UC Berkeley
Department of Electrical Engineering and Computer Science

EECS 227A
Nonlinear and Convex Optimization

Problem Set 6
Fall 2009

Issued: Thursday, November 5 **Due:** Thursday, November 19, 2009

Problem 6.1
B & V, Problem 5.21

Problem 6.2
B & V, Problem 5.31

Problem 6.3
B & V, Problem 5.39

Problem 6.4
B & V, Problem 5.43

Problem 6.5
Consider a problem of the form $p^* = \min_x \ f(Ax + b) + \frac{1}{2}\|x\|_2^2$ where $f : \mathbb{R}^m \to \mathbb{R}$ is a convex function (whose epigraph is a closed set), and $A \in \mathbb{R}^{m \times n}$ is a given matrix. The purpose of this exercise is to show that the optimal value p^* is a convex function of "kernel matrix" $K := AA^T \in \mathbb{R}^{m \times m}$. (This fact has important computational consequences for classification and regression problems.)

(a) Form a dual for the problem. *Hint:* introduce extra variables and constraints, and use the conjugate of f in your expression of the dual.

(b) Show that strong duality holds, using the result of Exercise 5.25 from BV, and prove that the function p^* is convex in K.

(d) Show that the primal problem is attained, and its solution unique. Show how to obtain the primal solution once a dual solution is found. *Hint:* use the result of [BV,§5.5.5].

(c) Express the dual in the following cases:

(i) *Support vector machines classification,* in which $b = \vec{1}$ and $f(r) = \sum_{i=1}^m \max\{0, r_i\}$.

(ii) *Least-squares regression* in which $f(r) = \frac{1}{2}\|r\|_2$.

(iii) *Least-norm regression* in which $f(r) = \|r\|$, where $\|\cdot\|$ is a norm. (Use the notation $\|\cdot\|_*$ for the dual norm.) In the case when $K \succ 0$, express the dual problem as the minimum distance (in some appropriately defined weighted Euclidean norm) to a set.

Problem 6.6

In this exercise, we explore the use of Newton's method for solving constrained optimization problems, as well as their Lagrangian duals. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and twice continuously differentiable, and consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to} \qquad Ax = b, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$. A Newton step for this problem can be derived as follows. Given a fixed $\bar{x} \in \mathbb{R}^n$ that is feasible (i.e., $A\bar{x} = b$), consider the quadratic approximation $f(\bar{x}+z) \approx f(\bar{x}) + \nabla f(\bar{x})^T z + \frac{1}{2} z^T \nabla^2 f(\bar{x}) z$. We interpret $z \in \mathbb{R}^n$ as the perturbation from the feasible $\bar{x}$ to another feasible point $x^{n+1} = \bar{x} + z$, so that we must have $Az = 0$. Thus, the Newton step z^* can be obtained by solving the constrained quadratic program of minimizing

$$g(z) = \nabla f(\bar{x})^T z + \frac{1}{2} z \nabla^2 f(\bar{x}) z \qquad \text{subject to } Az = 0. \tag{2}$$

If z^* is the optimal solution, then Newton step has the form $x^{n+1} = x^n + \alpha^n z^*$. With this background, consider the following questions:

(a) Let $\lambda \in \mathbb{R}^m$ be a Lagrange multiplier associated with the constraint $Az = 0$. Show that the Lagrangian conditions for the optimal (z^*, λ^*) to the QP (2) can be written in matrix form as:

$$\begin{bmatrix} \nabla^2 f(\bar{x}) & A^T \\ A & 0_{m \times m} \end{bmatrix} \begin{bmatrix} z^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -\nabla f(x^n) \\ 0 \end{bmatrix}. \tag{3}$$

Conclude that if the matrix

$$P(\bar{x}) = \begin{bmatrix} \nabla^2 f(\bar{x}) & A^T \\ A & 0_{m \times m} \end{bmatrix}$$

is invertible, then the Newton update is given by $x^{n+1} = x^n + \alpha^n z^*$ where

$$z^* = \begin{bmatrix} I_{n \times n} & 0_{n \times m} \end{bmatrix} P^{-1}(\bar{x}) \begin{bmatrix} -\nabla f(x^n) \\ 0 \end{bmatrix}.$$

(b) Consider the *constrained entropy maximization* problem of finding the distribution with maximum entropy subject to linear constraints. It is a problem of the form (1), with

$$f(x) = \sum_{i=1}^{n} x_i \log x_i, \qquad \text{dom}(f) = \{x \in \mathbb{R}^n \mid x \succ 0\}$$

Using the above expression for the Newton step, implement Newton's method (using Armjio rule for step size choice) to solve the problem with problem data A, b and initial condition x^0 given on the website. Plot the difference $\log |f(x^n) - p^*|$ versus iteration number n.

(c) Now consider a Lagragian dual formulation of the problem. Show that the dual problem is to perform an unconstrained maximization of the dual function

$$q(\lambda) = -b^T \lambda - \sum_{i=1}^{n} \exp(-1 - a_i^T \lambda)$$

where $\lambda \in \mathbb{R}^m$ and $a_i \in \mathbb{R}^m$ is the i^{th} column of A.

(d) Implement Newton's method with Armijo rule to solve the dual problem (with the same problem data as in (b)), and verify that the optimal dual value q^* is equal to the optimal primal value p^*. Plot the difference $\log |q(\lambda^n) - q^*|$ versus iteration number n.