

ON USE OF DISCRETE LAPLACE OPERATOR FOR PRECONDITIONING KERNEL MATRICES

JIE CHEN*

Abstract. This paper studies 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 technique is to differentiate the kernel 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 ridding 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, covariance 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 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 the direct method approaches (for recent developments of compression based direct methods see [17, 4] among others); rather, iterative methods are in concern. 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 and finite domain Ω in \mathbb{R}^d , consider the matrix $\Phi \in \mathbb{R}^{n \times n}$ defined with entries

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

for a set of points $X = \{\mathbf{x}_i \in \mathbb{R}^d, i = 1, \dots, n\} \subset \bar{\Omega}$ and a kernel function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ which is even, that is, $\phi(-\mathbf{x}) = \phi(\mathbf{x})$. We are interested in the asymptotics (in particular the condition number κ) of Φ as $n \rightarrow \infty$. 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, or simulating stochastic processes using increasingly dense sampling.

Suppose the kernel ϕ admits a Fourier transform (the case when ϕ admits only a

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

generalized Fourier transform is discussed later). Denote by $\hat{\phi}$ the transform, that is,

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

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

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

Here, the upright bold face letter \mathbf{i} denotes the imaginary unit, and it is to be distinguished with the italic bold face letter \mathbf{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 \mathbf{n} to compactly denote the entries n_1, \dots, n_d , and naturally we let $n = n_1 \times \cdots \times n_d$. It would be more convenient to index the points by an integer vector, such as \mathbf{j} , which takes values from $\mathbf{0}$ to $\mathbf{n} - \mathbf{1}$. Denote by $\mathbf{x} \circ \mathbf{y}$ and \mathbf{x}/\mathbf{y} the element-wise multiplication and division of two vectors \mathbf{x} and \mathbf{y} , respectively. Then the Fourier transform (1.1) leads to [12, 11]

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

with

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

Then, (1.2) is equivalent to

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

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

$$\kappa(\Phi) \leq \frac{\sup |\hat{\phi}_{\mathbf{n}}|}{\inf |\hat{\phi}_{\mathbf{n}}|}. \quad (1.6)$$

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

Central to this paper is the manipulation of the Fourier transform $\hat{\phi}$ (or $\hat{\phi}_{\mathbf{n}}$) to suppress the growth of the condition number as n (or \mathbf{n}) increases. The essential idea is to take Laplacians on the kernel, and this applies to kernels with a Fourier transform that behaves like a power function. Section 2 offers an overview of the theory for not

only the case of the existence of a Fourier transform and a regular grid, but also the case of generalized Fourier transforms and the case without a regular grid. Detailed analysis is then provided in subsequent sections. Even for the case of a regular grid, which brings about a multi-level 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 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))

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

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

A study of (1.1) enables generalizations of the preconditioning technique to non-Toeplitz matrices, when one has a set of scattered 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 transform of the original function at the discretized locations. Thus, this work bridges 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.

2. Laplacian preconditioning. The spectrum of Φ has a close connection with the 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

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

This, in fact, is a bilinear form $\mathbf{a}^T \hat{\Phi} \mathbf{a}$, 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

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

Naturally, one would expect that the spectrum of Φ is in some sense similar to that of $\hat{\Phi}$, that is, the set $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}(2\pi\mathbf{k}/\mathbf{n})\}$. 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 \mathbf{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 $(2\pi)^d \hat{\phi}_{\mathbf{n}}(\boldsymbol{\omega})$ in $[0, 2\pi)^d$. Then, we have

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

which is consistent with (1.6).

There are several ways to make use of the result of equal distributions. One way is to differentiate ϕ such that $\hat{\phi}$ is correspondingly modified. Taking the Laplacian of both sides of (1.1) $2s$ times gives

$$\Delta^{2s}\phi(\mathbf{x}) = \int_{\mathbb{R}^d} \|\boldsymbol{\omega}\|^{4s} \hat{\phi}(\boldsymbol{\omega}) \exp(\mathbf{i}\boldsymbol{\omega}^T \mathbf{x}) d\boldsymbol{\omega}. \quad (2.2)$$

