Classical integration rules and Wiener measure

Amarjit S. Budhiraja Andrew J. Sommese

Abstract

In this paper precise calculations are made of the square of the error of classical integration rules averaged over spaces of differentiable functions. Precisely, let $\mathcal{C}_0^N$ denote the space of N-times continuously differentiable functions on $[0,1]$ with $f^{(j)}(0) = 0$ for $j \leq N$. Let μ_N denote the measure induced by the N-times integrated Wiener process on $\mathcal{C}_0^N$. For a given set of nodes $0 = x_0 < x_1 < \cdots < x_K = 1$ and weights $\mathbf{w} := (w_{ij};\ \ i = 0, \cdots K; j = 0, \cdots, N)$ define the integration rule $f \to A_{\mathbf{w}}(f) := \sum_{i,j} w_{i,j} f^{(j)}(x_i)$, to approximate $I(f) := \int_0^1 f(x)\mathrm{d}x$. A study is made of the average square error $\int_{\mathcal{C}_0^N} |I(f) - A_{\mathbf{w}}(f)|^2 \mathrm{d}\mu_N(f)$; for what weights $\mathbf{w}$ the minimum is achieved; for what sets of nodes this minimum is minimized; and what the minimum is in this case.

AMS Subject Classification. 65D30, 60G15.

Keywords. integration rules, trapezoidal rule, Newton-Cotes rules, Wiener measure, integrated Wiener process, Hermite interpolation.

Contents

1 Introduction **1**

2 Background Material **4**
- 2.1 Function Spaces and Measures Derived from Wiener Measure . . . 4
- 2.2 Measures on $\mathcal{C}^N := C^N[0,1]$ . . . 4
- 2.3 Change of Variables Formula . . . 5

3 Integration Rules on $\mathcal{C}_0$ **6**
- 3.1 Reduction from a Composite Rule to a Simple Rule . . . 7
- 3.2 Some Specific Calculations . . . 9
  - 3.2.1 The trapezoidal rule and Simpson's rule . . . 9
  - 3.2.2 Newton-Cotes' rules . . . 9
- 3.3 The Optimal Integration Rule . . . 10

4 Integration Rules and Measures on Spaces of Differentiable Functions **13**
- 4.1 Optimal Integration Rule on $\mathcal{C}_0^N$ and Hermite Interpolation . . . 13
- 4.2 The Average Square Error for the Optimal Rule . . . 18

1 Introduction

For a classical integration rule of order $n-1$ applied to a sufficiently differentiable function, there is an estimate for the error of the approximation to the actual integral in terms of the n-th derivative of the function. For example, given a twice continuously differentiable function f on the interval $[0,1]$, and using the composite trapezoidal rule [StoB93],

$$\left|\int_0^1 f(x)\mathrm{d}x - \left(\frac{f(0)+\sum_{i=1}^{K-1} 2f\left(\frac{i}{K}\right)+f(1)}{2}\right)\right| \le \frac{\max_{[0,1]}|f^{(2)}|}{12K^2}.$$

For merely continuous functions, there are no strong error estimates that are true for each individual function. In this article, we study the L^2-error of the integration method for differentiable functions *averaged* with respect to a natural measure on the space of such functions. For example, using Wiener measure [Wie23, KarShr91] μ on $\mathcal{C}$, the continuous real valued functions on $[0,1]$, we have

$$\int_{\mathcal{C}}\left|\int_0^1 f(x)\mathrm{d}x - \left(\frac{f(0)+\sum_{i=1}^{K-1} 2f\left(\frac{i}{K}\right)+f(1)}{2}\right)\right|^2 \mathrm{d}\mu(f) = \frac{1}{12K^2}.$$

More precisely, given a partition $x_0 = 0 < x_1 \cdots x_{K-1} < x_K = 1$ of the interval $[0,1] \subset \mathbb{R}$, and an integer $N \ge 0$, we ask what integration method minimizes the L^2-error under the assumption that we know the first N derivatives of the function at the points, $x_0, \ldots, x_K$. We further ask what is this error explicitly; and for a given K and N, which choice of $K-1$ points in the open interval (a,b) minimize this error. More generally we are interested in these same questions where the underlying assumption is that we know possibly different numbers of derivatives at nodes on the interval.

The study of such errors averaged over an appropriate function space measure seems to have been begun by Sul'din [Su59, Su60]. Since then there have been a number of papers touching on different aspects of this study, cf., Smale [Sm85], Gao [G89, G91], and Choi [C94].

There are a number of issues that immediately come up as to what spaces of functions, measures, and norms should be used. The use of measures built out of a Wiener process is very natural. From Donsker's invariance principle [Bill68, Theorem 2.10.1], a path space analogue of the classical central limit theorem, Brownian motion and Wiener measure are the limiting stochastic process and measure when the uncertainties in the physical problem under consideration arise as a net result of many small and not strongly dependent random effects. This partially accounts for the pervasive use of Wiener measure as the basis for many useful applied real world models, e.g., prices of securities in financial markets. Thus our results say something about the average of a given integration method applied to a random sample of functions, e.g., of price histories securities in a fixed time interval.

Moreover, Wiener measure has many canonical properties, which effectively means that many different set of assumptions lead to it, e.g., [Str93, Chapter IV]. We use L^2 norms since the orthogonal increment property of the Wiener process and its simple covariance structure enable us to carry out many calculations explicitly.

For spaces of functions we use the space $C^N[0,1]$ of N times continuously differentiable functions on the interval $[0,1]$. We will denote $C^0[0,1]$, also by $C[0,1]$. A natural measure on the subset of $C^N[0,1]$ consisting of functions with $f^{(j)}(0) = 0$ for $j \le N$ is the measure associated to N times integrated Brownian motion. There are choices involved in extending the measure to all of $C^N[0,1]$, but for the case of main interest, N less than or equal to the order of the method, the natural choices lead to the same L^2-norm for error, see, §2.2.

Let us give a more detailed description of the results of this paper.

In §2, we summarize some facts about the function space measures we use in the rest of the paper.

In §3 we work only with continuous functions, because the combinatorics are such that we can give a very complete picture. We consider an integration rule $A(f) := \sum_{i=0}^{n} w_i f(x_i)$ for $f \in C[0,1]$

with $0 \le x_0, \ldots, x_n \le 1$ a strictly increasing sequence of points and $w_i \in \mathbb{R}$. We let

$$A_m(f) := \sum_{j=1}^{m} \left[\sum_{i=0}^{n} w_i f\left(a + \frac{(b-a)}{m}(j-1+x_i)\right)\right] \frac{(b-a)}{m}$$

be the associated m-fold composite rule on $[a,b]$. We show in Corollary 3.5 that if A is of order $N \ge 2$, i.e., if $A(p) = \int_0^1 p(t)\mathrm{d}t$ for all polynomials $p(t)$ of degree $\le N \ge 2$, then with the Wiener measure, μ on $C[a,b]$,

$$\int_{C[a,b]} \left| \int_a^b f(x)\mathrm{d}x - A_m(f) \right|^2 \mathrm{d}\mu(f) = \frac{(b-a)^3}{m^2} \left(\int_{C[0,1]} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3} \right).$$

There is also a slightly more complicated formula in the case when the order $N = 1$. Using this formula it is easy to compute the exact average square error for classical rules such as the Newton-Cotes rules. We carry out the calculations of this average square error, and show that it equals $\frac{1}{12K^2}$ in the case of the composite trapezoidal rule and $\frac{1}{9K^2}$ in the case of the composite Simpson's rule, where $K := nm$ is one less than the number of points where the function f is evaluated. In Table 1, we list the average square errors for the Newton-Cotes rules with $n \le 10$.

We next show in Theorem 3.8 that for a given set of node points for the rule $0 \le x_0 < \cdots < x_K \le 1$ the smallest average square error occurs for the composite trapezoidal rule, and moreover once the trapezoidal rule is fixed then varying over the sets of node points $0 \le x_0 \le \cdots \le x_K \le 1$, the smallest average error is obtained when the nodes are equidistant, i.e, $x_i = a + \frac{(b-a)i}{K}$.

