EECS 127 - EECS 227 AT 5/14/2015

L. El Ghaoui

Final

NAME: **SID:**

The exam lasts 3 hours. The maximum number of points is 40. Notes are not allowed except for a two-sided cheat sheet of regular format.

This booklet is 11 pages total, with extra blank pages allotted throughout, and 2 blank pages at the end, left for you to write your answers.

There are **four** separate problems, arranged in increasing order of difficulty. All the questions in this exam can be solved independently of each other.

The breakdown of points is as follows. Note that Parts 1 & 2, which focus on linear algebra, will be graded according to a clobber policy, as mentioned in class.

Part	a	b	c	d	e	total
1	3	3	4			10
2	2	2	3	3		10
3	1	3	2	3	1	10
4	2	4	2	2		10
						40

1. (*10 points*) Consider the set in $\mathbb{R}^{2n}$

$$\mathcal{E} := \left\{ z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ F(x,y) := \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}^T \begin{pmatrix} A_{xx} & A_{xy} & a_x \\ A_{xy}^T & A_{yy} & a_y \\ a_x^T & a_y^T & \alpha \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \leq 0 \right\},$$

where $A_{xx}, A_{xy}, A_{yy} \in \mathbb{R}^{n\times n}$, $a_x, a_y \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, and the $2n \times 2n$ matrix

$$A := \begin{pmatrix} A_{xx} & A_{xy} \\ A_{xy}^T & A_{yy} \end{pmatrix}$$

is symmetric and positive-definite. We define $a = (a_x, a_y) \in \mathbb{R}^{2n}$.

(a) *(3 points)* Express the function F, evaluated at a point $z = (x,y)$, as

$$F(x,y) = (z - z_0)^T Q (z - z_0) - \gamma$$

for some appropriate matrix Q, vector z_0, and scalar γ, which you will determine.

(b) *(3 points)* What is the shape of $\mathcal{E}$ in $\mathbb{R}^{2n}$? Express your answer in geometric terms, assuming that some relevant eigenvalue decomposition is available. In particular, explain why the condition $\alpha > a^T A^{-1} a$ ensures that $\mathcal{E}$ is empty.

(c) *(4 points)* Likewise, determine the projection of $\mathcal{E}$ onto the space of x-variables. *Hint:* a point $x \in \mathbb{R}^n$ belongs to the projection if and only if $\min_y F(x,y) \leq 0$.

2. (*10 points*) We consider regularized supervised learning problems of the form

$$P(\lambda) \; : \; \min_w \; f_\lambda(w) := \mathcal{L}(X^T w) + \frac{\lambda}{2}\|w\|_2^2,$$

where $\lambda > 0$ is a regularization parameter, $X \in \mathbb{R}^{n \times m}$ is a matrix of data points, and $\mathcal{L}$ is a differentiable, convex loss function, which we assume is decomposable: for any given $z \in \mathbb{R}^m$:

$$\mathcal{L}(z) = \sum_{i=1}^{m} l_i(z_i),$$

where l_i are given (differentiable, convex) functions.

We consider a gradient method to solve the problem, which, for a single value of λ, involves iterates $w_\lambda(k)$, $k = 0, 1, 2, \ldots$, obtained via the gradient recursion

$$w_\lambda(k+1) = w_\lambda(k) - t\nabla f_\lambda(w(k)), \quad k = 0, 1, 2, 3, \ldots \tag{1}$$

where $t > 0$ is a (fixed) step size. In this exercise, we would like to solve the problems $P(\lambda)$ for a sequence of values $\lambda \in \lambda_1, \ldots, \lambda_p$. We explore the use of matrix-matrix products to run the different gradient algorithms faster than running the recursions (1) one after the other in sequence. (Matrix-matrix products are highly optimized memory-wise, and usually are much faster than a corresponding sequence of matrix-vector products, for example.)

(a) *(2 points)* Show that forming the gradient of $\mathcal{L}$ at a point $z \in \mathbb{R}^m$ is a component-wise operation, namely

$$\nabla \mathcal{L}(z) = \begin{pmatrix} \nabla l_1(z_1) \\ \vdots \\ \nabla l_m(z_m) \end{pmatrix}.$$

Hint: approximate $\mathcal{L}(z + \delta) - \mathcal{L}(z)$ for any $z, \delta \in \mathbb{R}^m$, with δ small.

(b) *(2 points)* Show that for any given w, and $\lambda > 0$, $\nabla f_\lambda(w) = X\nabla\mathcal{L}(X^T w) + \lambda w$. *Hint:* again use a first-order expansion of f_λ.

