ORF 523	LECTURE 14	Spring 2017, Princeton University
Instructor: A.A. Ahmadi		
Scribe: G. Hall		Thursday, April 20, 2017

When in doubt on the accuracy of these notes, please cross check with the instructor's notes, on `aaa.princeton.edu/orf523`*. Any typos should be emailed to gh4@princeton.edu.*

In nonconvex optimization, it is common to aim for locally optimal solutions since finding global solutions can be too computationally demanding. In this lecture, we show that aiming for local solutions can be too ambitious also. In particular, we will show that even testing if a given candidate feasible point is a local minimum of a quadratic program (QP) (subject to linear constraints) is NP-hard. This goes against the somehow wide-spread belief that local optimization is easy.

We present complexity results for deciding both strict and nonstrict local optimality. In Section 1, we show that testing strict local optimality in unconstrained optimization is hard, even for degree-4 polynomials. We then show in Section 2 that testing if a given point is a local minimum of a QP is hard. The key tool used in deriving this latter result is a nice theorem from algebraic combinatorics due to Motzkin and Straus.

1 Strict local optimality in unconstrained optimization

In this section, we show that testing strict local optimality in the unconstrained case is hard even for low-degree polynomials.

Recall the definition of a strict local minimum: A point $\bar{x} \in \mathbb{R}^n$ is an unconstrained strict local minimum of a function $p : \mathbb{R}^n \to \mathbb{R}$ if $\exists \epsilon > 0$ such that $p(\bar{x}) < p(x)$ for all $x \in B(\bar{x}, \epsilon), x \neq \bar{x}$, where $B(\bar{x}, \epsilon) := \{x| \; ||x - \bar{x}|| \leq \epsilon\}$.

Denote by STRICT LOCAL-4 the following decision problem: Given a polynomial p of degree 4 (with rational coefficients) and a point $\bar{x} \in \mathbb{R}^n$ (with rational coordinates), is $\bar{x}$ an unconstrained strict local minimum of p?

Theorem 1. *STRICT LOCAL-4 is NP-hard.*

Proof: In the previous lecture, we showed that POLYPOS-4 is NP-hard. Recall that POLYPOS-4 is the following problem: Given a polynomial p of degree 4, decide if $p(x) > 0 \; \forall x \in \mathbb{R}^n$.

We will show that POLYPOS-4 reduces to STRICT LOCAL-4. Given a polynomial p of degree 4 with rational coefficients, we want to construct a degree-4 polynomial q with rational coefficients, and a rational point $\bar{x}$ such that

$$p(x) > 0, \ \forall x \in \mathbb{R}^n \Leftrightarrow \bar{x} \text{ is a strict local min for } q.$$

To obtain q, we will derive the "homogenized version" of p. Given $p := p(x)$ of degree d, we define its homogenized version as

$$p_h(x, y) := y^d p\left(\frac{x}{y}\right). \tag{1}$$

This is a homogeneous polynomial in $n+1$ variables $x_1, \ldots, x_n, y$. Here is an example in one variable:

$$\begin{aligned} p(x) &= x^4 + 5x^3 + 2x^2 + x + 5 \\ p_h(x, y) &= x^4 + 5x^3y + 2x^2y^2 + xy^3 + 5y^4. \end{aligned}$$

Note that p_h is indeed homogeneous as it satisfies $p_h(\alpha x, \alpha y) = \alpha^d p_h(x, y)$. Moreover, observe that we can get the original polynomial p back from p_h simply by setting $y = 1$:

$$p_h(x, 1) = p(x).$$

The following lemma illustrates why we are considering the homogenized version of p.

Lemma 1. *The point $x = 0$ is a strict local minimum of a homogeneous polynomial q if and only if $q(x) > 0, \forall x \neq 0$.*

