Abstract

Minimizing a convex function over a convex set in n-dimensional space is a basic, general problem with many interesting special cases. Here, we present a simple new algorithm for convex optimization based on sampling by a random walk. It extends naturally to minimizing quasi-convex functions and to other generalizations.

Solving Convex Programs by Random Walks

DIMITRIS BERTSIMAS and SANTOSH VEMPALA
M.I.T.

May 8, 2004

1 Introduction

The problem of minimizing a convex function over a convex set in $\mathbb{R}^n$ is a common generalization of well-known geometric optimization problems such as linear programming as well as a variety of combinatorial optimization problems including matchings, flows and matroid intersection, all of which have polynomial-time algorithms. As such, it represents a frontier of polynomial-time solvability and occupies a central place in the theory of algorithms.

In his groundbreaking work, Khachiyan [1979] showed that the Ellipsoid method [Yudin and Nemirovski 1976] solves linear programs in polynomial time. Subsequently, Karp and Papadimitriou [1982], Padberg and Rao [1981], and Grötschel, Lovász and Schrijver [Grötschel et al. 1981] independently discovered the wide applicability of the Ellipsoid method to combinatorial optimization problems. This culminated in the book by the last set of authors [Grötschel et al. 1988], in which it is shown that the Ellipsoid method solves the problem of minimizing a convex function over a convex set in $\mathbb{R}^n$ specified by a *separation oracle*, i.e., a procedure which given a point x either reports that the set contains x or returns a halfspace that separates the set from x. For the special case of linear programming, the oracle simply checks if the query point satisfies all the constraints of the linear program, and if not, reports a violated constraint; another well-known special case is *semi-definite* programming. Vaidya improved the complexity via a more sophisticated algorithm [Vaidya 1996].

In this paper, we present a simple new algorithm for the problem, based on random sampling. Its complexity is optimal in terms of the number of oracle queries, but the overall running time can be higher than that of previous algorithms (see Table 1 at the end of Section 2 for a precise comparison). The key component of the algorithm is sampling a convex set by a random walk. Random walks have long been studied for their mathematical appeal, but of late they have also played a crucial role in the discovery of polynomial-time algorithms. Notable applications include estimating the volume of a convex set [Dyer et al. 1991] and computing the permanent of a non-negative matrix [Jerrum et al. 2001]. They have also been used in machine learning and online algorithms [Kalai and Vempala 2002]. Our algorithm is a novel application of random walks to the field of optimization. For the analysis, we prove a generalization of Grunbaum's theorem [Grunbaum 1960] about cutting convex sets (see Section 4.1); this might be of interest in other settings.

So far, we have assumed that the convex set of interest has an efficient separation oracle. A natural question is whether optimization can be solved using a significantly weaker oracle, namely a *membership* oracle (which reports whether a query point is in the set or not, but provides no other information). One of the main results in [Grötschel et al. 1988] is that a linear function can be optimized over a convex set K given only by a membership oracle, provided K is "centered", i.e., we are also given a point y and a guarantee that a ball of some radius r around y is contained in K. The algorithm is intricate and involves a sophisticated variant of the Ellipsoid method, called the *shallow-cut* Ellipsoid (note that feasibility is trivial since we are given a feasible point). Our algorithm provides a simple solution to this problem. In fact, as we show in Section 6, it solves the following generalization: let K be the intersection of two convex sets K_1 and K_2, where K_1 is a "centered" convex set with a membership oracle and K_2 is a convex set with a separation oracle; find a point in K if one exists. The generalization includes the special case of minimizing a quasi-convex function over a centered convex set given by a membership oracle. This problem (and its special case mentioned above) are not known to be solvable using the Ellipsoid method or Vaidya's algorithm.

2 The algorithm

In this section, we present an algorithm for the *feasibility* problem: find a point in a given convex set specified by a separation oracle. This can be used to minimize any *quasi-convex* function. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasi-convex if for any real number t, the set $\{x \in \mathbb{R}^n \mid f(x) \le t\}$ is convex. Note that any convex function is also quasi-convex. The problem of minimizing a quasi-convex function is easily reduced to the feasibility problem: to minimize a quasi-convex function $f(x)$ we simply add the constraint $f(x) \leq t$ and search (in a binary fashion) for the optimal t.

In the description below, we assume that the convex set K is contained in the axis-aligned cube of width R centered at the origin; further if K is non-empty then it contains a cube of width r (see e.g. [Bertsimas and Tsitsiklis 1997, Grötschel et al. 1988] for a justification). The choice of cubes here is somewhat arbitrary; we could instead use balls, for example. The parameter L is equal to $\log \frac{R}{r}$.

Algorithm.
Input: `A separation oracle for a convex set` K `and a number` L.
Output: `A point in` K `or a guarantee that` K `is empty.`

1. `Let` P `be the axis-aligned cube of width` R `with center` $z = 0$.
2. `Check if` z `is in` K. `If so then report` z `and stop. If not, then let` $a^T x \leq b$ `be the halfspace containing` K `reported by the oracle. Set` $H = \{x \mid a^T x \leq a^T z\}$.
3. `Set` $P = P \cap H$. `Pick` N `random points` $y_1, y_2, \ldots, y_N$ `from` P. `Set` z `to be their average:` $z = \frac{1}{N} \sum_{i=1}^{N} y_i$.
4. `Repeat steps 2 and 3 at most` $2nL$ `times. Report` K `is empty.`

Roughly speaking, the algorithm is computing an approximate centroid in each iteration[1]. The number of samples required in each iteration, N, is $O(n)$ for an arbitrary convex set in $\mathbb{R}^n$ and $O(\log^2 m)$ if K is a polyhedron with m inequalities (i.e., a linear program).

The idea behind the algorithm is that the volume of the enclosing polytope P is likely to drop by a constant factor in each iteration. We prove this in Section 4.2 (see Lemma 5) and derive as a consequence that if the algorithm does not stop in $2nL$ iterations, then K must be empty with high probability (Theorem 6). Thus, the total number of calls to the separation oracle is at most $2nL$. This matches the (asymptotic) bound for Vaidya's

[1]The idea of an algorithm based on computing the *exact* centroid was suggested in 1965 by Y. Levin [Levin 1965], but is computationally intractable.

