Preprint ANL/MCS-P2076-0412

ON THE USE OF DISCRETE LAPLACE OPERATOR FOR PRECONDITIONING KERNEL MATRICES

JIE CHEN*

Abstract. This paper presents a preconditioning strategy applied to certain types of kernel matrices that are increasingly ill-conditioned. The ill-conditioning of these matrices is tied to the unbounded variation of the Fourier transform of the kernel function. Hence, the basic idea is to differentiate the kernel in order to suppress the variation. The idea resembles some existing preconditioning methods for Toeplitz matrices, where the theory heavily relies on the underlying fixed generating function. The theory does not apply to the case of a fixed domain with increasingly fine discretizations because the generating function depends on the grid size. For this case, we prove equal distribution results on the spectrum of the resulting matrices. Furthermore, the proposed preconditioning technique also applies to non-Toeplitz matrices, thus eliminating the reliance on a regular grid structure of the points. The preconditioning strategy can be used to accelerate an iterative solver for solving linear systems with respect to kernel matrices.

Key words. Laplace operator, preconditioning, kernel matrix, Toeplitz matrix, stiffness matrix

AMS subject classifications. 65F08, 65C60

1. Introduction.

Matrices generated by kernel functions are widely seen in scientific computing and engineering applications, such as statistical analysis, electronic structure calculations, and solving of integral equations. Such matrices are often dense (essentially full), unless the kernel function has a finite support, and/or sufficiently small values are precluded. Hence, kernel matrices pose significant challenges for solving the respective linear systems. This paper does not discuss direct method approaches (for recent developments of compression-based direct methods, see [22, 4] among others); rather, iterative methods are the focus. The complexity of the former is dimension dependent, whereas the latter enjoys a linear complexity provided that matrix-vector multiplications can be efficiently carried out and the number of iterations grows "very slowly" with the matrix size. In this paper, we are interested in improving the conditioning of the matrix to encourage convergence and reduce iteration counts.

Formally, given a fixed, finite and open domain Ω in $\mathbb{R}^d$, consider the matrix $\Phi \in \mathbb{R}^{n\times n}$ defined with entries

$$\Phi_{ij} = \phi(\boldsymbol{x}_i - \boldsymbol{x}_j)$$

for a set of points $X = \{\boldsymbol{x}_i, i = 1, \ldots, n\} \subset \bar{\Omega}$ and a kernel function $\phi : \mathbb{R}^d \to \mathbb{R}$ which is even, that is, $\phi(-\boldsymbol{x}) = \phi(\boldsymbol{x})$. We are interested in the asymptotics (in particular the condition number κ) of Φ as $n \to \infty$. The scenario of a fixed, finite domain with increasingly dense points is not rare. Examples in practice include solving equations in a domain with increasingly fine discretizations and simulating stochastic processes using increasingly dense sampling.

Suppose the kernel ϕ admits a Fourier transform (the case when ϕ admits only a generalized Fourier transform is discussed later). Denote by $\hat{\phi}$ the transform, that is,

$$\phi(\boldsymbol{x}) = \int_{\mathbb{R}^d} \hat{\phi}(\boldsymbol{\omega}) \exp(\mathbf{i}\, \boldsymbol{\omega}^T \boldsymbol{x})\, d\boldsymbol{\omega}. \tag{1.1}$$

*Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439. Email: jiechen@mcs.anl.gov.

Then, for any vector $\boldsymbol{a}$, the bilinear form $\boldsymbol{a}^T\Phi\boldsymbol{a}$ can be written as

$$\sum_{i,j=1}^{n} a_i a_j \phi(\boldsymbol{x}_i - \boldsymbol{x}_j) = \int_{\mathbb{R}^d} \hat{\phi}(\boldsymbol{\omega}) \left| \sum_{i=1}^{n} a_i \exp(\mathbf{i}\,\boldsymbol{\omega}^T \boldsymbol{x}_i) \right|^2 d\boldsymbol{\omega}. \tag{1.2}$$

Here, the upright boldface letter $\mathbf{i}$ denotes the imaginary unit, and it is to be distinguished with the italic boldface letter $\boldsymbol{i}$ meaning a vector. Since ϕ is even, Φ is symmetric; hence, the condition number $\kappa(\Phi)$ is the ratio between the largest and the smallest absolute eigenvalues of Φ. Further, Φ is positive definite if and only if $\hat{\phi}$ is positive almost everywhere.

The analysis of the integral (1.2) is made easy if one assumes that the point set X forms a regular grid; this assumption is used to illustrate the ill-conditioning of Φ. Without loss of generality we assume that $\bar{\Omega} = [0,1]^d$ and the grid has size $n_1 \times \cdots \times n_d$. We use the vector $\boldsymbol{n}$ to compactly denote the entries $n_1, \ldots, n_d$, and we let $n = n_1 \times \cdots \times n_d$. It is more convenient to index the points by an integer vector, such as $\boldsymbol{j}$, which takes values from $\mathbf{0}$ to $\boldsymbol{n} - \mathbf{1}$. Denote by $\boldsymbol{x} \circ \boldsymbol{y}$ and $\boldsymbol{x}/\boldsymbol{y}$ the elementwise multiplication and division of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Then the Fourier transform (1.1) leads to [16, 15]

$$\phi(\boldsymbol{j}/\boldsymbol{n}) = \int_{[0,2\pi)^d} \hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) \exp(\mathbf{i}\,\boldsymbol{\omega}^T \boldsymbol{j})\, d\boldsymbol{\omega} \tag{1.3}$$

with

$$\hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) \equiv n \sum_{\boldsymbol{l} \in \mathbb{Z}^d} \hat{\phi}\left(\boldsymbol{n} \circ (\boldsymbol{\omega} + 2\pi\boldsymbol{l})\right), \qquad \boldsymbol{\omega} \in [0, 2\pi)^d. \tag{1.4}$$

Then, (1.2) is equivalent to

$$\sum_{\mathbf{0} \le \boldsymbol{i}, \boldsymbol{j} \le \boldsymbol{n}-\mathbf{1}} a_{\boldsymbol{i}} a_{\boldsymbol{j}} \phi(\boldsymbol{i}/\boldsymbol{n} - \boldsymbol{j}/\boldsymbol{n}) = \int_{[0,2\pi)^d} \hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) \left| \sum_{\mathbf{0} \le \boldsymbol{j} \le \boldsymbol{n}-\mathbf{1}} a_{\boldsymbol{j}} \exp(\mathbf{i}\,\boldsymbol{\omega}^T \boldsymbol{j}) \right|^2 d\boldsymbol{\omega}. \tag{1.5}$$

By choosing the vector $\boldsymbol{a}$ with a unit norm, an immediate consequence of (1.5) is that if $|\hat{\phi}_{\boldsymbol{n}}|$ is bounded away from 0 and ∞, then the condition number

$$\kappa(\Phi) \le \frac{\sup |\hat{\phi}_{\boldsymbol{n}}|}{\inf |\hat{\phi}_{\boldsymbol{n}}|}. \tag{1.6}$$

In other words, κ depends on the variation of $\hat{\phi}_{\boldsymbol{n}}$. For example, when ϕ is positive, radially symmetric, and decreasing and when $\boldsymbol{n}$ is sufficiently large, (1.4) implies that the ratio on the right-hand side of (1.6) is in the order $\hat{\phi}(\mathbf{0})/\hat{\phi}(\boldsymbol{n})$. Then Φ is increasingly ill-conditioned even if $\hat{\phi}$ decays at a polynomial rate.

Central to this paper is the manipulation of the transform $\hat{\phi}$ to suppress the growth of the condition number as n increases. The essential idea is to take Laplacians on the kernel ϕ, and this applies to kernels with a Fourier transform that behaves like a power function. Section 2 offers an overview of the theory, not only for the case of the existence of a Fourier transform and a regular grid, but also for the case of generalized Fourier transforms and the case without a regular grid. Detailed analysis is provided in subsequent sections. Even for the case of a regular grid, which brings about a multilevel Toeplitz matrix, one should consider the implication of a fixed

and finite domain that makes existing theory on Toeplitz systems (see, for example, [10, 1]) and preconditioners for Toeplitz systems [2, 3] not immediately applicable. Existing theory (e.g., that of Szegö's) is generally based on a regular grid with fixed spacing but growing size. Thus, the analysis is made convenient by an underlying generating function that is independent of the grid size. On the other hand, when we recast the Fourier transform relation (1.1) into a Fourier series (cf. (1.3))

$$\hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \phi(\boldsymbol{j}/\boldsymbol{n}) \exp(-\mathbf{i}\,\boldsymbol{\omega}^T\boldsymbol{j}), \tag{1.7}$$

one sees that the "generating function" $\hat{\phi}_{\boldsymbol{n}}$ is dependent on $\boldsymbol{n}$. The sequence $\{\hat{\phi}_{\boldsymbol{n}}\}$ does not converge to some limit independent of $\boldsymbol{n}$, which hinders the applicability of existing theory.

Recently, a notion of *locally Toeplitz* [20, 12, 13] is studied. Informally, a locally Toeplitz matrix has slowly varying diagonals, hence it is approximately Toeplitz in a local sense. Often, it arises from a discretization of variable-coefficient differential operators; thus, a (possibly nonconstant) function defined on the domain is associated with the generating function. The theory for locally Toeplitz matrices generalizes Szegö's theory for Toeplitz matrices when characterizing the distributions of eigenvalues and singular values. Furthermore, the class of these matrices is closed under addition, multiplication, and inversion (certain assumptions need be imposed for the latter two operations). Although the theory of locally Toeplitz matrices do not directly apply here, as the "generating functions" do not converge for our setting, the idea of *approximating class of sequences* [11, 14] of matrices underlying the theory is illuminating. Our analysis of the regular grid case coincides with this idea.

A study of (1.1) enables generalizations of the preconditioning technique to non-Toeplitz matrices, when one has a set of scatted points possibly without structures. It is nontrivial to approximate derivatives on points without structures. The major technique considered here is a discretization of the Green's identity, so that second order derivatives are represented as a linear transformation of the original function at the discretized locations. Thus, this work provides a connection between the preconditioning theory and the theory of finite-element methods. It is not surprising to see that the derived linear transformation has a close connection with the stiffness matrix, which occurs when one discretizes an elliptic equation.

The proposed preconditioning technique requires the knowledge of only the decay rate but not the exact expression of $\hat{\phi}$. A convenience brought about by this technique is that one can estimate the decay rate by investigating, in the log scale, the fast Fourier transform of ϕ sampled on a regular grid. Because a Laplace operator acts locally, applying the preconditioner amounts to performing sparse matrix-vector multiplications, which is efficient. Furthermore, an interesting idea is that the sparsity pattern of this preconditioner can be naturally used for defining a sparse approximate inverse preconditioner. The comparison of these two types of preconditioners is out of the scope of this paper, but this idea can be used as a rationale for justifying the use of a neighborhood graph (or powers of the graph) to define the sparsity pattern for an approximate inverse preconditioner.

2. Laplacian preconditioning theory: overview. The spectrum of Φ is closely connected with the Fourier transform $\hat{\phi}$. Consider the regular grid case. By

discretizing the region $[0, 2\pi)^d$, the right-hand side of (1.5) is approximated by

$$\frac{(2\pi)^d}{n} \sum_{\mathbf{0} \le \boldsymbol{k} \le \boldsymbol{n}-\mathbf{1}} \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n}) \left| \sum_{\mathbf{0} \le \boldsymbol{j} \le \boldsymbol{n}-\mathbf{1}} a_{\boldsymbol{j}} \exp(\mathbf{i}\,(2\pi \boldsymbol{k}/\boldsymbol{n})^T \boldsymbol{j}) \right|^2.$$

This, in fact, is a bilinear form $\boldsymbol{a}^T \hat{\Phi} \boldsymbol{a}$, where $\hat{\Phi}$ is defined as $U^H \Lambda U$, with $U \in \mathbb{C}^{n \times n}$ being unitary, $\Lambda \in \mathbb{R}^{n \times n}$ being diagonal, and

$$U_{\boldsymbol{kj}} = \exp(\mathbf{i}\,(2\pi \boldsymbol{k}/\boldsymbol{n})^T \boldsymbol{j})/\sqrt{n}, \qquad \Lambda_{\boldsymbol{kk}} = (2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n}).$$

Naturally, one would expect that the spectrum of Φ is in some sense similar to that of $\hat{\Phi}$, that is, the set $\{(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n})\}$. In Section 3, we prove that they are *equally distributed* after a concurrent scaling by n. The definition of equal distribution resembles the determination of the equivalence of two random variables by equating all their moments. Therefore, with large samples, the histograms are sufficiently close. In our setting, this means that when $\boldsymbol{n}$ is large, the shape of the discrete surface of the eigenvalues (arranged in some way) looks almost the same as that of the surface $(2\pi)^d \hat{\phi}_{\boldsymbol{n}}$ in $[0, 2\pi)^d$. Then, we have

$$\kappa(\Phi) \approx \frac{\max |\hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n})|}{\min |\hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n})|}, \tag{2.1}$$

which is consistent with (1.6).

