

ERROR LOCALIZATION OF BEST L_1 POLYNOMIAL APPROXIMANTS*

YUJI NAKATSUKASA[†] AND ALEX TOWNSEND[‡]

Abstract. An important observation in compressed sensing is that the ℓ_0 minimizer of an underdetermined linear system is equal to the ℓ_1 minimizer when there exists a sparse solution vector and a certain restricted isometry property holds. Here, we develop a continuous analogue of this observation and show that the best L_0 and L_1 polynomial approximants of a polynomial that is corrupted on a set of small measure are nearly equal. We demonstrate an error localization property of best L_1 polynomial approximants and use our observations to develop an improved algorithm for computing best L_1 polynomial approximants to continuous functions.

Key words. polynomial approximation, best L_1 , compressed sensing, best L_0 , restricted isometry property, error localization

AMS subject classifications. 65F15, 15A18, 15A22

DOI. 10.1137/19M1242860

1. Introduction. In compressed sensing, the ℓ_0 minimizer of an underdetermined linear system $Ax = b$ can be exactly recovered by the ℓ_1 minimizer when the ℓ_0 minimizer is sufficiently sparse and A satisfies some regularity conditions [11, 15, 18]. Similarly, when an acquired signal is sparsely corrupted, one can exactly recover the original signal by minimizing the ℓ_1 error, under suitable assumptions [13]. In this paper, we investigate a continuous analogue of this phenomenon and show that the best L_0 and L_1 polynomial approximants of corrupted polynomials (see Definition 1.1) are equal, under suitable assumptions (see section 2). We also make precise a related observation that the best L_1 error can be concentrated to intervals of small measure, showing that they can be advantageous compared to minimax approximants for certain applications (see [30]).

Let $f : [-1, 1] \rightarrow \mathbb{R}$ be a continuous function and $n \geq 0$ an integer. The best L_1 polynomial approximant, $p_n^{L_1}$, of degree $\leq n$ to f exists, is unique [27, Thm. 14.3], and satisfies

$$(1.1) \quad \|f - p_n^{L_1}\|_1 = \min_{p \in \mathcal{P}_n} \|f - p\|_1, \quad \|f - p\|_1 = \int_{-1}^1 |f(x) - p(x)| dx,$$

where \mathcal{P}_n is the space of polynomials of degree $\leq n$. While the minimax approximant, $p_n^{L_\infty}$, is the best approximant in the sense that $\|f - p_n^{L_\infty}\|_\infty = \min_{p \in \mathcal{P}_n} \|f - p\|_\infty$, where $\|\cdot\|_\infty$ is the maximum norm, we know by the equioscillation theorem that the maximum deviation is attained $\geq n + 2$ times [27, Thm. 7.2]. On the other hand, it can frequently be observed that $|f(x) - p_n^{L_1}(x)| \ll \|f - p_n^{L_\infty}\|_\infty$ for most, but not all,

*Received by the editors February 7, 2019; accepted for publication (in revised form) November 5, 2020; published electronically February 1, 2021.

<https://doi.org/10.1137/19M1242860>

Funding: The work of the authors was partially supported by an extended collaboration visit in December 2018 to the National Institute of Informatics in Tokyo, where the majority of this research took place. The work of the first author was supported by the JSPS grants 17H01699 and 18H05837. The work of the second author was supported by the National Science Foundation grant 1818757.

[†]Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK (Yuji.Nakatsukasa@maths.ox.ac.uk).

[‡]Department of Mathematics, Cornell University, Ithaca, NY 14853 USA (townsend@cornell.edu).

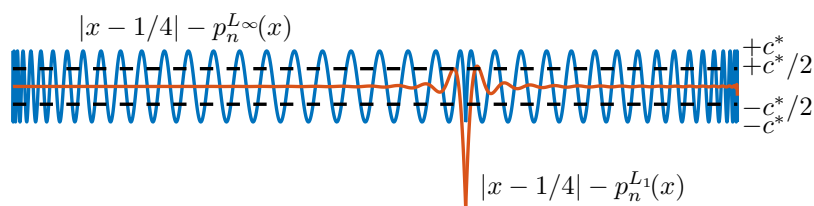


FIG. 1.1. The errors $f(x) - p_n^{L_\infty}(x)$ (blue line) and $f(x) - p_n^{L_1}(x)$ (red line) for $f(x) = |x - 1/4|$ on $[-1, 1]$ when $n = 80$. While $f(x) - p_n^{L_\infty}(x)$ has a smaller absolute maximum on $[-1, 1]$, we find that $|f(x) - p_n^{L_1}(x)| \leq c^*/2$ for most x in $[-1, 1]$, where $c^* = \|f - p_n^{L_\infty}\|_\infty$. Similar illustrations can be found in [31, Chap. 16] and [30].

$x \in [-1, 1]$ (see Figure 1.1 and section 4). To make this observation precise, we define the set¹

$$(1.2) \quad \Omega_n = \left\{ x \in [-1, 1] : |f(x) - p_n^{L_1}(x)| \geq \frac{1}{2} \|f - p_n^{L_\infty}\|_\infty \right\}.$$

For any $x \in [-1, 1] \setminus \Omega_n$ we know that $p_n^{L_1}(x)$ is a better approximation to $f(x)$ than $p_n^{L_\infty}(x)$. By the definition of $p_n^{L_\infty}$, Ω_n is not the empty set, but we often observe that $|\Omega_n| \rightarrow 0$ as $n \rightarrow \infty$ (see section 4). For example, in section 4 we prove that $|\Omega_n| = \mathcal{O}(n^{-2} \log n)$ for $f(x) = \sqrt{1 - x^2}$ and $|\Omega_n| = \mathcal{O}(n^{-1})$ for $f(x) = |x|$. In such cases we say that the error $f - p_n^{L_1}$ is “highly localized.” This property of best L_1 approximation seems to be underappreciated and is related to observations from compressed sensing.

The highly localized nature of $f - p_n^{L_1}$ means that best L_1 polynomial approximation is ideal for recovering functions that have been arbitrarily corrupted on a set of small measure.

DEFINITION 1.1. For $0 \leq s < 1$, we say that a function $f : [-1, 1] \rightarrow \mathbb{R}$ is a s -corrupted function if f can be written as

$$f(x) = g(x) + \omega(x),$$

where $g : [-1, 1] \rightarrow \mathbb{R}$ is a continuous function, $\omega(x)$ is a measurable function with $|\text{supp}(\omega)| \leq s$, and $|\text{supp}(\omega)|$ denotes the Lebesgue measure of the support of ω on $[-1, 1]$. Note that the support of ω , denoted by $\text{supp}(\omega)$, is a closed subset of $[-1, 1]$.

If $g = p_m$ is a polynomial of degree $\leq m$ in Definition 1.1, then we say that f is a corrupted polynomial. If, in addition, $s < \min(1, 1/(4n^2))$ for some integer $n \geq m$, then one finds that the best L_1 polynomial approximant of degree $\leq n$ to f is unique and $p_n^{L_1} = p_m$ (see Corollary 2.4). This means that the best L_1 approximation exactly recovers a corrupted polynomial with arbitrary corruption, provided that the corruption has small enough support.

Figure 1.2 illustrates the four regimes that one typically observes with best L_1 approximants of degree $\leq n$ of $f = p_m + \omega$: (a) If $n < m$, then $p_n^{L_1} \neq p_m$, but $p_n^{L_1}$ is a near-best approximant to p_m (see section 3). (b) If n is small and $n \geq m$, then one gets exact recovery as $p_n^{L_1} = p_m$ (see Corollary 2.4). (c) If n is a little larger, then $p_n^{L_1}$ tries to fit corruptions near ± 1 but not the corruptions away from ± 1 . (d) When n is large, $p_n^{L_1}$ tries to fit all the corruption, resulting in an overfit.

¹The constant of $1/2$ in the definition of Ω_n (see (1.2)) is an arbitrary choice as any constant in $(0, 1)$ would do, with very minor changes to the results that we derive.

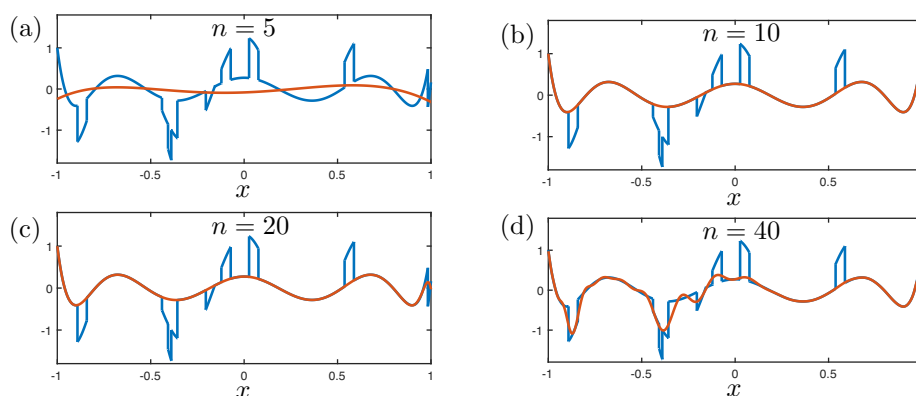


FIG. 1.2. Best L_1 polynomial approximants of degree $n = 5$ (see (a)), $n = 10$ (see (b)), $n = 20$ (see (c)), and $n = 40$ (see (d)) to a s -corrupted Legendre polynomial of degree 8 with $s \approx 0.349$. For this example, we find the following: When $n < 8$, $p_n^{L_1}$ does not recover the polynomial before it was corrupted (see (a)). When $8 \leq n \leq 15$, $p_n^{L_1}$ perfectly recovers the polynomial before it was corrupted (see (b)). When $16 \leq n \leq 26$, $p_n^{L_1}$ tries to fit corruptions near ± 1 but not corruptions away from ± 1 (see (c)). When $n > 27$, $p_n^{L_1}$ tries to fit corruptions away from ± 1 too (see (d)).

We go on to derive an efficient algorithm for the recovery of p_m from f by showing that the continuous optimization problem in (1.1) for $p_n^{L_1}$ can be reduced to a linear programming problem, provided that a sampling condition is satisfied (see Theorem 2.1). This observation results in a computationally efficient algorithm for the exact recovery of corrupted polynomials (see subsection 2.3).

