MINIMUM-ENERGY MEASURES FOR SINGULAR KERNELS

Luc Pronzato, Anatoly A. Zhigljavsky

MINIMUM-ENERGY MEASURES FOR SINGULAR KERNELS

LUC PRONZATO* AND ANATOLY ZHIGLJAVSKY†

Abstract. We develop algorithms for energy minimization for kernels with singularities. This problem arises in different fields, most notably in construction of space-filling sequences of points where singularity of kernels guarantees a strong repelling property between these points. Numerical algorithms are based on approximating singular kernels by non-singular ones, subsequent discretisation and solving non-singular discrete problems. For approximating singular kernels, we approximate an underlying completely monotone (briefly, CM) function with singularity by a bounded CM function with controlled accuracy. Theoretical properties of the suggested approximation are studied and some numerical results are shown.

Keywords: energy minimization, singular kernels, Riesz kernel, Riesz potential, space-filling design
AMS subject classifications: 62K99, 65D30, 65D99.

1. Introduction: kernels and energies. This section introduces the main concepts and formulates the most important results required for the following sections. It mostly follows [6], where the reader may find some proofs and more details. Subsection 1.12 discusses the motivation behind this research and describes the structure of the rest of the paper.

1.1. Main notation.

$\mathscr{X}$: a compact subset of $\mathbb{R}^d$; $d \geq 1$.
$\|\cdot\|$: the Euclidean norm.
$\mathscr{M}$: the set of finite signed Borel measures on $\mathscr{X}$.
$\mathscr{M}(q)$: the set of signed measures with total mass q, $\mathscr{M}(q)=\{\mu\in\mathscr{M} : \mu(\mathscr{X})=q\}$.
$\mathscr{M}^+(1)$: the set of Borel probability measures on $\mathscr{X}$.
$\mathscr{M}^+$: the set of finite positive measures on $\mathscr{X}$.
K: a kernel; that is, a continuous symmetric function $K : \mathscr{X}\times\mathscr{X} \to \mathbb{R}\cup\{\infty\}$; K is (uniformly) bounded if $K(\mathbf{x},\mathbf{x}) < \infty$ for all $\mathbf{x}\in\mathscr{X}$; K is singular if $K(\mathbf{x},\mathbf{x}) = +\infty$ for at least one $\mathbf{x}\in\mathscr{X}$. Further conditions on K will be specified in Section 1.2.
$\mathscr{E}_K(\nu)$, $\nu\in\mathscr{M}$: the energy of ν,

$$\mathscr{E}_K(\nu) = \int_{\mathscr{X}}\int_{\mathscr{X}} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x})\mathrm{d}\nu(\mathbf{x}')\,. \tag{1.1}$$

$\mathscr{M}_K$: the set of measures with finite energy,

$$\mathscr{M}_K = \{\nu\in\mathscr{M} : |\mathscr{E}_K(\nu)| < +\infty\}\,. \tag{1.2}$$

$P_\nu(\mathbf{x})$: the potential of $\nu\in\mathscr{M}$ at $\mathbf{x}$,

$$P_\nu(\mathbf{x}) = \int_{\mathscr{X}} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x}')\,,\quad \mathbf{x}\in\mathscr{X}\,. \tag{1.3}$$

$\gamma_K(\mu,\nu)$: the MMD (Maximum-Mean Discrepancy) between measures $\mu,\nu\in\mathscr{M}$,

$$\gamma_K(\mu,\nu) = \mathscr{E}_K^{1/2}(\nu-\mu) = \left(\int_{\mathscr{X}^2} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}(\nu-\mu)(\mathbf{x})\,\mathrm{d}(\nu-\mu)(\mathbf{x}')\right)^{1/2}. \tag{1.4}$$

†Cardiff University, UK; e-mail: Zhigljavskyaa@cardiff.ac.uk (corresponding author)
*Université Côte d'Azur, CNRS, I3S, France

CM (completely monotone) function: $f:(0,\infty)\to\mathbb{R}^+$ is CM if $f\in C^\infty(0,\infty)$ and

$$(-1)^k f^{(k)}(t)\geq 0\,,\quad \forall t\in(0,\infty);\quad k=0,1,\ldots \tag{1.5}$$

BF (Bernstein function): $g:(0,\infty)\to\mathbb{R}^+$ is a BF if $g\in C^\infty(0,\infty)$ and g' is CM; see [10, p.15].

1.2. Kernels of interest.

Definition 1.1. *A kernel K is* Strictly Positive Definite *(SPD) on $\mathscr{M}$ if K is bounded and for all $n\in\mathbb{N}$ and all pairwise different $\mathbf{x}_1,\ldots,\mathbf{x}_n\in\mathscr{X}$, the matrix $\mathbf{K}_n$ with elements $\{\mathbf{K}_n\}_{i,j}=K(\mathbf{x}_i,\mathbf{x}_j)$ $(i,j=1,\ldots,n)$ is positive definite.*

Definition 1.2. *A kernel K is* Integrally Strictly Positive Definite *(ISPD) on $\mathscr{M}$ if $\mathscr{E}_K(\nu)>0$ for any nonzero measure $\nu\in\mathscr{M}$.*

Definition 1.3. *A kernel K is* Conditionally Integrally Strictly Positive Definite *(CISPD) on $\mathscr{M}$ when it is ISPD on $\mathscr{M}(0)$; that is, when $\mathscr{E}_K(\nu)>0$ for all nonzero signed measures $\nu\in\mathscr{M}$ such that $\nu(\mathscr{X})=0$.*

An ISPD kernel is CISPD. A bounded ISPD kernel is SPD and defines an RKHS (Reproducing Kernel Hilbert Space) $\mathcal{H}_K$. For CISPD kernels K, the energy $\mathscr{E}_K(\nu)$ can be negative. As for singular kernels K the energy $\mathscr{E}_K(\cdot)$ can be infinite, we may have $\mathscr{M}_K\neq\mathscr{M}$, where $\mathscr{M}_K$ is defined in (1.2). If K is SPD, then $\mathscr{M}_K=\mathscr{M}$ and the potential $P_\nu(\mathbf{x})$, see (1.3), is well defined and finite for any $\nu\in\mathscr{M}$ and $\mathbf{x}\in\mathscr{X}$. However, there always exists $\nu\in\mathscr{M}_K$ such that $P_\nu(\mathbf{x}_0)$ is infinite for some $\mathbf{x}_0\in\mathscr{X}$ when K is singular.

1.3. Examples of kernels.

Example 1.1 (Kernels constructed through CM functions)**.** We consider two general ways of constructing kernels $K(\cdot,\cdot)$ via CM functions.

Let $f(\cdot)$ be a non-constant completely monotone (CM) function, see (1.5).

(a) For $x,x'\in\mathbb{R}$, define

$$K(x,x')=f(|x-x'|)\,. \tag{1.6}$$

As follows from Theorem 2.4 of Section 2.1, if $f(\cdot)$ is CM and also belongs to $L_1((0,\infty))$ then the kernel (1.6) is ISPD.

(b) For $\mathbf{x},\mathbf{x}'\in\mathbb{R}^d$, define

$$K(\mathbf{x},\mathbf{x}')=f(\|\mathbf{x}-\mathbf{x}'\|^2)\,. \tag{1.7}$$

The connection between K being SPD and f being CM is clarified in Theorem 2.3 of Section 2.1.

◁

In (1.6) and (1.7), the values of f and its derivatives at 0 may not be defined; in these cases, the function f is singular at 0 and hence the corresponding kernels K are singular. Examples of univariate CM functions with singularity at 0 are provided in Section 3.4.

Example 1.2 (Some important bounded ISPD kernels)**.**

- The squared exponential kernel $K_\beta(\mathbf{x},\mathbf{x}')=\exp(-\beta\,\|\mathbf{x}-\mathbf{x}'\|^2)$, $\beta>0$.
- The isotropic Matérn kernels $K_{\kappa,\beta}$ with shape parameter κ, in particular

$$K_{3/2,\beta}(\mathbf{x},\mathbf{x}')=(1+\sqrt{3}\beta\,\|\mathbf{x}-\mathbf{x}'\|)\,\exp(-\sqrt{3}\beta\,\|\mathbf{x}-\mathbf{x}'\|)\quad\text{(Matérn 3/2)}$$

and the exponential kernel $K_{1/2,\beta}(\mathbf{x},\mathbf{x}')=\exp(-\beta\|\mathbf{x}-\mathbf{x}'\|)$, $\beta>0$.

