Optimal sampling rates for approximating analytic functions from pointwise samples

BEN ADCOCK[*]

*Department of Mathematics, Simon Fraser University, 8888 University Drive Burnaby,
BC V5A 1S6, Canada*
*Corresponding author: ben_adcock@sfu.ca

RODRIGO B. PLATTE

*School of Mathematical and Statistical Sciences, Arizona State University,
P.O. Box 871804, Tempe, AZ 85287-1804, USA*

AND

ALEXEI SHADRIN

*DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road,
Cambridge, CB3 0WA, UK*

We consider the problem of approximating an analytic function on a compact interval from its values at $M + 1$ distinct points. When the points are equispaced, a recent result (the so-called *impossibility theorem*) has shown that the best possible convergence rate of a stable method is root-exponential in M, and that any method with faster exponential convergence must also be exponentially ill conditioned at a certain rate. This result hinges on a classical theorem of Coppersmith & Rivlin concerning the maximal behavior of polynomials bounded on an equispaced grid. In this paper, we first generalize this theorem to arbitrary point distributions. We then present an extension of the impossibility theorem valid for general nonequispaced points and apply it to the case of points that are equidistributed with respect to (modified) Jacobi weight functions. This leads to a necessary sampling rate for stable approximation from such points. We prove that this rate is also sufficient, and therefore exactly quantify (up to constants) the precise sampling rate for approximating analytic functions from such node distributions with stable methods. Numerical results—based on computing the maximal polynomial via a variant of the classical Remez algorithm—confirm our main theorems. Finally, we discuss the implications of our results for polynomial least-squares approximations. In particular, we theoretically confirm the well-known heuristic that stable least-squares approximation using polynomials of degree $N < M$ is possible only once M is sufficiently large for there to be a subset of N of the nodes that mimic the behavior of the Nth set of Chebyshev nodes.

Keywords: polynomial approximation; discrete approximation; exponential convergence; least-squares.

1. Introduction

The concern of this paper is the approximation of an analytic function $f : [-1, 1] \rightarrow \mathbb{C}$ from its values on an arbitrary set of $M + 1$ points in $[-1, 1]$. It is well known that if such points follow a Chebyshev distribution then f can be stably (up to a log factor in M) approximated by its polynomial interpolant with a convergence rate that is geometric in the parameter M. Conversely, when the points are equispaced, polynomial interpolants do not necessarily converge uniformly on $[-1, 1]$ as $M \rightarrow \infty$, an

effect known as Runge's phenomenon. Such an approximation is also exponentially ill conditioned in M, meaning that divergence is witnessed in finite precision arithmetic even when theoretical convergence is expected.

Many different numerical methods have been proposed to *overcome* Runge's phenomenon by replacing the polynomial interpolant by an alternative approximation (see Boyd & Ong, 2009; Platte *et al.*, 2011; Platte & Klein, 2016 and references therein). This raises a fundamental question: how successful can such approximations be? For equispaced points, this question was answered recently in Platte *et al.* (2011). Therein it was proved that no stable method for approximating analytic functions from equispaced nodes can converge better than root-exponentially fast in the number of points, and moreover any method that converges exponentially fast must also be exponentially ill conditioned.

A well-known method for approximating analytic functions is polynomial least squares fitting with a polynomial of degree $N < M$. Although a classical approach, this technique has become increasingly popular in recent years as a technique for computing so-called polynomial chaos expansions with application to uncertainty quantification (see Cohen *et al.*, 2013; Migliorati *et al.*, 2014; Chkifa *et al.*, 2015; Narayan & Zhou, 2015 and references therein), as well as in data assimilation in reduced-order modeling (see Binev *et al.*, 2016; DeVore *et al.*, 2016). A consequence of the impossibility theorem of Platte *et al.* (2011) is that polynomial least squares is an optimal stable and convergent method for approximating one-dimensional analytic functions from equispaced data, provided the polynomial degree N used in the least squares fit scales like the square root of the number of grid points $M + 1$ (see Adcock, 2017). Using similar ideas, it has also recently been shown that polynomial least squares is also an optimal, stable method for extrapolating analytic functions (see Demanet & Townsend, 2016).

1.1   *Contributions*

The purpose of this paper is to investigate the limits of stability and accuracy for approximating analytic functions from arbitrary sets of points. Of particular interest is the case of points whose behavior lies between the two extremes of Chebyshev and equispaced grids. Specifically, suppose a given set of points exhibits some clustering near the endpoints, but not the characteristic quadratic clustering of Chebyshev grids. Generalizing that of Platte *et al.* (2011), our main result precisely quantifies both the best achievable error decay rate for a stable approximation and the resulting ill-conditioning if one seeks faster convergence.

This result follows from an extension of a classical theorem of Coppersmith & Rivlin on the maximal behavior of a polynomial of degree N bounded on a grid of M equispaced points (see Coppersmith & Rivlin, 1992). In Theorem 3.1 we extend the lower bound proved in Coppersmith & Rivlin (1992) to arbitrary sets of points. We next present an abstract impossibility result (Lemma 4.1) valid for arbitrary sets of points. To illustrate this result in a concrete setting we then specialize to the case of nodes that are equidistributed with respect to modified Jacobi weight functions. Such weight functions take the form

$$\mu(x) = g(x)(1 - x)^{\alpha}(1 + x)^{\beta},$$

where $\alpha, \beta > -1$ and $c_1 \leq g(x) \leq c_2$ almost everywhere, and include equispaced $\left(\mu(x) = \frac{1}{2}\right)$ and Chebyshev $\left(\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}\right)$ weight functions as special cases. In our main result, Theorem 4.2, we prove an extended impossibility theorem for the corresponding nodes. Generalizing Platte *et al.* (2011), two important consequences of this theorem are as follows:

(i) If $\gamma = \max\{\alpha, \beta\} > -1/2$, any method that converges exponentially fast in M with geometric rate, i.e. the error decays like $\mathcal{O}(\rho^{-M})$ for some $\rho > 1$, must also be exponentially ill conditioned in M at a geometric rate.

(ii) The best possible convergence rate for a stable approximation is subgeometric with index $\nu = \frac{1}{2(\gamma+1)}$. That is, the error is at best $\mathcal{O}(\rho^{-M^\nu})$ for some $\rho > 1$ as $M \to \infty$.

We also give a full characterization of the trade-off between exponential ill-conditioning and exponential convergence at subgeometric rates lying strictly between $\nu = \frac{1}{2(\gamma+1)}$ and $\nu = 1$.

Although not a result about polynomials *per se*, this theorem is closely related to the behavior of discrete least-squares fitting with polynomials of degree $N < M$. Indeed, the quantity we estimate in Theorem 3.1 is equivalent (up to a factor of $\sqrt{M}$) to the infinity-norm condition number of such an approximation. By using polynomial least squares as our method, in Proposition 5.5 we show that the rate described in (ii) is not only necessary for stable recovery but also sufficient. Specifically, when the polynomial degree is chosen as

$$N \asymp M^\nu, \qquad \nu = \frac{1}{2(\gamma+1)}, \tag{1.1}$$

the polynomial least-squares approximation is stable and converges like $\mathcal{O}(\rho^{-M^\nu})$ for all functions analytic in an appropriate complex region. The fact that such a method is optimal in view of the generalized impossibility theorem goes some way towards justifying the popularity of discrete least-squares techniques.

Besides these results, in Section 6 we also introduce an algorithm for computing the maximal polynomial for an arbitrary set of nodes. This algorithm, which is based on a result of Schönhage (1961), is a variant of the classical Remez algorithm for computing best uniform approximations (see, for example, Powell, 1981; Pachón & Trefethen, 2009). We use this algorithm to present numerical results in the paper confirming our various theoretical estimates.

Finally, let us note that one particular consequence of our results is a confirmation of a popular heuristic for polynomial least-squares approximation (see, for example, Boyd & Xu, 2009). Namely, the number of nonequispaced nodes required to stably recover a polynomial approximation of degree N is of the same order as the number of nodes required for there to exist a subset of those nodes of size $N+1$ that mimics the distribution of the Chebyshev nodes $\{\cos(n\pi/N)\}_{n=0}^N$. In Proposition 5.6 we show that the same sufficient condition for boundedness of the maximal polynomial also implies the existence of a subset of N of the original $M+1$ nodes that interlace the Chebyshev nodes. In particular, for nodes that are equidistributed according to a modified Jacobi weight function one has this interlacing property whenever the condition $M \asymp N^{2(\gamma+1)}$ holds, which is identical to the necessary and sufficient condition (1.1) for stability of the least-squares approximation.

2. Preliminaries

Our focus in this paper is on functions defined on compact intervals, which we normalize to the unit interval $[-1, 1]$. Unless otherwise stated, all functions will be complex valued, and in particular, polynomials may have complex coefficients. Throughout the paper, $-1 = x_0 < \cdots < x_M = 1$ will denote the nodes at which a function $f : [-1, 1] \to \mathbb{C}$ is sampled. We include both endpoints $x = \pm 1$ in this set for convenience. All results we prove remain valid (with minor alterations) for the case when either or both endpoints are excluded.

We require several further pieces of notation. Where necessary throughout the paper $N \leq M$ will denote the degree of a polynomial. We write $\|f\|_\infty$ for the uniform norm of a function $f \in C([-1, 1])$ and $\|f\|_{M,\infty} = \max_{m=1,\dots,M} |f(x_m)|$ for the discrete uniform seminorm of f on the grid of points. We write $\langle \cdot, \cdot \rangle$ for the Euclidean inner product on $L^2(-1, 1)$ and $\|\cdot\|_2$ for the Euclidean norm. Correspondingly, we let $\langle f, g \rangle_M = \frac{2}{M+1} \sum_{m=0}^{M} f(x_m)\overline{g(x_m)}$ and $\|f\|_{M,2} = \sqrt{\langle f, f \rangle_M}$ be the discrete semiinner product and seminorm, respectively.

We will say that a sequence a_n converges to zero exponentially fast if $|a_n| = \mathcal{O}\left(\rho^{-n^r}\right)$ for some $\rho > 1$ and $r > 0$. If $r = 1$ then we say the convergence is geometric, and if $0 < r < 1$ or $r > 1$ then it is subgeometric or supergeometric, respectively. When $r = 1/2$ we also refer to this convergence as root-exponential. Given two non-negative sequences a_n and b_n we write $a_n \asymp b_n$ as $n \to \infty$ if there exist constants $c_1, c_2 > 0$ such that $c_1 b_n \leq a_n \leq c_2 b_n$ for all large n. Finally, we will on occasion use the notation $A \lesssim B$ to mean that there exists a constant $c > 0$ independent of all relevant parameters such that $A \leq cB$.

2.1 *The impossibility theorem for equispaced points*

We first review the impossibility theorem of Platte *et al.* (2011). Let $\{x_m\}_{m=0}^{M} = \{-1 + 2m/M\}_{m=0}^{M}$ be a grid of $M + 1$ equispaced points in $[-1, 1]$ and suppose that $F_M : C([-1, 1]) \to C([-1, 1])$ is a family of mappings such that $F_M(f)$ depends only on the values of f on this grid. We define the (absolute) condition numbers as

$$\kappa(F_M) = \sup_{f \in C([-1,1])} \lim_{\delta \to 0^+} \sup_{\substack{h \in C([-1,1]) \\ 0 < \|h\|_{M,\infty} \leq \delta}} \frac{\|F_M(f + h) - F_M(f)\|_\infty}{\|h\|_{M,\infty}}. \tag{2.1}$$

Suppose that $E \subseteq \mathbb{C}$ is a compact set. We now write $B(E)$ for the Banach space of functions that are continuous on E and analytic in its interior with norm $\|f\|_E = \sup_{z \in E} |f(z)|$.

THEOREM 2.1 (Platte *et al.* (2011)). Let $E \subseteq \mathbb{C}$ be a compact set containing $[-1, 1]$ in its interior and suppose that $\{F_M\}_{M=1}^{\infty}$ is an approximation procedure based on equispaced grids of $M + 1$ points such that for some $C, \rho > 1$ and $1/2 < \tau \leq 1$ we have

$$\|f - F_M(f)\|_\infty \leq C\rho^{-M^\tau} \|f\|_E, \qquad M = 1, 2, \dots,$$

for all $f \in B(E)$. Then the condition numbers (2.1) satisfy

$$\kappa(F_M) \geq \sigma^{M^{2\tau-1}}$$

for some $\sigma > 1$ and all large M.

Specializing to $\tau = 1$, this theorem states that exponential convergence at a geometric rate implies exponential ill-conditioning at a geometric rate. Conversely, stability of any method is possible only when $\tau = 1/2$, which corresponds to root-exponential convergence in M.

