Clenshaw-Curtis Type Rules for Statistical Integrals

Hans-Werner van Wyk, Max Gunzburger, John Burkardt

November 16, 2016

Abstract

In statistics, many commonly encountered quantities take the form of density weighted integrals. This paper treats their numerical estimation within the Chebyshev approximation framework. In particular, we discuss how a generic one dimensional density function can be incorporated into the construction of Clenshaw-Curtis type quadrature rules, either through an adjustment of the quadrature weights or by generating a set of quadrature nodes that satisfies the optimal spacing property in terms of the density-weighted uniform error. We consider a variety of density functions, including those that are piecewise continuous, or have unbounded support. The accompanying numerical experiments illustrate the behavior and performance of the resulting quadrature rules and offer a comparison with a variety of existing approaches for estimating density weighted integrals.

1 Introduction

For physical systems that are modeled by differential- or partial differential equations, but whose input parameters can only be determined in a statistical sense, stochastic simulations provide a means of computing statistical quantities of interest related to the model response, such as its expectation or variance. Usually, the uncertain input parameters can be approximated as functions of some random vector $\mathbf{Y}$ whose distribution over its range Γ is determined by a joint density function $\rho : \Gamma \to [0, \infty)$. The statistical quantity of interest Q then takes the form

$$Q = \int_\Gamma f(\mathbf{y})\rho(\mathbf{y})d\mathbf{y}, \tag{1}$$

where $f(\mathbf{y})$ is related to the system's response to a given realization of the random input. To be effective, numerical quadrature schemes that approximate Q should not only be able to accommodate the potential high dimensionality and/or unboundedness of the stochastic domain Γ, but should also aim to minimize the number of function evaluations $f(\mathbf{y})$, since these involve computing the numerical solution of a partial differential equation at a substantial computational cost.

Unlike f, the density function ρ is usually known *a priori* and the comparative cost of its evaluation is negligible. Moreover, the full incorporation of this readily available information into the construction of a quadrature rule will almost certainly improve its accuracy. Gaussian quadrature rules can accommodate the density ρ naturally through construction of the orthogonal interpolating basis [1, 25]. This paper treats the interpolation and numerical quadrature of such weighted integrals within the *Chebyshev* approximation framework. In particular we discuss two modifications of the standard Clenshaw-Curtis type rules, through either an adjustment of the quadrature weights, or through the placement of quadrature nodes, that lead to the efficient approximation of ρ-weighted integrals for a general set of densities ρ.

Although this paper is motivated in part by applications in uncertainty quantification, our results are also applicable in a variety of other settings, including optimization and control, mathematical finance, and reduced order modeling, where interpolation and numerical quadrature in the presence of density functions play an important role, for instance in the construction of surrogate models.

We restrict our attention to integrals of functions that depend on a single random variable Y whose distribution over the interval Γ is determined by a generic density function $\rho : \Gamma \to [0, \infty)$. For multidimensional integrals over hyperrectangles, i.e. $\Gamma = \prod_{k=1}^{K} \Gamma_k$ where Γ_k is an interval in $\mathbb{R}$ for $k = 1, \ldots, K$, efficient sparse grid quadrature rules ([7, 16]) can be constructed from one dimensional rules over each interval Γ_k. If in addition the random vector $\mathbf{Y} = [Y_1, ..., Y_K]^T$ has independent components, then $\rho(\mathbf{y}) = \prod_{k=1}^{K} \rho_k(y_k)$ for any point $\mathbf{y} = (y_1, ..., y_K) \in \Gamma$, and hence the sparse grid's constituent one dimensional rules can be defined with respect to the weighted measure $\rho_k dy_k$. The growth in the number of sparse grid quadrature nodes with increasing dimension places a limit on the order of its constituent one dimensional rules. In this work we therefore consider only moderately sized rules with up to about 30 nodes, for which no stability issues arise.

Let $\Gamma = [a, b]$ be a bounded interval, $\rho : [a, b] \to [0, \infty)$ be a piecewise continuous probability density function, and $f : [a, b] \to \mathbb{R}$ be continuous. Conditional on the change of variable $y = \frac{b+a}{2} + \frac{b-a}{2}x$, we may assume without loss of generality that

the integral we seek to approximate is of the form

$$Q = \int_{-1}^{1} f(x)\rho(x)dx. \tag{2}$$

We first discuss the case $\rho(x) = 1$. Let $I_n[f]$ denote the order $(n-1)$ Lagrange interpolant of f over a given set of nodes $\{x_i\}_{i=1}^n$, i.e.

$$I_n[f](x) := \sum_{i=1}^{n} f(x_i)l_i(x), \tag{3}$$

where each cardinal basis function $l_i(x)$ is the uniquely defined order $(n-1)$ polynomial satisfying $l_i(x_j) = \delta_{ij}$ for $i, j = 1, \ldots, n$ (δ_{ij} is the Kronecker delta function). The corresponding quadrature rule then takes the form

$$\int_{-1}^{1} f(x)dx \approx \int_{-1}^{1} I_n[f](x)dx = \sum_{i=1}^{n} w_i f(x_i), \tag{4}$$

where $w_i := \int_{-1}^{1} l_i(x)dx$ represents the quadrature weights for $i = 1, \ldots, n$. Clenshaw-Curtis (CC) type rules are special interpolatory quadrature rules associated with Chebyshev polynomials. Among many other useful properties, the Clenshaw-Curtis rules of order $2^k + 1$, $k = 0, 1, \ldots$ are nested, allowing for the re-use of quadrature nodes under refinement.

Various strategies exist for incorporating a nonuniform density function ρ. A simple method for generating a CC-type quadrature rule for (2) for non-uniform ρ, is to form the interpolant $I_n[f\rho](x)$, leading to the approximation

$$\int_{-1}^{1} f(x)\rho(x)dx \approx \sum_{i=1}^{n} w_i f(x_i), \tag{5}$$

where $w_i = \rho(x_i)\int_{-1}^{1} l_i(x)dx$. While the multiplication of f by ρ may have a mollifying influence on the integrand, the extent of this effect is impossible to predict if the precise form of f is unknown *a priori*. Moreover, using the same order of approximation for f and ρ does not exploit the ready accessibility of ρ.

Another common approach to approximating (2) involves first transforming it to the unit interval $[0, 1]$ through the use of the cumulative density function $F(x) := \int_{-1}^{x} \rho(\tau)d\tau$. Let $y = F(x)$ and note that $\frac{dF}{dx}(x) = \rho(x)$, yielding

$$\int_{-1}^{1} f(x)\rho(x)dx = \int_{0}^{1} \tilde{f}(y)dy, \tag{6}$$

where $\tilde{f}(y) = f(F^{-1}(y))$. This transform ostensibly eliminates $\rho(x)$ from the integral, and allows for the use of standard Clenshaw-Curtis or Fejér rules to approximate (6). However, the integrand $\tilde{f}$ that replaces f depends implicitly on ρ. The inverse cumulative distribution function (ICDF) F^{-1} increases sharply in regions where ρ is small, with jump discontinuities over intervals where $\rho(x) = 0$, and since $\frac{d\tilde{f}}{dy}(y) = f'(F^{-1}(y))\frac{dF^{-1}}{dy}(y)$, this can result in sharp increases in $\tilde{f}$, which has a direct bearing on its integrability (see Equation (7)).

In the following sections we outline two alternative approaches to generating CC-type numerical quadrature rules for the ρ-weighted integral (2). The first method, discussed in Section 2, uses the standard CC-type quadrature nodes described above but adjusts the quadrature weights to account for the density function ρ. The second method, treated in Section 3, is based on a generalization of the optimal spacing property of Chebyshev nodes to ρ-weighted uniform seminorm. Unlike in the case of Chebyshev nodes, no explicit formulas for such nodes exist in this generalized setting. We use the Remez algorithm to approximate them.

2 Incorporating the Density via Quadrature Weights

The following three variants of quadrature rules based on Clenshaw-Curtis type nodes are commonly used: the Clenshaw-Curtis (CC) rule (see [8]), Fejér's first rule (F_1), and Fejér's second (F_2) rule (see [11]), whose nodes are given respectively by $x_i^{cc} = \cos(\frac{(i-1)\pi}{n-1})$, $x_i^{f_1} = \cos(\frac{(i-\frac{1}{2})\pi}{n})$, and $x_i^{f_2} = \cos(\frac{i\pi}{n+1})$ for $i = 1, \ldots, n$. For each of these sets of nodes, the Lagrange interpolant in 3 can also be expressed in terms the first or second Chebyshev polynomials, given by $T_k(x) := \cos(k \arccos(x))$ and $U_k(x) := \sin((k+1)\arccos(x))/\sqrt{1-x^2}$ respectively, in a way that lays bare its connection with trigonometric interpolation. Indeed, the change of coordinates $x = \cos\theta$ renders the function $f(x) = f(\cos\theta)$ periodic over the interval $[0, \pi]$, allowing it to be readily approximated by trigonometric series. For example, the interpolant $I_n[f](x)$ over the CC nodes, can then be written in terms of θ as

$$I_n^{cc}[f](\cos\theta) = \sum_{k=0}^{n-1} a_k^{cc} \cos(k\theta) = \sum_{k=0}^{n-1} a_k^{cc} T_k(x),$$

where the k^{th} coefficient a_k^{cc} is computed as the trapezoidal approximation of the projection of $f(\cos\theta)$ onto $\cos(k\theta)$ (see [6]). Similarly, $I_n^{f_1}[f](x) = \sum_{k=0}^{n-1} a_k^{f_1} T_k(x)$ and $I_n^{f_2}[f] = \sum_{k=0}^{n-1} a_k^{f_2} U_k(x)$, where $a_k^{f_1}$ and $a_k^{f_2}$ are computed analogously for $k = 0, \ldots, n-1$. In [12] and [24], the authors exploit these relations to construct fast

and stable methods for computing CC type weights through the use of fast Fourier transforms (FFTs).

