpha,I}^* := \left(\sum_{\tau=0}^{\alpha-1} \left(\int_I F^{(\tau)}(x) dx \right)^2 + \int_I |F^{(\alpha)}(x)|^2 dx \right)^{1/2}$$

and for $\lambda \in (1/2, 1)$,

$$\|F\|_{\alpha,\lambda,\text{decay}} := \sup_{x \in \mathbb{R}, \tau \in \{0, \dots, \alpha-1\}} \left| e^{(1-\lambda)x^2/2} F^{(\tau)}(x) \right|,$$

allowing this quantity to be infinite. It turns out that $f \in \mathcal{H}_\alpha$ implies $\|f\rho\|_\alpha^* < \infty$, and $\|f\rho\|_{\alpha,\lambda,\text{decay}} < \infty$, as we showed in [9, Proof of Theorem 4.5].

Lemma 3.5 (Norm inequality). *Let $\alpha \geq 1$ and $\lambda \in (1/2, 1)$ be given. For $f \in \mathcal{H}_\alpha$, let $F(x) := f(x)\rho(x)$. Then, we have*

$$\|F\|_\alpha^* + \|F\|_{\alpha,\lambda,\text{decay}} \leq C_{\alpha,\lambda}^* \|f\|_\alpha$$

with a constant $C_{\alpha,\lambda}^* > 0$ depending on α and λ but independent of f .

Moreover, following [9] we use the following periodization (11) as an auxiliary variable. We refer to [9, Lemma 4.1] for a proof.

Lemma 3.6 (Auxiliary periodic function and its properties). *Let $\alpha \geq 1$ and $T > 0$ be given. For $f \in \mathcal{H}_\alpha$, let $F(x) := f(x)\rho(x)$ and define $G : [-T-d, T+d] \rightarrow \mathbb{R}$ by*

$$(11) \quad G(x) := F(x) - \sum_{\tau=1}^{\alpha} \frac{B_\tau^{[-T,T]}(x)}{\tau!} \left(\int_{-T}^T F^{(\tau)}(y) dy \right),$$

for $x \in [-T-d, T+d]$, with an arbitrarily small $d > 0$, where $B_\tau^{[-T,T]}$ denotes the scaled Bernoulli polynomial of degree τ on $[-T, T]$, i.e.,

$$B_\tau^{[-T,T]}(x) = (2T)^{\tau-1} B_\tau \left(\frac{x+T}{2T} \right)$$

with B_τ being the standard Bernoulli polynomial of degree τ . Then, the following holds:

- (1) The function G preserves the integral of F on $[-T, T]$:

$$\int_{-T}^T G(x) dx = \int_{-T}^T F(x) dx = \int_{-T}^T f(x)\rho(x) dx.$$

- (2) The function G is $(\alpha-1)$ -times continuously differentiable on $(-T-d, T+d)$ with $G^{(\alpha-1)}$ being absolutely continuous on $[-T, T]$, and satisfies $G^{(\tau)}(-T) = G^{(\tau)}(T)$ for all $\tau = 0, \dots, \alpha-1$.
- (3) The norm of G is bounded above by that of F :

$$\|G\|_{\alpha,[-T,T]}^* \leq \|F\|_{\alpha,[-T,T]}^*,$$

where we note that $\|G\|_{\alpha,[-T,T]}^*$ is well defined, since the continuous differentiability and the absolute continuity above implies the weak differentiability of $G^{(\tau)}$ on $[-T, T]$ for all $\tau \leq \alpha-1$.

The functions defined above will be used to analyze the integration error on $[-T, T]$. In the proof, we decompose the integrand $f(x)\rho(x)$ into two parts; the periodic part which is represented by $G(x)$, and the non-periodic part. We use $\|f\rho\|_\alpha^*$ to control the integration error of the periodic part, while the error of the non-periodic part is controlled by $\|f\rho\|_{\alpha,\lambda,\text{decay}}$, as can be seen in the proof of Lemma 3.8 below. The cut-off parameter T in (10) is given in such a way that these two errors are balanced. To control the error on the tail regions $(-\infty, -T]$ and $[T, \infty)$, we only need $\|f\|_{L_p^2}$, as shown in the proof of Lemma 3.9 below.