The result of equal distributions can be used in several ways. One way is to differentiate ϕ such that $\hat{\phi}$ is correspondingly modified. Taking the Laplacian of both sides of (1.1) s times gives

$$\Delta^s \phi(\boldsymbol{x}) = \int_{\mathbb{R}^d} \|\boldsymbol{\omega}\|^{2s} \hat{\phi}(\boldsymbol{\omega}) \exp(\mathbf{i}\,\boldsymbol{\omega}^T \boldsymbol{x})\, d\boldsymbol{\omega}. \tag{2.2}$$

For example, consider the Matérn kernel [16, 6, 21] whose Fourier transform

$$\hat{\phi}(\boldsymbol{\omega}) \asymp (1 + \|\boldsymbol{\omega}\|)^{-4\tau}, \qquad \tau > 0. \tag{2.3}$$

Then $\|\boldsymbol{\omega}\|^{2s} \hat{\phi} \asymp \|\boldsymbol{\omega}\|^{2s} (1 + \|\boldsymbol{\omega}\|)^{-4\tau}$, which decays slower than $\hat{\phi}$. This immediately reduces the growth of the condition number of Φ: from the approximation (2.1), $\kappa(\Phi) \approx O\left(\|\boldsymbol{n}\|^{4\tau}\right) = O\left(n^{4\tau/d}\right)$. After taking the Laplacian Δ^s, the growth is reduced to $O(n^{(4\tau-2s)/d})$.

For preconditioning, the task is to obtain a new matrix that approximates the kernel matrix defined by $\Delta^s \phi$ via linear transformations of Φ. Section 4 considers the regular grid case, where Δ is naturally approximated by second-order finite difference, denoted by D. We show that the spectrum of the new matrix by applying D^s and the set $\{(2\pi)^d \hat{\phi}^{[s]}_{\boldsymbol{n}}(2\pi \boldsymbol{k}/\boldsymbol{n})\}$ are equally distributed, where $\hat{\phi}^{[s]}_{\boldsymbol{n}}$ corresponds to $\mathrm{D}^s\,\phi$ just as $\hat{\phi}_{\boldsymbol{n}}$ corresponds to ϕ. Strictly speaking, s has to be strictly less than $2\tau - d/2$. Otherwise, (2.2) does not hold because $\|\boldsymbol{\omega}\|^{2s} \hat{\phi}$ is not integrable, and a presumption in the analysis in Section 4 is also not satisfied. Empirically, however, when s passes $2\tau - d/2$, the growth of the condition number of the new matrix still follows the rate $O(n^{(4\tau-2s)/d})$. This probably requires a different analysis argument. For example, when τ is an integer and when $s/2 = \tau$, Stein et. al. [18] prove that the condition number of the new matrix is $O(1)$ (that is, independent of n), by directly upper and

lower bounding $\hat{\phi}_{\boldsymbol{n}}^{[s]}$. As a rule of thumb, for any $\tau > 0$, one chooses s to be the even integer closest to $\tau/2$ to obtain the best preconditioning result.

The second way is to consider some ϕ that does not admit a Fourier transform, but rather, a generalized Fourier transform. To avoid the technicalities of generalized functions [8], we consider restricting the validity of (1.2) to some subspace instead. For example, consider the power function

$$\phi(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{\alpha}, & \alpha/2 \notin \mathbb{N} \\ \|\boldsymbol{x}\|^{\alpha} \log \|\boldsymbol{x}\|, & \alpha/2 \in \mathbb{N} \end{cases}$$

for some $\alpha > 0$. Standard theory on conditional positive-definite functions [19] shows that (1.2) holds with $\hat{\phi}(\boldsymbol{\omega}) = c\|\boldsymbol{\omega}\|^{-\alpha-d}$ for all sets of $\{a_i\}$ that satisfy

$$\sum_{i=1}^{n} a_i P(\boldsymbol{x}_i) = 0, \tag{2.4}$$

where P is any polynomial of degree at most $t = \lfloor \alpha/2 \rfloor$ and where c is some constant independent of $\boldsymbol{\omega}$ (see [6, 17]). The function $\hat{\phi}$ is known as the generalized Fourier transform of ϕ, and the vectors $\boldsymbol{a}$ form a subspace. Section 5 gives some results useful for supporting the applicability of Laplacian preconditioning in this case. In particular, every vector in the range of the discrete Laplace operator D satisfies (2.4) for P up to degree 1, and recursively applying the operator yields vectors that satisfy (2.4) for higher-order polynomials. In other words, restricting the validity of (1.2) to some subspace does not impose additional constraints compared with the case when ϕ admits a Fourier transform $\hat{\phi}$. Thus, the preconditioning strategy is also applicable here. In particular, since $\hat{\phi}$ in this case is asymptotically the same as the one in (2.3) for $\boldsymbol{\omega}$ away from the origin, we similarly apply the discrete operator D^s to suppress the growth of the condition number. In the ideal case, when $2s$ matches $\alpha + d$ (the exponent of $\|\boldsymbol{\omega}\|$ in $\hat{\phi}$), in the generalized function sense the left-hand side of (2.2) is the delta function, which in the discrete sense results in the identity matrix. Then, letting s to be the even integer closest to $(\alpha + d)/2$ gives the best preconditioning result.

The third way to fully exploit the Laplacian preconditioning idea is to eliminate the reliance on a regular grid. It is nontrivial to discretize the Laplace operator on a set of scattered points without a regular grid structure. In Section 6, with a reasonable assumption that a finite-element mesh of the points is available, we construct a discrete Laplace operator based on a discretization of the Green's identity. The resulting operator has a close connection with the stiffness matrix in finite-element analysis. We show that the discretization error decreases linearly with the size of the finite elements. Thus, in the limit, the discrete Laplace operator applied to the kernel matrix is equivalent to the Laplacian applied to the kernel function, hence reducing the condition number of the kernel matrix in a manner similar to the case of a regular grid.

The practical performance of the preconditioning strategies discussed above is shown by using two examples, the Matérn and the power-law kernel, in Section 7. We note that the applicability of the proposed strategies is not restricted to these two kernels. As long as the Fourier transform of the kernel exhibits a decay, one can apply the Laplacian a few times, where the number of times matches the decay rate divided by 2, to precondition the kernel matrix. This number has to be an even

integer because the discrete Laplacian matrix needs to be applied to both sides of the matrix simultaneously to preserve symmetry. This will be more clear when one reads (4.2).

3. Spectrum of Toeplitz matrix in a fixed and finite domain. In the regular grid case, the kernel matrix Φ is multilevel Toeplitz. There exist rich results about the asymptotics of Toeplitz matrices, one of the most well known of which is Szegö's theorem [10], which implies that the distribution of the eigenvalues approaches the generating function. However, such results are based on the presumption of a fixed sequence of values that define the matrix. These values are the Fourier coefficients of the generating function, and they are independent of the grid size. In fixed domain asymptotics, on the other hand, the sequence of values that define the matrix, $\{\phi(\boldsymbol{j}/\boldsymbol{n})\}_{\boldsymbol{j}}$, is clearly dependent on $\boldsymbol{n}$. To handle this situation, we use the asymptotic equivalence of matrices to show equal distribution results. To emphasize the dependence on $\boldsymbol{n}$, we use $\Phi_{\boldsymbol{n}}$ to denote the kernel matrix.

DEFINITION 3.1. *Two sets of real numbers* $\{a_j^{(n)}\}_{j=1,\dots,n}$ *and* $\{b_j^{(n)}\}_{j=1,\dots,n}$ *are* equally distributed *in the interval* $[M_1, M_2]$ *if for any continuous function* $F : [M_1, M_2] \to \mathbb{R}$*,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} [F(a_j^{(n)}) - F(b_j^{(n)})] = 0.$$

DEFINITION 3.2. *Two sequences of Hermitian matrices* $\{A_n\}$ *and* $\{B_n\}$ *are* asymptotically equivalent *(denoted as* $A_n \sim B_n$*) if*

(i) A_n *and* B_n *are both uniformly bounded in* 2*-norm, that is, when* n *is sufficiently large,* $\|A_n\|_2, \|B_n\|_2 \le M < \infty$ *for some* M *independent of* n*, and*

(ii) $\lim_{n\to\infty} \|A_n - B_n\|_F/\sqrt{n} = 0$.

We use $\lambda(\cdot)$ to denote an eigenvalue. The following result is proved in [9].

LEMMA 3.3. *If* $\{A_n\}$ *and* $\{B_n\}$ *are asymptotically equivalent, then* $\{\lambda_j(A_n)\}$ *and* $\{\lambda_j(B_n)\}$ *are equally distributed.*

Remark. To emphasize the regular grid structure, the lemma still holds when the indices n and j are replaced by $\boldsymbol{n}$ and $\boldsymbol{j}$, respectively.

Remark. The asymptotic equivalence in Definition 3.2 is based on a Hermitian assumption, as it is sufficient in this paper for ϕ being real-valued. To relax the restriction, one may consider the notion of p-equivalence, which concerns singular values. For this, the Schatten p-norm is used. Specifically, denoting by $\sigma(\cdot)$ a singular value, the Schatten p-norm $\|\cdot\|_{S,p}$ of a matrix A for $p \in [1, \infty]$ is defined as

$$\|A\|_{S,p} = \left(\sum_j \sigma_j(A)^p \right)^{1/p}.$$

Connecting to the usual notation of an induced p-norm, one sees that $\|A\|_{S,\infty} = \|A\|_2$, $\|A\|_{S,2} = \|A\|_F$, and $\|A\|_{S,1}$ is the trace norm of A. Two sequences of matrices $\{A_n\}$ and $\{B_n\}$ are said to be *p-equivalent* (denoted as $A_n \sim_p B_n$) if $\lim_{n\to\infty} \|A_n - B_n\|_{S,p}/n^{1/p} = 0$. Thus, Definition 3.2 (when A_n and B_n are positive definite) is a special case of 2-equivalence for Hermitian matrices. Furthermore, unlike Definition 3.2, the definition of p-equivalence does not require the uniform boundedness of the norms of A_n and B_n. The p-equivalence belongs to a broader sense of matrix closeness, namely *approximating class of sequences* (a.c.s.). We say

that $\{(B_m)_n\}$ is a.c.s. of $\{A_n\}$ if for all sufficiently large m, there exists $n(m)$ such that for all $n > n(m)$, $A_n - (B_m)_n$ is a low-rank plus small-norm matrix, where the rank divided by the matrix size, and the Schatten ∞-norm, are independent of n and tend to zero as $m \to \infty$. Thus, if $\{(B_m)_n\}$ and $\{A_n\}$ are p-equivalent, then $\{(B_m)_n\}$ is a.c.s. of $\{A_n\}$. The notion of a.c.s. can be used in the characterization of singular value distribution: If the singular values of $(B_m)_n$ are equally distributed with an underlying generating function f_m regularly sampled, and if f_m converges in measure to f, then the singular values of A_n are equally distributed with f regularly sampled [11, 14]. The theory presented in this section, on the other hand, does not require that f_m converges to some f.

Base on Lemma 3.3, the following is the major result of this section.

THEOREM 3.4. *Let $\phi \in L^1 \cap L^2$. Then the set of eigenvalues of $\Phi_{\boldsymbol{n}}/n$ and the set $\{(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})/n\}$ are equally distributed, with $\hat{\phi}_{\boldsymbol{n}}$ defined in* (1.4).

Proof. We repeat the Fourier series (1.7) with (1.3) here, by changing the notation $\phi(\boldsymbol{j}/\boldsymbol{n})$ to $\phi_{\boldsymbol{j}}^{(\boldsymbol{n})}$ for clarity:

$$\hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \phi_{\boldsymbol{j}}^{(\boldsymbol{n})} \exp(-\mathbf{i}\, \boldsymbol{\omega}^T \boldsymbol{j}) \quad \text{with} \quad \phi_{\boldsymbol{j}}^{(\boldsymbol{n})} = \int_{[0,2\pi)^d} \hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}) \exp(\mathbf{i}\, \boldsymbol{\omega}^T \boldsymbol{j})\, d\boldsymbol{\omega}. \tag{3.1}$$

Then the matrix $\Phi_{\boldsymbol{n}}$ is defined by the values $\{\phi_{\boldsymbol{j}}^{(\boldsymbol{n})}\}$. We shall construct circulant matrices $\tilde{C}_{\boldsymbol{n}}$ and $C_{\boldsymbol{n}}$ such that the eigenvalues of $C_{\boldsymbol{n}}$ are $(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})$ and that $\Phi_{\boldsymbol{n}}/n \sim \tilde{C}_{\boldsymbol{n}}/n \sim C_{\boldsymbol{n}}/n$. Then the theorem holds.

First, let the $(\boldsymbol{j}, \boldsymbol{k})$ entry of $\tilde{C}_{\boldsymbol{n}}$ be defined as $\tilde{c}_{\boldsymbol{j}-\boldsymbol{k}}^{(\boldsymbol{n})}$ with $\tilde{c}_{\boldsymbol{j}}^{(\boldsymbol{n})} := \sum_{\boldsymbol{k} \in S_{\boldsymbol{j}}} \phi_{\boldsymbol{j}+\boldsymbol{k}\circ\boldsymbol{n}}^{(\boldsymbol{n})}$, where the index set $S_{\boldsymbol{j}}$ consists of vectors whose entries are chosen from the following rule:

$$k_\ell \in \begin{cases} \{0\}, & j_\ell = 0 \\ \{0, -1\}, & j_\ell > 0 \\ \{0, 1\}, & j_\ell < 0. \end{cases}$$

Further, let the $(\boldsymbol{j}, \boldsymbol{k})$ entry of $C_{\boldsymbol{n}}$ be defined as $c_{\boldsymbol{j}-\boldsymbol{k}}^{(\boldsymbol{n})}$ with $c_{\boldsymbol{j}}^{(\boldsymbol{n})} := \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \phi_{\boldsymbol{j}+\boldsymbol{k}\circ\boldsymbol{n}}^{(\boldsymbol{n})}$. Using the Fourier series (3.1), one then sees that the eigenvalues of $C_{\boldsymbol{n}}$ are $(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})$.

