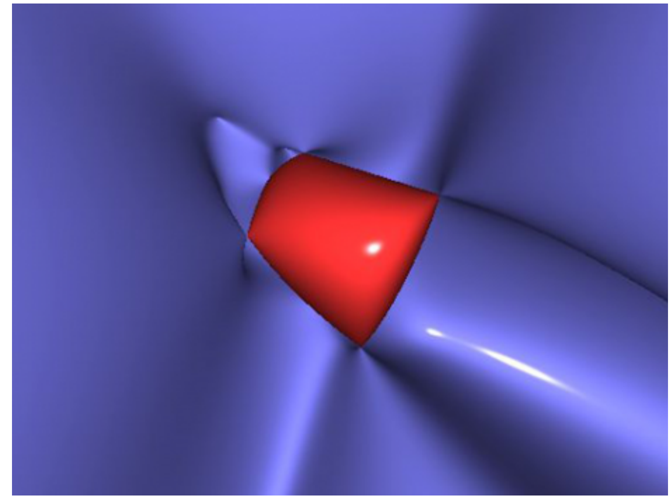
Polynomial Optimization, Efficiency through Moments and Algebra



Online learning weeks of the POEMA network

**Didier Henrion**

# Polynomial optimization (revisited)

27 May 2020

This first lecture on polynomial optimization illustrates the main steps of the moment-SOS aka Lasserre hierarchy

Given a **nonlinear nonconvex** problem:

1. Reformulate it as a **linear** problem (at the price of enlarging or changing the space of solutions)

2. Solve approximately the linear problem with a hierarchy of tractable **convex** relaxations (of increasing size)

3. Ensure **convergence**: either the original problem is solved at a finite relaxation size, or its solution is approximated with increasing quality

At each step, **conic duality** is an essential ingredient

# Conic duality

## Dual spaces and cones

Given a vector space $V$, its **dual space** $V'$ is the set of continuous linear functionals on $V$, with duality pairing $\langle x, y \rangle \in \mathbb{R}$ defined for all $x \in V$, $y \in V'$

Given a cone $K \subset V$, its **dual cone** $K' \subset V'$ is the set of positive continuous linear functionals on $K$ i.e.

$$K' := \{ y \in V' \, : \, \langle x, y \rangle \geq 0, \, \forall x \in K \}$$

$K'$ is always a **convex** cone, i.e. it is closed under addition and multiplication by a positive constant

## Familiar self-dual convex cones

The linear cone

$$\{x \in \mathbb{R}^n : x_1 \geq 0, \ldots, x_n \geq 0\}$$

the quadratic cone

$$\{x \in \mathbb{R}^n : x_1 \geq \sqrt{x_2^2 + \cdots + x_n^2}\}$$

and the semidefinite cone

$$\{X \in \mathbb{S}^n : \langle y, Xy \rangle \geq 0, \forall y \in \mathbb{R}^n\}$$

are all self-dual convex cones

## Infinite-dimensional convex cones

Given a compact set $X \subset \mathbb{R}^n$, the set of positive (non-negative) continuous **functions** on $X$ is an infinite-dimensional convex cone

$$C(X)_+ := \{f \in C(X) : f(x) \geq 0,\ \forall x \in X\}.$$

By a Riesz representation theorem, its dual is the cone of Borel regular positive measures (for short, **measures**)

$$C(X)'_+ := \{\mu \in C(X)' : \mu(A) \geq 0,\ \forall A \in \mathscr{B}(X)\}$$

and the duality pairing is integration

$$\langle f, \mu \rangle := \int_X f(x) d\mu(x)$$

for all $f \in C(X)$, $\mu \in C(X)'$

## Challenging finite-dimensional convex cones

The set of polynomials of degree at most $d$ which are positive on compact $X \subset \mathbb{R}^n$ is a finite-dimensional convex cone

$$P(X)_d := \{p \in \mathbb{R}[x]_d : p(x) = \sum_a p_a x^a \geq 0,\ \forall x \in X\} \subset \mathbb{R}^{\binom{n+d}{n}}$$

