A CHRISTOFFEL FUNCTION WEIGHTED LEAST SQUARES ALGORITHM FOR COLLOCATION APPROXIMATIONS

AKIL NARAYAN, JOHN D. JAKEMAN, AND TAO ZHOU

ABSTRACT. We propose, theoretically investigate, and numerically validate an algorithm for the Monte Carlo solution of least-squares polynomial approximation problems in a collocation framework. Our investigation is motivated by applications in the collocation approximation of parametric functions, which frequently entails construction of surrogates via orthogonal polynomials. A standard Monte Carlo approach would draw samples according to the density defining the orthogonal polynomial family. Our proposed algorithm instead samples with respect to the (weighted) pluripotential equilibrium measure of the domain, and subsequently solves a weighted least-squares problem, with weights given by evaluations of the Christoffel function. We present theoretical analysis to motivate the algorithm, and numerical results that show our method is superior to standard Monte Carlo methods in many situations of interest.

1. INTRODUCTION AND MAIN RESULTS

We consider the polynomial approximation of a function $f : \mathbb{R}^d \to \mathbb{R}$ using a least-squares collocation method. We are particularly interested in the case when the argument to f is a finite-dimensional random variable z (denoted lowercase throughout) with associated probability density function w. In this case, approximation of $f(z)$ is typically carried out in a w-weighted norm and can be constructed using a Monte Carlo procedure. This problem is particularly germane for parametric uncertainty quantification where f is usually a parameterized function with random parameter z [34]. Constructing a polynomial surrogate is a standard approach and is frequently explored via generalized Polynomial Chaos where f is expanded in a basis whose polynomial elements are orthogonal under the weight w [33, 36]. Using a collocation procedure to construct this polynomial is advantageous in practical large-scale simulations [25, 35].

While interpolatory approaches [2, 24] and compressive sampling or ℓ^1 regularization techniques [15, 27] are effective, the least-squares ℓ^2 regularization procedure is one of the simplest strategies that offers an attractive balance between cost and accuracy. Many existing methods for least-squares regression in this context concentrate on Monte Carlo approaches where the random variable ensemble $\{z_i\}_i$ is

sampled iid according to the weight function w [13, 22, 23, 29]. Alternative methods include the use of deterministic point constructions [40] or strategies involving subsampling from a "good" high-cardinality mesh [39].

This paper presents analysis and computational results for a type of weighted Monte Carlo approach for least-squares polynomial approximation that we call Christoffel Least Squares (CLS). The CLS prescription has two simple ingredients: given a probability weight/density w, we sample iid with respect to the (weighted) pluripotential equilibrium measure (*not* iid from w), and the weights are evaluations of the Christoffel function from the w-orthogonal polynomial family. The concrete procedures are shown in Algorithms 2 and 3. If one writes the least-squares problem in matrix formulation as an algebraic problem, weighting by the Christoffel function is equivalent to normalizing the system matrix so that each row has the same discrete ℓ^2 norm. The CLS algorithm is applicable for bounded and unbounded domains, with tensor-product or more general nontensor-product weights and domains.

Our analysis for the CLS method for polynomial approximation is based on the general least-squares theory presented in [14]. Given an N-dimensional subspace P of $L_w^2(D)$ for some closed set $D \subset \mathbb{R}^d$, let $\phi_n(z)$ denote any orthonormal family for P. We let $K(z)$ denote the "diagonal" of the reproducing kernel of P in L_w^2:

$$K(z) = \sum_{n=1}^{N} \phi_n^2(z).$$

Note that, fixing D and w, K does not depend on which orthonormal basis for P is chosen.[1] The analysis in [14] shows that a Monte Carlo least-squares approximation method with samples chosen iid from w is stable and accurate with high probability if the number of samples S satisfies

$$(1.1) \qquad \frac{S}{N \log S} \gtrsim C \frac{\|K\|_\infty}{N}$$

where C is a universal constant and $\|K\|_\infty \triangleq \max_{z \in D} K(z)$. Since K is a reproducing kernel diagonal and w is a probability density, then a lower bound on the value for $\|K\|_\infty$ is N:

$$N = \int_D \sum_{n=1}^{N} \phi_n^2(z) w(z) \mathrm{d}z \leq \left[\max_{z \in D} \sum_{n=1}^{N} \phi_n^2(z) \right] \int_D w(z) \mathrm{d}z = \|K\|_\infty;$$

i.e., the best (smallest) possible value of $\|K\|_\infty/N$ in (1.1) is unity. However, with P a total-degree polynomial space, for many weights w of interest the actual value of this quantity is very large and is quite sensitive to the choice of w (see Figure 1), and therefore makes the requirement for stability computationally onerous.

The CLS algorithm we present in this paper mitigates this situation by leveraging the fact that, for polynomials, the asymptotic behavior of the total-degree space reproducing kernel diagonal is known in great generality. (In this paper, "asymptotic" means with respect to the polynomial degree.) Let P_k denote the space of polynomials of degree k or less over $D \subset \mathbb{R}^d$, so that $N = \dim P_k = \binom{d+k}{d}$. We let K_k denote the L_w^2 reproducing kernel diagonal of P_k. The quantity N/K_k

---

[1]With $\boldsymbol{\phi}$ a vector containing the ϕ_n, then $K(z) = \boldsymbol{\phi}^T \boldsymbol{\phi}$. Thus, any change-of-basis $\boldsymbol{\psi} \leftarrow \boldsymbol{U}\boldsymbol{\phi}$ via any orthogonal matrix $\mathbf{U}$ preserves $K(z) = \boldsymbol{\psi}^T \boldsymbol{\psi} = \boldsymbol{\phi}^T \boldsymbol{\phi}$.

is the (normalized) Christoffel function from the theory of orthogonal polynomials (e.g., [26]), and is the eponymn of the CLS algorithm. If D is compact with nonvanishing interior and positive d-dimensional Lebesgue measure, and w is continuous on the interior of D and admits an orthogonal polynomial family, then

$$(1.2) \qquad \lim_{k \to \infty} \frac{N}{K_k(z)} = \frac{w(z)}{v(z)},$$

almost everywhere in D, where $v(z)$ is the Lebesgue weight function (a probability density) of the pluripotential equilibrium measure of D [6]. For example, in $d = 1$ dimension on the interval $D = [-1, 1]$, $v(z)$ is the arcsine or "Chebyshev" density. The utility of this statement for least-squares approximations is that the non-polynomial functions

$$\psi_n = \sqrt{\frac{N}{K_k(z)}} \phi_n(z),$$

form a basis for approximation in the space $\frac{1}{\sqrt{K_k}} P_k$, and are orthogonal in an L^2 space with the modified weight function $w \frac{K_k}{N}$. Owing to (1.2), the ψ_n are therefore approximately orthonormal with respect to v, and so an "approximate" reproducing kernel diagonal is given by

$$\widetilde{K}_k = \sum_n \psi_n^2 = \frac{N\, K_k}{K_k} = N,$$

and therefore this attains the the optimal (smallest) supremum value of N. Therefore, if we instead perform a Monte Carlo approximation with the ψ_n, sampling from v, then it may be possible to obtain the optimal sample-count stability criterion from (1.1) for most weights w of interest. This, in a nutshell, is the CLS algorithm. Although we have framed this discussion for compact domains and total-degree polynomial spaces, the CLS method may be applied for general polynomial subspaces on conic unbounded domains with exponential weights.

We present theoretical analysis following the results in [14] that crystallizes the motivation above, and accompanying numerical simulations show that the CLS algorithm significantly outperforms standard MC methods in many (but not all) scenarios of practical interest. For a general polynomial subspace P and its associated reproducing kernel diagonal K, the CLS algorithm performs approximation on the L^2 space weighted with $\widetilde{w} = \frac{N}{K(z)} v(z)$. Thus, the theory depends on a measure of discrepancy between $\widetilde{w}$ and w. One such measure on the space P is independent of the function being approximated: it is the $\widetilde{w}$-Gramian of the w-orthonormal basis ϕ_n, an $N \times N$ matrix $\mathbf{R}$ with entries:

$$(R)_{m,n} = \int_D \phi_m(z) \phi_n(z)\, \widetilde{w}(z) \mathrm{d}z.$$

Let Π denote the L_w^2-orthogonal projector onto P, where L_w^2 is the w-weighted L^2 space on D with norm $\|f\|^2 = \int_D f^2(z) w(z) \mathrm{d}z$. Similarly, let $\widetilde{\Pi}$ denote the $L_{\widetilde{w}}^2$-orthogonal projector onto P. A second w versus $\widetilde{w}$ discrepancy measure is data-dependent, the error in projections:

$$d(f) = \left\| \widetilde{\Pi}f - \Pi f \right\|_w,$$

which vanishes for any $f \in P$, or for any f when $\widetilde{w} = w$.

The following is one of our major theoretical results, summarizing our Theorem 5.2, and frames accuracy in terms of $d(f)$ and spectral quantities of $\mathbf{R}$.

Theorem. *Let D be compact. The CLS algorithm, i.e., discrete least-squares approximation by sampling iid from the equilibrium measure v and weighting with the inverse kernel diagonal $N/K(z)$, is stable with high probability if, for any $r > 0$, the number of samples S satisfies*

$$\frac{S}{N \log S} \geq C \frac{1 + r}{\lambda_{\min}(\mathbf{R})}$$

where C is an absolute constant. Let f be a function satisfying $|f| \leq L$, and let $\widetilde{\Pi}^S f$ denote the S-sample CLS estimator of f on P. Then under the sampling criterion above,

$$\mathbb{E}\left[\left\|f - T_L(\widetilde{\Pi}^S f)\right\|_w^2\right] \leq \|f - \Pi f\|_w^2 + \frac{\varepsilon(S)}{\lambda_{\min}(\mathbf{R})} \|f - \Pi f\|_{\widetilde{w}}^2 + \frac{8L^2}{S^r} + 4\kappa^2(\mathbf{R}) d^2(f)$$

where $T_L(x) = \operatorname{sgn}(x) \min\{|x|, L\}$ is a truncation function, and $\varepsilon(S) \sim \frac{1}{r \log S} \to 0$ as $S \to \infty$.

Above, $\lambda_{\min}(\mathbf{R})$ is the minimum eigenvalue and $\kappa(\mathbf{R})$ the 2-norm condition number of $\mathbf{R}$. Our numerical results for one dimension indicate that both of these quantities are very well-behaved for general w (see Figure 3) and thus are a significant improvement over the standard MC approach criterion (1.1). The scalar r introduced in the theorem above is a tunable factor that quantifies the oversampling rate, and is not a novel by-product of the CLS procedure or its theory. The projection discrepancy term $d(f)$ in the conclusion of the theorem above does not vanish as $S \to \infty$, and thus this theory appears uncompetitive with established theory for a "standard" Monte Carlo (MC) method. (We review some of this theory in Section 4.1.) However, our empirical investigation in Section 5.1.1 indicates that the error bound above for the CLS procedure has the same magnitude as the MC bounds. Although we cannot yet rigorously show the comparability of the CLS versus MC theory, our numerical results in Section 6 indicate that the CLS algorithm is frequently superior to a standard least-squares MC approach for polynomial approximation.

Although we have described only the bounded-domain case above, the CLS algorithm is also applicable on unbounded domains with exponential weights. The unbounded case presents no great difficulty in terms of analytical results comparable to the bounded case (see Theorems 5.3 and 5.4), but the implementation is less straightforward because an explicit formula for the sampling measure (the weighted pluripotential equilibrium measure) is not yet known for weights of interest. Nevertheless, we conjecture the forms of these weights and our simulations yield results that support our conjectures; see Table 2 and Section 6.

2. Setup

Let $D \subset \mathbb{R}^d$ be the domain and let $w : D \to \mathbb{R}$ be a weight function. We assume that the pair (D, w) is "admissible", by which we mean it falls into one of the following categories:

- (*bounded*) D is a compact set of nonzero d-dimensional Lebesgue measure and nonempty interior, and w is a continuous function on the interior of D such that $0 < \int_D p^2(z) w(z) \mathrm{d}z < \infty$ for any nontrivial algebraic polynomial p.

- (*unbounded*) D is an origin-centered unbounded conic domain (i.e., if $z \in D$, then $cz \in D$ for all $c \geq 0$) with nonzero d-dimensional Lebesgue measure, and $w = \exp(-2Q(z))$, with Q satisfying (i) $\lim_{|z|\to\infty} Q(z)/|z| > 0$, and (ii) there is a constant $t \geq 1$ such that

$$(2.1) \qquad Q(cz) = c^t Q(z), \qquad\qquad \forall\ z \in D,\ c > 0.$$

The condition (2.1) states that $Q = -\frac{1}{2}\log w$ is a t-homogeneous function. The unbounded case with the homogeneity condition on Q includes the following general family of weights:

$$Q(z) = \|z\|_p^t, \qquad\qquad \|z\|_p \triangleq \left(\sum_{k=1}^d |z_k|^p\right)^{1/p},$$

for any $p \geq 1$, which includes as special cases the one-sided exponential weight $w(z) = \exp\left(-\sum_j z_j\right)$ on $D = [0,\infty)^d$, and the Gaussian density function $w(z) = \exp\left(-\sum_j z_j^2\right)$ on $D = \mathbb{R}^d$. We will frequently write $|z|$ to mean $\|z\|_2$.

When citing results from pluripotential theory we will identify the function $\exp(-Q)$ as a "pluripotential weight function", which will be used a theoretical tool. For the CLS-unbounded case, note that specification of Q unique defines w. Whereas, for the CLS-bounded case we will take $Q = 0$ (for any admissible w). Such correspondences between w and Q are made in this paper to be consistent with notational conventions in pluripotential theory.

2.1. Orthogonal polynomials.

If (D, w) is admissible, then an $L_w^2(D)$ orthogonal polynomial family exists (see, e.g., [16, 20]). For a multi-index $\alpha \in \mathbb{N}_0^d$ we let ϕ_α denote the family of polynomials *orthonormal* under the w-weighted L^2 norm on D:

$$\langle \phi_\alpha, \phi_\beta \rangle_w = \int_D \phi_\alpha \phi_\beta w \, dz = \delta_{\alpha,\beta},$$

with $\|\cdot\|_w$ the corresponding induced norm on L_w^2. We implicitly assume that $\deg \phi_\alpha = |\alpha| \triangleq \alpha_1 + \cdots + \alpha_d$, with α_j the components of α. We use $\Lambda \subset \mathbb{N}_0^d$ to denote a general multi-index set, with $P \triangleq \mathrm{span}\{\phi_\alpha \mid \alpha \in \Lambda\}$ and the associated L_w^2-orthogonal projection onto P:

$$\Pi f \triangleq \sum_{\alpha \in \Lambda} c_\alpha \phi_\alpha, \qquad\qquad c_\alpha = \int_D f(z)\phi_\alpha(z) dz.$$

For notational simplicity, we suppress the explicit dependence of P and Π on the index set Λ.

Given any N-dimensional subspace P of L_w^2, in what follows we will study the "diagonal" of its reproducing kernel in L_w^2. This quantity is given by

$$(2.2) \qquad K(z) = \sum_{\alpha \in \Lambda} \phi_\alpha^2(z) = \sum_{n=1}^N \phi_n^2(z),$$

and is not dependent on the choice of basis ϕ_n. The above equation implicitly assumes a linear ordering of the elements in P:

$$\{\phi_\alpha\}_{\alpha \in \Lambda} \Longleftrightarrow \{\phi_n\}_{n=1}^N.$$

We will occasionally make use of this identification for notational convenience; the ordering of the indices in Λ with respect to $1, \ldots, N$ is irrelevant in our context.