Remark 3.7. *The method of periodizing a function by adding Bernoulli polynomials in (11) was first introduced by Korobov [10], and also studied by Zaremba [26], in the context of numerical integration. This method was originally proposed for*

numerically integrating non-periodic functions on a multidimensional box. This periodization is referred to as *Bernoulli polynomial method* in [20, Section 2.12]. Recently, a new proof strategy was developed for a class of quasi-Monte Carlo methods in [18], where this periodized function was used as an auxiliary function only appearing in a proof, but not in the algorithm. In particular, [18] showed that $G(x)$ is an orthogonal projection of $F(x)$ in a suitable sense, from one kind of Sobolev space to a corresponding periodic Sobolev space, in a multidimensional setting. This implies the norm inequality shown in Lemma 3.6.3. In [9] the authors applied this proof strategy for the Gaussian Sobolev spaces.

3.2.2. *Proof of Theorem 3.3.* Let us prove Theorem 3.3. It follows from the proof of Lemma 3.2 and Jensen's inequality that, for any $f \in \mathcal{H}_\alpha$, the mean-squared error of $A_{n,T}(f)$ is bounded above as

$$\begin{aligned}
& \mathbb{E} \left[\left(A_{n,T}^{M_\star, \delta}(f) - I(f) \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} f(\xi_j^*) \rho(\xi_j^*) + \Phi(-T)f(\xi_{\text{left}}^*) + (1 - \Phi(T))f(\xi_{\text{right}}^*) \right. \right. \\
&\quad \left. \left. - \int_{-T}^T f(x)\rho(x) dx - \int_{-\infty}^{-T} f(x)\rho(x) dx - \int_T^{\infty} f(x)\rho(x) dx \right)^2 \right] \\
&\leq 3 \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} f(\xi_j^*) \rho(\xi_j^*) - \int_{-T}^T f(x)\rho(x) dx \right)^2 \right] \\
(12) \quad &+ 3 \mathbb{E} \left[\left(\Phi(-T)f(\xi_{\text{left}}^*) - \int_{-\infty}^{-T} f(x)\rho(x) dx \right)^2 \right] \\
&+ 3 \mathbb{E} \left[\left((1 - \Phi(T))f(\xi_{\text{right}}^*) - \int_T^{\infty} f(x)\rho(x) dx \right)^2 \right].
\end{aligned}$$

Thus it suffices to give an upper bound on each term of (12). In Lemma 3.8 below, we show an upper bound on the first term. And then in Lemma 3.9, we show upper bounds on the second and third terms.

Lemma 3.8 (Upper bound on the RMSE for $[-T, T]$). *Let $\alpha \geq 1$, $n \geq 4$, and $T > 0$ be given. Fix $\lambda \in (1/2, 1)$ arbitrarily. For any $f \in \mathcal{H}_\alpha$, the first term of (12) is bounded as*

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} f(\xi_j^*) \rho(\xi_j^*) - \int_{-T}^T f(x)\rho(x) dx \right)^2 \right] \\
&\leq (C_{\alpha, \lambda}^*)^2 \|f\|_\alpha^2 \left(\frac{2\alpha^2 \max\{1, (2T)^{2\alpha}\}}{e^{(1-\lambda)T^2}} + \frac{2^{2\alpha+5}T^{2\alpha+1}}{\pi^{2\alpha}\eta^{2\alpha+1}} \right),
\end{aligned}$$

with $C_{\alpha, \lambda}^* > 0$ being the same constant as in Lemma 3.5.

Proof. Let us consider the auxiliary periodic function $G : [-T, T] \rightarrow \mathbb{R}$ defined in (11). From Lemma 3.6.1, we have

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} f(\xi_j^*) \rho(\xi_j^*) - \int_{-T}^T f(x) \rho(x) dx \right)^2 \right] \\
&= \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} (F(\xi_j^*) - G(\xi_j^*)) + \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx \right)^2 \right] \\
&\leq 2 \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} (F(\xi_j^*) - G(\xi_j^*)) \right)^2 \right] \\
(13) \quad &+ 2 \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx \right)^2 \right].
\end{aligned}$$

For the first term of (13), it follows from the periodization (11) that, for both M_\star and δ given,

$$\begin{aligned}
& \left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} (F(\xi_j^*) - G(\xi_j^*)) \right)^2 \\
&= 4T^2 \left(\frac{1}{M_\star} \sum_{j=0}^{M_\star-1} \sum_{\tau=1}^{\alpha} \frac{B_\tau^{[-T, T]}(\xi_j^*)}{\tau!} \left(\int_{-T}^T F^{(\tau)}(y) dy \right) \right)^2 \\
&\leq 4T^2 \left(\sum_{\tau=1}^{\alpha} \frac{(2T)^{\tau-1}}{2} \left| F^{(\tau-1)}(T) - F^{(\tau-1)}(-T) \right| \right)^2 \\
&\leq \left(\alpha \max\{1, (2T)^\alpha\} \|F\|_{\alpha, \lambda, \text{decay}} e^{-(1-\lambda)T^2/2} \right)^2 \\
(14) \quad &\leq \alpha^2 (C_{\alpha, \lambda}^*)^2 \|f\|_\alpha^2 \max\{1, (2T)^{2\alpha}\} e^{-(1-\lambda)T^2},
\end{aligned}$$

