QUADRATURE ERRORS, DISCREPANCIES, AND THEIR RELATIONS TO HALFTONING ON THE TORUS AND THE SPHERE[*]

MANUEL GRÄF[†], DANIEL POTTS[†], AND GABRIELE STEIDL[‡]

Abstract. This paper deals with continuous-domain quantization, which aims to create the illusion of a gray-value image by appropriately distributing black dots. For lack of notation, we refer to the process as halftoning, which is usually associated with the quantization on a discrete grid. Recently a framework for this task was proposed by minimizing an attraction-repulsion functional consisting of the difference of two continuous, convex functions. The first one of these functions describes attracting forces caused by the image gray values, the second one enforces repulsion between the dots. In this paper, we generalize this approach by considering quadrature error functionals on reproducing kernel Hilbert spaces (RKHSs) with respect to the quadrature nodes, where we ask for optimal distributions of these nodes. For special reproducing kernels these quadrature error functionals coincide with discrepancy functionals, which leads to a geometric interpretation. It turns out that the original attraction-repulsion functional appears for a special RKHS of functions on $\mathbb{R}^2$. Moreover, our more general framework enables us to consider optimal point distributions not only in $\mathbb{R}^2$ but also on the torus $\mathbb{T}^2$ and the sphere $\mathbb{S}^2$. For a large number of points the computation of such point distributions is a serious challenge and requires fast algorithms. To this end, we work in RKHSs of bandlimited functions on $\mathbb{T}^2$ and $\mathbb{S}^2$. Then the quadrature error functional can be rewritten as a least squares functional. We use a nonlinear conjugate gradient method to compute a minimizer of this functional and show that each iteration step can be computed in an efficient way by fast Fourier transforms at nonequispaced nodes on the torus and the sphere. Numerical examples demonstrate the good quantization results obtained by our method.

Key words. halftoning, dithering, stippling, Poisson disk distribution, point distributions, quadrature rules, discrepancies, optimization methods on Riemannian manifolds, CG method, nonequispaced fast Fourier transform, spherical Fourier transform

AMS subject classifications. 65T40, 65K10, 53B21, 49M15, 33C55

DOI. 10.1137/100814731

1. Introduction. Halftoning is a method for creating the illusion of a continuous tone image having only a small number of tones available. It is usually associated with the quantization on a discrete grid. In this paper, we use the name halftoning for continuous-domain quantization; cf. [48, 28]. We focus just on two tones, black and white, and ask for appropriate distributions of the black "dots." Applications of halftoning include printing and geometry processing [54] as well as sampling problems occurring in rendering [60], relighting [36], and artistic nonphotorealistic image visualization [4, 49]. Halftoning has been an active field of research for many years.

Dithering methods which place the black dots at image grid points include, e.g., ordered dithering [7, 47], error diffusion [22, 32, 52, 53], global or direct binary search [1, 6], and structure-aware halftoning [45]. Ostromoukhov [44] extended the error

diffusion idea by proposing a stencil design which takes the blue noise spectrum into account and depends on the gray values being processed. This approach was further improved by Chang, Alain, and Ostromoukhov [11] who proposed to use thresholds and error diffusion stencils depending on the local frequency, orientation, and contrast in the image. This is realized by using five lookup tables created manually in a calibration step. For a related, patented technique, see [31].

In this paper we consider algorithms which use the continuous domain as possible black dot positions instead of the image grid. More precisely, consider a gray-value image $u : \mathcal{G} \to [0,1]$ on a (squared) grid $\mathcal{G} := G \times G$, where $G := \{\frac{1}{2n} + \frac{i}{n} : i = 0, \ldots, n-1\}$. Since "black" is 0 and "white" 1, we will later use the corresponding weight distribution $w := 1 - u$. Now we intend to find the positions $p_i \in [0,1]^2$, $i = 1, \ldots, M$, of M black dots which create the illusion of the original gray-value image u; see Figure 1 for an illustration.

For this purpose stochastic point distributions were frequently used in computer graphics. Among them the *Poisson point distribution* stands out for its superior blue noise characteristics. The blue noise analysis of dithered patterns was first described by Ulichney [58, 59] and is widely accepted for evaluating dithering and screening methods. For a comparison of various methods for generating Poisson point distributions, we refer the reader to [39]. An artistic point distribution method on the continuous domain, called *stippling* was proposed by Secord [49]. Like some of the Poisson point distribution methods, it is based on weighted centroidal Voronoi tessellation and Lloyd's iterative algorithm [27, 41]. Recall that Lloyd's algorithm tries to find a useful minimizer of the error functional

$$L((p_k, V_k)_{k=1}^M) := \sum_{k=1}^M \int_{V_k} w(x)|x - p_k|^2 \, dx$$

over any set of points $\{p_k\}_{k=1}^M$ belonging to $\bar{\Omega} \subset \mathbb{R}^2$ and any tessellation $\{V_k\}_{k=1}^M$ of $\bar{\Omega}$. A necessary condition for L to be minimized is that the V_k's are the Voronoi regions corresponding to the p_k's, and, simultaneously, the p_k are the centroids of the corresponding V_k's. For a comprehensive overview and references up to 1999, see [17]. Recently, efficient centroidal Voronoi tessellation were also developed for *mesh surfaces* by Liu et al. [40]. A capacity-constrained variant of Lloyd's algorithm was proposed by Balzer, Schlömer, and Deussen [4]. Note that the method in this paper can also be considered as capacity-constrained, since it minimizes a certain discrepancy. Balzer's approach aims to overcome the drawback of Lloyd's algorithm which introduces regularity artifacts if not stopped at a suitable iteration step. Very recently an interesting approach for generating stochastic blue-noise distributions that formulates the problem as sampling a statistical mechanics interacting particle model was developed by Fattal [21]. In particular, his strategy generates a high-quality stochastic point distribution in a time that is linear in the number of points.

In [55], a novel halftoning framework was proposed, where the vector $\boldsymbol{p} := (p_i)_{i=1}^M \in \mathbb{R}^{2M}$ of the black dot positions was determined to be a minimizer of the functional

$$(1.1) \qquad E(\boldsymbol{p}) := \sum_{i=1}^M \sum_{x \in \mathcal{G}} w(x)\|p_i - x\|_2 - \frac{\lambda}{2} \sum_{i,j=1}^M \|p_i - p_j\|_2.$$

Here $\lambda := \frac{1}{M} \sum_{x \in \mathcal{G}} w(x)$ is an equilibration parameter between the "opposite" functionals. The intention for considering minimizers of this functional as "good" black

dot positions comes originally from electrostatic principles used in [48] for halftoning. The black points are considered as small particles of equal size moving in an environment, e.g., under a glass pane above the image w. The particles are *attracted* by the image forces $w(x)$ at the points $x \in \mathcal{G}$. On the other hand, there is a force of *repulsion* between the particles modeled by the negative sign of the second sum which becomes minimal if the sum of distances between the particles are maximized. In [55] a circumstantial numerical comparison to other methods was performed which showed that the method achieves unsurpassed quality, has a blue noise spectrum which can keep up with state-of-the-art techniques, and performs superior with respect to Gaussian scale space properties. Furthermore, the use of the nonequispaced fast Fourier transform lowers the complexity of $\mathcal{O}(M^2)$ per iteration achieved in [48] to $\mathcal{O}(M \log M)$. Very recently, a new parallel implementation for GPUs was proposed in [28] which reduces the runtime again substantially.

FIG. 1. *Left: Original* 256×256 *image. Right: Stippling result by minimizing* (1.1) *with* $m = 30150$ *points using the technique from* [55].

In this paper, we deal with the continuous version of the above attraction-repulsion functional

$$(1.2) \qquad E(\boldsymbol{p}) := \sum_{i=1}^{M} \int_{[0,1]^2} w(x)\|p_i - x\|_2 \, \mathrm{d}x - \frac{\lambda}{2} \sum_{i,j=1}^{M} \|p_i - p_j\|_2,$$

where $w : [0,1]^2 \to [0,1]$ is defined on the whole square $[0,1]^2$ and $\lambda := \frac{1}{M} \int_{[0,1]^2} w(x) \, \mathrm{d}x$. We can also consider (1.2) with more general functions $\varphi : [0,\infty) \to \mathbb{R}$ and obtain

$$(1.3) \qquad E_\varphi(\boldsymbol{p}) := \frac{\lambda}{2} \sum_{i,j=1}^{M} \varphi(\|p_i - p_j\|_2) - \sum_{i=1}^{M} \int_{[0,1]^2} w(x)\varphi(\|p_i - x\|_2) \, \mathrm{d}x.$$

In (1.2) the function $\varphi(r) = -r$ was used. In [48] the function $\varphi(r) = -\log(r)$ with a modification near zero was applied. Further, the authors in [55] also mentioned $\varphi(r) = -r^\tau$, $0 < \tau < 2$, and $\varphi(r) = r^{-\tau}$, $\tau > 0$, with a modification near zero as possible choices. For monotone decreasingly, convex φ with $\varphi(0) < \infty$ it is easily shown that the minimizers of (1.3) exist; cf. [55].

In this paper we look at the halftoning problem from different points of view. The framework arises primarily from approximation theory but touches many different areas in mathematics as well. The proposed setting is quite general and enables us to consider in some sense optimal point distributions not only in $\mathbb{R}^2$ but also on the torus $\mathbb{T}^2$ and the sphere $\mathbb{S}^2$, or other compact manifolds in $\mathbb{R}^d$. Let us remark that even in the seemingly easiest case with $w \equiv 1$ the search for optimal point configurations is a very tough problem, at least in more than one dimension. For example on the sphere with the Coulomb potential $\varphi(r) = r^{-1}$ we are confronted with the Thomson problem [56], which asks for the ground states of a given number of electrons on the sphere. This famous problem originated many publications concerning the computation [63], asymptotic [37], and characteristic [10] of optimal distributions on the sphere, to name but a few. Another interesting application of our halftoning procedure on the sphere may be found in methods for solving partial differential equations arising in geoscience [23].

In the following, we consider worst case quadrature errors on reproducing kernel Hilbert spaces (RKHSs) in dependence on the quadrature nodes and ask for optimal node distributions. In the literature this was mainly done for constant weights $w \equiv 1$. A weighted setting appears in connection with the so-named "importance sampling"; see [43] and the references therein. It turns out that the attraction-repulsion functional (1.2) leads to the same optimal point distributions as the quadrature error functional for a certain RKHS of functions on $\mathbb{R}^2$ with the Euclidean distance kernel. For special reproducing kernels we show that the quadrature error functionals coincide with discrepancy functionals. This adds another interesting point of view which is closely related to the notation of "capacity constraints" in halftoning; see [3, 4]. As was already mentioned, the main challenge for computing optimal point distributions is the design of fast algorithms. Here, we present an algorithm which works on RKHSs of bandlimited functions on $\mathbb{T}^2$ and $\mathbb{S}^2$. We show that the quadrature error functional can be rewritten as a least squares functional. Then we propose a nonlinear conjugate gradient (CG) method for computing a minimizer. Indeed, on $\mathbb{S}^2$ we apply the CG method on manifolds; see [19, 51]. This method was also successfully used for the approximation of spherical designs in [26]. We show how each step within the CG method can be realized in an efficient way by fast Fourier transforms at nonequispaced nodes on the torus (NFFT) and the sphere (NFSFT), respectively. Finally, we provide proof-of-concept numerical examples based on the NFFT library [34].

Our paper is organized as follows. In section 2, we introduce the worst case quadrature errors on RKHSs in dependence on the quadrature nodes and show that the attraction-repulsion functional (1.2) appears as a special case. The relation to discrepancy functionals is proved in section 3. Furthermore, we provide certain discrepancy kernels on $\mathbb{S}^1$, $\mathbb{T}^2$, and $\mathbb{S}^2$ and compare them numerically with the corresponding restrictions of the negative Euclidean distance kernel $-\|x - y\|_2$. Section 4 deals with the efficient computation of optimal point distributions. In subsection 4.1, the functionals are considered on RKHSs of bandlimited functions on $\mathbb{S}^1$, $\mathbb{T}^2$, and $\mathbb{S}^2$, which in that case can be rewritten as least squares functionals. We show that the evaluation of these functionals as well as the computation of their gradients and vector multiplications with their Hessians are realized in a fast way by using NFFTs/NFSFTs. Subsection 4.2 provides the CG algorithms with respect to our setting. Finally, we present halftoning examples on $\mathbb{T}^2$ and $\mathbb{S}^2$ in section 5.

2. Quadrature errors in RKHSs and halftoning. In this section, we consider worst case quadrature errors in RKHSs. We show that for special RKHSs the

minimizers of the corresponding error functional coincide with those of the halftoning functional (1.3). Of course our more general setting can be used as a starting point for various applications. In this paper, we will use it to design halftoning procedures on the torus $\mathbb{T}^2$ and the sphere $\mathbb{S}^2$, which can be easily generalized to halftoning on arbitrary compact manifolds $\mathcal{X} \subset \mathbb{R}^d$.

2.1. Quadrature error in RKHSs. In what follows, we restrict our attention to $\mathcal{X} \in \{\mathbb{R}^d, \mathbb{S}^1, \mathbb{T}^2, \mathbb{S}^2\}$ and remark that arbitrary measurable sets are possible. Recall, that the d-dimensional sphere and torus is defined by

$$\mathbb{S}^d := \{\boldsymbol{x} \in \mathbb{R}^{d+1} \ : \ \|\boldsymbol{x}\|_2 = 1\} \qquad \text{and} \qquad \mathbb{T}^d := \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 \subset \mathbb{R}^{2d},$$

respectively. A symmetric function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be *positive semidefinite* if for any $M \in \mathbb{N}$ points $x_1, \dots, x_M \in \mathcal{X}$ and any $a := (a_1, \dots, a_M)^{\mathrm{T}} \neq 0$ the relation

$$(2.1) \qquad\qquad a^{\mathrm{T}} \left(K(x_i, x_j) \right)_{i,j=1}^M a \geq 0$$

holds true and *positive definite* if we have strict inequality in (2.1). A (real) *reproducing kernel Hilbert space* (RKHS) is a Hilbert space H_K with inner product $\langle f, g \rangle_{H_K}$, $f, g \in H_K$, having a *reproducing kernel*, i.e., a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ which fulfills

$$K_x := K(\cdot, x) \in H_K \quad \forall\, x \in \mathcal{X},$$
$$(2.2) \qquad f(x) = \langle f, K(\cdot, x) \rangle_{H_K} \quad \forall x \in \mathcal{X} \quad \text{and} \quad \forall f \in H_K.$$

An equivalent definition of a RKHS says that it is a Hilbert space on which the point evaluation functionals are continuous. To every RKHS there corresponds a unique positive semidefinite kernel and conversely given a positive semidefinite function K there exists a unique RKHS of real-valued function having K as its reproducing kernel; see [62, Theorem 1.1.1].

If, additionally, $\mathcal{X}$ is compact and equipped with a finite measure $\mu_{\mathcal{X}}$ we can expand by Mercer's theorem any continuous kernels K into an μ_X-almost everywhere absolutely and uniformly convergent series

$$K(x, y) = \sum_{l=1}^{\infty} \lambda_l \psi_l(x) \overline{\psi_l(y)} = \sum_{l=1}^{\infty} \lambda_l \overline{\psi_l(x)} \psi_l(y)$$

of orthonormal eigenfunctions $\psi_l \in L_2(\mathcal{X})$ with associated eigenvalues $\lambda_l > 0$ of the integral operator T_K given by

$$T_K f(x) := \int_{\mathcal{X}} K(x, y) f(y) \, \mathrm{d}\mu_{\mathcal{X}}(y).$$

For more information on RKHSs, we refer the reader to [2].

