Scribe: Joyjit Roy

Polynomial Optimization

Today, we will be looking at the problem of **globally** optimizing polynomials, either with no constraints or with polynomial constraints. Let's start with some definitions.

- A monomial $\mathbf{z}^\alpha$ for $\mathbf{z} \in \mathbb{R}^n$ and $\alpha \in \mathbb{Z}_+^n$ is used to denote the polynomial

$$\mathbf{z}^\alpha := z_1^{\alpha_1} z_2^{\alpha_2} \dots z_n^{\alpha_n}.$$

- For $\alpha \in \mathbb{Z}_+^n$, we define $|\alpha|$ as follows

$$|\alpha| = \sum_{i=1}^n \alpha_i.$$

Any sum over $|\alpha| \leq m$ will henceforth mean a sum over all $\alpha \in \mathbb{Z}_+^n$ satisfying the inequality.

We will be interested in minimizing polynomials of the form

$$\mathbf{p}(\mathbf{z}) = \sum_{|\alpha| \leq d} p_\alpha \mathbf{z}^\alpha \tag{1}$$

where without loss of generality we assume $p_\mathbf{0} = 0$. We will use $\mathbf{p}$ to denote both the polynomial in (1) as well as the vector

$$\mathbf{p} = (p_\alpha)_{|\alpha| \leq d}.$$

We can assume (at least in the unconstrained case) that $d = 2m$ as otherwise the "leading term" in $\mathbf{p}$ is of odd degree and the one-dimensional restriction of $\mathbf{p}$ along some suitable line goes to $-\infty$ as $\mathbf{z}$ goes to $-\infty$ along that line.

The set of poynomials approximates the set of all continuous functions on compact sets so minimizing polynomials globally cannot be a trivial problem. To get an idea of how hard the problem we are dealing with is, we compare it to Integer Programming problems. Note that obtaining a feasible solution to

$$\mathbf{A}\mathbf{z} = \mathbf{b}, \quad \mathbf{z} \in \{0, 1\}^n$$

is equivalent to minimizing

$$\|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2 + \sum_{i=1}^n (z_i - z_i^2)^2.$$

So even for $m = 2$ (quartics), the polynomial optimization problem is NP-hard.

Obviously, $\min_{\mathbf{z}} \mathbf{p}(\mathbf{z})$ is not a convex problem but we can reformulate it as a convex problem as follows:

$$\min \left\{ \int \mathbf{p}(\mathbf{z})\, d\mu(\mathbf{z}) : \mu \text{ is a probability distribution on } \mathbb{R}^n \right\}.$$

The set above is unfortunately infinite-dimensional, but note that

$$\int \mathbf{p}(\mathbf{z})\, d\mu(\mathbf{z}) = \sum_{|\alpha|\leq 2m} p_\alpha \int_{\mathbb{R}^n} \mathbf{z}^\alpha\, d\mu(\mathbf{z}) = \sum_{|\alpha|\leq 2m} p_\alpha y_\alpha,$$

where $(y_\alpha)_{|\alpha|\leq 2m}$ is the moment vector (up to degree $2m$) of μ.

This gives us a finite-dimensional problem of choosing $\mathbf{y} = (y_\alpha)$ to minimize a linear function $\sum_{|\alpha|\leq 2m} p_\alpha y_\alpha$ such that $\mathbf{y}$ has $y_{\mathbf{0}} = 1$ (as μ is a probability measure) and $\mathbf{y}$ lies in the **moment cone** of moments of measures on $\mathbb{R}^n$.

Now let us **relax** this last constraint. For $\mathbf{y}$ to lie in the moment cone, we must have $y_{2\mathbf{e_1}} \geq 0$ as the second moment of a univariate random variable is always nonnegative. More generally, for any polynomial $\mathbf{q}$, we must have $\int \mathbf{q}^2(\mathbf{z})\, d\mu(\mathbf{z}) \geq 0$. In particular, for any polynomial $\mathbf{q} = \sum_{|\alpha|\leq m} q_\alpha \mathbf{z}^\alpha$, we must have

$$\sum_{|\alpha|\leq 2m} \Big(\sum_{\substack{\beta+\gamma=\alpha \\ |\beta|\leq m \\ |\gamma|\leq m}} q_\beta q_\gamma \Big) y_\alpha \geq 0.$$

