On the Hardness and Smoothed Complexity of Quasi-Concave Minimization

Jonathan A. Kelner
MIT Department of Mathematics and CSAIL
Cambridge, MA 02139
kelner@mit.edu

Evdokia Nikolova *
MIT CSAIL
Cambridge, MA 02139
nikolova@mit.edu

Abstract

In this paper, we resolve the smoothed and approximative complexity of low-rank quasi-concave minimization, providing both upper and lower bounds. As an upper bound, we provide the first smoothed analysis of quasi-concave minimization. The analysis is based on a smoothed bound for the number of extreme points of the projection of the feasible polytope onto a k-dimensional subspace, where k is the rank (informally, the dimension of nonconvexity) of the quasi-concave function. Our smoothed bound is polynomial in the original dimension of the problem n and the perturbation size ρ, and it is exponential in the rank of the function k. From this, we obtain the first randomized fully polynomial-time approximation scheme for low-rank quasi-concave minimization under broad conditions. In contrast with this, we prove $\log n$-hardness of approximation for general quasi-concave minimization. This shows that our smoothed bound is essentially tight, in that no polynomial smoothed bound is possible for quasi-concave functions of general rank k.

The tools that we introduce for the smoothed analysis may be of independent interest. All previous smoothed analyses of polytopes analyzed projections onto two-dimensional subspaces and studied them using trigonometry to examine the angles between vectors and 2-planes in $\mathbb{R}^n$. In this paper, we provide what is, to our knowledge, the first smoothed analysis of the projection of polytopes onto higher-dimensional subspaces. To do this, we replace the trigonometry with tools from random matrix theory and differential geometry on the Grassmannian. Our hardness reduction is based on entirely different proofs that may also be of independent interest: we show that the stochastic 2-stage minimum spanning tree problem has a supermodular objective and that supermodular minimization is hard to approximate.

*Supported in part by American Foundation for Bulgaria Fellowship.

1. Introduction

Concave minimization is a fundamental and extensively studied problem in optimization. Its applications range from solving directly practical problems, such as resource allocation and network design, to providing solution techniques for broad classes of optimization problems, including multiplicative programming, d.c. programming, and others [10]. However, given that the minimization of even quadratic concave functions over the unit hypercube is NP-hard [10], the main research focus has been to suggest efficient heuristics or to solve special cases of the problem. To this end, researchers have directed their attention to low-rank concave functions [13, 19, 14]. We shall define these more precisely in the next section, but, informally, these are functions defined on $\mathbb{R}^n$ whose nonconvexity lies within a low-dimensional subspace. However, even for this more restricted class of problems, no efficient approximation algorithms exist to date, and unanalyzed heuristics are required even for low-rank concave quadratic functions [13]. Furthermore, nothing is known about whether this problem can be approximated or solved in a smoothed or average-case setting.

In this paper, we resolve the smoothed and approximative complexity of low-rank quasi-concave minimization, providing both upper and lower bounds. A quasi-concave function attains its minimum at an extreme point of the feasible set [2]; thus an exact algorithm for finding it is to evaluate the function at all extreme points, which can be found via an enumeration method. Since polytopes typically have an exponential number of vertices, algorithms based on enumeration invariably have a worst-case exponential running time; on the other hand the heuristics used so far (*e.g.,* cut-

ting plane or tabu-search methods [19, 13]) remain with both unknown approximation guarantees as well as running time. In the case of a low-rank quasi-concave function, the set of potential minima is reduced further to the extreme points on the projection (shadow) of the feasible set onto the low-rank subspace corresponding to the function. In the worst case, however, the projection can still have exponentially many vertices, even when the polytope is integral and the subspace is two-dimensional [8, 1].

Our main result is a bound on the expected number of extreme points on the shadow of an integral polytope that is polynomial in the original dimension n and perturbation size ρ and exponential in the nonconvexity rank k. Thus, when the rank is constant, our result implies a smoothed polynomial-time algorithm for finding the minimum of a quasi-concave function. Using this, we also construct a fully polynomial-time approximation scheme under broad conditions. Our smoothed bound is stronger than previous smoothed analyses in that it perturbs only the objective function (which, in this case, amounts to perturbing the subspace onto which we're projecting) and not the feasible polytope. This is crucial for applications in combinatorial optimization [17] where the vertices of the polytope correspond to (integer) combinatorial objects. Further, we show that the smoothed bound is essentially tight by exhibiting $(\log n)$-hardness of approximation for general concave minimization.

Ours is the first smoothed analysis of the projection of a polytope onto a space of dimension greater than two. In the two-dimensional case, the smoothed analyses proceeded by using trigonometry to explicitly analyze the intersections and projections of vectors and 2-planes in $\mathbb{R}^n$. In the higher-dimensional case, this does not suffice, and the explicit calculations become intractable. Instead, one needs to apply tools from random matrix theory and differential geometry of the Grassmannian. We believe that the introduction of these tools is of independent interest and will aid in the further study of high-dimensional projections of polytopes.

Quasi-concave minimization is equivalent to quasi-convex maximization from a continuous optimization standpoint—in each case the optima are extreme points of the feasible set and consequently our positive smoothed results hold in both. Curiously, this symmetry is broken on the hardness of approximation side. In a corresponding discrete setting, supermodular maximization seems to be much harder ($\log n$-hard) to approximate than submodular minimization, which has a constant-factor approximation [6]. In particular, the hardness of approximation implies that no smoothed polynomial bound is possible.

We defer a detailed discussion of our results to the next section.

Related Work The literature on concave minimization is extensive; we refer the reader to the *Handbook of Global Optimization* [10] for a comprehensive list of references. As with general concave minimization, the research on low-rank concave minimization has been restricted primarily to heuristic algorithms with unknown worst-case approximation guarantees though with seemingly better practical performance [19, 13, 14].

The framework of smoothed analysis was started by Spielman and Teng [20, 21] in order to explain why some algorithms with high worst-case running times perform well in practice. In this sense, our smoothed analysis is a natural follow up on the experimental studies that have demonstrated a good practical performance for low-rank concave minimization problems [19, 13]. The smoothed framework has since been applied to a wide array of problems.

Analyzing the smoothed complexity of general low-rank quasi-concave minimization in this paper was prompted by the smoothed analysis for a special quasi-concave minimization problem in the context of stochastic shortest paths [17]. We refer the reader to the latter as both a motivating example and application of our general result. Somewhat unexpectedly the smoothed analysis techniques for solving the nonconvex problems in this paper are inspired by the recent techniques of Kelner and Spielman [12] for linear programming. We note that the latter has polynomial smoothed bounds *and* polynomial exact algorithms at the same time, while the low-rank nonconvex setting is vastly more general and does not have any known efficient algorithms.