- The generalized multiquadric kernel

$$K_{\alpha,\epsilon}(\mathbf{x},\mathbf{x}') = (\|\mathbf{x}-\mathbf{x}'\|^2+\epsilon)^{-\alpha/2}\,,\ \epsilon>0\,,\ \alpha>0\,.\ \triangleleft$$

Example 1.3 (Bounded CISPD kernels)**.** Consider the kernels defined by

$$K^{(\alpha)}(\mathbf{x},\mathbf{x}') = -\|\mathbf{x}-\mathbf{x}'\|^{\alpha}\,,\ \alpha>0\,, \tag{1.8}$$

which are CISPD for $\alpha\in(0,2)$, see [15], and the related distance-induced kernels

$$K'^{(\alpha)}(\mathbf{x},\mathbf{x}') = \|\mathbf{x}\|^{\alpha}+\|\mathbf{x}'\|^{\alpha}-\|\mathbf{x}-\mathbf{x}'\|^{\alpha}\,,\ \alpha>0\,.$$

$K'^{(\alpha)}$ is CISPD for $\alpha\in(0,2)$, but is not SPD (in particular, $K'^{(\alpha)}(\mathbf{0},\mathbf{0})=0$). In [15], $\mathscr{E}_{K'^{(\alpha)}}$ is called energy distance for $\alpha=1$ and generalized energy distance for general $\alpha\in(0,2]$. For $\alpha=1$ and $\mathscr{X}=[0,1]$ the kernel $K(x,x')=1-K^{(1)}(x,x')=1-|x-x'|$ is ISPD. $\triangleleft$

Example 1.4 (Singular ISPD kernels)**.** The Riesz kernels:

$$K_{(\alpha)}(\mathbf{x},\mathbf{x}') = \|\mathbf{x}-\mathbf{x}'\|^{-\alpha}\,,\ \alpha>0\,, \tag{1.9}$$

If $0<\alpha<d$ then $K_{(\alpha)}$ is ISPD. When $\alpha\geq d$, $\mathscr{E}_{K_{(\alpha)}}(\mu)$ is infinite for any $\mu\in\mathscr{M}$. $\triangleleft$

Example 1.5 (Singular CISPD kernel)**.** The logarithmic kernel:

$$K_{(0)}(\mathbf{x},\mathbf{x}') = -\log\|\mathbf{x}-\mathbf{x}'\|\,. \tag{1.10}$$

This kernel is sometimes considered as a member of the Riesz family of kernels (1.9), as $\alpha\to 0$. Since $K_{(0)}(\mathbf{x},\mathbf{x}')$ tends to $-\infty$ when $\|\mathbf{x}-\mathbf{x}'\|$ tends to $+\infty$, it can only be used in the case when $\mathscr{X}$ is compact. The kernel $K_{(0)}$ is CISPD, see [2, p. 80]. $\triangleleft$

1.4. Strict convexity of energy.

Lemma 1.1.

(a) *K is ISPD if and only if $\mathscr{M}_K$ is convex and $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}_K$.*

(b) *Assume that K is bounded. Then, K is CISPD if and only if $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}(1)$.*

For a proof, see Lemmas 3.1 and 3.2 in [6]. Lemma 1.1 does not cover the case of singular CISPD kernels where a similar result can be proved. For instance, in view of [9, Sect. I.3], the energy $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}(1)\cap\mathscr{M}_K$ for the logarithmic kernel (1.10).

In the remaining part of the paper we assume that K is such that $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}(1)\cap\mathscr{M}_K$, which is true in particular under the assumptions of Lemma 1.1.

1.5. MMD as Bregman divergence and Jensen difference. For $\mu,\nu\in\mathscr{M}_K$, denote by $F_K(\mu;\nu)$ the directional derivative of $\mathscr{E}_K(\cdot)$ at μ in the direction ν:

$$F_K(\mu;\nu) = \lim_{\alpha\to 0^+}\frac{\mathscr{E}_K[(1-\alpha)\mu+\alpha\nu]-\mathscr{E}_K(\mu)}{\alpha}\,.$$

Straightforward calculation gives

$$F_K(\mu;\nu) = 2\left[\int_{\mathscr{X}^2}K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x})\mathrm{d}\mu(\mathbf{x}')-\mathscr{E}_K(\mu)\right]\,. \tag{1.11}$$

The strict convexity of $\mathscr{E}_K(\cdot)$ implies that $\mathscr{E}_K(\nu) \geq \mathscr{E}_K(\mu) + F_K(\mu,\nu)$ for any $\mu,\nu \in \mathscr{M}_K$, with equality if and only if $\nu = \mu$.

The Bregman divergence (associated with the functional $\mathscr{E}_K(\cdot)$) between measures in $\mathscr{M}_K$ and between probability measures in $\mathscr{M}^+(1) \cap \mathscr{M}_K$, is

$$B_K(\mu,\nu) = \mathscr{E}_K(\nu) - [\mathscr{E}_K(\mu) + F_K(\mu,\nu)]\,;$$

see [5, 8] for a general definition of Bregman divergence. By direct calculation, $B_K(\mu,\nu) = B_K(\nu,\mu) = \mathscr{E}_K(\nu-\mu)$, which allows us to define the squared MMD $\gamma_K^2(\mu,\nu)$ as the Bregman divergence $B_K(\mu,\nu)$.

The Jensen difference (associated with $\mathscr{E}_K(\cdot)$) is

$$\Delta_J(\mu,\nu) = (1/2)[\mathscr{E}_K(\mu) + \mathscr{E}_K(\nu)] - \mathscr{E}_K[(\mu+\nu)/2]\,,$$

see [7] for a general definition of Jensen differences. Direct calculation gives

$$\gamma_K^2(\mu,\nu) = \mathscr{E}_K(\nu-\mu) = 4\,\Delta_J(\mu,\nu)\,.$$

Since $\mathscr{E}_K(\cdot)$ is a strictly convex functional on $\mathscr{M}(1) \cap \mathscr{M}_K$, both representations of MMD imply that $\gamma_K(\cdot,\cdot)$ defines a proper metric on the space of signed measures $\mathscr{M}(1) \cap \mathscr{M}_K$ and the space of probability measures $\mathscr{M}^+(1) \cap \mathscr{M}_K$.

For other interpretations of the MMD $\gamma_K(\cdot,\cdot)$, see [6] and references therein.

1.6. Minimum-energy probability measure.

From (1.3) and (1.11), the potential $P_\mu(\mathbf{x})$ associated with μ at $\mathbf{x} \in \mathscr{X}$, can be written as

$$P_\mu(\mathbf{x}) = \frac{1}{2}\,F_K(\mu;\delta_\mathbf{x}) + \mathscr{E}_K(\mu)\,, \tag{1.12}$$

where $\delta_\mathbf{x}$ is the delta-measure concentrated at $\mathbf{x}$.

Theorem 1.1. [6, Th. 3.1] *Assume that $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}^+(1) \cap \mathscr{M}_K$. Then, (a) there always exists a unique minimum-energy probability measure, and (b) $\mu_K^+ \in \mathscr{M}^+(1)$ is the minimum-energy probability measure on $\mathscr{X}$ if and only if*

$$\forall\,\mathbf{x} \in \mathscr{X}\,,\;\; P_{\mu_K^+}(\mathbf{x}) \geq \mathscr{E}_K(\mu_K^+)\,.$$

Statement (a) follows from the fact that $\mathscr{X}$ is compact and therefore the set $\mathscr{M}^+(1)$ is vagely compact. Moreover, a measure $\mu_K^+ \in \mathscr{M}^+(1)$ is the minimum-energy probability measure if and only if $F_K(\mu_K^+;\nu) \geq 0$ for all $\nu \in \mathscr{M}^+(1)$, or equivalently, since ν is a probability measure, if and only if $F_K(\mu_K^+;\delta_\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathscr{X}$. In view of (1.12), this is equivalent to the statement (b) of the theorem.

Note that, by construction, $\int_\mathscr{X} P_{\mu_K^+}(\mathbf{x})\,\mathrm{d}\mu_K^+(\mathbf{x}) = \mathscr{E}_K(\mu_K^+)$, implying $P_{\mu_K^+}(\mathbf{x}) = \mathscr{E}_K(\mu_K^+)$ on the support of the minimum-energy probability measure μ_K^+.

1.7. Minimum-energy signed measure.

Theorem 1.2. [6, Th. 3.2] *When $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}(1) \cap \mathscr{M}_K$, $\mu_K^* \in \mathscr{M}(1)$ is the minimum-energy signed measure with total mass one on $\mathscr{X}$ if and only if*

$$\forall\,\mathbf{x} \in \mathscr{X}\,,\;\; P_{\mu_K^*}(\mathbf{x}) = \mathscr{E}_K(\mu_K^*)\,. \tag{1.13}$$

For a proof we just need to note that μ_K^* is the minimum-energy measure in $\mathscr{M}(1)$ if and only if $F_K(\mu_K^*;\nu) = 0$ for all $\nu \in \mathscr{M}(1)$ and this condition is equivalent

to $F_K(\mu_K^*;\delta_{\mathbf{x}}) = 0$ for all $\mathbf{x} \in \mathscr{X}$. In view of (1.12) this is exactly the condition (1.13).

The minimum-energy signed measure μ_K^* may not exist, especially when the kernel K is differentiable; such cases include the squared exponential kernel, the genealized multiquadric kernel and isotropic Matérn kernels with shape parameter $\kappa > 1$, see Example 1.2.

1.8. Sufficient conditions for the minimum-energy signed measure to be a probability measure.

Theorem 1.3. [6, Th. 3.3] *Assume that K is ISPD and translation invariant, with $K(\mathbf{x},\mathbf{x}') = F(\mathbf{x}-\mathbf{x}')$ and F continuous, twice differentiable except at the origin, with Laplacian $\Delta_F(\mathbf{x}) = \sum_{i=1}^d \partial^2 F(\mathbf{x})/\partial x_i^2 \geq 0$, $\mathbf{x} \neq \mathbf{0}$. Then there exists a unique minimum-energy signed measure μ_K^* in $\mathscr{M}(1)$, and μ_K^* is a probability measure.*

The weaker condition $F(x-x') = f(|x-x'|)$ with $f(\cdot)$ convex on $(0,\infty)$ is sufficient when $d = 1$. When $d \geq 2$ with $F(\mathbf{x}-\mathbf{x}') = f(\|\mathbf{x}-\mathbf{x}'\|)$, $f(\cdot)$ must have a singularity at 0 to have $\Delta_F(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq 0$. For the Riesz kernels $K_{(\alpha)}$ of (1.9), we have $\Delta(\|\mathbf{x}\|^{-\alpha}) = \alpha(\alpha+2-d)/\|\mathbf{x}\|^{\alpha+2}$, $\mathbf{x} \neq \mathbf{0}$. Therefore, the conditions of the theorem are met when $d > 2$ and $\alpha \in (0, d-2]$. For the logarithmic kernel (1.10), we have $\Delta(-\log\|\mathbf{x}\|) = (2-d)/\|\mathbf{x}\|^2$, $\mathbf{x} \neq \mathbf{0}$ and the conclusions of the theorem remain valid for $d = 1, 2$, although the kernel is only CISPD.

