ACCURACY ANALYSIS OF THE PROXY POINT METHOD WITH APPLICATIONS TO SOME TOEPLITZ MATRICES*

MIKHAIL LEPILOV† AND JIANLIN XIA‡

Abstract. For some kernel matrices, low-rank approximations can be quickly obtained via analytic techniques. One important class of analytic methods is based on the use of proxy points. Accuracy analysis for various proxy point methods has often been heuristic in nature, other than for certain special kernels. For more general cases, the methods lack an explicit number or location of proxy points required to yield a particular approximation accuracy. In this work, we carry out new analysis of a proxy point method that is applicable to general complex-analytic kernels. An intuitive way of choosing proxy points is used to show explicit error bounds. Such bounds decay exponentially with regard to the number of proxy points. This also leads to convenient estimates of numerical ranks of relevant kernel matrices. To showcase the utility of this new analysis, we apply it to design a new sublinear-time hierarchically semiseparable approximation method for certain Toeplitz matrices, including ones that frequently arise from real-world applications. This allows, for example, inversion of such matrices with lower computational complexity compared with existing direct methods. Some extensions of these ideas are also discussed.

Key words. kernel matrix, proxy point method, approximation error, rank-structured matrix, Toeplitz matrix, sublinear complexity

AMS subject classifications. 65E99, 65F55, 65G99

1. Introduction.

Low-rank matrix approximation is a common task in many areas of mathematics, computation, and engineering. By a low-rank approximation of an $m \times n$ matrix A, we mean a decomposition $A \approx UV^T$, such that the column size of U is much smaller than m and n, and such that the approximation satisfies a certain accuracy requirement. Such an approximation allows even very large matrices to be represented, up to a certain accuracy, by matrices on which operations may be carried out significantly faster. Broadly speaking, we may divide (deterministic) low-rank approximation techniques into two classes: algebraic methods and analytic methods. Examples of algebraic methods include truncated SVDs and rank-revealing QR factorizations. Such methods are applicable to any matrix but are typically too slow to apply to large matrices. Examples of analytic methods include Taylor series expansions, interpolations, and proxy point methods. Such methods can be typically applied with little or essentially no computational cost, but require the matrix under consideration to be a kernel matrix defined by a kernel with desirable analytic properties.

In this work, we focus on one proxy point method which uses a set of proxy points to quickly construct basis matrices in low-rank approximations of kernel matrices. The root of this idea may be traced back to the fast multipole method (FMM) [6] and its variants [16, 29]. The error introduced by an analytic low-rank approximation is governed by analytic properties of the kernel in question. For proxy point methods, most accuracy analysis has relied on heuristics, arguing exponential convergence in the number of proxy points. Such analysis may be found in [15, 29, 13]. This is contrasted with the type of analysis carried out for the proxy point method in [28], which relates the number and location of proxy points to accurate error bounds of

*Submitted for review.

Funding: The research of Mikhail Lepilov was supported in part by NSF grant DMS-2038118. The research of Jianlin Xia was supported in part by NSF grant DMS-2111007.

†Department of Mathematics, Emory University, Atlanta, GA 30306 (mlepilo@emory.edu).

‡Department of Mathematics, Purdue University, West Lafayette, IN 47907 (xiaj@purdue.edu).

the low-rank approximation of the kernel matrix. However, the analysis in this last reference is only applicable to kernels of the form $k(x, y) = 1/(x - y)^d$ for $x, y \in \mathbb{C}$ and some $d \in \mathbb{N}$.

Here, we generalize this latter type of analysis to any one-dimensional complex-analytic kernel, bounding the errors of the proxy point method. Since for well-separated sets, many commonly-used kernels are complex-analytic in each variable on a region containing the set (holding the other variable constant), this analysis applies widely to one-dimensional kernels. We also show that, if such a kernel satisfies a certain univalence criterion, we may select proxy points like in [28] so as to get a rigorous accuracy bound that decays exponentially with respect to the number of proxy points. This also yields a convenient numerical rank estimate for the associated kernel matrix.

In addition to its theoretical utility, such analysis is useful, for example, when performing interpolative decompositions [14] of matrices using function approximation by a proxy point method. In [27], for example, even though the error introduced in the function approximation step is used to provide a bound for the error incurred in the interpolative decomposition step, the function approximation error is not itself studied. In this way, we hope that this work helps to fill gaps in the existing literature on proxy point methods.

As another illustration of the utility of this analysis, and following our work in [12], we introduce a new sublinear-time hierarchically semiseparable (HSS) approximation algorithm for certain Toeplitz matrices arising from univalent maps applied to regular grids. Such matrices appear, say, as covariance matrices of Gaussian processes like in [30]. A general overview of fast direct computations with Gaussian process covariance matrices is given in [1]. In this context, the kernel matrix is obtained from applying a given positive-definite kernel to a regular 1-dimensional grid. Existing rank-structured (HSS or similar) approximation construction schemes for such matrices carry at least $O(n)$ costs [1, 2, 26], so our scheme represents a substantial speedup over existing methods.

Additionally, we propose an extension of the ideas to kernels of dimension greater than one. We also verify that our proxy point approximation error bounds are illustrative of real-world performance by carrying out several numerical tests. Tests on the accuracy bounds for the proxy point low-rank approximation applied to different sets of points and kernels are given. We also carry out tests of the above HSS compression scheme applied to specific matrices.

The rest of this paper is structured as follows. In Section 2, we go over the proxy point method and perform the new proxy point error analysis for general complex-analytic functions. We also show how to use this analysis to guarantee the efficacy of proxy point approximations to some Toeplitz matrices. In Section 3, we review HSS matrix approximation. The HSS construction algorithm for the Toeplitz matrices under our consideration is detailed in Section 4. In Section 5, we perform some numerical tests. Finally, we suggest some extensions of this work in Section 6.

Throughout the paper, we use the following notation.

- Let $c \in \mathbb{C}$ and $r > 0$. Then $\mathcal{B}(c, r)$ denotes the open ball in $\mathbb{C}$ with center c and radius r, $\mathcal{O}(\mathcal{B}(c, r))$ denotes the set of holomorphic functions on $\mathcal{B}(c, r)$, and $\mathbb{D}$ denotes $\mathcal{B}(0, 1)$.
- Let $i \leq j$. Then $\{i : j\}$ denotes the set $\{i, i + 1, \ldots, j\}$.
- Let $k : F \times G \to \mathbb{C}$ be a function and $X \subseteq F, Y \subseteq G$ be totally-ordered finite subsets of size r and s, respectively. Then $k(X, Y) = (k(x_i, y_j))_{r\times s}$ means the $r \times s$ matrix with (i, j) entry $k(x_i, y_j)$, where x_i is the ith element of X

and y_j is the jth element of Y.
- Let C be an $m \times n$ matrix and $M \subseteq \{1:m\}, N \subseteq \{1:n\}$. Then by $A_{M\times N}$ we mean the $|M| \times |N|$ submatrix of A picked by the row index set M and column index set N.

2. Accuracy analysis for the proxy point method.

For a kernel matrix $k(X,Y)$ defined by the evaluation of a kernel function $k(x,y)$ at two finite sets $X, Y \subseteq \mathbb{C}$, the proxy point method is a simple yet powerful way for finding a low-rank approximation. The idea is to pick an appropriate set of proxy points $Z \subseteq \mathbb{C}$ based on ideas from, for example, potential theory, function interpolation, or an integral representation [11, 16, 27, 28, 29]. A low-rank approximation then carries the form

$$k(X,Y) \approx UV^T, \quad \text{with} \quad V = k(Y,Z). \tag{2.1}$$

(Alternatively, the form may be UV^T with $U = k(X,Z)$, depending on the context.) In particular, it is shown in [28] that one way of approximating $k(x,y) = 1/(x-y)^d$, for $d \in \mathbb{N}$, with an integral representation is to use the Cauchy integral formula. This would then allow us to take Z to be a set of quadrature points. In this work, we expand this idea to more general kernels.

2.1. Accuracy of kernel function approximation.

We follow the strategy of [28] but derive the approximation error results for general complex-analytic kernels $k(x,y)$. Let $D = \mathcal{B}(c,r)$ and $E = \mathcal{B}(c,R)$ be open balls in $\mathbb{C}$, with $r < R$. Let $X = \{x_j\}_{j=1}^m \subseteq D$ and $Y = \{y_j\}_{j=1}^n$ be finite sets, and let $k : \mathbb{C} \times Y \to \mathbb{C}$ be a function such that, for each $y \in Y$, $k(z,y)$ is an analytic function of z on E. Then, for each $x \in X, y \in Y$, by the Cauchy integral formula, we have

$$k(x,y) = \frac{1}{2\pi i}\int_C \frac{k(\zeta,y)}{\zeta - x} d\zeta, \tag{2.2}$$

where C is the boundary of an open ball with center c and radius $\sqrt{Rr}$. See Figure 2.1. Here, we chose the radius $\sqrt{Rr}$ heuristically from the study in [28] (that was conducted for a special kernel only). We will see shortly that precisely this choice of radius allows the analysis of this section to give a good accuracy bound for our proxy point approximation to $k(x,y)$.

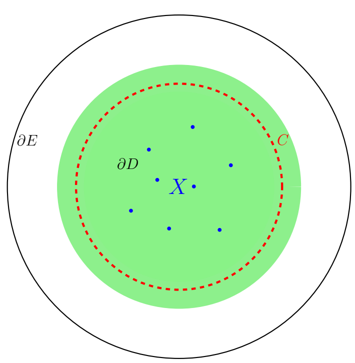


FIG. 2.1. *The contour C (dashed), the finite set X, and the boundaries of the open balls D and E. (The set Y is not pictured.) The shaded green region shows the disk F defined in Proposition 2.1, which is a compact region where the maximum of $k(z,y)$ (over all $y \in Y$) determines the approximation bound.*

Using the trapezoidal rule with p points to approximate (2.2), we then have

$$k(x,y) = \frac{\sqrt{Rr}}{p} \sum_{j=1}^{p} \left(\frac{1}{z_j - x} \right) \left(\omega^j k(z_j, y) \right) + \epsilon, \tag{2.3}$$

where

$$z_j = c + \sqrt{Rr}\omega^j, \quad \omega = e^{\frac{2\pi i}{p}}, \tag{2.4}$$

and the termwise error $\epsilon \in \mathbb{C}$ ideally has very small magnitude. The points z_j will serve as our proxy points. That is, $Z = \{z_j\}_{j=1}^{p}$ in (2.1). (2.3) provides a separable or degenerate expansion for $k(x,y)$. Hence, in this setup, for the low-rank approximation (2.3), we take

$$U = -\frac{\sqrt{Rr}}{p} \left(\frac{\omega^j}{x_i - z_j} \right)_{m \times p} = \frac{1}{p} \left(\frac{c - z_j}{x_i - z_j} \right)_{m \times p}, \quad V = k(Y, Z).$$

(Notice $m = |X|$ and $p = |Z|$.) This yields a proxy point low-rank approximation to $k(X,Y)$.

