Convergence rates of RLT and Lasserre-type hierarchies for the generalized moment problem over the simplex and the sphere

Felix Kirschner, Tilburg University

6.4.2021

Agenda

- $\mathcal{K} \subset \mathbb{R}^n$ compact
- $\mathcal{M}(\mathcal{K})$ the space of signed finite Borel measures supp. on $\mathcal{K}$
- $\mathcal{M}(\mathcal{K})_+$ the conv. cone of (pos.) finite Borel measures on $\mathcal{K}$
- $\mathcal{C}(\mathcal{K})$ denotes the space of continuous functions on $\mathcal{K}$
- $f_0, f_1, \dots, f_m \in \mathcal{C}(\mathcal{K})$ and $b_1, \dots, b_m \in \mathbb{R}$

GMP

$$\begin{aligned} \mathrm{val} = \inf_{\mu \in \mathcal{M}(\mathcal{K})_+} & \int_{\mathcal{K}} f_0(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) \\ \text{s.t.} & \int_{\mathcal{K}} f_i(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}) = b_i \ \forall i \in [m] \\ & \int_{\mathcal{K}} \mathrm{d}\mu(\mathbf{x}) \leq 1 \end{aligned}$$

Dual problem

The GMP is an infinite dimensional conic linear optimization problem
$\rightarrow$ well-understood duality theory.

Dual

$$\mathrm{val}' = \sup_{(y,t)\in\mathbb{R}^m\times\mathbb{R}_+} \sum_{i=1}^{m} y_i b_i - t$$

$$\text{s.t.} \quad f_0(\mathbf{x}) - \sum_{i=1}^{m} y_i f_i(\mathbf{x}) + t \geq 0 \; \forall \mathbf{x} \in \mathcal{K}$$

- Weak duality holds: $\mathrm{val}' \leq \mathrm{val}$
- Thus, the duality gap, i.e., $\mathrm{val} - \mathrm{val}'$ is always greater or equal to zero
- We will assume strong duality holds, i.e., the duality gap is zero

Theorem 1 (see, e.g. [3, Thm. 1.3])

Assume the primal of the GMP is feasible. Then it has an optimal solution (the inf is attained), and $\mathrm{val} = \mathrm{val}'$.

Define the bilinear form $\langle \cdot, \cdot \rangle : \mathcal{C}(\mathcal{K}) \times \mathcal{M}(\mathcal{K}) \to \mathbb{R}$ via

$$\mathcal{C}(\mathcal{K}) \times \mathcal{M}(\mathcal{K}) \ni (f, \mu) \mapsto \langle f, \mu \rangle = \int_{\mathcal{K}} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x}).$$

Theorem 2 (see, e.g. [4, Prop. 2.8])

If $b \in \mathrm{int}\,((\langle f_1, \mu \rangle, \langle f_2, \mu \rangle, \ldots, \langle f_m, \mu \rangle) : \mu \in \mathcal{M}(\mathcal{K})_+)$ then $\mathrm{val} = \mathrm{val}'$. If the common optimal value is finite, then the set of dual optimal solutions is nonempty and bounded.

Rational optimization

Consider the problem

$$p^* = \min_{\mathbf{x}\in\mathcal{K}} \frac{p(\mathbf{x})}{q(\mathbf{x})},$$

- with $p, q \in \mathbb{R}[\mathbf{x}]$ relatively prime and assume $q(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{K}$. (If q changes signs on $\mathcal{K}$, then $p^* = -\infty$)
- We make the stronger assumption $q(\mathbf{x}) \geq 1$ on $\mathcal{K}$

This can be modeled as a GMP:

$$\mathrm{val} = \inf_{\mu\in\mathcal{M}(\mathcal{K})_+} \left\{ \int_{\mathcal{K}} p(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) : \int_{\mathcal{K}} q(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) = 1 \right\}.$$

- Since we assume $q(\mathbf{x}) \geq 1$ on $\mathcal{K}$ we may add the (redundant) constraint $\int_{\mathcal{K}} \mathrm{d}\mu \leq 1$
- A dual optimal solution is given by $(y, t) = (p^*, 0)$

Idea of the relaxation

Recall the duality $\langle \cdot, \cdot \rangle : \mathcal{C}(\mathcal{K}) \times \mathcal{M}(\mathcal{K}) \to \mathbb{R}$ given by

$$(f, \mu) \mapsto \langle f, \mu \rangle = \int_{\mathcal{K}} f(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$$

is a bilinear form.
Hence, $\langle \cdot, \mu \rangle$ is a linear functional for every $\mu \in \mathcal{M}(\mathcal{K})$.

