Math 591 – Real Algebraic Geometry and Convex Optimization

Lecture 11: Theta bodies: approximating convex hulls of varieties

Cynthia Vinzant, Spring 2019

Note that maximizing a linear function over a set is equivalent to maximizing it over the convex hull of the set. That is, for a set $S \subset \mathbb{R}^n$ and a linear function $\ell : \mathbb{R}^n \to \mathbb{R}$,

$$\max_{p\in S} \ell(p) \;=\; \max_{p\in \mathrm{conv}(S)} \ell(p).$$

Today we will consider this problem in the case that S is the real variety $V_{\mathbb{R}}(I)$ of some ideal $I \subset \mathbb{R}[x_1, \ldots, x_n]$. Then to maximize a specific linear function, we have the relaxation:

$$\begin{aligned}\max_{p\in V_{\mathbb{R}}(I)} \ell(p) \;&=\; \min_{c\in\mathbb{R}} c \text{ such that } c - \ell(x) \geq 0 \text{ on } V_{\mathbb{R}}(I)\\ &\leq\; \min_{c\in\mathbb{R}} c \text{ such that } c - \ell(x) \in \mathrm{SOS}_{n,\leq 2d} + I,\end{aligned}$$

the last of which can be written as a semidefinite program. Today, we try to consider all linear functions simultaneously.

Note that the (closure of the) convex hull of $V_{\mathbb{R}}(I)$ can be written as the set of all points satisfying all valid linear inequalities on $V_{\mathbb{R}}(I)$:

$$\overline{\mathrm{conv}(V_{\mathbb{R}}(I))} \;=\; \{p \in \mathbb{R}^n : \ell(p) \geq 0 \text{ for all } \ell \in \mathbb{R}[x_1, \ldots, x_n]_{\leq 1} \text{ such that } \ell \geq 0 \text{ on } V_{\mathbb{R}}(I)\}$$

For any $k \in \mathbb{N}$, define the **k-th theta body** of the ideal I to be the convex subset of $\mathbb{R}^n$ defined by linear inequalities that can be certified to be nonnegative using sums of squares of degree $\leq 2k$:

$$\mathrm{TH}_k(I) \;=\; \{p \in \mathbb{R}^n : \ell(p) \geq 0 \text{ for all } \ell \in \mathbb{R}[x_1, \ldots, x_n]_{\leq 1} \text{ such that } \ell \in \mathrm{SOS}_{n,\leq 2k} + I\}$$

Any element of $\mathrm{SOS}_{n,\leq 2k} + I$ is nonnegative on $V_{\mathbb{R}}(I)$. As the sets of defining inequalities gets bigger, the set of points satisfying them become smaller:

$$\mathrm{TH}_1(I) \supseteq \mathrm{TH}_2(I) \supseteq \ldots \supseteq \mathrm{TH}_k(I) \supseteq \overline{\mathrm{conv}(V_{\mathbb{R}}(I))}.$$

Proposition. *(1) If $V_{\mathbb{R}}(I)$ is compact, then*

$$\bigcap_{k\in\mathbb{N}} \mathrm{TH}_k(I) \;=\; \overline{\mathrm{conv}(V_{\mathbb{R}}(I))}.$$

(2) If $V_{\mathbb{R}}(I)$ is finite then for some sufficiently large N,

$$\mathrm{TH}_N(I) \;=\; \mathrm{conv}(V_{\mathbb{R}}(I)).$$

Proof. (1) Recall that by Schmüdgen's Theorem, any polynomial that is strictly positive on $V_{\mathbb{R}}(I)$ belongs to $\mathrm{SOS}_n + I$. If some point $p \in \mathbb{R}^n$ does not belongs to $\overline{\mathrm{conv}(V_{\mathbb{R}}(I))}$, then there is some affine linear function $\ell \in \mathbb{R}[x_1, \ldots, x_n]_{\leq 1}$ so that $\ell > 0$ on $V_{\mathbb{R}}(I)$ and $\ell(p) < 0$. For some k, ℓ belongs to $\mathrm{SOS}_{n,\leq 2k} + I$, meaning that $p \notin \mathrm{TH}_k(I)$.

