16 Sum-of-squares relaxations for constrained problems: the case of the hypercube

In the last couple of lectures we looked at unconstrained polynomial optimisation, and at the problem of deciding global nonnegativity of a polynomial on $\mathbb{R}^n$. Today we will look at constrained polynomial optimisation. For concreteness we will look at the case of polynomial optimisation on the hypercube $X = \{-1, 1\}^n$.

We saw in Lecture 9 the maximum cut problem which is the problem of maximising a quadratic function (the Laplacian of a graph) on the hypercube:

$$\max \quad x^T L_G x \quad : \quad x \in \{-1, 1\}^n.$$

By the usual argument we can rewrite the maximum cut problem as:

$$\min \quad \gamma \quad : \quad \gamma - x^T L_G x \text{ is nonnegative on } \{-1, 1\}^n.$$

We are thus interested in understanding nonnegative polynomials on $\{-1, 1\}^n$.

One way to certify that a function f is nonnegative on $\{-1, 1\}^n$ is to try to express it in the following way:

$$f(x) = \sum_{i=1}^{l} q_i(x)^2 + \sum_{i=1}^{n} (x_i^2 - 1) h_i(x). \tag{1}$$

where q_i and h_i are arbitrary polynomials. It is clear that any f of the form (1) is nonnegative on $\{-1, 1\}^n$. For example consider the function $f(x) = 1 + x_1$. Clearly f is nonnegative on $\{-1, 1\}^n$ and one verify that we have the following certificate of nonnegativity $1 + x_1 = \frac{1}{2}(1 + x_1)^2 + (x_1^2 - 1) \cdot (-1/2)$.

Functions on the hypercube can be expressed in a specific basis, called the basis of *square-free monomials* (or *multilinear monomials*). A square-free monomial is a monomial of the form $x^S := \prod_{i \in S} x_i$ where $S \subseteq [n]$.

Proposition 16.1. *Any function $f : \{-1, 1\}^n \to \mathbb{R}$ can be expressed as*

$$f(x) = \sum_{S \subseteq [n]} f_S x^S \quad \forall x \in \{-1, 1\}^n \tag{2}$$

for some coefficients $(f_S)_{S \subseteq [n]}$.

Proof. Let $a \in \{-1, 1\}^n$ and let $\delta_a(x)$ be the function that takes value 1 at a and 0 elsewhere. Note that δ_a can be expressed as:

$$\frac{1}{2^n} \prod_{i=1}^{n} (1 + a_i x_i).$$

Expanding the product we see that δ_a is a linear combination of the square-free monomials. Finally since each function is a linear combination of the δ_as we get the desired result. □

Definition 16.1. We say that a function $f : \{-1, 1\}^n \to \mathbb{R}$ is *k-sos on* $\{-1, 1\}^n$ if it is a sum-of-squares of polynomials of degree at most k on $\{-1, 1\}^n$, i.e., if there exists polynomials $q_1, \ldots, q_l$ of degree at most k such that $f(x) = \sum_{i=1}^{l} q_i(x)^2$ for all $x \in \{-1, 1\}^n$.

Remark 1. *One can show (using e.g., division for multivariate polynomials) that f is k-sos on $\{-1,1\}^n$ if and only if it can expressed as (1) where $\deg q_i \le k$ for all $i = 1,\dots,l$ and $\deg h_i \le 2k-2$ for all $i = 1,\dots,n$ (assuming $\deg f \le 2k$).*

Example 16.1.
- *The function $f(x) = 1 + x_1$ is 1-sos on $\{-1,1\}^n$ because $1 + x_1 = \frac{1}{2}(1 + x_1)^2$ on $\{-1,1\}^n$.*
- *Any nonnegative function f on $\{-1,1\}^n$ is n-sos. Indeed we have $f = g^2$ on $\{-1,1\}^n$ where $g : \{-1,1\}^n \to \mathbb{R}$ is defined by $g(x) = \sqrt{f(x)}$. By Proposition 16.1 we know that g is a polynomial of degree at most n.*

Degree cancellations: There is an important difference that one must keep in mind between (i) sum-of-squares certificates on the hypercube, and (ii) global sum-of-squares certificates. We saw in Lecture 14 that if $f(x) = \sum_{i=1}^{l} q_i(x)^2$ for all $x \in \mathbb{R}^n$ then necessarily $\deg q_i \le (\deg f)/2$. When working on $\{-1,1\}^n$ however, such degree bounds on the q_i's do not hold anymore as there can be *degree cancellations*. This is already evident in the two examples above.

Exercise 16.1. *Show that any nonnegative polynomial of degree 1 on the hypercube is 1-sos.*

The next theorem shows that deciding whether a function $f : \{-1,1\}^n \to \mathbb{R}$ is k-sos is a semidefinite feasibility problem.

Theorem 16.1. *A function $f : \{-1,1\}^n \to \mathbb{R}$ is k-sos on $\{-1,1\}^n$ if and only if there exists a positive semidefinite matrix Q of size $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k}$ such that*

$$f_S = \sum_{\substack{U,V \subseteq [n] \\ |U|,|V| \le k \\ U \triangle V = S}} Q_{U,V}$$

where f_S is the coefficient of f in the expansion (2), and $U \triangle V$ is the symmetric difference of U and V, i.e., $U \triangle V = (U \setminus V) \cup (V \setminus U)$.