In §4 we study the average error on $C^N[0,1]$ assuming that the values of the function and some of its first N derivatives are known at the node points. We show in Theorem 4.2 that the least $L^2(\mu_N)$ error, where μ_N is the measure induced by the N times integrated Wiener process, is obtained when the integration rule is given by the integration of a piecewise polynomial function. In Corollary 4.3, we show that if the values of the function f and its first N derivatives are known at the node points, then the least $L^2(\mu_N)$ error is achieved when the integration rule is given by the integration between any two node points x_i, x_{i+1} of the unique interpolating polynomial $p(x)$ of degree $2N+1$, with $p^{(j)}(x_i) := f^{(j)}(x_i)$, $p^{(j)}(x_{i+1}) := f^{(j)}(x_{i+1})$, for $j \le N$, i.e.,

$$\begin{aligned} A(f) &:= \sum_{i=0}^{K-1} \int_{x_i}^{x_{i+1}} H_i(f)(y)\mathrm{d}y \\ &= \sum_{i=0}^{K-1} \sum_{j=0}^{N} (-1)^j \frac{\binom{2N+1-j}{N+1}}{\binom{2N+2}{N+1}(j+1)!} \left[f^{(j)}(x_{i+1}) + (-1)^j f^{(j)}(x_i) \right] (x_{i+1}-x_i)^{j+1}. \end{aligned} \tag{1}$$

We then turn to what the minimum $L^2(\mu_N)$ error is in this case. We show that for $N = 1$, the minimum $L^2(\mu_N)$ error is $\frac{1}{720K^4}$ with equality only for the rule given in equation 1, and the node points, $x_i := i/K$ for $i = 0, \ldots, K$. We also give a precise conjecture 4.5 for the minimum for all N, i.e., $\frac{1}{2(2N+3)!\binom{2N+1}{N}K^{2N+2}}$, with equality only for the rule given in equation 1, and the node points, $x_i := i/K$ for $i = 0, \ldots, K$.

The first author would like to thank National Science Foundation (NSF-DMI-9812857) and the University of Notre Dame Faculty Research Program for their support. The second author would like to thank the Duncan Chair of the University of Notre Dame and the Mathematical Sciences Research Institute in Berkeley for their support.

2 Background Material

In this section we collect some background material.

2.1 Function Spaces and Measures Derived from Wiener Measure

Let $C[0,\infty)$ be the space of continuous functions from $[0,\infty)$ to $\mathbb{R}$. Let $\mathcal{B}(C[0,\infty))$ denote the Borel σ- field for this space with the usual topology of uniform convergence on compacts. Let μ_∞ be the standard Wiener measure on $(C[0,\infty),\mathcal{B}(C[0,\infty)))$. Let $0\le a\le b<\infty$ be fixed and define: $\pi_{a,b}: C[0,\infty)\to C[a,b]$ as follows. For $f\in C[0,\infty)$:

$$(\pi_{a,b}f)(x) := f(x); \quad a\le x\le b.$$

Let $\mu_{a,b}$ be the push forward measure: $\mu_\infty o\pi_{a,b}^{-1}$ on $(C[a,b],\mathcal{B}(C[a,b]))$, where $\mathcal{B}(C[a,b])$ is the Borel σ field on $C[a,b]$. We will write $\mu_{0,1}$ as μ. The rest of §2.1 is not needed for §3.

Let $C^N[a,b]$ denote the space of N-times continuously differentiable functions on $[a,b]$ and let $C[a,b]$ be the space of continuous functions on $[a,b]$. We will sometimes denote $C^N[0,1]$ by $\mathcal{C}^N$ and $\mathcal{C}^0$ by $\mathcal{C}$. Also, let $\mathcal{C}_0^N$ be the space of N times continuously differentiable functions f on $[0,1]$ such that $f^{(j)}(0)=0$ for $j\le N$. Recall that $\mathcal{C}_0$ is the space of continuous functions f on $[0,1]$ such that $f(0)=0$.

Let μ, as before be the standard Wiener measure on $\mathcal{C}_0$. For $N\in\mathbb{N}$, define the map ϕ_N:

$$\phi_N:\mathcal{C}_0\to\mathcal{C}_0^N,$$

as follows. For $t_N\in[0,1]$ and $f\in\mathcal{C}_0$,

$$(\phi_N f)(t_N) := \int_0^{t_N}\left(\int_0^{t_{N-1}}\cdots\left(\int_0^{t_1} f(t_0)dt_0\right)\cdots\right)dt_{N-1}.$$

Denote by μ_N the push forward measure $\mu o\Phi_N^{-1}$ on $(\mathcal{C}_0^N,\mathcal{B}(\mathcal{C}_0^N))$, where $\mathcal{B}(\mathcal{C}_0^N)$ denotes the Borel σ-field on $\mathcal{C}_0^N$ corresponding to the uniform norm on the function value and its first N derivatives. Also let $\mu_0:=\mu$.

2.2 Measures on $\mathcal{C}^N := C^N[0,1]$

In this article we restrict attention to $\mathcal{C}_0^N$, the space of N times continuously differentiable functions f on $[0,1]$ such that $f^{(j)}(0)=0$ for $j\le N$. In this subsection we will show that, for natural extensions of the measures to $\mathcal{C}^N := C^N[0,1]$, we get the same results. Let $\mathcal{P}^N := P^N[0,1]$ denote the polynomials of degree $\le N$ on $[0,1]$. Every function $h\in\mathcal{C}^N$ can be written uniquely $h(x)=t_N(x)+f(x)$, where $t_N(x)\in\mathcal{P}^N$ is the Taylor polynomial of h of degree N expanded around 0, and $f(x)\in\mathcal{C}_0^N$. Now let $G:=\mathcal{C}_0^N\times\mathcal{P}^N$ and endow G with the usual product σ-field denoted by $\mathcal{G}$. Let ν_N be a probability measure on $(G,\mathcal{G})$ satisfying

$$\nu_N(A\times\mathcal{P}^N)=\mu_N(A), \tag{2}$$

for all $A\in\mathcal{B}(\mathcal{C}_0^N)$. Define $\theta: G\to\mathcal{C}^N$ as $\theta(f,p):=f+p$; $(f,p)\in\mathcal{C}_0^N\times\mathcal{P}^N$. Finally, let $\widetilde{\mu}_N:=\nu_N o\theta^{-1}$.

Lemma 2.1 *Let A be any integration rule of order $\ge N$ on functions $f\in\mathcal{C}^N$, i.e., $A(p)=\int_0^1 p(x)\mathrm{d}x$ for $p\in\mathcal{P}^N$. Let $\widetilde{\mu}_N$ be a probability measure on $(C^N,\mathcal{B}(C^N))$ as constructed above. Then:*

$$\int_{\mathcal{C}^N}\left|A(h)-\int_0^1 h(x)\mathrm{d}x\right|^2 d\widetilde{\mu}_N(h)=\int_{\mathcal{C}_0^N}\left|A(f)-\int_0^1 f(x)\mathrm{d}x\right|^2\mathrm{d}\mu_N(f).$$

Proof. Since $A(p) = \int_0^1 p(x)\mathrm{d}x$ for $p \in \mathcal{P}^N$, we have

$$\begin{aligned}\int_{\mathcal{C}^N} \left|A(h) - \int_0^1 h(x)\mathrm{d}x\right|^2 d\widetilde{\mu}_N(h) &= \int_{\mathcal{C}_0^N \times \mathcal{P}^N} \left|A(f+p) - \int_0^1 (f(x)+p(x))\mathrm{d}x\right|^2 d\nu_N(f,p) \\ &= \int_{\mathcal{C}_0^N \times \mathcal{P}^N} \left|A(f) - \int_0^1 f(x)\mathrm{d}x\right|^2 d\nu_N(f,p) \\ &= \int_{\mathcal{C}_0^N} \left|A(f) - \int_0^1 f(x)\mathrm{d}x\right|^2 d\mu_N(f),\end{aligned}$$

where the last step follows from (2). □

It is important to note, that when we talk about an integration rule of order $\geq N$ on $f \in \mathcal{C}_0^N$ of the form $\sum_{i=0}^{K}\sum_{j=0}^{N} w_{ij} f^{(j)}(x_i)$ for the set of nodepoints $x_0 = 0 < x_1 < \cdots < x_K = 1$, the weights w_{0k} for $k \leq N$ are determined by the other weights. Indeed we have that $\frac{1}{k+1} = A(x^k) = w_{0k}k! + \sum_{i=1}^K \sum_{j=k}^N w_{ij}\binom{j}{k}k!x_i^{j-k}$, i.e.,

$$w_{0k} = \frac{1}{(k+1)!} - \sum_{i=1}^{K}\sum_{j=k}^{N} w_{ij}\binom{j}{k} x_i^{j-k}.$$

for $k = 0, \ldots, N$. This observation is implicit in the proofs of the uniqueness statements of Theorem 3.8, Lemma 3.9, Corollary 4.3, Conjecture 4.5, and Theorem 4.6.

