Costly Zero Order Oracles

Renato Paes Leme RENATOPPL@GOOGLE.COM
Google Research NYC

Jon Schneider JSCHNEI@GOOGLE.COM
Google Research NYC

Editors: Jacob Abernethy and Shivani Agarwal

Abstract

We study optimization with an approximate zero order oracle where there is a cost $c(\epsilon)$ associated with querying the oracle with ϵ accuracy. We consider the task of reconstructing a linear function: given a linear function $f : X \subseteq \mathbb{R}^d \to [-1, 1]$ defined on a not-necessarily-convex set X, the goal is to reconstruct $\hat{f}$ such that $|f(x) - \hat{f}(x)| \leq \epsilon$ for all $x \in X$. We show that this can be done with cost $O(d \cdot c(\epsilon/d))$. The algorithm is based on a (poly-time computable) John-like theorem for simplices instead of ellipsoids, which may be of independent interest.

This question is motivated by optimization with threshold queries, which are common in economic applications such as pricing. With threshold queries, approximating a number up to precision ϵ requires $c(\epsilon) = \log(1/\epsilon)$. For this, our algorithm has cost $O(d \log(d/\epsilon))$ which matches the $\Omega(d \log(1/\epsilon))$ lower bound up to log factors.

1. Introduction

It is often the case that in order to evaluate a function $f(\cdot)$ you need to perform a costly procedure such as solving an optimization problem or taking the average of many samples of a random variable. In such scenarios, it tends to be possible to trade-off between accuracy and cost. Some examples (in the following examples, we assume $f(x) \in [-1, 1]$):

- if $f(x) = \mathbb{E}[Z(x)]$ and we can only sample $Z(x)$ we can get an ε approximation of $f(x)$ with $1 - \delta$ probability by taking the average of $O(\varepsilon^{-2} \log(1/\delta))$ samples.
- if $f(x) = \min_{y \in Y} g(x, y)$ for $g(x, \cdot)$ is convex, we compute $f(x)$ by minimizing $g(x, \cdot)$ via a variant of gradient descent, we can an ε approximation of f in $O(1/\varepsilon^\alpha)$ where the constant α depends on the properties of g (smooth vs non-smooth) and on the algorithm used (accelerated or not).
- if we have access only to *threshold queries*, i.e., we can only ask if $f(x) \leq p$ or not, then we can get an ε-approximation of f using $O(\log(1/\varepsilon))$ queries.
- if we have access to threshold queries, but we need to choose the thresholds in a non-adaptive manner (i.e., choose all the thresholds at the same time) then we can get an ε-approximation using $O(1/\varepsilon)$ queries.

Oracle model In our model we will assume that a function $f : X \subset \mathbb{R}^d \to [-1, 1]$ is equipped with an oracle that given $x \in X$ and $\varepsilon > 0$ returns $y \in \mathbb{R}$ such that $|f(x) - y| \leq \varepsilon$ (note that we only allow queries to points in X and not more generally to points in $\mathbb{R}^d$). Each such query to the oracle has some cost $c(\varepsilon)$, where the function $c(\varepsilon)$ depends on the underlying model (for example, for threshold queries, $c(\varepsilon) = \log(1/\varepsilon)$; for stochastic zeroth order queries, $c(\varepsilon) = \varepsilon^{-2}$). For simplicity we will assume that $c(\varepsilon)$ satisfies $c(O(\varepsilon)) = O(c(\varepsilon))$ (this holds for all polynomial/sub-polynomial $c(\varepsilon)$).

Learning task We will consider the task of reconstructing a linear function. Assume $f : X \to [-1, 1]$ is linear (with X not necessarily convex). Our goal is to produce a function $\hat{f} : X \to \mathbb{R}$ such that

$$|f(x) - \hat{f}(x)| \leq \varepsilon, \forall x \in X$$

Our initial motivation for studying this problem is how to learn customer preferences through pricing queries. Assume a point $x \in X$ corresponds to a product represented by a vector of features. We can interact with the customer by offering product $x \in X$ at price $p \in \mathbb{R}$ and observing if the customer purchases it or not. This corresponds to a threshold query. How many queries are necessary to reconstruct the customer's preference?

While one main application is in pricing, threshold queries are also common in other domains such as reliability testing (given a certain force, will a material break or not) and determining effectiveness of drugs/vaccines (given a certain dose of a medication, will the patient be cured). In such cases, there is usually a threshold separating the success and failure cases and the feedback is binary.

Main Results and Techniques The natural approach for reconstructing $f(x) = \langle v, x\rangle$ is to choose d points $x_1, \ldots, x_d$ that are linearly independent in X, use our costly zero-order oracle to obtain an estimate of $f(x_i)$ for each x_i, and then write x in the basis given by x_i to compute $f(x)$ for any x. Indeed, if we choose each x_i to be the unit vector e_i (under the assumption $e_i \in X$), this method works – by measuring each $f(e_i)$ to within ε/d, we can reconstruct $f(x)$ for any $x \in [0, 1]^d$ to within ε, for a total cost of $dc(\varepsilon/d)$.

This problem becomes harder when our set X of valid queries does not necessarily contain a "good" basis like the unit vectors e_i. The difficulty with this idea is that if the matrix with rows x_i is ill-conditioned, the error can be blown up by the condition number of the matrix, which can be arbitrarily bad if the choice of x_i is naive. The problem can therefore be thought of as follows: given an arbitrary (not even necessarily convex) set X, how can we find a good basis for X?