Interchanging the summations above, we have that for any $\mathbf{q} = (\mathbf{q}_\alpha)_{|\alpha|\leq m}$,

$$\sum_{|\beta|\leq m} \sum_{|\gamma|\leq m} q_\beta (y_{\beta+\gamma}) q_\gamma \geq 0. \tag{2}$$

With this in mind, we define $\mathbf{M}(\mathbf{y})$ to be the symmetric matrix with rows and columns indexed by β with $|\beta| \leq m$ and γ with $|\gamma| \leq m$ respectively, and the entry $y_{\beta+\gamma}$ in the corresponding row and column. We demonstrate the idea with a simple example.

Example 1 *If $m = n = 2$, we have*

$$\mathbf{M}(\mathbf{y}) = \begin{array}{c} \\ 1 \\ z_1 \\ z_2 \\ z_1^2 \\ z_1 z_2 \\ z_2^2 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & z_1 & z_2 & z_1^2 & z_1 z_2 & z_2^2 \end{array} \\ \begin{pmatrix} y_{00} & y_{10} & y_{01} & y_{20} & y_{11} & y_{02} \\ y_{10} & y_{20} & y_{11} & y_{30} & y_{21} & y_{12} \\ y_{01} & y_{11} & y_{02} & y_{21} & y_{12} & y_{03} \\ y_{20} & y_{30} & y_{21} & y_{40} & y_{31} & y_{22} \\ y_{11} & y_{21} & y_{12} & y_{31} & y_{22} & y_{13} \\ y_{02} & y_{12} & y_{03} & y_{22} & y_{13} & y_{04} \end{pmatrix} \end{array}.$$

Note that y_{ij} denotes the α-th element of $\mathbf{y} = (y_\alpha)$ for $\alpha = (i, j)$.

Inequality (2) essentially implies that $\mathbf{M}(\mathbf{y}) \succeq 0$. So we have the following relaxation for the Polynomial Optimization problem

$$
(D) \qquad \begin{array}{lrcl} \min_{\mathbf{y}} & \sum\limits_{|\alpha|\le 2m} p_\alpha y_\alpha & & \\ & y_{\mathbf{0}} & = & 1, \\ & \mathbf{M}(\mathbf{y}) & \succeq & 0. \end{array}
$$

If we substitute 1 for $y_{\mathbf{0}}$ above , we get an SDP in the dual form.

Note that the dimension of $\mathbf{y}$ is $\binom{2m+n}{n} - 1$ and the order of $\mathbf{M}(\mathbf{y})$ is $\binom{m+n}{n}$. The dimension of the matrix is simply the number of nonnegative integer solutions of $\alpha_1 + \alpha_2 + \ldots + \alpha_n \le m$ and the fact that this number is equal to $\binom{m+n}{n}$ can be seen if we look at the bijection, defined as in the following example for $m = 10, n = 3$, where we have "chosen" the $n = 3$ larger lines out of the $m + n = 13$ lines.

$$
\underbrace{|\ \ |\ \ |}_{m-|\alpha|=3} \quad \overbrace{\Big|}^{1} \quad \underbrace{|\ \ |}_{\alpha_1=2} \quad \overbrace{\Big|}^{2} \underbrace{|\ \ |\ \ |\ \ |}_{\alpha_2=4} \quad \overbrace{\Big|}^{3} \quad \underbrace{|}_{\alpha_3=1}
$$

is equivalent to the monomial

$$
z_1^2 z_2^4 z_3.
$$

Observe that globally minimizing the polynomial is equivalent to minimizing $\int \mathbf{p}(\mathbf{z})\ d\mu(\mathbf{z})$ over Dirac measures μ. Instead we allow μ to be any probability measure and then relax the conditions for $\mathbf{y}$ to belong to the moment cone.

