MA/OR/ST 706: Nonlinear Programming
Final Exam
Instructor: Dr. Kartik Sivaramakrishnan
Date: April 29, 2008

Solutions To the Final Exam
Prepared By Kartik

INSTRUCTIONS

1. Please write your name and student number clearly on the front page of the exam.

2. This has to be your own work. Cheating on the exam is not tolerated, and will fetch
you a zero for the test.

3. TIME LIMIT: 3 hours

4. There are 6 pages and 5 questions on the exam. Each question appears on a different
page. Read each question carefully.

5. The exam is worth 160 points. The distribution of points is clearly indicated on the
exam.

6. Solve each problem in sufficient detail in the space provided. Please use both sides of
each page as needed.

7. Write clearly, including all the steps to the final solution. If I can't read it, you won't
get credit.

8. Sources: Nocedal & Wright (2nd edition), class notes and homeworks, and solutions
to homeworks and midterm exam.

9. You may use an electronic calculator on the exam.
1. [35 points] Consider the problem

\[
\min \ f(x) \\
\text{s.t.} \quad Ax \leq b
\]

where \( x \in \mathbb{R}^n, \ A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m, \) and \( f(x) \) is a twice continuously differentiable function. Suppose that \( \bar{x} \) is a feasible solution such that \( A_1 \bar{x} = b_1 \) and \( A_2 \bar{x} < b_2 \) where \( A^T = \begin{bmatrix} A_1^T & A_2^T \end{bmatrix} \) and \( b^T = \begin{bmatrix} b_1^T & b_2^T \end{bmatrix} \). Assuming \( A_1 \) has full row rank, consider the matrix

\[
P = I - A_1^T (A_1 A_1^T)^{-1} A_1.
\]

(a) [5 points] Show that \( P \) projects any vector \( u \) onto the null space of \( A_1 \). In other words, show that \( w = Pu \) is in the null space of \( A_1 \). Also, show that \( P^2 = P \) and \( P^T = P \).

(b) [10 points] Let \( \bar{d} = -P \nabla f(\bar{x}) \). Show that if \( \bar{d} \neq 0 \), it is an improving direction; that is, \( \bar{x} + \lambda \bar{d} \) is feasible in (1) and \( f(\bar{x} + \lambda \bar{d}) < f(\bar{x}) \) for \( \lambda > 0 \) and sufficiently small.

(c) [10 points] Suppose that \( \bar{d} = 0 \) and that \( \bar{u} = -(A_1 A_1^T)^{-1} A_1 \nabla f(\bar{x}) \geq 0 \), where \( \bar{u} \) is the Lagrangian vector corresponding to the constraints \( A_1 x \leq b_1 \). Show that \( \bar{x} \) and \( \bar{u} \) satisfy the KKT conditions.

(d) [10 points] Show that \( \bar{d} = -P \nabla f(\bar{x}) \) is of the form \( \lambda \bar{d} \) for some \( \lambda > 0 \), where \( \bar{d} \) is an optimal solution to the following problem

\[
\min \ \nabla f(\bar{x})^T d \\
\text{s.t.} \quad A_1 d = 0 \\
d^T d \leq 1.
\]

1(a) Consider

\[
A_1 \bar{u} = A_1 Pu = A_1 \left( I - A_1^T (A_1 A_1^T)^{-1} A_1 \right) u
= A_1 u - (A_1 A_1^T)^{-1} (A_1 A_1^T) u
= A_1 u - A_1 u = 0
\]

\( \bar{u} \) is in the null space of \( A_1 \)

\[
P^2 = \left( I - A_1^T (A_1 A_1^T)^{-1} A_1 \right) \left( I - A_1^T (A_1 A_1^T)^{-1} A_1 \right)
= I - A_1^T (A_1 A_1^T)^{-1} A_1 - A_1^T (A_1 A_1^T)^{-1} A_1 - A_1^T (A_1 A_1^T)^{-1} (A_1 A_1^T) (A_1 A_1^T)^{-1} A_1
= I - A_1^T (A_1 A_1^T)^{-1} A_1 = P
\]

Finally

\[
P^T = \left( I - A_1^T (A_1 A_1^T)^{-1} A_1 \right)^T = 2 \left( I - A_1^T (A_1 A_1^T)^{-1} A_1 \right) = P
\]

