ME 555 Sample Exam 3 Solution - April 2013
Closed Book Closed Notes

Problem 1

Check if the following statements are true or not. Explain concisely.

(a) For the function $f = |x|$, Newton's method will converge to the optimum $x_* = 0$, when the starting point is close to the solution.

(b) The performance of the gradient method is very sensitive to scaling.

(c) Quasi-Newton methods always find the global unconstrained optimum.

(d) Using quasi-Newton methods, like BFGS, in SQP without line search guarantees global convergence.

(e) You can increase exactness of an inexact line search by making the acceptability interval larger.

(f) A line search uses quadratic interpolation with information from a single point is the same as Newton's method.

Problem 1 Solution

(a) False. For pure Newton's method, the Hessian at $x \neq 0$ will be zero thus its inverse does not exist.

(b) True.

(c) False. Quasi-Newton offers low computational cost and positive definite Hessian, therefore makes the Newton's method globally convergent. However, finding a global optimum is never guaranteed.

(d) False. We need line search to choose an appropriate step size in order to keep $\mathbf{x}_k$ not close to the feasible domain.

(e) False. For a given tracking scheme of α, increasing the interval will only reduce the number of iterations of the line search.

(f) True.

Problem 2

A problem in the form

$$\min f(x) \text{ subject to } h(x) = 0$$

is scaled, becoming of the form

$$\min K_1 f(x) \text{ subject to } K_2 h(x) = 0$$

What is the relationship between the Lagrange multipliers of h in the unscaled and scaled problems? What does this relationship imply for sensitivity analysis?

Problem 2 Solution

The KKT conditions for the two problems require

$$\nabla f_u + \lambda_u \nabla h_u = 0,$$

and

$$K_1 \nabla f_s + \lambda_s K_2 \nabla h_s = 0.$$

Notice that the two problems have the same solution for x. Therefore we have $\nabla f_u = \nabla f_s$ and $\nabla h_u = \nabla h_s$. Thus

$$\nabla h_u(-K_1 \lambda_u + K_2 \lambda_s) = 0,$$

or, for arbitrary ∇h_u, we need

$$\lambda_s = K_1 / K_2 \lambda_u.$$

Problem 3

For the linearly constrained problem (where $\mathbf{A}$ and $\mathbf{b}$ are a matrix and a vector of parameters, respectively):

$$\min \quad f(x), \text{ subject to } \mathbf{Ax} = \mathbf{b}$$

(a) Derive the expression for the reduced gradient.

(b) State the general iterative formula for taking a step in the reduced space using a quasi-Newton method (you do not need to state the actual update formula).

(c) State all steps and associated iterative formulas in a reduced gradient algorithm designed for this type of problem.

(d) Derive the Lagrange-Newton Equations for this problem. This involves the following steps: (i) State the Lagrangian function. (ii) State the KKT conditions for the Lagrangian. (iii) Apply Newton's equation-solving method to the Lagrangian KKT stationary conditions.

(e) Compare your answer in (c) with a typical GRG algorithm and briefly explain what simplifications occurred due to the linearity of constraints.

(f) Compare your answer in (d) with a typical SQP algorithm and briefly explain what simplifications occurred due to the linearity of constraints.

Problem 3 Solution

(a) The reduced gradient is defined as

$$\frac{\partial z}{\partial d} = \frac{\partial f}{\partial d} + \frac{\partial f}{\partial s}\frac{\partial s}{\partial d}$$

Partitioning $x^T = [s, d]$ and $A = [A_s, A_d]$, the linear equality constraints can be rewritten as

$$A_s s + A_d d = b,$$

or

$$s = A_s^{-1}(b - A_d d),$$

when A_s is invertible. This further gives us the relationship

$$\partial s = -A_d A_s^{-1} \partial d.$$

Therefore we have

$$\frac{\partial z}{\partial d} = \frac{\partial f}{\partial d} - \frac{\partial f}{\partial s} A_d A_s^{-1}.$$

(b) See textbook or notes

(c) See textbook or notes

(d) (a) Lagrangian $L = f + \lambda^T(Ax - b)$

(b) From KKT conditions, we have: $\nabla L_x = \nabla f + \lambda^T A = 0$, $\nabla L_\lambda = Ax - b = 0$. The goal is then to find solutions x and λ to meet these two equations.

(c) In order to apply Newton-Ralphson to the above equations, we need to linearize ∇L_x:

$$(\nabla L)_{k+1}^T = (\nabla L)_k^T + (\nabla^2 L)_k [\partial x_k, \partial \lambda_k]^T = 0,$$

where $(\nabla L)_k = [\nabla f + \lambda^T A, h^T]_k$ and $(\nabla^2 L)_k = [\nabla^2 f \ A^T; \ A \ 0]_k$. Therefore

$$[\nabla^2 f \ A^T; \ A \ 0]_k [x_{k+1} - x_k; \lambda_{k+1} - \lambda_k] = -[\nabla^T f + A\lambda; h]_k$$

are the Lagrange-Newton Equations.

(e) When only linear equality constraints exist, we don't need to use Newton-Ralphson for finding feasible ∂s in GRG.

(f) When only linear equality constraints exist, x goes back and stays as a feasible solution after one iteration. Therefore no need to add penalties in the merit function for line search.