To understand (D) and how tight a relaxation it gives, we consider its dual:

$$
(P) \qquad \begin{array}{lrcl} \max_{\mathbf{X}} & -1.x_{\mathbf{00}} & & \\ & \sum\limits_{\substack{\beta+\gamma=\alpha \\ |\beta|\le m \\ |\gamma|\le m}} x_{\beta\gamma} & = & p_\alpha, \quad |\alpha| \le 2m \text{ (except } \alpha = \mathbf{0}) \\ & \mathbf{X} & \succeq & 0. \end{array}
$$

For any feasible solution $\mathbf{X}$ to (P), we can factor

$$
\mathbf{X} = \sum_i \lambda_i (\mathbf{q}^i)(\mathbf{q}^i)^T
$$

and hence

$$
x_{\beta\gamma} = \sum_i \lambda_i (q^i_\beta)(q^i_\gamma),
$$

where all the λ_i's are nonnegative. So we have,

$$
p_\alpha = \sum_i \lambda_i \Big(\sum_{\substack{\beta+\gamma=\alpha \\ |\beta|\le m \\ |\gamma|\le m}} q^i_\beta q^i_\gamma \Big) = \sum_i \lambda_i r^i_\alpha,
$$

where $\mathbf{r}^i(\mathbf{z}) = (\mathbf{q}^i(\mathbf{z}))^2$. This leads to the fact that $\mathbf{p}(\mathbf{z}) + x_{\mathbf{00}}$ is $\sum_i \lambda_i(\mathbf{q}^i(\mathbf{z}))^2$, which is a **sum of squares** (SOS)!

Hence (P) is equivalent to finding the minimum $x_{\mathbf{00}}$ such that $\mathbf{p}(\mathbf{z}) + x_{\mathbf{00}}$ is an SOS. So the difference between our original problem and the SDP relaxation is the "difference" between the cone of polynomials which are nonnegative everywhere and the smaller cone of polynomials that can be written as a sum of squares. This "difference" is related to Hilbert's 17'th problem. In fact it is known (even Hilbert knew!) that the two cones are equivalent when

- $n = 1$,
- $m = 1$, or
- $n = m = 2$,

and in no other cases.

We look at a few examples to illustrate this difference.

Example 2 *For even n, the polynomial*

$$z_1^n + z_2^n + \ldots + z_n^n - nz_1z_2\ldots z_n$$

is an SOS, leading to a simple proof of the arithmetic-geometric mean (AM-GM) inequality for even n. The case for odd n is a simple extension.

Example 3 *The polynomial*

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

is always nonnegative (a simple consequence of the AM-GM inequality) but is not an SOS. This is the famous Motzkin polynomial.

Actually Hilbert posed the following question: If $\mathbf{p}$ is psd, then is it the sum of squares of rational functions (which are nothing but quotients of polynomials), i.e., can we write $\mathbf{r}(\mathbf{z})\mathbf{p}(\mathbf{z})$ as an SOS where $\mathbf{r}(\mathbf{z})$ is a square, or equivalently an SOS? This was answered affirmatively by E. Artin in 1926. For example, for the Motzkin polynomial, $(x^2 + y^2)p_M(x, y, z)$ is an SOS. In fact, if $\mathbf{p}(\mathbf{z})$ is **positive** on $\mathbb{R}^n$, then $(z_1^2 + z_2^2 + \ldots + z_n^2)^r\mathbf{p}(\mathbf{z})$ is an SOS for some finite r. This gives us a hierarchy of SDP problems, converging to the value of the original polynomial optimization problem.

Constrained Polynomial Optimization

We can also look at the problem of minimizing $\mathbf{p}(\mathbf{z})$ subject to $\mathbf{z} \in K$, where $K = \{\mathbf{q}_1(\mathbf{z}) \leq 0, \ldots, \mathbf{q}_r(\mathbf{z}) \leq 0\}$ where we assume that $\{\mathbf{z} : \mathbf{q}_1(\mathbf{z}) \leq 0\}$ is compact, and all functions are polynomials. The following theorem will prove useful in this regard.

Theorem 1 *If $\mathbf{p}(\mathbf{z})$ is positive on K, then we can write*

$$\mathbf{p}(\mathbf{z}) + \sum_{i=1}^{r} \mathbf{u}_i(\mathbf{z})\mathbf{q}_i(\mathbf{z}) = \mathbf{r}(\mathbf{z}),$$

where the $\mathbf{u}_i$'s and $\mathbf{r}$ are SOS.

This gives us a short certificate of $\mathbf{p}$'s nonnegativity and leads to SDP relaxations. Also, observe that the $\mathbf{u}_i$'s are like Lagrange multipliers.

Note that to get an SDP problem, we need to put a bound on the degrees of the $\mathbf{u}_i$'s and $\mathbf{r}$.