Note

We are looking for a linear operator $L^* \in (\mathbb{R}[\mathbf{x}])^*$ s.t.

$$L^*(\mathbf{x}^\alpha) = \langle \mathbf{x}^\alpha, \mu^* \rangle = y_\alpha$$

is the moment sequence of the optimal measure μ^*.

$\to$ approximate L^* by $L^{(r)}$ for $r \to \infty$.

Our setting for now

- $\mathcal{K} = \Delta_{n-1} = \{\mathbf{x} \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$
- $f_0, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ homogeneous of degree d
  - This is without loss of generality
- $\int_{\Delta_{n-1}} \mathrm{d}\mu^*(\mathbf{x}) \leq 1$ for μ^* optimal

The last bullet implies that we know an upper bound on the measure of Δ_{n-1} of the optimal solution. Since if we do, we may scale everything so that $\int_{\Delta_{n-1}} \mathrm{d}\mu(\mathbf{x}) \leq 1$ holds.

LP relaxation

Let $\mathcal{L}^{(r)} = \{L^{(r)} : \mathbb{R}[\mathbf{x}]_r \to \mathbb{R},\ \text{linear operator}\ \}$

$$\begin{aligned}
\underline{f}_{\text{LP}}^{(r)} = \min_{L^{(r)} \in \mathcal{L}^{(r)}} \ & L^{(r)}(f_0) = \sum_{\alpha \in \mathbb{N}_d^n} f_{0,\alpha} L^{(r)}(\mathbf{x}^\alpha) \\
\text{s.t.}\ \ & L^{(r)}(f_i) = b_i \ \forall i \in [m] \\
& L^{(r)}(1) \leq 1 \\
& L^{(r)}(\mathbf{x}^\alpha) \geq 0 \quad \forall\, |\alpha| \leq r \\
& L^{(r)}(\mathbf{x}^\alpha) = L^{(r)}\left(\mathbf{x}^\alpha \sum_{i=1}^n x_i\right) \quad \forall\, |\alpha| \leq r - 1.
\end{aligned}$$

Why is this an LP?

- Introduce variables y_α for $L^{(r)}(\mathbf{x}^\alpha)$ for each $|\alpha| \leq r$. The resulting program is an LP with $\binom{n+r}{r}$ variables
- $L^{(r)}(\mathbf{x}^\alpha) \geq 0 \quad \forall\, |\alpha| \leq r$
  - Necessary condition for a positive measure μ on the simplex $\int_{\Delta_{n-1}} \mathbf{x}^\alpha \mathrm{d}\mu(\mathbf{x}) \geq 0$
- $L^{(r)}(\mathbf{x}^\alpha) = L^{(r)}\left(\mathbf{x}^\alpha \sum_{i=1}^n x_i\right) \quad \forall\, |\alpha| \leq r - 1$
  - Because we want $L^{(r)}(p) = L^{(r)}(q)$ if $p(\mathbf{x}) = q(\mathbf{x})$ for all $\mathbf{x} \in \Delta_{n-1}$
- Note that for μ^* optimal setting $L^{(r)}(\cdot) = \langle \cdot, \mu^* \rangle$ is feasible, thus $\underline{f}^{(r)}_{\mathrm{LP}} \leq \mathrm{val}.$

Lemma 1

Let $r, k \in \mathbb{N}$ with $k \leq r$ and let $L^{(r)}$ be the solution to the linear relaxation. Then for all $\mathbf{x}^\gamma$ with $\gamma \in \mathbb{N}^n$ and $|\gamma| \leq r - k$ we have

$$L^{(r)}(\mathbf{x}^\gamma) = L^{(r)}\left(\mathbf{x}^\gamma \left(\sum_{i=1}^{n} x_i\right)^k\right)$$

