OpenAccess

ORIGINAL PAPER

The complexity of optimizing over a simplex, hypercube or sphere: a short survey

Etienne de Klerk

Abstract   We consider the computational complexity of optimizing various classes of continuous functions over a simplex, hypercube or sphere. These relatively simple optimization problems arise naturally from diverse applications. We review known approximation results as well as negative (inapproximability) results from the recent literature.

Keywords   Computational complexity · Global optimization · Linear and semidefinite programming · Approximation algorithms

Mathematics Subject Classification (2000)   68Q25 · 90C30 · 90C60

1 Introduction

Consider the generic global optimization problem:

$$\underline{f} := \min \{ f(x) \ : \ x \in K \}, \tag{1}$$

for some continuous $f : K \mapsto \mathbb{R}$ and compact convex set $K \subset \mathbb{R}^n$, and let

$$\bar{f} := \max \{ f(x) \ : \ x \in K \}.$$

In this short survey we will consider the computational complexity of computing or approximating $\underline{f}$ (or $\bar{f}$) in the case where K is one of the following three sets:

E. de Klerk (✉)
Department of Econometrics and Operations Research, Faculty of Economics and Business Studies,
Tilburg University, 5000 LE Tilburg, The Netherlands
e-mail: e.deklerk@uvt.nl

- the standard (or unit) simplex:

$$\Delta_n := \left\{ x \in \mathbb{R}^n \; : \; \sum_{i=1}^{n} x_i = 1, \; x \geq 0 \right\},$$

- the unit hypercube $[0, 1]^n$,
- the unit sphere: $S_n := \{x \in \mathbb{R}^n \; : \; \|x\| = 1\}$.

Problem (1) arises naturally from diverse applications for these choices of K, and we only mention a few.

For the simplex, and quadratic f, the applications include finding maximum stable sets in graphs, portfolio optimization, game theory, and population dynamics problems [see the review paper by Bomze (2002) and the references therein]. A recent application is the estimation of crossing numbers in certain classes of graphs (de Klerk et al. 2006). Another application is testing matrix copositivity—recall that an $n \times n$ matrix A is called copositive if $\min_{x \in \Delta_n} x^T A x = 0$. Recent seminal work on certifying matrix copositivity was done by Parrilo (2000), and the topic has also enjoyed attention in the more general setting of copositive programming; see e.g. Burer (2006) and the references therein.

The following example is of optimization of a general non-polynomial f over the simplex, that occurs in multivariate interpolation, and in finite element methods (see Hesthaven 1998).

Example 1.1 Given is a finite set of interpolation points $\Theta \subset \Delta_n$. Denote the fundamental Lagrange polynomial associated with an interpolation point $\theta \in \Theta$ by ℓ_θ^Θ. In other words, for $x \in \Theta$:

$$\ell_\theta^\Theta(x) = \begin{cases} 1 & \text{if } x = \theta \\ 0 & \text{else.} \end{cases}$$

For a given $g : \mathbb{R}^n \to \mathbb{R}$, the associated Lagrange interpolant of g with respect to Θ is:

$$L_\Theta(g)(x) := \sum_{\theta \in \Theta} g(\theta) \ell_\theta^\Theta(x).$$

Note that $L_\Theta(g)$ interpolates g at the points in Θ.

The associated Lebesgue constant is defined as

$$\Lambda(\Theta) := \max_{x \in \Delta_n} \sum_{\theta \in \Theta} \left| \ell_\theta^\Theta(x) \right|.$$

The Lebesgue constant is important in bounding the error of approximation, since one can show that

$$\|L_\Theta(g) - g\|_\infty \leq (1 + \Lambda(\Theta)) \|g - p^*\|_\infty,$$

where the norm is the supremum norm on Δ_n, and p^* is the best possible polynomial approximation of g of the same degree as $L_\Theta(g)$.

Thus, to compute the Lebesgue constant for Θ, we should maximize $f(x) = \sum_{\theta \in \Theta} |l_\theta(x)|$ over the simplex Δ_n.

For $K = [0, 1]^n$ and quadratic f, the examples include the maximum cut problem in graphs (see below). For general f, it includes many engineering design problems where simple upper and lower bounds on the variables are given, and no other constraints are present. These problems are sometimes referred to as "box constrained global optimization problems".

For $K = S_n$ (the sphere), problem (1) becomes a minimal eigenvalue problem for a quadratic form $f(x) = x^T Q x$, by the Rayleigh–Ritz theorem. For general quadratic functions it contains the trust region problem, that appears in many nonlinear programming algorithms as a sub-problem. For general forms (homogeneous polynomials), it contains the classical problem of deciding whether a given form is positive semidefinite; see e.g. Parrilo (2000).

