A Spline-Based Approach to Uncertainty Quantification and Density Estimation

Adi Ditkowski†, Gadi Fibich‡, and Amir Sagiv†‡

Abstract. The effect of model uncertainties and noise on a quantity of interest (“model output”) is often better described by its probability density function (PDF) than by its moments (mean, standard deviation, etc.). Although density estimation is a common uncertainty-quantification (UQ) task, the adequacy of approximation methods (surrogates) for density estimation has scarcely been analyzed before. In this paper, we first show that standard methods in uncertainty propagation (such as generalized polynomial chaos), which are highly accurate for moment estimation, sometimes fail to approximate the PDF even in the case of one-dimensional noise. Therefore, we develop a novel spline-based algorithm for this task. Our method offers significant advantages over existing methods for density estimation, including a guaranteed convergence rate which is at least cubic in the sampling resolution. This convergence rate is better than that of standard statistical density-estimation methods (such as histograms and kernel density estimators) at dimensions \(1 \leq d \leq \frac{5}{2} m\), where \(m\) is the spline order. Furthermore, although spectral methods approximate moments with exponential accuracy whereas splines approximate moments with polynomial accuracy, the spline-based approximation can be more accurate when the sample size is small and the quantity of interest has sharp-gradients regions. We also show how to approximate moments and PDFs of non-smooth quantities of interest, which is often prohibitive in spectral methods. Finally, we demonstrate our algorithm for problems in nonlinear optics and computational fluid dynamics.

Key words. Uncertainty Quantification, Density Estimation, Probability Density Function, Nonlinear Dynamics, Spline, Surrogate

AMS subject classifications. 65D07, 65Z05, 62G07, 78A60

1. Introduction. Uncertainties and noise are prevalent in mathematical models in all branches of science. In such cases, the solution of the (otherwise deterministic) model becomes random, and so one is interested in computing its statistics. This problem, sometimes known as forward uncertainty propagation (UQ), arises in various areas such as biochemistry [31, 33], fluid dynamics [6, 20, 29, 33], structural engineering [44], hydrology [7], and nonlinear optics [39].

In many applications, one is interested in computing the probability density function (PDF) of a “quantity of interest” (output) of the model [1, 6, 7, 20, 31, 39, 48]. Since standard forward UQ methods [21, 44] such as Stochastic Finite Element, or generalized Polynomial Chaos (gPC) [22, 34, 43, 54], hp-gPC [51], and Wiener-Haar expansion [30] can approximate moments with spectral accuracy [55, 56], these methods have been used for density estimation as well. In this paper, we point out that spectral methods are sometimes a poor choice for this task, and propose a novel spline-based method for density estimation. Perhaps surprisingly, this problem received little attention so far, despite its importance.

Our spline-based density-estimation method with cubic splines has a guaranteed convergence rate of at least \(h^3\), where \(h\) is the sampling resolution. More generally, with splines of order \(m\), the convergence rate is at least \(h^m\). These rates are superior to those of the standard kernel density estimators [47, 53], for noise dimension \(1 \leq d \leq \frac{5}{2} m\). Because we only rely on solving the underlying deterministic model (i.e., our method is non-intrusive), and because interpolation by spline is a standard numerical procedure, our proposed method is very easy to implement.
Traditionally, the error bounds of moment-estimation for spectral methods (e.g., gPC) are obtained asymptotically as $N$, the number of samples, goes to infinity. In many applications, however, the regime $N \gg 1$ is not feasible, as each solution of the deterministic model is computationally expensive (or involves a physical experiment). In these cases, the number of samples is limited (e.g., $N = 10$ or $N = 30$), and so spectrally-convergent methods might fail to attain the desired accuracy due to insufficient sampling resolution, even for one-dimensional noise. In contrast, the spline-based method approximates moments accurately even when the sample size is small. In addition, high-derivatives and discontinuities have little effect on our method’s accuracy, due to the fact that spline interpolation is predominantly local (see Sec. 4). Another advantage over gPC is that splines are not limited to a specific choice of sampling points.

The paper is organized as follows. Sec. 2 introduces the general settings and notations, and presents several density-estimation applications from the forward uncertainty propagation literature. Sec. 3 reviews standard statistical density-estimation methods (histogram, kernel density estimators) and the gPC method for moment- and density-estimation. In Sec. 4 we present our spline-based algorithm for moment- and density-estimation in the one-dimensional case. We then prove that the density-estimation error scales as $N^{-m}$, where $N$ is the number of samples and $m$ is the order of the splines (Theorem 4.9). Sec. 5 generalizes our algorithm to $d$-dimensional noises using tensor-product splines of order $m$. This section also contains our key theoretical result (Theorem 5.3), that the density-estimation error in the $d$-dimensional case scales as $N^{-\frac{m}{d}}$.

In Sec. 7 we compare numerically the moment-estimation and density-estimation accuracy of our spline-based method with that of gPC and KDE in one dimension. In addition, in Sec. 6.4 we show that both gPC and our spline-based method can approximate moments and the PDF of certain non-smooth quantities of interest. We conclude this section with two- and three-dimensional numerical examples (Sec. 6.5). In all cases, the density-estimation errors are consistent with our error estimates (Theorems 4.9 and 5.3). We use our method to compute the PDF of the rotation angle of the polarization ellipse in nonlinear optics (Sec. 7), and the PDF of the shock location in the Burgers equation (Sec. 8). In all these cases, we confirm that the spline-based density estimation converges at least at a cubic rate, and observe that the spline-based moments are more accurate than the gPC ones for small sample sizes. Sec. 9 concludes with open questions and future research directions.

### 2. Settings and computational goals

We consider initial value problems of the form

\[ u_t(t, x; \alpha) = Q(u, x; \alpha)u, \quad u(t = 0, x; \alpha) = u_0(x; \alpha), \]

where $x \in \mathbb{R}^d$, $Q$ is a possibly nonlinear differential operator, and $\alpha \in \Omega \subset \mathbb{R}^m$ is a random variable which is distributed according to a continuous weight function $c(\alpha)$ such that $\int c(\alpha) d\alpha = 1$. The randomness of $u(t, x; \alpha)$ is due to the dependence of $Q$ and/or $u_0$ on $\alpha$.

For a given a quantity of interest $f(\alpha) := f(u(t, x); \alpha)$, we may wish to perform:

1. **Moment estimation.** Compute the mean, variance, or standard deviation of $f(\alpha)$:

\[
(2.2) \quad E_\alpha[f] := \int_\Omega f(\alpha) c(\alpha) d\alpha, \quad \text{Var}[f] := \int_\Omega [f(\alpha)]^2 c(\alpha) d\alpha - [E_\alpha[f]]^2, \quad \sigma(f) := \sqrt{\text{Var}[f]}. \]

2. **Density estimation.** Compute the probability distribution function (PDF) of $f(\alpha)$.

\[
(2.3a) \quad p(y) := \frac{dP(y)}{dy}, \quad y \in \mathbb{R}, \]

where $P$ is the cumulative distribution function (CDF),

\[
(2.3b) \quad P(y) := \text{Prob}\{f(\alpha) < y\}.
\]
2.1. Applications. Two examples of density-estimation in UQ which will be discussed in this paper are the effect of shot-to-shot variation in nonlinear optics (Sec. 7) and hydrodynamical shock formation (Sec. 8). We briefly present two other examples of density estimation in the UQ literature, for which our method can also be applied:

1. **Out-of-equilibrium chemical reactions.** Belousov-Zhabotinsky type systems model out-of-equilibrium chemical reactions. One concrete system is the Oregonator \[18\]

\[
\begin{align*}
\frac{dX}{dt} &= k_1 Y - k_2 XY + k_3 X - k_4 X^2, \\
\frac{dY}{dt} &= -k_1 Y - k_2 XY + k_5 Z, \\
\frac{dZ}{dt} &= k_3 X - k_5 Z,
\end{align*}
\]

where \(X, Y,\) and \(Z\) are the concentrations of three different chemical species, and \(\{k_i\}_{i=1}^5\) are the rate-parameters, often estimated empirically \[33\]. For large values of \(t\), this system exhibits sustained, temporal oscillations with a frequency \(F = F(k_1, \ldots, k_5)\). To deal with an uncertainty in the parameters \(k_4\) and \(k_5\), the authors of \[31\] computed the moments of \(X, Y, Z\), and the PDF of the oscillations frequency \(F\). This is an example of (2.1)–(2.3) with \(\alpha = (k_4, k_5)\) and \(f = X, Y, Z\) and \(F\).

2. **Heat convection.** Consider the flow of a fluid in a two-dimensional box \(x = (x, y) \in [x_1, x_2] \times [y_1, y_2]\), which is modeled by the Navier-Stokes like equations

\[
\begin{align*}
\nabla \cdot u &= 0, \\
\frac{\partial \theta}{\partial t} + u \cdot \nabla \theta &= \nabla^2 \theta, \\
\frac{\partial u}{\partial t} + u \cdot \nabla u &= -\nabla p + \Pr \nabla^2 u + F(u, \theta),
\end{align*}
\]

where \(u(t, x; \alpha)\) is the fluid velocity, \(p(t, x; \alpha)\) is the pressure, \(\theta(t, x; \alpha)\) is the temperature, \(\Pr\) is the Prandtl number, and \(F\) is the buoyant force \[20\]. The temperature is a known constant \(\theta_0\) on one side of the box, but is random on the other side, i.e.,

\[
\theta(t, x_1, y) \equiv \theta_0, \quad \theta(t, x_2, y) = \theta_1(y; \alpha).
\]

The PDF of the pressure and of the velocity were computed in \[48\] when \(\theta_1(y; \alpha) = \theta_1(\alpha)\) and \(\alpha\) is uniformly distributed in \([\alpha_{\min}, \alpha_{\max}]\), and in \[20\] when \(\theta_1(y; \alpha)\) is a Gaussian random process.

3. Review of existing methods. We briefly present the standard methods in the literature for (2.1)–(2.3).

3.1. **Monte-Carlo method, the histogram method, and Kernel Density Estimators.** Given \(N\) independently and identically distributed (iid) samples \(\{\alpha_j\}_{j=1}^N\), the simplest moment estimator is the Monte-Carlo approximation \(E_\alpha[f] \approx 1/N \sum_{n=1}^N f(\alpha_n)\). The Monte Carlo method is intuitive and easy to implement. The main drawback of this method is its slow \(O(1/\sqrt{N})\) convergence rate. In cases where each computation of \(f(\alpha_j)\) is expensive (e.g., when it requires to solve numerically (2.1) with \(\alpha = \alpha_j\)), this slow convergence rate can make the Monte-Carlo method impractical.

Density estimation using \(N\) iid samples of \(f(\alpha)\), denoted by \(\{f_j\}_{j=1}^N\), is a fundamental problem in non-parametric statistics. A widely-used method for density estimation is the histogram method,
in which one partitions the range of \( f(\alpha) \) into \( L \) disjoint bins \( \{B_\ell \}_{\ell=1}^L \), and approximates the PDF \( p \) with the histogram estimator

\[
(3.1) \quad p_{\text{hist}}(y) := \frac{1}{N} \sum_{\ell=1}^L \left( \text{# of samples for which } f_j \in B_\ell \right) \cdot \mathbb{1}_{B_\ell}(y),
\]

where \( \mathbb{1}_{B_\ell} \) is the characteristic function of bin \( B_\ell \). An alternative approach (which, unlike the histogram method, can provide a smooth PDF) is the Kernel Density Estimator (KDE)

\[
(3.2) \quad p_{\text{kde}}(y) := \frac{1}{Nh} \sum_{j=1}^N K \left( \frac{y - f_j}{h} \right),\]

where \( h > 0 \) is the “window size” and \( K \) is the kernel function (e.g., \( K(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2} \)), see [47, 53].

The \( L^1 \) error of KDE method asymptotically scales as \( N^{-\frac{2}{3}} \) [13]. As with the Monte-Carlo method, this rate is too slow when each evaluation of \( f_j \) is computationally expensive.

### 3.2. Generalized Polynomial Chaos

The Monte-Carlo method, the histogram method, and KDE are all statistical methods, in the sense that they only rely on the sampled values \( \{f_j\}_{j=1}^N \). Much more information can be extracted from \( \{f_j\}_{j=1}^N \) if the two following conditions hold:

1. The “original” \( \{\alpha_j\}_{j=1}^N \) for which \( f(\alpha_j) = f_j \) are known.
2. \( f(\alpha) \) is smooth.

These two conditions often hold in the general settings of Sec. 2. In such cases, a powerful numerical approach, known as generalized Polynomial Chaos (gPC), can be applied [21, 22, 34, 54].

For clarity, we review the gPC method for a one-dimensional random variable \( \alpha \), i.e., \( \Omega \subseteq \mathbb{R} \). We define the set of orthogonal polynomials \( \{p_n(x)\}_{n=0}^\infty \) with respect to \( c(\alpha) \) by the conditions [45]

\[
(3.3) \quad \text{Deg}(p_n) = n, \quad \int_{\Omega} p_n(\alpha)p_m^*(\alpha)c(\alpha)d\alpha = \delta_{n,m}.
\]

This family of orthogonal polynomials constitutes an orthonormal basis of the space of square integrable functions, i.e., for all \( f \in L^2(\Omega, c) \),

\[
(3.4) \quad f(\alpha) = \sum_{n=0}^\infty \hat{f}(n)p_n(\alpha), \quad \hat{f}(n) := \int_{\Omega} f(\alpha)p_n(\alpha)c(\alpha)d\alpha, \quad n = 0, 1, \ldots .
\]

This expansion converges spectrally for the classical families of orthogonal polynomials, e.g., the Hermite and Legendre polynomials. Specifically, if \( f \) is analytic, the truncated expansion (3.4) has the exponential accuracy

\[
(3.5) \quad \left\| f(\alpha) - \sum_{n=0}^{N-1} \hat{f}(n)p_n(\alpha) \right\|_2 \sim Ce^{-\gamma N}, \quad N \gg 1,
\]

for some constants \( C, \gamma > 0 \) [46, 52, 54].

---

1. The mean \( L^2 \) error (the squared root of the “MISE”), also asymptotically scales as \( N^{-\frac{2}{3}} \) [47, 53].

2. In Sec. 6.4 we show how our method can be extended to non-smooth functions.

3. i.e., if \( f \) is in \( C^r \), then \( \{\hat{f}(n)\} \leq cn^{-r} \), and if \( f \) is analytic, then \( |\hat{f}(n)| \leq ce^{-\gamma n} \), for some \( c, \gamma > 0 \).
The expansion coefficients \( \{ \hat{f}(n) \} \), see (3.4), can be approximated using the Gauss quadrature formula
\[
E_\alpha[g] \approx \sum_{j=1}^{N} g(\alpha_j) w_j,
\]
where \( \{ \alpha_j \}_{j=1}^{N} \) are the distinct and real roots of \( p_N(\alpha) \), \( w_j := \int_{\Omega} l_j(\alpha) \, d\mu(\alpha) \) are the weights, and \( l_j(\alpha) \) are the Lagrange interpolation polynomials with respect to \( \{ \alpha_j \}_{j=1}^{N} \) [8], yielding
\[
(3.6) \quad \hat{f}(n) \approx \hat{f}_N(n) := \sum_{j=1}^{N} \hat{f}(\alpha_j) p_n(\alpha_j) \, w_j, \quad n = 0, 1, \ldots, N - 1.
\]

The gPC collocation approximation is defined by
\[
(3.7) \quad f^\text{gpc}_N(\alpha) := \sum_{n=0}^{N-1} \hat{f}_N(n) p_n(\alpha),
\]
where \( \{ \hat{f}_N(n) \}_{n=0}^{N-1} \) are given by (3.6), see [55].

The spectral \( L^2 \) accuracy of the gPC approximation implies a similar accuracy for the approximation in moments:

**Corollary 3.1.** Let \( f \) be analytic, and let \( f^\text{gpc}_N \) be its gPC collocation approximation of order \( N \), see (3.7). Then the moments (2.2) of \( f \) can be approximated by the respective moments of \( f^\text{gpc}_N \) with exponential accuracy as \( N \to \infty \).

**Proof.** See Appendix A.

For a smooth quantity of interest \( f \), this spectral convergence rate is superior to the Monte-Carlo \( 1/\sqrt{N} \) convergence rate, which explains the popularity of the gPC collocation method.

In [39] we used the gPC approximation for moments and density estimation:

Because of its spectral accuracy (Corollary 3.1), the number of sample points that is required for gPC to achieve a certain precision is considerably smaller than for Monte-Carlo. To the best of our knowledge, however, there is no convergence result for density estimation using gPC which is analogous to Corollary 3.1.

Algorithm 3.1 can also approximate non-smooth quantities of interest \( f(\alpha) \), as long as \( u(\cdot; \alpha) \) is smooth, see Sec. 7 and [39]. The choice of the histogram method in step 4 is discussed in Sec. 9.

The evaluation of \( \{ f(u^\text{gpc}_N(\cdot, \tilde{\alpha}_m)) \}_{m=1}^{M} \) in step 3 is computationally cheap, as it amounts to a substitution in a polynomial. Therefore, there is essentially no computational cost for choosing \( M \) to be sufficiently high for the histogram method. This algorithm is also non-intrusive, in the sense that it only requires direct simulations of the deterministic system (2.1) with specific \( \alpha_j \) values (as opposed to, e.g., Galerkin-type methods [11, 30, 56]).

Our choice of the Histogram method for density estimation will be explained later in Sec. 4.1.

**4. Spline-based UQ.** Despite its many advantages, the gPC-based Algorithm 3.1 has several drawbacks:

1. The spectral accuracy of the gPC method for moments estimation does not imply a similar accuracy for density estimation. Indeed, although the gPC approximation \( f^\text{gpc}_N(\alpha) \) is frequently used for computing the PDF of \( f(\alpha) \), see e.g., [6], it can be very inaccurate for this computational task. Intuitively, this is because the PDF generally depends on \( (\frac{df}{d\alpha})^{-1} \), see Lemma 4.8. Because polynomial approximations tend to be oscillatory, they “add” many artificial extremal points, which produce large deviations from the exact PDF. These oscillations are intrinsic to polynomial interpolation, and tend to further increase as the interpolated data become less smooth.
Algorithm 3.1 gPC-based estimation [39]

Let \( \{\alpha_j, w_j\}_{j=1}^N \) be the points and weights of the Gaussian quadrature rule of order \( N \) that correspond to the weight function \( c(\alpha) \), and let \( \{p_n(\alpha)\}_{n=0}^\infty \) be the respective orthogonal polynomials.

