ADAPTIVE LEJA SPARSE GRID CONSTRUCTIONS FOR STOCHASTIC COLLOCATION AND HIGH-DIMENSIONAL APPROXIMATION

AKIL NARAYAN AND JOHN D. JAKEMAN

ABSTRACT. We propose an adaptive sparse grid stochastic collocation approach based upon Leja interpolation sequences for approximation of parameterized functions with high-dimensional parameters. Leja sequences are arbitrarily granular (any number of nodes may be added to a current sequence, producing a new sequence) and thus are a good choice for the univariate composite rule used to construct adaptive sparse grids in high dimensions. When undertaking stochastic collocation one is often interested in constructing weighted approximation where the weights are determined by the probability densities of the random variables. This paper establishes that a certain weighted formulation of one-dimensional Leja sequences produces a sequence of nodes whose empirical distribution converges to the corresponding limiting distribution of the Gauss quadrature nodes associated with the weight function. This property is true even for unbounded domains. We apply the Leja-sparse grid approach to several high-dimensional and problems and demonstrate that Leja sequences are often superior to more standard sparse grid constructions (e.g. Clenshaw-Curtis), at least for interpolatory metrics.

1. INTRODUCTION

Stochastic collocation (SC) has become a standard tool for non-intrusively quantifying uncertainty in simulation models that are subject to a degree of uncertainty or randomness. Sources of uncertainty can include physical stochastic processes, parametric uncertainty or model form uncertainty [54, 40, 39]. If the sources of uncertainty can be parameterized by a set of random variables, the approaches used for uncertainty quantification (UQ) frequently reduce to computational methods that describe the behavior of a model with respect to those random variables.

In this paper we will focus on the use of stochastic collocation methods that utilize sparse grid interpolants [46, 56, 22, 9] to approximate the dependence of model output on unknown model parameters. Sparse grid stochastic collocation involves constructing an ensemble of random variables realizations, solving a deterministic physical system for each corresponding realization on the sparse grid, and using the resulting model output to build an approximation of the model response to the uncertain parameters. Once constructed, the sparse grid can be evaluated inexpensively to predict the variability of the physical model with respect to the random parameters.

Obtaining the ensemble of model solutions is usually the most expensive part of the collocation procedure since the model under consideration is often very complicated (e.g., complex geometry, multiscale features, stiff time-stepping, etc.). This dominance of model execution time on the computational expense of UQ motivates the need to build approximations of the model response that require as few model runs as possible. Sparse grids attempt to minimize the number or parameter realizations by generating ensembles that are geometrically sparse in the state space of the random variable. Most efficient sparse grid constructions build up the ensemble adaptively by concentrating model evaluations in dimensions of the paramter space where the sparse grid approximation is poor [23, 26, 19].

One of the crucial considerations when building a sparse grid is the choice of ensemble for the random variable. The sparse grid is constructed via the union of judicious tensorizations of one-dimensional ensemble grids, and so the identification of these univariate grids is of paramount importance. Common univariate choices are Clenshaw-Curtis or Chebyshev nodes and Gauss-quadrature-type nodes.

The desired characteristics in choosing a composite univariate rule for input into the sparse grid include:

- efficiency – high interpolation and/or quadrature accuracy with low cardinality sets
- robustness – consistent and increasing accuracy when the grid is refined
- monotonicity – fine-level grids are supersets of coarse-level grids
- granularity – the number of nodes needed for refinement of a grid is as small as possible

Efficiency and robustness are desired when using approximation grids in any context. The monotonicity property is motivated mainly by the high cost of solving deterministic physical simulations and the sparse grid algorithm: if monotonicity holds, then the tensorized sparse grid construction has many fewer total nodes. Granularity becomes important when several levels of refinement are necessary: it is much better to have the ability to add a small number nodes for each refinement step than to be required to, e.g., double the number of nodes.

In this paper we propose use of univariate Leja sequences for use in the sparse grid algorithm for interpolatory high-order and high-dimensional approximation. Leja rules are a sequence of interpolation/quadrature grids in one dimension that are strongly monotonic and granular: coarse grids are always strict subsets of fine grids, and refinement proceeds by adding a single node at a time. Leja sequences are very accurate as we show later, but they are not as efficient as Gauss-type rules in some cases (e.g., quadrature). Therefore, we argue that Leja sequences are a grid choice that serves as a good compromise of the above desired characteristics, in contrast to, e.g. a Gauss quadrature grid that is strong in efficiency and granularity, but very weak in monotonicity.

For many weight functions of interest (i.e. the random variable probability density function), we show that (our definition of) weighted Leja sequences produce a nodal sequence whose asymptotic distribution coincides with the asympototic distribution of Gauss quadrature nodes associated to the polynomial family orthogonal under the weight function. This result is known for the uniform-weight case; to our knowledge the weighted versions are novel. Additionally, we show that a contracted version of weighted Leja sequences are *asymptotically weighted Fekete*, meaning that their Vandermonde determinant grows comparably to the largest possible value (Fekete). That these results hold for unbounded domains is significant as it suggests that Leja sequences will be accurate for nested interpolatory approximation when the random variable state space is infinite.

In Section 2 we setup the problem and introduce notation and terminology. In Section 3 we introduce Leja sequences and formally present the above-mentioned properties. Section 4 develops the methodology for adaptive sparse grids. Finally, Section 5 shows that the Leja sparse grid algorithm produces results comparable to well-established sparse grid approximation methods, and in many cases is superior. The proof of our main result in Section 3 concerning the distribution of Leja sequence nodes is relatively involved, employing results from weighted potential theory; for this reason we leave this until the end in Section 7, serving somewhat as an appendix.

The first half of this paper (Section 3) proves certain results about one-dimensional weighted Leja sequences and compares them to other one-dimensional rules. The second half (Sections 4 and 5) uses these weighted Leja sequences for building adaptive multivariate sparse grids.

2. Setup

A model problem in the UQ community that serves as a motivating example is a parameterized elliptic equation where the parameters $\mathbf{Z}$ are random variables:

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

This model describes the steady-state temperature distribution u in a one-dimensional domain where the domain has diffusivity coefficient a and experiences an external heat source defined by f. Here x is a spatial variable taking values in a one-dimensional interval domain. The variable $\mathbf{Z}$ is a random vector with density function $\omega(\mathbf{Z})$ on domain $I_{\mathbf{z}}$ corresponding to a probability measure P on a complete probability space. The diffusion coefficient a is a model parameter that varies spatially, but is also influenced by uncertainty. Uncertainty in these kinds of models may arise from, e.g.,

imprecise knowledge of material parameters or external forcing. Under the assumption that the equation is well-posed almost surely, the solution $u(x, \mathbf{Z})$ is random, and essentially depends on $(1 + \dim \mathbf{Z})$ variables. One goal in the UQ community is efficient and accurate prediction of $u(x, \mathbf{Z})$ or some quantities of interest that depend on u (e.g. the temperature variance as a function of space x).

One popular technique is the generalized Polynomial Chaos (gPC) approach: we assume that the variation of u with respect to the random parameter $\mathbf{Z}$ can be described accurately by a finite-degree polynomial:

$$u(x, \mathbf{Z}) \simeq u_N(x, \mathbf{Z}) = \sum_{n=1}^{N} \widehat{u}_n(x) \phi_n(\mathbf{Z}),$$

where ϕ_n are polynomials that satisfy an orthogonality condition

$$\mathbb{E}[\phi_n(\mathbf{Z})\phi_m(\mathbf{Z})] = \int_{I_z} \phi_n(\mathbf{z})\phi_m(\mathbf{z})\omega(\mathbf{z})\mathrm{d}\mathbf{z} = \delta_{n,m},$$

and the $\widehat{u}_n$ are coefficient functions that depend only on the spatial variable x. The determination of the functions $\widehat{u}_n$ is the challenge, and one straightforward procedure is to use a probabilistic sampling strategy to compute these coefficients: let $\mathbf{z}_m$ for $1 \leq m \leq M$ be given samples of the variable $\mathbf{Z}$. For each $\mathbf{z}_m$, equation (1) is a deterministic differential equation, and any computational simulation or experimental setup may be used to obtain $u(x, \mathbf{z}_m)$. (Frequently, this solution is a finite-dimensional quantity rather than a function of a continuum variable x, but this distinction does not affect the main focus of our discussion.) Once these realizations of $u(x, \mathbf{Z})$ are collected, then we attempt to find $\widehat{u}_n(x)$ such that

$$\sum_{n=1}^{N} \phi_n(\mathbf{z}_m)\widehat{u}_n(x) \approx u(x, \mathbf{z}_m), \qquad\qquad m = 1, \ldots, M.$$

If one considers u as a scalar, then this is a linear algebra problem, seeking a vector $\widehat{\mathbf{u}}$ that solves:

$$\mathbf{A}\widehat{\mathbf{u}} \approx \mathbf{u}.$$

This problem may be solved by defining $\approx$ in any convenient fashion: interpolation, least-squares regression, quadrature, or compressive sampling. See, e.g. [52, 53, 14, 55]. Usually the particular choice made is dependent on the relationship between M and N (determined by the computational cost of computing each solution realization) and dependent on some *a priori* understanding of the accuracy for the choice.

One major difficulty with this approach is the selection of $\mathbf{z}_m$ when $\dim \mathbf{Z}$ is large. While the spatial variable x is usually restricted to have dimension less than or equal to 3, it is not uncommon to have 100 or more parameters as the components of $\mathbf{Z}$. High-dimensional approximation has been a persistent bottleneck in modern scientific computing. Methods that work very well for a small number of dimensions are rendered ineffective or impossible to implement in a large number of dimensions, owing to the curse of dimensionality: exponential dependence of functional complexity with respect to the parametric dimension (when the functional smoothness is fixed). Tensor product constructions and space-filling designs adopt this complexity with respect to dimension.

Attempting to circumvent the computationally onerous space-filling property is the main reason to consider alternatives such as sparse grids. The sparse grid construction still employs tensor product grids, but it does so in a way that attempts to control the cardinality of the mesh and delays the curse of dimensionality. Sparse grids are formed by (unions of) tensorized one-dimensional grids, and therefore an educated selection of the composite one-dimensional rules is of great importance.

In this paper we consider high-order interpolatory approximation using a sparse grid, and we employ weighted Leja grids as the one-dimensional composite rules. Weighted Leja grids are nested grids (they are a sequence), and we prove that the nodes distribute identically to the one-dimensional Gauss quadrature rules. Thus, Weighted Leja sequences distribute nodes in a way that emulates the Gauss quadrature rule, and have the advantage of being nested.

Gauss-Kronrod [29, 11, 20] and Gauss-Patterson [41] rules are likewise nested interpolatory schemes, but their computation is usually restricted to special weight functions because computation of the nested rules is relatively difficult. In contrast, weighted Leja sequences are simple and very easy to compute (exactly and approximately) even for exotic weight functions.

Our approach with weighted Leja sequences considers polynomial approximation on unbounded domains, but there are alternative high-order approaches. As described in [8], there are three popular approaches to high-order approximation on unbounded domains: (1) expansions on infinite domains using polynomial or non-polynomial complete basis sets [47, 37] (2) domain trunction, where the full domain I_z is replaced by a compact subset [7, 6], and (3) mapping techniques [51, 25, 24] where standard methods on compact intervals are "transplanted" onto an infinite interval via a domain mapping. Each of these methods can perform accurate approximation on unbounded domains depending on the application. Our approach is most closely related to (1), but in principle one may use Leja sequences for any of the above methods. However, this application is outside the scope of this paper.

For the remainder of this paper, we replace the uppercase variable $\mathbf{Z}$ (traditionally denoting a random quantity) with its lowercase counterpart $\mathbf{z}$: stochasticity does not affect our approach so in principle we may treat the random parameter as a deterministic parameter $\mathbf{z}$ with corresponding weight function ω.

3. Weighted Leja points

In this section we present and establish important properties of univariate Leja sequences. Consider approximation in the scalar variable z over the domain I in the presence of a weight function w. In the context of this paper, z represents one component of the vector-valued parameter $\mathbf{z}$, I is the one-dimensional restriction of $I_{\mathbf{z}}$ to the appropriate dimension, and $w(z)$ is the marginal density computed from the joint density $\omega(\mathbf{z})$.