It is worth emphasizing that the Lebesgue measure of the support of the corruption must be extremely small. For example, our theory only guarantees that a corrupted polynomial of degree 100 can be exactly recovered if it is corrupted on a set of measure $\leq 2.5 \times 10^{-5}$. Nevertheless, in practice, we observe that exact recovery is usually still possible when the corruption occurs on sets that have a much larger measure. Moreover, the distribution of the corruption in $[-1, 1]$ does matter. In particular, larger regions of corruption are allowed away from ± 1 , and we present an initial result in this direction (see Theorem A.1). For example, when $n = 100$ exact recovery is still guaranteed with any corruption interval of the form $[-s/2, s/2]$ with $s \leq 4 \times 10^{-4}$.

The error localization properties of best L_1 approximants lead to an iterative algorithm for computing $p_n^{L_1}$ given a continuous function $f : [-1, 1] \rightarrow \mathbb{R}$, based on a combination of linear programming and Newton's method (see section 5). This can be seen as an improvement on Watson's algorithm [20, 34]. Our algorithm allows for the zero set of $f - p_n^{L_1}$ to have a positive measure and heavily employs algorithmic advances over the last decade in polynomial rootfinding and adaptive Chebyshev interpolants [4, 25]. In particular, our implementation greatly benefits from the adaptive and robust algorithms for computing with functions in Chebfun.² It can accurately compute best L_1 approximants of degrees in the thousands (see section 5).

In addition to the L_1 -norm (see (1.1)), we also define the following for continuous functions $f : [-1, 1] \rightarrow \mathbb{R}$:

²Chebfun is an object-oriented software system written in MATLAB that provides an environment to compute with piecewise smooth functions [25]. It represents univariate functions defined on a finite interval by piecewise Chebyshev interpolants of adaptively selected degrees that are accurate to essentially machine precision [16].

$$(1.3) \quad \begin{aligned} \|f\|_\infty &= \max_{x \in [-1, 1]} |f(x)|, & \|f\|_{\ell_1} &= \sum_{j=0}^N w_j |f(x_j)|, \\ \|f\|_2 &= \sqrt{\int_{-1}^1 f(x)^2 dx}, & \|f\|_{\ell_0} &= \#\{j : |f(x_j)| > 0, \quad 0 \leq j \leq N\}, \end{aligned}$$

where $w_j \geq 0$ are weights so that $\sum_{j=0}^N w_j |f(x_j)| \rightarrow \int_{-1}^1 |f(x)| dx$ as $N \rightarrow \infty$. Despite the notation, $\|\cdot\|_{\ell_0}$ is not a norm. For completeness, we also define $\|f\|_0 = |\text{supp}(f)|$ as the Lebesgue measure of the support of f . We always take x_0, \dots, x_N in the discrete norms $\|f\|_{\ell_1}$ and $\|f\|_{\ell_0}$ to be the roots of the degree $N+1$ Chebyshev polynomial of the second kind U_{N+1} [24, Table 18.3.1]. That is,

$$(1.4) \quad x_j = \cos\left(\frac{(N+1-j)\pi}{N+2}\right), \quad 0 \leq j \leq N.$$

Accordingly, we take $w_j = \pi \sqrt{1 - x_j^2} / (N+2)$ in (1.3) so that the corresponding quadrature rule is related to the Gauss–Chebyshev rule. The Chebyshev polynomials of the second kind and their roots in (1.4) play a special role in best L_1 approximation [27, Chap. 14]. In particular, when $N = n$, the polynomial interpolant of f at the points in (1.4), i.e.,

$$(1.5) \quad p_n^{\text{cheb}}(x) = \sum_{j=0}^n f(x_j) \ell_j(x), \quad \ell_j(x) = \frac{\prod_{i=0, i \neq j}^n (x - x_i)}{\prod_{i=0, i \neq j}^n (x_j - x_i)},$$

is the best L_1 polynomial approximation of degree $\leq n$ to f if $f - p_n^{\text{cheb}}$ has exactly $n+1$ distinct zeros in $[-1, 1]$ [8, 26].

For an integer $n \geq 0$, we denote by $p_n^{L_\infty}$, $p_n^{L_2}$, $p_n^{\ell_1}$, and $p_n^{\ell_0}$ any best L_∞ , L_2 , ℓ_1 , and ℓ_0 polynomial of degree $\leq n$ to f , respectively. These polynomials are solutions to the following optimization problems:

$$(1.6) \quad \begin{aligned} p_n^{L_\infty} &= \arg \min_{q \in \mathcal{P}_n} \|f - q\|_\infty, & p_n^{\ell_1} &= \arg \min_{q \in \mathcal{P}_n} \|f - q\|_{\ell_1}, \\ p_n^{L_2} &= \arg \min_{q \in \mathcal{P}_n} \|f - q\|_2, & p_n^{\ell_0} &= \arg \min_{q \in \mathcal{P}_n} \|f - q\|_{\ell_0}. \end{aligned}$$

We also define $p_n^{L_0} = \arg \min_{q \in \mathcal{P}_n} \|f - q\|_0$, when best polynomial in this sense exists.

The paper is structured as follows. In section 2, we show that the exact recovery of an arbitrarily corrupted polynomial is possible provided that the support of the corruption has small enough measure. This leads to an efficient algorithm to achieve recovery. In section 3, we extend these ideas to the near-recovery of corrupted smooth functions. In section 4, we show that $|\Omega_n|$ is small precisely when $\|f - p_n^{L_1}\|_1 \rightarrow 0$ faster than $\|f - p_n^{L_\infty}\|_\infty \rightarrow 0$ as $n \rightarrow \infty$ and carefully consider two worked examples with error localization. Finally, in section 5, we present our iterative algorithm for computing best L_1 polynomial approximants of continuous functions.

2. Exact recovery of corrupted polynomials. In this section we suppose that $f : [-1, 1] \rightarrow \mathbb{R}$ is formed by an arbitrarily corrupted polynomial, i.e., $f = p_m + \omega$, where p_m is a polynomial of degree $\leq m$ and ω is a function with small support. We investigate the question, When is it possible to exactly recover p_m from knowledge of f ?

We show that for corrupted polynomials, we have $p_n^{L_0} = p_n^{\ell_0} = p_n^{\ell_1} = p_n^{L_1} = p_m$ provided that the support of ω is sufficiently small, $n \geq m$, and enough of the samples x_0, \dots, x_N in (1.4) lie outside of the support of ω .

THEOREM 2.1. *Let $f = p_m + \omega$ be a s -corrupted polynomial of degree $\leq m$. Then, the following statements hold when $n \geq m$:*

1. *If $s < 1$, then $p_n^{L_0} = p_m$.*
2. *If $(N - n)/2$, or fewer, of the samples x_0, \dots, x_N are in $\text{supp}(\omega)$, then $p_n^{\ell_0} = p_m$.*
3. *If k of x_0, \dots, x_N are in $\text{supp}(\omega)$, $N + 1 > 6(n + 1)k - 1$, and $N \geq n$, then $p_n^{\ell_1} = p_m$.*
4. *If $s < 1/(n + 1)^2$, then $p_n^{L_1} = p_m$.*

We prove the four statements in the theorem, in turn, in the next four subsections.

2.1. Exact recovery with best L_0 approximation. Intuitively, recovery of a corrupted function is ideal for best L_0 polynomial approximation as the Lebesgue measure of $\text{supp}(f - p_n^{L_0})$ is minimized. The polynomial approximant $p_n^{L_0}$ is so good at recovery that when $f = p_m + \omega$ we have $p_n^{L_0} = p_m$ provided that $\text{supp}(\omega)$ is less than half the interval and $n \geq m$.

To see this, note that $|\text{supp}(f - p_m)| = \text{supp}(\omega) = s < 1$. Suppose there is a polynomial q of degree $\leq n$ such that $|\text{supp}(f - q)| \leq |\text{supp}(f - p_m)|$. Then, $|\text{supp}(q - p_m)| \leq 2s < 2$, so q and p_m must coincide on a set of positive measure in $[-1, 1]$. Since q and p_m are polynomials and $n \geq m$, we have that $q = p_m$. We conclude that $p_n^{L_0} = p_m$ provided that $\text{supp}(\omega) < 1$ and $n \geq m$. This proves the first statement of Theorem 2.1.

2.2. Exact recovery with best ℓ_0 approximation. It can be algorithmically challenging to compute $p_n^{L_0}$, and it is reasonable to attempt recovery from $p_n^{\ell_0}$ instead, which involves a discrete optimization problem. The polynomial approximant $p_n^{\ell_0}$ is also ideal at recovering polynomials under the mild assumption that enough of the samples x_0, \dots, x_N (see (1.4)) lie outside of $\text{supp}(\omega)$.

To see this, suppose that $f = p_m + \omega$ and there is a polynomial q of degree $\leq n$ such that

$$(2.1) \quad \|f - q\|_{\ell_0} \leq \|f - p_m\|_{\ell_0} = k,$$

where $n \geq m$ and k is the number of samples x_0, \dots, x_N in $\text{supp}(\omega)$. If $k \leq (N - n)/2$, then $q - p_m \in \mathcal{P}_n$ is zero on at least $N + 1 - 2k \geq n + 1$ distinct points, and hence $q = p_m$ [27, p. 34]. By definition of $p_n^{\ell_0}$, we must have $p_n^{\ell_0} = p_m$. This proves the second statement of Theorem 2.1.

2.3. Exact recovery with best ℓ_1 approximation. The polynomial $p_n^{\ell_0}$ can be computationally prohibitive to compute if m is large. Fortunately, by using the restricted isometry property (RIP) from compressed sensing, one finds that $p_n^{\ell_0} = p_n^{\ell_1}$ when an oversampling condition is satisfied, along with some regularity assumptions. This means that $p_n^{\ell_1}$, which can be computed efficiently, can often be used for exact recovery [3].