Proof. ($\Leftarrow$) For any homogeneous polynomial q, we have $q(0) = 0$. Since by assumption, $q(x) > 0, \ \forall x \neq 0$, then $x = 0$ is a strict global minimum for q and hence also a strict local minimum for q.
($\Rightarrow$) If $x = 0$ is a strict local minimum of q, then $\exists \epsilon > 0$ such that $q(0) = 0 < q(x)$ for all $x \in B(0, \epsilon), x \neq 0$. By homogeneity, this implies that $q(x) > q(0) = 0, \ \forall x \in \mathbb{R}^n, \ x \neq 0$. Indeed, let $x \notin B(0, \epsilon)$, so $x \neq 0$. Then define

$$\tilde{x} := \epsilon \frac{x}{||x||}.$$

Notice that $\tilde{x}$ in $B(0, \epsilon)$ and $\frac{||x||}{\epsilon} > 1$. We get

$$q(x) = q\left(\frac{||x||}{\epsilon}\tilde{x}\right) = \frac{||x||^d}{\epsilon^d} q(\tilde{x}) > q(0).$$

$\square$

It remains to show that it is NP-hard to test positivity for degree-4 *homogeneous* polynomials. The proof we gave last lecture for NP-hardness of POLYPOS-4 (via a reduction from 1-IN-3-3-SAT or PARTITION for example) does not show this as it produced *non-homogeneous* polynomials. One would like to hope that the homogenization process in (1) preserves the positivity property. This is almost true but not quite. In fact, it is easy to see that homogenization preserves nonnegativity:

$$p(x) \geq 0, \ \forall x \Leftrightarrow p_h(x,y) \geq 0, \ \forall x, y.$$

Here's a proof:
($\Leftarrow$) If $p_h(x,y) \geq 0$ for all $x, y \Rightarrow p_h(x,1) \geq 0 \ \forall x \Rightarrow p(x) \geq 0 \ \forall x$.
($\Rightarrow$) By the contrapositive, suppose $\exists x, y$ s.t. $p_h(x,y) < 0$.

- If $y \neq 0$, $p_h(\frac{x}{y}, 1) < 0 \Rightarrow p(\frac{x}{y}) < 0$.
- If $y = 0$, by continuity, we perturb y to make it nonzero and we repeat the reasoning above.

However, the implications that we actually need are the following:

$$p(x) > 0 \ \forall x \Leftrightarrow p_h(x,y) > 0 \ \forall (x,y) \neq 0. \tag{2}$$

($\Leftarrow$) This direction is still true: $p_h(x,y) > 0, \ \forall (x,y) \neq 0 \Rightarrow p_h(x,1) > 0, \forall x \Rightarrow p(x) > 0 \ \forall x$.
($\Rightarrow$) This implication is also true if $y \neq 0$. Indeed, suppose $\exists (x,y)$ such that $p_h(x,y) = 0$, $y \neq 0$. Then, we rescale y to be 1:

$$0 = \frac{1}{||y||^d} p_h(x,y) = p_h\left(\frac{x}{||y||}, 1\right) = p\left(\frac{x}{||y||}\right)$$

and we get that $\exists \tilde{x} = \frac{x}{||y||}$ such that $p(\tilde{x}) = 0$.
However, the desired implication fails when $y = 0$. Here is a simple counterexample: Let

$$p(x_1, x_2) = x_1^2 + (1 - x_1 x_2)^2,$$

which is strictly positive $\forall x_1, x_2$. However, its homogenization

$$p_h(x_1, x_2, y) = x_1^2 y^2 + (y^2 - x_1 x_2)^2$$

has a zero at $(x_1, x_2, y) = (1, 0, 0)$.
At this point, we seem to be stuck. How can we get round this issue? Notice that we don't actually need to show that (2) is true for all polynomials. It is enough to establish

it for polynomials that appear in our reduction from 1-IN-3-3-SAT (indeed, our goal is to show that testing positivity for degree-4 homogeneous polynomials is harder than answering 1-IN-3-3-SAT).

Recall our reduction from 1-IN-3-3-SAT to POLYPOS (given here on one particular instance):

$$\phi = (x_1 \vee \bar{x_2} \vee x_3) \wedge (\bar{x_1} \vee \bar{x_2} \vee x_3) \wedge (x_1 \vee x_3 \vee x_4)$$

$$\downarrow$$

$$p(x) = \sum_{i=1}^{4} (x_i(1-x_i))^2 + (x_1 + (1-x_2) + x_3 - 1)^2 + \ldots + (x_1 + x_3 + x_4 - 1)^2.$$

Let us consider the homogeneous version of this polynomial [1]:

$$p_h(x,y) = \sum_{i=1}^{4} (x_i(y-x_i))^2 + (yx_1 + (y^2 - yx_2) + yx_3 - y^2)^2 + \ldots + (yx_1 + yx_3 + yx_4 - y^2)^2.$$