Proof. The proof is very similar to Theorem 14.2. Simply use the fact that $x^U x^V = x^{U \triangle V}$ on $\{-1,1\}^n$. ☐

Example: maximum cut Recall the maximum cut problem:

$$\begin{array}{ll} \text{maximise} & x^T L_G x \\ \text{subject to} & x \in \{-1,1\}^n \end{array} \quad = \quad \begin{array}{ll} \text{minimise} & \gamma \\ \text{subject to} & \gamma - x^T L_G x \text{ nonnegative on } \{-1,1\}^n \end{array} \tag{3}$$

where L_G is the Laplacian of the graph G. The semidefinite relaxation of the maximum cut problem that we defined in Lecture 9 takes the form (we have also written the dual minimisation problem; note that strong duality holds because, e.g., $X = I_n$ is strictly feasible for the maximisation problem):

$$\begin{array}{ll} \text{maximise} & \mathrm{Tr}(L_G X) \\ \text{subject to} & X \succeq 0, \quad X_{ii} = 1 \ \forall i = 1,\dots,n. \end{array} \quad = \quad \begin{array}{ll} \text{minimise} & \sum_{i=1}^n \lambda_i \\ \text{subject to} & \mathrm{diag}(\lambda) - Z = L_G, \quad Z \succeq 0. \end{array} \tag{4}$$

Consider now the following relaxation of (3) where we have replaced the nonnegativity constraint by a "1-sos" constraint

$$\begin{array}{ll} \text{minimise} & \gamma \\ \text{subject to} & \gamma - x^T L_G x \text{ is 1-sos on } \{-1,1\}^n. \end{array} \tag{5}$$

Using Theorem 16.1 we can express (5) as a semidefinite optimisation problem of size $\binom{n}{0} + \binom{n}{1} = 1 + n$. In fact if we write this problem explicitly we end up exactly with (4). To see why this is the case, we know from Theorem 16.1 that $\gamma - x^T L_G x$ is 1-sos on $\{-1,1\}^n$ if and only if there exists $Q \in \mathbf{S}_+^{1+n}$ whose rows and columns are indexed by subsets of cardinality at most 1, such that [1]

$$\begin{aligned} (S = \emptyset) \quad &: \quad \gamma - \mathrm{Tr}(L_G) = \sum_{U \subseteq [n]} Q_{U,U} \\ (S = \{i\}, i \in [n]) \quad &: \quad 0 = 2Q_{\emptyset,\{i\}} \\ (S = \{i,j\}, i \neq j) \quad &: \quad -2(L_G)_{ij} = 2Q_{\{i\},\{j\}}. \end{aligned}$$

In other words (5) is the same as:

$$\begin{array}{ll} \text{minimise} & \gamma \\ \text{subject to} & Q = \begin{bmatrix} Q_{\emptyset,\emptyset} & 0 \\ 0 & Z \end{bmatrix} \succeq 0 \\ & Z_{ij} = -(L_G)_{ij} \quad \forall i \neq j \\ & Q_{\emptyset,\emptyset} + \mathrm{Tr}(Z) = \gamma - \mathrm{Tr}(L_G). \end{array} \tag{6}$$

Since $Q_{\emptyset,\emptyset} \geq 0$ it is not difficult to see that the optimal solution of (6) will always have $Q_{\emptyset,\emptyset} = 0$. Thus the problem is equivalent to

$$\begin{array}{ll} \text{minimise} & \mathrm{Tr}(L_G + Z) \\ \text{subject to} & Z_{ij} = -(L_G)_{ij} \quad \forall i \neq j \\ & Z \succeq 0. \end{array} \tag{7}$$

It is easy to verify that (7) is the same as the minimisation problem in (4).

Sum-of-squares hierarchy for maxcut In general we can define a hierarchy of semidefinite relaxations for the maximum cut problem (3):

$$v_k = \min \quad \gamma \quad : \quad \gamma - x^T L_G x \text{ is } k\text{-sos on } \{-1,1\}^n.$$

One can verify that $v_1 \geq v_2 \geq \cdots \geq v_n = \mathrm{maxcut}(G)$ where $\mathrm{maxcut}(G)$ is the value of the maximum cut (i.e., the optimal value of (3)). The equality $v_n = \mathrm{maxcut}(G)$ follows from the fact that any nonnegative function on $\{-1,1\}^n$ is n-sos (see second bullet point of Example 16.1). We showed above that the value v_1 coincides with the value of the Goemans-Williamson relaxation which we proved in Lecture 10 satisfies $v_1 \geq \mathrm{maxcut}(G) \geq 0.878 v_1$.

[1] Note that the constant coefficient in $\gamma - x^T L_G x$ is $\gamma - \mathrm{Tr}(L_G)$ (and not just γ as I mistakenly wrote on the blackboard) since $x^T L_G x = \sum_{i=1}^n (L_G)_{ii} x_i^2 + \sum_{i \neq j} (L_G)_{ij} x_i x_j = \mathrm{Tr}(L_G) + \sum_{i \neq j} (L_G)_{ij} x_i x_j$ since $x_i^2 = 1$ on the hypercube.