A Leja sequence (unweighted) is classically defined [16, 30] as a sequence of points $z_n \in I = [-1, 1] \subset \mathbb{R}$ for $n = 1, 2, \ldots$, such that

$$(2) \qquad z_{N+1} = \operatorname*{argmax}_{z \in [-1,1]} \prod_{n=0}^{N} |z - z_n|,$$

where the starting point of the sequence z_0 is arbitrarily chosen in $[-1, 1]$. We note that it is only sensible to define the above Leja sequences on bounded domains. We list below the properties of one-dimensional (unweighted) Leja sequences:

- Leja sequences are not unique. The initial point z_0 may be arbitrarily chosen and the objective function being maximized need not have a unique maximizer.
- The Leja construction provides an interpolation *sequence*. Therefore if $\{z_1, \ldots, z_7\}$ are a Leja sequence, then we require only one more point z_8 to construct a higher-order interpolant. This addresses the granularity criterion for grids, and will be useful in minimizing the number of function evaluations necessary for approximation in high dimensions.
- Maximizing the objective function (2) is equivalent to a greedy determinant maximization (e.g., [15]). With $z_1, \ldots, z_{N-1}$ specified, let $V_N(z)$ be the $N \times N$ interpolatory Vandermonde-like matrix for the space Π_{N-1} at the collocation points $z_1, \ldots, z_{N-1}, z$. (The choice of basis for Π_{N-1} does not affect maximization.) Then (2) is equivalent to

$$z_{N+1} = \operatorname*{argmax}_{z \in [-1,1]} |\det V_N(z)|$$

  Thus one can view Leja sequences as a greedy D-optimal experimental design [17].
- The Lebesgue constant for interpolation on a Leja sequence grows subexponentially [48].
- Leja sequences are asymptotically Fekete [4]. (This is implied by the subexponentially growing Lebesgue constant.) Fekete points are those whose Vandermonde determinant is as large as possible. (These are known to be Gauss-Lobatto nodes in one dimension [18].) The
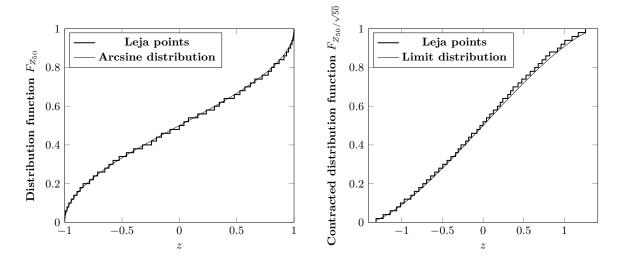
Figure 1. Theoretical asymptotic distribution of Gauss quadrature nodes for the family of polynomials orthogonal under weight ω, and empirical distributions for the first 50 associated Leja points (denoted Z_{50} in both cases). Left: uniform weight $w \equiv 1$ with limiting (arcsine) distribution. Right: Gaussian density weight $w(z) = \exp(-z^2)$ with a 50-point Leja sequence, contracted to the origin by a factor of $\sqrt{50}$, along with the contracted limiting distribution of the Gauss quadrature nodes. (See tables 1 and 2)

asymptotically Fekete property essentially means that the Vandermonde determinant of a Leja sequence grows on par with that of true Fekete nodes.

- Any Leja sequence asymptotically distributes according to the Chebyshev (arcsine) measure. (This is implied by the asymptotically Fekete property [4].)
- In practice, optimization of (2) over a discrete candidate set can be accomplished in computationally efficient ways [1] and with standard numerical linear algebra routines: E.g., the row permutation information from a row-pivoted LU matrix decomposition gives the Leja sequence order, e.g. [5].

The Leja sequences introduced above are quite useful for unweighted polynomial interpolation. However, in the UQ context we are usually interested in interpolation involving a weight function (here, the marginal density of the random variables). Therefore, we are also interested in *weighted* Leja sequences. Let $w(z)$ be a continuous and positive weight function on a univariate domain I, and let

$$(3) \qquad v(z) = \sqrt{w(z)}$$

be its square root. If I is unbounded, we assume that polynomials are dense in the space of continuous functions, measured in the v-weighted supremum norm. For example, if $v(z) \propto \exp(-|z|^\alpha)$, then we require $\alpha \geq 1$, e.g. [31]. This density assumption is necessary in our context: we cannot hope to form an accurate polynomial approximation without polynomial density.

A weighted Leja sequence can be constructed via the optimization

$$(4a) \qquad z_{N+1} = \operatorname*{argmax}_{z \in I} \sqrt{w(z)} \prod_{n=0}^{N} |z - z_n| = \operatorname*{argmax}_{z \in I} v(z) \prod_{n=0}^{N} |z - z_n|.$$

In the case where multiple choices of z maximize the objective, for concreteness we choose the one with smallest magnitude, i.e.,

$$(4b) \qquad z_{N+1} = \operatorname*{argmin}_{z} \left\{ |z| \mid z \in f^{-1}(W) \right\}, \qquad W = \max_{z \in I} f(z) \triangleq \max_{z \in I} v(z) \prod_{n=0}^{N} |z - z_n|.$$
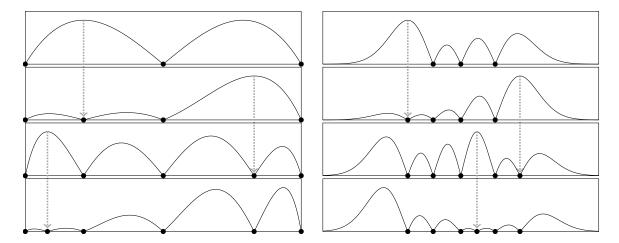
Figure 2. Sequential addition of nodes (top to bottom) via iteration on the Leja objective (4). Left: A Leja sequence on $[-1, 1]$ with $v = w \equiv 1$ (i.e., unweighted). Right: A weighted Leja sequence on $\mathbb{R}$ with $v^2 = w(z) = \exp(-z^2)$.

The sequence of points z_n produced by the above iteration is the central study of this paper, and we refer to this sequence as a weighted Leja sequence, or a w-weighted Leja sequence. This formulation still leaves a benign ambiguity when multiple maximizers differ only in sign. In this paper, Leja sequence optimization is one-dimensional, so in all that follows we optimize via (4) exactly (up to machine accuracy).

In general there is no standard choice of how to incorporate the weight function into a Leja objective; we have chosen $v = \sqrt{w}$, but alternatives have been proposed [15, 45]. Our choice above is motivated by the fact that under this formulation the sequence of points produced has the same asymptotic distribution as w-Gauss quadrature nodes. We illustrate this property now: In Figure 1 we show that the distribution of a univariate 50-point weighted Leja sequence seems to converge to the distribution of the Gauss points associated with the family of polynomial orthogonal under the weight w. Therefore, the objective (4) produces points that are 'approximately' Gauss nodes, with the additional benefit of being nested. This distribution property alone does not guarantee that Leja sequences are useful, but we provide several examples in this paper that suggest that Leja points have utility.

We emphasize that unlike the unweighted case (2), weighted Leja sequences are constructible on unbounded domains given our assumptions. In Figure 2 we show a graphical illustration of the iterative Leja procedure that produces a Leja sequence.

3.1. **Limits of weighted Leja sequences.** A significant concern for the unbounded case with the weighted Leja sequences we have introduced above is that they 'do the right thing', i.e. that they produce a set of nodes that is desirable from the approximation theory point of view.[1]

The purpose of this subsection is to show that the weighted Leja sequences we have proposed (4) produce points whose asymptotic distribution is identical to the Gauss quadrature nodes associated with the corresponding w-orthogonal polynomial family. To be precise, given a classical weight function w, it is known that there is another weight function $\widetilde{v}$ such that the empirical distribution of the w-Gauss quadrature nodes converges to the $\widetilde{v}$-weighted potential equilibrium measure; roughly speaking, $\widetilde{v} \sim \sqrt{w}$, modulo multiplicative polynomial factors. We show that for this same class of weight functions, the empirical measure of the Leja sequence converges to the same equilibrium

---

[1]Alternative propositions for weighted Leja points [45] construct points on a compact subset of the domain. These points are asymptotically optimal if we are interested in approximation with a basis $v^n p_n$. But because we are interested in approximation on an unbounded domain with just vp_n, we require samples to be produced on the entire domain.

measure. On unbounded domains with exponential weights, $\widetilde{v}$ is the square root of w. On bounded domains with Jacobi-type weights, $\widetilde{v}$ is the uniform weight.

A comprehensive discussion of potential theory and equilibrium measures can be found in [45]; here and later in Section 7, which contains proofs, we give a brief account of this topic. Let $\widetilde{v}$ be a weight function on I that we will precisely relate to w later; define $Q \triangleq -\log \widetilde{v}$. If $\widetilde{v}$ is admissible[2] on I, then there is a unique probability measure $\mu_{\widetilde{v}}$ that minimizes a weighted logarithmic energy. This measure $\mu_{\widetilde{v}}$ is the logarithmic potential equilibrium measure of the domain I in the presence of the external field (i.e. weight) Q. (See Section 7 for a more detailed discussion.) When $\widetilde{v}$ is the uniform measure on a compact interval I, then $\mu_{\widetilde{v}}$ is the arcsine, or Chebyshev measure.

For each $N \in \mathbb{N}$, let $\xi_{n,N}$ with $n = 1, \ldots, N$ denote the N zeros of the degree-N polynomial orthogonal under $w(z)$. I.e., $\xi_{n,N}$ are the N-point w-Gauss quadrature nodes. Let z_n denote any sequence of weighted Leja nodes given by (4). The $\xi_{n,N}$ are a triangular array ($n \leq N$) while the z_n are a sequence. We introduce a contraction factor k_n defined in the following theorem; this contraction factor is used to define the empirical (counting) measure for an N-point Gauss ($\xi_{n,N}$) and Leja (z_n) grid, respectively:

$$\nu_N^G = \frac{1}{N} \sum_{n=1}^{N} \delta_{(k_N \xi_{n,N})}, \qquad\qquad \nu_N^L = \frac{1}{N} \sum_{n=1}^{N} \delta_{(k_N z_n)},$$

where δ_z is the Dirac distribution centered at z. Our main result in this section shows that for most classical univariate weight functions of interest, ν_N^G and ν_N^L limit to the same measure.

Theorem 3.1. *Let $w(z)$ be a weight function on I.*

(1) *(Generalized Hermite)*
   Let $w(z) = z^{2\mu} \exp(-|z|^\alpha)$ for any $\alpha \geq 1$, $\mu > -\frac{1}{2}$ on $I = \mathbb{R}$. Define $k_n = n^{-1/\alpha}$ and $\widetilde{v} = \exp\left(-\frac{1}{2}|z|^\alpha\right)$.
(2) *(Laguerre)*
   Let $w(z) = z^s \exp(-|z|)$ for any $s > -1$ on $I = [0, \infty)$. Define $k_n = n^{-1}$ and $\widetilde{v}(z) = \exp\left(-\frac{1}{2}z\right)$.
(3) *(Jacobi)*
   Let $w(z) = (1-z)^\alpha (1+z)^\beta$ for any $\alpha, \beta > -1$ on $I = [-1,1]$. Define $k_n \equiv 1$ and $\widetilde{v}(z) \equiv 1$.

In all of the above cases, we have

$$\lim_{n\to\infty} \nu_n^G = \mu_{\widetilde{v}} = \lim_{n\to\infty} \nu_n^L, \tag{5}$$

where equality holds in the weak sense.

Remark 3.1. The "Hermite" result from the Theorem above that $\lim_n \nu_n^L \to \mu_{\widetilde{v}}$ also holds for $0 < \alpha < 1$. However, since polynomials are not dense for this weight function [31], it is unclear if one should use polynomial approximation in this case.

The above theorem states that the Leja sequence produced by (4) produces a nested mesh whose samples distribute precisely like w-Gaussian quadrature nodes. We emphasize that while this property is promising, it does not guarantee a good approximation: for example, one can generate a grid according to the acrsine measure on $I = [-1,1]$ whose Lebesgue constant does not grow subexponentially [4]. However, an unweighted Leja sequence is known to have subexponentially growing Lebesgue constant (which is 'good enough' in a sense for approximating very smooth functions). In practice, unweighted Leja sequence have logarithmically-growing Lebesgue constant. However, to our knowledge it is presently unknown if weighted Leja sequences have subexponentially growing weighted Lebesgue constant.

---

[2]v must be admissible in the sense of potential theory: (1) it is a non-negative upper semicontinuous function (2) $I \cap \widetilde{v}^{-1}((0, \infty])$ is nonpolar in the sense of potential theory (a polar is 'negligible' and has Lebesgue measure 0), and (3) if I is unbounded, $|z|\widetilde{v}(z) \to 0$ as $|z| \to \infty$

Class	Domain	Parameters	Weight	Contraction k_n	Equilibrium weight
Hermite	$I = \mathbb{R}$	$\alpha \in [1,\infty),\ \mu \in \left(-\frac{1}{2},\infty\right)$	$w(z) = z^{2\mu}\exp(-|z|^\alpha)$	$k_n = n^{-1/\alpha}$	$\widetilde{v}(z) = \exp\left(-\frac{1}{2}|z|^\alpha\right)$
Laguerre	$I = [0,\infty)$	$s \in (-1,\infty)$	$w(z) = z^s\exp(-z)$	$k_n = n^{-1}$	$\widetilde{v}(z) = \exp\left(-\frac{1}{2}z\right)$
Jacobi	$I = [-1,1]$	$\alpha,\beta \in (-1,\infty)$	$w(z) = (1-z)^\alpha(1+z)^\beta$	$k_n \equiv 1$	$\widetilde{v} \equiv 1$

Class	$\operatorname{supp}(\mu_{\widetilde{v}}) = [a,b]$	Distribution $F(t) = \mu_{\widetilde{v}}\left[(-\infty,t]\right]$	Density $f(t) = \frac{\mathrm{d}\mu_{\widetilde{v}}}{\mathrm{d}t}$
Hermite	$-a = b = \left[2^{\alpha-1}B\left(\frac{\alpha}{2},\frac{\alpha}{2}\right)\right]^{1/\alpha}$	See Table 2	$f(t) = \frac{\alpha}{\pi b^\alpha}\int_{|t|}^b \frac{u^{\alpha-1}}{\sqrt{u^2-t^2}}\mathrm{d}u$
Laguerre	$a = 0,\ b = 4$	$F(t) = \frac{2}{\pi}\arcsin\left(\frac{\sqrt{t}}{2}\right) + \frac{1}{2\pi}\sqrt{t(4-t)}$	$f(t) = \frac{1}{2\pi}\sqrt{\frac{4-t}{t}}$
Jacobi	$a = -1,\ b = 1$	$F(t) = \frac{1}{2} + \frac{1}{\pi}\arcsin t$	$f(t) = \frac{1}{\pi}\frac{1}{\sqrt{1-t^2}}$

Table 1. Asymptotic distributions and densities to which contracted weighted Leja sequences (and Gauss quadrature nodes) converge. Given a weighted Leja sequence z_n, the formulae to which the quantities in this table correspond is given by Theorem 3.1.