where we used $|B_\tau^{[-T, T]}(x)/\tau!| \leq (2T)^{\tau-1}/2$ for $x \in [-T, T]$, shown in [12], to get the first inequality, and applied Lemma 3.5 in the last inequality. Since the bound (14) does not depend on the realization of $M_\star(\omega)$ or on that of $\delta(\omega)$, the first term of (13) is bounded above by this bound multiplied by 2.

Let us move on to the second term of (13). Lemma 3.6.2 implies that G has the pointwise-convergent Fourier series

$$G(x) = \sum_{k \in \mathbb{Z}} \widehat{G}(k) \phi_k^{[-T, T]}(x),$$

where $\{\phi_k^{[-T, T]}(x) := \exp(2\pi i k(x+T)/(2T))/\sqrt{2T} \mid k \in \mathbb{Z}\}$ forms an orthonormal $L^2([-T, T])$ basis and $\widehat{G}(k)$ denotes the k -th Fourier coefficient

$$\widehat{G}(k) = \int_{-T}^T G(x) \overline{\phi_k^{[-T, T]}(x)} dx.$$

Besides, it follows from the square integrability of $G^{(\alpha)}$ that $G^{(\alpha)}$ has the L^2 -convergent Fourier series

$$G^{(\alpha)}(x) = \sum_{k \in \mathbb{Z}} \widehat{G^{(\alpha)}}(k) \phi_k^{[-T, T]}(x) = \sum_{k \in \mathbb{Z}} \left(\frac{2\pi i k}{2T} \right)^\alpha \widehat{G}(k) \phi_k^{[-T, T]}(x),$$

see [9, Proof of Lemma 4.1]. Therefore, regarding the second term of (13), for M_\star and δ both given,

$$\begin{aligned} \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx &= \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} \sum_{k \in \mathbb{Z}} \widehat{G}(k) \phi_k^{[-T, T]}(\xi_j^*) - \sqrt{2T} \widehat{G}(0) \\ &= \sum_{k \in \mathbb{Z}} \widehat{G}(k) \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} \phi_k^{[-T, T]}(\xi_j^*) - \sqrt{2T} \widehat{G}(0), \end{aligned}$$

holds, in which, for any $k \in \mathbb{Z}$, we have

$$\begin{aligned} \frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} \phi_k^{[-T, T]}(\xi_j^*) &= \frac{\sqrt{2T}}{M_\star} \sum_{j=0}^{M_\star-1} e^{2\pi i k (\xi_j^* + T)/(2T)} = \frac{\sqrt{2T}}{M_\star} \sum_{j=0}^{M_\star-1} e^{2\pi i k (j+\delta)/M_\star} \\ &= \frac{e^{2\pi i k \delta/M_\star} \sqrt{2T}}{M_\star} \sum_{j=0}^{M_\star-1} e^{2\pi i k (j/M_\star)} \\ (15) \quad &= \begin{cases} e^{2\pi i k \delta/M_\star} \sqrt{2T} & \text{if } k \equiv 0 \pmod{M_\star}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} &\mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx \right)^2 \right] \\ &= \mathbb{E} \left[\left(\sum_{\substack{k \in \mathbb{Z} \setminus \{0\} \\ k \equiv 0 \pmod{M_\star}}} e^{2\pi i k \delta/M_\star} \sqrt{2T} \widehat{G}(k) \right)^2 \right] \\ &= 2T \mathbb{E} \left[\sum_{\substack{k, \ell \in \mathbb{Z} \setminus \{0\} \\ k, \ell \equiv 0 \pmod{M_\star}}} e^{2\pi i (k-\ell)\delta/M_\star} \widehat{G}(k) \overline{\widehat{G}(\ell)} \right] \\ &= \frac{2T}{n-1-\lfloor n/2 \rfloor} \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \sum_{\substack{k, \ell \in \mathbb{Z} \setminus \{0\} \\ k, \ell \equiv 0 \pmod{m_\star}}} \widehat{G}(k) \overline{\widehat{G}(\ell)} \int_0^1 e^{2\pi i (k-\ell)\delta/m_\star} d\delta \\ &= \frac{2T}{n-1-\lfloor n/2 \rfloor} \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \sum_{\substack{k \in \mathbb{Z} \setminus \{0\} \\ k \equiv 0 \pmod{m_\star}}} |\widehat{G}(k)|^2 \\ &= \frac{2T}{n-1-\lfloor n/2 \rfloor} \sum_{k \in \mathbb{Z} \setminus \{0\}} |\widehat{G}(k)|^2 \sum_{m_\star=\lfloor n/2 \rfloor}^{n-2} \chi_{m_\star|k}, \end{aligned}$$