Since \( A_1 A_1^T \) is symmetric
(b) \( \bar{d} = -P v f(x) \)
\[= - (I - A^T (A A^T)^{-1} A) v f(x) \]

Now consider
\[v f(x)^T \bar{d} = -v f(x)^T P v f(x) \]
\[= -v f(x)^T P^2 v f(x) \]
\[= -v f(x)^T P^T P v f(x) \]
\[= -\|P v f(x)\|^2 = -\|\bar{d}\|^2 < 0 \]
\[y \quad \bar{d} \neq 0 \]

Also, \( A_1 \bar{d} = 0 \quad \Rightarrow \quad A_1 (x + \lambda \bar{d}) = b_1 \)

Moreover, \( A_2 \bar{x} < b_2 \quad \Rightarrow \quad A_2 (x + \lambda \bar{d}) < b_2 \)

\( x + \lambda \bar{d} \) is feasible and \( f(x + \lambda \bar{d}) < f(x) \)

for \( \lambda > 0 \) and sufficiently small.

1(c) The KKT conditions are
\[v f(x) = -A^T u - A^T \tilde{v} \]
\[
A_1 x \leq b_1 \\
A_2 x \leq b_2 \\
\tilde{v} (A_1 x - b_1) = 0 \quad \forall i \\
\tilde{v}_j (A_2 x - b_2) = 0 \quad \forall j \\
\tilde{v} f(x) = + A^T (A A^T)^{-1} A_1 v f(x) \]

or
\[(I - A^T (A A^T)^{-1} A) v f(x) = 0 \]

Moreover, if \( v \tilde{u} \neq 0 \) then

3(a) The KKT conditions for (1) are
\[v f(x) = -A^T u - 2\beta \bar{d} \]
\[A_1 \bar{d} = 0 \]
\[\bar{d}^T \bar{d} \leq 1 \]
\[\tilde{\beta} (\bar{d}^T \bar{d} - 1) = 0 \quad \text{and} \quad \tilde{\beta} > 0 \]

\[\Rightarrow \quad \bar{d} = 0 \]

It is clear that \( \bar{x}, \bar{u}, \tilde{\beta} = 0 \) satisfy the KKT conditions for (1).

We have \( \beta \neq 0 \) and so
\[\hat{d} = \frac{1}{2\beta} (-v f(x) - A^T \tilde{u}) \]

Also
\[\hat{A}_1 v f(x) = -(A A^T)^{-1} \tilde{u} \]

Also
\[\hat{d}^T \hat{d} = 1 \]

\[\hat{d} = \frac{-v f(x) - A^T \tilde{u}}{1 - v f(x) - A^T \tilde{u}} \]

\[\tilde{u} = -(A A^T)^{-1} A_1 v f(x) \]

\[\tilde{u} = \lambda \tilde{d} \quad \text{where} \quad \lambda = \|\tilde{d}\| \]
2. [30 points] Consider the convex entropy maximization problem

\[
\min \sum_{i=1}^{n} x_i \log x_i
\]
\[
\text{s.t.} \quad Ax \leq b
\]
\[
e^{T}x = 1
\]

where \( x \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, \) and \( e \in \mathbb{R}^n \) is the all-ones vector. Let \( \lambda \in \mathbb{R}^m \) and \( \nu \in \mathbb{R} \) be the Lagrangian multipliers corresponding to the first and the second set of constraints in (3).

(a) [20 points] Show that the Lagrangian dual problem can be written as

\[
\max_{\lambda, \nu} -b^T \lambda - \nu - e^{-\nu - 1} \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)
\]
\[
\text{s.t.} \quad \lambda \geq 0
\]

where \( a_i \in \mathbb{R}^m \) is the \( i \)th column of \( A \).