In this paper we will consider general index sets Λ, but some of our theoretical results focus on the multi-index set $\Lambda_k \triangleq \left\{ \alpha \in \mathbb{N}_0^d \,|\, |\alpha| \leq k \right\}$ corresponding to the degree-k polynomial space. We will use the notation P_k, Π_k, and K_k in the special case $\Lambda = \Lambda_k$. See Table 1 for a summary of notation.

2.2. Discrete least-squares approximation.
We consider the problem of least-squares approximation using discrete collocation samples onto a polynomial space P defined by general index set Λ. This regression problem is a discrete approximation to a continuous projection. We approximate the L_w^2-orthogonal projection of a function $f(z)$ onto P by sampling f at discrete locations.

The continuous projection onto P satisfies

$$(2.3) \qquad \Pi f \triangleq \underset{p \in P}{\operatorname{argmin}} \|f - p\|_{L_w^2} = \underset{p \in P}{\operatorname{argmin}} \mathbb{E}\left[f(z) - p(z)\right]^2,$$

where in the latter equality we consider z a random variable with density w. In practice this optimal projection can rarely be computed because of insufficient knowledge about f. An alternative approach is discrete approximation: compute the minimizer of a discretization of the continuous norm. If $\{z_s\}$ are iid samples of the random variable z, then an approximation to Πf from (2.3) can be computed by using S of these samples:

$$(2.4) \qquad \Pi^S f = \underset{p \in P}{\operatorname{argmin}} \frac{1}{S} \sum_{s=1}^{S} |f(z_s) - p(z_s)|^2.$$

If the samples z_s are drawn iid from w, then it is straightforward to see that $\lim_{S \to \infty} \Pi^S f = \Pi f$, with more precise conditions on accuracy in [14], whose main results are reiterated in Section 4.1. More generally, one can take $\{z_s\}$ to be iid samples, but drawn from a different density v, in which case a weighted formulation is required to approximate the w norm

$$(2.5) \qquad \widetilde{\Pi}^S f = \underset{p \in P}{\operatorname{argmin}} \frac{1}{S} \sum_{s=1}^{S} k_s \, |f(z_s) - p(z_s)|^2.$$

Using a change-of-measure argument, we see that if we choose $k_s = w(z_s)/v(z_s)$ and the support of v contains the support of w, then $\lim_{S \to \infty} \widetilde{\Pi}^S f = \Pi f$.

In either case (2.4) or (2.5), if the S samples z_s are given, we can formulate the algebraic version of these problems. For a fixed index set Λ, recall that $N = N(\Lambda)$ denotes the dimension of P. Let $\mathbf{V}$ be the $S \times N$ Vandermonde-like matrix for the basis ϕ_n with samples z_s: $(V)_{s,n} = \phi_n(z_s)$. We may express the approximation $\widetilde{\Pi}^S$ from (2.5) in the basis ϕ_α:

$$\widetilde{\Pi}^S f = \sum_{\alpha \in \Lambda} c_\alpha \phi_\alpha(z).$$

We collect the unknown coefficients c_α into the vector $\mathbf{c}$, and the function evaluations $f(z_s)$ into the vector $\mathbf{f} \in \mathbb{R}^S$. The solution to (2.5) is defined by the least-squares solution to the following k_s-weighted problem:

$$(2.6) \qquad \mathbf{c} = \underset{\mathbf{g} \in \mathbb{R}^N}{\operatorname{argmin}} \left\| \sqrt{\mathbf{K}} \mathbf{V} \mathbf{g} - \sqrt{\mathbf{K}} \mathbf{f} \right\|^2,$$

where $\mathbf{K}$ is an $S \times S$ diagonal matrix with entries $(K)_{s,s} = k_s$. Equivalently, we may seek the solution to the normal equations:

$$(2.7) \qquad \mathbf{Gc} = \frac{1}{S}\mathbf{V}^T\mathbf{Kf},$$

where $\mathbf{G}$ is an $N \times N$ Gramian matrix with the random entries

$$(2.8) \qquad (G)_{n,m} = \langle \phi_n, \phi_m \rangle_S, \qquad \langle g, h \rangle_S \triangleq \frac{1}{S}\sum_{s=1}^{S} k_s g(z_s)h(z_s).$$

In this paper, we seek to specify the measure from which the z_s are drawn, and subsequently the weights k_s. We will see that in the CLS algorithm, $k_s \approx v(z_s)/w(z_s)$, i.e., that our change of measure is not exactly faithful to w. The unweighted algorithm associated with (2.4), proceeding by choosing $k_s \equiv 1$ and taking the sampling density equal to orthogonality density $v = w$, is given in Algorithm 1.

Algorithm 1: Unweighted least squares with Monte Carlo (LSMC)

---

input : Weight/density function w with associated orthonormal family ϕ_α, index set Λ, function f

output: Expansion coefficients $\mathbf{c}$ to approximate $\Pi_\Lambda f$

1 Generate S iid samples $\{z_s\}$ from density w;
2 Assemble $\mathbf{f}$ with entries $(f)_s = f(z_s)$;
3 Form $S \times N(\Lambda)$ Vandermonde-like matrix $\mathbf{V}$ with entries $(V)_{s,n} = \phi_{\alpha(n)}(z_s)$;
4 Compute $\mathbf{c} = \mathrm{argmin}_{\mathbf{g}\in\mathbb{R}^N} \|\mathbf{Vg} - \mathbf{f}\|$.

TABLE 1. Notation used throughout this article

Symbol(s)	
ϕ_α, ϕ_n	L_w^2-orthonormal polynomials: $\deg \phi_\alpha = |\alpha|$. A bijection between $\alpha \in \Lambda$ and $n \in \{1, 2, \ldots, N\}$ is assumed.
Λ, N	General multi-index set, a subset of $\mathbb{N}_0^d$: $N = |\Lambda|$
Λ_k	The degree-k total-degree multi-index set: $\Lambda_k = \left\{ \alpha \in \mathbb{N}_0^d \,\middle|\, |\alpha| \leq k \right\}$
P (P_k)	Polynomial space spanned by ϕ_α for all $\alpha \in \Lambda$ (P_k corresponds to $\Lambda = \Lambda_k$)
Π (Π_k)	L_w^2-orthogonal projector onto P (respectively, onto P_k)
K (K_k)	Reproducing kernel "diagonal" of P in L_w^2 (respectively, of P_k in L_w^2)
$\widetilde{w}$	Equals equilibrium density weighted by normalized Christoffel function; see Section 5
$\widetilde{\Pi}$ ($\widetilde{\Pi}_k$)	$L_{\widetilde{w}}^2$-orthogonal projector onto P (respectively, onto P_k)
$\mathbf{R}$ ($\mathbf{R}_k$)	$L_{\widetilde{w}}^2$ Gramian of ϕ_α for $\alpha \in \Lambda$ (respectively, for $\alpha \in \Lambda_k$)

2.3. Equilibrium measures. We review some results in weighted pluripotential theory with the goal of introducing the (weighted) equilibrium measure. In the CLS algorithm, this measure will define the sampling density v from Section 2.2. Standard references for pluripotential theory are [7, 18] and Appendix B of [28].

In the following brief discussion of pluripotential theory, we need to define the weight function $\exp(-Q)$ that is derived from the function Q. The weight function $\exp(-Q(z))$ on D serves as the "pluripotential weight function". In general our discussion in this subsection does not require $\exp(-Q)$ to be related to the orthogonality weight function w. However, in the context of the CLS algorithm,

we will identify the pluripotential weight function $\exp(-Q)$ with $\sqrt{w}$ for the CLS-unbounded case, whereas in the CLS-bounded case we will always take $Q = 0$, and thus $\exp(-Q) \equiv 1$.

Consider the class of plurisubharmonic functions on $\mathbb{C}^d$ that grow at most logarithmically at infinity:

$$\mathcal{L} = \left\{ u \text{ plurisubharmonic on } \mathbb{C}^d \mid u(z) \leq \max\{\log(z), 0\} + C \right\},$$

where C is a constant that depends on u. Given a pluripotential-theoretic-admissible domain D and weight $\exp(-Q(z))$, the *weighted extremal function* is the function

$$V_{D,Q}(z) = \left\{ \sup u(z) \mid u \in \mathcal{L} \text{ satisfying } u \leq Q \text{ on } D \right\}.$$

This function may be represented as an upper envelope of the logarithm of polynomials, whence the connection to polynomial approximation can be established. The regularization of this function defined as $V_{D,Q}^*(z) = \limsup_{\xi \to z} V_{D,Q}(\xi)$ is an upper-semicontinuous function. The weighted pluripotential equilibrium measure is given by

$$\mu_{D,Q} = \frac{1}{(2\pi)^d} \left(dd^c V_{D,Q}^* \right)^d,$$

where $(dd^c u)^d$ is the complex Monge-Ampère operator applied to u. (In the previous equation the superscript d is the integer dimension, whereas the normal-text d is a complex differential operator.) The measure $\mu_{D,Q}$ is a probability measure and has compact support. Sharp conditions under which $\mu_{D,Q}$ is absolutely continuous with respect to Lebesgue measure are not known in general, but some sufficient conditions are given in [3, 7].

In one dimension, these concepts reduce to one-dimensional concepts from potential theory: $V_{D,Q}$ is the weighted complex Green's function, dd^c is (proportional to) the complex Laplacian, and $\mu_{D,Q}$ is the potential-theoretic (weighted) equilibrium measure.

For the bounded CLS-admissible case, we need the unweighted ($Q \equiv 0$) measure $\mu_D \equiv \mu_{D,0}$; we will denote its Lebesgue density as $d\mu_D(z) = v_D(z)dz$ (assuming such a density exists). The bounded CLS-admissible assumptions above guarantee in this case that D is potential-theoretic admissible and so μ_D is well defined, and is a probability measure on D.

For the unbounded CLS-admissible case, we will take $\exp(-Q) = \sqrt{w}$ so that the potential-theoretic quantity Q is given by $Q = -\frac{1}{2}\log w$. We denote the Lebesgue density of the corresponding weighted equilibrium measure as $d\mu_{D,Q}(z) = v_{D,Q}(z)dz$ (assuming such a density exists). The weighted equilibrium measure $\mu_{D,Q}$ here has compact support even though D is unbounded.

With regards to the CLS algorithm, we use μ_D as the sampling measure for the bounded case, and a scaled version of $\mu_{D,Q}$ as a sampling measure for the unbounded case. Section 3 gives examples of the equilibrium measure density v_D.

3. Christoffel Least Squares

This section describes the novel algorithmic content of this paper: the Christoffel Least-Squares (CLS) algorithm applied to Monte Carlo approximation of L^2 projections. Essentially, this algorithm solves the problem (2.5) and thus requires (a) specification of the weights k_s (i.e., the matrix $\mathbf{K}$) and (b) specification of the

sampling measure $v(z)$. The CLS algorithm takes on different formulations when the domain D is bounded or unbounded, but a common formula in both cases is the specification of the weights k_s. For a general index set Λ, the CLS algorithm chooses k_s to be quantities that scale each row of $\sqrt{\mathbf{K}}\mathbf{V}$ to have ℓ^2 norm equal to the constant N, i.e.,

$$(3.1) \qquad k_s = \frac{N}{\sum_{\alpha \in \Lambda} \phi_\alpha^2(z_s)} = \left(\frac{K(z_s)}{|\Lambda|} \right)^{-1}.$$

Thus, the weights k_s are evaluations of the normalized Christoffel function.

When D is bounded, the CLS algorithm chooses the sampling weight function as $v = v_D$, the density function of the (unweighted) pluripotential equilibrium measure of D. When D is an unbounded conic domain, the sampling weight function v is a scaled version of the $\sqrt{w}$-*weighted* pluripotential equilibrium measure of the domain D. Thus, the particular specification of the sampling measure differs when D is bounded versus unbounded.

Informally, the CLS algorithm is reasonable because it adheres to the change-of-measure argument following equation (2.5): if v is the suitable equilibrium measure density and we consider approximation with $\Lambda = \Lambda_k$, then

$$(3.2) \qquad k_s = \frac{N}{K_k(z_s)} \sim \frac{w(z_s)}{v(z_s)}.$$

Indeed, this is true for very general weights and domains and is a major result in weighted pluripotential theory, which we discuss in Sections 4 and 5. We also observe in Section 6 that, even for nontotal-degree spaces, the CLS approximation performs quite well.

3.1. Bounded domains.
Let $w(z)$ be an admissible weight function on a compact domain D. In this case we sample with the density given by the (unweighted) equilibrium measure

$$(3.3) \qquad v(z) = v_D(z) = \frac{\mathrm{d}\mu_D}{\mathrm{d}z},$$

which is a probability measure. (For our discussion, we assume v_D exists.) We emphasize that this weight v is independent of the orthogonality density w. The weights k_s are as given in (3.1), and indirectly build in the dependence on w through (3.2). The method is shown in Algorithm 2.

Algorithm 2: Christoffel Least Squares (CLS) on a compact domain D

---

    input : Weight/density function w with associated orthonormal family p_α, polynomial index set Λ and corresponding size N, function f

    output: Expansion coefficients $\mathbf{c}$ to approximate Πu

1 Generate S iid samples $\{z_s\}$ from equilibrium measure μ_D;

2 Assemble $\mathbf{u}$ with entries $(u)_s = u(z_s)$;

3 Compute LS weights $\mathbf{K}$ with entries $(K)_{s,s} = N/K(z_s)$ from (3.1);

4 Form $S \times N$ Vandermonde-like matrix $\mathbf{V}$ with entries $(V)_{s,n} = \phi_n(z_s)$;

5 Compute $\mathbf{c} = \operatorname{argmin}_{\mathbf{g} \in \mathbb{R}^N} \left\| \sqrt{\mathbf{K}}\mathbf{V}\mathbf{g} - \sqrt{\mathbf{K}}\mathbf{u} \right\|$.

The pluripotential equilibrium measure generalizes the univariate potential-theoretic measure. In one dimension on $D = [-1, 1]$, the measure μ_D is the arcsine measure with "Chebyshev" density $v_D(z) = \frac{1}{\pi\sqrt{1-z^2}}$. Thus, on an interval the CLS algorithm prescribes Chebyshev sampling regardless of the weight. This conforms with the colloquially well-known observation that the Chebyshev measure on an interval is somehow "universal"; e.g., see [27].

The equilibrium measure for $D = [-1, 1]^d$ is the product measure of the univariate measure. For more complicated multivariate domains, computing $v_D(z)$ is not trivial, but some special cases have explicit formulas. For example, if D is the unit ball in $\mathbb{R}^d$, $D = \{z \in \mathbb{R}^d \,|\, \|z\|_2 \leq 1\}$, the equilibrium measure μ_D has density $v_D(z) = \frac{2}{V_d} [1 - \|z\|_2]^{-1/2}$ with V_d the volume of D [10]. Analysis for more general convex, origin-symmetric domains is given in [3]. When D is the unit simplex, $D = \{z \in \mathbb{R}^d \,|\, z_j \geq 0, \ 1 - \|z\|_1 \geq 0\}$, the equilibrium density is $v_D(z) = C \left[(1 - \|z\|_1) \prod_{j=1}^d z_j \right]^{-1/2}$, with $C = \pi^{(d+1)/2}/\Gamma((d+1)/2)$, see, e.g., [38]. A fairly general result for convex sets is given in [12].