Problem 4

Consider the problem

$$
\begin{aligned}
&\min\ f = x_1^2 + x_2^2 - 3x_1x_2 \\
&\text{subject to} \\
&\quad g_1 = \frac{1}{6}x_1^2 + \frac{1}{6}x_2^2 - 1 \le 0, \quad g_2 = -x_1 \le 0, \quad g_3 = -x_2 \le 0.
\end{aligned}
$$

(a) Solve the problem analytically and provide a graphical sketch to aid visualization.

(b) Linearize f and g_1 about the point $\mathbf{x}_0 = (1,1)^T$.

(c) Solve the resulting linear programming problem (including g_2 and g_3) using monotonicity analysis.

(d) Confirm that point $\mathbf{x}_1 = (2,2)^T$ is a solution of the problem defined in (c). Linearize the original problem again at $\mathbf{x}_1$ and solve again.

(e) Steps (b)–(d) can form an algorithm that solves a nonlinear problem with a sequence of approximating subproblems. Describe formally what the steps for such an algorithm may be.

(f) List advantages and disadvantages (briefly) for such a *Sequential Linear Programming* (SLP) algorithm.

(g) Discuss (briefly) how the basic algorithm in (e) can be modified for better performance.

Problem 4 Solution

(a) This problem is solved in midterm 1.

(b) At any point $\mathbf{x}_0$, the linearized problem is

$$
\begin{aligned}
\min\ f' &= f_0 + \nabla f_0 \Delta\mathbf{x} \\
\text{subject to } g_1' &= g_{1,0} + \nabla g_{1,0}\Delta\mathbf{x} \le 0 \\
g_2 &= -x_1 \le 0 \\
g_3 &= -x_2 \le 0 \\
&\quad \text{where } \Delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_{i,0}
\end{aligned}
$$

At $\mathbf{x}_0 = (1,1)^T$ we have

$$
\begin{aligned}
f' &= (-1) + (-1)\Delta x_1 + (-1)\Delta x_2 = -1 - \Delta x_1 - \Delta x_2 \\
g_1' &= -4 + 2\Delta x_1 + 2\Delta x_2 \le 0
\end{aligned}
$$

In terms of x_1 and x_2, the LP subproblem is

$$\begin{aligned}
\min f' &= 1 - x_1 - x_2 \\
\text{subject to } g_1' &= -8 + 2x_1 + 2x_2 \le 0 \\
g_2 &= -x_1 \le 0 \\
g_3 &= -x_2 \le 0
\end{aligned}$$

(c) Solution of LP subproblem.
By MP1, g_1' is active, i.e., $x_{1*} + x_{2*} = 4$. Note that g_1 is multiply critical with respect to x_1 and x_2, and that $f_*' = -3$. Thus, the solution is a valley.

(d) The point $\mathbf{x}_1 = (2,\ 2)^T$ lies in the valley defined by $x_{1*} + x_{2*} = 4$. At $(2,\ 2)^T$ the new linearization gives

$$\begin{aligned}
\min f' &= 4 - 2x_1 - 2x_2 \\
\text{subject to } g_1' &= -14 + 4x_1 + 4x_2 \le 0 \\
g_2 &= -x_1 \le 0 \\
g_3 &= -x_2 \le 0
\end{aligned}$$

The solution is the new valley $2x_{1*} + 2x_{2*} = 7$.

(e) A sequential Linear Programming (SLP) algorithm may be as follows

(a) 1. Start at $\mathbf{x}_0$.

(b) 2. Linearize at $\mathbf{x}_0$ and sole the LP problem

$$\begin{aligned}
\min f' &= f_0 + \nabla f_0 \Delta \mathbf{x} \\
\text{subject to } h_i' &= h_{i,0} + \nabla h_{i,0} \Delta \mathbf{x} \le 0 \\
g_j' &= g_{j,0} + \nabla g_{j,0} \Delta \mathbf{x} \le 0
\end{aligned}$$

(c) 3. Let $\Delta \mathbf{x}_{*0}$ be the solution for the LP problem. Set $\mathbf{x}_1 = \mathbf{x}_0 + \Delta \mathbf{x}_{*0}$.

(d) 4. Check the termination criteria. If not met, return to step 2. linearizing at $\mathbf{x}_1$ instead of $\mathbf{x}_0$.

The algorithm generalizes by using k, k+1 instead of 0,1 for indices

(f)
- Advantages: Simplicity in formulation and solution of LP problems; standard LP codes can be used. Good convergence for problem with extensive monotonicities and constrained-bound solutions.
- Disadvantages: Slow or no convergence for non-convex, highly nonlinear problem: Interior solutions cannot be found since LP subproblem will always have constrained-bound solutions.

(g) There are two observations related to better performance: i) the LP solution may be infeasible with respect to the NLP problem; ii) the LP solution is always constrained-bound and may not correspond to an interior solution for the NLP problem. To address these problems, a modification to the simple algorithm is as follows:

Artificial simple bounds are imposed on each variable x_i,

$$x_{i,L} \leq x_i \leq x_{i,U}$$

where the parameters $x_{i,L}$ and $x_{i,U}$ are called "move limits" and are adjusted during iterations. The move limits create a hypercube around each linearization point. Adjustment of the move limits can force the solution of the LP problem to become NLP-feasible; further, small more limits can force an LP solution to be interior to the original NLP.

[This is an example of algorithm that employ a "trust region" strategy to determine a total move, rather than a line search , once a direction has been established.]

Note: Commercial-grade SLP algorithms are used mostly for solving structural optimization problems, particularly so-called size problems (rather than shape problems).