Weighted CC-type rules are constructed by interpolating only f at the CC-type nodes, which leads to the approximation

$$\int_{-1}^{1} f(x)\rho(x)dx \approx \int_{-1}^{1} I_n[f]\rho(x)dx = \sum_{i=1}^{n} w_i^{\rho} f(x_i),$$

where $w_i^{\rho} := \int_{-1}^{1} \rho(x) l_i(x)dx$ for $i = 1, \ldots, n$. This formulation effectively decouples the approximation of ρ from that of f. The quadrature weights w_i^{ρ} can be evaluated independently of f, either analytically (if possible), or through the use of unrelated high order quadrature methods taylored to ρ, including specialized methods, such as when ρ is highly oscillatory [10] or has an integrable singularity. Weighted rules also boast convergence rates that resemble those of CC rules for integrals with $\rho = 1$, as well as non-negative quadrature weights if $\rho(x) \geq 0$ for $x \in \Gamma$. In [23], the author generalizes FFT type the results of [24] to weighted CC-type rules. In this case, the weights can be computed using discrete Fourier transforms, based on the density's moments $\gamma_i = \int_{-1}^{1} \rho(x) T_i(x)dx$ for $i = 1, \ldots, n-1$.

To compute weighted rules more or less automatically for a wide array of different density functions, we make use of Chebyshev approximations in estimating both the density function ρ and the function f, albeit to different degrees. Let $P_k(x)$ be a generic Chebyshev polynomial of order k (CC, F_1, or F_2). We now expand ρ using the Chebyshev polynomials, while using the canonical basis to interpolate f, i.e.

$$I_{n_\rho}[\rho](x) = \sum_{k=0}^{n_\rho - 1} a_k^{\rho} P_k(x), \quad \text{and} \quad I_{n_f}[f](x) = \sum_{i=1}^{n_f} f(x_i) l_i(x).$$

Replacing f and ρ in (2) with their respective approximations yields the quadrature rule

$$\int_{-1}^{1} f(x)\rho(x)dx \approx \sum_{i=1}^{n_f} w_i^{\rho} f(x_i),$$

where the quadrature weight $w_i^{\rho} = \sum_{k=0}^{n_\rho - 1} a_k^{\rho} \int_{-1}^{1} l_i(x) P_k(x)dx$ can be computed exactly for $i = 1, ..., n$, either by an interpolatory quadrature rule of sufficiently high order, or explicitly. Indeed, expressing the canonical basis functions $l_i(x)$ as linear combinations of Chebyshev polynomials of either the first or second type reduces the evaluation of the integrals $\int_{-1}^{1} l_i(x) P_k(x)\ dx$ for $k = 0, ..., n-1$ and $i = 1, ..., n$ to integrals of the form $\int_{-1}^{1} P_{k'}^{f}(x) P_k^{\rho}(x)\ dx$, where P_k^f and P_k^{ρ} are the order k Chebyshev polynomials used to approximate f and ρ respectively. Explicit formulas for

the integrals of such products can be obtained by combining the known integrals of $T_k(x)$ and $U_k(x)$ with the product formulas [6]

$$\begin{aligned}
T_k(x)T_{k'}(x) &= \frac{1}{2}\left(T_{k-k'}(x) + T_{k+k'}(x)\right) \\
U_k(x)U_{k'}(x) &= \sum_{l=0}^{n} U_{k-k'+2l}(x) \\
T_k(x)U_{k'}(x) &= \begin{cases} \frac{1}{2}(U_{k+k'}(x) + U_{k-k'}(x)), & \text{if } k' \geq k-1 \\ \frac{1}{2}(U_{k+k'}(x) - U_{k'-k-2}(x)), & \text{if } k' \leq k-2 \end{cases},
\end{aligned}$$

where $U_{-1}(x) = 0$.

Example 1. In Figure 1, we compare the accuracy of the weighted Clenshaw-Curtis rule (the Fejér rules behave similarly) with approximations (6) and (5). For all three functions f, the weighted rule exhibits the same convergence behavior as the standard CC rule applied to the $\rho = 1$ case (included here as a reference), while the accuracy for both the ICDF transform and the product interpolant vary significantly. This confirms that, while f may be unknown, the influence of ρ on the error can be mitigated. A list of the density functions we use in this paper, is included in Appendix A.

2.1 Piecewise Continuous Densities

The above method can readily be extended to piecewise continuous densities. In particular, suppose the standard interval $[-1, 1]$ can be partitioned into sub-intervals $\{[a_j, b_j]\}_{j=1}^{m}$ so that ρ is continuous over each (a_j, b_j) for $i = 1, \ldots, m$. We can then form the order n_j^ρ Clenshaw-Curtis type interpolant of ρ over each sub-interval, giving rise to the piecewise defined polynomial interpolant

$$I_{n_\rho}^{\text{pw}}[\rho](x) = \sum_{j=1}^{m} \sum_{k=0}^{n_\rho - 1} a_{k,j}^{\rho} \mathbb{1}_{[a_j,b_j]}(x) P_{k,j}(x),$$

where $\mathbb{1}_{[a_j,b_j]}$ is the characteristic function over $[a_j, b_j]$, $j = 1, \ldots, m$. Combining this approximation with the same standard polynomial interpolation for f as before, we obtain

$$\int_{-1}^{1} f(x)\ \rho(x)\ dx \approx \int_{-1}^{1} I_{n_f}[f](x)\ I_{n_\rho}^{\text{pw}}[\rho](x)\ dx = \sum_{i=1}^{n_f} f(x_i) w_i,$$

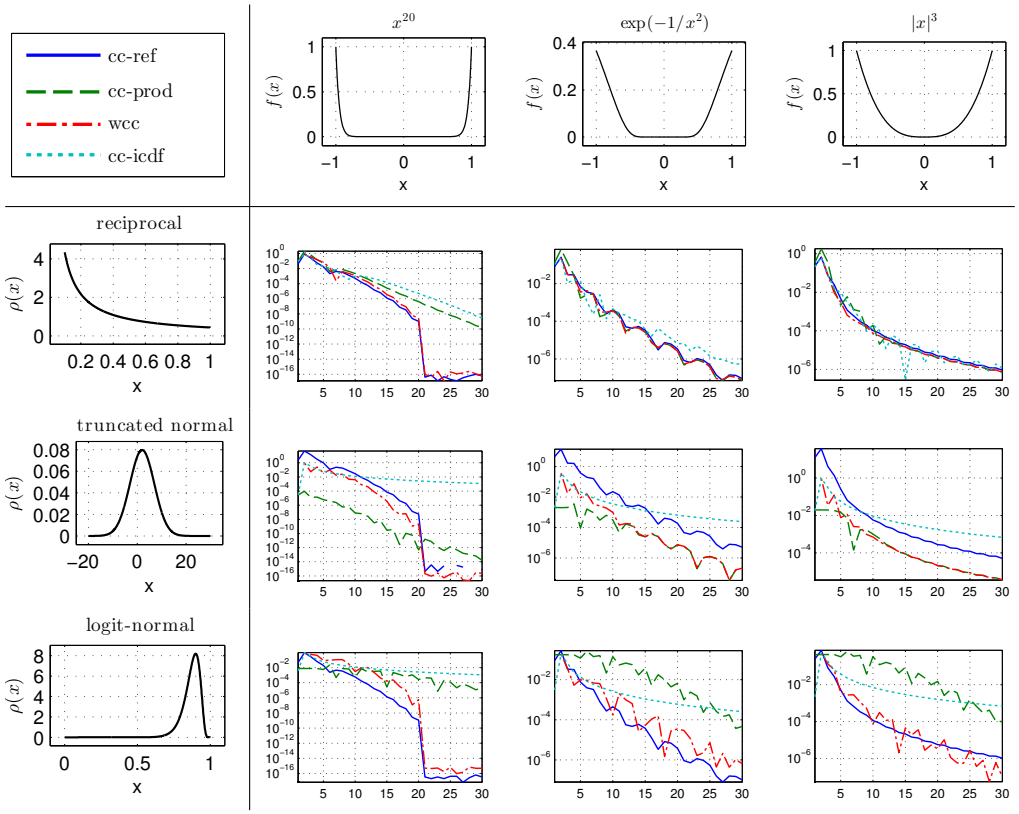


Figure 1: The absolute errors in approximating the integral (2) for different numbers of nodes, using either i) the weighted Clenshaw-Curtis rule (wcc), or applying the standard Clenshaw-Curtis rule to ii) the interpolant of the product (cc-prod), or iii) the ICDF transformed integral (cc-icdf). As a reference, we include the error associated with the standard CC rule applied to the associated $\rho = 1$ case (cc-ref).

where

$$w_i = \sum_{j=1}^{m} \sum_{k=0}^{n_\rho} a_{k,j}^{\rho} \int_{a_j}^{b_j} l_i(x) P_{k,j}(x)\, dx.$$

Again, the weights require the computation of integrals of products of pairs of Chebyshev polynomials, although in this case the factors are defined over different intervals. Evaluating these integral ultimately amounts to computing integrals of the form

$$\int_{a_j}^{b_j} P_{k'}(x)\, P_{k,j}(x)\, dx, \quad \text{for } k, k' = 0, \ldots, n-1.$$

This can be achieved by simply invoking the appropriate Gaussian quadrature rule.