For example, consider the Matérn kernel [12, 6, 16] whose Fourier transform

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

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

For the preconditioning purpose, the task is to obtain a new matrix that approximates the kernel matrix defined by $\Delta^{2s}\phi$ via linear transformations of Φ . Section 4 considers the regular grid case, where Δ is naturally approximated by second order finite difference, which is denoted by D . We show that the spectrum of the new matrix by applying D^{2s} and the set $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}^{[2s]}(2\pi\mathbf{k}/\mathbf{n})\}$ are equally distributed, where $\hat{\phi}_{\mathbf{n}}^{[2s]}$ corresponds to $D^{2s}\phi$ just as $\hat{\phi}_{\mathbf{n}}$ corresponds to ϕ . Strictly speaking, s has to be strictly less than $\tau - d/4$. Otherwise, (2.2) does not hold because $\|\boldsymbol{\omega}\|^{4s} \hat{\phi}$ is not integrable, and a presumption in the analysis in Section 4 is also not satisfied. However, empirically when s passes $\tau - d/4$ the growth of the condition number of the new matrix still follows the rate $O(n^{4(\tau-s)/d})$. This probably requires a different analysis argument. For example, when τ is an integer and when $s = \tau$, [14] proves that the condition number of the new matrix is $O(1)$ (that is, independent of n), by directly upper and lower bounding $\hat{\phi}_{\mathbf{n}}^{[2s]}$. As a rule of thumb, for any $\tau > 0$, one chooses $s = \text{round}(\tau)$ 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

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

for some $\alpha > 0$. Standard theory on conditional positive definite functions [15] shows that (1.2) holds with $\hat{\phi}(\boldsymbol{\omega}) = c\|\boldsymbol{\omega}\|^{-\alpha-d}$ for all sets of $\{a_i\}$ that satisfy

$$\sum_{i=1}^n a_i P(\mathbf{x}_i) = 0 \quad (2.4)$$

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, 13]). The function $\hat{\phi}$ is known as the generalized Fourier transform of ϕ , and the vectors \mathbf{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 ω away from the origin, we similarly apply the discrete operator D^{2s} to suppress the growth of the condition number. In the ideal case, when $4s$ matches $\alpha + d$ (the exponent of $\|\omega\|$ in $\hat{\phi}$), in the generalized function sense the left-hand side of (2.2) is the delta function which in some sense results in the identity matrix. In other words, letting $s = \text{round}((\alpha + d)/4)$ may give the best preconditioning result.

The third way to fully exploit the Laplacian preconditioning idea is to rid the regular grid reliance. 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. Hence, 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 that is similar to the case of a regular grid.

The practical performance of the preconditioning strategies discussed above is shown by using two example kernels, the Matérn and the power-law kernel, in Section 7.

3. Spectrum of Toeplitz matrix in a fixed and finite domain. In the regular grid case, the kernel matrix Φ is multilevel Toeplitz. There are rich results about the asymptotics of Toeplitz matrices, one of the most celebrating of which is the Szegő's theorem [10] that 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, $\{\phi(\mathbf{j}/\mathbf{n})\}_j$, is clearly dependent on \mathbf{n} . To handle this situation, we use the asymptotic equivalence of matrices to show equal distribution results. To emphasize the dependence on \mathbf{n} , we use $\Phi_{\mathbf{n}}$ 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 $F : [M_1, M_2] \rightarrow \mathbb{R}$,*

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

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

1. A_n and B_n are both uniformly bounded in 2-norm, that is, when n is sufficiently large, $\|A_n\|_2, \|B_n\|_2 \leq M < \infty$ for some M independent of n , and
2. $\lim_{n \rightarrow \infty} \|A_n - B_n\|_F / \sqrt{n} = 0$.

REMARK 3.3. To avoid the handling of Frobenius norm, one can replace condition 2 by a stronger condition $\lim_{n \rightarrow \infty} \|A_n - B_n\|_2 = 0$.

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

LEMMA 3.4. *If $\{A_n\}$ and $\{B_n\}$ are asymptotically equivalent, then $\{\lambda_j(A_n)\}$ and $\{\lambda_j(B_n)\}$ are equally distributed.*

REMARK 3.5. To emphasize the regular grid structure, the lemma still holds when the indices n and j are replaced by \mathbf{n} and \mathbf{j} , respectively.

The following is the major result of this section.