1.9. Separable kernels. Consider the case of the so-called separable (tensor product) kernels

$$K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^d K_i(x_i, x_i')$$

on $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_d$, where the K_i are univariate bounded (C)ISPD kernels. The following properties can be verified.

- If each $\mathscr{E}_{K_i}(\cdot)$ is strictly convex on $\mathscr{M}(1) \cap \mathscr{M}_{K_i}$ for $\mathscr{X}_i$, then $\mathscr{E}_K(\cdot)$ is strictly convex on $\mathscr{M}(1) \cap \mathscr{M}_K$ for $\mathscr{X}$; see [14].
- The minimum-energy probability measure μ_K^+ is the product of univariate minimum-energy probability measures $\mu_{K_i}^+$: $\mu_K^+(\mathrm{d}\mathbf{x}) = \prod_{i=1}^d \mu_{K_i}^+(\mathrm{d}x_i)$.
- If the minimum-energy signed measure $\mu_{K_i}^*$ for K_i on $\mathscr{X}_i$ exists for each i, then the minimum-energy signed measure μ_K^* for K on $\mathscr{X}$ exists and equals $\mu_K^*(\mathrm{d}\mathbf{x}) = \prod_{i=1}^d \mu_{K_i}^*(\mathrm{d}x_i)$.
- If, for each i, the minimum-energy signed measure $\mu_{K_i}^*$ for K_i on $\mathscr{X}_i$ exists and coincides with $\mu_{K_i}^+$, the minimum-energy probability measure for K_i on $\mathscr{X}_i$, then μ_K^* for K on $\mathscr{X}$ exists and coincides with μ_K^+, the minimum-energy probability measure for K on $\mathscr{X}$.

1.10. Numerical construction of minimum-energy signed measures for bounded kernels: discrete case. Assume that the set $\mathscr{X}$ is discrete: $\mathscr{X} = \mathscr{X}_N = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and the kernel K is SPD (and thus bounded). Let $\mathbf{1}_N = (1, 1, \ldots, 1)^T$ be the vector of ones of size N and $\mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\ldots,N}$. The energy of the measure ζ_N that assigns weights w_j to the points $\mathbf{x}_j$ $(j = 1, \ldots, N)$ is then $\mathscr{E}_K(\zeta_N) = \Phi(\mathbf{w}_N) = \mathbf{w}_N^T \mathbf{K} \mathbf{w}_N$, where $\mathbf{w}_N = (w_1, \ldots, w_N)^T$. The vector of weights corresponding to the minimum-energy signed measure ζ_N^* of total mass one can be easily computed:

$$\mathbf{w}_N^* = \mathbf{K}^{-1}\mathbf{1}_N / \left(\mathbf{1}_N^T \mathbf{K}^{-1} \mathbf{1}_N\right), \tag{1.14}$$

giving

$$\Phi(\mathbf{w}_N^*) = \min_{\mathbf{w}} \Phi(\mathbf{w}) = 1/\left(\mathbf{1}_N^T \mathbf{K}^{-1} \mathbf{1}_N\right) ,$$

where the minimum is taken over all vectors $\mathbf{w} = (w_1, \dots, w_N)^T$ with $\mathbf{1}_N^T \mathbf{w} = w_1 + \dots + w_N = 1$. The potential of ζ_N^* is the vector with constant entries $\mathbf{p}_{\mathbf{w}_N^*} = \mathbf{K}\mathbf{w}_N^* = \mathbf{1}_N/\left(\mathbf{1}_N^T \mathbf{K}^{-1} \mathbf{1}_N\right)$.

1.11. Numerical construction of optimal measures for SPD kernels: general case. We approximate a general set $\mathscr{X}$ with a finite set $\mathscr{X}_N = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ and in this way we approximate the original problem of finding the optimal measure ζ^* (either signed or probability) minimizing the energy $\mathscr{E}_K(\mu)$ with the discrete problem of minimizing $\Phi(\mathbf{w}) = \mathbf{w}^T \mathbf{K} \mathbf{w}$, where $\mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1,\dots,N}$. In both classes of discrete measures (signed or probability), this problem has a unique solution; in the class of signed measures, this solution is given by (1.14). As shown below, in many interesting cases the optimal (discrete) measure ζ_N^* is a probability measure, and it is easy to construct an accurate approximation of the optimal continuous measure ζ^* from ζ_N^*; for example, in the case $\mathscr{X} \subset \mathbb{R}$ we can build either piece-wise constant, or continuous piece-wise linear, approximations.

1.12. Motivation behind the research and the structure of the rest of the paper. Motivation behind this research is two-fold.

(1) For any PD kernel K and a probability measure ν, minus energy $-\mathscr{E}_K(\nu)$ is the so-called Rao's quadratic entropy introduced and studied by C.R. Rao, see e.g. [7, 8]. This entropy has wide-spread applications in many applied fields, especially in biology. Maximum-entropy (and hence minimum-energy) measures are very natural objects to study for different classes of kernels including singular ones. The authors have considered minimum-energy measures for the case of Riesz kernels in several previous papers, see for example, [11, 12, 6]. The present paper broadly extends this research.

(2) Consider the MMD $\gamma_K(\mu, \nu)$ defined in (1.4) in the case when the measure μ is uniform on $\mathscr{X}$. Methods of approximate computation of MMD $\gamma_K(\mu, \nu)$ and sequential minimization of this MMD (which is the energy for the measure $\mu - \nu$) with respect to ν in the case of singular kernels is important for the methodology of construction of space-filling sequences of points, as singularity of the kernel guarantees automatic repelling property of these points. This methodology is the subject matter of a recent paper [6] by the present authors.

The structure of the rest of the paper is as follows. Section 2 contains the main results and describes the principal algorithmic schemes. This section has five parts.

Section 2.1 summarizes properties of CM and Bernstein functions required later on. It also relates CM functions to PD and ISPD kernels. In Section 2.2 we develop our main approximation of a CM function with singularity at 0 by bounded CM functions and in Theorem 2.5 we study properties of the family of these functions. In Section 2.3 we discuss properties of the pre-Hilbert space associated with a singular kernel and relate this space to the RKHS of the PD kernels constructed from CM functions of Theorem 2.5. In Sections 2.4 and 2.5 we develop an algorithm for approximating minimum-energy measures of singular kernels. This algorithm is based on the following: (a) the methodology developed in Section 2.2; (b) discretization of the set $\mathscr{X}$; (c) method of numerical construction of minimum-energy signed measures for bounded kernels in discrete case, formulated in Section 1.10; and (d) using

piecewise constant functions for approximating densities of the minimum-energy measures. In Section 3, the methodology of Section 2 is further detailed and numerically investigated for the case of Riesz kernels on $\mathscr{X} = [0,1]$.

2. Approximation of a CM function with singularity at 0 by a sequence of bounded CM functions.

2.1. CM functions and Bernstein functions (BF). As formulated in Section 1.1, a function $f : (0,\infty) \to \mathbb{R}^+$ is completely monotone (CM) if $f \in C^\infty(0,\infty)$ and (1.5) holds. If f is a non-constant CM function, then the inequalities (1.5) are necessarily strict for all $t > 0$ and all $k = 1, 2, \ldots$; see [10, Remark 1.5].

In this section, we formulate several important auxiliary results on completely monotone and Bernstein functions.

Theorem 2.1. [10, Th. 1.4] **(Bernstein).** *f is a CM function if and only if it is the Laplace transform of a nonnegative Borel measure μ on $[0,\infty)$:*

$$f(x) = \int_0^\infty e^{-tx} \mu(\mathrm{d}t)\,. \tag{2.1}$$

Theorem 2.2. [10, Th. 3.2] **(Levy-Khinchine representation).** *A function $g : (0,\infty) \to \mathbb{R}^+$ is a BF if and only if*

$$g(x) = a + b\,x + \int_0^\infty (1 - e^{-tx})\,\nu(\mathrm{d}t) \tag{2.2}$$

where $a, b \geq 0$ and ν is a nonnegative measure on $(0,\infty)$ with $\int_0^\infty \min(1,t)\,\nu(\mathrm{d}t) < \infty$. The triplet (a,b,ν) uniquely determines g and vice versa.

If g is a BF of the form (2.2), then $f = g'$ is CM with the measure μ of (2.1) being

$$\mu(\mathrm{d}t) = b\,\delta_0(\mathrm{d}t) + t\,\nu(\mathrm{d}t)\,, \tag{2.3}$$

where $\delta_0(\mathrm{d}t)$ is the delta-measure concentrated at 0. According to [10, Proposition 3.4], for a completely monotone function f with measure μ, there exists a BF g such that $f = g'$ if and only if the measure μ of (2.1) satisfies

$$\int_0^\infty \frac{1}{1+t}\,\mu(\mathrm{d}t) < \infty\,. \tag{2.4}$$

In this case, we can set $g(t) = \int_0^t f(u)\mathrm{d}u$.

The relation between non-constant CM functions and SPD kernels is characterised by the following result, essentially proved by Shoenberg in [13].

Theorem 2.3. [16, Th. 7.13 & 7.14] *Let $\psi : [0,\infty) \to \mathbb{R}$ be a non-constant bounded function and define the kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by $K(\mathbf{x},\mathbf{x}') = \psi(\|\mathbf{x}-\mathbf{x}'\|^2)$, where $\|\cdot\|$ is the Euclidean norm. Then ψ is CM if and only if the kernel K is SPD for any $d = 1, 2, \ldots$*

Note that Theorem 2.1 and the fact that the exponential kernel $K_{1/2,\beta}(\mathbf{x},\mathbf{x}') = \exp(-\beta\|\mathbf{x}-\mathbf{x}'\|)$, $\beta > 0$, is ISPD, implies that if f is a non-constant, CM and bounded function, then the kernel $K(\mathbf{x},\mathbf{x}') = f(\|\mathbf{x}-\mathbf{x}'\|)$ is bounded and ISPD (and therefore SPD). Also, as shown in [10], having the kernel $K(\mathbf{x},\mathbf{x}') = -\psi(\|\mathbf{x}-\mathbf{x}'\|^2)$ CISPD for a continuous ψ requires ψ to be a BF.