2. Background and theorem statements

2.1. Concave and Quasi-concave functions

In this section, we define quasi-concave functions and extreme points, and state the problem of quasi-concave minimization and its hardness. Let C be a convex set.

Definition 2.1. *A function $f : C \to (-\infty, \infty)$ is* concave *if for all $x, y \in C$ and $\alpha \in [0, 1]$,*

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y).$$

A function $f : C \to (-\infty, \infty)$ is quasi-concave *if all of*

its upper level sets $L_\gamma = \{x \mid x \in C,\ f(x) \geq \gamma\}$ *are convex.*

Definition 2.2. *A point* x *is an* extreme *point of the set* C *if it cannot be represented as a convex combination of two other points in* C*, namely* $x = \alpha y + (1-\alpha)z$ *for* $y, z \in C$, $\alpha \in (0,1) \quad \Rightarrow \quad x = y = z$.

The problem of (quasi-)concave minimization asks for the minimum of a (quasi-)concave function subject to a set of linear constraints (which determine the *feasible set* for the problem).

Theorem 2.3. *[11, 2] Let* $C \subset \mathbb{R}^n$ *be a compact convex set. A quasi-concave function* $f : C \to \mathbb{R}$ *that attains a minimum over* C*, attains the minimum at some extreme point of* C.

Concave minimization is NP-hard even when restricted to concave quadratic functions over the hypercube [18]. In Section 3, we show that it is $(\log n)$-hard to approximate over the hypercube.

2.2. Low-Rank Quasi-Concave Minimization

In this section, we define the rank of a function and the shadow of the feasible set onto a subspace, and we state some properties of the global minima of low-rank quasi-concave functions.

Definition 2.4. *We say that the function* $f : \mathbb{R}^n \to \mathbb{R}$ *has* rank k *if it can be written in the form* $f(x) = \hat{f}(a_1^T x, a_2^T x, ..., a_k^T x)$ *for some function* $\hat{f} : \mathbb{R}^k \to \mathbb{R}$ *and linearly independent vectors* $a_1, ..., a_k \in \mathbb{R}^n$.

For low-rank functions the rank k is much smaller than n, so that the function can be thought of as living in the lower dimensional subspace spanned by the vectors $a_1, ..., a_k$.[1] We call this subspace the *rank* or *projection subspace* corresponding to the function f. Further, we can project the feasible set in $\mathbb{R}^n$ onto the rank subspace. We will call this projection the *(k-dimensional) shadow*. The following claim is readily verified:

Claim 2.5. *The minimum of* $\hat{f}$ *over the shadow of the polytope* Q *in the rank subspace* $\mathbb{R}^k$ *is the same as the minimum of* f *over* Q *in* $\mathbb{R}^n$*, and it is attained at an extreme point of the shadow.*

If the feasible polytope Q is given as a polynomial set of constraints and the quasi-concave function f is specified by oracle queries, we can find the minimum by enumerating the vertices of the shadow (and keeping track of their corresponding vertices in Q). This can be done for example by a depth-first search exploration of the associated shadow graph, where we can find all neighbors of a vertex in the shadow by checking all neighbors of the corresponding vertex in the original polytope.

We can use similar enumeration approaches for more general models of specification of the objective function and the feasible polytope; this is not the focus of our work so we limit our discussion here.

2.3. Smoothed framework for Low-Rank Quasi-concave problems

In this section, we describe our model for perturbing low-rank functions.

2.3.1 Sampling from the Grassmannian

To properly handle our probability calculations, we shall need to define precisely the spaces and distributions from which we are sampling.

Definition 2.6. *The* Grassmannian $\mathrm{Gr}(n,k)$ *is the set of* k*-dimensional subspaces of* $\mathbb{R}^n$.

Rather than sampling from the Grassmannian directly, we shall draw our samples from the space $O(n)$ of orthogonal $n \times n$ matrices and take the subspace spanned by the first k columns.

Definition 2.7. *Let* $\Pi_k : \mathbb{R}^{n^2} \to \mathbb{R}^{nk}$ *be the map that takes an* $n \times n$ *matrix to its first* k *columns, and let* $\overline{\Pi}_k : \mathbb{R}^{n^2} \to \mathrm{Gr}(n,k)$ *be the map that takes a matrix to the subspace spanned by its first* k *columns.*

The space of orthogonal matrices has a standard measure, known as the *Haar measure* [4], that properly encapsulates what we mean by a "uniformly random" orthogonal matrix. It is the measure induced by treating the space of orthogonal matrices as a subset of $\mathbb{R}^{n^2}$, and it is the unique rotation-invariant measure on $O(n)$. The Haar measure and the map $\overline{\Pi}_k$ then induce a measure on $\mathrm{Gr}(n,k)$. When we take integrals over $O(n)$ or $\mathrm{Gr}(n,k)$, it shall be with respect to these measures.

2.3.2 Perturbing a Low-Rank Function

We defined a function $f : \mathbb{R}^n \to \mathbb{R}$ to have rank k if it can be written in the form $f(x) = \hat{f}(a_1^T x, a_2^T x, ..., a_k^T x)$ for some function $\hat{f} : \mathbb{R}^k \to \mathbb{R}$ and linearly independent vectors $a_1, ..., a_k \in \mathbb{R}^n$. Equivalently, we can define a rank-k function as a function $\hat{f} : \mathbb{R}^k \to \mathbb{R}$ and a k-dimensional subspace S.

[1] Our definition of rank may differ by 1 from other definitions that have appeared in the literature [19].

Consequently, we can describe a perturbation in two equivalent ways, as a perturbation of the subspace or of the function. For notational simplicity, we shall choose the former option, although our results could easily be restated in terms of the latter. We stress again that in previous smoothed analysis both the function and feasible set are perturbed, while here we only perturb one part of the problem and leave the other unchanged.

In order to define a perturbation of a subspace, we shall need a notion of the distance between two matrices, which we take to be the spectral norm of their difference:

$$\mathrm{dist}(Q,T) = \max_{\|x\|=1} (Q-T)x,$$

where $\|\cdot\|$ is the L_2 norm.

As mentioned above, we shall sample from the space of orthogonal matrices and then obtain elements of the Grassmannian $\mathrm{Gr}(n,k)$ by keeping only the first k columns. It will thus suffice for us to define a ρ-perturbation of an orthogonal matrix, as this will induce a similar notion on the Grassmannian, and thus on the space of low-rank quasi-concave functions.