called the cone of **positive polynomials**

By the Riesz-Haviland theorem, its dual is the cone of **moments** of degree at most $d$ of measures on $X$

$$P(X)'_d := \{y \in \mathbb{R}^{\binom{n+d}{n}} : y_a = \int_X x^a d\mu(x),\ \mu \in C(X)'_+\}$$

and the duality pairing is integration

$$\langle p, y \rangle := \sum_a p_a y_a = \sum_a p_a \int_X x^a d\mu(x) = \int_X \sum_a p_a x^a d\mu(x) = \int_X p(x) d\mu(x)$$

for all $p \in P(X)_d$, $y \in P(X)'_d$

# Conic optimization

## Conic primal

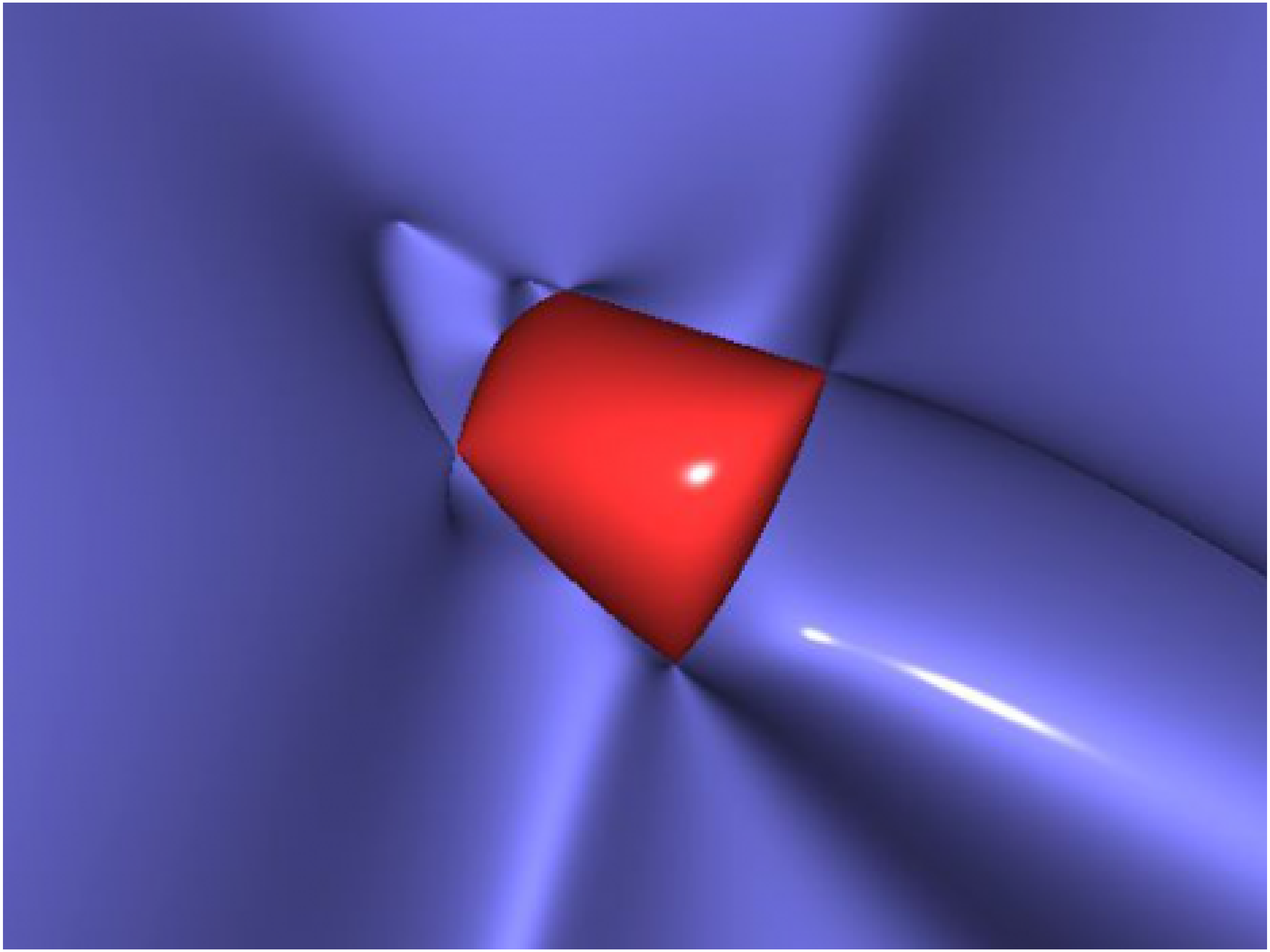
Conic optimization consists of minimizing a linear function on an affine section of a convex cone