We also have the following property for unbounded kernels constructed from CM functions.

Theorem 2.4. [3, Corollary 8] *Let $f\colon (0,\infty)\to\mathbb{R}$ be an L_1 CM function, possibly with singularity at zero. Then the kernel $K:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ defined by $K(x,x')=f(|x-x'|)$ is ISPD.*

Unlike Theorem 2.3, Theorem 2.4 deals with one-dimensional case only. The authors are unaware of any generalization of Theorem 2.4 to kernels in $\mathbb{R}^d$.

2.2. Approximating family. Assume that a non-constant function f is CM with the measure μ of (2.1) satisfying (2.4). Then $f=g'$, where $g(t)=\int_0^t f(u)\mathrm{d}u$ is BF. The value $f(0)$ may be undefined; that is, $f(0+)=\lim_{t\to 0+} f(t)\le+\infty$. Our aim is to construct a family of CM functions f_ε such that $f_\varepsilon(0+)=\lim_{t\to 0+} f_\varepsilon(t)<\infty$ for all $\varepsilon>0$ and $\lim_{\varepsilon\to 0} f_\varepsilon(t)=f(t)$ for all $t>0$.

The family is given by

$$f_\varepsilon(t)=\begin{cases} f(t)=g'(t), & \varepsilon=0\\ \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(s)\,\mathrm{d}s=\frac{1}{\varepsilon}\left[g(t+\varepsilon)-g(t)\right], & \varepsilon>0\,. \end{cases} \tag{2.5}$$

It satisfies the following important properties.

Theorem 2.5. *Let f be a non-constant CM function with the measure μ of (2.1) satisfying (2.4). Consider the family of functions (2.5), where $g(t)=\int_0^t f(u)\mathrm{d}u$. Then*

(i) *the functions f_ε are CM for all $\varepsilon\ge 0$,*

(ii) *the functions $f_\delta-f_\varepsilon$ are CM for all $0\le\delta<\varepsilon$,*

(iii) *functions $\varepsilon f_\varepsilon-\delta f_\delta$ are CM for all $0<\delta<\varepsilon$.*

Proof. (i) Consider the form (2.2) for the function g. Then, for any $\varepsilon>0$, the function f_ε from (2.5) can be written as

$$f_\varepsilon(x)=b+\frac{1}{\varepsilon}\int_0^\infty e^{-tx}(1-e^{-\varepsilon t})\,\nu(\mathrm{d}t)\,. \tag{2.6}$$

Another form of (2.6) is

$$f_\varepsilon(x)=\int_0^\infty e^{-tx}\mu_\varepsilon(\mathrm{d}t)\,, \tag{2.7}$$

where $\mu_\varepsilon(\mathrm{d}t)=b\,\delta_0(\mathrm{d}t)+h_\varepsilon(t)\,\nu(\mathrm{d}t)$ and

$$h_\varepsilon(t)=\begin{cases} t & \text{for } \varepsilon=0\,,\\ \frac{1}{\varepsilon}\,(1-e^{-\varepsilon t}) & \text{for } \varepsilon>0\,; \end{cases} \tag{2.8}$$

the expression for $h_0(t)$ follows from (2.3).

Since $h_\varepsilon(t)>0$ for all $\varepsilon\ge 0$ and $t>0$, Theorem 2.1 implies that the functions f_ε are CM for all $\varepsilon\ge 0$.

(ii) Assume $0<\delta<\varepsilon$ and consider (2.7) and similar representation for f_δ. Then for all $t>0$ we have

$$f_\delta(x)-f_\varepsilon(x)=\int_0^\infty e^{-tx}\left[\mu_\delta-\mu_\varepsilon\right](\mathrm{d}t) \tag{2.9}$$

where

$$\left[\mu_\delta-\mu_\varepsilon\right](\mathrm{d}t)=\left[h_\delta(t)-h_\varepsilon(t)\right]\nu(\mathrm{d}t)\,. \tag{2.10}$$

The measure $[h_\delta(t)-h_\varepsilon(t)]\,\nu(\mathrm{d}t)$ is positive since, for any $t>0$, we have $h_\delta(t)-h_\varepsilon(t)>0$. Indeed, for any fixed $t>0$, the function $h_\varepsilon(t)$, considered as a function of

$\varepsilon \in (0,\infty)$, is strictly positive and strictly decreasing; the former has been noted in the proof of (i) and the latter follows from

$$\frac{\partial h_\varepsilon(t)}{\partial \varepsilon} = \frac{(1+\varepsilon t)\, e^{-\varepsilon t} - 1}{\varepsilon^2}$$

and the easily verifiable fact that the function $\omega(s) = (1+s)e^{-s} - 1$ is strictly negative for all $s > 0$.

Consider now the case $\delta = 0$. The expressions (2.9) and (2.10) are still valid but now we need to justify that

$$h_0(t) - h_\varepsilon(t) = t - \frac{1}{\varepsilon}\left(1 - e^{-\varepsilon t}\right) = \frac{t\varepsilon + e^{-\varepsilon t} - 1}{\varepsilon}$$

is positive for all $\varepsilon > 0$ and all $t > 0$. This follows from the fact that the function $\omega_0(s) = s + e^{-s} - 1$ is strictly positive for all $s > 0$.

(iii) Using (2.8), we obtain similarly to (2.9) and (2.10):

$$\delta f_\delta(x) - \varepsilon f_\varepsilon(x) = \int_0^\infty e^{-tx} \left[\delta\mu_\delta - \varepsilon\mu_\varepsilon\right](\mathrm{d}t)$$

with

$$\begin{aligned}\left[\delta\mu_\delta - \varepsilon\mu_\varepsilon\right](\mathrm{d}t) &= b(\varepsilon-\delta)\delta_0(\mathrm{d}t) + \left[\delta h_\delta(t) - \varepsilon h_\varepsilon(t)\right]\nu(\mathrm{d}t) \\ &= b(\varepsilon-\delta)\delta_0(\mathrm{d}t) + \left[e^{-\varepsilon t} - e^{-\delta t}\right]\nu(\mathrm{d}t)\,.\end{aligned}$$

In view of Theorem 2.2, $b \geq 0$. As $0 < \delta < \varepsilon$, we have $b(\varepsilon - \delta) \geq 0$ and $e^{-\varepsilon t} - e^{-\delta t} > 0$ for all $t > 0$. Therefore, the measure $\left[\delta\mu_\delta - \varepsilon\mu_\varepsilon\right](\mathrm{d}t)$ is positive and hence, by Theorem 2.1, the function $\delta f_\delta - \varepsilon f_\varepsilon$ is CM. □

Four simple examples of CM functions and approximating families are given below. Note that in examples (a) and (b) the function $f(t)$ is bounded at 0 and $0 < \alpha < 1$ in examples (b)–(d).

(a) $f(t) = e^{-\lambda t}$, $\lambda > 0$: $f_\varepsilon(t) = c_{\varepsilon,\lambda} f(t)$ with $c_{\varepsilon,\lambda} = (1 - e^{-\lambda\varepsilon})/(\lambda\varepsilon)$.
(b) $f(t) = 1/(1+t)^{1+\alpha}$: $f_\varepsilon(t) = \left[(1+t)^{-\alpha} - (1+t+\epsilon)^{-\alpha}\right]/(\alpha\,\epsilon)$.
(c) $f(t) = 1/t^{1-\alpha}$: $f_\varepsilon(t) = ((t+\varepsilon)^\alpha - t^\alpha)/(\alpha\varepsilon)$.
(d) $f(t) = 1/(t^{1-\alpha}(1+t)^{1+\alpha})$: $f_\varepsilon(t) = \left[(1+1/(t+\varepsilon))^{-\alpha} - (1+1/t)^{-\alpha}\right]/(\alpha\varepsilon)$.

2.3. Reproducing kernel pre-Hilbert space associated with a singular kernel. Let $\mathscr{X} = [0,1]$ and f be a non-degenerate CM function with singularity at 0. Assume, like in Section 2.2, that the measure μ of (2.1) satisfies (2.4). Define the kernel $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ by the formula (1.6).

Consider the family of functions $\{f_\varepsilon\}_{\varepsilon>0}$ defined by (2.5). In view of Theorem 2.5 all these functions are CM and bounded. Using the construction of Theorem 2.4, create the SPD kernels

$$K_\varepsilon(x,x') = f_\varepsilon(|x - x'|)\,. \tag{2.11}$$

With each kernel $K_\varepsilon(\cdot,\cdot)$, $\varepsilon > 0$, we associate $\mathcal{H}(K_\varepsilon)$, the RKHS (Reproducing Kernel Hilbert Space). All these RKHS $\mathcal{H}(K_\varepsilon)$, $\varepsilon > 0$, are equivalent since, according to (ii) and (iii) of Theorem 2.5, $f_\delta - f_\varepsilon$ and $(\varepsilon/\delta)f_\varepsilon - f_\delta$ are CM functions for $0 < \delta < \varepsilon$. Here we have also used the following result [1, Th. 12]: $\mathcal{H}(K) \subset \mathcal{H}(K')$ if and only if there exists constant $c > 0$ such that $cK' - K$ is a positive definite kernel.

