PSET 2

Due on **Wednesday Oct 9, 2024 at 11:59 pm ET**

COMPSCI 2243: Algorithms for Data Science, Fall 2024 (Sitan Chen) Version: **1.0**

Submitted by <**name**>

This problem set will cover concepts from the second unit on sum-of-squares optimization.
The three questions have been labeled with the date of the lecture in which the relevant material is covered, to help you budget your time. The questions are meant to be challenging, so as with previous psets, do not feel discouraged if you get stuck and are unable to solve some of them. If you find that you are running low on time to finish all the problems, our recommendation is to try to aim for breadth rather than depth – e.g., it is better to complete a few parts of each of the three questions, than to completely solve one of the three questions and skip the others.

1 (40 PTS.) DOTTING i'S AND CROSSING t'S WITH SUM-OF-SQUARES (9/18)

Motivation: Pseudo-distributions can be quite intimidating at first glance; this exercise is designed to get you more comfortable with these objects by digging into some of the technical steps that were swept under the rug in lecture. For example, in class, it was claimed but not completely proven that the set of pseudo-distributions over solutions to a system of polynomial inequalities is a finite-dimensional convex object that one can optimize over. In this exercise, we will work out the gory details for this claim, as well as miscellaneous other technical claims about pseudo-distributions that were asserted without proof.

Let $t \geqslant 2$ be an even integer, and let $z_1, \ldots, z_n$ denote scalar-valued SoS variables. Let $\mathbb{R}[z_1, \ldots, z_n]$ denote the ring of polynomials in $z_1, \ldots, z_n$ with real-valued coefficients. For $p \in \mathbb{R}[z_1, \ldots, z_n]$, we will often denote $p(z_1, \ldots, z_n)$ by the shorthand $p(z)$. Let $\mathcal{S}_t$ denote the ordered tuples $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geqslant 0}$ such that $\sum_i \alpha_i \leqslant t$. Given $\alpha \in \mathcal{S}_t$, let $z_\alpha \triangleq \prod_i z_i^{\alpha_i}$. Let $N_t \triangleq |\mathcal{S}_t|$.

Let $p_1, \ldots, p_m \in \mathbb{R}[z_1, \ldots, z_n]$ be of degree at most t. Recall that a functional $\tilde{\mathbb{E}} : \mathbb{R}_{\leqslant t}[z_1, \ldots, z_n] \to \mathbb{R}$ is said to be a *degree-t pseudo-expectation* over $z_1, \ldots, z_n$ that satisfies the constraints

$$p_1(z) \geqslant 0, \ldots, p_m(z) \geqslant 0 \tag{1}$$

if the following properties are satisfied:

- **Normalization**: $\tilde{\mathbb{E}}[1] = 1$,
- **Linearity**: $\tilde{\mathbb{E}}[a \cdot p + b \cdot q] = a \cdot \tilde{\mathbb{E}}[p] + b \cdot \tilde{\mathbb{E}}[q]$ for all $a, b \in \mathbb{R}$ and polynomials $p, q \in \mathbb{R}[z_1, \ldots, z_n]$ of degree at most t,
- **Positivity**: $\tilde{\mathbb{E}}[p(z)] \geqslant 0$ for all polynomials p which are degree-$\leqslant t$ sums of squares, that is, $p(z) = \sum_i q_i(z)^2$ for some polynomials q_i of degree at most $t/2$.
- **Respecting constraints**:

$$\tilde{\mathbb{E}}\Big[\sum_{S \subseteq [m]} f_S(z) \cdot \prod_{j \in S} p_j(z)\Big] \geqslant 0\,. \tag{2}$$

  for any polynomials $\{f_S\}_{S \subseteq [m]}$ which are sums of squares such that the polynomial $f_S(z) \cdot \prod_{i \in S} p_i(z)$ has degree at most t for every $S \subseteq [m]$. For any polynomial inequality of the form Eq. (2), we say that it has a *degree-$\leqslant t$ SoS proof* using the constraints $p_1(z) \geqslant 0, \ldots, p_m(z) \geqslant 0$. Formally, the proof is specified by the polynomials $\{f_S\}_{S \subseteq [m]}$. You may also find **these notes** helpful for these definitions.

1.A. (5 PTS.) Define the *moment matrix* $M \in \mathbb{R}^{N_{t/2} \times N_{t/2}}$ for $\tilde{\mathbb{E}}$ to be the matrix whose rows and columns are indexed by ordered tuples in $\mathcal{S}_{t/2}$, where

$$M_{\alpha,\beta} = \tilde{\mathbb{E}}[z_\alpha z_\beta]\,.$$

Prove that the "positivity" condition for $\tilde{\mathbb{E}}$ is satisfied if and only if the moment matrix M is a positive semidefinite matrix (you may assume linearity is satisfied).