2.3 Change of Variables Formula

In §3 we are interested in studying the L^2 error, with respect to the measure $\mu_{a,b}$, for a given integration rule on $C[a,b]$. The following change of variables formula will allow us to study this L^2 error for a given integration rule on $C[a,b]$ in terms of a related integration rule on $\mathcal{C}_0 := \{f \in C[0,1] | f(0) = 0\}$. Henceforth for $f \in C[a,b]$, we write $\int_a^b f(x)\mathrm{d}x$ as $I_a^b(f)$. We will usually abbreviate $I_0^1(f)$ to $I(f)$.

Theorem 2.2 *Let $a = x_0 < x_1 < \cdots < x_K = b$ be a fixed partition of $[a,b]$. Let $\{w_j\}_{0\leq j\leq k}$ be a given set of weights. Let $A(\cdot)$ be an integration rule of order $N \geq 0$, on $C[a,b]$, defined as follows. For $f \in C[a,b]$*

$$A(f) := \sum_{j=0}^{K} w_j f(x_j).$$

Then

$$\int_{C[a,b]} [I_a^b(f) - A(f)]^2 d\mu_{a,b}(f) = (b-a)^3 \int_{\mathcal{C}_0} [I(f) - \widetilde{A}(f)]^2 d\mu(f),$$

where for $f \in \mathcal{C}_0$

$$\widetilde{A}(f) := \sum_{j=0}^{k} \widetilde{w}_j f(y_j);$$

$\widetilde{w}_j = \frac{w_j}{b-a}$ *and* $y_j = \dfrac{x_j - a}{b-a}$; $j = 0, \cdots, K$.

Proof. Let $\{W(t)\}_{0\le t<\infty}$ be a standard Wiener process on some probability space. Then

$$\begin{aligned}\int_{C[a,b]}[I_a^b(f)-A(f)]^2 d\mu_{a,b}(f) &= \mathbb{E}\left[\int_a^b W(s)\mathrm{d}s-\sum_{j=0}^K w_j W(x_j)\right]^2\\ &= \mathbb{E}\Bigg[\int_a^b (W(s)-W(a))\mathrm{d}s+W(a)(b-a)\\ &\qquad -\sum_{j=0}^K w_j(W(x_j)-W(a))-W(a)\sum_{j=0}^K w_j\Bigg]^2\\ &= \mathbb{E}\left[\int_a^b (W(s)-W(a))\mathrm{d}s-\sum_{j=0}^K w_j(W(x_j)-W(a))\right]^2,\end{aligned}$$

where the last equality follows on noticing that $\sum_{j=0}^K w_j=b-a$. Next note that

$$\int_a^b (W(s)-W(a))\mathrm{d}s=(b-a)\int_0^1[W(a+s(b-a))-W(a)]\mathrm{d}s.$$

Furthermore

$$\sum_{j=0}^K w_j(W(x_j)-W(a))=\sum_{j=0}^k w_j(W(a+y_j(b-a))-W(a)).$$

Define, for $s\in[0,1]$

$$\widetilde{W}(s):=\frac{W(a+s(b-a))-W(a)}{\sqrt{b-a}}.$$

With this notation

$$\begin{aligned}\int_{C[a,b]}[I_a^b(f)-A(f)]^2 d\mu_{a,b}(f) &= (b-a)^3\mathbb{E}\left[\int_0^1 \widetilde{W}(s)\mathrm{d}s-\sum_{j=0}^K \widetilde{w}_j\widetilde{W}(y_j)\right]^2\\ &= (b-a)^3\mathbb{E}\left[\int_0^1 W(s)\mathrm{d}s-\sum_{j=0}^K \widetilde{w}_j W(y_j)\right]^2\\ &= (b-a)^3\int_{\mathcal{C}_0}[I(f)-\widetilde{A}(f)]^2 d\mu(f),\end{aligned}$$

where the next to last equality above follows on observing that the probability law of the family $\{\widetilde{W}(s);\ 0\le s\le 1\}$ is same as that of the family $\{W(s); 0\le s\le 1\}$. □

3 Integration Rules on $\mathcal{C}_0$

We recall that μ is the standard Wiener measure on

$$\mathcal{C}_0:=\{f\in C[0,1]\big|f(0)=0\}$$

and for $f\in\mathcal{C}_0$, $I(f):=\int_0^1 f(x)\mathrm{d}x$. Let $A(f):=\sum_{i=0}^n w_i f(x_i)$, where

1. the x_i are a strictly increasing sequence of real numbers with $x_0 \geq 0$ and $x_n \leq 1$; and
2. the w_i are real numbers.

Given a rule $f \to A(f) := \sum_{i=0}^n w_i f(x_i)$ as above and $m \in \mathbb{N}_0$, define the m-times composite rule

$$A_m(f) := \sum_{j=1}^m \mathcal{A}_j(f) = \sum_{j=1}^m \left[\sum_{i=0}^n \frac{w_i}{m} f\left(\frac{j-1+x_i}{m}\right)\right].$$

Note that for $m = 1$, $\mathcal{A}_1 = A$. Moreover if A is of order N, then so is $\mathcal{A}_m$ for any m. The composite rule associated to $A(f) = \dfrac{f(0)+f(1)}{2}$ is the usual composite trapezoidal rule analyzed in §3.2.1, and the composite rule associated to $A(f) = \dfrac{f(0)+4f(\frac{1}{2})+f(1)}{6}$ is the usual composite Simpson's rule analyzed in §3.2.2.

The goal of this section is to compute

$$\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) \tag{3}$$

3.1 Reduction from a Composite Rule to a Simple Rule

In this section we make calculations with few assumptions on the weights and nodes. In particular we give Theorem 3.4 which reduces calculations for a composite method based on a given simple rule to that for the simple rule. Let $A(\cdot)$ be an integration rule as above. Then

$$\begin{aligned}
\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f) &= \int_{\mathcal{C}_0} I(f)^2 \mathrm{d}\mu(f) - 2\int_{\mathcal{C}_0} I(f)A(f)\mathrm{d}\mu(f) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) &(4)\\
&= 2\int_0^1 \int_0^y x \mathrm{d}x\mathrm{d}y - 2A(x) + A(x^2) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) &(5)\\
&= \frac{1}{3} - 2A(x) + A(x^2) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f). &(6)
\end{aligned}$$

An immediate consequence of this calculation is:

Lemma 3.1 *If A is an integration rule of order $N \geq 1$, then*

$$\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f) = \frac{1}{3} - 2A(x) + A(x^2) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f).$$

Furthermore, if $N \geq 2$, then

$$\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f) = \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3}.$$

The above lemma shows that the computation of $\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f)$ comes down to the computation of $\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f)$.