Note that in the setup for (2.2) above and in Figure 2.1, the geometric role of the set Y is not explicitly mentioned or shown, even though the exact locations of the points in Y are a key part of the analysis to follow. Indeed, in the existing literature, Y is typically assumed to be well separated from X and the distance between X and Y is used for certain bounds or heuristics [12, 28]. Here, the geometric information of the set Y factors into the above by determining the domain of analyticity (in z) of $k(z,y)$, which in turn determines R and r, given c. It also determines the disk F, shown in Figure 2.1 and used in to quantify the growth of k in the bound of Proposition 2.1.

For example, consider $k(x,y) = 1/(x-y)$ and $c = 0$. If the closest point to c in Y is $y_j = i (= \sqrt{-1})$, then we may pick $E = \mathcal{B}(0,1)$ and therefore obtain $R = 1$. On the other hand, if the closest point in Y to c is $y_j = 2$, we may pick $E = \mathcal{B}(0,2)$ and therefore $R = 2$. (r is chosen to be the smallest radius such that $\mathcal{B}(0,r)$ contains all the points in X. We will show in Proposition 2.1 that, if we maximize the ratio R/r, the analysis to follow provides the tightest proxy point approximation bound given a tame growth of $k(z,y)$ in z on F, which matches the heuristic explored in [28].

Our first goal is to find a bound for $|\epsilon|$ in (2.3). To do so, we give the following result. Its proof is based on classical techniques, including those for the proof of [18, Theorem 2.2], with some modifications for our context.

Proposition 2.1. *Let $D = \mathcal{B}(c,r)$ and $E = \mathcal{B}(c,R)$ be open balls in $\mathbb{C}$, with $r < R$, and let $X \subseteq D$ and Y be finite sets. Given p, let each z_j for $1 \le j \le p$ be defined as in (2.4), and let $k : \mathbb{C} \times Y \to \mathbb{C}$ be a function such that, for each $y \in Y$, $k(z,y)$ is an analytic function of z on E. Then for each $x \in X, y \in Y$, we have*

$$|\varepsilon| = \left| k(x,y) - \frac{\sqrt{Rr}}{p} \sum_{j=1}^{p} \left(\frac{1}{z_j - x} \right) \left(\omega^j k(z_j, y) \right) \right| \le \alpha \frac{\max_{z \in \partial F} |k(z,y)|}{(R/r)^{p/4} - 1},$$

where $\alpha = 2\frac{(R/r)^{1/4}}{(R/r)^{1/4} - 1}$ and $F = \mathcal{B}(c, r\,(R/r)^{3/4})$.

Proof. Fix $x \in X$ and $y \in Y$. First, by the parametrization $\gamma(t) = c + e^{ti}\sqrt{Rr}$ of the contour C in (2.2), we may write

$$k(x, y) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{k(c + \sqrt{Rr}e^{ti}, y)(i\sqrt{Rr}e^{ti})}{c + \sqrt{Rr}e^{ti} - x} dt.$$

Define $k_{x,y} : \mathcal{B}(0, \sqrt{R/r}) \setminus \overline{\mathcal{B}(0, \sqrt{r/R})} \to \mathbb{C}$ by

$$k_{x,y}(z) = \frac{k(c + z\sqrt{Rr}, y)(z\sqrt{Rr})}{c + z\sqrt{Rr} - x}. \tag{2.5}$$

Then

$$k(x, y) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{k_{x,y}(e^{ti}) i e^{ti}}{e^{ti}} dt = \frac{1}{2\pi i} \int_{\gamma_0} \frac{k_{x,y}(\zeta)}{\zeta} d\zeta \equiv a_0, \tag{2.6}$$

where $\gamma_0 = e^{ti}$ and a_0 denotes the 0th Laurent coefficient of $k_{x,y}$.

Consider the Laurent expansion of $k_{x,y}(z)$ at 0:

$$k_{x,y}(z) = \sum_{l=-\infty}^{\infty} a_l z^l, \tag{2.7}$$

which, by our assumption on k, is valid everywhere that $k_{x,y}$ is defined. We show that the sum of some Laurent coefficients $|a_l|$ can be used to bound the error incurred in applying the trapezoidal quadrature rule to (2.6) as in (2.3). In fact, (2.3), (2.4), and (2.5) mean

$$\varepsilon = k(x, y) - \frac{1}{p} \sum_{j=1}^{p} k_{x,y}(\omega^j).$$

For some steps below, we need a compact subregion on which the expansion (2.7) holds, because we will require absolute convergence, and because we will need to be able to bound certain quantities using values taken on its boundary. In particular, we define the compact annulus $A = \overline{\mathcal{B}\left(0, (R/r)^{1/4}\right)} \setminus \mathcal{B}\left(0, (r/R)^{1/4}\right)$. The radii here are chosen to scale with R/r. (Note that taking the radii to be $(R/r)^s$ and $(r/R)^s$ for any $s \in (1, 1/2)$ would work just as well, and it would just slightly alter the error bound). Now, since each $\omega^j \in A$ and A is compact, we have

$$\frac{1}{p} \sum_{j=1}^{p} k_{x,y}\left(\omega^j\right) = \frac{1}{p} \sum_{j=1}^{p} \sum_{l=-\infty}^{\infty} a_l \omega^{jl} = \frac{1}{p} \sum_{l=-\infty}^{\infty} a_l \sum_{j=1}^{p} \omega^{jl} = \sum_{l=-\infty}^{\infty} a_{pl},$$

where the last line follows from the fact that $\sum_{j=1}^{p} \omega^{jl}$ is p if l is a multiple of p and is 0 otherwise. Hence, by (2.6), we get

$$|\varepsilon| = \left| a_0 - \sum_{j=1}^{p} \frac{1}{p} k_{x,y}\left(\omega^j\right) \right| = \left| a_0 - \sum_{l=-\infty}^{\infty} a_{pl} \right| \leq \sum_{l=-\infty}^{-1} |a_{pl}| + \sum_{l=1}^{\infty} |a_{pl}|. \tag{2.8}$$

Next, we bound the magnitude of the Laurent coefficient a_{pl} of $k_{x,y}$ using R, r, p, and the maximum of k over $F = \mathcal{B}(c, r\,(R/r)^{3/4})$. (Recall that F is the green shaded region in Figure 2.1.) To do so, note that $k_{x,y}(z) = \frac{z-c}{z-x} k(\sqrt{Rr}z + c, y)$. Therefore, defining $F' = \overline{F} \setminus \mathcal{B}\left(c, r(R/r)^{1/4}\right)$, we have

$$\max_{z\in A} |k_{x,y}(z)| = \left(\max_{z\in F'} |(z-c)/(z-x)|\right)\left(\max_{z\in F'} |k(z,y)|\right)$$
$$= \frac{r(R/r)^{1/4}}{r((R/r)^{1/4}-1)} \max_{z\in F'} |k(z,y)| = \frac{\alpha}{2} \max_{z\in F'} |k(z,y)|.$$

Hence, for each $l \in \mathbb{Z}$ with $l \neq 0$, we have

$$|a_l| \leq \max\left(\left|\frac{1}{2\pi}\int_{|\zeta|=(R/r)^{1/4}} \frac{k_{x,y}(\zeta)}{\zeta^{l+1}} d\zeta\right|, \left|\frac{1}{2\pi}\int_{|\zeta|=(r/R)^{1/4}} \frac{k_{x,y}(\zeta)}{\zeta^{l+1}} d\zeta\right|\right)$$
$$\leq \frac{\max_{z\in A} |k_{x,y}(z)|}{\left((R/r)^{1/4}\right)^{|l|}} \leq \frac{\alpha}{2} \frac{\max_{z\in F'} |k(z,y)|}{\left((R/r)^{1/4}\right)^{|l|}}$$
$$\leq \frac{\alpha}{2} \frac{\max_{z\in F} |k(z,y)|}{\left((R/r)^{1/4}\right)^{|l|}} \leq \frac{\alpha}{2} \frac{\max_{z\in \partial F} |k(z,y)|}{\left((R/r)^{1/4}\right)^{|l|}},$$

where the last two inequalities follow from the maximum modulus principle and the fact that $k(z,y)$ is holomorphic in z on E. Combining this with (2.8), we get

$$|\varepsilon| \leq \frac{\alpha}{2}\left(2\sum_{l=1}^{\infty} \frac{\max_{z\in\partial F} |k(z,y)|}{\left((R/r)^{1/4}\right)^{pl}}\right) = \alpha \frac{\max_{z\in\partial F} |k(z,y)|}{(R/r)^{p/4}-1}.$$ ∎

For a thorough discussion of similar bounds, see [18]. However, note that the bounds given there and elsewhere in the numerical analysis literature do not simultaneously and explicitly bound the proxy point error for all values of an enclosed set X for each $y \in Y$. Hence, we may use our new result to bound the entrywise error for a kernel matrix. This feature will allow us to use this bound to guarantee applicability of the HSS construction method in Sections 4.1 and 4.2. This proof also provides justification for the heuristic, noted above and shown in [28] for the Cauchy kernel, that in the setup of this section we should pick C to have radius $\sqrt{Rr}$.

