ECE 287A: Convex Optimization and Applications **Fall 2006**

Homework 4

Lecturer: Gert Lanckriet

Due: Tuesday 11/28/06 (in class)

1. *(4 points)* Consider the problem

$$\max_x H(Ax + b),$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and $H : \mathbb{R}^m \mapsto \mathbb{R}$ is the harmonic mean, defined as

$$H(v) = \frac{1}{\frac{1}{v_1} + \ldots + \frac{1}{v_m}},$$

with $\mathrm{dom} H = \{v \in \mathbb{R}^m \mid v \succ 0\}$. Recast this problem as a semidefinite program.

2. *(4 points)* Consider the optimization problem

$$\min_x \sum_{i=1}^{m} \phi((Ax - b)_i) + C||x||_2^2,$$

where $A \in \mathbb{R}^{m \times n}(m \leq n)$, $b \in \mathbb{R}^m$ and $C > 0$ given, and ϕ convex, with domain the whole real line. Derive a Lagrange dual to the problem (introducing equality constraints and involving ϕ's conjugate function) to prove convexity of the optimal value of the problem as a function of $Q = AA^T$.

3. The maximum eigenvalue $\lambda_{\max}(A)$ of a matrix $A \in \mathcal{S}^n$ can be characterized as the optimal value of an SDP minimization problem.

   (a) *(1 point)* Formulate this SDP.

   (b) *(2 points)* Derive the dual of this SDP. Do we have strong duality?

   (c) *(2 points)* Can you relate the dual to another variational characterization of $\lambda_{\max}(A)$?

4. Given $Q \in \mathcal{S}^n_{++}$, $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$, we consider the two problems: the *penalized* problem

$$p_1 := \max_x c^T x - (1/2) x^T Q x, \quad \text{s.t.} \;\; Ax \preceq 0,$$

and the *constrained* problem

$$p_2 := \max_x c^T x \quad \text{s.t.} \;\; Ax \preceq 0, \; x^T Q x \leq 1.$$

(a) *(2 points)* Are these problems convex? How would you characterize each of them (as LP, QP, SOCP, QCQP, SDP, etc)?

(b) *(3 points)* Express the dual problem of the first problem (p_1). How would you characterize it (as LP, QP, SOCP, QCQP, SDP, etc)? Show that strong duality holds. (*hint*: check that both the primal and dual solution are attained and achieve the same value)

(c) *(3 points)* Express the dual problem of the second problem (p_2). How would you characterize it (as LP, QP, SOCP, QCQP, SDP, etc)? Show that strong duality holds. (*hint*: check that both the primal and dual solution are attained and achieve the same value)

(d) *(2 points)* How are the optimal objective values p_1 and p_2 related? How can solving either one of the dual problems yield the solutions to both problems? Is there a relation between x_1* and x_2^*, optimal solutions for the penalized respectively the constrained problem?

(e) *(2 points)* Show that, provided that the optimal values of either the penalized or constrained problem is positive, both problems are equivalent to solving the *non-convex* problem

$$\max_x c^T x \quad \text{s.t.} \quad Ax \preceq 0, \; x^T Q x = 1.$$

(f) *(3 points)* Replace the factor (1/2) by β in the formulation of the penalized problem and the constraint $x^T Q x \leq 1$ by $x^T Q x \leq \gamma$ in the constrained problem. Under what condition(s) is solving the penalized problem equivalent to solving the constrained problem?

5. Consider the problem

$$\min_x \frac{||Ax - b||_1^2}{1 - ||x||_\infty},$$

where $x \in \mathbb{R}^n$ and the data are $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The constraint $||x||_\infty < 1$ is implicit. You can assume that $b \notin \mathcal{R}(A)$, in which case the constraint $||x||_\infty < 1$ can be replaced with $||x||_\infty \leq 1$.

(a) *(3 points)* Is the problem, exactly as stated (and for all problem data), convex? If not, is it quasiconvex? Justify your answer.