We begin by examining the case where X is convex. By measuring along the basis formed by the axes of the John ellipsoid of X, we show it is possible to reconstruct $f(x)$ with total cost $dc(\varepsilon/d^{1.5})$.

Theorem 1 *If X is convex, there is an efficient (polynomial-time) procedure with cost $O(d \cdot c(\varepsilon/d^{1.5}))$ for reconstructing a linear function with a costly zero order oracle.*

We then consider the case where X is not-necessarily convex: for example, an arbitrary point set, or the union of several convex bodies (here we assume we have access to X through an oracle that given a vector $w \in \mathbb{R}^d$ returns a point $x \in \operatorname{argmax}_{x \in X} \langle w, x\rangle$). In this setting we can no longer use the same strategy of Theorem 1 of measuring along the axes of the John ellipsoid of (the convex hull of) X, since not all points in the convex hull of X can be directly queried.

To overcome this difficulty, we design a variant of John's theorem for simplices, showing that for any set of points X, it is always possible to efficiently find a simplex T whose vertices are contained in X such that a $\mathsf{poly}(d)$-dilation of T contains X. This theorem is the main technical result of the paper, and may be of independent interest.

Theorem 2 *Given a (not necessarily convex) set X there is a polynomial time algorithm that selects a simplex T with vertices in X such that X is contained in a translate of $-4dT$.*

By measuring along the basis formed by the vertices of this simplex, we show it is possible to efficiently reconstruct $f(x)$ with total cost $dc(\varepsilon/d)$ (note that this is a strict improvement over Theorem 1).

Theorem 3 *For any (not necessarily convex) set X, there is an efficient (polynomial-time) procedure with cost $O(d \cdot c(\varepsilon/d))$ for reconstructing a linear function over X with a costly zero order oracle.*

Finally, we show that for the specific case of threshold queries where $c(\varepsilon) = \log \varepsilon^{-1}$, we show that these algorithms are nearly optimal: both algorithms incur total cost $O(d \log(d/\varepsilon))$, but there is an information-theoretic lower bound showing that any algorithm must incur total cost at least $\Omega(d \log(1/\varepsilon))$ even when X is just the unit d-dimensional ball (Lemma 18). For other cost functions mentioned above (e.g. $c(\varepsilon) = \varepsilon^{-1}$ and $c(\varepsilon) = \varepsilon^{-2}$), the above algorithms are optimal up to a polynomial factor in d – understanding how to close this gap is an interesting open question.

1.1. Related Work

The geometric techniques we use for the reconstruction problem are closely related to the *largest simplex problem* introduced by Khachiyan (1995) and recently studied by Summa et al. (2014) and Nikolov (2015), who show approximation algorithms for maximum volume problem. Instead of the volume, we work here with with a local notion of optimality: we look for simplices where it is impossible to improve the volume by an α factor by changing any single vertex.

Learning from threshold queries is a prominent component of much recent work in revenue optimization and contextual pricing Cohen et al. (2016); Javanmard and Nazerzadeh (2016); Javanmard (2017); Mao et al. (2018); Paes Leme and Schneider (2018); Shah et al. (2019). Unlike in this paper, the majority of this work focuses on settings where we cannot query arbitrary points x, but instead where a stream of points $x_1, x_2, \ldots$ are generated online (either stochastically or adversarially) and the learner is only allowed to query those. Moreover, the typical objective is revenue and not reconstruction.

2. Warm up: convex case

Notation Given a set $Z \subset \mathbb{R}^d$ we will denote its *convex hull* by $\mathsf{Conv}(Z)$. We will call a convex hull of $d+1$ points $z_0, z_1, \ldots, z_d$ a *simplex* with vertices $z_0, z_1, \ldots, z_d$.

We will begin with the case where our set X of potential queries is convex. Here we will use John's Ellipsoid to construct a good basis of points to query.

Theorem 4 (John) *Given a convex full-dimensional set X, there is a point $u \in X$ and d orthogonal vectors $a_1, \ldots, a_d \in \mathbb{R}^d$ such that:*

$$u + E \subseteq X \subseteq u + dE \quad \textit{for} \quad E = \left\{ \sum_i z_i a_i; \sum_i z_i^2 \leq 1 \right\}$$

Lemma 5 *Let $f : X \subset \mathbb{R}^d \to \mathbb{R}$ be a linear function over a convex set and $u, a_1, \ldots, a_d$ be as in John's Theorem. Given an estimates $y_0, y_1, \ldots, y_d$ of $f(u), f(u + a_1), \ldots, f(u + a_d)$ with precision $\varepsilon/(3d^{1.5})$ then we can estimate any point in X within precision ε.*

Proof By John's Theorem, any point $x \in X$ can be written as $x = u + \sum_i d a_i z_i = u(1 - \sum_i d z_i) + \sum_i d(u + a_i) z_i$ with $\sum_i z_i^2 \leq 1$. Hence $\sum_i |z_i| \leq \sqrt{d}$.

Given estimates y_i such that $|y_0 - f(u)| \leq \varepsilon/(3d^{1.5})$ and $|y_i - f(u + a_i)| \leq \varepsilon/(3d^{1.5})$, then

$$|y_0(1 - \textstyle\sum_i dz_i) + \sum_i dz_i y_i - f(x)| \leq |1 - \sum_i dz_i| \cdot |y_0 - f(u)| + \sum_i d|z_i| \cdot |y_i - f(u + a_i)|$$
$$\leq \left(1 + d \sum_i |z_i|\right) \frac{\varepsilon}{3d^{1.5}} \leq \varepsilon.$$

■

Hence by querying u and $u + a_i$ with precision $\varepsilon/(3d^{1.5})$ we can obtain precise enough estimates that allow us to reconstruct f for every point of X (and thus prove Theorem 1).