Denote the scalar product in $\mathcal{H}(K_\varepsilon)$ by $\langle\cdot,\cdot\rangle_\varepsilon$. For a function $g \in \mathcal{H}(K_\varepsilon)$, consider the norms $\|g\|_\varepsilon = \sqrt{\langle g, g\rangle_\varepsilon}$ with $\varepsilon > 0$. As the set of potentials $P_\nu(\cdot)$ with $\nu \in \mathscr{M}$ is dense in $\mathcal{H}(K_\varepsilon)$, we may only consider functions $g = P_\nu(\cdot)$, $\nu \in \mathscr{M}$. In view of Theorem 2.5, the functions f_ε, $f_{\varepsilon'}$ and $f_{\varepsilon'} - f_\varepsilon$ are bounded and CM for $0 < \varepsilon' < \varepsilon$. Therefore, for functions $g = P_\nu(\cdot)$ with $\nu \in \mathscr{M}$ we have $\|g\|^2_\varepsilon = \mathscr{E}_{K_\varepsilon}(\nu)$, $\|g\|^2_{\varepsilon'} = \mathscr{E}_{K_{\varepsilon'}}(\nu)$ and $\|g\|^2_{\varepsilon'} - \|g\|^2_\varepsilon = \mathscr{E}_{K_{\varepsilon,\varepsilon'}}(\nu) \geq 0$, where $K_{\varepsilon,\varepsilon'}$ is the kernel $K_{\varepsilon,\varepsilon'}(x,x') = f_{\varepsilon'}(|x-x'|) - f_\varepsilon(|x-x'|)$.

As for all $g \in \mathcal{H}(K_\varepsilon)$ we have $\|g\|_\varepsilon \leq \|g\|_{\varepsilon'}$ for $0 \leq \varepsilon' \leq \varepsilon$, there is a limit

$$\|g\|_0 = \lim_{\varepsilon\to 0} \|g\|_\varepsilon \leq \infty\,.$$

We can then define the set

$$\mathcal{H}(K) = \{g \in \mathcal{H}(K_\varepsilon) \text{ for some } \varepsilon > 0\colon\ \|g\|_0 < \infty\}\,.$$

The scalar product in $\mathcal{H}(K)$ is defined as the limit: $\langle g, g'\rangle_0 = \lim_{\varepsilon\to 0}\langle g, g'\rangle_\varepsilon$ for $g, g' \in \mathcal{H}(K)$.

The set $\mathcal{H}(K)$ is not a Hilbert space as $\mathcal{H}(K)$ is not complete. It is, however, a pre-Hilbert space and can be referred to as "Reproducing kernel pre-Hilbert space associated with kernel (1.6)". Note that the set $\mathcal{H}(K)$ is not empty as it contains all potentials $P_\nu(\cdot)$, see (1.3), of measures $\nu \in \mathscr{M}_K$, where $\mathscr{M}_K$ is the set of signed measures with finite energy, see (1.2). Note also that, unlike the spaces $\mathcal{H}(K_\varepsilon)$ with $\varepsilon > 0$, the space $\mathcal{H}(K)$ does not contain potentials of delta-measures δ_x, $x \in \mathscr{X}$.

Similarly to the discussion above we can define a reproducing kernel pre-Hilbert space associated with kernel $K(\mathbf{x},\mathbf{x}') = f(\|\mathbf{x}-\mathbf{x}'\|^2)$, constructed in Theorem 2.3, through the RKHS of the bounded kernels $K_\varepsilon(\mathbf{x},\mathbf{x}') = f_\varepsilon(\|\mathbf{x}-\mathbf{x}'\|^2)$. Here $\mathbf{x}, \mathbf{x}' \in \mathscr{X}$, a compact subset of $\mathbb{R}^d$; $d \geq 1$.

2.4. Discrete approximations. In this section, we apply the methodology of Section 1.10 and construct discrete approximations to minimum-energy signed measures of total mass one on $\mathscr{X} = [0,1]$ for kernels K_ε defined by (2.11) with f_ε constructed as in Section 2.2.

Take $\varepsilon > 0$ and an integer N. Choose N design points $x_1, \ldots, x_N$ in $[0,1]$ (for example, set $x_k = (k-1)/(N-1)$, $k = 1, \ldots, N$), form the matrix $\mathbf{K}_{\varepsilon,N} = (f_\varepsilon(|x_i - x_j|))_{i,j=1,\ldots,N}$ and compute the minimum-energy signed measure $\zeta^*_{\varepsilon,N}$ in $\mathscr{M}(1)$ supported on the x_i. The optimal weights are given by (1.14),

$$\mathbf{w}^*_{\varepsilon,N} = \mathbf{K}^{-1}_{\varepsilon,N}\mathbf{1}_N \big/ \left(\mathbf{1}^T_N \mathbf{K}^{-1}_{\varepsilon,N}\mathbf{1}_N\right)\,, \tag{2.12}$$

and the minimum value of the discrete energy $\Phi_{\varepsilon,N}(\mathbf{w}) = \mathbf{w}^T\mathbf{K}_{\varepsilon,N}\mathbf{w}$ for measures of total mass one is

$$\mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N}) = \Phi_{\varepsilon,N}(\mathbf{w}^*_{\varepsilon,N}) = 1/\left(\mathbf{1}^T_N \mathbf{K}^{-1}_{\varepsilon,N}\mathbf{1}_N\right)\,.$$

For a fixed $\varepsilon > 0$, the kernel K_ε is ISPD, see Section 2.1. Since the function f_ε is convex, Theorem 1.3 implies that the minimum-energy signed measure $\mu^*_{K_\varepsilon}$ in $\mathscr{M}(1)$ exists and is a probability measure, obtained as the solution of condition (1.13) in Theorem 1.2. Therefore, if the design points x_i are such that $\max_{x\in[0,1]} \min_i |x - x_i| \to$

0 as $N \to \infty$, we have $\mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N}) \to \mathscr{E}_{K_\varepsilon}(\mu^*_{K_\varepsilon})$. This implies that

$$\begin{aligned}\gamma^2_{K_\varepsilon}(\zeta^*_{\varepsilon,N}, \mu^*_{K_\varepsilon}) &= \mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N} - \mu^*_{K_\varepsilon}) \\ &= \mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N}) + \mathscr{E}_{K_\varepsilon}(\mu^*_{K_\varepsilon}) - 2\sum_{i=1}^N w^*_{\varepsilon,N\,i} \int K_\varepsilon(x_i, x)\, \mu^*_{K_\varepsilon}(\mathrm{d}x) \\ &= \mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N}) - \mathscr{E}_{K_\varepsilon}(\mu^*_{K_\varepsilon}) \to 0 \text{ as } N \to \infty\,.\end{aligned}$$

Moreover, when the x_i are obtained from an infinite sequence $\{x_1, x_2, \ldots\}$ satisfying $\max_{x\in[0,1]} \min_i |x - x_i| \to 0$ as $N \to \infty$, the convergence is monotone; that is, $\gamma_{K_\varepsilon}(\zeta^*_{\varepsilon,N}, \mu^*_{K_\varepsilon}) \downarrow 0$. However, for any fixed N, $\mathscr{E}_{K_\varepsilon}(\zeta^*_{\varepsilon,N}) = \Phi_{\varepsilon,N}(\mathbf{w}^*_{\varepsilon,N})$ tends to ∞ as $\varepsilon \to 0$. As discrete measures cannot be used for kernels with singularities, we shall use an absolutely continuous version of $\zeta^*_{\varepsilon,N}$ having weights (2.12). The construction is described in the next section.

2.5. Approximations with piecewise constant densities. Let $0 \le x_1 < \ldots < x_N \le 1$ be the support points of a discrete probability measure ζ_N and $w_k \ge 0$ $(k = 1, \ldots, N)$ be the corresponding weights with $\sum_{k=1}^N w_k = 1$. Define the $N+1$ points z_i by

$$z_j = (x_j + x_{j+1})/2,\ j = 1, \ldots, N-1 \text{ and } z_0 = 0,\ z_N = 1\,.$$

We partition the interval $[0, 1)$ into N non-intersecting intervals $I_j = [z_{j-1}, z_j)$ $(j = 1, \ldots, N)$, with respective lengthes $l_j = z_j - z_{j-1}$. We have $l_j > 0$ for all $j = 1, \ldots, N$ and $\sum_j l_j = 1$. Define the piecewise constant function

$$p_N(t) = \begin{cases} w_j/l_j & \text{if } t \in I_j \qquad \text{for some } j = 1, \ldots, N \\ 0 & \text{if } t \notin [0,1)\,. \end{cases} \tag{2.13}$$

We have $p_N(t) \ge 0$ for all t and $\int_0^1 p_N(t)\mathrm{d}t = \sum_{j=1}^N w_j = 1$ and therefore p_N is a probability density function. We shall use it as a continuous approximation to ζ_N. Note that if $x_j = (j-1)/(N-1)$ $(j = 1, \ldots, N)$, then $l_1 = l_N = 1/[2(N-1)]$ whereas $l_i = 1/(N-1)$ for $i = 2, \ldots, N-1$.