THEOREM 3.6. *Let $\phi \in L^1 \cap L^2$. Then the set of eigenvalues of $\Phi_{\mathbf{n}}/n$ and the set $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}(2\pi\mathbf{j}/\mathbf{n})/n\}$ are equally distributed, with $\hat{\phi}_{\mathbf{n}}$ defined in (1.4).*

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

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

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

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

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

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

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

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

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

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

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

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

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

for all integers $s \geq 0$. This definition is consistent with (4.2) when the number in the superscript of Φ is even. Then we have a similar relation to (1.3):

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

where

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

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 $\Phi_{\mathbf{n}}^{[s]}/n$ and the set $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}^{[s]}(2\pi \mathbf{k}/\mathbf{n})/n\}$ are equally distributed, with $\hat{\phi}_{\mathbf{n}}^{[s]}$ defined in (4.5).*

Proof. Following a similar argument to the proof of Theorem 3.6, it suffices to show that both

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

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

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

since when taking $j = s$ this implies

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

and

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

We show (4.6) by induction.

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

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

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

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

The last equality to zero is because the limit of the term inside the square bracket as $\mathbf{n} \rightarrow \infty$ is 2 times the integral of $|\partial_p^3 f|$, which is finite. This shows the induction basis of (4.6), $j = 1$. Furthermore, for any $1 < j \leq s$,

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

with

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

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

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

$$\mathbf{a} = L^{[1]T} \dots L^{[s]T} \mathbf{u}$$

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

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

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

One may wish some equal distribution results for $\Phi_{\mathbf{n}}^{[2s]}$ 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 3.6 and 4.1—the Fourier series (3.1)—can not be established. Nevertheless, the growth of the condition number of $\Phi_{\mathbf{n}}^{[2s]}$ 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 \mathbf{k} and \mathbf{n} , $\mathrm{D}^s P(\mathbf{k}/\mathbf{n}) = 0$.*

Proof. Write

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

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

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

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. \square

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

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

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

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

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 \mathbf{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. \square

REMARK 5.3. 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 $\Delta u(\mathbf{x})$ by a linear combination of the $u(\mathbf{x}_i)$'s for a set of \mathbf{x}_i 's that surround \mathbf{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* which 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 $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_{d+1}} \in X$ in a non-degenerate 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

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

where \mathbf{n} is the outward unit normal of $\partial\Omega$. Here, we reuse the notation \mathbf{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 \mathbf{x}_i , let $v_i : \bar{\Omega} \rightarrow \mathbb{R}$ be a piecewise polynomial that is 1 at \mathbf{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

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

We also approximate ∇u as

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

Note that ∇v_i is not well defined for $\mathbf{x} \in [(\bigcup_E \partial E) \setminus \partial\Omega] \cup X$, 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,

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

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

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

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

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

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

then the above formula can be written in the matrix form:

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

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 refers 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

shall 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 nothing but the formula of integration by parts:

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

where we assume that the vertices x_1, \dots, 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 $\mathbf{x}_1, \dots, \mathbf{x}_{d+1}$. Define

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

Then the measure of E is

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

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

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

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

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

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

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

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

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

that maps the points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{d+1}$ to $\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_d$, where each \mathbf{e}_i is the i -th 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 \mathbf{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 \mathbf{e}_2 and 0 at other vertices of $T(E)$. It is easy to 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 thus conclude that

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

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

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

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

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

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

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 , it is clear that both are symmetric positive semi-definite. Further, M is non-singular 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 ,*

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

Furthermore, the condition number of M is upper bounded by

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

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

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

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

Proof. The first formula of Property (1) 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 $\mathbf{x} \in E \setminus \partial E$,

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

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

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

which proves the first equation of Property (2). The same technique also proves Property (3) by directly applying the definition of B .

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

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

and thus

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

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

$$\sum_i L_{ki} \mathbf{x}_i = \frac{1}{d} \sum_{E \ni \mathbf{x}_k} \text{meas}(E_s) \mathbf{n}^{E_s}. \quad (6.7)$$

When \mathbf{x}_k is not on the boundary of the mesh, all the elements E that contain \mathbf{x}_k form a polytope $\bigcup_{E \ni \mathbf{x}_k} E$ with boundary $\bigcup_{E \ni \mathbf{x}_k} E_s$. Then the above summation (effectively a closed surface integral of unit normal) is zero. This proves the second equation of Property (2).