2.2 *Coppersmith & Rivlin's bound*

The proof of Theorem 2.1, although it does not pertain to polynomials or polynomial approximation specifically, relies on a result of Coppersmith & Rivlin on the maximal behavior of polynomials bounded

on an equispaced grid. To state this result, we first introduce the following notation:

$$B(M,N) = \sup \left\{ \|p\|_\infty : p \in \mathbb{P}_N, \|p\|_{M,\infty} \le 1 \right\}. \tag{2.2}$$

Note that in the special case $M = N$, this is just the Lebesgue constant

$$B(N,N) = \Lambda(N) = \sup \left\{ \|F_N(f)\|_\infty : f \in C([-1,1]), \|f\|_\infty \le 1 \right\},$$

where $F_N(f)$ denotes the polynomial interpolant of degree N of a function f.

THEOREM 2.2 (Coppersmith & Rivlin (1992)). Let $\{x_m\}_{m=0}^M = \{-1 + 2m/M\}_{m=0}^M$ be an equispaced grid of $M+1$ points in $[-1,1]$ and suppose that $1 \le N \le M$. Then there exist constants $b \ge a > 1$ such that

$$a^{N^2/M} \le B(M,N) \le b^{N^2/M}.$$

Two implications of this result are as follows. First, a polynomial of degree N bounded on $M = \mathcal{O}(N)$ equispaced points can grow at most exponentially large in between those points. Second, one needs quadratically many equispaced points, i.e. $M \asymp N^2$, in order to prohibit growth of an arbitrary polynomial of degree N that is bounded on an equispaced grid. We remark in passing that when $M = N$, so that $B(N,N) = \Lambda(N)$ is the Lebesgue constant, one also has the well-known estimate $\Lambda(N) \sim \frac{2^{N+1}}{eN\log(N)}$ for large N (see, for example, Trefethen, 2013, Chpt. 15).

REMARK 2.3 Sufficiency of the scaling $M \asymp N^2$ is a much older result than Theorem 2.2, dating back to the works by Schönhage (1961) and Ehlich & Zeller (1964, 1965). Ehlich also proved unboundedness of $B(M,N)$ if $M = o(N^2)$ as $N \to \infty$ (see Ehlich, 1966). More recently, Rakhmanov (2007) has given a very precise analysis of not just $B(M,N)$ but also the pointwise quantity $B(M,N,x) = \sup\{|p(x)| : p \in \mathbb{P}_N, \|p\|_{M,\infty} \le 1\}$ for $-1 \le x \le 1$.

2.3 *Discrete least squares*

A simple algorithm that attains the bounds implied by Theorem 2.1 is discrete least-squares fitting with polynomials:

$$F_{M,N}(f) = \operatorname*{argmin}_{p \in \mathbb{P}_N} \sum_{m=0}^M |f(x_m) - p(x_m)|^2. \tag{2.3}$$

Here $N \le M$ is a parameter, which is chosen to ensure specific rates of convergence. The following result determines the conditioning and convergence of this approximation (note that this result is valid for arbitrary sets of points, not just equispaced).

PROPOSITION 2.4 Let $\{x_m\}_{m=0}^M$ be a set of $M+1$ points in $[-1,1]$ and suppose that $1 \le N \le M$. If $f \in C([-1,1])$ and $F_{M,N}(f)$ is as in (2.3) then the error

$$\|f - F_{M,N}(f)\|_\infty \le (1 + \kappa_{M,N}) \inf_{p \in \mathbb{P}_N} \|f - p\|_\infty,$$

where $\kappa_{M,N} = \kappa(F_{M,N})$ is the condition number of $F_{M,N}$. Moreover,

$$B(M,N) \leq \kappa_{M,N} \leq \sqrt{M+1}B(M,N), \tag{2.4}$$

where $B(M,N)$ is as in (2.2). Furthermore, if $M = N$, i.e. $F_N = F_{N,N}$ is the polynomial interpolant of degree N, then

$$\kappa_{N,N} = B(N,N) = \Lambda(N) \tag{2.5}$$

is the Lebesgue constant.

Although this result is well known, we include a short proof for completeness.

Proof. Since the points are distinct and $N \leq M$, the least-squares solution exists uniquely. Notice that the mapping $F_{M,N}$ is linear and a projection onto $\mathbb{P}_N$. Hence,

$$\kappa_{M,N} = \sup_{\substack{f \in C([-1,1]) \\ \|f\|_{M,\infty} \neq 0}} \frac{\|F_{M,N}(f)\|_\infty}{\|f\|_{M,\infty}}, \tag{2.6}$$

and consequently we have

$$\|f - F_{M,N}(f)\|_\infty \leq \|f - p\|_\infty + \|F_{M,N}(f - p)\|_\infty \leq (1 + \kappa_{M,N})\|f - p\|_\infty \quad \forall p \in \mathbb{P}_N.$$

It remains to estimate the condition number. Because $F_{M,N}(f)$ is a polynomial, it follows that

$$\kappa_{M,N} \leq B(M,N) \sup_{\substack{f \in C([-1,1]) \\ \|f\|_{M,\infty} \neq 0}} \frac{\|F_{M,N}(f)\|_{M,\infty}}{\|f\|_{M,\infty}}. \tag{2.7}$$

Now observe that

$$\|F_{M,N}(f)\|_{M,\infty}^2 \leq \sum_{m=0}^{M} |F_{M,N}(f)(x_m)|^2 = \frac{M+1}{2}\|F_{M,N}(f)\|_{M,2}^2.$$

Since $F_{M,N}(f)$ is the solution of a discrete least-squares problem it is a projection with respect to the discrete semiinner product $\langle \cdot, \cdot \rangle_M$. Hence, $\|F_{M,N}(f)\|_{M,2} \leq \|f\|_{M,2} \leq \sqrt{2}\|f\|_\infty$. Combining this with the previous estimate gives the upper bound $\kappa_{M,N} \leq \sqrt{M+1}B(M,N)$. For the lower bound, we use (2.6) and the fact that $F_{M,N}$ is a projection to deduce that

$$\kappa_{M,N} \geq \max_{\substack{p \in \mathbb{P}_N \\ \|p\|_{M,\infty} \neq 0}} \frac{\|p\|_\infty}{\|p\|_{M,\infty}} = B(M,N).$$

This completes the proof of (2.4). For (2.5) we merely use the definition of $\Lambda(N)$.   □

2.4 *Examples of nonequispaced points*

To illustrate our main results proved later in the paper, we shall consider points $-1 = x_0 < x_1 < \cdots < x_M = 1$ that are equidistributed with respect to so-called *modified Jacobi* weight functions. These are defined as

$$\mu(x) = g(x)(1 - x)^\alpha (1 + x)^\beta, \tag{2.8}$$

where $\alpha, \beta > -1$ and $g \in L^\infty(-1, 1)$ satisfies $c_1 \le g(x) \le c_2$ almost everywhere. Throughout, we assume the normalization

$$\int_{-1}^{1} \mu(x) \, dx = 1,$$

in which case the points $\{x_m\}_{m=0}^{M}$ are defined implicitly by

$$\frac{m}{M} = \int_{-1}^{x_m} \mu(x) \, dx, \quad m = 0, \ldots, M. \tag{2.9}$$

Ultraspherical weight functions are special cases of modified Jacobi weight functions. They are defined as

$$\mu(x) = c(1 - x^2)^\alpha, \qquad c = \left(\int_{-1}^{1} (1 - x^2)^\alpha \, dx \right)^{-1}, \tag{2.10}$$

for $\alpha > -1$. Within this subclass, we shall consider a number of specific examples:

(U)   ($\alpha = 0$) the uniform weight function $\mu(x) = 1/2$, corresponding to the equispaced points $x_m = -1 + 2\frac{m}{M}$.

(C1)   ($\alpha = -1/2$) The Chebyshev weight function of the first kind $\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}$, corresponding to the Chebyshev points. Note that these points are roughly equispaced near $x = 0$ and quadratically spaced near the endpoints $x = \pm 1$. That is, $|x_1 + 1|, |x_{M-1} - 1| = \mathcal{O}(M^{-2})$.

(C2)   ($\alpha = 1/2$) The Chebyshev weight function of the second kind $\mu(x) = \frac{2}{\pi}\sqrt{1 - x^2}$. Note that the corresponding points are roughly equispaced near $x = 0$, but are sparse near the endpoints. In particular, $|x_1 + 1|, |x_{M-1} - 1| = \mathcal{O}(M^{-1/2})$.

Recall that for (U) one requires a quadratic scaling $M \asymp N^2$ to ensure stability. Conversely, for (C1) any linear scaling of $M = cN$ with $c > 1$ suffices (see Remark 3.5). Since the points (C2) are so poorly distributed near the endpoints, we expect, and it will turn out to be the case, that a more severe scaling than quadratic is required for stability in this case.

We shall also consider two further examples:

(UC)   ($\alpha = -1/4$) The corresponding points cluster at the endpoints, although not quadratically. Specifically, $|x_1 + 1|, |x_{M-1} - 1| = \mathcal{O}(M^{-4/3})$.

(OC)   ($\alpha = -3/4$) The corresponding points overcluster at the endpoints: $|x_1 + 1|, |x_{M-1} - 1| = \mathcal{O}(M^{-4})$.

We expect (UC) to require a superlinear scaling of M with N for stability, although not as severe as quadratic scaling as in the case of (U). Conversely, in (OC) it transpires that linear scaling suffices, but unlike the case of (C1), the scaling factor c (where $M/N = c$) must be sufficiently large.
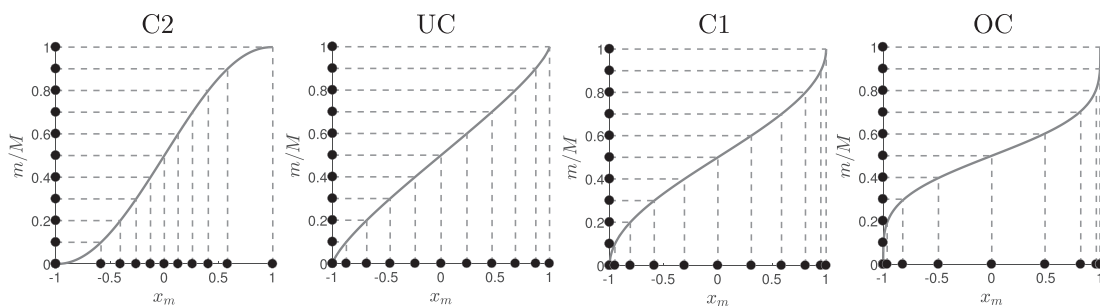
FIG. 1. Relationship between m/M and x_m given by (2.9) for the ultraspherical weight functions with $\alpha = 1/2$ (C2), $\alpha = -1/4$ (UC), $\alpha = -1/2$ (C1) and $\alpha = -3/4$ (OC). Here $M = 10$.

The node clustering for the above distributions is illustrated in Fig. 1. This figure also shows the corresponding cumulative distribution functions $\int_{-1}^{x} \mu(\xi)\,\mathrm{d}\xi$.

3. Maximal behavior of polynomials bounded on arbitrary grids

We now seek to estimate the maximal behavior of a polynomial of degree N that is bounded at arbitrary nodes $-1 = x_0 < x_1 < \cdots < x_M = 1$. As in Section 2.2, we define

$$B(M,N) = \sup\left\{\|p\|_\infty : p \in \mathbb{P}_N,\ |p(x_m)| \leq 1,\ m = 0,\ldots,M\right\}. \tag{3.1}$$

Once again we note that $B(N,N) = \Lambda(N)$ is the Lebesgue constant of polynomial interpolation.

