EECS 967 Optimization with Communications Applications
Fall Semester 2011
Midterm Exam 25 October 2011

NAME: __________________________ KUID: KF7Y

GENERAL INSTRUCTIONS

1. Put your KUID on each exam page, in case the pages get separated.

2. There are 80 points possible on this exam.

3. In case you have forgotten, the solutions to the quadratic equation $ax^2 + bx + c = 0$ are
   $$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

4. Show your work:
   Partial credit awarded for correct set-up. Even if your answer is correct, you will not receive full credit unless you show how you arrived at that answer. For proofs, give a reason justifying each step.

5. Show all work on each problem on the exam pages. I have tried to leave lots of room for you to show your work, and the back of the last page is blank for you to use for any problem. Please do not use any other pages unless absolutely necessary. If you use the back of the last page or other pages, clearly indicate on the problem page that there is additional work elsewhere (and where that work is).

6. Clearly indicate your answer to each part (underline, box, etc.).

7. When numerical values are requested, perform all calculations rather than leaving your answer as some complicated expression. Also, give your answers as decimal numbers (not fractions).

8. Some problems may give more information than is needed. It is also possible that not enough information has been given. If you have any question about any problem, ask me about it.

9. Stay calm. If you are having trouble with one problem or part of a problem, leave it and go on. You may be able to work the last parts of a problem without working the first parts.

10. If you get a numerical value that you know cannot be correct, but you can't find your mistake, say so and explain why the answer cannot be correct. Being wrong and knowing it is worth more than being wrong and oblivious!
1. Consider the following problem:

\[
\text{minimize: } f(x, y) = x^3 + 2x^2 + x + \frac{4}{3}y^3 + 2.5y^2 + y
\]

a. (8 points) Find all points that meet the first-order necessary conditions (FONC) for this problem.

b. (8 points) For each point that you found in part a, make the strongest possible statement concerning the point as a possible solution to this optimization problem.

c. (4 points) What can you say about the convexity of \(f(x, y)\)? Justify your answer.

There is more space for this problem on the next page.

\[
\text{FONC: } \nabla f(x, y) = (0, 0)
\]

\[
\nabla f(x, y) = (3x^2 + 4x + 1, 4y^2 + 5y + 1)
\]

\[
x = \frac{-4 \pm \sqrt{4^2 - (4)(3)}}{2(3)}
\]

\[
y = \frac{-5 \pm \sqrt{5^2 - (4)(4)}}{2(4)}
\]

\[
x = \frac{-4 \pm 2}{6} = -\frac{1}{3} \text{ or } -1
\]

\[
y = \frac{-5 \pm 3}{8} = -\frac{1}{4} \text{ or } -1
\]

Points that meet FONC:

\[
\begin{pmatrix}
-\frac{1}{3} \\
-\frac{1}{4}
\end{pmatrix},
\begin{pmatrix}
-\frac{1}{3} \\
-1
\end{pmatrix},
\begin{pmatrix}
-1 \\
-\frac{1}{4}
\end{pmatrix},
\begin{pmatrix}
-1 \\
-1
\end{pmatrix}
\]

or

\[
\begin{pmatrix}
-0.33 \\
-0.125
\end{pmatrix},
\begin{pmatrix}
-0.33 \\
-1
\end{pmatrix},
\begin{pmatrix}
-1 \\
-0.125
\end{pmatrix},
\begin{pmatrix}
-1 \\
-1
\end{pmatrix}
\]

b. Check 2nd order conditions: no constraints, so all points inferior.

\[
\mathbf{H} = \begin{pmatrix}
0 & 4 \\
0 & 8
\end{pmatrix}
\]

---

Prof. Petr  Copyright 2011 David W. Petr Fall 2011
More space for previous problem.

\[
\begin{bmatrix}
-0.33 \\
-0.25
\end{bmatrix}
\begin{bmatrix}
-2 + 4 & 0 \\
0 & -2 + 5
\end{bmatrix}
\begin{bmatrix}
0.2 \\
0.3
\end{bmatrix} = \begin{bmatrix}
4 \\
0.5
\end{bmatrix}
\begin{vmatrix}
0.2
\end{vmatrix} = 2 > 0
\begin{vmatrix}
0.3
\end{vmatrix} = 0 > 0

\text{(0.33) meets SOSC \Rightarrow it is a strict relative min pt.}

\[
\begin{bmatrix}
-0.33 \\
-1
\end{bmatrix}
\begin{bmatrix}
-2 + 4 & 0 \\
0 & -8 + 5
\end{bmatrix}
\begin{bmatrix}
0.2 \\
0.3
\end{bmatrix} = \begin{bmatrix}
-2 \\
-23
\end{bmatrix}
\begin{vmatrix}
0.2
\end{vmatrix} = -6 < 0
\begin{vmatrix}
0.3
\end{vmatrix} = -2 < 0