			Distribution $F(t) = \mu_{\widetilde{v}}\left[(-\infty,t]\right]$	Density $f(t) = \frac{\mathrm{d}\mu_{\widetilde{v}}}{\mathrm{d}t}$
			$t \in [-b(\alpha), b(\alpha)]$	$t \in [-b(\alpha), b(\alpha)]$
	$\alpha = 1$	$b = \pi$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + tf(t)$	$\frac{1}{\pi^2}\log\left[\frac{b+\sqrt{b^2-t^2}}{|t|}\right]$
	$\alpha = 2$	$b = \sqrt{2}$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + \frac{t}{2}f(t)$	$\frac{1}{\pi}\sqrt{b^2-t^2}$
$w(z) = \exp(-|z|^\alpha)$	$\alpha = 3$	$b = \left(\frac{\pi}{2}\right)^{1/3}$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + \frac{t}{3}f(t)$	$\frac{3}{\pi^2}\left[b\sqrt{b^2-t^2} + t^2\log\left[\frac{b+\sqrt{b^2-t^2}}{|t|}\right]\right]$
$b(\alpha) = \left[2^{\alpha-1}B\left(\frac{\alpha}{2},\frac{\alpha}{2}\right)\right]^{1/\alpha}$	$\alpha = 4$	$b = \left(\frac{4}{3}\right)^{1/4}$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + \frac{t}{4}f(t)$	$\frac{1}{\pi}\sqrt{b^2-t^2}\left(2t^2+b^2\right)$
	$\alpha = 5$	$b = \left(\frac{3\pi}{8}\right)^{1/5}$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + \frac{t}{5}f(t)$	$\frac{5}{3\pi^2}\left[b\left(3t^2+2b^2\right)\sqrt{b^2-t^2} + 3t^4\log\left[\frac{b+\sqrt{b^2-t^2}}{|t|}\right]\right]$
	$\alpha = 6$	$b = \left(\frac{16}{15}\right)^{1/6}$	$\frac{1}{2} + \frac{1}{\pi}\arcsin\left(\frac{t}{b}\right) + \frac{t}{6}f(t)$	$\frac{3}{8\pi}\sqrt{b^2-t^2}\left(8t^4+4b^2t^2+3b^4\right)$

Table 2. Explicit formulae for special cases of Hermite/exponential-type asymptotic distributions [50]. See Table 1. The distribution function $F(t)$ for $\alpha = 2$ is plotted in the right-hand pane of Figure 1. The densities $f(t)$ for these tabulated values of α are plotted in Figure 3.

The portion of (5) that relates the Gauss quadrature node distribution (zero distribution of orthogonal polynomials) to the measure $\mu_{\widetilde{v}}$ is well-known: [34, 38, 44, 50, 33, 45]. That the unweighted Leja node distribution converges to the arcsine measure is likewise well-known. Our novel contribution to result (5) is for the limit for the weighted Leja formulation (4).
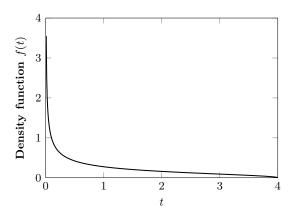
We give a summary of the weights w, contraction factors k_n, and some details about the asymptotic measures $\mu_{\widetilde{v}}$ in Table 1. The formulas for the exponential weight $w(z) = \exp(-|z|^\alpha)$ are not explicit for general α, so we explicitly compute and collect the density and distribution expressions for a selection of values for α in Table 2. Finally, the densities associated to $\mu_{\widetilde{v}}$ are plotted for these specials cases in Figure 3.

We note that by affine scaling, the limits of weighted Leja sequences in Tables 1 and 2 for other types of weights are readily derivable. I.e., suppose we have a new weight W defined on a new parameter Z that results from affine scaling of a canonical w, z pair from Theorem 3.1:

$$Z = T(z) \triangleq Az + B, \qquad\qquad W(Z) = w\left(T^{-1}(Z)\right),$$

for some $A, B \in \mathbb{R}$. Then the new limiting density $g(s)$ and distribution $G(s)$ for a W-weighted Leja sequence on the Z domain $T(I)$ can be expressed in terms of the weighted Leja asymptotics in Tables 1 and 2:

$$g(s) = \frac{1}{A}f\left(T^{-1}(s)\right), \qquad\qquad G(s) = F\left(T^{-1}(s)\right)$$
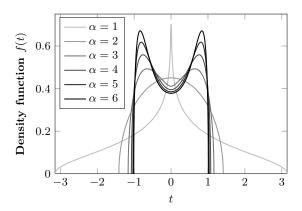
Figure 3. Plots of asymptotic densities for contracted weighted Leja (and Gauss) points. Left: Limiting density for the Laguerre weight $w(z) = \exp(-z)$. Right: Limiting density for the exponential weight $w(z) = \exp(-|z|^\alpha)$ for various α. See Table 1 for a summary and Table 2 for explicit formulas.

for $s \in T(I)$. The contraction factor for Z will be the same as it was for z.

We prove Theorem 3.1 by leveraging a significant result from potential theory: nodal sets that are 'asymptotically weighted Fekete' distribute according to the weighted potential equilibium measure [45, 49]. We prove that contracted versions of weighted Leja sequences are asymptotically weighted Fekete and essentially obtain (our novel contribution to) Theorem 3.1 as a corollary.

Let $V(\xi_0, \ldots, \xi_N)$ denote the modulus determinant of the polynomial Vandermonde matrix W on any array of points $\xi_0, \ldots, \xi_N$. W has entries $(W)_{j,k} = \xi_j^{k-1}$ for $j, k = 0, \ldots N$. Given a weight function $\widetilde{v}$ and a positive integer N define the maximum attainable value for the following weighted determinant:

$$\delta_{\widetilde{v}}^{(N)} = \left[\max_{(\xi_0, \ldots, \xi_N) \in I^{N+1}} V(\xi_0, \ldots, \xi_N) \prod_{j=0}^{N} \widetilde{v}^N(\xi_j) \right]^{2/(N^2+N)}$$

For any fixed N, a cardinality-N point set that achieves the maximum weighted determinant under the bracket is called a weighted Fekete set. It is known that the behavior of this maximum determinant value has a finite limit, the $\widetilde{v}$-weighted transfinite diameter of I:

$$\delta_{\widetilde{v}} = \lim_{N \to \infty} \delta_{\widetilde{v}}^{(N)}.$$

Any set of points whose asymptotic determinant limits to the transfinite diameter is called a set of asymptotically weighted Fekete points. In the cases of Theorem 3.1, the k_N-contracted w-weighted Leja points we proposed in (4) are asymptotically $\widetilde{v}$-weighted Fekete.

Theorem 3.2. *In all the cases of Theorem 3.1, the Leja sequence z_n defined by (4) produces a set of points whose k_N-contraction is asymptotically weighted Fekete:*

$$(6) \qquad \lim_{N \to \infty} \left[V(z_{0,N}, \ldots, z_{n,N}) \prod_{n=0}^{N} \widetilde{v}^N(z_{n,N}) \right]^{2/(N^2+N)} = \delta_{\widetilde{v}}, \qquad z_{n,N} = k_N z_n$$

where $\widetilde{v}$ is the weight function corresponding to the limit measure in Theorem 3.1.

Theorem 3.2 is a stronger result than Theorem 3.1 (see Lemma 7.1) and is the result that we spend the most effort proving. Once this is established, it is well-known that asymptotically weighted Fekete points distribute according to the weighted equilibrium measure $\mu_{\widetilde{v}}$. See Section 7 for details and the proof.

3.2. **Quadrature with Leja sequences.** The construction of Leja points is motivated mainly by interpolation; however quadrature/cubature in a multidimensional sparse grid framework is very desirable. To this end, one may simply explicitly integrate an interpolant on Leja points to construct a quadrature rule.

Consider a weighted Leja sequence z_n constructed using (4). We need only integrate the interpolant constructed from data on the z_n. Let $\{p_n\}$, $n = 0, \ldots, N-1$ denote family of polynomials orthonormal under $w(z)$. There are two observations we need: (i) if we assume that w is a probability density function, then $p_0 \equiv 1$, and (ii) even if w is known only empirically and does not have a representation in terms of classical functions, there are simple and accurate methods to construct the p_n in one dimension [28, 20, 36]. Given data f_n we wish to interpolate at the sites z_n, so we seek the coefficients c_n solving the linear problem

$$\mathbf{Vc} = \mathbf{f}, \qquad\qquad V_{n,m} = p_{m-1}(z_n).$$

Since $\mathbf{c} = \mathbf{V}^{-1}\mathbf{f}$, and

$$\int_\Omega \sum_{n=1}^N c_n p_{n-1}(z)\omega(z)\mathrm{d}z = c_1 \int_\Omega p_0(z)\omega(z)\mathrm{d}z = c_1,$$

then we immediately conclude that the first row of the matrix $\mathbf{V}^{-1}$ gives us quadrature weights w_n defining the Leja polynomial quadrature rule

$$Q^1_{N-1}f \simeq \sum_{n=1}^N w_n f(z_n).$$

The superscript '1' indicates that this quadrature rule applies to one-dimensional functions, and the subscript $N-1$ refers to the 'level' of the quadrature rule; both of these indicators unnecessary at the moment, but are meaningful in coming sections.

Naturally we wish to understand whether the Leja quadrature rules are useful in one dimension before proceeding to use them in higher-dimensional situations. We first verify that the quadrature rules are stable. The relative condition number of the quadrature rule Q^1_{N-1} is given by the ∞-norm of the $1 \times N$ matrix $\mathbf{W}$ with entries $W_{1,n} = w_n$. Thus, the condition number of the quadrature rule is

$$\kappa^1_{N-1} = \frac{\sum_{n=1}^N |w_n|}{\sum_{n=1}^N w_n} = \sum_{n=1}^N |w_n|,$$

where the last equality holds under the assumption that the p_n are orthonormal with respect to a probability density function w. The metric κ^1 indicates the presence of negative weights, which make the computation susceptible to catastrophic cancellation. The left-hand pane of Figure 4 graphs κ^1_{N-1} for three choices of w: the uniform density on $[-1,1]$, an oscillatory weight w on the same domain, and finally the Gaussian density function $w \propto \exp(-z^2)$ on $\mathbb{R}$. We see that the quadrature rules are all relatively well-conditioned.

The accuracy of the rules are considered in the following examples:

$$f_1(z) = \frac{1}{1 + 101\left(z - \frac{\pi}{4}\right)^2} \qquad\qquad f_2(z) = \cos\left(1 + 100z^3\right)$$

We perform one-dimensional global interpolation and refinement for these functions using the (uniform) Leja grids, nested Clenshaw-Curtis grids, and Legendre-Gauss-Patterson grids. Figure 5 shows results in the discrete maximum norm and the quadrature error. If $\widetilde{f}$ is the interpolatory approximation, then on a 10^4-sized Clenshaw-Curtis grid x_n with weights v_n, these metrics are defined
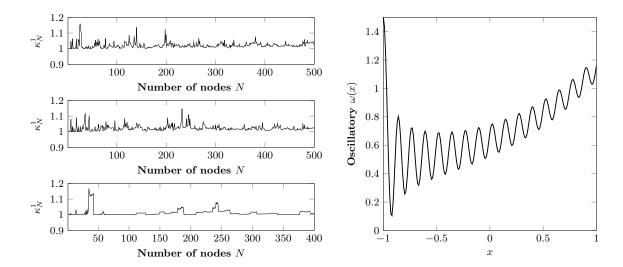
Figure 4. Left: Absolute condition number for Leja interpolatory quadrature rules for various weight functions ω: (top) $\omega \propto 1$ the uniform density, (middle) $w(z) \propto \frac{1}{2}I_0(1+z) + J_0(50+50z)$, where J_0 (I_0) is the (modified) Bessel function of the first kind, (bottom) $w(z) \propto \exp(-z^2)$ the Gaussian density. Right: plot of the second weight function, $w(z) \propto \frac{1}{2}I_0(1+z) + J_0(50+50z)$.

as

$$\sqrt{\sum_n v_n \left(f_1(x_n) - \widetilde{f}_1(x_n) \right)^2} \qquad \text{(Discrete } \ell^2 \text{ error)}$$

$$\max \left| f_1(x_n) - \widetilde{f}_1(x_n) \right| \qquad \text{(Discrete maximum error)}$$

$$\left| \sum_n v_n \left(f_1(x_n) - \widetilde{f}_1(x_n) \right) \right| \qquad \text{(Quadrature error)}$$

The left-hand pane of Figure 5 shows that the Leja grid is no less accurate than any of them in the maximum norm (being as accurate as the Clenshaw-Curtis grid). A discrete ℓ^2 error metric behaves similarly. The Leja sequence performs noticeably worse than the other two for the quadrature metric. We sacrifice quadrature accuracy in order to gain some dexterity in high-dimensional refinement: with the Clenshaw-Curtis or Legendre-Gauss-Patterson grid every refinement doubles the size of the (univariate) rule, whereas with the Leja procedure we can stop refinement at any size we choose. That the Gauss-Patterson grid performs so poorly for the maximum norm approximation can be explained by the fact that Gauss-Patterson nodes are constructed only to obtain a high degree of polynomial integration, not for interpolatory approximation. Gauss-Patterson grids that are formed for nested quadrature are not necessarily good for interpolation. This can be seem by comparing the left- and right-hand panels in Figure 5.