Table 2 summarizes some formulas for $v_{D,Q}$. In that table, the following notation for sets is used:

$$(3.4a) \qquad B^d = \left\{ z = (z_1, \ldots, z_d) \in \mathbb{R}^d \,\Big|\, |z|^2 = \sum_{j=1}^d z_j^2 \leq 1 \right\},$$

$$(3.4b) \qquad T^d = \left\{ z = (z_1, \ldots, z_d) \in \mathbb{R}^d \,\Big|\, z_j \geq 0 \text{ and } \|z\|_1 = \sum_{j=1}^d z_j \leq 1 \right\}.$$

3.2. Unbounded domains.
We now consider the case of unbounded $D \subset \mathbb{R}^d$, for which we recall the assumption (2.1), that $Q = -\frac{1}{2} \log w$ is a homogeneous function of order t.

The CLS algorithm in this unbounded case chooses the weights as in (3.1), but the sampling weight function is a scaled version of the $\sqrt{w}$-weighted equilibrium measure

$$(3.5) \qquad v(z) = k^{-d/t} v_{D,Q}\left(k^{-1/t} z\right) = k^{-d/t} \frac{\mathrm{d}\mu_{D,Q}}{\mathrm{d}z}\left(k^{-1/t} z\right), \qquad k = \max_{\alpha \in \Lambda} |\alpha|.$$

Again, for the purposes of discussion we assume that the Lebesgue density $v_{D,Q}$ exists. The measure $\mu_{D,Q}$ has compact support, which is why we require scaling by $k^{1/t}$, effectively expanding the support to contain the "important" parts of the domain. We give the method in Algorithm 3. In the unbounded case, the scaling depends on the index set Λ defining the polynomial space P.

To our knowledge, explicit formulae for multivariate *weighted* equilibrium measures $\mu_{D,Q}$ on real-valued sets D are currently unknown for many real-valued sets D, even the "canonical" ones considered here. This is the case even for the simple case $\sqrt{w} = \exp(-|z|^2)$ so that $Q(z) = |z|^2$ on $D = \mathbb{R}^d$. However, we *conjecture* the following density for the equilibrium measure in this case:

$$\left. \begin{array}{l} Q(z) = \sqrt{w(z)} = \exp(-|z|^2), \\ D = \mathbb{R}^d \end{array} \right\} \implies \frac{\mathrm{d}\mu_{D,Q}}{\mathrm{d}z} = v_{D,Q}(z) = C \left[1 - |z|^2 \right]^{d/2},$$

where C is a normalization constant, and $\mu_{D,Q}$ is supported only on the set B^d. Similarly, for the weight $\sqrt{w} = \exp\left(-\sum_{j=1}^d z_j\right)$, we *conjecture* the following equilibrium measure density:

$$\left. \begin{aligned} Q(z) &= \sqrt{w(z)} = \exp\left(-\textstyle\sum_{j=1}^d z_j\right), \\ D &= [0,\infty)^d \subset \mathbb{R}^d \end{aligned} \right\}$$

$$\implies \frac{\mathrm{d}\mu_{D,Q}}{\mathrm{d}z} = v_{D,Q}(z) = C\sqrt{\frac{\left(2 - \sum_{j=1}^d z_j\right)^d}{\prod_{j=1}^d z_j}},$$

where again C is a normalization constant, $\mu_{D,Q}$ is supported only on points $z \in 2T^d$. In general, even computing just the support of $\mu_{D,Q}$ is a nontrivial task [1]. Note, however, in $d = 1$ dimension, the weighted equilibrium measure is explicitly known for a wide class of weights on bounded and unbounded real-valued sets, e.g., [28]. Our conjectured densities for $d \geq 1$ above specialize with $d = 1$ to these known densities.

Numerical experiments we have conducted support our conjectures above, and examples shown in Section 6 are generated using these sampling schemes and show very good performance, which further support our conjectures. Section 6 also gives methodology for sampling from our conjectured densities.

Algorithm 3: Christoffel Least Squares (CLS) on an unbounded D

---

input : Weight/density function w with associated orthonormal family p_α, index set Λ and corresponding size N, function f

output: Expansion coefficients **c** to approximate Πu

---

1. Compute log-weight homogeneity factor t from a $w = \exp(-2Q)$ identification in (2.1);
2. Generate S iid samples $\{z_s\}$ from equilibrium measure $\mu_{D,Q}$;
3. Expand samples: $z_s \leftarrow k^{1/t} z_s$, where $k = \max_{\alpha \in \Lambda} |\alpha|$;
4. Assemble **f** with entries $(f)_s = f(z_s)$;
5. Compute LS weights **K** with entries $(K)_{s,s} = N/K(z_s)$ from (3.1);
6. Form $S \times N(\Lambda)$ Vandermonde-like matrix **V** with entries $(V)_{s,n} = \phi_{\alpha(n)}(z_s)$;
7. Compute $\mathbf{c} = \operatorname{argmin}_{\mathbf{g} \in \mathbb{R}^N} \left\| \sqrt{\mathbf{K}}\mathbf{V}\mathbf{g} - \sqrt{\mathbf{K}}\mathbf{f} \right\|$.

We have completed specification of the CLS algorithm. It is a straightforward weighted Monte Carlo approach if, given D and w, the equilibrium measure v_D or $v_{D,Q}$ is known and may be sampled. Our analysis presented later indicates that v is the *asymptotically* optimal measure, but is not strictly optimal for a fixed Λ. In general, one could consider sampling instead from the measure $\widetilde{w}(z) = w(z)\frac{K(z)}{N}$, and this is briefly explored in [17].

4. BACKGROUND

This section recalls the two cornerstone results we require: general discrete least-squares stability and accuracy from [14], and asymptotics of the Christoffel function from [5,6].

TABLE 2. Explicit CLS sampling strategies for particular bounded and unbounded scenarios. The sets B^d and T^d are defined in (3.4). Entries preceded with $(**)$ are *conjectures* only. (Our conjectures specialize to known, correct results in one dimension [28].) For unbounded domains, the sampling density shown is over a compact domain; degree-scaled sampling according to Algorithm 3 should be performed, which depends on Λ.

Domain D	Orthogonality weight w	Sampling density domain	Sampling density $v(y) = \frac{d\mu_{D,Q}}{dy}$
$[-1,1]^d$	Any admissible weight	$[-1,1]^d$	$\frac{1}{\pi^d \prod_{j=1}^d \sqrt{1-y_j^2}}$
B^d	Any admissible weight	B^d	$\frac{C}{\sqrt{1-|y|^2}}$
T^d	Any admissible weight	T^d	$C\sqrt{\frac{1-\sum_{j=1}^d y_j}{\prod_{j=1}^d y_j}}$
$\mathbb{R}^d$	$\exp(-|z|^2)$	$(**)\sqrt{2}B^d$	$(**)C\left[2-|y|^2\right]^{d/2}$
$[0,\infty)^d$	$\exp\left(-\sum_{j=1}^d z_j\right)$	$(**)4T^d$	$(**)C\sqrt{\frac{\left(4-\sum_{j=1}^d y_j\right)^d}{\prod_{j=1}^d y_j}}$

4.1. **Least squares stability and convergence.** This section summarizes the main results of [14]. For a weight function w on $D \subset \mathbb{R}^d$, let ϕ_n for $n = 1, 2, \ldots$ be any w-orthonormal system (not necessarily polynomials). To be consistent with previous notation, we let P denote the subspace spanned by these elements, even though this need not be a polynomial space. The diagonal of the reproducing kernel $K(z)$ is as before in (2.2) and, with P fixed, does not depend on the particular choice of basis.

If we sample S iid realizations according to the weight w and form a discrete least-squares problem, the S-dependent Gramian matrix $\mathbf{G}$ in (2.8) satisfies $\lim_{S\to\infty} \mathbf{G} = \mathbf{I}$, with $\mathbf{I}$ the $N \times N$ identity matrix. $\mathbf{G}$ is a random matrix, and one can use random matrix estimates to precisely specify the probability with which $\mathbf{G}$ is close to $\mathbf{I}$. In turn, this yields estimates on accuracy of the least-squares solution. In the following, $\|\|\cdot\|\|$ is the spectral norm of a matrix.

Theorem 4.1 ([14]). *Let P be any N-dimensional subspace of L_w^2. Assume that $\{z_s\}_{s=1}^S$ are S iid samples drawn from the density w, and that the number of samples satisfies*

$$(4.1) \qquad \frac{S}{N \log S} \geq \left[\frac{1+r}{c_\delta}\right] \frac{\|K(z)\|_\infty}{N}$$

with $c_\delta \triangleq \delta + (1-\delta)\log(1-\delta)$ for some $\delta \in (0,1)$ and $r > 0$. Then the discrete Gramian matrix $\mathbf{G}$ given by (2.8) satisfies the following stability condition:

$$(4.2) \qquad \Pr\left[\,\|\|\mathbf{G} - \mathbf{I}\|\| > \delta\,\right] \leq \frac{2}{S^r}.$$

Furthermore, letting Π^S be the unweighted Monte Carlo projection operator defined in (2.4), then the following convergence result holds for any $f \in L_w^2$ satisfying
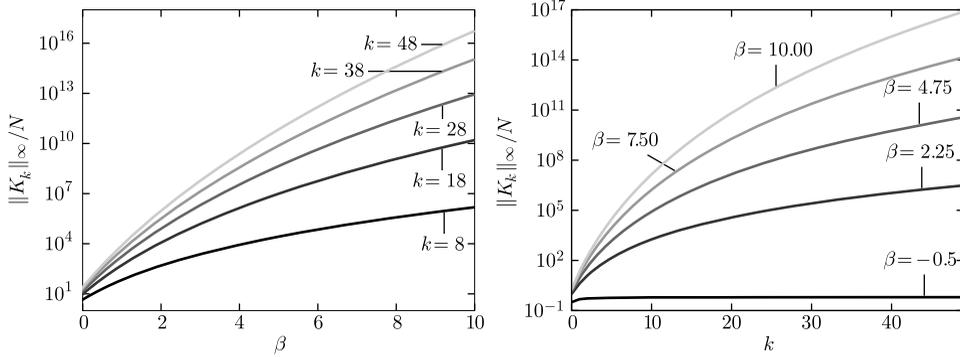
FIGURE 1. Values of the stability factor $\|K_k\|_\infty/N$ with $N = \dim P_k = k + 1$ for one-dimensional symmetric Jacobi polynomial families (parameters $\alpha = \beta$). Left: plots versus parameter β for various degrees k. Right: plots versus degree k for various parameters β.

$\|f\|_\infty \leq L$:

$$(4.3) \qquad \mathbb{E} \left\| f - T_L \left(\Pi^S f \right) \right\|_w^2 \leq [1 + \varepsilon(S)] \|g\|_w^2 + \frac{8L^2}{S^r},$$

where $g = (I - \Pi)f$ so that $\|g\|_w$ is the L_w^2 optimal error, $\varepsilon(S) \triangleq \frac{4c_\delta}{(1+r)\log S}$, and $T_L(x) = \operatorname{sgn}(x) \min\{|x|, L\}$ is a truncation function.

The critical term in the ensemble size condition (4.1) is the maximum of the reproducing kernel diagonal, $\|K\|_\infty/N$. Clearly this quantity depends on (i) the weight w and (ii) the space P.

When $D = [-1, 1]$ and $w = 1/\sqrt{1 - z^2}$ with P the degree-n polynomial space, a system of orthonormal polynomials is given by the Chebyshev polynomials $\phi_n(z) = T_n(z)$. The reproducing kernel diagonal for this one-dimensional basis satisfies

$$(4.4) \qquad \|K_k(z)\|_\infty = \sup_{z \in D} \sum_{j=0}^{k} T_j(z)^2 \equiv 2k + 1 = 2N - 1$$

for all $k \in \mathbb{N}$. Thus, so long as $S/(N \log S) \gtrsim C$ for an absolute constant C, then one can obtain stability from (4.2). In Section 6 showing our numerical experiments, we call the similar scaling $S \sim N \log N$ "log-linear" scaling of S with respect to N.

This optimal $\|K\|_\infty \sim N$ behavior only happens for very special weights. Consider the interval $D = [-1, 1]$ with weight function $w(z) = (1 - z^2)^\beta$ for $\beta \geq 0$ with corresponding orthonormal polynomial family ϕ_n. An understanding of how the choice of β affects the sampling criterion can be communicated by Figure 1. We show $\|K_k\|_\infty/N$ for various combinations of degree k and β values, with the conclusion that this quantity becomes extremely large when k or β is increased, and is very large even for moderate values of these parameters. Therefore, the restriction on the number of samples required for stability becomes onerous even on bounded sets in one dimension if one considers non-Chebyshev weights. Theoretical upper bounds for $\|K\|_\infty$ in the (tensor-product) multivariate case appear in [21]; their univariate behavior matches that shown in Figure 1.

Given the observations above, a standard MC approach is not tractable for all weights using this analysis, and in particular for polynomials on unbounded domains such as Hermite polynomials, one cannot directly use the above analysis if K is a polynomial kernel. (A straightforward remedy would be to instead use weighted polynomials, which is effectively what the CLS algorithm does.)

4.2. Christoffel function asymptotics.

For polynomials, the reproducing kernel diagonal quantity $K(z)$ is the inverse of the Christoffel function from the theory of orthogonal polynomials, about which much is known. The purpose of this section is to review formal results establishing the behavior in (3.2).

For unbounded domains we will need to discuss asymptotics of measures whose mass escapes to infinity; to make mathematically sensible statements in this case we adopt the strategy of "compressing" these measures by a scaling factor so that their mass remains on a compact interval. One way to implement this compression is to frame the discussion with respect to a *varying* weight w^k, i.e., w raised to the power k. The effect of using varying weights is that the measures we are interested in place their support on compact domains.

On unbounded domains D with a pluripotential-theoretic weight function $\rho = \exp(-Q)$ given, we consider polynomials orthonormal under the *varying* weight function ρ^{2k} for an integer k. For each $k \in \mathbb{N}$, let $\phi_\alpha^{(k)}$ be the family orthonormal under ρ^{2k}, i.e.,

$$(4.5) \qquad \int_D \phi_\alpha^{(k)}(z)\phi_\beta^{(k)}(z)\rho^{2k}(z)\mathrm{d}z = \delta_{\alpha,\beta}.$$

Then the reproducing kernel diagonal associated to P_k with the varying weight function ρ^{2k} is given by

$$K_k^{(k)} = \sum_{|\alpha| \leq k} \left(\phi_\alpha^{(k)}(z) \right)^2.$$

Note that $K_k^{(1)} \equiv K_k$, the standard L_w^2-kernel diagonal of P_k.

Asymptotics for both K (on bounded domains) and $K_k^{(k)}$ (unbounded domains) are known in many cases. In one dimension on $D = [-1, 1]$, the relation

$$(4.6) \qquad \lim_{k \to \infty} \frac{K_k(z)}{N} = \frac{1}{\pi w(z)\sqrt{1 - z^2}}$$

holds for almost every $z \in D$, for any w that is bounded and continuous. Similar results for $K_k^{(k)}$ on unbounded domains hold. (See, e.g., [9, 26, 30, 31].) Indeed, for more general w and multidimensional D, the above result holds if one replaces $\frac{1}{\pi\sqrt{1-z^2}}$ above by the Lebesgue density of the pluripotential equilibrium measure [10, 11, 19, 37, 38].

