Robust Fourier and Polynomial Curve Fitting

Venkatesan Guruswami
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
Email: guruswami@cmu.edu

David Zuckerman
Department of Computer Science
University of Texas at Austin
Austin, TX, USA
Email: diz@cs.utexas.edu

***Abstract*—We consider the robust curve fitting problem, for both algebraic and Fourier (trigonometric) polynomials, in the presence of outliers. In particular, we study the model of Arora and Khot (STOC 2002), who were motivated by applications in computer vision. In their model, the input data consists of ordered pairs $(x_i, y_i) \in [-1,1] \times [-1,1]$, $i = 1, 2, \ldots, N$, and there is an unknown degree-d polynomial p such that for all but ρ fraction of the i, we have $|p(x_i) - y_i| \leqslant \delta$. Unlike Arora-Khot, we also study the trigonometric setting, where the input is from $\mathbb{T} \times [-1,1]$, where $\mathbb{T}$ is the unit circle. In both scenarios, the i corresponding to errors are chosen randomly, and for such i the errors in the y_i can be arbitrary. The goal is to output a degree-d polynomial q such that $\|p - q\|_\infty$ is small (for example, $O(\delta)$). Arora and Khot could achieve a polynomial-time algorithm only for $\rho = 0$. Daltrophe et al. observed that a simple median-based algorithm can correct errors if the desired accuracy δ is large enough. (Larger δ makes the output guarantee easier to achieve, which seems to typically outweigh the weaker input promise.)**

We dramatically expand the range of parameters for which recovery of q is possible in polynomial time. Specifically, we show that there are polynomial-time algorithms in both settings that recover q up to ℓ_∞ error $O(\delta^{.99})$ provided

1) **$\rho \leqslant \frac{c_1}{\log d}$ and $\delta \geqslant 1/(\log d)^c$, or**
2) **$\rho \leqslant c_1 \frac{\log \log d}{\log^2 d}$ and $\delta \geqslant 1/d^c$.**

Here c is any constant and c_1 is a small enough constant depending on c. The number of points that suffices is $N = \tilde{O}(d)$ in the trigonometric setting for random x_i or arbitrary x_i that are roughly equally spaced, or in the algebraic setting when the x_i are chosen according to the Chebyshev distribution, and $N = \tilde{O}(d^2)$ in the algebraic setting with random (or roughly equally spaced) x_i.

***Index Terms*—Error-correction; Polynomial regression; Reed-Solomon codes.**

I. Introduction

The curve fitting problem is, roughly, to construct a curve that passes near many input points. The curve is typically constrained to be of a certain form, such as having low degree. This problem is fundamental and has applications in many areas, including statistics, computer vision, and Fourier analysis. In statistics, it is commonly referred to as linear or polynomial regression, and is one of the most basic tools for estimating the relationships between random variables. This is used extensively in machine learning; see for example the book [Zie11]. In computer vision, boundaries of objects can often be modeled as low-degree algebraic curves, so researchers try to fit curves to their estimates of boundaries.

For periodic functions or functions on the circle, it is more natural and useful to consider low-degree trigonometric polynomials. A trigonometric polynomial in θ is a linear function in $\sin(k\theta)$ and $\cos(k\theta)$ for integral k; the largest such k in absolute value is the degree. Thus low-degree trigonometric polynomials correspond to Fourier series approximations. This again has extensive uses.

An important issue with curve fitting is the presence of outliers. Ordinary least squares tends to be influenced significantly by outliers, whereas often it is best to ignore them. Therefore, researchers have introduced robust methods that do just this. For example, RANSAC (random sample consensus) [FB81] is a popular method that works by sampling a small number of points and fitting a polynomial q to the sample. If there are no outliers in the sample, then the hope is that q fits the entire input, excluding outliers. The algorithm repeatedly samples sufficiently many times so that with high probability one sample will contain no outliers.

There are a couple of problems with this method and related techniques. First, if the degree d of the (algebraic or trigonometric) polynomial is large, then we need samples of size $d+1$ just for the error-free case, so we shouldn't be able to handle more than about $(\log d)/d$ fraction of errors. Second, even if d is small, there could be significant errors among the inliers (non-outliers). A small sample of the points may not give us enough information to determine the polynomial.

Another approach to robust curve fitting is to minimize the ℓ_1 error rather than ℓ_2 error. In particular, Lasso has been used, at least for robust linear regression [XCM10]. However, outliers can still have noticeable effects, and we don't know of any theorems along the lines we propose below.

Arora and Khot [AK03] studied a model allowing significant errors in the inliers as well as a significant number of outliers. In other words, given noisy evaluations of an unknown polynomial p, where a significant number of evaluations are allowed to have arbitrary noise, can we recover p up to some reasonable error? Arora and Khot considered only algebraic polynomials; we extend their framework to include

Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing, UC Berkeley.

Research of V. Guruswami was supported in part by NSF grant CCF-1422045.

Research of D. Zuckerman was supported in part by NSF Grants CCF-1218723 and CCF-1526952.

trigonometric polynomials as follows.

There is an unknown (algebraic or trigonometric) polynomial $p : \mathcal{D} \to [-1, 1]$ of degree d with $\|p\|_\infty \leqslant 1$.

Input: $(a_1, y_1), \ldots, (a_n, y_n) \in \mathcal{D} \times [-1, 1]$, where $|p(a_i) - y_i| \leqslant \delta$ for "most" i.

Output: Degree d (algebraic or trigonometric) polynomial q with $\|p - q\|_\infty$ as small as possible.

For algebraic polynomials, we follow Arora and Khot and use $\mathcal{D} = [-1, 1]$; for trigonometric polynomials, we use $\mathcal{D} = \mathbb{T}$, the unit circle. Before discussing the outlier model and the meaning of "most," it is instructive to investigate the outlier-free case, i.e., $|p(a_i) - y_i| \leqslant \delta$ for all i. For this discussion, view δ as a small constant. The first issue is which sets of domain elements $A = \{a_1, \ldots, a_n\}$ are allowed. The simplest natural setting may be equally spaced points. In this case, $\Theta(d)$ points are necessary and sufficient on the circle $\mathbb{T}$, but perhaps surprisingly $\Theta(d^2)$ are necessary and sufficient on the interval $[-1, 1]$. We know such a polynomial exists, by assumption, and Arora and Khot observed that we can set up a linear program to find it. The issue is that we want the input to determine this polynomial up to some reasonable error, like $O(\delta)$. Viewing the error as a polynomial and normalizing, this amounts to showing that if a low-degree polynomial is at most 1 on equally spaced points, then it is bounded over the entire domain. Such results were show by Ehlich and Zeller [EZ64] and Coppersmith and Rivlin [CR92] for the interval and Rakhmanov and Shekhtman [RS06] and Dubinin [Dub11] for the circle.