The relatively good behavior of a Leja grid is not useful unless we have a good error metric from the hierarchical surplus. Now consider function f_2, and the results in Figure 6, where we show the behavior of the surplus versus all the error metrics. The left-hand pane shows that the maximum Leja surplus is an excellent indicator of error in all three norms. The right-hand pane shows similar results for the Clenshaw-Curtis grid, but two observations are apparent: first, if using the surplus as a refinement technique, the surplus when $N = 257$ does not accurately reflect the actual error at the next level $N = 513$. This is true even if one were to employ a type of Richardson extrapolation to estimate the error. Second, the surplus indicator is a very conservative estimate of the error. In such a case, it is likely that we will refine more than is necessary in order to obtain an approximation.
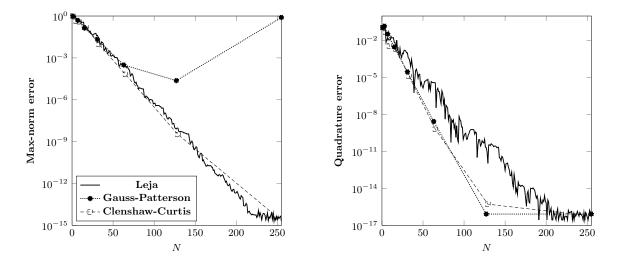
Figure 5. Left: Maximum errors for the one-dimensional Runge function f_1 with Leja, Clenshaw-Curtis, and Gauss-Patterson nested refinements. Right: Maximum hierarchical surplus at each level.
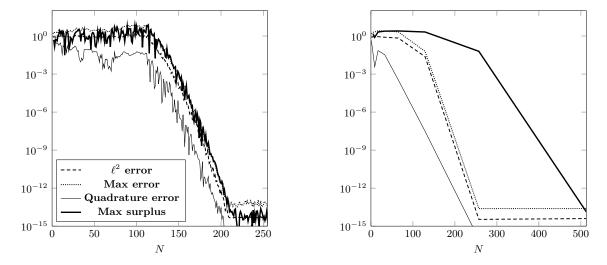

Figure 6. Left: Errors and the hierarchical surplus at each level for a one-dimensional Leja refinement of oscillatory function f_2. Right: Errors and surplus using a Clenshaw-Curtis refinement.

Both of these observations do not hold for the Leja surplus, which is sharper estimate of the error, and may be refined with arbitrary size.

3.3. **Barycentric Interpolation.** A numerically robust method for computing polynomial interpolants is furnished by the Barycentric interpolation formula (see [3] for an accessible introduction). This method is computationally efficient with respect to computing and evaluating the interpolant, and is stable so long as the interpolation problem itself is stable.

On a set of nodes $z_1, \ldots, z_N$, the degree-$(N-1)$ polynomial interpolant of a function $f(z)$ with data $f_n = f(z_n)$ is given by

$$(7) \qquad p(z) = \sum_{n=1}^{N} f_n \ell_n(z) = \frac{\sum_{n=1}^{N} \frac{b_n f_n}{z - z_n}}{\sum_{n=1}^{N} \frac{b_n}{z - z_n}},$$

where $\ell_n(z)$ are the cardinal Lagrange interpolation basis and b_n are the Barycentric weights, both defined as

$$\ell_n(z) = \prod_{m \neq n} \frac{z - z_m}{z_n - z_m}, \qquad\qquad b_n = \left[\prod_{m \neq n} (z_n - z_m) \right]^{-1}.$$

The Barycentric formula (7) allows evaluation of the interpolant $p(z)$ in only $\mathcal{O}(N)$ operations once the weights b_n are precomputed. In addition, the symmetry of this formula allows one to renormalize all the weights b_n by the same constant without affecting the interpolant. It is known that when the interpolation nodes z_n correspond to a well-conditioned interpolation operator, then the weights b_n all have comparable magnitude, leading to a well-conditioned numerical procedure.

With Leja sequences, we are essentially interested in v-weighted polynomial interpolation. (Recall (3).) This means that while we want to produce a polynomial interpolant f, we do so by interpolating the function vf using v-weighted polynomials. With this in mind, it is straightforward to show that the v-weighted analogue of (7) is the following polynomial formula:

$$(8) \qquad p(z) = \frac{1}{v(z)} \sum_{n=1}^{N} v_n f_n \ell^v(z) = \frac{\sum_{n=1}^{N} \frac{b_n^v v_n f_n}{v - v_n}}{\sum_{n=1}^{N} \frac{b_n^v v_n}{z - z_n}},$$

where $v_n \triangleq v(z_n)$, ℓ^v are v-weighted Lagrange polynomials, and b_n^v are v-weighted Barycentric weights:

$$\ell_n^v(z) = \frac{v(z)}{v_n} \ell_n(z), \qquad\qquad b_n^v = \frac{b_n}{v_n}.$$

For a sequence of Leja nodes generated according to (4), we use the Barycentric weights b_n^v given above to perform interpolation. We observe in practice that with this normalization that the weights b_n^v are all of comparable magnitude, just as we expect them to be for a well-conditioned interpolation problem.

Note that we do not necessarily avoid any troublesome numerical computations in the reformulated case (8); we have merely recast the problem into one that appears numerically well-conditioned. The actual process of interpolating f by an unweighted polynomial on an unbounded domain will still be mathematically ill-conditioned.

4. Sparse grids with univariate Leja rules

The approximation of a quantity depending on a finite number of Euclidean-like parameters is difficult when the parametric dimension d is large. It is well-known that an approximation to an r-times differentiable function converges with a rate of $\mathcal{O}(N^{-r/d})$ [2]. Let $f : \Gamma_{\boldsymbol{z}} \to \mathbb{R}$ for $\Gamma_{\boldsymbol{z}} \subset \mathbb{R}^d$ be a function that we wish to approximate and $\Gamma_{\boldsymbol{z}}$ be the domain of the possibly high-dimensional parameter $\boldsymbol{z}$ upon which the function depends. Spatial and temporal variables are modeled separately. For simplicity we assume that $\Gamma_{\boldsymbol{z}}$ is an isotropic tensor-product domain, so that the one-dimensional restricted variables z_k with $\boldsymbol{z} = (z_1, \ldots, z_d)$ all take values on the same restricted one-dimensional space.

When f is approximated via a sampling procedure, the curse of dimensionality is readily apparent: let $\Xi^1 \subset \mathbb{R}$ be an n-point nodal set in one dimension with associated quadrature weights w_k. A tensorization of this quadrature rule over d dimensions yields the nodal set

$$(9) \qquad \Xi^d = \bigotimes_{j=1}^{d} \Xi^1, \qquad\qquad N = |\Xi^d| = n^d.$$

The growth of the size of the tensorized set Ξ^d usually makes it infeasible for usage in high-dimensional approximation methods. (This is true both in cases when d is not small and fixed and n is increased, or when n is fixed and d is increased.) There are alternatives to tensor constructions, but approximation with any space-filling design requires $\mathcal{O}(n^d)$ samples, where n is the number of samples 'per dimension'.

An alternative to space-filling designs is the popular *sparse grid*, so named because of its geometrically dispersed distribution in $\Gamma_{\boldsymbol{z}}$. Like tensor constructions, sparse grids tensorize univariate nodal arrays, but sparse grids also attempt to delay the impact of the curse of dimensionality by taking only certain combinations of tensor products. We delay introduction of sparse grids until Section 4; for now we concentrate on motivating our choice of univariate rule: Leja sequences.

4.1. Common univariate rules.

The sparse grid construction requires specification of a univariate grid Ξ_l^1, for $l = 0, 1, \ldots$. Several choices for these univariate grids work well, and among the most popular are a Clenshaw-Curtis (CC) grid, or grids associated with Gauss quadrature rules [22]. For concreteness, we consider the one-dimensional finite interval $[-1, 1]$. Then, for example, we may choose CC and we have the following grid for any level l:

$$\Xi_l = \left\{ z_{l,n} \right\}_{n=1}^{N_l}, \qquad\qquad z_{l,N_l+1-n} = \cos\left(\frac{(n-1)\pi}{N_l - 1} \right),$$

for $n = 1, \ldots, N_l$, and $N_l = 2^l + 1$. This particular choice for Ξ_l^1 is popular for two reasons: (i) $\Xi_l \subset \Xi_{l+1}$ allowing for *hierarchical* approximation and adaptive refinement with as few model evaluations as possible, and (ii) the sequence Ξ_l is known to be both an excellent interpolatory and quadrature grid. Of course, one apparent concern is that $|\Xi_{l+1}| - |\Xi_l| = N_{l+1} - N_l = 2^l$, which grows exponentially with the level l. This means that each stage of refinement requires addition of a large number of points. In general, a large number of nodes is not necessarily adverse so long as the resulting grids have some optimality regarding, e.g., maximum degree of polynomial integration [42].

An alternative univariate rule that is competitive for quadrature purposes is the Gauss-Patterson grid [41] wherein one constructs a grid that is a subset of a given Gauss quadrature grid, and satisfies some polynomial integration optimality conditions. However, it is not always possible to construct such grids depending on (i) the cardinality of the subset X_l and (ii) the weight function ω. Even when such construction is possible, construction of the Gauss-Patterson grid requires implementation of a nontrivial algorithm, and it is frequently easier to precompute and store the grids, making the method inflexible with respect to the choice of density ω.

There are several locally adaptive strategies for sparse grids that are also successful in combating the curse of dimensionality [27, 32]. One method that has enjoyed recent success is the locally adaptive, high order, generalized sparse grid construction [27]. In this setup, one uses a high-order Lagrange polynomial basis as the univariate building block for a local high-order polynomial approximation; because the approximation is local, targeted adaptive strategies that utilize the grid hierarchical surpluses are naturally applicable and effective. However, the adaptation is usually (locally) uniform and it is well-known that high-order polynomial approximation on a uniform grid raises computational challenges.

We propose use of (weighted) Leja sequences as univariate building blocks for an adaptive Smolyak sparse grid constructions. Leja sequences can easily add an arbitrary number of samples at each stage, and have good interpolatory and quadrature properties, making them excellent ingredients for the Smolyak algorithm. Leja sequences have been used a sparse grid building blocks before: [13, 12], but we believe this is the first investigation into adaptive hierarchical approximations for high-dimensional approximation.

Having discussed univariate Leja sequences at length in Section 3, we may now construct standard sparse grids using Leja sequences as building blocks. Let $\boldsymbol{z} = (z_1, \ldots, z_d) \in I_{\boldsymbol{z}} \subseteq \mathbb{R}^d$ be a random variable with probability density function $\omega(\boldsymbol{z}) : I_{\boldsymbol{z}} \to \mathbb{R}$, and assume that the components of X are mutually independent so that $I_{\boldsymbol{z}}$ is a tensor-product domain

$$I_{\boldsymbol{z}} = \bigotimes_{j=1}^{d} I_{z_j}.$$

We let $\omega_i : I_{z_j} \to \mathbb{R}$ denote the marginal PDF of z_i, so that $\omega(\boldsymbol{z}) = \prod_{j=1}^d \omega^j(z_j)$. Sparse grids [9] approximate f via a weighted linear combination of basis functions

$$I_n[f] := f_n = \sum_{k=1}^n v_k \, \Psi_k(\boldsymbol{z}) \tag{10}$$

The approximation is constructed on a set of anisotropic grids $\Xi_{\boldsymbol{\ell}}$ on the domain $I_{\mathbf{z}}$ where $\boldsymbol{\ell} = (l_1, \ldots, l_d) \in \mathbb{N}^d$ is a multi-index denoting the level of refinement of the grid in each dimension. These rectangular grids are Cartesian product of nested one-dimensional grid points $\Xi_l = \{\xi_{l,i} : i < 0 \le i \le m_l\}$

$$\Xi_{\boldsymbol{\ell}} = \Xi_{l_1} \times \cdots \Xi_{l_d}$$

The number of points m_l of a one-dimensional grid of a given level is dependent on the growth rate of the quadrature rule chosen.

The multivariate basis functions Ψ_k are a tensor product of one dimensional basis functions. Adopting the multi-index notation used above we have

$$\Psi_{\boldsymbol{\ell},\mathbf{i}}(\boldsymbol{z}) = \prod_{n=1}^d \psi_{\ell_n, i_n}(z_n) \tag{11}$$

where $\mathbf{i}$ determines the location of a given grid point. There is a one-to-one relationship between Ψ_k in (10) and $\Psi_{\boldsymbol{\ell},\mathbf{i}}$ and each $\Psi_{\boldsymbol{\ell},\mathbf{i}}$ is uniquely associated with a grid point $\boldsymbol{z}_{\boldsymbol{\ell},\mathbf{i}} = (\xi_{\ell_1,i_1}, \ldots, \xi_{\ell_d,i_d}) \in \Xi_{\boldsymbol{\ell}}$. Many different one-dimensional basis functions $\psi_{\ell_n, i_n}(\xi_n)$ can be used. In the following we employ one-dimensional Lagrange polynomials for the functions ψ_{ℓ_n, i_n}.