Example 2. The benefits of using a weighted Clenshaw-Curtis rule, as opposed to approaches (5) or (6), become abundantly clear when considering piecewise continuous or -smooth density functions. In this case, the discontinuities in the density (or in its derivatives) can be treated explicitly, once again ensuring a convergence rate in accordance with the $\rho = 1$ case (see Figure 2).

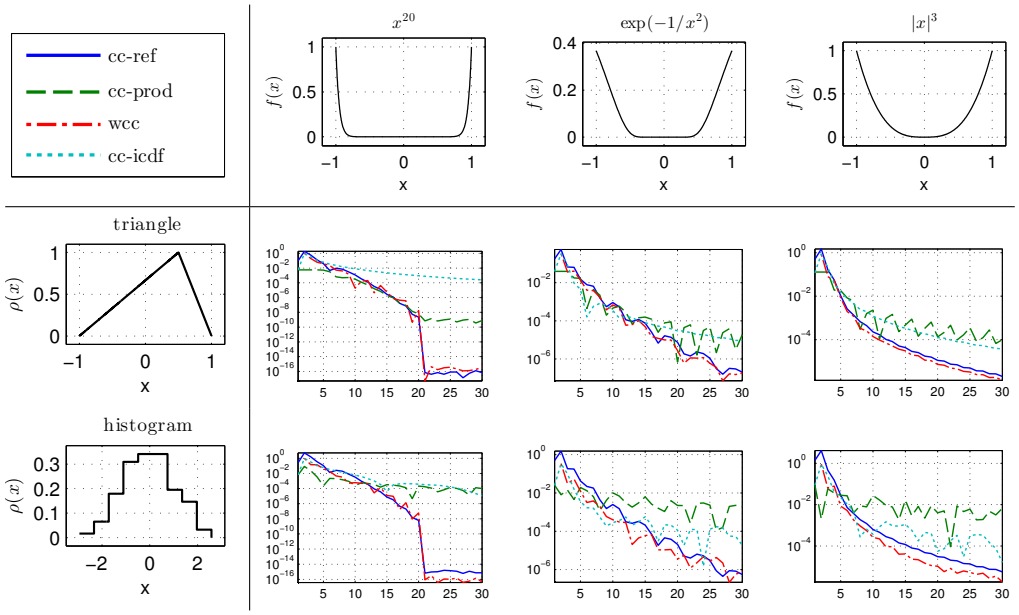


Figure 2: Absolute errors for piecewise defined density functions.

2.2 Unbounded Intervals

The construction of quadrature rules over unbounded domains, such as the infinite interval $\Gamma = (-\infty, \infty)$ or the semi-infinite one $\Gamma = [0, \infty)$, is usually accomplished either through the use of a polynomial basis that is intrinsic to the domain, such as Hermite or Laguerre polynomials, or by first transforming the unbounded interval to a bounded one, where a standard polynomial basis, such as Chebyshev polynomials, can be used. The ICDF is one such mapping that is specific to ρ. Applying its inverse to a set of quadrature nodes in $[0, 1]$ naturally spreads them over the unbounded interval in accordance with ρ. However, the caveats of this approach mentioned in Section 1 still apply.

Here, we make use of well-known family of algebraic mappings [3, 5, 4], defined for the semi-infinite domain by $\zeta(x) = L(1 + x)/(1 - x)$, $x \in [-1, 1)$, and for the

infinite domain by $\eta(x) = Lx/\sqrt{1-x^2}$, $x \in (-1,1)$, where $L > 0$ represents a scaling parameter that can be calibrated to different problems (we will use $L = 1$ throughout). These mappings give rise to rational basis functions over the unbounded interval that, among other useful properties, are preferrable to Hermite polynomials when interpolating functions that decay algebraically at infinity (see e.g. Chapter 17, [6]). This approach was employed in the Chebfun package [22], for the treatment integrals over unbounded intervals. To illustrate the construction of weighted CC-type quadrature rules over unbounded domains, we focus on the semi-infinite interval, the unbounded interval being similar. The change of variables $y = \zeta(x)$ yields

$$\int_0^\infty f(y)\rho(y)dy = \int_{-1}^1 f(\zeta(x))\frac{L\rho(\zeta(x))}{(1-x)^2}dx.$$

If the mapped density $\rho(\zeta(x))$ is bounded and converges to 0 at least quadratically as $x \to 1^-$, we can use techniques described above to compute the appropriate weights by evaluating $w_i^\rho = \int_{-1}^1 l_i(x)\frac{L\rho(\zeta(x))}{(1-x)^2}dx$, or approximating this integral numerically. Since the mapped density $\rho(\zeta(x))$ and the Jacobian of $\zeta(x)$ are non-negative, the resulting quadrature weights are guaranteed to be positive.

Another approach, which we do not discuss here, is to simply truncate the interval, i.e. disregard part of the domain beyond a given cut-off. If the integrand decays rapidly as $y \to \pm\infty$, the truncation will have a negligible effect on the error.

Example 3. For smooth functions, the weighted CC rule does as well as the case $\rho = 1$. For functions with convergence-limiting singularities such as $f(x) = \sin(x)/x$ and $1/(1+x^2)^{1/3}$, incorporating the density function improves accuracy, provided the density's support is concentrated away from the singularity, as is the case for the Weibull- or Pareto densities, but not for the Gamma(7,1) density. Here it becomes apparent how the introduction of a density function generally has a mollifying effect on the integrand.

3 Incorporating the Density through Node Placement

As the discussion in the previous section shows, a simple modification of the quadrature weights of traditional CC-type quadrature rules allows us to estimate density weighted integrals with almost the same accuracy as the corresponding rules for $\rho = 1$ and with little additional cost. Choosing the quadrature nodes independently of the density ρ may however be problematic, potentially resulting in a significant

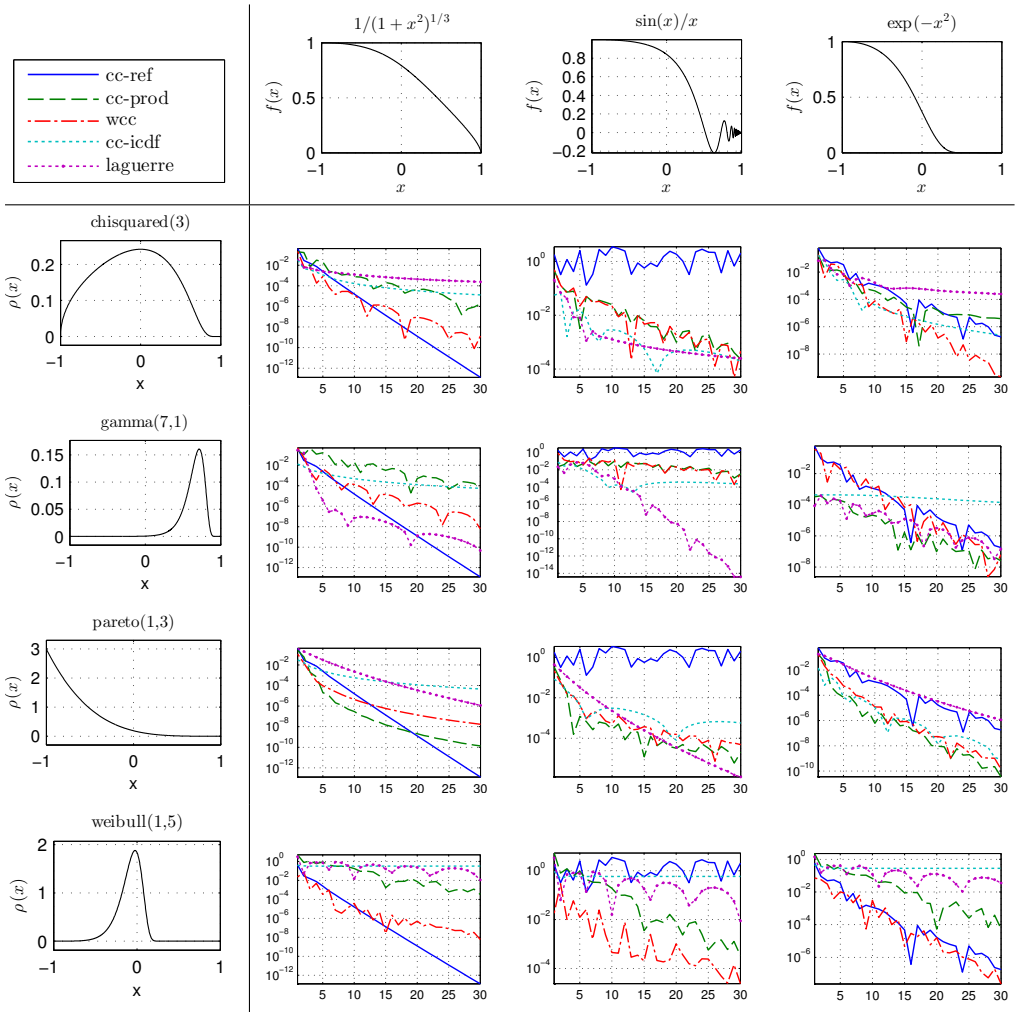


Figure 3: Absolute errors in approximating (1) for density functions with unbounded support.

number of quadrature points that lie in regions where ρ is small or even zero, i.e. where the random variable is unlikely or even impossible to occur. This in turn, can lead to evaluations of the integrand f that are costly, vis-à-vis their contribution to the overall accuracy, or at worst, points at which f is not even defined.