In fact, we observe that the points do not need to be equally spaced; they just need to form a dense enough cover, i.e., any point in the space must be close to some point in the set A. (This follows from an inspection of the proof for the circle in [Dub11], and a similar statement for the interval would then follow by a reduction.) We do need a cover because certain polynomials, like the Chebyshev polynomials, can be large on a small interval and close to zero otherwise. For intuition, note that even $(1-x^2)^d$ is small outside a small interval containing 0.

Arora and Khot assume the domain elements A are chosen uniformly at random. In this case $\Omega(d \log d)$ points are necessary on $\mathbb{T}$ and $\Omega(d^2 \log d)$ are necessary on $[-1, 1]$; otherwise the points probably will not form a fine enough cover. Arora and Khot showed a constructive upper bound of $O((d^2/\delta) \log(d/\delta))$ points, matching the lower bound except for terms involving δ.

Interestingly, if we can choose the points, then $\Theta(d)$ points are necessary and sufficient for both $\mathbb{T}$ and $[-1, 1]$. This follows because the algebraic case can be reduced to the trigonometric case by using the substitution $x = \cos\theta$. Thus, on the interval there will be many more points near ± 1 rather than near 0. Because of this reduction, our analysis will focus on the trigonometric case, and we will state the result for algebraic polynomials as a corollary. For these reasons, the Chebyshev distribution is useful, as sampling from this distribution on $[-1, 1]$ amounts to picking $\theta \in [0, \pi]$ uniformly and outputting $x = \cos\theta$.

We now discuss the outlier model. Simply limiting the number of outliers is not enough, because these outliers may all occur in some interval and then we lose control over the polynomial in that interval. We really need the inliers to form a suitable cover. Probably the most natural model here is that each a_i is chosen to be an outlier with some fixed probability ρ. If a point is chosen to be an outlier, then the corresponding y_i may be chosen arbitrarily by an adversary.

Arora and Khot showed that this is impossible when the outlier probability ρ is at least half. This is an information-theoretic result, in the sense that the polynomial is not uniquely determined up to small ℓ_∞ error, and in fact there can exponentially many candidate solutions which are pairwise far apart in ℓ_∞ norm (the "list-decoding regime" with large lists). They also gave an algorithm for very structured errors, even in the list-decoding regime, but it runs in exponential time. In this structured model, there are k unknown polynomials, and at each data point one polynomial p is picked randomly and y_i set to $p(x_i) + \beta$ with $|\beta| \leqslant \delta$ chosen adversarially. Their algorithm runs in time exponential in kd^2/δ.

Information-theoretically, the general, unstructured-error problem is possible to solve when the fraction of (randomly chosen) outliers is less than half. Although we don't achieve this information-theoretic bound, we do manage to achieve the first positive result in this model with outliers that is not exponential time. In particular, we give a polynomial-time algorithm that tolerates a fraction of about $1/\log d$ outliers.

Theorem 1. *(Informal for $\mathbb{T}$) Let $c \geqslant 1$ an arbitrary integer. There exists $c_1 > 0$ small enough compared to c, such that when $N \geqslant \tilde{\Omega}_c(d)$, the following holds. Let $A = \{a_1, \ldots, a_N\} \subset \mathbb{T}$ be any set of N roughly equally spaced points. In the model where for each $i \in [N]$ independently, y_i fails to satisfy $|y_i - p(a_i)| \leqslant \delta$ with probability at most ρ, there is a polynomial-time algorithm that with high probability recovers a trigonometric polynomial q of degree at most d satisfying $\|p - q\|_\infty \leqslant O(\delta^{0.99})$ provided*

1) $\rho \leqslant \frac{c_1}{\log d}$ *and* $\delta \geqslant \frac{1}{(\log d)^c}$, *or*
2) $\rho \leqslant c_1 \frac{\log \log d}{\log^2 d}$ *and* $\delta \geqslant \frac{1}{d^c}$.

Theorem 2. *(Informal for $[-1, 1]$) Let $c \geqslant 1$ an arbitrary integer. There exists $c_1 > 0$ small enough compared to c, such that when $N \geqslant \tilde{\Omega}(d^2)$ for roughly equally spaced (or uniformly random) x_i, or $N \geqslant \tilde{\Omega}_c(d)$ for x_i chosen according to the Chebyshev distribution, the following holds. In the model where for each $i \in [N]$ independently, y_i fails to satisfy $|y_i - p(x_i)| \leqslant \delta$ with probability at most ρ, there is a polynomial-time algorithm that with high probability recovers an algebraic polynomial f of degree at most d satisfying $\|p - f\|_\infty \leqslant O(\delta^{0.99})$ provided*

1) $\rho \leqslant \frac{c_1}{\log d}$ *and* $\delta \geqslant \frac{1}{(\log d)^c}$, *or*
2) $\rho \leqslant c_1 \frac{\log \log d}{\log^2 d}$ *and* $\delta \geqslant \frac{1}{d^c}$.

One simple approach, suggested e.g., by Daltrophe et al. [DDL12], is that if we look in a small enough interval, the value of the polynomial p will be close enough to constant that any outliers can be detected and removed. One can therefore take the median of values in this small interval to correct and find outliers ([DDL12] take the average). This simple approach works if δ is sufficiently large, since when δ is large, the output error $\|p-q\|_\infty$ can also be fairly large. Specifically, this median approach works when p doesn't vary by more than $O(\delta^{.99})$ on intervals where there are $c\log N$ points for a large enough constant c. Since the derivative of p is at most d on $\mathbb{T}$ and at most d^2 on $[-1,1]$, this means that for random or roughly equally-spaced points, medians work on $\mathbb{T}$ if $\delta^{.99} = \Omega(d(\log N)/N)$, and on $[-1,1]$ if $\delta^{.99} = \Omega(d^2(\log N)/N)$. Our approach works for smaller δ.

Another extreme case is when the good points have no error, i.e., the inlier noise $\delta = 0$. This is similar to the finite field setting, where Welch and Berlekamp [WB86] showed how to curve fit with the optimum number of outliers (i.e., optimally error-correct Reed-Solomon codes). Other authors, such as Arora-Khot and Daltrophe et al. [DDL12] specifically mention that the Welch-Berlekamp techniques appears to require $\delta = 0$. Kaltofen and Yang [KY13], [KY14] do manage to handle some noise in the inliers, but they require small relative error. They don't state a theorem along these lines, but they discuss handling relative error 10^{-7}. Regardless, handling relative error doesn't suffice for our purposes.

Despite the earlier lack of progress in applying Welch-Berlekamp to handle significant δ, we manage to use similar techniques to do just that. After giving some background and setting up the model, we give an overview of our techniques and then the proof itself in the next section.