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

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

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

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

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

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

Further, since the k -th row of B_{ki} is zero for \mathbf{x}_k not on the boundary, then using the second equation of Property (2), we complete the proof of the non-boundary case and conclude Property (4). \square

COROLLARY 6.2. *If the configuration of the points X is not degenerate, that is, there does not exist a nonzero vector \mathbf{b} such that $\mathbf{x}_i \cdot \mathbf{b}$ 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 j -th component of the \mathbf{x}_i 's, $j = 1, \dots, d$. \square

REMARK 6.3. 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 $[u(\mathbf{x}_i)]$ to zero. A proof is unclear for $d > 1$; however, the conjecture can be proved for the case $d = 1$ 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 it is clear that the dimension of the null space of $B + L$ is 2.

6.2. Discrete Laplace operator. Using Property (1) of Theorem 6.1, we have

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

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

$$M \cdot [\Delta u(\mathbf{x}_i)] \approx M' \cdot [\Delta u(\mathbf{x}_i)], \quad (6.8)$$

where M' is a diagonal matrix with

$$M'_{kk} = 3M_{kk}/2. \quad (6.9)$$

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

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

If we are only interested in Δu for the \mathbf{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 $[\Delta u(\mathbf{x}_i)]$, 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

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

Note that \tilde{B} is empty because of Property (3) of Theorem 6.1. Then, the matrix $(\tilde{M}')^{-1} \tilde{L}$ is the linear transformation that approximately maps $[u(\mathbf{x}_i)]$ to $[\Delta u(\mathbf{x}_i)]$.

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

DEFINITION 6.4. *The discrete Laplace operator D on an infinite mesh (without boundary) for an arbitrary function f is*

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

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

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 6.4. 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 6.7. *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 $\{\mathbf{x}_k\}$ and characterized by h , there exists a constant C independent of h (when h is sufficiently small), such that the M' -inner product*

$$\left| \left\langle [w(\mathbf{x}_k)], [\Delta u(\mathbf{x}_k) - D u(\mathbf{x}_k)] \right\rangle_{M'} \right| \leq C \cdot \text{tr}(M') \cdot h,$$

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

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

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

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

$$\begin{aligned} F &= \sum_{\mathbf{x}_k, \mathbf{x}_i \in \bar{\Omega}} M_{ki} w(\mathbf{x}_k) \Delta u(\mathbf{x}_k) - \int_{\Omega} w \Delta u \\ G &= - \int_{\Omega} \nabla w \cdot \nabla u + \int_{\Omega} \left(\sum_{\mathbf{x}_k \in \bar{\Omega}} w(\mathbf{x}_k) \nabla v_k \right) \left(\sum_{\mathbf{x}_i \in \bar{\Omega}} u(\mathbf{x}_i) \nabla v_i \right) \\ H &= \sum_{\mathbf{x}_k, \mathbf{x}_i \in \bar{\Omega}} L_{ki} w(\mathbf{x}_k) u(\mathbf{x}_i) - \sum_{\mathbf{x}_k \in \bar{\Omega}} \frac{3}{2} M_{kk} w(\mathbf{x}_k) \sum_{\mathbf{x}_i \in \bar{\Omega}} \frac{2L_{ki}}{3M_{kk}} u(\mathbf{x}_i), \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} \sup_{\Omega} |w\Delta u - w\Delta u| &\leq C_1 h^2 \cdot \sup_{\Omega} |\partial^2(w\Delta u)|, \\ \sup_{\Omega} |\nabla \underline{w} - \nabla w| &\leq C_2 h \cdot \sup_{\Omega} |\partial^2 w|, \end{aligned}$$

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 ,

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

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

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.