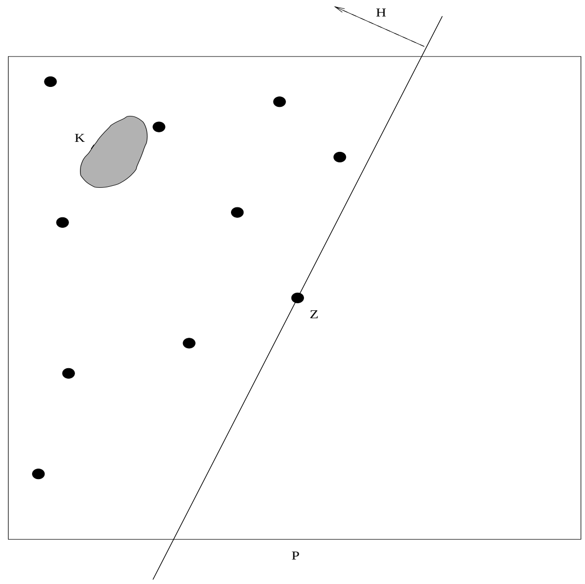

Figure 1: An illustration of the algorithm.

algorithm, and is in general the best possible [Yudin and Nemirovski 1976]. The Ellipsoid algorithm, in contrast, takes $O(n^2L)$ iterations and as many calls to the oracle.

Each iteration of our algorithm uses random samples from the current polytope P. The time per iteration depends on how quickly we can draw random samples from P. The problem of sampling from a convex set has received much attention in recent years [Dyer et al. 1991, Lovász and Simonovits 1993, Kannan et al. 1997, Lovász 1998, Kannan and Lovász 1999], in part because it is the only known way of efficiently estimating the volume of a convex set. The general idea is to take a random walk in the set. There are many ways to walk randomly; of these, the *ball* walk (go to a random point within a small distance) [Kannan and Lovász 1999] and *hit-and-run* (go to a random point along a random direction) [Lovász 1998] have the best known bounds on the number of steps needed to draw a random sample. The bounds on the number of steps depend on how "round" the convex set is. For a set that is close to *isotropic position* (see Section 4), $O(n^3)$ steps are enough to get one nearly random point from a "good" starting distibution. In our case, the initial convex set, the cube, is indeed in isotropic position. However this might not be the case after some iterations. As we describe in Section

Algorithm	Optimization/Feasibility with a Separation oracle	Optimization with a Membership oracle
Ellipsoid	$O(n^2LT + n^4L)$	$O(n^{10}LT + n^{12}L)$
Vaidya's	$O(nLT + n^{3.38}L)$	N/A
Random walk	$O(nLT + n^7L)$	$O(n^5LT + n^7L)$

Table 1: Complexity comparison

5, this problem can be tackled by computing an affine transformation that keeps the current polytope P in near-isotropic position. We also propose an alternative (in Section 5.1) that avoids this computation and instead incorporates the information about isotropy *implicitly* in the steps of a random walk by maintaining a set of points. The time per iteration of the algorithm is bounded by $O(n^4)$ steps of a random walk. Each step of the walk takes $O(n^2)$ arithmetic operations to implement, and hence the algorithm takes $O(n^6)$ arithmetic operations per iteration in addition to one oracle call.

In Section 6, we describe a variant of the algorithm for optimization when we are given a membership oracle. We conclude this section with a comparison of the complexities of the three algorithms given a separation oracle or given a membership oracle. The parameter T denotes the time taken by the relevant oracle to answer a query.

In practice, drawing random samples from convex sets might be much faster than the known worst-case bounds; also, sampling convex sets is an active research area and there might well be faster sampling methods in the future which would directly improve the complexity of the random walk algorithm.

3 Preliminaries

The following definitions will be useful throughout the analysis. We assume that all our convex sets are closed and bounded.

Definition 1 *A convex set K in $\mathbb{R}^n$ is said to be in* isotropic position *if its center of gravity is the origin, i.e., for a random point x in K,*

$$E_K(x) = 0$$

and its covariance matrix is the identity, i.e.,

$$E_K(xx^T) = I.$$

Equivalently, for any unit vector v ($||v|| = 1$),

$$\frac{1}{\mathrm{vol}(K)} \int_K (v^T x)^2 dx = 1.$$

In other words, for a set K in isotropic position, the average squared length in any direction is 1. In particular, this implies that

$$\frac{1}{\mathrm{vol}(K)} \int_K ||x||^2 dx = n.$$

For any full-dimensional convex set, there exists an affine transformation that puts the set in isotropic position. To bring a convex set K with center of gravity z into isotropic position, let $A = E_K((x-z)(x-z)^T)$ be its covariance matrix. Since K is full-dimensional, A is positive definite, and so there exists a matrix B such that $B^2 = A^{-1}$. Now define the transformation $K' = \{y : y = B(x-z), x \in K\}$. Then $E_{K'}(y) = 0$ and $E_{K'}(yy^T) = I$.

Definition 2 *We say that a convex set K is in* near-isotropic *position if for any unit vector v,*

$$\frac{1}{2} \le \frac{1}{\mathrm{vol}(K)} \int_K (v^T(x-\bar{x}))^2 dx \le \frac{3}{2}$$

where $\bar{x}$ is the centroid of K. Equivalently, the covariance matrix of the uniform distribution over K has eigenvalues between $\frac{1}{2}$ and $\frac{3}{2}$.

The *Minkowski sum* of two sets A, B in $\mathbb{R}^n$ is defined as

$$A + B = \{y + z \mid y \in A, z \in B\}.$$

For example, the Minkowski sum, $x + rB$, of a point x and a ball rB of radius r is the ball of radius r centered at x.

The following theorem is known as the Brunn-Minkowski inequality (see e.g. [Gardner 2002, Schneider 1993]).

Theorem 1 *Let A and B be full-dimensional convex subsets of $\mathbb{R}^n$. Then for any $\lambda \in [0,1]$,*

$$\mathrm{vol}(\lambda A + (1-\lambda)B)^{\frac{1}{n}} \ge \lambda \mathrm{vol}(A)^{\frac{1}{n}} + (1-\lambda)\mathrm{vol}(B)^{\frac{1}{n}}.$$

Some of our proofs utilize logconcave functions which we introduce next.