1.B. (9 PTS.) Write down a collection of linear functions $\ell_1, \ldots, \ell_a : \mathbb{R}^{N_{t/2} \times N_{t/2}} \to \mathbb{R}^{N_{t/2} \times N_{t/2}}$ such that the "respecting constraints" condition holds if and only if the matrices $\ell_1(M), \ldots, \ell_a(M)$ are positive semidefinite, where M denotes the moment matrix. (**Hint:** Start with the special case of $m = 1$.)

1.C. (12 PTS.) A *semidefinite program* is an optimization problem of the form

$$\min_X \langle C, X\rangle \ \text{ s.t. } \ \langle A_1, X\rangle = b_1, \ldots, \langle A_s, X\rangle = b_s, \text{ and } X \succeq 0\,,$$

where X is a $d \times d$ matrix-valued variable, the matrices $A_1, \ldots, A_s, C$ are fixed $d \times d$ matrices, and $b_1, \ldots, b_s$ are fixed scalar values. Here $\langle \cdot, \cdot \rangle$ denotes the trace inner product, e.g. $\langle C, X\rangle = \sum_{i,j} C_{ij} X_{ij}$, and the notation $X \succeq 0$ denotes that X must be positive semidefinite. Semidefinite programs, like linear programs, can be solved in polynomial time, e.g. via the ellipsoid algorithm or interior point methods.

Let $p_1, \ldots, p_m$ be as in Eq. (1), and let $q \in \mathbb{R}[z_1, \ldots, z_n]$ be some polynomial of degree at most t. Give a proof that the optimization problem

$$\min_{\tilde{\mathbb{E}}} \tilde{\mathbb{E}}[q(z)]\,,$$

where $\tilde{\mathbb{E}}$ ranges over degree-t pseudo-expectations satisfying Conditions 1-4 above, can be formulated as a semidefinite program.

The following questions are meant to verify some elementary facts about sum-of-squares proofs and pseudo-expectations that were used in lecture. As before, let $t \geqslant 2$ be even.

1.D. (5 PTS.) Let a, b be SoS variables. Give a degree-t SoS proof that $(a+b)^t \leqslant 2^{t-1}(a^t + b^t)$.

1.E. (9 PTS.) Let $a_1, \ldots, a_n, b_1, \ldots, b_n$ be SoS variables, and suppose $a_i^2 = a_i$ for all $i \in [n]$. Give a degree-$3t$ SoS proof that

$$\Big(\sum_i a_i b_i\Big)^t \leqslant \Big(\sum_i a_i\Big)^{t-1} \cdot \Big(\sum_i a_i b_i^t\Big)$$

For this part, you may assume that t is a power of two. We will provide **5** points of extra credit for a proof for general even t.

1.F. (0 PTS.) (**Ungraded**) Let $p \in \mathbb{R}[z_1, \ldots, z_n]$ be a sum of squares polynomial of degree at most ℓ. Prove that for any degree-ℓt pseudo-expectation $\tilde{\mathbb{E}}$ over $z_1, \ldots, z_n$,

$$\tilde{\mathbb{E}}[p(z)^{t-2}] \leqslant \tilde{\mathbb{E}}[p(z)^t]^{(t-2)/t}.$$

Solution:

1.A.

1.B.

1.C.

1.D.

1.E.

1.F.

1.G.

2 (60 PTS.) ROBUST MEAN ESTIMATION WITH SOS (9/23)

Motivation: It turns out the ideas from the SoS analyses for robust regression and mixtures of Gaussians also yield a novel algorithm for *outlier-robust mean estimation*. While there are by now more practical algorithms for this, in this problem we will explore a less practical approach that has the benefit of being conceptually simple in the framework of SoS and of being a good exercise for solidifying one's understanding of how to apply SoS to statistical problems.

Setup: Let $0 < \eta < 1/2$ be a known parameter, and let q be an arbitrary unknown distribution over $\mathbb{R}^d$ with mean μ^* and identity covariance, for $\mu^* \in \mathbb{R}^d$ unknown. Let $x_1, \ldots, x_n$ be samples generated as follows. One starts with i.i.d. samples $x_1^*, \ldots, x_n^*$ from q, and a random $\eta \cdot n$-sized subset of the points are given to an all-powerful adversary, who then arbitrarily modifies these points. We do not see $x_1^*, \ldots, x_n^*$. Given only $x_1, \ldots, x_n$, our goal is to estimate μ^* to within small additive error.

