First-Stage Comprehensive Examination in Continuous Optimization

Examiners: T. Coleman and S. Vavasis
Department of Combinatorics and Optimization, University of Waterloo

2012 June 20, 9:00 am to 12:00 noon

1. Assume that $f : \mathbf{R}^n \to \mathbf{R}$ is a twice continuously differentiable function of $\mathbf{x}$.

   (a) Show that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ is a necessary but not sufficient condition for $\bar{\mathbf{x}}$ to be a local minimizer of f.

   (b) Define condition (1) at a point $\mathbf{x}$:
   $$\nabla f(\bar{\mathbf{x}}) = \mathbf{0}, \quad \nabla^2 f(\bar{\mathbf{x}}) \text{ is positive semidefinite.} \tag{1}$$
   Prove or disprove: (1) is necessary and sufficient for $\bar{\mathbf{x}}$ to be a local minimizer of f.

   (c) Let $\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^n$ be a vector-valued function where component i of $\mathbf{F}$ is $F_i : \mathbf{R}^n \to \mathbf{R}$, F_i is continuously differentiable $i = 1, \ldots, n$. Let $f(\mathbf{x}) \triangleq \|\mathbf{F}(\mathbf{x})\|_2^2$. Prove or disprove: If $\bar{\mathbf{x}}$ satisfies $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ then either $\bar{\mathbf{x}}$ is a global minimizer of f or the Jacobian of $\mathbf{F}$ at $\bar{\mathbf{x}}$ is singular.

2. Let $\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^n$ be a vector-valued function where component i of $\mathbf{F}$ is $F_i : \mathbf{R}^n \to \mathbf{R}$, F_i is twice continuously differentiable, $i = 1, \ldots, n$.

   (a) What is Newton's method for the problem: solve $\mathbf{F}(\mathbf{x}) = \mathbf{0}$? What is the computational cost of Newton's method?

   (b) One popular approach to avoiding the expense of computing the Jacobian of $\mathbf{F}$ at each iteration is to use the rank-one secant (or Broyden) update to a current nonsingular Jacobian approximation B_c:
   $$B_+ = B_c + \frac{(\mathbf{y} - B_c\mathbf{s})\mathbf{s}^T}{\mathbf{s}^T\mathbf{s}}, \tag{2}$$
   where $\mathbf{s}$ is the step to the new point $\mathbf{x}_+$ from the current point $\mathbf{x}_c$, $\mathbf{s} = \mathbf{x}_+ - \mathbf{x}_c$, and $\mathbf{y}$ is the difference in the function values, $\mathbf{y} = \mathbf{F}(\mathbf{x}_+) - \mathbf{F}(\mathbf{x})$. Derive (2) using an optimization argument and the quasi-Newton condition: $B_+\mathbf{s} = \mathbf{y}$. Why is (2) not often used in the case $\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^m$, $m > n$?

3. Consider solving the nonlinear equality-constrained problem

$$\min_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{c}(\mathbf{x}) = \mathbf{0}\}, \tag{3}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is twice continuously differentiable, and $\mathbf{c} : \mathbf{R}^n \to \mathbf{R}^m$, $m < n$, and each component function $c_i : \mathbf{R}^n \to \mathbf{R}$ is twice continuously differentiable. One approach to (3) is to minimize the penalty function

$$p(\mathbf{x}, \mu) = f(\mathbf{x}) + \frac{1}{2\mu}\|\mathbf{c}(\mathbf{x})\|_2^2, \tag{4}$$

with respect to $\mathbf{x}$ for a selection of positive values of μ, usually decreasing to zero.

(a) For a fixed value of μ, what is the gradient and Hessian of (4) with respect to $\mathbf{x}$?

(b) Show that as $\mu \to 0$, $\nabla^2_{\mathbf{xx}} p(\mathbf{x}, \mu)$ approaches singularity. Why does this (apparently) pose a problem?

(c) Suggest a possible way to overcome the asymptotic ill-conditioning problem indicated by (b) but without abandoning the use of the penalty function (4).

(d) Indicate why a descent direction algorithm for (4) may require many steps to converge to a minimizer of (4) if starting from a point $\bar{\mathbf{x}}$ where $\mathbf{c}(\bar{\mathbf{x}}) = \mathbf{0}$ and μ is small ($\bar{\mathbf{x}}$ is not necesssarily close to a local minimizer of (4)).

(e) Suppose (4) is replaced with

$$p(\mathbf{x}, \mu) = f(\mathbf{x}) + \frac{1}{2\mu}\|\mathbf{c}(\mathbf{x})\|_1 \tag{5}$$

as an approach to solve (3). What is an advantage of using (5) over (4)? Disadvantage? If (3) has a (feasible) *global* minimizer, is (5) bounded below for sufficiently small but finite μ? Explain.

4. Consider the optimization problem of

$$\min \|A\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{x}\|_1, \tag{6}$$

where A is a given $m \times n$ matrix whose rank is m, i.e., its rows are linearly independent, and $\|\mathbf{x}\|_1$ as usual stands for $|x_1| + \cdots + |x_n|$.

(a) This problem may be rewritten as

$$\begin{array}{lll} \min & \|A\mathbf{x} - \mathbf{b}\|_2^2 + y_1 + \cdots + y_n, & \\ \text{subject to} & y_i \geq x_i, & i = 1, \ldots n, \\ & y_i \geq -x_i, & i = 1, \ldots n. \end{array} \tag{7}$$

Demonstrate the equivalence between (6) and (7).

(b) Both problems (6) and (7) are convex optimization problems. Explain why.

(c) Write down the Lagrangian function, the dual function, and the dual optimization problem for the reformulation (7). Be sure to include constraints in the dual optimization problem to eliminate the possibility that the dual objective function takes on negative infinite values over its feasible region.

5. Consider again the optimization problem (6) and its reformulation (7).

   (a) For (6), write down a condition for $\mathbf{x}^*$ to be a global optimizer in terms of the subdifferential of the objective function. Determine the subdifferential.

   (b) For (7), write down the KKT conditions for $\mathbf{x}^*$ to be a global optimizer.

   (c) It is no surprise that there is a close relationship between the conditions in (a) and (b). Determine that relationship.

6. Consider the problem of minimizing $\|\mathbf{x} - \mathbf{x}_0\|_2$ subject to $\mathbf{x} \in C$, where $\mathbf{x}_0$ is a given point in $\mathbf{R}^n$ and $C \subset \mathbf{R}^n$ is a nonempty closed convex bounded set.

   (a) This problem has a unique solution. Why? [Hint: use compactness for existence and convexity for uniqueness.]

   (b) Show that the hypothesis that C is bounded can be dropped, and the theorem is still true.

   (c) Show via counterexamples that the hypothesis that C is closed and the hypothesis that C is convex are both necessary (cannot be dropped).