2 Models of computation and complexity

Most of the optimization problems we will consider will be NP-hard (in a sense to be explained below), and we will therefore be interested in approximating the optimal values as well as possible in polynomial time.

One encounters two basic models of computation in the optimization literature, namely the bit (or Turing machine) model (see e.g. Garey and Johnson 1979), and the real number (or BSS) model (Blum et al. 1989). Both models may be extended to include randomized algorithms, by allowing a unbiased coin-toss as an operation, see e.g. Motwani and Raghavan (1995).

2.1 The real number (BSS) model in continuous optimization

In the real number model, the input is a finite set of real numbers, and an arithmetic operation between two real numbers requires one unit of time. A typical extension of this model, used in optimization, is the so-called *black box optimization model*. Here, typically the time required for evaluation of f is viewed as requiring one unit of time as well [see e.g. Chap. 6 in Vavasis (1991) as well as Nemirovsky and Yudin (1983)]. This framework allows one to treat very general optimization problems. The black box models falls in the framework of information-based complexity, see e.g. Traub et al. (1988). In particular, *admissible information operations* are specified beforehand, like evaluating f or a sub-gradient of f at unit cost. The complexity of approximating the solution to a problem in this model is expressed in terms of the dimension n, the required accuracy, and the number of admissible information operations that were performed in obtaining the approximation.

2.2 The Turing (bit) model in continuous optimization

In the bit model, the input of a Turing machine must be encoded as a string of symbols. This is clearly restrictive when dealing with a large class of objective functions f. Thus

we are more-or-less limited to polynomials and rational functions which may be coded in terms of their coefficients in some basis, provided these coefficients are rational.

In this paper, when dealing with the bit model, we will only consider the class of polynomials with rational coefficients. We do note that it is possible to deal with more general functions by using the most general theory of polynomial-time computable functions (see e.g. Weirauch 1998), but this extension in beyond the scope of this survey.

2.3 Complexity of continuous optimization

We now consider the input data that defines an instance of problem (1) for the two models of computation. Clearly, an instance is determined by the function f and the set K.

In the black box model, f enters the input as a black box subroutine that may be called at unit cost. The set K is represented in the input by the number of variables n.

In the bit model, we need an encoding of f in terms of an alphabet. Since we will only deal with polynomials in the bit model, this encoding can be a binary encoding of the coefficients of the polynomial in the standard monomial basis.

Recall that the definition of the complexity class NP is the class of "yes/no" decision problems where a "yes-instance" has a certificate that may be verified in polynomial time. In combinatorial optimization, one does not need to distinguish between optimization problems and "yes/no" decision problems from the complexity viewpoint, but in continuous optimization the situation is more complicated.

If problem (1) is such that suitable necessary and sufficient optimality conditions are known via duality theory, then one may reformulate problem (1) as a "yes/no" decision problem. For example, one may formulate a linear programming (LP) problem (i.e. problem (1) in the special case where f is linear and K a polytope) as: "do there exist primal and dual feasible solutions with zero duality gap?" In this sense, LP is in NP in both the bit and real number models. In fact, the LP problem is known to be in P in the bit model by a celebrated result of Khachiyan (1979). Likewise, convex quadratic programming (where f is convex quadratic and K a polytope) is known to be in P in the bit model (see e.g. Chap. 3 in Vavasis 1991). However, the complexity of LP is not known in the real number model.

In general, though, problem (1) does not allow such a reformulation as a decision problem, and the complexity of continuous optimization is therefore often studied in the following "decision version" setting.

Definition 2.1 (*Decision version of problem* (1)) Given a value γ, as well as input f and K (coded as described above), does there exist a $\bar{x} \in K$ such that $f(\bar{x}) \leq \gamma$?

Notice that the value γ is part of the input of this decision problem, and it must be rational in the bit model.

When we make a statement of the type that problem (1) is NP-hard for some class of functions and some choice of K, we usually refer to the decision version of the problem.

This decision problem is clearly in NP for the real number model, but the situation is much more difficult for the bit model. It is known, but already difficult to show, that general quadratic programming is in NP in the bit model (see e.g. Chap. 4 in Vavasis 1991). The author is not aware of more general results of this type.