In the following, we let $\mathcal{X}$ be equipped with the canonical Lebesgue surface measure $\mu_{\mathcal{X}}$ and use the abbreviation $\mathrm{d}y = \mathrm{d}\mu_{\mathcal{X}}(y)$. Furthermore, we consider a nontrivial, measurable function $w : \mathcal{X} \to [0, \infty)$ with the property

$$(2.3) \qquad\qquad h_w(x) := \int_{\mathcal{X}} w(y) K(x, y) \, \mathrm{d}y \ \in H_K,$$

i.e.,

$$\|h_w\|_{H_K}^2 = \left\langle \int_{\mathcal{X}} w(y) K(\cdot, y) \, \mathrm{d}y, h_w \right\rangle_{H_K} = \int_{\mathcal{X}} w(y) \langle K(\cdot, y), h_w \rangle_{H_K} \, \mathrm{d}y$$

$$(2.4) \qquad = \int_{\mathcal{X}} w(y) h_w(y) \, \mathrm{d}y = \int_{\mathcal{X}} \int_{\mathcal{X}} w(x) w(y) K(x, y) \, \mathrm{d}x \mathrm{d}y < \infty.$$

We remark that for compact sets $\mathcal{X}$ and continuous kernels K the condition (2.4) is satisfied. In the case $\mathcal{X} = \mathbb{R}^d$, which is only of interest in subsection 2.2, we suppose that w has compact support. For appropriately chosen points $p_j \in \mathcal{X}$, $j = 1, \ldots, M$, we are interested in approximating the integrals

$$I_w(f) := \int_{\mathcal{X}} f(x)w(x)\, \mathrm{d}x \quad \text{for } f \in H_K$$

by a quadrature rule

$$(2.5) \qquad Q(f, \boldsymbol{p}) := \lambda \sum_{i=1}^{M} f(p_i), \qquad \lambda := \frac{1}{M} \int_{\mathcal{X}} w(x)\, \mathrm{d}x,$$

where $\boldsymbol{p} := (p_1, \ldots, p_M) \in \mathcal{X}^M$. This quadrature rule appears to play a key role in our paper. In the literature mainly the case $w \equiv 1$ was considered; see [43] and the references therein. The *worst case quadrature error* is given by

$$(2.6) \qquad \mathrm{err}_K(\boldsymbol{p}) := \sup_{\substack{f \in H_K \\ \|f\|_{H_K} \leq 1}} |I_w(f) - Q(f, \boldsymbol{p})| = \|I_w - Q(\cdot, \boldsymbol{p})\|,$$

where the latter norm is the operator norm of the linear functionals on H_K. In particular, we see that $I_w(f) = Q(f, \boldsymbol{p})$, for some $\boldsymbol{p} \in \mathcal{X}^M$ and all $f \in H_K$, if and only if $\mathrm{err}_K(\boldsymbol{p}) = 0$.

The following theorem shows a relation between this error functional and the halftoning functional (1.3).

THEOREM 2.1. *Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive semidefinite function and H_K the associated RKHS. Then the relation*

$$\mathrm{err}_K(\boldsymbol{p})^2 = 2\lambda E_K(\boldsymbol{p}) + \|h_w\|^2_{H_K}$$

holds true, where

$$(2.7) \qquad E_K(\boldsymbol{p}) := \frac{\lambda}{2} \sum_{i,j=1}^{M} K(p_i, p_j) - \sum_{i=1}^{M} \int_{\mathcal{X}} w(x)K(p_i, x)\, \mathrm{d}x.$$

In particular, for compact sets $\mathcal{X}$ the minimizers of err_K and E_K exist and coincide.

Proof. We rewrite I_w as

$$I_w(f) = \int_{\mathcal{X}} \langle f, K(\cdot, x) \rangle_{H_K} w(x)\, \mathrm{d}x$$
$$= \left\langle f, \int_{\mathcal{X}} w(x)K(\cdot, x)\, \mathrm{d}x \right\rangle_{H_K} = \langle f, h_w \rangle_{H_K},$$

so that by (2.2) and (2.5)

$$I_w(f) - Q(f, \boldsymbol{p}) = \left\langle f, h_w - \lambda \sum_{i=1}^{M} K(\cdot, p_i) \right\rangle_{H_K}$$

and consequently

$$\mathrm{err}_K(\boldsymbol{p}) = \left\| h_w - \lambda \sum_{i=1}^{M} K(\cdot, p_i) \right\|_{H_K}.$$

Now the squared worst case error reads

$$\operatorname{err}_K(\boldsymbol{p})^2 = \|h_w\|_{H_K}^2 - 2\lambda \left\langle h_w, \sum_{i=1}^M K(\cdot, p_i) \right\rangle_{H_K} + \lambda^2 \sum_{i,j=1}^M K(p_i, p_j)$$

$$= \|h_w\|_{H_K}^2 - 2\lambda \sum_{i=1}^M h_w(p_i) + \lambda^2 \sum_{i,j=1}^M K(p_i, p_j)$$

$$(2.8) \qquad = \|h_w\|_{H_K}^2 - 2\lambda \sum_{i=1}^M \int_{\mathcal{X}} w(x) K(p_i, x)\, \mathrm{d}x + \lambda^2 \sum_{i,j=1}^M K(p_i, p_j)$$

and the minimizers of this functional coincide with those of E_K.   □

By the following proposition, slight modifications of the kernel do not change the minimizers of the functional.

PROPOSITION 2.2. *Let* $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *be a symmetric function and let* $\tilde{K}(x, y) :=$ $aK(x, y) + b(K(x, 0) + K(0, y)) + c$ *with* $a > 0$ *and* $b, c \in \mathbb{R}$. *Then the functionals* $E_{\tilde{K}}$ *and* E_K *differ only by the factor* a *and an additive constant* $C \in \mathbb{R}$, *i.e.,* $E_{\tilde{K}} = aE_K + C$.

Proof. By (2.7) and the definition of λ, we obtain with certain constants $\tilde{C}, C \in \mathbb{R}$ independent of $\boldsymbol{p}$ that

$$E_{\tilde{K}}(\boldsymbol{p}) = \frac{\lambda}{2} \sum_{i,j=1}^M \left(aK(p_i, p_j) + b\left(K(p_i, 0) + K(0, p_j) \right) \right)$$

$$- \sum_{i=1}^M \int_{\mathcal{X}} w(x) \left(aK(p_i, x) + b\left(K(p_i, 0) + K(0, x) \right) \right) \mathrm{d}x + \tilde{C}$$

$$= \frac{a\lambda}{2} \sum_{i,j=1}^M K(p_i, p_j) - a \sum_{i=1}^M \int_{\mathcal{X}} w(x) K(p_i, x)\, \mathrm{d}x$$

$$+ \frac{b\lambda}{2} 2M \sum_{i=1}^M K(p_i, 0) - b \sum_{i=1}^M K(p_i, 0) \int_{\mathcal{X}} w(x)\, \mathrm{d}x + C$$

$$= \frac{a\lambda}{2} \sum_{i,j=1}^M K(p_i, p_j) - a \sum_{i=1}^M \int_{\mathcal{X}} w(x) K(p_i, x)\, \mathrm{d}x + C = aE_K(\boldsymbol{p}) + C.$$ □

The functional E_K looks as that in (1.3) if K is a radial kernel $K(x, y) = \varphi(\|x - y\|_2)$. Therefore, we ask for radial kernels which are positive semidefinite in the next subsection.

2.2. Relation to halftoning functionals on $\mathbb{R}^d$. In this subsection, we consider the relation between E_K and (1.3) in detail. To this end, in this subsection we let $\mathcal{X} := \mathbb{R}^d$ and let $w : \mathbb{R}^d \to [0, \infty)$ be a nontrivial, continuous function with compact support in $[0, 1]^d$.

A kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called a *radial kernel* if $K(x, y) = \varphi(\|x - y\|_2)$ for some function $\varphi : [0, \infty) \to \mathbb{R}$. We are looking for positive semidefinite, radial kernels K. Note that since $\mathbb{R}^l$ is a subspace of $\mathbb{R}^d$ for $l \leq d$ positive semidefiniteness of a kernel on $\mathbb{R}^d$ implies its positive semidefiniteness on $\mathbb{R}^l$. Positive definite, radial kernels on $\mathbb{R}^d$ are given for instance by the inverse multiquadrics $K(x, y) := (\varepsilon^2 + \|x - y\|_2^2)^{-\tau}$,

$\varepsilon > 0, \tau > d/2$, which are related to $\varphi(r) = r^{-\tau}$ in (1.3). For other examples, see [64, 20].

For the kernel in (1.2) we have to consider *conditionally positive definite, radial functions* $\Phi(x) := \varphi(\|x\|_2)$ *of order* 1. These functions are defined to be continuous with the property that for any $M \in \mathbb{N}$ points $x_1, \ldots, x_M \in \mathbb{R}^d$ the relation

$$a^{\mathrm{T}} \left(\Phi(x_i - x_j) \right)_{i,j=1}^M a > 0 \quad \forall a = (a_1, \ldots, a_M)^{\mathrm{T}} \neq 0 \quad \text{with} \quad \sum_{i=1}^M a_i = 0$$

holds true. For conditionally positive definite, radial functions of higher order which are not relevant in this paper, we refer to [64]. Examples of conditionally positive definite, radial functions of order 1 in $\mathbb{R}^d$ are

$$\Phi(x) = -\|x\|_2^\tau, \quad 0 < \tau < 2,$$
$$\Phi(x) = -(\varepsilon^2 + \|x\|_2^2)^\tau, \quad 0 < \tau < 1 \quad \text{(multiquadrics)}.$$

Of course our dithering functional (1.2) is exactly $E_{\Phi(x-y)}$ for the first function with $\tau = 1$ in $\mathbb{R}^2$, while the multiquadrics are related to (1.3) with $\varphi(r) = -r^\tau$. Unfortunately the above kernels $\Phi(x - y)$ are not positive semidefinite. However, given a conditionally positive definite radial kernel Φ of order 1, then the kernel

$$K_\Phi(x, y) := \Phi(x - y) - \Phi(y) - \Phi(x) + \Phi(0) + 1$$

is again a positive semidefinite kernel, which gives rise to a RKHS H_{K_Φ}. These spaces can be characterized as in [64, Theorem 10.18–10.21] and by Proposition 2.2 we see that $E_{\Phi(x-y)}$ and E_{K_Φ} have the same minimizers. Hence, the minimizers of the original attraction-repulsion functional (1.2) can be considered as optimal quadrature points in the Hilbert space H_{K_Φ}.

Remark 2.3 (halftoning on $\mathbb{R}^1$). For $\mathcal{X} = \mathbb{R}$, the minimizers of (1.3) with $\varphi(r) = -r$ can be described analytically. In one dimension we can suppose that the point positions are ordered by $p_1 \leq \cdots \leq p_M$ such that our functional simplifies to the convex functional

$$E_\varphi(\boldsymbol{p}) = \sum_{i=1}^M \int_0^1 w(x)|p_i - x| \, \mathrm{d}x - \frac{\lambda}{2} \sum_{i,j=1}^M |p_i - p_j|$$
$$= \sum_{i=1}^M \int_0^1 w(x)|p_i - x| \, \mathrm{d}x + \lambda \sum_{i=1}^M (M - (2i - 1))p_i.$$

We observe that the right-hand side is a continuously differentiable function on $\mathbb{R}^M$, so that a sufficient condition for an ordered minimizer $\hat{\boldsymbol{p}}$ of the functional E_φ is given by

$$0 = \frac{\partial}{\partial p_i} E_\varphi(\hat{\boldsymbol{p}}) = \lambda(M - (2i - 1)) + \int_0^{\hat{p}_i} w(x) \, \mathrm{d}x - \int_{\hat{p}_i}^1 w(x) \, \mathrm{d}x$$
$$= \lambda(M - (2i - 1)) + \int_0^{\hat{p}_i} w(x) \, \mathrm{d}x - \left(\lambda M - \int_0^{\hat{p}_i} w(x) \, \mathrm{d}x \right),$$

which leads to the relations

$$\int_0^{\hat{p}_i} w(x) \, \mathrm{d}x = \lambda\left(i - \frac{1}{2} \right), \quad i = 1, \ldots, M.$$

In other words, the first point $\hat{p}_1$ is determined by $\int_0^{\hat{p}_1} w(x)\,\mathrm{d}x = \lambda/2$ and the other points by $\int_{\hat{p}_i}^{\hat{p}_{i+1}} w(x)\,\mathrm{d}x = \lambda$, $i = 1, \ldots, M-1$. In particular, for continuous functions w with $w(x) > 0$, $x \in [0,1]$, the minimizer $\hat{p}$ is uniquely determined with the points $\hat{p}_i$, $i = 1, \ldots, M$, lying in $[0,1]$.

We note further that there is an interesting connection to the Sobolev spaces

$$H_{K^\beta} := \{ f : [0,1] \to \mathbb{R} : f(\beta) = 0, f \text{ abs. continuous}, f' \in L_2([0,1]) \}$$

anchored at $\beta \in [0,1]$ which were considered in [43]. These RKHSs have the reproducing kernels

$$K^\beta(x,y) := \frac{1}{2}\left(|x - \beta| + |y - \beta| - |x - y| \right).$$

Using similar arguments as in the proof of Proposition 2.2 one can check that our functional E_φ and the functionals E_{K^β}, $\beta \in [0,1]$, have the same minimizers.

3. Discrepancies. The quadrature errors considered in the previous section are closely related to discrepancies which add a geometric point of view. We consider in the following $\mathcal{X} \in \{\mathbb{S}^1, \mathbb{T}^2, \mathbb{S}^2\}$ with the canonical Lebesgue surface measure $\mu_{\mathcal{X}}$ and geodesic metric $d_{\mathcal{X}}$. Let $D := \mathcal{X} \times [0, R]$ and let $\mathcal{B}(c, r) := \{x \in \mathcal{X} : d_{\mathcal{X}}(c, x) \leq r\}$ be the ball centered at $c \in \mathcal{X}$ with radius $r \leq R$. By $\mathbf{1}_{\mathcal{B}(c,r)}$ we denote the characteristic function of $\mathcal{B}(c, r)$. Then we define the L_2-*discrepancy* as

$$(3.1) \qquad \mathrm{disc}_2^{\mathcal{B}}(\boldsymbol{p}) := \left(\int_D \left(\int_{\mathcal{X}} w(x) \mathbf{1}_{\mathcal{B}(t)}(x)\,dx - \lambda \sum_{i=1}^M \mathbf{1}_{\mathcal{B}(t)}(p_i) \right)^2 \mathrm{d}t \right)^{\frac{1}{2}},$$

where we have set $t := (c, r) \in D$ and $\mathrm{d}t := \mathrm{d}c\,\mathrm{d}r$. For an interpretation of the L_2-discrepancy (3.1) we note that the expression in the inner brackets relates the integral of w on $\mathcal{B}(c, r)$ with the number of points contained in $\mathcal{B}(c, r)$ for fixed $(c, r) \in D$ and that the discrepancy is then the squared error of their differences taken over all $t \in D$. This point of view is closely related to *capacity-constrained methods* used in [3, 4].

In what follows, we will see that the so-defined L_2-discrepancy $\mathrm{disc}_2^{\mathcal{B}}$ can be considered as a quadrature error in a particular RKHS; cf. [43] and the references therein. Therefore, we define the *discrepancy kernel*

$$(3.2) \quad K_{\mathcal{B}}(x,y) := \int_0^R \int_{\mathcal{X}} \mathbf{1}_{\mathcal{B}(c,r)}(x) \mathbf{1}_{\mathcal{B}(c,r)}(y)\,\mathrm{d}c\,\mathrm{d}r = \int_0^R \mu_{\mathcal{X}}(\mathcal{B}(x,r) \cap \mathcal{B}(y,r))\,\mathrm{d}r$$

and conclude by the relation

$$a^{\mathrm{T}} \left(K_{\mathcal{B}}(x_i, x_j) \right)_{i,j=1}^M a = \int_0^R \int_{\mathcal{X}} \left(\sum_{j=1}^M a_j \mathbf{1}_{\mathcal{B}(c,r)}(x_j) \right)^2 \mathrm{d}c\,\mathrm{d}r \geq 0$$

that $K_{\mathcal{B}}$ is a positive semidefinite kernel. Hence, the kernel $K_{\mathcal{B}}$ gives rise to a RKHS $H_{K_{\mathcal{B}}}$ and the relation between the discrepancy $\mathrm{disc}_2^{\mathcal{B}}$ and the quadrature error $\mathrm{err}_{K_{\mathcal{B}}}$ is given by the following theorem.