Definition 3 *A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is* logconcave *iff for any two points $a, b \in \mathbb{R}^n$ and any $\lambda \in (0, 1)$,*

$$f(\lambda a + (1 - \lambda)b) \geq f(a)^\lambda f(b)^{1-\lambda}.$$

In other words, a nonnegative function f is logconcave if its support is convex and $\log f$ is concave. For example, a function that is constant over a bounded convex set and zero outside the set is logconcave. Another example is a Gaussian density function. It can be easily verified from the definition above that the product of two logconcave functions is also logconcave (but their sum is not).

Finally, we will use the following simple fact from linear algebra. When we apply a linear transformation A to a compact convex set K, then the volume of the resulting set AK is $|det(A)|\mathrm{vol}(K)$.

4 Analysis

In this section, we prove that with high probability, the algorithm needs only $O(n)$ random points per iteration and $O(nL)$ iterations. Section 5 describes how to efficiently obtain the random points in each iteration.

4.1 Geometric properties

The following theorem was proved by Grunbaum.

Theorem 2 *[Grunbaum 1960] For a convex set K in $\mathbb{R}^n$, any halfspace that contains the centroid of K also contains at least $1/e$ of the volume of K.*

Here, we prove a generalization.

Theorem 3 *Let K be a convex set in isotropic position and z be a point at distance t from its centroid. Then any halfspace containing z also contains at least $\frac{1}{e} - t$ of the volume of K.*

Theorem 3 with $t = 0$ clearly implies Grunbaum's theorem for isotropic convex sets. Further, any convex set K can be mapped to an isotropic one by an affine transformation which changes the volume of any subset by the same factor; so the ratio of two volume remains unchanged, and in particular for any halfspace H, the ratio $\mathrm{vol}(H \cap K)/\mathrm{vol}(K)$ is preserved under an affine transformation. Thus Grunbaum's theorem is implied in general.

The proof of Theorem 3 will rely on the next lemma. The function defined in the lemma is the distribution induced by cross-sectional areas perpendicular to some vector a.

Lemma 4 *For an isotropic convex set K and any unit vector $a \in \mathbb{R}^n$, define a function $f : \mathbb{R} \to \mathbb{R}_+$ as*

$$f(y) = \frac{1}{\mathrm{vol}(K)} \int_{x \in K, a^T x = y} dx.$$

Then

$$\max_{y \in \mathbb{R}} f(y) < 1.$$

Proof
Fix a unit vector $a \in \mathbb{R}^n$. Let f_K be the one-dimensional marginal distribution

$$f_K(y) = \frac{1}{\mathrm{vol}(K)} \int_{x \in K, a^T x = y} dx.$$

In other words, $f_K(y)$ is the $(n-1)$-dimensional volume of K intersected with the hyperplane $a^T x = y$, as a fraction of the volume of K. Define

$$\bar{y}_K = \int_{\mathbb{R}} y f_K(y)\, dy \quad \text{and} \quad I(K) = \int_{\mathbb{R}} (y - \bar{y}_K)^2 f_K(y) dy.$$

Note that $I(K)$ is the variance (or moment of inertia) of K along the vector a. By translating K, we can assume without loss of generality that $\bar{y}_K = 0$. Further, by a rotation, we can assume that $a = (1, 0, \ldots, 0)$, i.e., the unit vector along x_1.

We will, in fact, prove a slightly stronger statement: for any convex body K for which f_K is defined as above and which satisfies

$$\int_{\mathbb{R}} y f_K(y)\, dy = 0 \quad \text{and} \quad \int_{\mathbb{R}} y^2 f_K(y)\, dy = 1, \tag{1}$$

we have $\max_{\mathbb{R}} f_K(y) < 1$.

To this end, let K be a convex body satisfying (1) for which $\max f_K(y)$ is the maximum possible. We will show that K can be assumed to be an isotropic cone. To see this, first consider the set K' obtained by replacing each cross-section $K \cap \{x | x_1 = y\}$ by an $(n-1)$-dimensional ball of area $f_K(y)$ and centered at the point $(y, 0, \ldots, 0)$. Clearly, the set K' has the same volume as K. Further, if we let $r(y)$ denote the radius of the $(n-1)$-dimensional ball at y, then r is a concave function. This follows from the

Brunn-Minkowski inequality (Theorem 1): Let A_1 be the cross-section at y_1, A_2 the cross-section at y_2 and A the cross-section at $\lambda y_1 + (1-\lambda)y_2$ for some $\lambda \in [0,1]$. Then,

$$\mathrm{vol}(A)^{\frac{1}{n-1}} \geq \mathrm{vol}(\lambda A_1 + (1-\lambda)A_2)^{\frac{1}{n-1}} \geq \lambda \mathrm{vol}(A_1)^{\frac{1}{n-1}} + (1-\lambda)\mathrm{vol}(A_2)^{\frac{1}{n-1}}.$$

This implies that $r(\lambda y_1 + (1-\lambda)y_2) \geq \lambda r(y_1) + (1-\lambda) r(y_2)$, i.e., r is concave. Hence the set K' is convex. Further, $f_{K'} = f_K$ and so (1) continues to be satisfied. So, without loss of generality, we can assume that K is symmetric about the x_1 axis.

If K is a cone, we are done. If not, let y^* be a point where f_K is maximum and suppose that $y^* \geq 0$ (the other case is symmetric). Divide K into three

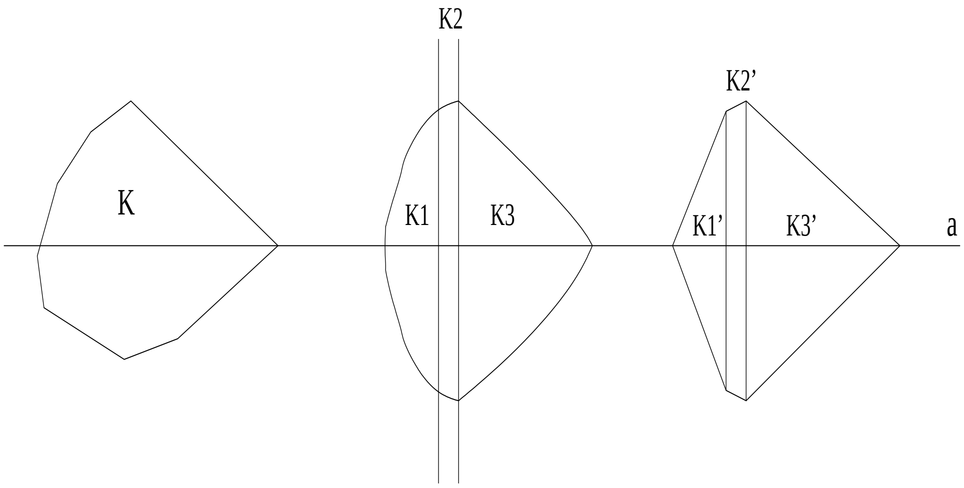

Figure 2: The proof of Lemma 4.