\text{(0.33) fails SOSC \Rightarrow it is not a rel. min. pt.}

\[
\begin{bmatrix}
-1 \\
-0.25
\end{bmatrix}
\begin{bmatrix}
-6 + 4 & 0 \\
0 & -2 + 5
\end{bmatrix}
\begin{bmatrix}
0.2 \\
0.3
\end{bmatrix} = \begin{bmatrix}
-2 \\
-23
\end{bmatrix}
\begin{vmatrix}
0.2
\end{vmatrix} = -2 < 0
\begin{vmatrix}
0.3
\end{vmatrix} = -2 < 0

\text{(1) fails SOSC \Rightarrow it is not a rel. min. pt.}

C. Since \( f(x,y) \) not pos. def. everywhere, it is not convex.
2. (10 points) We wish to minimize a particular function \( f(x_1, x_2) \) subject to \( x_1 \geq 0, \ x_2 \geq 0 \). The gradient and Hessian of \( f \) are:

\[
\nabla f = (2x_1 - 8 \ 6 - 2x_2) \quad \text{and} \quad F(x) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}
\]

For each of the points \( x = (0 \ 3)^T \) and \( x = (4 \ 0)^T \), evaluate all appropriate optimality conditions and make the strongest possible statement about each point as a possible solution to this optimization problem.

There is more space for this problem on the next page.

Neither point is an interior point.

\( (0) \):
 allowed directions are \((d_1, d_2)\) with \( d_1 \geq 0 \).

**FONC:** \( \nabla f(0)^T d \geq 0 \) for all allowed \( d \).

\[
(2(0)-8 \ (6-2(3))(d_1) = \begin{pmatrix} -8 \ 0 \end{pmatrix}(d_1) = -8d_1 \leq 0 \text{ since } d_1 \geq 0
\]

Conclusion: \((0)\) fails FONC \(\Rightarrow\) it is not a rel min pt.

\( (3) \):
 allowed directions are \((d_1, d_2)\) with \( d_2 \geq 0 \).

**FONC:** \( \nabla f(3)^T d \geq 0 \) for all allowed \( d \).

\[
(2(3)-8 \ (6-2(3))(d_1) = \begin{pmatrix} 0 \ 6 \end{pmatrix}(d_1) = 6d_2 \geq 0 \text{ since } d_2 \geq 0
\]

\( (3) \) meets FONC, so check SONE.

**SONE:**

1) \( \nabla f(3)^T d \geq 0 \) for all allowed \( d \) (FONC) \(\checkmark\)

2) \( \nabla f(3)^T d = 0 \Rightarrow d^T F d \geq 0 \)

From above, \( \nabla f(3)^T d \) = \( 6d_2 \) (next page)
More space for previous problem.

So, \( \forall t \in (0, b) \), \( d_1 = 0 \Rightarrow d_2 = 0 \).

\[
\begin{pmatrix}
  d_1 & 0 \\
  0 & 2
\end{pmatrix}
\begin{pmatrix}
  d_1 \\
  0
\end{pmatrix}
= \begin{pmatrix}
  2 d_1 & 0 \\
  0 & 2
\end{pmatrix}
\begin{pmatrix}
  d_1 \\
  0
\end{pmatrix}
= 2d_1^2 \geq 0 \quad \forall d_1
\]

and hence \((4)\) meets SNC.

There is no SOSC for non-inferior points.

Conclusion: \((4)\) meets both FORC and SOSC, and so may be a rel. min. of \( f(x_1, x_2) \).
3. (15 points) You are given a nonlinear function $f(x)$ to be minimized and a general iterative descent direction algorithm as follows. At step $k$ of the algorithm, the current point is $x_k$. A direction $d_k$ is chosen, and the next point $x_{k+1}$ is given by $x_{k+1} = x_k + \alpha d_k$ for some $\alpha > 0$.

Derive the condition that $d_k$ must meet in order to guarantee that $d_k$ is a descent direction, that is, to guarantee that there exists an $\alpha > 0$ such that $f(x_{k+1}) < f(x_k)$. Justify all steps of your derivation.

$$\phi(\alpha) = f(x_k + \alpha d_k) = f(x_{k+1})$$

Now $f(x_{k+1}) < f(x_k)$ is equivalent to $\phi(\alpha) < \phi(0)$ for some $\alpha > 0$.