Independent and Subsequent Work: In independent work, Chen et al. [CKPS16] give an efficient algorithm that takes as input samples from $p(x)+g(x)$, with p a degree-d polynomial and $\|g\|_2 \leqslant \delta$, and outputs a degree-d polynomial q with $\|p-q\|_2 = O(\delta)$. This doesn't imply our result, as outliers could cause $\|g\|_2$ to be quite large. Our result doesn't imply their theorem either, but for some parameters it can give a stronger ℓ_∞ guarantee for a weaker quantitative bound. To see this, observe that for uniformly random samples x_i, we have $\Pr[|g(x_i)| \geqslant s\delta] \leqslant 1/s^2$. Therefore, if $1/s^2$ is at most the error ρ allowed in our theorems, and $s\delta$ is suitably large, then we can obtain q with $\|p-q\|_\infty = O((s\delta)^{.99})$. We would need $s = O(\sqrt{\log d})$ or $O(\log d)$.

In exciting subsequent work, Eric Price used different techniques to obtain an essentially optimal algorithm for our setting [Pri16]. His algorithm works for any fraction of outliers $\rho < 1/2$, achieves error $O(\delta)$, and doesn't require any lower bound on δ. In the algebraic setting, he requires $O(d^2)$ uniformly random samples, or $O(d\log d)$ samples from the Chebyshev distribution.

II. Fitting a Fourier polynomial to noisy data

We consider bounded functions defined on the unit circle $\mathbb{T}$, whose points we will parameterize by angles $\theta \in [0, 2\pi)$. The distance $d_{\mathbb{T}}(\theta,\alpha)$ between two points $\theta,\alpha \in \mathbb{T}$ is given by the length of the arc on the unit circle between those angles.

We call a subset $A \subset \mathbb{T}$ an ϵ-cover if for all $\theta \in \mathbb{T}$, there exists $\alpha \in A$ such that $d_{\mathbb{T}}(\theta,\alpha) \leqslant \epsilon/2$. Equivalently, each interval (arc) of length ϵ contains a point of A. By dropping points, it is easy to see that an ϵ-cover may be converted to a 2ϵ-cover so that distances between consecutive points is in the range $[\epsilon/2, 2\epsilon]$. This is summarized in the following lemma, whose proof we defer to the appendix.

Lemma II.1. *An ϵ-cover on the circle $\mathbb{T}$ may be efficiently converted to a 2ϵ-cover such that distances between consecutive points lie in the range $[\epsilon/2, 2\epsilon]$. If the new cover has N points, then distances between consecutive points lie in the range $[\pi/(2N), 8\pi/N]$.*

Throughout this section, $p : \mathbb{T} \to [-1,1]$ will denote an unknown trigonometric polynomial of degree at most d (i.e., $p(\theta) = \sum_{j=0}^{d}(r_j\cos(j\theta) + s_j\sin(j\theta))$ for some real coefficients r_j, s_j). The goal is to reconstruct p, given its noisy evaluations at a sufficiently large number N of roughly equally spaced points $0 \leqslant a_0 < a_1 < \cdots < a_{N-1} < 2\pi$ on the circle satisfying

$$\pi/(2N) \leqslant d_{\mathbb{T}}(a_i, a_{i+1}) \leqslant 8\pi/N \qquad (1)$$

for $1 \leqslant i < N$, where we let $a_N = a_0$. We denote $[N] = \{0,1,\ldots,N-1\}$, and $A = \{a_0, a_1, \ldots, a_{N-1}\}$ to be the set of evaluation points.

Specifically, we are given a collection of pairs $(a_i, y_i) \in \mathbb{T}\times[-1,1]$ where y_i is close to $p(a_i)$ (within δ for some accuruacy parameter δ) on many points, but on the erroneous points we have no guarantee on the relation of y_i to $p(a_i)$ (these points are outliers with arbitrary errors).

Let us a call a point a_i *accurate* if $|p(a_i) - y_i| \leqslant \delta$, and an *outlier* otherwise. Let $\mathcal{A} = \{a_i \in A \mid a_i \text{ is accurate}\}$ and $\mathcal{O} = \{a_i \in A \mid a_i \text{ is an outlier}\}$ denote the corresponding subsets.

Let us define a point a_i to be *r-reliable*, or simply *reliable* when the integer parameter r is clear from context, if a_i is accurate, and $r-1$ points on either side of a_i are also accurate (i.e., a_i is in the middle of an accurate run of $(2r-1)$ points $a_{i-(r-1)}, ,..., a_{i-1}, a_i, a_{i+1}, \ldots, a_{i+r-1}$, where the operations on the subscript are understood to be mod N). The corresponding set of reliable points is denoted $\mathcal{R} = \{a_i \in A \mid a_i \text{ is reliable}\}$. The notion of reliable points will play an important role in our algorithm design and analysis.

A. Classical Approximation Theorems and Outlier-Free Curve-Fitting

We will use two classical tools from the theory of trigonometric polynomials: Jackson's theorem on approximating arbitrary functions by low-degree trigonometric polynomials, and discrete Remez-type inequalities to bound the norm of a trigonometric polynomial based on its values at a fine enough discrete cover of points. For a function $f : \mathbb{T} \to [-1, 1]$, we define $\|f\|_\infty = \sup\{|f(\theta)| : \theta \in \mathbb{T}\}$.

Proposition II.2 (Jackson approximation theorem for trigonometric polynomials)**.** *[Jac30] Let $f : \mathbb{T} \to [0, 1]$ be a continuous function with*

$$|f(x) - f(y)| \leqslant \Lambda |x - y|$$

for some Lipschitz constant $\Lambda < \infty$, uniformly over all $x, y \in \mathbb{T}$. Then, for $m \geqslant 1$, there exists a degree m trigonometric polynomial p, taking values in the range $[0, 1]$, such that

$$\|f - p\|_\infty \leqslant c_1 \cdot \frac{\Lambda}{m}$$

for some absolute constant $c_1 < \infty$.

For more on Jackson's theorem and other topics in approximation theory, see for example [Che82].

Proposition II.3 (Discrete Remez inequality)**.** *Let p be a trigonometric polynomial of degree m. Let $A \subset \mathbb{T}$ be a finite $\frac{2\pi}{M}$-cover for some integer $M \geqslant 2m$. Suppose $|p(\alpha)| \leqslant \gamma$ for all $\alpha \in A$. Then,*

$$\sup\{|p(\theta)| : \theta \in \mathbb{T}\} \leqslant \frac{\gamma}{\cos(\pi m/(2M))} \tag{2}$$

In particular, if we take $M = 2m$, then $\|p\|_\infty \leqslant \sqrt{2}\gamma < 2\gamma$.

