Optimization of a Convex Program with a Polynomial Perturbation

Ravi Kannan
Microsoft Research Bangalore

Luis Rademacher*
College of Computing
Georgia Institute of Technology

Abstract

We consider the problem of minimizing a convex function plus a polynomial p over a convex body K. We give an algorithm that outputs a solution x whose value is within $\epsilon \operatorname{range}_K(p)$ of the optimum value, where $\operatorname{range}_K(p) = \sup_{x \in K} p(x) - \inf_{x \in K} p(x)$. When p depends only on a constant number of variables, the algorithm runs in time polynomial in $1/\epsilon$, the degree of p, the time to round K and the time to solve the convex program that results by setting $p = 0$.

Keywords: nonlinear programming, approximation algorithms, polynomial optimization

1 Introduction

We give an algorithm to minimize approximately $f(x)+p(x)$ over a convex body K in $\mathbb{R}^n$ where f is any convex function and p is any polynomial in a constant number of variables. Our solution x satisfies

$$f(x) + p(x) \leq f(x^*) + p(x^*) + \epsilon \operatorname{range}_K(p)$$

where ϵ is a given error parameter, x^* is an optimum solution and

$$\operatorname{range}_K(p) = \sup_{x \in K} p(x) - \inf_{x \in K} p(x).$$

The algorithm runs in time

$$\left(O\left(\frac{kd^2}{\sqrt{\epsilon}} \right) \right)^k T(f, K) + \tilde{T}(K)$$

*Corresponding author. Email: `lrademac@cc.gatech.edu`. Address: 266 Ferst Drive, Atlanta GA 30332, USA.

where k is the number of variables that appear in p, d is the degree of p, $T(f, K)$ is the time to solve the convex program $\min_{x\in K} f(x)$ and $\tilde{T}(K)$ is the time to put K in near-isotropic position (discussed in Section 2).

In situations where p is a "small perturbation", the range of p is small and hence the error the algorithms makes. Also clearly the algorithm generalizes traditional convex optimization.

This paper is inspired by a result of Vavasis [7] for the case when p is quadratic with slightly different error bounds.

2 Preliminaries

For a bounded set $C \subseteq \mathbb{R}^n$, let w_C denote the width of C, that is

$$w_C = \inf_{h\in S_{n-1}} (\sup_{x\in C} h^T x - \inf_{x\in C} h^T x).$$

Let B_n denote the n-dimensional Euclidean unit ball. The volume of B_n is

$$\frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}.$$

A convex body K is said to be in *isotropic position* if its center of gravity is the origin and for any unit vector v, one has

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

It is known that for any convex body there is an affine transformation which puts the body in isotropic position.

We will now quote definitions and results from [2] about the existence of an efficient randomized algorithm that puts a convex body in near isotropic position.

For a convex body K, let $b(K)$ denote its center of gravity.

Definition 1 (2.4 in [2])**.** *We say that a convex body K is in θ-nearly isotropic position ($0 < \theta \le 1$), if*

$$\|b(K)\| \le \theta,$$

and for every vector $v \in \mathbb{R}^n$,

$$(1-\theta)\|v\|^2 \le \frac{1}{\mathrm{vol}(K)} \int_{K-b(K)} (v^T x)^2\, dx \le (1+\theta)\|v\|^2$$

Theorem 2 (Corollary 5.2 in [2])**.** *Let $\theta < 1/2$. If K is in θ-near isotropic position, then*

$$(1-2\theta)B_n \subseteq K \subseteq (1+2\theta)(n+1)B_n.$$

The following result is a specialization of Theorem 2.5 from [2]. There are some mistakes in the statement of that theorem, but a correct statement can be obtained by looking at the results on which it depends (Lemma 5.18 and Theorem 5.20 from [2])

Theorem 3. *There exists a randomized algorithm that, when given numbers $0 < r \leq R$ and access to a membership oracle of a convex body $K \subseteq \mathbb{R}^n$ with $rB_n \subseteq K \subseteq RB_n$, finds an affine transformation A for which AK is in 1/4-nearly isotropic position with probability at least $1 - \delta$. The number of oracle calls is*

$$O^*\left(n^5\left(\log\frac{R}{r}\right)^3\right).$$

Here O^* means that logarithmic factors in n and $1/\delta$ are ignored, where δ is the probability of failure. The improved random walks in [4] imply that one can approximately round a convex body in time $O^*(n^4 \log^c(R/r))$ for some constant c, but we use the results from [2] because they are in a form that is slightly closer to what we need.