Example 3.2 Let A be the trapezoid rule, $A(f) := \dfrac{f(0)+f(1)}{2}$. Then $\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) = \dfrac{1}{4}$, and thus,

$$\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f) = \frac{1}{3} - 2A(x) + A(x^2) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) = \frac{1}{3} - 1 + \frac{1}{2} + \frac{1}{4} = \frac{1}{12}.$$

Example 3.3 Let A be Simpson's rule, $A(f) := \dfrac{f(0) + 4f\left(\frac{1}{2}\right) + f(1)}{6}$. Then

$$\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) = \left[\frac{4}{6}\right]^2 \left[\frac{1}{2}\right] + 2\left[\frac{4}{6}\right]\left[\frac{1}{6}\right]\left[\frac{1}{2}\right] + \left[\frac{1}{6}\right]^2 = \frac{13}{36},$$

and thus,

$$\int_{\mathcal{C}_0} |I(f) - A(f)|^2 \mathrm{d}\mu(f) = \frac{1}{3} - 2A(x) + A(x^2) + \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) = \frac{1}{3} - 1 + \frac{1}{3} + \frac{13}{36} = \frac{1}{36}.$$

We now show how to reduce the computation of $\int_{\mathcal{C}_0} A_m(f)^2 \mathrm{d}\mu(f)$ for a composite rule to the computations for the rule.

Theorem 3.4 *Let A be an integration rule of order $N \geq 1$ and let A_m be the m-fold composite rule constructed from A. Then*

$$\int_{\mathcal{C}_0} A_m(f)^2 \mathrm{d}\mu(f) = \frac{1}{m^2}\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2} + \frac{1}{3}.$$

Proof. Note that if $j < k$ then

$$\int_{\mathcal{C}_0} \mathcal{A}_j(f)\mathcal{A}_k(f)\mathrm{d}\mu(f) = \frac{1}{m}\sum_{i=1}^{n} \frac{w_i}{m}\left(\frac{j-1+x_i}{m}\right) = \frac{1}{m^3}\left(j - 1 + \frac{1}{2}\right) = \frac{1}{m^3}\left(j - \frac{1}{2}\right)$$

Thus $\int_{\mathcal{C}_0} A_m(f)^2 \mathrm{d}\mu(f)$

$$\begin{aligned}
&= \sum_{j=1}^{m}\int_{\mathcal{C}_0} \mathcal{A}_j^2 \mathrm{d}\mu(f) + 2\sum_{j=1}^{m}\left[\frac{1}{m^3}\left(j - \frac{1}{2}\right)\right](m-j) \\
&= \sum_{j=1}^{m}\int_{\mathcal{C}_0} \mathcal{A}_j^2 \mathrm{d}\mu(f) - \frac{1}{2m} + \frac{1}{6m^2} + \frac{1}{3} \\
&= \sum_{j=1}^{m}\left[\sum_{i=1}^{n}\frac{w_i^2}{m^2}\left(\frac{j-1+x_i}{m}\right) + 2\sum_{1\leq i<k\leq n}\frac{w_i w_k}{m^2}\left(\frac{j-1+x_i}{m}\right)\right] - \frac{1}{2m} + \frac{1}{6m^2} + \frac{1}{3} \\
&= \left[\sum_{i=1}^{n}\frac{w_i^2}{m^2}\left(\frac{m-1}{2} + x_i\right) + 2\sum_{1\leq i<k\leq n}\frac{w_i w_k}{m^2}\left(\frac{m-1}{2} + x_i\right)\right] - \frac{1}{2m} + \frac{1}{6m^2} + \frac{1}{3} \\
&= \frac{m-1}{2}\left[\sum_{i=1}^{n}\frac{w_i}{m}\right]^2 + \frac{1}{m^2}\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{2m} + \frac{1}{6m^2} + \frac{1}{3} \\
&= \frac{m-1}{2}\frac{1}{m^2} + \frac{1}{m^2}\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{2m} + \frac{1}{6m^2} + \frac{1}{3} \\
&= \frac{1}{m^2}\int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2} + \frac{1}{3}.
\end{aligned}$$

□

Combining Theorem 3.4 with Lemma 3.1 we get the following corollary.

Corollary 3.5 *Let A be an integration rule of order $N \geq 1$ and let A_m be the m-fold composite rule constructed from A. Then*

$$\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) = -\frac{1}{3} + A_m(x^2) + \frac{1}{m^2} \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2}.$$

If $N \geq 2$ then

$$\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) = \frac{1}{m^2} \int_{\mathcal{C}_0} A(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2}.$$

3.2 Some Specific Calculations

In this section we compute the L^2 errors with respect to the Wiener measure for a few specific methods.

3.2.1 The trapezoidal rule and Simpson's rule

Example 3.6 (Composite Trapezoid Rule) Let A_m denote the m-fold composite trapezoidal rule. Noting that $A_m(x^2) = \frac{1}{3} + \frac{1}{6m^2}$, it follows from Corollary 3.5 and Example 3.2 that

$$\begin{aligned}\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) &= -\frac{1}{3} + \frac{1}{3} + \frac{1}{6m^2} + \frac{1}{m^2} \int_{\mathcal{C}_0} \mathcal{A}(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2} \\ &= \frac{1}{6m^2} + \frac{1}{m^2} \int_{\mathcal{C}_0} \mathcal{A}(f)^2 \mathrm{d}\mu(f) - \frac{1}{3m^2} \\ &= \frac{1}{6m^2} + \frac{1}{4m^2} - \frac{1}{3m^2} \\ &= \frac{1}{12m^2}.\end{aligned}$$

Note that the m-fold composite trapezoidal rule has $K+1$ nodes with $K = m$. Thus we have $\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) = \frac{1}{12K^2}$.

Example 3.7 (Composite Simpson Rule) Let $\mathcal{A}_m$ denote the m-fold composite Simpson rule. Using Corollary 3.5 and Example 3.3, we have $\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) = \frac{13}{36m^2} - \frac{1}{3m^2} = \frac{1}{36m^2}$. Note that the m-fold composite Simpson rule has $K+1$ nodes with $K = 2m$. Thus we have $\int_{\mathcal{C}_0} |I(f) - A_m(f)|^2 \mathrm{d}\mu(f) = \frac{1}{9K^2}$.

3.2.2 Newton-Cotes' rules

Calculations for other Newton-Cotes' rules are carried out similarly. Table 1 has the values for the first ten Newton-Cotes' rules using Maple. We have n equal to the number of nodes minus one. The third column is the value multiplied by n^2.

Note that the weights in Newton-Cotes are positive for $n \leq 7$ and $n = 9$. It is probably true that $n = 9$ is the last Newton-Cotes formula with all weights positive. The last column grows quite fast. Generally, the values for odd n are increasing and the values for even n are increasing, though the series for odd n is less than that for the preceding even n. For $n = 20$, the value in the last column is $\approx$ 177,908.

Table 1: Average Square Error for Newton-Cotes Formulae

n	$\int_{\mathcal{C}_0} \lvert I(f)-A(f)\rvert^2 \mathrm{d}\mu(f)$	$n^2\left(\int_{\mathcal{C}_0} \lvert I(f)-A(f)\rvert^2 \mathrm{d}\mu(f)\right)$		
1	$\frac{1}{12}$	$\frac{1}{12}$	$\approx$	0.083
2	$\frac{1}{36}$	$\frac{1}{9}$	$\approx$	0.111
3	$\frac{1}{96}$	$\frac{3}{32}$	$\approx$	0.094
4	$\frac{13}{1620}$	$\frac{52}{405}$	$\approx$	0.128
5	$\frac{169}{41472}$	$\frac{4225}{41472}$	$\approx$	0.102
6	$\frac{293}{50400}$	$\frac{293}{1400}$	$\approx$	0.209
7	$\frac{21211}{8294400}$	$\frac{1039339}{8294400}$	$\approx$	0.125
8	$\frac{2670323}{267907500}$	$\frac{42725168}{66976875}$	$\approx$	0.638
9	$\frac{33233407}{12042240000}$	$\frac{897301989}{4014080000}$	$\approx$	0.224
10	$\frac{554021687}{16295634432}$	$\frac{13850542175}{4073908608}$	$\approx$	3.400

3.3 The Optimal Integration Rule

In this subsection we show that in the class of integration rules with $K+1$ nodes, composite trapezoidal rule gives the least L^2 error.