Definition 2.8. *A ρ-perturbation of an orthogonal matrix Q_0 is an orthogonal matrix sampled from the distribution on orthogonal matrices whose density at a matrix Q is proportional to*

$$\exp\{\mathrm{dist}(Q^{-1}Q_0, \mathrm{Id})/\rho\}.$$

2.4. Our results

Our main theorem states that there is an expected polynomial-time smoothed algorithm for low-rank quasi-concave minimization over integral polytopes:

Theorem 2.9. *Let $P \subseteq \mathbb{R}^n$ be a polytope with integer vertices whose coordinates are all less than some constant γ and a polynomial number of facets, and let $f : P \to \mathbb{R}$ be a quasi-concave function of constant rank k given as an oracle. There exists an algorithm that minimizes a ρ-perturbation of f in expected time*

$$\mathrm{poly}(n, \gamma, 1/\rho),$$

where the degree of the polynomial depends on k.

By the enumeration techniques described in Section 2.2, this will follow immediately from our main technical lemma, which is a bound on the expected number of extreme points in the projection of the polytope onto the rank subspace of the objective function; the bound is polynomial in the perturbation size ρ and the original dimension n, but exponential in k.

Lemma 2.10. *Let $P \subseteq \mathbb{R}^n$ be a polytope with integer vertices whose coordinates are all less than some constant γ (and with* no *restriction on the number of facets), and let $Q_0 \subseteq \mathbb{R}^n$ be a k-dimensional vector subspace. The expected number of vertices on the projection of P onto a ρ-perturbation of Q_0 has at most*

$$\mathrm{poly}(n, \gamma, 1/\rho)$$

vertices, where the degree of the polynomial depends on k.

We also show that general concave minimization is $(\log n)$-hard to approximate and consequently that there can be no smoothed expected polynomial-time algorithm for arbitrary quasi-concave functions. The proof is based on a result of independent interest that supermodular minimization is $(\log n)$-hard to approximate.

3. Hardness of Approximation of Concave Minimization

In this section, we prove that minimizing a supermodular function and consequently, minimizing a concave function, is $(\log n)$-hard to approximate. The proof is by a gap-preserving reduction from the two-stage stochastic Minimum Spanning Tree (MST) problem considered by Flaxman, Frieze and Krivelevich [7], which is $(\log n)$-hard to approximate [7].

The two-stage stochastic MST problem is defined as follows. Edges in the graph have given weights in the first stage and random weights from known distributions in the second stage. The goal is to choose an optimal set of edges in the first stage that can be completed with edges in the second stage to form a spanning tree of minimum expected cost. More formally, denote the cost of edge e in the first and second stages by $C_1(e)$ and $C(e)$ respectively. Note that a subset of edges S chosen in the first stage uniquely determines (up to equal cost) the set of edges T_S chosen in the second stage which complete the MST. Our problem is to compute $\min_{S \subset E} \left\{ h(S) = C_1(S) + \mathbf{E}[C(T_S)] \right\}$, where E is the set of edges of the graph. We will show that the function $h(S)$ is supermodular. To simplify notation, we shall use the shorthand $S + e := S \cup \{e\}$ and $S - e := S \backslash \{e\}$. We also use $A \subset B$ to mean non-strict set inclusion.

Lemma 3.1. *The function $h(S) = C_1(S) + \mathbf{E}[C(T_S)]$ is supermodular.*

Proof. The function h is supermodular if and only if the marginal contribution of an edge to the function value is

bigger whenever the edge is added to a bigger set [15], that is

$$h(A+e)-h(A) \leq h(B+e)-h(B)$$

for all $A \subset B \subset E$ and $e \in E$, $e \notin A, B$. Assume without loss of generality that A, B as well as $A+e$, $B+e$ do not contain cycles. We have $C_1(A+e) - C_1(A) = C_1(e) = C_1(B+e) - C_1(B)$. By linearity of expectation, it suffices to show that

$$C(T_{A+e}) - C(T_A) \leq C(T_{B+e}) - C(T_B),$$

or equivalently

$$C(T_B) - C(T_{B+e}) \leq C(T_A) - C(T_{A+e}), \quad (1)$$

for every realization of the second stage costs C. Without loss of generality up to equality of the tree costs we can assume that $T_B \subset T_A$ since $A \subset B$. We now need to consider several cases for the edge e.

(i) $e \in T_A, T_B$. Then both sides of the inequality (1) are zero.

(ii) $e \in T_A, e \notin T_B$. Since $(A+e) \cup (T_A - e) = A \cup T_A$ is a valid spanning tree and $(A+e) \cup T_{A+e}$ is the minimal spanning tree corresponding to choice set $A+e$ from the first stage, we have $C(T_{A+e}) \leq C(T_A - e) = C(T_A) - C(e)$. Similarly, $C(T_B) \leq C(T_{B+e}+e)$ since $B \cup T_B$ is the cheapest spanning tree which contains set B and $B \cup (T_{B+e}+e) = (B+e) \cup T_{B+e}$ is another spanning tree containing set B. Therefore

$$C(T_B) - C(T_{B+e}) \leq C(e) \leq C(T_A) - C(T_{A+e}),$$

and we are done.

(iii) $e \notin T_A, T_B$. Without loss of generality $T_{A+e} \subset T_A$ and $T_{B+e} \subset T_B$. Let $T_{A+e} = T_A - e'$ and $T_{B+e} = T_B - e''$ where e' and e'' are the heaviest edges in the cycles formed in the spanning trees $A \cup T_A$ and $B \cup T_B$ respectively when edge e is added. (Note: this assumes that $A+e$, $B+e$ do not contain cycles so any cycles formed would be in T_A, T_B.) Since $T_B \subset T_A$, the heaviest edge e'' in the cycle in T_B is no heavier than e' in the cycle in T_A. Therefore,

$$\begin{aligned} C(T_B) - C(T_{B+e}) &= C(e'') \leq C(e') \\ &= C(T_A) - C(T_{A+e}). \end{aligned}$$

This completes the proof. □

We emphasize that the proof of Lemma 3.1 does not depend in any way on the distributions of the second stage costs, in particular they can be correlated or independent. The dependency requirement is crucial in the proof that the two-stage MST problem is $(\log n)$-hard to approximate [7]. The lemma thus gives the desired gap-preserving reduction for supermodular minimization from the two-stage MST problem;this may be of independent interest to the study of two-stage stochastic optimization, as it provides a connection between the latter and (non-monotone) supermodular and concave minimization.