Lemma 2

Consider the GMP and let $(y, t) \in \mathbb{R}^m \times \mathbb{R}_+$. Then the pair (y, t) is dual optimal only if

$$0 = \min_{\mathbf{x} \in \mathcal{K}} \left(f_0(\mathbf{x}) - \sum_{i=1}^{m} y_i f_i(\mathbf{x}) + t\right).$$

Theorem (Main tool)

Suppose $f \in \mathbb{R}[\mathbf{x}]$ homogeneous of degree d

$$f(\mathbf{x}) = \sum_{|\alpha|=d} f_\alpha \mathbf{x}^\alpha$$

is strictly positive on the simplex Δ_{n-1}. Let $\varepsilon = \min_{\mathbf{x}\in\Delta_{n-1}} f(\mathbf{x})$. Then,

$$(x_1 + \cdots + x_n)^k f(\mathbf{x})$$

has only positive coefficients if

- **Pólya** (1928): k is large enough
- **Powers, Reznick** (2000): $k > \binom{d}{2}\frac{B(f)}{\varepsilon} - d$, where $B(f) = \max_{|\alpha|=d} \frac{\alpha!}{d!}|f_\alpha|$
- **de Klerk et al.** (2006) $B(f) = \max_{|\alpha|=d} \frac{\alpha!}{d!} f_\alpha$

The key ingredient is that our $L^{(r)}$ is "compatible" with this positivity certificate.

- Firstly, $L^{(r)}(f) = L^{(r)}(f(x_1 + \cdots + x_n)^k)$ if $k + d \leq r$
- Secondly, suppose $g(\mathbf{x}) = \sum_{|\alpha|=k+d} c_\alpha \mathbf{x}^\alpha$ with $c_\alpha \geq 0$ for all $|\alpha| = k + d$, then $L^{(k+d)}(g) = \sum_{|\alpha|=k+d} c_\alpha L^{(k+d)}(\mathbf{x}^\alpha) \geq 0$

Theorem

Let val be the optimal value of the GMP for input data

- $\mathcal{K} = \Delta_{n-1}$
- $f_0, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ hom. of deg. d
- $b_i \in \mathbb{R}$ for $i \in [m]$
- $(\bar{y}, t) \in \mathbb{R}^m \times \mathbb{R}_+$ dual optimal

Set $f_{m+1}(\mathbf{x}) = (\sum_{i=1}^n x_i)^d = 1$ for all $\mathbf{x} \in \Delta_{n-1}$ and further set $\bar{y}_{m+1} = -t$. Then, setting $y_0 = 1$ and $y_i = -\bar{y}_i$ we have

$$0 \leq \mathrm{val} - \underline{f}_{\mathrm{LP}}^{(r)} \leq \frac{\sum_{i=0}^{m+1} B(y_i f_i) d(d-1)}{2(r-1) - d(d-1)}$$

Proof.

Let $L^{(r)}$ be an optimal solution to the relaxation and fix some $\varepsilon > 0$. Then,

$$
\begin{aligned}
0 \le \mathrm{val} - \underline{f}_{\mathrm{LP}}^{(r)} &= \mathrm{val} - L^{(r)}\left(\sum_{i=1}^m \bar{y}_i f_i - t + f_0 - \sum_{i=1}^m \bar{y}_i f_i + t\right) \\
&= \mathrm{val} - \sum_{i=1}^m \bar{y}_i L^{(r)}(f_i) + tL^{(r)}(1) - L^{(r)}\left(f_0 - \sum_{i=1}^m \bar{y}_i f_i + t\right) \\
&\le \mathrm{val} - \sum_{i=1}^m \bar{y}_i b_i + t - L^{(r)}\left(f_0 - \sum_{i=1}^m \bar{y}_i f_i + t\right) \\
&\le -L^{(r)}\left(f_0 - \sum_{i=1}^m \bar{y}_i f_i + t + \varepsilon\right) + \varepsilon
\end{aligned}
$$

Thus,

$$0 \leq \text{val} - \underline{f}_{\text{LP}}^{(r)} \leq \varepsilon - L^{(r)} \left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t + \varepsilon \right)$$

