Convex Optimization
Homework 3

Spring 1400
Due date: Farvardin 27th

---

1. Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0 \text{ holds for at least } k \text{ values of } i \end{array}$$

with variable $x \in \mathbf{R}^n$, where the objective f_0 and the constraint functions $f_i, i = 1, \ldots, m$ (with $m \geq k$), are convex. Here we require that only k of the constraints hold, instead of all m of them. In general this is a hard combinatorial problem; the brute force solution is to solve all $\begin{pmatrix} m \\ k \end{pmatrix}$ convex problems obtained by choosing subsets of k constraints to impose, and selecting one with smallest objective value.

In this problem we explore a convex restriction that can be an effective heuristic for the problem.

(a) Suppose $\lambda > 0$. Show that the constraint

$$\sum_{i=1}^{m} (1 + \lambda f_i(x))_+ \leq m - k$$

guarantees that $f_i(x) \leq 0$ holds for at least k values of i. ($(u)_+$ means $\max\{u, 0\}$.) Hint. For each $u \in \mathbf{R}$, $(1 + \lambda u)_+ \geq 1(u > 0)$, where $1(u > 0) = 1$ for $u > 0$, and $1(u > 0) = 0$ for $u \leq 0$

(b) Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \sum_{i=1}^{m} (1 + \lambda f_i(x))_+ \leq m - k \\ & \lambda > 0 \end{array}$$

with variables x and λ. This is a restriction of the original problem: If (x, λ) are feasible for it, then x is feasible for the original problem. Show how to solve this problem using convex optimization. (This may involve a change of variables.)

(c) Apply the method of part (b) to the problem instance

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \text{ holds for at least } k \text{ values of } i \end{array}$$

with $m = 70, k = 58$, and $n = 12$. The vectors b, c and the matrix A with rows a_i^T are given in the file satisfy_some_constraints_data.*
Report the optimal value of λ, the objective value, and the actual number of constraints that are satisfied (which should be larger than or equal to k). To determine if a constraint is satisfied, you can use the tolerance $a_i^T x - b_i \leq \epsilon^{\text{feas}}$, with $\epsilon^{\text{feas}} = 10^{-5}$. A standard trick is to take this tentative solution, choose the k constraints with the smallest values of $f_i(x)$, and then minimize $f_0(x)$ subject to these k constraints (i.e., ignoring the other $m-k$ constraints). This improves the objective value over the one found using the restriction. Carry this out for the problem instance, and report the objective value obtained.

2. The relative entropy between two vectors $x, y \in \mathbf{R}^n_{++}$ is defined as

$$\sum_{k=1}^{n} x_k \log (x_k / y_k)$$

This is a convex function, jointly in x and y. In the following problem we calculate the vector x that minimizes the relative entropy with a given vector y, subject to equality constraints on x :

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{n} x_k \log (x_k/y_k) \\ \text{subject to} & Ax = b \\ & \mathbf{1}^T x = 1 \end{array}$$

The optimization variable is $x \in \mathbf{R}^n$. The domain of the objective function is $\mathbf{R}^n_{++}$. The parameters $y \in \mathbf{R}^n_{++}$, $A \in \mathbf{R}^{m\times n}$, and $b \in \mathbf{R}^m$ are given. Derive the Lagrange dual of this problem and simplify it to get

$$\text{maximize} \quad b^T z - \log \sum_{k=1}^{n} y_k e^{a_k^T z}$$

(a_k is the k th column of A).

3. Consider the problem of projecting a point $a \in \mathbf{R}^n$ on the unit ball in ℓ_1-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.

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

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_k^2 = 1, \quad k = 1, \ldots, n \end{array} \qquad \text{(eq1)}$$

where $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$. We assume that $\operatorname{rank}(A) = n$, i.e., $A^T A$ is nonsingular. One possible application of this problem is as follows. A signal $\hat{x} \in \{-1, 1\}^n$ is sent over a noisy channel, and received as $b = A\hat{x} + v$ where $v \sim \mathcal{N}\left(0, \sigma^2 I\right)$ is Gaussian noise. The solution of (eq1) is the maximum likelihood estimate of the input signal $\hat{x}$, based on the received signal b.

(a) Derive the Lagrange dual of (eq1) and express it as an SDP.