The multi-dimensional basis (11) spans the discrete space $V_{\boldsymbol{\ell}} \subset L^2(I_{\mathbf{z}})$

$$V_{\boldsymbol{\ell}} = \operatorname{span}\{\Psi_{\boldsymbol{\ell},\mathbf{i}} : \mathbf{i} \in \mathcal{K}_{\boldsymbol{\ell}}\} \qquad \mathcal{K}_{\boldsymbol{\ell}} = \{\mathbf{i} \in \mathbb{N}_0^d : i_k = 0, \ldots, m_{l_k}, k = 1, \ldots, d\}$$

These discrete spaces can be further decomposed into hierarchical difference spaces

$$W_{\boldsymbol{\ell}} = V_{\boldsymbol{\ell}} \setminus V_{\boldsymbol{\ell}} \bigoplus_{n=0}^d V_{\boldsymbol{\ell}-\mathbf{e}_n}$$

The subspaces $W_{\boldsymbol{\ell}}$ consists of all basis functions $\Psi_{\boldsymbol{\ell},\mathbf{i}} \in V_{\boldsymbol{\ell}}$ which are not included in any of the spaces $V_{\mathbf{k}}$ smaller than $V_{\boldsymbol{\ell}}$, i.e. with $\mathbf{k} < \boldsymbol{\ell}$ with $<$ the lexicographic partial ordering on multi-indices. These hierarchical difference spaces can be used to decompose the input space such that

$$V_{\boldsymbol{\ell}} = \bigoplus_{\mathbf{k} \le \boldsymbol{\ell}} W_{\boldsymbol{\ell}} \quad \text{and} \quad L^2(I_{\mathbf{z}}) = \bigoplus_{k_1=0}^\infty \cdots \bigoplus_{k_d=0}^\infty W_{\mathbf{k}} = \bigoplus_{\mathbf{k} \in \mathbb{R}^d} W_{\mathbf{k}}$$

For numerical purposes we must truncate the number of difference spaces used to construct V. Traditional isotropic sparse grids can be obtained by all hierarchical subspaces $W_{\boldsymbol{\ell}}$ with and index set that satisfy

$$\mathcal{L} = \{\boldsymbol{\ell} : |\boldsymbol{\ell}|_1 \le l\} \tag{12}$$

Given a truncation, such as the *a priori* one above or one which has been determined adaptively, f can be approximated by

$$f_n = \sum_{\boldsymbol{\ell} \in \mathcal{L}} f_{\boldsymbol{\ell}}, \qquad\qquad f_{\boldsymbol{\ell}} = \sum_{\mathbf{i} \in \mathcal{I}_{\boldsymbol{\ell}}} v_{\boldsymbol{\ell},\mathbf{i}} \, \Psi_{\boldsymbol{\ell},\mathbf{i}}(\boldsymbol{z}) \tag{13}$$

where $\mathcal{I}_{\boldsymbol{\ell}} = \{\mathbf{i} : \Psi_{\boldsymbol{\ell},\mathbf{i}} \in W_{\boldsymbol{\ell}}\}$.

Here we note that the $v_{\boldsymbol{\ell},\mathbf{j}}$ are the coefficient values of the hierarchical product basis, also known as the hierarchical surplus. The surpluses are simply the difference between the function value and the sparse grid approximation at a point, not already in the sparse grid. That is

$$v_{\boldsymbol{\ell},\mathbf{j}} = f(\boldsymbol{z}_{\boldsymbol{\ell},\mathbf{i}}) - f_n(\boldsymbol{z}_{\boldsymbol{\ell},\mathbf{i}}), \qquad\qquad \mathcal{L} \cap \boldsymbol{\ell} = \emptyset$$

The particular choice of sparse grid in this paper is one constructed with univariate hierarchical Leja points: the sequence of points that we use for each dimension to evaluate the surplus and construct

the interpolant is a univariate Leja sequence. In this way, the point sets $z_{\ell,\mathbf{i}}$ are nested, and the number of points to add at each level can be as small or large as we wish. (I.e. the Leja choice allows a great deal of granularity for refinement.)

4.2. **Dimension adaptivity.** The dimensional adaptivity of our algorithm in this section is based on the idea presented in [22]. We begin with a low-level isotropic sparse grid approximation with a set of subspaces representing the current approximation $\mathcal{L}$ and the set of active subspaces $\mathcal{A}$ that indicate the levels for potential refinement. Often $\mathcal{L} = W_{\mathbf{0}}$ and $\mathcal{A} = \{W_{\mathbf{e}_k}, k = 1 \ldots, d\}$. We then choose $W_{\boldsymbol{\ell}} \in \mathcal{A}$ with the largest error indicator $\gamma_{\boldsymbol{\ell}}$ and refine that subspace. Here we define the error indicator $\gamma_{\boldsymbol{\ell}}$ as

$$(14) \qquad \gamma_{\boldsymbol{\ell}} = \int_{\Gamma_{\boldsymbol{z}}} (f_{\boldsymbol{\ell}})^2 d\omega(\boldsymbol{z}) - \left(\int_{\Gamma_{\boldsymbol{z}}} (f_{\boldsymbol{\ell}}) d\omega(\boldsymbol{z}) \right)^2, \qquad\qquad \eta = \sum_{\boldsymbol{\ell} \in \mathcal{A}} \gamma_{\boldsymbol{\ell}}$$

The indicator $\gamma_{\boldsymbol{\ell}}$ measures the contribution of the subspace $\boldsymbol{\ell}$ to the variance of f_n and the global indicator η measures the contribution of all active subspaces to the variance of f_n. These indicators are calculated by transforming the Lagrange interpolant on each *hierarchical* subspace into a Polynomial Chaos Expansion that is orthogonal to the (possibly mixed) distribution weight $\omega(\boldsymbol{z})$. The cost of this transformation is linear in terms of the number of subspace points [10]. The chosen subspace for refinement with index $\boldsymbol{\ell}$ is refined by adding all indices $W_{\mathbf{k}}$ with $\mathbf{k} = \boldsymbol{\ell} + \mathbf{e}_n, n = 1, \ldots, d$ that satisfy the following admissibility criterion

$$(15) \qquad\qquad \boldsymbol{\ell} - \mathbf{e}_k \in \mathcal{L} \text{ for } 1 \le k \le d, \, l_k > 1$$

The active set $\mathcal{A}$ is then rebuilt by adding each subspace corresponding to the indices from (15). This process continues until a computational budget limiting the number of model samples (grid points) is reached or a global error indicator drops below a predefined threshold. Pseudo-code for the dimension adaptive algorithm is shown in Algorithm 1.

The INDICATOR and TERMINATE routines in Algorithm 1 control which subspaces are added to the sparse grid via the use of a subspace error and global error metric. The indicators respectively provide estimates of the contribution of a subspace to reducing the error in the interpolant, and the error in the entire interpolant.

Algorithm 1 INTERPOLATE$[f(\mathbf{z}),\mathcal{L},\mathcal{A},\tau,n] \to f_n$

1: For a given $\mathcal{L}$ the points in the sparse grid are $\Xi := \bigcup_{\boldsymbol{\ell} \in \mathcal{L}} \Xi_{\boldsymbol{\ell}}$.
2: The number of sparse grid points are $N = \#\Xi$
3: **while** NOT TERMINATE$[\mathcal{A},N,\tau,n]$ **do**
4:     $W := \mathrm{argmax}_{W_{\boldsymbol{\ell}} \in \mathcal{A}} \gamma_{\boldsymbol{\ell}}$ % Determine the subspace with the highest priority
5:     $\mathcal{A} := \mathcal{A} \setminus W$ % Remove W from the active set
6:     $\mathcal{L} := \mathcal{L} \cup W$
7:     $\mathcal{J} := \text{REFINE}[W,\mathcal{L}]$ % Find all admissible forward neighbors of W
8:     $\gamma_{\boldsymbol{\ell}} := \text{INDICATOR}[W_{\boldsymbol{\ell}}] \, \forall \, W_{\boldsymbol{\ell}} \in \mathcal{J}$ % Calculate the priority of the neighbors
9:     $\mathcal{A} := \mathcal{A} \cup \mathcal{J}$ % Add the forward neighbors to the active index set
10: **end while**

---

5. Numerical examples

We consider several multidimensional examples below that compare the Smolyak-Leja algorithm with a more standard Clenshaw-Curtis-Smolyak algorithm. Effectively, we see that the Leja construction is competitive (usually superior) to Clenshaw-Curtis when an interpolation metric is used. However, they appear suboptimal when a quadrature metric is evaluated. This is not surprising as Leja sequences are constructed with the goal of interpolation and not necessarily for quadrature.

Throughout these examples we compute discrete ℓ_2 errors ε_{ℓ_2} using 100,000 random samples taken in a Monte-Carlo fashion from the distribution of the input variable $\boldsymbol{z}$. We also report absolute errors

in the sparse grid mean ε_μ and variance ε_{σ^2} where the exact moments are computed using a high-resolution sparse grid that was refined so that its ℓ_2 error was in the order of machine precision. The error metric ε_{ℓ^2} is simply the discrete ℓ^2 error (RMSE).

5.1. **Random oscillator.** This section investigates the relative performance of the sparse grids when approximating the output from a model of linear oscillator subject to external forcing with six unknown parameters. That is,

$$\text{(16)} \qquad \frac{d^2x}{dt^2}(t, \mathbf{z}) + \gamma \frac{dx}{dt} + kx = f \cos(\omega t),$$

subject to the initial conditions

$$\text{(17)} \qquad x(0) = x_0, \quad \dot{x}(0) = x_1,$$

where we assume the damping coefficient γ, spring constant k, forcing amplitude f and frequency ω, and the initial conditions x_0 and x_1 are all uncertain. We solve (16) analytically to allow us to avoid consideration of discretization errors in our investigation.

Let us choose our quantity of interest to be the position $x(T)$ of the osciallator at $T = 20$ seconds and let $\mathbf{z} = (\gamma, k, f, \omega, x_0, x_1)$ where $\gamma \in [0.08, 0.12]$, $k \in [0.03, 0.04]$, $f \in [0.08, 0.12]$, $\omega \in [0.8, 1.2]$, $x_0 \in [0.45, 0.55]$, $x_1 \in [-0.05, 0.05]$. For this choice of random parameters any parameter realization in $I_\mathbf{z}$ will produce an underdamped harmonic oscillator.

Figure 7 compares the ℓ_2 accuracy in the sparse grid interpolants obtained using Clenshaw-Curtis nodes and Leja nodes. Although both univariate rules have similar interpolation properties in one-dimension, the one-at-a-time nestedness of the Leja rule produces, in this higher-dimensional setting, an approximation that is significantly more accurate than the approximation based upon the Clenshaw-Curtis quadrature rule.
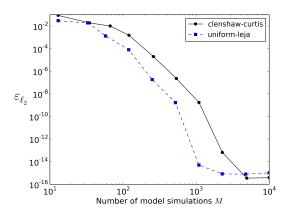

Figure 7. Convergence of RMSE in the sparse grid approximation of the oscillator position with respect to the number of model evaluations.

5.2. **Borehole model.** For the next numerical demonstration, consider the following model of water flow through a borehole

$$\text{(18)} \qquad f(z) = \frac{2\pi T_u(H_u - H_l)}{\log(r/r_w)\left(1 + \frac{2\pi T_u}{\log(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l}\right)}$$

where the unkonwn parameters are uniform random variables $\mathbf{z} = (z_1, \ldots, z_8)$ with the following bounds: $z_1 := r_w \in [0.05, 0.15]$ (meters) denotes the radius of borehole, $z_2 := r \in [100, 50000]$ (meters) the radius of influence, $z_3 := T_u \in [63070, 115600]$ (meters^2/years) the transmissivity of upper aquifer, $z_4 := H_u \in [990, 1110]$ (meters) the potentiometric head of upper aquifer, $z_5 := T_l \in [63.1, 116]$ (meters^2/years) the transmissivity of lower aquifer, $z_6 := H_l \in [700, 820]$ (meters) the

potentiometric head of lower aquifer, $z_7 := L \in [1120, 1680]$ (meters) the length of borehole, and $z_8 := K_w \in [9855, 12045]$ (meters/year) the hydraulic conductivity of borehole.

Figure 8 compares the accuracy of sparse grids based upon the univariate Clenshaw-Curtis and Leja nodes. Again the Leja interpolation sequence produces a more accurate interpolant for a given number of function evaluations, but suffers when evaluating quadrature quantities such as the mean shown in the left-hand pane. In this case as with many others, the Clenshaw-Curtis quadrature rule produces a more accurate estimate of the mean of the function. This statement is consistent with the one-dimensional results shown in Figure 5.
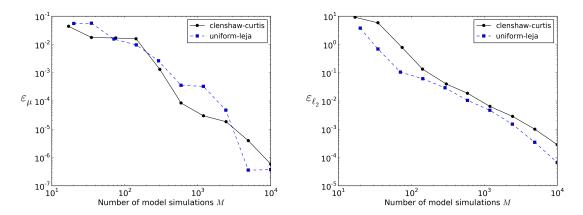

Figure 8. Convergence of the mean and RMSE in the sparse grid approximation of the borehole model with respect to the number of model evaluations.