3.1 *Lower bound for B(M, N)*

Our first main result of the paper is a generalization of the lower bound of Coppersmith & Rivlin (Theorem 2.2) to arbitrary nodes. Before stating this, we need several definitions. First, given $M \geq N \geq 1$ and nodes $-1 = x_0 < x_1 < \cdots < x_M = 1$, we define

$$Q_-(K,N) = \frac{\pi}{8}\left(\frac{2N^2}{\pi^2}\right)^{K-1}\frac{1}{\Gamma(K+1/2)^2}\prod_{n=1}^{K-1}(1+x_n), \qquad 2 \leq K \leq N,$$

$$Q_+(K,N) = \frac{\pi}{8}\left(\frac{2N^2}{\pi^2}\right)^{K-1}\frac{1}{\Gamma(K+1/2)^2}\prod_{n=1}^{K-1}(1-x_{M-n}), \qquad 2 \leq K \leq N,$$

$$Q_-(1,N) = Q_+(1,N) = 1.$$

Second, let $-1 < y_0 < \cdots < y_{N-1} < 1$ be the zeros of the Nth Chebyshev polynomial $q(x) = \cos(N\arccos(x))$:

$$y_n = -\cos\left(\frac{(2n+1)\pi}{2N}\right), \qquad n = 0,\ldots,N-1. \tag{3.2}$$
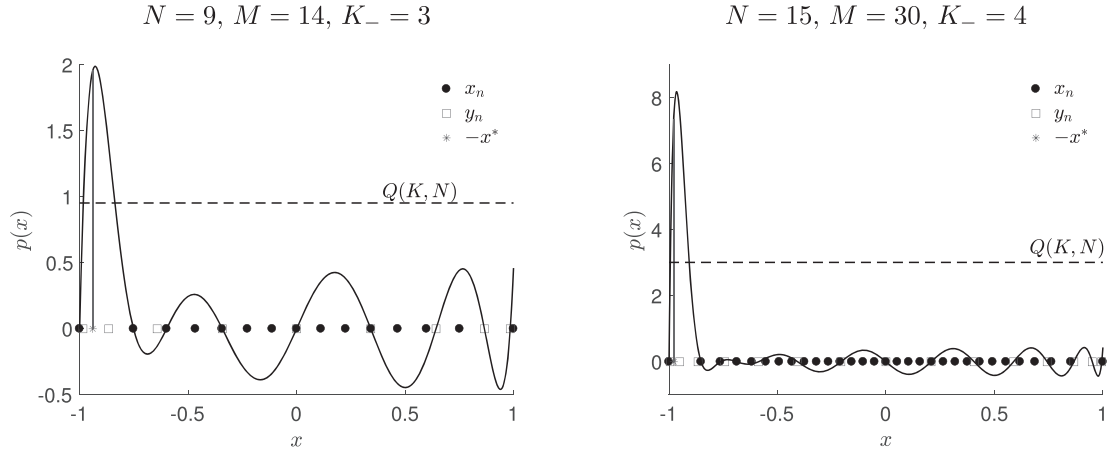
We now have the following theorem.

FIG. 2. Plots of the polynomial p used in the proof of Theorem 3.1 for two sets of points. In both cases the points x_m were generated with $\gamma = 1/2$: case (C2) in Section 2.4. For reference the lower bound $Q(K, N)$ is also included.

THEOREM 3.1 Let $M \geq N \geq 1$, $-1 = x_0 < x_1 < \cdots < x_M = 1$ and suppose that there exist integers $2 \leq K_- \leq N$ and $2 \leq K_+ \leq N$ such that

$$0 \geq x_n > y_n, \qquad n = 1, \ldots, K_- - 1 \tag{3.3}$$

and

$$0 \leq x_{M-n} < y_{N-n}, \qquad n = 1, \ldots, K_+ - 1, \tag{3.4}$$

respectively, where the y_n are as in (3.2). If either K_- or K_+ fails to exist, set $K_- = 1$ or $K_+ = 1$. Then the constant $B(M, N)$ defined in (3.1) satisfies

$$B(M, N) \geq \max \{Q_-(K_-, N), Q_+(K_+, N)\}. \tag{3.5}$$

Proof. We first show that $B(M, N) \geq Q_-(K_-, N)$. If $K_- = 1$ there is nothing to prove, hence we now assume that $2 \leq K_- \leq N$. Let $q(x) = \cos(N \arccos(x))$ be the Nth Chebyshev polynomial and define $p \in \mathbb{P}_N$ by

$$p(x) = \frac{1}{2} q(x) \prod_{n=0}^{K_- - 1} \frac{x - x_n}{x - y_n}. \tag{3.6}$$

Figure 2 illustrates the behavior of p. We first claim that

$$|p(x_m)| \leq 1, \quad m = 0, \ldots, M. \tag{3.7}$$

Clearly, for $m = 0, \ldots, K_- - 1$ we have $p(x_m) = 0$. Suppose now that $K_- \leq m \leq M$. Then, since $|q(x)| \leq 1$,

$$|p(x_m)| \leq \frac{1}{2} \prod_{n=0}^{K_- - 1} \left| \frac{x_m - x_n}{x_m - y_n} \right|.$$

By definition, we have $x_m > x_n$ for $n = 0, \ldots, K_- - 1$. Also, by (3.3),

$$x_m > x_{K_- - 1} \geq y_{K_- - 1} \geq y_n, \quad n = 0, \ldots, K_- - 1,$$

and therefore

$$|p(x_m)| \leq \frac{1}{2} \prod_{n=0}^{K_- - 1} \frac{x_m - x_n}{x_m - y_n}.$$

For $n = 1, \ldots, K_- - 1$, (3.3) gives $\frac{x_m - x_n}{x_m - y_n} \leq 1$. Also, since $x_m \geq y_{K_- - 1}$ and $y_1 > -1$ we have

$$\frac{x_m - x_0}{x_m - y_0} \leq \frac{y_{K_- - 1} + 1}{y_{K_- - 1} - y_0} = 1 + \frac{1 + y_0}{y_{K_- - 1} - y_0} = 1 + \frac{\sin^2(\pi/(4N))}{\sin(K_- \pi/(2N)) \sin((K_- - 1)\pi/(2N))}.$$

Recall that $2t/\pi \leq \sin(t) \leq t$ for $0 \leq t \leq \pi/2$. Hence,

$$\frac{x_m - x_0}{x_m - y_0} \leq 1 + \frac{\pi^2}{16 K_-(K_- - 1)} \leq 2,$$

and therefore

$$|p(x_m)| \leq \frac{1}{2} \prod_{n=0}^{K_- - 1} \frac{x_m - x_n}{x_m - y_n} \leq 1.$$

This completes the proof of claim (3.7).

We now wish to estimate $\|p\|_\infty$ from below. Following Fig. 2, we choose the point $-x^* = -\cos(\pi/N)$ midway between the endpoint $x = -1$ and the leftmost node y_1. Since $|q(-x^*)| = 1$ we derive from (3.6) that

$$\|p\|_\infty \geq |p(-x^*)| = \frac{1}{2} \prod_{n=0}^{K_- - 1} \left| \frac{x^* + x_n}{x^* + y_n} \right|.$$

Notice that $x^* + y_n > 0$ for $n = 1, \ldots, K_- - 1$ and therefore $x^* + x_n > 0$ for $n = 1, \ldots, K_- - 1$ by (3.3). Hence,

$$\|p\|_\infty \geq \frac{1}{2} \left| \frac{x^* + x_0}{x^* + y_0} \right| \prod_{n=1}^{K_- - 1} \frac{x^* + x_n}{x^* + y_n} \geq \frac{1}{2} \left| \frac{x^* + x_0}{x^* + y_0} \right| \prod_{n=1}^{K_- - 1} \frac{1 + x_n}{1 + y_n}, \tag{3.8}$$

where in the second step we use (3.3) and the fact that $-y_n < x^* \leq 1$ and $x_n > y_n$. Note that

$$1 + y_n = 2 \sin^2 \left(\frac{(2n + 1)\pi}{4N} \right)^2 \leq \frac{(2n - 1)^2 \pi^2}{8 N^2}$$

and that

$$\frac{|x^* + x_0|}{|x^* + y_0|} = \frac{1 - x^*}{\cos(\pi/(2N)) - x^*} \geq 1.$$
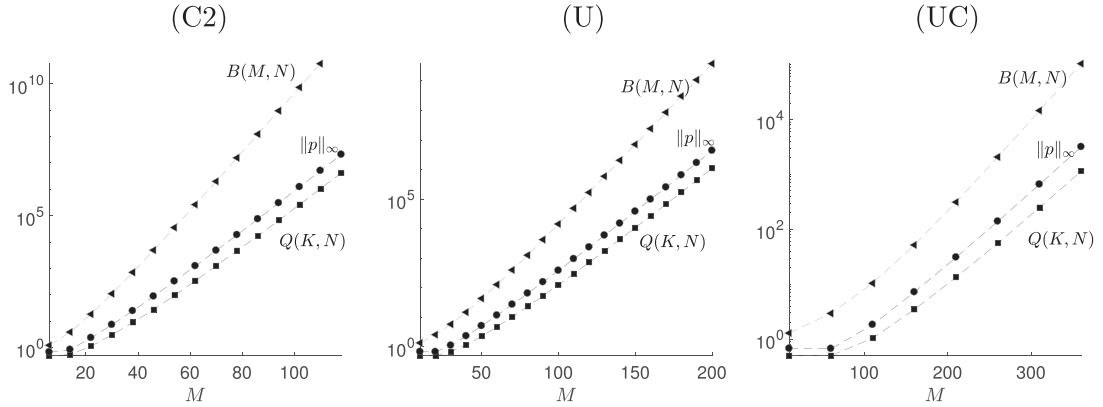
FIG. 3. Semilog plot of the growth of $B(M, N)$, $\|p\|_\infty$ and $Q(K, N)$ for several values of M and three node density functions described in Section 2.4: (C2), (U) and (UC). In all cases $N = M/2$. The computation of the quantity $B(M, N)$ is described in Section 6.

Therefore, we deduce that

$$\|p\|_\infty \geq \frac{1}{2} \left(\prod_{n=1}^{K_- - 1} \frac{8N^2}{(2n-1)^2 \pi^2} \right) \left(\prod_{n=1}^{K_- - 1} (1 + x_n) \right)$$

$$= \frac{1}{2} \left(\frac{8N^2}{\pi^2} \right)^{K_- - 1} \frac{\pi}{4_-^K \Gamma(K_- + 1/2)^2} \prod_{n=1}^{K_- - 1} (1 + x_n)$$

$$= Q_-(K_-, N),$$

which gives $B(M, N) \geq Q_-(K_-, N)$ as required. In order to prove $B(M, N) \geq Q_+(K_+, N)$ we repeat the same arguments, working from the right endpoint $x = +1$. □

Figure 3 shows the growth of $B(M, N)$, $Q(K, N)$ and the norm of the polynomial used to prove Theorem 3.1. In these examples, the nodes x_m were generated using the density functions (C2), (U) and (UC). In all cases the polynomial degree was chosen as $N = M/2$. Notice that the exponential growth rate of $\|p\|_\infty$ is well estimated by $Q(K, N)$, while both quantities underestimate the rate of growth of $B(M, N)$.

3.2 *Lower bound for modified Jacobi weight functions*

Theorem 3.1 is valid for arbitrary sets of points $\{x_m\}_{m=0}^M$. In order to derive rates of growth, we now consider points equidistributed according to modified Jacobi weight functions. For this, we first recall the following bounds for the gamma function (see, for example, Abramowitz & Stegun, 1974, Eqn. (6.1.38)):

$$\sqrt{2\pi} z^{z+1/2} e^{-z} \leq \Gamma(z+1) \leq \sqrt{2\pi} e^{1/12} z^{z+1/2} e^{-z}, \qquad z \geq 1. \tag{3.9}$$

COROLLARY 3.2 Let $\mu(x) = g(x)(1-x)^\alpha(1+x)^\beta$, where $\alpha, \beta > -1$ and $c_1 \leq g(x) \leq c_2$ almost everywhere. If $\gamma = \max\{\alpha, \beta\} > -1/2$ then there exist constants $C > 0$ and $\sigma > 1$ depending on α and β such that

$$B(M,N) \geq C\sigma^\nu, \qquad \nu = \left(N^{2(\gamma+1)}/M\right)^{\frac{1}{2\gamma+1}},$$

for all $1 \leq N \leq M$.