In this section we construct a set of nodes whose spread reflects the distribution of the underlying random variable, by seeking to reproduce the optimal spacing property of Clenshaw Curtis nodes (see Theorem 4.12, in [2]) for a ρ-weighted L^∞ seminorm error. Unlike in the unweighted case, no explicit formulas exist for the

positions of these optimal nodes. Here we resort to a weighted form of the well-known Remez algorithm ([19]) to approximate them numerically.

Let $\Gamma = [a, b]$ be a closed bounded interval and consider the interpolation of $f \in C^n([a, b])$ by the polynomial $p \in \mathcal{P}^{n-1}([a, b])$ through a set of points $\{x_i\}_{i=1}^n$. Recall that the approximation error is given explicitly by the expression

$$f(x) - p(x) = \frac{f^{(n)}(\xi_x)}{n!} \prod_{i=1}^{n} (x - x_i) \tag{7}$$

for some $\xi_x \in [a, b]$. As is readily apparent from (7), a point set $\{x_i\}_{i=1}^n$ that is optimally spaced in the sense of minimizing the quantity $\Theta(x; \{x_1, \ldots, x_n\}) := \prod_{i=1}^n (x - x_i)$ uniformly, i.e.

$$\min_{x_1,\ldots,x_n} \|\Theta(\cdot; \{x_1, \ldots, x_n\})\|_\infty \tag{8}$$

is likely to produce an interpolant with good uniform accuracy for a *generic*, sufficiently differentiable function f. The set of points that satisfy (8) are precisely the Clenshaw-Curtis points.

To reflect the presence of the density ρ, we propose measuring the approximation error $f(x)-p(x)$ in the ρ-weighted uniform semi-norm $|\cdot|_{\rho,\infty}$, which we define for any $f \in C([a, b])$ by $|f|_{\rho,\infty} := \sup_{[a,b]} |f(x)|\rho(x)$. In regions where the random variable is unlikely to lie, $\rho(x)$ is small and therefore the mismatch between p and f is less prone to affect the overall error, whereas regions in which $\rho(x)$ is relatively large carry a more substantial weight. Throughout we allow ρ to be piecewise continuous over $[a, b]$ but assume that it is also bounded above, which ensures $|f|_{\rho,\infty} < \infty$ for all $f \in C([a, b])$. Since modifying ρ at the points of discontinuity does not affect its underlying probability measure (and hence (1)), we may assume that $|f(x)\rho(x)|$ attains both of its global extrema, either within the sub-intervals in which ρ is continuous or at one of the one-sided limits at the sub-interval endpoints. The fact that $\int_a^b \rho(x)dx = 1$ guarantees $\rho(x) > 0$ within some open interval in $[a, b]$. Note that unless ρ is strictly positive over the entire interval $[a, b]$, the condition $|f|_{\rho,\infty} = 0$ does not necessarily imply $f = 0$ in which case $|\cdot|_{\rho,\infty}$ does not define a norm on $C([a, b])$. It does however define a norm on the space $\mathcal{P}^n([a, b])$.

Multiplying both sides of (7) by $\rho(x)$ yields the weighted difference

$$(f(x) - p(x))\rho(x) = \frac{f^{(n)}(\xi_x)}{n!} \left(\prod_{i=1}^{n} (x - x_i) \right) \rho(x). \tag{9}$$

As before, we seek a set of nodes that are optimally spaced, albeit in a weighted sense. Unlike for the unweighted case, the minimization problem

$$\min_{x_1,\ldots,x_n} |\Theta(\cdot;\{x_1,\ldots,x_n\})|_{\rho,\infty} = \min_{x_1,\ldots,x_n} \max_{x\in[a,b]} \rho(x) \prod_{i=1}^{n} |x - x_i| \tag{10}$$

no longer necessarily has a closed form solution and must be solved computationally. To this end, we reformulate (10) as a best polynomial approximation problem in the seminorm $|\cdot|_{\rho,\infty}$ and estimate its solution by means of a modified form of the Remez algorithm (see [19, 20, 21]). Recall the traditional best minimax approximation problem for a function $f \in C([a,b])$: Find $p^* \in \mathcal{P}^n([a,b])$ so that

$$\|f - p^*\|_\infty \leq \|f - p\|_\infty \text{ for all } p \in \mathcal{P}^n([a,b]). \tag{11}$$

It is readily apparent that solving problem (8) is equivalent to finding the best minmax approximation of a monic polynomial in $\mathcal{P}^n([a,b])$ by polynomials in $\mathcal{P}^{n-1}([a,b])$.

A unique solution p^* of (11) is known to exist and is characterized, by virtue of the Chebyshev equioscillation theorem, by the existence of $n+2$ equioscillation points $z_1,\ldots,z_{n+2}$ for which

$$f(z_i) - p^*(z_i) = \sigma(-1)^i \inf_{p\in\mathcal{P}_n} \|f - p\|_\infty \tag{12}$$

holds, where $\sigma = \pm 1$, depending on f and n. In a seminal series of papers, Remez developed an iterative algorithm to estimate this set of equioscillation points. The method received widespread interest in the 1950's with the increased availability of computational resources and became a fundamental tool in digital signal processing and filter design in the 1970's, forming the basis for the Parks McClellan algorithm [18]. In [18], the authors relied on a weighted version of the equioscillation theorem for trigonometric polynomials. To the best of our knowledge, no proof of the density-weighted equioscillation theorem has been given for regular polynomials. In the following sections we extend the weighted equioscillation theorem to polynomials and use the result as the foundation for a weighted version of the Remez algorithm for polynomials. The paper [17] provides an excellent introduction to the Remez algorithm and its approximation through barycentric interpolation in the unweighted case.

3.1 Best Approximation in the Weighted Uniform Error

The ρ-weighted best approximation problem for $f \in C([a,b])$ is formulated as follows: Find $p^* \in \mathcal{P}^n([a,b])$ so that

$$|f - p^*|_{\rho,\infty} = \inf_{p\in\mathcal{P}^n} |f - p|_{\rho,\infty}. \tag{13}$$

In the following, we generalize the Chebyshev equioscillation theorem to characterize the solution of (13). For notational convenience, we let $d_n := \inf_{p\in\mathcal{P}^n} |f-p|_{\rho,\infty}$. Since ρ is bounded above, the embedding $(C([a,b]), \|\cdot\|_\infty) \hookrightarrow (C([a,b]), |\cdot|_{\rho,\infty})$ is continuous, allowing us to readily extend the Weierstrass approximation theorem to $|\cdot|_{\rho,\infty}$. We set out to prove the following theorem.

Theorem 1 (Weighted Equioscillation Theorem)**.** *For any $f \in C([a,b])$ and $n \geq 0$, there exists a unique minimizer $p^* \in \mathcal{P}^n([a,b])$ of (13). Moreover, p^* is uniquely characterized by the existence of a sequence of at least $n+2$ uniformly alternating points $a \leq z_1 < \ldots < z_{n+2} \leq b$ satisfying,*

$$(f(z_j) - p^*(z_j))\rho(z_j) = \sigma(-1)^j |f-p^*|_{\rho,\infty} \quad \textit{for} \quad j = 1,\ldots,n+2, \tag{14}$$

where $\sigma = \pm 1$, depending on f and n.

Our proof follows the standard arguments (see [14]). After establishing the existence of minimizer, we prove a version of the de la Valeé-Poussin theorem, which asserts that a uniformly alternating sequence gives rise to the best approximating polynomial. To prove the converse, we first show that the optimal polynomial approximant generates a uniformly alternating set of length 2 and then demonstrate that unless the set is of length $n+2$ the approximation can always be improved. Uniqueness is trivial.

Lemma 1. *For every $n \in \mathbb{N}$, there exists a polynomial $p^* \in \mathcal{P}^n([a,b])$, so that*

$$|f-p^*|_{\rho,\infty} = d_n$$

Proof. The proof is almost identical to that in [14]. We omit it here in the interest of brevity. □

The de la Valée-Poussin Theorem, which relates the best approximation error d_n to the oscillations of any (not necessarily uniform) alternating sequence, readily extends to the weighted problem.

Theorem 2 (de la Valée-Poussin)**.** *Let p be any order n polynomial and let $a \leq z_1 < \ldots < z_{n+2} \leq b$ be any alternating sequence, i.e. points satisfying*

$$(f(z_i) - p(z_i))\rho(z_i) = (-1)^i e_i, \ i = 1,2,\ldots,n+2, \tag{15}$$

where the e_i's are either all strictly negative or all strictly positive. Then

$$\min_{1\leq i\leq n+2} \rho(z_i)|f(z_i)-p(z_i)| \leq d_n \leq \max_{1\leq i\leq n+2} \rho(z_i)|f(z_i)-q(z_i)| \tag{16}$$

for any other polynonial $q \in \mathcal{P}^n([a,b])$.