1: For \( j = 1, \ldots, N \), solve (2.1) with \( \alpha = \alpha_j \) to obtain \( u(t, x; \alpha_j) \).
2: Approximate

\[
    u(t, x; \alpha) \approx u^{\text{gPC}}(t, x; \alpha),
\]

where

\[
    u^{\text{gPC}}(t, x; \alpha) := \sum_{n=0}^{N-1} \hat{u}_N(t, x; n)p_n(\alpha)
\]

and

\[
    \hat{u}_N(u, x; n) = \sum_{j=1}^{N} p_n(\alpha_j)u(t, x; \alpha_j)w_j, \quad n = 0, \ldots, N - 1.
\]

3: Approximate \( f(\tilde{\alpha}_m) \approx f(u^{\text{gPC}}(\cdot, \tilde{\alpha}_m)) \) on a sample of \( M \gg N \) points \( \{\tilde{\alpha}_m\}_{m=1}^M \) which are i.i.d. according to \( c(\alpha) \).
4: if goal is moment estimation: then
5: Use the trapezoidal integration rule with \( \{f(\tilde{\alpha}_m)\}_{m=1}^M \) to approximate \( \mathbb{E}_n[f] \).
6: else if goal is density estimation: then
7: Use the histogram method (3.1) with \( \{f(\tilde{\alpha}_m)\}_{m=1}^M \) to estimate the PDF of \( f \).
8: end if

2. The spectral convergence of the gPC method is attained only asymptotically as the number of sample points \( N \) becomes sufficiently large. For small or moderate values of \( N \), however, its accuracy may be quite poor, due to insufficient resolution, and the global nature of spectral approximation.
3. The sample points \( \{\alpha_j\}_{j=1}^N \) of the gPC method are predetermined by the quadrature rule.

Therefore, if one wants to adaptively improve the accuracy, one cannot use the samples from the “old” low-resolution grid in the “new” high-accuracy approximation.

To overcome these limitations, we first note that \( f^{\text{gPC}}(\alpha) \) is an interpolating polynomial of \( f \):

Lemma 4.1. The gPC collocation approximation (3.7) is the interpolating polynomial of \( f \) of order \( N - 1 \) at the Gauss quadrature points \( \{\alpha_j\}_{j=1}^N \).

Proof. See Appendix B.

This suggests that other interpolants of \( f(\alpha) \) can be used in Algorithm 3.1. In what follows, we argue that for our computational tasks, splines provide a better way to approximate \( f(\alpha) \).

We recall that splines are functions that are piecewise polynomial of degree \( m \), with \( k < m \) smooth derivatives. Given an interval \( \Omega = [\alpha_{\min}, \alpha_{\max}] \) and a grid \( \alpha_{\min} = \alpha_1 < \alpha_2 < \cdots < \alpha_N = \alpha_{\max} \), the interpolating cubic spline \( s_N(\alpha) \) is a \( C^2 \), piecewise-cubic polynomial that interpolates \( f(\alpha) \) at \( \{\alpha_j\}_{j=1}^N \), endowed with two additional boundary conditions. Three standard choices are:

1. The natural cubic spline, for which \( \frac{d}{d\alpha} s^{\text{spline}}_N(\alpha_1) = \frac{d}{d\alpha} s^{\text{spline}}_N(\alpha_N) = 0 \).
2. The “not-a-knot” spline, for which \( \frac{d}{d\alpha} s^{\text{spline}}_N(\alpha_j) \) is continuous at \( \alpha_2 \) and \( \alpha_{N-1} \).
3. The clamped spline, for which \( \frac{d}{d\alpha} s^{\text{spline}}_N(\alpha_1) = \frac{d}{d\alpha} f(\alpha_j) \) for \( j = 1, N \).

Our decision to use splines is motivated by the following reasons:
1. The error of spline interpolation is guaranteed to be “small” for any sample size, in the following sense:

**Theorem 4.2 ([24, 3]).** Let \( f \in C^4([\alpha_{\min}, \alpha_{\max}]) \), and let \( f_N^{\text{spline}} \) be its “not-a-knot”, clamped or natural cubic spline interpolant. Then

\[
(4.1) \quad \| (f(\alpha) - f_N^{\text{spline}}(\alpha))^{(j)} \|_{L^\infty[\alpha_{\min}, \alpha_{\max}]} \leq C_{\text{spl}}^{(j)} \| f^{(4)} \|_\infty h_{\text{max}}^{-j}, \quad j = 0, 1, 2,
\]

where \( C_{\text{spl}}^{(j)} > 0 \) is a universal constant that depends only on the type of boundary condition and \( j \), and \( h_{\text{max}} = \max_{1 < j \leq N} |\alpha_j - \alpha_{j-1}| \).

2. Spline interpolation is predominantly local. To see this, let us first recall a classical result of Birkhoff and de Boor:

**Theorem 4.3 ([5, 9]).** Let \( s_i(\alpha) \) be the natural cubic spline that satisfies \( s_i(\alpha_k) = \delta_{i,k} \), where \( 1 \leq i, k \leq N \) and \( \alpha_{\min} = \alpha_1 < \alpha_2 < \cdots < \alpha_N = \alpha_{\max} \) is given. Then

\[
\max_{\alpha \notin (\alpha_{i-k}, \alpha_{i+k})} |s_i(\alpha)| \leq A 2^{-k}, \quad 1 < i < N,
\]

where \( A > 0 \) is a constant that depends on the global mesh ratio \( \frac{\max_{1 \leq i \leq N} \alpha_j - \alpha_{j-1}}{\min_{1 < k \leq N} \alpha_k - \alpha_{k-1}} \).

Therefore, the natural cubic spline \( f_N^{\text{spline}}(\alpha) \) is essentially a local approximation:

**Corollary 4.4.** Denote the natural cubic spline \( f_N^{\text{spline}} = f_N^{\text{spline}}(\alpha; f_1, \ldots, f_N) \) to emphasize the dependence of the spline interpolation on the sampled values. Then

\[
\max_{\alpha \notin (\alpha_{i-k}, \alpha_{i+k})} \left| \frac{\partial f_N^{\text{spline}}(\alpha; f_1, \ldots, f_N)}{\partial f_i} \right| \leq A 2^{-k}, \quad 1 < i < N, \quad 1 \leq k \leq N,
\]

where \( A > 0 \) is given by Theorem 4.3.

**Proof.** The function \( S(\alpha) = \sum_{i=1}^{N} f_i s_i(\alpha) \), where \( s_i(\alpha) \) are defined in Theorem 4.3, is a \( C^2 \) cubic spline, which by definition satisfies \( S(\alpha_i) = f_i \), and \( \frac{d}{d\alpha} S(\alpha_1) = \frac{d}{d\alpha} S(\alpha_N) = 0 \). By the uniqueness of the natural cubic spline, \( S(\alpha) = f_N^{\text{gPC}}(\alpha) \), so, \( \frac{\partial f_N^{\text{spline}}(\alpha; f_1, \ldots, f_N)}{\partial f_i} = s_i(\alpha) \). Hence, by Theorem 4.3, the corollary is proven.

**Remark 4.5.** For a proof of an equivalent result, which is valid also for the “not-a-knot” boundary condition, see Appendix C.

Thus, although \( f_N^{\text{spline}}(\alpha) \) depends on \( \{f_1, \ldots, f_N\} \), it predominantly depends on the few values \( f_j \) for which \( \alpha_j \) is adjacent to \( \alpha \). Therefore, large derivatives and discontinuities of \( f(\alpha) \) may impair the accuracy of \( f_N^{\text{spline}}(\alpha) \) only locally.\(^5\) This is in contrast to gPC (and polynomial interpolation in general), where discontinuities and large derivatives of \( f \) decrease the approximation accuracy across the entire domain.

Unlike gPC, splines can be constructed using any choice of sampling points.

In light of these considerations, we propose to replace the gPC interpolant with a spline:

---

\(^5\)For a review of cubic splines that are strictly local, see [4].
Algorithm 4.1 Spline-based estimation
Let $\Lambda = \{\alpha_1, \ldots, \alpha_N\}$ be a uniform grid on $[\alpha_{\min}, \alpha_{\max}]$.

1: For each $\alpha_j \in \Lambda$, solve (2.1) with $\alpha = \alpha_j$ to obtain $u(t, x; \alpha_j)$.
2: Approximate $u(t, x; \alpha) \approx u_N^{\text{spline}}(t, x; \alpha)$, where $u_N^{\text{spline}}$ is a cubic spline interpolant on $\Lambda$.
3: Approximate $f(\tilde{\alpha}_m) \approx f(u_N^{\text{gpc}}(\cdot, \tilde{\alpha}_m))$ on a sample of $M \gg N$ points $\{\tilde{\alpha}_m\}_{m=1}^M$ which are i.i.d. according to $c(\alpha)$.
4: if goal is moment estimation: then
5: Use the trapezoidal integration rule with $\{f(\tilde{\alpha}_m)\}_{m=1}^M$ to approximate $E_{\alpha}[f]$.
6: else if goal is density estimation: then
7: Use the histogram method (3.1) with $\{f(\tilde{\alpha}_m)\}_{m=1}^M$ to approximate the PDF of $f$.
8: end if

Remark 4.6. See Appendix E for a MATLAB implementation of this algorithm.

Which cubic spline should be used in line 2? If $f'(\alpha_{\min})$ and $f'(\alpha_{\max})$ are known, then one should use the clamped cubic spline (or the natural cubic spline if these derivatives are zero). When the boundary derivatives are unknown, however, the “not-a-knot” interpolating cubic spline should be used (as indeed was done in this manuscript). See [4] for further discussion.

Algorithm 4.1 is identical to Algorithm 3.1, except for two substantial points:
1. The sampling grid is uniform, rather than the Gauss quadrature grid.\(^6\)
2. The gPC interpolant $u_N^{\text{gpc}}$ is replaced by a cubic spline interpolant $u_N^{\text{spline}}$.

Remark 4.7. This method is not to be confused with spline-smoothing, in which one approximates the PDF $p$ with splines [15, 50]. Thus, Algorithm 4.1 approximates $u$ with a spline, but the resulting approximation of the PDF $p$ is not a spline.

4.1. Accuracy of Algorithm 4.1 for density estimation. The density estimation error of Algorithm 4.1 has two components - the error of the spline approximation (line 3) and that of the histogram method (line 7).

The accuracy of the histogram method in line 7 depends on the number of bins $L$ and on the number of samples $M$ at lines 3 and 7. If the number of bins is chosen to be

$$L_{\text{opt}} = K_f M^{-\frac{1}{3}}, \quad K_f = \left(\frac{\|f'\|_2^2 \max f - \min f}{6}\right)^{\frac{1}{2}},$$

the mean squared $L^2$ error (MISE) of the histogram method decays as $M^{-\frac{2}{3}}$ [53].\(^8\) Because the computational cost of increasing $L$ and $M$ is negligible, they can be set sufficiently large so that the accuracy of Algorithm 4.1 mainly depends on the difference between the PDFs of $f$ and $f_N^{\text{spline}}$, denoted by $p$ and $\tilde{p}$ respectively. We motivate the choice of the histogram method to estimate the density by four factors:

1. Implementing the histogram method is straightforward, and can be done with a few lines of code (see Appendix E).

\(^6\)Algorithm 4.1 can be performed with any choice of grid points. For clarity, we present it only with a uniform grid.

\(^7\)In terms of density estimators, this can be explained by the following argument. Denote by $p, p_N$, and $\hat{p}_{N,M}$ the density of $f, f_N$ and the density estimator of Algorithm 3.1 or 4.1, respectively. Then the approximation error (in any norm) satisfies $\|p - \hat{p}_{N,M}\| \leq \|p - p_N\| + \|p_N - \hat{p}_{N,M}\|$. The second term vanishes as $M \to \infty$ and $L$ is given by (4.2), in which case the density estimation error is roughly the bias incurred from approximating $f$ by $f_N$.

\(^8\)In practice, $f$ and $f'$ are often unknown, and so $K_f$ needs to be estimated.
2. The accuracy of the histogram method can be improved and controlled by varying the number of samples $M$, with a negligible computational cost.

3. The histogram method can be used even when the quantity of interest $f$ is not smooth.

4. The histogram method can be used for a multi-dimensional random parameter $\alpha$.

In principle, we could have used the explicit relation (4.3) to compute the PDF. Because this approach does not have the above advantages, however, the histogram method was chosen.

### 4.1.1. Accuracy of spline-based density estimation.

To explain why splines are a good choice for density estimation, we first prove Lemma 4.8.

**Lemma 4.8.** Let $f$ be a real, piecewise monotone, continuously differentiable function on $[a, b]$, where $-\infty \leq a < b \leq \infty$, and let $\mu$ be an absolutely continuous probability measure on $[a, b]$, i.e., there is $c \in L^1([a, b])$ such that $d\mu(\alpha) = c(\alpha)d\alpha$. Then

\[
p(y) = \sum_{f(\alpha_j) = y} \frac{c(\alpha_j)}{|f'(\alpha_j)|},
\]

where $p(y)$ is the PDF of $f$.

**Proof.** See Appendix D.

Therefore, if $f'(\alpha)$ is bounded away from zero, then $p$ is smooth. As noted, however, the gPC polynomial interpolant $f_{N}^{gpc}(\alpha)$ tends to be oscillatory, and so it might add artificial external points where $\frac{d}{d\alpha} f_{N}^{gpc}(\alpha) = 0$, see e.g., Fig. 2(c). At every such point where $\frac{d}{d\alpha} f_{N}^{gpc}(\alpha) = 0$, the PDF approximation becomes unbounded, and so a large error in the PDF estimation occurs. This is seldom the case with the spline interpolant, which due to its local nature (see Lemma 4.4) does not produce numerical oscillations throughout its domain $\Omega$. Indeed, the natural cubic spline $f_{N}^{spline}(\alpha)$ has the “minimum curvature” property [36], which implies that it oscillates “very little” about the original function. This notion is made precise by the following result:

**Theorem 4.9.** Let $f \in C^4([\alpha_{\min}, \alpha_{\max}])$ with $|f'(\alpha)| \geq a > 0$, let $\alpha$ be distributed by $c(\alpha)d\alpha$, where $c \in C^1([\alpha_{\min}, \alpha_{\max}])$, and let $p$ and $\tilde{p}$ be the PDFs of $f(\alpha)$ and of its natural, “not-a-knot”, or clamped cubic spline interpolant on a uniform grid of size $N$, respectively. Then

\[
\|p - \tilde{p}\|_1 \leq KN^{-3}, \quad N > \sqrt{\frac{2C_{\text{spl}}^{(1)} \|f'(\alpha)\|_{L^\infty}}{a}} (\alpha_{\max} - \alpha_{\min}),
\]

where $C_{\text{spl}}^{(1)}$ is given by Theorem 4.2 and $K$ depends only on $f(\alpha)$, $c(\alpha)$, and $|\alpha_{\max} - \alpha_{\min}|$.

**Proof.** See Appendix F.

**Remark 4.10.** If $f$ is only piecewise $C^4$, the $N^{-3}$ convergence is guaranteed when the grid points include the discontinuity points of $f(\alpha)$, since the proof can be repeated in each interval on which the function is $C^4$ in the same way.

**Remark 4.11.** Although Theorem 4.9 applies only to functions whose derivatives are bounded away from 0, in practice we observe cubic convergence for non-monotone functions as well (see Sec. 7). Whether Theorem 4.9 generalizes to non-monotone cases is unclear.

In our numerical simulations, see Figs. 2, 4, 8, and 9, we observe that the cubic convergence is often reached well before $N$ satisfies (4.4). We also observe that the density approximation error $\|p - \tilde{p}\|_1$ decays at a faster than cubic rate. A possible explanation for this observation is provided by
Lemma 4.12. Assume the conditions of Theorem 4.9, and let $J_N$ be the number of times that \( \frac{d}{d\alpha}(f(\alpha) - f_N^{\text{spline}}(\alpha)) \) changes its sign on \([\alpha_{\min}, \alpha_{\max}]\). If $J_N = O(N^r)$ for $0 \leq r \leq 1$, then \( \|p - \tilde{p}\|_1 \leq KN^{-4+r} \). Specifically, if $J_N$ is uniformly bounded for all $N \in \mathbb{N}$, then \( \|p - \tilde{p}\|_1 \leq KN^{-4} \).

Proof. See Appendix G.

4.2. Accuracy of moment estimation. Similarly to density estimation, the error of the moment estimation of Algorithm 4.1 comes from both the numerical integration (line 5) and interpolation (line 2). The trapezoidal rule integration error can be made sufficiently small by increasing the number of samples $M$ at line 3, at a negligible computational cost. Moreover, if $c(\alpha) \equiv 1$, the integration over $f_N^{\text{spline}}$ can be done exactly.\(^9\) Hence, the moment estimation error of Algorithm 4.1 is determined by the accuracy of the spline interpolation:

**Corollary 4.13.** Let $f \in C^4([\alpha_{\min}, \alpha_{\max}])$, let $f_N^{\text{spline}}$ be the natural, “not-a-knot”, or clamped cubic spline interpolant of $f$, and let $\alpha$ be distributed by $c(\alpha)d\alpha$, where $c(\alpha) \geq 0$, and $\int_{\alpha_{\min}}^{\alpha_{\max}} c(\alpha)d\alpha = 1$. Then

\[
|E_\alpha[f] - E_\alpha[f_N^{\text{spline}}]| \leq C_{\text{spl}}^{(0)} \|f\|_\infty h_{\text{max}}^4,
\]

where $C_{\text{spl}}^{(0)}$ and $h_{\text{max}}$ are defined in Theorem 4.2.