7.1. Matérn kernel. The Matérn kernel [12, 6, 16] 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

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

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 $\Phi_{\mathbf{n}}$ for $\mathbf{n} = [32; 32]$, $\nu = 3$ and $\ell = 0.1$. Overlapped with the eigenvalues are the sorted values (red) of $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}(2\pi\mathbf{j}/\mathbf{n})\}$. For computational convenience, the computation of $\hat{\phi}_{\mathbf{n}}$ based on (1.3) truncates the infinite series and sums over $-\mathbf{3} \leq \mathbf{l} \leq \mathbf{2}$ 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 once yields the equal distribution between $\{\lambda_{\mathbf{j}}(\Phi_{\mathbf{n}}^{[s]})/n\}$ and $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}^{[s]}(2\pi\mathbf{j}/\mathbf{n})/n\}$ for $s = 1, 2$. To illustrate, Figure 7.1(b) plots the sorted values of the two sets (without the concurrent scaling by n) for the case $s = 1$. Note that when $s \neq 0$, $\hat{\phi}_{\mathbf{n}}^{[s]}(\mathbf{0}) = 0$, 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 $\Phi_{\mathbf{n}}^{[s]}$ is of interest. Figure 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 outliers the majority of the two sets overlap. In other words, these plots seem to suggest equal distributions for $s > 2$.

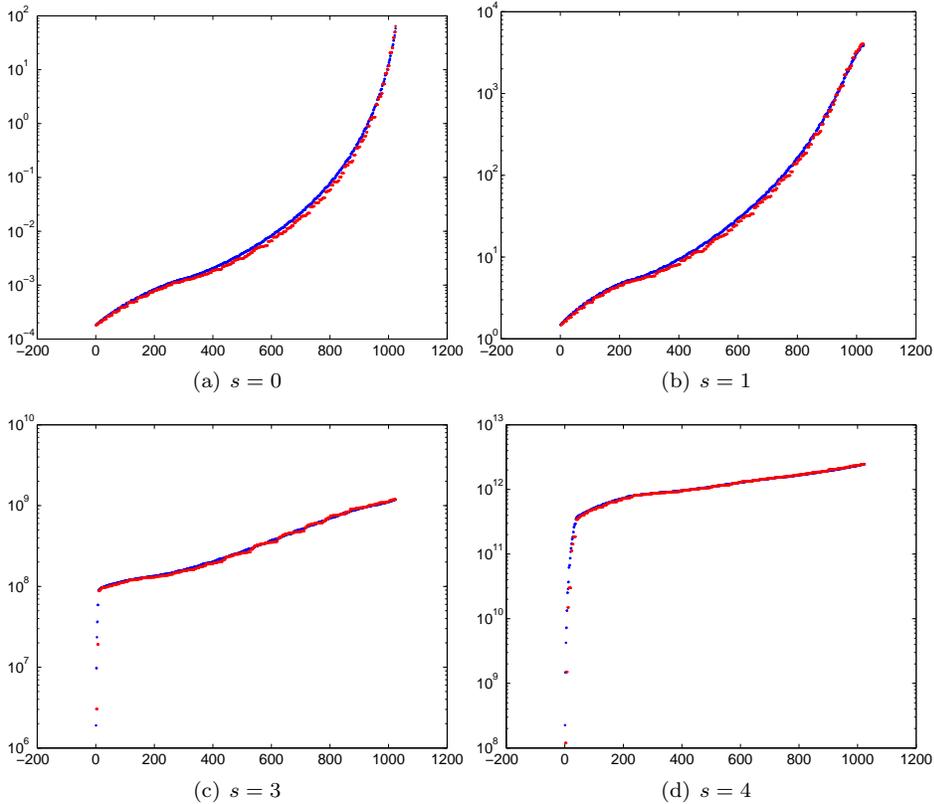


FIG. 7.1. Sorted eigenvalues of $\Phi_{\mathbf{n}}^{[s]}$ (blue) and sorted values of $(2\pi)^d \hat{\phi}_{\mathbf{n}}(2\pi \mathbf{j}/\mathbf{n})$ (red) for Matérn kernel.

To investigate the trend of the conditioning, Figure 7.2(a) plots the condition number of $\Phi_{\mathbf{n}}^{[s]}$ versus n . For simplicity, we set the two components of \mathbf{n} to be equal in the computation. The two extreme eigenvalues of $\Phi_{\mathbf{n}}^{[s]}$ 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 $\|\omega\|$ in $\hat{\phi}$, and thus the bounded result is guaranteed according to [14].