An inequality of this form, with a different constant multiplying γ in (2), can be deduced from Bernstein's bound on the derivatives of trigonometric polynomials. For example, see the analogous statement about algebraic polynomials in Rivlin's book [Riv81], pages 37-38, giving a proof due to Ehlich and Zeller [EZ64]. For the case when A consists of M equally spaced points on the circle, inequalities of the above form, with different constants multiplying γ in (2), have been shown in a few places, including [RS06], [Dub11]. An inspection of the proof in [Dub11] shows that it works for a $\frac{2\pi}{M}$-cover as well.

The above immediately implies an efficient algorithm, based on a natural linear program, to interpolate (a global approximation of) a polynomial from its values at a fine enough set of points when there are no outliers.

Lemma II.4 (Outlier-free interpolation)**.** *There exists an absolute constant $\kappa \in (0, 1)$ for which the following holds. There is an algorithm based on linear programming that, given as input $\{(a_i, y_i) \in \mathbb{T} \times [-1, 1] \mid i \in [N]\}$ where the a_i's form a $\frac{\kappa}{d}$-cover and $|y_i - p(a_i)| \leqslant \delta \; \forall i \in [N]$ for some degree $\leqslant d$ trigonometric polynomial, finds a trigonometric polynomial q of degree at most d satisfying $\|p - q\|_\infty \leqslant 2\delta$ using $\mathrm{poly}(d)$ operations.*

Note that the above implies one can interpolate a trigonometric polynomial from its values at $O(d)$ (roughly) equally spaced points. In contrast, for algebraic polynomials defined on $[-1, 1]$, one needs $\Omega(d^2)$ such equally spaced points [AK03]. However, if we choose the points carefully, then one can make do with $O(d)$ points, ensuring that there are enough points close to the boundaries.

B. Informal description of idea

Our approach is based on a (non-trivial) adaptation of the Welch-Berlekamp approach to error-correct Reed-Solomon codes. In that setting, we are given N pairs $(a_i, y_i) \in \mathbb{F}^2$ where the a_i's are distinct elements of a finite field $\mathbb{F}$, and the goal to recover an unknown degree d polynomial p such that $p(a_i) \neq y_i$ for at most ρ fraction of pairs (in this case, even the erroneous positions can be chosen adversarially). If $\rho N < (N - d)$, such a polynomial p if it exists is uniquely specified. If we define the *error-locator polynomial*

$$E_0(X) = \prod_{i : p(a_i) \neq y_i} (X - a_i) \; , \tag{3}$$

then the bivariate polynomial $R_0(X, Y) \stackrel{\text{def}}{=} E_0(X)(Y - p(X))$ satisfies $R_0(a_i, y_i) = 0 \; \forall i$. The approach then is to interpolate a nonzero polynomial $R(X, Y) = E(X)Y - Q(X)$ with appropriate degree restrictions on E and Q that satisfies $E(a_i)y_i - Q(a_i) = 0$ for all i, and then argue that there is no choice but for these coefficient polynomials E, Q to satisfy $\frac{Q(X)}{E(X)} = p(X)$, allowing for simple recovery of the polynomial p.

In the continuous setting, we will try to mimic this strategy. Specifically, we will attempt to argue that (with high probability) there exists an error-locating trigonometric polynomial E_0 such that (i) $|E_0(a_i)|$ is very small on the outliers, where $|p(a_i) - y_i| > \delta$, (ii) E_0 has low-degree, (iii) E_0 is bounded in absolute value by 1 everywhere, and (iv) E_0 is "large" often enough on accurate points a_i where $|p(a_i) - y_i| \leqslant \delta$. Note that in the finite field case, there is no notion of small/large, and one just had to meet requirement (i), which is trivially done by taking E_0 as in (3).

To show the existence of such an E_0, let us first drop the low-degree requirement, and imagine a piecewise linear error-indicator function $f : \mathbb{T} \to [0, 1]$ such that $f(a_i) = 0$ for outliers a_i, $f(a_i) = 1$ for the accurate points a_i, and for values between the a_i's we interpolate f linearly by a line segment joining $(a_i, f(a_i))$ with $(a_{i+1}, f(a_{i+1}))$ otherwise. Thus f is like a "mountain range" with valleys at the outliers, and peaks at the other evaluation points. We can try to approximate f by a low-degree polynomial via classical results in approximation theory, specifically Jackson's inequality (Proposition II.2). The degree of the approximating polynomial scales with the slope of our function f (which is $\Omega(N)$) and inverse of the approximation error (which we would like to be $\approx \delta$). Thus, used naively, this will yield an E_0 with degree $\gg N$ which is useless for our purposes.

Therefore, we relax the requirement on f and require instead a function $g : \mathbb{T} \to [0, 1]$ that it is large only on accurate points that are in the middle of long run of $2r$ accurate points; these are the points we called "reliable." In other words, g is an indicator function for reliable vs. outlier, and we don't care about its value at accurate points that are not reliable. This allows us to reduce the slope of g, and therefore the degree of the approximating polynomial E_0, by a factor of r.

If we naively use Jackson's theorem to approximate g within accuracy $\approx \delta$, then to keep the degree of E_0 smaller than N, we would need $r \geqslant 1/\delta$. In this case reliable points become very rare, occurring with probability $s = \exp(-1/\delta)$. Since we don't know the locations of reliable points but only their frequency, our efficient algorithm is only able to guarantee that average value of E_0 on large intervals is $\approx s$. We can then deduce that E_0 is at least s on one of these reliable points α, but this is much smaller than our desired $\Omega(1)$ value. If we use $Q(\alpha_i)/E_0(\alpha_i)$ as an approximation of $p(\alpha_i)$ (for some reliable α_i), the division by the potentially very small $E_0(\alpha_i)$ blows up the error by an $\exp(1/\delta)$ factor, completely destroying the original accuracy $|Q(\alpha_i) - p(\alpha_i)E_0(\alpha)| \leqslant O(\delta)$.

To get around this, we use Jackson's theorem to get a coarse approximation, within error 0.1 say, via a polynomial $F_0 : \mathbb{T} \to [0,1]$ of degree $O(N/r)$. The value of F_0 on outliers could now be ≈ 0.1, which is too high. To fix this, we raise F_0 to a large enough power $\ell \approx \log(1/\delta)$ to get our final choice of "approximator" E_0. This will ensure that $E_0(a_i) \leqslant \delta$ if a_i is an outlier, but also might decrease the value E_0 at reliable points; however, for suitable ℓ, we can ensure $E_0(a_i) \geqslant \delta^{0.1}$ for reliable a_i.

To accommodate the factor ℓ-fold degree increase caused by powering, we need the run length parameter $r \geqslant \log(1/\delta)$. Furthermore, even for random errors, there might be a run of $\approx \log N$ outliers, and to accommodate this in our analysis, the overall degree has to be $O(N/\log N)$ requiring an even larger run length $r \approx \log N \cdot \log(1/\delta)$. A careful choice of parameters in this approach shows that we can recover a good global approximation to the unknown p within accuracy $\delta^{0.99}$ if each a_i is an outlier independently with probability $\rho \leqslant 1/\log d$.