(b) [10 points] We can simplify the dual problem (4) by maximizing over the dual variable \( \nu \) analytically. Why is this possible? Show that this gives the dual problem

\[
\max_{\lambda} -b^T \lambda - \log \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)
\]
\[
\text{s.t.} \quad \lambda \geq 0.
\]

\[2 (a)\]

The Lagrangian function

\[ L(x, \lambda, \nu) = \sum_{i=1}^{n} x_i \log x_i - \lambda^T (b - Ax) - \nu (1 - e^T x) \]

Dual function

\[ Q(\lambda, \nu) = \min_{x} L(x, \lambda, \nu) = \min_{\lambda} \left\{ \sum_{i=1}^{n} x_i \log x_i - \lambda^T (b - Ax) - \nu (1 - e^T x) \right\} \]

The KKT conditions are

\[ (1 + \log x_i) + (a_i^T \lambda) + \nu = 0 \]
\[ \log x_i = -(\nu + 1 - a_i^T \lambda) \]
\[ x_i = e^{-(\nu + 1 - a_i^T \lambda)} \]
\[ i = 1, 2, \ldots, n \]

Since the problem is convex, the KKT conditions are necessary and sufficient conditions for optimality.

\[ Q(\lambda, \nu) = \frac{1}{e} \sum_{i=1}^{n} e^{-(\nu + 1 - a_i^T \lambda)} \left( \frac{\sum_{i=1}^{n} e^{-a_i^T \lambda}}{e} \right) + \nu \frac{1}{e} \sum_{i=1}^{n} e^{-(\nu + 1 - a_i^T \lambda)} \]
\[ = -b^T \lambda - \nu - e^{-(\nu + 1)} \left( \frac{\sum_{i=1}^{n} e^{-a_i^T \lambda}}{e} \right) \]
The dual problem is

$$\max_{\lambda \geq 0, \nu} q(\lambda, \nu) = \max_{\lambda \geq 0} \left( -b^T \lambda - \nu \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)$$

2(b) Since \( \nu \) does not appear in the constraints, we can maximize over \( \nu \) analytically to simplify the dual problem.

Differentiating \( q(\lambda, \nu) \) w.r.t. \( \nu \), we have

$$-1 + e^{-(\nu - 1)} \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right) = 0$$

$$e^{-(\nu - 1)} \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right) = 1$$

$$e^{-(\nu - 1)} = \frac{1}{\left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)}$$

$$e^{(\nu + 1)} = \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)$$

$$\nu = -1 + \log \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)$$

The dual problem is

$$\max_{\lambda \geq 0} -b^T \lambda \quad \text{subject to} \quad \nu = -1 + \log \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right)$$

$$\min_{\lambda \geq 0} -b^T \lambda - \log \left( \sum_{i=1}^{n} e^{-a_i^T \lambda} \right) - 1$$
3. [30 points] Consider the following convex quadratic program

$$\min \frac{1}{2} x^T G x + c^T x$$

$$\text{s.t.} \quad Ax \geq b$$
$$Ax = b$$ \hfill (6)

where $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $\bar{A} \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^m$, $\bar{b} \in \mathbb{R}^p$, and $G$ is a symmetric positive semidefinite matrix of size $n$.

(a) [10 points] Write the KKT conditions for (6).

(b) [10 points] We will replace (6) with the following sequence of barrier subproblems

$$\min \frac{1}{2} x^T G x + c^T x - \mu \sum_{i=1}^{m} \log y_i$$

$$\text{s.t.} \quad Ax - y = b$$
$$\bar{A}x = \bar{b}$$
$$y \succ 0$$ \hfill (7)

where $\mu > 0$ is an appropriate barrier parameter. What are the KKT conditions for (7)? Use the KKT conditions to derive an analogue of the generic primal-dual step (16.58 - see page 481 of Nocedal & Wright) for this problem.

(c) [10 points] Suggest an efficient way to solve the linear system in part (b).

Hint: See equations (16.61) and (16.62) on page 482 of Nocedal & Wright.
The KKT conditions can be written as

\[ F(x, y, \lambda, \beta, \mu) = 0 \text{ where } F(x, y, \lambda, \beta, \mu) = \begin{bmatrix} Gx + c - A^T \lambda - A^T \beta \\ Ax - y - b \\ A^T x - \bar{b} \\ \lambda \vee e - \mu e \end{bmatrix} \]

where \( e = \begin{bmatrix} \vdots \\ Y \end{bmatrix}, \quad Y = \begin{pmatrix} y_1 & \cdots & y_m \end{pmatrix} \text{ and } \lambda = \begin{pmatrix} \lambda_1 \vdots \lambda_m \end{pmatrix} \)