For the purpose of preconditioning, the discrete Laplace operator is applied to both sides of the kernel matrix simultaneously. Hence, only $\Phi_{\mathbf{n}}^{[s]}$ for even s can be obtained through linear transformations from $\Phi_{\mathbf{n}}$. 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 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

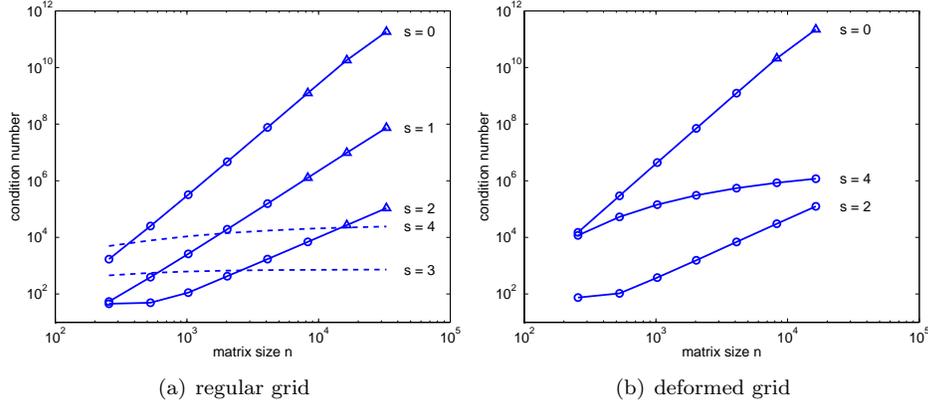


FIG. 7.2. Condition number of $\Phi_n^{[s]}$ by varying n . Matérn kernel, $\nu = 3$.

it is not repeated here. 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. Comparing 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 as 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 for regular grid. 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 $\text{round}(\nu + d/2)$. 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, but according to the trend, it is unclear if 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.

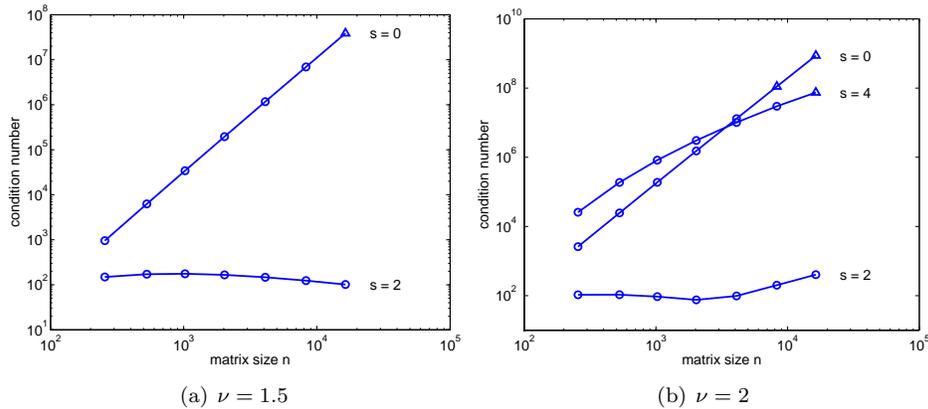


FIG. 7.3. Condition number of $\Phi_n^{[s]}$ by varying n . Matérn kernel; deformed grid.

7.2. Power-law kernel. The power-law kernel [6, 13], parameterized by $\alpha > 0$, is

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

The coefficients in front of $\|\mathbf{x}\|$ alternate signs whenever $\alpha/2$ crosses an integer value. The kernel is so defined such that for any vector \mathbf{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 $\{\lambda_j(\Phi_{\mathbf{n}}^{[s]})\}$ (blue) overlapped by the sorted set $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}^{[s]}(2\pi\mathbf{j}/\mathbf{n})\}$ (red), for a fixed grid $\mathbf{n} = [32; 32]$ but different parameters $\alpha = 2, 3$ and various s values. The sets of values are sorted in their increasing order, but since the vertical axis of the plots is in the log scale, the absolute values are plotted instead. So 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 $\{\lambda_j(\Phi_{\mathbf{n}}^{[s]})\}$ are distributed similarly with $\{(2\pi)^d \hat{\phi}_{\mathbf{n}}^{[s]}(2\pi\mathbf{j}/\mathbf{n})\}$. In plots (a) and (b), it is expected that there are 3 blue dots at the upper-left corner, corresponding to 3 negative eigenvalues. Two eigenvalues are close and they overlap so visually one sees only two blue dots.