**Proof.** By Theorem 4.2, \( \|f - f_N^{\text{spline}}\|_\infty \leq C_{\text{spl}}^{(0)} \|f^{(4)}\|_\infty h_{\text{max}}^4 \). Hence,

\[
\left| \int_{\alpha_{\min}}^{\alpha_{\max}} (f(\alpha) - f_N^{\text{spline}}(\alpha))c(\alpha)d\alpha \right| \leq \|f - f_N^{\text{spline}}\|_\infty \int_{\alpha_{\min}}^{\alpha_{\max}} c(\alpha)d\alpha \leq \|f - f_N^{\text{spline}}\|_\infty \cdot 1 \leq C_{\text{spl}}^{(0)} \|f^{(4)}\|_\infty h_{\text{max}}^4.
\]

Typically, $C_{\text{spl}}^{(0)} < 1$. For example, for the natural and “not-a-knot” cubic spline, $C_{\text{spl}}^{(0)}$ is equal to $\frac{5}{384}$ and $\frac{1}{25}$, respectively [24, 4]. On a uniform grid, $h_j = \frac{\alpha_{\max} - \alpha_{\min}}{N-1}$ for $1 < j \leq N$, and so $E_\alpha[f] - E_\alpha[f_N^{\text{spline}}] = O(N^{-4})$.

As $N \to \infty$, the polynomial convergence rate of the spline approximation (Corollary 4.13) is outperformed by gPC’s spectral convergence rate (Corollary 3.1). Quite often, however, the spline approximation is more accurate for moderate $N$ values. To see that, note that by (3.3), (3.6), and (3.7),

\[
E_\alpha[f_N^{\text{gpc}}] = \sum_{j=1}^{N} f(\alpha_j)w_j,
\]

which is the Gauss quadrature rule. Hence, if $f \in C^{2N}$, then

\[
E_\alpha[f] - E_\alpha[f_N^{\text{gpc}}] = \frac{f^{(2N)}(\xi)}{k_N^2(2N)!}, \quad \xi \in (\alpha_{\min}, \alpha_{\max}),
\]

where $k_N$ is the leading coefficient of $p_N(\alpha)$ [8]. If for small $N$, $\|f^{(2N)}\|_\infty$ increases faster than $k_N^2(2N)!$, the error initially increases with $N$. In these cases, the exponential convergence is only achieved at large $N$.\(^{10}\) Even when gPC does converge exponentially, i.e., $|E_\alpha[f] - E_\alpha[f_N^{\text{gpc}}]| \leq Ke^{-\gamma N}$, if $\gamma$ is small, the error of the spline approximation may be smaller for moderate values of $N$, see e.g., Fig. 1(c).

\(^9\)When $f$ is sufficiently smooth and $\alpha$ is uniformly distributed, one can approximate $E_\alpha[f] \approx E_\alpha[f_N^{\text{spline}}]$, and compute the right-hand side explicitly (in MATLAB, this can be done using the fnint command).

\(^{10}\)For example, if the numerator grows as $K^{2N}$, the error only decays for $N > K$. 

10
To conclude, unlike for gPC, the accuracy of spline-based moment approximation is guaranteed also with few samples.

5. Multi-dimensional noises. To generalize the spline-based density-estimation approach (Algorithm 4.1) to the case where \( \mathbf{\alpha} \in \Omega = [0,1]^d \), we use tensor-product splines, which are defined in the following way. Let \( m \geq 1 \), let \( f(\mathbf{\alpha}) \in C^{m+1}(\Omega) \), let \( \Lambda \) be the one-dimensional grid \( 0 = \alpha_1 < \cdots < \alpha_n = 1 \), and let \( \Lambda^d \) be the respective \( d \)-dimensional tensor-product grid. An \( m \)-th degree tensor-product spline interpolant of \( f \) is a function \( s(\mathbf{\alpha}) \in C^{m-1}(\Omega) \) that interpolates \( f \) on \( \Lambda^d \) and reduces to a one-dimensional \( m \)-th degree spline on every line on \( \Lambda^d \), see [41] for a more precise definition. The multidimensional extension of Algorithm 4.1 for density estimation is

Algorithm 5.1 Multidimensional spline-based density estimation

Let \( \Lambda^d = \{\alpha_1, \ldots, \alpha_N\}_D^D \) be a tensor-product uniform grid on \([0,1]^d\).
1. For each \( \mathbf{\alpha}_j \in \Lambda^d \), solve (2.1) with \( \alpha = \mathbf{\alpha}_j \) to obtain \( u(t,x;\mathbf{\alpha}_j) \).
2. Approximate \( u(t,x;\mathbf{\alpha}) \approx u_N^{\text{spline}}(t,x;\mathbf{\alpha}) \), where \( u_N^{\text{spline}} \) is a tensor-product spline interpolant of order \( m \) on \( \Lambda^d \).
3. Approximate \( f(\tilde{\mathbf{\alpha}}_m) \approx f(u_N^{\text{gpc}}(\cdot,\tilde{\mathbf{\alpha}}_m)) \) on a sample of \( M \gg N \) points \( \{\tilde{\mathbf{\alpha}}_m\}_{m=1}^M \) which are i.i.d. according to \( c(\mathbf{\alpha}) \).
4. Use the histogram method (3.1) with \( \{f(\tilde{\mathbf{\alpha}}_m)\}_{m=1}^M \) to approximate the PDF of \( f \).

As in the one-dimensional Algorithm 4.1, the analysis of the density-estimation error in Algorithm 5.1 is based on two components:

1. A pointwise error bound for tensor-product spline interpolants, due to Schultz:

   Theorem 5.1 ([38, 41]). Let \( \Omega = [0,1]^d \), \( f \in C^{m+1}(\Omega) \), and let \( s(\mathbf{\alpha}) \) be its \( m \)-th degree tensor-product spline interpolant. Then for any \( \mathbf{\alpha} \in \Omega \),

   \[
   |D^j(f - s)| < C_m h^{m+1-j}, \quad j = 0,1, \ldots m - 1, \tag{5.1}
   \]

   where \( D^j \) is any \( j \)-th order derivative,\(^{13} \) \( C_m = C_m(\|D^{m+1}f\|_{\infty}) \) depends only on the \( L^\infty \) norms of the \( m+1 \) order derivatives of \( f \), and \( h = \max_{1 \leq j < n} |\alpha_j + 1 - \alpha_j| \).

2. A multi-dimensional generalization of Lemma 4.8.\(^{14} \)

   Lemma 5.2. Let \( \Omega \subset \mathbb{R}^d \) be a Jordan set, denote by \( |\cdot| \) the Euclidean norm in \( \mathbb{R}^d \), let \( f \) be piecewise-differentiable with \( |\nabla f| \neq 0 \) on \( \Omega \), let \( \mathbf{\alpha} \) be an absolutely-continuous random variable in \( \Omega \), i.e., \( d\mu(\mathbf{\alpha}) = c(\mathbf{\alpha})d\mathbf{\alpha} \) for some non-negative \( c \in L^1(\Omega) \), and denote the PDF associated with \( f(\mathbf{\alpha}) \) by \( p \). Then

   \[
   p(y) = \frac{1}{\mu(\Omega)} \int_{f^{-1}(y)} \frac{c(\mathbf{\alpha})}{|\nabla f(\mathbf{\alpha})|} d\sigma, \tag{5.2}
   \]

   where \( d\sigma \) is a \( (d-1) \) dimensional surface element on \( f^{-1}(y) \).

---

\(^{13}\) i.e., when \( d - 1 \) coordinates of \( \mathbf{\alpha} \) are fixed in \( \Lambda \).

\(^{14}\) \( s(\mathbf{\alpha}) \) is unique when endowed with sufficiently many boundary conditions, see the discussion on the one-dimensional case in Sec. 4. Theorem 5.1 holds for many possible choices of boundary conditions, including the not-a-knot conditions which we have also used in our simulations.

\(^{13}\) More explicitly, \( D^j = \prod_{k=1}^d (\partial_{\alpha_k})^{\ell_k} \) where \( \ell_1 + \cdots + \ell_d = j \), and each \( \ell_k \) is a non-negative integer.

\(^{14}\) When \( \Omega \subset \mathbb{R}^d \) is a one-dimensional interval, Lemma 5.2 reduces to Lemma 4.8. Indeed, since \( |f'| \neq 0 \) on \( \Omega \) then \( f \) is piece-wise monotonic, and so \( f^{-1}(y) \) consists of a finite number of points. In addition, the surface element \( d\sigma \) is a point-mass distribution. Hence, (5.2) reduces to (4.3).
The generalization of Theorem 4.9 to the case of multidimensional noise is as follows:

**Theorem 5.3.** Let \( \Omega = [0,1]^d \), let \( m \geq 1 \), let \( f \in C^{m+1}(\Omega) \), let \( s \) be the \( m \)-degree tensor-product spline interpolant of \( f \), let \( \alpha \) be uniformly distributed in \( \Omega \), and let \( p \) and \( \hat{p} \) be the PDFs of \( f \) and \( s \), respectively. If \( \kappa := \min_{\Omega} |\nabla f| > 0 \), then for sufficiently small \( h \),

\[
\|p - \hat{p}\|_1 \leq K h^m,
\]

for some constant \( K > 0 \), where \( h \) is defined in Theorem 5.1.

**Proof.** See Appendix I.

Theorem 5.3 can be extended to any approximation \( \hat{f} \) of \( f \) and to any bounded domain \( \Omega \subseteq \mathbb{R}^d \), provided that the bound (5.1) holds for \( j = 0 \) and \( j = 1 \).

The total number of sample points in the special case where \( \Lambda \) is the uniform one-dimensional grid on \([0,1]\) is \( N = n^d \sim h^{-d} \). Therefore,

**Corollary 5.4.** Let \( \Lambda \) be the uniform grid on \([0,1]\). Then under the conditions of Theorem 5.3, then for sufficiently large \( N \),

\[
\|p - \hat{p}\|_1 \leq KN^{-\frac{m}{d}},
\]

for some constant \( K > 0 \).

As noted in Sec. 3.1, the \( L^1 \) error of the KDE method asymptotically scales as \( N^{-\frac{3}{5}} \) [13]. Therefore, by Corollary 5.4, Algorithm 5.1 outperforms KDEs for dimensions \( d \leq \frac{5}{2}m \).

6. Simulations. In this section, we compute the density and the moments of the function

\[
f(\alpha) = \tanh(9\alpha) + \frac{\alpha^2}{2}, \quad \alpha \in [-1,1],
\]

which is smooth but has a narrow high-derivative region.\(^{15}\)

![Figure 1](image.png)

**Figure 1.** (a) \( f(\alpha) \) (solid), see (6.1), and its spline interpolant (dashes) are nearly indistinguishable, whereas the gPC interpolant (dots) oscillates “around” \( f \). Both interpolants use \( N = 12 \) grid points. (b) \( L^2 \) error of both interpolants as a function of the number of samples. (c) Error of the standard deviation when it is approximated using Monte-Carlo method (dash-dot), the gPC-based method (dots) and the spline-based method (dashes).

6.1. Interpolation. With \( N = 12 \) samples, the spline interpolant \( f_N^{\text{spline}} \) of (6.1) is nearly indistinguishable from \( f \), whereas the gPC interpolant \( f_N^{\text{gPC}} \) slightly oscillates “around” \( f \), see Fig.

\(^{15}\)The \( \frac{\alpha^2}{2} \) term was added so that \( \frac{df}{d\alpha} \) is bounded away from zero, in order to prevent singularities in the PDF, see Sec. 6.3.
Although \( f_N^{\text{gpc}} \) converges exponentially to \( f \) in \( L^2 \), see Fig. 1(b), its \( L^2 \) approximation error \( \|f - f_N\|_2 = \left( \int_{-1}^{1} |f(\alpha) - f_N(\alpha)|^2 \, d\alpha \right)^{\frac{1}{2}} \) with few samples (10 \( \leq N \leq 40 \)) is larger than that of the spline interpolant by more than an order of magnitude. With sufficiently many samples \( (N > 70) \), however, the gPC approximation exponential convergence outperforms the spline's polynomial convergence rate. This example shows that with few samples, the occurrence of a “jump” in \( f \) hurts the accuracy of the gPC interpolant. Spline interpolation, on the other hand, is less sensitive to the “jump”, because it “confines” the approximation error induced by the jump to the jump interval (roughly \( \alpha \in (-0.1, 0.1) \)), see Lemma 4.4.

### 6.2. Moment approximation.

The interpolation accuracy is relevant to moment approximation, because a small \( L^2 \) error implies a small moment-approximation error (Lemma A.1). For example, Fig. 1(c) shows the standard deviation error \( |\sigma(f) - \sigma(f_N)| \), see (6.1), when \( \alpha \) is uniformly distributed in \([-1, 1]\). As expected, the spline-based method (Algorithm 4.1) is more accurate than the gPC-based method (Algorithm 3.1) with few samples, but the gPC is more accurate with sufficiently many samples. A purely statistical approach such as Monte-Carlo converges poorly compared to both the spline and gPC approach, with about 10% error with \( N \leq 100 \) sample points.

![Figure 2](attachment:image.png)

**Figure 2.** The PDF of \( f(\alpha) \), see (6.1), where \( \alpha \) is uniformly distributed in \([-1, 1]\). (a) exact PDF (solid) and its approximation by the gPC-based Algorithm 3.1 (dots) with \( N = 18 \) sample points. (b) Same, with the spline-based Algorithm 4.1 (dashes). The two lines are nearly indistinguishable. (c) Derivatives of \( f \) (solid), \( f_N^{\text{spline}} \) (dashes) and \( f_N^{\text{gpc}} \) (dots). (d) \( L^1 \) error of the PDF approximations as a function of the number of sample points, for the KDE (dash-dot), gPC-based approximation (dots), the spline-based approximation (dashes), and its power-law fit 103.2\( N^{-3.29} \) (solid).

### 6.3. Density estimation.

Consider the PDF induced by \( f(\alpha) \), see (6.1), when \( \alpha \) is uniformly distributed in \([-1, 1]\). The PDF computed by the gPC-based Algorithm 3.1 with \( N = 18 \) sample points deviates considerably from the exact PDF, see Fig. 2(a), whereas the PDF computed by the spline-based Algorithm 4.1 with \( N = 18 \) sample points is nearly indistinguishable from the exact
PDF, see Fig. 2(b). This is consistent with our discussion in Sec. 4. Indeed, the derivative of the spline interpolant \( \frac{d}{d\alpha} f_N^{\text{spline}} \) approximates \( f'(\alpha) \) with cubic accuracy, whereas the derivative of the gPC interpolant \( \frac{d}{d\alpha} f_N^{\text{gPC}} \) has many artificial extremal points where \( \frac{d}{d\alpha} f_N^{\text{gPC}}(\alpha) = 0 \), but \( \frac{d}{d\alpha} f(\alpha) \neq 0 \), see Fig. 2(c).

The \( L^1 \) distance \( \|p - \tilde{p}\|_1 \) between the exact PDF \( p \) and its approximation \( \tilde{p} \) is presented in Fig. 2(d). For \( 10 \leq N \leq 100 \) the spline-based approximation is more accurate than the gPC-based one by nearly two orders of magnitude. This is in contrast to moment estimation, see Fig. 1(c), in which the gPC approximation becomes more accurate for \( N \geq 40 \). Furthermore, we observe numerically that the spline-based method converges even faster than the \( N^{-3} \) rate predicted by Theorem 4.9. The KDE approximation has roughly 10\% error for \( N \leq 100 \). Other frequently-used distances between distributions, such as the Hellinger distance \( \frac{1}{\sqrt{2}} \|\sqrt{p} - \sqrt{\tilde{p}}\|_2 \) [28] and the Kullback-Leibler (KL) Divergence\(^{18} \) [27]

\[
\int_{-\infty}^{\infty} p(y) \log \left( \frac{p(y)}{\tilde{p}(y)} \right) dy ,
\]

produce similar results (data not shown).

6.4. Density estimation of non-smooth functions. Let

\[
g(\alpha) = f(\alpha) \mod (0.7),
\]

where \( f \) is given by (6.1).\(^{19} \) Because (6.3) is non-smooth, with few samples neither the spline, nor the gPC interpolant are even remotely close to \( g(\alpha) \), see Fig. 3. Therefore, to approximate the PDF associated with \( g(\alpha) \), we first use Algorithms 3.1 and 4.1 to approximate \( f(\alpha) \approx f_N(\alpha) \). Since \( f \) is smooth, both approximations are reasonable with few samples, see Fig. 1. Next, we approximate \( g(\alpha_m) \approx f_N(\alpha_m) \mod (0.7) \), and compute the PDF of \( g \) using the histogram method on a high-resolution sampling grid (\( M = 2 \cdot 10^6 \)). We again stress that evaluating \( f_N \) is computationally cheap, and therefore can be easily done with such a large sample. As in the smooth case, see Fig. 2, the PDF approximated by the gPC-based Algorithm 3.1 with \( N = 18 \) sample points has large deviations and converges poorly, see Fig. 4(a), whereas the PDF approximated by the spline-based Algorithm 4.1 with \( N = 18 \) sample points is nearly identical to the exact PDF, see Fig. 4(b). Indeed the \( L^1 \) error of spline-based PDF is smaller than that of the gPC-based PDF by at least an order of magnitude, for \( 20 < N < 50 \), see Fig. 4(c). Although Theorem 4.9 applies only to \( C^4 \) functions, we observe numerically that the convergence rate of the spline-based PDF is faster than \( N^{-3} \). The KDE approximation for the PDF of \( g(\alpha) \) is poorer than that of the spline-based and gPC-based approximations.

6.5. Multidimensional noise. To numerically confirm the error bound of the density estimation (Algorithm 5.1) for \( d > 1 \), we first consider the two-dimensional function

\[
f_{2d}(\alpha_1, \alpha_2) = \tanh(6\alpha_1 \alpha_2 + \alpha_1/2) + (\alpha_1 + \alpha_2)/3 .
\]

\(^{16} \)The MATLAB code that generates this PDF approximation is given in Appendix E.

\(^{17} \)The poor accuracy of the KDE method is due to the fact that the KDE does not use the “functional information” \( \{f_j = f(\alpha_j)\}_{j=1}^N \), but only the set \( \{f_j\}_{c=1}^N \).

\(^{18} \)Intuitively, the \( d_{KL} \) measures the entropy added, or conversely, the information lost, in approximating \( p \) by \( \tilde{p} \).

