ORF 523	PROBLEM SET 6	Spring 2021, Princeton University
Instructor: A.A. Ahmadi		
AIs: A. Chaudhry, C. Dibek	UCAs: C. Becker, K. Leung	
Due on April 27, 2021, at 1:30pm EST, on Gradescope		

Problem 1: Equivalence of decision and search for some problems in NP

1. Suppose you had a blackbox that given a 3SAT instance would tell you whether it is satisfiable or not. How can you make polynomially many calls to this blackbox to find a satisfying assignment to any satisfiable instance of 3SAT?

2. Suppose you had a blackbox that given a graph G and an integer k would tell you whether G has a stable set of size larger or equal to k. How can you make polynomially many calls to this blackbox to find a maximum stable set of a given graph?

Problem 2: Complexity of rank-constrained SDPs

Consider a family of decision problems indexed by a positive integer k:

RANK-k-SDP
Input: Symmetric $n \times n$ matrices $A_1, \ldots, A_m$ with entries in $\mathbb{Q}$, scalars $b_1, \ldots, b_m \in \mathbb{Q}$.
Question: Is there a real symmetric matrix X that satisfies the constraints

$$\mathrm{Tr}(A_i X) = b_i, i = 1, \ldots, m,\ X \succeq 0,\ \mathrm{rank}(X) = k?$$

Show that RANK-k-SDP is NP-hard for any integer $k \geq 1$.

(Hint: First show NP-hardness for $k = 1$, then see how you can modify your construction so that it would work for any other k.)

Problem 3: Complexity of testing monotonicity

A polynomial $p(x) := p(x_1, \ldots, x_n)$ is nondecreasing with respect to a variable x_i if

$$\frac{\partial p}{\partial x_i}(x) \geq 0, \forall x \in \mathbb{R}^n.$$

Show that the problem of deciding whether a degree-d polynomial[1] with rational coefficients is nondecreasing with respect to a particular variable (e.g., x_1) is

(i) in P if d is less than 5,[2]

(ii) NP-hard if d is greater than or equal to 5.

[1]Here, the degree of a polynomial is equal to the highest degree of its monomials with a nonzero coefficient.
[2]You can take as given that positive semidefiniteness of an $r \times r$ matrix can be checked in time $O(r^3)$.

Problem 4: Monotone and convex regression

In the previous problem, we saw that deciding whether a polynomial is monotone is NP-hard. The same claim holds for checking convexity of polynomials (of degree $2d \geq 4$). This suggests that optimizing over monotone or convex polynomials will naturally also be NP-hard. Nonetheless, in this problem, we explore some ways to perform this task.

1. In the file `regression_data.mat`, you are given 20 points (x_i, f_i) in $\mathbb{R}^2$ where $(x_i)_{i=1,\dots,20}$ are the entries of the vector `xvec` and $(f_i)_{i=1,\dots,20}$ are the entries of the vector `fvec`.

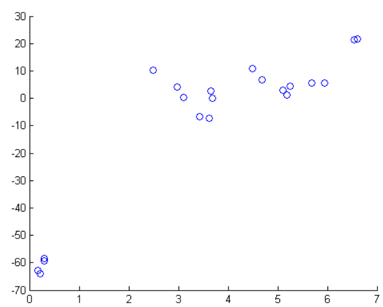

Figure 1: Figure generated by `scatter(xvec,fvec)`

The goal is to fit a polynomial of degree 7

$$p(x) = c_0 + c_1 x + \dots c_7 x^7 \tag{1}$$

to the data to minimize least square error:

$$\min_{c_0, c_1, \dots, c_7} \sum_{i=1}^{20} (p(x_i) - f_i)^2. \tag{2}$$

The data comes from noisy measurements of an unknown function that is a priori known to be nondecreasing (e.g., the number of calories you intake as a function of the number of Big Macs you eat).

(a) If the underlying function is truly monotone and the noise is not too large, one may hope that least squares would automatically respect the monotonicity constraint. Solve (2) to see if this is the case. Plot the optimal polynomial you get and report the optimal value.

(b) Resolve (2) subject to the constraint that the polynomial (1) is nondecreasing. Plot the optimal polynomial you get and report the optimal value.

2. In the file `regression_data.mat`, you are also given 30 points (x_i^1, x_i^2, g_i) in $\mathbb{R}^3$ where $(x_i^1)_{i=1,\ldots,30}$ are the entries of the vector `x1vec`, $(x_i^2)_{i=1,\ldots,30}$ are the entries of `x2vec` and $(g_i)_{i=1,\ldots,30}$ are the entries of the vector `gvec`. You can see these points below.

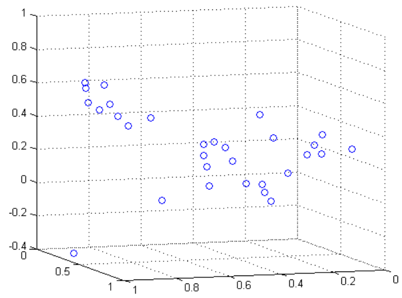

Figure 2: Figure generated by `scatter3(x1vec,x2vec,gvec)`

The goal in this case is to fit a polynomial of degree 4

$$p := p(x_1, x_2) = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_1^2 + c_4 x_1 x_2 + c_5 x_2^2 + \ldots c_{15} x_2^4$$

to the data to minimize least square error:

$$\min_{c_0, c_1, \ldots, c_{15}} \sum_{i=1}^{30} (p(x_i) - g_i)^2. \tag{3}$$

This time, the unknown underlying function is known to be convex; we want this property to be preserved in our regression.

(a) Solve (3) and plot the resulting polynomial together with the data points. Report the optimal value of the problem (denoted by η^*). Is the optimal polynomial convex?