Let us try once again to establish the claim we were after: $p(x) > 0 \ \forall x \Leftrightarrow p_h(x,y) > 0 \ \forall (x,y) \neq 0$. We have already shown that ($\Leftarrow$) holds and that ($\Rightarrow$) holds when $y \neq 0$. Consider now the case where $y = 0$ (which is where the previous proof failed). Here, $p_h(x,0) = \sum_i x_i^4 > 0 \ \forall x \neq 0$. $\square$

2 Local optimality in constrained quadratic optimization

Recall the quadratic programming problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & p(x) := x^T Q x + c^T x + d \\ \text{s.t. } \ & Ax \leq b. \end{aligned} \tag{3}$$

A point $\bar{x} \in \mathbb{R}^n$ is a local minimum of p subject to the constraints $Ax \leq b$ if $\exists \epsilon > 0$ such that $p(\bar{x}) \leq p(x)$ for all $x \in B(\bar{x}, \epsilon)$ s.t. $Ax \leq b$.

Let LOCAL-2 be the following decision problem: Given rational matrices and vectors (Q, c, d, A, b) and a rational point $\bar{x} \in \mathbb{R}^n$, decide if $\bar{x}$ is a local min for problem (3).

[1] Convince yourself that the homogenization of the product of two polynomials is the product of their homogenizations.

Theorem 2. *LOCAL-2 is NP-hard.*

The key result in establishing this statement is the following theorem by Motzkin and Straus [1].

Theorem 3 (Motzkin-Straus, 1965)**.** *Let G=(V,E) be a graph with $|V| = n$ and denote by $\omega(G)$ the size of its largest clique. Let*

$$f(x) := -\sum_{\{i,j\}\in E} x_i x_j$$

then

$$f^* := \min_{x\in\Delta} f(x) = \frac{1}{2\omega} - \frac{1}{2}, \tag{4}$$

where Δ is the simplex in dimension n, i.e.,

$$\Delta := \{(x_1, \ldots, x_n)| \sum_i x_i = 1, x_i \geq 0, i = 1, \ldots, n\}.$$

Notice that this optimization problem is a quadratic program with linear constraints.

<u>Proof:</u> The proof we present here is based on [2].

- We first show that $f^* \leq \frac{1}{2\omega} - \frac{1}{2}$. To see this, take

$$x_i = \begin{cases} \frac{1}{\omega} \text{ if } i \in \text{largest clique} \\ 0 \text{ otherwise} \end{cases},$$

  then

$$f(x) = \frac{-1}{\omega^2}\left(\frac{\omega(\omega-1)}{2}\right) = -\frac{1}{2} + \frac{1}{2\omega}.$$

- Let's show now that $f^* \geq \frac{1}{2\omega} - \frac{1}{2}$. We prove this by induction on n.

  Base case $(n = 2)$:

  - If the two nodes are not connected, then $f^* = 0$ as there are no edges. Moreover, $\omega = 1$ so $\frac{1}{2\omega} - \frac{1}{2} = 0$, which proves the claim.

- If the two nodes are connected, then $\omega = 2$ and

$$f^* = \min_{\{x_1+x_2=1,\ x_1\geq 0,\ x_2\geq 0\}} -x_1x_2.$$

The solution to this problem is

$$x_1^* = x_2^* = 1/2$$

(this will be shown in a more general case in (5)). This implies that $f^* = -\frac{1}{4}$. But $\frac{1}{2\omega} - \frac{1}{2} = -\frac{1}{4}$.

Induction step: Let's assume $n > 2$ and that the result holds for any graph with at most $n-1$ nodes. Let x^* be the optimal solution to (4). We cover three different cases.