2.4 Notions of approximation

One has to be careful when defining the notion of an approximation to an optimal solution of problem (1). The reason is that we usually do not know the range $\bar{f} - \underline{f}$ of function values on K in advance. If $\bar{f} - \underline{f}$ is small compared to a given ϵ, then it is not satisfactory to only compute some $x \in K$ with the property that $f(x) - \underline{f} < \epsilon$. It is therefore better to find an $x \in K$ such that $f(x) - \underline{f} < \epsilon(\bar{f} - \underline{f})$, since we then know that $f(x)$ belongs to the ϵ fraction of lowest function values.

The next definition is based on this idea, and has been used by several authors, including Ausiello et al. (1980), Bellare and Rogaway (1995), Bomze and De Klerk (2002), de Klerk et al. (2007b), Nesterov et al. (2000), and Vavasis (1992).

Definition 2.2 A value ψ_ϵ is called a $(1 - \epsilon)$-approximation of $\underline{f}$ for a given $\epsilon \in [0, 1]$ if

$$|\psi_\epsilon - \underline{f}| \le \epsilon(\bar{f} - \underline{f}). \tag{2}$$

If we replace condition (2) by the condition

$$|\psi_\epsilon - \underline{f}| \le \epsilon, \tag{3}$$

then we speak of a $(1 - \epsilon)$-approximation of $\underline{f}$ *in the weak sense*.

The following definitions are essentially from de Klerk et al. (2007b), and are consistent with the corresponding definitions in combinatorial optimization (see e.g. Garey and Johnson 1979, Sect. 6.1).

Definition 2.3 (*Polynomial time approximation algorithm*) Fix $\epsilon > 0$ and a class of continuous functions on K, say $\mathcal{F}$. An algorithm $\mathcal{A}_\epsilon$ is called a polynomial time $(1 - \epsilon)$-approximation algorithm for problem (1) for the function class $\mathcal{F}$, if the following holds:

1. For any instance $f \in \mathcal{F}$, $\mathcal{A}_\epsilon$ takes f and K as input and computes an $x_\epsilon \in K$ such that $f(x_\epsilon)$ is a $(1 - \epsilon)$-approximation of $\underline{f}$;
2. the number of operations required for the computation of x_ϵ is bounded by a polynomial in n, and

   - in the bit model: in the bit size of f;
   - in the real number model: in the number of operations performed in evaluating f and its subgradients/derivatives where applicable;
   - in the black box model: in the number of admissible information operations (like evaluating f) that are performed.

Definition 2.4 (*PTAS*) We say that problem (1) allows a polynomial time approximation scheme (PTAS) for a function class $\mathcal{F}$, if for every $\epsilon > 0$, it has a polynomial time $(1 - \epsilon)$-approximation algorithm.

Definition 2.5 (*FPTAS*) An algorithm $\mathcal{A}$ is called a fully polynomial time approximation scheme (FPTAS) for problem (1) for the function class $\mathcal{F}$, if the following holds:

1. For any instance $f \in \mathcal{F}$ and any $\epsilon > 0$, $\mathcal{A}$ takes f, ϵ, and K as input and computes an $x_\epsilon \in K$ such that $f(x_\epsilon)$ is a $(1 - \epsilon)$-approximation of $\underline{f}$;
2. the number of operations required for the computation of x_ϵ is bounded by a polynomial in n, $1/\epsilon$, and
   - in the bit model: in the bit size of f;
   - in the real number model: in the number of operations performed in evaluating f and its subgradients/derivatives where applicable;
   - in the black box model: in the number of admissible information operations (like evaluating f) that are performed.

These definitions can be adapted in an obvious way for maximization problems, or if the approximations are in the weak sense of (3).

3 Inapproximability results

We first review negative approximation results for problem (1). We will see that, in a well-defined sense, optimization over the hypercube is much harder than over the simplex, while the complexity of optimization over a sphere is somewhere in between. In what follows, if the model of computation is not specified, then the bit model is assumed.

3.1 The case of the simplex

If $K = \Delta_n$, then computing $\underline{f}$ is an NP-hard problem, already for quadratic polynomials, as it contains the maximum stable set problem as a special case. (Recall that a stable set of a graph is a subset of vertices such that no two vertices in this subset are adjoint, and that a maximum stable set is a stable set of maximum cardinality.)

Indeed, let G be a graph with adjacency matrix A and let I denote the identity matrix; then the maximum size $\alpha(G)$ of a stable set in G can be expressed as

$$\frac{1}{\alpha(G)} = \min_{x \in \Delta_{|V|}} x^T (I + A) x$$

by a theorem of Motzkin and Straus (1965). Moreover, this problem cannot have a FPTAS, unless all problems in NP can be solved in randomized polynomial time; this is due to a inapproximability result for the maximum stable set problem by Håstad (1999).