(b) Find a convex polynomial p of degree no more than 4 such that its least squares error

$$\eta := \sum_{i=1}^{30} (p(x_i) - g_i)^2$$

satisfies $\eta < 1.75\eta^*$. (Hint: Use the SOS relaxation and take advantage of the built-in functions of YALMIP such as `hessian, jacobian`, etc.)

Problem 5: What is the probability that Zoom's stock goes bust?

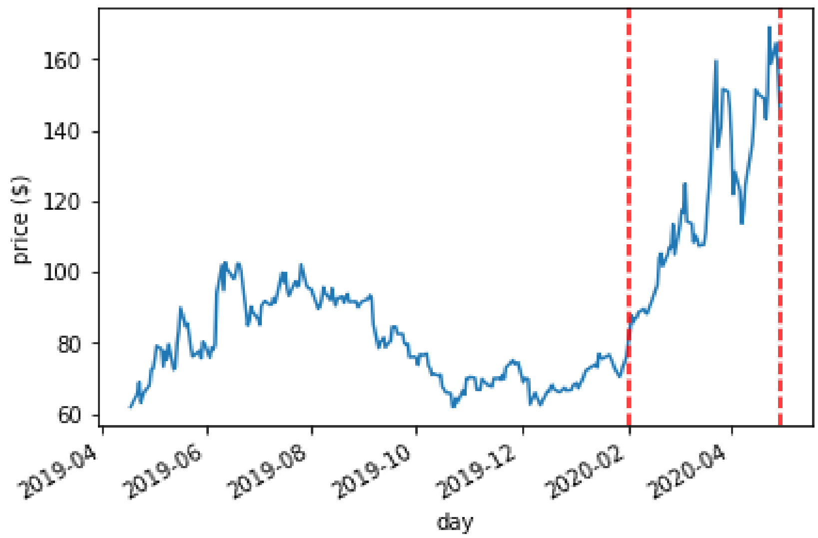

Figure 3: Zoom's stock price

You have noticed that Zoom has been growing in popularity (see Figure 3), and you wonder whether you should buy their stock. You download daily stock prices for the duration of three months starting from February 1, 2020 (excluding non-trading days) and compute the daily returns as

$$r_i = \frac{P_i - P_{i-1}}{P_i}, \quad i = 1, \dots, 61,$$

where P_i is the price of the stock on day i. You assume that the daily returns r_i are independent copies of a random variable r with unknown distribution supported on the interval $[-0.4, 0.4]$ (i.e., the daily returns never fall below -40% or go above 40%). From the data and for $k = 1, \dots, 4$, you compute the empirical means m_k of the k-th moment $\mathbb{E}[r^k]$ of r:

$$m_1 = 0.0068, m_2 = 0.0034, m_3 = 2 \times 10^{-6}, m_4 = 5 \times 10^{-5}. \tag{4}$$

Given that you are risk averse, you decide that you should buy Zoom's stock only if the probability that daily returns go below -0.1 is small. The problem, however, is that you do not know how to compute this probability as you don't know the distribution of the daily returns. You decide instead to compute the *worst-case* probability over all distributions whose first 4 moments are within 10% of those you have computed from data.

1. Let

$$\begin{aligned}
\alpha := \inf_{q,r,s,\gamma} \quad & \gamma \\
\text{s.t.} \quad & q(x) = \sum_{k=0}^{4} q_k x^k \;\; \text{is a degree-4 (univariate) polynomial,} \\
& r(x), s(x) \;\; \text{are quadratic polynomials that are sos,} \\
& q_0 + \sum_{k=1}^{4} q_k m'_k \leq \gamma \;\forall m'_k \in [0.9\, m_k, 1.1\, m_k] \text{ for } k = 1, \ldots, 4, \\
& q(x) - \left(0.4^2 - x^2\right) s(x) \text{ is sos,} \\
& q(x) - 1 - (0.4 + x)(-0.1 - x) r(x) \text{ is sos.}
\end{aligned}$$

   Show that

$$\mathbb{P}(r \in [-0.4, -0.1]) \leq \alpha,$$

   if the probability is calculated with respect to any distributions on r whose first 4 moments are within 10% of your empirical moments in (4).

   Hint: Use the basic fact that for any interval $[a, b]$, $\mathbb{P}(r \in [a, b]) = \mathbb{E}[1_{[a,b]}(r)]$, where $1_{[a,b]}(r)$ is equal to 1 if $r \in [a, b]$ and 0 otherwise.

2. Compute α to 4 digits after the decimal point.

3. You wonder if the bound α you got from the above problem is overly pessimistic. Find a discrete distribution of returns (i.e., points $x_1, \ldots, x_N \in [-0.4, 0.4]$ and probabilities $p_1, \ldots, p_N \in [0, 1]$ summing to one) such that

   i) The moments of your discrete distribution are within 10% of the empirical moments of r, i.e., $|\sum_{i=1}^N p_i x_i^k - m_k| \leq \frac{m_k}{10}, \quad k = 1, \ldots, 4,$

   ii) The probability assigned by your discrete distribution to the interval $[-0.4, -0.1]$ is equal to α; i.e.,

$$\sum_{i \in I} p_i = \alpha, \text{ where } I = \{i \in \{1, \ldots, N\} \mid x_i \in [-0.4, -0.1]\}.$$ [3]

   Hint: If q^* is the quartic polynomial that your solver returns for part 1, a plot of $q^* - 1_{[-0.4,-0.1]}$ can help you find the points $x_1, \ldots, x_N$.

[3] To avoid numerical issues, any discrete distribution that assigns to the interval $[-0.4, -0.1]$ a probability larger or equal than 0.99α is acceptable.