- $\Rightarrow f := \sum_{i=0}^{m+1} y_i f_i + \varepsilon \left(\sum_{i=1}^{n} x_i\right)^d$ is homogeneous of degree d
- Moreover, $\min_{\mathbf{x} \in \Delta_{n-1}} f(\mathbf{x}) = \varepsilon$
- $\rightarrow$ with main tool $\exists k : f(\mathbf{x}) \left(\sum_{i=1}^{n} x_i\right)^k = \sum_{\beta \in \mathbb{N}^n_{d+k}} c_\beta x^\beta$ with all $c_\beta \geq 0$.
- $k > \binom{d}{2} \frac{B(f)}{\varepsilon} - d$
- Recall $B(f) = \max_{|\alpha|=d} \frac{\alpha!}{d!} f_\alpha$
- $B(f) \leq \sum_{i=0}^{m+1} B(y_i f_i) + \varepsilon$

We find for r large enough, i.e.

$$\left\lceil \binom{d}{2} \frac{\sum_{i=0}^{m} B(y_i f_i) + \varepsilon}{\varepsilon} \right\rceil = k + d = r_\varepsilon$$

that

$$\begin{aligned} 0 \leq \mathrm{val} - \underline{f}_{\mathrm{LP}}^{(r_\varepsilon)} &\leq \varepsilon - L^{(r_\varepsilon)}(f) \\ &= \varepsilon - L^{(r_\varepsilon)} \left(f \left(\sum_{i=1}^{n} x_i \right)^k \right) \\ &= \varepsilon - \sum_{\beta \in \mathbb{N}^n_{d+k}} c_\beta L^{(r_\varepsilon)}(\mathbf{x}^\beta) \\ &\leq \varepsilon \\ &\leq \frac{\sum_{i=0}^{m+1} B(y_i f_i) d(d-1)}{2(r_\varepsilon - 1) - d(d-1)} \end{aligned}$$

Conceptually, the proof works like this:

- Let r be given and $L^{(r)}$ be a solution to a given relaxation
- We know $0 \leq \mathrm{val} - \underline{f}^{(r)} \leq -L^{(r)}\left(f_0 - \sum_{i=1}^{m} \bar{y}_i f_i + t\right)$
- Also $\min_{\mathbf{x} \in \mathcal{K}} f_0(\mathbf{x}) - \sum_{i=1}^{m} \bar{y}_i f_i(\mathbf{x}) + t = 0$
- Suppose we have a positivity certificate $\Gamma_{\varepsilon_r}(\mathbf{x})$, s.t.
  - $f_0(\mathbf{x}) - \sum_{i=1}^{m} \bar{y}_i f_i(\mathbf{x}) + t + \varepsilon_r = \Gamma_{\varepsilon_r}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{K}$
  - $L^{(r)}(\Gamma_{\varepsilon_r}) \geq 0$ if $r \geq \deg(\Gamma_{\varepsilon_r}) \in O({\varepsilon_r}^{-\gamma})$ for some $\gamma \in \mathbb{R}_+$
- Then $\mathrm{val} - \underline{f}^{(r)} \leq \varepsilon_r \in O(r^{-1/\gamma})$

Thus, this proof technique may be employed in different settings

- $\mathcal{K} = \mathcal{S}^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x}||_2^2 = 1\}$
- $f_0, f_1, \ldots, f_m$ homogeneous polynomials of even degree $2d$
- $[\mathbf{x}]_r$ monomial basis vector up to degree r

The Lasserre hierarchy of semidefinite relaxations of the GMP over the sphere is given by

$$\begin{aligned} \underline{f}_{\mathrm{SDP}}^{(2r)} = \min\ & L^{(2r)}(f_0) \\ \text{s.t.}\quad & L^{(2r)}(f_i) = b_i \quad \forall i \in [m] \\ & L^{(2r)}(1) \leq 1 \\ & L^{(2r)}\left([\mathbf{x}]_r[\mathbf{x}]_r^T\right) \succeq 0 \\ & L^{(2r)}(\mathbf{x}^\alpha) = L^{(2r)}\left(\mathbf{x}^\alpha ||\mathbf{x}||_2^2\right) \quad \forall \quad |\alpha| \leq 2r-2, \end{aligned}$$

where the $L^{(2r)}$ operator is now applied entry-wise to matrix-valued functions, where needed.