Since K is a specialization of $K^{(k)}$ when $Q \equiv 0$, it is sufficient to consider asymptotics of $K^{(k)}$, which is the goal of the following general result.

Theorem 4.2 ([4–6]). *Let D be a potential-theoretic admissible domain in $\mathbb{R}^d$ equipped with a smooth weight $\mathrm{d}V(z) = q(z)\mathrm{d}z$, along with a bounded and continuous weight function ρ such that $\rho(z)\mathrm{d}V(z)$ defines an orthonormal polynomial family ϕ_α in $L_{q\rho}^2(D)$. With $K_k^{(k)}$ the $L_{q\rho^{2k}}^2$ reproducing kernel diagonal of the total-degree*

polynomial space P_k, then the following convergence holds weakly:

$$\lim_{k \to \infty} \frac{1}{N} \rho^{2k}(z) K_k^{(k)}(z) \mathrm{d}V(z) = \mathrm{d}\mu_{D,Q}(z).$$

Note that the above is a very general result, but special cases of this (especially in one dimension) have been known long before the references cited in the theorem; see., e.g., [30] and references therein. We will specialize the theorem above to two cases: (i) on bounded D, we set $\rho \equiv 1$ and $\mathrm{d}V(z) = w(z)\mathrm{d}z$, (ii) on unbounded D with $w = \exp(-2Q)$, we set $\rho = \exp(-Q)$ and $\mathrm{d}V(z) = \mathrm{d}z$.

Corollary 4.1. *The following two cases are specializations of Theorem 4.2. For both cases below, we assume that the density $v_{D,Q}$ exists.*

(1) *Let w be any continuous weight function admitting an orthogonal polynomial basis on a compact, connected set D. With $\rho \equiv 1$ and $\mathrm{d}V = w(z)\mathrm{d}z$ then*

(4.7) $$\lim_{k \to \infty} \frac{1}{N} K_k(z) = \frac{\mathrm{d}\mu_D}{\mathrm{d}V} = \frac{v_D(z)}{w(z)},$$

with K_k the L_w^2 reproducing kernel diagonal for P_k.

(2) *Let $D \subset \mathbb{R}^d$ be an unbounded convex cone with $w = \exp(-2Q)$, and let $\rho(z) = \sqrt{w}(z)$ and $\mathrm{d}V(z) = \mathrm{d}z$. Then*

(4.8) $$\lim_{k \to \infty} \frac{1}{N} \rho^{2k}(z) K_k^{(k)}(z) = \frac{\mathrm{d}\mu_{D,Q}}{\mathrm{d}z} = v_{D,Q}(z),$$

with $K_k^{(k)}$ the $L_{w^k}^2$ reproducing kernel diagonal for P_k.

4.3. **Optimal measures.** The discussion of this section is not directly related to the goal of this paper, but this brief diversion provides the following useful message: If one chooses to perform an unweighted Monte Carlo least-squares approximation (Algorithm 1) with a polynomial subspace P_k, then sampling according to the equilibrium measure ($v = v_D$ for bounded D, $v = v_{D,Q}(k^{1/t}z)$ for unbounded D) gives the k-asymptotically optimal sampling criterion in the sense of Theorem 4.1. This observation is essentially a corollary on a convergence result of "optimal measures" as presented in [8]. The outline of this section is as follows:

- For any given probability measure[2] μ, the standard Monte Carlo procedure (Algorithm 1) has a required sample count criterion for stability and accuracy (Theorem 4.1).
- One can ask "for which μ is the procedure most stable?" and, because of (4.1), this stability can be quantified by minimizing the maximum of the μ-reproducing kernel diagonal K. The most stable measure will depend on the polynomial space P considered, and it is natural to consider total degree spaces P_k.
- This leads to a definition of an "optimal" measure μ_k corresponding to each degree, which is considered in [8].
- Any sequence of optimal measures converges to the equilibrium measure. This motivates the message that, *asymptotically* in k, the equilibrium measure is the most stable measure from which to perform discrete least-squares polynomial approximation.

---

[2]In this paper we are mainly interested in measures with Lebesgue densities w, but here we generalize by considering measures μ that are not necessarily absolutely continuous with respect to Lebesgue measure.

We note that the above does not imply anything about approximation quality; in practice one is interested in approximation in a specific norm and is not terribly concerned about the defined optimal measures above. We only seek to point out that, asymptotically in k, the most stable recontruction procedure would result from reconstruction according to the equilibrium measure, and therefore is a heuristic motivation for the CLS sampling choice.

On a compact domain $D \subset \mathbb{R}^d$, consider approximation on the total degree space P_k. Let $\phi_\alpha(z; \mu)$ denote the orthonormal polynomials for P_k under the L^2 norm with the probability measure μ on D. We are interested in choosing μ to optimize stability for a discrete least-squares problem. Thus, we consider the maximum value of the reproducing kernel diagonal as a function of the measure μ:

$$\kappa_k(\mu) = \max_{z \in D} \sum_{|\alpha| \leq k} \phi_\alpha^2(z; \mu), \qquad \int_D \phi_\alpha(z; \mu)\phi_\beta(z; \mu)\mathrm{d}\mu(z) = \delta_{\alpha,\beta}.$$

If we ask for the μ that minimizes $\kappa_k(\mu)$ over all probability measures, this leads to the notion of *optimal measures* as defined in [8]. Following the work in [8], a measure μ_k is *optimal* for D and P_k if, for all probability measures μ on D,

$$\kappa_k(\mu_k) \leq \kappa_k(\mu).$$

Algorithm 1 is most efficient for approximation with P_k in the sense of Theorem 4.1 and condition (4.1) when the weight w corresponds to the measure μ_k, because this choice of measure produces the minimal value of $\kappa_k = \|K_k\|_\infty$.

The notion of an optimal measure changes slightly for the weighted, unbounded case: we essentially need to build the exponentially-decaying part of the weight (i.e., $\exp(-2Q)$) into the least-squares problem in an intrinsic way, and then ask for the (optimal) measure that results in the most stable procedure for this exponentially-weighted least-squares problem. We let the exponentially-decaying part of the problem be defined by $\sqrt{w} = \rho = \exp(-Q)$ on an unbounded D. By building in the weights ρ^{2k} into the least-squares formulation, then the appropriate version of the quantity $\|K_k\|_\infty$ is given by $\sup_{z \in D} \rho^{2k}(z) \sum_\alpha \left[\phi_\alpha^{(k)}\right]^2(z)$, where the $\phi^{(k)}$ are ρ^{2k}-orthonormal as defined in (4.5). Similarly to the bounded case, we proceed to replace the measure $\rho^{2k}(z)\mathrm{d}z$ in (4.5) with $\rho^{2k}(z)\mathrm{d}\mu(z)$ for some probability measure μ on D, and define the resulting kernel maximum:

$$\kappa_{k,\rho}(\mu) = \max_{z \in D} \rho^{2k}(z) \sum_{|\alpha| \leq k} \left[\phi_\alpha^{(k)}(z; \mu)\right]^2.$$

The measure μ_k is an optimal measure for D with weight ρ if, for all probability measures μ on D:

$$\kappa_{k,\rho}(\mu_k) \leq \kappa_{k,\rho}(\mu).$$

Again this notion of an optimal measure indicates which sampling measure produces the smallest sampling size requirement in (4.1).

In either the bounded or unbounded case we want to sample from μ_k to achieve an optimal stability factor. The following main result from [8] indicates that, as the polynomial degree k tends to infinity, any sequence of (weighted) optimal measures converges to the (weighted) equilibrium measure.

Theorem 4.3 ([8]). *Let μ_k be an optimal measure for P_k on D with weight $\rho = \exp(-Q)$. We have* (i) *for each μ_k, $\kappa_{k,\rho} = N$ $\mu_{D,Q}$-a.e., and* (ii) $\lim_{k \to \infty} \mu_k = \mu_{D,Q}$ *weakly.*

Note that the above result holds also in the unweighted case $Q \equiv 0$. While computing μ_k for each k will not be tractable in most situations, the result indicates that the optimal sampling measure for these least-squares problems must *asymptotically* be μ_D in the bounded domain case, or a scaled version of $\mu_{D,Q}$ for the unbounded case. This result does not imply any optimality for a fixed k, and so in principle sampling with the equilibrium measure may be quite suboptimal if k is small enough so that μ_k deviates significantly from $\mu_{D,Q}$.

Note also that Theorem 4.3 indicates that the stability factor κ_k *asymptotically* attains its optimal (minimal) value of N, which, according to Theorem 4.1 results in *asymptotically* simple log-linear scaling of S with respect to N, the best possible sample count criterion in the sense of Theorem 4.1. Therefore, if k could be taken very large, then sampling with the equilibrium measure will eventually produce a near-optimal $S \log S \gtrsim N$ sampling criterion for stability. While this seems promising since the asymptotic result is dimension-independent, it is not clear how large k must be relative to d to see this asymptotic behavior. In addition, it is computationally infeasible to use large k for high-dimensional simulations since $\dim P_k \sim k^d$.

5. Asymptotics of the Christoffel Least Squares algorithm

This section concentrates on showing that the limiting behavior of Algorithms 2 and 3 is stable and accurate. Our estimates depend on a discrepancy measure between the orthogonality weight w and the effective CLS weight $\widetilde{w} \triangleq v\frac{N}{K}$, with v the sampling density prescribed in Sections 3.1 or 3.2. Our convergence analysis is less constructive than the stability analysis, because the former depends on a w versus $\widetilde{w}$ reprojection error, which is not easily computable.

We recall some of our notation in Table 1 for clarity: Λ is a multi-index set that defines a polynomial subspace P, its L_w^2 reproducing kernel diagonal K, the L_w^2-orthogonal projector Π whose range is P, the $L_{\widetilde{w}}^2$-orthogonal projector $\widetilde{\Pi}$ whose range is P, and the $L_{\widetilde{w}}^2$ Gramian matrix of the ϕ_n, $\mathbf{R}$. When $\Lambda = \Lambda_k$ corresponds to the polynomial space of total degree k, we use the abbreviated versions P_k, Π_k, $\widetilde{\Pi}_k$, and $\mathbf{R}_k$.

5.1. CLS for bounded domains.
We assume the pair (D, w) are bounded-admissible in the sense of Section 2. Our results in this section and the one immediately following are essentially adaptations of the results reproduced in Sections 4.1 and 4.2. This section deals with compact domains.

The CLS framework is a weighted least-squares formulation; alternatively, we may consider it an unweighted least-squares problem with the nonpolynomial functions

$$(5.1) \qquad \psi_\alpha(z) = \frac{\sqrt{N}\phi_\alpha(z)}{\sqrt{\sum_{\alpha \in \Lambda} \phi_\alpha^2(z)}},$$

followed by sampling with the equilibrium measure weight $v(z)\mathrm{d}z = \mathrm{d}\mu_D(z)$. This essentially uses a modified weight function that approximates w:

$$\widetilde{w}(z) = \frac{N}{K(z)}v_D(z).$$

The functions ψ_α are not exactly orthogonal with respect to the weight $v_D(z)$, and we will need a quantification of this nonorthogonality behavior as a function of the
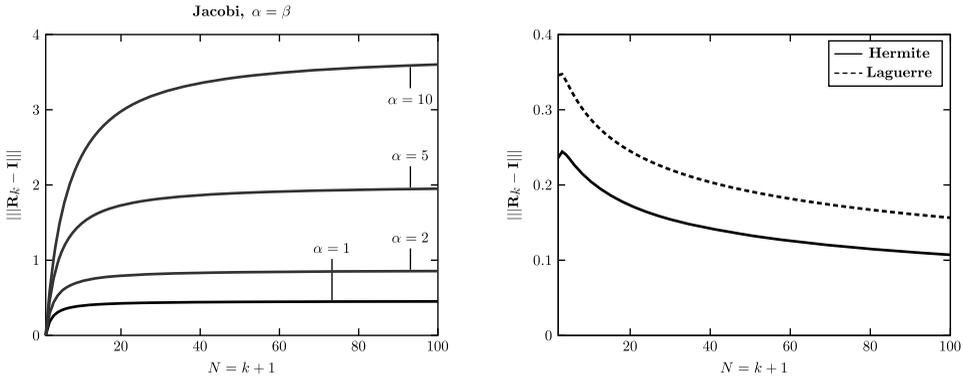
FIGURE 2. Deviation of the matrix $\mathbf{R}_k$ from the $N \times N$ identity for one-dimensional domains D. Left: bounded domain $D = [-1, 1]$ with $\mathbf{R}_k$ defined in (5.2) with one-dimensional orthogonality density $w(z) = (1 - z^2)^\alpha$. Right: unbounded domain D with weights $w(z) = \exp(-z^2)$ and $w(z) = \exp(-z)$ defining $\mathbf{R}_k$ as in (5.14).

index set Λ. For a given w and fixed index set Λ, define

$$(5.2) \qquad (R)_{\alpha,\beta} = \int_D \psi_\alpha(z)\psi_\beta(z)\mathrm{d}\mu_D = \int_D \phi_\alpha(z)\phi_\beta(z)\widetilde{w}(z)\mathrm{d}z,$$

so that the $N \times N$ matrix $\mathbf{R}$ is defined. We emphasize that the functions ψ_α and the weight $\widetilde{w}$ depend on Λ. We use the notation $\mathbf{R} = \mathbf{R}_k$ to denote the special case of $\Lambda = \Lambda_k$. Owing to asymptotics of the Christoffel function, any fixed (α, β) entry of the matrix $\mathbf{R}_k$ converges to the corresponding entry of the identity matrix.

Proposition 5.1. *For any fixed α, β, the quantity in (5.2) for the polynomial space P_k satisfies*

$$\lim_{k \to \infty} (R_k)_{\alpha,\beta} = \delta_{\alpha,\beta}.$$

Proof. The entries of $\mathbf{R}_k$ are given by

$$(R_k)_{\alpha,\beta} = \int_D \phi_\alpha(z)\phi_\beta(z)\frac{N}{K_k(z)}v(z)\mathrm{d}z.$$

The result (4.7) implies that $\widetilde{w} = \frac{N}{K_k(z)}v(z)$ converges to $w(z)$ weakly, so that for fixed α, β,

$$(R_k)_{\alpha,\beta} \to \int_D \phi_\alpha\phi_\beta w \,\mathrm{d}z = \delta_{\alpha,\beta}. \qquad \square$$

The above is an asymptotic result indicating that individual terms of the matrix $\mathbf{R}_k$ behave like terms of the identity. However, this cannot be used to conclude that $\mathbf{R}_k$ is close to the identity matrix in, e.g., the induced ℓ^2 norm for increasing k since the size of $\mathbf{R}_k$ also increases with k. To illustrate this, we compile results for the one-dimensional domains $D = [-1, 1]$ with symmetric Jacobi weights $w(z) = (1 - z^2)^\alpha$ in the left-hand pane of Figure 2. These results alone cannot even be used to conclude that $d = 1$ cases for $\mathbf{R}_k$ are well-conditioned. However, one can combine Figure 2 with Figure 3 to see that in fact the $\mathbf{R}_k$ are relatively well-behaved.