Before stating the result, recall that ZPP is the class of problems that may be solved by a randomized (Las Vegas) algorithm running in expected polynomial time; for a more precise definition see Motwani and Raghavan (1995).

Theorem 3.1 (Håstad) *Unless NP = ZPP, one cannot approximate $\alpha(G)$ to within a factor $|V|^{(1-\epsilon)}$ for any $\epsilon > 0$.*

Corollary 3.1 *Unless NP = ZPP, there is no FPTAS for problem* (1) *for the class of quadratic functions and $K = \Delta_n$.*

3.2 The case of the hypercube

If $K = [0, 1]^n$ and f quadratic, then problem (1) contains the maximum cut problem in graphs as a special case. Indeed, for a graph $G = (V, E)$ with Laplacian matrix L, the size of the maximum cut is given by (see Goemans and Williamson 1995; Nesterov 1998):

$$|\text{maximum cut}| = \max_{x \in [-1,1]^{|V|}} \frac{1}{4} x^T L x = \max_{x \in [0,1]^{|V|}} \frac{1}{4} (2x - e)^T L (2x - e), \quad (4)$$

where e is the all-ones vector. The maximum cut problem does not allow a PTAS, as the following theorem shows.

Theorem 3.2 (Håstad 2001) *Unless $P = NP$, there is no polynomial time 16/17-approximation algorithm for the maximum cut problem* (4).

It follows that problem (1) does not allow a PTAS for any class of functions that includes the quadratic polynomials if $K = [0, 1]^n$.

A related negative result is due to Bellare and Rogaway (1995), who proved that if $P \neq NP$ and $\epsilon \in (0, 1/3)$, there is no polynomial time $(1 - \epsilon)$-approximation algorithm in the weak sense for the problem of minimizing a polynomial of total degree $d \geq 2$ over all sets of the form $K = \{x \in [0, 1]^n \mid Ax \leq b\}$.

A general (negative) result in the black box model is the following.

Theorem 3.3 (Theorem 6.1 in Vavasis 1991) *For $K = [0, 1]^n$ and the class of f with Lipschitz constant L, the worst case complexity of computing a $(1 - \epsilon)$-approximation of $\underline{f}$ in the weak sense is between*

$$\left(\frac{L}{2\epsilon}\right)^n \quad and \quad \left(\frac{L}{2\epsilon} + 1\right)^n$$

in the black box model.

More results of this type may be found in Novak (1988) and Horn (2006).

3.3 The case of the sphere

Nesterov (2003) showed that maximizing a cubic form (homogeneous polynomial) on the unit sphere is an NP-hard problem, using a reduction from the maximum stable set problem.

Theorem 3.4 (Nesterov) *Consider a graph $G = (V, E)$ with stability number $\alpha(G)$. One has*

$$\sqrt{1 - \frac{1}{\alpha(G)}} = 3\sqrt{3} \max_{\|x\|^2 + \|y\|^2 = 1} \sum_{\substack{i < j \\ \{i,j\} \notin E}} y_{ij} x_i x_j.$$

Note that this indeed involves maximizing a (square free) form of degree 3 in the variables x and y over the unit sphere. Also note that the number of variables is polynomial in $|V|$, since the x variables correspond to the vertices of G, and the y variables correspond to the edges of the complement of G.

In view of the inapproximability result for the maximum stable set problem in Theorem 3.1, we have the following corollary.

Corollary 3.2 *Unless $NP = ZPP$, there is no FPTAS for minimizing square free degree 3 forms over the unit sphere.*

4 Approximation results

4.1 The case of the simplex

We now consider the complexity of (approximately) solving problem (1) for $K = \Delta_n$.

Easy cases

Problem (1) can be solved in polynomial time or allows a PTAS for the following classes of functions:

- f is concave; in this case the global minimum of f is attained at one of the n extreme points of Δ_n;
- f is convex quadratic; here problem (1) is a special case of the general convex quadratic programming problem, that may be solved in polynomial time in the bit model; see e.g. Chap. 3 in Vavasis (1991);
- in the black box model, if f is convex with Lipschitz constant L and both f and its (sub)gradient may be evaluated with unit cost (by a black box) at any $x \in \Delta_n$, then the ellipsoid algorithm of Yudin and Nemirovski (see e.g. Nemirovsky and Yudin 1983) yields an $\bar{x} \in \Delta_n$ such that $f(\bar{x}) - \underline{f} \leq \epsilon$ in at most

$$2n(n + 1)(\ln(L/\epsilon) + \ln n) + 2$$

iterations. Each iteration requires one function evaluation and one (sub)gradient evaluation (see Sect. 6.3 in Vavasis 1991).