Lemma

Let $L : \mathbb{R}[\mathbf{x}]_{2k} \to \mathbb{R}$ *be a linear operator and suppose* $L\left([\mathbf{x}]_k[\mathbf{x}]_k^T\right) \succeq 0$*, where the operator is applied entrywise to the matrix* $[\mathbf{x}]_k[\mathbf{x}]_k^T$*. Then,* $L(\sigma) \geq 0$ *for all* $\sigma \in \Sigma[\mathbf{x}]_k$*.*

The Lasserre hierarchy for the problem of minimizing a polynomial p on the sphere is defined as

$$p_r = \max\left\{\gamma \in \mathbb{R} : p(\mathbf{x}) - \gamma = \sigma(\mathbf{x}) + (1 - ||\mathbf{x}||_2^2)h(\mathbf{x}) \right.$$
$$\left. \text{for } \sigma \in \Sigma[\mathbf{x}]_r, h \in \mathbb{R}[\mathbf{x}]_{2r-2}\right\}.$$

This scheme provides increasing lower bounds on $p_{\min} = \min_{\mathbf{x} \in \mathcal{S}^{n-1}} p(\mathbf{x})$, i.e., $p_{r-1} \leq p_r \leq p_{\min}$.

Fang and Fawzi proved the following theorem in [1].

Theorem

Assume $p(\mathbf{x})$ is a homogeneous polynomial of degree $2d$ in n variables with $d \leq n$, and let $p_{\min}$ (resp. $p_{\max}$) denote the minimum (resp. maximum) of p on $\mathcal{S}^{n-1}$. Then for any $r \geq C_d n$

$$1 \leq \frac{p_{\max} - p_r}{p_{\max} - p_{\min}} \leq 1 + C_d' \left(\frac{n}{r}\right)^2$$

for some constants C_d', C_d that depend only on d.

Essentially, if $p_{\min} = 0$ and $p_{\max} = 1$ this implies for $r \geq C_d n$ that $p + C_d' \left(\frac{n}{r}\right)^2 = \sigma(\mathbf{x}) + (1 - ||\mathbf{x}||_2^2)h(\mathbf{x})$ for $\sigma \in \Sigma[\mathbf{x}]_r$ and $h \in \mathbb{R}[\mathbf{x}]_{2r-2}$.

Theorem

- Let val be the optimal value of the GMP
- input data $\mathcal{K} = \mathcal{S}^{n-1}, f_0, f_1, \ldots, f_m \in \mathbb{R}[\mathbf{x}]$ homogeneous of even degree $2d$ and $d \leq n$
- $b_1, \ldots, b_m \in \mathbb{R}$ and $(\bar{y}, t)$ dual optimal
- $f_{m+1}(\mathbf{x}) := (\sum_{i=1}^{n} x_i^2)^d = 1$ for every $\mathbf{x} \in \mathcal{S}^{n-1}$
- For $i = \{0, 1, \ldots, m+1\}$ define $f_{\max}^{i,y_i} = \max_{\mathbf{x} \in \mathcal{S}^{n-1}} y_i f_i(\mathbf{x})$ for $y_i = -\bar{y}_i$ for $i \in [m]$ and $y_{m+1} = t$, $y_0 = 1$.

There exist constants C_d, C'_d, only dependent on d, such that if $r \geq C_d n$ we have

$$0 \leq \mathrm{val} - \underline{f}_{\mathrm{SDP}}^{(2r)} \leq \frac{C'_d n^2 \sum_{i=0}^{m+1} f_{\max}^{i,y_i}}{r^2}.$$

[1] K. Fang and H. Fawzi, The sum-of-squares hierarchy on the sphere and applications in quantum information theory, Mathematical Programming (2020)

[2] E. de Klerk, M. Laurent, and P. Parrilo. A PTAS for the minimization of polynomials of fixed degree over the simplex. Theoretical Computer Science, 361(2-3):210–225 (2006)

[3] J. B. Lasserre, Moments, Positive Polynomials and Their Applications, Imperial College Press, 10 (2009)

[4] A. Shapiro, On duality theory of conic linear problems, Semi-Infinite Programming, pages 135–165, 12 (2000)

Thank you for your attention.

Further reading: arXiv:2103.02924

Questions?