4. Semidefinite program relaxations of integer programs.

Lecturer: Massimo Lauria

http://www.csc.kth.se/~lauria/sos14/

Disclaimer: this lecture note has not yet been reviewed by the mail lecturer. It is released as it is for the convenience of the students.

We continue the discussion on how to improve the initial relaxation made of an integer program. Here, we focus on semidefinite program (SDP) relaxations.

Motivation for SDP relaxations

SDP relaxations are more complicated to construct than linear relaxations, but more powerful. To see this, we consider the maximum cut problem.

Maximum cut problem[1]

[1] Poljak and Tuza. Maximum cuts and largest bipartite subgraphs. *Combinatorial Optimization*, 20, 1995

We consider a finite, undirected and loopless graph $G(V,E)$, where V denotes the set of vertexes and E denotes the set of edges. Each edge is weighted according to some function $c : E \to \mathbb{R}$.

Let $S \subseteq V$, the *cut* is defined as the edges with one end in S and the other end in $V \backslash S$ and the *weight of the cut* is the sum of the weights corresponding to the edges belonging to the cut.

The maximum cut problem is equivalent to the problem of finding the set S that maximizes the weight of the induced cut. It is NP-complete[2].

Several approximation algorithms have been developed to be able to solve it with a high *performance guarantee*[3] in polynomial time. The approximation algorithm with the highest performance guarantee is an SDP relaxation that achieves $\rho \approx 0.878$.

However, any Sherali-Adams relaxation of degree $t = n^{\delta(\epsilon)}$ can only achieve a performance guarantee of $1/2 + \epsilon$[4].

[2] Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*. Plenum Press, 1972

[3] The performance guarantee is defined as

$$\rho = \frac{\text{optimal value of relaxed problem}}{\text{optimal value of original problem}},$$

that is, the inverse of the integrality gap.

[4] Citation needed

Small example of SDP relaxation

We want to show that a polynomial p is non-negative over some polytope and write $p \geq 0$. If the polynomial p is expressible as a sum of squared polynomials, then we can find such representation in a tractable manner by reformulating the problem as a SDP relaxation[5]. Let us see a small example. Consider

$$F(x,y) = 2x^4 + 2x^3y - x^2y^2 + 5y^4$$

and assume it is expressible as a sum of squared polynomials. These have degree at least 2, half the degree of F, and furthermore since F is homogeneous

[5] Parrilo

they have degree exactly 2. The monomials in (x, y) of degree 2 are x^2, y^2 and xy.

We can express sums of squares of polynomials with the following matrix operation:

$$\begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix}^T \underbrace{\begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}}_{Q} \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix} =$$

$$q_{11}x^4 + q_{22}y^4 + (q_{33} + 2q_{12})x^2y^2 + 2q_{13}x^3y + 2q_{23}xy^3. \qquad (1)$$

We match the coefficients in (1) with those in $F(x, y)$, that is,

$$F(x,y) = \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix}^T Q \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix}$$

for

$$Q = \begin{pmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{pmatrix}.$$

To construct the sum of squared polynomials yielding $F(x, y)$, we perform a Cholesky factorization[6] of the matrix Q in the following way:

[6] Need SDP?

$$F(x,y) = \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix}^T U^T U \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix} =$$

$$\begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix}^T \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}^T \frac{1}{\sqrt{2}} \begin{pmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix} =$$

$$\left(\frac{1}{\sqrt{2}} (2x^2 - 3y^2 + xy) \right)^2 + \left(\frac{1}{\sqrt{2}} (y^2 + 3xy) \right)^2.$$

We have reformulated $F(x, y)$ as a sum of squared polynomials and, consequently, we have shown that $F(x, y) \geq 0$.

It is important to note that not every polynomial $p \geq 0$ on $\mathbb{R}^n$ can be expressed as a sum of squared polynomials if $n \geq 2$. An example of this is the Motzkin polynomial[7],

$$M(x, y, z) = x^6 + y^4z^2 + y^2z^4 - 3x^2y^2z^2.$$

[7] Motzkin. The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*. Academic Press, 1967

Moment matrix[8]

[8] Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001

We consider polynomials $p(x) : \mathbb{R}^n \to \mathbb{R}$ of degree t. We can express them as

$$p(x) = \sum_{\alpha \in [t]^n} p_\alpha x^\alpha, \text{ with } x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \text{ and } \sum_{i=1}^{n} \alpha_i \leq t,$$

where

$$1, x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_1x_n, \ldots, x_n^2, \ldots, x_1^t, \ldots, x_n^t \qquad (2)$$

is a basis of $p(x)$ and p is the coefficient vector of $p(x)$ with respect to (2).