where, in the last line, $\chi_{m_\star|k}$ is equal to 1 if m_\star divides k and is equal to 0 otherwise. The last sum over m_\star counts the number of divisors of k in $\{\lfloor n/2 \rfloor, \dots, n-2\}$, which we denote by $\tau_{\lfloor n/2 \rfloor}^{n-2}(k)$. Then, Hölder's inequality leads to

$$\begin{aligned} & \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx \right)^2 \right] \\ &= \frac{2T}{n-1-\lfloor n/2 \rfloor} \sum_{k \in \mathbb{Z} \setminus \{0\}} |\widehat{G}(k)|^2 \tau_{\lfloor n/2 \rfloor}^{n-2}(k) \\ &= \frac{2T}{n-1-\lfloor n/2 \rfloor} \sum_{k \in \mathbb{Z} \setminus \{0\}} |\widehat{G}(k)|^2 \left(\frac{2\pi k}{2T} \right)^{2\alpha} \left(\frac{2T}{2\pi k} \right)^{2\alpha} \tau_{\lfloor n/2 \rfloor}^{n-2}(k) \\ &\leq \frac{2T}{n-1-\lfloor n/2 \rfloor} \left(\sum_{k \in \mathbb{Z} \setminus \{0\}} |\widehat{G}(k)|^2 \left(\frac{2\pi k}{2T} \right)^{2\alpha} \right) \sup_{k \in \mathbb{Z} \setminus \{0\}} \left(\frac{2T}{2\pi k} \right)^{2\alpha} \tau_{\lfloor n/2 \rfloor}^{n-2}(k) \\ &\leq \frac{2T^{2\alpha+1}}{\pi^{2\alpha}(n-1-\lfloor n/2 \rfloor)} \|G\|_{\alpha, [-T, T]}^2 \sup_{k \in \mathbb{Z} \setminus \{0\}} \frac{\tau_{\lfloor n/2 \rfloor}^{n-2}(k)}{k^{2\alpha}}, \end{aligned}$$

where we used Parseval's identity for the last inequality; see [9, Proof of Lemma 4.1]. We trivially have $\tau_{\lfloor n/2 \rfloor}^{n-2}(k) = 0$ for any $|k| < \lfloor n/2 \rfloor$. For the case $|k| \geq \lfloor n/2 \rfloor$, since m_\star divides k if and only if k/m_\star is an integer, $\tau_{\lfloor n/2 \rfloor}^{n-2}(k)$ is bounded above by the number of integers contained in the interval

$$\left[\frac{|k|}{n-2}, \frac{|k|}{\lfloor n/2 \rfloor} \right].$$

Therefore, we have

$$\tau_{\lfloor n/2 \rfloor}^{n-2}(k) \leq \frac{|k|}{\lfloor n/2 \rfloor} - \frac{|k|}{n-2} + 1 \leq \frac{|k|}{n/2-1} - \frac{|k|}{n-2} + 1 = \frac{|k|}{n-2} + 1.$$

Using this bound, noting $\alpha \geq 1$, the supremum over k above is bounded by

$$\sup_{k \in \mathbb{Z} \setminus \{0\}} \frac{\tau_{\lfloor n/2 \rfloor}^{n-2}(k)}{k^{2\alpha}} \leq \sup_{k \geq \lfloor n/2 \rfloor} \frac{1}{k^{2\alpha}} \left(\frac{k}{n-2} + 1 \right) = \frac{1}{\lfloor n/2 \rfloor^{2\alpha}} \left(\frac{\lfloor n/2 \rfloor}{n-2} + 1 \right) \leq \frac{2^{2\alpha+1}}{n^{2\alpha}},$$

which, together with Lemma 3.6.3 and Lemma 3.5, yields

$$\begin{aligned} \mathbb{E} \left[\left(\frac{2T}{M_\star} \sum_{j=0}^{M_\star-1} G(\xi_j^*) - \int_{-T}^T G(x) dx \right)^2 \right] &\leq \|G\|_{\alpha, [-T, T]}^2 \frac{2^{2\alpha+2} T^{2\alpha+1}}{\pi^{2\alpha}(n-1-\lfloor n/2 \rfloor)n^{2\alpha}} \\ &\leq (C_{\alpha, \lambda}^*)^2 \|f\|_{\alpha}^2 \frac{2^{2\alpha+4} T^{2\alpha+1}}{\pi^{2\alpha} n^{2\alpha+1}}. \end{aligned}$$