In Figure 7.5(a) and (b) plot the growth of the condition number of $\Phi_{\mathbf{n}}^{[s]}$ as n increases. We varied s from 0 to $\text{round}((\alpha + d)/2)$, 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. In the case $\alpha = 2$ and $s = 2$ (such that $2s = (\alpha + d)/2$), the condition number is finitely bounded.

In Figure 7.5(c) and (d) 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), which shows the effectiveness of the discrete Laplace operator for preconditioning.

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 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

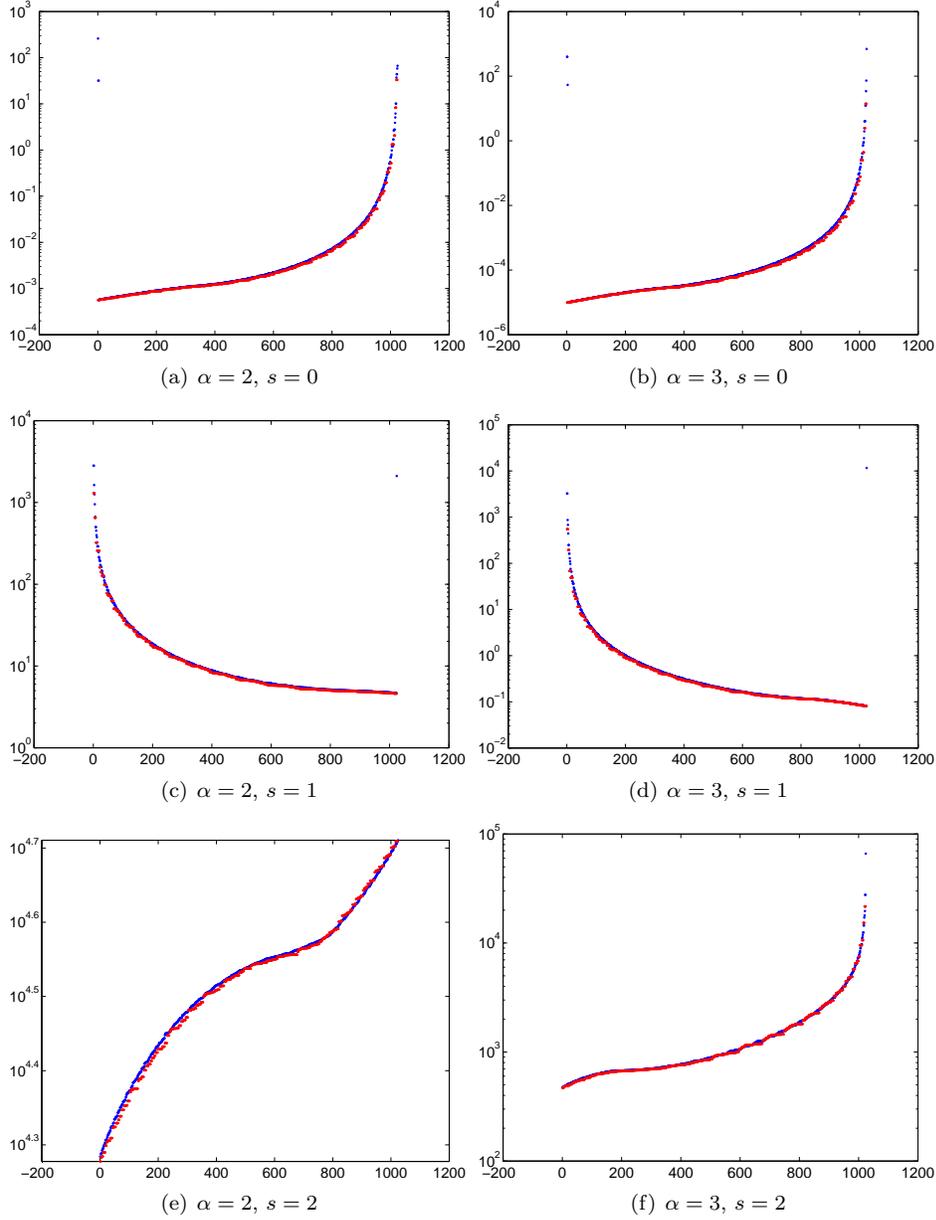
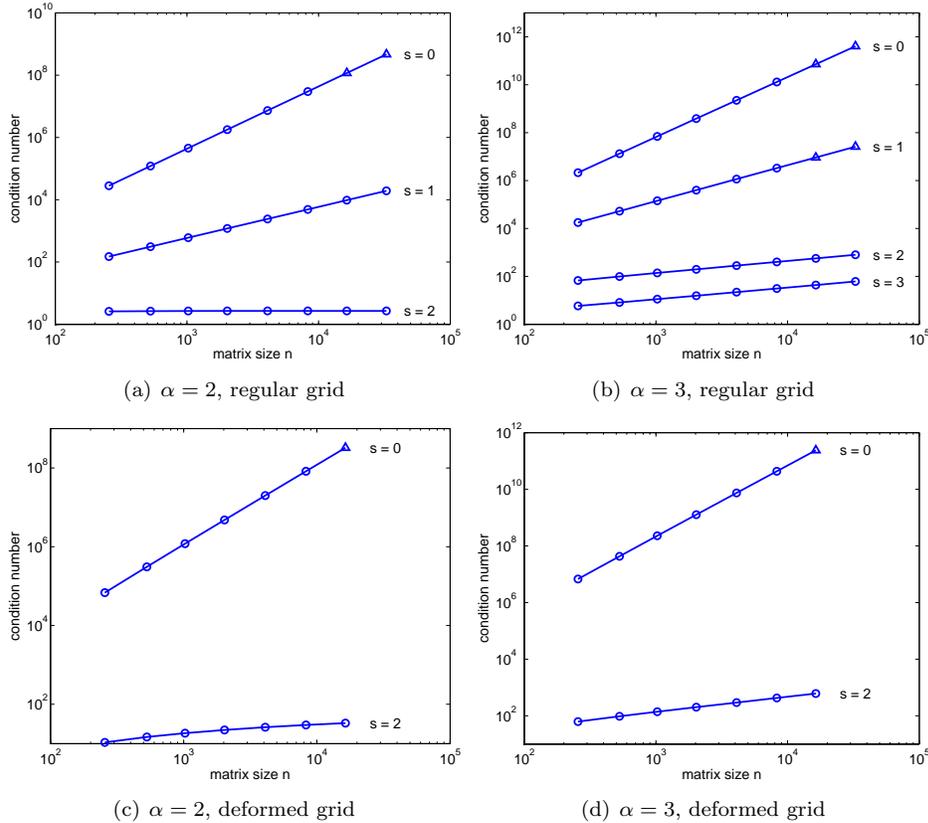