The algorithm uses linear programming to find polynomials Q, E of appropriate degree such that $|Q(a_i) - y_i E(a_i)| \leqslant \delta$ for each $i \in [N]$. Then, for points where $E(a_i)$ is above some threshold τ, it uses $Q(a_i)/E(a_i)$ as a good approximation to $p(a_i)$, and then recovers an approximation to p itself via outlier-free curve fitting on those points.

C. A low-degree error locator

We now formalize the above discussion and record the construction of a low-degree polynomial E_0 that is small on outliers, and much larger on reliable points.

In all statements that follow, $O(\cdot)$ is meant to hide only absolute constants; dependency on any of our parameters will be spelled out explicitly. We also use $A \lesssim B$ to denote $A \leqslant O(B)$.

By definition, if $\alpha \in \mathcal{R}$ and $\beta \in \mathcal{O}$, then

$$d_{\mathbb{T}}(\alpha, \beta) \geqslant r \cdot \frac{\pi}{2N} \ . \tag{4}$$

This implies that one can approximate the indicator function of being reliable vs. an outlier with a degree that is a factor $\Omega(r)$ smaller than what is required to approximate the indicator function of accurate vs. outlier.

Lemma II.5. *Let $\gamma \in (0, 1/4)$. There exists a trigonometric polynomial $F_0 : \mathbb{T} \to [0,1]$ of degree at most $O\left(\frac{N}{r\gamma}\right)$ satisfying $0 \leqslant F_0(\beta) \leqslant \gamma$ whenever $\beta \in A$ is an outlier, and $1 - \gamma \leqslant F_0(\alpha) \leqslant 1$ whenever $\alpha \in A$ is reliable.*

Proof. The proof follows by applying Proposition II.2 to a function piecewise linear function $g : \mathbb{T} \to [0,1]$ that satisfies $g(\alpha) = 1$ if $\alpha \in \mathcal{R}$; $g(\beta) = 0$ if $\beta \in \mathcal{O}$, and defined by linear interpolation for points outside $\mathcal{R} \cup \mathcal{O}$. The Lipschitz constant of g is at most $\frac{2N}{\pi r}$ by virtue of (4). ☐

Corollary II.6. *Let $\delta > 0$ be sufficiently small, and let $\gamma \in (\delta, 1/4)$. There exists a degree*

$$\Delta \lesssim \frac{N \cdot \log(1/\delta)}{r \cdot \gamma \log(1/\gamma)} \tag{5}$$

trigonometric polynomial $E_0 : \mathbb{T} \to [0,1]$ such that $0 \leqslant E_0(\beta) \leqslant \delta/2$ whenever $\beta \in \mathcal{O}$, and $E_0(\alpha) \geqslant \delta^{\gamma}$ whenever $\alpha \in \mathcal{R}$.

Proof. Take $E_0 = F_0^{\ell}$ for a large enough power

$$\ell = \left\lceil \frac{\ln(2/\delta)}{\ln(1/\gamma)} \right\rceil \ .$$

Since $\gamma^{\ell} \leqslant \delta/2$, we have $E_0(\beta) \leqslant \delta/2$ when $\beta \in \mathcal{O}$. Using $(1-\gamma) \geqslant e^{-8\gamma/7}$ for $\gamma < 1/4$, one can easily check that $(1-\gamma)^{\ell} \geqslant \delta^{\gamma}$ for small enough δ. ☐

D. Assumptions on the noise

For greater modularity, we will analyze our algorithm under some abstract assumptions on the errors, which we detail now before presenting the algorithm in the next subsection. We will assume the accurate and reliable points satisfy the following conditions, for parameters t, r, b, η:

1) The accurate points form a t/N-cover, for some constant $t < \infty$
2) Every set of b consecutive points has at least ηb r-reliable points, for some parameters $b < \infty$ and $\eta > 0$.

A qualitative assumption like the first one is necessary to have any hope of recovering a global approximation of p. The second assumption on reliable points arises because of the specific nature of our algorithm.

We will give a robust curve fitting algorithm that succeeds when all these parameters obey some constraints that will come out of our algorithm and its analysis. Imagine η is some absolute constant. The quality of data we are fitting improves as t and b get smaller, and r gets larger. When t is small, the accurate points occur regularly, and when b gets smaller, the window we need to look at to find a good frequency of runs of r accurate points shrinks. We will later see that when outliers are picked randomly at a small enough error rate, accurate points will occur regularly so t will be small. Further, r-reliable points for reasonably large r will also occur with good frequency in windows of modest size b.

E. Linear programming based algorithm

We now describe the algorithm to recover a good approximation to the unknown polynomial p.

Robust Fourier Curve Fit Algorithm:

INPUT: N pairs $(a_i, y_i) \in \mathbb{T} \times [-1, 1]$ with a_i's satisfying (1), and y_i's satisfying the assumptions made in Section II-D concerning the relation to the evaluations $p(a_i)$.

DESIRED OUTPUT: A trigonometric polynomial q of degree at most d such that $\|q - p\| \leqslant \delta^{\Omega(1)}$

1) Set up a linear program to find polynomials E, Q with $\deg(E) \leqslant \Delta$ and $\deg(Q) \leqslant \Delta + d$ (for Δ obeying (5)) that satisfy the following conditions (all of which are linear in the coefficients of E and Q), for every $i \in [N]$:

$$|Q(a_i) - y_i E(a_i)| \leqslant \delta \tag{6}$$

$$\sum_{j=0}^{b-1} E(a_{i+j}) \geqslant \delta^\gamma \cdot \eta b \tag{7}$$

$$E(a_i) \in [0, 1] \tag{8}$$

where in (7), $i + j$ is computed mod N.

2) Compute $T \subseteq [N]$ as $T = \{i \mid E(a_i) \geqslant \tau\}$ for threshold

$$\tau \stackrel{\text{def}}{=} \delta^\gamma \eta \ .$$

For each $i \in T$, compute $z_i = Q(a_i)/E(a_i)$.

3) Compute, using the linear programming based algorithm of Lemma II.4, a trigonometric polynomial q of degree d such that $|q(a_i) - z_i| \leqslant 4\delta^{1-\gamma}\eta^{-1}$ for $i \in T$. If no such polynomial exists, return Failure.

Before embarking on the analysis of the algorithm, which will also fix the dependencies between the various parameters, let us quickly check the feasibility of the LP in Step 1 of the algorithm.

Lemma II.7. *Under the assumptions on the error model made in Section II-D, there exist trigonometric polynomials E, Q of degrees at most Δ and $\Delta + d$ satisfying the constraints* (6) *and* (7).

