Convex Optimization
Homework 4

Spring 1401
Due date: 2nd of Ordibehesht

---

1. Consider the optimization problem

$$\begin{aligned} &\text{minimize } x^2+1 \\ &\text{subject to } (x-2)(x-4) \le 0 \end{aligned}$$

with variable $x \in \mathbf{R}$.

(a) Give the feasible set, the optimal value, and the optimal solution.

(b) Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian $L(x, \lambda)$ versus x for a few positive values of λ. Verify the lower bound property ($p^\star \ge \inf_x L(x, \lambda)$ for $\lambda \ge 0$). Derive and sketch the Lagrange dual function g.

(c) State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution $\lambda^\star$. Does strong duality hold?

(d) Let $p^\star(u)$ denote the optimal value of the problem

$$\begin{aligned} &\text{minimize } x^2+1 \\ &\text{subject to } (x-2)(x-4) \le u \end{aligned}$$

as a function of the parameter u. Plot $p^\star(u)$. Verify that $dp^\star(0)/du = -\lambda^\star$.

2. *Numerical perturbation analysis example.* Consider the quadratic program

$$\begin{aligned} &\text{minimize } x_1^2 + 2x_2^2 - x_1x_2 - x_1 \\ &\text{subject to } x_1 + 2x_2 \le u_1 \\ &\qquad\qquad x_1 - 4x_2 \le u_2 \\ &\qquad\qquad 5x_1 + 76x_2 \le 1 \end{aligned} \tag{1}$$

with variables x_1, x_2, and parameters u_1, u_2.

(a) Solve this QP, for parameter values $u_1 = -2$, $u_2 = -3$, to find optimal primal variable values x_1^* and x_2^*, and optimal dual variable values λ_1^*, λ_2^* and λ_3^*. Let p^* denote the optimal objective value. Verify that the KKT conditions hold for the optimal primal and dual variables you found (within reasonable numerical accuracy).

(b) We will now solve some perturbed versions of the QP, with

$$u_1 = -2 + \delta_1, \quad u_2 = -3 + \delta_2, \tag{2}$$

where δ_1 and δ_2 each take values from $\{-0.1, 0, 0.1\}$. (There are a total of nine such combinations, including the original problem with $\delta_1 = \delta_2 = 0$.) For each combination of δ_1 and δ_2, make a prediction p^*_{pred} pred of the optimal value of the perturbed QP, and compare it to p^*_{exact} exact, the exact optimal value of the perturbed QP (obtained by solving the perturbed QP). Put your results in the two righthand columns in a table with the form shown below. Check that the inequality $p^*_{\text{pred}} \le p^*_{\text{exact}}$ exact holds.

δ_1	δ_2	$p^\star_{\rm pred}$	$p^\star_{\rm exact}$
0	0		
0	-0.1		
0	0.1		
-0.1	0		
-0.1	-0.1		
-0.1	0.1		
0.1	0		
0.1	-0.1		
0.1	0.1		

3. *Robust LP with polyhedral cost uncertainty.* We consider a robust linear programming problem, with polyhedral uncertainty in the cost:

$$\begin{array}{ll} \text{minimize} & \sup_{c \in \mathcal{C}} c^T x \\ \text{subject to} & Ax \succeq b, \end{array}$$

with variable $x \in \mathbf{R}^n$, where $\mathcal{C} = \{c | Fc \preceq g$. You can think of x as the quantities of n products to buy (or sell, when $x_i < 0$), $Ax \succeq b$ as constraints, requirements, or limits on the available quantities, and $\mathcal{C}$ as giving our knowledge or assumptions about the product prices at the time we place the order. The objective is then the worst possible (*i.e.*, largest) possible cost, given the quantities x, consistent with our knowledge of the prices.

In this exercise, you will work out a tractable method for solving this problem. You can assume that $\mathcal{C} \neq \emptyset$, and the inequalities $Ax \succeq b$ are feasible.

(a) Let $f(x) = \sup_{c \in \mathcal{C}} c^T x$ be the objective in the problem above. Explain why f is convex.

(b) Find the dual of the problem

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & Fc \preceq g \leq 0 \end{array}$$

with variable c. (The problem data are x, F, and g.) Explain why the optimal value of the dual is $f(x)$.