2.2. Accuracy for kernel matrix low-rank approximations. The termwise error bound for each element $k(x,y)$ allows us to obtain an absolute 2-norm error bound for the approximation to the matrix $k(X,Y)$. Furthermore, if k satisfies a univalence condition, we may obtain a relative 2-norm error bound for the matrix $k(X,Y)$ that guarantees exponential convergence in the number of proxy points.

PROPOSITION 2.2. *Let $D, E, X, Y, r, R, k, F, \alpha$, and each z_j for $1 \leq j \leq p$ be as in Proposition 2.1, and define*

$$U = \sqrt{Rr}\begin{pmatrix} \frac{1}{z_1-x_1} & \frac{1}{z_2-x_1} & \cdots & \frac{1}{z_p-x_1} \\ \frac{1}{z_1-x_2} & \frac{1}{z_2-x_2} & \cdots & \frac{1}{z_p-x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{z_1-x_l} & \frac{1}{z_2-x_l} & \cdots & \frac{1}{z_p-x_l} \end{pmatrix} \operatorname{diag}(\omega, \omega^2, \ldots, \omega^p),$$

$$V = \begin{pmatrix} k(y_1,z_1) & k(y_2,z_1) & \cdots & k(y_m,z_1) \\ k(y_1,z_2) & k(y_2,z_2) & \cdots & k(y_m,z_2) \\ \vdots & \vdots & \ddots & \vdots \\ k(y_1,z_p) & k(y_2,z_p) & \cdots & k(y_m,z_p) \end{pmatrix},$$

where $x_1, \ldots, x_l$ are the elements of X and $y_1, \ldots, y_m$ are the elements of Y. Then

$$\left\|k(X,Y) - UV^T\right\|_2 \leq lm\alpha \frac{\max_{y\in Y, z\in\partial F}(k(z,y))}{(R/r)^{p/4}-1}. \tag{2.9}$$

Furthermore, if in addition, $l \geq 2$, c is one of the points in X, and if $k(z,y)$ is bounded and univalent as a function of z on E for each $y \in Y$, then

$$\frac{\left\|k(X,Y) - UV^T\right\|_2}{\|k(X,Y)\|_2} \leq \frac{lm(1+\alpha\beta)}{(R/r)^{p/4}-1},$$

where $\beta = (R/r)^{3/4}\left(\frac{1+(r/R)}{1-(r/R)^{1/4}}\right)^2$.

Proof. The first result is obvious, since by Proposition 2.1,

$$\left\|k(X,Y) - UV^T\right\|_2 \leq \left\|k(X,Y) - UV^T\right\|_F \leq lm\alpha \frac{\max_{y\in Y, z\in\partial F}(k(z,y))}{(R/r)^{p/4}-1}. \tag{2.10}$$

To see the second result, we first bound the function maximum on the right-hand side of (2.10). The condition that $k(z,y)$ is univalent in z on E allows us to bound its growth away from c by the distance from c. In particular, we use the growth theorem for univalent maps on the unit disk that take the value 0 and have derivative equal to 1 at the origin; such maps are called regular univalent maps. We define a regular map $g_y(z)$ on the unit disk that takes values related to $k(z,y)$, use the growth theorem to bound its growth away from 0, and then use this to bound the growth of $k(z,y)$ away from c. More precisely, for each $y \in Y$, define the functions $f_y, h_y, g_y : \mathbb{D} \to \mathbb{C}$ by

$$\begin{aligned}
f_y(z) &= k(z,y), \\
h_y(z) &= Rz + c, \text{ and} \\
g_y(z) &= \frac{(f_y \circ h_y)(z) - (f_y \circ h_y)(0)}{(f_y \circ h_y)'(0)},
\end{aligned}$$

respectively. Then each g_y is regular and univalent, so by the growth theorem, we have $\frac{|z|}{(1+|z|)^2} \leq |g_y(z)| \leq \frac{|z|}{(1-|z|)^2}$. Thus, for $z \in \mathbb{D}$,

$$\frac{|z|\left|(f_y\circ h_y)'(0)\right|}{(1+|z|)^2} \leq |(f_y\circ h_y)(z) - (f_y\circ h_y)(0)| \leq \frac{|z|\left|(f_y\circ h_y)'(0)\right|}{(1-|z|)^2}. \tag{2.11}$$

Therefore, from the second inequality of Equation (2.11), for $z \in \partial\mathcal{B}\left(0,(r/R)^{1/4}\right)$, we have

$$\begin{aligned}
|(f_y\circ h_y)(z)| &\leq \frac{|z|\left|(f_y\circ h_y)'(0)\right|}{(1-|z|)^2} + |(f_y\circ h_y)(0)| \\
&\leq \frac{(r/R)^{1/4}}{(1-(r/R)^{1/4})^2}\left|(f_y\circ h_y)'(0)\right| + |(f_y\circ h_y)(0)| \\
&= \frac{(r/R)r^{1/4}}{(1-(r/R)^{1/4})^2}\left|f_y'(c)h_y'(0)\right| + |f_y(c)| \\
&\leq \frac{(r/R)^{1/4}}{(1-(r/R)^{1/4})^2}\left|f_y'(c)\right|\left|h_y'(0)\right| + |f_y(c)| \\
&= \frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}\left|f_y'(c)\right| + |f_y(c)|,
\end{aligned}$$

so $|k(z,y)| \leq \frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y)| + |k(c,y)|$ for all $z \in \partial F$ and $y \in Y$. (Here, $k_z(c,y)$ denotes the derivative of $k(z,y)$ as a function of z evaluated at c.) Defining $\gamma = \frac{1}{(R/r)^{p/4}-1}$, we thus have by Equation (2.10) that

$$\frac{\|k(X,Y)-UV\|_2}{\|k(X,Y)\|_2} \leq lm\gamma\left(\frac{\max_{y\in Y}\left(\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y)|+|k(c,y)|\right)}{\|k(X,Y)\|_2}\right)$$
$$\leq lm\gamma\left(\frac{\max_{y\in Y}\left(\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y)|+|k(c,y)|\right)}{\|k(X,Y)\|_1}\right)$$
$$= lm\gamma\left(\frac{\max_{y\in Y}\left(\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y)|+|k(c,y)|\right)}{\max_{y\in Y}\left(\sum_{j=1}^{l}|k(x_j,y)|\right)}\right). \tag{2.12}$$

Now, from the first inequality of Equation (2.11) and the fact that $l \geq 2$ and $c = x_{j_0}$ for some $1 \leq j_0 \leq l$, there exists a $1 \leq j_1 \leq l$ such that x_{j_1} is a distance r away from x_{j_0}. Hence, by the triangle inequality applied to $(f_y \circ h_y)(x_{j_0})$, $(f_y \circ h_y)(x_{j_1})$, and 0, we know since $\frac{(r/R)R|k_z(c,y)|}{1+(r/R)^2} \leq |k(x_{j_1},y) - k(x_{j_1},y)|$ that

$$(1/2)\frac{r\,|k_z(c,y)|}{(1+(r/R)^2)} \leq \max(|k(x_{j_1},y)|,|k(x_{j_2},y)|). \tag{2.13}$$

In particular, we then have the following bound on the denominator of (2.12):

$$\max_{y\in Y}\left(\max\left(|k(c,y)|,(1/2)\frac{r\,|k_z(c,y)|}{(1+(r/R)^2)}\right)\right) \leq \max_{y\in Y}\left(\sum_{j=1}^{l}|k(x_j,y)|\right). \tag{2.14}$$

Let $y_0 = \arg\max_{y\in Y}\left(\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y)|+|k(c,y)|\right)$. Combining (2.12), (2.13), and (2.14), we thus have

$$\frac{\left\|k(X,Y)-UV^T\right\|_2}{\|k(X,Y)\|_2} \leq lm\gamma\left(\frac{\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y_0)|+|k(c,y_0)|}{\max\left(|k(c,y_0)|,(1/2)\frac{r|k_z(c,y_0)|}{(1+(r/R)^2)}\right)}\right)$$
$$\leq lm\gamma\left(\frac{\frac{R^{3/4}r^{1/4}}{(1-(r/R)^{1/4})^2}|k_z(c,y_0)|}{(1/2)\frac{r|k_z(c,y_0)|}{(1+(r/R)^2)}}+\frac{|k_z(c,y_0)|}{|k_z(c,y_0)|}\right)$$
$$= lm\gamma(1+\alpha\beta).$$

The result then follows by our definition of γ. ∎

Hence, for a given 2-norm tolerance τ of the proxy point approximation to $k(X,Y)$, we only need to use $O(\log lm + \log\tau)$ proxy points, as long as the assumption on the analyticity of k holds, and as long as k grows sub-exponentially on the relevant domains. For a lot of functions k, this growth condition is not satisfied, and it is unclear *a priori* when the growth of k may be tame enough for Equation (2.9) to allow the feasibility of the proxy point method. But if k satisfies a certain univalence condition, then Proposition 2.2 guarantees slow growth and hence a relative error bound for the approximation to $k(X,Y)$ that decreases exponentially in the number of proxy points p.

2.3. Application: approximating some Toeplitz matrices. As a specific application of this, which we develop further in Section 4, we show that some off-diagonal blocks of Toeplitz matrices with certain kinds of Toeplitz vectors can be approximated efficiently by the proxy point method. To do so, we assume without loss of generality that n is divisible by 8. (This is purely for convenience of notation and is not a restriction on the applicability of the main ideas.) Let

$$T = \begin{pmatrix} t_0 & t_{-1} & \dots & t_{-(n-1)} \\ t_1 & t_0 & \dots & t_{-(n-2)} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_{n-2} & \dots & t_0 \end{pmatrix}, \tag{2.15}$$

be an $n \times n$ real- or complex-valued Toeplitz matrix whose entries are $t_i = f_1(i)$ for $-n \le i \le -1$ and $t_i = f_2(i)$ for $1 \le i \le n$, where $f_1 \in \mathcal{O}\left(\mathcal{B}\left(-n/2, n/2\right)\right)$ and $f_2 \in \mathcal{O}\left(\mathcal{B}\left(n/2, n/2\right)\right)$ are univalent. Such matrices occur in [3, 20], as well as in the Gaussian process literature mentioned in the introduction [1]. In [3], for example, f_1, f_2 are defined by $-f_1(z) = f_2(z) = (1-z)\log\left((z-1)/z\right) + (z+1)\log\left(z/(z+1)\right)$. Other commonly-used kernels that are univalent on the relevant domain include the Cauchy kernel and the Gaussian kernel with “large” (in this context, $O(n)$) length scale.