Next, we prove that the 2-norms of $\Phi_{\boldsymbol{n}}/n$, $\tilde{C}_{\boldsymbol{n}}/n$ and $C_{\boldsymbol{n}}/n$ are uniformly bounded. For any $\boldsymbol{\omega}$, there exists a constant M such that

$$\frac{1}{n} |\hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega})| \le \frac{1}{(2\pi)^d} \cdot \frac{1}{n} \sum_{\boldsymbol{j} \in \mathbb{Z}^d} |\phi_{\boldsymbol{j}}^{(\boldsymbol{n})}| \le M,$$

when each component of $\boldsymbol{n}$ is sufficiently large, because the kernel $\phi(\boldsymbol{x})$ is absolutely integrable. Then the 2-norm of $\Phi_{\boldsymbol{n}}/n$ is bounded by M by using (1.5), and the 2-norms of $\tilde{C}_{\boldsymbol{n}}/n$ and $C_{\boldsymbol{n}}/n$ are bounded by M by using (3.1).

Last, we show that $\|\Phi_{\boldsymbol{n}}/n - \tilde{C}_{\boldsymbol{n}}/n\|_F^2/n \to 0$ and $\|\tilde{C}_{\boldsymbol{n}}/n - C_{\boldsymbol{n}}/n\|_F^2/n \to 0$ as $\boldsymbol{n}$ approaches infinity. By definition,

$$\|\Phi_{\boldsymbol{n}} - \tilde{C}_{\boldsymbol{n}}\|_F^2 = \sum_{\boldsymbol{j}=-\boldsymbol{n}+\boldsymbol{1}}^{\boldsymbol{n}-\boldsymbol{1}} \left[\prod_{\ell=1}^{d} (n_\ell - |j_\ell|) \right] \left| \sum_{\boldsymbol{k} \in S_{\boldsymbol{j}} \setminus \{\boldsymbol{0}\}} \phi_{\boldsymbol{j}+\boldsymbol{k}\circ\boldsymbol{n}}^{(\boldsymbol{n})} \right|^2.$$

Then, $\|\Phi_{\boldsymbol{n}}/n - \tilde{C}_{\boldsymbol{n}}/n\|_F^2/n$ is upper bounded by

$$\frac{1}{n^2}\sum_{\boldsymbol{j}=-\boldsymbol{n}+\boldsymbol{1}}^{\boldsymbol{n}-\boldsymbol{1}}\left|\sum_{\boldsymbol{k}\in S_j\setminus\{\boldsymbol{0}\}}\phi^{(\boldsymbol{n})}_{\boldsymbol{j}+\boldsymbol{k}\circ\boldsymbol{n}}\right|^2 \le \frac{1}{n^2}\sum_{\boldsymbol{j}=-\boldsymbol{n}+\boldsymbol{1}}^{\boldsymbol{n}-\boldsymbol{1}}(2^d-1)\sum_{\boldsymbol{k}\in S_j\setminus\{\boldsymbol{0}\}}|\phi^{(\boldsymbol{n})}_{\boldsymbol{j}+\boldsymbol{k}\circ\boldsymbol{n}}|^2 \le \frac{(2^d-1)^2}{n^2}\sum_{\boldsymbol{j}=-\boldsymbol{n}+\boldsymbol{1}}^{\boldsymbol{n}-\boldsymbol{1}}|\phi^{(\boldsymbol{n})}_{\boldsymbol{j}}|^2.$$

On the other hand, $\|\tilde{C}_{\boldsymbol{n}}/n - C_{\boldsymbol{n}}/n\|_F^2/n$ is upper bounded by

$$\frac{1}{n^3}\sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{n}-\boldsymbol{1}}|\lambda_{\boldsymbol{j}}(\tilde{C}_{\boldsymbol{n}})-\lambda_{\boldsymbol{j}}(C_{\boldsymbol{n}})|^2 = \frac{1}{n^3}\sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{n}-\boldsymbol{1}}\left|\sum_{|\boldsymbol{k}|\ge\boldsymbol{n}}\phi^{(\boldsymbol{n})}_{\boldsymbol{k}}\exp(-\mathbf{i}\,(2\pi\boldsymbol{k}/\boldsymbol{n})^T\boldsymbol{j})\right|^2 = \frac{1}{n^2}\sum_{|\boldsymbol{k}|\ge\boldsymbol{n}}|\phi^{(\boldsymbol{n})}_{\boldsymbol{k}}|^2.$$

Because the kernel $\phi(\boldsymbol{x})$ is square integrable, the rightmost terms in the above two formulas vanish when $\boldsymbol{n}$ approaches infinity. □

4. Discrete Laplace operator on a regular grid. Consider a matrix L that has $(n_1-2)\times\cdots\times(n_d-2)$ rows and $n_1\times\cdots\times n_d$ columns. For row indices $\boldsymbol{i}=\boldsymbol{1},\ldots,\boldsymbol{n}-\boldsymbol{2}$ and column indices $\boldsymbol{j}=\boldsymbol{0},\ldots,\boldsymbol{n}-\boldsymbol{1}$, the entries are defined as

$$L_{\boldsymbol{ij}} = \begin{cases} n_p^2 & \text{if } \boldsymbol{j}=\boldsymbol{i}\pm\boldsymbol{e}_p \text{ for } p=1,\ldots,d \\ -2\sum_{p=1}^d n_p^2 & \text{if } \boldsymbol{j}=\boldsymbol{i} \\ 0 & \text{otherwise.}\end{cases} \tag{4.1}$$

An example of L for a 6×4 grid is the following, where $a=-104$, $b=36$ and $c=16$:

$$\left[\begin{array}{cccc|cccccc|cccccc|cccc}
c & & & & b & a & b & & & & c & & & & & & & & & \\
 & c & & & & b & a & b & & & & c & & & & & & & & \\
 & & c & & & & b & a & b & & & & c & & & & & & & \\
 & & & c & & & & b & a & b & & & & c & & & & & & \\
\hline
 & & & & c & & & & & & b & a & b & & & & c & & & \\
 & & & & & c & & & & & & b & a & b & & & & c & & \\
 & & & & & & c & & & & & & b & a & b & & & & c & \\
 & & & & & & & c & & & & & & b & a & b & & & & c
\end{array}\right].$$

Define, and denote by D, the discrete Laplace operator on a grid for an arbitrary function f:

$$\mathrm{D}\,f\left(\frac{\boldsymbol{k}}{\boldsymbol{n}}\right) := \sum_{p=1}^d n_p^2\left[f\left(\frac{\boldsymbol{k}+\boldsymbol{e}_p}{\boldsymbol{n}}\right)+f\left(\frac{\boldsymbol{k}-\boldsymbol{e}_p}{\boldsymbol{n}}\right)-2f\left(\frac{\boldsymbol{k}}{\boldsymbol{n}}\right)\right].$$

Clearly, L is the matrix form of D for a finite grid. One can then easily verify that the matrix $L\Phi L^T$ is multilevel Toeplitz and has entries $(L\Phi L^T)_{\boldsymbol{ij}} = \mathrm{D}^2\,\phi((\boldsymbol{i}-\boldsymbol{j})/\boldsymbol{n})$ for $\boldsymbol{i},\boldsymbol{j}=\boldsymbol{1},\ldots,\boldsymbol{n}-\boldsymbol{2}$.

To generalize these observations, let $L^{[s]}$, $s>0$, be a matrix with $(n_1-2s)\times\cdots\times(n_d-2s)$ rows and $(n_1-2s+2)\times\cdots\times(n_d-2s+2)$ columns. Its entries are defined in (4.1), with row indices $\boldsymbol{i}=\boldsymbol{s},\ldots,\boldsymbol{n}-\boldsymbol{s}-\boldsymbol{1}$ and column indices $\boldsymbol{j}=\boldsymbol{s}-\boldsymbol{1},\ldots,\boldsymbol{n}-\boldsymbol{s}$, where $\boldsymbol{s}$ is the vector with all entries being s. Define the matrix

$$\Phi^{[2s]} = L^{[s]}\cdots L^{[1]}\Phi {L^{[1]}}^T\cdots {L^{[s]}}^T \tag{4.2}$$

for $s = 1, 2, \ldots$. In addition, let $\Phi^{[0]} \equiv \Phi$. Then by induction the $(\boldsymbol{i}, \boldsymbol{j})$ entry of $\Phi^{[2s]}$ is $\mathrm{D}^{2s}\,\phi((\boldsymbol{i}-\boldsymbol{j})/\boldsymbol{n})$. We further extend the definition of $\Phi^{[\cdot]}$ as

$$\Phi^{[s]}(\boldsymbol{i}, \boldsymbol{j}) := \mathrm{D}^s\,\phi((\boldsymbol{i}-\boldsymbol{j})/\boldsymbol{n}) \tag{4.3}$$

for all integers $s \geq 0$. This definition is consistent with (4.2) when the number in the superscript of Φ is even. Then applying the definition of D to (1.3) we obtain

$$\mathrm{D}^s\,\phi(\boldsymbol{k}/\boldsymbol{n}) = \int_{[0,2\pi)^d} \hat{\phi}^{[s]}_{\boldsymbol{n}}(\boldsymbol{\omega}) \exp(\mathbf{i}\,\boldsymbol{\omega}^T\boldsymbol{k})\,d\boldsymbol{\omega}, \tag{4.4}$$

where

$$\hat{\phi}^{[s]}_{\boldsymbol{n}}(\boldsymbol{\omega}) \equiv \left[-4\sum_{p=1}^{d} n_p^2 \sin^2\left(\frac{\omega_p}{2}\right)\right]^s \hat{\phi}_{\boldsymbol{n}}(\boldsymbol{\omega}). \tag{4.5}$$

The following is the major result of this section.

THEOREM 4.1. *If all the partial derivatives of ϕ of order up to $2s+1$ belong to $L^1 \cap L^2$, then the set of eigenvalues of $\Phi^{[s]}_{\boldsymbol{n}}/n$ and the set $\{(2\pi)^d \hat{\phi}^{[s]}_{\boldsymbol{n}}(2\pi\boldsymbol{k}/\boldsymbol{n})/n\}$ are equally distributed, with $\hat{\phi}^{[s]}_{\boldsymbol{n}}$ defined in* (4.5).

Proof. Following a similar argument to the proof of Theorem 3.4, it suffices to show that both

$$\frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^s\,\phi(\boldsymbol{k}/\boldsymbol{n})| \qquad \text{and} \qquad \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^s\,\phi(\boldsymbol{k}/\boldsymbol{n})|^2$$

are finite as $\boldsymbol{n}$ approaches infinity; then the theorem holds. In turn, it suffices to show that for all $1 \leq j \leq s$,

$$\lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^j\,\phi(\boldsymbol{k}/\boldsymbol{n}) - \Delta^j\phi(\boldsymbol{k}/\boldsymbol{n})| = 0, \tag{4.6}$$

since when taking $j = s$ this implies

$$\lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^s\,\phi(\boldsymbol{k}/\boldsymbol{n})| = \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\Delta^s\phi(\boldsymbol{k}/\boldsymbol{n})| = \int |\Delta^s\phi| < \infty$$

and

$$\lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^s\,\phi(\boldsymbol{k}/\boldsymbol{n})|^2 = \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n}\sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\Delta^s\phi(\boldsymbol{k}/\boldsymbol{n})|^2 = \int |\Delta^s\phi|^2 < \infty.$$

We show (4.6) by induction.