Lemma 3.2. *Supermodular minimization is* $(\log n)$*-hard to approximate, assuming* $P \neq NP$.

Next, we show that quasi-concave minimization is also hard to approximate.

Theorem 3.3. *Quasi-concave minimization is* $(\log n)$*-hard to approximate, assuming* $P \neq NP$.

Proof. Given a supermodular set function h defined on the vertices of the unit hypercube in $\mathbb{R}^n$ (which we also call 0-1-vectors), its Lovász extension $f : \mathbb{R}^n_+ \to \mathbb{R}$ is defined as $f(x) = h(\mathbf{0}) + \sum_{i=1}^n \lambda_i h(b_i)$, where $\mathbf{0}$ is the zero vector and $x = \sum_{i=1}^n \lambda_i b_i$ is the unique representation of x via a basis of increasing 0-1-vectors $b_1 < b_2 < ... < b_n$ and $\lambda_i \geq 0$. Lovász showed that h is supermodular if and only if its extension f is concave [15].

We will show that there is a gap-preserving reduction of (quasi-)concave minimization from supermodular minimization.

By Theorem 2.3, the minimum of a supermodular set function and the minimum of its Lovász extension over the unit hypercube will coincide. Therefore, a γ-approximation of the minimum of a supermodular function is also a γ-approximation of the minimum of its corresponding Lovász extension.

Conversely, suppose we can approximate the minimum f_{min} of the Lovász extension f over the unit hypercube within a factor of γ, namely we can find $x \in [0,1]^n$ such that $f(x) \leq \gamma f_{min}$. It follows by the Caratheodory/Krein-Milman's theorem [2] that x is a convex combination of vertices of the hypercube, namely $x = \sum_i \lambda_i x_i$ where x_i are 0-1-vectors and $\sum_i \lambda_i = 1$, $\lambda_i > 0$. Since f is concave, $\gamma f_{min} \geq f(x) \geq \sum_i \lambda_i f(x_i)$, therefore for at least one hypercube vertex x_j, we have $f(x_j) \leq \gamma f_{min}$. Therefore x_j would give a γ-approximation of the supermodular function minimum, as desired. □

4. Smoothed Analysis

In this section, we shall prove our main technical lemma, Lemma 2.10. Our general method of proof is motivated by the techniques used by Kelner and Spielman to construct their polynomial-time simplex method for linear programming. However, while the overall approach of bounding the combinatorial complexity through volumetric arguments is the same, studying the projection onto a subspace of dimension greater than two inheres significant new technical difficulties. Resolving these requires us to analyze how the geometry of the Grassmannian or the orthogonal group changes under the change of coordinates induced by a fairly intricate matrix decomposition.

As it will simplify our notation slightly, we shall prove the lemma for the projection of a polytope onto a $(k+1)$-dimensional subspace instead of onto a k-dimensional one.

At a high level, our proof is quite simple. We shall begin by showing that every face of P has a fairly large volume. We shall then show that this remains true about the projection of P onto a ρ-perturbation of a $(k+1)$-dimensional subspace. As we have assumed that all of P's coordinates are bounded above by some constant γ, the surface area of the projection will be bounded above by some constant depending on k and γ. Since we will have a lower bound on the volume of each facet of the projection, this will imply an upper bound on the total number of facets, and thus on the total number of vertices, as desired.

4.1. The Volume of Faces of P

We begin by providing a lower bound on the volume of any face of P.

Lemma 4.1. *Let $P \subset \mathbb{R}^n$ be a polytope with integer vertices. The (k-dimensional) volume of every k-dimensional face of P is at least $1/k!$.*

Proof. Let F be the face in question, and suppose that F is the convex hull of the vertices $v_1, \ldots, v_s \in \mathbb{R}^n$. Since F is k-dimensional, there exists some set of of $(k+1)$ of these vertices that does not lie in any $(k-1)$-dimensional affine subspace. Without loss of generality, let this set consist of $v_1, \ldots, v_{k+1}$. Since $\mathrm{conv}(v_1, \ldots, v_{k+1}) \subseteq \mathrm{conv}(v_1, \ldots, v_s)$, it suffices to show that $\mathrm{Vol_k}(\mathrm{conv}(v_1, \ldots, v_{k+1})) \geq 1/k!$.

We thus aim to bound the volume of a k-dimensional simplex Σ with integer vertices. Translating the entire simplex if necessary, we can also assume that v_{k+1} is the origin. If V is the $n \times k$ matrix whose i^{th} column comprises the coordinates of v_i, we have that

$$\mathrm{Vol_k}(\Sigma) = \frac{1}{k!}\sqrt{\det(V^T V)}.$$

We have assumed that Σ has nonzero volume, so $\det(V^T V) \neq 0$. Since $V^T V$ is an integer matrix, its determinant is an integer. Its absolute value is thus at least 1, so $\mathrm{Vol}_k(\Sigma) \geq 1/k!$, as desired. □

4.2. The Volume of Faces of the Projection

We now aim to show that the projection of a given k-dimensional face F of an n-dimensional polytope P onto a perturbed $(k+1)$-dimensional subspace has a sufficiently large volume, contingent upon F appearing on the boundary of the projection. Conditioning on F appearing on the boundary of the projection is a rather subtle operation and will require quite a bit of work. In order to do so, we shall require some fairly intricate machinery from random matrix theory.

The proof of this result is rather technical. In order to provide some geometric intuition for why it should be true, we shall begin by considering some simpler variants of the problem. In the first such variant, we shall replace our perturbed subspace by a uniformly random one, and we shall remove the conditioning constraint that F appears on the boundary of the projection. We shall then add our conditioning constraint back in but still consider a uniformly random subspace. After that, we shall progress to our final, most general version.

4.2.1 Projecting onto a Uniform Subspace with no Conditioning

We begin with the following problem: we are given some k-dimensional face F embedded in $\mathbb{R}^n$, and we choose a uniformly random $(k+1)$-dimensional subspace S. The goal is to show that the ratio

$$\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)}$$

is unlikely to be less than some inverse polynomial bound, where π_S denotes orthogonal projection onto S.