(c) Use the expression for $f(x)$ found in part (b) in the original problem, to obtain a single LP equivalent to the original robust LP.

(d) Carry out the method found in part (c) to solve a robust LP with the data below. In Matlab:

```
rand('seed',0);
A = rand(30,10);
b = rand(30,1);
c_nom = 1+rand(10,1); % nominal c values
In Python:
import numpy as np
np.random.seed(10)
(m, n) = (30, 10)
A = np.random.rand(m, n); A = np.asmatrix(A)
b = np.random.rand(m, 1); b = np.asmatrix(b)
c_nom = np.ones((n, 1)) + np.random.rand(n, 1); c_nom = np.asmatrix(c_nom)
In Julia:
srand(10);
n = 10;
m = 30;
A = rand(m, n);
b = rand(m, 1);
c_nom = 1 + rand(n, 1);
```

Then, use $\mathcal{C}$ described as follows. Each c_i deviates no more than 25% from its nominal value, *i.e.*, $0.75c_{\text{nom}} \preceq c \preceq 1.25c_{\text{nom}}$, and the average of c does not deviate more than 10% from the average of the nominal values, *i.e.*, $0.9(\mathbf{1}^T c_{\text{nom}})/n \leq \mathbf{1}^T c/n \leq 1.1(\mathbf{1}^T c_{\text{nom}})/n$.

Compare the worst-case cost $f(X)$ and the nominal cost $c_{\text{nom}}^T x$ for x optimal for the robust problem, and for x optimal for the nominal problem (*i.e.*, the case where $\mathcal{C} = \{c_{\text{nom}}\}$). Compare the values and make a brief comment.

4. *Bandlimited signal recovery from zero-crossings.* Let $y \in \mathbf{R}^n$ denote a *bandlimited* signal, which means that it can be expressed as a linear combination of sinusoids with frequencies in a band:

$$y_t = \sum_{j=1}^{B} a_j \cos\left(\frac{2 * pi}{n}(f_{\min} + j - 1)t\right) + b_j \sin\left(\frac{2 * pi}{n}(f_{\min} + j - 1)t\right), \quad t = 1, 2, \ldots, n,$$

where $f_{\min}$ is lowest frequency in the band, B is the bandwidth, and $a, b \in \mathbf{R}^B$ are the cosine and sine coefficients, respectively. We are given $f_{\min}$ and B, but not the coefficients a, b or the signal y.

We do not know y, but we are given its sign $s = \mathbf{sign}(y)$, where $s_t = 1$ if $y_t \geq 0$ and $s_t = -1$ if $y_t < 0$. (Up to a change of overall sign, this is the same as knowing the 'zero-crossings' of the signal, *i.e.*, when it changes sign. Hence the name of this problem.)

We seek an estimate $\hat{y}$ of y that is consistent with the bandlimited assumption and the given signs.

Of course we cannot distinguish y and αy, where $\alpha > 0$, since both of these signals have the same sign pattern. Thus, we can only estimate y up to a positive scale factor. To normalize $\hat{y}$, we will require that $\|\hat{y}\|_1 = n$, *i.e.*, the average value of y_i is one. Among all $\hat{y}$ that are consistent with the bandlimited assumption, the given signs, and the normalization, we choose the one that minimizes $\|\hat{y}\|_2$.

(a) Show how to find $\hat{y}$ using convex or quasiconvex optimization.

(b) Apply your method to the problem instance with data in zero_*crossings_data*.*. The data files also include the true signal y (which of course you cannot use to find $\hat{y}$). Plot $\hat{y}$ and y, and report the relative recovery error, $\|y - \hat{y}\|_2 / \|y\|_2$. Give one short sentence commenting on the quality of the recovery.

5. Consider the problem of projecting a point $a \in \mathbf{R}^n$ on the unit ball in ℓ-norm:

$$\begin{array}{ll} \text{minimize} & (1/2)\|x - a\|_2^2 \\ \text{subject to} & \|x\|_1 \leq 1. \end{array}$$

Derive the dual problem and describe an efficient method for solving it. Explain how you can obtain the optimal x from the solution of the dual problem.