First, we know that $\|f - q_n\|_{\ell_0} = k$ for $q_n \in \mathcal{P}_n$ is equivalent to a vector \underline{y} having precisely k nonzero entries, where

$$(2.2) \quad \underline{y} = \underbrace{\begin{bmatrix} U_0(x_0) & \cdots & U_n(x_0) \\ U_0(x_1) & \cdots & U_n(x_1) \\ \vdots & \ddots & \vdots \\ U_0(x_N) & \cdots & U_n(x_N) \end{bmatrix}}_{= \Phi} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} - \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_N) \end{bmatrix}, \quad q_n(x) = \sum_{i=0}^n c_i U_i(x)$$

and $U_i(x)$ is the Chebyshev polynomial of the second kind of degree i [24, Table 18.3.1]. The problem of minimizing $\|\underline{y}\|_{\ell_0}$ over \mathcal{P}_n in (2.2) is solved by $p_n^{\ell_0}$ and can be written as

$$(2.3) \quad \min_{\underline{c} \in \mathbb{R}^{n+1}} \|\Phi \underline{c} - \underline{f}\|_{\ell_0}, \quad \underline{f} = [f(x_0) \quad \cdots \quad f(x_N)]^\top,$$

which is equivalent to the following diagonally scaled problem:

$$(2.4) \quad \min_{\underline{c} \in \mathbb{R}^{n+1}} \|D\Phi \underline{c} - D\underline{f}\|_{\ell_0}, \quad D = \sqrt{2/(N+2)} \operatorname{diag}\left(\sqrt{1-x_0^2}, \dots, \sqrt{1-x_N^2}\right).$$

By a technique described in [13, p. 4204], if $V \in \mathbb{R}^{(N+1) \times (N-n)}$ is a matrix whose columns form a basis for the left null space of $D\Phi$ so that $V^\top(D\Phi) = 0$, then (2.3) is also a constrained ℓ_0 minimization problem:

$$(2.5) \quad \min_{\underline{z} \in \mathbb{R}^{N+1}} \|\underline{z}\|_{\ell_0} \quad \text{subject to} \quad V^\top \underline{z} = -V^\top D\underline{f},$$

where $\underline{z} = D\Phi \underline{c} - D\underline{f}$. This problem is precisely the task of interest in the compressed sensing literature with a short-fat matrix V^\top and an unknown sparse vector \underline{z} .

The ℓ_0 minimization problem (2.5) is known to be NP-hard [18, sect. 2.3]. A practical remedy is to replace the ℓ_0 -norm with the ℓ_1 -norm. To understand when this gives the solution to the ℓ_0 problem, an important concept in compressed sensing is the RIP. We say that a matrix $A \in \mathbb{C}^{m \times r}$ satisfies the RIP if there exists a constant $0 < \delta_k < 1$ such that

$$(2.6) \quad (1 - \delta_k) \|\underline{x}\|_2^2 \leq \|A\underline{x}\|_2^2 \leq (1 + \delta_k) \|\underline{x}\|_2^2, \quad \|\underline{x}\|_2^2 = \sum_{i=1}^r |x_i|^2,$$

for every vector $\underline{x} \in \mathbb{C}^r$ that has at most k nonzero entries [13]. It is known that if V^\top satisfies the RIP with $\delta_k < \frac{1}{3}$, then the solution to (2.5) is exactly recovered (under the assumption that the ℓ_0 -minimizer $\leq k$ nonzero entries) by solving the ℓ_1 minimization problem [9]

$$(2.7) \quad \min_{\underline{z} \in \mathbb{R}^{N+1}} \|\underline{z}\|_{\ell_1} \quad \text{subject to} \quad V^\top \underline{z} = -V^\top D\underline{f}.$$

Here, (2.7) can be efficiently solved as a basis pursuit problem via the spectral projected-gradient L_1 (SPGL1) algorithm [32], especially since there is a fast matrix-vector product for V^\top based on the discrete sine transformation (see (2.10)).

Note that unlike $\|f\|_{\ell_1}$ in (1.3) for functions, the ℓ_1 -norm for vectors is simply the sum of the absolute values of the vector entries. The problem in (2.7) is equivalent to

$$(2.8) \quad \min_{\underline{c} \in \mathbb{R}^{n+1}} \|D(\Phi \underline{c} - \underline{f})\|_{\ell_1},$$

which in turn can be written as (recalling (1.3)) the best ℓ_1 approximation problem:

$$(2.9) \quad \min_{q \in \mathcal{P}_n} \|f - q\|_{\ell_1}.$$

We conclude that if the matrix V^\top satisfies the RIP with $\delta_k < \frac{1}{3}$, then we have $p_n^{\ell_1} = p_n^{\ell_0}$, where

$$p_n^{\ell_1}(x) = \sum_{j=0}^n c_j^* U_j(x)$$

and the vector c^* is the solution to (2.8).

We are left with the task of studying when the matrix V^\top in (2.5) satisfies the RIP with $\delta_k < \frac{1}{3}$. For the samples x_0, \dots, x_N that are given in (1.4), we have the discrete orthogonality condition $\sum_{\ell=0}^N U_i(x_\ell) U_j(x_\ell) (1 - x_\ell^2) = 0$ for $i \neq j$ [23, sect. 4.6.1] so that we can write down an explicit basis for the left null space of $D\Phi$ in (2.2). That is,

$$(2.10) \quad V = D \begin{bmatrix} U_{n+1}(x_0) & \dots & U_N(x_0) \\ \vdots & \ddots & \vdots \\ U_{n+1}(x_N) & \dots & U_N(x_N) \end{bmatrix} \in \mathbb{R}^{(N+1) \times (N-n)}.$$

It turns out that due to the choice of the diagonal matrix D in (2.4), the matrix V in (2.10) is formed from a subset of columns of an orthogonal matrix. Furthermore, the size of V^\top need not be extremely short-fat, as often required in compressed sensing. It is therefore possible to show that V^\top satisfies the RIP under a mild oversampling condition.

PROPOSITION 2.2. *If $N+2 > 2(n+1)k$ for some integer $k \geq 1$, then V^\top in (2.10) satisfies the RIP with $\delta_k = (2(n+1)/(N+2))k$.*

Proof. Let A be the $(N+1) \times (N+1)$ Chebyshev–Vandermonde matrix, i.e., $A_{ij} = U_j(x_i)$ for $0 \leq i, j \leq N$, where x_i is given in (1.4). Let D be a diagonal matrix with $D_{i,i} = \sqrt{2/(N+2)}\sqrt{1-x_i^2}$ for $0 \leq i \leq N$. By the discrete orthogonality properties of Chebyshev polynomials of the second kind [23, sect. 4.6.1], DA is an orthogonal matrix with

$$A^\top D = \begin{bmatrix} \Phi^\top D \\ V^\top \end{bmatrix},$$

where Φ and V are given in (2.2) and (2.10), respectively. Since $A^\top D$ has orthonormal columns, we find that

$$(2.11) \quad \left\| \begin{bmatrix} \Phi^\top D \\ V^\top \end{bmatrix} \underline{z} \right\|_2^2 = \|\Phi^\top D \underline{z}\|_2^2 + \|V^\top \underline{z}\|_2^2 = \|\underline{z}\|_2^2, \quad \underline{z} \in \mathbb{C}^{N+1}.$$

Since $\sqrt{1-x^2}|U_i(x)| \leq 1$ for $x \in [-1, 1]$ [24, eq. (18.14.7)], each entry of $A^\top D$ has absolute value $\leq \sqrt{2/(N+2)}$; it follows by Cauchy–Schwarz that each entry of $\Phi^\top D \underline{z}$ is bounded by $\sqrt{\frac{2k}{N+2}}$ where k is the number of nonzero entries in \underline{z} , so we have

$$(2.12) \quad \|\Phi^\top D \underline{z}\|_2^2 \leq \frac{2(n+1)k}{N+2} \|\underline{z}\|_2^2.$$

Therefore, from (2.11) and the trivial bound of $\|V^\top \underline{z}\|_2^2 \leq \|\underline{z}\|_2^2$, we conclude that

$$\left(1 - \frac{2(n+1)k}{N+2}\right) \|\underline{z}\|_2^2 \leq \|V^\top \underline{z}\|_2^2 \leq \|\underline{z}\|_2^2$$

for any vector $\underline{z} \in \mathbb{C}^{N+1}$ with at most k nonzero entries. The statement immediately follows from the definition of the RIP (see (2.6)). \square

Proposition 2.2 tells us that V^\top in (2.10) satisfies the RIP with $\delta_k < 1/3$ if $N+1 > 6(n+1)k-1$. Since k is the number of samples x_0, \dots, x_N that lie in $\text{supp}(\omega)$, it means that $p_n^{\ell_0} = p_n^{\ell_1}$ provided that the discrete problem is sufficiently oversampled. Since $k < (N+2)/(6(n+1))$ implies that $k \leq N-n$ when $k \geq 1$ and when $k = 0$ we need $N \geq n$, we conclude from subsection 2.2 that if $N+1 > 6(n+1)k-1$ and $N \geq n$, then $p_n^{\ell_1} = p_n^{\ell_0} = p_m$ when $n \geq m$. This proves the third statement of Theorem 2.1.

The polynomial $p_n^{\ell_1}$ can be computed by solving the basis pursuit problem in (2.7). This means that Proposition 2.2 gives us a practical and efficient algorithm for the exact recovery of corrupted polynomials with degrees in the thousands. Often, one does not know the degree of the corrupted polynomial or k . Since the oversampling condition $N+1 > 6(n+1)k-1$ penalizes taking unnecessarily large n , we recommend slowly increasing n , computing the error $f - p_n^{\ell_1}$, and stopping at the smallest n for which $\text{supp}(f - p_n^{\ell_1}) < 2$.

2.4. Exact recovery with best L_1 approximation. To begin to highlight the importance of error localization of best L_1 polynomial approximants, we now show that $p_n^{\ell_1}$ can also be used for the exact recovery of corrupted polynomials when the corruption has sufficiently small support. One can achieve this by demonstrating that a polynomial of degree $\leq n$ is not too concentrated in any small subset of $[-1, 1]$.

LEMMA 2.3. *Let $\Omega_s \subseteq [-1, 1]$ be a set of Lebesgue measure $s \geq 0$. For any $n \geq 0$, we have*

$$(2.13) \quad \int_{\Omega_s} |p(x)| dx \leq \frac{s(n+1)^2}{2} \int_{-1}^1 |p(x)| dx$$

for any polynomial p of degree $\leq n$.

Proof. This statement is proved in [5, sect. 4.2, Exercise 6]. \square

Lemma 2.3 tells us that polynomials of degree $\leq n$ cannot be too localized in a set of small measure. In particular, if $0 \leq |\Omega_s| < 1/(n+1)^2$, then

$$(2.14) \quad \int_{\Omega_s} |p(x)| dx \leq \int_{[-1,1] \setminus \Omega_s} |p(x)| dx, \quad p \in \mathcal{P}_n,$$

with equality if and only if p is the zero polynomial. A consequence of (2.14) is that a corrupted polynomial can be exactly recovered by best L_1 polynomial approximation.