5.3. **Heterogeneous diffusion equation.** In this section, we consider the heterogeneous diffusion equation in one-spatial dimension subject to uncertainty in the diffusivity coefficient. For $d \geq 1$ random dimensions:

$$-\frac{d}{dx}\left[a(x, \mathbf{z}) \frac{du}{dx}(x, \mathbf{z}) \right] = 1, \quad (x, \mathbf{z}) \in (0, 1) \times I_{\mathbf{z}} \tag{19}$$

subject to the physical boundary conditions

$$u(0) = 0, \quad u(1) = 0 \tag{20}$$

Furthermore assume that the random diffusivity satisfies

$$a(x, \mathbf{z}) = \bar{a} + \sigma_a \sum_{k=1}^{d} \sqrt{\lambda_k} \phi_k(x) z_k \tag{21}$$

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

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

The variability of the diffusivity field (21) is controlled by σ_a and the correlation length l_c which determines the decay of the eigenvalues λ_k. Here we wish to approximate the solution $u(1/3, \mathbf{z})$ when $d = 40$, $\sigma_a = 0.021$ and $l_c = 1/14$ and $z_k \in [-1, 1]$, $k = 1, \ldots, 40$ to be independent and uniformly distributed random variables. We solve (21) non-intrusively: for each node $\mathbf{z}_{\ell,1}$ on our sparse grid, we use a finite-element discretization in x to compute the solution. The comparison between Leja and Clenshaw-Curtis Smolyak construction is shown in Figure 9. In this case, the Leja construction only performs marginally better than the CC approach for interpolation, and exhibits a now-familiar difficulty with quadrature. We explain this difference in the following way: for this equation, we certainly have dimensional anisotropy because the eigenvalues λ_k decay. However, if we plot the
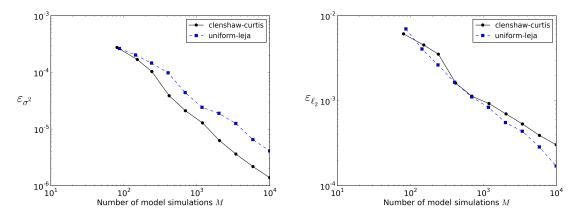
Figure 9. Convergence of variance and RMSE in the sparse grid approximation of the solution to the diffusion equation with respect to the number of model evaluations
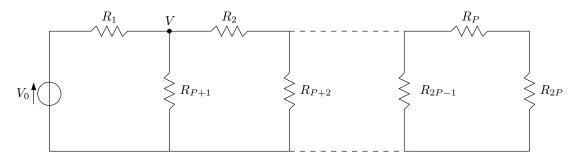

Figure 10. Resistor network

parameter indices l for the subspaces $W_{\boldsymbol{\ell}}$ that contribute significantly to the solution, we will see an ellipsoid shape in index space. Thus, extra refinement performed by CC in certain directions is not wasted because these degrees of freedom can properly resolve mixed terms in parameter space. In this example, the granularity offered by Leja sequences is not needed or utilized.

5.4. **Resistor network.** Consider the electrical resistor network shown in Figure 10 [43]. The network is comprised of $d = 2P$ resistances of uncertain ohmage and the network is driven by a voltage source providing a known potential V_0. We are interested in determining how the voltage V shown in Figure 10 depends on the $d = 2P$ resistances, which we take as random parameters that are independent and identically distributed Gaussian random variables with mean $\mu = 1$ and standard deviation $\sigma = 0.005$. Note for this value of σ, the probability that we encounter negative resistances is extremely small, and so apart from the obvious modeling error of possible negative resistances, no numerical difficulties are introduced. I.e. none of the sparse grid points or random samples used to generate the error resulted in a negative resistance. In this example we set $d = 40$ and set the reference potential $V_0 = 1$.

In this case our comparison is not with CC, but with a nested Genz-Keister rule, which is one of the the standard ways to peform nested interpolation and quadrature under a Gaussian weight [21]. Here our Leja rule is generated on an infinite domain with univariate weight function $w(z_j) = \exp(-(z_j - \mu)^2/(2\sigma^2))$. In Figure 11 we see significant interpolatory improvement with the Leja rule, and even the quadrature results are competitive in this example.
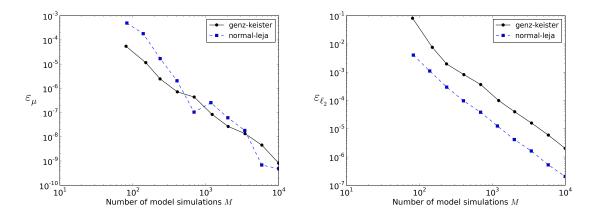
Figure 11. Convergence of the Mean and RMSE in the sparse grid approximation of V with respect to the number of model evaluations.

6. Summary

We have used Leja interpolatory grids as one-dimensional composite rules for adaptive Smolyak sparse grid construction. In one dimension, Leja rules are excellent interpolation grids and are a sequence, allowing one to generated nested rules with arbitrary granularity. We have shown that for several classical one-dimensional weight functions of interest, a corresponding weighted Leja rule produces a sequence whose empirical distribution asymptotically coincides with the limiting distribution for the Gauss quadrature nodes of the same family.

Using Leja rules to build up sparse grids in multiple dimensions grants the user a greater dexterity in adaptive refinement compared to more standard composite rules such as Clenshaw-Curtis. We have shown via several examples that the Leja rule can outperform standard high-order Smolyak constructions in interpolatory metrics, but are suboptimal when considering quadrature metrics. Design of Leja-like rules that are effective for approximating integrals will be the subject of future investigation.

7. Proof of Theorem 3.2

7.1. **Weighted potential theory.** In this section we give the proof of Theorems 3.1 and 3.2. Indeed, if we show Theorem 3.2, then well-established results imply Theorem 3.1. We recall our notation: w is a given weight function associated to z. v is the square root of w. $\widetilde{v}$ is related to w through Theorem 3.1, and is essentially the square root of w, ignoring polynomial factors.

The proof of Theorem 3.2 relies on results from weighted potential theory. Potential theory is frequently explored in the complex plane $\mathbb{C}$, but we will restrict ourselves to subsets of $\mathbb{C}$ lying on the real axis. An excellent exposition of univariate weighted potential theory with comprehensive historical references is given in [45].

Let domain $I \subseteq \mathbb{R}$ with non-negative weight function $\widetilde{v}$ be given. On rather mild assumptions on $\widetilde{v}$ and I then there exists a unique probability measure denoted $\mu_{\widetilde{v}}$ under which a weighted logarithmic energy for $Q = -\log \widetilde{v}$ is minimized:

$$
(22) \qquad \mu_{\widetilde{v}} = \operatorname*{argmin}_{\mu \,:\, \mu(I)=1} \int_I \int_I \log \frac{1}{|\eta - \xi| \widetilde{v}(\eta) \widetilde{v}(\xi)} \mathrm{d}\mu(\eta) \mathrm{d}\mu(\xi)
$$

$$
= \operatorname*{argmin}_{\mu \,:\, \mu(I)=1} \left[\int_I \int_I \log \frac{1}{|\eta - \xi|} \mathrm{d}\mu(\eta) \mathrm{d}\mu(\xi) + 2 \int_I Q(\xi) \mathrm{d}\mu(\xi) \right]
$$

A common physical analogy of the above is to find the minimum-energy electrostatic charge distribution (measure) in a region I when an external electrostatic field Q is applied. The measure $\mu_{\widetilde{v}}$ is

the weighted equilibrium measure of I in the presence of the field Q. Even if I is unbounded, $\mu_{\widetilde{v}}$ always has compact support.

A discrete version of the above optimization problem is furnished by the concept of Fekete points. An array of points which maximizes a weighted Vandermonde matrix determinant is called a set of Fekete points; this weighted determinant is a discrete, unnormalized Monte-Carlo-like estimate of the negative exponential of the integral in (22). A set of points $\xi_0, \ldots, \xi_n$ is an array of $\widetilde{v}$-weighted Fekete points if it satisfies

$$(23) \qquad \{\xi_0, \ldots \xi_n\} = \operatorname*{argmax}_{z_0, \ldots, z_n} V(z_0, \ldots, z_n) \prod_{j=0}^n \widetilde{v}^n(z_j) = \operatorname*{argmax}_{z_0, \ldots, z_n} \prod_{0 \le j < k \le n} |z_j - z_k| \widetilde{v}(z_j) \widetilde{v}(z_k),$$

where $V(z_0, \ldots, z_n)$ is modulus determinant of the $(n+1) \times (n+1)$ Vandermonde matrix W with entries $W_{r,s} = z_r^s$ for $0 \le r, s \le n$. Fekete sets are not necessarily unique, and are notoriously difficult to compute exactly. However, Fekete points are excellent interpolation/approximation nodal sets. The $n \to \infty$ limiting behavior of the weighted Vandermonde determinant for Fekete nodes is described by the weighted transfinite diameter $\delta_{\widetilde{v}}$ (= weighted logarithmic capacity [35]). Let V_n denote the maximum weighted determinant from (23) achieved by Fekete points. Then

$$(24) \qquad \lim_{n \to \infty} (V_n)^{1/m_n} = \delta_{\widetilde{v}},$$

where $m_n \triangleq 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. Any triangular array of nodes whose determinant behaves like (24) is called *asymptotically weighted Fekete*, alluding to the fact that the determinant is not exactly maximum, but is asymptotically comparable to Fekete points.

The connection between Fekete nodes and the equilibrium measure is established by the following result: if a triangular array of nodes $\{\xi_{j,n}\}_{j \le n}$ is asymptotically weighted Fekete, then its empirical measure distributionally converges to the weighted equilibrium measure:

Lemma 7.1 ([49, 45]). *Suppose $\xi_{j,n}$ is a triangular array of points satisfying*

$$\lim_{n \to \infty} \left[V(\xi_{0,n}, \ldots, \xi_{n,n}) \prod_{j=0}^n \widetilde{v}(\xi_{j,n}) \right]^{1/m_n} = \delta_{\widetilde{v}}.$$

Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^n \delta_{\xi_{j,n}} = \mu_{\widetilde{v}},$$

where δ_z is the Dirac mass centered at z.

Here (and elsewhere when discussing convergence of measures) equality is in the weak-$*$ sense. Our strategy is to first show Theorem 3.2, that a contracted version of the w-weighted Leja points from (4) are $\widetilde{v}$-weighted asymptotically Fekete. This will allow us to immediately use Lemma 7.1 to conclude Theorem 3.1.

To proceed, we will need the notion of Chebyshev constants. Given a potential-theoretic admissible weight $\widetilde{v}$ on I, the weighted Chebyshev constant of order n is

$$\tau_{\widetilde{v}}^{(n)} = \inf \left\{ \|\widetilde{v}^n(z) p_n(z)\|_I \mid p_n(z) = z^n + q_{n-1}(z) \; \forall \; q_{n-1} \in P^{n-1} \right\},$$

where $\| \cdot \|_I$ is the sup-norm on the domain I, and P^n is the space of polynomials of degree n or less. The sequence of constants $\left\{ \left(\tau_{\widetilde{v}}^{(n)} \right)^{1/n} \right\}$ is a decreasing sequence with a limit:

$$\inf_n \left(\tau_{\widetilde{v}}^{(n)} \right)^{1/n} = \lim_{n \to \infty} \left(\tau_{\widetilde{v}}^{(n)} \right)^{1/n} \triangleq \tau_{\widetilde{v}}$$

This limit is called the weighted Chebyshev constant [35], and coincides with the transfinite diameter $\delta_{\widetilde{v}}$ in the unweighted case, but is distinct in general weighted scenarios. Note that by definition,

$$(25) \qquad \|\widetilde{v}^n(z)(z^n + q_{n-1}(z))\|_I \ge (\tau_{\widetilde{v}})^n.$$

for any $q_{n-1} \in P^{n-1}$. In general, the relation between the Chebyshev constant $\tau_{\widetilde{v}}$ and the transfinite diameter $\delta_{\widetilde{v}}$ is given by

$$(26) \qquad \tau_{\widetilde{v}} = \delta_{\widetilde{v}} \exp\left(\int_S Q(z) \, \mathrm{d}\mu_{\widetilde{v}}(z) \right), \qquad\qquad S = \mathrm{supp}\, \mu_{\widetilde{v}}.$$

For the Jacobi, Hermite, and Laguerre cases mentioned in Theorem 3.1, our goal is to prove the result (6). Our contraction proofs are different for the cases of bounded I versus unbounded I. We first consider the bounded Jacobi case.