This can be guaranteed for some $\alpha > 0$ (perhaps $\alpha$ very small) if $\phi'(0) < 0$ (i.e., slope of $\phi(\alpha)$ at $\alpha = 0$ is negative).

$$\phi(\alpha) = f(x_k + \alpha d_k)$$

$$\phi'(\alpha) = \nabla f(x_k + \alpha d_k) \cdot d_k \quad \text{(Chain Rule)}.$$ 

$$\phi'(0) = \nabla f(x_k) \cdot d_k = g_k \cdot d_k < 0$$
4. (10 points) Discuss how the Trust Region approach to multi-dimensional unconstrained nonlinear optimization addresses the problems/limitations of the "pure" multi-dimensional Newton method.

Pure Newton: Model \( f(x) \) as quadratic thru \( x_k \).

\[
x_{k+1} = x_k - F_k^{-1} g_k
\]

where \( g_k = \nabla f(x_k) \) and \( F_k \) is \( f''(x_k) \), the Hessian.

Some problems:

a) \( F_k^{-1} \) may not exist.
b) If \( F_k^{-1} \) exists, \( f(x_{k+1}) \geq f(x_k) \) possible due to poor match between \( f(x) \) and quadratic model "far" from \( x_k \).

Trust Region: Model \( f(x) \) as quadratic w/ matrix \( B_k \) that approximates \( F_k \):

\[
M_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p
\]

Then, attempt to solve: \( \min_{p} M_k(p) \) s.t. \( p \leq \Delta_k \) (trust region radius).

a) With \( B_k = F_k \), inverse still may not exist, but TR gives alternatives in this case.
b) \( \Delta_k \) limits the "range" of model function application -- specifies a neighborhood where model function can be trusted to match actual obj. function reasonably well.
5. (10 points) In the linear conjugate gradient algorithm, one of the steps for iteration $k$ is

$$p_{k+1} = -r_{k+1} + \beta_k p_k$$

where $p_{k+1}$, $r_{k+1}$, and $p_k$ are all column vectors of dimension $n$, and the scalar $\beta_k$ is chosen to ensure that $p_{k+1}$ is conjugate to $p_k$ with respect to the symmetric $n \times n$ positive definite matrix $A$. Derive the expression for $\beta_k$.

$p_{k+1}$ conjugate to $p_k$ means $p_{k+1}^T A p_k = 0$

$$p_{k+1}^T A p_k = (-r_{k+1} + \beta_k p_k)^T A p_k = 0$$

$$\Rightarrow \beta p_k^T A p_k = r_{k+1}^T A p_k$$

$$\Rightarrow \beta = \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$$
6. Let \( A \) be a point-to-set mapping from \( E^2 \) to \( E^2 \) defined as:
\[
A(x_1, x_2) = \{(y_1, y_2) \in [-0.9 \cdot |x_1|, 0.9 \cdot |x_1|], y_2 \in [-0.9 \cdot |x_1|, 0.9 \cdot |x_1|]\}
\]

Note that both \( y_1 \) and \( y_2 \) depend only on \( x_1 \). Now define an iterative algorithm \( x_{k+1} \in A(x_k) \).
Further define the solution set to be \( \Gamma = \{(0, 0)\} \).

a. (5 points) State the conditions for \( Z(x) \) to be a valid descent function.

b. (10 points) Is \( Z(x) = x^T x \) a valid descent function for \( A \) and \( \Gamma \)? You must justify your answer.

1. i) For \( x_k \notin \Gamma \) and \( x_{k+1} \in A(x_k) \), \( Z(x_{k+1}) < Z(x_k) \).

ii) For \( x_k \in \Gamma \) and \( x_{k+1} \in A(x_k) \), \( Z(x_{k+1}) \leq Z(x_k) \).

b). For given \( \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \), \( A \) defines a square centered at \( (0, 0) \) with sides having length \((2 \cdot 0.9) \cdot (1 \cdot 1)\), so "new" \( x_1 \) is guaranteed to have magnitude \( \leq \) mag of "old" \( x_1 \) and the square shrinks in size with every iteration. In the limit, the square becomes the point \((0, 0)\), however, since the size of the square depends only on \( x_1 \), \( Z(x_{k+1}) > Z(x_k) \) possible.

Example: \( x_k = (0) \) and \( Z(x_k) = \begin{pmatrix} 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \end{pmatrix} \cdot 1 = 1 \)

But then \( x_{k+1} = \begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix} \in A(x_k) \), but \( Z(x_{k+1}) \) is \( 0.8 \cdot 0.8 \cdot 0.8 \cdot 0.8 = 0.64 + 0.64 = 1.28 < 1 = Z(x_k) \).

So, \( Z \) is not a descent function.
More space for any problem.