COROLLARY 2.4. *Let $f = p_m + \omega$ be a s -corrupted polynomial of degree $\leq m$ on $[-1, 1]$. Then, the best L_1 polynomial approximant of degree $\leq n$ to f is p_m if $n \geq m$ and $s < 1/(n+1)^2$.*

Proof. Let $\delta p \in \mathcal{P}_n$, and let $\Omega_s \subset [-1, 1]$ be the support of ω . Since $[-1, 1] = \Omega_s \cup ([-1, 1] \setminus \Omega_s)$, we have by the triangle inequality

(2.15)

$$\begin{aligned}
\|f - p_m - \delta p\|_1 &= \int_{\Omega_s} |f(x) - p_m(x) - \delta p(x)| dx + \int_{[-1,1] \setminus \Omega_s} |\delta p(x)| dx \\
&\geq \int_{\Omega_s} |f(x) - p_m(x)| dx - \int_{\Omega_s} |\delta p(x)| dx + \int_{[-1,1] \setminus \Omega_s} |\delta p(x)| dx \\
&\geq \|f - p_m\|_1,
\end{aligned}$$

where the last inequality follows from (2.14) as well as the fact that $f(x) - p_m(x) = 0$ for $x \in [-1, 1] \setminus \Omega_s$. An equality holds in (2.15) if and only if $\delta p = 0$. We conclude that p_m is the unique best L_1 polynomial approximant to f of degree $\leq n$. \square

This proves the fourth and final statement of Theorem 2.1 and explains regime (b) in Figure 1.2. It tells us that if a polynomial is corrupted on a subset of $[-1, 1]$ that has small enough Lebesgue measure, then the best L_1 polynomial approximant exactly recovers the polynomial. Figure 2.1 illustrates Corollary 2.4 for the corrupted polynomial $f = T_5 + \omega$, where T_5 is the degree 5 Chebyshev polynomial of the first kind and $\text{supp}(\omega) = [-.7, -.67] \cup [.9, .903]$. Using the fact that $p_n^{L_1} = p_n^{\ell_1}$, one can efficiently recover T_5 to within essentially machine precision. Numerically, we find that $\|p_n^{L_1} - T_5\|_\infty \approx 1.22 \times 10^{-15}$.

To highlight the importance of the L_1 -norm for Corollary 2.4, we consider the best polynomial approximants of degree ≤ 5 to f in the L_2 - and L_∞ -norm (see Figure 2.1 (right)). One finds that any corruption of arbitrarily small support prevents the best L_2 and L_∞ polynomial approximants from recovering the uncorrupted polynomial.

The bound on s of $s < 1/(n+1)^2$ in Corollary 2.4 is probably not sharp, though we know that it cannot be increased above $\pi^2/(2(n+2)^2)$ [5, sect. 4.2]. This means that the algebraic scaling with respect to n is definitive. In Appendix A, we extend Corollary 2.4 by demonstrating that the location of the support of the corruption in $[-1, 1]$ is important, and more is allowed provided that the corruption occurs away from ± 1 .

For concreteness, we have assumed that the sample points are the Chebyshev points given in (1.4). This choice is recommended when the samples can be taken at arbitrary points in $[-1, 1]$. However, in some cases, the sample points may be given a

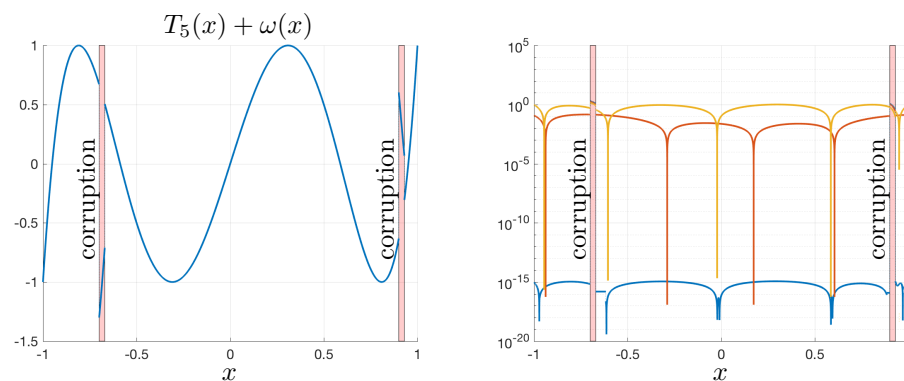


FIG. 2.1. Left: Corrupted polynomial $f = T_5 + \omega$, where T_5 is the degree 5 Chebyshev polynomial of the first kind and $\text{supp}(\omega) = [-.7, -.67] \cup [.9, .903]$ (shaded red). Right: The error $|f(x) - p_5^*(x)|$, where p_5^* is the best degree ≤ 5 polynomial approximant to f in the L_1 -norm (blue line), L_2 -norm (red line), and L_∞ -norm (yellow line). One can see that $|f(x) - p_5^{L_1}(x)|$ is essentially machine precision for $x \notin \text{supp}(\omega)$ whereas $p_5^{L_2}$ and $p_5^{L_\infty}$ do not recover T_5 .

priori and cannot be chosen. Most of our results carry over to such cases with minor modifications and assumptions on the distribution of sample points.

3. Near-recovery of corrupted smooth functions. When recovering a corrupted polynomial $f = p_m + \omega$, the degree of p_m is usually unknown, so we compute best L_1 polynomial approximants to f of degree $\leq n$ for a slowly increasing sequence of n , stopping when $\text{supp}(f - p_n^{L_1}) < 2$. For the majority of this process $n < m$, and one may wonder what $p_n^{L_1}$ is achieving in this regime (see Figure 1.2(a)). Similarly, if f is a corrupted smooth function $f = f_0 + \omega$, where f_0 is a continuous function (not necessarily a polynomial) on $[-1, 1]$, then one cannot hope for exact recovery using best L_1 polynomial approximation. Instead, we find that $p_n^{L_1}$ delivers a near-recovery of f_0 in the sense that $p_n^{L_1}$ is a near-best L_1 approximation to f_0 , provided that the support of the corruption is small and f_0 can be well-approximated by a degree $\leq n$ polynomial. We first show that the best L_1 approximations for f and f_0 are relatively close to each other.

THEOREM 3.1. *Let $f = f_0 + \omega$ be a s -corrupted function on $[-1, 1]$, where $f_0 : [-1, 1] \rightarrow \mathbb{R}$ is continuous and $p_n^{L_1}$ be a best L_1 polynomial approximant of degree $\leq n$ to f . If $s < 1/(n+1)^2$, then*

$$\|p_n^{L_1} - p_n^*\|_1 \leq \frac{4}{2 - s(n+1)^2} \|f_0 - p_n^*\|_1,$$

where p_n^* is the best L_1 approximant of degree $\leq n$ to f_0 on $[-1, 1]$.

Proof. Let $\delta p \in \mathcal{P}_n$ and $\Omega_s = \text{supp}(\omega)$. Since $[-1, 1] = \Omega_s \cup ([-1, 1] \setminus \Omega_s)$ and by the triangle inequality, we have

$$\begin{aligned} \|f - p_n^* - \delta p\|_1 &= \int_{\Omega_s} |f(x) - p_n^*(x) - \delta p(x)| dx + \int_{[-1, 1] \setminus \Omega_s} |f(x) - p_n^*(x) - \delta p(x)| dx \\ &\geq \int_{\Omega_s} (|f(x) - p_n^*(x)| - |\delta p(x)|) dx + \int_{[-1, 1] \setminus \Omega_s} (|\delta p(x)| - |f(x) - p_n^*(x)|) dx \\ &\geq \|f - p_n^*\|_1 - 2\|f_0 - p_n^*\|_1 + \int_{[-1, 1] \setminus \Omega_s} |\delta p(x)| dx - \int_{\Omega_s} |\delta p(x)| dx, \end{aligned}$$

where the last inequality holds since $[-1, 1] = \Omega_s \cup ([-1, 1] \setminus \Omega_s)$ and $f(x) = f_0(x)$ for $x \notin \Omega_s$. From (2.13), we find that

$$\int_{\Omega_s} |\delta p(x)| dx \leq \frac{s(n+1)^2}{2} \|\delta p\|_1, \quad \int_{[-1, 1] \setminus \Omega_s} |\delta p(x)| dx \geq \left(1 - \frac{s(n+1)^2}{2}\right) \|\delta p\|_1.$$

Hence, for any $\delta p \in \mathcal{P}_n$ we have the inequality

$$\|f - p_n^* - \delta p\|_1 \geq \|f - p_n^*\|_1 - 2\|f_0 - p_n^*\|_1 + \left(1 - \frac{s(n+1)^2}{2}\right) \|\delta p\|_1.$$

Finally, by setting $\delta p = p_n^{L_1} - p_n^*$ and noting that $\|f - p_n^{L_1}\|_1 \leq \|f - p_n^*\|_1$ we conclude that

$$-2\|f_0 - p_n^*\|_1 + \left(1 - \frac{s(n+1)^2}{2}\right) \|p_n^{L_1} - p_n^*\|_1 \leq 0.$$

The result follows by rearranging this inequality. \square

Theorem 3.1 shows that best L_1 polynomial approximation is useful for near-recovery of a corrupted smooth function. More precisely, when $s < 1/(n+1)^2$ we have

$$(3.1) \quad \|f_0 - p_n^{L_1}\|_1 \leq \left(1 + \frac{4}{2 - s(n+1)^2}\right) \min_{q_n \in \mathcal{P}_n} \|f_0 - q_n\|_1,$$

and we conclude that a best L_1 approximant of f recovers f_0 as best it can, up to a factor that depends on n and s .

The inequality in (3.1) also partially explains regime (a) in Figure 1.2. It provides theoretical justification that $p_5^{L_1}$ is a near-best polynomial approximant to P_8 in Figure 1.2. For the example in Figure 1.2, we observe this near-recovery phenomenon since

$$\|P_8 - p_5^{L_1}\|_1 \approx 0.450, \quad \min_{q_5 \in \mathcal{P}_5} \|P_8 - q_5\|_1 \approx 0.414,$$

where $p_5^{L_1}$ is the best L_1 approximant of degree ≤ 5 to the corrupted function.