Observe that for any three times differentiable function f, the Taylor expansion at $\boldsymbol{k}/\boldsymbol{n}$ with a remainder term gives

$$\mathrm{D}\,f(\boldsymbol{k}/\boldsymbol{n}) - \Delta f(\boldsymbol{k}/\boldsymbol{n}) = \sum_{p=1}^{d}\left[\frac{\partial_p^3 f((\boldsymbol{k}+\xi_p^1\boldsymbol{e}_p)/\boldsymbol{n})}{6n_p} - \frac{\partial_p^3 f((\boldsymbol{k}-\xi_p^2\boldsymbol{e}_p)/\boldsymbol{n})}{6n_p}\right],$$

where $\xi_p^1, \xi_p^2 \in (0,1)$ are both dependent on $\boldsymbol{k}$. Then

$$\begin{aligned}
&\lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}\, f(\boldsymbol{k}/\boldsymbol{n}) - \Delta f(\boldsymbol{k}/\boldsymbol{n})| \\
&\le \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \sum_{p=1}^{d} \frac{1}{6n_p} \Big(|\partial_p^4 f((\boldsymbol{k}+\xi_p^1 \boldsymbol{e}_n)/\boldsymbol{n})| + |\partial_p^4 f((\boldsymbol{k}+\xi_p^2 \boldsymbol{e}_n)/\boldsymbol{n})| \Big) \\
&= \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \sum_{p=1}^{d} \frac{1}{6n_p} \left[\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{n} |\partial_p^3 f((\boldsymbol{k}+\xi_p^1 \boldsymbol{e}_n)/\boldsymbol{n})| + \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \frac{1}{n} |\partial_p^3 f((\boldsymbol{k}+\xi_p^2 \boldsymbol{e}_n)/\boldsymbol{n})| \right] = 0.
\end{aligned} \tag{4.7}$$

The last equality is zero because the limit of the term inside the square bracket as $\boldsymbol{n} \to \boldsymbol{\infty}$ is two times the integral of $|\partial_p^3 f|$, which is finite. This shows the induction basis of (4.6), $j = 1$. Furthermore, for any $1 < j \le s$,

$$A_j := \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}^j\, \phi(\boldsymbol{k}/\boldsymbol{n}) - \Delta^j \phi(\boldsymbol{k}/\boldsymbol{n})| \le \mathrm{D}(A_{j-1}) + B_j$$

with

$$B_j := \lim_{\boldsymbol{n}\to\boldsymbol{\infty}} \frac{1}{n} \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |\,\mathrm{D}\, \Delta^{j-1} \phi(\boldsymbol{k}/\boldsymbol{n}) - \Delta \Delta^{j-1} \phi(\boldsymbol{k}/\boldsymbol{n})|.$$

Based on the induction assumption $A_{j-1} = 0$, we have $\mathrm{D}(A_{j-1}) = 0$. From (4.7), $B_j = 0$. Hence, $A_j = 0$, completing the induction. □

5. Bilinear form in a subspace. The continuous operator Δ^s maps any polynomial $P(\boldsymbol{x})$ of order at most $2s-1$ to zero. In parallel, this section shows that the discrete operator D^s maps the gridded signal $P(\boldsymbol{k}/\boldsymbol{n})$ to a zero signal. This result then implies that for any vector $\boldsymbol{p}$ with entries $\boldsymbol{p}_{\boldsymbol{k}} = P(\boldsymbol{k}/\boldsymbol{n})$, $L^{[s]} \cdots L^{[1]} \boldsymbol{p} = \boldsymbol{0}$. Therefore, for any $\boldsymbol{u}$, the vector

$$\boldsymbol{a} = {L^{[1]}}^T \cdots {L^{[s]}}^T \boldsymbol{u}$$

satisfies $\boldsymbol{a}^T \boldsymbol{p} = 0$. Such vectors $\boldsymbol{a}$ form a subspace. Then the bilinear form

$$\boldsymbol{a}^T \Phi \boldsymbol{a} = \boldsymbol{u}^T \Phi_{\boldsymbol{n}}^{[2s]} \boldsymbol{u}$$

is always bounded if $\hat{\phi}_{\boldsymbol{n}}^{[2s]}$ is bounded, even though the norm of the original matrix $\Phi_{\boldsymbol{n}}$ may not be.

One may wish some equal distribution results for $\Phi_{\boldsymbol{n}}^{[2s]}$ as in the preceding section. However, a difficulty is that when ϕ does not admit a Fourier transform, the essential ingredient in the proofs of Theorems 3.4 and 4.1—the Fourier series (3.1)—cannot be established. Nevertheless, the growth of the condition number of $\Phi_{\boldsymbol{n}}^{[2s]}$ does reduce empirically when s increases. The analysis probably needs a different argument, and it is not given in this paper.

THEOREM 5.1. *For any polynomial P of order at most $2s-1$ and all integer vectors $\boldsymbol{k}$ and $\boldsymbol{n}$, $\mathrm{D}^s\, P(\boldsymbol{k}/\boldsymbol{n}) = 0$.*

Proof. Write

$$P(\boldsymbol{k}/\boldsymbol{n}) = \sum_{j=0}^{2s-1} \sum_{s_1+\cdots+s_d=j} c(s_1,\ldots,s_d) \left(\frac{k_1}{n_1}\right)^{s_1} \cdots \left(\frac{k_d}{n_d}\right)^{s_d},$$

where $\{c(\cdot)\}$ is a set of coefficients. Then for any p where $s_p \geq 2$,

$$\begin{aligned}\mathrm{D}\, P\left(\frac{\boldsymbol{k}}{\boldsymbol{n}}\right) &= P\left(\frac{\boldsymbol{k}+\boldsymbol{e}_p}{\boldsymbol{n}}\right) + P\left(\frac{\boldsymbol{k}-\boldsymbol{e}_p}{\boldsymbol{n}}\right) - 2P\left(\frac{\boldsymbol{k}}{\boldsymbol{n}}\right) \\ &= \sum_{j=0}^{2s-1} \sum_{s_1+\cdots+s_d=j} c'(s_1,\ldots,s_d) \left(\frac{k_1}{n_1}\right)^{s_1} \cdots \left(\frac{k_p}{n_p}\right)^{s_p-2} \cdots \left(\frac{k_d}{n_d}\right)^{s_d},\end{aligned}$$

where $\{c'(\cdot)\}$ is another set of coefficients. When $s_p < 2$, the term $(k_p/n_p)^{s_p-2}$ is replaced by zero. Hence, the operator D reduces the order of P by 2. Then, after applying the operator s times, the result is zero. ∎

THEOREM 5.2. *Denote by $\mathbb{P}_t^d$ the space of polynomials of d variables of degree at most t. The space*

$$A_t^1 := \left\{ \boldsymbol{a} \in \mathbb{R}^n \;\middle|\; \sum_{\mathbf{0}\leq\boldsymbol{k}\leq\boldsymbol{n}-\mathbf{1}} a_{\boldsymbol{k}} P(\boldsymbol{k}/\boldsymbol{n}) = 0,\ \forall P \in \mathbb{P}_t^d, \right\}$$

has dimension $n - \binom{t+d}{d}$ when $n > \binom{t+d}{d}$, and the space

$$A_s^2 := \left\{ {L^{[1]}}^T \cdots {L^{[s]}}^T \boldsymbol{u} \;\middle|\; \boldsymbol{u} \in \mathbb{R}^{(n_1-2s)\cdots(n_d-2s)} \right\}$$

has dimension $(n_1 - 2s)\cdots(n_d - 2s)$.

Proof. The number of monomials of d variables of degree at most t is

$$\sum_{j=0}^{t} \binom{j+d-1}{d-1} = \binom{t+d}{d},$$

which is thus the dimension of $\mathbb{P}_t^d$. Then the entries of $\boldsymbol{a}$ have a degree of freedom $n - \binom{t+d}{d}$.

The dimension of A_s^2 is equal to the number of rows of $L^{[s]}\cdots L^{[1]}$, since the latter matrix has full row rank. ∎

Remark. In fact, A_s^2 is a subspace of A_{2s-1}^1 when $n > \binom{2s-1+d}{d}$.

6. Discrete Laplace operator on a finite-element mesh. This section concerns generalizing the Laplace operator for a set of scattered points. For a twice-differentiable u, the objective is to approximate $\Delta u(\boldsymbol{x})$ by a linear combination of the $u(\boldsymbol{x}_i)$'s for a set of $\boldsymbol{x}_i$'s that surround $\boldsymbol{x}$. For this, we assume that a mesh of the points can be constructed, so that neighboring information for every point is available. In what follows, by "mesh" we refer to a *finite-element mesh* that consists of a triangulation of X. Hence, the points X are the mesh vertices. In $\mathbb{R}^d$, each *finite element* E of the mesh is a d-simplex, defined as the convex hull of $d+1$ vertices $\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_{d+1}} \in X$ in a nondegenerate position; that is, the vertices do not lie on any subspace of $\mathbb{R}^d$ with a lower dimension. The union of E is the closed domain $\bar{\Omega} = \Omega \cup \partial\Omega$, where $\partial\Omega$ is the boundary.

For a twice differentiable u and a continuously differentiable v, consider the Green's identity

$$\int_\Omega (v\Delta u + \nabla v \cdot \nabla u) = \oint_{\partial\Omega} v\,(\nabla u \cdot \boldsymbol{n})$$

where $\boldsymbol{n}$ is the outward unit normal of $\partial\Omega$. Here, we reuse the notation $\boldsymbol{n}$ because its original meaning of the grid dimension is useless in this section. Discretizing the Green's identity requires a set of basis functions to represent u and v. For each vertex $\boldsymbol{x}_i$, let $v_i : \bar{\Omega} \to \mathbb{R}$ be a piecewise polynomial that is 1 at $\boldsymbol{x}_i$ and 0 at all other vertices. The simplest case is to let v_i be linear in each E, and we consider only this case. The span of the v_i's is the space of piecewise linear functions on the mesh. We can then approximate u and Δu as

$$u(\boldsymbol{x}) \approx \sum_{i=1}^n u(\boldsymbol{x}_i)\, v_i(\boldsymbol{x}), \qquad \Delta u(\boldsymbol{x}) \approx \sum_{i=1}^n \Delta u(\boldsymbol{x}_i)\, v_i(\boldsymbol{x}).$$

We also approximate ∇u as

$$\nabla u(\boldsymbol{x}) \approx \sum_{i=1}^n u(\boldsymbol{x}_i)\, \nabla v_i(\boldsymbol{x}).$$

Note that ∇v_i is not well defined for $\boldsymbol{x} \in [(\bigcup_E \partial E)\backslash\partial\Omega] \cup X$, which is the set of locations that are adjacent to two or more elements. However, this set has a measure of zero in $\mathbb{R}^d$ and thus does not contribute to the integral over Ω. For our purpose we can arbitrarily define, for example,

$$\nabla v_i(\boldsymbol{x}) = \frac{1}{|\{E \ni \boldsymbol{x}\}|} \sum_{E \ni \boldsymbol{x}} \nabla v_i(E), \quad \boldsymbol{x} \in [(\textstyle\bigcup_E \partial E)\backslash\partial\Omega] \cup X,$$

where $\nabla v_i(E) = \nabla v_i(\boldsymbol{x})$ is the constant gradient for all $\boldsymbol{x} \in E\backslash\partial E$.

Based on the above approximation, for every $v = v_k$, the Green's identity is then discretized as

$$\sum_{i=1}^n \left[\int_\Omega v_k v_i\right] \Delta u(\boldsymbol{x}_i) + \sum_{i=1}^n \left[\int_\Omega \nabla v_k \cdot \nabla v_i\right] u(\boldsymbol{x}_i) \approx \sum_{i=1}^n \left[\oint_{\partial\Omega} v_k\,(\nabla v_i \cdot \boldsymbol{n})\right] u(\boldsymbol{x}_i).$$

If we define the square matrices (row indexed by k and column indexed by i)

$$M = \left[\int_\Omega v_k v_i\right], \quad L = \left[-\int_\Omega \nabla v_k \cdot \nabla v_i\right], \quad B = \left[\oint_{\partial\Omega} v_k\,(\nabla v_i \cdot \boldsymbol{n})\right], \tag{6.1}$$

then the above formula can be written in matrix form:

$$M \cdot \Big[\Delta u(\boldsymbol{x}_i)\Big] \approx (B + L) \cdot \Big[u(\boldsymbol{x}_i)\Big]. \tag{6.2}$$

Note that we overload the notation L here; it is different from the one in the regular grid case (4.1). All the discussions of L in this section refer to the one defined in (6.1). In finite-element analysis, the matrix $-L$ is the *stiffness matrix*, and the matrix M is the *mass matrix*. Roughly speaking, the linear transformation $M^{-1}(B + L)$ acts like a Laplacian on u, in a discrete sense, but this is not the discrete Laplace operator we

will define. Properties of the matrices M, L, and B are studied in the next subsection before we propose a formal definition of the discrete Laplace operator.

The 1D case needs a special treatment. In $\mathbb{R}^1$, the Green's identity is simply the formula of integration by parts:

$$\int_{x_1}^{x_n} vu'' = vu'\Big|_{x_1}^{x_n} - \int_{x_1}^{x_n} v'u',$$

where we assume that the vertices $x_1, \ldots, x_n$ are ordered increasingly. Then following a similar argument as above, we see that the discretization of the formula also leads to (6.2), using the same definition of M and L, with the matrix B slightly modified to

$$B = \left[v_k v_i' \Big|_{x_1}^{x_n} \right].$$

6.1. Formulas and properties of M, L, and B. The computation of M and L is well known. For the sake of completeness we briefly derive the formulas. These formulas are important to the definition of the discrete Laplace operator.

