EE 127 / EE 227AT

Final Spring 2016

Please write your answers on these sheets, use the back sides if needed. Show your work. You can use a fact from the slides/book without having to prove it unless you are specifically asked to do so. Be organized and use readable handwriting. There is a page for scratch work at the end.

Exercise 1 (Cholesky decomposition.) (10 pts) Suppose that the square matrix B has the QR-factorization $B = QR$, where Q is orthogonal and R is upper triangular with positive diagonal terms, and another matrix $A = B^\top B$. Describe the Cholesky method for solving the system of equations $Ax = b$ for some vector b using this information.

Exercise 2 (Ridge regression.) (10 pts) Ridge regression involves solving optimization problems of the form $\min \|Ax - b\|_2^2 + \lambda \|x\|_2^2$, where λ is a positive regularization parameter and A and b are given. Write this problem as an equivalent least-squares problem (without regularization) such that standard least-squares methods can be used.

Exercise 3 (Risk minimization.) Let $c_j \in \mathbb{R}^n$ be a vector of costs faced in a future scenario j, with $j = 1, ..., m$, and $x \in \mathbb{R}^n$ be a decision vector to be optimized subject to the constraints $x \geq 0$ and $\sum_{i=1}^n x_i = 1$. Since we are unsure about the future scenario, we adopt a risk-based formulation where we aim to minimize the α-superquantile. This leads to the problem

$$\min f(\bar{x}) = \xi + \frac{1}{1-\alpha}\frac{1}{m}\sum_{j=1}^{m} \max\{0, c_j^\top x - \xi\}, \text{ subject to the constraints,}$$

where $\bar{x} = (\xi, x) \in \mathbb{R}^{n+1}$. This problem can be reformulated as a linear program with $n+m+1$ variables, $2m + n$ inequality constraints and one equality constraint. There are algorithms that can solve this linear program in computational effort that is proportional to $(m+n)^{3.5}$. In this question we contrast this effort with that of the subgradient method.

The subgradient method for this problem is as follows. Starting from an initial point $\bar{x}_0 \in \mathbb{R}^{n+1}$, the function f is minimized directly without reformulation using the recursion $\bar{x}_{k+1} = P(\bar{x}_k - s_k \nabla f(\bar{x}_k))$, $k = 0, 1, 2, ...$, where $\nabla f(\bar{x}_k)$ is a subgradient of f at $\bar{x}_k$, s_k is a step size, and $P(\cdot)$ is the projection onto the feasible set.

1. (10 pts) Since one can use a chain rule to obtain a subgradient of f at $\bar{x}_k$, it is sufficient to work out a formula for the subgradient of a function $g_j(\bar{x}) = \max\{0, c_j^\top x - \xi\}$ at $\bar{x}_k = (\xi_k, x_k)$. Write such a formula.

2. (5 pts) Suppose n is relatively small, but m is huge. Would you prefer the linear programming approach described above or the subgradient method for solving the above problem. Explain why.

Exercise 4 (Steepest descent method.) (10 pts) It is known that iterations of the steepest descent method generate a sequence of iterates x_k that are gradually closer to optimality with progress bounded by the expression

$$f(x_{k+1}) - p^* \leq c(f(x_k) - p^*),$$

where f is the (convex) function being minimized, p^* its minimum value, and $c \in (0, 1)$ is a constant. Derive a formula in terms of the initial error $\epsilon_0 = f(x_0) - p^*$, with x_0 being the initial solution, for how many iterations it takes this method to reach an x with $f(x) - p^* \leq \epsilon$, where $\epsilon > 0$.

Exercise 5 (Barrier method.) We consider the original problem $\min f_0(x)$ subject to $f_i(x) \leq 0$, $i = 1, ..., m$, and the corresponding logarithmic barrier problem $\min f_0(x) - (1/t) \sum_{i=1}^{m} \log(-f_i(x))$, where $t > 0$.

1. (10 pts) Write the KKT conditions for the original problem.

2. (5 pts) Write the KKT conditions for the barrier problem.

3. (5 pts) Given $t > 0$, let $x^*(t)$ be an optimal solution of the barrier problem and set $\lambda_i^*(t) = -1/(tf_i(x^*(t)))$, $i = 1, ..., m$. Suppose that $f_0, f_1, ..., f_m$ are convex. Show that $x^*(t)$ is a minimum solution of the Lagrangian of the original problem when the multipliers λ_i are set to $\lambda_i^*(t)$ for all $i = 1, ..., m$.

Exercise 6 (Support vector machine.) (10 pts) A Support Vector Machine approach to classification needs to consider the problem

$$\min_w \frac{1}{m}\sum_{i=1}^m \hat{E}(y_i\phi(x_i)^\top w) = \frac{1}{m}\sum_{i=1}^m \max\{0, 1 - y_i\phi(x_i)^\top w\}$$

where $\{x_i, y_i\}_{i=1}^m$ is given data and $\hat{E}(\alpha) = \max\{0, 1-\alpha\}$ is the hinge loss. Write this problem as a linear program.

Exercise 7 (Shape-Constrained Regression.) Suppose that we have some data $\{x_i, y_i\}_{i=1}^m$, with $x_i, y_i \in \mathbb{R}$. We would like to carry out least-squares regression on this data set, but need to ensure that the regression function is convex. We will achieve this using epi-splines. First, we discretize a sufficiently large part of the first axis by the points $m_0, m_1, ..., m_N$. Second, on each segment (m_{k-1}, m_k), we let the regression function be a second-order polynomial of the form $a_0^k + a_1^k x + a_2^k x^2$. The a coefficients are to be determined by the optimization. Consequently, the regression function $f_a(x)$ is a one-dimensional function defined on $[m_0, m_N]$, which is piecewise polynomial with coefficients determined by the various a-coefficients. Least-squares regression aims to find such a regression function that has a small error relative to the observed data.

1. (5 pts.) Write an objective function that expresses the least-squares criterion in this case. You can assume that the regression function is continuous.

2. (10 pts.) Write a set of constraints that ensures that the regression function is convex. Make sure you include constraints that enforce continuity of the regression function.

Exercise 8 (Control.) (5 pts.) We need to develop a control algorithm for a robot that moves at constant speed v in a two-dimensional space. The robot's motion is modeled as a Dubin's vehicle, i.e., $\dot{x}_1(t) = v \cos x_3(t)$, $\dot{x}_2(t) = v \sin x_3(t)$, and $\dot{x}_3(t) = u(t)$, where $x_1(t)$ and $x_2(t)$ are the coordinates in the plane at time t and $x_3(t)$ is the heading at time t. The control input at time t is $u(t)$. Consider Euler's method of solution of this differential equation with time discretization step Δt so that the discretized version of the optimal control problem will involve the state variables $x_i(k\Delta t)$, for $i = 1, 2, 3$ and $k = 0, 1, 2, ..., N$.

Using the state variables at the discretized points in time, write a set of constraints that ensures that the robot is never further away from a desired trajectory given by $\{(y_1(t), y_2(t)), t \geq 0\}$ than the robot's second coordinate at the same points in time. What type of constraints will this be?

Exercise 9 (Positive definiteness.) (5 pts) Consider an n-by-n symmetric matrix with smallest eigenvalue $\lambda_{\min} > n$. Suppose that this matrix is augment with one row at the bottom and one column to the right consisting exclusively of ones. The augmented matrix is then of dimension $(n+1)$-by-$(n+1)$. Recall that an m-by-m matrix of only ones has largest eigenvalue of m. Prove that the augmented matrix is positive definite.

(blank)