$$
\begin{array}{rcll}
p^* & = & \text{inf}_x & \langle c, x \rangle \\
 & & \text{s.t.} & Ax = b \\
 & & & x \in K
\end{array}
$$

for given $A : V \to U'$, $b \in U'$, $c \in V'$ - let us call this the **primal**

When $K$ is a linear, quadratic resp. semidefinite cone, this is called linear, quadratic resp. **semidefinite programming**, and the feasibility region (the affine section of the cone) is called a polyhedron, conic resp. **spectrahedron**

Spectrahedra are defined by **linear matrix inequalities, LMIs**



# Conic duality

Given a conic problem

$$\begin{array}{rcll} p^* & = & \inf_x & \langle c, x \rangle \\ & & \text{s.t.} & Ax = b \\ & & & x \in K \end{array}$$

with data $A : V \to U'$, $b \in U'$, $c \in V'$, define the Lagrangian

$$\ell(x, y, z) := \langle c, x \rangle - \langle Ax - b, y \rangle - \langle z, x \rangle$$

where $x \in V$ are the **primal** variables, and $y \in U$, $z \in V'$ are Lagrange multipliers, or **dual** variables

Define the dual Lagrange function

$$d(y, z) := \inf_{x \in V} \ell(x, y, z)$$

and observe that $p^* \geq d(y, z)$ for all $y \in U$, $z \in K'$

# Conic dual

The tightest lower bound on $p^*$ is obtained with the **dual**

$$\begin{array}{rcll} d^* & = & \text{sup}_{y,z} & d(y,z) \\ & & \text{s.t.} & z \in K' \end{array}$$

Rearrange the dual function as follows

$$d(y,z) = \inf_{x \in V} (\langle c, x\rangle - \langle Ax - b, y\rangle - \langle z, x\rangle) = \langle b, y\rangle + \inf_{x \in V} \langle c - A'y - z, x\rangle$$

(where $A' : U \to V'$ is the adjoint map to $A$) and observe that it is finite only if $z = c - A'y$ so that the dual can be formulated as a conic problem

$$\begin{array}{rcll} d^* & = & \text{sup}_{y} & \langle b, y\rangle \\ & & \text{s.t.} & c - A'y \in K' \end{array}$$

## Primal and dual conic problems

To summarize, these two problems are in duality

$$\begin{array}{rclll} p^* & = & \inf_x & \langle c, x\rangle \\ & & \text{s.t.} & Ax = b \\ & & & x \in K \end{array} \qquad \begin{array}{rclll} d^* & = & \sup_y & \langle b, y\rangle \\ & & \text{s.t.} & z = c - A'y \\ & & & z \in K' \end{array}$$

Weak duality $p^* \geq d^*$ always holds, and strong duality also called **no duality gap** $p^* = d^*$ holds under some assumptions

If $x^*$ is primal optimal and $y^*$ is dual optimal, then $x^*$ and $z^* := c - A'y^*$ are **complementary**: $\langle x^*, z^*\rangle = 0$

If there is no duality gap then primal and dual optimality holds if and only if complementarity holds - one can then expect a good behavior of **numerical optimization algorithms**