\(^{19} \)This example is motivated by our study of the NLS [39], where the cumulative phase \( \psi(t; \alpha) = \arg [\psi(t; 0; \alpha)] \) is smooth, but the quantity of interest, the angle \( \phi \mod (2\pi) \), is discontinuous. See Sec. 7. for another optics application which motivates this example.
where $\alpha_1$ and $\alpha_2$ are independent and uniformly distributed in $[-1, 1]$. As in the one-dimensional example, see (6.1), $f_2d$ is analytic with high-gradients regions, see Fig. 5(a). The spline-based PDF approximation with $N = 8^2$ sample points is very close to the exact PDF of $f(\alpha_1, \alpha_2)$, whereas the gPC-based PDF deviates from it substantially (Fig. 5(b)). The convergence rate of Algorithm 5.1 with cubic splines is $N^{-2.15}$ (Fig. 5(c)), which is consistent with the theoretical $N^{-2/3}$ error bound (Corollary 5.4). The convergence rates of both the KDE and the gPC methods are considerably slower for “small” sample sizes ($N \leq 200$).

Next, consider the three-dimensional function

$$ f_{3d}(\alpha_1, \alpha_2, \alpha_3) = \tanh(8\alpha_1 + 5\alpha_2 + 10\alpha_3) + (\alpha_1 + \alpha_2 + \alpha_3)/3, $$

where $\alpha_1$, $\alpha_2$, and $\alpha_3$ are independent and uniformly distributed in $[-1, 1]$. The spline-based PDF with $N = 10^3$ sample points approximates the exact PDF well, see Fig. 6(a), and its convergence rate is $N^{-1.1}$ (see Fig. 6(b)), which is consistent with the theoretical $N^{-1}$ convergence rate (Corollary 5.4). For comparison, the fitted convergence rate of the KDE is $N^{-0.39}$, which is consistent with the theoretical $N^{-2/5}$ rate [13]. Therefore, the spline-based method is more accurate than the KDE for sufficiently many samples ($N > 10^6$). For smaller values of $N$ (e.g., $N = 216$), however, the KDE achieves a slightly better accuracy than the spline-based method. This can be explained by what is known as the “curse of dimensionality”. Thus, in the three-dimensional tensor-grid spline, $N = 216$ sample points correspond to a mere six sample points in each dimension, which leads to insufficient resolution. The KDE method, on the other hand, does not approximate the underlying function $f_{3d}$, and is therefore “indifferent” to the noise dimension. See Sec. 9 for further discussion.
Figure 5. (a) Contours of the function $f_{2d}(\alpha)$, see (6.4). (b) The PDF of $f_{2d}(\alpha)$ (solid), its approximation by the spline-based Algorithm 4.1 (dashes), and by the gPC-based Algorithm 3.1 (dots). Here $\alpha$ is uniformly distributed in $[-1, 1]$, and both approximations use $N = 64$ sample points. (c) $L^1$ error of the PDF approximations as a function of the number of sample points, for the KDE (dash-dots), gPC-based approximation (dots-squares), the spline-based approximation (circles). The solid line is the power-law fit $1208N^{-2.15}$ (solid).

Figure 6. (a) The PDF of $f_{3d}(\alpha)$, see (6.5), where $\alpha$ is uniformly distributed in $[-1, 1]^3$ (solid) and its approximation by the spline-based Algorithm 4.1 (dashes) with $N = 8^3$ sample points. (b) $L^1$ error of the PDF approximations as a function of the number of sample points, for the KDE (dash-dots), the gPC-based PDF (rectangles), the spline-based PDF (circles), and its power-law fit $354N^{-1.11}$ (solid).

7. Application 1 - nonlinear Schrödinger equation. The one-dimensional coupled nonlinear Schrödinger equation (CNLS)

\begin{equation}
    i \frac{\partial A_{\pm}(t, x)}{\partial t} + \frac{\partial^2 A_{\pm}}{\partial x^2} + \frac{2}{3} |A_{\pm}|^2 + 2 |A_{\mp}|^2 A_{\pm} = 0,
\end{equation}

where $0 < \epsilon \ll 1$, $t \geq 0$, and $x \in \mathbb{R}$, describes the propagation of elliptically polarized, ultra-short pulses in optical fibers [2], of elliptically polarized continuous-wave (CW) beams in a bulk medium [35, 42], Stokes and anti-Stokes radiation in Raman amplifiers [37], and rogue water-waves formation at the interaction of crossing seas [1]. We consider (7.1) with an elliptically-polarized Gaussian input pulse with a random amplitude [35, 42]

\begin{equation}
    \left( \begin{array}{c}
        A_+ \\
        A_-
    \end{array} \right) = (1 + 0.1\alpha) \left( \begin{array}{c}
        8 \\
        4
    \end{array} \right) e^{-x^2},
\end{equation}

where $A_+$ and $A_-$ are the clockwise and counter-clockwise circularly-polarized components, respectively. The on-axis ellipse rotation angle is defined as

\begin{equation}
    \theta(t; \alpha) := (\varphi_+(t; \alpha) - \varphi_-(t; \alpha)) \mod (2\pi),
\end{equation}
where $\varphi_{\pm}(t; \alpha) := \arg[A_{\pm}(t, 0; \alpha)]$ are the on-axis phases of the components. The distribution of $\theta(t; \alpha)$ indicates to what extent the ellipse rotation angle is "deterministic".\(^{20}\)

\[ \theta(0.15; \alpha) = \frac{1}{N} \int_{-\pi}^{\pi} e^{i\theta(t; \alpha)} d\alpha, \quad \sigma^\circ \theta = \sqrt{-2\ln |E^\circ \theta|}. \]

The advantage of splines over gPC with few samples for moment approximation can be seen in Table 1. The approximation of $E^\circ \theta(0.15; \alpha)$ using the spline approximation with $N = 32$

\(^{20}\)We solve the CNLS using a fourth-order, compact finite-difference scheme for the spatial discretization, and a predictor-corrector Crank-Nicolson scheme for the temporal integration of the semi-discrete problem [17].

\(^{21}\)Because we have no explicit solution for $\theta(t; \alpha)$, the errors in this section are measured by comparison with $\theta_{513}^{\text{sl}}(0.15, \alpha)$ with $N = 513$ sample points. We verified that $\|\theta_{513}^{\text{sl}}(0.15, \alpha) - \theta_{513}^{\text{gPC}}(0.15, \alpha)\|_2 \approx 5 \times 10^{-5}$, which is an order of magnitude smaller than the approximation errors noted in the text.

\(^{22}\)To motivate why a different definition for circular moments is needed, consider $y \sim U(-\pi, \pi)$ and $z \sim U(0, 2\pi)$. If we consider $y$ and $z$ as angles, or points on the circle, they are identical. Using the conventional mean definition, however, yields $E[y] = 0$, but $E[z] = \pi$. 

**Figure 7.** The polarization angle $\theta(t = 0.15; \alpha)$ for solutions of the CNLS (7.2) with $\epsilon = 10^{-5}$, and an elliptically polarized Gaussian initial condition (7.2). (a) Spline interpolation (dashes) and gPC interpolation (dots), with $N = 64$ sample points. The two lines are nearly indistinguishable. (b) Pointwise error of the gPC interpolant. (c) Same for the spline interpolant.

**Interpolation.** For a given sample grid $\{\alpha_j\}_{j=1}^N$, we compute $\theta(t; \alpha_j)$ for each $1 \leq j \leq N$ by solving (7.1)–(7.2) and using (7.3). Fig. 7(a) shows the spline and gPC interpolants of $\theta(t = 0.15; \alpha)$ with $N = 64$ points.\(^{21}\) While these interpolants seem nearly identical, the spline interpolant is more accurate than the gPC interpolant by more than an order of magnitude (cf. Figs. 7(b) and 7(c)). Indeed, the $L^2$ error of the gPC interpolant (0.17%) is an order of magnitude larger than that of the spline interpolant (0.017%).

**Density estimation.** The gPC-based approximation with $N = 64$ differs substantially from the exact PDF, see Fig. 8(a). In contrast, the spline-based approximated PDF with $N = 64$ sample points is indistinguishable from the exact PDF, see Fig. 8(b). Indeed, the KL divergence of the gPC-based approximation, see (6.2), is about 16,000 times larger than that of the spline-based approximation, and the $L^1$ error is 200 times larger (46% vs. 0.2%). With $N = 32$, the spline-based is 32 times more accurate than the gpc-approximated PDF, in term of KL divergence, and 11 time more accurate in terms of the $L^1$ error (41% vs. 4.5%). The $L^1$ error of the spline-based PDF decays as $N^{-3.76}$, see Fig. 8(c), in accordance with Theorem 4.9. Since the PDF of $\theta(0.15; \alpha)$ has discontinuities and high derivatives, spline smoothing techniques and KDE methods with smooth kernels were not considered in this case.

**Moment approximation.** The mean and standard deviation of circular quantities can be defined as $[32]^{22}\)

(7.4) 

$$E^\circ \theta = \frac{1}{N} \int_{\theta_{\alpha}^{-}}^{\theta_{\alpha}^{+}} e^{i\theta(t; \alpha)} d\alpha, \quad \sigma^\circ \theta = \sqrt{-2\ln |E^\circ \theta|}.$$
Figure 8. Same settings as in Fig. 7. The PDF of $\theta(0.15, \alpha)$, where $\alpha \sim U(-1,1)$. (a) Exact PDF (solid), and gPC-based approximation using $N = 64$ sample points (dots). (b) Same with the spline-based approximation (dashes). The two lines are indistinguishable. (c) $L^1$ error of the spline-based PDF as a function of $N$ (circles) and the power-law fit $1.35 \cdot 10^4 N^{-3.76}$ (solid).

is 4 times more accurate than that of the gPC; with $N = 64$ it is 14 times more accurate. The approximation of the standard deviation using the spline-based method with $N = 32$ is 12 times more accurate than the gPC; with $N = 64$ it is 33 times more accurate than the gPC-based approximation.

| $N$ | gPC error | spline error | $\text{spline error} / \text{gPC error}$ |
|-----|-----------|--------------|--------------------------------------|
| 32  | 2.2%      | 0.54%        | 4                                    |
| 64  | 0.089%    | 0.006%       | 14                                   |
| 32  | 0.64%     | 0.054%       | 12                                   |
| 64  | 0.031%    | 0.0009%      | 33                                   |

Table 1 Approximation error of the circular mean and standard deviation, see (7.4), of $\theta(0.15, \alpha)$, see (7.3), with gPC- and spline-based approximations, using $N$ sample points.

8. Application 2 - inviscid Burgers equation. The inviscid Burgers equation

$$u_t(t, x) + \frac{1}{2} (u^2)_x = \frac{1}{2} (\sin^2(x))_x, \quad x \in [0, \pi], \quad t \geq 0,$$

with the initial and boundary conditions $u(0, x) = u_0(x)$ and $u(t, 0) = u(t, \pi) = 0$ models isentropic gas flow in a dual-throat nozzle. Solutions of this equation can develop a static shock wave at a lateral location $x = X_s$ [40]. Following [6], we consider the case in which $\alpha$ is a random variable with a known distribution, $u_0(x) = u_0(x; \alpha)$ is random, and we wish to compute the PDF of $X_s$ using Algorithms 3.1 and 4.1. In general, to do that requires, for each $1 \leq j \leq N$, to compute $X_s(\alpha_j)$ by solving (8.1) with $\alpha_j$. For the special initial condition

$$u_0(x) = \alpha \sin(x),$$

however, the shock location is explicitly given by [6]

$$\alpha = - \cos(X_s).$$

This explicit expression allows us to sample $X_s(\alpha)$ without solving (8.1).

Consider the case where

$$\alpha = \begin{cases} 
\frac{1 + \sqrt{1 + 4 \nu^2}}{2 \nu} & \text{if } \nu \neq 0, \\
0 & \text{if } \nu = 0
\end{cases}$$


and \( \nu \sim \mathcal{N}(0, \sigma) \), i.e., it is normally distributed with a zero mean. Because \( \alpha \) is not distributed by a classical, standard measure, there is no obvious choice of quadrature points to sample by, nor is there a “natural” orthogonal polynomials basis to expand the solution by. Therefore, the gPC approach cannot be straightforwardly applied. We can, however, apply the gPC approach to this problem by denoting \( X_s(\nu) = X_s(\alpha(\nu)) \), and approximating \( X_s(\nu) \) using the Hermite polynomials (which are orthogonal with respect to the normal distribution). The gPC-based approximated PDF with \( N = 7 \) sample points differs considerably from the exact PDF, see Fig. 9(a). In contrast, the spline-based approximated PDF can be directly applied to \( X_s(\alpha) \), and it is nearly indistinguishable from the exact PDF already with \( N = 7 \) sample points, see Fig. 9(b). In general, the spline-based PDF approximation is more accurate than the gPC-based approximation by more than one order of magnitude for \( 5 < N < 50 \), see Fig. 9(c). The \( L^1 \) error of the spline-based PDF is observed numerically to decay as \( N^{-3.11} \), in accordance with Theorem 4.9.

We repeated these simulations for the case with \( \alpha \sim B(r, s) \), where \( B(r, s) \) is the Beta distribution on \([-1, 1]\). The spline-based approximations are nearly identical to the exact PDF, whereas the gPC method were less accurate by an order of magnitude with few samples (results not shown).

9. Discussion. In this paper, we introduced a spline-based method for density and moment estimation. The advantages of this method are:

1. Our \( m \)-th order spline-based method approximates the density at a guaranteed convergence rate of \( N^{-\frac{d}{2m}} \), where \( N \) is the sample size and \( d \) is the noise dimension. Thus, our method outperforms KDEs for noise dimensions \( 1 \leq d \leq \frac{5}{2}m \).
2. It provides reasonable approximations for the density and moments using small sample sizes.
3. Its accuracy is relatively unimpaired by the presence of large derivatives.
4. It is non-intrusive, i.e., it is based solely on solving the underlying deterministic model.
5. It is easy to implement.
6. It is applicable with any choice of sample points.
7. It can be applied to non-smooth quantities of interest.

\footnote{Nevertheless, even for non-standard distributions, the expansion of \( \alpha \) by a classical orthogonal-polynomials basis can still converge spectrally, under certain conditions \cite{14}.}

\footnote{Indeed, in \cite{6} the authors use the gPC-Galerkin method with the Hermite polynomials \cite{22, 56}.}

\footnote{The PDF of the Beta distribution on \([0, 1]\) is \( p(\alpha) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \alpha^{r-1}(1-\alpha)^{s-1} \).}
When \( f \in C^{m+1} \), it is tempting to use splines of order \( m > 3 \) for density estimation, in order to attain faster than cubic convergence rate. If one generalizes Algorithm 4.1 to splines of order \( m \) then, similarly to Theorem 4.9, a convergence of order \( N^{-m} \) is guaranteed. Even if \( f \) is analytic, however, it is not advisable to take a large \( m \), for two reasons. First, for \( s(\alpha) \) to be monotone (and so, by Lemma 4.8 for the PDF to be continuous), \( N \) should scale as \( \sqrt[N]{\|f^{(m+1)}\|_{\infty}} \), see (F.1). Therefore, for a large \( m \), high-order convergence might only be attained for very large sample sizes. Second, the density approximation error depends linearly on \( \|f^{(m+1)}\|_{\infty} \), see \( F \), and so it might “blow-up” exponentially with \( m \). To conclude, although we do not know whether the optimal spline order is \( m = 3 \), an arbitrarily high-order spline should not be used.

When approximating a \( d \)-dimensional function with a resolution \( h \) at each dimension, the total number of samples \( N \) scales as \( h^{-d} \). As a result, for a prescribed accuracy, the computational cost grows exponentially with the dimension (the “curse of dimensionality”). In other words, for a given \( N \), the accuracy decays exponentially with the dimension. Indeed, this is consistent with the \( N^{-\frac{d}{2}} \) error estimate of the spline-based Algorithm 5.1 (Corollary 5.4). In contrast, the KDE method, which is a standard nonparametric statistical density estimator, converges at a rate of \( N^{-\frac{1}{2}} \), regardless of \( d \). Hence, our method will outperform KDE for “low” dimensions \((d < \frac{5}{2}m)\), but may become inferior to KDE at higher dimensions.

A popular approach for moment estimation of high-dimensional noise is the use of sparse sampling grids [21, 54]. Recently, a spline approximation based on sparse grids was used in the context of forward uncertainty propagation [49]. Most sparse-grid methods, however, are designed with moment estimation in mind. As we have seen, even in the one-dimensional case (see Sec. 4.1), an accurate moment approximation does not necessarily imply an accurate density estimation. Whether sparse-grids methods can be adapted to density estimation remains an open question. the proof of Theorem 5.3 in Appendix 1, however, suggests sufficient conditions by which new approximation methods can be tested for efficient density estimation: (1) The settings should be such that Lemma 5.2 applies, and (2) the approximation method should have a pointwise error bounds similar to Theorem 5.1.

In this paper we showed that spline-based density estimation is better than gPC-based density estimation, because it does not produce numerous artificial extremal points (see Lemma 4.8). An interpolating cubic spline, however, might still produce artificial extremal points, though not as much as the gPC polynomial. To absolutely prevent artificial extremal points from being produced, it may be better to use spline interpolants [19] and quasi-interpolants [10] which are monotonicity-preserving (i.e., splines which are monotone wherever the sampled data is monotone). Hence, although these methods have the same order of error (with respect to \( h \)) as spline interpolation, they may provide better approximations for small samples, as they are guaranteed not to produce artificial extremal points. We leave it to future research to check whether monotonicity-preserving interpolants provide more accurate PDF approximations than a standard interpolating cubic spline.

10. Acknowledgments. The authors thank Y. Harness, B. Brill, R. Kats, F. Abramovich, and D. Levin for useful comments and conversations.

Appendix A. Proof of Corollary 3.1. We begin our proof with the following Lemma:

**Lemma A.1.** Let \((\Omega, \mu)\) be a probability space, denote \( \| \cdot \|_p := \| \cdot \|_{L^p(\Omega)} \), and let \( f, g \in L^2 \cap L^1 \).

Then

\[
\begin{align*}
(A.1a) \quad & |E_\alpha[f] - E_\alpha[g]| \leq \|f - g\|_2, \\
(A.1b) \quad & |\text{Var}(f) - \text{Var}(g)| \leq (\sigma(f) + \sigma(g)) \cdot \|f - g\|_2, \\
(A.1c) \quad & |\sigma(f) - \sigma(g)| \leq \|f - g\|_2.
\end{align*}
\]
Proof. For all \( f, g \in L^2 \),
\[
|E_\alpha[f] - E_\alpha[g]| \leq \int_\Omega |f(\alpha) - g(\alpha)| \, d\mu(\alpha) = \int_\Omega 1 \cdot |f(\alpha) - g(\alpha)| \, d\mu(\alpha) \leq \|1\|_2 \cdot \|f - g\|_2 = \|f - g\|_2,
\]
where in the second inequality we used the Cauchy-Schwarz inequality. Thus, we proved (A.1a).