THEOREM 3.1. *Let $K_{\mathcal{B}}$ be defined by (3.2) and let $H_{K_{\mathcal{B}}}$ be the associated RKHS of functions on $\mathcal{X}$. Then $\mathrm{err}_{K_{\mathcal{B}}}$ given by (2.6) and $\mathrm{disc}_2^{\mathcal{B}}$ determined by (3.1) coincide:*

$$\mathrm{err}_{K_{\mathcal{B}}}(\boldsymbol{p}) = \mathrm{disc}_2^{\mathcal{B}}(\boldsymbol{p}).$$

Proof. Using the definition (3.2) of $K_{\mathcal{B}}$ and (3.1) we obtain

$$
\begin{aligned}
(\mathrm{disc}_2^{\mathcal{B}}(\boldsymbol{p}))^2 &= \int_D \left(\int_{\mathcal{X}} w(x) 1_{\mathcal{B}(t)}(x) \mathrm{d}x \right)^2 \mathrm{d}t - 2\lambda \int_D \int_{\mathcal{X}} w(x) 1_{\mathcal{B}(t)}(x) \mathrm{d}x \left(\sum_{i=1}^{M} 1_{\mathcal{B}(t)}(p_i) \right) \mathrm{d}t \\
&\quad + \lambda^2 \sum_{i,j=1}^{M} \int_D 1_{\mathcal{B}(t)}(p_i) 1_{\mathcal{B}(t)}(p_j) \mathrm{d}t \\
&= \int_D \left(\int_{\mathcal{X}} w(x) 1_{\mathcal{B}(t)}(x) \, dx \right)^2 \mathrm{d}t - 2\lambda \sum_{i=1}^{M} \int_{\mathcal{X}} w(x) \int_D 1_{\mathcal{B}(t)}(x) 1_{\mathcal{B}(t)}(p_i) \mathrm{d}t \, \mathrm{d}x \\
&\quad + \lambda^2 \sum_{i,j=1}^{M} K_{\mathcal{B}}(p_i, p_j) \\
&= \int_D \left(\int_{\mathcal{X}} w(x) 1_{\mathcal{B}(t)}(x) \mathrm{d}x \right)^2 \mathrm{d}t \\
&\quad - 2\lambda \sum_{i=1}^{M} \int_{\mathcal{X}} w(x) K_{\mathcal{B}}(p_i, x) \mathrm{d}x + \lambda^2 \sum_{i,j=1}^{M} K_{\mathcal{B}}(p_i, p_j).
\end{aligned}
$$

Finally, we see by (2.4) that

$$
\begin{aligned}
\int_D \left(\int_{\mathcal{X}} w(x) 1_{\mathcal{B}(t)}(x) \, dx \right)^2 \mathrm{d}t &= \int_{\mathcal{X}} \int_{\mathcal{X}} w(x) w(y) \int_D 1_{\mathcal{B}(t)}(x) 1_{\mathcal{B}(t)}(y) \mathrm{d}t \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{\mathcal{X}} \int_{\mathcal{X}} w(x) w(y) K_{\mathcal{B}}(x,y) \mathrm{d}t \, \mathrm{d}x \, \mathrm{d}y \; = \; \|h_w\|^2_{H_{K_{\mathcal{B}}}},
\end{aligned}
$$

and we are done by (2.8). □

Next we want to examine the relation between the negative Euclidean distance kernel $K(x,y) = \Phi(x-y) = -\|x-y\|_2$ considered in subsection 2.2 and the discrepancy kernel $K_{\mathcal{B}}$ defined in (3.2) for $\mathcal{X} \in \{\mathbb{S}^1, \mathbb{T}^2, \mathbb{S}^2\}$.

Kernels on $\mathbb{S}^1$. The circle $\mathbb{S}^1$ is naturally embedded in $\mathbb{R}^2$ by

$$
\mathbb{S}^1 = \{x := (2\pi)^{-1}\big(\cos(2\pi\alpha), \sin(2\pi\alpha)\big)^{\mathsf{T}} \in \mathbb{R}^2 : \alpha \in [0,1)\},
$$

where the correspondence between $x \in \mathbb{S}^1$ and $\alpha \in [0,1)$ is one-to-one. The geodesic distance is given for $x := \big(\cos(2\pi\alpha), \sin(2\pi\alpha)\big)^{\mathsf{T}}$ and $y := \big(\cos(2\pi\beta), \sin(2\pi\beta)\big)^{\mathsf{T}}$ by

$$
\mathrm{d}_{\mathbb{S}^1}(x,y) = (2\pi)^{-1} \arccos\big(\cos(2\pi(\alpha-\beta))\big) = \min\{|\alpha-\beta|, 1-|\alpha-\beta|\} \le \frac{1}{2}.
$$

The restriction of the negative Euclidean distance kernel on $\mathbb{S}^1$ is

$$
\begin{aligned}
\Phi(x-y) = -\|x-y\|_2 &= -(2\pi)^{-1}\sqrt{(\cos(2\pi\alpha) - \cos(2\pi\beta))^2 + (\sin(2\pi\alpha) - \sin(2\pi\beta))^2} \\
&= -\frac{1}{\pi}|\sin(\pi(\alpha-\beta))| = -\frac{1}{\pi}|\sin(\pi \mathrm{d}_{\mathbb{S}^1}(x,y))|.
\end{aligned}
$$

For the discrepancy kernel we use the balls $\mathcal{B}(c,r)$, $c \in \mathbb{S}^1$, $0 \le r \le 1/2$, and obtain

with $d := \mathrm{d}_{\mathbb{S}^1}(x, y)$ that

$$
\begin{aligned}
K_{\mathcal{B}}(x, y) &= \int_0^{\frac{1}{2}} \mu_{\mathbb{S}^1}\big(\mathcal{B}(x, r) \cap \mathcal{B}(y, r)\big) \, \mathrm{d}r \\
&= \int_{\frac{d}{2}}^{\frac{1}{2}} 2\left(r - \frac{d}{2}\right) \mathrm{d}r + \int_{\frac{1}{2} - \frac{d}{2}}^{\frac{1}{2}} 2\left(r - \frac{1}{2} + \frac{d}{2}\right) \mathrm{d}r \\
&= \frac{1}{4} + \frac{1}{2} d(d - 1).
\end{aligned}
$$

We note that Wahba's smoothing spline kernel R of order 1 on $\mathbb{S}^1$ (cf., [62, p. 21]) is given by

$$
\begin{aligned}
R(x, y) &= 1 + 2 \sum_{k=1}^{\infty} \frac{1}{(2\pi k)^2} \cos(2\pi k d) \\
&= \frac{13}{12} + \frac{1}{2} d(d - 1) \\
&= \frac{5}{6} + K_{\mathcal{B}}(x, y)
\end{aligned}
$$

so that $E_{K_{\mathcal{B}}}$ and E_R have the same minimizers. The kernel R is the reproducing kernel of the Hilbert space consisting of the functions

$$
f(x) = \tilde{f}(\alpha) := f_0 + 2 \sum_{k=1}^{\infty} f_k \cos(2\pi k \alpha) \quad \text{with} \quad \sum_{k=1}^{\infty} k^2 f_k^2 < \infty,
$$

$x := \big(\cos(2\pi\alpha), \sin(2\pi\alpha)\big)^{\mathrm{T}}$, with inner product

$$
\langle f, g \rangle_{H_R} = \int_0^1 \tilde{f}(\alpha) \, \mathrm{d}\alpha \int_0^1 \tilde{g}(\alpha) \, \mathrm{d}\alpha + \int_0^1 \tilde{f}'(\alpha) \tilde{g}'(\alpha) \, \mathrm{d}\alpha.
$$

The kernels $\Phi(x - (-(2\pi)^{-1}, 0)^{\mathrm{T}})$ and $K_{\mathcal{B}}(x, (-(2\pi)^{-1}, 0)^{\mathrm{T}})$ as functions of α are plotted in Figure 2. Since adding a constant and multiplying the kernel by a positive constant does not change the local minimizers of our functional (2.7), we compare the different kernels after an appropriate affine scaling. That is the maximum and the minimum of the scaled kernels are set without loss of generality to 1 and 0, respectively. Figure 2 shows that both kernels are quite similar.

By the following remark, the minimizers of E_K for the discrepancy kernel K can be characterized analytically. The arguments are similar as those in Remark 2.3.

Remark 3.2 (halftoning on $\mathbb{S}^1$). The discrepancy kernel reads up to a constant as $K(x, y) = \frac{1}{2}([x - y]_1^2 - [x - y]_1)$, where

$$
[x - y]_1 := \begin{cases} |x - y| & \text{if } |x - y| \leq 1/2, \\ 1 - |x - y| & \text{otherwise,} \end{cases} \qquad x, y \in [0, 1].
$$

We are looking for minimizers of

$$
\frac{\lambda}{2} \sum_{i,j=1}^{M} \big([p_i - p_j]_1^2 - [p_i - p_j]_1\big) - \sum_{i=1}^{M} \int_0^1 w(x)\big([p_i - x]_1^2 - [p_i - x]_1\big) \, \mathrm{d}x.
$$

If $|x - y| \geq 1/2$, we obtain that

$$
[x - y]_1^2 - [x - y]_1 = 1 - 2|x - y| + |x - y|^2 - 1 + |x - y| = |x - y|^2 - |x - y|.
$$

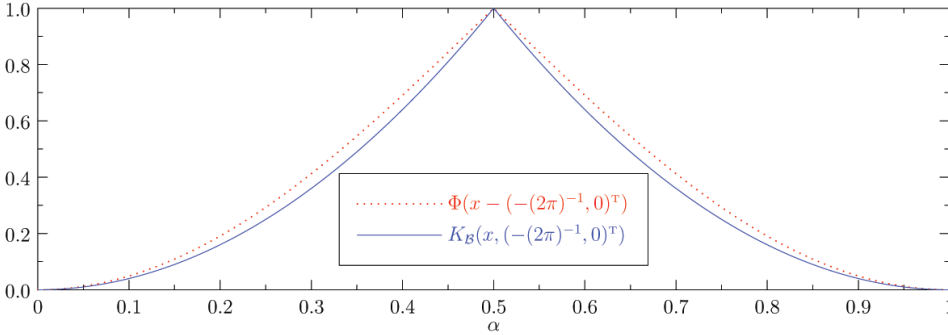FIG. 2. *Scaled kernels* $\Phi(x - (-(2\pi)^{-1}, 0)^T)$ *and* $K_{\mathcal{B}}(x, (-(2\pi)^{-1}, 0)^T)$ *on* $\mathbb{S}^1$ *as functions of* α.

Thus, we can replace $[\cdot]_1$ by $|\cdot|$ in the functional. Now an ordering of the point positions $0 \le p_1 \le \cdots \le p_M \le 1$ results as in Remark 2.3 in

$$\frac{\lambda}{2} \sum_{i,j=1}^{M} (p_i - p_j)^2 + \lambda \sum_{i=1}^{M} (M - (2i-1)) p_i - \sum_{i=1}^{M} \int_0^1 w(x)(p_i - x)^2 \, \mathrm{d}x + \sum_{i=1}^{M} \int_0^1 w(x)|p_i - x| \, \mathrm{d}x.$$

Setting the gradient to zero we obtain

$$\int_0^{\hat{p}_i} w(x) \, \mathrm{d}x = \lambda\left(i - \frac{1}{2}\right) + \lambda \sum_{j=1}^{M} \hat{p}_j - \int_0^1 x w(x) \, \mathrm{d}x, \quad i = 1, \ldots, M.$$

Subtracting the ith equation from the $(i+1)$st one, we see that the points have to fulfill $\int_{\hat{p}_i}^{\hat{p}_{i+1}} w(x) \, \mathrm{d}x = \lambda$, $i = 1, \ldots, M - 1$. It follows that for the constant weight $w \equiv 1$ the point $\hat{p}_1$ can be chosen arbitrarily in $[0, 1/M]$.

   Kernels on $\mathbb{T}^2$. The torus $\mathbb{T}^2$ is naturally embedded in $\mathbb{R}^4$ by

$$\mathbb{T}^2 = \{x := (2\pi)^{-1}\big(\cos(2\pi\alpha_1), \sin(2\pi\alpha_1), \cos(2\pi\alpha_2), \sin(2\pi\alpha_2)\big)^T \in \mathbb{R}^4 : \alpha_1, \alpha_2 \in [0, 1)\}$$

with geodesic distance

$$\mathrm{d}_{\mathbb{T}^2}(x, y) = \sqrt{\mathrm{d}_{\mathbb{S}^1}(\alpha_1, \beta_1)^2 + \mathrm{d}_{\mathbb{S}^1}(\alpha_2, \beta_2)^2}$$

for

$$x := (2\pi)^{-1}\big(\cos(2\pi\alpha_1), \sin(2\pi\alpha_1), \cos(2\pi\alpha_2), \sin(2\pi\alpha_2)\big)^T,$$
$$y := (2\pi)^{-1}\big(\cos(2\pi\beta_1), \sin(2\pi\beta_1), \cos(2\pi\beta_2), \sin(2\pi\beta_2)\big)^T.$$

The restriction of the negative Euclidean distance kernel is

$$\Phi(x - y) := -\|x - y\|_2 = -\frac{1}{\pi}\sqrt{\sin^2(\pi(\alpha_1 - \beta_1)) + \sin^2(\pi(\alpha_2 - \beta_2))}$$

$$= -\frac{1}{\pi}\sqrt{1 - \cos(\pi(\alpha_1 + \alpha_2 - \beta_1 - \beta_2))\cos(\pi(\alpha_1 - \alpha_2 - \beta_1 + \beta_2))}.$$

Since the torus is flat, the balls $\mathcal{B}(x, r)$ on $\mathbb{T}^2$ with radius $r \le R \le 1/2$ can be considered as two-dimensional Euclidean balls $\tilde{\mathcal{B}}((\alpha_1, \alpha_2), r)$. In the Euclidean plane

$\mathbb{R}^2$ the area of intersection of two balls of radius r with distance d between their centers is

$$a(r,d) := \begin{cases} 2r^2 \arccos(d/(2r)) - d\sqrt{r^2 - d^2/4}, & r \geq d/2, \\ 0 & \text{else.} \end{cases}$$

With the integral
(3.3)
$A_R(d)$
$$:= \int_0^R a(r,d)\,\mathrm{d}r$$
$$= \begin{cases} \frac{R}{3}(2R^2 \arccos(d/(2R)) - d\sqrt{4R^2 - d^2}) - \frac{d^3}{12}\log\left(d/(2R + \sqrt{4R^2 - d^2})\right), & d \leq 2R, \\ 0 & \text{else,} \end{cases}$$

we obtain the kernel

$$K_{\mathcal{B}}(x,y) = \int_0^R \mu_{\mathbb{T}^2}(\mathcal{B}_r(x) \cap \mathcal{B}_r(y))\mathrm{d}r = \sum_{i=1}^4 \int_{d_i/2}^R a(r,d_i)\,\mathrm{d}r = \sum_{i=1}^4 A_R(d_i),$$

where
(3.4)
$$d_1 := \sqrt{|\alpha_1 - \beta_1|^2 + |\alpha_2 - \beta_2|^2}, \quad d_2 := \sqrt{(|\alpha_1 - \beta_1| - 1)^2 + |\alpha_2 - \beta_2|^2},$$
$$d_3 := \sqrt{|\alpha_1 - \beta_1|^2 + (|\alpha_2 - \beta_2| - 1)^2}, \quad d_4 := \sqrt{(|\alpha_1 - \beta_1| - 1)^2 + (|\alpha_2 - \beta_2| - 1)^2}.$$

For an illustration of the above relations see Figure 3, where one easily observes that this kernel is the periodization of the radial kernel $A_{1/2}(d)$. We remark that this kernel cannot be written in the form $K_{\mathcal{B}}(x,y) = \tilde{K}(d_{\mathbb{T}^2}(x,y))$, hence it is not rotationally invariant. Figure 4 shows on the left-hand side the appropriately scaled kernels $\Phi(x)$ and $K_{\mathcal{B}}(x,0)$ as functions of (α_1, α_2) which have nearly the same shape.