Now that we have obtained upper bounds on the two terms of (13), we complete the proof. \square

The benefit of choosing the number of nodes M_\star randomly manifests itself in the proof of Lemma 3.8. As indicated in (15), the trapezoidal rule with a fixed number of nodes m does not give exact values for integrating Fourier modes $\phi_k^{[-T, T]} = \exp(2\pi i k(x+T)/(2T))/\sqrt{2T}$, $k \neq 0$, if k is a multiple of m . Thus, if the number of

nodes is fixed, due to the amplitude of the k -th Fourier coefficient of the periodized function G with k being a multiple of m , the best RMSE attainable is already of order $m^{-\alpha}$ and not less (cf. the proof of [9, Lemma 4.1]). Although this rate is optimal in the sense of the (deterministic) worst-case error (3) (cf. [9]), in view of the lower bound in Theorem 2.1 we need a half rate extra for $A_{n,T}$ to be optimal in the sense of the worst-case RMSE. By choosing M_* randomly, on average we can avoid the situation where k is a multiple of m , which leads to an improved, optimal rate of the RMSE.

Not all quadratures require randomization of the number of nodes to achieve the optimality. Indeed, using (scaled) higher-order scrambled digital nets from [6] for the nodes ξ_j^* , instead of the equispaced points as in (6), following a similar argument as above, the same rate (up to a logarithmic factor) turns out to be achievable. Note however that constructing higher-order nets requires the smoothness parameter α as an input to attain the optimal rate of the RMSE for smooth non-periodic functions, which makes it hard for the resulting randomized algorithm to be free from α in the construction. Algorithm 3.1 compares favorably in this regard, as we discuss later in Remark 3.11.

Lemma 3.9 (Upper bound on the RMSE for tails). *Let $\alpha \geq 1$, $n \geq 4$, and $T > 0$ be given. For any $f \in \mathcal{H}_\alpha$, the second and third terms of (12) are bounded as*

$$\mathbb{E} \left[\left(\Phi(-T)f(\xi_{\text{left}}^*) - \int_{-\infty}^{-T} f(x)\rho(x) dx \right)^2 \right] \leq \|f\|_\alpha^2 \frac{1}{\sqrt{2\pi T}e^{T^2/2}},$$

and

$$\mathbb{E} \left[\left((1 - \Phi(T))f(\xi_{\text{right}}^*) - \int_T^\infty f(x)\rho(x) dx \right)^2 \right] \leq \|f\|_\alpha^2 \frac{1}{\sqrt{2\pi T}e^{T^2/2}},$$

respectively.

Proof. We give a proof only for the bound on the second term of (13) since the bound on the third term can be proven in the same way.

From the proof of Lemma 3.2

$$\mathbb{E} [\Phi(-T)f(\xi_{\text{left}}^*)] = \int_{-\infty}^{-T} f(x)\rho(x) dx,$$

holds, and thus

$$\begin{aligned} & \mathbb{E} \left[\left(\Phi(-T)f(\xi_{\text{left}}^*) - \int_{-\infty}^{-T} f(x)\rho(x) dx \right)^2 \right] \\ &= \mathbb{E} \left[(\Phi(-T)f(\xi_{\text{left}}^*))^2 \right] - \left(\int_{-\infty}^{-T} f(x)\rho(x) dx \right)^2 \\ &\leq \mathbb{E} \left[(\Phi(-T)f(\xi_{\text{left}}^*))^2 \right] = \Phi(-T) \int_{-\infty}^{-T} |f(x)|^2 \rho(x) dx \\ &\leq \Phi(-T) \|f\|_{L^2_p}^2 \leq \Phi(-T) \|f\|_\alpha^2, \end{aligned}$$

in which we have

$$\begin{aligned}\Phi(-T) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-T} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_T^{\infty} e^{-x^2/2} dx \\ &\leq \frac{1}{\sqrt{2\pi}} \int_T^{\infty} \frac{x}{T} \cdot e^{-x^2/2} dx = \frac{e^{-T^2/2}}{T\sqrt{2\pi}}.\end{aligned}$$