For \( h \in L^2 \cap L^1 \), let \( \tilde{h} := h - E_\alpha[h] \). By definition, \( \text{Var}(h) = \|\tilde{h}\|_2^2 \) and \( \sigma(h) = \|\tilde{h}\|_2 \). Hence,
\[
|\text{Var}(f) - \text{Var}(g)| = |E_\alpha[f^2 - \tilde{g}^2]| = \left| \int_\Omega (f - \tilde{g})(\tilde{f} + \tilde{g}) \, d\mu(\alpha) \right| \leq \|\tilde{f} + \tilde{g}\|_2 \cdot \|f - g\|_2.
\]
(A.2)

In addition, \( \|\tilde{h}\|_2^2 = \text{Var}(h) = E_\alpha[h^2] - E_\alpha[h]^2 \leq E_\alpha[h^2] = \|h\|_2^2 \), and so \( \|\tilde{h}\|_2 \leq \|h\|_2 \). Applying this inequality with \( h = f - g \) to (A.2) yields (A.1b). Finally, by (A.1b),
\[
|\sigma(f) - \sigma(g)| = \left| \frac{\sigma^2(f) - \sigma^2(g)}{\sigma(f) + \sigma(g)} \right| \leq \frac{\sigma(f) + \sigma(g)}{\sigma(f) + \sigma(g)} \|f - g\|_2 = \|f - g\|_2.
\]
which proves (A.1c).

Therefore, in the case of gPC, let \( g = f_N^\text{gPC} \), the colocation gPC approximation of \( f \), see (3.7). Since \( f_N^\text{gPC} \) converges exponentially to \( f \) in the \( L^2 \) norm \([54, 25]\), Lemma A.1 implies that the moments of \( f_N^\text{gPC} \) converge exponentially to the moments of \( f \).

### Appendix B. Proof of Lemma 4.1.

Let \( \{p_n(\alpha)\}_{n=0}^\infty \) be the family of orthogonal polynomials with respect to the weight function \( c(\alpha) \), and let \( \{\alpha_1, \ldots, \alpha_N\} \) be the roots of \( p_N(\alpha) \). Denote \( f_j := f(\alpha_j) \) for \( j = 1, \ldots, N \).

Let \( \Pi(\alpha) = \sum_{n=0}^{N-1} \gamma_n p_n(\alpha) \) be the interpolating polynomial of \( f \) at the quadrature points, where \( \{\gamma_0, \ldots, \gamma_{N-1}\} \) are some unknown coefficients. The interpolation conditions \( \Pi(\alpha_j) = f_j \) for \( j = 1, \ldots, N \) can be written in a matrix form as
\[
(A_\gamma = f),
\]
where \( A_{i,j} = p_{j-1}(\alpha_i) \), \( \gamma = (\gamma_0, \ldots, \gamma_{N-1})^T \), and \( f = (f_1, \ldots, f_N)^T \). We define the matrix \( B \) by
\[
B_{i,j} = w_{j} A_{i,j} = w_{j} p_{i-1}(\alpha_j). \quad \text{We claim that} \quad B = A^{-1}. \quad \text{Indeed, for all} \quad 1 \leq i, j \leq N,
\]
\[
(BA)_{i,j} = \sum_{k=1}^{N} B_{i,k} A_{k,j} = \sum_{k=1}^{N} w_k p_{i-1}(\alpha_k) p_{j-1}(\alpha_k) \cdot
\]
Because the Gauss quadrature rule is exact for polynomials of degree \( \leq 2N - 1 \) \([8]\),
\[
\sum_{k=1}^{N} w_k p_{i-1}(\alpha_k) p_{j-1}(\alpha_k) = \int_\Omega p_{i-1}(\alpha) p_{j-1}(\alpha) c(\alpha) d\alpha = \delta_{i,j},
\]
where the last equality is due of the orthogonality of \( p_{i-1} \) and \( p_{j-1} \). By (B.1),
\[
\gamma_n = (A^{-1} f)_{n+1} = \sum_{k=1}^{N} B_{n+1,k} f_k = \sum_{k=1}^{N} w_k p_n(\alpha_k) f_k, \quad n = 0, 1, \ldots, N - 1.
\]
Hence \( \Pi(\alpha) = \sum_{n=0}^{N-1} \left( \sum_{k=1}^{N} w_k p_n(\alpha_k) f(\alpha_k) \right) p_n(\alpha) = f_N^\text{gPC}(\alpha) \), see (3.6) and (3.7).

### Appendix C. An alternative to Corollary 4.4.

We prove a slightly different bound than that of Corollary 4.4:
Consider a uniform grid $\alpha_1 = \alpha_{\min} < \alpha_1 \cdots < \alpha_N = \alpha_{\max}$, where $\alpha_{i+1} - \alpha_i = h$ for all $i = 1, \ldots, N-1$, and let $f_N^{\text{spline}} = f_N^{\text{spline}}(\alpha_1, f_1, \ldots, f_N)$ be either the natural cubic spline, the clamped cubic spline, or the "not-a-knot" interpolating spline. Then

$$\max_{\alpha \in (\alpha_{i-k}, \alpha_{i+k})} \left| \frac{\partial f_N^{\text{spline}}(\alpha; f_1, \ldots, f_N)}{\partial f_i} \right| \leq C \delta^k, \quad 1 < i < N, \quad 1 \leq k \leq N.$$  

For the case of the natural cubic spline, $C \leq 140$ and $\delta < 0.27$.

**Proof.** Define the B-splines on the (unbounded) uniform grid by [36]

$$B_i(\alpha) := \frac{1}{h^3} \begin{cases} 
(\alpha - \alpha_{i-2})^3 & \text{if } \alpha \in [\alpha_{i-2}, \alpha_{i-1}) \\
\frac{1}{3}h^3(\alpha - \alpha_{i-1}) + 3h(\alpha - \alpha_{i-1})^2 - 3(\alpha - \alpha_{i-1})^3 & \text{if } \alpha \in [\alpha_{i-1}, \alpha_i) \\
\frac{1}{3}h^3(\alpha - \alpha_{i+1} - \alpha) + 3h(\alpha_{i+1} - \alpha)^2 - 3(\alpha_{i+1} - \alpha)^3 & \text{if } \alpha \in [\alpha_i, \alpha_{i+1}) \\
(\alpha_{i+2} - \alpha)^3 & \text{if } \alpha \in [\alpha_{i+1}, \alpha_{i+2}) \\
0 & \text{otherwise},
\end{cases}$$

see Fig. 10. By definition, B-splines are $C^2$ cubic splines. One can express any $C^2$ cubic spline as a linear combination of B-splines. Specifically $f_N^{\text{spline}}(\alpha) = \sum_{i=0}^{N+1} c_i B_i(\alpha)$ [36]. We first prove this for the natural cubic spline, for which $\frac{d}{\alpha} f_N^{\text{spline}}(\alpha_1) = \frac{d}{\alpha} f_N^{\text{spline}}(\alpha_N) = 0$. In this case, by (C.1), $B_0'(\alpha_1) = -B_2'(\alpha_1)$, and $B'_1(\alpha_1) = 0$. Hence,

$$0 = \frac{d}{\alpha} f_N^{\text{spline}}(\alpha_1) = c_0 B_0'(\alpha_1) + c_1 B_1'(\alpha_1) + c_2 B_2'(\alpha_1) = \frac{1}{h}(c_2 - c_0),$$

and so $c_0 = c_2$ and similarly $c_{N-1} = c_{N+1}$. Therefore, $f_N^{\text{spline}}(\alpha) = \sum_{i=1}^{N} c_i B_i(\alpha) + c_2 B_0(\alpha) + c_{N+1} B_{N+1}(\alpha)$. The interpolation conditions $f_N^{\text{spline}}(\alpha_j) = f(\alpha_j)$ for $1 \leq j \leq N$ can now be written in a matrix form $Bc = f$, where $c = (c_1, \ldots, c_N)^T, f = (f(\alpha_1), \ldots, f(\alpha_N))^T$, and

$$B = \begin{pmatrix} B_1(\alpha_1) & B_0(\alpha_1) + B_2(\alpha_1) & 0 & \cdots & \cdots & 0 \\
B_1(\alpha_2) & B_2(\alpha_2) & B_3(\alpha_2) & 0 & \cdots & \cdots \\
0 & B_2(\alpha_3) & B_3(\alpha_3) & B_4(\alpha_3) & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & B_{N-1}(\alpha_N) + B_{N+1}(\alpha_N) & B_N(\alpha_N) \end{pmatrix}$$

Since $B_i(\alpha_i) = 4, B_i(\alpha_{i-1}) = B_i(\alpha_{i+1}) = 1$, and $B_i(\alpha_n) = 0$ for $|n - i| > 1$, we have that

$$B = \begin{pmatrix} 4 & 2 & 0 & \cdots & \cdots & 0 \\
1 & 4 & 1 & 0 & \cdots & 0 \\
0 & 1 & 4 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & 2 & 4 \end{pmatrix}$$

Since $B$ is strongly diagonally dominant, it is invertible, and so $c = B^{-1}f$ is unique. Therefore

$$f_N^{\text{spline}}(\alpha) = b(\alpha)B^{-1}f,$$

where $b(\alpha) = (B_1(\alpha), B_0(\alpha) + B_2(\alpha), B_3(\alpha), \ldots, B_{N-2}(\alpha), B_{N-1}(\alpha) + B_{N+1}(\alpha), B_N(\alpha))$.  

22
Let $\alpha \in [\alpha_{n-1}, \alpha_n]$ for some $1 < n \leq N$. By (C.3),

\[(C.4) \quad \frac{\partial f_N^{\text{spline}}(\alpha)}{\partial f_i} = (b(\alpha)B^{-1})_i = \sum_{\ell=1}^{N} (b(\alpha))_\ell (B^{-1})_{\ell,i}.
\]

To bound the elements of $B^{-1}$, we recall

**Theorem C.2** (Demko, Moss and Smith [12]). Let $A$ be an invertible, positive-definite, $m$-banded matrix. Then

\[|A^{-1}_{i,n}| \leq C\delta|i-n|,
\]

where

\[C = \max \left\{ \frac{1}{c_{\min}}, \frac{(1+\sqrt{q})^2}{2c_{\max}} \right\}, \quad \delta = \left( \frac{\sqrt{q} - 1}{\sqrt{q} + 1} \right)^{\frac{2}{m}}, \quad q = \frac{c_{\max}}{c_{\min}},
\]

and $c_{\min}$ and $c_{\max}$ are the minimal and maximal eigenvalues of $A$, respectively.

The matrix $B$, see (C.2), is invertible and banded with $m = 2$. Since $B$ is strongly diagonally dominated (SDD), it is also positive-definite. Thus, $B$ satisfies the requirements of Theorem C.2. Therefore, $(B)_{i,\ell} \leq C\delta|i-\ell|$. Hence, by (C.4),

\[\left| \frac{\partial f_N^{\text{spline}}(\alpha)}{\partial f_i} \right| \leq \sum_{\ell=1}^{N} |(b(\alpha))_\ell| C\delta|\ell-i|.
\]

Next, we show that $0 < \delta < 1$ by applying Geršggorin’s Circle Theorem

**Theorem C.3** ([23]). Let $A$ be a rectangular matrix. Denote by $D_i$ the closed disc in the complex plane centered at $A_{i,i}$ and with radius $R_i := \sum_{k \neq i} |A_{i,k}|$. Then every eigenvalue of $A$ lies inside one of the discs $D_i$.

In the case where $A = B$, we have $R_i = 2$ and $B_{i,i} = 4$ for all $1 \leq i \leq N$. Because $B$ is positive definite, its eigenvalues are real and bounded between 2 and 6; in particular, $2 \leq c_{\min}, c_{\max} \leq 6$.

Hence, by Theorem C.2, it follows that $\delta \leq \left( \frac{\sqrt{3} - 1}{\sqrt{3} + 1} \right)^{\frac{2}{2}} \approx 0.27$, and similarly $C \leq 0.65$.

Because the support of each $B_\ell(\alpha)$ is $[\alpha_{\ell-2}, \alpha_{\ell+2}]$, see Fig. 10, then $(b(\alpha))_\ell \neq 0$ only for $\ell = n-2, n-1, n$, and $n+1$. By (C.1), $0 < B_\ell(\alpha) \leq 4$ for all $\ell$.

Let us consider the case where $5 \leq n \leq N - 3$.\(^{26}\) Then

\[\left| \frac{\partial f_N^{\text{spline}}(\alpha)}{\partial f_i} \right| \leq \sum_{\ell=n-2}^{n+1} |(b(\alpha))_\ell| C\delta|\ell-i| \leq \sum_{\ell=n-2}^{n+1} 4 \cdot C\delta|\ell-i| \leq 4 \cdot 4 \cdot C\delta^{-2}\delta^{n-i},
\]

\(^{26}\)The proof for other values of $n$ is similar.
Taking $k = |n - i|$, one has that

$$\max_{\alpha \not\in (\alpha_{i-k}, \alpha_{i+k})} \left| \frac{\partial f_N^{\text{spline}}(\alpha)}{\partial f_i} \right| \leq \tilde{C} \delta^k,$$

where $\tilde{C} = 16C\lambda^{-2} \leq 16 \cdot 0.65\lambda^{-2} \leq 140$, which completes the proof.

A similar proof holds for the “not-a-knot” and the clamped boundary conditions. For example, in the “not-a-knot” interpolation, $\lim_{\alpha \to \alpha_i} f^{\text{spline}} N = \lim_{\alpha \to \alpha_i} f^{\text{spline}} N$ for $j = 2, N - 1$. By direct computation, see (C.1), this yields $c_0 - 4c_1 + 6c_2 - 4c_3 + c_4 = 0$. As in the above proof, we take $c_0 = 4c_1 - 6c_2 + 4c_3 - c_4$, and similarly $c_{N+1} = 4c_N - 6c_{N-1} + 4c_{N-2} - c_{N-3}$, and obtain a slightly different $B$ matrix,

$$B^{\text{not-a-knot}} = \begin{pmatrix}
8 & -5 & 4 & -1 & \cdots & 0 \\
1 & 4 & 1 & 0 & \cdots & 0 \\
0 & 1 & 4 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & 1 & 4 & 1 \\
0 & \cdots & -1 & 4 & -5 & 8
\end{pmatrix}.$$

This matrix is still banded, and because its symmetric part $\frac{1}{2} (B + B^T)$ is diagonally dominant, it is also positive definite [26]. Therefore, the proof follows in a similar manner.

Remark C.4. In the case of a nonuniform grid, the definition of the B-splines slightly changes, and so does the matrix $B$, see (C.2). Nevertheless, $B$ remains regular and banded with $m = 2$. Therefore, Theorem C.2 applies and the proof remains essentially the same.

Appendix D. Proof of Lemma 4.8.

When $f$ is strictly increasing, its CDF is given by

$$P(y) = \int_{\alpha_{\min}}^{f^{-1}(y)} c(\alpha) d\alpha.$$

By the Leibniz rule and the inverse function theorem,

$$p(y) = \frac{dP(y)}{dy} = c \left( f^{-1}(y) \right) \left( f^{-1} \right)' = c \left( f^{-1}(y) \right) \frac{1}{f' \left( f^{-1}(y) \right)}.$$

Similarly, if $f$ is monotonically decreasing, then $P(y) = \int_{f^{-1}(y)}^{\alpha_{\max}} c(\alpha) d\alpha$, and so

$$p(y) = -\frac{c \left( f^{-1}(y) \right)}{f' \left( f^{-1}(y) \right)}.$$

Note that since $f' < 0$, then $p(y) \geq 0$. Finally, if $f$ is piecewise monotonic, we apply this method separately on each sub-interval on which it is monotonic, and sum up the contributions.

Appendix E. Sample MATLAB code for Algorithm 4.1. The following MATLAB code generates the dashed curve in Fig. 2(b).
\[
\alpha_{\text{min}} = -1; \quad \alpha_{\text{max}} = 1; \quad N = 18; \quad \% \text{sample size}
\]
\[
f = @ (x) \tan \hskip 1pt h (9 * x) + x / 2;
\]
\[
\% \text{define the initial sample on the grid } [\alpha_{\text{1}}, \ldots, \alpha_{\text{N}}]
\]
\[
samplingGrid = \text{linspace} (\alpha_{\text{min}}, \alpha_{\text{max}}, N);
\]
\[
samples = f (samplingGrid); \quad \% \text{step 1}
\]
\[
\% \text{define the refined sample grid } [\tilde{\alpha}_{\text{1}}, \ldots, \tilde{\alpha}_{\text{M}}]
\]
\[
M = 2 e 6;
\]
\[
denseGrid = \text{linspace} (\alpha_{\text{min}}, \alpha_{\text{max}}, M);
\]
\[
fN . \text{spline} = \text{spline} (samplingGrid, samples, denseGrid); \% \text{steps 2+3}
\]
\[
\% \text{When f is given explicitly, the optimal number of bins (L)}
\]
\[
Cf = 1.69; \quad L = C f * M^{-1} (1/3);
\]
\[
\% \text{step 4 - histogram of fN on denseGrid, not normalized}
\]
\[
[\text{histogram, binsEdges}] = \text{hist} (fN . \text{spline}, L);
\]
\[
\text{binWidth} = \left( \max (\text{binsEdges}) - \min (\text{binsEdges}) \right) / L;
\]
\[
\% \text{normalize the histogram so that it would be a PDF}
\]
\[
\text{pdf} = \text{histogram} / \left( \sum (\text{histogram} \times \text{binWidth}) \right);
\]
\[
\text{plot} (\text{binsEdges}, \text{pdf})
\]

**Appendix F. Proof of Theorem 4.9.** Without loss of generality, we can assume that \( f' ( \alpha ) \geq a > 0 \). For brevity, denote \( s ( \alpha ) = f_{N}^{\text{spline}} ( \alpha ) \), \( h = h_{\text{max}} \), and \( \| f^{(4)} \|_{\infty} = \| f^{(4)} \|_{L^{\infty}[\alpha_{\text{min}}, \alpha_{\text{max}}]} \). In general, \( s ( \alpha ) \) can be non-monotone. By Theorem 4.2, however, \( | s' ( \alpha ) - f' ( \alpha ) | < C_{\text{spl}}^{(1)} \| f^{(4)} \|_{\infty} h^{3} \).