For our implementations it will be necessary to approximate the kernels by a trigonometric polynomial. Therefore, we expand the kernel $K_{\mathcal{B}}$ in a Fourier series

$$(3.5) \qquad K_{\mathcal{B}}(x,y) = \sum_{(n_1,n_2)^{\mathrm{T}} \in \mathbb{Z}^2} \hat{K}_{\mathcal{B}}(n_1,n_2)\mathrm{e}^{-2\pi\mathrm{i}(n_1,n_2)^{\mathrm{T}}\cdot(\alpha_1-\beta_1,\alpha_2-\beta_2)^{\mathrm{T}}}.$$

Since the kernel $K_{\mathcal{B}}$ is the periodization of the radial kernel $A_{1/2}(d)$ the Fourier coefficients of $K_{\mathcal{B}}$ are given by

$$\hat{K}_{\mathcal{B}}(n_1,n_2) = \hat{k}_{\mathcal{B}}\left(\sqrt{n_1^2 + n_2^2}\right),$$

where the function

$$\hat{k}_{\mathcal{B}}(0) = 2\pi \int_0^1 x A_{1/2}(x)\mathrm{d}x = \frac{\pi^2}{160},$$

$$\hat{k}_{\mathcal{B}}(r) = 2\pi \int_0^1 x A_{1/2}(x) J_0(2\pi x r)\mathrm{d}x = \frac{\pi^2}{160}\,_2F_3\left(\frac{3}{2}, \frac{5}{2}; 2, 3, \frac{7}{2}; -\pi^2 r^2\right), \qquad r > 0,$$

can be expressed via the *Bessel function of first kind* J_0 and the *hypergeometric function* $_2F_3$.
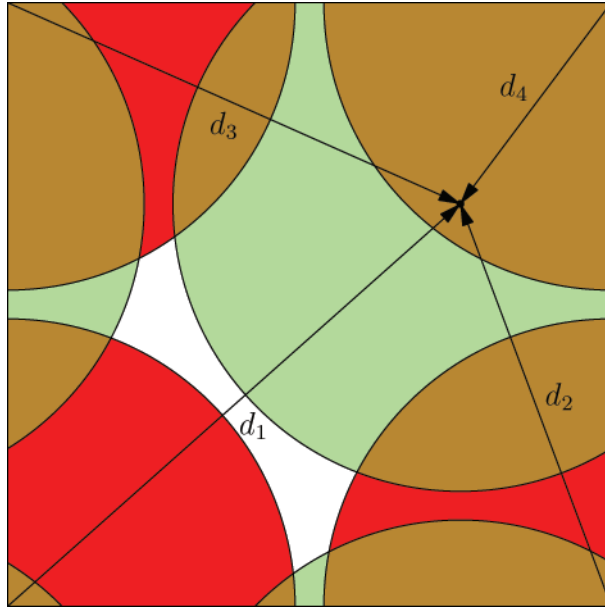
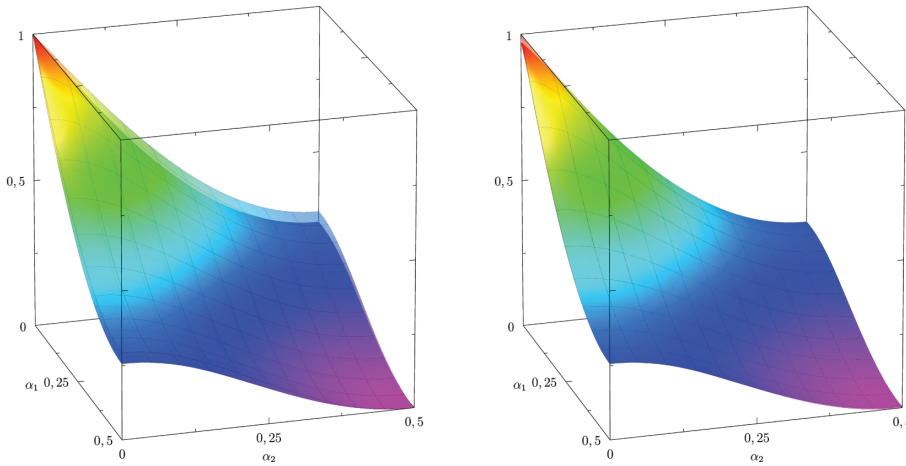FIG. 3. *Visualization of the intersection of two balls on the torus* $\mathbb{T}^2$; *cf.* (3.3)–(3.4).

FIG. 4. *Left: Scaled kernels* $\Phi(x)$ *and* $K_{\mathcal{B}}(x,0)$ *on* $\mathbb{T}^2$. *Right: Scaled kernels* $K_{\mathcal{B}}(x,0)$ *and* $\tilde{K}_{\mathcal{B}}(x,0)$ *as functions of* (α_1, α_2).

The right-hand side of Figure 4 depicts the scaled versions of the kernel $K_{\mathcal{B}}(x,0)$ and the truncation of $\tilde{K}_{\mathcal{B}}(x,0)$ with bandwidth $N = 40$, where one observes a notable difference only at the origin.

Kernels on $\mathbb{S}^2$. The sphere $\mathbb{S}^2$ is embedded in $\mathbb{R}^3$ by using spherical coordinates

$$x = x(\theta, \varphi) := (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)^{\mathsf{T}}, \qquad (\varphi, \theta) \in [0, 2\pi) \times [0, \pi].$$

The geodesic distance is given by

$$\mathrm{d}_{\mathbb{S}^2}(x,y) = \arccos(x \cdot y), \qquad x, y \in \mathbb{S}^2.$$

The restricted distance kernel has the form

$$\Phi(x - y) = -\|x - y\|_2 = -2\sin(\mathrm{d}_{\mathbb{S}^2}(x, y)/2).$$

On the sphere there is no special direction. Hence, the discrepancy kernel $K_{\mathcal{B}}$ obtained from the spherical caps $\mathcal{B}(c, r)$, $c \in \mathbb{S}^2$ with radius $r \leq \pi$ is rotationally invariant, i.e., $K_{\mathcal{B}}(x, y) = \tilde{K}_{\mathcal{B}}(\mathrm{d}_{\mathbb{S}^2}(x, y))$. For computing the function $\tilde{K}_{\mathcal{B}}$ we need the area of intersection of two spherical caps with center distance d and radius r which is given by

$$a(r, d) = \begin{cases} 0, & 0 \leq r \leq d/2, \\ 4\left[\arccos\left(\sin(d/2)/\sin r\right) - \cos r \arccos\left(\tan(d/2)\cot r\right)\right], & d/2 < r < \pi/2, \\ 4r - 2d, & r = \pi/2, \\ 4\left[\arccos\left(\sin(d/2)/\sin r\right) - \cos r \arccos\left(\tan(d/2)\cot r\right)\right], & \pi/2 < r < \pi - d/2, \\ -4\pi\cos r, & \pi - d/2 \leq r < \pi. \end{cases}$$

Then the discrepancy kernel is given by

$$K_{\mathcal{B}}(x, y) = \tilde{K}_{\mathcal{B}}(\mathrm{d}_{\mathbb{S}^2}(x, y)) = \int_0^\pi a(r, \mathrm{d}_{\mathbb{S}^2}(x, y))\, \mathrm{d}r.$$

Figure 5 shows a plot of the scaled distance and discrepancy kernels as functions of $\mathrm{d}_{\mathbb{S}^2}(x, (0, 0, 1)^{\mathrm{T}})$.
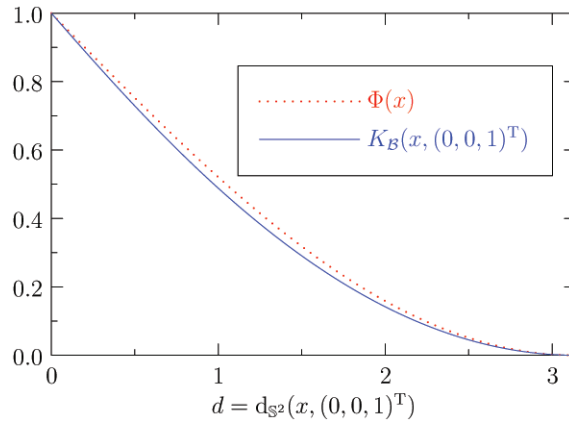

FIG. 5. *Scaled kernels $\Phi(x)$ and $K_{\mathcal{B}}(x, (0, 0, 1)^T)$ on $\mathbb{S}^2$ as functions of $\mathrm{d}_{\mathbb{S}^2}(x, (0, 0, 1)^T)$.*

4. Computation of minimizers on $\mathbb{S}^1$, $\mathbb{T}^2$, and $\mathbb{S}^2$. In this section, we present algorithms for the efficient computation of local minimizers $\hat{p}$ of functionals E_K for given functions w on $\mathcal{X} \in \{\mathbb{S}^1, \mathbb{T}^2, \mathbb{S}^2\}$. The case $\mathcal{X} = \mathbb{S}^1$ is only included for convenience. In the next section, we will use the resulting coordinates $\hat{p}_i$, $i = 1, \ldots, M$, as point positions for halftoning of w on the torus and the sphere. Our algorithms rest upon bandlimited kernels $K = K_N$ which approximate the distance/discrepancy kernels from the previous section. First, we reformulate E_{K_N} as a nonlinear least squares functional $\mathcal{E}_N$. We will see that the evaluation of this functional, its gradient, and the vector multiplication with its Hessian at any point $p \in \mathcal{X}^M$ can be realized in an efficient way. This will be used for the nonlinear CG method to compute a local minimizer of $\mathcal{E}_N$.

4.1. A least squares setting. Let $\mathcal{X} \in \{\mathbb{S}^1, \mathbb{T}^2, \mathbb{S}^2\}$ and let $\{\psi_l : l \in \mathbb{N}\}$ be an orthonormal basis of $L_2(\mathcal{X})$. Then any real-valued function $w \in L_2(\mathcal{X})$ can be written in the form

$$(4.1) \qquad w(x) = \sum_{l=1}^{\infty} \hat{w}_l \psi_l(x), \quad \hat{w}_l = \langle w, \psi_l \rangle_{L_2} = \int_{\mathcal{X}} w(x) \overline{\psi_l(x)} \, dx.$$

We will work in spaces of bandlimited functions

$$\Pi_N(\mathcal{X}) := \operatorname{span}\{\psi_l : l = 1, \ldots, d_N\}$$

of dimension $d_N := \dim \Pi_N(\mathcal{X})$. More precisely, we will use the following settings:

$$\begin{aligned}
&\Pi_N(\mathbb{S}^1) \\
&:= \operatorname{span}\{\, e^{-2\pi i n(\cdot)} \, : \, n = -N/2, \ldots, N/2\}, && d_N = N+1, \\
&\Pi_N(\mathbb{T}^2) \\
&:= \operatorname{span}\{\, e^{-2\pi i n(\cdot)} \, : \, n = (n_1, n_2)^{\mathrm{T}}, n_j = -N/2, \ldots, N/2, \, j = 1, 2\}, && d_N = (N+1)^2, \\
&\Pi_N(\mathbb{S}^2) \\
&:= \operatorname{span}\{Y_n^k \, : \, n = 0, \ldots, N; k = -n, \ldots, n\}, && d_N = (N+1)^2,
\end{aligned}$$

where N is supposed to be even in the first two cases. Here Y_n^k denotes the *spherical harmonics* of degree n and order k (cf. [42]),

$$Y_n^k(x) = Y_n^k(\theta, \varphi) := \sqrt{\frac{2n+1}{4\pi}} P_n^{|k|}(\cos\theta) \, e^{ik\varphi},$$

where the *associated Legendre functions* $P_n^k : [-1, 1] \to \mathbb{R}$ and the *Legendre polynomials* $P_n : [-1, 1] \to \mathbb{R}$ are given by

$$P_n^k(x) := \left(\frac{(n-k)!}{(n+k)!}\right)^{1/2} \left(1-x^2\right)^{k/2} \frac{d^k}{dx^k} P_n^k(x), \qquad n \in \mathbb{N}_0, \; k = 0, \ldots, n,$$

$$P_n(x) := \frac{1}{2^n n!} \frac{d^n}{dx^n} \left(x^2 - 1\right)^n, \qquad n \in \mathbb{N}_0.$$

We will apply bandlimited kernels of the form

$$(4.2) \qquad\qquad K_N(x, y) := \sum_{l=1}^{d_N} \lambda_l \psi_l(x) \overline{\psi_l(y)}$$

with $\lambda_l > 0$. Note that these kernels are reproducing kernels for the RKHSs $H_{K_N} := \Pi_N(\mathcal{X})$ with the inner product

$$\langle f, g \rangle_{H_{K_N}} = \sum_{l=1}^{d_N} \frac{\hat{f}_l \overline{\hat{g}_l}}{\lambda_l}.$$

We are interested in minimizers of E_{K_N} for given $w \in L_2(\mathcal{X})$. For the efficient computation of these minimizers it is useful to rewrite the functional as a weighted least squares problem.

THEOREM 4.1. *Let the kernel K_N be given by (4.2) and let $H_{K_N} := \Pi_N(\mathcal{X})$ be the associated RKHS. Then the relation $\mathrm{err}_{K_N}(\boldsymbol{p})^2 = \mathcal{E}_N(\boldsymbol{p})$ holds true, where*

$$(4.3) \qquad \mathcal{E}_N(\boldsymbol{p}) := \sum_{l=1}^{d_N} \lambda_l \left| \lambda \sum_{i=1}^{M} \overline{\psi_l(p_i)} - \hat{w}_l \right|^2 = \| \boldsymbol{\Lambda}^{\frac{1}{2}} F(\boldsymbol{p}) \|_2^2$$

with $\boldsymbol{\Lambda} := \mathrm{diag}(\lambda_l)_{l=1}^{d_N}$ and $F(\boldsymbol{p}) = (F_l(\boldsymbol{p}))_{l=1}^{d_N}$, $F_l(\boldsymbol{p}) := \lambda \sum_{i=1}^{M} \overline{\psi_l(p_i)} - \hat{w}_l$. In particular, the functionals E_{K_N} and $\mathcal{E}_N$ have the same minimizers.