Proof. By definition, the points $\{x_m\}_{m=0}^M$ satisfy

$$\frac{m}{M} = \int_{-1}^{x_m} g(x)(1-x)^\alpha(1+x)^\beta \, dx.$$

Without loss of generality, suppose that $\beta \geq \alpha$. Let m be such that $x_m \leq 0$. Then

$$\frac{m}{M} = \int_{-1}^{x_m} g(x)(1-x)^\alpha(1+x)^\beta \, dx \lesssim \int_{-1}^{x_m} (1+x)^\beta \, dx \lesssim (1+x_m)^{1+\beta},$$

and therefore

$$1 + x_m \gtrsim \left(\frac{m}{M}\right)^{\frac{1}{1+\beta}}, \qquad x_m \leq 0. \tag{3.10}$$

We now apply Theorem 3.1 and (3.9) to get

$$\begin{aligned}
B(M,N) \geq Q(K,N) &\gtrsim \left(\frac{2N^2}{\pi^2 M^{\frac{1}{1+\beta}}}\right)^{K-1} \frac{\Gamma(K)^{\frac{1}{1+\beta}}}{\Gamma(K+1/2)^2} \\
&\gtrsim \left(\frac{2N^2}{\pi^2 M^{\frac{1}{1+\beta}}}\right)^{K-1} \frac{\left((K/2)^{K-1} e^{1-K}\right)^{\frac{1}{1+\beta}}}{K^{2K} e^{2(1-K)}} \\
&= K^{-2} \left(\frac{2^{\frac{\beta}{1+\beta}} e^{\frac{1+2\beta}{1+\beta}} N^2}{\pi^2 M^{\frac{1}{1+\beta}} K^{\frac{1+2\beta}{1+\beta}}}\right)^{K-1},
\end{aligned} \tag{3.11}$$

where K is any value such that $x_n \geq y_n$ for $n = 1, \ldots, K-1$. We need to determine the range of K for which this holds. From (3.10) we find that $x_n \geq y_n$ provided

$$\left(\frac{n}{M}\right)^{\frac{1}{1+\beta}} \gtrsim \frac{n^2}{N^2},$$

or equivalently

$$n \leq c \left(\frac{N^{2(\beta+1)}}{M}\right)^{\frac{1}{2\beta+1}} = c\nu,$$

where $c > 0$ is some constant. Hence, (3.11) holds for

$$K \in \{2, \ldots, 1 + \lfloor c\nu \rfloor\}. \tag{3.12}$$

We next pick a constant sufficiently small so that $0 < d \leq 2c/3$ and

$$\frac{2^{\frac{\beta}{1+\beta}} e^{\frac{1+2\beta}{1+\beta}}}{\pi^2 (2d)^{\frac{1+2\beta}{1+\beta}}} \geq 2. \tag{3.13}$$

Consider the case where

$$\nu > 2/d.$$

We set

$$K = 1 + \left\lceil \frac{d}{2}\nu \right\rceil,$$

and notice that

$$K \leq 1 + 1 + \frac{d}{2}\nu < d\nu + \frac{d}{2}\nu = \frac{3d}{2}\nu \leq c\nu$$

due to the assumptions on d. Hence, this value of K satisfies (3.12). We next apply (3.11), the bounds $K \geq d\nu/2$ and $K \leq 2d\nu$ and (3.13) to deduce that

$$B(M,N) \gtrsim \nu^{-2} \left(2^{d/2} \right)^{\nu} \geq \rho^{\nu}$$

for some $\rho > 1$. This holds for all $\nu > 2/d$. But since d is a constant, we deduce that $B(M,N) \gtrsim \rho^{\nu}$ for all $\nu \geq 1$. This completes the proof. $\qquad\square$

This result shows that if $M = \mathcal{O}(N^{\tau})$ for some $0 < \tau \leq 2(\gamma + 1)$ then the maximal polynomial grows at least exponentially fast with rate

$$r = \frac{2(\gamma + 1) - \tau}{2\gamma + 1}.$$

In particular, the scaling $M \asymp N^{2(\gamma+1)}$, $N \to \infty$ is necessary for boundedness of the maximal polynomial. In Section 5.1 we will show that this rate is also sufficient.

It is informative to relate this result to several of the examples introduced in Section 2.4. First, if $\gamma = 0$, i.e. case (U), we recover the lower bound of Theorem 2.2. Conversely, for (C2) we have

$$B(M,N) \geq C\sigma^{\sqrt{N^3/M}}.$$

Thus, the cubic scaling $M \asymp N^3$ is necessary for stability. Finally, for the points (UC) we have

$$B(M,N) \geq C\sigma^{\left(M^{3/2}/N \right)^2},$$

which implies a necessary scaling of $M \asymp N^{3/2}$. Note that Corollary 3.2 says nothing about the case $-1 < \gamma \leq -1/2$. We discuss the case $\gamma = -1/2$ further in Remark 3.5.

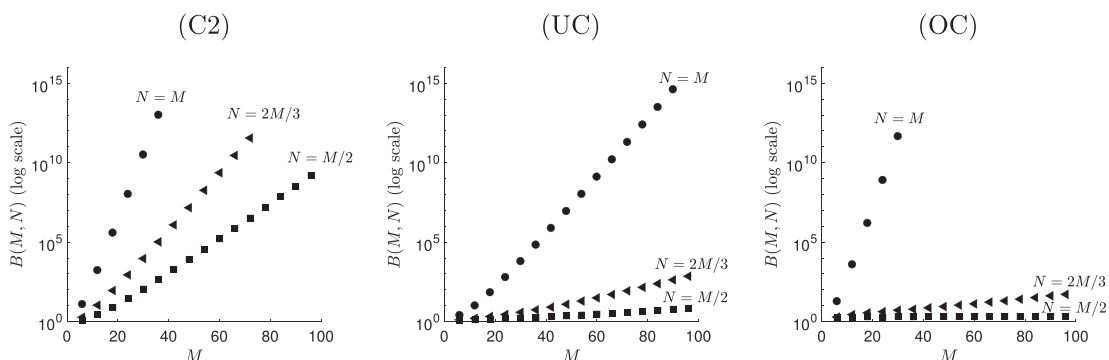The case of linear oversampling ($M \asymp N$) warrants closer inspection.

FIG. 4. The growth of $B(M, N)$ as a function of M for the node densities (C2), (UC) and (OC). In each case $B(M, N)$ is plotted for $N = M$, $N = 2M/3$, and $N = M/2$.

COROLLARY 3.3 Let $\mu(x)$, C and σ be as in Corollary 3.2 and $c \geq 1$. Then

$$B(\lfloor cN \rfloor, N) \geq C \left(\sigma^{c^{-\frac{1}{2\gamma+1}}} \right)^N.$$

In particular, the Lebesgue constants $\Lambda(N) = B(N, N)$ satisfy

$$\Lambda(N) \geq C\sigma^N.$$

In other words, whenever the points cluster more slowly than quadratically at one of the endpoints (recall that $\gamma = \max\{\alpha, \beta\} > -1/2$ in the above result), linear oversampling (including the case $M = N$, i.e. interpolation) necessarily leads to exponential growth of the maximal polynomial at geometric rate. Figure 4 illustrates this for the cases (C2) and (UC). Interestingly, the case (OC), although not covered by this result, also exhibits exponential growth, whenever the oversampling factor is below a particular threshold.

REMARK 3.4 In this paper we are primarily interested in lower bounds for $B(M, N)$, since this is all that is required for the various impossibility theorems. However, in the case of ultraspherical weight functions an upper bound can be derived from results of Rakhmanov. Specifically, let $\gamma = \alpha = \beta > -1/2$ and $x_1, \ldots, x_M$ be a set of M points that are equispaced with respect to the ultraspherical weight function (2.10). Then Rakhmanov (2007, Thm. 1(a)) gives

$$\max_{-r \leq x \leq r} |p(x)| \leq C \max\{|p(x_m)|, m = 1, \ldots, M\} \qquad \forall p \in \mathbb{P}_N,$$

for some constant $C = C_r$ depending only on r, where $r^2 < 1 - (N/M)^\tau$ and $\tau = \frac{2}{2\gamma+1}$.[1] The Remez inequality (see, for example, Borwein & Erdélyi, 1995, Thm. 5.1.1) now gives

$$\max_{|x|\leq 1} |p(x)| \leq CT_N(1/r) \max\{|p(x_m)|, m = 1, \ldots, M\} \qquad \forall\, p \in \mathbb{P}_N,$$

where T_N is the Nth Chebyshev polynomial. Suppose that $N/M \leq 1/2$. Then

$$T_N(1/r) < \left(1/r + \sqrt{1/r^2 - 1}\right)^N < \left(1 + (N/M)^\tau + 2(N/M)^{\tau/2}\right)^N \leq \exp(c(N/M)^{\tau/2}N),$$

for some $c > 0$. Using the definition of τ, we obtain

$$\|p\|_\infty \leq \tilde{C}\tilde{\sigma}^\nu \max\{|p(x_m)|, m = 1, \ldots, M\} \qquad \forall\, p \in \mathbb{P}_N,$$

for $\tilde{C} > 0$ and $\tilde{\sigma} > 1$, where

$$\nu = \left(\frac{N^{2(\gamma+1)}}{M}\right)^{\frac{1}{2\gamma+1}}.$$

The exponent ν is exactly the same as in the lower bound for $B(M, N)$ (Corollary 3.2). In other words, the two-sided bounds of Coppersmith & Rivlin (see Theorem 2.2) for equispaced nodes extend to nodes equidistributed according to ultraspherical weight functions.

REMARK 3.5 Corollaries 3.2 and 3.3 do not apply to the nodes (C1). It is, however, well known (see, for example, Trefethen, 2013, Thm. 15.2) that

$$\Lambda(N) \sim \frac{2}{\pi} \log(N), \quad N \to \infty$$

in this case. Furthermore, a classical result of Ehlich & Zeller (1964) gives

$$B^*(M, N) \leq \frac{1}{\cos\left(\frac{\pi N}{2M}\right)},$$

where

$$B^*(M, N) = \sup\{\|p\|_\infty : p \in \mathbb{P}_N, |p(y_m)| \leq 1, \; m = 1, \ldots, M\},$$

and $y_m = \cos\left(\frac{2m-1}{2M}\right)$, $m = 1, \ldots, M$.[2] In particular this result implies boundedess of $B^*(M, N)$ in the case of linear oversampling, i.e. $M \geq (1 + \epsilon)N$ for any $\epsilon > 0$.

---

[1] Rakhmanov's result excludes the endpoints $x = \pm 1$ from this set, whereas in our results we include these points. However, as noted earlier, our main theorems would remain valid (with minor changes) if these points were excluded.

[2] Similar to the previous footnote, $B^*(M, N)$ excludes the endpoints $x = \pm 1$.

4. An extended impossibility theorem

For $\theta > 1$ let $E_\theta \subseteq \mathbb{C}$ be the Bernstein ellipse with parameter θ. That is, the ellipse with foci at ± 1 and semiminor and semimajor axis lengths summing to θ. Given a domain $E \subseteq \mathbb{C}$ we let $D(E)$ be the set of functions that are analytic on E. The following lemma—whose proof follows the same ideas as Theorem 2.1—shows how the condition number of an exponentially convergent method for approximating analytic functions can be bounded in terms of the quantity $B(M, N)$ for suitable N.

LEMMA 4.1 (Abstract impossibility lemma).   Given points $-1 = x_0 < x_1 < \cdots < x_M = 1$, let $F_M : C([-1,1]) \to C([-1,1])$ be an approximation procedure such that $F_M(f)$ depends only on the values $\{f(x_m)\}_{m=0}^M$ for any $f \in C([-1, 1])$. Suppose that

$$\|f - F_M(f)\|_\infty \leq C\rho^{-M^\tau} \|f\|_E \tag{4.1}$$

for all $f \in D(E)$, where $E \subseteq \mathbb{C}$ is a compact set containing $[-1, 1]$ in its interior, and $C > 0$, $\rho > 1$ and $\tau > 0$ are constants that are independent of f and M. If $E \in \mathbb{N}$ satisfies

$$N < \frac{M^\tau \log(\rho) - \log(2C)}{\log(\theta)},$$

where $\theta > 1$ is such that the $E \subseteq E_\theta$, then the condition number $\kappa(F_M)$ defined in (2.1) satisfies

$$\kappa(F_M) \geq \tfrac{1}{2} B(M, N),$$

for $B(M, N)$ is as in (3.1).

Proof. Let $p \in \mathbb{P}_N$. Since p is entire, (4.1) gives

$$\|F_M(p)\|_\infty \geq \|p\|_\infty - C\rho^{-M^\tau} \|p\|_E.$$

Also, due to a classical result of Bernstein (1912, Sec. 9) (see also Platte *et al.*, 2011, Lem. 1), one has $\|p\|_E \leq \|p\|_{E_\theta} \leq \theta^N \|p\|_\infty$. Hence,

$$\|F_M(p)\|_\infty \geq \left(1 - C\theta^N \rho^{-M^\tau}\right) \|p\|_\infty \geq \frac{1}{2}\|p\|_\infty.$$

It now follows that

$$\kappa(F_M) \geq \max_{\substack{p \in \mathbb{P}_{N-1} \\ \|p\|_{M,\infty} \neq 0}} \frac{\|F_M(p)\|_\infty}{\|p\|_{M,\infty}} \geq \frac{1}{2} \max_{\substack{p \in \mathbb{P}_{N-1} \\ \|p\|_{M,\infty} \neq 0}} \frac{\|p\|_\infty}{\|p\|_{M,\infty}} = \frac{1}{2} B(M, N),$$

as required.   □

   In order to demonstrate this result in a concrete setting we specialize to the case of points equidistributed with respect to modified Jacobi weight functions. This leads to the following theorem, which is the second main result of the paper.