For this to work it is important that X is convex. If X is not convex, we can still apply the previous lemma to $\mathsf{Conv}(X)$ but the points u and $u + a_i$ might not be in X – however, by Caratheodory's theorem we can always write them as a convex combination of $d + 1$ points in X. The following lemma uses this fact to give a simple procedure with total cost $O(d^2 c(\varepsilon/(3d^{1.5}))$.

Lemma 6 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a linear function and let $x_1, x_2, \ldots, x_n$ be n points in $\mathbb{R}^d$. Given approximations $y_1, y_2, \ldots, y_n$ to $f(x_i)$ which satisfy $|f(x_i) - y_i| \leq \varepsilon$, any point $x \in \mathsf{Conv}(\{x_1, x_2, \ldots, x_n\})$ can be approximated within ε error.*

Proof If $x = \sum_{i=1}^n \lambda_i x_i$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$ then $|\sum_i \lambda_i y_i - f(x)| \leq \varepsilon$ since

$$|\textstyle\sum_i \lambda_i y_i - f(x)| = |\sum_i \lambda_i y_i - \sum_i \lambda_i f(x_i)| \leq \sum_i \lambda_i |y_i - f(x_i)| \leq \varepsilon$$

■

If the points u and $u + a_i$ are not in X we can write each of them as a convex combination of $d + 1$ points in X using Caratheodory's Theorem since those points are in the convex hull. If we obtain a $\varepsilon/(3d^{1.5})$ approximation for each of those points, we can obtain the same approximation for u and $u + a_i$ using Lemma 6 and then apply 5 to reconstruct f. The problem with this approach is that this requires us to approximate f on $O(d^2)$ points, leading to total cost of $O(d^2 c(\varepsilon/d^{1.5}))$. For $c(\varepsilon) = \log 1/\varepsilon$, this is polynomially worse in d than the $dc(\varepsilon/d^{1.5})$ bound of Theorem 5 (let alone the $dc(\varepsilon/d)$ total cost if the unit vectors are present). In the next section, we will see how to achieve a $O(d \cdot c(\varepsilon/d))$ bound for general non-convex sets (thus achieving a nearly optimal bound for threshold queries).

3. Non-convex case

We now recover a $O(d \cdot c(\varepsilon/d))$ bound for non-convex sets using a different technique based on developing a John's theorem for simplices.

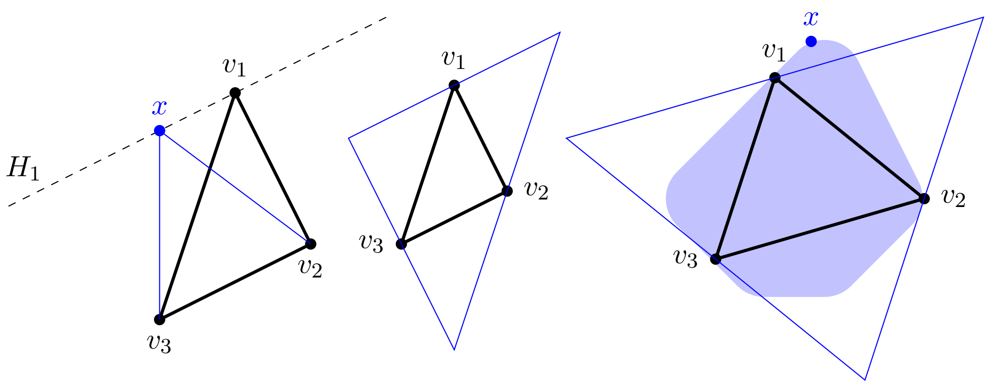

Figure 1: Outer dual construction

Lemma 7 *Given a simplex T with vertices in X and a point $u \in T$ that satisfies $S \subseteq u + \rho(T - u)$, for some $|\rho| \geq 1$ (i.e., X is contained in some possibly reflected dilation of T by a factor of ρ) it is possible to solve the reconstruction problem for X with $O(d \cdot c(2|\rho|/\varepsilon))$ cost.*

Proof Let $v_1, v_2, \ldots, v_{d+1}$ be the vertices of T. By assumption $v_i \in X$ for all i, so we can query those points with accuracy precision $\varepsilon/(2\rho)$; in total this requires $(d+1) \cdot c(2\rho/\varepsilon) = O(d \log(2\rho/\varepsilon))$ queries. Since $u \in T$, note that we can write u as a convex combination of the vectors v_i, and therefore calculate $f(u)$ to within precision $\varepsilon/(2\rho)$ (by Lemma 6).

Let $T' = u + \rho(T - u)$; by assumption, $X \subseteq T'$. We will show we can calculate the value of $f(x)$ for any $x \in T'$ to within precision ε. To do this, it suffices (by Lemma 6) to compute the vertices of T' to within precision ε. The vertices of T' are given by $v'_i = \rho v_i + (1 - \rho)u$. Since $|\rho| + |1 - \rho| \leq 2\rho$, from our estimates of $f(v_i)$ and $f(u)$ we can therefore estimate $f(v'_i)$ to within $2\rho \cdot \varepsilon/(2\rho) = \varepsilon$, as desired. ■

We now show that if we take the maximum volume simplex (MVS) with vertices in X, then this simplex satisfies the conditions of Lemma 7 with $\rho = d$. By Lemma 7, this implies an algorithm for reconstruction that uses $O(d \log(d/\varepsilon)) = O(d \log(d/\varepsilon))$ queries.