We may consider, for example, a certain off-diagonal block of T to be a kernel matrix corresponding to the kernel $k(x, y) = f_2(x - y)$:

$$T_{[n/2+n/4+1, n-n/4]\times[1,n/2]} = k(X, Y),$$

where $X = [n/2 + n/4 + 1, n - n/4]$ and $Y = [1, n/2]$. By our assumption on f_2, we are able to use the proxy point method with center $3n/4$ and radius $n/(4\sqrt{2})$ to get an approximation for $k(X, Y)$. Note that here, $R = n/2$ and $r = n/4$, so $R/r = 2$. Ensuring this separation between X and Y, and hence the analyticity of f_2, is the reason why we did not pick $X = [n/2 + 1, n]$ and attempt to approximate the entire bottom-left subblock of T. Using Equation (2.9), together with the function bound in Proposition 2.3 below, guarantees that we would need $O(\log n)$ proxy points to get a given approximation accuracy for large n.

Proposition 2.3. *Let f be holomorphic, bounded, and univalent on $\mathcal{B}\left(n/2, n/2\right)$. Then for $z \in \partial\mathcal{B}\left((n+1)/2, n/2 - 1\right)$,*

$$|f(z)| \le (n/2)^3 \left|f'\left(n/2\right)\right| + \left|f\left(n/2\right)\right|.$$

Proof. This is an adaptation of the proof of Proposition 2.2; we modify it here to explicitly relate n to the case of the Toeplitz matrix above. Define the functions $h : \mathbb{D} \to \mathbb{C}$ and $g : \mathbb{D} \to \mathbb{C}$

$$h(z) = (n/2)\, z + n/2, \text{ and}$$
$$g(z) = \frac{(f \circ h)(z) - (f \circ h)(0)}{(f \circ h)'(0)}.$$

Then g is schlicht, so by the growth theorem, we have $|g(z)| \le \frac{|z|}{(1-|z|)^2}$. Thus, for $z \in \partial\mathcal{B}\left(\left(\frac{2}{n}\right)\frac{n+1}{2}, \left(\frac{2}{n}\right)\left(\frac{n}{2} - 1\right)\right)$,

$$\left|(f \circ h)(z) - (f \circ h)(0)\right| \le \frac{|z|\left|(f \circ h)'(0)\right|}{(1-|z|)^2}.$$

Therefore, we have

$$
\begin{aligned}
\left|(f\circ h)(z)\right| &\le \frac{|z|\left|(f\circ h)'(0)\right|}{(1-|z|)^2}+\left|(f\circ h)(0)\right| \\
&\le (n/2)^2\left|(f\circ h)'(0)\right|+\left|(f\circ h)(0)\right| \\
&= (n/2)^2\left|f'(n/2)\,h'(0)\right|+\left|f(n/2)\right| \\
&\le (n/2)^2\left|f'(n/2)\right|\left|h'(0)\right|+\left|f(n/2)\right| \\
&= (n/2)^2(n/2)\left|f'(n/2)\right|+\left|f(n/2)\right|,
\end{aligned}
$$

so the result follows from the definition of h. ∎

Plugging the above bound into Equation (2.9) and using the fact that, in this case, $R/r = 2$, we get the following bound for the 2-norm error incurred using a proxy point approximation $k(X, Y) \approx UV$ with p points:

$$\|k(X,Y)-UV\|_2 \le \left(n^2/8\right)\left(\frac{2^{5/4}}{2^{1/4}-1}\right)\left(\frac{1}{2^{p/4}-1}\right)\left((n/2)^3\left|f'(n/2)\right|+\left|f(n/2)\right|\right).$$

Therefore, a given error tolerance requires $O(\log n)$ proxy points. In the following two sections, we further develop this idea to construct an HSS approximation to T with a computational cost that is sublinear in n.

3. Review of HSS matrix approximation.

Next, we review the data structure known as a hierarchically semiseparable (HSS) matrix form. Here we only give a brief outline; more details can be found in [25].

DEFINITION 3.1. *Let M be a matrix. Assume without loss of generality that M is square with row/column size n equal to a power of two, and let $L < \log_2(n)$. Recursively partition in two the set of row/column indices of M for a total of $2^L - 1$ subsets. Specifically, for each $0 \le l \le L$, partition $[1 : n]$ into the 2^l sets*

$$\mathcal{I}_l = \left\{\left[1 : \frac{n}{2^l}\right], \left[\frac{n}{2^l}+1 : \frac{n}{2^{l-1}}\right], \ldots, \left[(2^l-1)\frac{n}{2^l}+1 : n\right]\right\}.$$

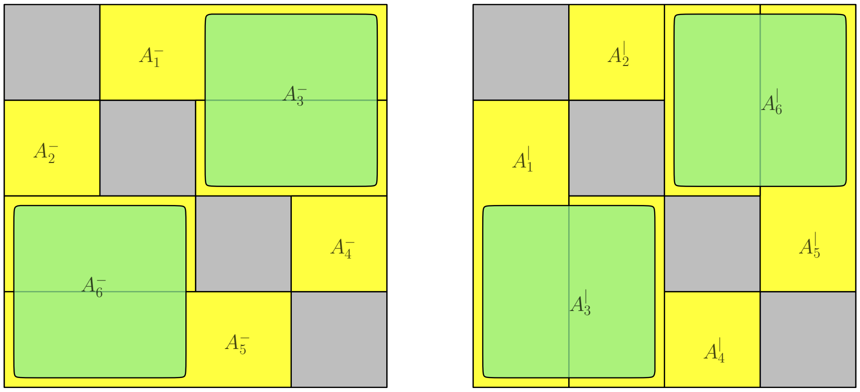
Let $\mathcal{I} = \bigcup_{j=0}^{L} \mathcal{I}_l$, and impose a partial order on $\mathcal{I}$ by set inclusion. We call $\mathcal{I}$ the L-level HSS index set *of M. Then its Hasse diagram $\mathcal{T}$ is a perfect binary tree, called the* HSS tree *of M. Now, for each $1 \le j \le 2^L - 1$, define $\mathbf{i}_j \in \mathcal{I}$ to be the element corresponding to the jth vertex of $\mathcal{T}$ in its postordered traversal. For each $1 \le j \le 2^L - 1$, define $M_j^- = M_{\mathbf{i}_j \times [1:n]\setminus \mathbf{i}_j}$ and $M_j^| = M_{[1:n]\setminus \mathbf{i}_j \times \mathbf{i}_j}$; these are called the* jth HSS block row *and* jth HSS block column*, respectively. (See Figure 3.1 for an example when $L = 2$.) The* HSS rank *of M is the maximum rank, over all $1 \le j \le 2^L - 1$, of M_j^- and $M_j^|$.*

An L-level HSS form *for M is a 6-tuple $\{\mathbf{D}, \mathbf{U}, \mathbf{R}, \mathbf{V}, \mathbf{W}, \mathbf{B}\}$, where:*

- *$\mathbf{U} = \{U_j\}_{1\le j\le 2^L-2}$, $\mathbf{V} = \{V_j\}_{1\le j\le 2^L-2}$, and $\mathbf{B} = \{B_j\}_{1\le j\le 2^L-2}$ are sets of matrices;*
- *$\mathbf{D} = \{D_j\}_{j\in\mathbf{I}}$ is a set of matrices, where $\mathbf{I}$ is the set of postordered indices of leaves of $\mathcal{T}$;*
- *and $\mathbf{R} = \{R_j\}_{j\in\mathbf{J}}$ and $\mathbf{W} = \{W_j\}_{j\in\mathbf{J}}$ are sets of matrices, where $\mathbf{J}$ is the set of postordered indices of vertices of $\mathcal{T}$ of depth at least two;*

such that

1. *$D_j = M_{\mathbf{i}_j\times\mathbf{i}_j}$ for $j \in \mathbf{I}$;*



(a) HSS block rows (b) HSS block columns

Fig. 3.1. *The HSS block rows and columns of M where $L = 2$. The labeled green blocks with rounded corners correspond to the HSS tree depth $l = 1$; the labeled yellow blocks with sharp corners correspond to the HSS tree depth $l = 2$.*

2. $M_{\mathbf{i}_j \times \mathbf{i}_{\mathrm{sib}(j)}} = U_j B_j V_{\mathrm{sib}(j)}^T$ *for* $1 \le j \le 2^L - 2$, *where* B_j *is full-rank and* $\mathrm{sib}(j)$ *is the postordered index of the sibling of* j*;*
3. *and* $U_j = \begin{pmatrix} U_{c_1(j)} R_{c_1(j)} \\ U_{c_2(j)} R_{c_2(j)} \end{pmatrix}$ *and* $V_j = \begin{pmatrix} V_{c_1(j)} W_{c_1(j)} \\ V_{c_2(j)} W_{c_2(j)} \end{pmatrix}$ *for* $1 \le j \le 2^L - 2$, *where* $c_1(j)$ *and* $c_2(j)$ *denote the postordered indices of the left and right children of the postordered* j*th vertex of* $\mathcal{T}$*, respectively.*

Collectively, all of the matrices mentioned in this definition are called *HSS generators* of M. Note that we can find generators whose sizes can all be bounded by the HSS rank of M [25]; this is the main point constructing the HSS form of M and the reason for the efficiency of HSS algorithms. Figure 3.2 illustrates the various relationships of the HSS generators of M.