(c) *(3 points)* Define the matrices $W(k) = [w_{\lambda_1}(k), \ldots, w_{\lambda_p}(k)] \in \mathbb{R}^{n \times p}$, $k = 0, 1, 2, \ldots$. Show that the recursion (1) can be written

$$W(k+1) = W(k) - t\left(X\nabla\mathcal{L}(X^T W(k)) + W(k)D\right), \quad k = 0, 1, 2, 3, \ldots$$

for an appropriate diagonal matrix D, which you will determine, and with the convention that $\nabla\mathcal{L}$ applies in column-wise fashion on a matrix input: for a matrix $Z = [z_1, \ldots, z_p] \in \mathbb{R}^{m \times p}$, we set $\nabla\mathcal{L}(Z) = [\nabla\mathcal{L}(z_1), \ldots, \nabla\mathcal{L}(z_p)] \in \mathbb{R}^{m \times p}$.

(d) *(3 points)* In some variants of the gradient method, the step size depends on the regularization parameter λ and/or on the iteration count k, which we denote by $t_\lambda(k)$. How would the matrix recursion above be modified in that case?

3. (10 points) *Risk Budgeting.* We are given a symmetric, $n \times n$ positive-definite matrix C, and a vector $\theta \in \mathbb{R}^n_{++}$, with $\mathbf{1}^T\theta = 1$, where $\mathbf{1}$ is the n-vector of ones. We assume that a factorization of the form $C = R^T R$, with $R \in \mathbb{R}^{n \times n}$, is available.

   We consider a *risk budgeting* problem arising in financial optimization, which consists in finding a vector $x \in \mathbb{R}^n$ such that

   $$x > 0, \;\; x_i(Cx)_i = \theta_i(x^T Cx), \;\; i = 1, \dots, n. \tag{2}$$

   (In case you are curious: the term "risk budgeting" refers to the fact that each so-called "partial risk" $x_i(Cx)_i$ is assigned a fixed proportion θ_i of the total risk, defined as the variance $x^T Cx$, itself the sum of the partial risks.)

   (a) *(1 point)* Is the problem of finding a risk budgeting portfolio, as defined above, or determine there is no such portfolio, convex? Justify your answer carefully.

   (b) *(3 points)* Consider a generic constraint on a triple (z, u, v), with z a vector and u, v scalars, of the form

   $$u \geq 0, \;\; v \geq 0, \;\; uv \geq z^T z.$$

   Prove that the above so-called "rotated second-order cone" constraint can be written as

   $$\left\| \begin{pmatrix} 2z \\ u - v \end{pmatrix} \right\|_2 \leq u + v.$$

   Make sure to carefully handle the signs of u, v.

   (c) *(2 points)* Consider the constraints

   $$x > 0, \;\; x_i(Cx)_i \geq \theta_i(x^T Cx), \;\; i = 1, \dots, n. \tag{3}$$

   Show how they can be written as second-order cone constraints on x.

   (d) *(3 points)* Show that any point $x \in \mathbb{R}^n$ that satisfies (3) also satisfies constraints (2). *Hint:* proceed by contradiction, and think about summing constraints (3).

   (e) *(1 point)* Consider the problem of maximizing an expected return $\hat{r}^T x$, subject to the risk budget constraints. Express the problem as an SOCP.

4. (*10 points*) Let $A \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$ and $\mu > 0$. Consider the problem

$$p^* = \min_x \|Ax - y\|_1 + \mu \|x\|_\infty.$$

In the sequel, for $j \in \{1, \ldots, n\}$, we denote by a_j the j-th column of A, so that $A = [a_1, \ldots, a_n]$, and define

$$\|A\|_1 := \sum_{j=1}^n \|a_j\|_1.$$

(a) *(2 points)* Express the problem as an LP.

(b) *(4 points)* Show that a dual to the problem can be written as

$$d^* = \max_u -u^T y \ : \ \|u\|_\infty \leq 1, \ \ \|A^T u\|_1 \leq \mu.$$

Hint: use the fact that, for any vector z:

$$\max_{u \,:\, \|u\|_1 \leq 1} u^T z = \|z\|_\infty, \quad \max_{u \,:\, \|u\|_\infty \leq 1} u^T z = \|z\|_1.$$

In the sequel, you may assume that strong duality holds.

(c) *(2 points)* Show that the condition "$\|A^T u\|_1 < \mu$ for every u with $\|u\|_\infty \leq 1$" ensures that $x = 0$ is optimal.

(d) *(2 points)* Show that the condition in the previous part holds if $\mu > \|A\|_1$.

EXTRA SPACE

EXTRA SPACE