(1) Suppose $x_i^* = 0$ for some i. Remove node i from G and obtain a new graph G' with $n-1$ nodes. Consider the optimization problem (4) for G'. Denote by f' its objective function and by x' its optimal solution. Then

$$f(x^*) \geq f'(x').$$

This can be seen by taking $x' = \tilde{x}$, where $\tilde{x}$ contains the entries of x^* with the i^{th} entry removed. We know $f'(x') \geq \frac{1}{2\omega'} - \frac{1}{2}$ by induction hypothesis, where ω' is the size of the largest clique in G'. Notice also that $\omega \geq \omega'$ as all cliques in G' are also cliques in G. Hence

$$f^* = f(x^*) \geq f'(x') = \frac{1}{2\omega'} - \frac{1}{2} \geq \frac{1}{2\omega} - \frac{1}{2}.$$

(2) Suppose $x_i^* > 0$ for all i and $G \neq K_n$, where K_n is the complete graph on n nodes. Again, we want to prove that $f^* \geq \frac{1}{2\omega} - \frac{1}{2}$. We are going to need an optimality condition from a previous lecture, which we first recall. Consider the optimization problem

$$\begin{aligned} &\min g(x) \\ &\text{s.t. } Ax = b. \end{aligned}$$

If a point $\bar{x}$ is locally optimal, then $\exists \mu \in \mathbb{R}^m$ s.t. $\nabla g(x) = A^T\mu$. This necessary condition is also sufficient (for global optimality) if g is convex.

In our case, our constraint space is the simplex, hence we can write our constraints $e^T x = 1,\ x \geq 0$. The necessary optimality condition then translates to x^* satisfying

$$\nabla f(x^*) = \mu e,$$

in other words, all entries of $\nabla f(x^*)$ are the same. Notice that we have not included the constraints $x \geq 0$ in the optimality condition. Indeed, necessity of the optimality condition means that if the condition is violated at x^*, then there exists a feasible descent direction at x^*. By continuity, the constraints $\{x_i^* > 0\}$ will hold on a small ball around x^*. Therefore, locally we only need to worry about the constraint $e^T x = 1$.

Since $G \neq K_n$, at least one edge is missing. W.l.o.g., let's assume that this edge is $(1, 2)$. Then

$$\frac{\partial f}{\partial x_1}(x^*) = -\sum_{j \in \mathcal{N}_1} x_j^* = \frac{\partial f}{\partial x_2}(x^*) = -\sum_{j \in \mathcal{N}_2} x_j^*.$$

This implies that

$$f(x_1^* + t, x_2^* - t, x_3^*, \ldots, x_n^*) = f(x^*),\ \forall t.$$

Indeed, expanding out $f(x_1^* + t, x_2^* - t, x_3^*, \ldots, x_n^*)$ we get

$$\begin{aligned}
&f(x_1^* + t, x_2^* - t, x_3^*, \ldots, x_n^*) \\
&= -\sum (\text{terms without } x_1, x_2) - \sum_{j \in N_1} (x_1^* + t) x_j^* - \sum_{j \in N_2} (x_2^* - t) x_j^* \\
&= f(x^*) + t \sum_{j \in N_1} x_j^* - t \sum_{j \in N_2} x_j^* \\
&= f(x^*)
\end{aligned}$$

For some t, we can make either $x_1^* + t$ or $x_2^* - t = 0$. (Notice that by doing this, we remain on the simplex, with the same objective value). Hence, we are back to the previous case.

(3) In this last case, $x_i^* > 0,\ \forall i$ and $G = K_n$. Then

$$f(x) = -\sum_{\{i,j\}} x_i x_j = \frac{(x_1^2 + \ldots + x_n^2) - (x_1 + \ldots + x_n)^2}{2}$$

and

$$\min_{x \in \Delta} f(x) = \frac{1}{2} \min_{x \in \Delta} (x_1^2 + \ldots + x_n^2) - \frac{1}{2}.$$

We claim that the minimum of $g(x) = x_1^2 + \ldots + x_n^2$ over Δ is

$$x^* = (1/n, \ldots, 1/n). \tag{5}$$

To see this, consider the optimality condition seen in the previous case, which is now sufficient, as g is convex. Clearly, $x^* \in \Delta$ and

$$\nabla g(x^*) = 2 \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix} = \mu e$$

for $\mu = \frac{2}{n}$, which proves the claim. Finally, as $\omega = n$, we obtain $f^* = \frac{1}{2\omega} - \frac{1}{2}$. $\square$