THEOREM 4.2 (Impossibility theorem for modified Jacobi weight functions). For $M \in \mathbb{N}$, let $\{x_m\}_{m=0}^{M}$ be equidistributed according to a modified Jacobi weight function (2.8) with parameters $\alpha, \beta > -1$ (see (2.9)). Let $F_M : C([-1, 1]) \to C([-1, 1])$ be a family of approximation procedures such that $F_M(f)$ depends only on the values $\{f(x_m)\}_{m=0}^{M}$ for any $f \in C([-1, 1])$ and $M \in \mathbb{N}$. Suppose that

$$\|f - F_M(f)\|_\infty \leq C \rho^{-M^\tau} \|f\|_E \tag{4.2}$$

for all $f \in D(E)$, where $E \subset \mathbb{C}$ is a compact set containing $[-1, 1]$ in its interior, and $C > 0$, $\rho > 1$ and $\tau > 0$ are independent of f and M. If

$$\gamma = \max\{\alpha, \beta\} > -1/2$$

and

$$\tau > \frac{1}{2(\gamma + 1)},$$

then the condition numbers $\kappa(F_M)$ satisfy

$$\kappa(F_M) \geq \sigma^{M^\nu}$$

for some $\sigma > 1$ and all large M, where

$$\nu = \frac{2(\gamma + 1)\tau - 1}{2\gamma + 1}.$$

Proof. We apply Lemma 4.1 and Corollary 3.2. □

This result is best summarized by the statements in following corollary.

COROLLARY 4.3 Consider the setup of Theorem 4.2. If $\gamma = \max\{\alpha, \beta\} > -1/2$ then the following holds:

 (i) If $F_M(f)$ converges exponentially fast with geometric rate for all $f \in B(E)$ (i.e. $\tau = 1$ in (4.2)) then the condition numbers $\kappa(F_M) \geq \sigma^M$ grow exponentially fast with geometric rate.

(ii) The best possible rate of exponential convergence of a stable method F_M is subgeometric with index $\frac{1}{2(\gamma+1)}$.

Proof. We set $\tau = 1$ (part (i)) or $\tau = \frac{1}{2(\gamma+1)}$ (part (ii)) in Theorem 4.2. □

Note that by letting $\gamma = 0$ (i.e. equispaced points) we recover the original impossibility theorem (Theorem 2.1). It is of interest to note that geometric convergence necessarily implies geometrically large condition numbers, regardless of the endpoint behavior of the sampling points whenever $\gamma = \max\{\alpha, \beta\} > -1/2$. Conversely, this result says nothing about points that cluster quadratically or faster at $x = \pm 1$, which corresponds to the case $-1 < \gamma \leq -1/2$. Indeed, we shall prove in the next section that geometric convergence is possible in this setting with a stable approximation.

5. Optimality of the approximation rate and discrete least squares

Theorem 4.2 gives a necessary relation between the rate of exponential convergence and the rate of exponential ill-conditioning. For example, as asserted in Corollary 4.3, stable approximation necessarily implies subgeometric convergence with index $\frac{1}{2(\gamma+1)}$. We now show that there exists an algorithm that achieves these rates: namely, polynomial least-squares fitting. In particular, if the polynomial degree is chosen as

$$N \asymp M^{\nu}, \qquad \nu = \frac{1}{2(\gamma+1)},$$

one obtains a stable approximation which converges exponentially with rate $\frac{1}{2(\gamma+1)}$.

5.1 *A sufficient condition for boundedness of the maximal polynomial*

We commence with a sufficient condition for boundedness of the quantity $B(M, N)$.

LEMMA 5.1 Let $-1 = x_0 < \cdots < x_M = 1$ be arbitrary and suppose that $N\zeta < 1$, where

$$\zeta = \max_{m=0,\dots,M-1} \int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-x^2}}\, \mathrm{d}x. \tag{5.1}$$

If $B(M, N)$ is as in (3.1), then

$$B(M,N) \leq \frac{1}{1-N\zeta}.$$

Proof. Let $p \in \mathbb{P}_N$ with $|p(x_m)| \leq 1$, $m = 0, \dots, M$ and suppose that $-1 \leq x \leq 1$ with $x_m \leq x \leq x_{m+1}$ for some $m = 0, \dots, M - 1$. Then

$$|p(x)| \leq |p(x_m)| + \int_{x_m}^{x} |p'(z)|\, \mathrm{d}z \leq 1 + \sup_{-1 \leq z \leq 1} |\sqrt{1-z^2}p'(z)| \int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-z^2}}\, \mathrm{d}z.$$

Bernstein's inequality states that $|\sqrt{1-z^2}p'(z)| \leq N\|p\|_\infty$ for any $-1 \leq z \leq 1$ (see Borwein & Erdélyi, 1995, Thm. 5.1.7). Hence,

$$|p(x)| \leq 1 + N\zeta\|p\|_\infty.$$

Since $-1 \leq x \leq 1$ was arbitrary the result now follows. $\qquad\square$

Note that this lemma makes no assumption on the points $\{x_m\}_{m=0}^{M}$. The following result estimates the constant ζ for points arising from modified Jacobi weight functions.

LEMMA 5.2 Let μ be a modified Jacobi weight function (2.8) with parameters $\alpha, \beta > -1$. If ζ is as in (5.1) then

$$\zeta \leq CM^{-\frac{1}{2(1+\gamma)}},$$

where $\gamma = \max\{\alpha, \beta, -1/2\}$ and $C > 0$ is a constant depending on α and β only.

Proof. We consider the following four cases:

(i) $-1 < \alpha, \beta \le -1/2$;

(ii) $-1 < \alpha \le -1/2, \beta > -1/2$;

(iii) $\alpha > -1/2, -1 < \beta \le -1/2$;

(iv) $\alpha, \beta > -1/2$.

Case (i): Suppose first that $-1 < \alpha, \beta \le -1/2$. Then

$$\int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-x^2}} \, dx \lesssim \int_{x_m}^{x_{m+1}} \mu(x) \, dx = \frac{1}{M}.$$

Hence, $\zeta \lesssim 1/M$ in this case.

Case (ii): Now suppose that $-1 < \alpha \le -1/2$ and $\beta > -1/2$. Then

$$\int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-x^2}} \, dx \lesssim \int_{x_m}^{x_{m+1}} g(x)(1-x)^\alpha (1+x)^{-1/2} \, dx.$$

Recall from (3.10) that

$$1 + x_m \gtrsim \left(\frac{m}{M}\right)^{\frac{1}{1+\beta}}, \qquad m = 0, \ldots, M,$$

and therefore $1 + x_1 \gtrsim M^{-\frac{1}{1+\beta}}$. For $m = 1, \ldots, M-1$ we have

$$\int_{x_m}^{x_{m+1}} g(x)(1-x)^\alpha (1+x)^{-1/2} \, dx \le (1+x_m)^{-1/2-\beta} \int_{x_m}^{x_{m+1}} g(x)(1-x)^\alpha (1+x)^\beta \, dx$$

$$= (1+x_m)^{-1/2-\beta} \frac{1}{M}$$

$$\le (1+x_1)^{-1/2-\beta} \frac{1}{M}$$

$$\lesssim M^{-\frac{1}{2(1+\beta)}}.$$

Now let $m = 0$. For this we first notice that $x_1 \le 0$ whenever $M \gtrsim 1$. Therefore, we have

$$\int_{x_0}^{x_1} g(x)(1-x)^\alpha (1+x)^{-1/2} \, dx \lesssim \sqrt{1+x_1} \lesssim M^{-\frac{1}{2(1+\beta)}}.$$

Hence, we deduce that $\zeta \lesssim M^{-\frac{1}{2(1+\beta)}}$ for this case as well.

Case (iii): This is identical to the previous case and thus omitted.

Case (iv): For $m = 1, \ldots, M - 2$ we have

$$\int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-x^2}} \, dx \leq (1 - x_{m+1})^{-1/2-\alpha}(1 + x_m)^{-1/2-\beta} \int_{x_m}^{x_{m+1}} g(x)(1-x)^\alpha(1+x)^\beta \, dx$$

$$\leq (1 - x_{M-1})^{-1/2-\alpha}(1 + x_1)^{-1/2-\beta} \frac{1}{M}$$

$$\lesssim M^{1 - \frac{1}{2(1+\alpha)} - \frac{1}{2(1+\beta)}}$$

$$\leq M^{-\frac{1}{2(1+\gamma)}},$$

where in the final step we use the fact that $\alpha, \beta > -1/2$. For $m = 0$, recalling that $x_1 \leq 0$ for all large M, we have

$$\int_{x_0}^{x_1} \frac{1}{\sqrt{1-x^2}} \, dx \lesssim \sqrt{1 + x_1} \lesssim M^{-\frac{1}{2(1+\beta)}},$$

and similarly for $m = M - 1$, noting that $x_{M-1} \geq 0$ for all large M gives

$$\int_{x_{M-1}}^{x_M} \frac{1}{\sqrt{1-x^2}} \, dx \lesssim \sqrt{1 - x_M} \lesssim M^{-\frac{1}{2(1+\alpha)}}.$$

We therefore deduce that $\zeta \lesssim M^{-\frac{1}{2(1+\gamma)}}$ as required.   □

This lemma immediately leads to the following result.

PROPOSITION 5.3 (Necessary and sufficient condition for boundedness of $B(M, N)$).   For $N, M \in \mathbb{N}$, let $\{x_m\}_{m=0}^M$ be equidistributed according to a modified Jacobi weight function (2.8) with parameters $\alpha, \beta > -1$. Then $B(M, N) \lesssim 1$ if and only if

$$M \asymp N^{2(\gamma+1)}.$$

Proof. We combine Corollary 3.2 and Lemma 5.2.   □

The proposition is illustrated in Fig. 5 for cases (C2), (U), (UC) and (OC). It plots the smallest values of M such that $B(M, N) \leq 10$ for given values of N. Notice that the relationship between the computed values of M and N is in good agreement with the asymptotic relation $M \asymp N^{2(\gamma+1)}$. The constants used to define the dashed lines in this figure were chosen by trial and error. Similar agreement is shown in Fig. 6, where the contour levels of $B(M, N)$ for cases (U), (UC) and (OC) are presented. Notice that for the (OC) case $B(M, N)$ remains bounded with $M = CN$, but its values very quickly increase from 10 to more than 10^{13} when C is decreased below 1.65.

REMARK 5.4 For equispaced points, the sufficiency of the rate $M \asymp N^2$ is a classical result (see Remark 2.3). More recently, similar sufficient conditions have appeared for the case where the sampling points are drawn randomly and independently according to the measure $\mu(x) \, dx$. For example, Cohen *et al.* (2013) prove that $M \asymp (N \log(N))^2$ uniformly distributed points are sufficient for L^2/ℓ^2-stability (note that we consider L^∞/ℓ^∞-stability in this paper), whereas only $M \asymp N \log(N)$ points are required when drawn from the Chebyshev distribution. In the multivariate setting, similar results have been
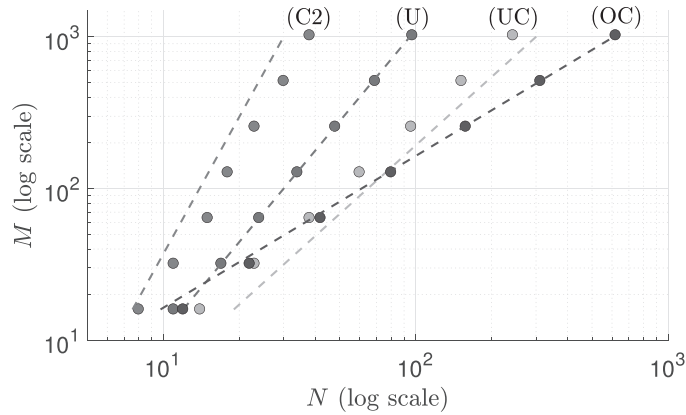
FIG. 5. Log-log plot of N and M, where M is the smallest integer such that $B(M, N) \leq 10$. Circle markers show the computed values and dashed lines represent the theoretical estimate in Proposition 5.3. The dashed lines corresponding to (U), (C2) and (UC) are given by $M = (N/3)^{2(\gamma+1)}$, while the dashed line corresponding to (OC) is $M = 1.65N$.

proved in the works by Migliorati (2013) and Migliorati *et al.* (2014) for quasi-uniform measures. Up to the log factors and the different norms used, these are the same as the rate prescribed in Proposition 5.3, which is both sufficient and necessary.

5.2 *Application to polynomial least squares*

We now apply Proposition 5.3 to show that polynomial least squares achieves the optimal approximation rate of a stable approximation (up to a small algebraic factor in M) specified by the generalized impossibility theorem (Theorem 4.2).