(2, sketch) If I equals the ideal of all polynomials vanishing on $V_{\mathbb{R}}(I)$, i.e. $I = \mathcal{I}(V_{\mathbb{R}}(I))$, then we saw that every polynomial nonnegative on $V_{\mathbb{R}}(I)$ can be written as a sum of squares of degree $\leq 2{\cdot}N$ modulo I, where $V_{\mathbb{R}}(I)$ has N points. Restricting to affine linear polynomials then shows that $\mathrm{TH}_N(I) = \mathrm{conv}(V_{\mathbb{R}}(I))$.

If I is not all $\mathcal{I}(V_{\mathbb{R}}(I))$, then one has to work a little harder. Can you prove it? $\square$

Theta bodies are useful because they are *projection of spectrahedra*, meaning that one can optimize over them use semidefinite programming.

To write this down explicitly we work in the quotient ring $\mathbb{R}[x_1,\ldots,x_n]/I$. Elements have the form $f+I$ where $f\in\mathbb{R}[x_1,\ldots,x_n]$ and two polynomial f,g define the same element of the quotient ring $f+I=g+I$ if and only if $f-g\in I$. Then addition and multiplication in $\mathbb{R}[x_1,\ldots,x_n]/I$ work the way you would want:

$$(f+I)+(g+I)=(f+g)+I \quad \text{and} \quad (f+I)\cdot(g+I)=f\cdot g+I.$$

We can define the *degree* of an element $f+I$ of $\mathbb{R}[x_1,\ldots,x_n]/I$ as the minimal degree of a representative, i.e.

$$\deg(f+I) \;=\; \min_{h\in I}\deg(f+h).$$

For any collection $\mathcal{B}\subset\mathbb{R}[x_1,\ldots,x_n]/I$, let $\mathcal{B}_k$ denote subset of elements of degree $\leq k$.

Given the ideal I, we will assume that we have an $\mathbb{R}$-basis $\mathcal{B}$ for $\mathbb{R}[x_1,\ldots,x_n]/I$ satisfying

(i) $\mathcal{B}_1=\{1+I,x_1+I,\ldots,x_n+I\}$ and

(ii) if $f_i+I,f_j+I\in\mathcal{B}_d$, then $f_i\cdot f_j+I$ belongs to the $\mathbb{R}$-span of $\mathcal{B}_{2d}$.

For those familiar with the theory of term orders and Gröbner bases, one could take $\mathcal{B}$ to be the standard monomials of I with respect to a graded term order. In particular, we can take the elements of $\mathcal{B}$ to have monomial representatives, $x^\alpha+I$.

Example. For $I=\langle x_1^2-x_1,x_2^2-x_2\rangle\subset\mathbb{R}[x_1,x_2]$, the quotient $\mathbb{R}[x_1,x_2]/I$ is a four-dimensional $\mathbb{R}$-vector space. We can take the basis

$$\mathcal{B} \;=\; \{1+I,x_1+I,x_2+I,x_1x_2+I\},$$

which satisfies the conditions (i) and (ii) above.

Example. For $I=\langle x_1^2+x_2^2-1\rangle\subset\mathbb{R}[x_1,x_2]$, the quotient $\mathbb{R}[x_1,x_2]/I$ is an infinite-dimensional $\mathbb{R}$-vector space. We claim that

$$\mathcal{B} \;=\; \{x_1^k+I : k\in\mathbb{N}\}\cup\{x_1^kx_2+I : k\in\mathbb{N}\},$$

is a basis for the quotient which satisfies the conditions (i) and (ii) above. Note that any monomial $x_1^ax_2^b+I$ with $b\geq 2$ can be written as a linear combination of elements from $\mathcal{B}$ by replacing x_2^2 with $1-x_1^2$ as many times as needed. To check condition (ii) in a specific instance, $x_2+I\in\mathcal{B}$, and its square can be written as an $\mathbb{R}$-linear combination of elements in $\mathcal{B}_2$:

$$(x_2+I)(x_2+I)=x_2^2+I=(1+I)-(x_1^2+I).$$

Then we can write $\mathrm{TH}_k(I)$ as the projection of a spectrahedron as follows. Let $y\in\mathbb{R}^{\mathcal{B}_{2k}}$ be a vector of real numbers whose entries are indexed by elements of $\mathcal{B}_{2k}$. Define the $|\mathcal{B}_k|\times|\mathcal{B}_k|$ matrix $\mathcal{M}_{\mathcal{B}_k}(y)$ with rows indexed by the elements of $\mathcal{B}_k$ with entries

$$\left(\mathcal{M}_{\mathcal{B}_k}(y)\right)_{f_i+I,f_j+I} \;=\; \sum_{f_\ell+I\in\mathcal{B}_k} c_\ell y_\ell$$

where

$$(f_i+I)\cdot(f_j+I)=\sum_{f_\ell+I\in\mathcal{B}_k} c_\ell(f_\ell+I).$$

is the (unique) linear representation guaranteed by property (ii) of $\mathcal{B}$.

Proposition. $\mathrm{TH}_k(I) \;=\overline{\{\pi_{(x_1+I,\ldots,x_n+I)}(y)\in\mathbb{R}^n : y\in\mathbb{R}^{\mathcal{B}_{2k}},\; y_{1+I}=1,\; \mathcal{M}_{\mathcal{B}_k}(y)\succeq 0\}}.$

The proof of this relies on the duality between the cone of sums of squares $\mathrm{SOS}_{n,\leq 2k} + I$ and the cone of "moments" $\{y \in \mathbb{R}^{\mathcal{B}_{2k}} : \mathcal{M}_{\mathcal{B}_k}(y) \succeq 0\}$. We leave out the details here though.

Example. Consider $I = \langle x_1^2 + x_2^2 - 1 \rangle$. Then

$$\mathcal{B}_4 = \{1 + I, x_1 + I, x_2 + I, x_1^2 + I, x_1x_2 + I, x_1^3 + I, x_1^2x_2 + I, x_1^4 + I, x_1^3x_2 + I\},$$

to which we can associate the vector

$$y = (y_\emptyset, y_1, y_2, y_{11}, y_{12}, y_{111}, y_{112}, y_{1111}, y_{1112}) \ \in \ \mathbb{R}^9.$$

Then $\mathcal{M}_{\mathcal{B}_1}(y)$ is a 3×3 matrix whose rows and columns are indexed by the set of elements $\mathcal{B}_1 = \{1 + I, x_1 + I, x_2 + I\}$:

$$\mathcal{M}_{\mathcal{B}_1}(y) \ = \ \begin{pmatrix} y_\emptyset & y_1 & y_2 \\ y_1 & y_{11} & y_{12} \\ y_2 & y_{12} & y_\emptyset - y_{11} \end{pmatrix}.$$

Here the $(3,3)$ entry was calculated by writing

$$(x_2 + I)(x_2 + I) = (1 + I) - (x_1^2 + I) \quad \to \quad y_\emptyset - y_{11}.$$

Then the proposition above states that

$$\mathrm{TH}_1(I) \ = \ \overline{\left\{ (y_1, y_2) \in \mathbb{R}^2 : \exists y_{11}, y_{12} \in \mathbb{R} \text{ s.t. } \begin{pmatrix} 1 & y_1 & y_2 \\ y_1 & y_{11} & y_{12} \\ y_2 & y_{12} & 1 - y_{11} \end{pmatrix} \succeq 0 \right\}}.$$

This is the projection of a spectrahedron in $\mathbb{R}^4$.