(b) *(3 points)* Explain how to solve the problem. Your method can involve an SDP solver, an SOCP solver, an LP solver, or any combination. You can include a one-parameter bisection, if necessary (e.g., solve the problem by bisection on a parameter, where each iteration consists of solving an SDP feasibility problem). Give the best method you can. In judging best, we use the rule that using bisection is worse than solving *one* LP, SOCP or SDP and that an LP solver is simpler and thus better than an SOCP solver, which is considered simpler and better than an SDP solver.

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

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_k^2 = 1, \quad k=1,\ldots,n, \end{array} \tag{4.1}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. We assume that $\mathrm{rank}(A) = n$, i.e., A^TA 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 $v \sim \mathcal{N}(0, \sigma^2 I)$ is Gaussian noise. The solution of (4.1) is the maximum likelihood estimate of the input signal $\hat{x}$, based on the received signal b.

(a) *(2 points)* Derive the Lagrange dual of (4.1) and express it as an SDP.

(b) *(3 points)* Derive the dual of the SDP in part (a) and show that it is equivalent to

$$\begin{array}{ll} \text{minimize} & \mathrm{tr}(A^TAZ) - 2b^TAz + b^Tb \\ \text{subject to} & \mathrm{diag}(Z) = 1, \\ & \begin{pmatrix} Z & z \\ z^T & 1 \end{pmatrix} \succeq 0. \end{array} \tag{4.2}$$

Interpret this problem as a relaxation of (4.1). Show that if

$$\mathrm{rank}\begin{pmatrix} Z & z \\ z^T & 1 \end{pmatrix} = 1 \tag{4.3}$$

at the optimum of (4.2), then the relaxation is exact, i.e., the optimal values of problems (4.1) and (4.2) are equal, and the optimal solution z of (4.2) is optimal for (4.1). This suggests a heuristic for rounding the solution of the SDP (4.2) to a feasible solution of (4.1), if (4.3) does not hold. We compute the eigenvalue decomposition

$$\begin{pmatrix} Z & z \\ z^T & 1 \end{pmatrix} = \sum_{i=1}^{n+1} \lambda_i \begin{pmatrix} v_i \\ t_i \end{pmatrix} \begin{pmatrix} v_i \\ t_i \end{pmatrix}^T,$$

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

$$\begin{pmatrix} Z & z \\ z^T & 1 \end{pmatrix} \approx \lambda_1 \begin{pmatrix} v_1 \\ t_1 \end{pmatrix} \begin{pmatrix} v_1 \\ t_1 \end{pmatrix}^T.$$

(Here we assume the eigenvalues are sorted in decreasing order). Then we take $x = \mathrm{sign}(v_1)$ as our guess of a good solution of (4.1).

(c) *(3 points)* We can also give a probabilistic interpretation of the relaxation (4.2). Suppose we interpret z and Z as the first and second moments of a random vector $v \in \mathbb{R}^n$ (i.e., $z = \mathbf{E}v, Z = \mathbf{E}vvT$). Show that (4.2) is equivalent to the problem

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

where we minimize over all possible probability distributions of v.

This interpretation suggests another heuristic method for computing suboptimal solutions of (4.1) based on the result of (4.2). We choose a distribution with first and second moments $\mathbf{E}v = z, \mathbf{E}vv^T = Z$ (for example, the Gaussian distribution $\mathcal{N}(z, Z - zz^T)$). We generate a number of samples $\tilde{v}$ from the distribution and round them to feasible solutions $x = \text{sign}(\tilde{v})$. We keep the solution with the lowest objective value as our guess of the optimal solution of (4.1).

(d) *(6 points)* Write code (use Matlab and one of the SDP solvers) to solve the dual problem (4.2). 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 following heuristics and compare the results.

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

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

(2) $x^{(b)} = \text{sign}(z)$ where z is the optimal value of the variable z in the SDP (4.2).

(3) $x^{(c)}$ is computed from a rank-one approximation of the optimal solution of (4.2), as explained in part (b) above.

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