Proof. We rewrite the function in (4.3) as

$$\mathcal{E}_N(\boldsymbol{p}) = \lambda^2 \sum_{l=1}^{d_N} \lambda_l \left| \sum_{i=1}^{M} \left(\overline{\psi_l(p_i)} - \frac{\hat{w}_l}{M\lambda} \right) \right|^2$$

$$= \lambda^2 \sum_{l=1}^{d_N} \lambda_l \sum_{i=1}^{M} \left(\overline{\psi_l(p_i)} - \frac{\hat{w}_l}{M\lambda} \right) \sum_{j=1}^{M} \left(\psi_l(p_j) - \frac{\overline{\hat{w}_l}}{M\lambda} \right)$$

$$= \lambda^2 \sum_{l=1}^{d_N} \lambda_l \sum_{i=1}^{M} \sum_{j=1}^{M} \left(\frac{|\hat{w}_l|^2}{M^2\lambda^2} - \frac{\hat{w}_l}{M\lambda} \psi_l(p_i) - \frac{\overline{\hat{w}_l}}{M\lambda} \overline{\psi_l(p_j)} + \overline{\psi_l(p_i)} \psi_l(p_j) \right)$$

$$= \sum_{l=1}^{d_N} \lambda_l |\hat{w}_l|^2 - 2\lambda \mathrm{Re} \left(\sum_{l=1}^{d_N} \lambda_l \hat{w}_l \sum_{i=1}^{M} \psi_l(p_i) \right) + \lambda^2 \sum_{i=1}^{M} \sum_{j=1}^{M} K_N(p_i, p_j).$$

Using the relation $h_w(x) = \sum_{l=1}^{d_N} \lambda_l \hat{w}_l \psi_l(x)$ (cf. (2.3)) we further conclude that

$$\mathcal{E}_N(\boldsymbol{p}) = \|h_w\|_{H_{K_N}}^2 - 2\lambda \sum_{i=1}^{M} \int_{\mathcal{X}} w(x) K_N(p_i, x) \mathrm{d}x + \lambda^2 \sum_{i=1}^{M} \sum_{j=1}^{M} K_N(p_i, p_j),$$

which yields the assertion. ☐

Remark 4.2 (relation to spherical designs). By Theorem 4.1 we have that $I_w(f) = Q(f, \boldsymbol{p})$ for some $\boldsymbol{p} \in \mathcal{X}^M$ and all $f \in H_{K_N}$ if and only if $\mathrm{err}_{K_N}(\boldsymbol{p})^2 = \mathcal{E}_N(\boldsymbol{p}) = 0$.

Consider the case $\mathcal{X} = \mathbb{S}^2$, $w \equiv 1 = \sqrt{4\pi} Y_0^0$, and

$$(4.4) \qquad K_N(x, y) := \sum_{n=0}^{N} \sum_{k=-n}^{n} \lambda_n Y_n^k(x) \overline{Y_n^k(y)} = \sum_{n=0}^{N} \lambda_n \frac{2n+1}{4\pi} P_n(x \cdot y).$$

A set $\{p_i : i = 1, \ldots, M\}$ satisfying

$$\int_{\mathbb{S}^2} f(x) \, \mathrm{d}x = \frac{4\pi}{M} \sum_{i=1}^{M} f(p_i) \qquad \forall f \in \Pi_N(\mathbb{S}^2)$$

is called *spherical N-design*. The concept of spherical N-designs was introduced by Delsarte, Goethals, and Seidel [15]. Up to now there is no theoretical result which proves the existence of an N-design with $M = (N+1)^2$ nodes for arbitrary $N \in \mathbb{N}$. But recently, in [12] it was verified that for $N = 1, \ldots, 100$, spherical N-designs with $(N+1)^2$ nodes exist using the characterization of fundamental spherical N-designs and interval arithmetic. For further recent developments on spherical N-designs and

related topics we refer to the survey article [5]. Finally, we remark that the equivalence between spherical N-designs and the relation

$$\mathcal{E}_N(\boldsymbol{p}) = \lambda^2 \sum_{n=1}^{N} \sum_{k=-n}^{n} \lambda_n \left| \sum_{i=1}^{M} Y_n^k(p_i) \right|^2 = 0$$

was applied by Sloan and Womersley in [50].

The rest of this section describes how to compute for given $\hat{w}_l$, $l = 1, \ldots, d_N$, local minimizers of $\mathcal{E}_N(\boldsymbol{p})$ in an efficient way also for large numbers M of point positions. As our algorithm of choice we present the nonlinear CG method in the next subsection. The efficient computation of each CG step rests upon

- algorithms for the fast evaluation of bandlimited functions on $\mathbb{S}^1, \mathbb{T}^2$, and $\mathbb{S}^2$,
- simple representations of the gradient and the Hessian of $\mathcal{E}_N$.

In the following, we describe these two items in more detail. The evaluation of bandlimited functions

$$f(p_i) = \sum_{l=1}^{d_N} \hat{f}_l \psi_l(p_i), \quad i = 1, \ldots, M,$$

can be written in matrix-vector form as

$$\boldsymbol{f} = \boldsymbol{A}_N \hat{\boldsymbol{f}},$$

where $\boldsymbol{f} := (f(p_i))_{i=1}^{M}$, $\hat{\boldsymbol{f}} := (\hat{f}_l)_{l=1}^{d_N}$ appropriately ordered, and

(4.5)
$$\boldsymbol{A}_N := \begin{cases} \boldsymbol{F}_N = \left(\mathrm{e}^{-2\pi i n p_i} \right)_{i=1,\ldots,M; n=-N/2,\ldots,N/2} \in \mathbb{C}^{M,N+1} & \text{for } \mathbb{S}^1, \\ \boldsymbol{F}_{2,N} = \left(\mathrm{e}^{-2\pi i (n_1,n_2)^{\mathrm{T}} \cdot p_i} \right)_{i=1,\ldots,M; n_i=-N/2,\ldots,N/2, i=1,2} \in \mathbb{C}^{M,(N+1)^2} & \text{for } \mathbb{T}^2, \\ \boldsymbol{Y}_N = (Y_k^n(p_i))_{i=1,\ldots,M; n=0,\ldots,N, |k|\leq n} \in \mathbb{C}^{M,(N+1)^2} & \text{for } \mathbb{S}^2. \end{cases}$$

Recently fast algorithms for the matrix-vector multiplication with $\boldsymbol{A}_N$ and $\overline{\boldsymbol{A}}_N^{\mathrm{T}}$ were proposed. More precisely, the algorithms for the first two cases $\mathbb{S}^1$ and $\mathbb{T}^2$, called *nonequispaced fast Fourier transform*, (NFFT) or *unequally spaced fast Fourier transform* can be found, e.g., in [8, 18, 34, 46]. The algorithms on the sphere $\mathbb{S}^2$, called *nonequispaced fast spherical Fourier transform* (NFSFT), were developed in [35, 38]; see also [16, 30]. In our numerical examples we have applied the software package [33].

PROPOSITION 4.3. *For a prescribed accuracy $\epsilon > 0$ the matrix-vector multiplication with the matrices $\boldsymbol{A}_N$ and $\overline{\boldsymbol{A}}_N^{T}$ given in (4.5) can be computed by using the NFFT/NFSFT with the following arithmetic complexity:*

(4.6)
$$\begin{array}{ll} \mathcal{O}(N \log N + M \log(1/\epsilon)) & \text{for } \mathbb{S}^1, \\ \mathcal{O}(N^2 \log N + M \log^2(1/\epsilon)) & \text{for } \mathbb{T}^2, \\ \mathcal{O}(N^2 \log^2 N + M \log^2(1/\epsilon)) & \text{for } \mathbb{S}^2. \end{array}$$

Using the NFFT/NFSFT algorithms the same complexity as in Proposition 4.3 is achieved for the evaluation of $\mathcal{E}_N(\boldsymbol{p})$ as the following corollary states.

COROLLARY 4.4 (efficient evaluation of $F(\boldsymbol{p})$ and $\mathcal{E}_N(\boldsymbol{p})$). *For a given point $\boldsymbol{p} \in \mathcal{X}^M$ and $\hat{w}_l$, $l = 1, \ldots, d_N$, the computation of $F(\boldsymbol{p})$ and $\mathcal{E}_N(\boldsymbol{p})$ can be realized with the arithmetic complexity given in Proposition* 4.3.

Proof. We conclude from the relation

$$\left(\sum_{i=1}^{M}\overline{\psi_l(p_i)}\right)_{l=1}^{d_N} = \overline{\boldsymbol{A}}_N^{\mathrm{T}}\boldsymbol{e}, \qquad \boldsymbol{e} := (1,\dots,1)^{\mathrm{T}} \in \mathbb{R}^M,$$

that the vector on the left-hand side can be computed with the arithmetic complexity provided by Proposition 4.3. The remaining operations do not increase this complexity. $\square$

Next we consider the gradient of $\mathcal{E}_N$. By $\nabla_{\mathcal{X}}\psi$ we denote the gradient of ψ on $\mathcal{X}$ and by $\nabla_{\mathcal{X}}^i \mathcal{E}_N$ the gradient of $\mathcal{E}_N$ with respect to the ith component vector p_i. Then straightforward computation shows that the gradient $\nabla = \nabla_{\mathcal{X}^M}$ of $\mathcal{E}_N$ at $\boldsymbol{p} \in \mathcal{X}^M$ is given by $\nabla \mathcal{E}_N(\boldsymbol{p}) = \left(\nabla_{\mathcal{X}}^i \mathcal{E}_N(\boldsymbol{p})\right)_{i=1}^{M}$, where

$$\nabla_{\mathcal{X}}^i \mathcal{E}_N(\boldsymbol{p}) = 2\lambda\mathrm{Re}\left[\sum_{l=1}^{d_N}\lambda_l\underbrace{\left(\lambda\sum_{j=1}^{M}\overline{\psi_l(p_j)} - \hat{w}_l\right)}_{F_l(\boldsymbol{p})}\nabla_{\mathcal{X}}\psi_l(p_i)\right].$$

Hence, the gradient can be written as

(4.7) $$\nabla \mathcal{E}_N(\boldsymbol{p}) = 2\mathrm{Re}\left[\overline{J_F(\boldsymbol{p})}^{\mathrm{T}}\boldsymbol{\Lambda}\,F(\boldsymbol{p})\right],$$

where

$$J_F(\boldsymbol{p}) := \left(\left(\nabla_{\mathcal{X}}^i F_l(p_i)\right)^{\mathrm{T}}\right)_{l=1,i=1}^{d_N,M} = \lambda\left(\nabla_{\mathcal{X}}\overline{\psi_l(p_i)}\right)_{l=1,i=1}^{d_N,M}$$

denotes the *Jacobian matrix* of F. For our three settings the gradients are specified as follows.

Gradient on $\mathbb{S}^1$. For

$$\mathcal{E}_N(\boldsymbol{p}) = \sum_{n=-N/2}^{N/2}\lambda_n\left|\lambda\sum_{i=1}^{m}\mathrm{e}^{-2\pi\mathrm{i}np_i} - \hat{w}_n\right|^2$$

we obtain with $\nabla_{\mathbb{S}^1}\psi_n(p_i) = -2\pi\mathrm{i}n\,\mathrm{e}^{-2\pi\mathrm{i}np_i}$ that

(4.8)
$$\overline{J_F(\boldsymbol{p})}^{\mathrm{T}} = \lambda\left(-2\pi\mathrm{i}n\,\mathrm{e}^{-2\pi\mathrm{i}np_i}\right)_{i=1,n=-N/2}^{M,N/2} = \lambda\boldsymbol{F}_N\boldsymbol{D}_N, \quad \boldsymbol{D}_N := \mathrm{diag}\left(-2\pi\mathrm{i}n\right)_{n=-N/2}^{N/2}.$$

Gradient on $\mathbb{T}^2$. For

$$\mathcal{E}_N(\boldsymbol{p}) = \sum_{n_1,n_2=-N/2}^{N/2}\lambda_n\left|\lambda\sum_{i=1}^{M}\mathrm{e}^{-2\pi\mathrm{i}n\cdot p_i} - \hat{w}_n\right|^2, \qquad n := (n_1,n_2)^{\mathrm{T}},$$

we get with $\nabla_{\mathbb{T}^2}\psi_n(p_i) = -2\pi\mathrm{i}n\,\mathrm{e}^{-2\pi\mathrm{i}n\cdot p_i}$ and an appropriate ordering that

(4.9) $$\overline{J_F(\boldsymbol{p})}^{\mathrm{T}} = \lambda(\boldsymbol{I}_2 \otimes \boldsymbol{F}_{2,N})\begin{pmatrix}\boldsymbol{I}_N \otimes \boldsymbol{D}_N \\ \boldsymbol{D}_N \otimes \boldsymbol{I}_N\end{pmatrix},$$

where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix and $\otimes$ the Kronecker product.

Gradient on $\mathbb{S}^2$. On the sphere we will work only with kernels of the form (4.4) such that

$$\mathcal{E}_N(\boldsymbol{p}) = \sum_{n=0}^{N} \sum_{k=-n}^{n} \lambda_n \left| \lambda \sum_{i=1}^{m} \overline{Y_n^k(p_i)} - \hat{w}_n^k \right|^2 .$$

Then we have to clarify the definition of the gradient of a function on $\mathbb{S}^2$. To this end, let $T_x \mathbb{S}^2 := \{\boldsymbol{v} \in \mathbb{R}^3 : \langle \boldsymbol{v}, \boldsymbol{x} \rangle = 0\}$ be the *tangent space* at a point $x \in \mathbb{S}^2$. For $x := x(\theta, \varphi) \in \mathbb{S}^2 \backslash \{(0,0,\pm1)^{\mathsf{T}}\}$ this tangent space is spanned by the orthonormal vectors $e_\varphi := (-\sin\varphi, \cos\varphi, 0)^{\mathsf{T}}$ and $e_\theta := (\cos\theta\cos\varphi, \cos\theta\sin\varphi, -\sin\theta)^{\mathsf{T}}$. Then the *spherical gradient operator* is defined as

$$\nabla_{\mathbb{S}^2} := e_\varphi \frac{1}{\sin\theta} \frac{\partial}{\partial\varphi} + e_\theta \frac{\partial}{\partial\theta};$$

cf. [24, 57]. Note that this is the orthogonal projection of $\nabla\tilde{\psi}(x) \in \mathbb{R}^3$ onto $T_x \mathbb{S}^2$, where $\tilde{\psi}$ denotes an extension of ψ to $\mathbb{R}^3$. In particular, the derivatives of the spherical harmonics can be computed by

$$\frac{\partial}{\partial\varphi} Y_n^k(\theta, \varphi) = \mathrm{i}k Y_n^k(\theta, \varphi),$$

$$\sin\theta \frac{\partial}{\partial\theta} Y_n^k(\theta, \varphi) = n \underbrace{\sqrt{\frac{(n+1)^2 - k^2}{(2n+1)(2n+3)}}}_{a_{n+1}^k} Y_{n+1}^k(\theta, \varphi)$$

$$- (n+1) \underbrace{\sqrt{\frac{n^2 - k^2}{(2n+1)(2n-1)}}}_{b_{n-1}^k} Y_{n-1}^k(\theta, \varphi),$$

where $Y_{n-1}^k := 0$ for $|k| > n-1$; see [61, p. 146]. Using this relation we obtain for $p_i = x(\theta_i, \varphi_i) \in \mathbb{S}^2 \backslash \{(0,0,\pm1)^{\mathsf{T}}\}$ that

$$\nabla_{\mathbb{S}^2} Y_n^k(p_i) = \frac{1}{\sin\theta_i} \mathrm{i}k\, Y_n^k(\varphi_i, \theta_i)\, e_{\varphi_i} + \frac{1}{\sin\theta_i} \left(a_{n+1}^k Y_{n+1}^k(\varphi_i, \theta_i) - b_{n-1}^k Y_{n-1}^k(\varphi_i, \theta_i) \right) e_{\theta_i}$$

and consequently

(4.10)

$$\nabla_{\mathbb{S}^2}^i \mathcal{E}_N(\boldsymbol{p}) = \frac{2\lambda}{\sin\theta_i} \mathrm{Re} \left[\sum_{n=0}^{N} \sum_{k=-n}^{n} Y_n^k(\varphi_i, \theta_i)\, \mathrm{i}k\, \lambda_n \overline{F_n^k(\boldsymbol{p})} \right] e_{\varphi_i}$$

$$+ \frac{2\lambda}{\sin\theta_i} \mathrm{Re} \left[\sum_{n=0}^{N+1} \sum_{k=-n}^{n} Y_n^k(\varphi_i, \theta_i) \left(a_n^k \lambda_{n-1} \overline{F_{n-1}^k(\boldsymbol{p})} - b_n^k \lambda_{n+1} \overline{F_{n+1}^k(\boldsymbol{p})} \right) \right] e_{\theta_i}$$

$$= x_{\varphi_i} e_{\varphi_i} + x_{\theta_i} e_{\theta_i},$$

where $F_{N+1}^k(\boldsymbol{p}) = F_{N+2}^k(\boldsymbol{p}) = 0$. Hence, the coordinate vectors $\boldsymbol{x}_\varphi := (x_{\varphi_i})_{i=1}^{M}$ and $\boldsymbol{x}_\theta := (x_{\theta_i})_{i=1}^{M}$ can be computed by

(4.11)
$$\begin{pmatrix} \boldsymbol{x}_\varphi \\ \boldsymbol{x}_\theta \end{pmatrix} = 2\lambda \mathrm{Re} \left[(\boldsymbol{I}_2 \otimes \boldsymbol{S}^{-1}) \begin{pmatrix} \boldsymbol{Y}_N \boldsymbol{D}_{N,\varphi} \\ \boldsymbol{Y}_{N+1} \tilde{\boldsymbol{D}}_{N,\theta} \end{pmatrix} \boldsymbol{\Lambda}\, F(\boldsymbol{p}) \right],$$

where $\boldsymbol{S} := \mathrm{diag}\,(\sin\theta_i)_{i=1}^M$, $\boldsymbol{D}_{N,\varphi}$ is the diagonal matrix determined by the first summand in (4.10), and $\tilde{\boldsymbol{D}}_{N,\theta}$ is the matrix with at most two nonzero entries in each row corresponding to the second summand in (4.10).