To do this, we introduce the concept of the *outer dual* of a simplex T.

Definition 8 *Given $d + 1$ points, $v_1, v_2, \ldots, v_{d+1} \in \mathbb{R}^d$, define the* signed volume

$$\mathsf{Vol}_\Delta(v_1, v_2, \ldots, v_{d+1}) = \frac{1}{d!} \det \begin{bmatrix} v_1 - v_2 \\ v_1 - v_3 \\ \vdots \\ v_1 - v_{d+1} \end{bmatrix}.$$

Note that up to sign, $\mathsf{Vol}_\Delta(v_1, v_2, \ldots, v_{d+1})$ equals the volume of the simplex T with vertices v_i.

Note that the set of points x such that $\mathsf{Vol}_\Delta(x, v_2, \ldots, v_{d+1}) = \mathsf{Vol}_\Delta(v_1, v_2, \ldots, v_{d+1})$ forms a hyperplane in $\mathbb{R}^d$ (in fact, the hyperplane passing through v_1 parallel to the facet spanned by $v_2, v_3, \ldots, v_{d+1}$). See Figure 1a. Taking the intersection of all such hyperplanes, we get a simplex containing T which we call the *outer dual* of T (Figure 1b).

Definition 9 (Outer dual) *Given a simplex T in $\mathbb{R}^d$ with vertices $v_1, v_2, \ldots, v_{d+1}$, define the* outer dual $D(T)$ *of T to be the simplex formed by the intersection of the $d+1$ halfspaces*

$$H_i = \left\{ x \in \mathbb{R}^d \text{ s.t. } \frac{\mathsf{Vol}_\Delta(x, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_{d+1})}{\mathsf{Vol}_\Delta(v_i, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_{d+1})} \leq 1 \right\}.$$

Lemma 10 *Let T be the maximum volume simplex with vertices in X. Then X is contained in a translate of $-dT$.*

Proof We first claim that X is a subset of $D(T)$. To see this, assume there is an $x \in X$ that does not belong to $D(T)$. Then x violates one of the constraints H_i (Figure 1c). But then (by the definition of H_i), the simplex with vertices $\{x, v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{d+1}\}$ has volume strictly greater than $\mathsf{Vol}(T)$, contradicting that T is the maximum volume simplex.

We next claim that $D(T)$ is a translate of $-dT$. To show this, note that both the definition of $D(T)$ and the statement of this claim are invariant under affine transformations (in particular, ratios of signed volumes are invariant under affine transformations). It thus suffices to prove this for any simplex T in $\mathbb{R}^d$. We will choose the simplex with vertices $v_i = e_i$ for $1 \leq i \leq d$ (where e_i is the ith unit vector) and $v_{d+1} = 0$. For this T, $H_i = \{x \mid x_i \leq 1\}$ for $1 \leq i \leq d$, and $H_{d+1} = \{x \mid \sum x_i \geq 0\}$. Let $\mathbf{1} = \sum_i e_i$ be the all-ones vector; then these $d+1$ hyperplanes form a simplex $D(T)$ with vertices $\mathbf{1}$ and $\mathbf{1} - de_i$ for each $1 \leq i \leq d$, confirming that $D(T)$ is a translate of $-dT$.

The lemma follows from the above two claims. ■

Unfortunately, finding the maximum volume simplex with vertices in a set X (even for a finite point set X) is known to be computationally hard (Summa et al., 2014). To this end, we prove a variant of Lemma 10 which holds for simplices which are "close" to being maximal in volume. Here the precise definition of "close" is important. The most natural choice might be the definition of an α-*maximum-volume* simplex, whose volume is within a factor α of the maximum volume simplex, but unfortunately the best known efficient algorithms can only find such a simplex where α is exponential in d (Khachiyan, 1995; Summa et al., 2014; Nikolov, 2015). Instead we will work with a local version of maximal volume, where we will say a simplex is α-optimal if it is impossible to improve its volume by a factor of more than α by modifying a single vertex.

Definition 11 (α-optimal) *A simplex T' is X-adjacent to a simplex T if it is possible to obtain T' from T by replacing a single vertex v in T with a vector x in X. A simplex T with vertices in X is α-optimal if for any X-adjacent simplex T', $\mathsf{Vol}(T') \leq \alpha \mathsf{Vol}(T)$.*

We can similarly define a relaxation of the outer dual of T.

Definition 12 (α-outer dual) *Given a simplex T in $\mathbb{R}^d$ with vertices $v_1, v_2, \ldots, v_{d+1}$, define the* α-outer dual $D_\alpha(T)$ *of T to be the simplex formed by the intersection of the $d+1$ halfspaces*

$$H_i^\alpha = \left\{ x \in \mathbb{R}^d \text{ s.t. } \frac{\mathsf{Vol}_\Delta(x, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_{d+1})}{\mathsf{Vol}_\Delta(v_i, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_{d+1})} \leq \alpha \right\}.$$

Given an α-optimal simplex, we can repeat the proof of Lemma 10 to show that X is is contained in a translate of $-\alpha(d+1)T$.