parts:

$$K_1 = K \cap \{x | x_1 \leq 0\}, \quad K_2 = K \cap \{x | 0 \leq x_1 \leq y^*\}, \quad K_3 = K \cap \{x | x_1 \geq y^*\}.$$

We will now use the following claim to simplify K (which essentially says that moving mass away from the center of gravity can only increase the moment of inertia).

Claim. Suppose a set C' is obtained from a set C by moving mass away from the centroid of C or keeping it at the same distance along the x_1 axis, i.e., for any $t \geq 0$, $\mathrm{vol}(C' \cap \{x : |x_1| \geq t\}) \geq \mathrm{vol}(C \cap \{x : |x_1| \geq t\})$. Then $I(C') \geq I(C)$.

To prove the claim, assume that the centroid of C is 0. Let the coordinate of a point x change from x_1 to $x_1 + g(x)$ where $g(x)$ is a real number with the same sign as x_1. Note that $\mathsf{E}_C(x_1) = 0$ by assumption.

$$\begin{aligned} I(C') = \mathsf{Var}_{C'}(x_1) &= \mathsf{Var}_C(x_1 + g(x)) \\ &= \mathsf{E}_C((x_1 + g(x))^2) - \mathsf{E}_C(x_1 + g(x))^2 \\ &= \mathsf{Var}_C(x_1) + \mathsf{Var}_C(g(x)) + 2\mathsf{E}_C(x_1 g(x)) \\ &\geq I(C) \end{aligned}$$

since $x_1 g(x) \geq 0$ for every point x.

We now make the following changes to our set K. Replace K_1 by a cone K_1' of the same volume with base area equal to $f_K(0)$. Replace K_2 by a truncated cone K_2' of height y^*, top area $f_K(0)$ and base area $f_K(y^*)$. Finally replace K_3 by a cone K_3' of base area $f_K(y^*)$ and $\mathrm{vol}(K_3') = \mathrm{vol}(K_3) + \mathrm{vol}(K_2) - \mathrm{vol}(K_2')$. Let $K' = K_1' \cup K_2' \cup K_3'$ be the new convex set. See Figure 2 for an illustration.

During the first step, by the claim, the moment of inertia can only increase. After the first step, the new center of gravity along the x_1 axis can only have moved to the left (since mass moves to the left). Thus the next two steps also move mass away from the center of gravity, and we can apply the claim. At the end, we get

$$I(K') \geq I(K), \quad \mathrm{vol}(K) = \mathrm{vol}(K'), \quad \max f_K(y) = \max f_{K'}(y).$$

If $I(K') > I(K)$, then we can scale down K' along the x_1 axis and scale it up perpendicular to the x_1 axis, in such a way as to maintain the volume, achieve $I(K') = I(K)$, and have $\max f_{K'}(y) > \max f_K(y)$. This is a contradiction. Thus, we can assume that K has the form of K', i.e., $K_i = K_i'$.

At this point, one could perform a somewhat tedious (but routine) computation to show that a cone of the same volume as K and base area $f_K(y^*)$ has a larger moment of inertia than K, leading to the same contradiction unless K itself is a cone. The following observation, again using the claim, avoids this computation. For a point $y \in [0, y^*]$ along the x_1 axis, consider the subsets $S_1 = \{x | x_1 \leq -y\}$, $S_2 = \{x | -y \leq x_1 \leq y\}$ and $S_3 = \{x | x_1 \geq y\}$. Now leave the set S_3 unchanged, replace the set S_2 by the truncated cone S_2' with base area $f_K(y)$ and top area $f_K(-y)$ and the set S_1 by the cone S_1' with base area $f_K(-y)$ and volume $\mathrm{vol}(K) - \mathrm{vol}(S_3) - \mathrm{vol}(S_2')$, so that the total volume is unchanged. This is illustrated in Figure 3. If $K_1 \cup K_2$ is not a cone, then there is some $y > 0$ for which we can do the above replacement. But then, by construction, we are moving mass away from the center (mass

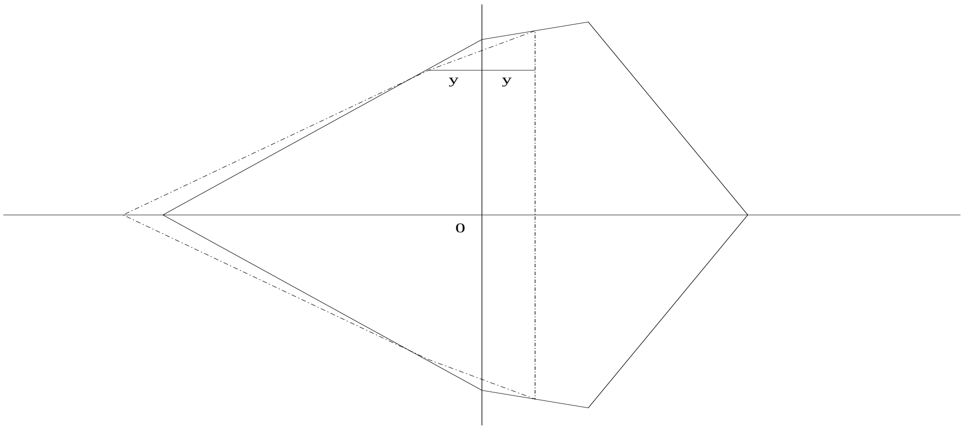

Figure 3: The proof of Lemma 4 continued.

at distance at most y along the x_1 axis is moved to distance at least y), and so by the claim, the moment of inertia can only go up. Hence, we can assume that $K_1 \cup K_2$ is a cone of base area $f_K(y^*)$ and thus K is the union of two cones with the same base. Now it is straightforward to check that a cone of the same volume as K and base area $f_K(y^*)$ has a larger moment of inertia, and so K itself must be a cone.