Proof. Suppose to the contrary that $d_n < |e_i|$ for all $i = 1, \ldots, n+2$. By Lemma 1, there then exists a polynomial $q \in \mathcal{P}^n([a,b])$ satisfying $|f-q|_{\rho,\infty} = d_n < |e_i|$. Let $r(x) = p(x) - q(x)$ and suppose $e_i > 0$, the case $e_i < 0$ being similar. Then

$$\begin{aligned} r(z_i)\rho(z_i) &= (f(z_i) - q(z_i))\rho(z_i) - (f(z_i) - p(z_i))\rho(z_i) \\ &= (f(z_i) - q(z_i))\rho(z_i) - (-1)^{i+1}e_i, \end{aligned}$$

which implies $r(x)\rho(x)$, and hence $r(x)$, changes signs $n+2$ times. This means the polynomial r has $n+1$ zeros and must therefore be identically zero, which implies $q(x) = p(x)$ on $[a,b]$. □

As a consequence, if the sequence $\{z_i\}_{i=1}^{n+2}$ in Theorem 2 is *uniformly* alternating with levelled error $e > 0$, it follows

$$d_n \leq |f-q|_{\rho,\infty} \leq \left(\frac{|f-q|_{\rho,\infty}}{e}\right) d_n, \tag{17}$$

where $\frac{|f-q|_{\rho,\infty}}{e} \geq 1$. If moreover $e = |f-q|_{\rho,\infty}$, then $|f-q|_{\rho,\infty} = d_n$ and hence q minimizes (13). This proves Theorem 1 in one direction. It now remains to show that this minimizer gives rise to a uniformly alternating sequence of length $n+2$ that satisfies (14). We first prove that there exists such a sequence of length 2.

Lemma 2. *If $p \in \mathcal{P}^n([a,b])$ minimizes* (13)*, then there exists a uniformly alternating sequence of length 2, with uniform levelled error $e = |f-p|_{\rho,\infty}$.*

Proof. Under the assumptions made above, the expression $(f(x)-p(x))\rho(x)$ achieves both its minimum m_0 and its maximum m_1 respectively at some points z_1 and z_2 in $[a,b]$. We show that z_1, z_2 forms an alternating sequence for f, p by proving that $m_0 = -m_1$. Assume to the contrary that $m_0 + m_1 \neq 0$. We define the new polynomial $q(x) = p(x) - \frac{m_0+m_1}{2\rho_{\max}}$, where $\rho_{\max}$ is the maximum of ρ on $[a,b]$, and show that $|f-q|_{\rho,\infty} < |f-p|_{\rho,\infty}$, contradicting the fact that p is the polynomial of best approximation. We first treat the case $|f-p|_{\rho,\infty} = m_1$, the case $|f-p|_{\rho,\infty} = -m_0$ being similar. Then $m_0 > -m_1$ and hence

$$\begin{aligned} \rho(x)(f(x)-q(x)) &= \rho(x)(f(x)-p(x)) - \frac{m_0+m_1}{2\rho_{\max}}\rho(x) \geq m_0 - \frac{m_0+m_1}{2\rho_{\max}}\rho(x) \\ &\geq m_0 - \frac{m_0+m_1}{2} = \frac{m_0-m_1}{2} > -m_1 \end{aligned}$$

for every $x \in [a,b]$. To prove $(f(x)-q(x))\rho(x)$ has an upper bound that is strictly less than m_1, we consider the subset $A = \{x \in [a,b] : (f(x)-p(x))\rho(x) \geq \frac{m_1}{2}\}$. Since

$f - p_n$ is continuous, it has an upper bound M on $[a, b]$. Consequently, $M\rho(x) \geq (f(x) - p(x))\rho(x) \geq \frac{m_1}{2}$ and hence $\rho(x) \geq \frac{m_1}{2M} > 0$ on A. It follows that for all $x \in A$,

$$\begin{aligned}(f(x) - q(x))\rho(x) &= (f(x) - p(x))\rho(x) - \frac{m_0 + m_1}{2\rho_{\max}}\rho(x) \\ &< (f(x) - p(x))\rho(x) \leq m_1,\end{aligned}$$

since $m_1 + m_0 > 0$, while for $x \in [a, b]\backslash A$,

$$(f(x) - p(x))\rho(x) - \frac{m_0 + m_1}{2\rho_{\max}}\rho(x) \leq \frac{m_1}{2} - \frac{m_0 + m_1}{2\rho_{\max}}\rho(x) \leq \frac{m_1}{2} < m_1.$$

Thus, $|f - q|_{\rho,\infty} < m_1 = |f - p|_{\rho,\infty}$. □

The following Lemma shows that if the approximation of $f \in C([a, b])$ by $p \in \mathcal{P}^n([a, b])$ gives rise to a uniformly alternating sequence $\{z_i\}_{i=1}^m$ of length $m < n + 2$ with levelled error $e = |f - p|_{\rho,\infty}$, then p is not optimal. To this end, we partition the interval $[a, b]$ into sections, each of which contains only a single point z_i of the alternating sequence. By possibly extending the alternating set, we can further ensure that no section contains both an upper point, a point at which the upper limit is achieved, and a lower point, one at which the lower limit is achieved (see [14]). With a possible modification of ρ at its points of discontinuity, we may further subdivide the interval into closed subintervals in which $\rho(x)(f(x)-p(x))$ is uniformly continuous. Then there exists a $\delta > 0$ so that

$$|\rho(x)(f(x) - p(x)) - \rho(y)(f(y) - p(y))| < e/2 \text{ whenever } |x - y| < \delta.$$

This gives a lower bound on the distance between the points in the alternating sequence and therefore an upper bound on the number of sections.

Lemma 3. *Let $f \in C([a, b])$, $p \in \mathcal{P}^n([a, b])$ and $\{z_i\}_{i=0}^m$ be a sectioned, uniformly alternating set with levelled error satisfying $e = |f - p|_{\rho,\infty} > 0$. If $m < n + 2$, then there exists a polynomial $q \in \mathcal{P}^n([a, b])$ so that $|f - (p + q)|_{\rho,\infty} < e$.*

Proof. Let $s_1 < s_2 < \ldots < s_n$ be the section endpoints that are interior to $[a, b]$, i.e. the sections are $[a, s_1], [s_1, s_2], \ldots, [s_{m-1}, s_m], [s_m, b]$. Let $r(x) = \sigma \prod_{i=1}^m (x - s_i)$, where $\sigma = \pm 1$ is chosen to ensure that $r(x) > 0$ is positive on the interior of the upper sections (sections containing an upper point) and $r(x) < 0$ on the interior of the lower sections (sections containing a lower point). Since none of the upper sections contain lower points and vice versa, there exists an $\varepsilon > 0$ so that $-e + \varepsilon \leq \rho(x)(f(x) - p(x)) \leq e$ for every x in an upper section and $-e \leq \rho(x)(f(x) - p(x)) \leq e - \varepsilon$ for

every x in a lower section. Choose a scaling $k > 0$ so that $|k\rho(x)r(x)| < \varepsilon$ for all $x \in [a, b]$ and let $q(x) = kr(x)$. Since ρ is bounded above, such a constant exists. We now prove that $-e < \rho(x)(f(x) - (p(x) + q(x))) < e$ for all $x \in [a, b]$. We only show how this holds for any x in a lower section, the proof for upper sections being analogous. Since the inequality is trivially satisfied when $\rho(x) = 0$, we assume henceforth that $\rho(x) > 0$. If $x \neq s_i$ for any $i = 1, \ldots, m$ then $q(x) < 0$ and therefore $\rho(x)(f(x) - p(x) - q(x)) > \rho(x)(f(x) - p(x)) \geq -e$. If $x = s_i$ for some $i = 1, \ldots, m$, then x cannot be a lower point (by construction). Moreover, $q(x) = 0$ and hence $\rho(x)(f(x) - p(x) - q(x)) = \rho(x)(f(x) - p(x)) > -e$. To establish the upper bound, we note that since $|\rho(x)q(x)| < \varepsilon$, we must also have

$$\begin{aligned}\rho(x)(f(x) - p(x) - q(x)) &\leq |\rho(x)(f(x) - p(x) - q(x))| \\ &< |\rho(x)(f(x) - p(x))| + |\rho(x)q(x)| < e - \varepsilon + \varepsilon = e.\end{aligned}$$

□

We are now in a position to prove the weighted version of the Chebyshev equioscillation theorem.