Proof. Take $E = E_0$ as guaranteed by Corollary II.6, and $Q = pE_0$. For each $i \in [N]$, $|Q(a_i) - y_i E(a_i)| = |E(a_i)||p(a_i) - y_i|$. If a_i is accurate then $|p(a_i) - y_i| \leqslant \delta$ and $E(a_i) \leqslant 1$, whereas if a_i is an outlier then $E(a_i) \leqslant \delta/2$ and $|p(a_i) - y_i| \leqslant 2$. So (6) is satisfied. As for (7), every set of b consecutive points has at least ηb reliable points, and for $a_i \in \mathcal{R}$ we have $E(a_i) \geqslant \delta^\gamma$. ☐

F. Analysis

Throughout this section, to avoid repeating this, we assume that the actual evaluations $p(a_i)$ of the unknown polynomial p, and their noisy versions y_i, satisfy the two assumptions on errors from Section II-D, namely that the correct points form a t/N-cover, and every b consecutive points has a fraction η of reliable points.

Lemma II.8. *Assume that $\Delta + d \leqslant \frac{\alpha_0 N}{t}$ for some small enough absolute constant $\alpha_0 > 0$. Then the polynomials Q and E found in Step 1 of the algorithm satisfy $\|Q - pE\|_\infty \leqslant 4\delta$.*

Proof. If a_i is correct, then by triangle inequality $|Q(a_i) - p(a_i)E(a_i)|$ is at most

$$|Q(a_i) - y_i E(a_i)| + |y_i - p(a_i)||E(a_i)| \leqslant 2\delta \ .$$

Since the correct points form a t/N-cover, by Proposition II.3, if $\Delta + d \leqslant \frac{\alpha_0 N}{t}$ for some absolute constant $\alpha_0 > 0$ (in fact we can take $\alpha_0 = \pi$), then we must have $|Q(\theta) - p(\theta)E(\theta)| \leqslant 4\delta$ for every $\theta \in \mathbb{T}$. ☐

Lemma II.9. *Assume that $b \leqslant \beta_0 N/d$ for some small enough absolute constant $\beta_0 > 0$, and that $\|Q - pE\|_\infty \leqslant 4\delta$. Then Step 3 succeeds in finding a degree d trigonometric polynomial q and any such polynomial q satisfies*

$$\|p - q\|_\infty \leqslant 16 \cdot \eta^{-1}\delta^{1-\gamma} \ .$$

Proof. Note that for any $i \in T$, we have $|p(a_i) - z_i| \leqslant \frac{4\delta}{\delta^\gamma \eta}$, so Step 3 will succeed in finding a polynomial q such that $|q(a_i) - z_i| \leqslant 4\delta^{1-\gamma}\eta^{-1}$ for $i \in T$. Thus $|q(a_i) - p(a_i)| \leqslant 8\delta^{1-\gamma}\eta^{-1}$ for $i \in T$.

The condition (7) implies that any set of b consecutive evaluation points contains an a_j with $E(a_j) \geqslant \delta^\gamma$. This together with (1) implies that the set $\{a_i \mid i \in T\}$ computed in Step 2 is an $\frac{8\pi b}{N}$-cover. For $b \leqslant \beta_0 N/d$ for small enough $\beta_0 > 0$ (in fact one can take $\beta_0 = 1/8$), $\{a_i \mid i \in T\}$ forms a sufficiently fine cover, allowing us to apply Proposition II.3 and conclude that $\|q - p\|_\infty \leqslant 16\delta^{1-\gamma}\eta^{-1}$. ☐

Summarizing the analysis, the requirements on the parameters in Lemmas II.8 and II.9 which suffice for recovery of q with $\|q - p\|_\infty$ small are

$$r \cdot \gamma \log(1/\gamma) \gtrsim t \log(1/\delta) \qquad \text{and} \qquad N \gtrsim bd \ . \tag{9}$$

G. Random errors

We now compute the values of t and b (for the assumptions in Section II-D) when the evaluation points are made outliers in an i.i.d fashion at a small enough noise rate.

Lemma II.10. *For the noise model where each $a_i \in A$ is an outlier independently with probability ρ, with probability at least $1 - 1/N$ the set of accurate points is a t/N-cover when $t = C\frac{\log N}{\log(1/\rho)}$ for a large enough absolute constant C.*

Proof. If the set $\mathcal{A}$ of accurate points is not a t/N-cover, there must be at least $\frac{t}{8\pi} > \frac{t}{30}$ consecutive a_i's which are outliers. The probability of this happening is at most $N \cdot \rho^{t/30}$, by a trivial union bound over all N sets of consecutive points starting at any of the a_i's. If $t = C\frac{\log N}{\log(1/\rho)}$ for a large enough absolute constant C, then this bound is at most $1/N$. ☐

Lemma II.11. *Let $\rho \in (0, 1/2]$ and integer $b \geqslant 64r(1 - \rho)^{-4r} \ln N$. For the noise model where each $a_i \in A$ is an outlier independently with probability ρ, the following holds with probability at least $1 - 1/N$:*

Every set of b consecutive points in A contains at least $(1-\rho)^{4r}b/4$ points that fall in the subset $\mathcal{R}$ of r-reliable points.

Proof. Partition each interval of length b into $\ell \geqslant 16(1-\rho)^{-4r}\ln N$ disjoint intervals of length $4r$. The probability that a particular interval of length $4r$ contains only accurate points is $q = (1-\rho)^{4r}$. Call such an interval *always-accurate*. An always-accurate interval contains at least $2r$ reliable points. Fix any interval J of length b containing ℓ subintervals of length $4r$. The expected number of always-accurate subintervals is $\mu = \ell q \geqslant 16\ln N$, so by a Chernoff bound,

$$\Pr[J \text{ fails to contain } \mu/2 \text{ always-accurate subintervals}] \leqslant \exp(-\mu/8) \leqslant 1/N^2 .$$

By a union bound, with probability $1-1/N$, all intervals of length b contain at least $\mu/2$ always-accurate subintervals. Thus, every b consecutive points contains at least $(\mu/2)(2r) = r\ell q = b(q/4)$ accurate points, as required. □

Thus, we may take

$$b = \Theta(r(1-\rho)^{-4r}\log N) \qquad \text{and} \qquad \eta \geqslant \frac{1}{4}(1-\rho)^{4r} \quad (10)$$

in the claim that every set of b consecutive points has at least ηb reliable points.