We introduce the auxiliary variable $y_{00\ldots0} = 1$ and perform the variable substitution $x_1^i x_2^j \cdots x_n^l = y_{ij\ldots l}$. The moment matrix $M_t(y)$ has rows and columns labelled according to the basis (2) and its elements are the corresponding $y_{ij\ldots l}$. For simplicity, we only show the case for $n = 2$ and $t = 2$,

$$M_2(y) = \begin{array}{c} \\ 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1x_2 \\ x_2^2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{array} \\ \begin{pmatrix} 1 & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{30} & y_{22} & y_{13} & y_{04} \end{pmatrix} \end{array}.$$

Consider the squared polynomial

$$p^2(x) = (\sum_\alpha p_\alpha x^\alpha)^2 = \sum_\alpha \sum_{\alpha'} p_\alpha x^\alpha p_{\alpha'} x^{\alpha'}.$$

It can be expressed, using the moment matrix, in the following way:

$$p^2(x) = p_\alpha^T M_t(y) P_\alpha.$$

Consequently, we have that $p^2(x) \geq 0 \Leftrightarrow M_t(y) \succeq 0$.

Lovász-Schrijver hierarchy with SDP

LS_+[9] is a lift and project relaxation as its linear version. We are still limited to start each iteration with linear inequalities and we are only allowed to introduce quadratic inequalities. In fact, the only difference is that the matrix constraint $Y \succeq 0$ has been added.

[9] The Lovász-Schrijver hierarchy with SDP.

Geometric interpretation

As for the Lovász-Schrijver hierarchy, we introduce the auxiliary variable $x_0 = 1$ and perform the variable substitution $x_i x_j = y_{ij}$ for every $i, j = 0, \ldots, n$. We let any terms of degree zero be represented by y_{00}. We enforce the following constraints on Y:

1. $Y = Y^T$,
2. $y_{0i} = y_{i0} = y_{ii}, \quad$ for $i = 0, \ldots, n$,
3. $Y^{(i)} \in P \quad$ for $i = 0, \ldots, n$,
4. $Y^{(0)} - Y^{(i)} \in P \quad$ for $i = 0, \ldots, n$.
5. $Y \succeq 0$.

Constraints 1 and 2 are due to that $x_i x_j = x_j x_i$ and $x_i - x_i^2 = 0$. Constraints 3 and 4 stem from

$$\sum_{j=1}^{n} A^{(j)} y_{ij} - b y_{i0} \leq 0, \quad \text{for } i = 0, \ldots, n,$$

and

$$\sum_{j=1}^{n} A^{(j)} (y_{0j} - y_{ij}) - b(y_{00} - y_{i0}) \leq 0, \quad \text{for } i = 0, \ldots, n,$$

respectively. The matrix constraint 5 is equivalent to $p^2(Y) \geq 0$ for all polynomials $p(\cdot)$ of degree one as discussed previously.

Operator and rank

The operation $N_+(P)$ is the projection over $1 = Y_{00}, x_1 = Y_{01}, \ldots, x_n = Y_{0n}$. The rank t of a proof in LS_+ is the same as for Lovász-Schrijver hierarchy. Note that, if P is the polytope corresponding to the set of initial inequalities then $N_+(P)$ is determined by inequalities derivable in rank one.

The projection $N_+^t(P)$ can be solved in $n^{\mathcal{O}(t)}$ given a polynomial time separator of P. We can use the same separator as for $N^t(P)$, with the additional requirement that $Y \succeq 0$.

Example 1. Consider the maximum independent set problem on a graph $G(V, E)$

$$\begin{aligned} \text{maximize} \quad & \sum_{u \in V} x_u \\ \text{subject to} \quad & x_u + x_v \leq 1, \quad (u, v) \in E \\ & x_u \in \{0, 1\}. \end{aligned} \tag{3}$$

Over the complete graph of order n we want to show that the objective value of (3) is at most one, this is equivalent to showing that $\sum x_u \leq 1$. If we are using LS relaxations, then we need rank n. Using LS_+, we can do this in rank one.

The axioms of the proof are $1 - x_u - x_v \geq 0$, $x_u^2 - x_u = 0$, and we can derive

$$\sum_u \sum_{v \neq u} (1 - x_u - x_v) x_u + \sum_u (x_u^2 - x_u)(n-2) + (1 - \sum_u x_u)^2 \geq 0$$
$$\Rightarrow 1 - \sum_u x_u \geq 0.$$

Positivstellensatz proof systems[10]

Positivstellensatz proof systems are, as the name implies, based on the Positivstellensatz[11].

We consider a system of constraints

$$f_1 = 0, \ldots, f_k = 0,\ h_1 \geq 0, \ldots, h_m \geq 0, \tag{4}$$

[10] Dima Grigoriev and Nicolai Vorobjov. Complexity of null-and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1):153–160, 2001

[11] Bochnak, Coste, and Roy. *Geometrie Algebrique Reelle*. Springer-verlag, 1987

where f_i and h_i are polynomials of $x_1, \ldots, x_n$, and we require $0 \leq x_i \leq 1$ and $x_i^2 - x_i = 0$. We introduce the following pair of polynomials:

$$f = \sum_{s=1}^{k} f_s g_s, \quad h = \sum_{I \subseteq \{1,\ldots,m\}} \Big(\prod_{i \in I} h_i\Big)\Big(\sum_j e_{I,j}^2\Big), \tag{5}$$

where $e_{I,j}$ and g_s are arbitrary polynomials. If $f + h = -1$, we know that system (4) has no feasible solution because our initial constraints imply that $f = 0$ and $g \geq 0$. Such a polynomial pair is called the *refutation* for (4). The degree of the refutation is given by

$$\max_{s,I,j}\{\deg(f_s, g_s), \deg(e_{I,j}^2 \prod_{i \in I} h_i)\}. \tag{6}$$

Positivstellensatz calculus[12]