Turning to stability, since $\mathbf{R}$ is not the identity, we can only expect the CLS normal equations matrix $\mathbf{G}$ to converge to $\mathbf{R}$ as the sample count increases. (Recall the definition of $\mathbf{G}$ from (2.7).) The same arguments as in [14] may be applied to conclude an analogue of the stability result in Theorem 4.1: A sampling size criterion implies that the discrete Gramian $\mathbf{G}$ of the CLS procedure is close to $\mathbf{R}$ with high probability.

Theorem 5.1. *For a compact domain D and admissible weight w with index set Λ, consider the CLS algorithm, Algorithm 2. If the number of samples S satisfies*

$$(5.3) \qquad \frac{S}{N \log S} \geq \left[\frac{1+r}{c_\delta} \right] \frac{1}{\lambda_{\min}(\mathbf{R})},$$

for $\delta \in (0,1)$ and $r > 0$, with c_δ defined in (4.1), then

$$\Pr \left[\frac{\|\|\mathbf{G} - \mathbf{R}\|\|}{\|\|\mathbf{R}\|\|} > \delta \right] \leq \frac{2}{S^r}.$$

Proof. The proof of Theorem 4.1 in [14] is easily amended for our purposes. Since $\mathbf{R}$ is symmetric positive-definite, its symmetric positive-definite square root $\mathbf{R}^{1/2}$ is well defined. Then

$$\|\|\mathbf{G} - \mathbf{R}\|\| = \left\|\left\| \mathbf{R}^{1/2} \left(\mathbf{R}^{-1/2} \mathbf{G} \mathbf{R}^{-1/2} - \mathbf{I} \right) \mathbf{R}^{1/2} \right\|\right\|$$
$$\leq \|\|\mathbf{R}\|\| \left\|\left\| \mathbf{R}^{-1/2} \mathbf{G} \mathbf{R}^{-1/2} - \mathbf{I} \right\|\right\|,$$

and so

$$(5.4) \qquad \Pr \left[\frac{\|\|\mathbf{G} - \mathbf{R}\|\|}{\|\|\mathbf{R}\|\|} > \delta \right] \leq \Pr \left[\left\|\left\| \mathbf{R}^{-1/2} \mathbf{G} \mathbf{R}^{-1/2} - \mathbf{I} \right\|\right\| > \delta \right].$$

The CLS Gramian matrix $\mathbf{G}$ can be decomposed into a sum of independent matrices

$$\mathbf{G} = \sum_{s=1}^{S} \mathbf{Y}_s, \qquad\qquad (\mathbf{Y}_s)_{\alpha,\beta} = \frac{N}{S} \left[\frac{\phi_\alpha(z_s) \phi_\beta(z_s)}{K(z_s)} \right].$$

Defining $\mathbf{X}_s = \mathbf{R}^{-1/2} \mathbf{Y}_s \mathbf{R}^{-1/2}$, then $\mathbf{R}^{-1/2} \mathbf{G} \mathbf{R}^{-1/2} = \sum_s \mathbf{X}_s$. The spectral norm of $\mathbf{X}_s$ satisfies

$$\|\|\mathbf{X}_s\|\| \leq \left\|\left\|\mathbf{R}^{-1}\right\|\right\| \|\|\mathbf{Y}_s\|\|.$$

Since $\mathbf{Y}_s$ is a rank-1 matrix formed from the outer product of $\left(\sqrt{\frac{N}{SK(z_s)}} \phi_\alpha(z_s) \right)_\alpha$ with itself, then

$$\|\|\mathbf{X}_s\|\| \leq \frac{N}{S \lambda_{\min}(\mathbf{R})} \frac{\sum_{\alpha \in \Lambda} \phi_\alpha^2(z_s)}{K(z_s)} = \frac{N}{S \lambda_{\min}(\mathbf{R})}$$

with probability 1. The summed expected value of $\mathbf{X}_s$ yields the identity matrix: $\sum_{s=1}^{S} \mathbb{E} \mathbf{X}_s = \mathbf{R}^{-1/2} \mathbf{R} \mathbf{R}^{-1/2} = \mathbf{I}$.

Now we can use the matrix Chernoff bound from [32]: for any collection of independent random matrices $\mathbf{X}_s$ satisfying $\|\!|\!|\mathbf{X}_s|\!|\!\| \leq M$, then

$$\Pr\left[\lambda_{\min}\left(\sum_{s=1}^{S} \mathbf{X}_s\right) \leq (1-\delta)\lambda_{\min}\left(\sum_{s=1}^{S} \mathbb{E}\mathbf{X}_s\right)\right]$$

$$\leq N\exp\left(-\frac{c_\delta\lambda_{\min}\left(\sum_{s=1}^{S}\mathbb{E}\mathbf{X}_s\right)}{M}\right),$$

$$\Pr\left[\lambda_{\max}\left(\sum_{s=1}^{S} \mathbf{X}_s\right) \geq (1+\delta)\lambda_{\max}\left(\sum_{s=1}^{S} \mathbb{E}\mathbf{X}_s\right)\right]$$

$$\leq N\exp\left(-\frac{c_\delta\lambda_{\max}\left(\sum_{s=1}^{S}\mathbb{E}\mathbf{X}_s\right)}{M}\right),$$

where $c_\delta = \delta + (1-\delta)\log(1-\delta) \in (0,1)$. We use $\mathbf{G} = \sum_s \mathbf{X}_s$ and $\sum_s \mathbb{E}\mathbf{X}_s = \mathbf{I}$ along with the bound $M = \frac{N}{S\lambda_{\min}(\mathbf{R})}$. Thus we have

$$\Pr\left[\left|\!\left|\!\left|\mathbf{R}^{-1/2}\mathbf{G}\mathbf{R}^{-1/2} - \mathbf{I}\right|\!\right|\!\right| \geq \delta\right]$$

(5.5)
$$\leq N\left[\exp\left(-\frac{c_\delta S\lambda_{\min}(\mathbf{R})}{N}\right) + \exp\left(-\frac{c_\delta S\lambda_{\min}(\mathbf{R})}{N}\right)\right]$$

$$\leq 2S\exp\left(-\frac{c_\delta S\lambda_{\min}(\mathbf{R})}{N}\right).$$

If we require (5.3), then

(5.6)
$$\exp\left(-\frac{c_\delta S\lambda_{\min}(\mathbf{R})}{N}\right) \leq \exp\left(-(1+r)\log S\right) = \frac{1}{S^{1+r}}.$$

Combining (5.4), (5.5), and (5.6) yields the result. $\qquad\square$

Remark 5.1. The above result is a stability estimate for a general weighted least-squares approach for any biased weight $\widetilde{w} \neq w$.

We emphasize that we have only established stability of the least-squares problem relative to $\mathbf{R}$. The required sample count for stability no longer depends on the normalized polynomial reproducing kernel $K(z)/N$, but instead on $\lambda_{\min}^{-1}(\mathbf{R})$, which is a stability measure for $\mathbf{R}$. (Compare (4.1) with (5.3).) In Figure 3 we plot the inverse of the minimum eigenvalue for one-dimensional cases: symmetric Jacobi polynomials (with parameter $\beta > -1$) and also for Hermite and Laguerre polynomials.

For most of these one-dimensional cases of interest with the space P_k, the factor $1/\lambda_{\min}(\mathbf{R}_k)$ is less than 2. This is in stark contrast to the results in Figure 1 where the stability factor of $\frac{\|K\|_\infty}{N}$ of (4.1) from similar one-dimensional scenarios is extremely large.

While these results are promising in one dimension for a large, fixed degree k, they will likely deteriorate as the dimension d is increased with a fixed k. Similarly, Figures 3 and 2 can be used to conclude that $\mathbf{R}_k$ is quite well-conditioned for classical one-dimensional problems, but this is unlikely to persist for large dimensions.
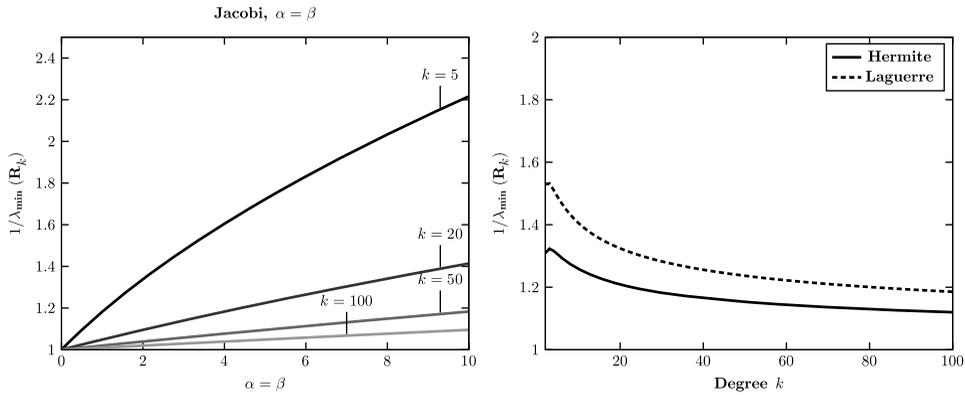
FIGURE 3. Inverse of the minimum eigenvalue of the matrix $\mathbf{R}_k$ for various one-dimensional polynomial families. Left: Jacobi polynomials with symmetric parameters $\alpha = \beta$. Right: Hermite and Laguerre polynomials.

We are able to prove a convergence result if $|f(z)| \leq L$ for all $z \in D$. The random function $\widetilde{\Pi}^S f$ is the CLS-bounded discrete projection. We introduce a truncation of this discrete projection:

$$(5.7) \qquad \widetilde{f}(z) = T_L \left[\widetilde{\Pi}^S f \right],$$

where $T_L(x) = \operatorname{sgn}(x) \min \{|x|, L\}$. We need an additional discrepancy measure between w and $\widetilde{w}$ that depends on the function f being approximated. We define this as the L_w^2 error between $\widetilde{\Pi} f$ and Πf:

$$(5.8) \qquad d(f) \triangleq d(w, \widetilde{w}, P; f) = \left\| \widetilde{\Pi} f - \Pi f \right\|_w = \left\| \widetilde{\Pi} (I - \Pi) f \right\|_w.$$

Note that in all the follows we avoid explicit indication that d depends on w, $\widetilde{w}$, and the polynomial space P to limit notational clutter. Also note that, when $f \notin P$, then the size of $d(f)$, relative to the optimal error $\|f - \Pi f\|_w$, is bounded by the operator norm of $\widetilde{\Pi}$ on the kernel of Π in L_w^2:

$$d(f) \leq \|(I - \Pi)f\|_w \sup_{f \in \ker \Pi} \frac{\left\| \widetilde{\Pi} (I - \Pi) f \right\|_w}{\|(I - \Pi)f\|_w} = \left\| \widetilde{\Pi} \right\|_{P^\perp} \|(I - \Pi)f\|_w.$$

Above, $P^\perp$ is the orthogonal complement of P in L_w^2. It is unclear whether or not the above operator norm of $\widetilde{\Pi}$ can be computed or estimated in general situations.

Following the arguments in [14], we can bound the error for the truncated CLS estimator.

Theorem 5.2. *For a given function f, let Πf be the L_w^2 projection onto a polynomial space P. If the number of samples S in the CLS-bounded algorithm satisfies (5.3), then the mean-square L_w^2 error of the truncated CLS approximation $\widetilde{f}$ defined in (5.7) satisfies*

$$(5.9\text{a}) \quad \mathbb{E}\left[\|f - \widetilde{f}\|_w^2 \right] \leq \|f - \Pi f\|_w^2 + \frac{\varepsilon(S)}{\lambda_{\min}(\mathbf{R})} \|f - \Pi f\|_{\widetilde{w}}^2 + \frac{8L^2}{S^r} + 4\kappa^2(\mathbf{R})d^2(f)$$

with $\varepsilon(S) \triangleq \frac{2-2\log 2}{(1+r)\log S} \to 0$ as $S \to \infty$, and $\kappa(\mathbf{R}) = \lambda_{\max}(\mathbf{R})/\lambda_{\min}(\mathbf{R})$ the 2-norm condition number of $\mathbf{R}$.

Proof. Our proof follows that of Theorem 2 in [14]. Under the sampling condition (5.3) with $\delta = \frac{1}{2}$, we have the following inequality with probability at least $1 - \frac{2}{S^r}$:

(5.10)
$$\lambda_{\max}\left(\mathbf{G}^{-1}\right) \le \lambda_{\max}\left(\mathbf{R}^{-1/2}\right)\lambda_{\max}\left(\mathbf{R}^{1/2}\mathbf{G}^{-1}\mathbf{R}^{1/2}\right)\lambda_{\max}\left(\mathbf{R}^{-1/2}\right) \le \frac{2}{\lambda_{\min}(\mathbf{R})}.$$

We denote the probabilistic set under which this happens as Ω_+, and Ω_- the set under which this fails. Then

$$\mathbb{E}\left[\|f - \widetilde{f}\|_w^2\right] \le \mathbb{E}\left[\|f - \widetilde{\Pi}^S f\|_w^2 \,|\, \Omega_+\right] + \frac{8L^2}{S^r}$$

where we have used the fact that $\|T_L[f]\|_w \le \|f\|_w$, and $T_L[f] = f$ if $|f| \le L$. Let $g = f - \Pi f$. We note that since g is L_w^2-orthogonal to Πf, and $\widetilde{\Pi}^S$ is the identity on P, then

$$f - \widetilde{\Pi}^S f = g - \widetilde{\Pi}^S g,$$

so that

$$\left\|g - \widetilde{\Pi}^S g\right\|_w^2 = \|g\|_w^2 + \left\|\widetilde{\Pi}^S g\right\|_w^2 = \|g\|_w^2 + \|\mathbf{b}\|^2,$$

where the coefficients b_j in the vector $\mathbf{b}$ are the polynomial coefficients recovered from the CLS approximation on the function g. Thus, we have

(5.11a)
$$\mathbb{E}\left[\|f - \widetilde{f}\|_w^2\right] \le \|g\|_w^2 + \frac{8L^2}{S^r} + \mathbb{E}\left[\|\mathbf{b}\|^2 \,|\, \Omega_+\right].$$

We have (5.10), and using the normal equations (2.7) on the event Ω_+ (in (2.7), replace $\mathbf{f} \leftarrow \mathbf{g}$ and $\mathbf{c} \leftarrow \mathbf{b}$), we have

(5.11b)
$$\|\mathbf{b}\|^2 \le \frac{4}{S^2\lambda_{\min}^2(\mathbf{R})}\|\mathbf{V}^T\mathbf{K}\mathbf{g}\|^2 = \frac{4}{\lambda_{\min}^2(\mathbf{R})}\sum_{n=1}^N \left\langle \phi_n(z), \frac{N}{K(z)}g(z)\right\rangle_S^2$$

where $\langle \cdot, \cdot \rangle_S$ is defined in (2.8). Letting z_s denote random iid variables distributed according to the sampling density v, then each summand on the right-hand side above has expected value given by

$$\mathbb{E}\left(\frac{1}{S}\sum_{s=1}^S \frac{N}{K(z_s)}\phi_n(z_s)g(z_s)\right)^2 = \frac{1}{S^2}\sum_{s,r=1}^S \mathbb{E}\frac{N^2}{K(z_s)K(z_r)}\phi_n(z_s)\phi_n(z_r)g(z_s)g(z_r)$$

$$= \frac{S}{S^2}\mathbb{E}\frac{N^2}{K^2(z_1)}\phi_n^2(z_1)g^2(z_1) + \frac{S(S-1)}{S^2}\left[\mathbb{E}\frac{N}{K(z_1)}\phi_n(z_1)g(z_1)\right]^2.$$