Theorem 3.8 *Let $a = x_0 < x_1 < \cdots < x_K = b$ be a fixed partition of $[a,b]$. For a given set of weights $\mathbf{w} := (w_0, w_1, \cdots, w_K)$ define the integration rule $A_{\mathbf{w}}$ on $C[a,b]$ as follows. For $f \in C[a,b]$:*

$$A_{\mathbf{w}}(f) := \sum_{j=0}^{K} w_j f(x_j).$$

Then

$$\inf_{\mathbf{w}} \int_{C[a,b]} [I_a^b(f) - A_{\mathbf{w}}(f)]^2 d\mu_{a,b}(f) = \int_{C[a,b]} [I_a^b(f) - A(f)]^2 d\mu_{a,b}(f);$$

if and only if

$$A(f) = \sum_{j=0}^{K} \frac{\Delta_j}{2}(f(x_{j+1}) + f(x_j)),$$

where

$$\Delta_j = x_{j+1} - x_j.$$

Moreover

$$\int_{C[a,b]} [I_a^b(f) - A_{\mathbf{w}}(f)]^2 d\mu_{a,b}(f) \geq \frac{(b-a)^3}{12K^2},$$

with equality if and only if $x_j = a + \dfrac{(b-a)j}{K}$ for $j = 0, \ldots K$, and $A_{\mathbf{w}}(f)$ is the composite trapezoidal rule, i.e.,

$$A_{\mathbf{w}}(f) := \left(f(x_0) + \sum_{j=1}^{K-1} 2f(x_j) + f(x_K) \right) \frac{(b-a)}{2K}.$$

As an immediate consequence of Theorem 2.2 we are reduced to showing the following lemma.

Lemma 3.9 *Let* $0 = x_0 < x_1 < \cdots < x_K = 1$ *be a given partition of* $[0,1]$. *For a given set of weights:* $\mathbf{w} := (w_0, w_1, \cdots, w_K)$ *define the integration rule on* $\mathcal{C}_0$ *as follows. For* $f \in \mathcal{C}_0$

$$A_{\mathbf{w}}(f) := \sum_{j=0}^{K} w_j f(x_j).$$

Then

$$\inf_{\mathbf{w}} \int_{\mathcal{C}_0} [I(f) - A_{\mathbf{w}}(f)]^2 \mathrm{d}\mu(f) = \int_{\mathcal{C}_0} [I(f) - A(f)]^2 \mathrm{d}\mu(f),$$

where

$$A(f) = \sum_{j=0}^{K} \frac{\Delta_j}{2} [f(x_{j+1}) + f(x_j)],$$

and $\Delta_j := x_{j+1} - x_j$. *Moreover*

$$\int_{\mathcal{C}_0} [I(f) - A_{\mathbf{w}}(f)]^2 \mathrm{d}\mu(f) \geq \frac{1}{12K^2},$$

with equality if and only if $A_{\mathbf{w}}(f)$ *is the composite trapezoidal rule, i.e.,*

$$A_{\mathbf{w}}(f) := \left(f(0) + \sum_{j=1}^{K-1} 2f\left(\frac{j}{K}\right) + f(1) \right) \frac{1}{2K}.$$

Proof. In order to prove the first conclusion of the lemma it suffices to show that

$$\mathbb{E}_{\mu}\left[I(f) \mid f(x_0), \cdots, f(x_K)\right] = A(f),$$

where $\mathbb{E}_{\mu}$ denotes the expected value corresponding to the measure μ. Observe that

$$\begin{aligned} \mathbb{E}_{\mu}[I(f) \mid f(x_0), \cdots, f(x_K)] &= \sum_{j=0}^{K-1} \mathbb{E}_{\mu}[\int_{x_j}^{x_{j+1}} f(x)\mathrm{d}x \mid f(x_0), \cdots, f(x_K)] \\ &= \sum_{j=0}^{K-1} \int_{x_j}^{x_{j+1}} \mathbb{E}_{\mu}[f(x) \mid f(x_0), \cdots, f(x_K)]\, \mathrm{d}x. \end{aligned}$$

Next note that for $x \in [x_j, x_{j+1})$, $\mathbb{E}_{\mu}[f(x) \mid f(x_0), \cdots, f(x_K)]$

$$\begin{aligned} &= f(x_j) + \mathbb{E}_{\mu}[f(x) - f(x_j) \mid f(x_0), \cdots, f(x_K)] \\ &= f(x_j) + \mathbb{E}_{\mu}[f(x) - f(x_j) \mid f(x_0), f(x_1) - f(x_0), \cdots, f(x_K) - f(x_{K-1})] \\ &= f(x_j) + \mathbb{E}_{\mu}[f(x) - f(x_j) \mid f(x_{j+1}) - f(x_j)] \\ &= f(x_j) + \frac{x - x_j}{\Delta_j}(f(x_{j+1}) - f(x_j)), \end{aligned}$$

where the third equality follows from the independent increment property of the Wiener process. Hence

$$\mathbb{E}_{\mu}[I(f) \mid f(x_0), \cdots, f(x_K)] = \sum_{j=0}^{K-1} \int_{x_j}^{x_{j+1}} \left(f(x_j) + \frac{x - x_j}{\Delta_j}(f(x_{j+1}) - f(x_j)) \right) \mathrm{d}x$$

$$= \sum_{j=0}^{K-1} \left(\Delta_j f(x_j) + \frac{\Delta_j}{2}(f(x_{j+1}) - f(x_j)) \right)$$

$$= \sum_{j=0}^{K-1} \frac{\Delta_j}{2}(f(x_{j+1}) + f(x_j)).$$

This proves the conclusion of the theorem for arbitrary node points x_i.

The final assertion is a minimization statement about the explicit function:

$$F_K(x_1, \ldots, x_{K-1}) := \int_{\mathcal{C}_0} \left(\int_0^1 f(x)\mathrm{d}x - \sum_{j=0}^{K-1} \frac{x_{j+1} - x_j}{2}(f(x_{j+1}) + f(x_j)) \right)^2 \mathrm{d}\mu(f).$$

By Example 3.2.1, $F_K\left(\frac{1}{K}, \ldots, \frac{K-1}{K}\right) = \frac{1}{12K^2}$. Therefore it suffices to show that there is a unique minimum of $F_K(x_1, \ldots, x_{K-1})$ on the region $0 \le x_1 \le \cdots \le x_{K-1} \le 1$, and it occurs at $(x_1, \ldots, x_{K-1}) = (\frac{1}{K}, \ldots, \frac{K-1}{K})$. First note that

$$\begin{aligned} F_K(x_1, \ldots, x_{K-1}) = \frac{1}{3} &- \sum_{j=0}^{K-1} \left(x_j + x_{j+1} - \frac{x_j^2}{2} - \frac{x_{j+1}^2}{2} \right)(x_{j+1} - x_j) \\ &+ \sum_{j=0}^{K-1} (3x_j + x_{j+1}) \left(\frac{x_{j+1} - x_j}{2} \right)^2 \\ &+ \sum_{j=0}^{K-1} (x_j + x_{j+1})(x_{j+1} - x_j)(1 - x_{j+1}) \end{aligned}$$

Noting that F_K with two variables equal, e.g., $x_i = x_{i+1}$ is simply the F_{K-1} with one of the variables omitted, we are reduced by induction to checking the local minima of F_K in the region $x_1 < \cdots < x_{K-1}$. To this end note that

$$\frac{\partial F_K(x_1, \ldots, x_{K-1})}{\partial x_j} = \frac{1}{4}(x_{j-1} - x_{j+1})(x_{j-1} - 2x_j + x_{j+1}).$$

In the region $x_1 < \cdots < x_{K-1}$, we conclude that any relative minima satisfies $x_{j-1} - 2x_j + x_{j+1} = 0$ for $j = 0, \ldots, K-1$. This system of linear equations has the unique solution $(x_1, \ldots, x_{K-1}) = (\frac{1}{K}, \ldots, \frac{K-1}{K})$. To see this, note that $(x_1, \ldots, x_{K-1}) = (\frac{1}{K}, \ldots, \frac{K-1}{K})$ is a solution, and further, with the tridiagonal $(K-1) \times (K-1)$-matrix T with -2 along the diagonal, and 1 on the off diagonals, e.g., for $K \ge 7$,

$$T := \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & \cdots & 0 & 0 & 1 & -2 \end{pmatrix},$$

it follows by using difference equations applied to the expansion by minors that $\det T = (-1)^{K-1}K \neq 0$. That it is a relative minima follows from the fact that by induction, $F_K \ge \frac{1}{12(K-1)^2}$ on the boundary of the region. □

4 Integration Rules and Measures on Spaces of Differentiable Functions

We will consider in this section integration rules on $\mathcal{C}_0^N$ and the $L^2(\mu_N)$ errors associated with such rules. We follow the notation of §2.1.