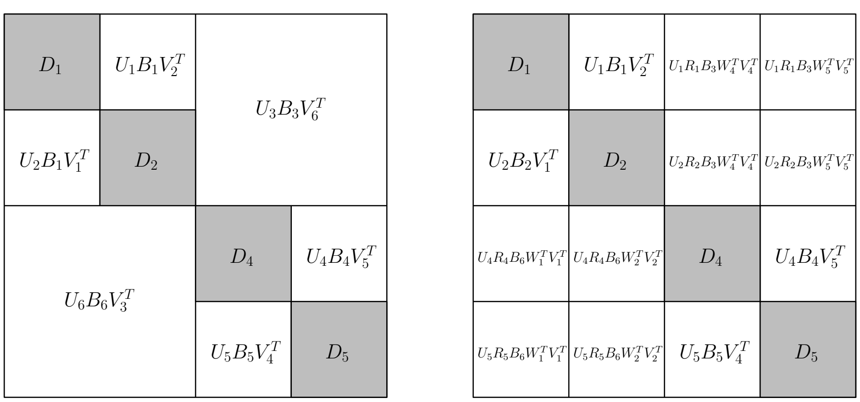


Fig. 3.2. *The HSS generator products of M placed into the blocks of M that they generate.*

Finally, we say M has *numerical HSS rank k with respect to a tolerance τ* if the numerical rank of M_j^- and $M_j^|$ with respect to a tolerance τ is at most k over all $1 \le j \le 2^L - 1$. We define an *L-level rank-k HSS approximation* of M to be an

L-level HSS form of M where we replace condition 2 in Definition 3.1 above with the following:

$2'$ $M_{\mathbf{i}_j \times \mathbf{i}_{\text{sib}(j)}} \approx U_j B_j V_{\text{sib}(j)}^T$ for $1 \le j \le 2^L - 2$, where B_j is a $k \times k$ matrix.

In the case that M is a Toeplitz matrix, for $1 \le j \le 2^L - 2$, existing methods of constructing any one of U_j, V_j, R_j, W_j, or B_j in general scale linearly in n, for at least some j.[22] In the next section, we outline an algorithm to construct any such generator with sublinear cost. This is useful depending on how the HSS form of M is subsequently used. For example, our method confers a speedup if only part of the output of a matrix-vector multiplication with M is needed.

4. Sublinear Toeplitz kernel HSS generator construction. In this section, we detail our sublinear HSS construction algorithm for Toeplitz matrices arising from univalent maps applied to a regular grid in one dimension. The combination of ideas necessary for this method was first explored in [12]. The approximation construction algorithm is detailed in 4.1 and 4.2. The analysis, started in Section 2, of the number of proxy points necessary for a good approximation is continued in 4.3.

To understand the utility of the new scheme, it is worth briefly reviewing existing Toeplitz methods. Over the past six decades, many algorithms have been devised that exploit the additional structure of Toeplitz matrices to perform various matrix operations faster than the counterpart "naive" algorithms applicable to general matrices. For example, so-called "fast" (faster than cubic time in the size of the matrix) and "superfast" (faster than quadratic time in the size of the matrix) algorithms have been devised to solve Toeplitz systems [9, 5, 7]. The central idea of such algorithms over the past few decades has become to apply fast Fourier transforms (FFTs) and solve the equivalent system in the frequency space. The resulting Cauchy-like matrix turns out to both be quickly solved by Gaussian elimination and to have low off-diagonal rank; hence, it can be quickly approximated by structured matrices [5, 17, 2]. Similarly, in digital signal processing, it has become well-known that the multiplication of Toeplitz convolution matrices with a given signal can be accelerated by applying FFTs and performing the equivalent operation in the frequency domain [10, 4].

After certain speedups that may be obtained using randomized techniques, the dominant cost in such structured matrix frequency-domain Toeplitz solution and multiplication algorithms becomes the application of FFTs [25, 14, 26]. Hence, in theory, general HSS algorithms can potentially achieve a speedup for matrix operations whenever a matrix is both Toeplitz and has low off-diagonal rank before the application of FFTs [25]. In such algorithms, the dominant cost becomes the construction of the structured approximant; thus, bringing this cost down is a worthwhile endeavor. In this work, we show that for Toeplitz matrices whose Toeplitz vector is generated by a univalent map applied to the positive integers, we are able to reduce the HSS construction time cost from $O\left(r^2 n\right)$ [22, 23] to $O\left(\log^5(n)\right)$ in the size n of a square matrix with off-diagonal rank bound r. While the new algorithm is less widely applicable, it may nevertheless be applied to certain important classes of matrices, such as those arising as covariance matrices of Gaussian processes [1, 30], or from a convolution of a digital signal with a large Gaussian filter [4]. In addition, since this new scheme does not rely on Fourier space representation, it has the advantage of preserving the rank structure of any diagonal or rank-structured summand that may be added to the Toeplitz matrix, such as when localizing eigenvalues [21, 19].

The first key idea in our new construction scheme is the use of the proxy point method in the process of obtaining an interpolative decomposition (also known as skeletonization) of the HSS blocks, as was done previously in [15]. The second key idea is the reuse of the resulting approximate basis matrix factors for all the HSS blocks at a given HSS depth, as was done previously in [12]. Here is where we use our new analysis from Section 2 to guide the process of obtaining these approximate basis factors, as well as to understand when the construction scheme is applicable. In the case that the proxy point method is used to approximate off-diagonal blocks of Toeplitz matrices with Toeplitz vector generated by a complex-analytic univalent map, this error is then shown to increase slowly enough in n to allow our construction algorithm to be performed in sublinear time relative to n. While we do not perform an operation count to justify this here, since our algorithm is almost identical to the one outlined in [12], the analysis from Section 5 of that paper applies to the algorithm outlined in this section.

Let T be a Toeplitz matrix defined by the Toeplitz vector

$$(t_{-(n-1)}, t_{-(n-2)}, \ldots, t_{-1}, t_0, t_1, \ldots, t_{n-2}, t_{n-1}),$$

as in Equation (2.15), and similarly define f_1 and f_2 as in Section 2. To more easily illustrate the application of this method, we will deal with the symmetric case $t_{-i} = t_i$ (so $f_1(-i) = f_2(i)$) for $i = 0, \ldots, n-1$; define $f(z) = f_1(-z)$. The non-symmetric case is handled similarly (see Section 4.2). Since we are constructing generators for approximations to the off-diagonal blocks of T, we may again assume without loss of generality that $t_0 = 0$. Furthermore, since this algorithm is meant to apply to large matrices, we may assume that n is a power of two greater than 8.

4.1. Constructing the HSS row generators. Let $L \leq \log_2(n) - 2$ be the number levels in the desired HSS approximation to T. Let r be a bound for the numerical HSS rank of T; we assume specifically that r is $O(\log n)$. The analysis in Section 4.3 can actually be used to give a bound for r. In particular, we can show that r is $O(\log^2 n)$; see Section 6.

For each $1 \leq i, j \leq n$ with $i \neq j$, we have $T_{i,j} = f(|j-i|)$. Hence, we may consider an HSS block T_j^- to be the kernel matrix $k(\mathbf{i}_j, [1:n] \setminus \mathbf{i}_j)$, where k is defined by $k(x,y) = f(|x-y|)$. Directly finding a low-rank factorization for T_j^-, for example as when $j = 1$ in the first step in the HSS construction algorithm in [25], is already prohibitively expensive with at least $O(n)$ flops. Instead, we may follow a similar list of steps as in [12, Section 3.2]:

- If j is not leaf of $\mathcal{T}$, we assume we have performed this list of steps on its children $c_1(j)$ and $c_2(j)$ to obtain sets of indices $\mathbf{i}'_{c_1(j)}, \mathbf{i}'_{c_2(j)} \subseteq \mathbf{i}_j$. If j is a leaf, we define $c_1(j) = c_2(j) = j$ and $\mathbf{i}'_j = \mathbf{i}_j$. Then, we define $\bar{\mathbf{i}}'_j = \mathbf{i}'_{c_1(j)} \cup \mathbf{i}'_{c_2(j)}$ and apply a proxy point approximation to $\left(T_j^-\right)_{\bar{\mathbf{i}}'_j \times [1:n-|\mathbf{i}_j|]}$. However, since we only assumed that f is analytic on $\mathcal{B}(n/2, n/2)$, by Equation (2.9), the ratio R/r in this case could be as large $1/n$, and therefore the number of proxy points p required to obtain a reasonably good approximation may be prohibitively large. Hence, we first separate $\mathbf{i}_j$ into the "near-field" and "far-field" subsets $\hat{\mathbf{i}}_j$ and $\tilde{\mathbf{i}}_j = \mathbf{i}_j \setminus \hat{\mathbf{i}}_j$, respectively, where $\hat{\mathbf{i}}_j$ is the subset of $\mathbf{i}_j$ consisting of its first and last $|\mathbf{i}_j|/4$ values, respectively, ordered the usual way. We then define $\hat{\mathbf{i}}'_j = \hat{\mathbf{i}}_j \cap \bar{\mathbf{i}}'_j$, $\tilde{\mathbf{i}}'_j = \tilde{\mathbf{i}}_j \cap \bar{\mathbf{i}}'_j$, $T_{j,1}^- = k\left(\hat{\mathbf{i}}'_j, [1:n] \setminus \mathbf{i}_j\right)$, and $T_{j,2}^- = k\left(\tilde{\mathbf{i}}'_j, [1:n] \setminus \mathbf{i}_j\right)$; and we apply a proxy point approximation to only the far-field subblock: $T_{j,2}^- \approx \tilde{U}_j \tilde{V}_j$. For this approximation, we use a circular contour with center $(1/2)\left(\min(\mathbf{i}_j) + \max(\mathbf{i}_j)\right)$ and radius $(\sqrt{2}/2)\left(\max(\mathbf{i}_j) - \min(\mathbf{i}_j) + 1\right)$ to obtain $R/r = 2$. (See Figure 4.1 and Figure 4.2.)