Hence

\[
(F.1) \quad s' ( \alpha ) \geq \frac{a}{2} > 0, \quad N > \sqrt[3]{2 C_{\text{spl}}^{(1)} \| f^{(4)} \|_{\infty}} (\alpha_{\text{max}} - \alpha_{\text{min}}),
\]

and so \( s ( \alpha ) \) is monotonically increasing and invertible for sufficiently large \( N \).\footnote{In the numerical example (6.1), this lower bound is roughly \( N > 30 \).} Because \( s ( \alpha ) \) interpolates \( f ( \alpha ) \), and because both functions are monotone, then \( \text{range} (s) = \text{range} (f) \). Since \( s, f \in C^{1} \) and are invertible, by Lemma 4.8

\[
(F.2) \quad \| p - \tilde{p} \|_{1} = \int_{f(\alpha_{\min})}^{f(\alpha_{\max})} | p ( y ) - \tilde{p} ( y ) | \, dy = \int_{f(\alpha_{\min})}^{f(\alpha_{\max})} \left| c \left( f^{-1} ( y ) \right) - c \left( s^{-1} ( y ) \right) \right| \, dy.
\]

Denote \( y = f ( \alpha ) \) and \( \alpha_{\star} := \alpha_{\star} ( \alpha ) = s^{-1} ( f ( \alpha ) ) \). Then by a change of variable

\[
(F.3) \quad \| p - \tilde{p} \|_{1} = \int_{\alpha_{\min}}^{\alpha_{\max}} \left| \frac{c ( \alpha )}{f' ( \alpha )} - \frac{c ( \alpha_{\star} )}{s' ( \alpha_{\star} )} \right| f' ( \alpha ) \, d \alpha = \int_{\alpha_{\min}}^{\alpha_{\max}} \left| s' ( \alpha_{\star} ) c ( \alpha ) - f' ( \alpha ) c ( \alpha_{\star} ) \right| \frac{1}{s' ( \alpha_{\star} )} \, d \alpha.
\]

For all \( \alpha \in [\alpha_{\min}, \alpha_{\max}] \),

\[
| s' ( \alpha_{\star} ) c ( \alpha ) - f' ( \alpha ) c ( \alpha_{\star} ) | \leq c ( \alpha ) | s' ( \alpha_{\star} ) - s' ( \alpha ) | + c ( \alpha ) | s' ( \alpha ) - f' ( \alpha ) | + f' ( \alpha ) | c ( \alpha ) - c ( \alpha_{\star} ) |.
\]

Because \( s' ( \alpha ) \) and \( c ( \alpha ) \) are differentiable,

\[
(F.4) \quad | s' ( \alpha_{\star} ) c ( \alpha ) - f' ( \alpha ) c ( \alpha_{\star} ) | \leq D | \alpha - \alpha_{\star} | + \| c \|_{\infty} | f' ( \alpha ) - s' ( \alpha ) |,
\]
where $D = \|c\|_\infty : \|s''\|_\infty + \|c'\|_\infty : \|f\|_\infty$.

By Lagrange’s mean-value theorem, there exists $\beta$ between $\alpha$ and $\alpha_*$ such that

$$s(\alpha) - s(\alpha_*) = s'(\beta)(\alpha - \alpha_*).$$

On the other hand, since $\alpha_* = s^{-1}(f(\alpha))$, then $s(\alpha_*) = f(\alpha)$, and so

$$s(\alpha) - s(\alpha_*) = f(\alpha) - f(\alpha).$$

Therefore $\alpha - \alpha_* = \frac{s(\alpha) - f(\alpha)}{s'(\beta)}$. By (F.1), $s'(\beta) \geq \frac{a}{2}$, and by Theorem 4.2, we have $|f(\alpha) - s(\alpha)| \leq C_{spl} \|f(4)\|_\infty h^4$. Hence,

$$\frac{|\alpha - \alpha_*|}{a} \leq 2C_{spl}\|f(4)\|_\infty h^4.$$

By Theorem 4.2, $|f'(\alpha) - s'(\alpha)| \leq C_{spl}\|f(4)\|_\infty h^3$. Hence (F.4) reads

$$F.5 \quad |s'(\alpha_*)c(\alpha) - f'(\alpha)c(\alpha_*)| \leq K_1h^3 + K_2h^4,$$

where $K_1 = C_{spl}\|c\|_\infty \|f(4)\|_\infty$ and $K_2 = \frac{2aC_{spl}\|f(4)\|_\infty D}{\frac{a}{2}}$, see (F.1), and (F.5) in (F.3), for sufficiently large $N$ such that $h = \frac{\alpha_{\max} - \alpha_{\min}}{N-1} < 1$ we have that

$$\|p - \tilde{p}\|_1 = \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{2(K_1 + K_2)}{a}h^3 d\alpha = \frac{2(K_1 + K_2)}{a}(\alpha_{\max} - \alpha_{\min})h^3 \leq \frac{K}{N^3},$$

where $K = \frac{2(K_1 + K_2)}{a}(\alpha_{\max} - \alpha_{\min})^4$.

**Remark F.1.** If $f'(\alpha) = 0$ for some values of $\alpha$, the approximation $\tilde{p}$ is not guaranteed to converge in the $L^1$ norm. By (F.5), however, we can guarantee a third-order convergence for the pointwise error $p(y) - \tilde{p}(y)$, for every real number $y$ such that $f'(\alpha)$ does not vanish on $\{\alpha \mid f(\alpha) = y\}$.

**Appendix G. Proof of Lemma 4.12.**

Similarly to the proof of (F.4),

$$|s'(\alpha_*)c(\alpha) - f'(\alpha)c(\alpha_*)| \leq D|\alpha - \alpha_*| + c(\alpha)|f'(\alpha) - s'(\alpha)|.$$

Because $|\alpha - \alpha_*| \leq K_2h^4$, then by (F.3),

$$G.1 \quad \|p - \tilde{p}\|_1 \leq \frac{2K_2}{a}h^4 + \int_{\alpha_{\min}}^{\alpha_{\max}} |f'(\alpha) - s'(\alpha)|c(\alpha) d\alpha.$$

Since $f'(\alpha) - s'(\alpha)$ is continuous on $[\alpha_{\min}, \alpha_{\max}]$, it vanishes and changes its sign only at $J_N < \infty$ points, denoted by $\alpha_{\min} = \gamma_0 < \gamma_1 < \cdots < \gamma_{J_N} = \alpha_{\max}$. Using integration by parts, the last integral reads

$$\int_{\alpha_{\min}}^{\alpha_{\max}} |f'(\alpha) - s'(\alpha)|c(\alpha) d\alpha = \eta \sum_{j=1}^{J-1} (-1)^j \int_{\gamma_j}^{\gamma_{j+1}} (f'(\alpha) - s'(\alpha)) c(\alpha) d\alpha$$

28By the same argument as (F.1), for a fixed $\epsilon > 0$ there exists a sufficiently large $N_0$ such that $s''(\alpha) \leq f''(\alpha) + \epsilon$ for all $N > N_0$. Therefore $\max \|s''\|_\infty \leq \max \|f''\|_\infty + \epsilon$, and so $D$ is independent of $N$, and depends only on $f(\alpha)$, $c(\alpha)$, $\alpha_{\min}$ and $\alpha_{\max}$. 28
\[ \eta \sum_{j=1}^{J_N-1} (-1)^j \left[ c(\gamma_{j+1}) (f(\gamma_{j+1}) - s(\gamma_{j+1})) - c(\gamma_j) (f(\gamma_j) - s(\gamma_j)) - \int_{\gamma_j}^{\gamma_{j+1}} (f(\alpha) - s(\alpha)) c'(\alpha) \, d\alpha \right], \]

where \( \eta = \text{sign} \left[ f'(\alpha_{\text{min}}) - s'(\alpha_{\text{min}}) \right] \). By Theorem 4.2,

\[ |c(\gamma_j) (f(\gamma_j) - s(\gamma_j))| \leq \|c\|_{\infty} C_{\text{spl}}^{(0)} \|f^{(4)}\|_{\infty} h^4, \quad 1 \leq j \leq J_N, \]

and

\[ \left| \int_{\gamma_j}^{\gamma_{j+1}} (f(\alpha) - s(\alpha)) c'(\alpha) \, d\alpha \right| \leq \|c\|_{\infty} (\gamma_{j+1} - \gamma_j) C_{\text{spl}}^{(0)} \|f^{(4)}\|_{\infty} h^4, \quad 1 < j \leq J_N. \]

Substituting these bounds in (G.1) yields

\[ \|p - \bar{p}\|_1 \leq \frac{2K_2}{\alpha} h^4 + K_3 h^4 + K_4 J_N h^4, \]

where \( K_3 = \|c\|_{\infty} |\alpha_{\max} - \alpha_{\min}| C_{\text{spl}}^{(0)} \|f^{(4)}\|_{\infty} \) and \( K_4 = 2 \|c\|_{\infty} C_{\text{spl}}^{(0)} \|f^{(4)}\|_{\infty} \). In the case of a uniform grid, the first two terms are \( O(N^{-4}) \), and the last term is \( O(N^{-4}J_N) \), which completes the proof.

**Appendix H. Proof of Lemma 5.2.**

For any \( y \in \mathbb{R} \), the CDF of \( f \) is

\[ P(y) = \text{Prob} \{ f(\alpha) \leq y \} = \frac{1}{\mu(\Omega)} \int_{D(y)} d\mu(\alpha) = \frac{1}{\mu(\Omega)} \int_{D(y)} c(\alpha) \, d\alpha, \]

where

\[ D(y) := \{ \alpha \in \Omega \mid f(\alpha) \leq y \}. \]

To compute the PDF \( p(y) := \frac{d}{dy} P(y) \), we recall the co-area formula:

**Lemma H.1 ([16]).** Let \( A \subseteq \mathbb{R}^d \) be a Jordan set, let \( u : A \to \mathbb{R} \) be Lipschitz and piecewise differentiable such that \( u^{-1}(z) \subseteq A \) is a \((d-1)\) dimensional manifold for all \( z \in \mathbb{R} \), and let \( g \in L^1(A) \). Then

\[ \int_A g(\alpha) |\nabla u(\alpha)| \, d\alpha = \int_{z \in u(A)} dz \int_{u^{-1}(z)} g(\alpha) \, d\sigma, \]

where \( d\sigma \) is the \((d-1)\) dimensional surface element of \( u^{-1}(z) \).

We apply the co-area formula to the right-hand-side of (H.1) by substituting \( A = D(y), g = \frac{c}{|\nabla f|} \) and \( u = f \) in (H.3). The use of (H.3) is justified since

1. \( D(y) \) is bounded, since \( \Omega \) is bounded. We can therefore show that \( D(y) \) is Jordan by proving that \( m(\partial D(y)) = 0 \), where \( m \) is the Lebesgue measure in \( \mathbb{R}^d \). Since \( \partial D(y) \subseteq f^{-1}(y) \cup \partial \Omega \), it is sufficient to show that each of these sets is of measure zero. Indeed, \( \Omega \) is Jordan, and so \( m(\partial \Omega) = 0 \). In addition, since \( |\nabla f| \neq 0 \) on \( f^{-1}(y) \), by the implicit function theorem \( f^{-1}(y) \) is a \((d-1)\) dimensional manifold, and so \( m(f^{-1}(y)) = 0 \).
2. \( f \) is piecewise-differentiable by the conditions of the Lemma 5.2. Furthermore, because \( f \) is piecewise-differentiable on a compact set \( \Omega \), it is also Lipschitz.
3. Since $f$ is continuous and $|\nabla f| \neq 0$ on $\Omega$, then $\frac{1}{|\nabla f|}$ is bounded from above. Therefore, since $c \in L^1$, so $g = \frac{c}{|\nabla f|}$.


Thus, by Lemma H.1 and (H.1),

\begin{equation}
P(y) = \frac{1}{\mu(\Omega)} \int_{D(y)} c(\alpha) \, d\alpha = \frac{1}{\mu(\Omega)} \int_{-\infty}^{y} dz \int_{f^{-1}(z)} \frac{c}{|\nabla f|} \, d\sigma.
\end{equation}

The outer integral on the right-hand-side is over $(-\infty, y)$ since $f(D(y)) \subseteq (-\infty, y)$, see (H.2).

Finally, since $p(y) = \frac{d}{dy} P(y)$, differentiating the last integral using the (one-dimensional) Leibnitz integral rule yields (5.2).

**Appendix I. Proof of Theorem 5.3.**

Since $f \in C^{m+1}(\Omega)$ and $\Omega$ is compact, $f$ is also Lipschitz. Hence, Lemma 5.2 can be applied with $m([0,1]^d) = 1$ and $c(\alpha) \equiv 1$, yielding

\begin{equation}
\|p - \tilde{p}\|_1 = \int_{-\infty}^{\infty} I(y) \, dy,
\end{equation}

where $\sigma$ is the $d$-1-dimensional surface measures induced by the Lebesgue measure.

The outline of the proof is as follows:

1. For a fixed $y$ in the image of $s(\alpha)$, we construct a cover \{\$A_j(y)\}$ of $s^{-1}(y)$.
2. We then construct a set of maps $\phi_j : A_j(y) \rightarrow f^{-1}(y)$, which are characterized in Lemma I.1.
3. We construct a disjoint cover $\tilde{A}_j \subseteq A_j(y)$ for $1 \leq j \leq d$. Lemma I.3 proves that $\{\phi_j(\tilde{A}_j)\}$ are mutually disjoint, up to an $O(h^m)$ error, and almost cover $f^{-1}(y)$, up to an $O(h^m)$ error.
4. By an inclusion-exclusion argument and the implicit function theorem, we split the integral of $I(y)$ to $d$ integrals over compact domains in $\mathbb{R}^{d-1}$.
5. By Theorem 5.1, and similarly to the proof of the one-dimensional counterpart (Theorem 4.9), we bound each of the integrals obtained in step 3. Thus, we obtain a pointwise bound on $p(y) - \tilde{p}(y)$.
6. Finally, we use compactness of $\Omega$ and the fact that $f, s \in C^1(\Omega)$ to bound $\|p - \tilde{p}\|_1$.

**Step 1.** For brevity, denote by $\partial_{\alpha_j} = \frac{\partial}{\partial \alpha_j}$ the partial derivative along the $j$-th axis for $1 \leq j \leq d$. Fix $y$, and let $A_j = A_j(y) \subseteq s^{-1}(y)$ be defined by

\begin{equation}
A_j := \{ \alpha \in s^{-1}(y) \mid |\partial_{\alpha_j} f(\alpha)| > \frac{\kappa_j}{d} \} \quad j = 1, \ldots, d.
\end{equation}

Since $|\nabla f| = \sqrt{\sum_{j=1}^{d} (\partial_{\alpha_j} f)^2} \geq \kappa_\ell$ on $\Omega$, for every $\alpha \in s^{-1}(y)$ at least one component of $\nabla f(\alpha)$ satisfies $|\partial_{\alpha_j} f| \geq \frac{\kappa_j}{d}$.\footnote{Since $\kappa_\ell \leq \sqrt{\sum_{j=1}^{d} (\partial_{\alpha_j} f)^2} \leq \sqrt{d} \max_{j=1,\ldots,d} |\partial_{\alpha_j} f|$, then $\max_{j=1,\ldots,d} |\partial_{\alpha_j} f| \geq \frac{\kappa_\ell}{\sqrt{d}} \geq \frac{\kappa_j}{d}.$}

Hence, $\alpha \in A_j(y)$ for some $1 \leq j \leq d$, and so

\begin{equation}
s^{-1}(y) = \cup_{j=1}^{d} A_j(y).
\end{equation}

**Step 2.** Next, we prove the existence of the maps $\phi_j : A_j \rightarrow f^{-1}(y)$.

**Lemma I.1.** Let $\alpha \in A_j(y)$ and let $h$ be defined as in Theorem 5.1. Then for a sufficiently small $h > 0$, there exists a real number $\delta = \delta(\alpha)$ such that

\begin{equation}
\phi_j(\alpha) \in \Omega_h = \Omega \setminus \{ f^{-1}(c) \}
\end{equation}



1. \( \alpha + \delta(\alpha) e_j \in f^{-1}(y) \), where \( e_j \) is the unit vector in the direction of the \( j \)-th axis.

2. The maps

\[
\phi_j(\alpha) := \alpha + \delta(\alpha) e_j, \quad j = 1, \ldots, d.
\]

are injective from \( A_j = A_j(y) \) to \( f^{-1}(y) \).

3. For every \( \alpha \in A_j \),

\[
\delta(\alpha) = O(h^{m+1}).
\]

4. For every \( E \subseteq A_j \),

\[
|\sigma(E) - \sigma(\phi_j(E))| = O(h^m),
\]

where as in (I.1), \( \sigma \) is the \( d-1 \) dimensional surface measure induced by the Lebesgue measure on \( \Omega \).

**Proof.**

1. We prove this for the case where \( y > f(\alpha) \) and \( \partial_{\alpha_j} f(\alpha) > 0 \) on \( \Omega \). The proofs for the three other cases are similar. Since \( f \in C^{m+1}(\Omega) \) and \( \Omega \) is compact, all the second derivatives of \( f \) are bounded, and so \( |\partial_{\alpha_j} f| < M_2 < \infty \) on \( \Omega \). Hence, since \( \partial_{\alpha_j} f(\alpha) > \frac{\kappa_j}{2} \), there exists a segment \( L = L(\alpha) = \{ \alpha + \xi e_j, \ |\xi| < \xi_{\max} \} \), where \( \xi_{\max} \) depends only on \( M_2 \), such that \( \partial_{\alpha_j} f \geq \frac{\kappa_j}{2} \) on \( L \). Therefore \( f(\alpha + \xi_{\max} e_j) > f(\alpha) + \frac{\kappa_j}{2} \xi_{\max} \). By the mean-value theorem, \( f \) attains on \( L \) all values in \( [f(\alpha), f(\alpha) + \frac{\kappa_j}{2} \xi_{\max}] \).