FIG. 7.4. Sorted eigenvalues of $\Phi_{\mathbf{n}}^{[s]}$ (blue) and sorted values of $(2\pi)^d \hat{\phi}_{\mathbf{n}}(2\pi\mathbf{j}/\mathbf{n})$ (red) for power-law kernel. Absolute values are plotted since the vertical axis is in log scale.

points. For some applications such as in statistics, this means some boundary information is discarded. With increasingly fine discretizations/dense sampling, statistical 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. Nevertheless, in the future we plan to investigate effective preconditioners that are full rank linear transformations.

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

Acknowledgments. The author is indebted to Mihai Anitescu and Michael Stein for their helpful discussions. 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. GRENNANDER AND G. SZEGÖ, *Toeplitz Forms and Their Applications*, American Mathematical Soc., 1984.
- [11] M. L. STEIN, *Fixed-domain asymptotics for spatial periodograms*, Journal of the American

- Statistical Association, 90 (1995), pp. 1277–1288.
- [12] ———, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, 1999.
 - [13] ———, *Equivalence of gaussian measures for some nonstationary random fields*, Journal of Statistical Planning and Inference, 123 (2004), pp. 1–11.
 - [14] 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.
 - [15] J. STEWART, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math, 6 (1976), pp. 409–434.
 - [16] H. WENDLAND, *Scattered Data Approximation*, Cambridge University Press, 2005.
 - [17] 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.

<p>The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”) under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.</p>
--