Lemma 13 *Let T be an α-optimal simplex in X. Then X is contained in a translate of $-\alpha(d+1)T$.*

Proof We similarly prove two claims: that X is a subset of $D_\alpha(T)$, and that $D_\alpha(T)$ is a translate of $-((d+1)\alpha - 1)T$.

We begin by showing that X is a subset of $D_\alpha(T)$. Again, this follows from the definition of D_α: if there is an $x \in X$ that does not belong to $D_\alpha(T)$, then x violates one of the constraints H_i^α, and the simplex with vertices $\{x, v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{d+1}\}$ has volume strictly greater than $\alpha \mathsf{Vol}(T)$, contradicting that T is α-optimal.

Similarly, since the definition of $D_\alpha(T)$ is still invariant under affine transformations, to show that $D_\alpha(T)$ is a translate of $-((d+1)\alpha-1)T$ it suffices to show it for a specific T. We again choose the simplex with vertices $v_i = e_i$ for $1 \leq i \leq d$ (where e_i is the ith unit vector) and $v_{d+1} = 0$. For this choice of T, $H_i^\alpha = \{x \,|\, x_i \leq \alpha\}$ for $1 \leq i \leq d$, and $H_{d+1}^\alpha = \{x \,|\, \sum x_i \geq -(\alpha - 1)\}$. These $d+1$ hyperplanes form a simplex $D_\alpha(T)$ with vertices $\alpha\mathbf{1}$ and $\alpha\mathbf{1} - ((d+1)\alpha - 1)e_i$ for each $1 \leq i \leq d$, confirming that $D_\alpha(T)$ is a translate of $-((d+1)\alpha-1)T$. It follows that X is contained within a translate of $-(d+1)\alpha T$, as desired. ■

It now suffices to find any α-optimal simplex where $\alpha = \mathsf{poly}(d)$. An α-maximum-volume simplex (i.e. a simplex whose volume is within a factor α of the maximum volume simplex in X) is α-optimal, but as mentioned earlier, the best known algorithms can only find α-maximum-volume simplices with α exponential in d. We will demonstrate an efficient procedure that starts with any α-maximum-volume simplex and transforms it into a 2-optimal simplex in time polynomial in $\log \alpha$.

Lemma 14 *There is an efficient algorithm which, given an α-maximum-volume simplex T in X, returns a 2-optimal simplex T' in X, and runs in time polynomial in $\log \alpha$.*

Proof To begin, note that given any d points $v_2, \ldots, v_{d+1}$, we can efficiently find the point $x \in X$ which maximizes $\mathsf{Vol}_\Delta(x, v_2, \ldots, v_{d+1})$: this x is exactly the x in X which maximizes $\langle x, v^\perp \rangle$ where $v^\perp$ is a vector perpendicular to the hyperplane spanned by $v_2, \ldots, v_{d+1}$. We can find this x with one call to our optimization oracle for X.

Now, by running this for each vertex of a simplex T, this gives us an efficient procedure that takes a simplex T and either: a) certifies that T is 2-optimal or b) returns a new simplex T' whose volume is at least twice the volume of T. If we start with an α-maximum-volume simplex T, we can iterate this procedure at most $\log \alpha$ times before it must result in a 2-optimal simplex T'. ■

By Lemma 14, it suffices to start with any α-maximum-volume simplex in X where $\alpha = \exp(\mathsf{poly}(d))$. To find such a simplex, we can certainly use any of the known algorithms (Khachiyan, 1995; Summa et al., 2014; Nikolov, 2015) for approximating the maximum volume simplex. However, these algorithms generally assume X is provided as the convex hull of a finite set of points, and it is not necessarily straightforward to adapt these methods to general sets X with an optimization oracle. Instead, we will construct an α-maximum-volume simplex as follows.

Lemma 15 *Given a set X and an optimization oracle for X, one can efficiently find a d^{3d}-maximum-volume simplex T in X.*

Proof We will show we can efficiently compute the maximum volume simplex T contained in the John ellipsoid E of $\mathsf{Conv}(X)$. By Lemma 10, E is contained within a translate of d^2T. Since (by the definition of John ellipsoid) X is contained within a translate of dE, this means that X is contained within a translate of d^3T, and therefore T is at least a (d^{3d})-maximum-volume simplex in X.

To find T, let A be the linear map which maps the unit ball B to E (this is equivalent to finding the principal axes of E, so we can efficiently find A). Let T_B be the maximum volume simplex contained in the unit ball (this is just an appropriately scaled regular simplex). Then, since ratios of volumes are preserved under linear transformations, AT_B is the maximum volume simplex contained in E. ■

Combining such results we obtain the following John-like theorem:

Theorem 16 (Restatement of Theorem 2) *Given a (not necessarily convex) set X there is a polynomial time algorithm that selects a simplex T with vertices in X such that X is contained in a translate of $-4dT$.*

Proof By Lemma 15, we can find a d^{3d}-maximum-volume simplex T in X. By Lemma 14, we can then find a 2-optimal simplex T' in X. By Lemma 13, X is contained within a translate of $-4dT$. ■

Corollary 17 *There is a procedure with cost $O(d \cdot c(\varepsilon/d))$ to reconstruct a linear function using a costly zero order oracle.*

Proof This follows directly from Theorem 16 together with Lemma 7. ■

4. Lower bounds

The cost of the procedure given in Lemma 5 is nearly optimal for threshold queries with cost $c(\varepsilon) = O(\log(1/\varepsilon))$. For these queries, the procedure has cost $O(d \log(d/\varepsilon))$ while no algorithm can obtain better than $\Omega(d \log(1/\varepsilon))$ by a simple information theoretic argument sketched in the following lemma.