Finally, an isotropic cone has height $h = n + 1\sqrt{\frac{n+2}{n}}$ and volume Ah/n where A is the area of the base. Hence,

$$\max f_K(y) = \frac{A}{Ah/n} = \frac{n}{h} = \frac{n}{n+1}\sqrt{\frac{n}{n+2}}.$$

Using the lemma, we can prove Theorem 3.

Proof(of Theorem 3)
Let K be a convex set in isotropic position and let z be a point at distance t from its centroid. Consider a hyperplane through z with normal vector a such that $||a|| = 1$. Let f be defined as in Lemma 4. By the Lemma, $\max_{y\in\mathbb{R}} f(y) < 1$. Without loss of generality, assume that $a^T z \geq 0$. The fraction of the volume of K that is cut off by the halfspace $a^T x \geq a^T z$ is at

least

$$\begin{aligned}
\int_{a^T z}^{\infty} f(y)dy &= \int_0^{\infty} f(y)dy - \int_0^{a^T z} f(y)dy \\
&\geq \frac{1}{e} - \int_0^{a^T z} f(y)dy \quad \text{(using Theorem 2)} \\
&\geq \frac{1}{e} - |a^T z| \quad \text{(using Lemma 4)} \\
&\geq \frac{1}{e} - ||z|| = \frac{1}{e} - t.
\end{aligned}$$

4.2 Number of iterations

The next theorem, whose proof uses Theorem 3, is the key to bounding the number of iterations. In this section, by high probability we mean probability higher than any fixed constant.

Lemma 5 *The volume of P drops by a factor of $\frac{2}{3}$ with high probability in each iteration.*

Proof
We will prove that for any convex set K, if z is the average of sufficiently many random samples from K, then any halfspace through z cuts off a constant fraction of the volume of K. To this end we can assume without loss of generality that K is in isotropic position. This is because of two facts: (i) as shown earlier, any convex set can be brought into isotropic position by an affine transformation and (ii) on applying an affine transformation A to K, the volume scales by $det(A)$, i.e., $\text{vol}(AK) = |det(A)|\text{vol}(K)$; so affine transformations preserve ratios of volumes.
Let $y_1, y_2, \ldots, y_N$ be the N samples drawn uniformly from K. Let

$$z = \frac{1}{N}\sum_{i=1}^{N} y_i.$$

Then, using the isotropy of K,

$$\mathsf{E}(y_i) = 0 \quad \text{and} \quad \mathsf{E}(||y_i||^2) = n.$$

Thus,

$$\mathsf{E}(z) = 0 \quad \text{and} \quad \mathsf{E}(||z||^2) = \frac{n}{N}.$$

Since $\mathsf{E}(||z||^2) = \frac{n}{N}$ by choosing $N = O(n)$ we can have $||z||$ smaller than any constant. By Theorem 3 any halfspace $a^T x \geq a^T z$ passing through z cuts off at least $\frac{1}{e} - ||z||$ fraction of the volume of K. We choose N so that $\frac{1}{e} - ||z|| \geq \frac{1}{3}$ with high probability. Then any halfspace through z cuts off at least $\frac{1}{3}$ of the volume of K.

We remark that in the proof above we only need the random samples to be pairwise independent. This is because the only quantity we needed to bound was the distance of z from the origin for an isotropic convex set, and for this we only used the variance of $||z||$. The bounds for the variance only need pairwise independence.

Theorem 6 *If the algorithm does not find a feasible point in $2nL$ iterations, then with high probability, the given convex set is empty.*

Proof
The initial volume of P is R^n. If the set K is nonempty, then it has volume at least r^n. Thus the number of iterations in which the volume drops by $2/3$ is at most

$$\log_{\frac{3}{2}} \left(\frac{R^n}{r^n} \right) = n \log_{\frac{3}{2}} \frac{R}{r}$$

and thus with high probability, the total number of iterations is at most $2nL$.

4.3 A better bound for linear programming

For the case of linear programming, when the target convex set K is a polyhedron defined by m linear inequalities, a smaller number of samples can be used in each iteration. The next lemma proves that any single violated inequality is detected with high probability. The main idea is that since the separation oracle will simply return one of only m hyperplanes in case of a violation, one can apply the "union bound" to show that with high probability a small sample will detect a violation. This is in contrast to the general case where the separation oracle might return one of infinitely many possibilities.

Lemma 7 *Let K be a convex set and z be the average of N random samples from K. Let a be a fixed unit vector. Then the probability that the halfspace $a^T x \leq a^T z$ cuts off at least $\frac{1}{3}$ of K is at least $1 - 2^{-c\sqrt{N}}$ for some constant c.*

Proof
Assume without loss of generality that P is in isotropic position. Let $Y_i = a^T y_i$ for $i = 1, \ldots, N$. Then, Y_i's are independent random variables with distribution given by the logconcave function

$$f(y) = \frac{1}{\mathrm{vol}(K)} \int_{x \in K, a^T x = y} dx.$$

We have

$$\mathsf{E}(Y_i) = 0 \quad \text{and} \quad \mathsf{E}(Y_i^2) = 1.$$

Let

$$Z = a^T z = \frac{1}{N} \sum_{i=1}^{N} Y_i.$$

Then,

$$\mathsf{E}(Z) = 0 \quad \text{and} \quad \mathsf{E}(Z^2) = \frac{1}{N}.$$

Now the distribution of Z is the convolution of the distributions of the Y_i's and is also logconcave [Prekopa 1973]. Thus, there is a constant D such that for any $t \geq 1$,

$$P\left(|Z| > \frac{t}{\sqrt{N}}\right) \leq e^{-Dt}.$$

As a consequence, for $N > 1/c^2$,

$$Pr(|a^T z| > c) \leq e^{-cD\sqrt{N}}.$$

By choosing c small enough, we get that the halfspace $a^T x \leq a^T z$ cuts off at least $\frac{1}{3}$ of the volume of P with probability at least $1 - e^{-c'\sqrt{N}}$ for some absolute constant c'.