Unlike corrupted polynomials (see section 2), f_0 cannot be exactly recovered by $p_n^{\ell_1}$. Nonetheless, we find that $p_n^{\ell_1}$ is often still a near-best approximant to f_0 , i.e., $p_n^{\ell_1} \approx p_n^*$. By interpreting $f_0 - p_n^{\ell_1}$ as noise, we observe that ℓ_1 minimization gives a stable signal recovery in the presence of noise, a phenomenon that is appreciated in the classical compressed sensing context [12]. Making this observation precise in our setting is left as an open problem. Since by Theorem 3.1 we also have $p_n^{L_1} \approx p_n^*$; it follows that $p_n^{\ell_1} \approx p_n^{L_1}$ and $p_n^{\ell_1}$ is an excellent initial guess for Newton's method for computing $p_n^{L_1}$ (see subsection 5.3).

3.1. Related studies. The contents of sections 2 and 3 can be regarded as contributions in compressed sensing, and several related studies are available in the literature. For (exact and near-exact) recovery of corrupted functions with ℓ_1 minimization, examples include the papers by Adcock, Brugiapaglia, and Webster [1], and Shin and Xiu [29]. Unlike this work, these papers consider recovering high-dimensional functions, describing probabilistic methods by taking random samples. Here we focus on univariate polynomials, reveal connections between L_0, L_1, ℓ_0 , and ℓ_1 minimizers, and derive a deterministic recovery algorithm (under assumptions on the size of sampled corruption k) with ℓ_1 minimization. Few of the results in this paper appear to be trivially generalizable to the higher-dimensional setting; this is left as an interesting open problem.

In the more classical setting of recovering a discrete signal (rather than a function) from a corrupted vector of observations, numerous contributions are available in the literature. See, for example, [10, 13, 21, 35] and the references therein. Ideas in compressed sensing have also been applied for general high-dimensional function approximation [2, 14].

4. Error localization of best L_1 polynomial approximants. In sections 2 and 3 we saw that $p_n^{L_1}$ can be used for recovering corrupted polynomials and smooth functions. This is fundamentally due to the error localization properties of best L_1 polynomial approximation. The error localization properties of $p_n^{L_1}$ are also important when approximating continuous functions $f : [-1, 1] \rightarrow \mathbb{R}$ that one might not necessarily view as corrupted functions. We observe that continuous functions with singularities often have $|f(x) - p_n^{L_1}(x)| \ll \|f - p_n^{L_1}\|_\infty$ for most $x \in [-1, 1]$.

To make this precise, recall the definition of Ω_n in (1.2). By definition of Ω_n , we find that $\|f - p_n^{L_1}\|_1 \geq \frac{|\Omega_n|}{2} \|f - p_n^{L_\infty}\|_\infty$ and thus,

$$(4.1) \quad 0 < |\Omega_n| \leq 2 \|f - p_n^{L_1}\|_1 / \|f - p_n^{L_\infty}\|_\infty.$$

Therefore, the measure of Ω_n is bounded above by the disparity between the magnitude of $\|f - p_n^{L_1}\|_1$ and $\|f - p_n^{L_\infty}\|_\infty$. If $\|f - p_n^{L_1}\|_1 \rightarrow 0$ asymptotically faster than $\|f - p_n^{L_\infty}\|_\infty \rightarrow 0$ as $n \rightarrow \infty$, then the error $f(x) - p_n^{L_1}(x)$ must be highly localized for sufficiently large n . An upper bound on $|\Omega_n|$ follows from an upper bound on $\|f - p_n^{L_1}\|_1$ and a lower bound on $\|f - p_n^{L_\infty}\|_\infty$.

4.1. Error localization of best L_1 approximants to $\sqrt{1-x^2}$. Consider the function $f(x) = \sqrt{1-x^2}$, which is continuous on $[-1, 1]$ with square root singularities at ± 1 . Here, we show that $|\Omega_n| = \mathcal{O}(n^{-2} \log n)$ proving that $p_n^{L_1}(x)$ is a better pointwise estimate to $f(x)$ than $p_n^{L_\infty}(x)$ for all $x \in [-1, 1]$ except for a set of measure $\mathcal{O}(n^{-2} \log n)$.

By [17, Lem. 4], we know that when n is an even integer we have $p_n^{L_1} = p_n^{\text{cheb}}$ for $\sqrt{1-x^2}$, where p_n^{cheb} is the degree n Chebyshev interpolant of $\sqrt{1-x^2}$ (see (1.5)). This allows us to derive an explicit expression for $\|f - p_n^{L_1}\|_1$ by using an explicit formula for $\|f - p_n^{\text{cheb}}\|_1$ [8]. By applying the formula in [8] to $\sqrt{1-x^2}$, we find that

$$\|f - p_n^{L_1}\|_1 = 2 \left| \sum_{\nu=0}^{\infty} (2\nu+1)^{-1} b_{(\nu+1)(n+2)-1} \right|, \quad b_j = \begin{cases} -\frac{8}{(j-1)(j+1)(j+3)\pi}, & j = \text{even}, \\ 0, & j = \text{odd}. \end{cases}$$

Here, the values of b_j are derived as the expansion coefficients of $\sqrt{1-x^2}$ in a Chebyshev series of the second kind. That is,

$$\sqrt{1-x^2} = \sum_{j=0}^{\infty} b_j U_j(x), \quad b_j = \frac{2}{\pi} \int_{-1}^1 (1-x^2) U_j(x) dx, \quad j \geq 0.$$

Since $|b_j| \leq 16(j+1)^{-3}/\pi$ for $j > 0$, we can bound $\|f - p_n^{L_1}\|_1$ by

$$\|f - p_n^{L_1}\|_1 \leq \frac{32}{\pi(n+2)^3} \sum_{\nu=0}^{\infty} \frac{1}{(2\nu+1)(\nu+1)^3} \leq \frac{64}{\pi(n+1)^3},$$

where the last inequality uses the crude bounds of $\sum_{\nu=0}^{\infty} (2\nu+1)^{-1}(\nu+1)^{-3} \leq 2$ and $n+2 \geq n+1$.

We now seek a lower bound on $\|f - p_n^{L_\infty}\|_\infty$. Let $p_n^{\text{proj}}(x) = \sum_{j=0}^n a_j T_j(x)$ be the Chebyshev expansion of the first kind for $\sqrt{1-x^2}$ that is truncated after $n+1$ terms. The values of a_j are simple to calculate: $a_{2j-1} = 0$ for all integers j , and

$$a_0 = \frac{1}{\pi} \int_{-1}^1 T_0(x) dx = \frac{2}{\pi}, \quad a_{2j} = \frac{2}{\pi} \int_{-1}^1 T_{2j}(x) dx = \frac{4}{(1-4j^2)\pi}, \quad j \geq 1.$$

Assuming n is an even integer, we find that

$$p_n^{\text{proj}}(1) = \sum_{j=0}^n a_j = \frac{2}{\pi} + \frac{4}{\pi} \sum_{j=1}^{n/2} \frac{1}{1-4j^2} = \frac{2}{\pi} \left(1 - \frac{n}{n+1} \right) = \frac{2}{\pi(n+1)}.$$

Thus, $\|f - p_n^{\text{proj}}\|_\infty \geq 2/(\pi(n+1))$ for an even integer n . By [22, Cor. 4.1], we know that

$$\|f - p_n^{\text{proj}}\|_\infty \leq (1 + \sigma_n) \|f - p_n^{L_\infty}\|_\infty, \quad \sigma_n = \frac{1}{\pi} \int_0^\pi \frac{|\sin(n+1/2)\theta|}{\sin(\theta/2)} d\theta.$$

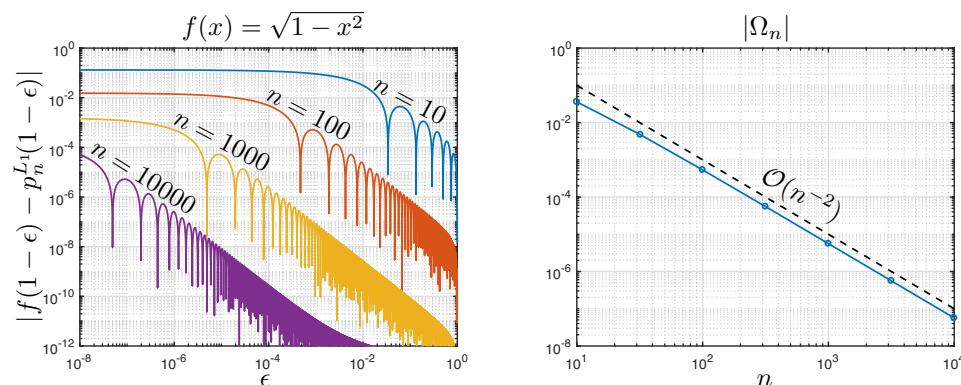


FIG. 4.1. Left: The error $|f(x) - p_n^{L1}(x)|$ for $f(x) = \sqrt{1-x^2}$ with $n = 10, 100, 1000$, and 10000 , shown on the interval $[0, 1)$. Right: It is observed that $|\Omega_n| = \mathcal{O}(n^{-2})$, showing that the error $|f(x) - p_n^{L1}(x)|$ is highly localized. In particular, we find that $|f(x) - p_n^{L1}(x)| < \frac{1}{2} \|f - p_n^{L\infty}\|_\infty$ for all $x \in [-1, 1]$ except for a set of measure $< 10^{-5}$ near $x = \pm 1$ when $n = 1000$.

We conclude from (4.1) that for $f(x) = \sqrt{1-x^2}$ we have

$$|\Omega_n| \leq \frac{128}{\pi(n+1)^3} \frac{\pi(n+1)(1+\sigma_n)}{2} = \frac{64(1+\sigma_n)}{(n+1)^2} = \mathcal{O}(n^{-2} \log n),$$

where the final equality holds since it is known that $\sigma_n \sim 4\pi^{-2} \log n$ [22, eq. 20].

Figure 4.1 (left) shows the error $|f(x) - p_n^{L1}(x)|$ for $x \in [0, 1)$ demonstrating that it is localized near $x = \pm 1$. The measure of $|\Omega_n|$ is shown in Figure 4.1 (right) where it is numerically observed that $|\Omega_n| = \mathcal{O}(n^{-2})$. When $n = 1000$, we find that $|f(x) - p_n^{L1}(x)| < \frac{1}{2} \|f - p_n^{L\infty}\|_\infty$ for all $x \in [-1, 1]$ except for a set of measure $< 10^{-5}$.