Finally we need to pick parameters carefully subject to the requirements (9) and (10), and $t \gtrsim \frac{\log N}{\log(1/\rho)}$ from Lemma II.10. We will make the following choices which can be checked to satisfy all these requirements for suitable constants e_1, e_2, e_3, e_4, assuming the degree d is assumed to be sufficiently large.[1] We will make two choices of ρ, depending on error we are willing to tolerate in the final approximation. For $\delta \geqslant 1/(\log d)^c$, we will take $\rho = c_1/\log d$, and if greater accuracy $\delta \geqslant 1/d^c$ is desired, we will take $\rho = c_1 \log\log d/(\log d)^2$, where c is an arbitrary constant, and c_1 is small enough as a function of c (and e_2, which can be taken to be an absolute constant).

$$\gamma = 1/100 \quad (11)$$

$$t = e_1 \frac{\log d}{\log\log d} \text{for } e_1 = 2C \text{ where } C \text{ is from Lemma II.10}$$

$$r = \left\lfloor e_2 \frac{\log d}{\log\log d}\log(1/\delta)\right\rfloor \text{ for } e_2 = e_2(e_1) \text{ large enough}$$

$$\rho = \frac{e_3}{\log d} \text{when } \delta \geqslant \tfrac{1}{(\log d)^c} \quad (\text{or } \rho = \tfrac{e_3 \log\log d}{\log^2 d} \text{ when } \delta \geqslant \tfrac{1}{d^c})$$

$$\text{for } e_3 = e_3(c, e_2) \text{ small enough}$$

$$N = \left\lfloor e_4 d \frac{\log^2 d}{\log\log d}\log(1/\delta)\right\rfloor \text{ for } e_4 = e_4(e_2) \text{ large enough} \quad (12)$$

The only thing that might require some justification is the bound on N. We know that

$$N \gtrsim bd \gtrsim dr(1-\rho)^{-4r}\log N \gtrsim dr \cdot e^{8\rho r}\log N . \quad (13)$$

For $\delta \geqslant 1/(\log d)^c$, taking $\rho = c_1/\log d$ for small enough c_1, we can ensure $e^{8\rho r} \leqslant \delta^{-1/(c\log\log d)} \leqslant 2$. Simmilarly, for $\delta \geqslant 1/d^c$, taking $\rho = c_1\log\log d/(\log d)^2$, we can ensure $e^{8\rho r} \leqslant \delta^{-1/(c\log d)} \leqslant 2$. Thus the bound (13) on N is satisfied if $\frac{N}{\log N} \gtrsim d\frac{\log d}{\log\log d}\log(1/\delta)$ for some c_2 large enough compared to c. As $\log\log(1/\delta) \ll \log d$ for either of our assumptions on δ, this condition is in turn met by the choice in (12).

Also as we ensured $e^{8\rho r} \leqslant 2$,

$$\eta \geqslant \frac{1}{4}(1-\rho)^{4r} \geqslant \frac{1}{4}e^{-8\rho r} \geqslant \frac{1}{8} . \quad (14)$$

We are finally ready to claim our result for recovering from randomly inflicted outliers.

Theorem 1. *Let $c \geqslant 1$ an arbitrary integer. There exists $c_1 > 0$ small enough compared to c and C large enough, such that when d is a sufficiently large integer and $N \geqslant Cd\frac{\log^2 d}{\log\log d}\log(1/\delta)$, the following holds. In the model where for each $i \in [N]$ independently, y_i fails to satisfy $|y_i - p(a_i)| \leqslant \delta$ with probability at most ρ, the Robust Fourier Curve Fit Algorithm runs in polynomial time and with high probability recovers a trigonometric polynomial q of degree at most d satisfying $\|p-q\|_\infty \leqslant O(\delta^{0.99})$ provided*

1) $\rho \leqslant \frac{c_1}{\log d}$ *and* $\delta \geqslant 1/(\log d)^c$, *or*
2) $\rho \leqslant c_1\frac{\log\log d}{\log^2 d}$ *and* $\delta \geqslant 1/d^c$.

In the model where the a_i are chosen uniformly on $\mathbb{T}$, we need $N \geqslant Cd\frac{\log^3 d}{\log\log d}\log(1/\delta)$.

Proof. We pick parameters t, r, b as above, and since the conditions of Lemmas II.8 and II.9 are met for our choices, and the noise assumptions of Section II-D hold with high probability, the algorithm finds a degree d polynomial q such that $\|p-q\|_\infty \leqslant 16\eta^{-1}\delta^{1-\gamma}$. Using the lower bound on η from (14), and $\gamma = 1/100$, we get $\|p-q\| \leqslant 128\delta^{0.99}$.[2] For random a_i, a coupon collector-type argument yields that N points gives a $O\left(\frac{\log N}{N}\right)$-cover

The polynomial runtime of the algorithm is evident, as the main computation is solving two linear programs, one in Step 1 and another for outlier-free interpolation in Step 3. The linear programs can be solved in time polynomial in N and the number of bits of precision with which the input (a_i, y_i) is given. □

III. ROBUST POLYNOMIAL CURVE FITTING OVER $[-1,1]$

We curve fit polynomials over $[-1,1]$ by reducing to curve fitting over the circle $\mathbb{T}$. We convert between the two domains with the substitutions $x_i = \cos a_i$. For this, we need the Chebyshev polynomials.

[1] We fix $\gamma = 1/100$ only for simplicity, one can take it smaller to improve the approximation guarantee of the final polynomial as guaranteed by Lemma II.9

[2] The 0.99 can of course be replaced by any constant bounded away from 1 with a change in choice of γ.

Definition III.1. *The Chebyshev polynomials* $\{T_d\}$ *of the first kind is defined by* $\cos(d\theta) = T_d(\cos\theta)$.

Robust Polynomial Curve Fit Algorithm:

INPUT: N pairs $(x_i, y_i) \in [-1,1] \times [-1,1]$ with y_i's satisfying the assumptions made in Section II-D concerning the relation to the evaluations $p(a_i)$.

DESIRED OUTPUT: An algebraic polynomial f of degree at most d such that $\|f-p\| \leqslant \delta^{\Omega(1)}$

1) Convert algebraic data points (x_i, y_i) to trigonometric data points (a_i, y_i) over the domain $[0,\pi] \times [-1,1]$, where $a_i = \arccos(x_i)$.
2) For each point $(a_i, y_i) \in [0,\pi] \times [-1,1]$, add the point $(2\pi - a_i, y_i)$ to get points over the full circle.
3) Run the Robust Fourier Curve Fit Algorithm on these $2N$ points to get a trigonometric polynomial q of degree d;

$$q(\theta) = \sum_{j=0}^{d} q_j \cos(j\theta) + s_j \sin(j\theta).$$

4) Output the algebraic polynomial $f(x) = \sum_{j=0}^{d} q_j T_j(x)$, where the T_j are the Chebyshev polynomials.

The following lemma shows that when the trigonometric polynomial q fits well, so does the algebraic polynomial f.

Lemma III.2. *Suppose* $|q(a_i)-y_i| \leqslant \alpha$ *and* $|q(2\pi-a_i)-y_i| \leqslant \alpha$. *Then* $|f(x_i)-y_i| \leqslant \alpha$.