Proof. If f is a polynomial of degree n or less, then the proof is trivial. Assume $f \notin \mathcal{P}^n([a, b])$, in which case $d_n > 0$. As mentioned earlier, Theorem 2 asserts that if $\rho(x)(f(x) - p(x))$ has a uniformly alternating sequence of size $n + 2$ that satisfies 14, then $|f - p|_{\rho,\infty} = d_n$. To prove the converse, assume p is a polynomial of best approximation to f. Lemma 2 asserts that there exists a uniformly alternating sequence of length 2 for $\rho(x)(f(x) - p(x))$ with levelled error $e = |f - p|_{\rho,\infty}$. According to Lemma 3, its length must be at least $n + 2$ to guarantee optimality. To show uniqueness, suppose that both q and p are best approximating polynomials in $\mathcal{P}^n([a, b])$ for f. Then

$$\left| f - \frac{p + q}{2} \right|_{\rho,\infty} \leq \frac{1}{2}|f - p|_{\rho,\infty} + \frac{1}{2}|f - q_n|_{\rho,\infty} = d_n.$$

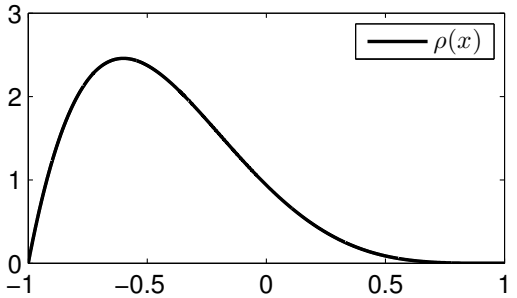
By the first part of this proof, there must be an alternating set $\{z_i\}_{i=1}^{n+2}$ at which

$$\rho(z_i)\left(\frac{f(z_i) - p(z_i)}{2}\right) + \rho(z_i)\left(\frac{f(z_i) - q(z_i)}{2}\right) = \pm d_n.$$

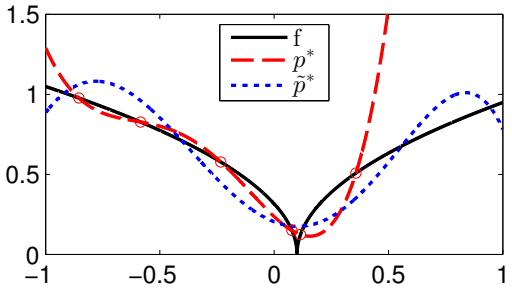
Since $\rho(z_i) > 0$ for $i = 1, \ldots, n + 2$, this is only possible if $p(z_i) = q(z_i)$ for $i = 1, \ldots, n + 2$ and hence $p = q$. □

As an illustration of how the density ρ affects the best approximation of a function, consider the following example.

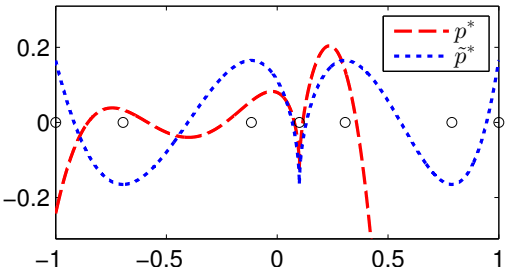
Example 4 (Best Approximation)**.** Let $f(x) = \sqrt{|x - 0.1|}$ and $\rho(x)$, depicted in Sub-figure 4a, be the density function corresponding to the random variable $X = -1+2Y$, where $Y \sim \text{Beta}(2, 5)$. Figure 4b shows the best polynomial approximations p^* and $\tilde{p}^*$ of f in $\mathcal{P}^5([-1, 1])$, using the ρ-weighted seminorm and the standard uniform norm respectively. The red circles in Figure 4b represent the interpolation points of p^*. In Figure 4c, the error $f(x) - \tilde{p}^*$ equioscillates at the points denoted by 'o', while $f(x) - p^*(x)$ grows in regions where ρ is small. On the other hand, Figure 4d shows how the weighted error $\rho(x)(f(x) - p^*(x))$ of p^* equioscillates at the circled equioscillation points (which have a more limited spread than those in Figure 4c), while that of $\tilde{p}^*$ is sub-optimal.



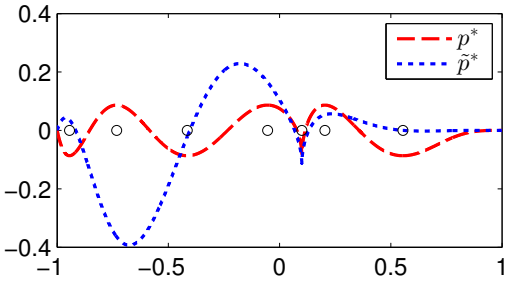
(a) The density function.



(b) f with best approximation p^*.



(c) The unweighted error $f(x) - p^*(x)$.



(d) Weighted error $\rho(x)(f(x) - p^*(x))$.

Figure 4: The best approximating polynomial of $f(x) = \sqrt{|x - 0.1|}$ in the weighted seminorm $|\cdot|_{\rho,\infty}$.

3.2 Point Selection Using the Remez Algorithm

The weighted Remez algorithm approximates the set of equioscillation points (and hence the best polynomial approximation) iteratively from an initial set of $n + 2$

trial points. At the kth step, a given a set of trial points $\{z_1^k, \ldots, z_{n+2}^k\}$ is used to determine the unique polynomial p_k and uniform levelled error $e_k > 0$ so that

$$\rho(z_i^k)(f(z_i^k) - p_k(z_i^k)) = \sigma_i e_k, \tag{18}$$

for $i = 1, \ldots, n+2$, where $\sigma_i = \lambda(-1)^i$ and $\lambda = \pm 1$. These are then used to adjust the set of trial points, to obtain a new set $\{z_1^{k+1}, \ldots, z_{n+2}^{k+1}\}$ at which the error oscillates (it need not equioscillate) and so that $e_k < \min_i \rho(z_i^{k+1})|f(z_i^{k+1}) - p_k(z_i^{k+1})|$. The latter step can be achieved by means of the Remez exchange algorithm and ensures that $e_k < e_{k+1}$. Indeed, applying the de la Valée-Poussin Theorem to p_k, alternating on the set $\{z_1^{k+1}, \ldots, z_{n+2}^{k+1}\}$, and using $q = p_{k+1}$ gives

$$\begin{aligned} e_k &< \min_{0 \le i \le n+1} \rho(z_i^{k+1})|f(z_i^{k+1}) - p_k(z_i^{k+1})| \\ &\le \max_{0 \le i \le n+1} \rho(z_i^{k+1})|f(z_i^{k+1}) - p_{k+1}(z_i^{k+1})| = e_{k+1}. \end{aligned}$$

This is guarantees that the error $|f - p^k|_{\rho,\infty}$ decreases monotonically to d_n.

Our implementation of the weighted Remez algorithm is based largely on that of [17], which treats the unweighted Remez algorithm within the setting of the Chebfun project [9]. We make a few notable adjustments which we discuss here briefly. Consider the estimation of the equioscillating polynomial $p(x)$ and levelled error $e > 0$ from a set $\{z_1^k, ..., z_{n+2}^k\}$, based on condition (18). Through the use of Lagrange bases centered on different subsets of the trial set, the authors in [17] derive an explicit formula for e. Trial polynomials, expressed in Chebyshev form can then be defined directly via (18). In this form, the trial polynomial can not only be evaluated efficiently by means of the barycentric interpolation formula, which is forward stable for point sets with small Lebesgue constants [13], but also effectively differentiated, and solved. In the weighted case, the same approach yields an explicit expression for

$$e = \frac{\sum_{j=1}^{n+2} w_j f(z_j^k)}{\sum_{j=1}^{n+2} \sigma_j w_j / \rho(z_j^k)}, \quad \text{where } w_j = \prod_{\substack{i=1 \\ i \ne j}}^{n+2} \frac{1}{2}(z_i^k - z_j^k)^{-1}$$

is the jth barycentric weight. However, depending on the density function, the trial set may no longer produce a small Lebesgue constant (see Remark 1 below). Moreover, polynomials that are optimal approximations in the weighted uniform seminorm usually deviate strongly in regions where ρ is small, by design. In our computations, we have found the modified Lagrange interpolant [13] to give more accurate results for a wide array of densities. As a consequence, we forgo the convenience of the Chebfun library and must resort to less efficient, brute force search methods in our implementation of the exchange algorithm and in locating the interpolation points.

Remark 1. Let $\{x_j\}_{j=1}^n \subset [a,b]$ be a set of quadrature nodes at which $\rho(x_j) > 0$, and $\{\ell_j\}_{j=1}^n$ be the associated Lagrange basis. It is well known that the ∞-norm of the interpolation operator $L : C([a,b]) \to \mathcal{P}^n([a,b])$ is given by the Lebesgue constant $\Lambda = \sup_{x\in[a,b]} \sum_{j=1}^n |\ell_j(x)|$. When $\|\cdot\|_\infty$ is replaced with the density-weighted seminorm $|\cdot|_{\rho,\infty}$, it can readily be seen that the analogous weighted Lebesgue constant takes the form

$$\Lambda^\rho = \sup_{x\in[a,b]} \sum_{j=1}^n \frac{|\ell_j(x)|}{\rho(x_j)}\rho(x).$$

While the point sets generated by the weighted Remez algorithm (as well as the weighted Léja nodes) exhibit low weighted Lebesgue constants, their unweighted Lebesgue constants are usually large, since the associated Lebesgue functions typically grow outside the support of ρ. This limits the use of fast numerical methods such as barycentric interpolation. The development of efficient interpolation algorithms that are stable for point sets with low *weighted* Lebesgue constants would be an interesting avenue for further research.

Remark 2. Our numerical experiments suggest that the optimally spaced nodes also give rise to quadrature weights that are positive, although we have so far been unable to prove this assertion.