Corollary 8 *Let the target convex set be the intersection of m halfspaces. Then with $N = O(\log^2 m)$, the volume of P drops by a factor of $\frac{2}{3}$ in each iteration with high probability.*

Proof
As in the proof of Lemma 5, let $y_1, \ldots, y_N$ be random variables denoting the samples from the current polytope P and let z be their average. In Lemma 5, we showed that for *any* unit vector a, the hyperplane normal to a and passing through z is likely to cut off at least $\frac{1}{3}$ of the volume of P. Suppose that the target set K is the intersection of m halfspaces defined by hyperplanes with normal vectors $a_1, \ldots, a_m$ (with $||a_i|| = 1$). Then we only

need to show that each of the hyperplanes $a_i^T x = a_i^T z$ cuts off at least $\frac{1}{3}$ of P. This is because the separation oracle for the target convex set (which simply checks the m halfspaces and reports one that is violated) will return a hyperplane parallel to one of these.

By Lemma 7, any single halfspace $a^T x \leq a^T z$ cuts off $\frac{1}{3}$ of P with probability at least $1 - 2^{-c\sqrt{N}}$ for some constant c. Setting $N = O(\log^2 m)$ implies that with high probability any one of m halfspaces will cut off a constant fraction of P.

5 Sampling and Isotropy

In each iteration, we need to sample the current polytope. For this we take a random walk. There are many ways to walk randomly but the two ways with the best bounds on the *mixing time* are the ball walk and hit-and-run. They are both easy to describe.

Ball walk

1. Choose y uniformly at random from the ball of radius δ centered at the current point x.

2. If y is in the convex set then move to y; if not, try again.

Hit-and-run

1. Choose a line ℓ through the current point x uniformly at random.

2. Move to a point y chosen uniformly from $K \cap \ell$.

The mixing time of the walk depends on how close the convex set is to being in isotropic position (recall definition 2). In addition to isotropy, the starting point (or distribution) of a random walk plays a role in the bounds on the mixing time. The best bounds available are for a *warm start* – a starting distribution is already close to the stationary distribution π. The *total variation* distance between two distributions σ and π with support K is defined as:

$$||\sigma - \pi|| = \sup_{A \subseteq K} |\sigma(A) - \pi(A)|.$$

The following result is paraphrased from [Lovász 1998].

Theorem 9 *[Lovász 1998] Let K be a near-isotropic convex set, σ be any distribution on it with the property that for any set S with $\pi(S) \leq \epsilon$, we* *have $\sigma(S) \leq 10\epsilon$. Then after $\Omega(\frac{n^3}{\epsilon^2}\log\frac{1}{\epsilon})$ steps of hit-and-run starting at σ, the distribution obtained has total variation distance less than ϵ from the uniform distribution.*

A similar statement holds for ball walk with step size $\delta = \Theta(\frac{1}{\sqrt{n}})$ (see [Kannan and Lovász 1999]). One advantage of hit-and-run is that there is no need to choose the "step size" δ.

The problem with applying this directly in our algorithm to the current polytope P is that after some iterations P may not be in near-isotropic position. One way to maintain isotropic position is by using random samples to calculate an affine transformation. It was proven in [Kannan et al. 1997] that for any convex set, $O(n^2)$ samples allow us to find an affine transformation that brings the set into near-isotropic position. This was subsequently improved to $O(n\log^2 n)$ [Rudelson 1999]. The procedure for a general convex set K is straightforward:

1. Let $y_1, y_2, \dots, y_N$ be random samples from K.
2. Compute

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad \text{and} \quad Y = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(y_i - \bar{y})^T. \tag{2}$$

The analysis of this procedure rests on the following consequence of a more general lemma due to Rudelson [Rudelson 1999].

Lemma 10 *[Rudelson 1999] Let K be a convex body in $\mathbb{R}^n$ in isotropic position and $\eta > 0$. Let $y_1, \dots, y_N$ be independent random points uniformly distributed in K, with*

$$N \geq c\frac{np}{\eta^2}\log\frac{n}{\eta^2}\max\{p, \log n\}.$$

where c is a fixed absolute constant and p is any positive integer. Then

$$\mathsf{E}\left(\left\|\frac{1}{N}\sum_{i=1}^{N} y_i y_i^T - I\right\|^p\right) \leq \eta^p$$

Corollary 11 *Let K be a convex set and $\bar{y}$ and Y be defined as in (2). There is an absolute constant c such that for any integer $p \geq 1$, and $N \geq cpn\log n\max\{p, \log n\}$, with probability at least $1 - \frac{1}{2^{p-1}}$, the set*

$$K' = \{x | Y^{\frac{1}{2}}x + \bar{y} \in K\}$$

is in near-isotropic position.

Proof

Without loss of generality, we can assume that K is in isotropic position. Let

$$Y_0 = \frac{1}{N}\sum_{i=1}^{N} y_i y_i^T.$$

Then applying Lemma 10 we have

$$\mathsf{E}(||Y_0 - I||^p) \leq \eta^p.$$

Hence,

$$\begin{aligned} Pr(||Y_0 - I|| > 2\eta) &= Pr(||Y_0 - I||^p > (2\eta)^p) \\ &\leq \frac{\mathsf{E}[||Y_0 - I||^p]}{(2\eta)^p} \\ &\leq \frac{1}{2^p} \end{aligned}$$

Now $Y = Y_0 - \bar{y}\bar{y}^T$, and so $||Y - I|| \leq ||Y_0 - I|| + ||\bar{y}||^2$. The random variable $\bar{y}$ has a logconcave distribution and $\mathsf{E}(||\bar{y}||^2) = n/N$. So, there is a constant D such that for any $t \geq 1$,

$$Pr\left(||\bar{y}|| > t\sqrt{\frac{n}{N}}\right) \leq e^{-Dt}.$$

Using $\eta = 1/8$ and $t = \sqrt{N/4n}$, we get that with $N \geq cpn \log n \max\{p, \log n\}$ points (this is a different constant c from that in Lemma 10),

$$Pr\left(||Y - I|| > \frac{1}{2}\right) \leq \frac{1}{2^{p-1}}.$$

Next we evaluate $\max v^T Y v$ and $\min v^T Y v$ over unit vectors v.

$$\begin{aligned} v^T Y v &= v^T I v + v^T (Y - I) v \\ &\leq 1 + ||Y - I||. \end{aligned}$$

$$\begin{aligned} v^T Y v &= v^T I v + v^T (Y - I) v \\ &\geq 1 - ||Y - I||. \end{aligned}$$

Thus with probability $1 - \frac{1}{2^{p-1}}$, $v^T Y v$ is between $\frac{1}{2}$ and $\frac{3}{2}$ for any unit vector v, and so K' is in near-isotropic position.