In summary we obtain the following corollary.

COROLLARY 4.5 (efficient evaluation of $\nabla\mathcal{E}_N(\boldsymbol{p})$). *For a given point $\boldsymbol{p} \in \mathcal{X}^M$ and given $\hat{w}_l$, $l = 1,\ldots,d_N$, the gradient $\nabla\mathcal{E}_N(\boldsymbol{p})$ can be computed with the arithmetic complexity given in Proposition 4.3.*

Proof. The proof follows by Corollary 4.4 and the relation (4.7) together with (4.8), (4.9), and (4.11). ∎

Finally, we are interested in the Hessian $\mathrm{H} = \mathrm{H}_{\mathcal{X}^M}$ of $\mathcal{E}_N$. By $\mathrm{H}_{\mathcal{X}}\psi$ we denote the Hessian of ψ on $\mathcal{X}$. By straightforward computation we obtain that

$$\mathrm{H}\mathcal{E}_N(\boldsymbol{p}) = \left(\mathrm{H}^{i,j}\mathcal{E}_N(\boldsymbol{p})\right)_{i,j=1}^M,$$

where

$$(4.12) \qquad \mathrm{H}^{i,j}\mathcal{E}_N(\boldsymbol{p}) = 2\lambda^2\mathrm{Re}\left[\sum_{l=1}^{d_N}\lambda_l\,\nabla_{\mathcal{X}}\psi_l(p_i)\left(\nabla_{\mathcal{X}}\overline{\psi_l(p_j)}\right)^{\mathrm{T}}\right]$$

$$(4.13) \qquad + \delta_{i,j}2\lambda\mathrm{Re}\left[\sum_{l=1}^{d_N}\lambda_l\underbrace{\left(\lambda\sum_{m=1}^M\overline{\psi_l(p_m)} - \hat{w}_l\right)}_{F_l(\boldsymbol{p})}\mathrm{H}_{\mathcal{X}}\psi_l(p_i)\right].$$

For the proposed optimization methods one can also use an approximation $\tilde{\mathrm{H}}\mathcal{E}_N$ of the Hessian of $\mathcal{E}_N$, which involves only the first summand (4.12) in the above expression, i.e., the diagonal part is neglected so that

$$\tilde{\mathrm{H}}\mathcal{E}_N(\boldsymbol{p}) := 2\lambda^2\mathrm{Re}\left(\sum_{l=1}^{d_N}\lambda_l\,\nabla_{\mathcal{X}}\psi_l(p_i)\left(\nabla_{\mathcal{X}}\overline{\psi_l(p_j)}\right)^{\mathrm{T}}\right)_{i,j=1}^M$$

$$(4.14) \qquad = 2\mathrm{Re}\left[\overline{J_F(\boldsymbol{p})}^{\mathrm{T}}\boldsymbol{\Lambda}\,J_F(\boldsymbol{p})\right].$$

This matrix does not depend on the values $\hat{w}_l$. Note that the approximate Hessian is also used in the Gauss–Newton method for solving nonlinear least squares problems; see [9, sect. 9.2]. Let us specify the Hessian for our three settings.

Hessian on $\mathbb{S}^1$. Since $\mathrm{H}_{\mathbb{S}^1}\left(e^{-2\pi i n p_i}\right) = (2\pi i n)^2\,e^{-2\pi i n p_i}$, we obtain together with (4.14) and (4.8) that

$$\mathrm{H}\mathcal{E}_N(\boldsymbol{p}) = 2\lambda^2\mathrm{Re}\left[\boldsymbol{F}_N\boldsymbol{D}_N^2\boldsymbol{\Lambda}\overline{\boldsymbol{F}_N}^{\mathrm{T}}\right] + 2\lambda\mathrm{diag}\left(\mathrm{Re}\left[\boldsymbol{F}_N\boldsymbol{D}_N^2\boldsymbol{\Lambda}F(\boldsymbol{p})\right]\right).$$

Hessian on $\mathbb{T}^2$. We have that

$$\mathrm{H}_{\mathbb{T}^2}\left(e^{-2\pi i n\cdot p_i}\right) = -4\pi^2\begin{pmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{pmatrix}e^{-2\pi i n\cdot p_i}.$$

Hence, the block-diagonal part (4.13) of the Hessian is given by

$$2\lambda\mathrm{Re}\left[(\boldsymbol{I}_2\otimes\boldsymbol{F}_{2,N})\begin{pmatrix} \boldsymbol{I}_N\otimes\boldsymbol{D}_N^2 & \boldsymbol{D}_N\otimes\boldsymbol{D}_N \\ \boldsymbol{D}_N\otimes\boldsymbol{D}_N & \boldsymbol{D}_N^2\otimes\boldsymbol{I}_N \end{pmatrix}(\boldsymbol{I}_2\otimes\boldsymbol{\Lambda}F(\boldsymbol{p}))\right]$$

after a corresponding sorting. Thus, the multiplication of the first (4.12) and second part (4.13) of the Hessian with a vector can be realized in a fast way by applying (4.14) and (4.9).

Hessian on $\mathbb{S}^2$. The Hessian $\mathrm{H}_{\mathbb{S}^2}\psi(x)$ is a linear operator on $\mathrm{T}_x\mathbb{S}^2$. For $x := x(\theta, \varphi)$ we consider the corresponding matrix with respect to the basis $\{e_\theta, e_\varphi\}$. Then the Hessian reads as

$$\sin^2\theta\mathrm{H}_{\mathbb{S}^2} = \left(\sin\theta\frac{\partial}{\partial\theta}, \frac{\partial}{\partial\theta}\right)^{\mathrm{T}}\left(\sin\theta\frac{\partial}{\partial\theta}, \frac{\partial}{\partial\theta}\right) - \cos\theta\begin{pmatrix} \sin\theta\frac{\partial}{\partial\theta} & \frac{\partial}{\partial\varphi} \\ \frac{\partial}{\partial\varphi} & -\sin\theta\frac{\partial}{\partial\theta} \end{pmatrix};$$

cf. [57]. Using the relations for the derivatives of the spherical harmonics again, we conclude that the block-diagonal part (4.13) of the Hessian can be expressed by

$$(4.15) \qquad\qquad 2\lambda\mathrm{Re}\left[(\boldsymbol{I}_2 \otimes \boldsymbol{S}^{-2})\boldsymbol{M}(\boldsymbol{I}_2 \otimes \boldsymbol{\Lambda}F(\boldsymbol{p}))\right],$$

where

$$\boldsymbol{M} := \begin{pmatrix} \boldsymbol{Y}_{N+2}\tilde{\boldsymbol{D}}_{N+1,\theta}\tilde{\boldsymbol{D}}_{N,\theta} - \boldsymbol{C}\boldsymbol{Y}_{N+1}\tilde{\boldsymbol{D}}_{N,\theta} & \boldsymbol{Y}_{N+1}\tilde{\boldsymbol{D}}_{N,\theta}\boldsymbol{D}_{N,\varphi} - \boldsymbol{C}\boldsymbol{Y}_N\boldsymbol{D}_{N,\varphi} \\ \boldsymbol{Y}_{N+1}\tilde{\boldsymbol{D}}_{N,\theta}\boldsymbol{D}_{N,\varphi} - \boldsymbol{C}\boldsymbol{Y}_N\boldsymbol{D}_{N,\varphi} & \boldsymbol{Y}_N\boldsymbol{D}_{N,\varphi}\boldsymbol{D}_{N,\varphi} + \boldsymbol{C}\boldsymbol{Y}_{N+1}\tilde{\boldsymbol{D}}_{N,\theta} \end{pmatrix}$$

with $\boldsymbol{C} := \mathrm{diag}(\cos\theta_i)_{i=1}^M$ after a corresponding sorting. Hence, the multiplication of the first part (4.12) and second one (4.13) of the Hessian with a vector can be realized in a fast way by applying (4.14) and (4.9).

We summarize our findings (4.12)–(4.15) in the following corollary.

COROLLARY 4.6 (efficient vector multiplication with $\mathrm{H}\mathcal{E}_N(\boldsymbol{p})$). *For a given point $\boldsymbol{p} \in \mathcal{X}^M$ and given $\hat{w}_l$, $l = 1, \ldots, d_N$, the multiplication of a vector with the Hessian $\mathrm{H}\mathcal{E}_N(\boldsymbol{p})$ can be computed with the arithmetic complexity given in Proposition 4.3.*

4.2. Nonlinear CG algorithm. Among the various minimization strategies for weighted least squares functionals as the method of steepest descent, the Newton method, the Levenberg–Marquardt algorithm, and the nonlinear CG algorithm we restrict ourselves to the latter one. The reason for the choice of the CG method is that for large numbers of points it provides a reasonable tradeoff between computational effort and convergence towards a local minimizer. More precisely, using the CG method only matrix-vector multiplications with the Hessian are needed instead of solving a huge linear equation system. Furthermore, in the neighborhood of a local minimizer superlinear convergence is proved for the CG method under relative mild regularity assumptions; cf. [51]. As an example for the suitability of the nonlinear CG algorithm we refer to [26], where numerical spherical designs for high polynomial degrees are computed with quadrature error near machine precision. Since we mainly follow the lines of [26] we only briefly sketch the approach for our halftoning setting. For $\mathcal{X} = \mathbb{S}^1$ and $\mathcal{X} = \mathbb{T}^2$ we apply the CG algorithm in the Euclidean space. A good survey for various nonlinear CG methods in Euclidean space is given in [29].

Algorithm. CG ALGORITHM IN THE EUCLIDEAN SPACE $\mathbb{R}^{dM}$.
Initialization: $\boldsymbol{p}^{(0)}$, $\boldsymbol{h}^{(0)} := \nabla\mathcal{E}_N(\boldsymbol{p}^{(0)})$, $\boldsymbol{d}^{(0)} = -\boldsymbol{h}^{(0)}$, $k := 0$.
For $r = 0, 1, \ldots$ repeat until a convergence criterion is reached:
  1. Determine the step size $\alpha_r > 0$ by the search of a minimizer along $\boldsymbol{p}^{(r)} + t\boldsymbol{d}^{(r)}$, i.e., $(\boldsymbol{d}^{(r)})^{\mathrm{T}}\nabla\mathcal{E}_N(\boldsymbol{p}^{(r)} + \alpha_r\boldsymbol{d}^{(r)}) = 0$;
  2. $\boldsymbol{p}^{(r+1)} := \boldsymbol{p}^{(r)} + \alpha_r\boldsymbol{d}^{(r)}$;
  3. $\boldsymbol{h}^{(r+1)} := \nabla\mathcal{E}_N(\boldsymbol{p}^{(r+1)})$;

4. Compute

$$\beta_r := \frac{\langle \boldsymbol{h}^{(r+1)}, \mathrm{H}\mathcal{E}_N(\boldsymbol{p}^{(r+1)})\boldsymbol{d}^{(r)}\rangle}{\langle \boldsymbol{d}^{(r)}, \mathrm{H}\mathcal{E}_N(\boldsymbol{p}^{(r+1)})\boldsymbol{d}^{(r)}\rangle};$$

5. $\boldsymbol{d}^{(r+1)} := -\boldsymbol{h}^{(r+1)} + \beta_r \boldsymbol{d}^{(r)}$;
6. If $r + 1 \equiv k \mod dM$ or $(\boldsymbol{d}^{(r+1)})^{\mathrm{T}}\boldsymbol{h}^{(r+1)} \geq 0$, set $\boldsymbol{d}^{(r+1)} := -\boldsymbol{h}^{(r+1)}$, $k := r + 1$.

There exist other variants of the CG algorithm presented here, which differ by the choice of β_r in step 4. The above method is the one for exact conjugacy proposed by Daniel in [13], which uses the second order derivative information provided by the Hessian. In contrast, the choices of β_k proposed by Fletcher–Reeves or Polak–Ribière can be considered as finite difference approximations for the Hessian. Furthermore, the line search in the first step of the generic CG algorithm is idealized and for the most instances impractically to compute. Hence, usually inexact line search rules are applied. Since we can incorporate information of the second order derivatives we approximate the step size by the first one-dimensional Newton step, which is determined by

$$(4.16) \qquad \alpha_r := -\frac{\langle \boldsymbol{d}^{(r)}, \boldsymbol{h}^{(r)}\rangle}{\langle \boldsymbol{d}^{(r)}, \mathrm{H}\mathcal{E}_N(\boldsymbol{p}^{(r)})\boldsymbol{d}^{(r)}\rangle}.$$

We remark that the CG method with the step size rule (4.16) converges to a local minimizer with positive definite Hessian if the starting point is sufficiently close to that minimizer; cf. [14]. Thus, the CG algorithm presented here is just a local optimization method and global convergence, i.e., convergence for every starting point $\boldsymbol{p}^{(0)}$, is not necessarily assured. However, in our setting the numerical results indicate the suitability of the simple CG method given above. For other variants of the CG method, e.g., that of Fletcher–Reeves, global convergence to stationary points is proved if the step size α_r satisfies a Wolfe condition; cf. [29].