Lemma 18 *Any reconstruction procedure must incur cost at least $\Omega(d \log(1/\varepsilon))$ for threshold queries, i.e., for $c(\varepsilon) = \log(1/\varepsilon)$.*

Proof Let X be the ball of radius 1. We will construct $k = (1/\varepsilon)^{\Omega(d)}$ linear functions $f_i(x) = v_i^\top x$ with $f_i(x) = v_i^\top x$ such that $\max_{x \in X} |f_i(x) - f_j(x)| > \varepsilon$. To do that observe that we can choose $k = (1/\varepsilon)^{\Omega(d)}$ points $v_1, \dots, v_k$ with $\|v_i\| = 1$ such that $\|v_i - v_j\|_2 > \varepsilon$ for any $i \neq j$ (using standard ε-net arguments).

Note that $\max_{x \in X} |f_i(x) - f_j(x)| = \|v_i - v_j\|_2 > \varepsilon$. For this reason, the reconstruction algorithm must be able to distinguish between all functions f_i. Since the cost of the query for $c(\varepsilon) = \log(1/\varepsilon)$ corresponds to the number of bits needed to represent the result (in particular, consider the implementation of this oracle which returns the answers to the $\log(1/\varepsilon)$ threshold queries binary search would perform), we must need at least $\log((1/\varepsilon)^{O(d)}) = O(d \log(1/\varepsilon))$ total cost. ■

Against deterministic, non-adaptive algorithms, we can show a similar lower bound of $\Omega(dc(10\varepsilon))$ for any cost function c. (Note that all our algorithms in Sections 2 and 3 are deterministic and non-adaptive).

Lemma 19 *Any deterministic non-adaptive reconstruction procedure must incur cost at least* $\Omega(dc(10\varepsilon))$.

Proof Again, let X be the ball of radius 1 in $\mathbb{R}^d$. Note that any deterministic and non-adaptive algorithm must always make the same set of queries to the oracle; let us write these queries as $(x_1, \varepsilon_1), (x_2, \varepsilon_2), \dots, (x_n, \varepsilon_n)$, where (x_i, ε_i) means that the algorithm queries the oracle for the value of $f(x_i)$ to within error ε_i. Let k be the number of queries i where $\varepsilon_i \leq 10\varepsilon$ (and without loss of generality, let these k queries be (x_1, ε_1) through (x_k, ε_k)). We will prove that we must have $k \geq d$, from which the theorem follows.

Assume $k < d$. We will construct an instance on which this reconstruction procedure fails. Specifically, we will show how to construct two linear functions f and f' and a sequence of responses y_i to the queries such that:

- For each i, $|y_i - f(x_i)| \leq \varepsilon_i$ and $|y_i - f'(x_i)| \leq \varepsilon_i$.
- There exists an $x \in X$ such that $|f(x) - f'(x)| > 2\varepsilon$.

The first condition guarantees the response y_i to query (x_i, ε_i) is valid for both functions f and f'. The second condition guarantees that any correct reconstruction procedure must be able to distinguish between these two functions.

Consider the k points $x_1, x_2, \dots x_k$. Since $k < d$, these points lie on some hyperplane H of dimension $d' < d$. Choose a unit vector u that is perpendicular to H. Consider the two linear functions $f(x) = 2\varepsilon\langle u, x\rangle$ and $f'(x) = -2\varepsilon\langle u, x\rangle$. Note that $|f(u) - f'(u)| = 4\varepsilon$, satisfying the second point above. In addition, note that $|f(x)| \leq 2\varepsilon$ and $|f'(x)| \leq 2\varepsilon$ for all x.

Now consider what happens if we set $y_i = 0$ for all i (i.e. our oracle always responds 0). For $i \leq k$, note that $f(x_i) = f'(x_i) = 0$ (since $x_i \in H$), so $|y_i - f(x_i)| = |y_i - f'(x_i)| = 0$, and the first condition is satisfied. On the other hand, for $i > k$, we know that $\varepsilon_i > 10\varepsilon$, and therefore $|f(x) - y_i| = |f(x)| \leq \varepsilon_i$ (and likewise, $|f'(x) - y_i| \leq \varepsilon_i$), and again the first condition is satisfied. This implies the theorem statement. ■

5. Open Problems

In this paper, we demonstrate an algorithm which solves the reconstruction problem for arbitrary sets with total cost $O(dc(\varepsilon/d))$. This is nearly tight for threshold queries (where $c(\varepsilon) = \log 1/\varepsilon$), but not necessarily for other cost functions. The main question left open is whether it is possible to improve the asymptotic total cost of our algorithms, possibly to the lower bound of $\Omega(dc(\varepsilon))$ in Lemma 19.

One approach for improving our algorithms is to improve the $O(d)$ bound in our analogue of John's theorem (Theorem 2). For simplices it turns out that this bound is asymptotically tight (see Section 5.1 below), but we can ask the same question for convex combinations of larger numbers of points.