In fact, using $N = O^*(n)$ points, the probability of failure in each iteration can be made so low (inverse polynomial) that the probability of failure over the entire course of the algorithm is smaller than any fixed constant. In order to keep the sampling efficient, we calculate such an affine transformation in each iteration, using samples from the previous iteration. However, we do not need to apply the transformation to the current polytope P. Instead, we could keep P as it is and use the following modified random walk: From a point x,

1. Choose a point y uniformly at random from $x + \delta Y^{\frac{1}{2}} B_n$.

2. If y is in the convex set then move to y; if not, try again.

A similar modification (choose the line ℓ from $Y^{\frac{1}{2}} B_n$ rather than B_n) can be made to hit-and-run also. In this paper, we do not deal with the technical difficulties caused by near-uniform sampling of nearly-independent points. For a detailed discussion of those issues, the reader is referred to [Kannan et al. 1997].

Theorem 12 *Each iteration of the algorithm can be implemented in $O^*(n^4)$ steps of a random walk in P.*

Proof
Our initial convex set, the cube, is in isotropic position and it is easy to sample from it. Given a warm start in a current polytope P in near-isotropic position, we take $O^*(n^3 N)$ steps of a random walk to get $2N$ nearly random samples (using Theorem 9). We compute the average z of N of these, and if z is not in the target convex set, then we refine P with a new halfspace H to get P'. Of the remaining N points, at least a constant fraction (say $\frac{1}{4}$) are in P' with high probability. Further, their distribution is nearly uniform over P'. Using these points, we estimate $Y^{\frac{1}{2}}$ by the formula of Corollary 11. With high probability this is a nonsingular matrix. By Corollary 11, we only need $N = O^*(n)$ samples in each iteration. The reason for discarding the subset of samples used in computing the average is to avoid correlations with future samples.

5.1 An isotropic variant

Here we describe an alternative method which implicitly maintains isotropy. It has the advantage of completely avoiding the computation of the linear transformation $Y^{\frac{1}{2}}$.

Instead of walking with a single point, consider the following *multi-point* walk: Maintain m points $v_1, v_2, \ldots, v_m$. For each $j = 1, \ldots, m$,

1. Choose a direction $\ell = \sum_{i=1}^{m} \alpha_i v_i$ where $\alpha_1, \ldots, \alpha_m$ are drawn independently from the standard normal distribution.

2. Move v_j to a random point in K along ℓ.

Theorem 13 *Suppose the multi-point walk is started with $m = \Omega(n \log^2 n)$ points drawn from a distribution σ that satisfies $\sigma(S) \leq 10\epsilon$ for any subset S with $\pi(S) \leq \epsilon$. Then their joint distribution after $\Omega(m\frac{n^3}{\epsilon^2} \log \frac{1}{\epsilon})$ steps has total variation distance less than ϵ from the distribution obtained by picking m random points uniformly and independently from π.*

Proof
The random walk described above is invariant under affine transformations, meaning there is a one-to-one map betweeen a walk in K and a walk in an affine transformation of K; hence we can assume that K is isotropic. Now consider a slightly different walk where we keep m points, pick one and make a standard hit-and-run step, i.e., the direction is chosen uniformly. The corresponding Markov chain has states for each m-tuple of points from K, and since each point is walking independently, it follows from Theorem 9 that the chain has the bound described above.

Compare this to the Markov chain for our random walk. The stationary distribution is the same (our Markov chain is symmetric). The main observation is that for $v_1, \ldots, v_m$ picked at random and $m = \Omega(n \log^2 n)$, Corollary 11 imples that with high probability, the matrix of inertia has eigenvalues bounded by constants from above and below. We can choose m so that this probability is $1 - \frac{1}{n^{10}}$. Thus, for all but a $\frac{1}{n^{10}}$ fraction of states, the probability of each single transition of our Markov chain is within a constant factor of the first Markov chain. So the conductance of any subset of probability greater than $\frac{1}{n^8}$ (say) is within a constant factor of its conductance in the first chain. Since we start out with a warm distribution (i.e., a large subset), this implies the bound on the mixing time.

Thus the entire convex programming algorithm now consists of the following: We maintain $2N$ random points in the current polytope P. We use N of them to generate the query point and refine P. Among the rest, we keep those that remain in P, and continue to walk randomly as described above till we have $2N$ random points again.

6 A generalization

In this section, we consider the problem of optimization over a convex set given by a weaker oracle, namely a membership oracle. In [Grötschel et al. 1988], it is shown that this is indeed possible using a sophisticated variant of the ellipsoid algorithm, provided we have a *centered* convex set K. That is, in addition to the membership oracle, we are given a point $y \in K$, and a guarantee that a ball of radius r around y is contained in K. The algorithm and proof there are quite nice but intricate[2].

Our algorithm provides a solution for a more general problem: K is the intersection of two convex sets K_1 and K_2. We are given a separation oracle for K_1, a membership oracle for K_2 and a point in K_2. The problem is to minimize a convex function over K.

Theorem 14 *Let $K = K_1 \cap K_2$ where K_1 is given by a separation oracle and K_2 by a membership oracle. Let R be the radius of an origin-centered ball containing K. Then, given a point $y \in K_2$ such that the ball of radius r around y is contained in K_2 (i.e., K_2 is centered), the optimization problem over K can be solved in time polynomial in n and $\log \frac{R}{r}$.*

Proof
It is clear that optimization is reducible to feasibility, i.e., we just need to find a point x in K or declare that K is empty. We start with P as K_2 intersected with the ball of radius $2R$ with center y and apply our algorithm. The key procedure, a random walk in P, needs a point in P and a membership oracle for P, both of which we have initially and at each iteration. Using $O^*(n)$ samples we can bring P into near-isotropic position. Also, each subsequent query point z is guaranteed to be in K_2, since it is the average of points from K_2. Thus, the algorithm needs at most $O(n \log \frac{R}{r})$ iterations and $O(n \log \frac{R}{r})$ calls to the separation oracle for K_1. Each iteration (after the first) makes $O^*(n^4)$ calls to the membership oracle for K_2.