3.2.1 Nested Points

Due to the importance of nesting in the context of both high dimensional- and adaptive quadrature methods and the need for high granularity, i.e. the ability to add just a small number of additional points at each refinement level, quadrature rules based on the Léja nodes have steadily been gaining traction (see [15]). These are generated sequentially, the first being chosen arbitrarily, while each successive node is placed where Θ attains its maximum, i.e. x_{i+1} satisfies

$$x_{i+1} = \text{argmax}_{x\in[a,b]} |\Theta(x; \{x_1, \ldots, x_i\})|.$$

Although the Léja nodes are not optimally spaced, they are relatively simple to compute and are not only nested but also highly granular (see [15]).

The above procedure can readily be extended to construct nested rules in a greedy way. Consider an existing quadrature rule with nodes $y_1, \ldots, y_m$ and suppose we want to add an additional n points $x_1, \ldots, x_n$ so as to minimize

$$\min_{x_1,\ldots,x_n} |\Theta(x; \{y_1, \ldots, y_m, x_1, \ldots, x_n\})|_{\rho,\infty}. \tag{19}$$

To obtain this new set of nodes, we apply the Remez algorithm to find the best polynomial approximation to $f(x) = x^n \prod_{i=1}^{m}(x - y_i)$ using polynomials of the form $p(x)\prod_{i=1}^{n}(x - y_i)$, where $p \in \mathcal{P}^{n-1}([a, b])$. In this case, the weighted error $|\Theta|_{\rho,\infty}$ takes the form

$$\begin{aligned} &|\Theta(\cdot; y_1, \ldots, y_m, x_1, \ldots, x_n)|_{\rho,\infty} \\ &= \max_{-1\le x\le 1} \rho(x) \left| (x^n - p(x)) \prod_{i=1}^{m}(x - y_i) \right| = \max_{-1\le x\le 1} \tilde{\rho}(x)|x^n - p(x)|, \end{aligned}$$

where $\tilde{\rho}(x) := \rho(x)\left(\prod_{i=1}^{n}|x - y_i|\right)$. As before, we need to find the best polynomial approximation to x^n, albeit now with respect to a different weighted error.

Example 5. Here we compare the distribution of quadrature nodes as well as the 'spread' function Θ of the Léja points with those obtained by using the Remez algorithm to greedily generate a sequence of nested points $x_1, x_2, \ldots$, each successive member x_{i+1} of which solves

$$\min_{x_{i+1}} |\Theta(\cdot; \{x_1, \ldots, x_i\})|_{\rho,\infty}.$$

As in Example 4, ρ is given by the Beta(2,5) density function. Figure 5 shows how the points obtained from the greedy algorithm tend to cluster in the region where ρ is large, while the Léja points, chosen successively at the point of highest error, have a more even spread.

3.2.2 Accuracy of Quadrature Rules Based on ρ-Optimal Point Sets

In the following numerical tests, we compare the accuracy of quadrature rules based on the ρ-optimal nodes (or on the ρ-weighted Léja nodes) with that of the ρ-weighted Clenshaw-Curtis type rules discussed in Section 2. Figure 6 shows the error for densities with support over bounded domains, while Figure 7 shows analogous results for densities defined on semi-infinite domains. As before, we use the algebraic map introduced in Section 2 to transform the semi-infinite intervals to the standard interval [-1,1]. The weighted Clenshaw-Curtis type rules are based on the Féjer type II nodes to avoid evaluating functions at infinity. The results of our experiments lead us to make the following observations. The quadrature rules based on the weighted Léja nodes exhibit errors that are strikingly similar to those associated with the ρ-optimal point set. These errors also tend to be at least as small as those of the weighted Clenshaw-Curtis type rules. Moreover, for functions with singularities, the rules based on ρ-optimal node placement have much better convergence whenever

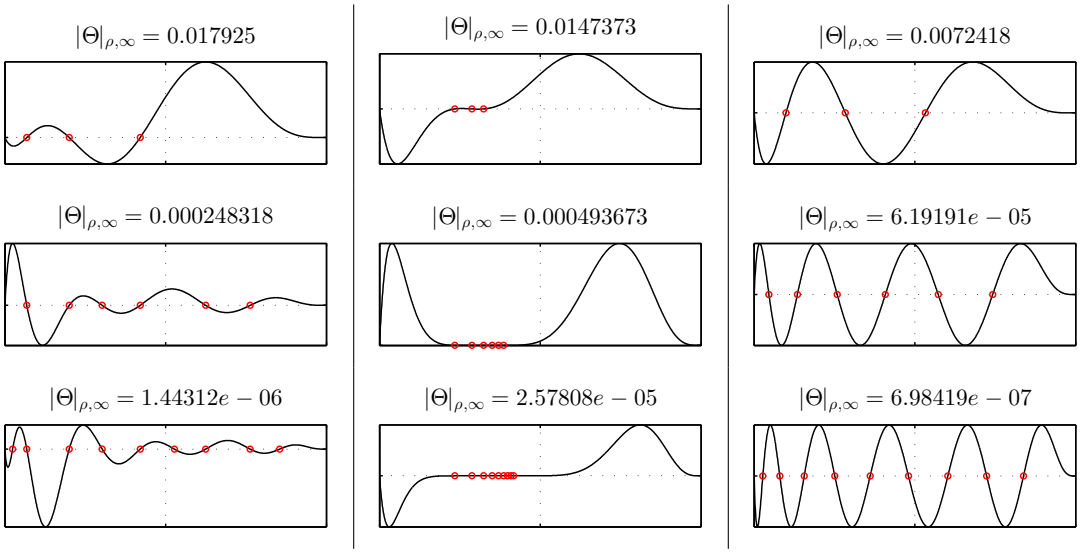


Figure 5: The function Θ for the nested Léja points (left column), for the nested points generated by the greedy strategy outlined above (center column), and for the ρ-optimal points (right column). The number of points in rows 1,2, and 3 are 3,6, and 9 respectively.

the support of the density is located away from the singularity. As an example in the bounded case, consider the logit normal density function. It is small near $x = 0.5$, leading to much improved estimates of the integrals of functions such as $f(x) = \exp(-1/(x-0.5)^2)$ and $f(x) = |x-0.5|^3$, compared to the rule based on the standard point set. As an example in the semi-infinite case, compare the errors of standard weighted rule with the ρ-optimal node rules for the gamma, lognormal, and pareto densities, when $f(x) = \sin x/x$ or $f(x) = x^2$. Both these functions are singular near the right endpoint when mapped to $[-1, 1]$. The mapped gamma density has support near $x = 1$ and the mapped pareto density decays slowly as $x \to 1^-$, resulting in errors determined largely by the singularities of $f(x)$ (although still better than the weighted F_2 rule for $f(x) = x^2$). The lognormal density on the other hand, decays rapidly as $|x| \to 1$, leading to excellent convergence rates for the ρ-optimal node rules.

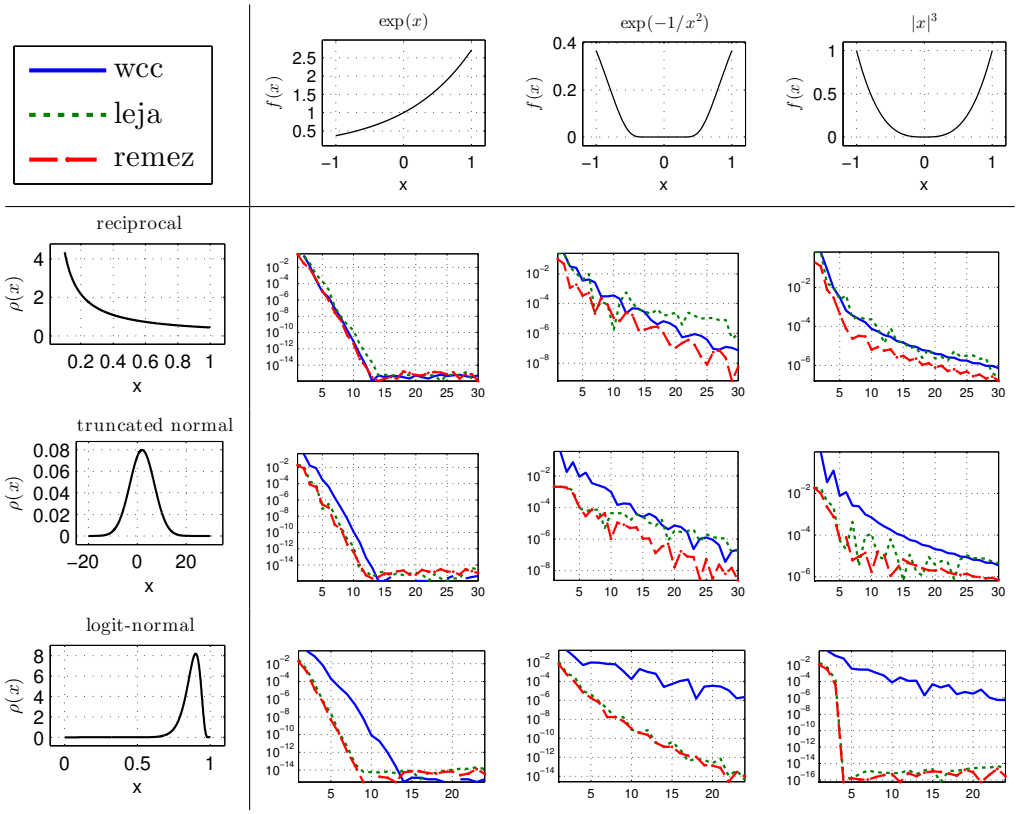


Figure 6: Absolute quadrature errors for the weighted CC-type, -Remez, and -Léja rules on bounded domains.