3 Optimizing a Convex Program with a Polynomial Perturbation

Formally, for a convex body $K \subseteq \mathbb{R}^n$, a convex function $f : \mathbb{R}^n \to \mathbb{R}$ and a polynomial p that depends on only k variables, let problem (P) be:

$$\begin{aligned} \min g(x) = f(x) + p(x) \\ \text{subject to } x \in K \end{aligned} \qquad (P)$$

The algorithm that we propose to solve (P) is essentially an enumeration over a suitable grid in the space of the variables that appear in the polynomial. It is possible to guarantee a certain quality of approximation because of a known bound on the gradient of a polynomial in a convex set as a function of the range of the polynomial (Theorem 5).

Let $\operatorname{proj} : \mathbb{R}^n \to \mathbb{R}^k$ be the orthogonal projection onto the subspace of the variables that appear in p. Consider a covering $\mathcal{C}$ of $\frac{3}{2}(k+1)B_k$ with cubes having side

$$s = \frac{\pi}{8d^2}\sqrt{\frac{\epsilon}{k}}. \tag{1}$$

The algorithm will put $\operatorname{proj}(K)$ in 1/4-nearly isotropic position by an affine transformation A (see Section 2), which implies that $A\operatorname{proj}(K) \subseteq \frac{3}{2}(k+1)B_k$, and $\mathcal{C}$ is also a covering of $A\operatorname{proj}(K)$. We will take $\mathcal{C}$ to be the set of all cubes with centers in $s\mathbb{Z}^k$ that intersect $\frac{3}{2}(k+1)B_k$. These cubes are contained in $(\frac{3}{2}(k+1) + s\sqrt{k})B_k$, and thus

$$|\mathcal{C}| \leq \left(\frac{O(\sqrt{k})}{s}\right)^k \tag{2}$$

(using that the volume of each cube is s^k and the volume of the ball given in Section 2).

The algorithm uses the following piecewise linear approximation of p: For every cube C with center y_C, define

$$\phi_C(x) = p(A^{-1}y_C) + \nabla p(A^{-1}y_C)^T(x - A^{-1}y_C).$$

Algorithm 1

1. Compute an affine transformation $A : \mathbb{R}^k \to \mathbb{R}^k$ such that $A\,\mathrm{proj}(K)$ is in 1/4-nearly isotropic position.
2. For every cube $C \in \mathcal{C}$, compute an optimal solution x_C to
$$\min_{x \in \mathrm{proj}^{-1} A^{-1}(C) \cap K} f(x) + \phi_C(x).$$
If this problem is infeasible, remove C from $\mathcal{C}$.
3. Let $C^* = \mathrm{argmin}_{C \in \mathcal{C}}\, f(x_C) + p(x_C)$. Output x_{C^*}.

Let $T(f, K)$ be the time to solve the convex program obtained from P when $p = 0$. This can be done in polynomial time with mild assumptions on K and f and their representation [1, 3]. Let $\tilde{T}(K)$ be the time to put the projection of K in 1/4-nearly isotropic position. If K is given by a membership oracle and numbers $0 < r \leq R$ so that $rB_n \subseteq K \subseteq RB_n$, then an efficient algorithm is guaranteed by Theorem 3.

Theorem 4. *Let x^* be an optimum solution of Problem P. Algorithm 1 outputs $x \in K$ satisfying*

$$f(x) + p(x) \leq f(x^*) + p(x^*) + \epsilon\,\mathrm{range}_K(p)$$

in time

$$\left(O\left(\frac{kd^2}{\sqrt{\epsilon}}\right)\right)^k T(f, K) + \tilde{T}(K).$$

Proof. By means of Theorem 2 we have that $B_k/2 \subseteq A\,\mathrm{proj}(K)$, which implies that the width of $A\,\mathrm{proj}(K)$ is at least 1/2. With this, we will now analyze the error that the algorithm makes when replacing p by ϕ_C. We apply Corollary 6 and Taylor's theorem on the multivariate polynomial $q : \mathbb{R}^k \to \mathbb{R}$, $q(x) = p(A^{-1}x)$ satisfying $\mathrm{range}_{A\,\mathrm{proj}(K)}(q) = \mathrm{range}_K(p)$ to get for any cube C and for all $x \in \mathrm{proj}^{-1} A^{-1}(C) \cap K$:

$$\begin{aligned}
|\phi_C(x) - p(x)| &\leq \frac{1}{2} \sup_{z \in K, h \in S_{n-1}} |h^T \nabla^2 q(Az) h| \, \|A\,\mathrm{proj}(x) - y_C\|^2 \\
&\leq \frac{16}{\pi^2} \frac{d^4}{w^2_{A\,\mathrm{proj}(K)}} \mathrm{range}_{A\,\mathrm{proj}(K)}(q) s^2 k \\
&\leq \frac{64}{\pi^2} d^4\, \mathrm{range}_K(p) s^2 k \\
&\leq \epsilon\, \mathrm{range}_K(p).
\end{aligned}$$