6. We consider the non-convex least-squares approximation problem with binary constraints

$$\begin{array}{ll} \text{minimize} & ||Ax - b||_2^2 \\ \text{subject to} & x_k^2 = 1, \quad k = 1, 2, \ldots, n \end{array} \tag{3}$$

where $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$. We assume that $\mathbf{rank}(A) = n$, i.e., $A^T A$ is nonsingular. One possible application of this problem is as follows. A signal $\hat{x} \in \{-1, 1\}$ is sent over a noisy channel, and received as $b = A\hat{x} + v$ where $v\ N(0, \sigma^2 I)$ is Gaussian noise. The solution of (3) is the maximum likelihood estimate of the input signal $\hat{x}$, based on the received signal b.

(a) Derive the Lagrange dual of (3) and express it as an SDP.

(b) Derive the dual of the SDP in part (a) and show that it is equivalent to

$$\begin{aligned} &\text{minimize } tr(A^TAZ) - 2b^TAz + b^Tb \\ &\text{subject to } \mathbf{diag}(Z) = 1 \\ &\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \succeq 0 \end{aligned} \tag{4}$$

Interpret this problem as a relaxation of (3). Show that if

$$\mathbf{rank}(\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix}) = 1 \tag{5}$$

at the optimum of (4), then the relaxation is exact, i.e., the optimal values of problems (3) and (4) are equal, and the optimal solution z of (4) is optimal for (3). This suggests a heuristic for rounding the solution of the SDP (4) to a feasible solution of (3), if (5) does not hold. We compute the eigenvalue decomposition

$$\mathbf{rank}(\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix}) = \sum_{i=1}^{n+1} \lambda_i \begin{bmatrix} v_i \\ t_i \end{bmatrix} \left(\begin{bmatrix} v_i \\ t_i \end{bmatrix} \right)^T$$

where $v_i \in \mathbf{R}^n$ and $t_i \in \mathbf{R}$, and approximate the matrix by a rank-one matrix

$$\begin{bmatrix} Z & z \\ z^T & 1 \end{bmatrix} \approx \lambda_1 \begin{bmatrix} v_1 \\ t_1 \end{bmatrix} \left(\begin{bmatrix} v_1 \\ t_1 \end{bmatrix} \right)^T \tag{6}$$

(Here we assume the eigenvalues are sorted in decreasing order). Then we take $x = \mathbf{sign}(v_1)$ as our guess of good solution of (3).

(c) We can also give a probabilistic interpretation of the relaxation (4). Suppose we interpret z and Z as the first and second moments of a random vector $v \in \mathbf{R}^n$ (i.e., $z = \mathbf{E}v, Z = \mathbf{E}vv^T$). Show that (4) is equivalent to the problem

$$\begin{aligned} &\text{minimize } \mathbf{E}||Av - b||_2^2 \\ &\text{subject to } \mathbf{E}v_k^2 = 1, \quad k = 1, 2, \ldots, n \end{aligned}$$

where we minimize over all possible probability distributions of v. This interpretation suggests another heuristic method for computing suboptimal solutions of (3) based on the result of (4). We choose a distribution with first and second moments $\mathbf{E}v = z, \mathbf{E}vvT = Z$ (for example, the Gaussian distribution $N(z, Z - zz^T)$). We generate a number of samples $\hat{v}$ from the distribution and round them to feasible solutions $x = \mathbf{sign}(\hat{v})$. We keep the solution with the lowest objective value as our guess of the optimal solution of (3).

(d) Solve the dual problem (4) using CVX. Generate problem instances using the Matlab code:

```
randn('state',0)
m = 50;
n = 40;
A = randn(m,n);
xhat = sign(randn(n,1));
b = A*xhat + s*randn(m,1);
```

for four values of the noise level $s : s = 0 : 5, s = 1, s = 2, s = 3$. For each problem instance, compute suboptimal feasible solutions x using the the following heuristics and compare the results.

i. $x^{(a)} = \mathbf{sign}(x_{ls})$ where x_{ls} is the solution of the least-squares problem

$$\text{minimize } ||Ax - b||_2^2$$

ii. $x^{(b)} = \mathbf{sign}(z)$ where z is the optimal value of the variable z in the SDP (4).

iii. $x^{(c)}$ is computed from a rank-one approximation of the optimal solution of (4), as explained in part (b) above.

iv. $x^{(d)}$ is computed by rounding 100 samples of $N(z, Z - zz^T)$, as explained in part (c) above.

Good Luck!