This remains true if K is the unit hypercube or unit ball. In summary, we may state this as a theorem.

Theorem 4.1 (Yudin-Nemirovski) *In the black box model, problem* (1) *allows a FPTAS in the weak sense for the class of convex, Lipschitz continuous functions with a given (fixed) Lipschitz constant, for $K = \Delta_n$, $K = [0, 1]^n$ or K the unit ball in $\mathbb{R}^n$.*

Results for polynomial f

Bomze and De Klerk (2002) showed that, for $K = \Delta_n$ and f quadratic, problem (1) allows a PTAS. One of the PTAS algorithms that they considered is particularly simple: it evaluates f on the regular grid

$$\Delta(n, m) := \left\{ x \in \Delta_n \ : \ mx \in \mathbb{N}_0^n \right\},$$

and returns the lowest value. In other words, it computes the value

$$f_{\Delta(n,m)} := \min_{x \in \Delta(n,m)} f(x).$$

Note that $|\Delta(n, m)| = \binom{n+m-1}{m}$ which is a polynomial in n for fixed m.
    Bomze and De Klerk (2002) showed the following.

Theorem 4.2 (Bomze, De Klerk) *Let f be quadratic. One has*

$$f_{\Delta(n,m)} - \underline{f} \le \frac{1}{m}(\bar{f} - \underline{f}) \tag{5}$$

for any $m \ge 1$.

Corollary 4.1 (Bomze, De Klerk) *There exists a PTAS for minimizing quadratic polynomials over the unit simplex.*

    This PTAS result was extended to polynomials of fixed degree by de Klerk et al. (2007b). Earlier, related results were obtained by Faybusovich (2004).

Theorem 4.3 (de Klerk et al. 2007b) *Let $f(x)$ be a form of degree d and $r \ge 0$ an integer. Then,*

$$f_{\Delta(n,r+d)} - \underline{f} \le (1 - w_r(d)) \binom{2d-1}{d} d^d (\bar{f} - \underline{f}), \tag{6}$$

where

$$w_r(d) := \frac{(r+d)!}{r!(r+d)^d} = \prod_{i=1}^{d-1} \left(1 - \frac{i}{r+d} \right). \tag{7}$$

One can verify that

$$1 - \binom{d}{2}\frac{1}{r+d} \le w_r(d) \le 1,$$

which implies that $\lim_{r\to\infty} w_r(d) = 1$.

Corollary 4.2 (de Klerk et al. 2007b) *Fix $d \in \mathbb{N}$. There exists a PTAS for minimizing forms of degree d over the unit simplex.*

The optimization algorithm that evaluates f on a regular grid was studied in detail by Nemirovsky and Yudin (1983) in a more general context in the black box model. In particular, they showed that this algorithm has the best worst case complexity in some situations. The interested reader is referred to Nemirovsky and Yudin (1983) and to Traub et al. (1988), Chap. 8.

There also exist more sophisticated (and practical) PTAS algorithms for minimizing forms of fixed degree over the simplex that employ linear or semidefinite programming. For example, the authors of de Klerk et al. (2007b) consider

$$f_{\min,\mathrm{LP}}^{(r)} := \max \lambda \ \text{ s.t. the polynomial } \left(\sum_{i=1}^{n} x_i\right)^r \left(f(x) - \lambda\left(\sum_{i=1}^{n} x_i\right)^d\right)$$
$$\text{has nonnegative coefficients} \quad (r = 0, 1, \ldots)$$

Note these bounds may be computed using LP, and this computation may be performed in polynomial time when r and d are fixed. Moreover, it is shown in de Klerk et al. (2007b) that

$$0 \le \underline{f} - f_{\min,\mathrm{LP}}^{(r)} \le \left(\frac{1}{w_r(d)} - 1\right)\binom{2d-1}{d}d^d(\bar{f} - \underline{f}), \quad (r = 0, 1, \ldots),$$

where $w_r(d)$ was defined in (7). Thus the values $f_{\min,\mathrm{LP}}^{(r)}$ also provide a PTAS, since

$$\lim_{r\to\infty} w_r(d) = 1.$$

One may also define semidefinite programming (SDP) based bounds that are at least as strong as the LP ones:

$$f_{\min,\mathrm{SDP}}^{(r)} := \max \lambda \ \text{ s.t. the polynomial } \left(\sum_{i=1}^{n} x_i^2\right)^r \left(f(x \circ x) - \lambda\left(\sum_{i=1}^{n} x_i^2\right)^d\right)$$
$$\text{is a sum of squares of polynomials} \quad (r = 0, 1, \ldots),$$

where "$\circ$" is the component-wise (Hadamard) product. Note that $\underline{f} \ge f_{\min,\mathrm{SDP}}^{(r)} \ge f_{\min,\mathrm{LP}}^{(r)}$.