Now, by Theorem 5.1, since \( \alpha \in s^{-1}(y) \) and since \( y > f(\alpha) \),

\[
y - f(\alpha) = s(\alpha) - f(\alpha) \leq C_m h^{m+1}.
\]

Hence, for \( h \) sufficiently small, \( y \in [f(\alpha), f(\alpha) + \frac{\kappa_j \xi_{\max}}{2}] \), and so there exists a point \( \alpha + \delta(\alpha) e_j \in L \) such that \( f(\alpha + \delta(\alpha) e_j) = y \).

2. Assume by negation that \( \phi_j \) is not injective. Then there exist \( \alpha^1, \alpha^2 \in A_j \) such that \( \phi_j(\alpha^1) = \phi_j(\alpha^2) = \lambda \). Since \( \phi_j \) only changes the \( j \)-th coordinate, see (I.4), we can regard \( s \) and \( f \) as single-variable functions of the \( j \)-th coordinate \( \alpha_j \). Since \( \phi_j(\alpha^1) = \phi_j(\alpha^2) = \lambda \), from the proof of item (1) in this lemma it follows that \( \lambda \in L(\alpha^1) \cap L(\alpha^2) \).

Hence, the segment between \( \alpha^1 \) and \( \alpha^2 \) is contained in \( L(\alpha^1) \cup L(\alpha^2) \), where we know that \( |\partial_{\alpha_j} f| < \frac{\kappa_j}{2} \).

By Theorem 5.1, this means that if \( h \) is sufficiently small, \( |\partial_{\alpha_j} f| > \frac{\kappa_j}{2} \) on the segment between \( \alpha^1 \) and \( \alpha^2 \). This leads to a contradiction, since on the one hand \( \alpha^1, \alpha^2 \in A_j(y) \subseteq s^{-1}(y) \), and so \( s(\alpha^1) = s(\alpha^2) = y \), but on the other hand \( s(\alpha) \) is strictly monotone on the segment between \( \alpha^1 \) and \( \alpha^2 \).

3. Since \( f \in C^2 \), and by (I.4),

\[
\partial_{\alpha_j} f(\phi_j(\alpha)) - \partial_{\alpha_j} f(\alpha) = \partial_{\alpha_j} f(\alpha + \delta(\alpha) e_j) - \partial_{\alpha_j} f(\alpha) = O(\delta(\alpha)).
\]

In addition, by Lagrange mean-value theorem, for any \( \alpha \in s^{-1}(y) \),

\[
s(\alpha) - f(\alpha) = y - f(\alpha) = f(\alpha + \delta(\alpha) e_j) - f(\alpha) = \partial_{\alpha_j} f(\alpha + \zeta e_j) \cdot \delta(\alpha), \quad 0 \leq \zeta \leq \delta.
\]

Hence, using Theorem 5.1, and since \( |\partial_{\alpha_j} f| \geq \frac{\kappa_j}{2} \) on the segment between \( \alpha \) and \( \phi_j(\alpha) \) (see proof of item 1 in this lemma), we have that

\[
|\delta(\alpha)| = \left| \frac{s(\alpha) - f(\alpha)}{\partial_{\alpha_j} f(\alpha + \zeta e_j)} \right| \leq C_m h^{m+1} \frac{\kappa_j}{2} = O(h^{m+1}).
\]

29
4. For brevity of notations and without loss of generality, fix \( j = d \), and let \( E \subseteq A_d \). In this case, \( \partial_{\alpha_d}s \neq 0 \) on \( E \), and so by the implicit function theorem there exists a function \( S \), such that if \( s(\alpha_1, \ldots, \alpha_d) = y \), then \( \alpha_d = S(\alpha_1, \ldots, \alpha_{d-1}) \). The domain of \( S \) is

\[
G_E := \{ (\alpha_1, \ldots, \alpha_{d-1}) \mid \exists \alpha_d \in [0,1] \text{ s.t. } (\alpha_1, \ldots, \alpha_d) \in E \}.
\]

In particular, if \((\alpha_1, \ldots, \alpha_{d-1}) \in G_E\), then \( s(\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1})) = y \). Therefore

\[
\sigma(E) = \int_E 1 \, d\sigma = \int_{G_E} \sqrt{1 + |\nabla S|^2} \, d\alpha_1 \cdots d\alpha_{d-1}.
\]

Furthermore, by the implicit function theorem, \( \partial_{\alpha_d}S = -\frac{\partial_{\alpha_d}s}{\nabla s} \) for \( 1 \leq j < d \), and so

\[
\sqrt{1 + |\nabla S|^2} = \sqrt{1 + \sum_{j=1}^{d-1} \frac{\left( \partial_{\alpha_j}s \right)^2}{\left( \partial_{\alpha_d}s \right)^2} + \sum_{j=1}^{d-1} \left( \frac{\partial_{\alpha_j}s}{\partial_{\alpha_d}s} \right)^2} = \frac{1}{\left| \partial_{\alpha_d}s \right|} |\nabla s|.
\]

Hence,

(I.10a) \[ \sigma(E) = \int_{G_E} \frac{|\nabla s|}{\left| \partial_{\alpha_d}s \right|} \, d\alpha_1 \cdots d\alpha_{d-1}. \]

Next, since \( |\partial_{\alpha_d}f| \geq \frac{\kappa}{2} \) on \( \phi_d(E) \) (see proof of item 1 in this lemma), we similarly apply the implicit function on \( \phi_d(E) \): there exists function \( F : G_{\phi_d(E)} \to \mathbb{R} \) where \( G_{\phi_d(E)} \subset \mathbb{R}^{d-1} \), such that \( f(\alpha_1, \ldots, \alpha_{d-1}, F(\alpha_1, \ldots, \alpha_{d-1})) = y \). Hence, since \( \phi_d(E) \subset f^{-1}(y) \),

(I.10b) \[ \sigma(\phi_d(E)) = \int_{\phi_d(E)} 1 \, d\sigma = \int_{G_{\phi_d(E)}} \frac{|\nabla f|}{|\partial_{\alpha_d}f|} \, d\alpha_1 \cdots d\alpha_{d-1}. \]

Next, by item 2 of this lemma, then \( \phi_d \) induces a bijection \( \varphi_d : G_E \to G_{\phi_d(E)} \). But, because \( \phi_d \) only alters the \( \alpha_d \) coordinate, \( \varphi_d = \text{Id} \), and so \( G_E = G_{\phi_d(E)} \). Using this equality and (I.10) yields

(I.11) \[ |\sigma(E) - \sigma(\phi_d(E))| = \left| \int_{G_E} \left( \frac{|\nabla f(\phi_d(\beta))|}{|\partial_{\alpha_d}f(\phi_d(\beta))|} - \frac{|\nabla s(\beta)|}{|\partial_{\alpha_d}s(\beta)|} \right) \, d\alpha_1 \cdots d\alpha_{d-1} \right| \]

\[ \leq \int_{G_E} \frac{|\nabla f(\phi_d(\beta))| \cdot |\partial_{\alpha_d}s(\beta)| - |\nabla s(\beta)| \cdot |\partial_{\alpha_d}f(\phi_d(\beta))|}{|\partial_{\alpha_d}f(\phi_d(\beta))| \cdot |\partial_{\alpha_d}s(\beta)|} \, d\alpha_1 \cdots d\alpha_{d-1}, \]

where for brevity, we denote \( \beta : = (\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1})) \in E \) and note that by (I.4)

\( (\alpha_1, \ldots, \alpha_{d-1}, F(\alpha_1, \ldots, \alpha_{d-1})) = \phi_d(\beta) \).

To bound the right-hand-side of (I.11), note that since \( |\partial_{\alpha_d}f| > \frac{\kappa}{2} \) on \( E \), and since by Theorem 5.1, \( |\partial_{\alpha_d}s - \partial_{\alpha_d}f| \leq C_{m,h}^{\alpha_1} \), then for a sufficiently small \( h \), \( |\partial_{\alpha_d}s| > \frac{\kappa}{2} \) on \( E \). Substituting these bounds in (I.11) yields

(I.12) \[ |\sigma(E) - \sigma(\phi_d(E))| \leq \frac{2d}{\kappa^2} \int_{G_E} \left| \nabla f(\phi_d(\beta)) \cdot |\partial_{\alpha_d}s(\beta)| - |\nabla s(\beta)| \cdot |\partial_{\alpha_d}f(\phi_d(\beta))| \right| \, d\alpha_1 \cdots d\alpha_{d-1}. \]

\[ \]
Therefore, we can rewrite and bound the right-hand-side integrand by
\[(I.13)\]
\[
\left| |\nabla f(\phi_d(\beta))| \cdot |\partial_{\alpha_d}s(\beta)| - |\nabla s(\beta)| \cdot |\partial_{\alpha_d}f(\phi_d(\beta))| \right| \leq \\
|\partial_{\alpha_d}s(\beta)| \cdot \left( |\nabla f(\phi_d(\beta))| - |\nabla f(\beta)| \right) + |\nabla f(\beta)| \cdot \left( |\partial_{\alpha_d}s(\beta)| - |\partial_{\alpha_d}f(\phi_d(\beta))| \right) + |\partial_{\alpha_d}f(\phi_d(\beta))| \cdot \left( |\nabla f(\beta)| - |\nabla s(\beta)| \right).
\]

Since \( s, f \in C^2(\Omega) \) and \( \Omega \) is compact, \( \partial_{\alpha_d}s, \partial_{\alpha_d}f \) and \( \nabla f \) are bounded on \( \Omega \). Furthermore, since \( s, f \in C^2 \), the first and second term in the right-hand-side of \((I.13)\) are \( O(\delta) \), and so by \((I.5)\) both of these terms are \( O(h^{m+1}) \). In addition, by Theorem 5.1 the third and fourth term on the right-hand-side of \((I.13)\) are \( O(h^m) \). Hence, the left-hand-side of \((I.13)\) is \( O(h^m) \), and so finally, \((I.12)\) reads
\[
|\sigma(E) - \sigma(\phi_d(E))| \leq \frac{2d^2}{K_1} \int_{G_E} K h^m d\alpha_1, \ldots, d\alpha_d \leq K h^m,
\]
for some constant \( K > 0 \).

We finish this step by noting that Lemma I.1 would still hold if we interchange \( f \) and \( s \). Hence,

**Corollary I.2.** There exists sets \( B_j \subseteq f^{-1}(y) \) such that \( f^{-1}(y) = \bigcup_{j=1}^d B_j \) and maps \( \tilde{\phi}_j : B_j \to s^{-1}(y) \) for which items 1-4 of Lemma I.1 holds, interchanging \( f \) and \( s \).

**Step 3.** Next, we re-partition \( s^{-1}(y) \) into disjoint sets \( \{\tilde{A}_j\}_{j=1}^d \) where \( \tilde{A}_j \subseteq A_j \) for every \( 1 \leq j \leq d \). Let \( \tilde{A}_1 := A_1 \), and define
\[(I.14)\]
\[
\tilde{A}_j := A_j \setminus \left( \bigcup_{k=1}^{j-1} \tilde{A}_k \right), \quad 1 < j \leq d.
\]

Since by construction, \( \bigcup_{j=1}^d \tilde{A}_j = \bigcup_{j=1}^d A_j \), and since by \((I.3)\) \( \bigcup_{j=1}^d A_j = s^{-1}(y) \), then
\[
\bigcup_{j=1}^d \tilde{A}_j = s^{-1}(y).
\]

Hence, since the sets \( \{\tilde{A}_j\}_{j=1}^d \) are disjoint, we can rewrite the first component of \( I(y) \), see \((I.1)\), as
\[(I.15)\]
\[
\int_{s^{-1}(y)} \frac{1}{|\nabla s|} d\sigma = \sum_{j=1}^d \int_{\tilde{A}_j} \frac{1}{|\nabla s|} d\sigma.
\]

To prove a counterpart of \((I.15)\) for \( \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma \), we first prove the following Lemma:

**Lemma I.3.** Let \( \sigma \) be the surface measure on \( f^{-1}(y) \), let \( \{\tilde{A}_j\}_{j=1}^d \) be defined by \((I.15)\) and \( \{\phi_j\}_{j=1}^d \) be defined by \((I.4)\).

**1.** For any \( 1 \leq k, j \leq d \) with \( k \neq j \), then
\[(I.16)\]
\[
\sigma \left( \phi_j(\tilde{A}_j) \cap \phi_k(\tilde{A}_k) \right) = O(h^m).
\]

**2.**
\[(I.17)\]
\[
\sigma \left( f^{-1}(y) \setminus \bigcup_{j=1}^d \phi_j(\tilde{A}_j) \right) = O(h^m).
\]
1. Fix the indices \( j \neq k \) and denote for brevity \( D_{jk} = \phi_j(\tilde{A}_j) \cap \phi_k(\tilde{A}_k) \). Let \( \beta \in D_{jk} \). By injectivity of \( \phi_j \) and \( \phi_k \) (see Lemma I.1), There exist unique points \( \alpha^{(j)} \in \tilde{A}_j \) and \( \alpha^{(k)} \in \tilde{A}_k \) such that \( \phi_j(\alpha^{(j)}) = \phi_k(\alpha^{(k)}) = \beta \). By definition (I.4),

\[
\beta - \alpha^{(j)} = \delta(\alpha^{(j)}) \hat{e}_j, \quad \beta - \alpha^{(k)} = \delta(\alpha^{(k)}) \hat{e}_k.
\]

Since \( \hat{e}_j \perp \hat{e}_k \) and since by (I.5) \( \delta(\alpha^{(j)}), \delta(\alpha^{(k)}) = O(h^{m+1}) \), then\(^{31}\)

\[
|\alpha^{(j)} - \alpha^{(k)}| = O(h^{m+1}).
\]

Next, denote the geodesic distance on \( s^{-1} \) by \( | \cdot |_s \). Since \( s \in C^1 \), then \( |\nabla s| \) is bounded from above on \( \Omega \) and so \( |\alpha^{(j)} - \alpha^{(k)}|_s = O(h^{m+1}) \) as well. But since the interiors of \( \tilde{A}_j \) and \( \tilde{A}_k \) are disjoint, then the geodesic path between \( \alpha^{(j)} \) and \( \alpha^{(k)} \) must pass through a point \( \alpha^* \in \partial \tilde{A}_j \cap \partial \tilde{A}_k \). Hence,

\[
\phi_j^{-1}(D_{jk}) \subseteq E_{jk}(h) := \left\{ \alpha \in s^{-1}(y) \mid \inf_{\alpha^* \in \partial \tilde{A}_j \cap \partial \tilde{A}_k} |\alpha - \alpha^*|_s \leq K h^{m+1} \right\},
\]

for some \( K > 0 \). It is therefore sufficient to show that \( \sigma(E_{jk}(h)) = O(h^m) \) for \( 0 < h \ll 0 \). By construction, \( \partial \tilde{A}_j \cap \partial \tilde{A}_k \subseteq \cup_{j=1}^{d} \partial A_j \). Since \( f \in C^1 \), then \( \sigma(\cup_{j=1}^{d} \partial A_j) = 0 \) and so by monotonicity of measure \( \sigma(\partial \tilde{A}_j \cap \partial A_k) = 0 \) as well.\(^{32}\) Furthermore \( \partial A_j \cap \partial A_k \), is a finite union of smooth subsurface of \( s^{-1}(y) \), each of finite \((d-2)\)-dimensional surface measure.\(^{33}\) Finally, since \( \partial \tilde{A}_j \cap \partial \tilde{A}_k \) is compact in the topology of the smooth \((d-1)\)-dimensional manifold \( s^{-1}(y) \) (it is bounded and close), and since \( E_{jk}(h) \) is of geodesic radius \( Kh^{m+1} \) from \( \partial \tilde{A}_j \cap \partial \tilde{A}_k \), then \( \sigma(E_{jk}) = O((h^{m+1})^{(d-1)}) \leq O(h^m) \).

Hence,

\[
\sigma(\phi_j^{-1}(D_{jk})) \leq \sigma(E_{jk}(h)) = O(h^m).
\]

In addition, since \( \phi_j \) is injective, \( \phi_j(\phi_j^{-1}(D_{jk})) = D_{jk} \). Hence, by taking \( E = \phi_j^{-1}(D_{jk}) \) in (I.6) yields

\[
|\sigma(\phi_j^{-1}(D_{jk})) - \sigma(D_{jk})| = |\sigma(E) - \sigma(\phi_j(E))| \leq O(h^m).
\]

Combined with (I.19) this proves that \( \sigma(D_{jk}) = O(h^m) \), as required.

2. Since \( \cup_{j=1}^{d} \phi_j(\tilde{A}_j) \subseteq f^{-1}(y) \), then

\[
\sigma \left( \bigcup_{j=1}^{d} \phi_j(\tilde{A}_j) \right) \leq \sigma \left( f^{-1}(y) \right).
\]

On the other hand, by item (I.16) and by the inclusion-exclusion argument

\[
\sigma \left( \bigcup_{j=1}^{d} \phi_j(\tilde{A}_j) \right) = \sum_{j=1}^{d} \sigma(\phi_j(\tilde{A}_j)) - \sum_{j_1, j_2} \sigma(\phi_{j_1}(\tilde{A}_{j_1}) \cap \phi_{j_2}(\tilde{A}_{j_2})) + \cdots + (-1)^{d+1} \sigma(\phi_1(\tilde{A}_1) \cap \cdots \cap \phi_d(\tilde{A}_d)) = \sum_{j=1}^{d} \sigma(\tilde{A}_j) + O(h^m) = \sum_{j=1}^{d} \sigma(\tilde{A}_j) + O(h^m).
\]
where the last equality is due to (I.6). Hence,

(I.20b) \[ \sigma \left( \bigcup_{j=1}^{d} \phi_j(\tilde{A}_j) \right) = \sum_{j=1}^{d} \sigma(\tilde{A}_j) + O(h^m) = \sigma(\bigcup_{j=1}^{d} \tilde{A}_j) + O(h^m) = \sigma(s^{-1}(y)) + O(h^m), \]

where the second equality follows from the fact that the sets \( \{\tilde{A}_j\}_{j=1}^{d} \) are disjoint, and the third equality follows from \( \bigcup_{j=1}^{d} \tilde{A}_j = s^{-1}(y) \).