4.1 Optimal Integration Rule on $\mathcal{C}_0^N$ and Hermite Interpolation

Suppose that:

$$0 = x_0 < x_1 < x_2 \cdots < x_K = 1$$

is a given partition on $[0,1]$. We will like to find the integration rule on $\mathcal{C}_0^N$ which minimizes the $L^2(\mu_N)$ error and which is based on

$$\left\{ f^{(j)}(x_i);\ \ i,j \in \mathbb{N}_0;\ \ i \le K; j \le N \right\},$$

where f is an arbitrary element of $\mathcal{C}_0^N$.

Let $W(\cdot)$ be a standard Brownian motion on $[0,1]$, and for $j \in \{1, \cdots, N\}$ let $X^{(j)}(\cdot)$ be a stochastic process defined inductively as

$$X^{(j)}(t) := \int_0^t X^{(j-1)}(s)\mathrm{d}s,$$

where $X^{(0)}(\cdot) := W(\cdot)$. Clearly, μ_N is the law of $X^{(j)}(\cdot)$, i.e. for any continuous bounded real valued function G on $\mathcal{C}_0^N$

$$\int_{\mathcal{C}_0^N} G(f) d\mu_N(f) = \mathbb{E}\left(G\left(X^{(N)}(\cdot)\right)\right).$$

We let

$$R^{(N)}(u,v) := \mathbb{E}\left(X^{(N)}(u) X^{(N)}(v)\right).$$

Lemma 4.1 *With the above notation we have that for $n \in \mathbb{N}_0$*

$$R^{(n)}(u,v) = \begin{cases} \frac{(-1)^n}{(2n+1)!}\left[\sum_{k=0}^n \binom{2n+1}{k} u^{2n+1-k}(-v)^k\right] & \text{for } u \le v \\ \\ \frac{(-1)^n}{(2n+1)!}\left[\sum_{k=0}^n \binom{2n+1}{k} (-u)^k v^{2n+1-k}\right] & \text{for } v \le u \end{cases}$$

Proof. The proof is via an inductive argument. The result is clearly true when $n = 0$. Next suppose that the result is true for $n-1$ for some $n \in \mathbb{N}$. Then for $u \le v$, $\mathbb{E}[X^{(n)}(u)X^{(n)}(v)]$

$$\begin{aligned} &= \int_0^v \left(\int_0^u \mathbb{E}[X^{(n-1)}(s)X^{(n-1)}(t)]\mathrm{d}s \right) dt \\ &= 2\int_0^u \left(\int_0^t \mathbb{E}[X^{(n-1)}(s)X^{(n-1)}(t)]\mathrm{d}s \right) \mathrm{d}t + \int_u^v \left(\int_0^u \mathbb{E}[X^{(n-1)}(s)X^{(n-1)}(t)]\mathrm{d}s \right) \mathrm{d}t. \end{aligned}$$

Thus

$$\begin{aligned} R^{(n)}(u,v) &= 2\int_0^u \left(\int_0^t \frac{(-1)^{n-1}}{(2n-1)!} \left[\sum_{k=0}^{n-1} \binom{2n-1}{k} s^{2n-1-k}(-t)^k \right] \mathrm{d}s \right) dt \\ &+ \int_u^v \left(\int_0^u \frac{(-1)^{n-1}}{(2n-1)!} \left[\sum_{k=0}^{n-1} \binom{2n-1}{k} s^{2n-1-k}(-t)^k \right] \mathrm{d}s \right) dt \end{aligned}$$

$$= \frac{(-1)^{n-1}}{(2n-1)!}\sum_{k=0}^{n-1}\binom{2n-1}{k}(-1)^k\frac{2u^{2n+1}}{(2n-k)(2n+1)}$$
$$+ \frac{(-1)^{n-1}}{(2n-1)!}\sum_{k=0}^{n-1}\binom{2n-1}{k}(-1)^k\frac{u^{2n-k}[v^{k+1}-u^{k+1}]}{(2n-k)(k+1)}.$$

Rearranging the terms above we have that

$$R^{(n)}(u,v) = \frac{(-1)^{n-1}}{(2n-1)!}\sum_{k=0}^{n-1}\binom{2n-1}{k}(-1)^k\frac{u^{2n+1}}{(2n-k)}\left[\frac{2}{2n+1}-\frac{1}{k+1}\right]$$
$$+ \frac{(-1)^{n-1}}{(2n-1)!}\sum_{k=0}^{n-1}\binom{2n-1}{k}(-1)^k\frac{u^{2n-k}v^{k+1}}{(2n-k)(k+1)}$$
$$= \frac{(-1)^{n-1}}{(2n-1)!}u^{2n+1}\left[\sum_{k=0}^{n-1}\binom{2n-1}{k}(-1)^k\frac{2k+1-2n}{(2n-k)(2n+1)(k+1)}\right] \quad (7)$$
$$+ \frac{(-1)^n}{(2n+1)!}\sum_{k=1}^{n}\binom{2n+1}{k}(-1)^k u^{2n+1-k}v^k. \quad (8)$$

Finally, the expression in (7) can be written as:

$$\frac{(-1)^{n-1}}{(2n-1)!}u^{2n+1}\left[-\sum_{k=0}^{n-1}\binom{2n-1}{k}\frac{(-1)^k}{(2n+1)(k+1)}+\sum_{k=0}^{n-1}\binom{2n-1}{k}\frac{(-1)^k}{(2n-k)(2n+1)}\right]$$
$$= \frac{(-1)^{n-1}}{(2n-1)!}u^{2n+1}\left[\frac{-1}{2n(2n+1)}\sum_{k=0}^{n-1}(-1)^k\binom{2n}{k+1}+\frac{1}{2n(2n+1)}\sum_{k=0}^{n-1}(-1)^k\binom{2n}{k}\right]$$
$$= \frac{(-1)^n}{(2n+1)!}u^{2n+1}\left[1-2\sum_{k=0}^{n-1}(-1)^k\binom{2n}{k}-(-1)^n\binom{2n}{n}\right]$$
$$= \frac{(-1)^n}{(2n+1)!}u^{2n+1}\left[1+(-1)^n 2\binom{2n-1}{n-1}-(-1)^n\binom{2n}{n}\right]$$
$$= \frac{(-1)^n}{(2n+1)!}u^{2n+1}.$$

Using the above observation in (7) we have the result. □

We now show that the optimal $L^2(\mu_N)$ estimate is obtained by using an interpolating polynomial. More precisely,

Theorem 4.2 *Let $0 = x_0 < x_1 < \cdots < x_K = 1$ be a given partition of $[0,1]$. Let*

$$S \subset \{0,1,\cdots,K\}\times\{0,1,\cdots,N\}$$

be given. For a given set of weights $\mathbf{w} := (w_{ij};(i,j)\in S)$ define the integration rule

$$A_{\mathbf{w}}(f) := \sum_{(i,j)\in S} w_{ij}\left(\frac{d}{dx}\right)^j f(x_i);$$

where f is an arbitrary element of $\mathcal{C}_0^N$. Then

$$\inf_{\mathbf{w}}\int_{\mathcal{C}_0^N}[I(f)-A_{\mathbf{w}}(f)]^2 d\mu_N(f) = \int_{\mathcal{C}_0^N}[I(f)-A(f)]^2 d\mu_N(f),$$

where

$$A(f) := \sum_{k=0}^{K-1} \int_{x_k}^{x_{k+1}} P_k(f)(y)\mathrm{d}y,$$

and $P_k(f)(y); y \in [x_k, x_{k+1}]$ *is a polynomial of degree at most* $2N+1$. *Furthermore,* $P_k(f)$ *satisfies*

$$\frac{d^j}{dx^j} P_k(f)(x_k) = f^{(j)}(x_k); \quad \forall j \;\; s.t. \;\; (k,j) \in S$$

and

$$\frac{d^j}{dx^j} P_k(f)(x_{k+1}) = f^{(j)}(x_{k+1}); \quad \forall j \;\; s.t. \;\; (k+1,j) \in S.$$