Results for non-polynomial f

Recently, de Klerk et al. (2007a) derived approximation results for (not necessarily polynomial) functions that meet a Lipschitz condition of given order.

Once again, the underlying algorithm is simply the evaluation of f on a suitable regular grid.

Before defining this class of functions, recall that the modulus of continuity of f on a compact convex set K is defined by

$$\omega(f, \delta) := \max_{\substack{x, y \in K \\ \|x-y\| \le \delta}} |f(x) - f(y)|.$$

We now define the class $\mathrm{Lip}_L(\alpha)$ of functions that meet the Lipschitz condition of given order $\alpha > 0$ with respect to a given constant $L > 0$:

$$\mathrm{Lip}_L(\alpha) := \left\{ f \in C(\Delta_n) \ : \ \omega(f, \delta) \le \delta^\alpha L \right\}. \tag{8}$$

This condition is also called a Hölder continuity condition. (Some authors reserve the term "Lipschitz" for the case $\alpha = 1$.)

Theorem 4.4 (De Klerk, Elabwabi, Den Hertog) *Let $\epsilon > 0$, $\alpha > 0$, and $L > 0$ be fixed and assume $f \in Lip_L(\alpha)$ [see (8)].*
Then, for

$$m = \left\lceil \left(\frac{2L}{\epsilon} \right)^{\frac{1}{2\alpha}} \right\rceil,$$

one has

$$f_{\Delta(n,m)} - \underline{f} \le \epsilon,$$

where $f_{\Delta(n,m)} := \min_{x \in \Delta(n,m)} f(x)$ requires a number of function evaluations that is bounded by a polynomial in n.

This theorem implies a PTAS *in the weak sense* for minimizing functions from the class $\mathrm{Lip}_L(\alpha)$ over Δ_n, for fixed L and α, in the black box model.

If we compare this result with the corresponding result for the hypercube (Theorem 3.3), we see that the worst-case complexity is exponential in n for the hypercube, but polynomial in n for the simplex. This illustrates again the stark difference in complexity between the two cases.

Note that Theorem 4.4 does not imply Corollary 4.2. de Klerk et al. (2007a) also identified a further class of functions that allow a PTAS in the black box model. This class includes the polynomials of fixed degree, and is defined in terms of suitable bounds on the higher order derivatives of f. The interested reader is referred to de Klerk et al. (2007a) for more details.

It is still an open question to completely classify the classes of functions that allow a PTAS.

4.2 The case of the hypercube

For the maximum cut problem there is a celebrated polynomial time 0.878-approximation algorithm due to Goemans and Williamson (1995), who suggested the following SDP relaxation of the maximum cut problem (4):

$$|\text{maximum cut}| \leq \overline{OPT} := \max_{X} \left\{ \frac{1}{4}\text{trace}(LX) \mid \text{diag}(X) = e, \ X \succeq 0 \right\}, \quad (9)$$

where e is the all-ones vector, and $X \succeq 0$ means that X is a symmetric positive semidefinite matrix.

Goemans and Williamson (1995) also devised a randomized rounding scheme that uses the optimal solution of (9) to generate cuts in the graph. Their algorithm produces a cut of cardinality at least $0.878 \ \overline{OPT} \geq 0.878|\text{maximum cut}|$.

Related approximation results for quadratic optimization over a hypercube were given by Nesterov (1998) and Nesterov et al. (2000). In particular, Nesterov showed the following.

Theorem 4.5 (Nesterov 1998) *There exists a* (*randomized*) *polynomial time* $2/\pi$ *approximation algorithm for the problem of maximizing a convex quadratic function over* $[0, 1]^n$.

Notice that the objective function in the maximum cut problem (4) is convex quadratic, since the Laplacian matrix of a graph is always positive semidefinite. Thus the theorem by Nesterov covers a larger class of problems than maximum cut, but the $2/\pi$ constant is significantly lower than the 0.878 obtained by Goemans and Williamson.