This proves the result. \square

Applying the results from Lemma 3.8 and 3.9 to (12), we obtain

$$\begin{aligned}\mathbb{E} \left[\left(A_{n,T}^{M_{\star},\delta}(f) - I(f) \right)^2 \right] \\ \leq 3(C_{\alpha,\lambda}^*)^2 \|f\|_{\alpha}^2 \left(\frac{2\alpha^2 \max\{1, (2T)^{2\alpha}\}}{e^{(1-\lambda)T^2}} + \frac{2^{2\alpha+5}T^{2\alpha+1}}{\pi^{2\alpha}n^{2\alpha+1}} \right) + \|f\|_{\alpha}^2 \frac{6}{\sqrt{2\pi}Te^{T^2/2}},\end{aligned}$$

so that

$$\begin{aligned}e^{\text{rmse}}(A_{n,T}, \mathcal{H}_{\alpha}) &= \sup_{\substack{f \in \mathcal{H}_{\alpha} \\ \|f\|_{\alpha} \leq 1}} \left(\mathbb{E} \left[\left(A_{n,T}^{M_{\star},\delta}(f) - I(f) \right)^2 \right] \right)^{1/2} \\ &\leq \left(3(C_{\alpha,\lambda}^*)^2 \left(\frac{2\alpha^2 \max\{1, (2T)^{2\alpha}\}}{e^{(1-\lambda)T^2}} + \frac{2^{2\alpha+5}T^{2\alpha+1}}{\pi^{2\alpha}n^{2\alpha+1}} \right) + \frac{6}{\sqrt{2\pi}Te^{T^2/2}} \right)^{1/2}.\end{aligned}$$

Choosing the cut-off parameter T according to (10), we have $T > 1$ and

$$(16) \quad e^{\text{rmse}}(A_{n,T}, \mathcal{H}_{\alpha}) \leq \frac{T^{\alpha+1/2}}{n^{\alpha+1/2}} \left(3(2^{2\alpha+1}\alpha^2 + \pi^{-2\alpha}2^{2\alpha+5})(C_{\alpha,\lambda}^*)^2 + \frac{6}{\sqrt{2\pi}} \right)^{1/2}$$

$$(17) = \frac{(\ln n)^{\alpha/2+1/4}}{n^{\alpha+1/2}} \sqrt{\left(\frac{2\alpha+1}{1-\lambda} \right)^{\alpha+1} \left(3(2^{2\alpha+1}\alpha^2 + \pi^{-2\alpha}2^{2\alpha+5})(C_{\alpha,\lambda}^*)^2 + \frac{6}{\sqrt{2\pi}} \right)},$$

which completes the proof of Theorem 3.3.

Remark 3.10. *The quadrature for the tail $\Phi(-T)f(\xi_{\text{left}}^*) + (1 - \Phi(T))f(\xi_{\text{right}}^*)$ in the algorithm (6) are utilized only to make the estimator unbiased and thus to obtain the empirical error estimator (9). Indeed, the rate $O(n^{-\alpha-1/2}(\ln n)^{\alpha/2+1/4})$ as in Theorem 3.3 can be established even without these nodes, since under the decay of the integrands assumed, the dominant error comes from the integration error on $[-T, T]$ as in Lemma 3.8.*

Remark 3.11. *In Theorem 3.3, the cut-off parameter T depends on the smoothness parameter α . This may be a problem in practice, since the smoothness of the target integrand may be unknown. One way to make the algorithm independent of α , as we did in [9, Corollary 4.4], is to replace α in (10) with any slowly increasing function $\gamma : \mathbb{N} \rightarrow [0, \infty)$, such as $\gamma(n) = \ln(n)$ or $\gamma(n) = \max\{\ln(\ln(n)), 0\}$. The resulting randomized trapezoidal rule does not require any information about α , but still achieves the optimal rate up to a factor of $(\gamma(n) \ln n)^{\alpha/2+1/4}$.*

4. NUMERICAL EXPERIMENTS

We conclude this paper with numerical experiments for our randomized trapezoidal rule $A_{n,T}$. In what follows, all computations are carried out in double precision arithmetic using MATLAB 2020a. Let us consider the following three test