Summing over n, we have

$$\mathbb{E}\|\mathbf{V}^T\mathbf{K}\mathbf{g}\|^2 = \frac{1}{S}\mathbb{E}\frac{N^2}{K^2(z_1)}K(z_1)g^2(z_1) + \frac{S-1}{S}\sum_{n=1}^N\left[\mathbb{E}\frac{N}{K(z_1)}\phi_n(z_1)g(z_1)\right]^2$$

$$= \underbrace{\frac{N}{S}}_{(a)}\underbrace{\mathbb{E}\frac{N}{K(z_1)}g^2(z_1)}_{(b)} + \frac{S-1}{S}\underbrace{\sum_{n=1}^N\left[\int_D \phi_n(z)g(z)\widetilde{w}\mathrm{d}z\right]^2}_{(c)}.$$

Term (a) on the right-hand side can be bounded by using condition (5.3), so that (a) $= \frac{N}{S} \leq \frac{\lambda_{\min}(\mathbf{R})c_\delta}{(1+r)\log S}$. Term (b) is equal to $\|g\|_{\widetilde{w}}^2$. To bound (c), consider the continuous projection $\widetilde{\Pi}g = \sum_{n=1}^N d_n\phi_n$, with the d_n solving the expected value of the normal equations (2.7):

$$\mathbf{Rd} = \mathbf{h}, \qquad\qquad (h)_n = \int_D g(z)\phi_n(z)\widetilde{w}(z)\mathrm{d}z.$$

We have

$$(\mathrm{c}) = \sum_{n=1}^N \left[\int_D \phi_n(z)g(z)\widetilde{w}\mathrm{d}z\right]^2 = \|\mathbf{h}\|^2 \leq \|\|\mathbf{R}\|\|^2 \|\mathbf{d}\|^2 \leq \lambda_{\max}^2(\mathbf{R})\|\widetilde{\Pi}g\|_w^2.$$

Thus we have

(5.11c) $$\mathbb{E}\|\mathbf{V}^T\mathbf{Kg}\|^2 \leq \lambda_{\min}(\mathbf{R})\frac{c_\delta}{(1+r)\log\delta}\|g\|_{\widetilde{w}}^2 + \lambda_{\max}^2(\mathbf{R})\|\widetilde{\Pi}g\|_w^2.$$

Combining (5.11a), (5.11b), and (5.11c), we have

$$\mathbb{E}\left[\|f-\widetilde{f}\|_w^2\right] \leq \|g\|_w^2 + \frac{4c_\delta}{(1+r)\lambda_{\min}(\mathbf{R})\log S}\|g\|_{\widetilde{w}}^2 + \frac{8L^2}{S^r} + 4\kappa^2(\mathbf{R})\|\widetilde{\Pi}g\|_w^2,$$

where the appropriate value of c_δ from (4.1) for $\delta = \frac{1}{2}$ should be used. Noting that the term $\left\|\widetilde{\Pi}g\right\|_w^2$ is precisely $d^2(f)$ proves the result. $\qquad\square$

The above result is suboptimal in the sense that as $S \to \infty$, the error converges to a value that deviates from the optimal value of $\|g\|_w^2$. The suboptimal term involving $d^2(f)$ is a term that stems from the discrepancy between w and $\widetilde{w}$. Indeed, $d(f) = \widetilde{\Pi}(I-\Pi)f$, which vanishes when w and $\widetilde{w}$ coincide (or when f is a polynomial in P regardless of $\widetilde{w}$). A second apparently deviation of the CLS result above from the standard Monte Carlo estimate (4.3) is that the second term on the right-hand side of (5.9a) is proportional to $\|f-\Pi f\|_{\widetilde{w}}$ rather than $\|f-\Pi f\|_w$.

5.1.1. *Size of the CLS discrepancy terms.* In this section we give empirical evidence to suggest that the standard Monte Carlo error estimate from Theorem 4.1 and the CLS error estimate derived in Theorem 5.2 give comparable bounds. The different error terms in the standard Monte Carlo convergence result (4.3) and the CLS convergence result (5.9a) are, respectively,

$$\varepsilon(S)\|g\|_w^2 \quad \text{versus} \quad \frac{\varepsilon(S)}{\lambda_{\min}(\mathbf{R})}\|g\|_{\widetilde{w}}^2 + 4\kappa^2(\mathbf{R})d^2(f).$$

Above, $g = f - \Pi f$ is the truncation error between f and its L_w^2 projection onto the polynomial space P. To investigate how suboptimal the above terms from CLS algorithm are, we separate dependence on the sample count S by assuming that $\varepsilon(S) = 1$. In this case, a measure of suboptimality of the CLS theory is given by the ratio of the above terms:

(5.12) $$\Delta(f) \triangleq \frac{1}{\lambda_{\min}(\mathbf{R})}\frac{\|g\|_{\widetilde{w}}^2}{\|g\|_w^2} + 4\kappa^2(\mathbf{R})\frac{d^2(f)}{\|g\|_w^2},$$

with $\Delta \approx 1$ indicating that the CLS and standard Monte Carlo error terms are of roughly the same magnitude. Values satisfying $\Delta \gg 1$ indicate that the CLS convergence terms are quite suboptimal. We consider the case $D = [-1, 1]$ with the weight $w = 1$, so that the ϕ_n basis elements are orthonormal Legendre polynomials.
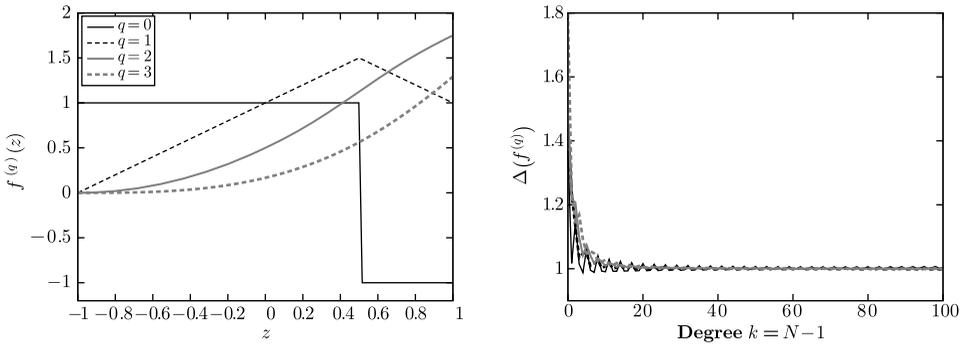
FIGURE 4. Left: 4 test functions defined by (5.13) whose smooth-
ness increases with q. Right: Size of the theoretical CLS bound
(5.9a) relative to the size of the standard Monte Carlo bound (4.3),
measured by the parameter Δ defined in (5.12). Values $\Delta \approx 1$ in-
dicate that the CLS bound is comparable to the standard Monte
Carlo bound.

We choose four test functions $f^{(q)}$ for $q = 0, 1, 2, 3$, where q is an indicator of the
smoothness of these functions:

$$f^{(0)}(z) = \begin{cases} 1, & z \le \frac{1}{2}, \\ -1, & z > \frac{1}{2}, \end{cases}$$

(5.13)
$$f^{(q+1)}(z) = \int_{-1}^{z} f^{(q)}(x)\mathrm{d}x.$$

These functions have q derivatives in L_w^2 and have $\mathcal{O}(1)$ values on $[-1, 1]$. In
Figure 4 we show values of Δ for these four test functions, and see that $\Delta \approx 1$ in
most scenarios, with $\Delta < 2$ in all cases tested. This suggest that the CLS error
bound provided by (5.9a) is, in practice, as sharp as that provided by (4.3), at
least for our choice of w and $\widetilde{w}$. Similar tests in $d = 1$ dimension yield similar
results. Thus, while the CLS error bound in Theorem 5.2 initially appears weaker
than the standard Monte Carlo estimate in 4.1, our limited testing indicates that
they are comparable. We close this section by noting that the first term in (5.12)
involving $\|g\|_{\widetilde{w}}/\|g\|_w$ is the major contributer to Δ, having value approximately
1 in our tests. This is consistent with expectations since $\widetilde{w} \approx w$. Therefore, the
second term involving $d^2(f)$ appeared to have comparatively little contribution to
the value of Δ.

5.2. **Unbounded domains.** We assume the pair (D, w) is unbounded-admissible
in the sense of Section 2. Our results are essentially identical to the bounded domain
case, but some of the definitions change. The CLS method in this case is given by
Algorithm 3, with the sampling density v given by (3.5).

   The Christoffel-weighted functions ψ_α are identical to the bounded case given
by (5.1). The surrogate weight $\widetilde{w}$ is defined as

$$\widetilde{w}(z) = v(z)\frac{N}{K(z)} = k^{-d/t}\frac{N}{K(z)}v_{D,Q}\left(k^{-1/t}z\right)$$

where $k = \max_{\alpha \in \Lambda} |\alpha|$. We let $A \triangleq \operatorname{supp} \mu_{D,Q}$ be the support of the equilibrium measure, and the analogous definition of (5.2) in the unbounded case is

(5.14)

$$(R)_{\alpha,\beta} = k^{-d/t} \int_{k^{1/t}A} \psi_\alpha(z)\psi_\beta(z)\mathrm{d}\mu_{D,Q}(k^{-1/t}z) = \int_{k^{1/t}A} \phi_\alpha(z)\phi_\beta(z)\widetilde{w}(z)\mathrm{d}z.$$

Before continuing, we need a result that relates polynomials orthonormal under w^k to those orthonormal under w, assuming $w = \exp(-2Q)$ with Q a homogeneous function.

Lemma 5.1. *Let a weight function $w = \exp(-2Q)$ with Q satisfying (2.1) be given with homogeneity exponent t on an unbounded conic domain D. Let ϕ_α be a polynomial family that is L^2 orthonormal under weight w. Then a family of polynomials $\phi_\alpha^{(k)}$ that is orthonormal under w^k is*

$$\phi_\alpha^{(k)}(z) = k^{d/2t}\phi_\alpha\left(k^{1/t}z\right).$$

Proof. Since ϕ_α is orthonormal under w, then

$$\int_D \phi_\alpha(z)\phi_\beta(z)w(z)\mathrm{d}z = \delta_{\alpha,\beta}.$$

By assumption (2.1), $w^k(z) = w\left(k^{1/t}z\right)$. Then making the substitution $z \leftarrow k^{1/t}z$ in the relation above yields

$$k^{d/t}\int_D \phi_\alpha(k^{1/t}z)\phi_\beta(k^{1/t})w^k(z)\mathrm{d}z = \delta_{\alpha,\beta}.$$

Since $\phi_\alpha(Cz)$ is still a polynomial of degree α for any constant C, this proves the result. $\qquad\square$

Again, the matrix $\mathbf{R}_k$ corresponding to the total-degree polynomial space P_k is reasonably well-behaved with respect to the identity, as can be seen from the right-hand pane of Figure 2, and owing to the result (4.8) from Corollary 4.1, individual entries of $\mathbf{R}_k$ converge to the Kronecker delta.

Proposition 5.2. *Fix multi-indices α and β. Then the entries of $\mathbf{R}_k$ in (5.14) obey*

$$\lim_{k\to\infty}(R_k)_{\alpha,\beta} = \delta_{\alpha,\beta}.$$

Proof. Since $\rho = \sqrt{w}$ is negative-log-homogeneous of degree t, then $\rho^n(z) = \rho(zn^{1/t})$, or in other words, $w^k\left(k^{-1/t}z\right) = w(z)$. Let $\phi_\alpha^{(k)}$ be the polynomial family orthogonal under w^k. By Lemma 5.1, we have

$$k^{-d/t}K_k^{(k)}\left(k^{-1/t}z\right) = k^{-d/t}\sum_{|\alpha|\leq k}\left[\phi_\alpha^{(k)}\left(k^{-1/t}z\right)\right]^2$$

$$= k^{-d/t}\sum_{|\alpha|\leq k}\left[k^{d/2t}\phi_\alpha\left(z\right)\right]^2 = K_k(z).$$

With $\rho^{2k}(z) = w^k(z)$, we can use (4.8) to conclude:

$$\frac{Nv_{D,Q}(k^{-1/t}z)}{K_k^{(k)}\left(k^{-1/t}z\right)w^k\left(k^{-1/t}z\right)} \to 1$$

weakly on compact sets. Now from (5.14) the entries of $\mathbf{R}_k$ are given by

$$
\begin{aligned}
(R_k)_{\alpha,\beta} &= k^{-d/t} \int_{k^{1/t}A} \phi_\alpha(z)\phi_\beta(z) \frac{N}{K_k(z)} v_{D,Q}\left(k^{-1/t}z\right) \mathrm{d}z \\
&= k^{-d/t} \int_{k^{1/t}A} \phi_\alpha(z)\phi_\beta(z) \frac{N v_{D,Q}\left(k^{-1/t}z\right)}{k^{-d/t} K_k^{(k)}\left(k^{-1/t}z\right)} \mathrm{d}z \\
&= \int_{k^{1/t}A} \phi_\alpha(z)\phi_\beta(z) w(z) \frac{N v_{D,Q}\left(k^{-1/t}z\right)}{K_k^{(k)}\left(k^{-1/t}z\right) w^k\left(k^{-1/t}z\right)} \mathrm{d}z \\
&= \int_D \phi_\alpha(z)\phi_\beta(z) w(z) \left[\frac{N v_{D,Q}\left(k^{-1/t}z\right)}{K_k^{(k)}\left(k^{-1/t}z\right) w^k\left(k^{-1/t}z\right)} \mathbb{1}_{k^{1/t}A}(z) \right] \mathrm{d}z.
\end{aligned}
$$

Since the term in brackets weakly converges to 1 on any compact set, we have

$$
(R_k)_{\alpha,\beta} \to \int_D \phi_\alpha\phi_\beta w \,\mathrm{d}z = \delta_{\alpha,\beta}. \qquad \square
$$

It is clear that the results for the bounded case in Theorem 5.1 may be extended to the unbounded case.

Theorem 5.3. *For D a conic unbounded domain with weight decomposition $w = \rho^2$ satisfying (2.1), define $\mathbf{R}_k$ through (5.14). With CLS Algorithm 3 operating under the condition (5.3), then*

$$
\Pr\left[\frac{\|\mathbf{G} - \mathbf{R}_k\|}{\|\mathbf{R}_k\|} \geq \delta \right] \leq \frac{2}{S^r},
$$

for any $\delta \in (0,1)$ and $r > 0$.

This theorem is the unbounded analogue of Theorem 5.1. As before, the minimum eigenvalue of $\mathbf{R}_k$ will play a role in determining the sample count requirement through (5.3).

A convergence result for the unbounded case that mirrors Theorem 5.2 may likewise be proven using the same method.