4.3 The case of the sphere

The complexity of optimization over the sphere is still relatively poorly understood, compared to the simplex or hypercube. To be more precise, there still is a big gap between approximation and inapproximability results.

Easy case: quadratic optimization over the unit ball

Consider the problem of minimizing a quadratic function $f(x) = x^T B x + 2a^T x + \alpha$ over the unit ball $\|x\| \leq 1$. By the S-procedure of Yakubovich (1973) (see also Polyak 1998, and the recent survey Pólik and Terlaky 2007) this problem may be rewritten as a semidefinite program as follows.

$$\min_{\|x\| \leq 1} f(x) = \max_{\beta, \tau \geq 0} \beta$$

subject to

$$\begin{bmatrix} B & a \\ a^T & \alpha - \beta \end{bmatrix} \succeq \tau \begin{bmatrix} -I & 0 \\ 0^T & 1 \end{bmatrix}.$$

In fact, one may show (see Sect. 4.3 in Vavasis 1991), that the problem is polynomially solvable in the bit model. [This is a stronger statement, since the complexity of SDP is only known to be either in $NP \cap co - NP$ or not in $NP \cup co - NP$ in the bit model, and to be in $NP \cap co - NP$ in the real number model Ramana (1997).]

Results for polynomial objective functions

The result in Corollary 4.2 implies that there exists a PTAS for minimizing *even* forms of fixed degree on the unit sphere. (Recall that a form is called even if all exponents are even.)

Recently, Barvinok (2007) has proved another partial result: one may derive a randomized PTAS for maximizing a form on the sphere for a special class of forms called (δ, N)-focused forms.

Definition 4.1 (Barvinok) Assume f is a form of degree d. Fix a number $0 < \delta \leq 1$ and a positive integer $N > d$. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is (δ, N)-focused if there exist N unit vectors $c_1, \ldots, c_N$ in $\mathbb{R}^n$ such that $c_i^T c_j \geq \delta$ for all $i, j \in \{1, \ldots, N\}$, such that the form f can be written as

$$f(x) = \sum_{\substack{I \subset \{1,\ldots,N\} \\ |I|=d}} \alpha_I \prod_{i \in I} c_i^T x,$$

for some nonnegative scalars α_I.

Theorem 4.6 (Barvinok) *There exists an absolute constant $\gamma > 0$ with the following property. For any $\delta > 0$, for any positive integer N, for any (δ, N)-focused form $f : \mathbb{R}^n \to \mathbb{R}$ of degree d, for any $\epsilon > 0$, and any positive integer*

$$k \geq \gamma \epsilon^{-2} \delta^{-2} \ln(N + 2),$$

the inequality

$$(1 - \epsilon)^{d/2} \max_{x \in S_n \cap L} f(x) \leq \left(\frac{k}{n}\right)^{d/2} \bar{f} \leq (1 - \epsilon)^{-d/2} \max_{x \in S_n \cap L} f(x)$$

holds with probability at least $2/3$ for a random k-dimensional subspace $L \subset \mathbb{R}^n$.

One may solve the problem $\max_{x \in S_n \cap L} f(x)$ in polynomial time using techniques from computational algebraic geometry, since L is of fixed dimension, and therefore the number of variables in the resulting optimization problem is in fact fixed. Thus we obtain the following corollary.

Corollary 4.3 (Barvinok) *Fix $\delta > 0$, $N \in \mathbb{N}$, and $d \in \mathbb{N}$. There exists a (randomized) PTAS for minimizing (δ, N)-focused forms of degree d over the unit sphere.*

It is still an open question whether this result can be extended to all forms of fixed degree.

5 Conclusion and discussion

Approximation algorithms have been studied extensively for combinatorial optimization problems, but have not received the same attention for NP-hard continuous optimization problems. Indeed, most of the early, seminal results on approximation in continuous optimization have focused on the convex case, and most of the results described in this survey for non-convex global optimization were obtained in the last decade.

There is also not much computational experience yet with approximation algorithms for nonlinear programming in the non-convex case. The only significant exception so far is for semidefinite programming relaxations for quadratic optimization on the simplex or the hypercube.

It is therefore the hope of this author that this relatively young research area will attract both theoretically and computationally minded researchers.