<u>Proof of Theorem 2:</u>

The goal is to show that it is NP-hard to certify local optimality when minimizing a (non-convex) quadratic function subject to affine inequalities.
We start off by formulating a decision version of the Motzkin-Straus theorem: Given an integer k,

$$\omega(G) \geq k \Leftrightarrow f^* < \frac{1}{2k-1} - \frac{1}{2}.$$

Indeed,

- If $\omega(G) \geq k \Rightarrow f^* = \frac{1}{2\omega} - \frac{1}{2} \leq \frac{1}{2k} - \frac{1}{2} < \frac{1}{2k-1} - \frac{1}{2}$.
- If $\omega(G) < k \Rightarrow \omega(G) \leq k-1 \Rightarrow f^* = \frac{1}{2\omega} - \frac{1}{2} \geq \frac{1}{2k-2} - \frac{1}{2} \geq \frac{1}{2k-1} - \frac{1}{2}$.

Recall that given an integer k, deciding whether $\omega(G) \geq k$ is an NP-hard problem, as it is equivalent to STABLE SET on $\bar{G}$ (and we already gave a reduction 3SAT $\to$ STABLESET).
Define now

$$g(x) := f(x) - \left(\frac{1}{2k-1} - \frac{1}{2}\right).$$

Then, for a given k, deciding whether $\omega(G) \geq k$ is equivalent to deciding whether

$$\min_{x \in \Delta} g(x) < 0.$$

To go from this problem to local optimality, we try once again to make the objective homogeneous. Define

$$h(x) := f(x) - \left(\frac{1}{2k-1} - \frac{1}{2}\right)(x_1 + \ldots + x_n)^2.$$

We have

$$\min_{x\in\Delta} g(x) < 0 \Leftrightarrow \min_{x\in\Delta} h(x) < 0 \Leftrightarrow \min_{\{x_i\geq 0,\ i=1,\ldots,n\}} h(x) < 0,$$

where the last implication holds by homogeneity of h. As $h(0) = 0$ and h is homogeneous, this last problem is equivalent to deciding whether $x = 0$ is a local minimum of the nonconvex QP with affine inequalities:

$$\begin{aligned} &\min\ \ h(x) \\ &\text{s.t.}\ \ x_i \geq 0,\ \ i = 1, \ldots, n. \end{aligned} \tag{6}$$

Hence, we have shown that given an integer k, deciding whether $\omega(G) \geq k$ is equivalent to deciding whether $x = 0$ is a local minimum for (6), which shows that this latter problem is NP-hard. □

2.1 Copositive matrices

Definition 1 (Copositive matrix)**.** *A matrix $M \in S^{n\times n}$ is copositive if $x^T Mx \geq 0$, for all $x \geq 0$ (i.e., all vectors in $\mathbb{R}^n$ that are elementwise nonnegative).*

A sufficient condition for M to be copositive is

$$M = P + N,$$

where $P \succeq 0$ and $N \geq 0$ (i.e., all entries of N are nonnegative). This can be checked by semidefinite programming.

Notice that as a byproduct of the previous proof, we have shown that it is NP-hard to decide whether a given matrix M is copositive. To see this, consider the matrix M associated to the quadratic form h (i.e., $h(x) = x^T Mx$). Then,

$$x = 0 \text{ is a local minimum for (6) } \Leftrightarrow M \text{ is copositive.}$$

Contrast this complexity result with the "similar-looking" problem of checking whether M is positive semidefinite, i.e.,

$$x^T Mx \geq 0, \forall x.$$

Although checking copositivity is NP-hard, checking positive semidefiniteness of a matrix can be done in time $O(n^3)$.

2.2 Local optimality in unconstrained optimization

In Section 1, we showed that checking strict local optimization for degree-4 polynomials is hard. We now prove that the same is true for checking local optimality using a simple reduction from checking matrix copositivity.

Indeed, it is easy to see that a matrix M is copositive if and only if the homogeneous degree-4 polynomial

$$p(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{pmatrix}^T M \begin{pmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{pmatrix}$$

is globally nonnegative; i.e., it satisfies $p(x) \geq 0, \forall x$. By homogeneity, this happens if and only if $x = 0$ is a local minimum for the problem of minimizing p over $\mathbb{R}^n$.

References

[1] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of turán. *Canad. J. Math*, 17(4):533–540, 1965.

[2] S.A. Vavasis. *Nonlinear Optimization: Complexity Issues*. Oxford University Press, Inc., 1991.