Theorem 5.4. *Let $f \in L_w^2$ be given. If the number of samples S in the CLS-unbounded algorithm satisfies (5.3), then the mean-square L_w^2 error of the truncated CLS approximation $\widetilde{f} = T_L \widetilde{\Pi}^S f$ satisfies*

$$
(5.15) \quad \mathbb{E}\left[\|f - \widetilde{f}\|_w^2 \right] \leq \|f - \Pi f\|_w^2 + \frac{\varepsilon(S)}{\lambda_{\min}(\mathbf{R})} \|f - \Pi f\|_{\widetilde{w}}^2 + \frac{8L^2}{S^r} + 4\kappa^2(\mathbf{R})d^2(f)
$$

with $\varepsilon(S) \triangleq \frac{2 - 2\log 2}{(1+r)\log S} \to 0$ as $S \to \infty$, $\kappa(\mathbf{R}) = \lambda_{\max}(\mathbf{R})/\lambda_{\min}(\mathbf{R})$ the 2-norm condition number of $\mathbf{R}$, and $d(f)$ as in (5.8).

Just as with the bounded case, this result is influenced by w versus $\widetilde{w}$ discrepancy terms; the empirical observations in Section 5.1.1 regarding the size of these additional terms holds in this case as well.

6. EXAMPLES

In the following section we investigate the stability and convergence properties of the CLS algorithm. The method we compare against will be a standard unweighted Monte Carlo method, Algorithm 1. We are interested primarily in investigating how linear and log-linear sampling rates of S versus the approximation space dimension

N affect stable and accurate reconstruction. In our figures and results, we will use "MC" to denote an unweighted Monte Carlo procedure (i.e., as specific in Algorithm 1), and the notation "CLS" to denote the result of the Christoffel Least Squares algorithm (i.e., either Algorithm 2 or 3).

The sampling strategies we use are from Table 2. Note that for unbounded domains, our sampling strategies are only *conjectures* because explicit formulae for weighted equilibrium measures in these cases are currently unknown.

Sampling from the "Hermite" distribution on $\mathbb{R}^d$. For the "Hermite" case with density $w = \exp(-z^2)$, the following is one way to sample from $v_{D,Q}$ shown in Table 2: let $W \in \mathbb{R}^d$ be a d-variate standard normal random variable. The random variable $\frac{W}{\|W\|_2}$ is uniformly distributed on the surface of the unit ball ∂B^d. Thus, we need only find an appropriate random variable R whose distribution matches the marginal distribution of $\|z\|$. Since marginalizing a spherically symmetric density on the unit ball to the radial coordinate introduces an r^{d-1} factor, then the marginal density for R has the form

$$\rho_R(r) = Cr^{d-1}(2 - r^2)^{d/2}, \qquad\qquad 0 \le r \le \sqrt{2},$$

where C is a normalization constant. However, with the change of variables $R^2 \leftarrow P$, we see that $\frac{1}{\sqrt{2}}P$ has Beta distribution with parameters $\alpha = \frac{d}{2}$ and $\beta = \frac{d}{2} + 1$. Therefore, the following prescription generates samples Z according to the conjectured equilibrium measure:

(1) Generate a d-variate standard normal random variable W.
(2) Generate a $\text{Beta}\left(\frac{d}{2}, \frac{d}{2} + 1\right)$ random variable P.
(3) Set $Z = \sqrt{2P}\frac{W}{\|W\|_2}$.

Sampling from the "Laguerre" distribution on $[0, \infty)^d$. For the "Laguerre" case with density $w = \exp(-\sum_j z_j)$, we need to sample from the appropriate density $v_{D,Q}$ in Table 2. However, we note that the form

$$\frac{\mathrm{d}\mu_{D,Q}}{\mathrm{d}y} = C\left(4 - \sum_{j=1}^{d} y_j\right)^{d/2} \prod_{j=1}^{d} (y_j)^{-1/2}$$

is the density for a $(d + 1)$-dimensional Dirichlet distribution on the variables $(y_1, \ldots, y_d, 4 - \|y\|_{\ell^1})$ with the $d + 1$ parameters $\left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{d}{2} + 1\right)$. Therefore the following prescription generates samples Z according to the conjectured equilibrium measure:

(1) Generate a $(d + 1)$-variate Dirichlet random variable W with parameters $\left(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{d}{2} + 1\right)$.
(2) Truncate the last $((d + 1)$'th) entry of W.
(3) Set $Z = 4W$.

6.1. Matrix stability. In this section we investigate the condition number $\kappa(\sqrt{\mathbf{K}}\mathbf{V}) = \frac{\sigma_{\max}(\sqrt{\mathbf{K}}\mathbf{V})}{\sigma_{\min}(\sqrt{\mathbf{K}}\mathbf{V})}$ of the weighted design matrix $\mathbf{V}$ from both the CLS and the unweighted MC methods, where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values of a matrix, respectively. Because the design matrices for both algorithms are random matrices, we report the *mean* condition number over a size-100 ensemble of tests.
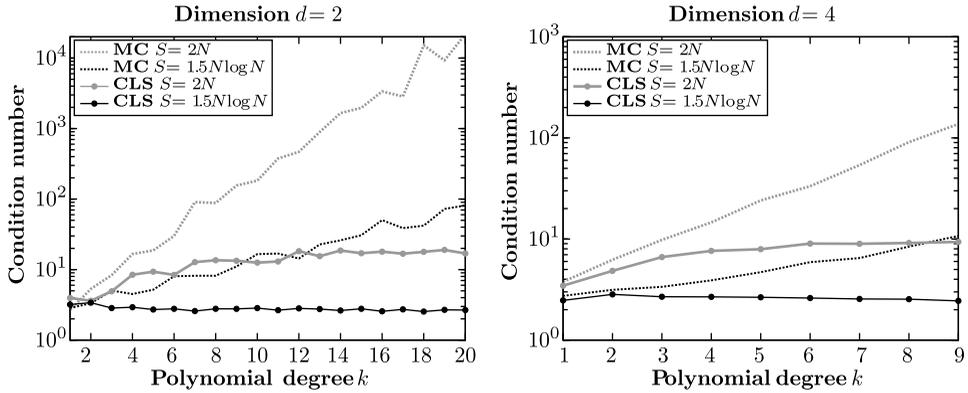
Figure 5. Condition number with respect to the polynomial degree k in the 2-dimensional (left) and 4-dimensional (right) total degree polynomial spaces (uniform measure). Sampling is shown for rates $S = 2N$ and $S = 1.5N \log N$.

6.1.1. *Bounded domains.*

We first consider the uniform distribution where Legendre polynomials are used. In Figure 5 we show the condition number with respect to the polynomial degree k for total degree spaces Λ_k. The left plot shows two-dimensional results while the right plot shows four-dimensional results. Both plots show results for linear scaling of sample count, i.e., $S = 2N$, and for *log-linear* dependence $S = 1.5N \log N$. The CLS algorithm is much more stable compared to the standard MC method. Moreover, the log-linear scaling admits decay properties of the condition number with respect to the polynomial order k with the CLS sampling strategy. In contrast, the linear rule admits a growth of the condition number with respect to the polynomial order k, for both kinds of the design points.

6.1.2. *Unbounded domains.*

For stability on unbounded domains, we will consider the Gaussian density function $\exp(-\sum z_j^2)$ corresponding to Hermite polynomials, and an exponential density function $\exp(-\sum z_j)$ corresponding to Laguerre polynomials. In Figure 6, we report the condition number of the design matrix with respect to the polynomial degree k in both the 2-dimensional total degree space and the 4-dimensional total degree space with log-linear scaling $S = N \log N$. The left-hand figure shows the Hermite results, and the right-hand figure shows the Laguerre results. Again, our approach works much better, but we see that the increased dimensionality of the problem makes the CLS algorithm more ill-conditioned in the Gaussian case.

In Figure 7, we test how the dimension d affects the condition number for the CLS algorithm. In the left plot, we report the numerical condition number for Legendre approach with $S = N \log N$ for $d = 2, 4, 6$. The dimension has little effect on the condition number, and the approach remains stable with the same dependence. In the right plot, we provide results for Laguerre polynomials. For this unbounded case, the dimension parameter d appears to affect stability only weakly, just as with the bounded (Legendre) case.
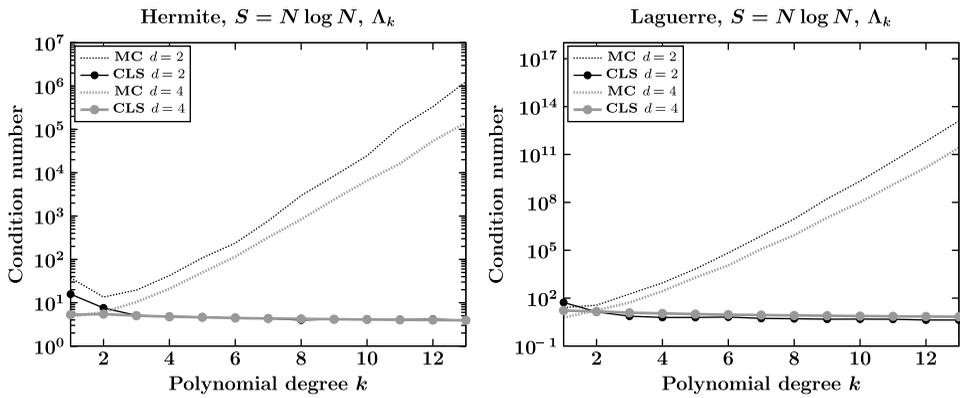
FIGURE 6. Condition number against polynomial degree n for the total degree polynomial space. Left: Gaussian density (Hermite polynomials). Right: Exponential density (Laguerre polynomials).
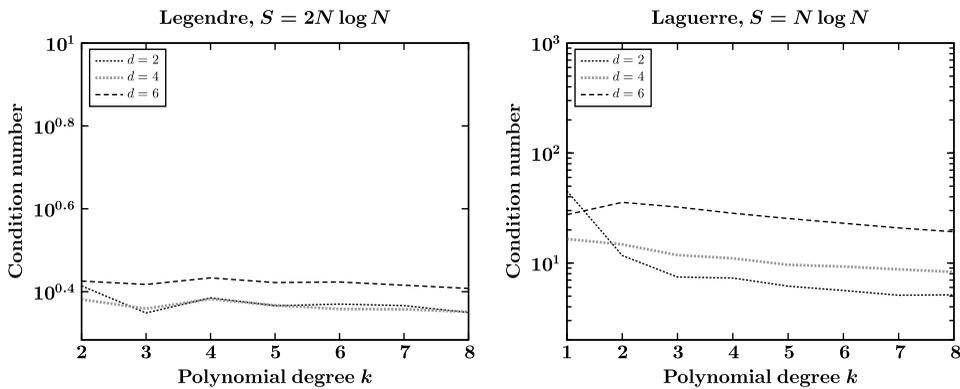

FIGURE 7. CLS Condition numbers against polynomial degree k for different dimensions. Left: Uniform density (Legendre polynomials), $S = 2N \log N$. Right: Exponential density (Laguerre polynomials), $S = N \log N$.

6.1.3. ℓ^p polynomial spaces.

Up until this point we have only provided numerical examples using total-degree polynomial spaces $\Lambda_k = \{\phi_\alpha : |\alpha| \le k\}$. In the following we will consider the effect of using polynomial spaces whose indices are defined by ℓ^p contour lines (with $p < 1$) on the stability of the design matrix. We define the polynomial space of strength $p > 0$ as $\Lambda_k^p = \{\phi_\alpha : \|\alpha\|_p \le k\}$, where $\|\cdot\|_p$ is the discrete ℓ^p norm, and setting $p = 1$ reverts to a total-degree space.

Figure 8 plots the condition number of the design matrices against polynomial degree n for 10-dimensional total-degree and ℓ^p ($p = 2/5$) Laguerre polynomial spaces. In 10 dimensions the CLS algorithm produces larger condition numbers than MC for a given total-degree space for low polynomial degree. This is in contrast to the lower dimensional results shown in Figure 7. However, CLS is again more stable than MC when we use the space $\Lambda_k^{2/5}$. The cardinality of these
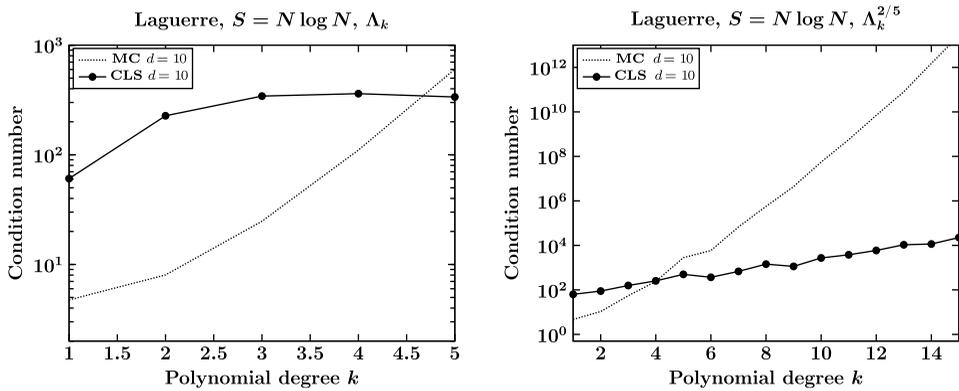
FIGURE 8. Condition number against polynomial degree k for total-degree Λ_k (left) and ℓ^p polynomial space $\Lambda_k^{2/5}$ (right) 10-dimensional Laguerre polynomial spaces

spaces grows much slower than the cardinality of the total-degree spaces. This slower growth allows us to provide numerical results that consider a much larger range of degrees which are computationally unfeasible using total-degree spaces. Since the benefit of the Christoffel function is asymptotic in the degree, we believe that the inclusion of terms that are high-degree in one variable and low-degree in the others (as is the case for these ℓ^p spaces) results in better performance of the CLS algorithm. We also note that the right-hand window of Figure 7 is the more practical case in high-dimensional approximation: using ℓ^p-type index sets.

6.2. **Least-squares accuracy.** In this section we will compare the CLS and MC algorithms in terms of their ability to approximate a number of test functions. In all examples that follow we report the *mean* condition number over a size-20 ensemble of tests.

6.2.1. *Algebraic function.* In Figure 9 (left), we report the convergence rate of the least-squares projection for Legendre approximation in the 2-dimensional total degree space, for the test function $f(z) = \exp\left(-\sum_{i=1}^{d} z_i^2\right)$. We measure accuracy using the discrete ℓ_2 norm which is computed using $10,000$ random samples drawn from the probability measure of orthogonality. The CLS algorithm is very stable and the error in the approximation can be driven to machine accuracy. In contrast the MC strategy becomes unstable as the polynomial degree is increased. Furthermore, MC sampling requires more samples to achieve a given error tolerance. In Figure 9 (right), we consider the Hermite approximation for the test function $f(z) = \exp\left(-\sum_{i=1}^{d} z_i\right)$, in the 3-dimensional total degree space. Again, our approach remains stable, while the MC sampling strategy becomes unstable as the polynomial degree is increased. However, in this case the CLS estimator has noticeably worse accuracy.
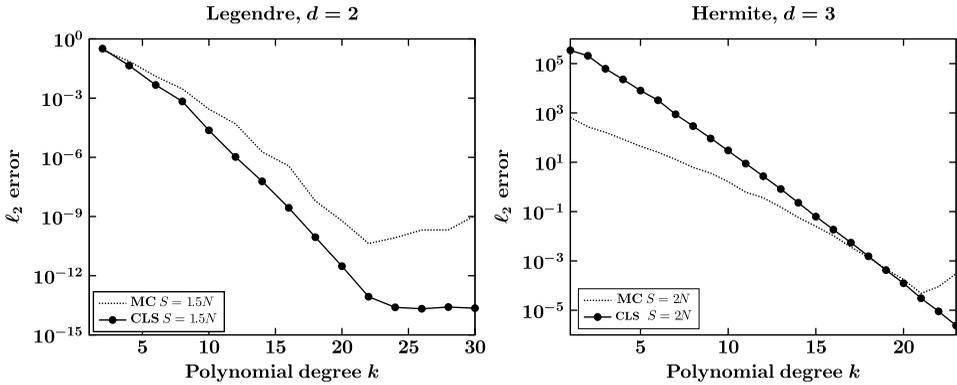
FIGURE 9. Approximation error against polynomial degree k. Left: Legendre approximation of $f(Y) = \exp\left(-\sum_{i=1}^{d} Y_i^2\right)$. Right: Hermite approximation of $f(Y) = \exp\left(-\sum_{i=1}^{d} Y_i\right)$.