Applying Newton's method we have

\[
\begin{bmatrix} G \quad 0 - A^T \\ A - I \quad 0 \quad 0 \\ A \quad 0 \quad 0 \quad 0 \\ 0 \quad A \quad 0 \quad 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta \lambda \\ \Delta \beta \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_2 \\ -x_3 \\ -x_4 \end{bmatrix}
\]

where

\[ x_1 = Gx + c - A^T \lambda - A^T \beta \]
\[ x_2 = Ax - y - b \]
\[ x_3 = A^T x - \bar{b} \] and \[ x_4 = \lambda \vee e - \mu e \]

NOTE: WE WILL ASSUME THAT \( A \) and \( \bar{A} \) have full row rank and \( Y \) and \( \lambda \) are positive definite

Then find \( \Delta y \) using (3)

Then find \( \Delta \lambda \) using (2)

Then find \( \Delta x \) using (1)

Find \( \Delta \beta \)

\[
\begin{align*}
\Delta y + \lambda^{-1} \lambda \Delta \lambda &= -x_4 \\
A \Delta x + \lambda^{-1} \lambda \Delta \lambda &= -(x_2 + \lambda^{-1} x_4)
\end{align*}
\]

This gives

\[
(y^T \lambda) A \Delta x + \lambda \Delta \lambda = (y^T \lambda)(x_2 + \lambda^{-1} x_4)
\]

Then find \( \Delta x \) using (1)

\[
\begin{align*}
\Delta x &= -(G + A^T (y^T \lambda) A)^{-1} A \Delta \beta \\
&= -(G + A^T (y^T \lambda) A)^{-1} (x_1 + A^T (y^T \lambda)(x_2 + \lambda^{-1} x_4))
\end{align*}
\]

Find \( \Delta \beta \)

Solve

\[
(A (G + A^T (y^T \lambda) A)^{-1} - \lambda_2) \Delta \beta = \bar{A} (G + A^T (y^T \lambda) A)^{-1} (x_1 + A^T (y^T \lambda)(x_2 + \lambda^{-1} x_4)) - x_3 \text{ FOR } \Delta \beta
\]
4. [30 points] Consider the problem

\[
\begin{align*}
\text{max} & \quad x_1 x_2 \\
\text{s.t.} & \quad 1 - x_1^2 - x_2^2 \geq 0.
\end{align*}
\]

(a) [5 points] Is the objective function of (8) a convex function? Is the constraint region of (8) a convex set? Give reasons for your answer.

(b) [10 points] Find all the KKT points for (8). Does the LICQ hold at these KKT points?

(c) [10 points] Test the second order necessary conditions for the KKT points in part (b).

(d) [5 points] Which of the KKT points in part (b) are maxima for (8). Give reasons for your answer.

4 (a) The objective function of (8) is not convex since

\[ a = \begin{bmatrix}
0 & -1 \\
-1 & 0
\end{bmatrix} \]

which is indefinite.

The constraint set \( x_1^2 + x_2^2 \leq 1 \)

is the unit ball

that is a convex set.

The problem (8) is a non-convex problem.

\[ \begin{cases} 
\text{If } \lambda = 0, \text{ then we have} \\
\quad x_1 = x_2 = 0
\end{cases} \]

\[ \begin{cases} 
\text{If } \lambda \neq 0 \text{ then} \\
\quad x_2 = \frac{x_1}{2x_2}, \\
\quad x_2 = x_1^2, \\
\quad x_1 = \pm x_1
\end{cases} \]

4 (b) The KKT conditions are

\[
\begin{align*}
-x_2 &= \lambda \\
-x_1 &= 2x_1
\end{align*}
\]

\[
\begin{align*}
x_1^2 + x_2^2 &= 1 \\
\lambda (x_1^2 + x_2^2 - 1) &= 0 \\
\lambda &> 0
\end{align*}
\]

5 It however holds at \((\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})\) and \((-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})\)
(c) For \( x^* = (0, 0) \)

We have

\[ C(x^*, x^*) = \mathbb{R}^n \]

\[ V_{xx} L(x, \lambda) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \]

Moreover,

\[ V_{xx} L(x, \lambda) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \]

\[ (x^* = (0, 0), \lambda^* = 0) \]

which is indefinite.

\[ \therefore x^* = (0, 0) \text{ is a SADDLE POINT} \]

For \( x^* = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \) and \( x^* = \left( \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) \)

\[ V_{xx} L(x, \lambda) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \]

which is PSD

\[ (x^* = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \lambda^* = \frac{1}{2}) \]

Moreover, in true case

\[ C(x^*, x^*) = \sum d \in \mathbb{R}^2 d_1 + d_2 = 0 \]

Since \( V_{xx} L(x, \lambda) \) is PSD, it is also true that

\[ dt \cdot V_{xx} L(x, \lambda) d > 0 \quad \forall d \in C(x^*, x^*) \]

4(d): Both \( x^* = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \) and \( \left( \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) \)

are LOCAL MAXIMA.