Positivstellensatz calculus is the dynamic version of Positivstellensatz. Again our input is the system of constraints (4), but now we let f defined in (5) instead be an arbitrary polynomial derived from $f_1 = 0, \ldots, f_k = 0$ and $x_i^2 - x_i = 0$ using the polynomial calculus inference rules[13]. If $f + h = -1$, we call the polynomial pair the *Positivstellensatz calculus refutation* of (5). The degree of the refutation is the maximum degree of $e_{I,j}^2 \prod_{i \in I} h_i$ and of the derivation for f.

[12] Dima Grigoriev and Nicolai Vorobjov. Complexity of null-and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1):153–160, 2001

[13] The polynomial calculus inference rules are

$$\frac{p \quad q}{\alpha p + \beta q}, \quad \frac{p}{xp}$$

for any polynomials p, q, coefficients α, β and variable x.

Sum-of-squares hierarchy

The Lasserre[14] hierarchy is a weaker version of the Positivstellensatz proof system where we do not allow to multiply constraints $h_j \geq 0$ together. We consider the same system (4). We introduce the following polynomial:

$$p = u_0 + \sum_j u_j h_j + \sum_i g_i f_i, \tag{7}$$

where u_j are sum-of-squares and g_i are arbitrary polynomials. The degree of a proof is the maximal degree of u_0, $u_j h_j$ and $g_i f_i$.

[14] also known as sum-of-squares

Geometrical interpretation

We start with $h_1 \geq 0, \ldots, h_m \geq 0$ and a fixed t. As it was the case with Sherali-Adams we lift the problem to a larger space and we implicitly enforce $x_i^2 = x_i$ constraints by considering multilinearized polynomials, thus we have the moment matrix $M^t = (y_{A \cup B})$ indexed by sets $|A|, |B| \leq t$. The constraints are to be *localized on moment matrices*, that is $M^t(h_i \circ Y)$. If $h_i = \sum_S \alpha_S \prod_{i \in S} x_i$ and we substitute every monomial $\prod_{i \in S} x_i$ for y_S, then the constraint becomes $M^t(h_i \circ Y) = \sum_S \alpha_S (y_{A \cup B \cup S})$.

We want to find a y such that $M_t(y) \succeq 0$ and $M_t(h_j y) \succeq 0$ where the last matrix inequality is equivalent to $h_j p^2(y)$ for all multilinear polynomials p of degree at most t.

We let $Q_t(K)$ denote the projection over $1 = y_\emptyset$ and $x_i = y_{\{i\}}$ for all i, where K is the set defined by the inequalities $h_1 \geq 0, \dots, h_m \geq 0$.

This setting does not allow equality constraints, but we can deal with them by splitting each equality $f_i = 0$ into $f_i \geq 0$ and $f_i \leq 0$.

Lasserre relaxations are stronger than Sherali-Adams. The following is an example of how to derive SA inequalities by using the Lasserre setup. Say we have the constraint $h(x) = x \geq 0$ and we want to derive the constraint $x(1-y)(1-z) \geq 0$.

The moment matrix of the constraint h is

$$M^2(h \circ Y) = \begin{array}{c} \\ x \\ xy \\ xz \\ xyz \end{array} \begin{array}{c} \begin{array}{cccc} x & xy & xz & xyz \end{array} \\ \begin{pmatrix} y_x & y_{xy} & y_{xz} & y_{xyz} \\ y_{xy} & y_{xy} & y_{xyz} & y_{xyz} \\ y_{xz} & y_{xyz} & y_{xz} & y_{xyz} \\ y_{xyz} & y_{xyz} & y_{xyz} & y_{xyz} \end{pmatrix} \end{array}. \tag{8}$$

The condition $M^2(h \circ Y) \succeq 0$ is equivalent to that $\alpha^T M^2(h \circ Y)\alpha \geq 0$ for any vector α, in particular it holds for the vector $(1 \quad -1 \quad -1 \quad 1)^T$ corresponding to $(1-y)(1-z) = (1-y-z+yz)$. Performing the calculation yields

$$y_x - y_{xy} - y_{xz} + y_{xyz} \geq 0 \tag{9}$$

which is the constraint we would have obtained applying the Sherali-Adams method.

References

[BCR87] Bochnak, Coste, and Roy. *Geometrie Algebrique Reelle*. Springer-verlag, 1987.

[GV01] Dima Grigoriev and Nicolai Vorobjov. Complexity of null-and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1):153–160, 2001.

[Kar72] Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*. Plenum Press, 1972.

[Las01] Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[Mot67] Motzkin. The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*. Academic Press, 1967.

[PT95] Poljak and Tuza. Maximum cuts and largest bipartite subgraphs. *Combinatorial Optimization*, 20, 1995.