Question 1 *Given a convex set K, what are the values of v and s such that it is possible to find a set $V \subseteq K$ with $|V| = v$ such that for $P = \mathsf{Conv}(V)$ we have $P \subseteq K \subseteq x + sP$?*

A feasible (v, s) pair would lead to an algorithm for linear reconstruction with cost $O(v \cdot c(\varepsilon/s))$. Notably, for $v > d + 1$, we can hope for values of s asymptotically less than d. Even for very simple sets X that contain a basis of unit vectors (e.g. the unit ball, or the unit hypercube $[0,1]^d$), it is not clear whether it is possible to achieve a total cost of $O(dc(\varepsilon))$ – the naive approach of "estimate a basis of unit vectors" results in a cost of $O(d \cdot c(\varepsilon/s))$ (for the unit ball, this can be improved to $O(d \cdot c(\varepsilon/\sqrt{d}))$ following the ideas of Lemma 6).

For the hypercube $K = [0,1]^d$, the following lemma says we cannot hope to find such a set with $v = O(d)$ and $s < \sqrt{d}$.

Lemma 20 *In the setting of Question 3 if $K = [0,1]^d$, we have: $s \geq \Omega(d^{1/2} v^{-1/d})$. In particular for $v = O(d)$ points the best scaling is at least $s \geq \Omega(\sqrt{d})$.*

Proof Hadamard's maximum determinant problem (Hadamard, 1893) is the problem of finding d points in $[0,1]^d$ maximizing the determinant of the matrix obtained by taking such points as rows. Hadamard showed that any such determinant is upper-bounded in magnitude by $d^{d/2}/2^d$, with this bound being achieved for dimensions that are powers of 2. This implies in particular that the volume of any simplex in $[0,1]^d$ is at most $\frac{1}{d!} \cdot \frac{d^{d/2}}{2^d} = \exp(-\frac{1}{2} d \log(d) + O(d))$.

If $V \subseteq [0,1]^d$ and $|V| = v$, we can decompose $\mathsf{Conv}(V)$ into most $v - d + 1$ simplices. Hence $\mathsf{Vol}(\mathsf{Conv}(V)) \leq v \cdot \exp(-\frac{1}{2} d \log(d) + O(d))$. Since $s\mathsf{Conv}(V)$ contains $[0,1]^d$ we have:

$$s^d \cdot v \cdot \exp\left(-\frac{1}{2} d \log(d) + O(d)\right) \geq 1.$$

This implies the theorem statement. ■

Question 2 *Does there exist an algorithm for the hypercube $K = [0,1]^d$ with total cost $d \cdot c(\varepsilon/\sqrt{d})$? Does there exist a matching lower bound?*

Finally, all of the algorithms presented in Sections 2 and 3 are deterministic and non-adaptive in the sense that they identify all points to be queried and precision of each query at the beginning of the procedure. Our lower bound in Lemma 19 of $O(d \cdot c(\varepsilon))$ applies only to deterministic and non-adaptive algorithms. This naturally raises the question of whether it is possible to harness the additional power that comes with adaptivity and randomization.

Question 3 *Is there a gap between adaptive and non-adaptive algorithms for the problem of linear reconstruction using a costly zero-order oracle? In particular, when the domain is either the ball or the cube, can we get $O(d \cdot c(\varepsilon))$ total cost? Is it possible to extend the lower bound in Lemma 19 to adaptive / randomized algorithms?*

Interestingly, for adaptive algorithms in the slightly more powerful model of arbitrary threshold queries, we can perform linear reconstruction with $O(d \cdot \log(1/\varepsilon))$ total cost (whereas the best we know how to do in the regular setting is $O(d \cdot \log(d/\varepsilon))$. In the threshold queries model we can ask any query of the form "is $f(x) < p$?" at cost 1 (for $x \in X$ and $p \in [0,1]$). Note that while a zero-order oracle with cost $c(\varepsilon) = \log(1/\varepsilon)$ can be simulated with threshold queries at the same cost, it is unclear if the opposite is possible.

Theorem 21 *Let $X = \{x \in \mathbb{R}^d; \|x\|_2 \leq 1\}$ and let $f(x) = \langle v, x \rangle$ for some $v \in \mathbb{R}^d$ with $\|v\|_2 \leq 1$. It is possible to obtain a vector $\hat{v}$ such that $\|v - \hat{v}\|_2 \leq \varepsilon$ with $O(d \log(1/\varepsilon))$ threshold queries to function f.*

Proof Consider the following algorithm: initialize $A = [d]$ as the set of active coordinates and let $t = 1$. While A is non-empty, we query

$$f(e_i) \overset{?}{\leq} \frac{2^t}{\sqrt{d}}, \forall i \in A$$

where e_i are the coordinate vectors. For the vectors for which the query returned true, we remove i from A and use binary search starting from the range $\left[0, 2^t/\sqrt{d}\right]$ to find the value of $f(e_i)$ within $\varepsilon/\sqrt{d}$ precision.

After the algorithm finishes, we know have an estimate $\hat{v}_i$ for each coordinate such that $|\hat{v}_i - \hat{v}| \leq \varepsilon/\sqrt{d}$ and hence $\|v - \hat{v}\|_2 \leq \varepsilon$. We only need to bound the total number of threshold queries used.