In this case, all that matters about F is the k-dimensional affine subspace T that it spans, and we are just trying to provide lower bounds on the determinant of the projection map from T onto S. When k equals 1, this amounts to understanding the length of the projection of a uniformly random unit vector onto some fixed

unit vector, which is a very well-analyzed question. In this case, the expected length grows asymptotically like $1/\sqrt{n}$. In the higher dimensional case at hand, this is a well-studied question in random matrix theory (see [16], for example) whose answer is given by

$$\frac{\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{n-d+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n+1}{2}\right)} \approx C(d)\frac{1}{n^{d/2}}.$$

4.2.2 Projecting onto a Uniform Subspace with Conditioning

When we project our polytope onto a lower-dimensional subspace, not all of the faces appear on the boundary of the projection. In order to analyze the expected volumes of the faces that do appear, we must therefore condition our probability estimates on the event that a given face appears on the boundary of the projection. It is here that we must deviate from previous smoothed analyses and introduce more involved methods from geometry and random matrix theory.

Let $P \subseteq \mathbb{R}^n$ be the set of points that satisfy a collection of m linear constraints,

$$P = \{x \,|\, a_i \cdot x \leq 1,\ i = 1, \ldots, m\}.$$

For a k-dimensional face F of P and a $(k+1)$-dimensional subspace S, let $A_F(S)$ be the event that F appears as a facet of the projection of P onto S. The following theorem provides a geometric criterion for when $A_F(S)$ occurs:

Theorem 4.2. *Let $a_{i_1}, \ldots, a_{i_s}$ be the constraints that are satisfied with equality at all points of F, and let $C = \mathrm{pos}(a_{i_1}, \ldots, a_{i_k})$ be their positive hull. The event $A_F(S)$ occurs if and only if the intersection $C \cap S \neq \emptyset$.*

Proof. This follows from standard convex geometry, contained for example in [3]. □

We note that for a generic polytope, there will be $s = n - k$ constraints that are tight at a k-dimensional face, so the cone C will be $(n-k)$-dimensional. The $(n-k)$-dimensional subspace spanned by C will thus generically intersect the random $(k+1)$-dimensional subspace S in a 1-dimensional ray, and the question is whether this ray lies in C. This will occur with a positive probability that is proportional to the k-dimensional volume of the intersection of C with the unit sphere. In order to condition on $A_F(S)$, we must therefore only consider subspaces S that intersect C. In all that follows, we shall assume that $n > 2k+2$. Since we are assuming k is a constant and studying the asymptotics as n gets large, this inheres no loss of generality.

Our goal is to bound the probability

$$\Pr_{S \in \mathrm{Gr}(n,k+1)}\left[\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)} \leq \epsilon \,|\, A_F(S)\right].$$

By rotating our coordinate system if necessary, we can take F parallel to the subspace $\mathrm{span}(e_1, \ldots, e_k)$, where the e_i are our unit basis vectors. If V is an $n \times (k+1)$ matrix whose columns form an orthonormal basis for S, and $\mathcal{F} = \begin{bmatrix} \mathrm{Id}_k \\ \mathbf{0}_{n-k} \end{bmatrix}$, then

$$\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)} = \sqrt{\det\left((\mathcal{F}^T V)(\mathcal{F}^T V)^T\right)},$$

and therefore

$$\Pr_{S \in \mathrm{Gr}(n,k+1)}\left[\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)} \leq \epsilon \,|\, A_F(S)\right] \quad (2)$$
$$= \frac{\displaystyle\int_{Q \in O(n),\, \overline{\Pi}_{k+1}(Q) \cap C \neq \emptyset} \ \int_{\sqrt{\det((\mathcal{F}^T \Pi_{k+1}(Q))(\mathcal{F}^T \Pi_{k+1}(Q))^T)} \leq \epsilon} 1 \, dQ}{\displaystyle\int_{Q \in O(n),\, \overline{\Pi}_{k+1}(Q) \cap C \neq \emptyset,} 1 \, dQ},$$

where C is defined as in the statement of Theorem 4.2, and dQ denotes the Haar measure on $O(n)$.

Right now, Q is being represented as an $n \times n$ matrix. In this representation, it is rather difficult to describe which matrices meet the other conditions of integration, which makes it difficult to evaluate the integrals. To remedy this, we shall change to a new coordinate system on $O(n)$ that is more conducive to this task. To do so, we shall make use of a matrix decomposition known as the *generalized cosine-sine (CS) decomposition*.

Theorem-Definition 4.3 ([9])**.** *Let*

$$Q = \begin{array}{c} \begin{array}{ccc} \text{sizes} & k & n-k \end{array} \\ \left(\begin{array}{c|c} k+1 & Q_{11} & Q_{12} \\ n-k-1 & Q_{21} & Q_{22} \end{array}\right) \end{array}$$

be an $n \times n$ orthogonal matrix. For any $k \leq n/2$, there exists a unique decomposition:

$$Q = \begin{array}{c} \begin{array}{ccc} \text{sizes} & k+1 & n-k-1 \end{array} \\ \left(\begin{array}{c|c} k+1 & U & 0 \\ n-k-1 & 0 & V \end{array}\right) \end{array}$$
$$\times \begin{array}{c} \begin{array}{ccccc} \text{sizes} & k & 1 & k & n-2k-1 \end{array} \\ \left(\begin{array}{c|ccc} k & C & 0 & -S & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \hline k & S & 0 & C & 0 \\ n-2k-1 & 0 & 0 & 0 & \mathrm{Id}_{n-2k-1} \end{array}\right) \end{array}$$

$$\times \quad \begin{pmatrix} \text{sizes} & k & n-k \\ k & W^T & 0 \\ n-k & 0 & Z^T \end{pmatrix}$$

where

- *U, V, W, and Z are orthogonal matrices,*
- *S and C are positive diagonal matrices,*
- *the diagonal elements $c_1, \ldots, c_p$ of C are nondecreasing, and*
- $C^2 + S^2 = \mathrm{Id}_k$.

We call this decomposition the generalized CS decomposition of Q. *The angles $\theta_1, \ldots, \theta_k$ such that $\cos(\theta_i) = c_i$ are called the* principal angles *between the subspace $\Pi_{k+1}(Q)$ and the subspace $\Pi_k(\mathcal{F})$.*

Let

$$\lambda_i = \cos^2(\theta_i).$$

We shall change coordinates so that our variables comprise $\lambda_1, \ldots, \lambda_k$ along with the orthogonal matrices U, V, W, and Z.

Theorem 4.4. *The Jacobian of the change of variables described above is given by:*[2]

$$dQ \quad = \quad \prod_{i<j}(\lambda_j - \lambda_i) \prod_{i=1}^{k} (1-\lambda_i)^{(n-2k-2)/2} \times d\lambda_1 \ldots d\lambda_p \, dU \, dV \, dW \, dZ,$$

where dU, dV, dW, and dZ are the Haar measures on the appropriate-dimensional spaces of orthogonal matrices.