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

$$\begin{array}{ll} \text{minimize} & \operatorname{tr}\left(A^T A Z\right) - 2b^T A z + b^T b \\ \text{subject to} & \operatorname{diag}(Z) = 1 \\ & \left[\begin{array}{cc} Z & z \\ z^T & 1 \end{array}\right] \succeq 0 \end{array} \qquad \text{(eq2)}$$

Interpret this problem as a relaxation of (eq1). Show that if

$$\operatorname{rank}\left(\left[\begin{array}{cc} Z & z \\ z^T & 1 \end{array}\right]\right) = 1 \qquad \text{(eq3)}$$

at the optimum of (eq2), then the relaxation is exact, i.e., the optimal values of problems (eq1) and (eq2) are equal, and the optimal solution z of (eq2) is optimal for (eq1). This suggests a heuristic for rounding the solution of the SDP (eq2) to a feasible solution of (eq1) if (eq3) does not hold. We compute the eigenvalue decomposition

$$\left[\begin{array}{cc} Z & z \\ z^T & 1 \end{array}\right] = \sum_{i=1}^{n+1} \lambda_i \left[\begin{array}{c} v_i \\ t_i \end{array}\right] \left[\begin{array}{c} v_i \\ t_i \end{array}\right]^T$$

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

$$\left[\begin{array}{cc} Z & z \\ z^T & 1 \end{array}\right] \approx \lambda_1 \left[\begin{array}{c} v_1 \\ t_1 \end{array}\right] \left[\begin{array}{c} v_1 \\ t_1 \end{array}\right]^T$$

(Here we assume the eigenvalues are sorted in decreasing order). Then we take $x = \operatorname{sign}(v_1)$ as our guess of good solution of (eq1)

(c) We can also give a probabilistic interpretation of the relaxation (eq2). Suppose we interpret z and Z as the first and second moments of a random vector $v \in \mathbf{R}^n$ $\left(i.e., z = \mathbf{E}v, Z = \mathbf{E}vv^T\right)$. Show that (eq2) is equivalent to the problem

$$\text{subject to} \quad \mathbf{E}v_k^2 = 1, \quad k = 1, \ldots, n$$

where we minimize over all possible probability distributions of v. This interpretation suggests another heuristic method for computing suboptimal solutions of (eq1) based on the result of (eq2) . We choose a distribution with first and second moments $\mathbf{E}v = z, \mathbf{E}vv^T = Z$ (for example, the Gaussian distribution $\mathcal{N}\left(z, Z - zz^T\right)$) . We generate a number of samples $\tilde{v}$ from the distribution and round them to feasible solutions $x = \operatorname{sign}(\tilde{v})$. We keep the solution with the lowest objective value as our guess of the optimal solution of (eq1)

(d) Solve the dual problem (eq2) 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. $x^{(a)} = \operatorname{sign}(x_{\text{ls}})$ where x_{ls} is the solution of the least-squares problem

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

ii. $x^{(b)} = \operatorname{sign}(z)$ where z is the optimal value of the variable z in the SDP (eq2).

iii. $x^{(c)}$ is computed from a rank-one approximation of the optimal solution of (eq2) as explained in part (b) above.

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

5. Consider the optimization problem

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

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

(a) Analysis of primal problem: Give the feasible set, the optimal value, and the optimal solution.

(b) Lagrangian and dual function: 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 ($p^\star \geq \inf_x L(x, \lambda)$ for $\lambda \geq 0$). Derive and sketch the Lagrange dual function g.

(c) Lagrange dual problem: 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) Sensitivity analysis. Let $p^\star(u)$ denote the optimal value of the problem

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

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

6. Express the dual problem of

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & f(x) \leq 0 \end{array}$$

with $c \neq 0$, in terms of the conjugate f^*. Explain why the problem you give is convex. We do not assume f is convex.

7. A Boolean linear program is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0, 1\}, \quad i = 1, \ldots, n \end{array}$$

and is, in general, very difficult to solve. In exercise 4.15 we studied the LP relaxation of this problem,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \leq x_i \leq 1, \quad i = 1, \ldots, n \end{array} \qquad \text{(eq)}$$

which is far easier to solve, and gives a lower bound on the optimal value of the Boolean LP. In this problem we derive another lower bound for the Boolean LP, and work out the relation between the two lower bounds.

(a) The Boolean LP can be reformulated as the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i (1 - x_i) = 0, \quad i = 1, \ldots, n \end{array}$$

which has quadratic equality constraints. Find the Lagrange dual of this problem. The optimal value of the dual problem (which is convex) gives a lower bound on the optimal value of the Boolean LP. This method of finding a lower bound on the optimal value is called Lagrangian relaxation.

(b) Show that the lower bound obtained via Lagrangian relaxation, and via the LP relaxation (eq), are the same. Hint. Derive the dual of the LP relaxation (eq).