To simplify notation, we assume without loss of generality that the vertices of an element E are $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d+1}$. Define

$$Q = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{d+1} \\ 1 & 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}.$$

Then the measure of E is

$$\operatorname{meas}(E) = \frac{|\det(Q)|}{d!}. \tag{6.3}$$

Let R_i denote the matrix by replacing the component $\boldsymbol{x}_i$ in Q by $\boldsymbol{x}$. Then the basis function

$$v_i(\boldsymbol{x}) = \frac{\det(R_i)}{\det(Q)}, \qquad \text{for } \boldsymbol{x} \in E,$$

with partial derivatives $\partial_j v_i = (Q^{-1})_{ij}$. Therefore

$$\int_E \nabla v_k \cdot \nabla v_i = \frac{|\det(Q)|}{d!} \sum_{j=1}^{d} (Q^{-1})_{kj} (Q^{-1})_{ij}. \tag{6.4}$$

In $\mathbb{R}^1$, with simple algebraic calculations we have

$$\int_E v_k v_i = \frac{|\det(Q)|}{6} \qquad k \neq i, \qquad \int_E v_k v_k = \frac{|\det(Q)|}{3} \qquad \text{in 1D.}$$

In high dimensions, without loss of generality we assume that $k = 2$, $i = 3$. We consider the linear transformation

$$T(\boldsymbol{x}) = [\boldsymbol{x}_2 - \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{d+1} - \boldsymbol{x}_1]^{-1} (\boldsymbol{x} - \boldsymbol{x}_1)$$

that maps the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{d+1}$ to $\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$, where each $\boldsymbol{e}_i$ is the ith column of the identity matrix. We call the simplex defined by the latter set of points the *canonical element*, and denote it by $T(E)$. Then by a change of variables,

$$\int_E v_2 v_3 = |\det(T^{-1})| \int_{T(E)} \tilde{v}_2 \tilde{v}_3,$$

where $\tilde{v}_2$ is the piecewise linear function that is equal to 1 at $\boldsymbol{e}_1$ and 0 at other vertices of $T(E)$, and similarly $\tilde{v}_3$ is the piecewise linear function that is equal to 1 at $\boldsymbol{e}_2$ and 0 at other vertices of $T(E)$. One can easily verify that the integral $\int_{T(E)} \tilde{v}_2\tilde{v}_3$ is equal to $1/(d+2)!$. When $k = i$, say both are equal to 2, then applying the same transformation T the integral becomes $\int_{T(E)} \tilde{v}_2\tilde{v}_2$, which can be verified to be $2/(d+2)!$. Since $|\det(T^{-1})|$ is equal to $|\det(Q)|$, we conclude that

$$\int_E v_k v_i = \frac{|\det(Q)|}{(d+2)!} \qquad k \neq i, \qquad \int_E v_k v_k = \frac{2\,|\det(Q)|}{(d+2)!}. \tag{6.5}$$

This formula is consistent with the case $d = 1$.

Last, we consider the boundary integral $\oint v_k\,(\nabla v_i \cdot \boldsymbol{n})$ on $E \cap \partial\Omega$. Since $E \cap \partial\Omega$ may contain more than one face of E, we use E_s to denote each individual face, and we let $\boldsymbol{x}_{i'}$ be the only vertex that belongs to E but not E_s. Observe that $\nabla v_i \cdot \boldsymbol{n}$ is a constant; thus,

$$\oint_{E_s} v_k\,(\nabla v_i \cdot \boldsymbol{n}) = (\nabla v_i \cdot \boldsymbol{n}) \oint_{E_s} v_k, \qquad \boldsymbol{x}_k \in E_s.$$

Furthermore, the normal $\boldsymbol{n}$ is equal to $-\nabla v_{i'}/\|\nabla v_{i'}\|$, and $\oint_{E_s} v_k$ is equal to $\mathrm{meas}(E)$ divided by the distance from $\boldsymbol{x}_{i'}$ to E_s, which is $\mathrm{meas}(E) \cdot \|\nabla v_{i'}\|$. Therefore,

$$\oint_{E_s} v_k\,(\nabla v_i \cdot \boldsymbol{n}) = -\frac{\nabla v_i \cdot \nabla v_{i'}}{\|\nabla v_{i'}\|} \cdot \mathrm{meas}(E) \cdot \|\nabla v_{i'}\| = -\frac{|\det(Q)|}{d!} \sum_{j=1}^{d} (Q^{-1})_{i'j}(Q^{-1})_{ij}. \tag{6.6}$$

One can show that (6.6) is also valid for the case $d = 1$.

In summary, the matrices L, M, and B are computed based on (6.4), (6.5), and (6.6), respectively, followed by an assembly of all the quantities computed on each element E or element face E_s on the boundary. By investigating the bilinear forms with respect to L and M, one can see that both are symmetric positive semi-definite. Further, M is nonsingular because of the first item of the following theorem.

THEOREM 6.1. *The matrices M, L, and B have the following properties.*

(i) *For every k,*

$$2\sum_{i\neq k} M_{ki} = M_{kk} = \frac{2}{d(d+1)} \sum_{E \ni \boldsymbol{x}_k} \mathrm{meas}(E).$$

Furthermore, the condition number of M is upper bounded by

$$3 \cdot \frac{\max_k\{\sum_{E\ni\boldsymbol{x}_k} \mathrm{meas}(E)\}}{\min_k\{\sum_{E\ni\boldsymbol{x}_k} \mathrm{meas}(E)\}}.$$

(ii) *For every k, $\sum_i L_{ki} = 0$, and for every $\boldsymbol{x}_k \notin \partial\Omega$, $\sum_i L_{ki}\,\boldsymbol{x}_i = \boldsymbol{0}$.*

(iii) *For every k, $\sum_i B_{ki} = 0$, and for every $\boldsymbol{x}_k \notin \partial\Omega$, B_{ki} is zero for all i.*

(iv) *For every k, $\sum_i (B+L)_{ki}\,\boldsymbol{x}_i = \boldsymbol{0}$.*

Proof. The first formula of Property (i) is obvious in light of (6.3) and (6.5). Then the bound of the condition number of M is a direct consequence of the Gershgorin's circle theorem.

For any element E and any point $\boldsymbol{x} \in E\backslash\partial E$,

$$\sum_{\boldsymbol{x}_i \in E} \nabla v_i(\boldsymbol{x}) = 0,$$

because $\sum_{\boldsymbol{x}_i \in E} v_i$ is the constant function with value 1. Therefore,

$$\sum_{\boldsymbol{x}_i \in E} \int_E \nabla v_k \cdot \nabla v_i = 0,$$

which proves the first equation of Property (ii). The same technique also proves Property (iii) by directly applying the definition of B.

For a specific j, the vector $\sum_{\boldsymbol{x}_i \in E} (\boldsymbol{x}_i)_j \nabla v_i$ consists of the first d entries of the jth row of QQ^{-1}, the identity matrix. Therefore, $\sum_{\boldsymbol{x}_i \in E} (\boldsymbol{x}_i)_j \nabla v_i = \boldsymbol{e}_j$. Hence,

$$\sum_{\boldsymbol{x}_i \in E} \left(\int_E \nabla v_k \cdot \nabla v_i \right) (\boldsymbol{x}_i)_j = \int_E \nabla v_k \cdot \left(\sum_{\boldsymbol{x}_i \in E} (\boldsymbol{x}_i)_j \nabla v_i \right) = \int_E (\nabla v_k)_j,$$

and thus

$$\sum_{\boldsymbol{x}_i \in E} \left(\int_E \nabla v_k \cdot \nabla v_i \right) \boldsymbol{x}_i = \int_E \nabla v_k = \operatorname{meas}(E)\, \nabla v_k.$$

Let a face E_s be defined by all the vertices of E but $\boldsymbol{x}_k$. Then E_s has a unit normal $\boldsymbol{n}^{E_s}$ pointing outward, and further, $\operatorname{meas}(E)\, \nabla v_k = -1/d \cdot \operatorname{meas}(E_s)\, \boldsymbol{n}^{E_s}$. Thus,

$$\sum_i L_{ki}\, \boldsymbol{x}_i = \frac{1}{d} \sum_{E \ni \boldsymbol{x}_k} \operatorname{meas}(E_s)\, \boldsymbol{n}^{E_s}. \tag{6.7}$$

When $\boldsymbol{x}_k$ is not on the boundary of the mesh, all the elements E that contain $\boldsymbol{x}_k$ form a polytope $\bigcup_{E \ni \boldsymbol{x}_k} E$ with boundary $\bigcup_{E \ni \boldsymbol{x}_k} E_s$. Then the above summation (effectively a closed surface integral of unit normal) is zero. This proves the second equation of Property (ii).

When $\boldsymbol{x}_k$ is on the boundary of the mesh, the boundary of the polytope $\bigcup_{E \ni \boldsymbol{x}_k} E$ consists of two parts, $\bigcup_{E \ni \boldsymbol{x}_k} E_s$ as defined above, and $\bigcup_{E \ni \boldsymbol{x}_k} E \cap \partial\Omega$. Using a similar idea as above, fixing an element E, for each $E_s \subset E \cap \partial\Omega$, we have

$$\sum_{\boldsymbol{x}_i \in E} \left[\oint_{E_s} v_k\, (\nabla v_i \cdot \boldsymbol{n}) \right] \boldsymbol{x}_i = \boldsymbol{n} \oint_{E_s} v_k = \frac{1}{d} \cdot \operatorname{meas}(E_s)\, \boldsymbol{n}.$$

Note that the normal $\boldsymbol{n}$ here is orthogonal to E_s and pointing outward. Therefore,

$$\sum_i B_{ki}\, \boldsymbol{x}_i = \frac{1}{d} \sum_{E \ni \boldsymbol{x}_k} \operatorname{meas}(E_s)\, \boldsymbol{n}^{E_s}.$$

Then together with (6.7) and using once again the fact of zero surface integral, we see that for any $\boldsymbol{x}_k$ on the boundary,

$$\sum_i (B + L)_{ki}\, \boldsymbol{x}_i = \boldsymbol{0}.$$

Further, since the kth row of B_{ki} is zero for $\boldsymbol{x}_k$ not on the boundary, then using the second equation of Property (ii), we complete the proof of the nonboundary case and conclude Property (iv). ∎

COROLLARY 6.2. *If the configuration of the points X is not degenerate, that is, there does not exist a nonzero vector $\boldsymbol{b}$ such that $\boldsymbol{x}_i \cdot \boldsymbol{b}$ are the same for all i, then the dimension of the null space of $B + L$ is at least $d + 1$.*

Proof. The $d+1$ linearly independent vectors mapped to zero by $B+L$ are the vector of all 1's and the vectors of the jth component of the $\boldsymbol{x}_i$'s, $j = 1, \ldots, d$. ∎

Remark. We conjecture that the dimension of the null space is exactly $d+1$. A proof of the conjecture will need to verify that for any function u that is not affine, $B+L$ will not map the vector $[u(\boldsymbol{x}_i)]$ to zero. A proof is unclear for $d > 1$; however, the conjecture can be proved for the case $d = 1$ by using a different technique. The argument uses the fact that the top and the bottom row of $B+L$ are zero and the rest of the rows form a tridiagonal structure. Hence the dimension of the null space of $B+L$ is 2.

6.2. Discrete Laplace operator. Using Property (i) of Theorem 6.1, we have

$$\sum_k M_{ki} \Delta u(\boldsymbol{x}_i) \approx \sum_k \frac{3}{2} M_{kk} \Delta u(\boldsymbol{x}_k),$$

because M_{ki} is nonzero only when $\boldsymbol{x}_i$ is connected to $\boldsymbol{x}_k$, that is, $\Delta u(\boldsymbol{x}_i) \approx \Delta u(\boldsymbol{x}_k)$. This leads to

$$M \cdot \Big[\Delta u(\boldsymbol{x}_i)\Big] \approx M' \cdot \Big[\Delta u(\boldsymbol{x}_i)\Big], \tag{6.8}$$

where M' is a diagonal matrix with

$$M'_{kk} = 3M_{kk}/2. \tag{6.9}$$

Then combining (6.2) and (6.8), we have

$$M' \cdot \Big[\Delta u(\boldsymbol{x}_i)\Big] \approx (B+L) \cdot \Big[u(\boldsymbol{x}_i)\Big].$$

If we are interested in Δu for only the $\boldsymbol{x}_i$'s not on the boundary $\partial\Omega$, we can remove in the above formula the rows and columns of M', the entries of the vector $[\Delta u(\boldsymbol{x}_i)]$, and the rows of $B+L$, that correspond to the points on $\partial\Omega$. Denoting by $\tilde{M}'$, $\tilde{B}$, and $\tilde{L}$ the smaller matrices after the removal, we have

$$\Big[\Delta u(\boldsymbol{x}_i)\Big]_{\boldsymbol{x}_i \notin \partial\Omega} \approx (\tilde{M}')^{-1} (\tilde{B} + \tilde{L}) \cdot \Big[u(\boldsymbol{x}_i)\Big].$$

Note that $\tilde{B}$ is empty because of Property (iii) of Theorem 6.1. Then, the matrix $(\tilde{M}')^{-1}\tilde{L}$ is the linear transformation that approximately maps $[u(\boldsymbol{x}_i)]$ to $[\Delta u(\boldsymbol{x}_i)]$.

As a matter of formality, we introduce notation because boundary points are removed every time a Laplacian is applied. Let $\bar{\Omega}^{[0]} \equiv \bar{\Omega}$, $\partial\Omega^{[0]} \equiv \partial\Omega$ and $X^{[0]} \equiv X$. Recursively for $s = 1, 2, \ldots$, we define $X^{[s]} = X^{[s-1]} \backslash \{\boldsymbol{x}_i \in \partial\Omega^{[s-1]}\}$, the set of points not on the boundary of $\bar{\Omega}^{[s-1]}$, and we let $\bar{\Omega}^{[s]}$ be the domain defined by the submesh of the point in $X^{[s]}$, and $\partial\Omega^{[s]}$ be the boundary of $\bar{\Omega}^{[s]}$. Then we arrive at the following definition.