For $k = 2$, we find that

$$\mathcal{M}_{\mathcal{B}_2}(y) \ = \ \begin{pmatrix} y_\emptyset & y_1 & y_2 & y_{11} & y_{12} \\ y_1 & y_{11} & y_{12} & y_{111} & y_{112} \\ y_2 & y_{12} & y_\emptyset - y_{11} & y_{112} & y_1 - y_{111} \\ y_{11} & y_{111} & y_{112} & y_{1111} & y_{1112} \\ y_{12} & y_{112} & y_1 - y_{111} & y_{1112} & y_{11} - y_{1111} \end{pmatrix}.$$

And the second theta body of I is $\mathrm{TH}_2(I) =$

$$\overline{\left\{ (y_1, y_2) \in \mathbb{R}^2 : \exists y_{11}, \ldots, y_{1112} \in \mathbb{R} \text{ s.t. } \begin{pmatrix} 1 & y_1 & y_2 & y_{11} & y_{12} \\ y_1 & y_{11} & y_{12} & y_{111} & y_{112} \\ y_2 & y_{12} & 1 - y_{11} & y_{112} & y_1 - y_{111} \\ y_{11} & y_{111} & y_{112} & y_{1111} & y_{1112} \\ y_{12} & y_{112} & y_1 - y_{111} & y_{1112} & y_{11} - y_{1111} \end{pmatrix} \succeq 0 \right\}}.$$

For finite varieties, there is a nice characterization of when the first theta body is exact. For it, we need the following definition:

Definition. A polytope $P \subset \mathbb{R}^n$ is called 2-level if for every facet F of P, all the vertices of P belong to F or to a unique translate of the affine span of F.

For example the cube $[0,1]^n$ is a 2-level, but a pentagon in the plane is not. Gouveia, Parrilo, ad Thomas showed that these characterize finite sets for which the first theta body is exact.

Theorem (Gouveia, Parrilo, Thomas, 2010, [1]). *Let $S \subset \mathbb{R}^n$ and let $I = \mathcal{I}(S) \subset \mathbb{R}[x_1, \ldots, x_n]$ be the ideal of polynomial vanishing on S. Then $\mathrm{TH}_1(I)$ equals $\mathrm{conv}(S)$ if and only if $\mathrm{conv}(S)$ is a 2-level polytope.*

Sketch of $\Rightarrow$. Let $\ell_1 \geq 0, \ldots, \ell_m \geq 0$ be a minimal set of affine linear inequalities defining $P = \mathrm{conv}(S)$. Then for each i, $F_i = \{p : \ell_i(p) = 0\} \cap P$ is a facet of P.

If $\mathrm{TH}_1(I) = \mathrm{conv}(S)$, then we can write ℓ_i as a sum of squares of degree two mod I:

$$\ell_i \equiv \sum_k h_k^2 \quad \mod I \text{ where } h_k \in \mathbb{R}[x_1, \ldots, x_n]_{\leq 1}.$$

On the points $S \cap F_i$, $\ell_i = 0$ and any element of I also vanishes, implying that h_k must vanish also. Since h_k has degree ≤ 1, this implies that h_k vanishes on the whole affine span of $S \cap F_i$.

However the affine span of $S \cap F_i$ is the hyperplane defined by $\ell_i = 0$, implying that h_k is some multiple of ℓ_i, $h_k = \lambda_k \ell_i$. This implies that

$$\ell_i \equiv (\sum_k \lambda_k^2)\ell_i^2 \mod I \quad \Rightarrow \quad 0 \equiv \ell_i \cdot (c\ell_i - 1) \mod I,$$

where $c = \sum_k \lambda_k^2$. Therefore for every point $p \in S$, either p lies on the facet defined by $\ell_i = 0$ or on the translate defined by $\ell_i = 1/c$. This shows that P is 2-level. $\square$

This has a very nice connection with the set STAB_G of indicator sets of stable sets of a graph that we worked with last time.

Theorem (Gouveia, Parrilo, Thomas, 2010, [1]). $\mathrm{conv}(\mathrm{STAB}_G)$ *is a 2-level polytope if and only if the graph G is perfect.*

References

[1] J. Gouveia, P. Parrilo, and R. Thomas, *Theta bodies for polynomial ideals.* SIAM Journal of Optimization 20(4), pp. 2097 – 2118, 2010.