Denote by $\widetilde{\zeta}_N$ the measure having the density (2.13). Then its energy (1.1) for $K(x, y) = f(|x - y|)$ can be written as

$$\begin{aligned}\mathscr{E}_K(\widetilde{\zeta}_N) &= \int_0^1 \int_0^1 K(x,y) p_N(x) p_N(y) \mathrm{d}x\, \mathrm{d}y \\ &= 2\sum_{i=2}^N \frac{w_i}{l_i} \int_{I_i} \left[\sum_{j=1}^{i-1} \frac{w_j}{l_j} \int_{I_j} f(x-y)\mathrm{d}y \right] \mathrm{d}x + \sum_{i=1}^N \frac{w_i^2}{l_i^2} \int_{I_i}\int_{I_i} f(|x-y|)\mathrm{d}x\mathrm{d}y \\ &= \mathbf{w}_N^T \widetilde{\mathbf{K}}_N \mathbf{w}_N,\end{aligned}$$

where $\mathbf{w}_N = (w_1, \ldots, w_N)^T$ and $\widetilde{\mathbf{K}}_N$ is symmetric with

$$(\widetilde{\mathbf{K}}_N)_{i,i} = \frac{1}{l_i^2} \int_{z_{i-1}}^{z_i} \left[\int_{z_{i-1}}^{x} f(x-y)dy + \int_x^{z_i} f(y-x)\mathrm{d}y \right] \mathrm{d}x \quad \text{and} \tag{2.14}$$

$$(\widetilde{\mathbf{K}}_N)_{j,i} = \frac{1}{l_i l_j} \int_{z_{i-1}}^{z_i} \left[\int_{z_{j-1}}^{z_j} f(x-y)\mathrm{d}y \right] \mathrm{d}x \quad \text{for } j < i\,. \tag{2.15}$$

We can therefore obtain $\mathscr{E}_K(\widetilde{\zeta}_N)$ in closed-form when closed-form expressions for

$$\int_a^b \left[\int_a^t f(t-s)\mathrm{d}s\right]\mathrm{d}t \quad \text{and} \quad \int_c^d \left[\int_a^b f(t-s)\mathrm{d}s\right]\mathrm{d}t \tag{2.16}$$

are available for any $0 \le a \le b \le c \le d \le 1$. In that case, if there exists a minimum-energy measure $\mu_K^* \in \mathscr{M}(1)$ on $[0,1]$ for the kernel K, the minimum energy measure $\widetilde{\zeta}_N^*$ of total mass one, for K, among those having piecewise constant densities $p_N(t)$ on the partition above, gives an approximation of μ_K^*. The optimal measure $\widetilde{\zeta}_N^*$, with density $p_N^*(t)$, is characterized by the weights $\widetilde{\mathbf{w}}_N^* = \widetilde{\mathbf{K}}_N^{-1}\mathbf{1}_N/(\mathbf{1}_N^T\widetilde{\mathbf{K}}_N^{-1}\mathbf{1}_N)$. Also, developments similar to those in Section 2.4 give

$$\gamma_K^2(\widetilde{\zeta}_N^*, \mu_K^*) = \mathscr{E}_K(\widetilde{\zeta}_N^*) - \mathscr{E}_K(\mu_K^*)\,, \tag{2.17}$$

which is well defined for singular kernels.

The potential $\widetilde{P}_N(x)$ for the measure with density $p_N(t)$ can be computed in a similar way for x in any I_j, $j = 1, \ldots, N$. Indeed, for any $x \in I_j = [z_j, z_{j+1})$, we have

$$\begin{aligned}
\widetilde{P}_N(x) &= \int_0^1 K(x,y)p_N(y)\mathrm{d}y = \sum_{i=1}^N \frac{w_i}{l_i}\int_{I_i} f(|x-y|)\mathrm{d}y \\
&= \sum_{i=1}^{j-1}\frac{w_i}{l_i}\int_{z_i}^{z_{i+1}} f(x-y)\mathrm{d}y + \frac{w_j}{l_j}\int_{z_j}^{z_{j+1}} f(|x-y|)\mathrm{d}y + \sum_{k=j+1}^{N}\frac{w_k}{l_k}\int_{z_k}^{z_{k+1}} f(y-x)\mathrm{d}y \\
&= \sum_{i=1}^{j-1}\frac{w_i}{l_i}\int_{z_i}^{z_{i+1}} f(x-y)\mathrm{d}y + \sum_{k=j+1}^{N}\frac{w_k}{l_k}\int_{z_k}^{z_{k+1}} f(y-x)\mathrm{d}y \\
&\quad + \frac{w_j}{l_j}\left(\int_{z_j}^{x} f(x-y)\mathrm{d}y + \int_{x}^{z_{j+1}} f(y-x)\mathrm{d}y\right).
\end{aligned}$$

This expression for $\widetilde{P}_N(x)$ can be used in particular to check how close the potential $\widetilde{P}_N^*(x)$ associated with $p_N^*(t)$ is to being constant for $x \in [0,1]$.

The construction above can be applied to the discrete approximations $\zeta_{\varepsilon,N}^*$ of Section 2.4. We denote by $\widetilde{\zeta}_{\varepsilon,N}$ the measure with piecewise constant density and weights $\mathbf{w}_{\varepsilon,N}^*$ given by (2.12), and by $\widetilde{\zeta}_{\varepsilon,N}^*$ the measure with piecewise constant density having weights $\widetilde{\mathbf{w}}_{\varepsilon,N}^*$ obtained with the kernel $\widetilde{K}_{\varepsilon,N}$ computed with (2.14, 2.15) where f is replaced by f_ε, $\varepsilon \ge 0$. When $K(x,x') = f(|x-x'|)$ with f singular at zero and μ_K^* exists, we can compute the MMD discrepancies $\gamma_K(\widetilde{\zeta}_{\varepsilon,N}, \mu_K^*)$ and $\gamma_K(\widetilde{\zeta}_{\varepsilon,N}^*, \mu_K^*)$, see (2.17), and plot them as functions of ε.

3. Case study: Riesz kernels on $[0,1]$.

3.1. Riesz kernels and associated optimal measures. Consider the function $f(t) = t^{-\alpha}$ $(0 < \alpha < 1)$, on $t \in (0,1]$ and the associated kernel

$$K(t,s) = f(|t-s|) = |t-s|^{-\alpha}, \quad t,s \in [0,1]\,, \quad t \ne s. \tag{3.1}$$

In this case, the minimizing measure μ^* for the energy functional (1.1) is known. In fact, it is a probability measure with density

$$\phi_\alpha(t) = c_\alpha \cdot [t(1-t)]^{(\alpha-1)/2}\,, \quad t \in [0,1]\,, \quad c_\alpha = \frac{\Gamma(\alpha+1)}{\left[\Gamma(\frac{\alpha+1}{2})\right]^2}\,. \tag{3.2}$$

That is, μ^* has Beta-distribution with parameters $\frac{\alpha+1}{2}, \frac{\alpha+1}{2}$ on $[0,1]$.

We have:

$$\int_0^1 K(t,s)\phi_\alpha(s)\mathrm{d}s = \int_0^t \frac{1}{(t-s)^\alpha}\phi_\alpha(s)\mathrm{d}s + \int_t^1 \frac{1}{(s-t)^\alpha}\phi_\alpha(s)\mathrm{d}s = c_\alpha \frac{\pi}{\cos(\pi\alpha/2)} = \Phi^*_\alpha\,,$$

where

$$\Phi^*_\alpha = \frac{\pi\Gamma(\alpha+1)}{\cos(\pi\alpha/2)\,\left[\Gamma(\frac{\alpha+1}{2})\right]^2} = \min_{\mu\in\mathscr{M}(1)} \Phi(\mu)\,.$$

Values of Φ^*_α are plotted in Fig 3.1, left; normalised values $(\Phi^*_\alpha)^{1-\alpha}$ are plotted in Fig 3.1, right.

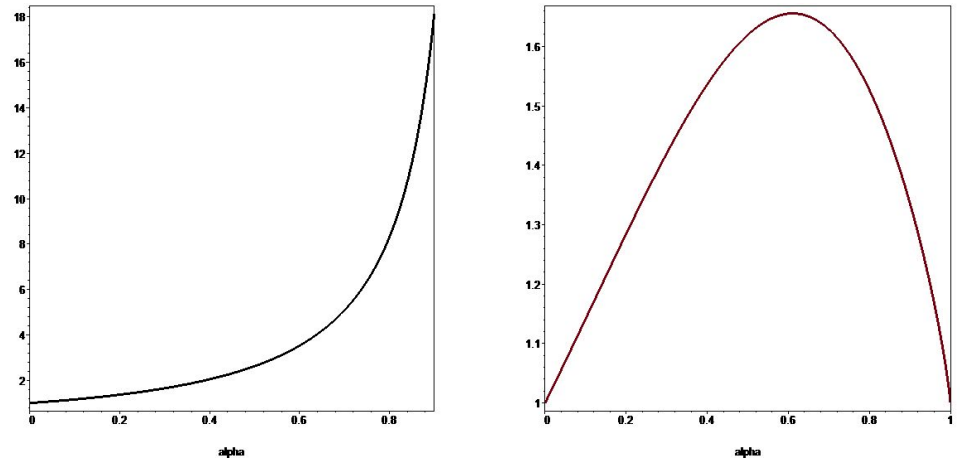

Figure 3.1: *Left: values of* Φ^*_α *for* $\alpha\in[0,0.9]$. *Right: values of* $(\Phi^*_\alpha)^{1-\alpha}$ *for* $\alpha\in[0,1]$.

Note that for the kernel (3.1) we have the following exact formulas for (2.16):

$$\int_a^b \left[\int_a^t (t-s)^{-\alpha}\mathrm{d}s\right]\mathrm{d}t = \frac{(b-a)^{2-\alpha}}{(1-\alpha)(2-\alpha)}\,;$$
$$\int_c^d \left[\int_a^b (t-s)^{-\alpha}\mathrm{d}s\right]\mathrm{d}t = \frac{(d-a)^{2-\alpha}+(c-b)^{2-\alpha}-(d-b)^{2-\alpha}-(c-a)^{2-\alpha}}{(1-\alpha)(2-\alpha)}\,.$$

To compute the potentials for measures $\mu=\mu_N$ having the density (2.13) we use

the following formulas for $x \in I_j = [z_j, z_{j+1})$:

$$P_{\mu_N}(x) = \int_0^1 K(x,y) p_N(y) \mathrm{d}y = \sum_{i=0}^{N} \frac{w_i}{l_i} \int_{I_i} f(|x-y|) \mathrm{d}y$$

$$= \sum_{i=0}^{j-1} \frac{w_i}{l_i} \int_{z_i}^{z_{i+1}} (x-y)^{-\alpha} \mathrm{d}y + \frac{w_j}{l_j} \int_{z_j}^{z_{j+1}} |x-y|^{-\alpha} \mathrm{d}y + \sum_{k=j+1}^{N} \frac{w_k}{l_k} \int_{z_k}^{z_{k+1}} (y-x)^{-\alpha} \mathrm{d}y$$

$$= \frac{1}{1-\alpha} \left[\sum_{i=0}^{j-1} \frac{w_i}{l_i} \left[(x-z_i)^{1-\alpha} - (x-z_{i+1})^{1-\alpha} \right] + \frac{w_j}{l_j} \left[(x-z_j)^{1-\alpha} + (z_{j+1}-x)^{1-\alpha} \right] \right.$$

$$\left. + \sum_{k=j+1}^{N} \frac{w_k}{l_k} \left[(z_{k+1}-x)^{1-\alpha} - (z_k - x)^{1-\alpha} \right] \right] .$$