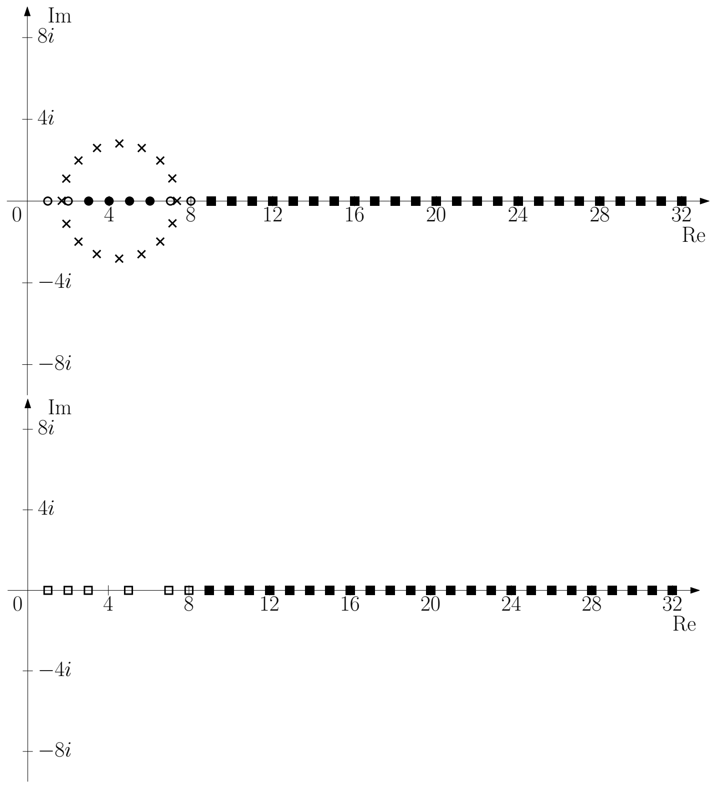


FIG. 4.1. *Top: near-field points* $\tilde{\mathbf{i}}_1'$ *(*$\circ$*), far-field points* $\hat{\mathbf{i}}_1'$ *(*$\bullet$*), proxy points (*$\times$*), and the points* $[9:32]$ *(*$\blacksquare$*) involved in the approximation of the leaf HSS block* $T_1^-|_{\tilde{\mathbf{i}}_1' \times [1:n-|\mathbf{i}_1|]} = k\,([1:8],[9:32])$ *for a matrix of size* $n = 32$*, number of HSS levels* $L = 2$*, and number of proxy points* $p = 16$*. Bottom: the resulting index set* $\mathbf{i}_1'$ *(*$\square$*). (These are "cartoon illustrations" and are not actual results from such an approximation applied to a subblock of an actual matrix* T*.)*

We thus have

$$\left(T_j^-\right)|_{\tilde{\mathbf{i}}_j' \times [1:n-|\mathbf{i}_j|]} = \Pi_i \begin{pmatrix} T_{j,1}^- \\ T_{j,2}^- \end{pmatrix} = \Pi_i \begin{pmatrix} I & 0 \\ 0 & \tilde{U}_i \end{pmatrix} \begin{pmatrix} T_{j,1}^- \\ \tilde{V}_i \end{pmatrix},$$

where Π_i is a permutation matrix.

- Next, we find a strong rank-revealing QR factorization

$$\tilde{U}_j = \overline{U}_j \left(\Pi_j'^T \tilde{U}_j\right)|_{[1:r]\times[1:p]},$$

where $\overline{U}_j = \begin{pmatrix} I & E_j \end{pmatrix}^T$ and Π_j' is a permutation matrix. In theory, any rank-revealing QR factorization may suffice, but in practice the SRRQR factorization results in greater numerical stability when working with E_j (and hence U_j); see [8] for details. We then have

$$T_{j,2}^- \approx \overline{U}_j \left(\Pi_j'^T \tilde{U}_j\right)_{[1:r]\times[1:p]} \tilde{V}_j \approx \overline{U}_j \left(\Pi_j'^T T_{j,2}^-\right)_{[1:r]\times[1:n]\setminus \mathbf{i}_j},$$

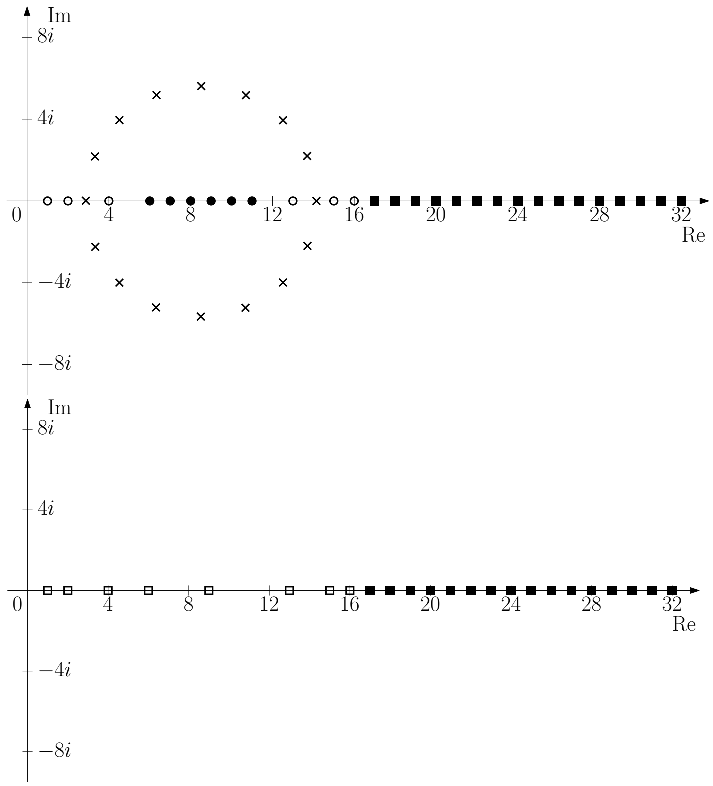


FIG. 4.2. *Top: near-field points* $\tilde{\mathbf{i}}_3'$ *(*$\circ$*), far-field points* $\hat{\mathbf{i}}_3'$ *(*$\bullet$*), proxy points (*$\times$*), and the points* $[17:32]$ *(*$\blacksquare$*) involved in the approximation of height-2 HSS block* $T_3^-|_{\tilde{\mathbf{i}}_3'\times[1:n-|\mathbf{i}_3|]} = k(\tilde{\mathbf{i}}_j', [17:32])$ *for a matrix of size* $n = 32$*, number of HSS levels* $L = 2$*, and number of proxy points* $p = 16$*. Bottom: the resulting index set* $\mathbf{i}_3'$ *(*$\square$*). (As noted in Figure 4.1 above, these are "cartoon illustrations" and are not reflective of actual numerical results.)*

so

$$\begin{aligned}\left(T_j^-\right)|_{\tilde{\mathbf{i}}_j'\times[1:n-|\mathbf{i}_j|]} &\approx \Pi_j \begin{pmatrix} T_{j,1}^- \\ T_{j,2}^- \end{pmatrix} \\ &\approx U_j \begin{pmatrix} \left(\Pi_j^T T_j^-|_{\tilde{\mathbf{i}}_j'\times[1:n]\setminus\mathbf{i}_j}\right)|_{\left[1:\left|\hat{\mathbf{i}}_j'\right|\right]\times[1:n-|\mathbf{i}_j|]} \\ \left(\Pi_j^T T_j^-|_{\tilde{\mathbf{i}}_j'\times[1:n]\setminus\mathbf{i}_j}\right)|_{\left[\left|\hat{\mathbf{i}}_j'\right|+1:\left|\hat{\mathbf{i}}_j'\right|+r\right]\times[1:n-|\mathbf{i}_j|]} \end{pmatrix} \\ &= U_j T^-|_{\mathbf{i}_j'\times[1,n]\setminus\mathbf{i}_j},\end{aligned}$$

where $\mathbf{i}_j' \subseteq \mathbf{i}_j$ is of size $\left|\hat{\mathbf{i}}_j'\right| + r$ and

$$U_j = \Pi_j \begin{pmatrix} I & 0 \\ 0 & \Pi_j' \begin{pmatrix} I \\ E_j \end{pmatrix} \end{pmatrix}.$$

Now, if j is a leaf, this last display is precisely the HSS generator. If j is not a leaf, we set $R_{c_1(j)} = U_j|_{\left(\mathbf{i}_j'\cap\mathbf{i}_{c_1(j)}\right)\times\left[1:\left|\tilde{\mathbf{i}}_j'\right|+r\right]}$ and $R_{c_2(j)} = U_j|_{\left(\mathbf{i}_j'\cap\mathbf{i}_{c_2(j)}\right)\times\left[1:\left|\tilde{\mathbf{i}}_j'\right|+r\right]}$.

4.2. Constructing the remaining HSS generators. Now, note that for each j at the leaf level in $\mathcal{T}$, each matrix $\left(T_j^-\right)_{(\mathbf{i}'_{c_1(j)}\cup\mathbf{i}'_{c_2(j)})\times[1:n-|\mathbf{i}_j|]}$ used to obtain the generator U_j yields the same U_j regardless of the specific value of j. Hence, $\mathbf{i}'_j$ is the same for any leaf-level j. Therefore we can show by induction on L that for each j at the same depth of $\mathcal{T}$, U_j and $\mathbf{i}'_j$ are the same. This shows that *we only need to perform the above steps once at each depth of* $\mathcal{T}$ to obtain all the HSS row generators U_j for a leaf-level j and R_j for j with $\operatorname{depth}(j) \le L-2$. Furthermore, because the above steps do not depend on the specific function $k(x, y) = f(|x-y|)$ as long as f satisfies the analyticity condition, *the above steps also construct the HSS column generators* V_j and W_j. So, we set $V_j = U_j$ for a leaf-level j and $W_j = R_j$ for j with $\operatorname{depth}(j) \le L-2$. This last fact shows why our assumption that $f_1 = f_2$ at the beginning of this section confers no loss of generality. Finally, for each $j \in \mathcal{T}$, we set $B_j = T_{\mathbf{i}'_j\times\mathbf{i}'_{\operatorname{sib}(j)}}$.

So far, we have not mentioned how many proxy points are required for the far-field approximation at each level in the above construction method; we will explore this issue in the next section. We note here, however, that if the number of proxy points is $O(\log n)$, then the flop count of this method is the same as that of the method in [12], for a total of $O(\log^5 n)$ flops. We will show that this is indeed the case in the next section whenever f satisfies the univalent condition in Proposition 2.3.