It is not difficult to see that

$$\det\left(\left(\mathcal{F}^T \Pi_{k+1}(Q)\right) \left(\mathcal{F}^T \Pi_{k+1}(Q)\right)^T \right) = \prod_{i=1}^{k} \lambda_i.$$

This allows us to reexpress the right-hand side of equation (2) as

$$\frac{\displaystyle\int_{\substack{\Pi_1(Z)\in C, \\ \prod_i \lambda_i \le \epsilon^2}} \prod_{i<j}(\lambda_j - \lambda_i) \prod_{i=1}^{k} (1-\lambda_i)^{(n-2k-2)/2} \times d\lambda_1 \ldots d\lambda_p \, dU dV dW dZ}{\displaystyle\int_{\Pi_1(Z)\in C} \prod_{i<j}(\lambda_j - \lambda_i) \prod_{i=1}^{k} \lambda_i (1-\lambda_i)^{(n-2k-2)/2} \times d\lambda_1 \ldots d\lambda_p \, dU dV dW dZ}$$

$$= \frac{\displaystyle\int_{\prod_i \lambda_i \le \epsilon^2} \prod_{i<j}(\lambda_j - \lambda_i) \prod_{i=1}^{k} (1-\lambda_i)^{(n-2k-2)/2} \times d\lambda_1 \ldots d\lambda_p}{\displaystyle\int \prod_{i<j}(\lambda_j - \lambda_i) \prod_{i=1}^{k} (1-\lambda_i)^{(n-2k-2)/2} \times d\lambda_1 \ldots d\lambda_p.}$$

Let $\zeta_i = 1 - \lambda_i$ and write $\vec{\zeta} = (\zeta_1, \ldots, \zeta_k)$ and $\vec{\zeta'} = (\zeta_1, \ldots, \zeta_{k-1})$ for brevity. By expanding the products in the numerator and denominator and collecting terms, we can rewrite the last expression in the string of equations above as

$$\frac{\displaystyle\int_{\prod_i (1-\zeta_i) \le \epsilon^2} \sum_{t \ge 0} \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}}{\displaystyle\int \sum_{t \ge 0} \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}}$$

for some multivariate functions $p_t(\vec{\zeta'})$. Interchanging the summation with the integration yields:

$$\begin{aligned}
&\frac{\sum_{t\ge 0} \int_{\prod_i (1-\zeta_i)\le \epsilon^2} \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}}{\sum_{t\ge 0} \int \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}} \\
&\le \max_{t\ge 0} \frac{\int_{\prod_i (1-\zeta_i)\le \epsilon^2} \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}}{\int \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}} \\
&\le \max_{t\ge 0} \frac{\int_{(1-\zeta_k)\le \epsilon^{2/k}} \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}}{\int \zeta_k^{t+(n-2k-2)/2} p_t(\vec{\zeta'}) d\vec{\zeta}} \\
&= \max_{t\ge 0} \frac{\int_{\zeta_k=1-\epsilon^{2/k}}^{1} \zeta_k^{t+(n-2k-2)/2} d\zeta_k}{\int_{\zeta_k=0}^{1} \zeta_k^{t+(n-2k-2)/2} d\zeta_k} \\
&\le 1 - (1-\epsilon^{2/k})^{(n-2k)/2} \\
&\le \frac{n-2k}{2} \epsilon^{2/k},
\end{aligned}$$

where the maxima are taken over all t for which $p_t \neq 0$.

[2] We do not believe that this calculation has previously occurred in the literature. Dumitriu asserted a value for it in her thesis [5], but the claimed Jacobian stated there was incorrect. Due to space constraints, we defer the proper computation to the full version of this abstract.

Remark 4.5. *For simplicity of exposition, we have provided crude bounds on this integral that suffice to prove Lemma 2.10. We defer more precise bounds to the full version of this abstract.*

4.3. The General Case

We now consider the fully general case described in Lemma 2.10, where we examine the projection of P onto a ρ-perturbation of an arbitrary $(k+1)$-dimensional subspace, and we take $\rho < 1/(n-2k)$. This results in a nonuniform probability density over the space of orthogonal matrices from which we draw Q; let μ be this probability density. We aim to compute

$$\begin{aligned}
&\Pr_{S\in \mathrm{Gr}(n,k+1)}\left[\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)} \le \epsilon \,|\, A_F(S)\right] \\
&\qquad = \frac{\displaystyle\int_{\substack{Q\in O(n),\,\overline{\Pi}_{k+1}(Q)\cap C\neq\emptyset,\\ \sqrt{\det((\mathcal{F}^T\Pi_{k+1}(Q))(\mathcal{F}^T\Pi_{k+1}(Q))^T)}\le\epsilon}} \mu(Q)\,dQ}{\displaystyle\int_{Q\in O(n),\,\overline{\Pi}_{k+1}(Q)\cap C\neq\emptyset} \mu(Q)\,dQ} \\
&\qquad = \frac{\displaystyle\int_{\substack{\Pi_1(Z)\in C,\\ \prod_i \lambda_i\le\epsilon^2}} \prod_{i<j}(\lambda_j-\lambda_i)\prod_{i=1}^{k}(1-\lambda_i)^{(n-2k-2)/2} \times \mu(\lambda_1,\ldots,\lambda_k,U,V,W,Z) \times d\lambda_1\ldots d\lambda_p\,dU\,dV\,dW\,dZ}{\displaystyle\int_{\Pi_1(Z)\in C} \prod_{i<j}(\lambda_j-\lambda_i)\prod_{i=1}^{k}\lambda_i(1-\lambda_i)^{(n-2k-2)/2} \times \mu(\lambda_1,\ldots,\lambda_k,U,V,W,Z) \times d\lambda_1\ldots d\lambda_p\,dU\,dV\,dW\,dZ}.
\end{aligned}$$