References

Ausiello G, D'Atri A, Protasi M (1980) Structure preserving reductions among convex optimization problems. J Comput Syst Sci 21:136–153

Barvinok (2007) A integration and optimization of multivariate polynomials by restriction onto a random subspace. Found Comput Math (to appear)

Bellare M, Rogaway P (1995) The complexity of approximating a nonlinear program. Math Program 69:429–441

Blum L, Shub M, Smale S (1989) On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. Bull Am Math Soc (NS) 21(1):1–46

Bomze IM (2002) Regularity versus degeneracy in dynamics, games, and optimization: a unified approach to different aspects. SIAM Rev 44(3):394–414

Bomze I, De Klerk E (2002) Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. J Glob Optim 24(2):163–185

Burer S (2006) On the copositive representation of binary and continuous nonconvex quadratic programs. Manuscript. http://dollar.biz.uiowa.edu/~burer/papers/021-copos.pdf

Faybusovich L (2004) Global optimization of homogeneous polynomials on the simplex and on the sphere. In: Floudas C, Pardalos P (eds) Frontiers in global optimization. Kluwer, Dordrecht, pp 109–121

Garey MR, Johnson DS (1979) Computers and intractability: a guide to the theory of NP-completeness. W.H. Freeman and Company, Publishers, San Fransisco

Goemans MX, Williamson DP (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. J ACM 42(6):1115–1145

Håstad J (1999) Clique is hard to approximate within $|V^{1-\epsilon}|$. Acta Math 182:105–142

Håstad J (2001) Some optimal inapproximability results. J ACM 48:798–859

Hesthaven JS (1998) From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. SIAM J Numer Anal 35(2):655–676

Horn M (2006) Optimal algorithms for global optimization in case of unknown Lipschitz constant. J Complex 22:50–70

Khachiyan L (1979) A polynomial time algorithm in linear programming. Sov Math Dokl 20:191–194

de Klerk E, Maharry J, Pasechnik DV, Richter B, Salazar G (2006) Improved bounds for the crossing numbers of $K_{m,n}$ and K_n. SIAM J Discrete Math 20:189–202

de Klerk E, den Hertog D, Elabwabi G (2007a) On the complexity of optimization over the standard simplex. Eur J Oper Res (to appear)

de Klerk E, Laurent M, Parrilo P (2007b) A PTAS for the minimization of polynomials of fixed degree over the simplex. Theor Comput Sci (to appear)

Motwani R, Raghavan P (1995) Randomized algorithms. Cambridge University Press, Cambridge

Motzkin TS, Straus EG (1965) Maxima for graphs and a new proof of a theorem of Túran. Can J Math 17:533–540

Nemirovsky AS, Yudin DB (1983) Problem complexity and method efficiency in optimization. Wiley–Interscience, New York

Nesterov YE (1998) Semidefinite relaxation and nonconvex quadratic optimization. Optim Methods Softw 9:141–160

Nesterov YE (2003) Random walk in a simplex and quadratic optimization over convex polytopes. CORE Discussion Paper 2003/71, CORE-UCL

Nesterov YE, Wolkowicz H, Ye Y (2000) Semidefinite programming relaxations of nonconvex quadratic optimization. In: Wolkowicz H, Saigal R, Vandenberghe L (eds) Handbook of semidefinite programming. Kluwer, Norwell, pp 361–419

Novak E (1988) Deterministic and stochastic error bounds in numerical analysis. In: Lecture Notes in Mathematics, vol. 1349. Springer, Heidelberg

Parrilo PA (2000) Structured semidefinite programs and semi-algebraic geometry methods in robustness and optimization. PhD Thesis, California Institute of Technology, Pasadena, California, USA. Available at: http://www.cds.caltech.edu/~pablo/

Pólik I, Terlaky T (2007) A survey of the S-lemma. SIAM Rev (to appear)

Polyak BT (1998) Convexity of quadratic transformations and its use in control and optimization. J Optim Theory Appl 99:553–583

Ramana M (1997) An exact duality theory for semidefinite programming and its complexity implications. Math Program Ser B 77(2):129–162

Traub JF, Wasilkowski GW, Woźniakowski H (1988) Infomation-based complexity. Academic Press, Inc., New York

Vavasis SA (1991) Nonlinear optimization: complexity issues. Oxford University Press, New York

Vavasis SA (1992) Approximation algorithms for concave quadratic programming. In: Floudas CA, Pardalos P (eds) Recent advances in global optimization. Princeton University Press, New Jersey, pp 3–18

Weirauch K (1998) Computable analysis. Springer, Heidelberg

Yakubovich VA (1973) The S-procedure and duality theorems for nonconvex problems of quadratic programming, vol. 1. Vestnik Leningrad. University, pp 81–87