To bring the initial set P into isotropic position, we can adapt the "chain of bodies" procedure of [Kannan et al. 1997]. Since K_2 is centered, it contains a ball of radius r around y. Scale up so that $y + rB_n$ is in isotropic position and call it Q (just for convenience; this has the effect of scaling up R also). Now consider $Q' = P \cap (y + 2^{\frac{1}{n}} rB_n)$, i.e., K_2 intersected with the ball of radius $2^{1/n}r$ around y. The volume of Q' is at most twice the volume of Q. This is because

$$Q' = P \cap (y + 2^{\frac{1}{n}} rB_n) \subseteq 2^{\frac{1}{n}}(P \cap (y + rB_n)) = 2^{\frac{1}{n}} Q.$$

Thus, a uniform sample from Q provides a warm start for sampling from Q'. Sample from Q', drawing each sample in $poly(n)$ time and use the samples

[2] It is not known whether Vaidya's algorithm can be extended to solve this problem.

to calculate an affine transformation that would put Q' in isotropic position. Apply the transformation to P. Reset Q to be the transformed Q', y to be its center of gravity and r to be the radius of the ball around y containing Q. Again let Q' be the intersection of P with the ball of radius $2^{1/n}r$ around y. After at most $O(n \log \frac{R}{r})$ such phases, P will be in near-isotropic position.

7 Extensions

The methods of this paper suggest the following *interior point* algorithm for optimization. Suppose we want to maximize $c^T x$ over a full-dimensional convex set K. We start with a feasible solution z_1 in K. Then add the constraint $c^T x \geq c^T z_1$ and let $K := K \cap \{x \mid c^T x \geq c^T z_1\}$. Next we generate N random points y_i in the set K and let z_2 be their average. If $c^T z_2 - c^T z_1 < \epsilon$ we stop; otherwise we set $K := K \cap \{x \mid c^T x > c^T z_2\}$, and continue.

The algorithms seems to merit empirical evaluation; in practice, it might be possible to sample more efficiently than the known worst-case bounds. The algorithm presents a strong motivation to find geometric random walks that mix faster.

Acknowledgement. We thank Adam Kalai, Ravi Kannan, Laci Lovász and Mike Sipser for useful discussions and their encouragement.

References

[Bertsimas and Tsitsiklis 1997] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, 1997.

[Dyer et al. 1991] M. Dyer, A. Frieze and R. Kannan, "A random polynomial time algorithm for estimating the volumes of convex bodies," *Journal of the ACM*, 38, 1–17, 1991.

[Gardner 2002] R. J. Gardner, "The Brunn-Minkowski inequality," *Bulletin of the AMS*, 39(3), 355-405, 2002.

[Grötschel et al. 1981] M. Grötschel, L. Lovász, and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization," *Combinatorica* 1, 169–197, 1981.

[Grötschel et al. 1988] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer, 1988.

[Grunbaum 1960] B. Grunbaum, "Partitions of mass-distributions and convex bodies by hyperplanes," *Pacific J. Math.* 10, 1257–1261, 1960.

[Jerrum et al. 2001] M. Jerrum, A. Sinclair, E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries," *Proc. 33rd Annual ACM STOC*, 712–721, 2001.

[Kalai and Vempala 2002] A. Kalai and S. Vempala, "Efficient algorithms for universal portfolios," *J. Machine Learning Research* 3, 423–440, 2002.

[Kannan and Lovász 1999] R. Kannan and L. Lovász, "Faster mixing via average conductance," *Proc. 31st Annual ACM STOC*, 282–287, 1999.

[Kannan et al. 1995a] R. Kannan, L. Lovász and M. Simonovits. "Isoperimetric problems for convex bodies and a localization lemma," *Discrete Computational Geometry* 13, 541–559, 1995.

[Kannan et al. 1997] R. Kannan, L. Lovász and M. Simonovits. "Random walks and and an $O^*(n^5)$ volume algorithm for convex sets," *Random Structures and Algorithms* 11, 1–50, 1997.

[Kannan et al. 1995b] R. Kannan, J. Mount, S. Tayur, "A randomized algorithm to optimize over certain convex sets," *Math. Oper. Res.* 20(3), 529–549, 1995.

[1982] R. M. Karp and C. H. Papadimitriou, "On linear characterization of combinatorial optimization problems," *Siam J. Comp.*, 11, 620–632, 1982.

[1979] L. G. Khachiyan, "A polynomial algorithm in linear programming," (in Russian), *Doklady Akedamii Nauk SSSR*, 244, 1093–1096, 1979 (English translation: *Soviet Mathematics Doklady*, 20, 191–194, 1979).

[Levin 1965] A. Yu. Levin, "On an algorithm for the minimization of convex functions," (in Russian), *Doklady Akedemii Nauk SSSR*, 160, 1244–1247, 1965 (English translation: *Soviet Mathematics Doklady*, 6, 286-290, 1965).

[Lovász 1998] L. Lovász, "Hit-and-run mixes fast," *Mathematical Programming* 86, 443–461, 1998.

[Lovász and Simonovits 1993] L. Lovász and M. Simonovits, "Random walks in convex bodies and an improved volume algorithm," *Random Structures and Algorithms* 4, 259–412, 1993.

[1981] M. W. Padberg and M. R. Rao, "The Russian method for linear programming III: Bounded integer programming," Research Report 81-39, New York University, Graduate School of Business Administration, New York, 1981.

[Prekopa 1973] A. Prekopa, "On logarithmic concave measures and functions," *Acta Sci. Math. Szeged*, 34, 335–343, 1973.

[Rudelson 1999] M. Rudelson, "Random vectors in the isotropic position," *J. Funct. Anal.* 164, 60–72, 1999.

[Schneider 1993] R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge University Press, Cambridge, 1993.

[Vaidya 1996] P. M. Vaidya, "A new algorithm for minimizing convex functions over convex sets," *Mathematical Programming*, 73, 291–341, 1996.

[Yudin and Nemirovski 1976] D. B. Yudin and A. S. Nemirovski, "Evaluation of the information complexity of mathematical programming problems," (in Russian), Ekonomika i Matematicheskie Metody 12, 128-142, 1976 (English translation: Matekon 13, 2, 3-45, 1976).