7.2. **Jacobi case.** We prove that the Leja maximization scheme (4) produces nodes whose limiting distribution is the *unweighted* equilibrium measure, or the arcsine measure. For the unweighted case we have $\widetilde{v} \equiv 1$, the contraction is $k_n \equiv 1$ (and so is omitted), and the transfinite diameter δ and the Chebyshev constant τ are identical:

$$(27) \qquad\qquad\qquad\qquad\qquad \delta = \tau$$

We begin by considering the proof assuming $\alpha, \beta \geq 0$. We have $v(z) = f^{(\alpha/2,\beta/2)}(z) = (1-z)^{\alpha/2}(1+z)^{\beta/2}$ for $\alpha, \beta \geq 0$ over $I = [-1,1]$. We will need the following constants:

$$C_1 = C_1(\alpha,\beta) \triangleq \left\| f^{(\alpha/2,\beta/2)} \right\|_\infty < \infty$$

$$C_2 = C_2(\alpha,\beta) \triangleq \left\| \frac{f^{(\lceil \alpha/2 \rceil, \lceil \beta/2 \rceil)}(z)}{f^{(\alpha/2,\beta/2)}(z)} \right\|_\infty < \infty$$

With these constants, we have

$$1 \geq \frac{v(z)}{C_1} = \frac{f^{(\alpha/2,\beta/2)}(z)}{C_1} \geq \frac{f^{(\lceil \alpha/2 \rceil, \lceil \beta/2 \rceil)}(z)}{C_1 C_2},$$

We now note that $\lceil \alpha/2 \rceil$ and $\lceil \beta/2 \rceil$ are integers, and for shorthand we write $\gamma = \lceil \alpha/2 \rceil + \lceil \beta/2 \rceil$. Then the right-hand side of the above equation is a monic polynomial (modulo sign) of degree γ. Therefore, we have

$$
\begin{aligned}
V(z_0, \ldots, z_n) = \prod_{j=1}^{n} \prod_{k=0}^{j-1} |z_j - z_k| &\geq \frac{1}{C_1^n} \prod_{j=1}^{n} v(z_j) \prod_{k=0}^{j-1} |z_j - z_k| \\
&= \frac{1}{C_1^n} \prod_{j=1}^{n} \left\| v(z) \prod_{k=0}^{j-1} |z - z_k| \right\|_I \\
&\geq \frac{1}{C_1^n C_2^n} \prod_{j=1}^{n} \left\| f^{(\lceil \alpha/2 \rceil, \lceil \beta/2 \rceil)}(z) \prod_{k=0}^{j-1} |z - z_k| \right\|_I \\
&\geq \frac{1}{C_1^n C_2^n} \prod_{j=1}^{n} \tau^{j+\gamma} = \frac{\tau^{m_n} \tau^{n\gamma}}{C_1^n C_2^n}
\end{aligned}
$$

Thus, we have

$$\tau^{m_n} \left[\frac{\tau^\gamma}{C_1 C_2} \right]^n \leq V(z_0, \ldots, z_n) \leq \max_{x_0, \ldots x_n \in \Xi} V(x_0, \ldots, x_n)$$

We raise all the above to the $1/m_n$ power, which yields

$$\lim_{n \to \infty} \left[V(z_0, \ldots, z_n) \right]^{1/m_n} \geq \tau = \delta$$

$$\lim_{n \to \infty} \left[V(z_0, \ldots, z_n) \right]^{1/m_n} \leq \delta$$

We have thus proven that Jacobi-weighted Leja sequences for $\alpha, \beta \geq 0$ are $\widetilde{v}$-asymptotically Fekete, i.e., we have proven (6).

For $-1 < \alpha, \beta < 0$, we proceed without loss under the assumption that both α and β are negative. In this case then $\xi_0 \neq \pm 1$ is arbitrarily chosen so that

$$z_1 = \operatorname*{argmax}_z v(z)|z - z_0|.$$

But since $\widetilde{v}$ is infinite at the endpoints ± 1, then the maximum is achieved at one of these points, say $z_1 = +1$. Then the Leja iteration continues:

$$z_2 = \operatorname*{argmax}_z v(z)|z - z_0||z - z_1|.$$

Now since $z_1 = +1$, then the last term is identical to $(1 - z)$, and we may absorb this term into the weight, giving a more explicit formula:

$$z_2 = \operatorname*{argmax}_z (1 - z)^{\alpha/2+1}(1 + z)^{\beta/2}|z - z_0|.$$

But since $\alpha/2 + 1 > 0$, then the maximum (infinity) is now achieved at $z = -1$, which becomes z_2. Now we choose z_3 and again combine the terms $|z - z_1|$ and $|z - z_2|$ into the weight function:

$$z_3 = \operatorname*{argmax}_z (1 - z)^{\alpha/2+1}(1 + z)^{\alpha/2+1} \prod_{\substack{k=0 \\ k \neq 1,2}}^{2} |z - z_k|$$

Proceeding in this way, future Leja points are chosen according to

$$z_n = \operatorname*{argmax}_z f^{(\alpha/2+1,\beta/2+1)}(z) \left(\prod_{\substack{k=0 \\ k \neq 1,2}}^{n-1} |z - z_k| \right)$$

In other words, we choose z_n as a Leja optimization with a new weight function $f^{(\alpha/2+1,\beta/2+1)}$, whose parameters are $\alpha/2 + 1 > 0$ and $\beta/2 + 1 > 0$. This effectively reduces the problem to the case where $\alpha, \beta > 0$. Then as before, this Leja sequence is asymptotically (unweighted) Fekete and so has empirical measure that converges to the arcsine measure μ.

7.3. Hermite case. In this section, $I = \mathbb{R}$ (and for shorthand write $\|\cdot\|_I = \|\cdot\|$) and $w(z) = z^{2\mu} \exp(-2|z|^\alpha)$. (Compared to Theorem 3.1, in this section we have redefined $z \leftarrow 2^{1/\alpha} z$ to make the computations cleaner.) The contraction factor is $k_n = n^{-1/\alpha}$. As usual, define $v(z) = \sqrt{w(z)}$. The limit weight $\widetilde{v}$ from Theorem 3.1 is $\widetilde{v}(z) = \exp(-|z|^\alpha)$. For this $\widetilde{v}$, it is known that the weighted Chebyshev constant is related to the weighted transfinite diameter by the relation

$$(28) \qquad\qquad \exp\left(-\frac{1}{2\alpha}\right) \tau_{\widetilde{v}} = \delta_{\widetilde{v}}.$$

See e.g., [34, 35]. We choose a weighted Leja sequence of points z_n according to (4). The negative log-weight of $\widetilde{v}$ is $Q = |z|^\alpha$, and it is a homogeneous function with homogeneity exponent α:

$$Q(cz) = c^\alpha Q(z), \qquad\qquad \forall\ c > 0, z \in \mathbb{C}$$

Then the following are easily verified: for any $n > 0$:

$$(29) \qquad\qquad \widetilde{v}\left(n^{1/\alpha}z\right) = \widetilde{v}^n(z), \qquad\qquad \widetilde{v}^n\left(n^{-1/\alpha}z\right) = \widetilde{v}(z)$$

Given the Leja sequence z_n from (4), then for each n we define progressively contracted versions of the grid:

$$z_{j,n} = n^{-1/\alpha}z_j, \qquad\qquad j, n = 1, \ldots$$

We will only need $z_{j,n}$ for $j \leq n$: we view $z_{j,n}$ as a triangular array of points with $j \leq n$. We note that one can easily transform between one contracted set of nodes and another:

$$z_{j,n} = \left(\frac{n-1}{n}\right)^{1/\alpha} z_{j,n-1},$$

which in turn implies that for our particular choice of family of weight functions:

$$\widetilde{v}^n\left(z_{j,n}\right) = \widetilde{v}^{n-1}\left(z_{j,n-1}\right) = \cdots = \widetilde{v}\left(z_{j,1}\right) = \widetilde{v}\left(z_j\right)$$

We wish to prove that the array $z_{j,n}$ is $\widetilde{v}$-weighted asymptotically Fekete. I.e., we wish to prove

$$(30) \qquad \lim_{n\to\infty}\left[\left|\det V\left(z_{0,n},\ldots,z_{n,n}\right)\right|\prod_{j=0}^{n}\widetilde{v}^n\left(z_{j,n}\right)\right]^{1/m_n} = \delta_{\widetilde{v}},$$

We first consider the case with the parameter $\mu > 0$. Using the explicit Vandermonde determinant formula (23) yields the following formula that we wish to prove:

$$(31) \qquad \lim_{n\to\infty}\left[\widetilde{v}^n(z_{0,n})\prod_{j=1}^{n}\widetilde{v}^n(z_{j,n})\prod_{k=0}^{j-1}\left|z_{j,n}-z_{k,n}\right|\right]^{1/m_n} = \delta_{\widetilde{v}}.$$

As with the Jacobi case, showing that the limit is $\leq \delta_{\widetilde{v}}$ is straightforward from the definition of $\delta_{\widetilde{v}}$, so we concentrate on the inequality $\geq$. We have

$$(32) \qquad \widetilde{v}^n(z_{0,n})\prod_{j=1}^{n}\widetilde{v}^n(z_{j,n})\prod_{k=0}^{j-1}\left|z_{j,n}-z_{k,n}\right| \overset{(29)}{=} \widetilde{v}(z_0)\prod_{j=1}^{n}n^{-j/\alpha}\underbrace{\widetilde{v}(z_j)\prod_{k=0}^{j-1}\left|z_j-z_k\right|}_{(a)}.$$

For the term (a), we note that $z_j \neq 0$ for $j \geq 1$ since $\mu > 0$. Therefore, we may write this term as

$$(a) = \underbrace{\left|z_j\right|^{-\mu}}_{(aa)}\underbrace{v(z_j)\prod_{k=0}^{j-1}\left|z_j-z_k\right|}_{(ab)}.$$

To compute lower bounds for terms (aa) and (ab) we will need the following notation: for the $\widetilde{v}$-weighted equilibrium measure, we have

$$(33) \qquad S = \operatorname{supp}\mu_{\widetilde{v}} = [-c,c], \qquad\qquad S_n = \left[-n^{1/\alpha}c, n^{1/\alpha}c\right].$$

The constant c is $2^{1/\alpha}b(\alpha)$, with $b(\alpha)$ given in the "Hermite" case of Table 1, or in Table 2.

To bound (aa), we know that z_j was computed from the optimization (4). This allows us to derive an upper bound for the magnitude of these weighted Leja points.

Lemma 7.2. *Let $M = \lceil\mu\rceil$. For each j, we have the following bound for term (aa):*

$$(34) \qquad \left|z_j\right|^{-\mu} \geq (M+j)^{-\mu/\alpha}c^{-\mu}$$

Proof. We make use of the following result [34] that compactifies the set on which the supremum norm of a weighted polynomial "lives" for our exponential weights:

$$(35a) \qquad \left\|\widetilde{v}(z)\left(z^n+q_{n-1}\right)\right\|_{\mathbb{R}} = \left\|\widetilde{v}(z)\left(z^n+q_{n-1}\right)\right\|_{S_n}, \qquad \forall\, q_{n-1}\in P_{n-1}$$

$$(35b) \qquad \left\|\widetilde{v}^n(z)\left(z^n+q_{n-1}\right)\right\|_{\mathbb{R}} = \left\|\widetilde{v}^n(z)\left(z^n+q_{n-1}\right)\right\|_{S}, \qquad \forall\, q_{n-1}\in P_{n-1}$$

where S and S_n are given by (33). Note that z_j is chosen as in (4) with

$$v(z_j)\prod_{k=0}^{j-1}\left|z_j-z_k\right| = \left\|v(z)\prod_{k=0}^{j-1}\left|z-z_k\right|\right\| = \left\|z^\mu\widetilde{v}(z)\prod_{k=0}^{j-1}\left|z-z_k\right|\right\|$$

If μ is an integer, then the argument under the norm is a $\widetilde{v}$-weighted monic polynomial of degree $\mu + j$. Thus, the extremum of the argument is achieved on the set $S_{\mu+j}$, which implies that the smallest-magnitude maximizer as stipulated in (4b) satisfies $z_j \in S_{\mu+j}$. The result (34) follows.

If μ is not an integer, then consider the function

$$f(z) = |z|^M \, \widetilde{v}(z) \prod_{k=0}^{j-1} |z - z_k|$$

We know that z_j is a maxmizer of $|z|^{\mu-M} f(z)$. Let z_* be the smallest-magnitude maximizer of $f(z)$. Suppose $|z_j| > |z_*|$; since $\mu - M < 0$ we have

$$|z_j|^{\mu-M} f(z_j) < |z_*|^{\mu-M} f(z_j) < |z_*|^{\mu-M} f(z_*),$$

which is a contradiction since z_j maximizes the norm of $z^{\mu-M} f(z)$. Therefore, $|z_j| \leq |z_*|$. But z_* maximizes f, which is a $\widetilde{v}$-weighted polynomial of degree $M + j$. Therefore, by (35a) we have $z_* \in S_{M+j}$. This in turn implies $z_j \in S_{M+j}$, and again (34) follows. $\square$

We likewise have a bound for the (ab) term:

Lemma 7.3. *For every j, we have the following bound for term (ab):*

$$(36) \qquad v(z_j) \prod_{k=0}^{j-1} |z_j - z_k| \geq \frac{j^{(j+\mu)/\alpha}}{c^{M-\mu}} \, (\tau_{\widetilde{v}})^{j+M} ,$$

where c is the constant from (33).