4 Conclusion

Density functions can readily be incorporated into the construction of CC-type rules, leading to more efficient quadrature schemes. An adjustment in the weights of the standard Clenshaw-Curtis type rules results in errors that are comparable to those in the $\rho = 1$ case. If, moreover, the quadrature nodes are chosen to optimize the ρ-weighted seminorm of the spread function Θ, they tend to concentrate in regions where the underlying random variable is likely to lie. This leads to further improvements in accuracy, especially when the density's support is away from the location of the integrand's singularities. One remaining numerical challenge is the efficient evaluation of polynomial interpolants of functions centered at the ρ-optimal nodes, where high lebesgue constants limit the use of traditional algorithms like the barycentric interpolation formula.

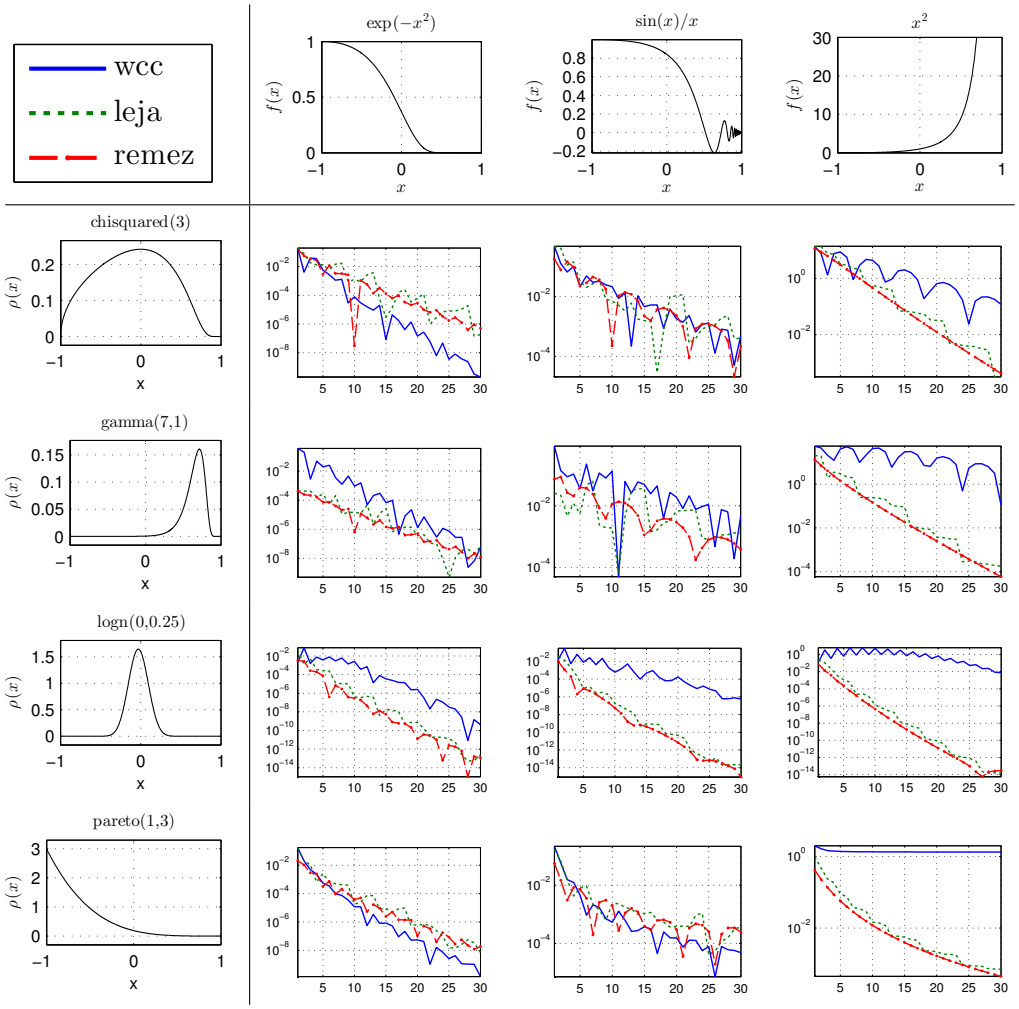


Figure 7: Absolute quadrature errors for the density weighted Féjer II, -Remez, and -Léja rules over semi-infinite domains.

A List of Density Functions

The following table offers an overview of the density functions used in this paper.

References

[1] Richard Askey and James Wilson. Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials. *Mem. Amer. Math. Soc.*,

Density	**Support**	$\rho(x)$	**Parameters**
Finite Support			
Beta(α, β)	$[0, 1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\alpha > 0, \beta > 0$
Reciprocal(a, b)	$[a, b]$	$\frac{1}{x(\ln(b)-\ln(a))}$	$0 < a < b < \infty$
TruncN(μ, σ, a, b)	$[a, b]$	$\frac{C(a,b)}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$-\infty \leq a < b \leq \infty$, $\mu \in \mathbb{R}, \sigma > 0$
LogitN(μ, σ)	$(0, 1)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\text{logit}(x)-\mu)^2}{2\sigma^2}} \frac{1}{x(1-x)}$	$\mu \in \mathbb{R}, \sigma > 0$
Semi-Infinite Support			
Exp(λ)	$[0, \infty)$	$\lambda e^{-\lambda x}$	$\lambda \in [0, \infty)$
$\chi^2(k)$	$[0, \infty)$	$\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$	$k \in \mathbb{N}$
Gamma(k, θ)	$(0, \infty)$	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$k, \theta \in [0, \infty)$
lnN(μ, σ^2)	$(0, \infty)$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{(\ln x-\mu)^2}{2\sigma^2}}$	$\mu \in \mathbb{R}, \sigma > 0$
Pareto(x_m, α)	$[x_m, \infty)$	$\frac{\alpha x_m^\alpha}{x^{\alpha+1}}, x \geq x_m$	$x_m, \alpha > 0$
Weibull(λ, k)	$[0, \infty)$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$	$\lambda, k > 0$

54(319):iv+55, 1985.

[2] Kendall Atkinson. *An Introduction to Numerical Analysis*. John Wiley & Sons Inc, 1989.

[3] John P Boyd. Exponentially convergent fourier-chebshev quadrature schemes on bounded and infinite intervals. *Journal of Scientific Computing*, 2(2):99–109, 1987.

[4] John P Boyd. Orthogonal rational functions on a semi-infinite interval. *Journal of Computational Physics*, 70(1):63–88, 1987.

[5] John P Boyd. Spectral methods using rational basis functions on an infinite interval. *Journal of Computational Physics*, 69(1):112–142, 1987.

[6] John P. Boyd. *Chebyshev and Fourier spectral methods*. Dover Publications, Inc., Mineola, NY, second edition, 2001.

[7] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[8] Charles W Clenshaw and Alan R Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2(1):197–205, 1960.

[9] T. A Driscoll, N. Hale, and L. N. Trefethen. *Chebfun Guide*. Pafnuty Publications, 2014.

[10] GA Evans and JR Webster. A comparison of some methods for the evaluation of highly oscillatory integrals. *Journal of Computational and Applied Mathematics*, 112(1):55–69, 1999.

[11] Leopold Fejér. Mechanische quadraturen mit positiven cotesschen zahlen. *Mathematische Zeitschrift*, 37(1):287–309, 1933.

[12] W Morven Gentleman. Implementing clenshaw-curtis quadrature, i methodology and experience. *Communications of the ACM*, 15(5):337–342, 1972.

[13] Nicholas J Higham. The numerical stability of barycentric lagrange interpolation. *IMA Journal of Numerical Analysis*, 24(4):547–556, 2004.

[14] Robert Mayans. The chebyshev equioscillation theorem. *Journal of Online Mathematics and Applications*, (1316), 2006.

[15] Akil Narayan and John Jakeman. Adaptive leja sparse grid constructions for stochastic collocation and high-dimensional approximation. *arXiv*, 2014.

[16] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[17] R. Pachón and L. N. Trefethen. Barycentric-Remez algorithms for best polynomial approximation in the chebfun system. *BIT Numerical Mathematics*, 49(4):721–741, 2009.

[18] T. Parks and J. McClellan. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circuit Theory*, 19(2):189–194, 1972.

[19] Eugene Y. Remez. Sur la détermination des polynômes d'approximation de degré donnée. *Comm. Soc. Math. Kharkov*, 10:41–63, 1934.

[20] Eugene Y. Remez. Sur le calcul effectif des polynômes d'approximation de tchebichef. *Comptes Rendus de l'Academie des Sciences*, 199:337–340, 1934.

[21] Eugene Y. Remez. Sur un procédé convergent d'approximations successives pour déterminer les polynômes d'approximation. *Comptes Rendus de l'Academie des Sciences*, 198(2063-2065), 1934.

[22] Mark Richardson. Approximating divergent functions in the chebfun system. Master's thesis, University of Oxford, 2009.

[23] Alvise Sommariva. Fast construction of fejér and clenshaw–curtis rules for general weight functions. *Computers & Mathematics with Applications*, 65(4):682–693, 2013.

[24] Jörg Waldvogel. Fast construction of the fejér and clenshaw–curtis quadrature rules. *BIT Numerical Mathematics*, 46(1):195–202, 2006.

[25] Dongbin Xiu and George Em Karniadakis. The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.