Let $a_i^* \triangleq \mathbf{1}i$ not corrupted.

To avoid worrying about concentration inequalities, in this exercise you may assume the following hold:

$$\left\| \frac{1}{(1-\eta)n} \sum_{i=1}^n a_i^* (x_i^* - \mu^*)(x_i^* - \mu^*)^\top - \mathbb{1} \right\|_{\mathsf{op}} \leqslant 0.1$$

$$\left\| \frac{1}{n} \sum_{i=1}^n a_i^* (x_i^* - \mu^*) \right\|_2 \leqslant \epsilon$$

for some small ϵ (because the set of uncorrupted points is a *random* $(1-\eta)n$-sized subset of the data, ϵ can be made arbitrarily small by taking n large enough).

2.A. (8 PTS.) Using SoS variables $a_1, \ldots, a_n$ and μ which respectively correspond to indicator variables for a $(1-\eta)n$-sized subset of $[n]$ and an estimate for the mean, write down an SoS program for solving this task. *For the purposes of this homework, you are allowed to "cheat" slightly and include a constraint quantified over all vectors. If you are interested, please see the last three pages of the* ***notes*** *for how to make this more formal.*

(**Hint:** Note that you do not necessarily need to minimize an objective function for this problem.)

2.B. (9 PTS.) Give a degree-$\leqslant 4$ SoS proof that

$$\left(\frac{1}{n}\sum_i a_i a_i^*\right) \cdot \|\mu - \mu^*\|^2 = \underbrace{\frac{1}{n}\sum_i a_i^* \langle x_i^* - \mu^*, \mu - \mu^* \rangle}_{A} + \underbrace{\frac{1}{n}\sum_i (a_i - 1) a_i^* \langle x_i^* - \mu^*, \mu - \mu^* \rangle}_{B} + \underbrace{\frac{1}{n}\sum_i (1 - a_i^*) a_i \langle x_i - \mu, \mu - \mu^* \rangle}_{C}.$$

(Note that the inner product in the third sum is slightly different from the one in the first two sums on the right-hand side.)

2.C. (5 PTS.) Give a degree-$\leqslant 4$ SoS proof that

$$\left(\frac{1}{n}\sum_i a_i a_i^*\right) \cdot \|\mu - \mu^*\|^2 \geqslant (1 - 2\eta) \cdot \|\mu - \mu^*\|^2 .$$

2.D. (6 PTS.) Give a degree-$\leqslant 2$ SoS proof that $A^2 \leqslant \epsilon^2 \cdot \|\mu - \mu^*\|^2$.

2.E. (6 PTS.) Give a degree-$\leqslant 4$ SoS proof that $B^2 \leqslant O(\eta) \cdot \|\mu - \mu^*\|^2$.

2.F. (6 PTS.) Give a degree-$\leqslant 6$ SoS proof that $C^2 \leqslant O(\eta) \cdot \|\mu - \mu^*\|^2$.

2.G. (5 PTS.) By combining Parts **2.C.** - **2.F.**, conclude that for any degree-6 pseudo-distribution $\tilde{\mathbb{E}}$ over solutions to the SoS program in Part **2.A.**, we have

$$\tilde{\mathbb{E}}[\|\mu - \mu^*\|^4] \leqslant (1 - 2\eta)^{-2} \cdot O(\epsilon^2 + \eta) \cdot \tilde{\mathbb{E}}[\|\mu - \mu^*\|^2] .$$

2.H. (5 PTS.) Use pseudo Cauchy-Schwarz (from Lecture 6) to conclude that $\tilde{\mathbb{E}}[\|\mu - \mu^*\|^2] \leqslant (1 - 2\eta)^{-2} \cdot O(\epsilon^2 + \eta)$.

2.I. (10 PTS.) Given a degree-6 pseudo-distribution $\tilde{\mathbb{E}}$ over solutions to the SoS program in Part **2.A.**, propose an estimator $\widehat{\mu}$ for μ^*. Use pseudo Cauchy-Schwarz to establish that $\|\widehat{\mu} - \mu^*\| \leqslant (1 - 2\eta)^{-1} \cdot O(\epsilon + \sqrt{\eta})$.

Solution:

2.A.

2.B.

2.C.

2.D.

2.E.

2.F.

2.G.

2.H.

2.I.