4.2. Error localization of best L_1 approximants to $|x|$. As a second example of error localization, consider $f(x) = |x|$ on $[-1, 1]$, which is continuously differentiable except at $x = 0$. The error formula for $\|f - p_n^{L1}\|_1$ with $f(x) = |x|$ is calculated in [8] and simplifies to

$$\|f - p_n^{L1}\|_1 \sim \frac{8}{\pi n^2} \left(\sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{(2\nu+1)^3} \right) = \frac{\pi^2}{4n^2}.$$

Moreover, it is known that $\|f - p_n^{L\infty}\|_\infty \sim \frac{\beta}{2n}$ for some $0.28016 < \beta < 0.28018$ [33]. We conclude from (4.1) that $|\Omega_n| \lesssim \frac{\pi^2}{\beta n}$ as $n \rightarrow \infty$. Figure 4.2 (left) shows the error $|f(x) - p_n^{L1}(x)|$ for $x \in [0, 1)$ demonstrating that it is highly localized and Figure 4.2 numerically confirms that $|\Omega_n| = \mathcal{O}(n^{-1})$.

5. A globally convergent algorithm for computing best L_1 polynomial approximants. We now turn to the algorithmic aspects of computing p_n^{L1} . We integrate our findings on exact recovery of corrupted polynomials and error localization into Watson's algorithm based on Newton's method [34]. An algorithm to compute best L_1 approximants with degrees in the thousands is developed based on recent advances in approximation theory such as stable polynomial interpolation, fast domain subdivision, and robust rootfinding implemented in Chebfun [16]. Figure 5.1 gives an overview of our algorithm.

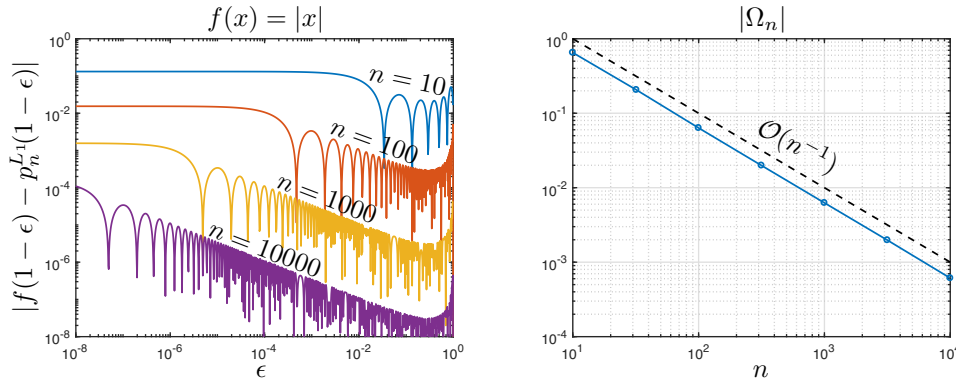


FIG. 4.2. Left: The error $|f(x) - p_n^{L1}(x)|$ for $f(x) = |x|$ with $n = 10, 100, 1000$, and 10000 shown on the interval $[0, 1)$. Right: It is observed that $|\Omega_n| = \mathcal{O}(n^{-1})$. In this example, we find that the error $|f(x) - p_n^{L1}(x)|$ is highly localized near $x = 0$ and $x = \pm 1$.

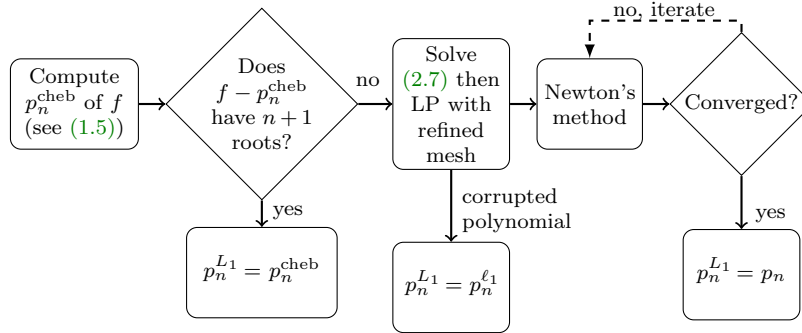


FIG. 5.1. Flowchart for our algorithm to compute the best L_1 polynomial approximant of degree $\leq n$ to a continuous function f on $[-1, 1]$.

5.1. Initial attempt: The Chebyshev interpolant. The polynomial interpolant p_n^{cheb} in (1.5) with $N = n$ can be computed in $\mathcal{O}(n \log n)$ operations [19], and the roots of $f - p_n^{\text{cheb}}$ on $[-1, 1]$ can be computed efficiently when f is a smooth function [7]. Since $p_n^{L1} = p_n^{\text{cheb}}$ when $f - p_n^{\text{cheb}}$ has exactly $n + 1$ roots in $[-1, 1]$ [26], we recommend that (1.5) is always computed to see if $p_n^{L1} = p_n^{\text{cheb}}$. When it is, p_n^{L1} is efficient to compute, and from practical experience it is relatively common for $p_n^{L1} = p_n^{\text{cheb}}$ (for example, see [17, Lem. 4]). This can happen also when f is a corrupted polynomial.

5.2. Test for corrupted polynomials and initial guess: Compute ℓ_1 minimizer. When $f - p_n^{\text{cheb}}$ has $> n + 1$ zeros in $[-1, 1]$, computing p_n^{L1} is more involved and, in general, requires an iterative procedure. In this case, we first solve the discrete ℓ_1 problem in (2.9) to obtain $p_n^{\ell_1}$. This has two purposes: (i) if f is a corrupted polynomial $f = p_m + \omega$ (see section 2), then $p_n^{\ell_1} = p_m = p_n^{L1}$, and (ii) if f is not a corrupted polynomial, then $p_n^{\ell_1} \approx p_n^{L1}$ [28, Thm. 3.9], which is then used as the initial guess for Newton's method (see subsection 5.3).

Specifically, we solve the LP in (5.2) with a large number of samples $N + 1$, taking x_0, \dots, x_N and $w_j = \pi \sqrt{1 - x_j^2} / (N + 2)$ as in (1.4). In our implementation we select

$N + 1 = \max(1000 + 50n, 5000)$. (This is an engineering choice that assumes the corruption k is small.) Recall from Theorem 2.1 that we want $N + 1 > 6(n + 1)k - 1$. The maximum value 5000 is set to keep the LP size $2(N + 1) + n + 1$ manageable.

Once $p_n^{\ell_1}$ is computed, we check whether f is a corrupted polynomial. This can be done by testing if $f(x_j) = p_n^{\ell_1}(x_j)$ holds at most of the sample points to within working precision. If not, then we improve the estimate $p_n^{\ell_1} \approx p_n^{L_1}$ by refining the LP mesh and then proceed to Newton's method.

5.2.1. Refinement: Reducing the discretization error. Underlying the minimization problem (2.7) is an approximate integration of a nondifferentiable function. Specifically,

$$(5.1) \quad \min_{p_n \in \mathcal{P}_n} \sum_{i=0}^N w_i |f(y_i) - p_n(y_i)|, \quad \int_{-1}^1 |f(x) - p_n(x)| dx \approx \sum_{i=0}^N w_i |f(y_i) - p_n(y_i)|.$$

Since $|f(x) - p_n(x)|$ is expected to be continuous but nondifferentiable at $\geq n + 2$ points, one expects the integration error in (5.1) to be large, and there is little benefit from using a high-order quadrature rules. Indeed, using $N + 1$ sample points, we find that the LP solution has accuracy $\|p_n^{\ell_1} - p_n^{L_1}\|_1 = \mathcal{O}(N^{-1})$, whether a high-order method (e.g., Clenshaw–Curtis) or a low-order method (such as the midpoint rule) is used. In more detail, the quadrature error in (5.1) is $\mathcal{O}(N^{-2})$, so the objective function value $\|f - p_n^{\ell_1}\|_1$ is within $\mathcal{O}(N^{-2})$ of optimal: $\|f - p_n^{\ell_1}\|_1 = \|f - p_n^{L_1}\|_1 + \mathcal{O}(N^{-2})$. However, this only implies $\|p_n^{L_1} - p_n^{\ell_1}\|_1 = \mathcal{O}(N^{-1})$, which is a common phenomenon in optimization: at a global (or local) minimum, an ϵ -perturbation in the solution results in $\mathcal{O}(\epsilon^2)$ perturbation in the objective value. This low accuracy of $p_n^{\ell_1}$ can cause convergence issues for Newton's method, when it is used as an initial guess.

To improve the discretization error in (5.1), we follow a three-step procedure: (1) We use the initial LP solution with N points to obtain an $\mathcal{O}(N^{-1})$ approximation to $p_n^{L_1}$, which we denote by \tilde{p}_n . (2) The roots $\{r_i\}_{i=1}^K$ of $f - \tilde{p}_n$ in $[-1, 1]$ are computed, which we expect to be $\mathcal{O}(N^{-1})$ approximations to the roots of $f - p_n^{L_1}$. Finally, (3) we solve another LP to obtain $p_n^{\ell_1}$, which is a better approximant to $p_n^{L_1}$ than \tilde{p}_n , with a discretization scheme that forms a finer mesh near the roots: We take $\approx N/2$ points on $\cup_{i=1}^K [r_i - \delta, r_i + \delta]$, where $\delta = 4/N$, taking equispaced points on each subinterval. We then take $\approx N/2$ more points on $[-1, 1]$, outside the subintervals, again uniformly, i.e., the grid is much coarser (see Figure 5.2 (right)). We take the weights w_j according to the midpoint rule. We thus take a mesh $\mathcal{O}(1/N^2)$ rather than $\mathcal{O}(1/N)$ near the roots, while still having a $\mathcal{O}(1/N)$ mesh elsewhere. This refinement of the quadrature rule is observed to improve the accuracy to $\|p_n^{\ell_1} - p_n^{L_1}\|_1 = \mathcal{O}(N^{-2})$, as the quadrature error at the roots has been improved from $\mathcal{O}(N^{-2})$ to $\mathcal{O}(N^{-4})$. We then solve (2.7) by a standard technique of casting it as linear programming [13], namely,

$$(5.2) \quad \begin{aligned} & \min_{u_0, \dots, u_N, v_0, \dots, v_N, c_0, \dots, c_n} \sum_{i=0}^N w_i (u_i + v_i) \\ & \text{subject to} \quad u_i \geq 0, \quad v_i \geq 0, \quad -v_i \leq f(y_i) - \sum_{j=0}^n c_j U_j(y_i) \leq u_i, \quad 0 \leq i \leq N. \end{aligned}$$

Note that we do not use SPGL1 or the Chebyshev points from (2.2) in the refinement stage. This is because SPGL1 requires the computation of the null space V^\top , which can be more expensive. Due to the sparsity structure of LP, we find that the MOSEK optimization toolbox [3] (using its MATLAB interface) offers an efficient solver.