We shall fix U, V, W, and Z arbitrarily and bound the resulting ratio of integrals over the λ_i. Once we fix U, V, W, and Z, our density depends only on the λ_i, and our expression becomes (in the notation of the previous section):

$$\begin{aligned}
&\frac{\displaystyle\int_{\prod_i\lambda_i\le\epsilon^2} \prod_{i<j}(\lambda_j-\lambda_i)\prod_{i=1}^{k}(1-\lambda_i)^{(n-2k-2)/2} \times \mu(\lambda_1,\ldots,\lambda_k)\,d\lambda_1\ldots d\lambda_p}{\displaystyle\int \prod_{i<j}(\lambda_j-\lambda_i)\prod_{i=1}^{k}\lambda_i(1-\lambda_i)^{(n-2k-2)/2} \times \mu(\lambda_1,\ldots,\lambda_k)\,d\lambda_1\ldots d\lambda_p} \\
&= \frac{\int_{\prod_i(1-\zeta_i)\le\epsilon^2}\sum_{t\ge0}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}}{\int\sum_{t\ge0}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}} \\
&= \frac{\sum_{t\ge0}\int_{\prod_i(1-\zeta_i)\le\epsilon^2}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}}{\sum_{t\ge0}\int\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}} \\
&\le \max_{t\ge0}\frac{\int_{\prod_i(1-\zeta_i)\le\epsilon^2}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}}{\int\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}} \\
&\le \max_{t\ge0}\frac{\int_{(1-\zeta_k)\le\epsilon^{2/k}}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}}{\int_{(1-\zeta_k)\le\rho}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu(\vec{\zeta})d\vec{\zeta}}.
\end{aligned}$$

Suppose that μ is the density for a ρ-perturbation of some fixed subspace S, given as the span of the first $(k+1)$ columns of an orthogonal matrix Q_0. For fixed values of $\zeta_1,\ldots,\zeta_{k-1}$, U, V, W, and Z, changing ζ_k by ϕ causes $\mathrm{dist}(Q^{-1}Q_0,\mathrm{Id})$ to change by at most ϕ. By our definition of a ρ-perturbation, this implies that

$$\mu(\zeta_1,\ldots,\zeta_{k-1},\zeta_k) \ge \mu(\zeta_1,\ldots,\zeta_{k-1},\zeta_k+\phi)/e$$

for any $\phi<\rho$. Consequently there exists some density $\mu'(\zeta_1,\ldots,\zeta_{k-1})$ dependent only on the first $k-1$ of the ζ_i such that

$$\mu'(\vec{\zeta}')/e \le \mu(\vec{\zeta}) \le \mu'(\vec{\zeta}').$$

This allows us to bound our probability by

$$\begin{aligned}
&\max_{t\ge0}\frac{e\int_{(1-\zeta_k)\le\epsilon^{2/k}}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu'(\vec{\zeta}')d\vec{\zeta}}{\int_{(1-\zeta_k)\le\rho}\zeta_k^{t+(n-2k-2)/2}p_t(\vec{\zeta}')\mu'(\vec{\zeta}')d\vec{\zeta}} \\
&= \max_{t\ge0}\frac{e\int_{(1-\zeta_k)\le\epsilon^{2/k}}\zeta_k^{t+(n-2k-2)/2}d\zeta_k}{\int_{(1-\zeta_k)\le\rho}\zeta_k^{t+(n-2k-2)/2}d\zeta_k} \\
&= e\max_{t\ge0}\frac{\zeta_k^{t+(n-2k)/2}\big|_{\zeta_k=1-\epsilon^{2/k}}^{1}}{\zeta_k^{t+(n-2k)/2}\big|_{\zeta_k=1-\rho}^{1}} \\
&\le e\frac{1-(1-\epsilon^{2/k})^{(n-2k)/2}}{1-(1-\rho/2)^{(n-2k)/2}} \\
&\le O(1)\frac{(n-2k)\epsilon^{2/k}}{(n-2k)\rho}, \quad \text{as } \rho<1/(n-2k) \\
&= O(1)\frac{\epsilon^{2/k}}{\rho},
\end{aligned}$$

where the maxima in all of the above expressions are taken over the set of t for which $p_t\neq0$.

4.4. Putting the Steps Together

In Section 4.1, we showed that every k-dimensional facet of P has volume at least $1/k!$. We then showed in Section 4.3 that

$$\Pr_{S\in\mathrm{Gr}(n,k+1)}\left[\frac{\mathrm{Vol}_k(\pi_S(F))}{\mathrm{Vol}_k(F)}\le\epsilon\,\middle|\,A_F(S)\right]\le O(1)\frac{\epsilon^{2/k}}{\rho}.$$

If Q is the projection of P onto S, combining the two results above shows that the probability that any given k-face of Q has volume less than $\epsilon/k!$, contingent upon this face appearing, is bounded above by $O(1)\epsilon^{2/k}/\rho$. This implies that the expected volume of any such face, contingent upon it appearing, is at least $(1/k!)(\rho/O(1))^{k/2}$.

The polytope Q is contained in a ball of radius $\gamma\sqrt{k}$. Its surface area is therefore bounded above by the surface area of this ball, which equals

$$2\frac{\pi^{(k+1)/2}}{\Gamma((k+1)/2)}(\gamma\sqrt{k})^k.$$

The surface area of Q equals the sum of the areas of its facets, so we obtain:

$$\begin{aligned}
&2\frac{\pi^{(k+1)/2}}{\Gamma((k+1)/2)}(\gamma\sqrt{k})^k \\
&\quad\geq \sum_F \mathbf{E}[\mathrm{Vol}_k(\pi_S(F))|A_F(S)] \cdot \Pr[A_F(S)] \\
&\quad\geq (1/k!)(\rho/O(1))^{k/2} \sum_F \Pr[A_F(S)] \\
&\quad= (1/k!)(\rho/O(1))^{k/2}\, \mathbf{E}[\text{number of facets of } Q].
\end{aligned}$$

Combining this with the fact that each facet of Q contains only polynomially many vertices (since k is a constant) yields Lemma 2.10. □

5. Approximating low-rank concave minimization

Here we show that the polynomial smoothed bounds from Section 4 can be used to obtain a randomized fully-polynomial additive approximation scheme for minimizing low-rank concave functions under certain conditions. Since the objective function may have a zero minimum, the concept of multiplicative approximation is not so meaningful. On the other hand the additive approximation implies an arbitrarily good multiplicative approximation for functions bounded below by a constant or inverse polynomial. We also show that no polynomial smoothed bounds are possible for general concave minimization.

Theorem 5.1. *There is a randomized fully-polynomial time algorithm with additive approximation ϵ for any $\epsilon > 0$, for low-rank quasi-concave minimization over a polynomially-bounded polytope, when the objective function satisfies a Lipschitz condition with respect to the L_1-norm[3] with a polynomially bounded Lipschitz coefficient $\varphi(n)$.*

Proof. Denote the given function f, then the Lipschitz condition implies

$$|f(x) - f(y)| \leq \varphi(n)|x - y|_1 \qquad \forall x, y.$$

Further, the points in the feasible set satisfy $|x|_1 \leq \psi(n)$ some polynomial ψ, where $|.|_1$ denotes the L_1-norm.

Consider the following randomized polynomial-time algorithm for minimizing f.