In the case $\mathcal{X} = \mathbb{S}^2$ we use the nonlinear CG algorithm on Riemannian manifolds $\mathcal{M}$; cf. [19, 51]. In Riemannian geometry the addition of a tangent vector from $\mathrm{T}_x\mathcal{M}$ to the base point $x \in \mathcal{M}$ as required in step 2 of the CG algorithm is replaced by the *exponential map* $\exp_x : \mathrm{T}_x\mathcal{M} \to \mathcal{M}$. Furthermore, the translation of tangent vectors which is needed in steps 4 and 5 of the CG algorithm is replaced by the concept of *parallel transport of a vector along geodesics*. In the particular situation we deal with $\mathcal{M} := (\mathbb{S}^2)^M$. In the following, we introduce the above concepts on $\mathbb{S}^2$ which generalize in a straightforward way to $(\mathbb{S}^2)^M$. For $x \in \mathbb{S}^2$ and $v \in \mathrm{T}_x\mathbb{S}^2$, we consider the unique geodesic curve $g_{x,v} : \mathbb{R} \to \mathbb{S}^2$ with $g_{x,v}(0) = x$ and $\dot{g}_{x,v}(0) = v$, which is given by

$$g_{x,v}(t) = \cos(t\|v\|_2)\, x + \sin(t\|v\|_2)\, \underbrace{\frac{v}{\|v\|_2}}_{\tilde{v}}, \quad t \in \mathbb{R}.$$

Note that the vectors $\tilde{v} := \frac{v}{\|v\|_2}$ and $x \times \tilde{v}$ form an orthonormal system of $\mathrm{T}_x\mathbb{S}^2$ and that any $w \in \mathrm{T}_x\mathbb{S}^2$ can be written as $w = \langle w, \tilde{v}\rangle\tilde{v} + \langle w, x \times \tilde{v}\rangle(x \times \tilde{v})$. The *exponential map* $\exp_x : \mathrm{T}_x\mathbb{S}^2 \to \mathbb{S}^2$ is explicitly parameterized by the geodesic due to $\exp_x(v) = g_{x,v}(1)$; see [57, p. 19]. Now the *parallel transport of a vector* $w \in \mathrm{T}_x\mathbb{S}^2$

along the geodesic $g_{x,v}$ is realized by

$$P_{g_{x,v}(t)}(w) := \langle w, \tilde{v} \rangle \frac{\dot{g}_{x,v}(t)}{\|v\|_2} + \langle w, x \times \tilde{v} \rangle \left(g_{x,v}(t) \times \frac{\dot{g}_{x,v}(t)}{\|v\|_2} \right)$$

$$= \langle w, \tilde{v} \rangle \left(\cos(\|v\|_2 t) \, \tilde{v} - \sin(\|v\|_2 t) \, x \right) + w - \langle w, \tilde{v} \rangle \tilde{v}, \quad t \in \mathbb{R}.$$

In particular, any geodesic $g_{x,v}$ parallel transports its own tangent vectors, i.e.,

$$P_{g_{x,v}(t)}(v) = \dot{g}_{x,v}(t), \quad t \in \mathbb{R}.$$

An illustration of the parallel transport is given in Figure 6. After these preliminaries the CG algorithm to minimize $\mathcal{E}_N$ on Riemannian manifolds, e.g., $\mathcal{M} := (\mathbb{S}^2)^M$, reads as follows.

Algorithm. CG ALGORITHM ON (dM)-DIMENSIONAL RIEMANNIAN MANIFOLDS.
Initialization: $\boldsymbol{p}^{(0)}$, $\boldsymbol{h}^{(0)} := \nabla \mathcal{E}_N(\boldsymbol{p}^{(0)})$, $\boldsymbol{d}^{(0)} = -\boldsymbol{h}^{(0)}$, $k := 0$.
For $r = 0, 1, \ldots$ repeat until a convergence criterion is reached:
1. Determine the step size $\alpha_r > 0$ by $\dot{\boldsymbol{g}}_{\boldsymbol{p}^{(r)}, \boldsymbol{d}^{(r)}}(\alpha_r)^{\mathrm{T}} \nabla \mathcal{E}_N(\boldsymbol{g}_{\boldsymbol{p}^{(r)}, \boldsymbol{d}^{(r)}}(\alpha_r)) = 0$;
2. $\boldsymbol{p}^{(r+1)} := \exp_{\boldsymbol{p}^{(r)}}(\alpha_r \boldsymbol{d}^{(r)})$;
3. $\boldsymbol{h}^{(r+1)} := \nabla \mathcal{E}_N(\boldsymbol{p}^{(r+1)})$;
4. Compute

$$\beta_r := \frac{\langle \boldsymbol{h}^{(r+1)}, \mathrm{H}\mathcal{E}_N(\boldsymbol{p}^{(r+1)}) \tilde{\boldsymbol{d}}^{(r)} \rangle}{\langle \tilde{\boldsymbol{d}}^{(r)}, \mathrm{H}\mathcal{E}_N(\boldsymbol{p}^{(r+1)}) \tilde{\boldsymbol{d}}^{(r)} \rangle}, \quad \tilde{\boldsymbol{d}}^{(r)} := \boldsymbol{P}_{\boldsymbol{g}_{\boldsymbol{p}^{(r)}, \boldsymbol{d}^{(r)}}(\alpha_r)}(\boldsymbol{d}^{(r)});$$

5. $\boldsymbol{d}^{(r+1)} := -\boldsymbol{h}^{(r+1)} + \beta_r \tilde{\boldsymbol{d}}^{(r)}$;
6. If $r + 1 \equiv k \mod dM$ or $(\boldsymbol{d}^{(r+1)})^{\mathrm{T}} \boldsymbol{h}^{(r+1)} \geq 0$, set $\boldsymbol{d}^{(r+1)} := -\boldsymbol{h}^{(r+1)}$, $k := r + 1$.

For an illustration of a single CG iteration on the sphere $\mathbb{S}^2$, see Figure 6. Again, in our numerical examples, step 1 is replaced by a one-dimensional Newton step, where the step size α_r is determined by an appropriate variant (4.16) on Riemannian manifolds. Moreover, on the sphere $\mathbb{S}^2$ we prefer the approximate Hessian $\tilde{\mathrm{H}}\mathcal{E}_N$ over the Hessian $\mathrm{H}\mathcal{E}_N$ since this leads to similar results but requires less computational effort; cf. [26].

Using the results from the previous subsection, we conclude that every CG iteration on $\mathcal{X}^M$ can be realized with the arithmetic complexity given in Proposition 4.3.

For the description of the generic algorithm we leave the stopping criteria open, since it depends on the specific application. In our first numerical example we show the efficiency of the proposed method by requiring the norm of the gradient to be less a given accuracy. In the ensuing examples we just show the results obtained after a prescribed number of iterations, since this seems to be sufficient for generating appealing point distributions in the setting of halftoning.

5. Numerical results. In the following, we present some numerical result on the torus $\mathbb{T}^2$ and the sphere $\mathbb{S}^2$. We apply the nonlinear CG methods on the functional $\mathcal{E}_N(\boldsymbol{p})$ for randomly distributed starting points $\boldsymbol{p}^{(0)} \in \mathcal{X}^M$. For the evaluation of the functional $\mathcal{E}_N(\boldsymbol{p})$, its gradient $\nabla \mathcal{E}_N(\boldsymbol{p})$, and the matrix times vector multiplication with its Hessain $\mathrm{H}\mathcal{E}_N(\boldsymbol{p})$ we have to determine the Fourier coefficients $\hat{w}_l$ and λ_l of the function $w : \mathcal{X} \to [0, \infty)$ (cf. (4.1)) and the bandlimited kernel $K_N : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (cf. (4.2)), respectively. If the Fourier coefficients $\hat{w}_l$ are not given explicitly, we compute
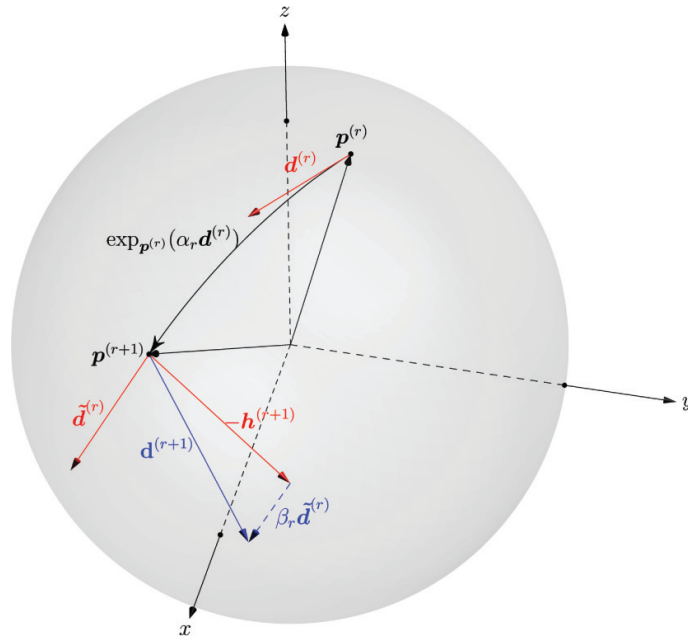
FIG. 6. *An iteration step of the nonlinear CG method on the sphere* $\mathbb{S}^2$.

them approximately by well-known quadrature rules on $\mathcal{X}$. More precisely, we sample the function w on sampling points $\boldsymbol{x} := (x_i)_{i=1}^{L} \in \mathcal{X}^L$ and obtain approximate Fourier coefficients

$$(5.1) \qquad \hat{w}_l := \sum_{i=1}^{L} \omega_i w(x_i)\overline{\psi_l(x_i)}, \qquad l = 1, \ldots, d_N,$$

where the weights ω_i are given such that

$$\int_{\mathcal{X}} f(x)\mathrm{d}x = \sum_{i=1}^{L} \omega_i f(x_i), \qquad f \in \Pi_N(\mathcal{X}).$$

We recall that the above sums can be evaluated in an efficient way by fast Fourier transforms.

The proposed algorithms are implemented in MATLAB R2010a, where the mex-interface to the NFFT library [34] is used. The internal parameters in this library are set as follows. In both routines NFFT for $\mathbb{T}^2$ and NFSFT for $\mathbb{S}^2$, we set the cutoff parameter $m = 9$. In the NFSFT we furthermore set the threshold parameter $\kappa = 1000$ and use the flags PRE_PSI and PRE_PHI_HUT. The computations are performed on an Intel(R) Core(TM) i7 CPU 920 with 12GB RAM.

Examples on $\mathbb{T}^2$. For all test images we determine the Fourier coefficients $\hat{w}_l$ by the above mentioned Gauss quadrature rule. The underlying kernel is given by the bandlimited version of the approximated discrepancy kernel $\tilde{K}_{\mathcal{B}}$ with certain bandwidths; cf. (3.5).

For a comparison of our results with other stippling and dithering methods, we refer to the extensive experiments provided in [55]. Note that, in contrast to [55], our assumption of periodic boundary conditions leads to some boundary artifacts.

FIG. 7. *Halftoning results on the torus* $\mathbb{T}^2$ *for the image in Figure* 1 *with* $M = 30150$ *points. Left: Kernel bandwidth* $N = 650$ *after* $r = 744$ *iterations. Right: Kernel bandwidth* $N = 1300$ *after* $r = 21157$ *iterations.*

Our *first example* deals with the left image of Figure 1, which we aim to approximate by $M = 30150$ points. In this example we stop the CG method if the condition

$$\|\nabla \mathcal{E}_N(\boldsymbol{p}^{(r+1)})\|_2 / \|\nabla \mathcal{E}_N(\boldsymbol{p}^{(0)})\|_2 \leq \varepsilon$$

is fulfilled for some prescribed accuracy $\varepsilon > 0$. We present two halftoning results for $M = 30150$ points, which are shown in Figure 7. For comparison reasons we might consider the optimization procedure as a dynamical system of M points, which converge to a local minimizer. Hence, for an increasing number of iterations the motion of the points decreases and the step length might be a good indicator for being near a local minimizer, where little progress can be made. The left image in Figure 7 shows the result for the kernel with bandwidth $N = 650$, where the prescribed accuracy $\varepsilon = $ 1e-3 is achieved after $r = 744$ iterations. At this stage the step length is $\|\alpha_r \boldsymbol{d}^{(r)}\|_2 \approx$ 1e-3. Thus, the point distribution is far from being a local minimizer of $\mathcal{E}_N$, but it is a good halftoning result, and the computation takes about 15 minutes. In contrast, the right image of Figure 7 shows the result for the kernel with bandwidth $N = 1300$ and prescribed accuracy $\varepsilon = $ 1e-10 obtained after $r = 21157$ iterations, where the step length is $\|\alpha_r \boldsymbol{d}^{(r)}\|_2 \approx$ 1e-8. In that case the point distribution shows more regular "hexagonal" patterns which is besides the low norm of the gradient a further indication to be very close to a local minimizer, where such "regularity artifacts" seem to occur; cf. [55]. However, in that case the computation takes about one day. This underlines the difficulty of finding highly accurate (local) minimizers of such a high dimensional, nonlinear, and nonconvex optimization problem. However, as the former example indicates, quite appealing point distributions are obtained after relatively few iterations, compared to the dimension of the problem.

In the *second example* we consider a Gaussian weight w. Figure 8 depicts our halftoning result for a kernel of bandwidth $N = 1300$ after $r = 20000$ iterations.

Examples on $\mathbb{S}^2$**.** In the following, we will use in our functional $\mathcal{E}_N$ the bandlimited version of the restricted kernel $\Phi(x - y) = -2\sin(\mathrm{d}_{\mathbb{S}^2}(x, y)/2)$, where the
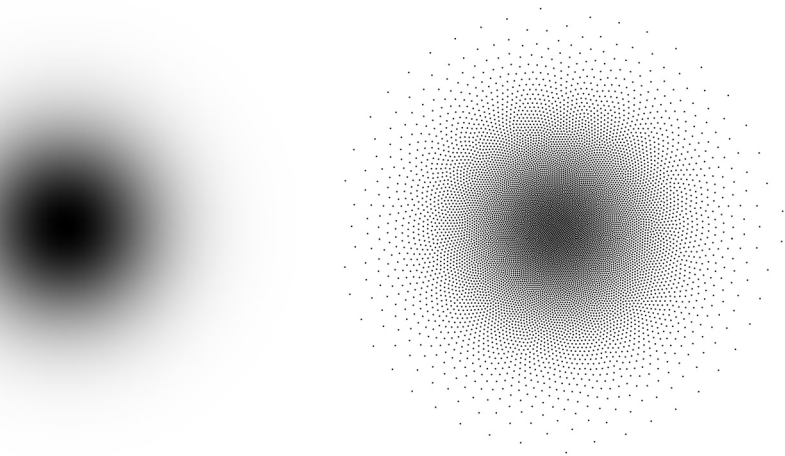
FIG. 8. *Left: Original Gaussian. Right: Halftoning result with $M = 10023$ points and kernel bandwidth $N = 1300$ after $r = 20000$ iterations.*

coefficients (cf. (4.4)) are explicitly given by

$$\lambda_n = \frac{16\pi}{(2n+3)(2n+1)(2n-1)}, \quad n \in \mathbb{N}_0.$$

In the CG algorithm we apply the approximate Hessian of $\tilde{\mathrm{H}}\mathcal{E}_N$; cf. (4.14).

The *first example* uses the topography map of the earth from MATLAB. This map consists of the earth's elevation data. Since the values range from -7473 to 5731, we have scaled it to the range 0 to 1, in order to avoid negative values. The data is sampled on the grid $\boldsymbol{x} := (x\,(\pi i/180, \pi j/180))_{i=1,j=1}^{180,360}$. For this grid we compute nonnegative quadrature weights $\omega_{i,j}$ for a polynomial degree $N = 179$ by the simple CG algorithm proposed in [25]. After applying the quadrature rule (5.1) we obtain a polynomial approximation $w = \sum_{n=0}^{179} \sum_{k=-n}^{n} \hat{w}_n^k Y_n^k$ of the earth's topography; see the left-hand side of Figure 9. We apply our algorithm to $M = 200000$ random points $\boldsymbol{p} \in \mathcal{X}^M$ with a kernel of bandwidth $N = 1000$ and obtain after $r = 3600$ iterations the right image in Figure 9. Here an iteration step takes about 1.5 min.

In our *second example*, we apply our halftoning procedure to three Gaussians on the sphere. More precisely, the weight function is determined by

$$w(x) = \sum_{i=1}^{3} \exp(-5\arccos(x \cdot q_i)^2),$$

where q_i are three orthonormal vectors. Figure 10, right-hand side, shows the result with $M = 5000$ points for a kernel of bandwidth $N = 400$ after $r = 1000$ iterations.