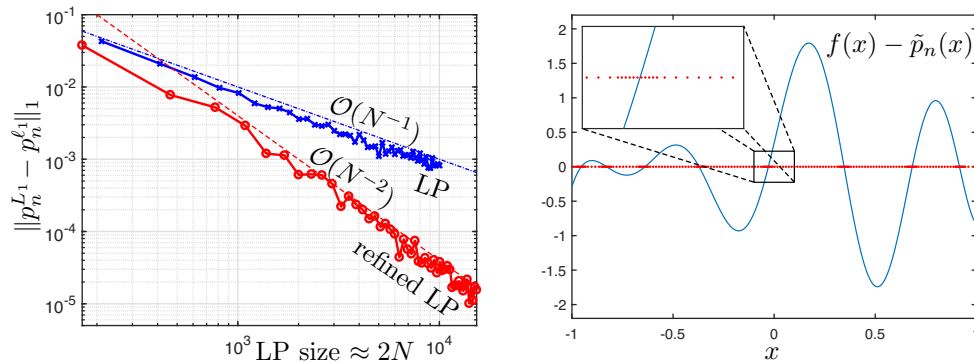


FIG. 5.2. Left: The error $\|p_n^{L_1} - p_n^{\ell_1}\|_1$ for $f(x) = \exp(x)\sin(10x)$ and $n = 10$ with and without refinement compared against the number of LP variables, which is roughly $2N$. Here, m is the number of sample points used to discretize the continuous L_1 optimization problem. Right: Sample points (red dots) used in the refined LP with $n = 5$. The mesh is much finer near the roots of $f - \tilde{p}_n$, so that the discretization error is significantly reduced. Here, \tilde{p}_n is the solution of the first (unrefined) LP.

In Figure 5.2 (left) we show the error $\|p_n^{L_1} - p_n^{\ell_1}\|_1$ with the LP solution for $10^2 \leq N \leq 10^4$, with and without the refinement. Note that the number of decision variables in LP (5.2) is $2(N+1) + n + 1$, with $4(N+1)$ inequality constraints.

5.3. Iterative procedure: Newton's method. To improve the initial guess obtained in subsection 5.2 we employ Newton's method based on the ideas in Watson's algorithm [34, sect. 4], which is a globally convergent (under mild assumptions) iterative method for computing $p_n^{L_1}$ when the set $S = \{x \in [-1, 1] : f(x) = p_n^{L_1}(x)\}$ has zero Lebesgue measure. We assume this below; otherwise f is a corrupted polynomial, which would be detected by (2.7) if the corruption is small.

When the set S has zero Lebesgue measure, an alternative characterization of $p_n^{L_1}$ is [27, Thm. 14.1]

(5.3)

$$\int_{-1}^1 s(x)q(x)dx = 0, \quad s(x) = \text{sign}(f(x) - p_n^{L_1}(x)) = \begin{cases} 1, & f(x) - p_n^{L_1}(x) \geq 0, \\ 0, & f(x) - p_n^{L_1}(x) = 0, \\ -1, & f(x) - p_n^{L_1}(x) < 0, \end{cases}$$

for all $q \in \mathcal{P}_n$. We propose to apply Newton's method to (5.3). By using the Chebyshev polynomials of the second kind as a basis for \mathcal{P}_n , we define a vector-valued operator $L : \mathbb{R}^{n+1} \mapsto \mathbb{R}^{n+1}$ given by

(5.4)

$$L[(c_0, \dots, c_n)^\top] = (\mu_0, \dots, \mu_n)^\top, \quad \mu_j = \int_{-1}^1 \text{sign}\left(f(x) - \sum_{j=0}^n c_j U_j(x)\right) U_j(x) dx.$$

We note that $L[(c_0^*, \dots, c_n^*)^\top] = \underline{0}$ if and only if $p_n^{L_1} = \sum_{j=0}^n c_j^* U_j$ from (5.3), and we propose to use Newton's method on L to find it.

Newton's method tells us to perform the following iteration:

$$(5.5) \quad \underline{c}^{(k+1)} = \underline{c}^{(k)} - J_k^{-1} L[\underline{c}^{(k)}], \quad (J_k)_{i,j} = \frac{\partial}{\partial c_j^{(k)}} \int_{-1}^1 \operatorname{sign}\left(f(x) - \sum_{t=0}^n c_t^{(k)} U_t(x)\right) U_i(x) dx.$$

Moreover, it can be shown that J_k can be expressed as [34]

$$(5.6) \quad J_k = 2V_k^\top \operatorname{diag}\left(\frac{1}{e'_k(r_1)}, \dots, \frac{1}{e'_k(r_K)}\right) V_k, \quad e_k(x) = f(x) - \sum_{t=0}^n c_t^{(k)} U_t(x),$$

where r_1, \dots, r_K are the roots of $e(x)$ and V_k is the Chebyshev–Vandermonde matrix at r_1, \dots, r_K , i.e., $(V_k)_{i,j} = U_j(r_i)$.

At the k th Newton iteration, we must calculate the roots of $e_k(x) = f(x) - \sum_{t=0}^n c_t^{(k)} U_t(x)$, evaluate μ_j for $0 \leq j \leq n$ and $e'_k(x)$ at r_1, \dots, r_K , form J_k using (5.6), and then solve an $(n+1) \times (n+1)$ dense linear system where the right-hand side is $L[\underline{c}]$. All these operations can be performed conveniently and robustly in Chebfun to an accuracy of essentially machine precision [16]. The dominant computation in each Newton's step lies either in the evaluation of μ_j in (5.4), which costs $\mathcal{O}(nm^2)$ where m is the Chebfun degree of f , or the linear system $\mathcal{O}(n^3)$, for a total of $\mathcal{O}(n^2(m+n))$ complexity. Typically Newton converges within a handful of iterations.

As Watson notes [34], a small modification for the formula for J_k in (5.6) is required when $e'_k(r_j) = 0$ for some r_j , e.g., set $J = I$, or when V is rank-deficient, e.g., set $J := J + \delta I$ for some small $\delta > 0$. Under mild restrictions, this modified Newton's method generically converges to $p_n^{L_1}$ at a quadratic rate [34].

5.4. Stopping criterion: Near-best condition. It is important to have a stopping criterion to determine when Newton's method in (5.5) should be terminated. The simplest criterion could be to stop computing iterates as soon as $\|\underline{c}^{(k+1)} - \underline{c}^{(k)}\|_2 < \epsilon \|\underline{c}^{(k)}\|_2$, where $\epsilon > 0$ is a small parameter. However, we prefer to stop Newton's method as soon as $\max_{0 \leq i \leq n} |(L[\underline{c}^{(k)}])_i| < \epsilon \|f\|_1$ because it leads to a near-best guarantee.

THEOREM 5.1. *Let $f : [-1, 1] \rightarrow \mathbb{R}$ be a continuous function and $\underline{c} \in \mathbb{R}^{n+1}$. If $\frac{2}{\pi}(n+2)^2 \max_{0 \leq i \leq n} |(L[\underline{c}])_i| < 1$, then*

$$(5.7) \quad \left\| f - \sum_{j=0}^n c_j U_j \right\|_1 \leq \left(\frac{1}{1 - \frac{2}{\pi}(n+2)^2 \max_{0 \leq i \leq n} |(L[\underline{c}])_i|} \right) \|f - p_n^{L_1}\|_1,$$

where U_j is the degree j Chebyshev polynomial of the second kind.

Proof. Let $p_n \in \mathcal{P}$, and define $s_p(x) = \pm \operatorname{sign}(f(x) - p_n(x))$ so that $\|f - p_n\|_1 = \int_{-1}^1 s_p(x)(f(x) - p_n(x))dx$. Then,

$$(5.8) \quad \|f - p_n^{L_1}\|_1 \geq \int_{-1}^1 s_p(x)(f(x) - p_n^{L_1}(x))dx = \|f - p_n\|_1 + \int_{-1}^1 s_p(x)(p_n(x) - p_n^{L_1}(x))dx.$$

Therefore, we find that $\|f - p_n\|_1 \leq \|f - p_n^{L_1}\|_1 + \int_{-1}^1 s_p(x)(p_n^{L_1}(x) - p_n(x))dx$. Expanding $p_n^{L_1} - p_n$ in a Chebyshev series, we find that

$$p_n^{L_1}(x) - p_n(x) = \sum_{i=0}^n a_i U_i(x), \quad a_i = \frac{2}{\pi} \int_{-1}^1 (p_n^{L_1}(x) - p_n(x)) U_i(x) \sqrt{1-x^2} dx.$$

Since $|U_i(x)| \leq (i+1)$ for $x \in [-1, 1]$ [24, eqs. (18.14.4) and (18.7.4)], we have

$$|a_i| \leq \frac{2}{\pi}(i+1)\|p_n^{L_1} - p_n\|_1 \leq \frac{4}{\pi}(i+1)\|f - p_n\|_1,$$

where the last inequality comes from the fact that $\|p_n^{L_1} - p_n\|_1 \leq \|p_n^{L_1} - f\|_1 + \|f - p_n\|_1 \leq 2\|f - p_n\|_1$. It follows that

$$\begin{aligned} (5.9) \quad \left| \int_{-1}^1 s_p(x) (p_n^{L_1}(x) - p_n(x)) dx \right| &= \sum_{i=0}^n |a_i| \left| \int_{-1}^1 s_p(x) U_i(x) dx \right| \\ &\leq \frac{2}{\pi}(n+2)^2 \|f - p_n\|_1 \max_{0 \leq i \leq n} \left| \int_{-1}^1 s_p(x) U_i(x) dx \right|, \end{aligned}$$

where the inequality holds since $\sum_{i=0}^n (i+1) = (n+1)(n+2)/2 \leq (n+2)^2/2$. By using (5.9) to bound the right-hand side of (5.8), the result follows by rearranging. \square

Theorem 5.1 shows that one can track the quantity $\max_{0 \leq i \leq n} |(L[\underline{c}^{(k)}])_i|$ for $k \geq 0$ to estimate how close the current Newton iterate is to computing $p_n^{L_1}$. In practice, we terminate Newton's method as soon as $\max_{0 \leq i \leq n} |(L[\underline{c}^{(k)}])_i| < 10^{-14}\|f\|_1$. It can happen that the initial guess in subsection 5.2 already satisfies the stopping criteria in which case no Newton iterations are computed.