4.3. Number of proxy points required. First, we fix some notation: let $\mathcal{T}, \mathcal{I}$ be the HSS tree and HSS index set of T, respectively, and let $j \in \mathcal{T}$ have corresponding index set $\mathbf{i}_j \in \mathcal{I}$. We define $\hat{\mathbf{i}}_j$ to be the subset of $\mathbf{i}_j$ missing its least and greatest $|\mathbf{i}_j|/4$ elements, ordered the usual way. We also define $\tilde{T}_n^{j,p}$ to be the p-point proxy point approximation (in the first variable) to the subblock $T|_{\hat{\mathbf{i}}_j\times[1:n]\setminus\mathbf{i}_j} = k(\hat{\mathbf{i}}_j, [1:n]\setminus\mathbf{i}_j)$ with center $(1/2)\left(\min(\mathbf{i}_j)+\max(\mathbf{i}_j)\right)$ and radius $(1/2)\left(\max(\mathbf{i}_j)-\min(\mathbf{i}_j)+1\right)$.

Next, we show with Example 1 that for general $f \in \mathcal{O}\left(\mathcal{B}\left(n/2, n/2\right)\right)$, this approximation need not have good convergence properties. This corresponds to the case that f grows rapidly away from $n/2$; this corresponds to the case that the function bound in Equation (2.9) is large.

EXAMPLE 1. *For $n \ge 8$, let $T_n \in \mathbb{R}_{n\times n}$ have entries $(T_n)_{i,j} = \cos\left((\pi/4)\,|j-i|\right)$, and let $\mathcal{I}_n = \{\mathbf{i}_{n,1}, \mathbf{i}_{n,2}, \mathbf{i}_{n,3}\}$ be its one-level HSS index set, indexed the usual way. Then the associated function $f(z) = f_1(z) = f_2(z) = \cos\left((\pi z)/4\right)$ is holomorphic on $\mathcal{B}\left(n/2, n/2\right)$. Table 4.1 shows the minimum number of points p required for $\tilde{T}_n^{1,p}$ to approximate $(T_n)\,|_{\hat{\mathbf{i}}_{n,1}\times[1:n]\setminus\mathbf{i}_{n,1}}$ to a given tolerance. Note that even for such small matrix sizes and large tolerance, the number of proxy points required already scales linearly with n. It is also worth noting that the rank of T_n is at most 8 for all n and every off-diagonal block.*

TABLE 4.1
The size n of the matrix T_n and the minimum number of proxy points p required to attain $\left\|(T_n)\,|_{\hat{\mathbf{i}}_{n,1}\times[1:n]\setminus\mathbf{i}_{n,1}} - \tilde{T}_n^{1,p}\right\|_F < 10^{-6}$.

n	16	24	32	40	48	56	64	72	80
p	21	27	34	39	47	53	59	65	72

The poor performance in Example 1 makes sense in light of Proposition 2.1: for each $y \in Y = [1:n]\setminus\mathbf{i}_{n,1} = [n/2+1:n]$, $k(z, y) = f\left(|z-y|\right)$ must not be too large in absolute value for all $z \in \partial F = \partial \mathcal{B}\left(n/4 + 1/2, \sqrt[4]{8}n/8\right)$ in order for a small number of proxy points to be sufficient. But in this case, we may observe that, if $y = n/2 + 1$, the maximum of $f(|y-z|) = \cos((\pi/4)|y-z|)$ along $z \in \partial F$ grows exponentially in n. In particular, even though cosine is bounded on the real line, its growth along the one-dimensional line $z(t) = t + it$ (for real t) is exponential. Hence, the growth of p with respect to n shown in Table 4.1 gives evidence that f with large values on $\mathcal{B}(n/2, n/2)$ may require a lot of proxy points for an accurate approximation.

On the other hand, if f is bounded on the real line and univalent on $\mathcal{B}(n/2, n/2)$, we show in Example 2 that we do seem to have good proxy point convergence for the HSS approximation outlined in Sections 4.1 and 4.2.

EXAMPLE 2. *For $n \geq 8$, let $T_n \in \mathbb{R}_{n\times n}$ have entries $(T_n)_{i,j} = \cos((\pi|j-i|)/n)$. Then the associated function $f(z) = f_1(z) = f_2(z) = \cos((\pi z)/n)$ is univalent on $\mathcal{B}(n/2, n/2)$ and bounded on the real line. Table 4.2 shows the minimum number of proxy points required for the sublinear HSS construction method to yield a given approximation tolerance for the topmost HSS row block.*

TABLE 4.2
The size n of the matrix T_n and the minimum value of p such that the L-level HSS approximation constructed in Sections 4.1 and 4.2 with p proxy points approximates the topmost HSS block of T_n to a relative Frobenius norm error 10^{-10}.

n	2048	4096	4096	8192	8192	8192	16384	16384	16384	16384
L	1	1	2	1	2	3	1	2	3	4
p	26	27	27	28	28	28	28	28	28	28

Example 2 gives numerical evidence that the proxy-point approximation has good enough convergence properties to be used in practice, even despite global HSS error accumulation. We now show that good proxy point convergence is true for general univalent f in this context, as well as in the general case of Proposition 2.2.

LEMMA 4.1. *Let $\mathcal{I}$ be an HSS index set for an $n \times n$ matrix, where n is a power of 2; let $\mathbf{i} \in \mathcal{I}$; and let l be the height of $\mathbf{i}$. Define $k(x,y) = f(|y-x|)$ for some $f \in \mathcal{O}(\mathcal{B}(n/2, n/2))$; let $x \in \hat{\mathbf{i}}$; let $y \in [1:n] \setminus \mathbf{i}$; and let $p \in \mathbb{N}$. Then*

$$\left| k(x,y) - \sum_{j=1}^{p} \left(\frac{\left(\sqrt[4]{8}\right) 2^{l-1}}{p} \right) \frac{\omega^j k(z_j, y)}{z_j - x} \right| < 14 \frac{\max_{z\in\partial F}(|f(y-z)|)}{2^{p/4} - 1}, \tag{4.1}$$

where $z_j = c + \left(\sqrt[4]{8}\right) 2^{l-1}\omega^j$, F is the open ball with center c and radius $\left(\sqrt[4]{8}\right) 2^{l-1}$, and $c = (1/2)(\max(\mathbf{i}) - \min(\mathbf{i}) + 1)$.

Proof. This is a straightforward application of Proposition 2.1, where we set $X = \hat{\mathbf{i}}$; $Y = [1:n] \setminus \mathbf{i}$; and D and E to be the open balls with center c and radii $R = 2^{l-1}$ and $r = 2^l$, respectively. We thus get $\alpha = 2\sqrt[4]{2}/(\sqrt[4]{2} - 1) < 14$. ∎

Therefore, by the maximum modulus principle and Lemma 4.1, if we find that $\max_{z\in\partial\mathcal{B}((n+1)/2, n/2-1)} |f(z)|$ has a sufficiently small bound with respect to n, we would need only $O(\log n) + |\log \epsilon|$ proxy points to obtain an entrywise proxy point approximation with tolerance ϵ at every height of the HSS tree. But note that we obtained exactly such a bound in Section 2 in Proposition 2.3 if f is univalent on $\mathcal{B}(n/2, n/2)$, and if f and its derivative does not grow too quickly quickly with respect to n along the real axis. Hence, we obtain the following absolute error bound for the proxy point approximation of an off-diagonal "far-field" row block:

COROLLARY 4.2. *Let $T \in \mathbb{C}$ be the $n \times n$ matrix with entries $T_{i,j} = f(|j-i|)$, where $f \in \mathcal{O}(\mathcal{B}(n/2, n/2))$ is injective on $\mathcal{B}(n/2, n/2)$. Let $\mathcal{I}$ be the HSS index set of T, and let $\mathbf{i}_j \in \mathcal{I}$. Then*

$$\left\| T|_{\hat{\mathbf{i}}_j \times [1:n]\setminus \mathbf{i}_j} - \tilde{T}^{j,p} \right\|_F \leq \left(\frac{7n^2}{2^{p/4+1} - 2} \right) \left((n^3/8) \left| f'(n/2) \right| + \left| f(n/2) \right| \right).$$

Proof. By Lemma 4.1, the maximum modulus principle, and Proposition 2.2, in that order, we have that for each $1 \leq u \leq |\hat{\mathbf{i}}_j|$ and $1 \leq v \leq |[1:n] \setminus \mathbf{i}_j|$,

$$\begin{aligned} \left| \left(T|_{\hat{\mathbf{i}}_j \times [1:n]\setminus \mathbf{i}_j} \right)_{u,v} - \left(\tilde{T}^{j,p} \right)_{u,v} \right| &< 14 \frac{\max_{y \in [1:n]\setminus \mathbf{i},\, z \in \partial F} \left(|f(y-z)| \right)}{2^{p/4} - 1} \\ &\leq 14 \frac{\max_{z \in \partial \mathcal{B}((n+1)/2, n/2-1)} \left(|f(z)| \right)}{2^{p/4} - 1} \\ &\leq \frac{14}{2^{p/4} - 1} \left((n^3/8) \left| f'(n/2) \right| + \left| f(n/2) \right| \right). \end{aligned}$$

Since $|\hat{\mathbf{i}}_j|, |[1, n] \setminus \mathbf{i}_j| \leq \frac{n}{2}$, the result follows by summing over all u and v. ∎

Thus, to obtain a given proxy point approximation tolerance ϵ for any level, we need $O(\log n) + O(|f(n/2)|) + O(|f'(n/2)|) + O(|\log \epsilon|)$ proxy points. In practice, f and its derivative are often bounded on the real line, as in Examples 3 and 4 below.

5. Discussion and numerical tests. First, we note that, although injectivity of f as defined in the previous section is a sufficient condition, it is not strictly necessary in practice to enable the use of our sublinear Toeplitz HSS construction algorithm. The point of the injectivity criterion is simply to allow, using Proposition 2.2, a sufficiently slow growth bound for f that depends only on its radius of analyticity. However, functions f that are not univalent on the relevant region can also grow sufficiently slowly in order for their related construction algorithm outlined in the previous section to work on the related Toeplitz matrix. Example 3 illustrates this.