PROPOSITION 5.5 For $M \in \mathbb{N}$, let $\{x_m\}_{m=0}^{M}$ be equidistributed according to a modified Jacobi weight function (2.8) with parameters $\alpha, \beta > -1$. For $1 \leq N \leq M$ let $F_{N,M}$ be the discrete least-squares
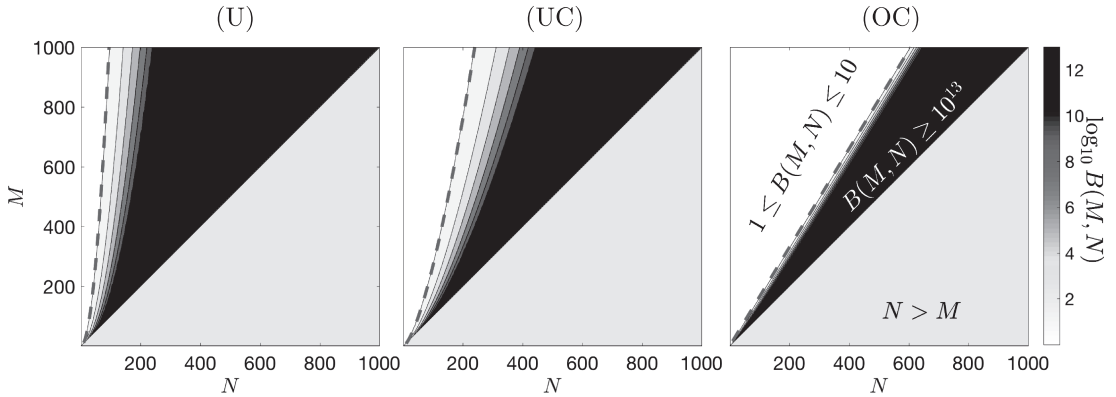

FIG. 6. Contour plot of $\log_{10} B(M, N)$ for node densities (U), (UC) and (OC). Black regions correspond to $B(M, N) > 10^{13}$ and white regions to $B(M, N) < 10$. Dashed lines are given by $M = (N/3)^2$ (U), $M = (N/2.4)^{3/2}$ (UC) and $M = 1.65N$ (OC).

approximation defined by (2.3). If

$$M \asymp N^{2(\gamma+1)}$$

then

$$1 \leq \kappa(F_{M,N}) \leq C\sqrt{M},$$

and

$$\|f - F_{N,M}(f)\|_\infty \leq \frac{C\sqrt{M}}{\theta - 1}\theta^{-M^{\frac{1}{2(\gamma+1)}}}\|f\|_{E_\theta}$$

for all $f \in D(E_\theta)$ and any $\theta > 1$, where $C > 0$ is a constant.

Proof. Proposition 2.4 gives $\kappa(F_{M,N}) \leq \sqrt{2}\sqrt{M}B(M,N)$ and

$$\|f - F_{M,N}(f)\|_\infty \leq 2\sqrt{2}\sqrt{M}B(M,N)\inf_{p \in \mathbb{P}_N}\|f - p\|_\infty.$$

The result follows immediately from Proposition 5.3 and the well-known error bound $\inf_{p \in \mathbb{P}_N}\|f - p\|_\infty \leq \frac{2}{\theta-1}\|f\|_{E_\theta}\theta^{-N}$ (see, for example, Trefethen, 2013, Thm. 8.2).   □

5.3 *The mock-Chebyshev heuristic*

A well-known heuristic is that stable approximation from a set of $M + 1$ points $-1 = x_0 < x_1 < \cdots < x_M = 1$ is possible only once there exists a subset of $N + 1$ of those points that mimic a Chebyshev grid (see, for example, Boyd & Xu, 2009). We now confirm this heuristic. Let

$$z_n = -\cos(n\pi/N), \qquad n = 0, \ldots, N \tag{5.2}$$

be a Chebyshev grid (note that the z_n are equidistributed according to the Chebyshev weight function $\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}$). Then we have the following proposition.

PROPOSITION 5.6 Let $-1 = x_0 < \cdots < x_M = 1$ be arbitrary and suppose that

$$N\zeta < \pi,$$

where ζ is as in (5.1). Then there exists a subset $\{x_n^*\}_{n=1}^N$ of $\{x_m\}_{m=0}^M$ such that

$$-1 = z_0 < x_1^* < z_1 < x_2^* < \cdots < z_{N-1} < x_N^* < z_N = 1, \tag{5.3}$$

where the z_n are as in (5.2). In particular, if $\{x_m\}_{m=0}^M$ are equidistributed according to a modified Jacobi weight function with parameters $\alpha, \beta > -1$ then there exists such a subset $\{x_n^*\}_{n=1}^N$ whenever

$$M \asymp N^{2(\gamma+1)}, \qquad \gamma = \max\{\alpha, \beta, -1/2\}.$$

Proof. Let $\theta_m = \cos^{-1}(-x_m) \in [0, \pi]$ so that

$$\theta_{m+1} - \theta_m = \int_{x_m}^{x_{m+1}} \frac{1}{\sqrt{1-x^2}} \, \mathrm{d}x \leq \zeta < \pi/N. \tag{5.4}$$

We now construct a subset $\{\phi_n\}_{n=1}^{N} \subseteq \{\theta_m\}_{m=0}^{M}$ such that

$$0 < \phi_1 < \frac{\pi}{N} < \phi_2 < \frac{2\pi}{N} < \cdots < \frac{(N-1)\pi}{N} < \phi_N < \pi.$$

First, let m_1 be the largest m such that $\theta_m < \pi/N$. Set $\phi_1 = \theta_{m_1}$. Next, observe that $\theta_{m_1+1} < \theta_{m_1} + \pi/N < 2\pi/N$. Hence, there exists at least one of the θ_m's in the interval $(\pi/N, 2\pi/N)$. Let m_2 be the largest m such that $\theta_m < 2\pi/N$ and set $\phi_2 = \theta_{m_2}$. We now continue in the same way to construct a sequence $\{\phi_n\}_{n=1}^{N}$ with the required property. Since the function $\cos^{-1}(-\theta)$ is increasing on $[0, \pi]$ it follows that the sequence $\{x_n^*\}_{n=1}^{N}$ with $x_n^* = \cos^{-1}(-\phi_n)$ satisfies (5.3). □

Recalling Lemma 5.1 we note that the same sufficient condition (up to a small change in the right-hand side) for boundedness of the maximal polynomial (for arbitrary points) also guarantees an interlacing property of a subset of N of those points with the Chebyshev nodes. In particular, if $M, N \to \infty$ with $N\zeta \leq c$, where $c < \pi$, the nodes $\{x_n^*\}_{n=1}^{N}$ equidistribute according to the Chebyshev weight function $\mu(x) = \frac{1}{\pi\sqrt{1-x^2}}$. Moreover, for modified Jacobi weight functions the sampling rate that guarantees the existence of this 'mock-Chebyshev' grid, i.e. $M \asymp N^{2(\gamma+1)}$, is identical to that which was found to be both necessary and sufficient for stable approximation via discrete least squares (recall Proposition 5.5).

6. Computation of $B(M, N)$ and the maximal polynomial

Let $-1 = x_0 < x_1 < \cdots < x_M = 1$ be a set of $M + 1$ points. We now introduce an algorithm for the computation of

$$B(M, N) = \max\left\{\|p\|_{\infty} : p \in \mathbb{P}_N, \ |p(x_m)| \leq 1, \ m = 0, \cdots, M\right\},$$

and the maximizing polynomial $p \in \mathbb{P}_N$. In fact, in order to compute this polynomial we will first compute the pointwise quantity

$$B(M, N, x) = \max\{|p(x)| : |p(x_m)| \leq 1, \ m = 0, \cdots, M, \ p \in \mathbb{P}_N\}, \quad -1 \leq x \leq 1. \tag{6.1}$$

As we prove below, $B(M, N, x)$ is a piecewise polynomial with knots at the points $\{x_m\}_{m=0}^{M}$. Hence, the maximal polynomial for $B(M, N)$ can be obtained by computing $B(M, N, x)$ in each subinterval and identifying the interval and corresponding polynomial in which the maximum is attained.

Our algorithm for computing (6.1) is a variant of the classical Remez algorithm for computing best uniform approximations (see, for example, Powell, 1981; Pachón & Trefethen, 2009). It is based on a result of Schönhage (1961).

6.1 *Derivation*

We first require some notation. Given a set $Y = \{y_n\}_{n=0}^{N}$ of $N + 1$ points, let

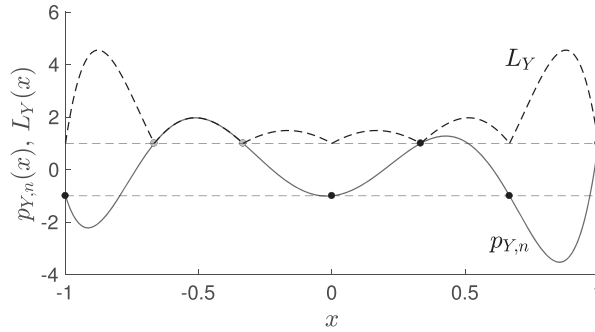$$\ell_{Y,n}(x) = \prod_{\substack{m=0 \\ m \neq n}}^{N} \frac{x - y_m}{y_n - y_m}, \qquad n = 0, \cdots, N$$

FIG. 7. The Lebsgue function L_Y for seven equispaced points and its polynomial representation $p_{Y,n}$ on the interval $[y_1, y_2]$, as defined in Lemma 6.1.

be the Lagrange polynomials, and

$$L_Y(x) = \max\{|p(x)| : p \in \mathbb{P}_N, |p(y_n)| \leq 1, n = 0, \cdots, N\} = \sum_{n=0}^{N} |\ell_{Y,n}(x)|$$

be the Lebesgue function of Y. Note the second equality is a straightforward exercise. We now require the following lemma.

LEMMA 6.1 Let $y_0 < y_1 < \cdots < y_N$. If $x \in [y_n, y_{n+1}]$ then

$$L_Y(x) = p_{Y,n}(x),$$

where $p_{Y,n} \in \mathbb{P}_N$ is the unique polynomial such that

$$p_{Y,n}(y_k) = \begin{cases} (-1)^{n-k}, & k = 0, \ldots, n, \\ (-1)^{n+1-k}, & k = n+1, \ldots, N, \end{cases} \tag{6.2}$$

Proof. Notice that, for $x \in [y_n, y_{n+1}]$, we have

$$\text{sign}\left(\ell_{Y,k}(x)\right) = \begin{cases} (-1)^{n-k}, & k = 0, \cdots, n, \\ (-1)^{n+1-k}, & k = n+1, \cdots, N, \end{cases}$$

Hence,

$$L_Y(x) = \sum_{k=0}^{N} |\ell_{Y,k}(x)| = \sum_{k=0}^{N} \text{sign}\left(\ell_{Y,k}(x)\right)\ell_{Y,k}(x) = \sum_{k=0}^{N} p_{Y,n}(y_k)\ell_{Y,k}(x) = p_{Y,n}(x),$$

as required.   □

This lemma is illustrated in Fig. 7, where L_Y and its polynomial representation $p_{Y,n}$ on the interval $[y_1, y_2]$ are plotted.

LEMMA 6.2 Let $y_0 < y_1 < \cdots < y_N$ and for $n = 0, \ldots, N - 1$ consider the polynomial $p_{Y,n}(x)$ defined in Lemma 6.1. Then

$$p_{Y,n}(x) < 1, \quad x \in (y_{n-1}, y_n) \cup (y_{n+1}, y_{n+2}).$$

Proof. Write $p = p_{Y,n}$ and $I_k = [y_k, y_{k+1}]$ for $k = 0, \ldots, N - 1$. Since $p(y) > 1$ for $y \in I_n \setminus \{y_n, y_{n+1}\}$ and $p(y_n) = p(y_{n+1}) = 1$ there is a point in I_n where p' vanishes. Additionally, $p'(y_{n+1}) < 0$. Suppose now that $p(y) \geq 1$ for some $y \in I_{n+1}$. Then the derivative p' will have at least two zeros in I_{n+1}. The polynomial p has at least $N - 1$ zeros on (y_0, y_N), and since $p(y_n) = p(y_{n+1}) = 1$ it has no zeros on I_n. Hence, it must have exactly one zero on each subinterval I_k for $k \neq n$ and one further zero outside (y_0, y_N). This implies there are at least $N - 1$ zeros of p outside $I_n \cup I_{n+1}$, and therefore p' has at least $N - 3$ zeros outside $I_n \cup I_{n+1}$. Adding the one zero in I_n and the two zeros in I_{n+1} implies that p' has at least N zeros. Since $p \in \mathbb{P}_N$ this is impossible. $\qquad \square$

We now produce our main result that will lead to the Remez-type algorithm. The following result is due to Schönhage (1961). Since the work by Schönhage (1961) is written in German and the relevant result ('Satz 3') is stated for equispaced points only, we reproduce the proof below.