Since they satisfy the 2nd order necessary conditions.
5. [35 points] Give an explicit closed form solution for the following problems

(a) [10 points] Consider the problem

\[
\begin{aligned}
\min & \quad c^T x \\
\text{s.t.} & \quad x^T A x \leq 1
\end{aligned}
\]

where \( c \in \mathbb{R}^n \) is nonzero vector and \( A \) is a symmetric positive definite matrix of size \( n \).

(b) [10 points] What is an optimal solution to (9) if \( A \) is not positive definite?

(c) [15 points] Consider the problem

\[
\begin{aligned}
\min & \quad x^T B x \\
\text{s.t.} & \quad x^T A x \leq 1
\end{aligned}
\]

where \( B \) is a symmetric positive semidefinite matrix of size \( n \) and \( A \) is a symmetric positive definite matrix of size \( n \).

5(a) Consider

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad x^T A x \leq 1
\end{align*}
\]

KKT conditions are

\[
\begin{align*}
C = -\lambda A x \\
\lambda (x^T A x - 1) = 0 \\
x^T A x \leq 1 \\
\lambda > 0
\end{align*}
\]

\[
\begin{align*}
\Rightarrow & \text{ Note that this problem is convex if } A \text{ is positive definite} \\
\Rightarrow & \text{ Since } c \neq 0, \lambda \neq 0 \\
\Rightarrow & \lambda x = \left( -\frac{1}{\lambda} \right) c \\
\Rightarrow & x = \left( -\frac{1}{\lambda} \right) A^{-1} c \text{ Since } A \text{ is nonsingular} \\
\Rightarrow & x^T A x = 1 \\
\Rightarrow & \frac{1}{\lambda^2} c^T A^{-1} A^{-1} c = 1 \\
\Rightarrow & \lambda = \sqrt{c^T A^{-1} c} \\
\text{ (positive square root since } \lambda > 0)
\end{align*}
\]

5(b) If \( A \) is not positive definite, then we will have multiple optimal solutions provided that \( c \in R^6(A) \) \( (R(A) \text{ is the range of the column space of } A) \).
If \( c \notin R(A) \), the problem is infeasible.

5(c) Consider

\[
\begin{align*}
\min & \quad x^T B x \\
\text{s.t.} \quad x^T A x & \leq 1
\end{align*}
\]

Since \( A \) is p.d., we can rewrite

\[ x^T A x \leq 1 \text{ as } y^T y \leq 1 \text{ where } y = A^{1/2} x \]

\[
\begin{cases}
 x = A^{-1/2} y \\
 1 - e
\end{cases}
\]

We can rewrite the original problem (10) as

\[
\begin{align*}
\min & \quad y^T (A^{-1/2} B A^{-1/2}) y \\
\text{s.t.} \quad y^T y & \leq 1
\end{align*}
\]

KKT conditions are

\[
(A^{-1/2} B A^{-1/2}) y = -\lambda y
\]

\[
\lambda (y^T y - 1) = 0
\]

and \( \lambda > 0 \)

Note that

\[
A^{-1/2} B A^{-1/2} \text{ is a symmetric matrix}
\]

If \( \lambda > 0 \), then \( y^T y = 1 \).

In this case, \( y \) is a normalized eigenvector of \( (-A^{-1/2} B A^{-1/2}) \) corresponding to the eigenvalue \( \lambda \).

Now \( (-A^{-1/2} B A^{-1/2}) \) is negative semi-definite and all its eigenvalues are non-positive.

\( \lambda = 0 \) and \( y \) is an eigenvector corresponding to a zero eigenvalue of \( (-A^{1/2} B A^{-1/2}) \) or \( (A^{1/2} B A^{-1/2}) \)

and \( X = A^{-1/2} y \)