EXAMPLE 3. *For $n \geq 8$, let $T_n \in \mathbb{R}_{n \times n}$ have entries $(T_n)_{i,j} = (|j-i| - n/2)^2$, so the associated function $f(z) = f_1(z) = f_2(z) = (z - n/2)^2$ is not univalent on $\mathcal{B}(n/2, n/2)$. Table 5.1 lists the relative approximation tolerance for various HSS approximations of T from Sections 4.1 and 4.2. (For the scheme as outlined there, we set the maximum off-diagonal rank to $r = 28$. This is sufficient, since each matrix involved has a relative off-diagonal numerical rank of 3 with respect to the tolerance 10^{-14}.) Note that relatively small values of p result in a good approximation.*

On the other hand, the conditions of Proposition 4.2 provides a wide class of functions for which our sublinear HSS construction algorithm is guaranteed to work.

EXAMPLE 4. *Since $f_1(z) = n/z$ and $f_2(z) = -n/z$ are univalent on $\mathcal{B}(n/2, n/2)$, the method from Sections 4.1 and 4.2 should work to find the HSS generators of T_n, the Cauchy kernel matrix evaluated at n equidistant points in $[-1, 1]$, in sublinear time. Table 5.2 lists the relative approximation tolerance for various HSS approximations to the matrix $T_n \in \mathbb{R}_{n \times n}$ with off-diagonal values $(T_n)_{i,j} = n/(j-i)$ and diagonal values equal to 0. The maximum relative off-diagonal numerical rank r is also listed; for this experiment, we set $r = 28$ for each matrix. It is worth noting that the accuracy bound*

TABLE 5.1
The relative Frobenius norm errors of the L-level HSS approximation to T_n from Sections 4.1 and 4.2 using p proxy points. The top and bottom tables show the errors using 32 and 48 proxy points at each level, respectively.

n	2048	2048	8192	8192	16384	16384
L	2	4	4	6	6	7
rel. err. (e10^{-13})	5.4863	2.9697	7.7119	3.3541	6.9370	3.4362

n	2048	2048	8192	8192	16384	16384
L	2	4	4	6	6	7
rel. err. (e10^{-13})	2.0441	9.3656	3.2532	1.0675	2.9239	1.0933

given in [28] may also be used in lieu of Proposition 2.1 for this particular kernel matrix to indicate applicability of the scheme from Section 4.

TABLE 5.2
The relative Frobenius norm errors of the L-level HSS approximation to T_n from Sections 4.1 and 4.2 using p proxy points, as well as the numerical HSS rank r of T_n with tolerance 10^{-14}. Again, the top and bottom tables show the errors using 32 and 48 proxy points at each level, respectively.

n	2048	2048	8192	8192	16384	16384
r	26	26	30	30	33	33
L	2	4	4	6	6	7
rel. err. (e10^{-14})	7.1041	5.9208	8.1024	6.1210	9.4705	6.1585

n	2048	2048	8192	8192	16384	16384
r	26	26	30	30	33	33
L	2	4	4	6	6	7
rel. err. (e10^{-14})	1.7926	1.1841	2.1102	1.2407	2.5062	1.2521

Again, we note that even after global error accumulation associated with an HSS tree of depth 6 and 7 in Examples 4 and 3, the relative error is still quite low. This gives evidence that the asymptotic error decay regime from Proposition 2.2 holds well enough in practice: note that the maximum of the function in Example 3 is even increasing on $\mathcal{B}(n/2, n/2)$ as n grows. This increase, however, is polynomial in n, and therefore so is the numerator of the bound given by Corollary 4.2. The denominator of this bound is exponential in p, which helps explain the quality of the approximation in Example 3.

6. Extensions.

In forthcoming studies, we can use the arguments of Section 4.3 to bound the numerical rank of certain classes of matrices. In particular, we could use control over the error in Proposition 2.1 to produce bounds similar to Corollary 4.2 and argue when a general one-dimensional kernel matrix may have low numerical rank. Furthermore, we may perform a more detailed analysis of the global error accumulated after all compression steps in Sections 4.1 and 4.2 are performed, including the SRRQR factorization steps. This gives additional motivation for proving an absolute bound in Proposition 2.1, Proposition 2.3, and Corollary 4.2, since relative bounds are harder to integrate into a global HSS error analysis.

Finally, we may also extend the bound of Proposition 2.1 to analytic functions of more than one (complex) variable. In particular, no part of the argument used in this proposition relies on complex analysis concepts that apply only in the one-variable case. Hence, we may explore generalizations of the complex-analytic low-rank approximations discussed here to more general Toeplitz matrices, as well as to non-Toeplitz matrices that are defined by analytic functions in other ways. When doing so, we may also combine the results of Section 4.3 with the hierarchical partitioning described in [24]. As mentioned above, this may again enable us to obtain off-diagonal rank bounds for classes of kernel matrices by certain multivariable analytic functions satisfying adequate growth bounds.

REFERENCES

[1] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D.W. Hogg, and M. O'Neil, *Fast direct methods for Gaussian processes*, IEEE Trans. Pattern Anal. Mach. Intell., 38 (2016), pp. 252–265.

[2] S. Chandrasekaran, M. Gu, X. Sun, J. Xia, and J. Zhu, *A superfast algorithm for Toeplitz systems of linear equations*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 1247–1266.

[3] R. Bilato, O. Maj, M. Brambilla, *An algorithm for fast Hilbert transform of real functions*, Adv. Comput. Math. 40 (2014), pp. 1159–1168.

[4] W. Burger and M. J. Burge, *Digital Image Processing: An Algorithmic Introduction Using Java, 2nd ed.*, Springer Publishing Company, Inc. (2016) pp. 496–498.

[5] I. Gohberg, T. Kailath, and V. Olshevsky, *Fast Gaussian elimination with partial pivoting for matrices with displacement structure*, Math. Comp., 64 (1995), pp. 1557–1576.

[6] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys., 73 (1987), pp. 325–348.

[7] M. Gu, *Stable and efficient algorithms for structured systems of linear equations*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 279–306.

[8] M. Gu and S. C. Eisenstat, *Efficient algorithms for computing a strong-rank revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.

[9] G. Heinig, *Inversion of generalized Cauchy matrices and other classes of structured matrices*, in Linear Algebra for Signal Processing, IMA Vol. Math. Appl. 69, Springer, New York, 1995, pp. 95–114.

[10] A. K. Jain *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc. (1989), pp. 25–26.

[11] R. Kress, *Linear Integral Equations, Third Edition*, Springer, 2014.

[12] M. Lepilov and J. Xia, *Rank-structured approximation of some Cauchy matrices with sublinear complexity*, Numer Linear Algebra Appl. 31(1) (2024), e2526.

[13] P.-D. Letourneau, C. Cecka, and E. Darve, *Cauchy fast multipole method for general analytic kernels*, SIAM J. Sci. Comput., 36 (2014), pp. A396–A426.

[14] E. Liberty, F. Woolfe, P. G. Martinsson, V. Rokhlin, and M. Tygert, *Randomized algorithms for the low-rank approximation of matrices*, Proc. Natl. Acad. Sci. USA, 104 (2007), pp. 20167–20172.

[15] P. G. Martinsson and V. Rokhlin, *A fast direct solver for boundary integral equations in two dimensions*, J. Comput. Phys., 205 (2005), pp. 1–23.

[16] P. G. Martinsson and V. Rokhlin, *An accelerated kernel-independent fast multipole method in one dimension*, SIAM J. Sci. Comput., 29 (2007), pp. 1160–1178.

[17] P. G. Martinsson, V. Rokhlin, and M. Tygert, *A fast algorithm for the inversion of general Toeplitz matrices*, Comput. Math. Appl., 50 (2005), pp. 741–752.

[18] L. N. Trefethen, J. A. C. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Rev., 56 (2014), pp. 385-458.

[19] J. Vogel, J. Xia, S. Cauley, V. Balakrishnan, *Superfast divide-and-conquer method and perturbation analysis for structured eigenvalue solutions*, SIAM J. Sci. Comput. 38 (2016), pp. A1358–A1382.

[20] H. Wang, K. Wang, T. Sircar, *A direct $O\left(N\log^2(N)\right)$ finite difference method for fractional diffusion equations*, J. Comp. Phys. 229 (2010), pp. 8095–8104.

[21] Y. Xi, J. Xia, R. Chan, *A fast randomized eigensolver with structured LDL factorization update*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 974–996.

[22] Y. Xi, J. Xia, S. Cauley, and V. Balakrishnan, *Superfast and stable structured solvers for Toeplitz least squares via randomized sampling*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 44–72.

[23] J. Xia, *On the complexity of some hierarchical structured matrix algorithms*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 388–410.

[24] J. Xia, *Multi-layer hierarchical structures*, CSIAM Trans. Appl. Math., 2 (2021), pp. 263–296.

[25] J. Xia, S. Chandrasekaran, M. Gu, and X. S. Li, *Fast algorithms for hierarchically semiseparable matrices*, Numer. Linear Algebra Appl., 17 (2010), pp. 953–976.

[26] J. Xia, Y. Xi, and M. Gu, *A superfast structured solver for Toeplitz linear systems via randomized sampling*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 837–858.

[27] X. Xing and E. Chow, *Interpolative decomposition via proxy points for kernel matrices*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 221–243.

[28] X. Ye, J. Xia, and L. Ying, *Analytical low-rank compression via proxy point selection*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 1059–1085.

[29] L. Ying, G. Biros, and D. Zorin, *A kernel-independent adaptive fast multipole algorithm in two and three dimensions*, J. Comput. Phys., 196 (2004), pp. 591–626.

[30] Y. Zhang, W.E. Leithead, and D.J. Leith, *Time-series Gaussian process regression based on Toeplitz computation of $O(N^2)$ operations and $O(N)$-level storage*, Proc. IEEE Conf. Decis. Control. 44 (2005), pp. 3711-3716.