LEMMA 6.3 Let $-1 = x_0 < x_1 < \cdots < x_M = 1$ and $B(M, N, x)$ be as in (6.1). If $x_m < x < x_{m+1}$ for some $m = 0, \ldots, M - 1$ then

$$B(M, N, x) = \min_Y L_Y(x),$$

where the minimum is taken over all sets $Y \subseteq \{x_m\}_{m=0}^M$ of size $|Y| = N + 1$ with $x_m, x_{m+1} \in Y$.

Proof. Consider a set $Y \subseteq \{x_m\}_{m=0}^M$ of size $|Y| = N + 1$ with $x_m, x_{m+1} \in Y$. Then, by definition, $B(M, N, x) \leq L_Y(x)$. Since there are only finitely many such Y, there is a Y^* with

$$L_{Y^*}(x) = \min_Y L_Y(x).$$

Let $Y^* = \{y_k\}_{k=0}^N$ and $p = p_{Y^*,n}$ be the polynomial defined in Lemma 6.1, where n is such that $y_n = x_m$. We now claim that

$$|p(x_j)| \leq 1, \quad j = 0, \cdots, M. \tag{6.3}$$

We shall prove this claim in a moment, but let us first note that this implies the lemma. Indeed, assuming (6.3) holds we have

$$p(x) \leq B(M, N, x) \leq L_{Y^*}(x) = p(x).$$

Hence, $B(M, N, x) = p(x) = L_{Y^*}(x)$ as required.

To prove the claim we argue by contradiction. Suppose that (6.3) does not hold and let j be such that $|p(x_j)| > 1$. Note that $x_j \notin Y^*$. There are now three cases:

Case 1. Suppose x_j lies between two adjacent points of Y^*, i.e.

$$y_k < x_j < y_{k+1} \tag{6.4}$$

for some $k = 0, \ldots, N - 1$. Since $\mathrm{sign}(p(y_{k+1})) = -\mathrm{sign}(p(y_k))$ there are two subcases:

$$\text{(a) } \mathrm{sign}(p(x_j)) = \mathrm{sign}(p(y_k)); \qquad \text{(b) } \mathrm{sign}(p(x_j)) = \mathrm{sign}(p(y_{k+1})).$$

Suppose that subcase (a) occurs. Exchange y_k with x_j and define the new set

$$\widehat{Y} = (Y^* \setminus \{y_k\}) \cup \{x_j\} = \{\hat{y}_i\}_{i=0}^N.$$

We now claim that $x_m, x_{m+1} \in \widehat{Y}$; in other words, $j \neq m, m+1$. First, notice that $j \neq m$ since (6.4) cannot hold when $k = n$. Second, $j = m + 1$ cannot hold either. Indeed, if $j = m + 1$ then $\text{sign}(p(x_j)) = \text{sign}(p(x_{m+1})) = 1$ and hence $p(x_j) > 1$. But by Lemma 6.2 we have $p(x) < 1$ for $y_{n+1} < x < y_{n+2}$, which is a contradiction.

Let ℓ_i and $\hat{\ell}_i$ be the Lagrange polynomials for Y^* and $\widehat{Y}$, respectively. Then, for $x_m < x < x_{m+1}$, we have

$$\text{sign}\left(\hat{\ell}_i(x)\right) = \text{sign}\left(\ell_i(x)\right) = \text{sign}\left(p(y_i)\right) = \text{sign}\left(p(\hat{y}_i)\right). \tag{6.5}$$

Hence, expanding p in the Lagrange polynomials $\hat{\ell}_i$, we obtain

$$L_{Y^*}(x) = p(x) = \sum_{i=0}^{N} p(\hat{y}_i)\hat{\ell}_i(x) = \sum_{i=0}^{N} |p(\hat{y}_i)||\hat{\ell}_i(x)| > \sum_{i=0}^{N} |\hat{\ell}_i(x)| = L_{\widehat{Y}}(x),$$

which contradicts the minimality of Y^*. Subcase (b) is treated in a similar manner.

Case 2. Suppose that $x_j > y_N$. In this case we have the two subcases

$$\text{(a) } \text{sign}(p(x_j)) = \text{sign}(p(y_N)); \qquad \text{(b) } \text{sign}(p(x_j)) = -\text{sign}(p(y_N)).$$

In subcase (a) we construct $\hat{Y}$ by replacing y_N with x_j and, similarly to case 1, arrive at a contradiction.

Now consider subcase (b). We first note that $x_m \neq y_0$. Indeed, if this were the case then p would have $N + 1$ zeros—namely, $N - 1$ zeros between y_1 and y_N, one zero between y_N and x_j and one zero to the left of y_0—which is a contradiction. Hence, we can exchange y_0 with x_j to construct a new set of points

$$\widehat{Y} = \left(Y^* \backslash \{y_0\}\right) \cup \{x_j\} = \{y_i\}_{i=1}^{N+1},$$

where $y_{N+1} = x_j$. The Lagrange polynomials on $\widehat{Y}$ satisfy

$$\text{sign}\left(\hat{\ell}_i(x)\right) = \text{sign}\left(\ell_i(x)\right) = \text{sign}\left(p(y_i)\right), \quad i = 1, \cdots, N$$

and

$$\text{sign}\left(\hat{\ell}_{N+1}(x)\right) = -\text{sign}\left(\ell_N(x)\right) = -\text{sign}\left(p(y_N)\right) = \text{sign}\left(p(y_{N+1})\right).$$

As before, it follows that $L_{Y^*}(x) > L_{\widehat{Y}}(x)$, contradicting the minimality of Y^*.

Case 3. This is similar to case 2, and hence omitted. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Note that a particular consequence of this lemma is that, as claimed, the function $B(M, N, x)$ is a polynomial on each subinterval $[x_m, x_{m+1}]$.

6.2 *A Remez-type algorithm for computing B(M, N, x)*

Lemma 6.3 not only gives an expression for $B(M, N, x)$, its proof also suggests a numerical procedure for its computation. The algorithm follows the steps of the proof and proceeds roughly as follows. First, a set Y of the form described in Lemma 6.3 is chosen and the polynomial $p = p_{Y,n}$ of Lemma 6.1 is computed. If (6.3) holds, then, as shown in the proof of Lemma 6.3, $p(x) = B(M, N, x)$. If not, then

a point $x^* \in \{x_m\}_{m=0}^M$ that maximizes $|p(x_m)|$ is found, and, following the proof once more, a suitable element y_k of Y is exchanged with x^* to construct a new set $\widehat{Y}$. This process is repeated until (6.3) holds.

ALGORITHM 6.4 (First Remez-type algorithm).

1. Pick a subset $Y \subseteq \{x_m\}_{m=0}^M$ with $|Y| = N + 1$ and $x_p, x_{p+1} \in Y$.

2. Compute the polynomial $p = p_{Y,n} \in \mathbb{P}_N$ satisfying (6.2), where n is such that $x_n = x_p$.

3. Find a point $x^* \in \{x_m\}_{m=0}^M$ with

$$|p(x^*)| \max_{m=0,\cdots,M} |p(x_m)|.$$

4. If $|p(x^*)| = 1$ then set $B(M, N, x) = p(x)$ and stop.

5. If $|p(x^*)| > 1$ then proceed as follows:

   (a) Suppose that $y_k < x^* < y_{k+1}$ for some $k = 0, \ldots, N - 1$. If $\text{sign}(p(x^*)) = \text{sign}(p(y_k))$ then replace y_k with x^* in the set Y; otherwise replace y_{k+1} with x^*.
   (b) Suppose that $x^* < y_0$. If $\text{sign}(p(x^*)) = \text{sign}(p(y_0))$ then replace y_0 with x^* in the set Y; otherwise replace y_N with x^*.
   (c) Suppose that $x^* > y_N$. If $\text{sign}(p(x^*)) = \text{sign}(p(y_N))$ then replace y_N with x^* in the set Y; otherwise replace y_0 with x^*.

6. Return to step 2.

This algorithm is guaranteed to converge in a finite number of steps. As shown in the proof of Lemma 6.3, the exchange performed in step 5 strictly decreases the value $L_Y(x)$. Since there are only finitely many possible sets Y, the algorithm must therefore terminate in finite time.

In practice, it is usually preferable to exchange more than one point x^* at a time. This leads to the following algorithm.

ALGORITHM 6.5 (Second Remez-type algorithm).

The algorithm is similar to Algorithm 6.4, except that in step 3 we find all extrema of $p(x)$ on the set $\{x_m\}_{m=0}^M$. Note that there are at least $N - 3$ and at most $N - 2$ such points.[3] Each subinterval $[y_k; y_{k+1}]$ contains at most one of these extrema. Hence, we now proceed with the exchange as in step 5 above for each such point.

6.3 Numerical results

The performance of the first and second Remez-type algorithms is presented in Fig. 8, where the maximal polynomial of degree 300 over the interval $[x_{200}, x_{201}]$ is computed for the (OC) case. The first Remez-type algorithm takes over 80 iterations to converge, while the second type computes the maximal polynomial in 18 iterations. In this experiment, the algorithm was started using the subset $Y \subseteq \{x_m\}_{m=0}^{500}$

---

[3] The polynomial p of degree N has a full set of N zeros, hence $N - 1$ extrema. Of those, one extremum is between x_m and x_{m+1}, and at most one extremum might be outside the interval under consideration. Hence, the number of extrema that may appear in the algorithm is $N - 2$ at most, and $N - 3$ at least.

[3] The polynomial p of degree N has a full set of N zeros, hence $N - 1$ extrema. Of those, one extremum is between x_m and x_{m+1}, and at most one extremum might be outside the interval under consideration. Hence, the number of extrema that may appear in the algorithm is $N - 2$ at most, and $N - 3$ at least.
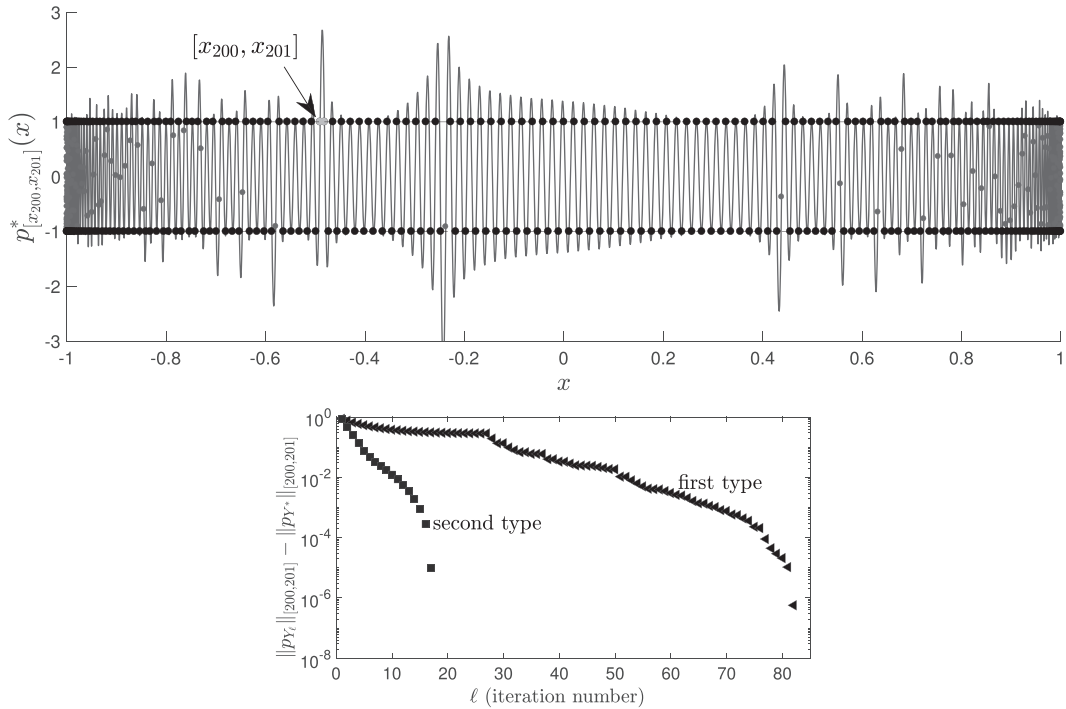
FIG. 8. Top: maximal polynomial of degree 300 over the interval $[x_{200}, x_{201}]$, where $\{x_m\}_{m=0}^M$ is the set of 500 points drawn from the (OC) density. Dot markers show the value of the polynomial at the grid x_m, with black dots corresponding to points for which the polynomial evaluates to ± 1. Bottom: convergence plot of the first and second Remez-type algorithms to this polynomial.

of 301 points closest to the Chebyshev points of the second kind, that is, the mock-Chebyshev subset. The iteration count can be significantly higher when the initial subset of points is selected at random. We also point out that the algorithm may fail to converge if at any iteration the interpolation set Y is very ill conditioned. In double precision the Lebesgue constant for Y must not exceed 10^{16}. Choosing mock-Chebyshev points to initialize the procedure, therefore, reduces the likelihood of failed iterations.