Repeat n times:
Perturb the function f to a new function $\bar{f}$ and run an algorithm for finding the minimum of $\bar{f}$ for 2Δ steps, where Δ is the expected smoothed polynomial bound for finding the minimum of f. If the algorithm terminates, return the minimum of $\bar{f}$.

With probability at least $1 - 1/2^n$, this algorithm will output $\bar{f}_{min} = \bar{f}(y) = f(Ry)$, which we will show is very close to f_{min}. By definition of the perturbation matrix R and the Lipschitz condition,

$$\begin{aligned}
|\bar{f}(x) - f(x)| = |f(Rx) - f(x)| &\leq \varphi(n)|(R - I)x|_1 \\
&\leq \varphi(n) dist(R, I)|x|_1 \leq \varphi(n)\psi(n)\rho,
\end{aligned}$$

for every x in the feasible set, where $dist(R, I) \leq \rho$ is the perturbation size. Thus for a perturbation $\rho \leq 1/n^l$, where l is equal to the degree of the polynomial $\varphi(n)\psi(n)$ plus $\frac{\log(1/\epsilon)}{\log n}$, we have that $|\bar{f}(x) - f(x)| \leq \epsilon$ for all x and in particular, $\bar{f}_{min} \leq f_{min} + \epsilon$. □

Note that for functions bounded below by 1, the above theorem at once implies a randomized fully-polynomial approximation scheme (with multiplicative approximation factor $1 + \epsilon$).

Corollary 5.2. *There is no smoothed polynomial-time algorithm for general quasi-concave minimization, assuming $RP \neq NP$.*

Proof. Suppose contrarily that there is a smoothed expected polynomial-time algorithm. We will show that this implies the existence of a randomized fully polynomial approximation scheme for finding the minimum of f, the Lovász extension of the objective function h from Lemma 3.1, thereby contradicting theorem 3.3.

[3] We use the L_1-norm to facilitate the corollary below, however the theorem holds for more general metrics.

More precisely, we will work with the specific objective function h which is used in showing that the random 2-stage MST problem is $(\log n)$-hard to approximate [7]. We note that a modified version of the edge costs in which infinite edge costs are replaced by costs of n^3 does not alter the proof of hardness of the 2-stage MST, and thus the important property of the expected spanning tree cost h for our purpose is that it takes values in a polynomially bounded interval: $1 \leq h(x) \leq n^4$ for all 0-1-vectors x.

Next, we show that the Lovász extension f of h satisfies the Lipschitz condition

$$f(x) - f(y) \leq 2n^4 |x - y|_1,$$

where $|.|_1$ is the L_1 norm. By the definition of Lovász extension, f is piecewise linear, thus the ratio

$$\frac{f(x) - f(y)}{|x - y|_1}$$

is maximized when x, y are both on one of the linear segments $f_0(x) = \sum_i c_i x_i$. Again from the definition of f, we have $c_1 = f(b_{j_1})$ and $c_i = f(b_{j_i}) - f(b_{j_{i-1}})$ for $i = 2, ..., n$ where $\{b_{j_1}, ..., b_{j_n}\}$ is the basis for representing x in the Lovász extension definition. Therefore $|c_i| \leq 2n^4$.

Finally,

$$\frac{f(x) - f(y)}{|x - y|_1} = \frac{\sum_i c_i x_i - \sum_i c_i y_i}{\sum_i |x_i - y_i|} \leq \max_i |c_i| \leq 2n^4.$$

Now applying Theorem 5.1 implies the result and we get the desired contradiction. □

The last corollary implies that our smoothed bounds are tight, namely no polynomial smoothed bounds for general concave minimization are possible.

Acknowledgments

We thank Alan Edelman, Michel Goemans, David Karger, Christopher Mihelich, Michael Mitzenmacher, Asu Ozdaglar, Raj Rao, and Dan Spielman for valuable suggestions.

References

[1] N. Amenta and G. Ziegler. *Deformed Products and Maximal Shadows of Polytopes*, volume 223 of *Contemporary Mathematics*, pages 57–90. American Mathematics Society, Providence, 1999.

[2] D. Bertsekas, A. Nedić, and A. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA, 2003.

[3] T. Bonnesen and W. Fenchel. *Theory of Convex Bodies*. B C S Associates, Moscow, Idaho, 1988.

[4] J. Conway. *A Course in Functional Analysis*. Springer-Verlag, 1990.

[5] I. Dumitriu. *Eigenvalue Statistics for Beta-Ensembles*. PhD thesis, MIT, 2003.

[6] U. Feige. Personal communication. 2006.

[7] A. D. Flaxman, A. Frieze, and M. Krivelevich. On the random 2-stage minimum spanning tree. *Random Structures & Algorithms*, 28(1):24–36, 2006.

[8] D. Goldfarb. Worst case complexity of the shadow vertex simplex algorithm. Technical report, Columbia University, 1983.

[9] G. Golub and C. V. Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[10] R. Horst and P. M. Pardalos. *Handbook of Global Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[11] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[12] J. A. Kelner and D. A. Spielman. A randomized polynomial-time simplex algorithm for linear programming. In *Proceedings of the Symposium on the Theory of Computing*, 2006.

[13] H. Konno, C. Gao, and I. Saitoh. Cutting plane/tabu search algorithms for low rank concave quadratic programming problems. *Journal of Global Optimization*, 13:225–240, 1998.

[14] H. Konno, T. Thach, and H. Tuy. *Optimization on Low Rank Nonconvex Structures*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[15] L. Lovász. Submodular functions and convexity. In *Mathematical Programming—The State of the Art, A. Bachem, M, Grőtschel and B. Korte, eds.*, pages 235–257. Springer-Verlag, 1983.

[16] M. Mehta. *Matrix Theory*. 1990.

[17] E. Nikolova, J. A. Kelner, M. Brand, and M. Mitzenmacher. Stochastic shortest paths via quasi-convex maximization. In *Lecture Notes in Computer Science 4168 (ESA 2006)*, pages 552–563, Springer-Verlag, 2006.

[18] P. M. Pardalos and G. Schnitger. Checking local optimality in constrained quadratic programming is np-hard. *Operations Research Letters*, 7:33–35, 1988.

[19] M. Porembski. Cutting planes for low-rank-like concave minimization problems. *Operations Research*, 52(6):942–953, 2004.

[20] D. A. Spielman and S.-H. Teng. Smoothed analysis of termination of linear programming algorithms. *Math. Prog.: B*, 97(1/2):375, 2003.

[21] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51:385–463, 2004.