Proof. Note that

$$\int_{\mathcal{C}_0^N} [I(f) - A_{\mathbf{w}}(f)]^2 d\mu_N(f) = \mathbb{E}\left[\int_0^1 X^{(N)}(s)\mathrm{d}s - A_{\mathbf{w}}(X^{(N)})\right]^2.$$

Furthermore

$$\inf_{\mathbf{w}} \mathbb{E}\left[\int_0^1 X^{(N)}(s)\mathrm{d}s - A_{\mathbf{w}}(X^{(N)})\right]^2 = \mathbb{E}\left[\int_0^1 X^{(N)}(s)\mathrm{d}s - \zeta\right]^2,$$

where

$$\begin{aligned} \zeta &= \mathbb{E}\left[\int_0^1 X^{(N)}(s)\mathrm{d}s \mid X^{(j)}(x_i); (i, N-j) \in S\right] \\ &= \sum_{k=0}^{K-1} \mathbb{E}\left[\int_{x_k}^{x_{k+1}} X^{(N)}(s)\mathrm{d}s \mid X^{(j)}(x_i); (i, N-j) \in S\right] \\ &= \sum_{k=0}^{K-1} \int_{x_k}^{x_{k+1}} \mathbb{E}\left[X^{(N)}(s) \mid X^{(j)}(x_i); (i, N-j) \in S\right] \mathrm{d}s. \end{aligned}$$

Thus it suffices to show that for all $k \in \{0, 1, \cdots, K-1\}$ and $s \in [x_k, x_{k+1}]$

$$\mathbb{E}\left[X^{(N)}(s) \mid X^{(j)}(x_i); (i,j) \in \widetilde{S}\right] = P_k(X^{(N)})(s), \tag{9}$$

where

$$\widetilde{S} := \{(i,j) : (i, N-j) \in S\}.$$

Since $(X^{(N)}(s), X^{(j)}(x_i); (i,j) \in \widetilde{S})$ is a Gaussian family for all $s \in [0,1]$ we have that the conditional expectation in (9) equals $u_S(x)A_S^{-1}v_S$, where A_S is a $|S| \times |S|$ matrix with a typical entry being $\mathbb{E}[X^{(j)}(x_i)X^{(j')}(x_{i'})]$; $(i,j),(i',j') \in \widetilde{S}$, v_S is a $|S| \times 1$ matrix with a typical entry equal to $X^{(j)}(x_i)$ and finally $u_S(x)$ is a $1 \times |S|$ matrix with a typical entry equal to $\mathbb{E}[X^{(N)}(s)X^{(j)}(x_i)]$; $(i,j) \in \widetilde{S}$. We will now show that for $s \in [x_k, x_{k+1}]$ all the entries of the row vector $u_S(x)$ are polynomials with degree at most $2N+1$. Note that for $j \in \{0, 1, \cdots, N\}$

$$\begin{aligned} X^{(N)}(s) &= \sum_{l=0}^{N-j} X^{(N-l)}(x_k)\frac{(s-x_k)^l}{l!} \\ &+ \int_{x_k}^{s}\left(\int_{x_k}^{t_{N-1}} \cdots \left(\int_{x_k}^{t_{j+1}} (X^{(j)}(t_j) - X^{(j)}(x_k))dt_j\right) \cdots\right) dt_{N-1}. \end{aligned}$$

Thus $\mathbb{E}(X^{(N)}(s)X^{(j)}(x_i))$

$$\begin{aligned} &= \sum_{l=0}^{N-j} \mathbb{E}\left(X^{(N-l)}(x_k)X^{(j)}(x_i)\right)\frac{(s-x_k)^l}{l!} \\ &+ \int_{x_k}^{s}\left(\int_{x_k}^{t_{N-1}}\cdots\left(\int_{x_k}^{t_{j+1}}\left(\mathbb{E}(X^{(j)}(t_j)X^{(j)}(x_i)) - \mathbb{E}(X^{(j)}(x_k)X^{(j)}(x_i))\right)dt_j\right)\cdots\right)dt_{N-1}. \end{aligned} \tag{10}$$

Now observe that $\mathbb{E}(X^{(j)}(t_j)X^{(j)}(x_i)) = R^{(j)}(x_i,t_j)$ and also that either $t_j \leq x_i$ or $t_j \geq x_i$ for all $t_j \in [x_k, x_{k+1}]$. Thus from Lemma 4.1 we have that for all $t_j \in [x_k, x_{k+1}]$, $\mathbb{E}(X^{(j)}(t_j)X^{(j)}(x_i))$ is a polynomial in t_j of degree $2j+1$. This yields that the expression in (10) is a polynomial, for $s \in [x_k, x_{k+1}]$, of degree at most:

$$2j + 1 + N - j \leq 2N + 1.$$

This establishes that $P_k(X^{(N)}(s))$ is a polynomial of degree at most $2N+1$ for $s \in [x_k, x_{k+1}]$ and $k \in \{0, 1, \cdots, K-1\}$. Finally note that for fixed $k \in \{0, 1, \cdots, K-1\}$ and j such that $(k, j) \in S$

$$\begin{aligned} \frac{d^j}{dx^j}P_k(X^{(N)})(x_k) &= \mathbb{E}\left(X^{(N-j)}(x_k) \mid X^{(j)}(x_i); (i,j) \in \widetilde{S}\right) \\ &= X^{(N-j)}(x_k). \end{aligned}$$

Similarly for j such that $(k+1, j) \in S$

$$\frac{d^j}{dx^j}P_k(X^{(N)})(x_{k+1}) = X^{(N-j)}(x_{k+1}).$$

□

Given $f \in \mathcal{C}_0^N$ and $i \in \{0, \cdots, K-1\}$ let $H_i(f)$ be the unique polynomial of degree $2N+1$ on $[x_i, x_{i+1}]$ such that

$$\left(\frac{d^j}{dx^j}H_i(f)\right)(x_i) = f^{(j)}(x_i)$$

and

$$\left(\frac{d^j}{dx^j}H_i(f)\right)(x_{i+1}) = f^{(j)}(x_{i+1}),$$

where $j \in \{0, \cdots, N\}$. Thus $H_i(f)$ is the Hermite interpolating polynomial over the interval $[x_i, x_{i+1}]$.

Corollary 4.3 *Let*

$$0 = x_0 < x_1 < x_2 \cdots < x_K = 1$$

be a given partition of $[0,1]$. *For a given set of weights* $\mathbf{w} := (w_{ij};\quad i = 0, \cdots K; j = 0, \cdots, N)$ *define the integration rule*

$$A_{\mathbf{w}}(f) := \sum_{i,j} w_{i,j} f^{(j)}(x_i),$$

where f *is an arbitrary element of* $\mathcal{C}_0^N$. *Then*

$$\inf_{\mathbf{w}} \int_{\mathcal{C}_0^N}[I(f) - A_{\mathbf{w}}(f)]^2 d\mu_N(f) = \int_{\mathcal{C}_0^N}[I(f) - A(f)]^2 d\mu_N(f),$$

where $A(f) :=$

$$\sum_{k=0}^{K-1} \int_{x_k}^{x_{k+1}} H_k(f)(y)\mathrm{d}y = \sum_{i=0}^{K-1} \sum_{j=0}^{N} (-1)^j \frac{\binom{2N+1-j}{N+1}}{\binom{2N+2}{N+1}(j+1)!} \left[f^{(j)}(x_{i+1}) + (-1)^j f^{(j)}(x_i) \right] (x_{i+1} - x_i)^{j+1}.$$

Proof. All that remains to be shown is the second equality in the description of $A(f)$. This follows from the following Lemma.