This implies $g(x_{C^*}) \leq OPT + \epsilon \operatorname{range}_K(p)$.

The bound on the running time follows from Equations (1) and (2). $\square$

The following result is a slight reformulation of a result by Skalyga [6] (who proved it with constant $4/\pi$, it was proved before with constant 2 by Wilhelmsen [8], [5, Section 6.1.5]). This is a multivariate generalization of a univariate inequality by A. A. Markov [5, Chapter 6, Theorem 1.2.1]

Theorem 5. *Let $K \subseteq \mathbb{R}^n$ be a convex body. Let $p : K \to \mathbb{R}$ be a multivariate polynomial of degree d. Then for any $x \in K$:*

$$\|\nabla p(x)\| \leq \frac{4}{\pi} \frac{d^2 \operatorname{range}_K(p)}{w_K}.$$

If we just used Theorem 5, we could only use a constant approximation in the role of our linear approximation ϕ_C to p and the dependence on ϵ of our algorithm would be $1/\epsilon$. But one can easily use Theorem 5 inductively to get a second order version of it, and be able to use a linear approximation ϕ_C:

Corollary 6. *Let $K \subseteq \mathbb{R}^n$ be a convex body. Let $p : K \to \mathbb{R}$ be a multivariate polynomial of degree d. Then for any $x \in K$ and $h \in S_{n-1}$:*

$$|h^T \nabla^2 p(x) h| \leq \frac{32}{\pi^2} \frac{d^4 \operatorname{range}_K(p)}{w_K^2}.$$

Proof. Use Theorem 5 twice on the polynomials $x \mapsto \nabla p(x)^T h$ and p to get:

$$\begin{aligned}
|h^T \nabla^2 p(x) h| &\leq \frac{4}{\pi} \frac{d^2 \operatorname{range}_K(\nabla p(\cdot)^T h)}{w_K} \\
&\leq \frac{8}{\pi} \frac{d^2 \sup_{x \in K} |\nabla p(x)^T h|}{w_K} \\
&\leq \frac{32}{\pi^2} \frac{d^4 \operatorname{range}_K(p)}{w_K^2}
\end{aligned}$$

$\square$

4 Discussion

Vavasis's paper is crucially based on a result similar to Theorem 5 for quadratic polynomials. Note that the above theorem essentially asserts that a degree d polynomial cannot be wild at any place. The only factor that can make its gradient too large is the width which we bound using isotropic position. An interesting open problem is to extend the results to a larger class of perturbations of convex minimization problems. In effect any function satisfying Theorem 5 type of conclusion would be amenable to this. In a sense with Theorem 5 and efficient methods to find near-isotropic position on hand, the current paper can be viewed as just using them together. But given that polynomials are very general and not too many clean generalizations of convex optimization are known, this records one such while raising the question of possibly others.

Acknowledgments

We would like to thank an anonymous referee for helpful comments and suggestions.

References

[1] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.

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

[3] L. Lovász and S. Vempala. Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 57–68, 2006.

[4] L. Lovász and S. Vempala. Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithm. *Journal of Computer and System Sciences*, 72(2):392–417, 2006.

[5] G. V. Milovanović, D. S. Mitrinović, and T. M. Rassias. *Topics in polynomials: extremal problems, inequalities, zeros.* World Scientific Publishing Co. Inc., River Edge, NJ, 1994.

[6] V. I. Skalyga. Estimates of the derivatives of polynomials on convex bodies. *Proc. Steklov Inst. Math*, 218:372–383, 1997.

[7] S. A. Vavasis. Approximation algorithms for indefinite quadratic programming. *Math. Program.*, 57:279–311, 1992.

[8] D. R. Wilhelmsen. A Markov inequality in several dimensions. *J. Approximation Theory*, 11:216–220, 1974.