The *final example* is motivated by applications in geoscience. In [23] one is concerned with the problem of solving partial differential equations on the sphere by the method of radial basis functions. There the authors present an algorithm for placing sampling nodes adequately for some given partial differential equation in order to increase the accuracy and stability of the solvers. In a particular test case the nodes are distributed accordingly to the function

$$(5.2) \qquad w(\theta, \varphi) = \begin{cases} \frac{\sqrt{3}}{2\sin\theta}\mathrm{sech}^2(3\sin\theta)\tanh(3\sin\theta), & 0 < \theta < \pi, \\ \frac{3\sqrt{3}}{2} & \text{else} \end{cases}$$

by a method based on electrostatic repulsion. In Figure 11 we see that our method produces similar point distributions as in [23]. Again, we compute the Fourier coefficients $\hat{w}_n^k$, $n = 0, \ldots, 179$, $k = -n, \ldots, n$, as in the previous example. The result for $M = 1849$ points and kernel bandwidth $N = 400$ after $r = 1000$ iterations is presented in Figure 11.
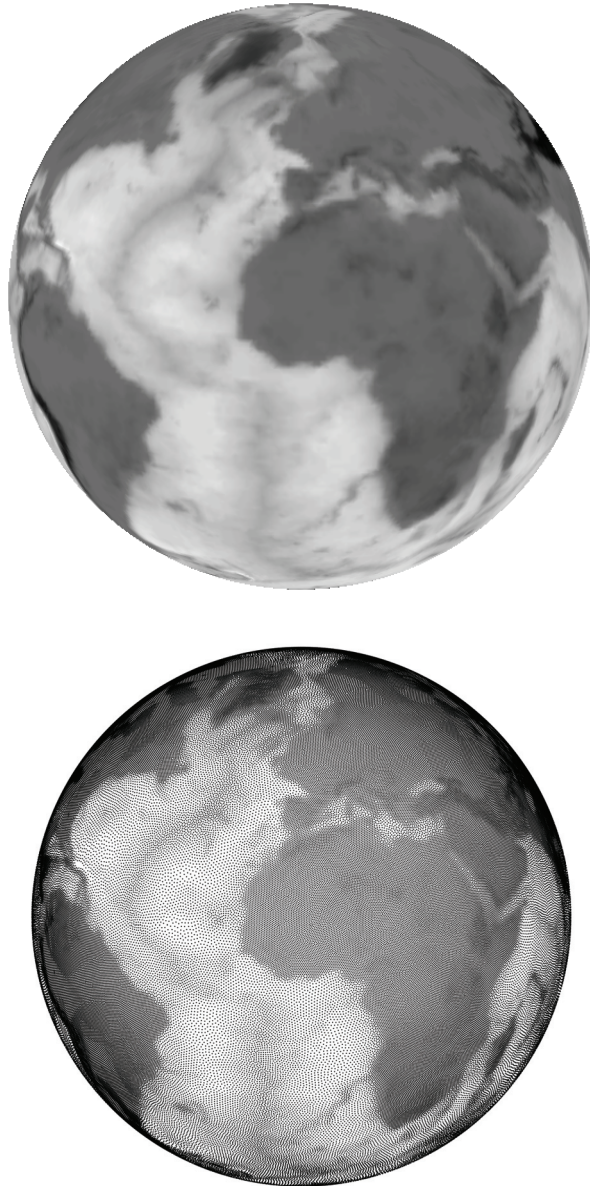

FIG. 9. *Top: Original image of the earth's topology adapted to our setting. Bottom: Halftoning result with with $M = 200000$ points for kernel bandwidth $N = 1000$ after $r = 3600$ iterations.*
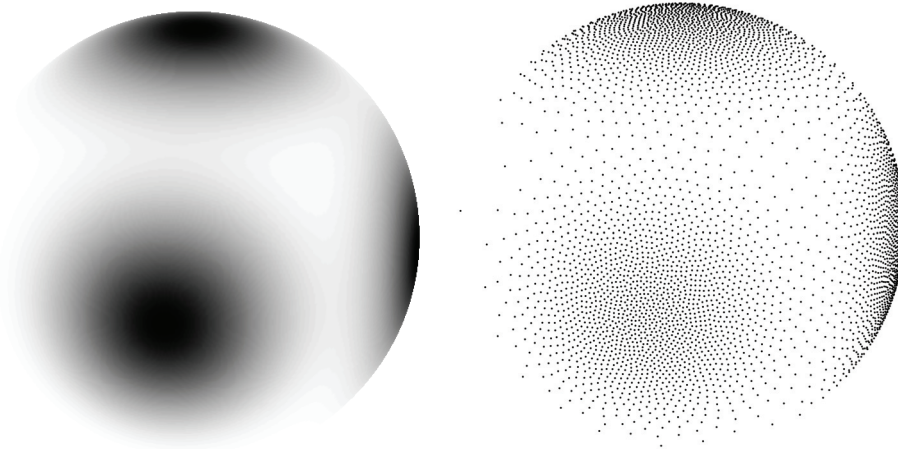
FIG. 10. *Left: Original Gaussians. Right: Halftoning result with $M = 1000$ points for kernel bandwidth $N = 400$ after $r = 1000$ iterations.*
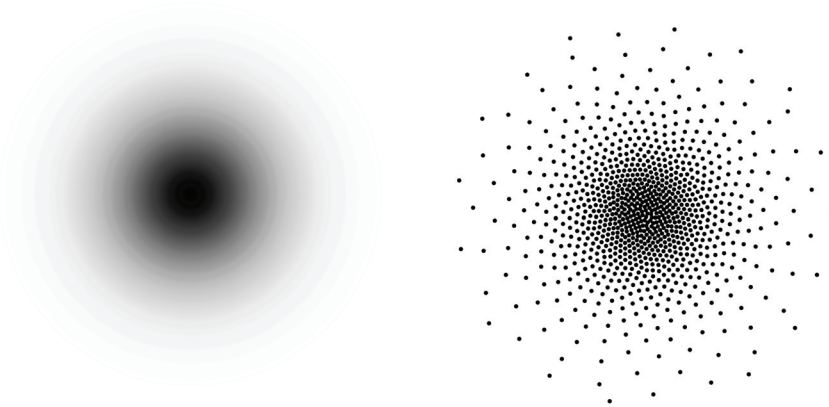

FIG. 11. *Left: Original weight w from* (5.2). *Right: Distribution of $M = 1849$ points on the sphere $\mathbb{S}^2$ according to this weight function by our halftoning algorithm with kernel bandwidth $N = 400$ after $r = 1000$ iterations.*

REFERENCES

[1] M. Analoui and J. P. Allebach, *Model-based halftoning using direct binary search*, Proc. SPIE, 1666 (1992), pp. 96–108.

[2] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc., 68 (1950), pp. 337–404.

[3] F. Aurehammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica, 20 (1998), pp. 61–76.

[4] M. Balzer, T. Schlömer, and O. Deussen, *Capacity-constrained point distributions: A variant of Lloyd's method*, ACM Trans. Graph., 28 (2009), 86.

[5] E. Bannai and E. Bannai, *A survey on spherical designs and algebraic combinatorics on spheres*, European J. Combin., 30 (2009), pp. 1392–1425.

[6] F. A. Baqai and J. P. Allebach, *Halftoning via direct binary search using analytical and stochastic printer models*, IEEE Trans. Image Process., 12 (2003), pp. 1–15.

[7] B. E. Bayer, *An optimum method for two-level rendition of continuous-tone pictures*, in Conference Record, IEEE International Conference on Communications, Volume 1, IEEE, Washington, DC, 1973, pp. 11–15.

[8] G. Beylkin, *On the fast Fourier transform of functions with singularities*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 363–381.

[9] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[10] M. J. Bowick, D. R. Nelson, and A. Travesset, *Interacting topological defects on frozen topographies*, Phys. Rev. B, 62 (2000), pp. 8738–8751.

[11] J. Chang, B. Alain, and B. Ostromoukhov, *Structure-aware error diffusion*, ACM Trans. Graph., 28 (2009), 162.

[12] X. Chen, A. Frommer, and B. Lang, *Computational existence proof for spherical t-designs*, Numer. Math., 117 (2011), pp. 289–305.

[13] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), pp. 10–26.

[14] J. W. Daniel, *Convergence of the conjugate gradient method with computationally convenient modifications*, Numer. Math., 10 (1967), pp. 125–131.

[15] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geometriae Dedicata, 6 (1977), pp. 363–388.

[16] J. R. Driscoll and D. Healy, *Computing Fourier transforms and convolutions on the 2–sphere*, Adv. in Appl. Math., 15 (1994), pp. 202–250.

[17] Q. Du, V. Faber, and M. Gunzburger, *Centroidal Voronoi tessellations: Applications and algorithms*, SIAM Rev., 41 (1999), pp. 637–676.

[18] A. Dutt and V. Rokhlin, *Fast Fourier transforms for nonequispaced data*, SIAM J. Sci. Comput., 14 (1993), pp. 1368–1393.

[19] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 303–353.

[20] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB*, World Scientific, Hackensack, NJ, 2007.

[21] R. Fattal, *Blue-noise sampling using kernel density model*, ACM Trans. Graph.—Proceedings of ACM SIGGRAPH 2011, ACM, New York, 2011, 48.

[22] R. W. Floyd and L. Steinberg, *An adaptive algorithm for spatial grey scale*, Proceedings of the Society of Information Display, 17 (1976), pp. 75–77.

[23] N. Flyer and E. Lehto, *Rotational transport on a sphere: Local node refinement with radial basis functions*, J. Comput. Phys., 229 (2010), pp. 1954–1969.

[24] W. Freeden, T. Gervens, and M. Schreiner, *Constructive Approximation on the Sphere*, Oxford University Press, New York, 1998.

[25] M. Gräf, S. Kunis, and D. Potts, *On the computation of nonnegative quadrature weights on the sphere*, Appl. Comput. Harmon. Anal., 27 (2009), pp. 124–132.

[26] M. Gräf and D. Potts, *On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms*, Numer. Math., 119 (2011), pp. 699–724.

[27] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Lecture Notes in Math. 1730, Springer-Verlag, Berlin, 2000.

[28] P. Gwosdek, C. Schmaltz, J. Weickert, and T. Teuber, *Fast electrostatic halftoning*, J. Real-Time Image Processing, to appear.

[29] W. W. Hager and H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pac. J. Optim., 2 (2006), pp. 35–58.

[30] D. M. Healy, P. J. Kostelec, and D. Rockmore, *Towards safe and effective high-order Legendre transforms with applications to FFTs for the 2-sphere*, Adv. Comput. Math., 21 (2004), pp. 59–105.

[31] P. W. M. Ilbery, *Force field halftoning*, U.S. patent 6124844, 2000.

[32] J. F. Jarvis, C. N. Judice, and W. H. Ninke, *A survey of techniques for the display of continuous tone pictures on bilevel displays*, Comput. Graphics and Image Processing, 5 (1976), pp. 13–40.

[33] J. Keiner, S. Kunis, and D. Potts, *NFFT 3.0, C subroutine library*, http://www.tu-chemnitz.de/~potts/nfft.

[34] J. Keiner, S. Kunis, and D. Potts, *Using NFFT3—a software library for various nonequispaced fast Fourier transforms*, ACM Trans. Math. Software, 36 (2009), 19.

[35] J. Keiner and D. Potts, *Fast evaluation of quadrature formulae on the sphere*, Math. Comput., 77 (2008), pp. 397–419.

[36] T. Kolling and A. Keller, *Efficient illumination by high dynamic range images*, in Proceedings of the 14th Eurographics Workshop on Rendering, ACM International Conference Proceeding Series, Volume 44, ACM, New York, 2003, pp. 45–50.

[37] A. B. J. Kuijlaars and E. B. Saff, *Asymptotics for minimal discrete energy on the sphere*, Trans. Amer. Math. Soc., 350 (1998), pp. 523–538.

[38] S. Kunis and D. Potts, *Fast spherical Fourier algorithms*, J. Comput. Appl. Math., 161 (2003), pp. 75–98.

[39] A. Lagae and P. Dutré, *A comparison of methods for generating poisson disk distributions*, Computer Graphics Forum, 27 (2008), pp. 114–129.

[40] Y. Liu, W. Wang, B. Lévy, F. Sun, D. M. Yan, L. Lu, and C. Yang, *On centroidal Voronoi tessellation—energy smoothness and fast computation*, ACM Trans. Graph., 28 (2009), 101.

[41] S. P. Lloyd, *Least square quantization in PCM*, IEEE Trans. Inform. Theory, 28 (1982), pp. 129–137.

[42] C. Müller, *Spherical Harmonics*, Lecture Notes in Math. 17, Springer-Verlag, Berlin, 1966.

[43] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems, Volume* II: *Standard Information for Functionals*, EMS Tracts Math. 12, EMS, Zürich, 2010.

[44] V. Ostromoukhov, *A simple and efficient error-diffusion algorithm*, in Proceedings of SIGGRAPH 2001, E. Fiume, ed., Computer Graphics Proceedings, Annual Conference Series, Los Angeles, CA, 2001, pp. 567–572.

[45] W.-M. Pang, Y. Qu, T.-T. Wong, D. Cohen-Or, and P.-A. Heng, *Structure-aware halftoning*, ACM Trans. Graph., 27 (2008), 89.

[46] D. Potts, G. Steidl, and M. Tasche, *Fast Fourier transforms for nonequispaced data: A tutorial*, in Modern Sampling Theory: Mathematics and Applications, J. J. Benedetto and P. J. S. G. Ferreira, eds., Birkhäuser Boston, Boston, 2001, pp. 247–270.

[47] W. Purgathofer, R. F. Tobler, and M. Geiler, *Forced random dithering: Improved threshold matrices for ordered dithering*, in Proceedings of the 1st IEEE International Conference on Image Processing, Volume 2, Austin, TX, 1994, pp. 1032–1035.

[48] C. Schmaltz, P. Gwosdek, A. Bruhn, and J. Weickert, *Electrostatic halftoning*, Computer Graphics Forum, 29 (2010), pp. 2313–2327.

[49] A. Secord, *Weighted Voronoi stippling*, in Proceedings of the 2nd International Symposium on Non-Photorealistic Animation and Rendering, ACM, New York, 2002, pp. 37–43.

[50] I. H. Sloan and R. S. Womersley, *A variational characterisation of spherical designs*, J. Approx. Theory, 159 (2009), pp. 308–318.

[51] S. T. Smith, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, Fields Inst. Commun. 3, AMS, Providence, RI, 1994, pp. 113–136.

[52] R. L. Stevenson and G. R. Arce, *Binary display of hexagonally sampled continuous-tone images*, J. Opt. Soc. Amer. A, 2 (1985), pp. 1009–1013.

[53] P. Stucki, *MECCA—A Multiple-error Correcting Computation Algorithm for Bilevel Hardcopy Reproduction*, Technical report RZ1060, IBM Research Lab, Zürich, Switzerland, 1981.

[54] V. Surazhsky, P. Alliez, and C. Gotsman, *Isotropic remeshing of surfaces: A local parameterization approach*, in Proceedings of the 12th International Meshing Roundtable, Santa Fe, NM, 2003, pp. 215–224.

[55] T. Teuber, G. Steidl, P. Gwosdek, C. Schmaltz, and J. Weickert, *Dithering by differences of convex functions*, SIAM J. Imaging Sci., 4 (2011), pp. 79–108.

[56] J. J. Thomson, *On the structure of the atom: An investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure*, Philos. Mag., 7 (1904), pp. 237–265.

[57] C. Udrişte, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Math. Appl. 297, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[58] R. Ulichney, *Digital Halftoning*, MIT Press, London, 1987.

[59] R. Ulichney, *Dithering with blue noise*, Proceedings of the IEEE, 76 (1988), pp. 56–79.

[60] D. Vanderhaeghe and V. Ostromoukhov, *Polyomino-based digital halftoning*, in IADIS International Conference on Computer Graphics and Visualization, P. Isaías, ed., Amsterdam, 2008, pp. 11–18.

[61] D. Varshalovich, A. Moskalev, and V. Khersonskii, *Quantum Theory of Angular Momentum*, World Scientific, Teaneck, NJ, 1988.

[62] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[63] D. J. Wales, H. McKay, and E. L. Altschuler, *Defect motifs for spherical topologies*, Phys. Rev. B, 79 (2009), 224115.

[64] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, Cambridge, UK, 2005.