DEFINITION 6.3. *The discrete Laplace operator* D *on an infinite mesh (without boundary) for an arbitrary function f is*

$$\mathrm{D}\, f(\boldsymbol{x}_k) := \sum_i \frac{2L_{ki}}{3M_{kk}} f(\boldsymbol{x}_i),$$

where L and M are defined in (6.1). *In the finite case, for $s = 1, 2, \ldots$, $L^{[s]}$ is the matrix representation of the operator* D *with entries $(L^{[s]})_{ki} = \frac{2}{3} L_{ki}/M_{kk}$ for all k*

and i where $\boldsymbol{x}_k \in X^{[s]}$ and $\boldsymbol{x}_i \in X^{[s-1]}$. The preconditioned kernel matrix $\Phi^{[2s]}$ is defined by reusing (4.2) *that was originally used for the regular grid case.*

The following result is parallel to Theorem 5.1 for the regular grid case. It is an immediate consequence of Property (ii) of Theorem 6.1.

THEOREM 6.4. *For any affine function P and mesh vertices $\boldsymbol{x}_k$, $\mathrm{D}\,P(\boldsymbol{x}_k) = 0$.*

6.3. Example. In $\mathbb{R}^1$, let $h_i = x_{i+1} - x_i$ denote the spacing between adjacent points, and let $g_i = h_{i-1} + h_i$. The modified stiffness matrix and the modified mass matrix are, respectively,

$$\tilde{L} = \begin{bmatrix} 1/h_1 & -g_2/(h_1h_2) & 1/h_2 & & & \\ & 1/h_2 & -g_3/(h_2h_3) & 1/h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1/h_{n-2} & -g_{n-1}/(h_{n-2}h_{n-1}) & 1/h_{n-1} \end{bmatrix}$$

and

$$\tilde{M}' = \begin{bmatrix} g_2/3 & h_2/6 & & & \\ h_2/6 & g_3/3 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & h_{n-2}/6 \\ & & & h_{n-2}/6 & g_{n-1}/3 \end{bmatrix}.$$

Then, for the kernel function $\phi(x) = |x|^3$,

$$\begin{aligned} \Phi^{[2]} &= L^{[1]}\Phi {L^{[1]}}^T \\ &= \frac{8}{9}\begin{bmatrix} 2/g_2 & h_2/(g_2g_3) & & & & \\ h_2/(g_2g_3) & 2/g_3 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & h_{n-2}/(g_{n-2}g_{n-1}) & \\ & & & h_{n-2}/(g_{n-2}g_{n-1}) & 2/g_{n-1} \end{bmatrix}. \end{aligned} \tag{6.10}$$

THEOREM 6.5. *The condition number of $\Phi^{[2]}$ in* (6.10) *is upper bounded by*

$$\frac{\sqrt{2}+1}{\sqrt{2}-1} \cdot \frac{\max\{g_i\}}{\min\{g_i\}}.$$

Proof. Consider the matrix in the square bracket of (6.10). If we multiply a diagonal matrix $\mathrm{diag}(\sqrt{g_i})$ to its both sides simultaneously and denote the resulting matrix A, then the condition number of $\Phi^{[2]}$ is upper bounded by $\kappa(A) \cdot \max\{g_i\}/\min\{g_i\}$.

The diagonal of the matrix A is a constant 2. The sum of the two off-diagonal elements on a row of A is

$$\frac{h_{i-1}}{\sqrt{g_{i-1}g_i}} + \frac{h_i}{\sqrt{g_ig_{i+1}}} \le \frac{\sqrt{h_{i-1}}}{\sqrt{g_i}} + \frac{\sqrt{h_i}}{\sqrt{g_i}} = \frac{\sqrt{h_{i-1}}+\sqrt{h_i}}{\sqrt{h_{i-1}+h_i}} \le \sqrt{2}.$$

Therefore, from the Gershgorin's circle theorem, the condition number of A is bounded by $(2+\sqrt{2})/(2-\sqrt{2})$. ∎

6.4. Analysis. Standard results of finite-element analysis can be borrowed to characterize the difference between the continuous Laplacian Δ and the discrete Laplace operator D defined in Definition 6.3. Let h_E and ρ_E denote the diameter of an element E and the supremum of the diameters of the spheres inscribed in E, respectively. Denote by h the maximum of h_E over all E, and by σ_h an upper bound of h_E/ρ_E. A family of finite-element meshes characterized by h is said to be *conforming* if σ_h is bounded away from infinity when h is sufficiently small. We have the following theorem.

THEOREM 6.6. *Given any $w \in C^3(\Omega)$ that vanishes on $\partial\Omega$ and any $u \in C^4(\Omega)$, if all the partial derivatives of w and u are finitely bounded, then for a family of conforming finite-element meshes on $\bar{\Omega}$ with vertices $\{\boldsymbol{x}_k\}$ and characterized by h, there exists a constant C independent of h (when h is sufficiently small) such that the M'-inner product*

$$\left| \Big\langle [w(\boldsymbol{x}_k)],\ [\Delta u(\boldsymbol{x}_k) - \mathrm{D}\, u(\boldsymbol{x}_k)] \Big\rangle_{M'} \right| \le C \cdot \mathrm{tr}(M') \cdot h,$$

where M' is the positive definite diagonal matrix defined in (6.9).

Proof. The inner product (without the absolute sign) is equal to

$$\sum_{\boldsymbol{x}_k \notin \partial\Omega} \frac{3}{2} M_{kk} w(\boldsymbol{x}_k) \Delta u(\boldsymbol{x}_k) - \sum_{\boldsymbol{x}_k \notin \partial\Omega} \frac{3}{2} M_{kk} w(\boldsymbol{x}_k)\, \mathrm{D}\, u(\boldsymbol{x}_k)$$

which can be rewritten as $F + G + H$, where

$$F = \sum_{\boldsymbol{x}_k, \boldsymbol{x}_i \in \bar{\Omega}} M_{ki} w(\boldsymbol{x}_k) \Delta u(\boldsymbol{x}_k) - \int_\Omega w \Delta u$$

$$G = -\int_\Omega \nabla w \cdot \nabla u + \int_\Omega \left(\sum_{\boldsymbol{x}_k \in \bar{\Omega}} w(\boldsymbol{x}_k) \nabla v_k \right) \left(\sum_{\boldsymbol{x}_i \in \bar{\Omega}} u(\boldsymbol{x}_i) \nabla v_i \right)$$

$$H = \sum_{\boldsymbol{x}_k, \boldsymbol{x}_i \in \bar{\Omega}} L_{ki} w(\boldsymbol{x}_k) u(\boldsymbol{x}_i) - \sum_{\boldsymbol{x}_k \in \bar{\Omega}} \frac{3}{2} M_{kk} w(\boldsymbol{x}_k) \sum_{\boldsymbol{x}_i \in \bar{\Omega}} \frac{2L_{ki}}{3M_{kk}} u(\boldsymbol{x}_i),$$

because the combination of the second term of F and the first term of G vanishes by the Green's identity and the combination of the second term of G and the first term of H vanishes by the definition of L. It is clear that H is zero. By the definition of M, the first term of F is equal to

$$\int_\Omega \left[\sum_{\boldsymbol{x}_k \in \bar{\Omega}} w(\boldsymbol{x}_k) \Delta u(\boldsymbol{x}_k) v_k \right] \left[\sum_{\boldsymbol{x}_i \in \bar{\Omega}} v_i \right] = \int_\Omega \sum_{\boldsymbol{x}_k \in \bar{\Omega}} w(\boldsymbol{x}_k) \Delta u(\boldsymbol{x}_k) v_k.$$

For any function f we denote by $\underline{f}$ the piecewise linear approximation which agrees with f at all the $\boldsymbol{x}_k$'s. Then

$$F = \int_\Omega \Big(\underline{w \Delta u} - w \Delta u \Big) \qquad \text{and} \qquad G = \int_\Omega \Big(\nabla \underline{w} \cdot \nabla \underline{u} - \nabla w \cdot \nabla u \Big).$$

Standard error analysis in conforming finite elements (see, e.g., [7]) states that there exists constants C_1 and C_2 that are independent of h such that

$$\sup_{\Omega} |\underline{w\Delta u} - w\Delta u| \le C_1 h^2 \cdot \sup_{\Omega} |\partial^2 (w\Delta u)|,$$
$$\sup_{\Omega} |\nabla \underline{w} - \nabla w| \le C_2 h \cdot \sup_{\Omega} |\partial^2 w|,$$

and similarly for $\nabla \underline{u} - \nabla u$. Then, absorbing C_1, C_2 with the supremums of the second order partial derivatives (because they are finite), we have for some constants $\tilde{C}_1$, $\tilde{C}_2$ and $\tilde{C}_3$ independent of h,

$$|F| \le \tilde{C}_1 \cdot \mathrm{meas}(\Omega) \cdot h^2 \qquad \text{and} \qquad |G| \le \tilde{C}_2 \cdot \mathrm{meas}(\Omega) \cdot h + \tilde{C}_3 \cdot \mathrm{meas}(\Omega) \cdot h^2.$$

The proof of the theorem is thus complete by noting that $\mathrm{tr}(M') = \frac{3}{d}\,\mathrm{meas}(\Omega)$. ∎

7. Example kernels. Two kernels are shown as examples in this section. The spectrum of the kernel matrices and the growth of the condition number are studied to empirically validate the preceding analysis. Experiments on the solution of the corresponding linear systems are presented to show the effectiveness of preconditioning.

7.1. Matérn kernel. The Matérn kernel [16, 6, 21] is one of the most popular kernels for data interpolation, for its flexibility in modeling the local smoothness of spatial/temporal data. Parameterized by $\nu, \ell > 0$, the Matérn kernel is defined as

$$\phi(\boldsymbol{x}) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|\boldsymbol{x}\|}{\ell} \right)^{\nu} \mathsf{K}_{\nu} \left(\frac{\sqrt{2\nu}\|\boldsymbol{x}\|}{\ell} \right),$$

where Γ is the Gamma function and K_ν is the modified Bessel function of the second kind of order ν. The kernel is infinitely differentiable everywhere except at the origin (where it is only $\lceil 2\nu - 1 \rceil$ times differentiable). The kernel admits a Fourier transform that is dimension dependent:

$$\hat{\phi}(\boldsymbol{\omega}) = \frac{(2\nu)^{\nu}\Gamma(\nu + d/2)}{\pi^{d/2}\ell^{2\nu}\Gamma(\nu)} \left(\frac{2\nu}{\ell^2} + \|\boldsymbol{\omega}\|^2 \right)^{-(\nu+d/2)}.$$

The transform $\hat{\phi}$ is positive, and hence the kernel matrix Φ is always positive definite. Figure 7.1(a) plots the sorted eigenvalues (blue) of $\Phi_{\boldsymbol{n}}$ for $\boldsymbol{n} = [32; 32]$, $\nu = 3$ and $\ell = 0.1$. Overlaped with the eigenvalues are the sorted values (red) of $\{(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})\}$. For computational convenience, the computation of $\hat{\phi}_{\boldsymbol{n}}$ based on (1.3) truncates the infinite series and sums over $-\mathbf{3} \le \boldsymbol{l} \le \mathbf{2}$ only. One clearly sees the similar distributions of the two sets.

The Fourier transform $\hat{\phi}$ is asymptotically equivalent to $(1 + \|\boldsymbol{\omega}\|)^{-(2\nu+d)}$. When $d = 2$ and $\nu = 3$, Theorem 4.1 implies that applying the discrete Laplace operator yields the equal distribution between $\{\lambda_{\boldsymbol{j}}(\Phi_{\boldsymbol{n}}^{[s]})/n\}$ and $\{(2\pi)^d \hat{\phi}_{\boldsymbol{n}}^{[s]}(2\pi \boldsymbol{j}/\boldsymbol{n})/n\}$ for $s = 1, 2$. To illustrate, in Figure 7.1(b) we plot the sorted values of the two sets (without the concurrent scaling by n) for the case $s = 1$. Note that when $s \neq 0$, $\hat{\phi}_{\boldsymbol{n}}^{[s]}(\mathbf{0}) = 0$, which means that one of the red dots cannot be displayed.