Appendix A. Corruption away from the endpoints. Lemma 2.3 shows that polynomials of degree $\leq n$ cannot be too concentrated in a set of measure $< 1/(n+1)^2$. An alternative analysis can use the following bound on the derivative of a polynomial [6, Chap. 5]:

$$|p'(x)| \leq \frac{n}{\sqrt{1-x^2}} \|p\|_\infty, \quad -1 < x < 1,$$

for any $p \in \mathcal{P}_n$. This inequality is useful when x is away from ± 1 and suggests that polynomials of degree $\leq n$ are less concentrated in the middle of $[-1, 1]$ compared to near ± 1 . This turns out to be the case.

THEOREM A.1. *Let $\Omega_s \subseteq [-1, 1]$ with Lebesgue measure $s \geq 0$, and suppose that $\zeta = \max\{|x| : x \in \Omega_s\}$ is such that $1 - \zeta \geq 1/n$. For $n \geq 1$, we have*

$$(A.1) \quad \int_{\Omega_s} |p(x)| dx \leq \frac{sn^{3/2}}{(1-\zeta^2)^{1/4}} \int_{-1}^1 |p(x)| dx$$

for any polynomial $p \in \mathcal{P}_n$.

Proof. Let $p \in \mathcal{P}_n$, and let $\|p\|_{\Omega_s}$ denote its absolute maximum in Ω_s . By Bernstein's inequality [6, Chap. 5] we have that $|p'(x)| \leq n\|p\|_\infty/\sqrt{1-x^2}$ for $x \in \Omega_s$ and $|p'(x)| \leq n^2\|p\|_\infty$ for $x \in [-1, 1]$. Let $x^* \in [-1, 1]$ be such that $|p(x^*)| = \|p\|_\infty$. Using these two inequalities, we observe that there is an interval $\mathcal{I} \subset [-1, 1]$ containing x^* of width at least $1/n^2$ for which $p(x)$ is of the same sign as $p(x^*)$. The area of the triangle of width $1/n^2$ and height $|p(x^*)|$ is $I_1 = |p(x^*)|/(2n^2)$. Next use the same argument for $x^{*,\Omega} \in \Omega_s$, such that $|p(x^{*,\Omega})| = \|p\|_{\Omega_s}$, to obtain a triangle with area $\|p\|_{\Omega_s}^2 \sqrt{1-\zeta^2}/(2\|p\|_\infty n)$. Note that since $1-\zeta \geq 1/n$, the two triangles can be chosen to not overlap. We can thus write $\int_{-1}^1 |p(x)| dx \geq I_1 + I_2$.

Since $\int_{\Omega_s} |p(x)| dx \leq s \|p\|_{\Omega_s}$, we find that

$$\int_{\Omega_s} |p(x)| dx \leq \frac{s}{\frac{X(p)}{2n^2} + \frac{\sqrt{1-\zeta^2}}{2nX(p)}} \int_{-1}^1 |p(x)| dx, \quad X(p) = \frac{\|p\|_{L_\infty}}{\|p\|_{\Omega_s}}.$$

The function $g(x) = x/(2n^2) + \sqrt{1-\zeta^2}/(2nx)$ on $x \geq 0$ is minimized at $x_* = \sqrt{n}(1-\zeta^2)^{1/4}$. The bound in (A.1) holds since $g(x) \geq g(x_*) = (1-\zeta^2)^{1/4}n^{-3/2}$ for any $x \geq 0$. \square

Arguing as in Corollary 2.4, Theorem A.1 means that a corrupted polynomial of degree $n \geq 1$ can be exactly recovered by $p_n^{L_1}$ when $s < (1-\zeta^2)^{1/4}n^{-3/2}/2$ and $1-\zeta \geq 1/n$. For sufficiently large n , this is a relaxation of the requirements for exact recovery in section 2 when the corruption is away from ± 1 (see Figure 1.2(c) and the localized error near $x = \pm 1$ in Figure 4.2). Other results in section 3 can be relaxed by using Theorem A.1 under the restriction that the corruption occurs away from ± 1 . In particular, one can show that if $\Omega_s = [-s/2, s/2]$ with $s = n^{-3/2}/32$, then

$$\int_{[-1,1] \setminus \Omega_s} |f(x) - p_n^{L_1}(x)| dx \leq 4 \|f_0 - p_n^*\|_1,$$

where p_n^* is the best L_1 polynomial approximation of f_0 on $[-1, 1]$.

Theorem A.1 also encourages us to wildly speculate (recalling the derivation of Corollary 2.4) that the error localization of $f - p_n^{L_1}$ is usually more concentrated for functions with endpoint singularities, i.e., $|\Omega_n| = \mathcal{O}(n^{-2})$, and less concentrated for functions with singularities away from ± 1 , i.e., $|\Omega_n| = \mathcal{O}(n^{-1.5})$ or even $\mathcal{O}(n^{-1})$.

Acknowledgments. We thank Laurent Demanet for discussing the implications of the Remez inequality with us. We also thank Vanni Noferini, who was present during the initial discussions of this work. We thank Nick Trefethen and Heather Wilber for reading a draft of this manuscript and improving the text.

REFERENCES

- [1] B. ADCOCK, A. BAO, AND S. BRUGIAPAGLIA, *Correcting for unknown errors in sparse high-dimensional function approximation*, Numer. Math., 142 (2019), pp. 667–711.
- [2] B. ADCOCK, S. BRUGIAPAGLIA, AND C. G. WEBSTER, *Compressed sensing approaches for polynomial approximation of high-dimensional functions*, in Compressed Sensing and Its Applications, Springer, Cham, 2017, pp. 93–124.
- [3] MOSEK, *MOSEK Optimization Toolbox for MATLAB Manual. Version 8.1.*, 2017, <http://docs.mosek.com/8.1/toolbox/index.html>.
- [4] Z. BATTLES AND L. N. TREFETHEN, *An extension of MATLAB to continuous functions and operators*, SIAM J. Sci. Comput., 25 (2004), pp. 1743–1770.
- [5] Y. BENYAMINI, A. KROÓ, AND A. PINKUS, *l^1 -approximation and finding solutions with small support*, Constr. Approx., 36 (2012), pp. 399–431.
- [6] P. BORWEIN AND T. ERDÉLYI, *Polynomials and Polynomial Inequalities*, Springer Science & Business Media, New York, 2012.
- [7] J. P. BOYD, *Computing zeros on a real interval through Chebyshev expansion and polynomial rootfinding*, SIAM J. Numer. Anal., 40 (2002), pp. 1666–1682.
- [8] H. BRASS, *A remark on best L^1 -approximation by polynomials*, J. Approx. Theory, 52 (1988), pp. 359–361.
- [9] T. T. CAI, *Sharp rip bound for sparse signal and low-rank matrix recovery*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 74–93.
- [10] E. CANDÈS, M. RUDELSON, T. TAO, AND R. VERSHYNIN, *Error correction via linear programming*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05), IEEE, 2005, pp. 668–681.
- [11] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

- [12] E. J. CANDÉS, J. K. ROMBERG, AND T. TAO, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
- [13] E. J. CANDÉS AND T. TAO, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.
- [14] A. COHEN AND R. DEVORE, *Approximation of high-dimensional parametric PDEs*, Acta Numer., 24 (2015), pp. 1–159.
- [15] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [16] T. A. DRISCOLL, N. HALE, AND L. N. TREFETHEN, *Chebfun Guide*, Pafnuty Publications Oxford, 2014.
- [17] H. FIEDLER AND W. B. JURKAT, *Best L^1 -approximation by polynomials*, II, J. Approx. Theory, 63 (1990), pp. 267–272.
- [18] S. FOUCART AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Springer, Cham, 2013.
- [19] W. M. GENTLEMAN, *Implementing Clenshaw–Curtis quadrature*, II. Computing the cosine transformation, Comm. ACM, 15 (1972), pp. 343–346.
- [20] K. GLASHOFF AND R. SCHULTZ, *Über die genaue Berechnung von besten L_1 -Approximierenden*, J. Approx. Theory, 25 (1979), pp. 280–293.
- [21] J. N. LASKA, M. A. DAVENPORT, AND R. G. BARANIUK, *Exact signal recovery from sparsely corrupted measurements through the pursuit of justice*, in 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, IEEE, 2009, pp. 1556–1560.
- [22] J. C. MASON, *Near-best L_p approximations by real and complex Chebyshev series*, IMA J. Numer. Anal., 3 (1983), pp. 493–504.
- [23] J. C. MASON AND D. C. HANDSCOMB, *Chebyshev Polynomials*, Chapman and Hall/CRC, Boca Raton, FL, 2002.
- [24] F. W. OLVER, D. W. LOZIER, R. F. BOISVERT, AND C. W. CLARK, *NIST Handbook of Mathematical Functions*, Cambridge University Press, Cambridge, UK, 2010.
- [25] R. PACHÓN, R. B. PLATTE, AND L. N. TREFETHEN, *Piecewise-smooth chebfuns*, IMA J. Numer. Anal., 30 (2010), pp. 898–916.
- [26] A. M. PINKUS, *On L_1 -Approximation*, Cambridge University Press, Cambridge, UK, 1989.
- [27] M. J. D. POWELL, *Approximation Theory and Methods*, Cambridge University Press, Cambridge, UK, 1981.
- [28] T. J. RIVLIN, *An Introduction to the Approximation of Functions*, Dover, New York, 2003.
- [29] Y. SHIN AND D. XIU, *Correcting data corruption errors for multivariate function approximation*, SIAM J. Sci. Comput., 38 (2016), pp. A2492–A2511.
- [30] L. N. TREFETHEN, *Six myths of polynomial interpolation and quadrature*, Math. Today (Southend-on-Sea), 47 (2011), pp. 184–188.
- [31] L. N. TREFETHEN, *Approximation Theory and Approximation Practice*, SIAM, Philadelphia, PA, 2013.
- [32] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput., 31 (2008), pp. 890–912, <https://doi.org/10.1137/080714488>.
- [33] R. S. VARGA AND A. J. CARPENTER, *On the Bernstein conjecture in approximation theory*, Constr. Approx., 1 (1985), pp. 333–348.
- [34] G. A. WATSON, *An algorithm for linear L_1 approximation of continuous functions*, IMA J. Numer. Anal., 1 (1981), pp. 157–167.
- [35] J. WRIGHT AND Y. MA, *Dense error correction via ℓ^1 -minimization*, IEEE Trans. Inform. Theory, 56 (2010), pp. 3540–3560.