To find the maximal polynomial over the whole interval the Remez procedure must be repeated for every subinterval $[x_m, x_{m+1}]$, unless the location where the maximum is achieved is known. In the case of equispaced nodes the maximum is known to be near the endpoints. In the (OC) case, the maximum is in the interior of the interval as illustrated in Fig. 9, but not necessarily in the subinterval closest to the center as shown in the bottom right panel. Although only moderate values of M and N were used in this figure, the Remez algorithm is able to compute maximal polynomials of much larger degrees, as shown in Figs 5 and 6.

7. Concluding remarks

We have presented a generalized impossibility theorem for approximating analytic functions from $M+1$ nonequispaced points. This follows from a new lower bound for the maximal behavior of a polynomial of degree N that is bounded on an arbitrary set of points. By specializing to modified Jacobi weight functions, we have derived explicit relationships between the parameter $\gamma = \max\{\alpha, \beta, -1/2\}$, the rate
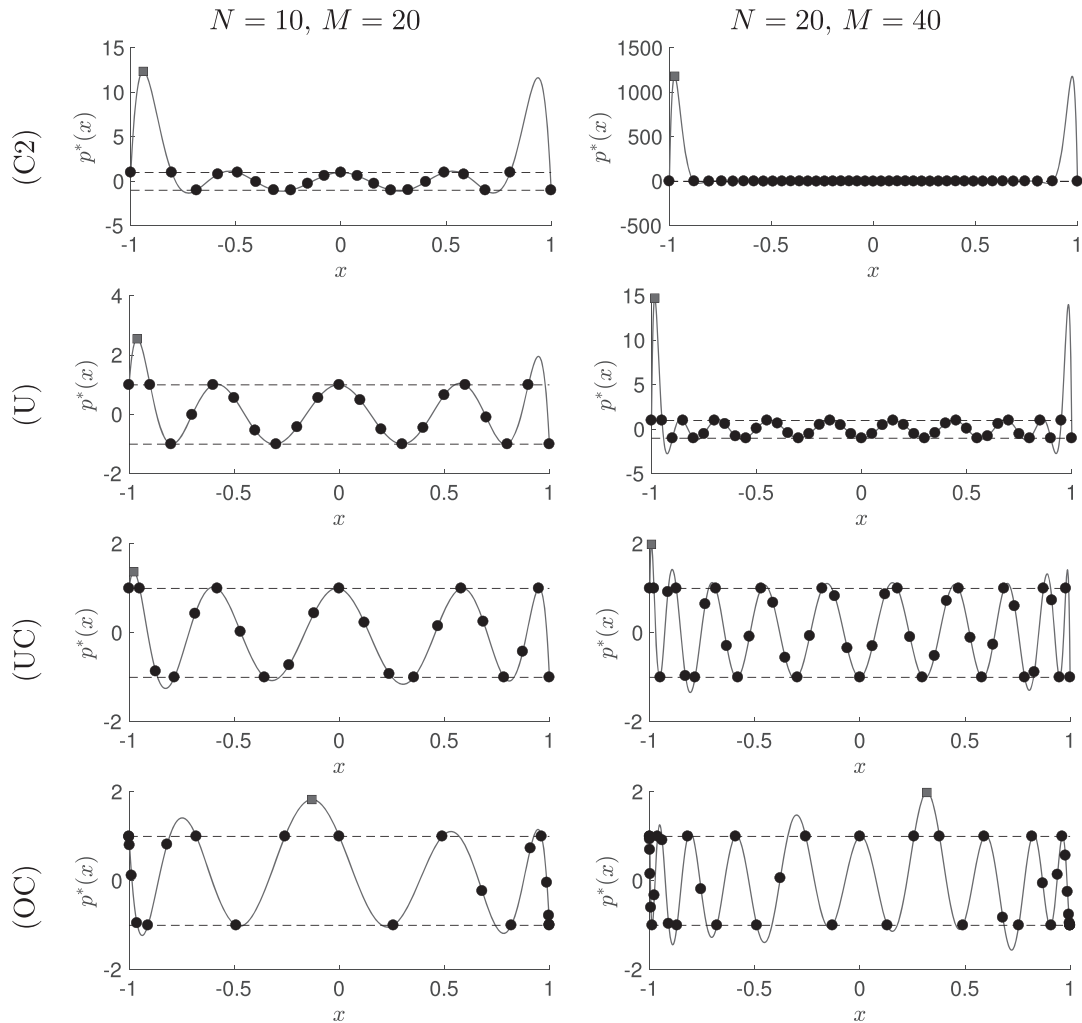
FIG. 9. Maximal polynomials for different node sets. Square markers show the maximum point for each case, while the dot markers show the value of the polynomials at the grid points.

of exponential convergence and the rate of exponential ill-conditioning. Polynomial least squares using a polynomial of degree N transpires to be an optimal stable method in view of this theorem. In particular, the sampling rate $M \asymp N^{2(\gamma+1)}$, where $\gamma = \max\{\alpha, \beta, -1/2\}$, is both sufficient and necessary for stable approximation with optimal convergence.

There are a number of directions for future investigation. First, we have derived an upper bound for $B(M, N)$ only in the case of ultraspherical weight functions (see Remark 3.4). We expect that the techniques of Rakhmanov (2007) can be extended to the modified Jacobi case whenever $\max\{\alpha, \beta\} > -1/2$. Second, we have observed numerically that there is an exponential blowup for $B(M, N)$ in the case $-1 < \gamma < -1/2$ when $M = cN$ for some $c > 0$ below a critical threshold. This remains to be proven. Third, we have mentioned in passing recent results on sufficient sampling rates when drawing random

points from measures associated with modified Jacobi weight functions. It would be interesting to see whether the techniques used in the paper could also establish the necessity (in probability) of those rates. Fourth, the extension of our results to analytic functions of two or more variables remains an open problem. Fifth, we note in passing that there is a related impossibility theorem for approximating analytic functions from their Fourier coefficients (see Adcock *et al.*, 2014b) (this is in some senses analogous to the case of equispaced samples). The extension to nonharmonic Fourier samples may now be possible using the techniques of this paper. Note that necessary and sufficient sampling conditions for this problem have been proven in Adcock *et al.* (2016) and Adcock *et al.* (2014a, 2015), respectively.

Finally, we make the following remark. The impossibility theorems proved here and originally in Platte *et al.* (2011) assert the lack of existence of stable numerical methods with rapid convergence. They say nothing about methods for which the error decays only down to some finite tolerance. If the tolerance can be set on the order of machine epsilon, the limitations of such methods in relation to methods that have theoretical convergence to zero may be of little consequence in finite precision calculations. Several methods with this behavior have been developed in previous works by Adcock *et al.* (2014c); Adcock & Platte (2016). The existence (or lack thereof) of impossibility theorems in this finite setting is an open question.

REFERENCES

ABRAMOWITZ, M. & STEGUN, I. A. (1974) *Handbook of Mathematical Functions*. New York: Dover.

ADCOCK, B. (2017) Infinite-dimensional ℓ^1 minimization and function approximation from pointwise data. *Constr. Approx.*, **45**, 343–390.

ADCOCK, B., GATARIC, M. & HANSEN, A. C. (2014a) On stable reconstructions from nonuniform Fourier measurements. *SIAM J. Imaging Sci.*, **7**, 1690–1723.

ADCOCK, B., GATARIC, M. & HANSEN, A. C. (2015) Recovering piecewise smooth functions from nonuniform Fourier measurements. *Proceedings of the 10th International Conference on Spectral and High Order Methods* (R. M. Kirby, M. Berzins & J. S. Hesthaven eds). Cham: Springer.

ADCOCK B., GATARIC, M. & ROMERO, J. L. (2016) Computing reconstructions from nonuniform Fourier samples: universality of stability barriers and stable sampling rates. arXiv:1606.07698.

ADCOCK, B., HANSEN, A. C. & SHADRIN, A. (2014b) A stability barrier for reconstructions from Fourier samples. *SIAM J. Numer. Anal.*, **52**, 125–139.

ADCOCK, B., HUYBRECHS, D. & MARTÍN-VAQUERO, J. (2014c) On the numerical stability of Fourier extensions. *Found. Comput. Math.*, **14**, 635–687.

ADCOCK, B. & PLATTE, R. (2016) A mapped polynomial method for high-accuracy approximations on arbitrary grids. *SIAM J. Numer. Anal.*, **54**, 2256–2281.

BERNSTEIN, S. (1912) *Sur l'Ordre de la Meilleure Approximation des Fonctions Continues par des Polynomes de Degré Donné*. *Mém. Acad. Roy. Belg.* 1–104.

BINEV, P., COHEN, A., DAHMEN, W., DEVORE, R. A., PETROVA, G. & WOJTASZCZYK, P. (2016) Data assimilation in reduced modeling. *SIAM/ASA J. Uncertain. Quantif.* In press.

BORWEIN, P. & ERDÉLYI, T. (1995) *Polynomials and Polynomial Inequalities*. New York: Springer.

BOYD, J. P. & ONG, J. R. (2009) Exponentially-convergent strategies for defeating the Runge phenomenon for the approximation of non-periodic functions. I. Single-interval schemes. *Commun. Comput. Phys.*, **5**, 484–497.

BOYD, J. P. & XU, F. (2009) Divergence (Runge phenomenon) for least-squares polynomial approximation on an equispaced grid and mock-Chebyshev subset interpolation. *Appl. Math. Comput.*, **210**, 158–168.

CHKIFA, A., COHEN, A., MIGLIORATI, G., NOBILE, F. & TEMPONE, R. (2015) Discrete least squares polynomial approximation with random evaluations-application to parametric and stochastic elliptic pdes. *ESAIM Math. Model. Numer. Anal.*, **49**, 815–837.

COHEN, A., DAVENPORT, M. A. & LEVIATAN, D. (2013) On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, **13**, 819–834.

COPPERSMITH, D. & RIVLIN, T. (1992) The growth of polynomials bounded at equally spaced points. *SIAM J. Math. Anal.*, **23**, 970–983.

DEMANET, L. & TOWNSEND, A. (2016) Stable extrapolation of analytic functions. arXiv:1605.09601.

DEVORE, R. A., PETROVA, G. & WOJTASZCZYK, P. (2016) Data assimilation and sampling in Banach spaces. arXiv:1602.06342.

EHLICH, H. (1966) Polynome zwischen Gitterpunkten. *Math. Zeit.*, **93**, 144–153.

EHLICH, H. & ZELLER, K. L. (1964) Schwankung von Polynomen zwischen Gitterpunkten. *Math. Zeit.*, **86**, 41–44.

EHLICH, H. & ZELLER, K. L. (1965) Numerische Abschätzung von Polynomen. *Z. Agnew. Math. Mech.*, **45**, T20–T22.

MIGLIORATI, G. (2013) Polynomial approximation by means of the random discrete L^2 projection and application to inverse problems for PDEs with stochastic data. *Ph.D. Thesis,* Politecnico di Milano, Milano.

MIGLIORATI, G., NOBILE, F., VON SCHWERIN E. & TEMPONE, R. (2014) Analysis of the discrete L^2 projection on polynomial spaces with random evaluations. *Found. Comput. Math.*, **14**, 419–456.

NARAYAN, A. & ZHOU, T. (2015) Stochastic collocation on unstructured multivariate meshes. *Commun. Comput. Phys.*, **18**, 1–36.

PACHÓN, R. & TREFETHEN, L. N. (2009) Barycentric–Remez algorithms for best polynomial approximation in the chebfun system. *BIT*, **49**, 721–741.

PLATTE, R. & KLEIN, G. (2016) A comparison of methods for recovering analytic functions from equispaced samples. In preparation.

PLATTE, R., TREFETHEN, L. N. & KUIJLAARS, A. (2011) Impossibility of fast stable approximation of analytic functions from equispaced samples. *SIAM Rev.*, **53**, 308–318.

POWELL, M. J. D. (1981) *Approximation Theory and Methods*. Cambridge: Cambridge University Press.

RAKHMANOV, E. A. (2007) Bounds for polynomials with a unit discrete norm. *Ann. Math.*, **165**, 55–88.

SCHÖNHAGE, A. (1961) Fehlerfortpflantzung bei Interpolation. *Numer. Math.*, **3**, 62–71.

TREFETHEN, L. N. (2013) *Approximation Theory and Approximation Practice*. Philadelphia: SIAM.