Theorem 4.1 does not apply for $s > 2$. Nevertheless, one can arbitrarily increase s, and the spectrum of $\Phi_{\boldsymbol{n}}^{[s]}$ is of interest. Figures 7.1(c) and (d) show the two sets of values (just like (a) and (b)) for $s = 3, 4$. One observes that except for a few

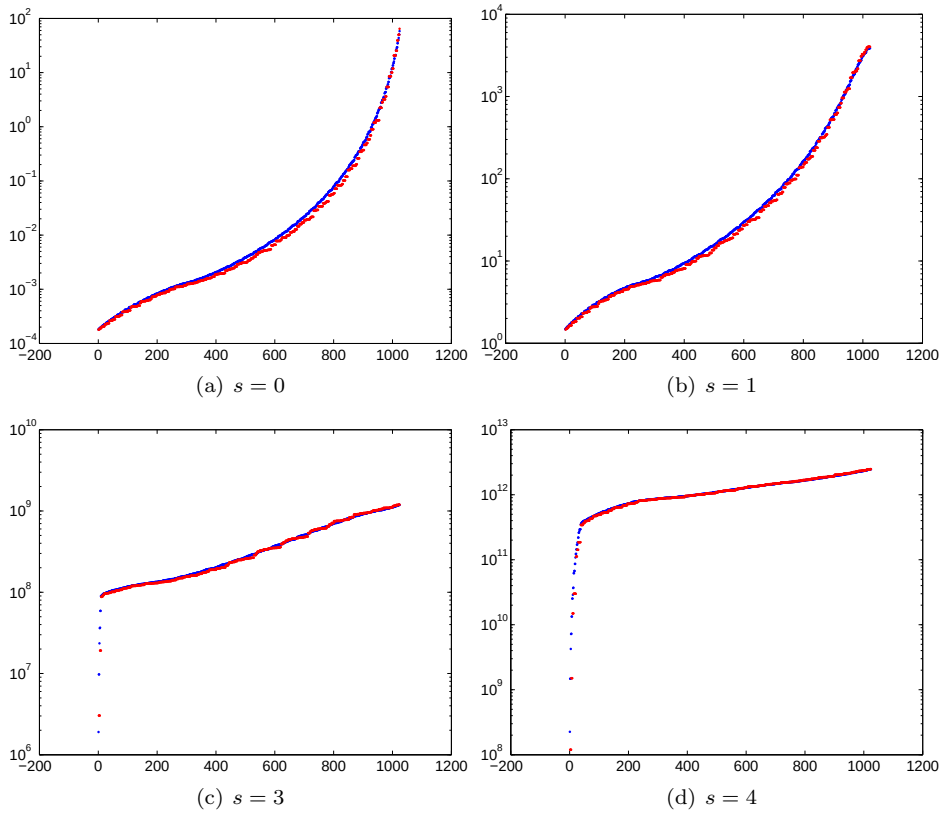


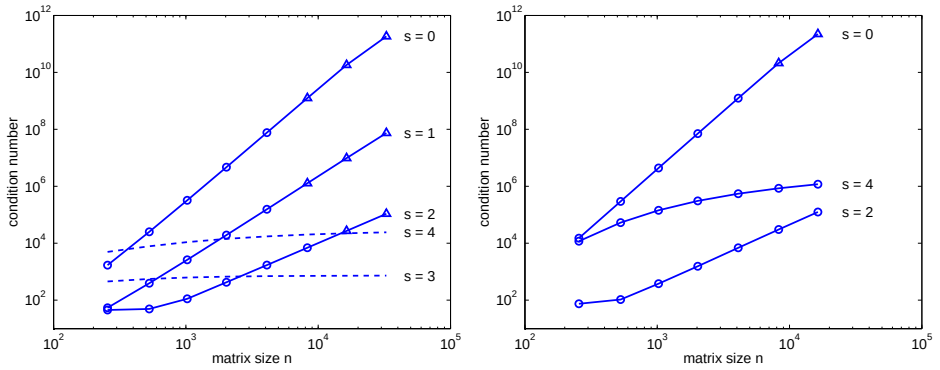
FIG. 7.1. *Sorted eigenvalues of* $\Phi_{\boldsymbol{n}}^{[s]}$ *(blue) and sorted values of* $(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})$ *(red) for Matérn kernel.*

outliers the majority of the two sets overlap. In other words, these plots suggest equal distributions for $s > 2$.

To investigate the trend of the conditioning, we plot in Figure 7.2(a) the condition number of $\Phi_{\boldsymbol{n}}^{[s]}$ versus n. For simplicity, we set the two components of $\boldsymbol{n}$ to be equal in the computation. The two extreme eigenvalues of $\Phi_{\boldsymbol{n}}^{[s]}$ were estimated by using the standard Lanczos algorithm. The triangles correspond to values that are not sufficiently accurate. The accurate values are strictly larger than the estimated values. Comparing the cases $s = 0, 1$, and 2, one sees that applying the discrete Laplace operator significantly reduces the condition number and suppresses its growth. Then for $s = 3, 4$, the condition number tends to be finitely bounded. The case $s = 4$ makes $2s$ match $2\nu + d$, the exponent of $\|\boldsymbol{\omega}\|$ in $\hat{\phi}$, and thus the bounded result is guaranteed according to [18].

For preconditioning, the discrete Laplace operator is applied to both sides of the kernel matrix simultaneously. Hence, only $\Phi_{\boldsymbol{n}}^{[s]}$ for even s can be obtained through linear transformations from $\Phi_{\boldsymbol{n}}$. According to the above results, $s = 4$ gives the optimal preconditioning performance.

To demonstrate the effectiveness of the discrete Laplace operator on a finite-element mesh, we consider a grid deformed from the regular grid (centered at the origin) by scaling the y-coordinates of the grid points by a quadratic function, which



(a) regular grid (b) deformed grid

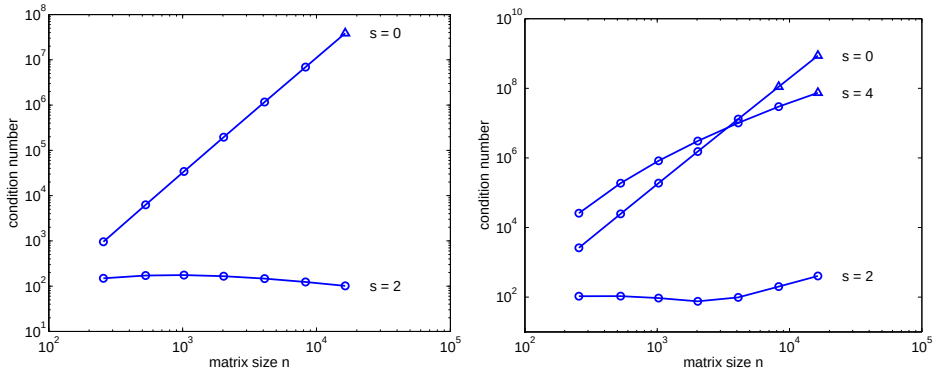
FIG. 7.2. *Condition number of* $\Phi_{\boldsymbol{n}}^{[s]}$ *by varying* $\boldsymbol{n}$. *Matérn kernel,* $\nu = 3$.

is 1 in the middle of the range of x and 0.5 at the extremes. Hence the deformed grid has an olive shape. The grid was used in [5] to model nonstationary stochastic processes, and it is not repeated here. In order to obtain a triangular mesh, for each grid cell the northeast and the southwest corners are connected.

Figure 7.2(b) plots the condition number of $\Phi^{[s]}$ versus n, for $s = 0, 2, 4$. One observes that the discrete Laplace operator helps significantly reduce the condition number. Compared with the regular grid case, the curves for $s = 0, 2$ look close to those shown in plot (a). The curve for $s = 4$ also shows a similar trend to that in plot (a); in particular, this curve seems finitely bounded as n increases.

So far, we have shown examples for $\nu = 3$, whereby there exists an integer s such that $2s = 2\nu + d$, where a bounded condition number is expected. In Figure 7.3 we plot the cases $\nu = 1.5$ and 2. We experimented with the choices of s such that the maximum is round$(\nu + d/2)$. For plot (a), the curve of $s = 2$ clearly shows the preconditioning effect and possible bounded result. For plot (b), using $s = 2$ yields a much smaller condition number for the grid sizes we have experimented with, but from the trend, it is unclear whether the curve of $s = 2$ will always stay under the one of $s = 4$. Nevertheless, both curves show reduction on the condition number as the grid size increases, and the reduction will be significant.

We show in Table 7.1 the convergence results of the conjugate gradient (CG) iterations for solving a linear system with respect to $\Phi^{[s]}$, by varying the grid configuration, the parameter ν, and the number n of grid points. The right-hand side of the linear system is generated by assuming a known solution vector (random), whose entries are iid standard normal variables, and we use a zero initial guess. The stopping criterion is a relative residual tolerance of `1.4901e-08` and a maximum of 2000 iterations. All the cases in the table correspond to the plots of Figures 7.2 and 7.3. One sees that the trend of the iteration numbers well match that of the condition numbers. In particular, without preconditioning ($s = 0$) CG is unable to converge for matrices of size larger than a few thousands; on the other hand, with an appropriate use of s, convergence is seen in less than 200 iterations for the largest grid. The only exception is the case of a deformed grid with $\nu = 3$ and $s = 4$. The iteration number grows to 1252 for the largest n. In fact, this is not surprising by inspecting Figure 7.2(b): the condition number is around 10^6. Nevertheless, according to the trend, it is not unreasonable to expect that the growth of the iteration number will



(a) $\nu = 1.5$ (b) $\nu = 2$

FIG. 7.3. *Condition number of $\Phi^{[s]}$ by varying n. Matérn kernel; deformed grid.*

TABLE 7.1

Number of CG iterations and relative solution error (underlined) for solving $\Phi^{[s]}\boldsymbol{x} = \boldsymbol{b}$ to relative residual tolerance `1.4901e-08`. Matérn kernel. The symbol "-" means convergence has not been seen within 2000 iterations, in which case the relative solution error is at the 2000th iteration.

Grid	ν	Fig.	s	$\log_2 n$ 8	9	10	11	12	13	14
Regular	3	7.2(a)	0	160	487	1396	-	-	-	-
				1e-6	8e-6	3e-5	2e-2	3e-1	7e-1	9e-1
			2	32	51	86	159	303	557	1047
				1e-8	6e-8	1e-7	2e-7	4e-7	1e-6	2e-6
			4	90	112	131	150	166	178	188
				2e-8	1e-8	2e-8	2e-8	2e-8	3e-8	4e-8
Deformed	3	7.2(b)	0	383	1425	-	-	-	-	-
				1e-6	2e-5	1e-2	2e-1	5e-1	8e-1	9e-1
			2	39	70	133	263	515	1025	1886
				3e-8	9e-8	2e-7	4e-7	1e-6	2e-6	5e-6
			4	104	185	295	467	685	984	1252
				2e-8	4e-8	6e-8	9e-8	8e-8	1e-7	1e-7
Deformed	1.5	7.3(a)	0	145	323	650	1367	-	-	-
				3e-7	1e-6	4e-6	1e-5	1e-3	5e-2	2e-1
			2	36	43	46	51	53	54	60
				3e-8	2e-8	1e-8	1e-8	3e-8	2e-8	2e-8
Deformed	2	7.3(b)	0	209	543	1240	-	-	-	-
				7e-7	4e-6	2e-5	2e-3	8e-2	3e-1	6e-1
			2	33	41	49	63	80	114	158
				2e-8	2e-8	3e-8	4e-8	6e-8	7e-8	9e-8
			4	139	306	639	1371	-	-	-
				7e-8	9e-8	1e-7	2e-7	2e-3	1e-2	9e-3

slow down and tend to stabilize instead of exploding.

7.2. Power-law kernel. The power-law kernel [6, 17], parameterized by $\alpha > 0$, is

$$\phi(\boldsymbol{x}) = \begin{cases} \Gamma(-\alpha/2)\|\boldsymbol{x}\|^\alpha, & \alpha/2 \notin \mathbb{N} \\ \frac{2(-1)^{\alpha/2+1}}{(\alpha/2)!}\|\boldsymbol{x}\|^\alpha \log\|\boldsymbol{x}\|, & \alpha/2 \in \mathbb{N}. \end{cases}$$

The coefficients in front of $\|\boldsymbol{x}\|$ alternate signs whenever $\alpha/2$ crosses an integer value. The kernel is so defined such that for any vector $\boldsymbol{a}$ with entries satisfying the condition (2.4), the expression (1.2) is valid with

$$\hat{\phi}(\boldsymbol{\omega}) = \frac{2^\alpha}{\pi^{d/2}}\Gamma\left(\frac{\alpha+d}{2}\right)\|\boldsymbol{\omega}\|^{-\alpha-d},$$

which is positive. Since the polynomial P in the condition (2.4) is required to have a degree at most $t = \lfloor \alpha/2 \rfloor$, it is expected that $\Phi^{[s]}$ with s being even and $s > t$ is positive definite. In fact, in the regular grid case, as long as $s > t$ (not necessarily being even), $\Phi^{[s]}$ is definite; it switches between positive and negative definiteness whenever s increases by 1. When $s \leq t$, $\Phi^{[s]}$ has $-2s + 2t + 1$ negative eigenvalues (the rests are positive) if s is even, and $\Phi^{[s]}$ has $-2s+2t+1$ positive eigenvalues (the rests are negative) if s is odd.