Since the left-hand-sides of (I.20a) and (I.20b) are identical, it follows that

(I.21a) \[ \sigma \left( s^{-1}(y) \right) + O(h^m) \leq \sigma \left( f^{-1}(y) \right). \]

Crucially, since by Corollary I.2, both Lemma I.1 and item 1 of this lemma remain valid if we interchange \( f \) and \( s \), we also have that

(I.21b) \[ \sigma \left( f^{-1}(y) \right) + O(h^m) \leq \sigma \left( s^{-1}(y) \right). \]

Combining the two inequalities of (I.21) yields that

(I.22) \[ |\sigma(f^{-1}(y)) - \sigma(s^{-1}(y))| = O(h^m). \]

Finally

\[ \sigma \left( f^{-1}(y) \setminus \bigcup_{j=1}^{d} \phi_j(\tilde{A}_j) \right) = \sigma \left( f^{-1}(y) \right) - \sigma \left( \bigcup_{j=1}^{d} \phi_j(\tilde{A}_j) \right) \leq |\sigma \left( f^{-1}(y) \right) - \sigma \left( s^{-1}(y) \right)| + O(h^m) = O(h^m), \]

where the inequality in the first line is due to (I.20b), and the last equality is due to (I.22).

**Step 4.** By (I.17), and since \( \frac{1}{|\nabla f|} \leq \frac{1}{\kappa_I} \), then

\[ \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma = \int_{\bigcup_{j=1}^{d} \phi_j(\tilde{A}_j)} \frac{1}{|\nabla f|} d\sigma + O(h^m). \]

Hence, by an inclusion-exclusion argument,

(I.23) \[ \int_{f^{-1}(y)} \frac{1}{|\nabla f|} d\sigma = O(h^m) + \sum_{j=1}^{d} \int_{\phi_j(\tilde{A}_j)} \frac{1}{|\nabla f|} d\sigma \]

\[ - \sum_{j_1 < j_2} \int_{\phi_{j_1}(\tilde{A}_{j_1}) \cap \phi_{j_2}(\tilde{A}_{j_2})} \frac{1}{|\nabla f|} d\sigma + \cdots + (-1)^{d-1} \int_{\phi_1(\tilde{A}_1) \cap \cdots \cap \phi_d(\tilde{A}_d)} \frac{1}{|\nabla f|} d\sigma. \]

But, by (I.16), we can reduce all of the higher-order terms to yield

(I.24) \[ \int_{f^{-1}(y)} \frac{1}{|\nabla s|} d\sigma = \sum_{j=1}^{d} \int_{\phi_j(\tilde{A}_j)} \frac{1}{|\nabla f|} d\sigma + O(h^m). \]

Hence, substituting (I.15) and (I.24) into (I.1) yields

(I.25) \[ I(y) \leq \sum_{j=1}^{d} \left| \int_{\phi_j(\tilde{A}_j)} \frac{1}{|\nabla f|} d\sigma - \int_{\tilde{A}_j} \frac{1}{|\nabla s|} d\sigma \right| + O(h^m). \]
Step 5. By (I.25), in order to show that \(I(y) = O(h^m)\), it is sufficient to prove that

\[
I_j(y) := \left| \int_{\phi_j(A_j)} \frac{1}{|\nabla f|} d\sigma - \int_{A_j} \frac{1}{|\nabla s|} d\sigma \right| = O(h^m), \quad 1 \leq j \leq d.
\]

This proof is similar to that of item 4 in Lemma I.1. For ease of notations, we assume without loss of generality that \(j = d\). In this case, \(\partial_{\alpha_d} s \neq 0\) on \(A_j\), and by the implicit function theorem there exists a function \(S\), such that if \(s(\alpha_1, \ldots, \alpha_d) = y\), then \(\alpha_d = S(\alpha_1, \ldots, \alpha_{d-1})\). The domain of \(S\) is

\[
G_{\tilde{A}_d} := \{ (\alpha_1, \ldots, \alpha_{d-1}) \mid \exists \alpha_d \in [0,1] \text{ s.t. } (\alpha_1, \ldots, \alpha_d) \in \tilde{A}_d \}.
\]

In particular, if \((\alpha_1, \ldots, \alpha_{d-1}) \in G_{\tilde{A}_d}\), then \(s(\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1})) = y\). Therefore

\[
\int_{\tilde{A}_d} \frac{1}{|\nabla s|} d\sigma = \int_{G_{\tilde{A}_d}} \frac{1}{\nabla s(\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1}))} \sqrt{1 + |\nabla S|^2} d\alpha_1 \ldots d\alpha_{d-1}.
\]

Furthermore, by the implicit function theorem, \(\partial_{\alpha_j} S = \frac{\partial s}{\partial \alpha_d} s\) for \(1 \leq j < d\), and so

\[
\sqrt{1 + |\nabla S|^2} = \sqrt{1 + \sum_{j=1}^{d-1} \left( \frac{\partial s}{\partial \alpha_d} s \right)^2} = \left( \frac{\partial s}{\partial \alpha_d} s \right)^2 + \sum_{j=1}^{d-1} \left( \frac{\partial s}{\partial \alpha_d} s \right)^2 = \frac{1}{\left| \frac{\partial s}{\partial \alpha_d} s \right|} |\nabla s|.
\]

Hence,

\[
(I.27a) \quad \int_{\tilde{A}_d} \frac{1}{|\nabla s|} d\sigma = \int_{G_{\tilde{A}_d}} \frac{1}{\left| \frac{\partial s}{\partial \alpha_d} s \right|} d\alpha_1 \ldots d\alpha_{d-1}.
\]

Similarly, since \(|\partial_{\alpha_d} f| \geq \frac{\kappa_d}{2\pi} > 0\) on \(\phi_j(\tilde{A}_j)\), applying the implicit function theorem to \(f\) yields a function \(F : G_{\phi_d(\tilde{A}_d)} \to \mathbb{R}\) where \(G_{\phi_d(\tilde{A}_d)} \subset \mathbb{R}^{d-1}\), such that \(f(\alpha_1, \ldots, \alpha_{d-1}, F(\alpha_1, \ldots, \alpha_{d-1})) = y\), and

\[
(I.27b) \quad \int_{\phi_d(\tilde{A}_d)} \frac{1}{|\nabla f|} d\sigma = \int_{G_{\phi_d(\tilde{A}_d)}} \frac{1}{\left| \frac{\partial f}{\partial \alpha_d} f \right|} d\alpha_1 \ldots d\alpha_{d-1}.
\]

Next, by item 2 of Lemma I.1, then \(\phi_d\) induces a surjective map \(\varphi_d : G_{\tilde{A}_d} \to G_{\phi_d(\tilde{A}_d)}\). But, because \(\phi_d\) only alters the \(\alpha_d\) coordinate, \(\varphi_d = \text{Id}\), and so \(G_{\tilde{A}_d} = G_{\phi_d(\tilde{A}_d)}\). Substituting this equality and (I.27) into (I.26) yields

\[
(I.28) \quad I_d(y) \leq \left| \int_{G_{\tilde{A}_d}} \left( \frac{1}{|\partial_{\alpha_d} f|} - \frac{1}{\left| \frac{\partial s}{\partial \alpha_d} s \right|} \right) d\alpha_1 \ldots d\alpha_{d-1} \right| \leq \int_{G_{\tilde{A}_d}} \frac{|\partial_{\alpha_d} f - \partial s|}{|\frac{\partial s}{\partial \alpha_d} s|} d\alpha_1 \ldots d\alpha_{d-1}.
\]

Bounding (I.28) is similar to its one-dimensional counterpart in Appendix F. Since \(|\partial_{\alpha_d} f| \geq \frac{\kappa_d}{d}\), and since by Theorem 5.1, \(|\partial s - \partial_{\alpha_d} f| \leq C_m h^m\), then for a sufficiently small \(h\), \(|\partial_{\alpha_d} s| > \frac{\kappa_d}{2\pi}\) on \(\phi_d(\tilde{A}_d)\). Substituting these bounds in (I.28) yields

\[
I_d(y) \leq \frac{2d^2}{\kappa_d^2} \int_{G_{\tilde{A}_d}} |\partial_{\alpha_d} s(\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1})) - \partial_{\alpha_d} f(\alpha_1, \ldots, \alpha_{d-1}, F(\alpha_1, \ldots, \alpha_{d-1}))| d\alpha_1 \ldots d\alpha_{d-1}.
\]

\[\text{As before, this follows for sufficiently small } h \text{ from the fact that } |\partial_{\alpha_d} f| \geq \frac{\kappa_d}{d} \text{ on } A_d(y), \text{ and from Theorem 5.1.} \]
Next, if we denote \( \beta = (\alpha_1, \ldots, \alpha_{d-1}, S(\alpha_1, \ldots, \alpha_{d-1})) \), then by (I.4) \[
\phi_d(\beta) = (\alpha_1, \ldots, \alpha_{d-1}, F(\alpha_1, \ldots, \alpha_{d-1})).
\]

Therefore, we can rewrite and bound the left-hand-side integrand by
\[
|\partial_{\alpha_d}s(\beta) - \partial_{\alpha_d}f(\phi_d(\beta))| \leq |\partial_{\alpha_d}s(\beta) - \partial_{\alpha_d}f(\beta)| + |\partial_{\alpha_d}f(\beta) - \partial_{\alpha_d}f(\phi_d(\beta))|.
\]

This bound is very similar to its one-dimensional counterpart in (F.3). The first term on the right-hand-side of (I.29) is \( O(h) \), see Theorem 5.1. In addition, since \( f \in C^2 \), the second term in the right-hand-side of (I.29) reads
\[
|\partial_{\alpha_d}f(\beta) - \partial_{\alpha_d}f(\phi_d(\beta))| \leq M_2|\beta - \phi_d(\beta)| = O(h^{m+1}),
\]
where, as before, \( M_2 = \max_{\Omega} |\partial^2_{\alpha_d}f| \) and the last equality is due to (I.7). Applying these bounds to (I.29) yields
\[
I_d(y) \leq \frac{2d^2}{K^2} K h^m \int_{\Gamma_{\alpha_d}} d\alpha_1 \cdots d\alpha_{d-1} = K h^m,
\]
for some constants \( \tilde{K}, K > 0 \). Moreover, since (I.30) holds for \( I_j(y) \) for all indices \( 1 \leq j \leq d \), then by (I.25)
\[
I(y) \leq \sum_{j=1}^d I_j(y) + O(h^m) \leq dK h^m + O(h^m).
\]

**Step 6.** Although \( \|p - \tilde{p}\|_1 = \int_{-\infty}^{\infty} I(y) \, dy \), since \( \Omega \) is compact and \( s \) and \( f \) are continuous, \( Q_1 \leq s, (\alpha), f(\alpha) \leq Q_2 \).

and so \( I(y) = 0 \) for \( y \notin [Q_1, Q_2] \). Hence, by (I.30)
\[
\|p - \tilde{p}\|_1 = \int_{-\infty}^{\infty} I(y) \, dy = \int_{Q_1}^{Q_2} I(y) \, dy \leq K h^m (Q_2 - Q_1).
\]
REFERENCES

[1] M.J. Ablowitz and T.P. Horikis. Interacting nonlinear wave envelopes and rogue wave formation in deep water. *Phys. Fluids*, 27:012107, 2015.

[2] G.P. Agrawal. *Nonlinear Fiber Optics*. 5th edition, Academic Press, 2012.

[3] R. K. Beatson. On the convergence of some cubic spline interpolation schemes. *SIAM J. Numer. Anal.* 23:903–912, 1986.

[4] R. K. Beatson and E. Chacko. Which cubic spline should one use? *SIAM J. Sci. Stat. Comput.* 13:1009–1024, 1992.

[5] G. Birkhoff and C. De Boor. Error bounds for spline interpolation. *J. Math. Mech.*, 13:827–836, 1964.

[6] Q.Y Chen, D. Gottlieb, and J.S. Hesthaven. Uncertainty analysis for the steady-state flows in a dual throat nozzle. *J. Comp. Phys.*, 204:378–398, 2005.

[7] I. Colombo, F. Nobile, G. Porta, A. Scotti, and L. Tamellini. Uncertainty Quantification of geochemical and mechanical compaction in layered sedimentary basins. *Comp. Meth. Appl. Mech. Eng.*, 328:122-146, 2018.

[8] P.J. Davis and P. Rabinowitz. *Numerical Integration*. Blaisdell, Waltham, Mass., 1967.

[9] C. De Boor. On cubic spline functions that vanish on all knots. *Adv. Math.*, 20:1-17, 1976.

[10] C. De Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

[11] B.J. Debusschere, H.N. Najm, P.P. Pébay, O.M. Knio, R. Ghanem, and O. Le Maıtre. Numerical challenges in the use of polynomial chaos representations for stochastic processes. *SIAM J. Sci. Comp.*, 26:698–719, 2004.

[12] S. Demko, W. F. Moss, and P. W. Smith. Decay rates for inverse of band matrices *Math. Comp.* 43:491–499, 1984.

[13] L. Devroye and L. Györfi. *Nonparametric Density Estimation - The L1 View*. Wiley, New York, 1985.

[14] A. Ditkowski and R. Kats. On spectral approximations with nonstandard weight functions and their implementations to generalized chaos expansions. *arXiv preprint arXiv:1611.00242*, 2016.

[15] R.L. Eubank. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, New York, 1999.

[16] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton FL, 1991.

[17] G. Fibich. *The Nonlinear Schrödinger Equation*. Springer, New York, 2015.

[18] R. J. Field and R. M. Noyes. Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *J. Chem. Phys.*, 60:1877–1884, 1974.

[19] F.N. Fritsch and R.E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 17:238–246, 1980.

[20] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *J. Comp. Phys.*, 225:652–685, 2007.

[21] R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of Uncertainty Quantification*. Springer, New York, 2017.

[22] R. Ghanem and P.D. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Springer-Verlag, New-York, 1991.

[23] G. H. Golub and C. F. Van Loan. *Matrix Computations*. JHU, Baltimore, MD, 2012.

[24] C. A. Hall and W. W. Meyer. Optimal error bounds for cubic spline interpolation. *J. Approx. Theory*, 16:105–122, 1976.

[25] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral Methods for Time-Dependent Problem* Cambridge, UK, 2007.

[26] C. R. Johnson. Positive definite matrices. *Amer. Math. Month.* 77:259–264, 1970.

[27] S. Kullback and R.A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[28] L. Le Cam and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Science & Business Media, New York, 2012.

[29] O. Le Maıtre and O.M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer, New-York, 2010.

[30] O.P. Le Maitre, O.M. Knio, H.N. Najm, and R. Ghanem. Uncertainty propagation using wiener–haar expansions. *J. Comp. Phys.*, 197:28–57, 2004.

[31] O.P. Le Maitre, O.M. Knio, H.N. Najm, and R. Ghanem. Asynchronous time integration for polynomial chaos expansion of uncertain periodic dynamics. *Discrete Contin. Dyn. Syst.*, 28:199–226, 2010.

[32] K.V. Mardia and P.E. Jupp. *Directional Statistics*. John Wiley & Sons, Chichester, UK, 2009.

[33] H.N. Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Ann. Rev. Fluid Mech.*, 41:35–52, 2009.

[34] A. O’Hagan. Polynomial chaos: A tutorial and critique from a statistician’s perspective. *SIAM/ASA J. Uncertainty Quantification*, 20:1–20, 2013.

[35] G. Patwardhan, X. Gao, A. Dutt, J. Ginsberg, and A.L. Gaeta. Loss of polarization in collapsing beams of
elliptical polarization. In Conference on Lasers and Electro-Optics (CLEO), pages 1–2. OSA, 2017.

[36] P.M. Prenter. Splines and Variational Methods. Courier Corporation, New-York, 2008.

[37] S. Randoux, N. Dalloz, and P. Suret. Intracavity changes in the field statistics of Raman fiber lasers. Opt. Lett., 36:790–792, 2011.

[38] J. R. Rice. Multivariate piecewise polynomial approximation. In Multivariate Approximation. edited by D.G. Handscomb. Academic Press, New York, 1978

[39] A. Sagiv, A. Ditkowski, and G. Fibich. Loss of phase and universality of stochastic interactions between laser beams. Opt. Exp., 25:24387–24399, 2017.

[40] M.D. Salas, S. Abarbanel, and D. Gottlieb. Multiple steady states for characteristic initial value problems. Appl. Numer. Math., 2:193–210, 1986.

[41] M. H. Schultz. \( L^\infty \)-Multivariate approximation theory. SIAM J. Numer. Anal. 6:161–183, 1969.

[42] A.H. Sheinfux, E. Schleifer, J. Papeer, G. Fibich, B. Ilan, and A. Zigler. Measuring the stability of polarization orientation in high intensity laser filaments in air. Applied Physics Letters, 101(20):201105, 2012.

[43] G. Stefanou. The stochastic finite element method: past, present and future. Comp. Meth. Appl. Mech. Eng., 198:1031–1051, 2009.

[44] B. Sudret and A. Der Kiureghian. Stochastic Finite Element Methods and Reliability: a State-of-the-Art Report. Department of Civil and Environmental Engineering, University of California Berkeley, Berkeley, CA, 2000.

[45] G. Szego. Orthogonal Polynomials, Colloquium Publications, Vol. 23. American Mathematical Soc., New York, 1939.

[46] L.N. Trefethen. Approximation Theory and Approximation Practice. SIAM, Philadelphia, PA, 2013.

[47] A. B. Tsybakov. Introduction to Nonparametric Estimation. Springer, New York, 2009.

[48] S. Ullmann. POD-Galerkin Modeling for Incompressible Flows with Stochastic Boundary Conditions. M.Sc. disseratation, Technical University of Darmstadt, 2014.

[49] Y. van Halder, B. Sanderse, B. Koren. An adaptive minimum spanning tree multi-element method for uncertainty quantification of smooth and discontinuous responses. arXiv preprint arXiv:1803.06833, 2018.

[50] G. Wahba. Spline Models for Observational Data. SIAM, 1990.

[51] X. Wan and G.E. Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. J. Comput. Phys., 209:617–642, 2005.

[52] H. Wang and S. Xiang. On the convergence rates of Legendre approximation. Math. Comp., 81:861–877, 2012.

[53] L. Wasserman. All of Statistics: A Concise Course in Statistical Inference. Springer Science & Business Media, New York, 2004.

[54] D. Xiu. Numerical Methods for Stochastic Computations: a Spectral Method Approach. Princeton University Press, Princeton, NJ, 2010.

[55] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput., 27:1118–1139, 2005.

[56] D. Xiu and G.E. Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput., 24:619–644, 2002.