Lemma 4.4 *Let* $p(y)$ *be a polynomial of degree* $2N+1$ *on an interval* $[a, b]$*. Then*

$$\int_a^b p(y)\mathrm{d}y = \sum_{j=0}^{N} (-1)^j \frac{\binom{2N+1-j}{N+1}}{\binom{2N+2}{N+1}(j+1)!} \left[p^{(j)}(b) + (-1)^j p^{(j)}(a) \right] (b-a)^{j+1}.$$

To prove the Lemma is suffices by changing variables to prove the lemma on the interval $[0, 1]$. Any polynomial can be written as a sum of degree $2N+1$ polynomials $p(y) = q(y) + r(y)$ where for $j \le N$: $q^{(j)}(0) = 0 = r^{(j)}(1)$; $q^{(j)}(1) = p^{(j)}(1)$; and $r^{(j)}(0) = p^{(j)}(0)$. Thus, it suffices to prove the result for polynomials of the form $q(y)$. We are therefore reduced to showing that for polynomials of degree $2N+1$ satisfying $p^{(j)}(0) = 0$ for $j \le N$, we have

$$\int_0^1 p(y)\mathrm{d}y = \sum_{j=0}^{N} (-1)^j \frac{\binom{2N+1-j}{N+1}}{\binom{2N+2}{N+1}(j+1)!} p^{(j)}(1).$$

Since $p(y) = y^{N+1} \sum_{k=0}^{N} c_k (y-1)^k$, we have

$$\sum_{k=0}^{N} c_k (y-1)^k = \left(\frac{1}{1+(y-1)} \right)^{N+1} \left(\sum_{j=0}^{2N+1} \frac{p^{(j)}(1)}{j!} (y-1)^j \right).$$

Expanding $\left(\frac{1}{1+(y-1)} \right)^{N+1}$ in a power series around 1, we get explicit expressions for the coefficients c_k, i.e.,

$$c_k := \sum_{j=0}^{k} (-1)^{k-j} \binom{N+k-j}{N} \frac{p^{(j)}(1)}{j!}.$$

Integrating $y^{N+1} \sum_{k=0}^{N} c_k (y-1)^k$ from 0 to 1 yields

$$\sum_{k=0}^{N} (-1)^k \frac{1}{(N+k+2)\binom{N+k+1}{k}} \left(\sum_{j=0}^{k} (-1)^{k-j} \binom{N+k-j}{N} \frac{p^{(j)}(1)}{j!} \right).$$

Changing the order of summation we obtain

$$\sum_{j=0}^{N} (-1)^j \left(\sum_{k=j}^{N} \frac{\binom{N+k-j}{N}}{(N+k+2)\binom{N+k+1}{k} j!} \right) p^{(j)}(1).$$

Summing the middle sum gives the desired expression. □

4.2 The Average Square Error for the Optimal Rule

Using Corollary 4.3 for any particular N and K, the computation of the infimum that is achieved with the optimal integration rule is completely straightforward using a symbolic processing program such as Maple. Unfortunately, increasingly complicated combinatorics make the study of the average square error difficult for arbitrary N and K. We have the following conjecture.

Conjecture 4.5 *Let $0 = x_0 < x_1 < x_2 \cdots < x_K = 1$ be a given partition of $[0,1]$. For a given set of weights $\mathbf{w} := (w_{ij};\ \ i = 0, \cdots K; j = 0, \cdots, N)$ define the integration rule*

$$A_{\mathbf{w}}(f) := \sum_{i,j} w_{i,j} \left(\frac{d}{\mathrm{d}x}\right)^j f(x_i),$$

where f is an arbitrary element of $\mathcal{C}_0^N$. Then

$$\int_{\mathcal{C}_0^N} [I(f) - A_{\mathbf{w}}(f)]^2 d\mu_N(f) \geq \frac{1}{2(2N+3)!\binom{2N+1}{N} K^{2N+2}},$$

with equality only if $A_{\mathbf{w}}(f)$ is as given in Corollary 4.3 with $x_i := \dfrac{i}{K}$ for $i = 0, \ldots, K$.

All that needs to be done to prove the conjecture is analyze the average square distance in Corollary 4.3 for general N and K. The above conjecture is based on many calculations for a variety of N and K

Theorem 3.8 proves the conjecture when $N = 0$. Here we give a proof of the conjecture when $N = 1$.

Theorem 4.6 *Let*

$$0 = x_0 < x_1 < x_2 \cdots < x_K = 1$$

be a given partition of $[0,1]$. For a given set of weights $\mathbf{w} := (w_{ij};\ \ i = 0, \cdots K; j = 0, \cdots, N)$ define the integration rule

$$A_{\mathbf{w}}(f) := \sum_{i,j} w_{i,j} \left(\frac{d}{\mathrm{d}x}\right)^j f(x_i),$$

where f is an arbitrary element of $\mathcal{C}_0^N$. Then

$$\int_{\mathcal{C}_0^N} [I(f) - A_{\mathbf{w}}(f)]^2 d\mu_N(f) \geq \frac{1}{720K^4},$$

with equality only if

$$A_{\mathbf{w}}(f) := \left[\sum_{i=0}^{K-1} (f(x_i) + f(x_{i+1}))\frac{(x_{i+1} - x_i)}{2}\right] - \left[\sum_{i=0}^{K-1} (f'(x_{i+1}) - f'(x_i))\frac{(x_{i+1} - x_i)^2}{12}\right],$$

and $x_i := \dfrac{i}{K}$.

Proof. The calculation follows exactly the lines of the proof of Lemma 3.9. We leave the reader to check that

$$F_K(x_1, \ldots, x_{K-1}) := \int_{\mathcal{C}_0^N} [I(f) - A_{\mathbf{w}}(f)]^2 d\mu_N(f) = \frac{1}{720} - \frac{1}{144} \sum_{i=0}^{K-1} x_i x_{i+1}(x_{i+1}^2 - x_i x_{i+1} + x_i^2)(x_{i+1} - x_i)$$

Thus for $i = 1$ to $i = K - 1$, we have

$$\frac{\partial F_K}{\partial x_i} = \frac{-1}{144}(x_{i+1} - x_{i-1})(x_{i-1} - 2x_i + x_{i+1})[(x_{i+1} - x_i)^2 + (x_i - x_{i-1})^2].$$

This allows us by induction to conclude the smallest value occurs when $x_{i-1} - 2x_i + x_{i+1} = 0$ for each i from 1 to $K - 1$. Thus for the minimum, we have $x_i = \frac{i}{K}$. The rest of the calculation is straightforward. □

References

[Bill68] P. Billingsley, *Convergence of Probability Measures*, (1968), John Wiley & Sons. Inc.

[C94] S.H. Choi, Probabilistic analysis of Simpson's quadrature, J. Complexity 10 (1994), 384–410.

[G89] F. Gao, A probabilistic theory of error estimation for automatic integration algorithms, NumerMath. 56 (1989), 309–329.

[G91] F. Gao, Probabilistic analysis of numerical integration algorithms, J. Complexity 7 (1991), 58–69.

[KarShr91] I. Karatzas and S. Shreve, *Brownian motion and Stochastic Calculus*, (1991), Springer-Verlag.

[Sm85] S. Smale, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. 13 (1985), 87–121.

[StoB93] J. Stoer and R. Bullirsch, *Introduction to Numerical Analysis*, 2nd Ed. (1993), Springer-Verlag, New York.

[Str93] D.W. Stroock, *Probability Theory, An Analytic View*, (1993), Cambridge University Press, Cambridge.

[Su59] A.V. Sul'din, Wiener measure and its application to approximation methods (Russian) Izv. Vysš. Učebn. Zaved. Matematika (1959) no. 6 (13), 145–158.

[Su60] A.V. Sul'din, Wiener measure and its applications to approximation methods. II. (Russian) Izv. Vysš. Učebn. Zaved. Matematika (1960) no. 5 (18), 165–179.

[Wie23] N. Wiener, Differential Space, J. Math. Phys. 3, 127–146.

Amarjit S. Budhiraja
budhiraja.1@nd.edu

Andrew J. Sommese
sommese@nd.edu
URL: `www.nd.edu/~sommese`

Department of Mathematics
University of Notre Dame
Notre Dame, Indiana 46556, U.S.A.
fax: U.S.A. + 219–631-6579