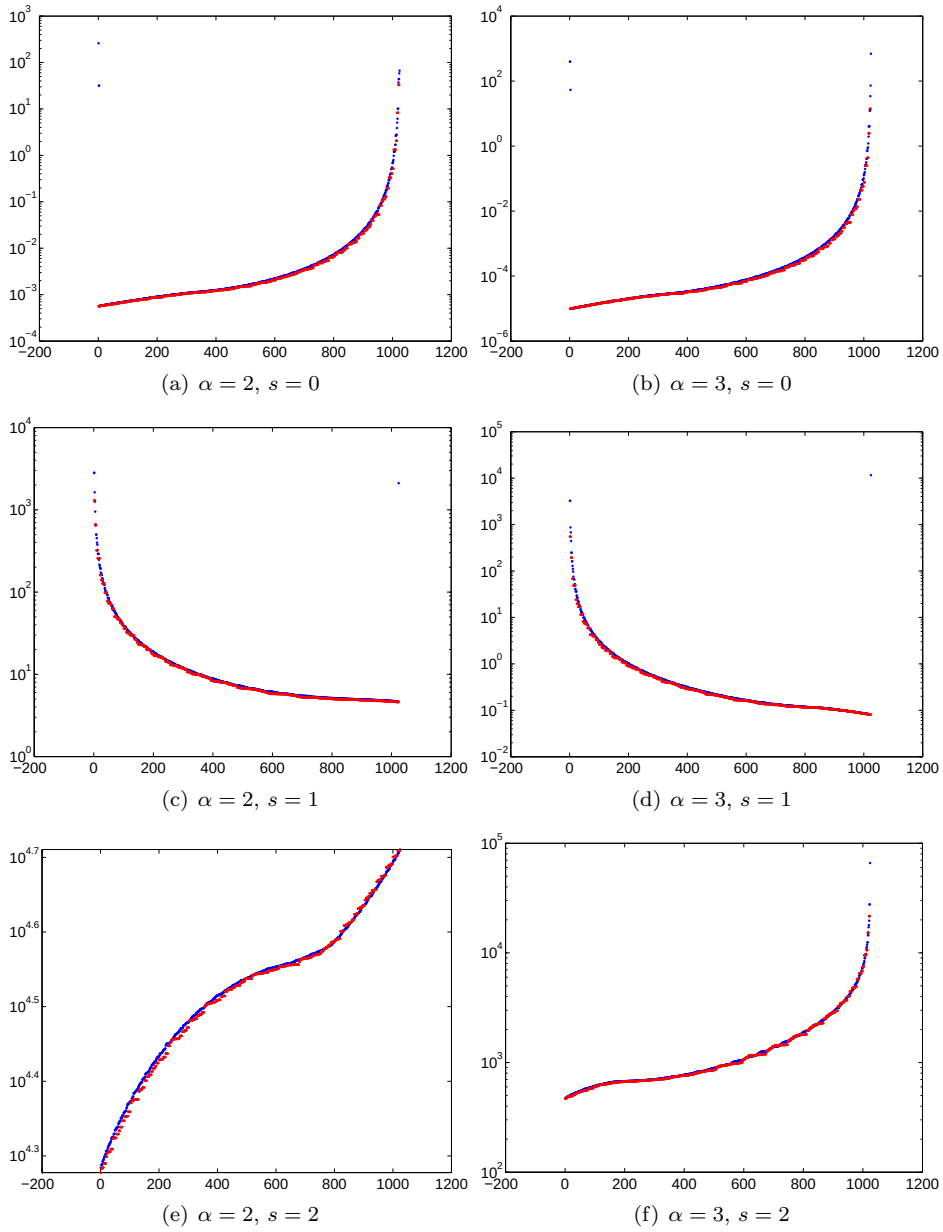
Figure 7.4 shows a series of plots of the sorted set $\{\lambda_{\boldsymbol{j}}(\Phi^{[s]}_{\boldsymbol{n}})\}$ (blue) overlapped by the sorted set $\{(2\pi)^d\hat{\phi}^{[s]}_{\boldsymbol{n}}(2\pi\boldsymbol{j}/\boldsymbol{n})\}$ (red), for a fixed grid $\boldsymbol{n} = [32; 32]$ but different parameters $\alpha = 2, 3$ and various s values. The sets of values are sorted in increasing order; but since the vertical axis of the plots is in the log scale, the absolute values are plotted instead. Hence, one should expect a "V" shape in each plot, with one side of "V" being extremely narrow since there are only $-2s + 2t + 1$ such values. One sees that a majority of the eigenvalues $\{\lambda_{\boldsymbol{j}}(\Phi^{[s]}_{\boldsymbol{n}})\}$ are distributed similarly with $\{(2\pi)^d\hat{\phi}^{[s]}_{\boldsymbol{n}}(2\pi\boldsymbol{j}/\boldsymbol{n})\}$. In plots (a) and (b), it is expected that there are three blue dots at the upper-left corner, corresponding to three negative eigenvalues. Two eigenvalues are close and they overlap, so visually one sees only two blue dots.

In Figures 7.5(a) and (b) we plot the growth of the condition number of $\Phi^{[s]}_{\boldsymbol{n}}$ as n increases. We varied s from 0 to round$((\alpha + d)/2)$, so that the maximum of $2s$ approximately agrees with the exponent in $\hat{\phi}$. One sees that as s increases, the growth of the condition number is progressively reduced. When $\alpha = 2$ and $s = 2$ (such that $2s = (\alpha + d)/2$), the condition number is finitely bounded.

In Figures 7.5(c) and (d) we plot the growth of the condition number for $\Phi^{[s]}$ on the deformed grid. In this case, $\Phi^{[s]}$ is defined for only even s. One observes that the plots look similar to (a) and (b).

We show in Table 7.2 the convergence results of the CG iterations for solving a linear system with respect to $\Phi^{[s]}$. The setting is similar to that of the experiments of the Matérn kernel. All the cases in the table correspond to the plots of Figure 7.5. Recall that $\Phi^{[s]}$ is positive definite only when $s > 0$, thus we only list the results of $s = 2$. The table clearly shows that preconditioning with an appropriate s is effective.

8. Discussion and conclusion. We have studied preconditioning kernel matrices by consecutively differentiating the kernel. The spectrum of the matrix in the regular grid case is revealed, and the effect of applying the discrete Laplace operator on the matrix is analyzed. The Laplacian preconditioning technique is generalized to scattered points without a regular grid structure, and a discrete Laplace operator



(a) $\alpha = 2,\ s = 0$ (b) $\alpha = 3,\ s = 0$

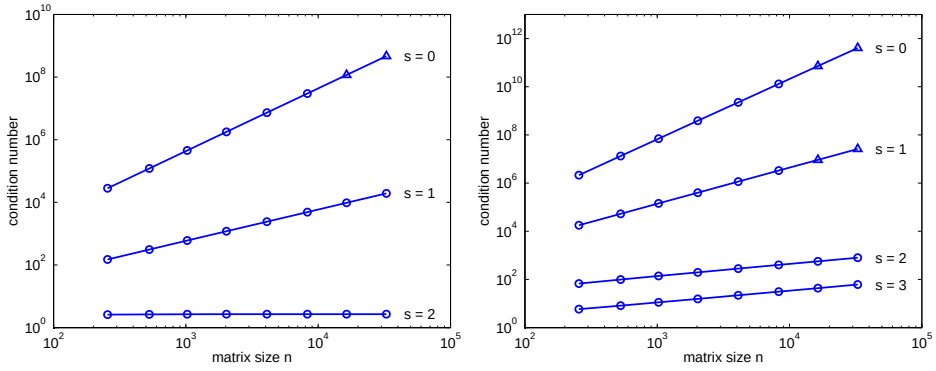
(c) $\alpha = 2,\ s = 1$ (d) $\alpha = 3,\ s = 1$

(e) $\alpha = 2,\ s = 2$ (f) $\alpha = 3,\ s = 2$

FIG. 7.4. *Sorted eigenvalues of* $\Phi_{\boldsymbol{n}}^{[s]}$ *(blue) and sorted values of* $(2\pi)^d \hat{\phi}_{\boldsymbol{n}}(2\pi \boldsymbol{j}/\boldsymbol{n})$ *(red) for power-law kernel. Absolute values are plotted since the vertical axis is in log scale.*

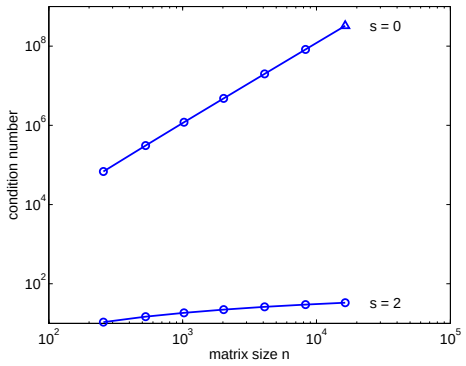
for this case is derived and analyzed. Numerical results confirm the preconditioning effect by applying the operators.

One shortcoming of the operator is that it reduces the size of the matrix, because there is not sufficient information to approximate the derivatives of the boundary points. For some applications such as in statistics, this means boundary information is not fully used. With increasingly fine discretizations/dense sampling, statistical

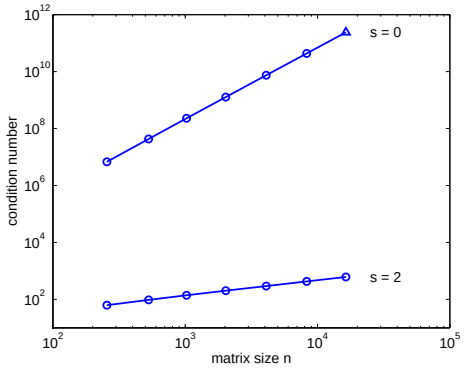


(a) $\alpha = 2$, regular grid

(b) $\alpha = 3$, regular grid



(c) $\alpha = 2$, deformed grid



(d) $\alpha = 3$, deformed grid

FIG. 7.5. *Condition number of* $\Phi^{[s]}$ *by varying* n. *Power-law kernel.*

TABLE 7.2
Number of CG iterations and relative solution error (underlined) for solving $\Phi^{[s]}\boldsymbol{x} = \boldsymbol{b}$ *to relative residual tolerance* `1.4901e-08`. *Power-law kernel.*

Grid	α	Fig.	s	$\log_2 n$						
				8	9	10	11	12	13	14
Regular	2	7.5(a)	2	14	14	14	14	14	14	14
				7e-9	9e-9	1e-8	1e-8	1e-8	1e-8	1e-8
Deformed	2	7.5(c)	2	23	28	33	37	42	45	49
				2e-8	1e-8	1e-8	2e-8	1e-8	2e-8	1e-8
Regular	3	7.5(b)	2	35	45	52	62	73	86	101
				3e-8	2e-8	4e-8	2e-8	2e-8	2e-8	2e-8
Deformed	3	7.5(d)	2	39	46	56	65	79	101	119
				3e-8	3e-8	2e-8	4e-8	3e-8	2e-8	5e-8

estimates from the remaining information may be asymptotically as efficient as those from the original information. Therefore, reducing the size of the kernel matrix is acceptable (see, e.g., [18]). Nevertheless, in the future we plan to investigate effective preconditioners that are full rank linear transformations.

Acknowledgments. The author is indebted to Mihai Anitescu and Michael Stein for their helpful discussions. The author also thanks the two anonymous referees whose comments have substantially improved the paper. This work was supported by the U.S. Department of Energy under Contract DE-AC02-06CH11357.

REFERENCES

[1] A. Böttcher and B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Springer, 1999.
[2] R. H. Chan and X.-Q. Jin, *An Introduction to Iterative Toeplitz Solvers*, SIAM, 2007.
[3] R. H. Chan and P. T. P. Tang, *Fast band-Toeplitz preconditioners for Hermitian Toeplitz systems*, SIAM J. Sci. Comput., 15 (1994), pp. 164–171.
[4] S. Chandrasekaran, P. Dewilde, M. Gu, W. Lyons, and T. Pals, *A fast solver for HSS representations via sparse matrices*, SIAM J. Matrix Anal. Appl., 29 (2006), pp. 67–81.
[5] J. Chen, M. Anitescu, and Y. Saad, *Computing $f(a)b$ via least squares polynomial approximations*, SIAM J. Sci. Comput., 33 (2011), pp. 195–222.
[6] J.-P. Chilès and P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty*, Wiley-Interscience, 1999.
[7] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, SIAM, 2002.
[8] I. M. Gel'fand and G. E. Shilov, *Generalized functions*, Academic Press, 1977.
[9] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*, Now Publishers Inc, 2006.
[10] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, American Mathematical Soc., 1984.
[11] S. Serra-Capizzano, *Distribution results on the algebra generated by Toeplitz sequences: a finite-dimensional approach*, Linear Algebra Appl., 328 (2001), pp. 121–130.
[12] ———, *Generalized locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations*, Linear Algebra Appl., 366 (2003), pp. 371–402.
[13] ———, *The GLT class as a generalized Fourier analysis and applications*, Linear Algebra Appl., 419 (2006), pp. 180–233.
[14] S. Serra-Capizzanoa and P. Sundqvist, *Stability of the notion of approximating class of sequences and applications*, J. Comput. Appl. Math., 219 (2008), pp. 518–536.
[15] M. L. Stein, *Fixed-domain asymptotics for spatial periodograms*, Journal of the American Statistical Association, 90 (1995), pp. 1277–1288.
[16] ———, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, 1999.
[17] ———, *Equivalence of gaussian measures for some nonstationary random fields*, Journal of Statistical Planning and Inference, 123 (2004), pp. 1–11.
[18] M. L. Stein, J. Chen, and M. Anitescu, *Difference filter preconditioning for large covariance matrices*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 52–72.
[19] J. Stewart, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math, 6 (1976), pp. 409–434.
[20] P. Tilli, *Locally Toeplitz sequences: spectral properties and applications*, Linear Algebra Appl., 278 (1998), pp. 91–120.
[21] H. Wendland, *Scattered Data Approximation*, Cambridge University Press, 2005.
[22] J. Xia, S. Chandrasekaran, M. Gu, and X. S. Li, *Superfast multifrontal method for large structured linear systems of equations*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1382–1411.