For each t there can be only $d/4^t$ coordinates with $w_i > 2^t/\sqrt{d}$ hence the set A has size at most $d/4^{t-1}$ at each t. For each coordinate the cost of estimating it is the log of the range divided by the precision, which leads to a total cost of:

$$\sum_{t=1}^{\frac{1}{2}\log_2(d)} \frac{d}{4^t} \log\left(\frac{2^t/\sqrt{d}}{\varepsilon/\sqrt{d}}\right) \leq \sum_{t=1}^{\infty} \frac{d}{4^t} \left[t \log(2) + \log\left(\frac{1}{\varepsilon}\right)\right] = O\left(d \log\left(\frac{1}{\varepsilon}\right)\right)$$

■

Question 4 *Is there a gap between the minimum cost achievable in the threshold query model and the minimum cost achievable in the noisy zeroth order oracle model with cost $c(\varepsilon) = \log(1/\varepsilon)$? Is it possible to achieve total cost $O(d \log(1/\varepsilon))$ for the ball in the oracle model? Is it possible to achieve total cost $O(d \log(1/\varepsilon))$ for arbitrary sets X in the threshold query model?*

5.1. John's theorem for simplices with positive dilation

Theorem 2 states that in any set of S points in $\mathbb{R}^d$, it is possible to find $d+1$ points forming a simplex T such that S is contained in a translate $-O(d)T$. Note that this requires a negative dilation factor. Although it does not immediately help design better algorithms for the reconstruction problem, it is natural to ask what is possible if we restrict ourselves to positive dilations only. Immediately, from the fact that $-T$ is contained within a translate of dT for any simplex T, Theorem 2 implies it is possible to find a simplex T such that S is contained in a translate of $O(d^2)T$. Interestingly, it is unclear if it is possible to do better.

Question 5 *Can we get a John's theorem for simplices with positive dilation factor $O(d)$? In other words, for every $X \subseteq \mathbb{R}^d$, does there exist a simplex T with vertices in X such that X is contained in some translate of $O(d)T$?*

The only lower bound we are aware of is d, achieved by a ball in $\mathbb{R}^d$ (this also serves as a lower bound for negative dilations).

Lemma 22 *Let B_d be a ball in $\mathbb{R}^d$. Then it is possible to find a simplex T_d such that $T_d \subseteq B_d \subseteq dT_d$. Moreover, d is the optimal factor for the ball.*

The lemma follows from a theorem by Fejes-Toth (1964) (see also Klamkin and Tsintsifas (1979)). For any given simplex T, it relates the radius of the maximum inscribed ball B_i (i.e. the ball inside T of maximum radius) and B_o the minimum circumscribed ball (i.e the ball containing T of minimum volume). The observation is that:

$$\frac{\text{radius}(B_o)}{\text{radius}(B_i)} \geq d$$

and the ratio is tight for the regular simplex, which can be defined as follows. For $d = 1$ define $V_1 = \{-1, 1\} \subseteq \mathbb{R}$ and recursively:

$$V_d = \{(1,0)\} \cup \left\{ \left(-\frac{1}{d}, x \cdot \sqrt{1 - \frac{1}{d^2}} \right); x \in V_{d-1} \right\}$$

The regular simplex is $\mathsf{Conv}(V_d)$. Using this, we can easily prove Lemma 22:

Proof The theorem by Fejes-Toth states that for the regular simplex T, there is a ball B such that $\frac{1}{d}B \subseteq T \subseteq B$ and hence $B \subseteq dT$. It also implies that if T is any simplex T in the ball, then if for scaling factor s we had $T \subseteq B \subseteq x + sT$ then we would have: $\frac{1}{s}(B - x) \subseteq T \subseteq B$ and hence $s \geq d$. ■

Even for $d = 2$, we do not know if this lower bound is tight.

Question 6 *For $d = 2$ is the circle the worst-case for John's theorem for simplices? In other words, given any set $X \in \mathbb{R}^2$, can you always find a triangle T with vertices in X so that X is contained in some translate of $2T$?*

Acknowledgments

We would like to thank Allen Liu for reading an earlier version of this manuscript and pointing out to us that our approach in Section 3 results in an algorithm with total cost $O(dc(\varepsilon/d))$ (instead of $O(dc(\varepsilon/d^2))$, as we had originally written).

References

Maxime C Cohen, Ilan Lobel, and Renato Paes Leme. Feature-based dynamic pricing. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 817–817. ACM, 2016.

L. Fejes-Toth. Regular figures. *Proceedings of the Edinburgh Mathematical Society (Pergamon Press)*, 14(2):174–175, 1964.

Jacques Hadamard. Resolution d'une question relative aux determinants. *Bull. des sciences math.*, 2: 240–246, 1893.

Adel Javanmard. Perishability of data: dynamic pricing under varying-coefficient models. *The Journal of Machine Learning Research*, 18(1):1714–1744, 2017.

Adel Javanmard and Hamid Nazerzadeh. Dynamic pricing in high-dimensions. *Working paper, University of Southern California*, 2016.

Leonid Khachiyan. On the complexity of approximating extremal determinants in matrices. *Journal of Complexity*, 11(1):138–153, 1995.

Murray S Klamkin and George A Tsintsifas. The circumradius-inradius inequality for a simplex. *Mathematics Magazine*, 52(1):20–22, 1979.

Jieming Mao, Renato Paes Leme, and Jon Schneider. Contextual pricing for lipschitz buyers. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 5648–5656, 2018.

Aleksandar Nikolov. Randomized rounding for the largest simplex problem. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 861–870. ACM, 2015.

Renato Paes Leme and Jon Schneider. Contextual search via intrinsic volumes. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 268–282, 2018.

Virag Shah, Ramesh Johari, and Jose Blanchet. Semi-parametric dynamic contextual pricing. In *Advances in Neural Information Processing Systems*, pages 2360–2370, 2019.

Marco Di Summa, Friedrich Eisenbrand, Yuri Faenza, and Carsten Moldenhauer. On largest volume simplices and sub-determinants. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 315–323. SIAM, 2014.