Proof. Since, by (4), z_j maximizes the left-hand side of (36), we are concerned with bounding

$$\left\| v(z) \prod_{k=0}^{j-1} |z - z_k| \right\| = \left\| z^\mu \widetilde{v}(z) \prod_{k=0}^{j-1} |z - z_k| \right\|.$$

We contract $z \leftarrow j^{-1/\alpha} z$ to obtain

$$(37a) \qquad \left\| z^\mu \widetilde{v}(z) \prod_{k=0}^{j-1} |z - z_k| \right\| = j^{(j+\mu)/\alpha} \left\| z^\mu \widetilde{v}^j(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\|.$$

We also have the following properties:

$$(37b) \qquad \min_{z \in S} z^{\mu-M} = c^{\mu-M}$$

$$(37c) \qquad \|\widetilde{v}\| = 1$$

Thus we have

$$
\begin{aligned}
v(z_j) \prod_{k=0}^{j-1} |z_j - z_k| &\overset{(4)}{=} \left\| v(z) \prod_{k=0}^{j-1} |z - z_k| \right\| \\
&\overset{(37a)}{=} j^{(j+\mu)/\alpha} \left\| z^{\mu} \widetilde{v}^{j}(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\| \\
&\geq j^{(j+\mu)/\alpha} \left\| z^{\mu} \widetilde{v}^{j}(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\|_{S} \\
&\overset{(37b)}{\geq} j^{(j+\mu)/\alpha} \frac{1}{c^{M-\mu}} \left\| z^{M} \widetilde{v}^{j}(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\|_{S} \\
&\overset{(37c)}{\geq} j^{(j+\mu)/\alpha} \frac{1}{c^{M-\mu}} \left\| z^{M} \widetilde{v}^{j+M}(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\|_{S} \\
&\overset{(35b)}{=} j^{(j+\mu)/\alpha} \frac{1}{c^{M-\mu}} \left\| z^{M} \widetilde{v}^{j+M}(z) \prod_{k=0}^{j-1} |z - z_{k,j}| \right\|_{\mathbb{R}} \\
&\overset{(25)}{\geq} j^{(j+\mu)/\alpha} \frac{1}{c^{M-\mu}} (\tau_{\widetilde{v}})^{j+M} .
\end{aligned}
$$

$\square$

With these two lemmas obtained, we can bound term (a) from (32):

$$
\begin{aligned}
(a) = |z_j|^{-\mu} v(z_j) \prod_{k=0}^{j-1} |z_j - z_k| \\
\overset{(34)}{\geq} \frac{1}{(M+j)^{\mu/\alpha} c^{\mu}} v(z_j) \prod_{k=0}^{j-1} |z_j - z_k| \\
\overset{(36)}{\geq} \frac{j^{(j+\mu)/\alpha}}{(M+j)^{\mu/\alpha} c^{M}} (\tau_{\widetilde{v}})^{j+M}
\end{aligned}
$$

We can therefore bound the entire weighted determinant from (32):

$$
\begin{aligned}
\widetilde{v}^{n}(z_{0,n}) \prod_{j=1}^{n} \widetilde{v}^{n}(z_{j,n}) \prod_{k=0}^{j-1} |z_{j,n} - z_{k,n}| &= \widetilde{v}(z_0) \prod_{j=1}^{n} n^{-j/\alpha} \underbrace{\widetilde{v}(z_j) \prod_{k=0}^{j-1} |z_j - z_k|}_{(a)} \\
&\geq \widetilde{v}(z_0) \prod_{j=1}^{n} n^{-j/\alpha} \frac{j^{(j+\mu)/\alpha}}{(M+j)^{\mu/\alpha} c^{M}} (\tau_{\widetilde{v}})^{j+M} \\
&= \widetilde{v}(z_0) \left[\prod_{j=1}^{n} \frac{j}{(M+j)} \right]^{\mu/\alpha} \left[\frac{\tau_{\widetilde{v}}}{c} \right]^{Mn} \left[\prod_{j=1}^{n} \left(\frac{j}{n} \right)^{j} \right]^{1/\alpha} (\tau_{\widetilde{v}})^{m_n}
\end{aligned}
$$

We now raise the result to the $1/m_n = \frac{2}{n(n+1)}$ power. The first two bracketed terms have the following limit:

$$\lim_{n\to\infty} \left[\prod_{j=1}^{n} \frac{j}{(M+j)} \right]^{\frac{2\mu}{\alpha n(n+1)}} = 1$$

$$\lim_{n\to\infty} \left[\frac{\tau_{\widetilde{v}}}{c} \right]^{\frac{2Mn}{n(n+1)}} = 1$$

The logarithm of the last bracketed term has the limit

$$\lim_{n\to\infty} \log \left[\left[\prod_{j=1}^{n} \left(\frac{j}{n} \right)^{j} \right]^{1/\alpha} \right]^{2/(n(n+1))} = \frac{2}{\alpha} \lim_{n\to\infty} \frac{1}{n+1} \sum_{j=1}^{n} \left(\frac{j}{n} \right) \log \left(\frac{j}{n} \right)$$

$$= \frac{2}{\alpha} \int_{0}^{1} x \log x \, \mathrm{d}x$$

$$= -\frac{1}{2\alpha}$$

Therefore, we have

$$\lim_{n\to\infty} \left[V(z_{0,n}, \ldots, z_{n,n}) \prod_{j=0}^{n} \widetilde{v}(z_{j,n}) \right]^{1/m_n} = \lim_{n\to\infty} \left[\widetilde{v}^n(z_{0,n}) \prod_{j=1}^{n} \widetilde{v}^n(z_{j,n}) \prod_{k=0}^{j-1} |z_{j,n} - z_{k,n}| \right]^{1/m_n}$$

$$\geq \tau_{\widetilde{v}} \exp \left(-\frac{1}{2\alpha} \right) = \delta_{\widetilde{v}},$$

and so we have proven (31), that the weighted Leja points are asymptotically weighted Fekete.

For the case $\mu < 0$, we may repeat arguments for the Jacobi $\alpha, \beta < 0$ case: $\mu < 0$ implies that node $z_1 = 0$ is chosen (assuming $z_0 \neq 0$), which then reverts the w-weighted Leja objective (4) to one where $\mu > 0$.

7.4. **Laguerre case.** The Laguerre result for weighted Leja sequences defined by (4) can easily be obtained by following the argument in Section 7.3, so we omit the details. We only mention that for weights of the form $\widetilde{v}(z) = \exp(-|z|)$ on $[0, \infty)$, we can directly obtain from (26):

$$\tau_{\widetilde{v}} = \sqrt{e} \delta_{\widetilde{v}}.$$

I.e., (28) holds with $\alpha = 1$ [33]. With this, the remainder of the proof follows in precisely the same fashion as the Hermite case.

References

[1] J. Baglama, D. Calvetti, and L. Reichel. Fast leja points. *Electronic Transactions on Numerical Analysis*, 7:124–140, 1998.

[2] R. E. Bellman. *Dynamic Programming*. Courier Dover Publications, 2003.

[3] J.-P. Berrut and L. N. Trefethen. Barycentric lagrange interpolation. *SIAM Review*, 46(3):501–517, January 2004.

[4] T. Bloom, L. Bos, C. Christensen, and N. Levenberg. Polynomial interpolation of holomorphic functions in $\mathbb{C}$ and $\mathbb{C}^N$. *Rocky Mountain Journal of Mathematics*, 22(2):441–470, June 1992.

[5] L. Bos, S. De Marchi, A. Sommariva, and M. Vianello. Computing multivariate fekete and leja points by numerical linear algebra. *SIAM Journal on Numerical Analysis*, 48(5):1984, 2010.

[6] J. P Boyd. The optimization of convergence for chebyshev polynomial methods in an unbounded domain. *Journal of Computational Physics*, 45(1):43–79, January 1982.

[7] J. P. Boyd. Chebyshev domain truncation is inferior to fourier domain truncation for solving problems on an infinite interval. *Journal of Scientific Computing*, 3(2):109–120, June 1988.

[8] J. P. Boyd. *Chebyshev and Fourier Spectral Methods: Second Revised Edition*. Dover Publications, 2 revised edition, December 2001.

[9] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13(-1):147–269, 2004.

[10] G. T. Buzzard. Efficient basis change for sparse-grid interpolating polynomials with application to t-cell sensitivity analysis. *Computational Biology Journal*, 2013, April 2013.

[11] D. Calvetti, G. H. Golub, W. B. Gragg, and L. Reichel. Computation of gauss-kronrod quadrature rules. *Mathematics of Computation*, 69(231):1035–1052., February 2000.

[12] J.-P. Calvi and P. Manh. Lagrange interpolation at real projections of leja sequences for the unit disk. *Proceedings of the American Mathematical Society*, 140(12):4271–4284, 2012.

[13] J.-P. Calvi and M. Phung Van. On the lebesgue constant of leja sequences for the unit disk and its applications to multivariate interpolation. *Journal of Approximation Theory*, 163(5):608–622, May 2011.

[14] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Foundations of Computational Mathematics*, 13(5):819–834, October 2013.

[15] S. De Marchi. On leja sequences: some results and applications. *Applied Mathematics and Computation*, 152(3):621–647, May 2004.

[16] A. Edrei. Sur les déterminants récurrents et les singularités d'une fonction donnée par son développement de taylor. *Composito Mathematica*, 7, 1940.

[17] V. V. Fedorov. *Theory of optimal experiments*. Academic Press, 1972.

[18] L. Fejér. Bestimmung derjenigen abszissen eines intervalles, fr welche die quadratsumme der grundfunktionen der lagrangeschen interpolation im intervalle ein mglichst kleines maximum besitzt. *Annali della Scuola Normale Superiore, Classe di Scienze*, 1(3):263–276, 1932.

[19] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, 2007.

[20] W. Gautschi. *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press, USA, June 2004.

[21] A. Genz and B. D. Keister. Fully symmetric interpolatory rules for multiple integrals over infinite regions with gaussian weight. *Journal of Computational and Applied Mathematics*, 71(2):299–309, July 1996.

[22] T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18(3):209–232, January 1998.

[23] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, SEP 2003.

[24] D. Gottlieb and S. A. Orszag. *Numerical Analysis of Spectral Methods : Theory and Applications*. Society for Industrial Mathematics, January 1987.

[25] C. E Grosch and S. A Orszag. Numerical solution of problems in unbounded regions: Coordinate transforms. *Journal of Computational Physics*, 25(3):273–295, November 1977.

[26] M. Hegland. Adaptive sparse grids. In K. Burrage and Roger B. Sidje, editors, *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001*, volume 44, pages C335–C353, April 2003.

[27] J.D. Jakeman and S.G. Roberts. Local and dimension adaptive stochastic collocation for uncertainty quantification. In Jochen Garcke and Michael Griebel, editors, *Sparse Grids and Applications*, volume 88 of *Lecture Notes in Computational Science and Engineering*, pages 181–203. Springer Berlin Heidelberg, 2013.

[28] J. Kautsky and G. H. Golub. On the calculation of jacobi matrices. *Linear Algebra and its Applications*, 52-53:439–455, July 1983.

[29] D. P. Laurie. Calculation of gauss-kronrod quadrature rules. *Math. Comput.*, 66(219):1133–1145, 1997.

[30] F. Leja. Sur certaines suites lies aux ensembles plans et leur application la reprsentation conforme. *Annales Polonici Mathematici*, 4(1), 1957.

[31] D. Lubinsky. A survey of weighted polynomial approximation with exponential weights. *Surveys in Approximation Theory*, 3:1–105, 2007.

[32] X. Ma and N. Zabaras. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *Journal of Computational Physics*, 229(10):3884–3915, May 2010.

[33] H. N. Mhaskar and E. B. Saff. Extremal problems for polynomials with laguerre weights. In C.K. Chui, L.L Schumaker, and J.D. Ward, editors, *Approximation Theory IV*, pages 619–624. Academic Press, New York, 1983.

[34] H. N. Mhaskar and E. B. Saff. Extremal problems for polynomials with exponential weights. *Transactions of the American Mathematical Society*, 285(1):203–234, September 1984.

[35] H. N. Mhaskar and E. B. Saff. Weighted analogues of capacity, transfinite diameter, and chebyshev constant. *Constructive Approximation*, 8(1):105–124, March 1992.

[36] A. Narayan and J.S. Hesthaven. Computation of connection coefficients and measure modifications for orthogonal polynomials. *BIT Numerical Mathematics*, 52(2):457–483, 2012.

[37] A. C. Narayan and J. S. Hesthaven. A generalization of the wiener rational basis functions on infinite intervals: Part iderivation and properties. *Mathematics of Computation*, 80:1557–1583, September 2011.

[38] P. G. Nevai and J. S. Dehesa. On asymptotic average properties of zeros of orthogonal polynomials. *SIAM Journal on Mathematical Analysis*, 10(6):1184–1192, 1979.

[39] L. W.-T. Ng and M. Eldred. Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. American Institute of Aeronautics and Astronautics, April 2012.

[40] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, January 2008.

[41] TNL Patterson. The optimum addition of points to quadrature formulae. *Mathematics of Computation*, 22(104):847–856, October 1968.

[42] K. Petras. Smolyak cubature of given polynomial degree with few nodes for increasing dimension. *Numerische Mathematik*, 93(4):729–753, February 2003.

[43] R. Pulch. Private communication.

[44] E. A. Rakhmanov. On asymptotic properties of polynomials orthogonal on the real axis. *Mathematics of the USSR-Sbornik*, 47(1):155, February 1984.

[45] E. Saff and V. Totik. *Logarithmic Potentials with External Fields*. Springer, Berlin, 1997.

[46] SA Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Dokl. Akad. Nauk SSSR*, volume 4, page 111, 1963.

[47] F. Stenger. Numerical methods based on whittaker cardinal, or sinc functions. *SIAM Review*, 23(2):165–224, April 1981.

[48] R. Taylor and V. Totik. Lebesgue constants for leja points. *IMA Journal of Numerical Analysis*, 30(2):462–486, April 2010.

[49] V. Totik. Fast decreasing polynomials via potentials. *Journal dAnalyse Mathematique*, 62(1):131–154, December 1994.

[50] J. L. Ullman. Orthogonal polynomials associated with an infinite interval. *The Michigan Mathematical Journal*, 27(3):353–363, 1980.

[51] J. A. C. Weideman and A. Cloot. Spectral methods and mappings for evolution equations on the infinite line. *Computer Methods in Applied Mechanics and Engineering*, 80(13):467–481, June 1990.

[52] D. Xiu. Fast numerical methods for stochastic computations: A review. 2009.

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

[54] D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, January 2005.

[55] L. Yan, L. Guo, and D. Xiu. Stochastic collocation algorithms using l1-minimization. *International Journal for Uncertainty Quantification*, 2(3):279–293, 2012.

[56] C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations, Proceedings of the Sixth GAMM-Seminar*, volume 31 of *Notes on Num, Fluid Mech.*, pages 241–251, Kiel, 1990. Vieweg-Verlag.