3.2. Efficiency of the uniform probability measure. Let μ_0 be the uniform probability distribution on $[0,1]$. We define the efficiency of μ_0 as

$$\mathrm{eff}(\mu_0) = \frac{\Phi_\alpha^*}{\Phi_\alpha(\mu_0)} = \frac{(1-\alpha)(2-\alpha)\cos(\pi\alpha/2)\ \left[\Gamma(\frac{\alpha+1}{2})\right]^2}{2\pi\Gamma(\alpha+1)} , \tag{3.3}$$

where

$$\Phi_\alpha(\mu_0) = \mathscr{E}_K(\mu_0) = \int_0^1 \int_0^1 |t-s|^{-\alpha} \mathrm{d}t \mathrm{d}s = \frac{2}{(1-\alpha)(2-\alpha)} \tag{3.4}$$

is the energy of the uniform measure. For all values of $\alpha \in [0,1)$ this efficiency is quite high, see Fig 3.2, left (the lowest value of this efficiency is $\simeq 0.98135$ which is achieved at $\alpha \simeq 0.36253$). The quality of approximations used for the uniform measure are therefore indicative of what is happening with approximations used for the optimal measures.

The potential of the uniform measure is

$$P_{\mu_0;\alpha}(t) = \int_0^1 |t-s|^{-\alpha} \mathrm{d}s = \int_0^t (t-s)^{-\alpha} \mathrm{d}s + \int_t^1 (s-t)^{-\alpha} \mathrm{d}s = \frac{t^{1-\alpha} + (1-t)^{1-\alpha}}{1-\alpha} .$$

This potential and its average value $\Phi_\alpha(\mu_0) = \int_0^1 P_{\mu_0;\alpha}(t) \mu_0(\mathrm{d}t)$ are plotted in Fig 3.2, right, for $\alpha = 0.5$. As can be seen from this figure, despite the uniform measure has high efficiency, there is still room for improvement, since the potential of the optimal measure is a constant function.

3.3. Approximation of f by f_ε. The Bernstein function associated with $f(t) = t^{-\alpha}$ is $g(t) = t^{1-\alpha}/(1-\alpha)$, $t \geq 0$, and the functions f_ε from (2.5) are

$$f_{\varepsilon,\alpha}(t) = \frac{(t+\varepsilon)^{1-\alpha} - t^{1-\alpha}}{\varepsilon(1-\alpha)} , \quad t \geq 0. \tag{3.5}$$

Let us study the quality of approximation of the kernel (3.1) by the family of kernels

$$K_{\varepsilon,\alpha}(t,s) = f_{\varepsilon,\alpha}(|t-s|), \quad t,s \in [0,1] , \tag{3.6}$$

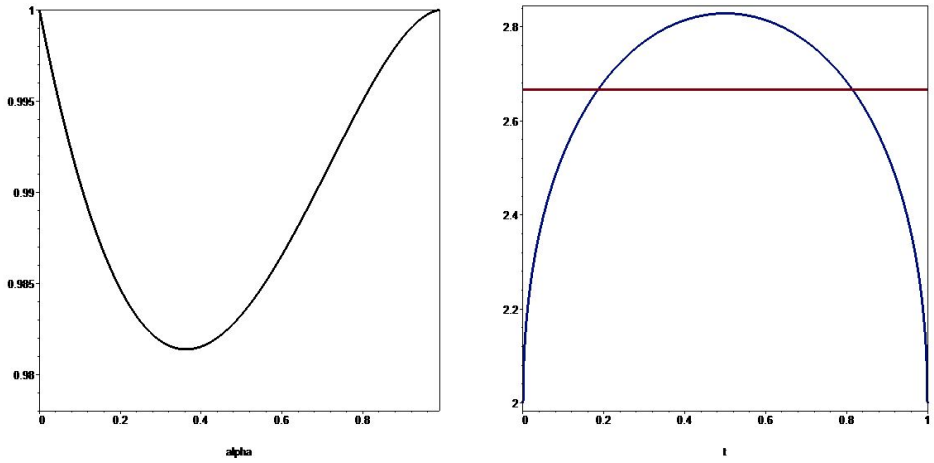

Figure 3.2: *Left: efficiency of the uniform measure computed by* (3.3) *for* $\alpha \in [0,1)$. *Right: Potential of the uniform measure* $P_{\mu_0;\alpha}(t)$ *and its average value* $\Phi_\alpha(\mu_0)$, *see* (3.4), *computed for* $\alpha = 0.5$.

where $f_{\varepsilon,\alpha}$ is given in (3.5). The energies of the uniform measure with respect to kernels (3.6) are

$$\Phi_{\alpha,\varepsilon}(\mu_0) = \int_0^1 \int_0^1 f_{\varepsilon,\alpha}(|t-s|)\mathrm{d}t\mathrm{d}s = 2\frac{(1+\varepsilon)^{3-\alpha} - \varepsilon^{3-\alpha} - (3-\alpha)\varepsilon^{2-\alpha} - 1}{\varepsilon(1-\alpha)(2-\alpha)(3-\alpha)}\,. \quad (3.7)$$

Since $f_{\varepsilon,\alpha}(t) < f(t) = t^{-\alpha}$ for all $\alpha \in (0,1)$ and $t > 0$, we have $\Phi_{\alpha,\varepsilon}(\mu_0) < \Phi_\alpha(\mu_0)$ for all $\alpha \in (0,1)$. Values of the ratio $\Phi_{\alpha,\varepsilon}(\mu_0)/\Phi_\alpha(\mu_0)$ are plotted in Fig 3.3, left. We can deduce from this figure that if α is not very close to 1 (that is, when singularity is not very strong) then $f_{\varepsilon,\alpha}$ can be considered as an accurate approximation to f, even for ε not very small. Note that the case when singularity of the kernel is strong (when α is close to 1) is not very interesting when K is used to model the covariance function of a random process, as this is very close to the case of no dependence (the white noise case), for which the minimum energy measure is the uniform measure.

Expanding rhs in (3.7) into a series we obtain

$$\Phi_{\alpha,\varepsilon}(\mu_0) = \Phi_\alpha(\mu_0)\left[1 - \varepsilon^{1-\alpha} + \varepsilon(1-\alpha/2)\right] + O(\varepsilon^2)\,, \quad \varepsilon \to 0\,.$$

The resulting approximation is very accurate for all $\alpha \in [0,1)$, even if ε is not very small. Already the very simple approximation

$$\Phi_{\alpha,\varepsilon}(\mu_0)/\Phi_\alpha(\mu_0) \simeq 1 - \varepsilon^{1-\alpha} \quad (\varepsilon \simeq 0) \quad (3.8)$$

is quite accurate, as can be seen from Fig 3.3, right.

The optimal density (3.2) and the approximation (2.13) obtained for the weights $\widetilde{\mathbf{w}}^*_{\varepsilon,N}$ on the uniform grid $x_k = k/N$ $(k = 0,1,\ldots,N = 200)$ are presented in Figs. 3.4 and 3.5 for different values of ε and α, illustrating the accuracy of the approximation.

Concluding this section we can state that if α is not too close to 1 (so that the singularity of the kernel is not too severe) then $f_{\varepsilon,\alpha}$ accurately approximates f, if ε is small enough.

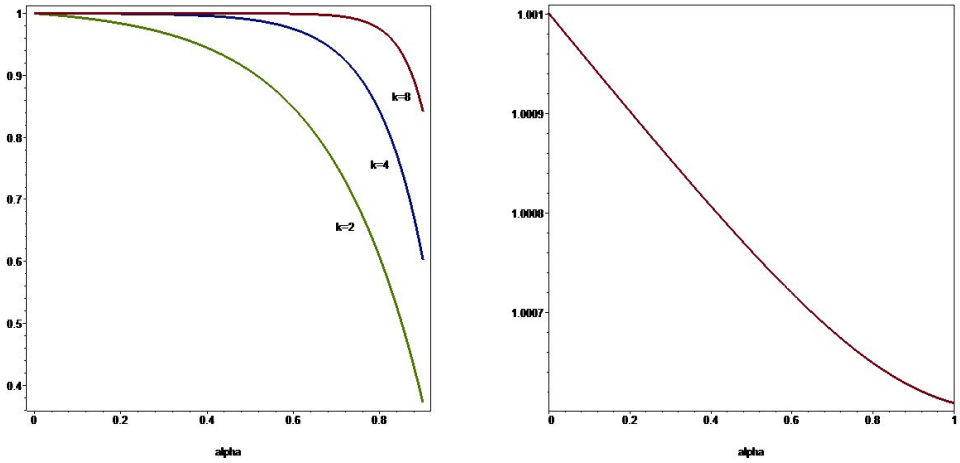

Figure 3.3: *Left: ratios* $\Phi_{\alpha,\varepsilon}(\mu_0)/\Phi_\alpha(\mu_0)$ *for* $\varepsilon = 10^{-k}$, $k = 2, 4, 8$. *Right: quality of approximation* (3.8)*: values of* $\Phi_{\alpha,\varepsilon}(\mu_0)/\left(\Phi_\alpha(\mu_0)(1-e^{1-\alpha})\right)$ *computed for* $\varepsilon = 0.001$ *and* $\alpha \in [0, 1)$.