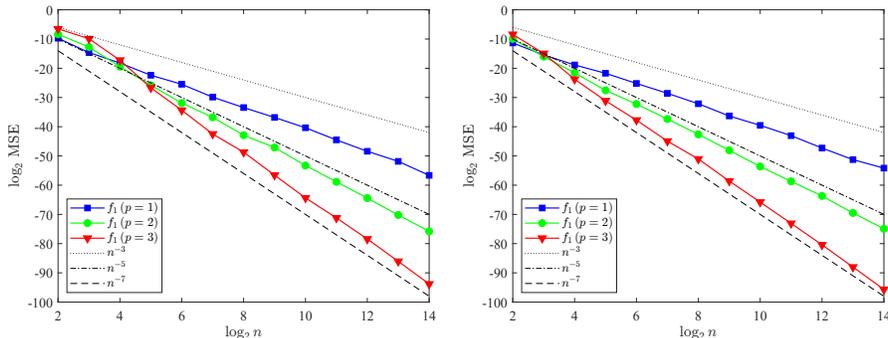


FIGURE 1. Mean-squared error for f_1 with various values of p . In the left panel, the cut-off parameter T is chosen as in (10); in the right panel, T is chosen independently of α .

functions with different smoothness:

$$f_1(x) = f_{1,p}(x) = (\max\{x, 0\})^p, \quad \text{with } p = 1, 2, 3,$$

$$f_2(x) = \begin{cases} e^{-1/(1-x^2)} & \text{for } |x| < 1, \\ 0 & \text{otherwise,} \end{cases}$$

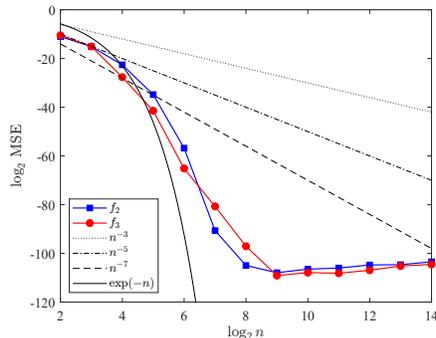
$$f_3(x) = (\tanh(x))^2.$$

The function $f_1 = f_{1,p}$ satisfies $f_1 \in \mathcal{H}_p$ and $f_1 \notin \mathcal{H}_{p+1}$, which we exploit to check how well our theory explains the error that the estimator estimates. The functions f_2 and f_3 go beyond our theory, in that the Sobolev class does not capture their smoothness; they are infinitely differentiable and bounded on \mathbb{R} and thus are in \mathcal{H}_α for any $\alpha \in \mathbb{N}$. Note that f_3 is analytic, while f_2 is not.

Throughout the following experiments, we set $\lambda = 0.51$. We estimate the mean-squared error by following (9) with $r = 50$ and n being powers of 2. The reason why we estimate the mean-squared error instead of the RMSE is only because of the fact that the square root of (9) is not an unbiased estimator of the RMSE. We also stress that, even though f_2 and f_3 go beyond our theory, the practical error estimator (9) is still valid.

For f_1 , we test two cut-off strategies to choose the value of T . One way is choosing T depending on α and we set $\alpha = p$ for $f_1 = f_{1,p}$; see (10) in Theorem 3.3. Another way is to use our α -free cut-off, in which we replace α in (10) with $\gamma(n) = \max\{\ln(\ln(n)), 0\}$; see Remark 3.11. Figure 1 presents error decays for f_1 with varying values of p , with two cut-off strategies outlined above. As expected from the theory, the convergence rate of n^{-2p-1} is achieved for each value of p . Moreover, we see that our α -free choice of the cut-off parameter does not deteriorate the performance of the randomized trapezoidal rule, which supports our theory.

For the functions f_2 and f_3 , we conduct the experiments only for the α -free choice of the cut-off parameter T . Figure 2 shows the results. Since the functions are infinitely smooth, the mean-squared error is expected to decay at least at the rate $n^{-2\alpha-1}$ for any finite α . In fact, we observe that the error decays super-algebraically fast until it drops down to around 2^{-100} . This error decay is consistent with our theory. Nevertheless, it does indicate that our theory may not be suitable for providing sharp error estimates for infinitely smooth functions. For such functions,

FIGURE 2. Mean-squared error for f_2 and f_3 .

it may be useful to employ function classes used to analyze e.g., analytic functions. We refer to [21, 22] for error estimates of a deterministic trapezoidal rule for analytic functions. Setting the theory aside, we again stress that the crucial advantage of our randomized rule is that it allows for an error estimation, which is quite hard for deterministic quadrature rules.