# Polynomial optimization and the Lasserre hierarchy

## POP (Polynomial Optimization Problem)

Given polynomials $p, g_1, \ldots g_m \in \mathbb{R}[x]$ of the indeterminate $x \in \mathbb{R}^n$, consider the **nonlinear nonconvex** global optimization problem

$$\begin{array}{rcll} v^* & = & \text{min}_x & p(x) \\ & & \text{s.t.} & x \in X \end{array}$$

defined on the bounded basic semialgebraic set

$$X := \{x \in \mathbb{R}^n : g_1(x) \geq 0, \ \ldots \ g_m(x) \geq 0\}$$

## Step 1 - Linear reformulation

## Primal linear reformulation

Instead of the POP

$$\begin{array}{rcll} v^* & = & \min_x & p(x) \\ & & \text{s.t.} & x \in X \end{array}$$

consider the linear problem (LP)

$$\begin{array}{rcll} p^* & = & \inf_\mu & \langle p, \mu \rangle \\ & & \text{s.t.} & \langle 1, \mu \rangle = 1 \\ & & & \mu \in C(X)'_+ \end{array}$$

**Exercise 1.1:** Prove that $v^* = p^*$ and that the LP has an optimal solution equal to the Dirac measure at any optimal solution of the POP

## Dual linear reformulation

Dual to the primal LP

$$p^* = \inf_\mu \langle p, \mu \rangle \quad \text{s.t.} \quad \langle 1, \mu \rangle = 1, \quad \mu \in C(X)'_+$$

is the LP

$$d^* = \sup_{v \in \mathbb{R}} v \quad \text{s.t.} \quad p - v \in C(X)_+$$

**Exercise 1.2:** Derive the dual LP from the primal LP using convex duality. Prove that strong duality holds i.e. $p^* = d^*$. Give a graphical interpretation to the dual LP

## Step 2 - Convex hierarchy

## Moments and positive polynomials

POP is replaced with a primal and a dual LP

$$\begin{array}{rcll} p^* & = & \inf_\mu & \langle p, \mu \rangle \\ & & \text{s.t.} & \langle 1, \mu \rangle = 1 \\ & & & \mu \in C(X)'_+ \end{array} \qquad \begin{array}{rcll} d^* & = & \sup_{v \in \mathbb{R}} & v \\ & & \text{s.t.} & p - v \in C(X)_+ \end{array}$$

or equivalently

$$\begin{array}{rcll} p^* & = & \inf_y & \langle p, y \rangle \\ & & \text{s.t.} & \langle 1, y \rangle = 1 \\ & & & y \in P(X)'_d \end{array} \qquad \begin{array}{rcll} d^* & = & \sup_{v \in \mathbb{R}} & v \\ & & \text{s.t.} & p - v \in P(X)_d \end{array}$$

since $p$ is a degree $d$ polynomial

## Approximating positive polynomials

The cone of positive polynomials

$$P(X)_d := \{p \in \mathbb{R}[x]_d : p(x) \geq 0, \forall x \in X\}$$

on the bounded basic semialgebraic set

$$X := \{x \in \mathbb{R}^n : g_k(x) \geq 0, k = 1, \ldots, m\}$$

is generally **intractable**, so we will approximate it

Denoting $g_0(x) := 1$ and enforcing (without loss of generality) $g_1(x) := R^2 - \sum_{i=1}^n x_i^2$ for $R$ large enough, consider for $r \geq d$

$$Q(X)_{d,r} := \{q \in \mathbb{R}[x]_d : q = \sum_{k=0}^{m} s_k g_k,\ \ s_k \in \Sigma,\ \ s_k q_k \in \mathbb{R}[x]_r\}$$

where $\Sigma$ denotes polynomial sums of squares (SOS), and observe that it is an **inner approximation**: $Q(X)_{d,r} \subset P(X)_d$