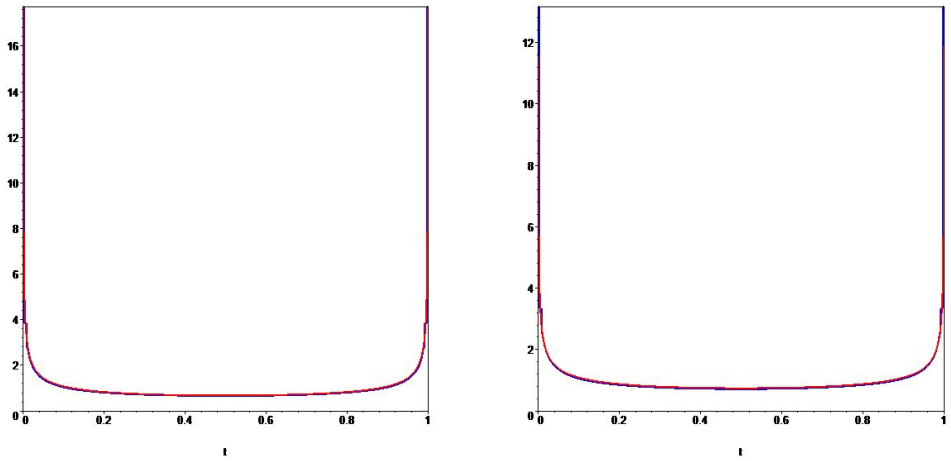

Figure 3.4: *Optimal densities* (3.2)*, red, and numerically computed densities* (2.13)*, blue, on the uniform grid* $x_k = k/N$ *(*$k = 0, 1, \ldots, N$*);* $N = 200$, $\varepsilon = 0.01$. *Left:* $\alpha = 0.1$, *efficiency* $\mathscr{E}_K(\mu_K^*)/\mathscr{E}_K(\widetilde{\zeta}_{\varepsilon,N}^*) \simeq 0.99939$. *Right:* $\alpha = 0.25$, *efficiency* $\mathscr{E}_K(\mu_K^*)/\mathscr{E}_K(\widetilde{\zeta}_{\varepsilon,N}^*) \simeq 0.99788$.

3.4. Other examples. A long list of Bernstein functions can be found in [10, Chapter 15]. In view of [4], among the first 50 Bernstein functions g_i from this list, the following cases the corresponding CM functions f_i (proportional to g_i') have infinite value at zero: 1, 7, 8, 9, 11, 12, 13, 16, 17, 19, 23, 25, 27, 31, 33, 34, 36, 38, 40, 41, 42, 43, 44, 45.

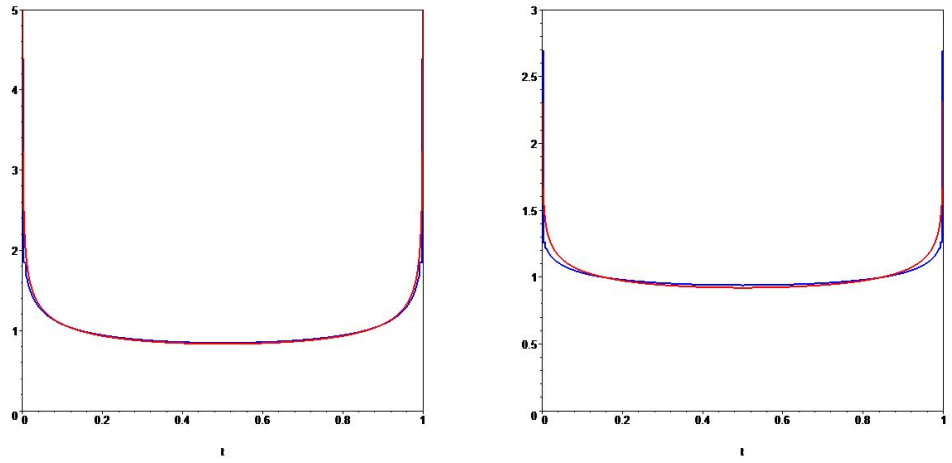

Figure 3.5: *Optimal densities* (3.2), *red, and numerically computed densities* (2.13), *blue, on the uniform grid* $x_k = k/N$ $(k = 0, 1, \ldots, N)$; $N = 200$, $\varepsilon = 0.001$. *Left:* $\alpha = 0.5$, *efficiency* $\mathscr{E}_K(\mu_K^*)/\mathscr{E}_K(\widetilde{\zeta}_{\varepsilon,N}^*) \simeq 0.99953$. *Right:* $\alpha = 0.75$, *efficiency* $\mathscr{E}_K(\mu_K^*)/\mathscr{E}_K(\widetilde{\zeta}_{\varepsilon,N}^*) \simeq 0.99869$.

Some of these functions are as follows ($t > 0$):

$$
\begin{aligned}
f_1(t) &= t^{-\alpha},\ 0 < \alpha < 1;\\
f_8(t) &= \frac{t^{-\alpha}}{(1+t)^{2-\alpha}},\ 0 < \alpha < 1;\\
f_{11}(t) &= \frac{\alpha t^{\alpha-1}(1-t^{\beta}) - \beta t^{\beta-1}(1-t^{\alpha})}{(1-t^{\alpha})^2},\ 0 < \alpha < \beta < 1\,;\\
f_{16}(t) &= \frac{\alpha_1 t^{-\alpha_1-1} + \ldots + \alpha_n t^{-\alpha_n-1}}{\left(t^{-\alpha_1} + \ldots + t^{-\alpha_n}\right)^2},\ 0 < \alpha_1, \ldots, \alpha_n \le 1;\\
f_{19}(t) &= \left(1 - (\lambda\sqrt{t} - 1)e^{-\lambda\sqrt{t}}\right)/\sqrt{t},\ \lambda > 0;\\
f_{23}(t) &= t\,(1 + 1/t)^{1+t} \log\,(1 + 1/t)\ .
\end{aligned}
$$

The families of functions $f_{i,\varepsilon}(t) = (g_i(t+\varepsilon) - g_i(t))/\varepsilon$ are constructed by (2.5). Below we give expressions for g_i. Note that there may be an extra multiplier and a different parametrization if these functions are compared against corresponding expressions in [10]. Since all functions $g_{i,\varepsilon}$ are normalized so that $g_i(0) = 0$ for all i, the values

$f_{i,\varepsilon}(0)$ are simply $f_{i,\varepsilon}(0) = g_i(\varepsilon)$; for small $\varepsilon > 0$ these values are large.

$$
\begin{aligned}
g_1(t) &= t^{1-\alpha}/(1-\alpha)\,,\ 0<\alpha<1\,;\\
g_8(t) &= \frac{t^{1-\alpha}}{(1-\alpha)(1+t)^{1-\alpha}}\,,\ 0<\alpha<1\,;\\
g_{11}(t) &= (t^\beta - t^\alpha)/(t^\alpha - 1),\ 0<\alpha<\beta<1\,;\\
g_{16}(t) &= 1/\left(t^{-\alpha_1}+\ldots+t^{-\alpha_n}\right),\ 0<\alpha_1,\ldots,\alpha_n\le 1;\\
g_{19}(t) &= 2\sqrt{t}\left(1+e^{-\lambda\sqrt{t}}\right),\ \lambda>0;\\
g_{23}(t) &= t\left(1+1/t\right)^{1+t}-1\,.
\end{aligned}
$$

Minimum-energy measures for the kernels constructed for all these (as well as many other) CM functions has been numerically constructed by the authors. Results (and figures of optimal densities) are quite similar to the ones provided above for the Riesz kernel.

Acknowledgements. The first author benefited from the support of the project INDEX (INcremental Design of EXperiments) ANR-18-CE91-0007 of the French National Research Agency (ANR). The second author thanks Dr Tomos Phillips for many relevant discussions.

REFERENCES

[1] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, Boston, 2004.
[2] N.S. Landkof. *Foundations of Modern Potential Theory*. Springer, Berlin, 1972.
[3] T.R.L. Phillips and K.M. Schmidt. On unbounded positive definite functions. *Mathematica Pannonica*, 26(2):33–51, 2018.
[4] T.R.L. Phillips, K.M. Schmidt, and A.A. Zhigljavsky. Extension of the Schoenberg theorem to integrally conditionally positive definite functions. *Journal of Mathematical Analysis and Applications*, 470(1):659–678, 2019.
[5] L. Pronzato, H.P. Wynn, and A.A. Zhigljavsky. Bregman divergences based on optimal design criteria and simplicial measures of dispersion. *Statistical Papers*, 60(2):195–214, 2019.
[6] L. Pronzato and A.A. Zhigljavsky. Bayesian quadrature and energy minimization for space-filling design. *arXiv preprint arXiv:1808.10722*, 2018.
[7] C. R. Rao. Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, 21(1):24–43, 1982.
[8] C.R. Rao and T.K. Nayak. Cross entropy, dissimilarity measures and characterizations of quadratic entropy. *IEEE Transactions on Information Theory*, 31(5):589–593, 1985.
[9] E.B. Saff and V. Totik. *Logarithmic Potentials with External Fields*. Springer, Berlin, 2013.
[10] R.L. Schilling, R. Song, and Z. Vondracek. *Bernstein Functions: Theory and Applications*. de Gruyter, Berlin/Boston, 2012.
[11] K. M. Schmidt and A. Zhigljavsky. A characterization of the arcsine distribution. *Statistics & Probability Letters*, 79(24):2451–2455, 2009.
[12] K. M. Schmidt and A. Zhigljavsky. An extremal property of the generalized arcsine distribution. *Metrika*, 76(3):347–355, 2013.
[13] I.J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536, 1938.
[14] Z. Szabó and B. Sriperumbudur. Characteristic and universal tensor product kernels. *Journal of Machine Learning Research*, 18:1–29, 2018.
[15] G. Székely and M.L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.
[16] H. Wendland. *Scattered Data Approximation*. Cambridge University Press, 2005.