REFERENCES

- [1] N. S. Bakhvalov. On the approximate calculation of multiple integrals (in Russian). *Vestnik MGU, Ser. Mat. Mekh. Astron. Fiz. Khim.*, 4:3–18, 1959. English translation in *J. Complexity*, 31:502–516, 2015.
- [2] N. S. Bakhvalov. An estimate of the mean remainder term in quadrature formulae (in Russian). *Zh. Vych. Mat.*, 1:64–77, 1961.
- [3] Z. I. Botev. The normal law under linear restrictions: simulation and estimation via minimax tilting. *J. R. Stat. Soc. B*, 79(1):125–148, 2017.
- [4] N. Chopin. Fast simulation of truncated Gaussian distributions. *Stat. Comput.*, 21(2):275–288, 2011.
- [5] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.
- [6] J. Dick. Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *Ann. Statist.*, 39(3):1372–1398, 2011.
- [7] J. Dick, T. Goda, and K. Suzuki. Component-by-component construction of randomized rank-1 lattice rules achieving almost the optimal randomized error rate. *Math. Comp.*, 91(338):2771–2801, 2022.
- [8] J. Dick, C. Irrgeher, G. Leobacher, and F. Pillichshammer. On the optimal order of integration in Hermite spaces with finite smoothness. *SIAM J. Numer. Anal.*, 56(2):684–707, 2018.
- [9] Y. Kazashi, Y. Suzuki, and T. Goda. Sub-optimality of Gauss–Hermite quadrature and optimality of trapezoidal rule for functions with finite smoothness. *SIAM J. Numer. Anal.*, 61(3):1426–1448, 2023.
- [10] N. M. Korobov. *Teoretiko-chislovye metody v priblizhenom analize*. Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1963.
- [11] P. Kritzer, F. Y. Kuo, D. Nuyens, and M. Ullrich. Lattice rules with random n achieve nearly the optimal $\mathcal{O}(n^{-\alpha-1/2})$ error independently of the dimension. *J. Approx. Theory*, 240:96–113, 2019.
- [12] D. H. Lehmer. On the maxima and minima of Bernoulli polynomials. *Amer. Math. Monthly*, 47:533–538, 1940.
- [13] J. F. Monahan. *Numerical Methods of Statistics*. Cambridge University Press, second edition, 2011.
- [14] E. Novak. *Deterministic and Stochastic Error Bounds in Numerical Analysis*. Springer-Verlag, Berlin, 1988.
- [15] E. Novak. On the power of adaption. *J. Complexity*, 12(3):199–237, 1996.

- [16] E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems, Volume I: Linear Information*. EMS, Zürich, 2008.
- [17] E. Novak and H. Woźniakowski. *Tractability of Multivariate Problems, Volume II: Standard Information for Functionals*. EMS, Zürich, 2010.
- [18] D. Nuyens and Y. Suzuki. Scaled lattice rules for integration on \mathbb{R}^d achieving higher-order convergence with error analysis in terms of orthogonal projections onto periodic spaces. *Math. Comp.*, 92(339):307–347, 2023.
- [19] A. B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910, 1997.
- [20] I. H. Sloan and S. Joe. *Lattice Methods for Multiple Integration*. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1994.
- [21] M. Sugihara. Optimality of the double exponential formula—functional analysis approach—. *Numer. Math.*, 75(3):379–395, 1997.
- [22] L. N. Trefethen and J. A. C. Weideman. The exponentially convergent trapezoidal rule. *SIAM Rev.*, 56(3):385–458, 2014.
- [23] M. Ullrich. A Monte Carlo method for integration of multivariate smooth functions. *SIAM J. Numer. Anal.*, 55(3):1188–1200, 2017.
- [24] G. W. Wasilkowski. Information of varying cardinality. *J. Complexity*, 2(3):204–228, 1986.
- [25] Y. Wu. A randomized trapezoidal quadrature. *Int. J. Comput. Math.*, 99(4):680–692, 2022.
- [26] S. K. Zaremba. La méthode des “bons treillis” pour le calcul des intégrales multiples. In *Applications of Number Theory to Numerical Analysis*, pages 39–119. Academic Press, New York, 1972.

SCHOOL OF ENGINEERING, UNIVERSITY OF TOKYO, 7-3-1 HONGO, BUNKYO-KU, TOKYO 113-8656, JAPAN.

Email address: `goda@frcer.t.u-tokyo.ac.jp`

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF STRATHCLYDE, GLASGOW G1 1XH, UK

Email address: `y.kazashi@strath.ac.uk`

DEPARTMENT OF MATHEMATICAL SCIENCES, NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY, SENTRALBYGG IL, ALFRED GETZ' VEI 1, GLØSHAUGEN, 7034 TRONDHEIM, NORWAY; DEPARTMENT OF MATHEMATICS AND SYSTEMS ANALYSIS, AALTO UNIVERSITY SCHOOL OF SCIENCE, ESPOO, FI-00076 AALTO, FINLAND

Email address: `yuya.suzuki@aalto.fi`