## Polynomial SOS

Observe that (the truncated quadratic module)

$$Q(X)_{d,r} := \{q \in \mathbb{R}[x]_d : q = \sum_{k=0}^{m} s_k g_k,\ \ s_k \in \Sigma,\ \ s_k g_k \in \mathbb{R}[x]_r\}$$

is a projection of the SOS cone

Observe also that by construction

$$Q(X)_{d,r} \subset Q(X)_{d,r+1}$$

**Exercise 1.3:** Prove that deciding whether a polynomial is SOS can be reduced to semidefinite programming

## Moment relaxations

Hence we have a hierarchy of tractable **inner approximations** for the cone of positive polynomials

$$Q(X)_{d,r} \subset Q(X)_{d,r+1} \subset P(X)_d$$

Using convex duality, we also have a hierarchy of tractable **outer approximations**, or **relaxations**, for the cone of moments

$$Q(X)'_{d,r} \supset Q(X)'_{d,r+1} \supset P(X)'_d$$

Elements of $Q(X)'_{d,r}$ are sometimes called pseudo-moments

> **Exercise 1.4:** Describe explicitly the moment cone relaxation $Q(X)'_{d,r}$ as the projection of a spectrahedron

## Moment-SOS hierarchy

Replace the intractable problems

$$p^* = \inf_y \langle p, y \rangle \quad \text{s.t.} \quad \langle 1, y \rangle = 1, \quad y \in P(X)'_d \qquad\qquad d^* = \sup_{v \in \mathbb{R}} v \quad \text{s.t.} \quad p - v \in P(X)_d$$

with the hierarchy of tractable problems for $r = d, d+1, \ldots$

$$p_r^* = \inf_y \langle p, y \rangle \quad \text{s.t.} \quad \langle 1, y \rangle = 1, \quad y \in Q(X)'_{d,r} \qquad\qquad d_r^* = \sup_{v \in \mathbb{R}} v \quad \text{s.t.} \quad p - v \in Q(X)_{d,r}$$

**Exercise 1.5:** Prove that strong duality holds: $v_r^* := p_r^* = d_r^*$

## Step 3 - Convergence

## Convergence

Integer $r$ is called the **relaxation order**

Since $Q(X)_{d,r} \subset Q(X)_{d,r+1} \subset P(X)_d$, we have a monotone non-decreasing sequence of lower bounds on the POP value:

$$v_r^* \leq v_{r+1}^* \leq v^*$$

**Theorem (Putinar 1993):** $\overline{Q(X)_{d,\infty}} = P(X)_d$

**Theorem (Lasserre 2001):** $v_\infty^* = v^*$

The moment-SOS hierarchy is known as the **Lasserre hierarchy**

## Finite convergence

**Theorem (Nie 2014):** Generically $\exists r < \infty$ such that $v_r^* = v^*$

In other words, a vanishing small random perturbation of the input data of a given POP ensures finite convergence of the Lasserre hierarchy

We also have sufficient linear algebra conditions to ensure finite convergence, certify global optimality and extract minimizers

## Software

The moment-SOS hierarchy is implemented in the **GloptiPoly** package for Matlab (2002)

Conic relaxations are solved by a dedicated solver, e.g. **SeDuMi**, **MOSEK** or **PENSDP**

Julia packages are currently being developed, e.g. **MomentOpt** by Tillmann Weisser or **MomentTools** by Lorenzo Baldi and Bernard Mourrain

# Summary

This first lecture on polynomial optimization illustrates the main steps of the moment-SOS aka Lasserre hierarchy

Given a **nonlinear nonconvex** problem:

1. Reformulate it as a **linear** problem (at the price of enlarging or changing the space of solutions)

2. Solve approximately the linear problem with a hierarchy of tractable **convex** relaxations (of increasing size)

3. Ensure **convergence**: either the original problem is solved at a finite relaxation size, or its solution is approximated with increasing quality

At each step, **conic duality** is an essential ingredient

## Teaser

In the next lectures we will bring some **dynamics** to polynomial optimization and the Lasserre hierarchy

First we will approximate the maximal positively invariant set for a discrete dynamical system

Then we will approximate the value function for the optimal control of ordinary differential equations

A key technical ingredient will be **occupation measures** and approximation of their support with **polynomial sublevel sets**