Proof. First note that $q(\theta) + q(2\pi-\theta) = 2\sum_{j=0}^{d} q_j \cos(j\theta)$. Now

$$\begin{aligned} f(x_i) &= \sum_{j=0}^{d} q_j T_j(\cos a_i) \\ &= \sum_{j=0}^{d} q_j \cos(j a_i) = \frac{q(a_i) + q(2\pi - a_i)}{2} \ . \end{aligned}$$

Therefore, $|f(x_i)-y_i| \leqslant (|q(a_i)-y_i| + |q(2\pi-a_i)-y_i|)/2$, which is at most α for $i \in S$. □

Theorem 2. *Let* $c \geqslant 1$ *an arbitrary integer. There exists* $c_1 > 0$ *small enough compared to* c *and* C *large enough, such that when* d *is a sufficiently large integer and* N *is a large enough function of* d *and* $\log(1/\delta)$ *(see below), the following holds. In the model where for each* $i \in [N]$ *independently,* y_i *fails to satisfy* $|y_i - p(a_i)| \leqslant \delta$ *with probability at most* ρ, *the Robust Polynomial Curve Fit Algorithm runs in polynomial time and with high probability recovers a polynomial* q *of degree at most* d *satisfying* $\|p-q\|_\infty \leqslant O(\delta^{0.99})$ *provided*

1) $\rho \leqslant \frac{c_1}{\log d}$ *and* $\delta \geqslant 1/(\log d)^c$, *or*
2) $\rho \leqslant c_1 \frac{\log\log d}{\log^2 d}$ *and* $\delta \geqslant 1/d^c$.

The bound on N *depends on the model:*

1) *When the* x_i *are evenly spread, or simply form an* $O(1/N)$*-cover, we need* $N \geqslant Cd^2 \frac{\log^4 d}{(\log\log d)^2} \log^2(1/\delta)$.
2) *When the* x_i *are chosen uniformly on* $[-1,1]$, *we need* $N \geqslant Cd^2 \frac{\log^5 d}{(\log\log d)^2} \log^2(1/\delta)$.
3) *When the* x_i *are chosen according to the Chebyshev distribution, we need* $N \geqslant Cd \frac{\log^2 d}{\log\log d} \log(1/\delta)$.

Proof. Lemma III.2 implies that if the trigonometric polynomial q fits $1-\rho$ fraction of points, then the algebraic polynomial f fits at least $1-2\rho$ fraction of points. To verify that the points form a suitable cover, note that an ϵ^2-cover in $[-1,1]$ becomes an $O(\epsilon)$-cover in $\mathbb{T}$. This is because $|\cos(\theta+\epsilon) - \cos(\theta)| = \Omega(\epsilon^2)$, which follows from the Taylor expansion of cosine. □

ACKNOWLEDGEMENTS

We thank Eric Price for helpful comments.

REFERENCES

[AK03] S. Arora and S. Khot. Fitting algebraic curves to noisy data. *Journal of Computer and System Sciences*, 67:325–340, 2003.

[Che82] E.W. Cheney. *Approximation Theory*. AMS Chelsea Publishing, 1982.

[CKPS16] X. Chen, D.M. Kane, E. Price, and Z. Song. Sparse Fourier transforms without a frequency gap. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[CR92] D. Coppersmith and T.J. Rivlin. The growth of polynomials bounded at equally spaced points. *SIAM J. Math. Anal.*, 23:970–983, 1992.

[DDL12] H. Daltrophe, S. Dolev, and Z. Lotker. Big data interpolation an efficient sampling alternative for sensor data aggregation. In *Algorithms for Sensor Systems, 8th International Symposium on Algorithms for Sensor Systems, Wireless Ad Hoc Networks and Autonomous Mobile Entities, ALGOSENSORS 2012*, pages 66–77, 2012.

[Dub11] V.N. Dubinin. Lower bound for the discrete norm of a polynomial on the circle. *Mathematical Notes*, 90, 2011.

[EZ64] H. Ehlich and K. Zeller. Schwankung von polynomen zwischen gitterpunkten. *Math. Z.*, 86:41–44, 1964.

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[Jac30] D. Jackson. The theory of approximation. *Amer. Math. Soc. Colloq. Publ.*, XI, 1930.

[KY13] E. Kaltofen and Z. Yang. Sparse multivariate function recovery from values with noise and outlier errors. In *International Symposium on Symbolic and Algebraic Computation, ISSAC'13*, pages 219–226, 2013.

[KY14] E. Kaltofen and Z. Yang. Sparse multivariate function recovery with a high error rate in the evaluations. In *International Symposium on Symbolic and Algebraic Computation, ISSAC'14*, pages 280–287, 2014.

[Pri16] E. Price. Robust polynomial regression. Unpublished manuscript, 2016.

[Riv81] T.J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, 1981.

[RS06] E. Rakhmanov and B. Shekhtman. On discrete norms of polynomials. *Journal of Approximation Theory*, 139:2–7, 2006.

[WB86] Lloyd R. Welch and Elwyn R. Berlekamp. Error correction of algebraic block codes. *US Patent Number 4,633,470*, December 1986.

[XCM10] H. Xu, C. Caramanis, and S. Mannor. Robust regression and lasso. *IEEE Transactions on Information Theory*, 56:3561–3574, 2010.

[Zie11] A. Zielesny. *From Curve Fitting to Machine Learning*, volume 18 of *Intelligent Systems Reference Library*. Springer-Verlag, 2011.

APPENDIX

We recall the statement about converting an ϵ-cover into a net where consecutive points are well separated.

Lemma II.1. An ϵ-cover on the circle $\mathbb{T}$ may be efficiently converted to a 2ϵ-cover such that distances between consecutive points lie in the range $[\epsilon/2, 2\epsilon]$. If the new cover has N points, then distances between consecutive points lie in the range $[\pi/(2N), 8\pi/N]$.

Proof. Let S be the initial ϵ-cover. To construct the desired cover T, begin by adding an arbitrary point s of S to T. We work modulo 2π. There exists a point in the interval $[s + \epsilon/2, s + 3\epsilon/2]$; add this to T. Continue in this manner: if t was the last point added to T, then there exists a point t' in $[t+\epsilon/2, t+3\epsilon/2]$; add t' to T, as long as we haven't wrapped around, i.e., as long as t' is not in the interval $[s-\epsilon/2, s+\epsilon/2]$. Stop once a point lands in this interval. Thus each distance between consecutive points is at least $\epsilon/2$, and all are at most $3\epsilon/2$, except possibly the distance between the last point and s, which is at most 2ϵ.

The number of points N satisfies $\pi/\epsilon \leqslant N \leqslant 4\pi/\epsilon$. Thus $\epsilon/2 \geqslant \pi/(2N)$ and $2\epsilon \leqslant 8\pi/N$, so the lemma follows. $\square$