6.2.2. *Diffusion equation.* Consider the heterogeneous diffusion equation in one-spatial dimension

$$(6.1) \quad -\frac{d}{dx}\left[a(x,\mathbf{z})\frac{du}{dx}(x,\mathbf{z})\right] = 1, \quad (x,\mathbf{z}) \in (0,1) \times I_{\mathbf{z}}, \quad u(0,\mathbf{z}) = 0, \quad u(1,\mathbf{z}) = 0$$

with an uncertain diffusivity coefficient that satisfies

$$(6.2) \quad a(x,\mathbf{z}) = \bar{a} + \sigma_a \sum_{k=1}^{d} \sqrt{\lambda_k}\phi_k(x)z_k,$$

where $\{\lambda_k\}_{k=1}^{d}$ and $\{\phi_k(x)\}_{k=1}^{d}$ are, respectively, the eigenvalues and eigenfunctions of the squared exponential covariance kernel

$$C_a(x_1, x_2) = \exp\left[-\frac{(x_1 - x_2)^2}{l_c^2}\right].$$

In the following we set $d = 2$, $\bar{a} = 1.0$, $\sigma = 0.1$, $l_c = 1$ and approximate the solution $u(1/3, \mathbf{z})$ when $\mathbf{z} = (z_1, z_2)$ are independent and normally-distributed random variables. [3]

Figure 10 (left) compares the convergence accuracy of the Hermite polynomial least squares projection of the quantity of interest $u(1/3, \mathbf{z})$ using the CLS and MC algorithms. The accuracy of the approximation obtained using CLS is stable, whereas the MC based approximation becomes unstable as the polynomial degree is increased.

6.2.3. *Resistor network.* Consider the electrical resistor network shown in Figure 11. The network is comprised of $d = 2P = 6$ resistances R_j of uncertain Ohmage and the network is driven by a voltage source providing a known potential

---

[3]We solve the model (6.1) using quadratic finite elements with a high enough spatial resolution to neglect discretization errors in our analysis.
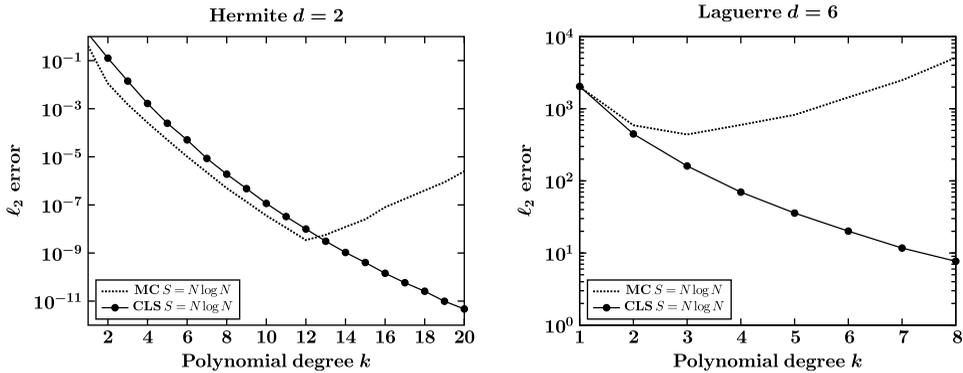
FIGURE 10. Approximation error against polynomial degree k. Left: Hermite approximation of the 2-dimensional diffusion equation. Right: Laguerre approximation of the 6-dimensional resistor network.
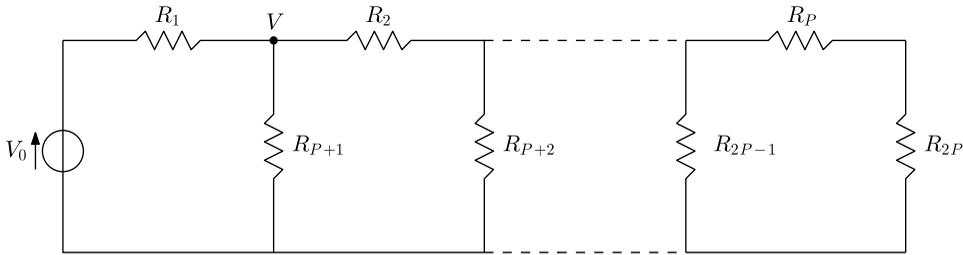

FIGURE 11. Resistor network

$V_0 = 1$. We are interested in using Laguerre polynomials to construct a least-squares approximation of the voltage V when the resistances are independent and identically distributed exponential random variables. As shown in all the previous examples the approximation obtained using CLS is stable for the ranges of degrees considered, whereas the MC based approximation becomes unstable as the polynomial degree is increased.

7. CONCLUSION

Monte Carlo approximation for discrete least-squares polynomial approximation is an effective tool for approximating high-dimensional functions, and of great interest is the number of samples required for stability and convergence. We have shown that the Christoffel Least Squares algorithm can effectively approximate functions on bounded and unbounded multivariate domains, with very general multi-index sets that define the approximation space. Our theoretical results suggest that the CLS algorithm is optimal when the polynomial degree is large; our numerical results validate that the method is either superior to or competitive with standard Monte Carlo techniques in many situations of interest.

We expect it is possible to improve several of the statements about convergence using more precise estimates of Christoffel functions, which is the subject of ongoing work.

Acknowledgments

References

[1] M. A. Alan, *Supports of weighted equilibrium measures and examples*, Potential Anal. **38** (2013), no. 2, 457–470, DOI 10.1007/s11118-012-9281-1. MR3015359

[2] V. Barthelmann, E. Novak, and K. Ritter, *High dimensional polynomial interpolation on sparse grids*, Adv. Comput. Math. **12** (2000), no. 4, 273–288, DOI 10.1023/A:1018977404843. Multivariate polynomial interpolation. MR1768951

[3] E. Bedford and B. A. Taylor, *The complex equilibrium measure of a symmetric convex set in* $\mathbf{R}^n$, Trans. Amer. Math. Soc. **294** (1986), no. 2, 705–717, DOI 10.2307/2000210. MR825731

[4] R. Berman, S. Boucksom, and D. Nyström, *Fekete points and convergence towards equilibrium measures on complex manifolds*, Acta Mathematica **207** (2011), no. 1, 1–27.

[5] R. J. Berman, *Bergman kernels and equilibrium measures for line bundles over projective manifolds*, Amer. J. Math. **131** (2009), no. 5, 1485–1524, DOI 10.1353/ajm.0.0077. MR2559862

[6] R. J. Berman, *Bergman kernels for weighted polynomials and weighted equilibrium measures of* $\mathbb{C}^n$, Indiana Univ. Math. J. **58** (2009), no. 4, 1921–1946, DOI 10.1512/iumj.2009.58.3644. MR2542983

[7] T. Bloom, *Weighted polynomials and weighted pluripotential theory*, Trans. Amer. Math. Soc. **361** (2009), no. 4, 2163–2179, DOI 10.1090/S0002-9947-08-04607-2. MR2465832

[8] T. Bloom, L. Bos, N. Levenberg, and S. Waldron, *On the convergence of optimal measures*, Constr. Approx. **32** (2010), no. 1, 159–179, DOI 10.1007/s00365-009-9078-7. MR2659752

[9] T. Bloom and N. Levenberg, *Asymptotics for Christoffel functions of planar measures*, J. Anal. Math. **106** (2008), 353–371, DOI 10.1007/s11854-008-0052-2. MR2448990

[10] L. Bos, *Asymptotics for the Christoffel function for Jacobi like weights on a ball in* $\mathbf{R}^m$, New Zealand J. Math. **23** (1994), no. 2, 99–109. MR1313445

[11] L. Bos, B. Della Vecchia, and G. Mastroianni, *On the asymptotics of Christoffel functions for centrally symmetric weight functions on the ball in* $\mathbf{R}^d$, Proceedings of the Third International Conference on Functional Analysis and Approximation Theory, Vol. I (Acquafredda di Maratea, 1996), Rend. Circ. Mat. Palermo (2) Suppl. **52, Vol. I** (1998), 277–290. MR1644555

[12] D. Burns, N. Levenberg, S. Ma'u, and Sz. Révész, *Monge-Ampère measures for convex bodies and Bernstein-Markov type inequalities*, Trans. Amer. Math. Soc. **362** (2010), no. 12, 6325–6340, DOI 10.1090/S0002-9947-2010-04892-5. MR2678976

[13] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, and R. Tempone, *Discrete least squares polynomial approximation with random evaluations—application to parametric and stochastic elliptic PDEs*, ESAIM Math. Model. Numer. Anal. **49** (2015), no. 3, 815–837, DOI 10.1051/m2an/2014050. MR3342229

[14] A. Cohen, M. A. Davenport, and D. Leviatan, *On the stability and accuracy of least squares approximations*, Found. Comput. Math. **13** (2013), no. 5, 819–834, DOI 10.1007/s10208-013-9142-3. MR3105946

[15] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of PDEs with stochastic inputs*, J. Comput. Phys. **230** (2011), no. 8, 3015–3034, DOI 10.1016/j.jcp.2011.01.002. MR2774328

[16] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann, *On the convergence of generalized polynomial chaos expansions*, ESAIM Math. Model. Numer. Anal. **46** (2012), no. 2, 317–339, DOI 10.1051/m2an/2011045. MR2855645

[17] J. Hampton and A. Doostan, *Coherence motivated sampling and convergence analysis of least-squares polynomial chaos regression*, arXiv: 1410.1931, 2014.

[18] M. Klimeck, *Pluripotential Theory*, Oxford University Press, Oxford, 1991.

[19] A. Kroó and D. S. Lubinsky, *Christoffel functions and universality in the bulk for multivariate orthogonal polynomials*, Canad. J. Math. **65** (2013), no. 3, 600–620, DOI 10.4153/CJM-2012-016-x. MR3043043

[20] D. S. Lubinsky, *A survey of weighted polynomial approximation with exponential weights*, Surv. Approx. Theory **3** (2007), 1–105. MR2276420

[21] G. Migliorati, *Multivariate Markov-type and Nikolskii-type inequalities for polynomials associated with downward closed multi-index sets*, J. Approx. Theory **189** (2015), 137–159, DOI 10.1016/j.jat.2014.10.010. MR3280676

[22] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, *Analysis of discrete L^2 projection on polynomial spaces with random evaluations*, Found. Comput. Math. **14** (2014), no. 3, 419–456, DOI 10.1007/s10208-013-9186-4. MR3201952

[23] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, *Approximation of quantities of interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces*, SIAM J. Sci. Comput. **35** (2013), no. 3, A1440–A1460, DOI 10.1137/120897109. MR3061475

[24] A. Narayan and D. Xiu, *Stochastic collocation methods on unstructured grids in high dimensions via interpolation*, SIAM J. Sci. Comput. **34** (2012), no. 3, A1729–A1752, DOI 10.1137/110854059. MR2970271

[25] A. Narayan and T. Zhou, *Stochastic collocation on unstructured multivariate meshes*, Commun. Comput. Phys. **18** (2015), no. 1, 1–36, DOI 10.4208/cicp.020215.070515a. MR3371550

[26] P. Nevai, *Géza Freud, orthogonal polynomials and Christoffel functions. A case study*, J. Approx. Theory **48** (1986), no. 1, 167, DOI 10.1016/0021-9045(86)90016-X. MR862231

[27] H. Rauhut and R. Ward, *Sparse Legendre expansions via ℓ_1-minimization*, J. Approx. Theory **164** (2012), no. 5, 517–533, DOI 10.1016/j.jat.2012.01.008. MR2903116

[28] E. B. Saff and V. Totik, *Logarithmic Potentials with External Fields*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 316, Springer-Verlag, Berlin, 1997. Appendix B by Thomas Bloom. MR1485778

[29] T. Tang and T. Zhou, *On discrete least-squares projection in unbounded domain with random evaluations and its application to parametric uncertainty quantification*, SIAM Journal on Scientific Computing (2014), A2272–A2295.

[30] V. Totik, *Asymptotics for Christoffel functions for general measures on the real line*, J. Anal. Math. **81** (2000), 283–303, DOI 10.1007/BF02788993. MR1785285

[31] V. Totik, *Asymptotics for Christoffel functions with varying weights*, Adv. in Appl. Math. **25** (2000), no. 4, 322–351, DOI 10.1006/aama.2000.0698. MR1788824

[32] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434, DOI 10.1007/s10208-011-9099-z. MR2946459

[33] N. Wiener, *The Homogeneous Chaos*, Amer. J. Math. **60** (1938), no. 4, 897–936, DOI 10.2307/2371268. MR1507356

[34] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, Princeton, NJ, 2010. MR2723020

[35] D. Xiu and J. S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput. **27** (2005), no. 3, 1118–1139 (electronic), DOI 10.1137/040615201. MR2199923

[36] D. Xiu and G. E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. **24** (2002), no. 2, 619–644 (electronic), DOI 10.1137/S1064827501387826. MR1951058

[37] Y. Xu, *Christoffel functions and Fourier series for multivariate orthogonal polynomials*, J. Approx. Theory **82** (1995), no. 2, 205–239, DOI 10.1006/jath.1995.1075. MR1343834

[38] Y. Xu, *Asymptotics of the Christoffel functions on a simplex in $\mathbf{R}^d$*, J. Approx. Theory **99** (1999), no. 1, 122–133, DOI 10.1006/jath.1998.3312. MR1696565

[39] T. Zhou, A. Narayan, and D. Xiu, *Weighted discrete least-squares polynomial approximation using randomized quadratures*, J. Comput. Phys. **298** (2015), 787–800, DOI 10.1016/j.jcp.2015.06.042. MR3374578

[40] T. Zhou, A. Narayan, and Z. Xu, *Multivariate discrete least-squares approximations with a new type of collocation grid*, SIAM Journal on Scientific Computing (2014), A2401–A2422.

DEPARTMENT OF MATHEMATICS AND SCIENTIFIC COMPUTING AND IMAGING (SCI) INSTITUTE, UNIVERSITY OF UTAH, UNIVERSITY OF UTAH, SALT LAKE CITY, UTAH 84112
  E-mail address: `akil@sci.utah.edu`

COMPUTER SCIENCE RESEARCH INSTITUTE, SANDIA NATIONAL LABORATORIES, 1450 INNOVATION PARKWAY, SE, ALBUQUERQUE, NEW MEXICO 87123
  E-mail address: `jdjakem@sandia.gov`

LSEC, INSTITUTE OF COMPUTATIONAL MATHEMATICS, ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE, CHINESE ACADEMY OF SCIENCES, BEIJING 100190, PEOPLE'S REPUBLIC OF CHINA
  E-mail address: `tzhou@lsec.cc.ac.cn`