EXAM OUTLINE

The final exam for this course takes place in the regular course classroom (MEB 238) on Monday, March 12, 8:30-10:20 am. You may bring 1 two-sided 8×11 page of notes to the final exam. This exam consists of 5 question areas each worth 60 points for a total of 300 points. The exam will be graded on a curve as will your total point score for this course.

The content of each question is as follows.

**Question 1:** In this question you will concern first- and second-order optimality conditions for unconstrained optimization problems as well as such related concepts as coercivity and convexity.

**Question 2:** The question concerns some property or derivation for quadratic functions \( q : \mathbb{R}^n \to \mathbb{R}, \) e.g.

\[
q(x) = \alpha + c^T x + \frac{1}{2} x^T Q x,
\]

\[
q(x) = \alpha + g^T (x - \bar{x}) + \frac{1}{2} (x - \bar{x})^T H (x - \bar{x}) , \quad \text{or}
\]

\[
q(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f(x) (y - x) ,
\]

where in the last possibility \( f : \mathbb{R}^n \to \mathbb{R} \) is a twice continuously differentiable (possibly non-quadratic) function. In this final case the expression given for \( q(y) \) is called the second-order Taylor expansion of \( f \) about \( x. \) If it so happens that \( f \) is quadratic, then \( q(y) = f(y) \) for all \( y \) regardless of the choice of \( x. \)

In addition, this question may also concern properties of symmetric matrices such as their eigenvalue decomposition, Cholesky factorization, and the relationships between the quadratic form determined by a symmetric matrix and its eigenvalues. It may also concern the optimization of a quadratic function, e.g. convexity, coercivity, critical points, optimal solution sets, optimal value, uniqueness of solutions, ... etc. In that last instance above, how is the associated optimization problem related to Newton’s method.

**Question 3:** This question concerns some property or derivation for linear least squares functions, e.g.

\[
\ell(x) = \frac{1}{2} \| Ax - b \|_2^2
\]

\[
\ell(x) = \frac{1}{2} \| A(x - \bar{x}) + c \|_2^2 ,
\]

\[
\ell_Q(x) = \frac{1}{2} \| Ax - b \|_Q^2 = \frac{1}{2} (Ax - b)^T (Ax - b), \quad \text{where} \quad Q \quad \text{is positive definite, or}
\]

\[
\ell(y) = \frac{1}{2} \| F(x) + F'(x)(y - x) \|_2^2 , \quad \text{where} \quad F : \mathbb{R}^n \to \mathbb{R}^m \quad \text{is smooth}.
\]

In addition, this question may also concern the optimization of such functions, e.g. convexity, coercivity, critical points, optimal solution sets, optimal value, uniqueness of solutions, ... etc. What is the relationship between quadratic functions and linear least squares functions? For example, it is easily shown that every linear least squares function can be written as a quadratic function, but under what conditions can a quadratic function be written as a linear least squares function? In the last instance above, how is the associated solution related to the Gauss-Newton method, and what is the Gauss-Newton method?
**Question 4:** This question concerns Newton, Newton-like, and Gauss-Newton methods for equation solving and optimization. In particular, how are these algorithms designed. What are appropriate line search methods? What are the details of these line search methods. In the context of Newton-Like methods, what are the motivations behind the derivations and properties of associated matrix secant methods (e.g. the Broyden and BFGS updates). I will NOT require that you memorize the various matrix secant formulas, but I may give them to you and base a question on them. Also, you need NOT memorize the Sherman-Morrison-Woodbury formula. If it is needed for any reason, I will give it to you.

**Question 5:** This question will focus on the first- and second-order optimality conditions for constrained optimization problems. Of particular importance will be Lagrangian duality and the ability to derive and apply the KKT conditions. Particular cases include linear and quadratic programming. But you should also be able to compute KKT points for concrete instances of low dimensional optimization problems.

**SAMPLE QUESTIONS**

See the sample questions for the midterm exam, but also consider the following questions. Obviously, a final exam would not contain questions of this depth and scope for each question. That is, the final will contain some easy and some hard questions. The questions listed below are all on the hard side with some very hard.

1. (a) Is the function \( f(x_1, x_2) = e^{(x_1 + x_2)^2} \) coercive? Is it convex? Justify your answers.
   (b) For what values of the parameters \( \alpha, \beta \in \mathbb{R} \) is the function
   \[
   f(x_1, x_2) = x_1^2 + 2\alpha x_1 x_2 + \beta x_2^2
   \]
a convex function. For what values is it coercive? Justify your answers.
   (c) Compute and classify all critical points of the function
   \[
   f(x_1, x_2, x_3) = 2(ax_1 - 1)^2 + (x_1 - x_2^2)^2 + (ax_2 + x_3)^2,
   \]
   where \( a > 0 \) is a given positive real number.

2. Consider the function
   \[
   f(x) = \frac{1}{2} x^T Q x + c^T x,
   \]
   where \( Q \in \mathbb{R}^{n \times n} \) is symmetric and \( c \in \mathbb{R}^m \).
   (a) Under what condition on the matrix \( Q \in \mathbb{R}^{n \times n} \) is \( f \) convex? Justify your answer.
   (b) If \( Q \) is symmetric and positive definite, show that there is a nonsingular matrix \( L \) such that \( Q = LL^T \).
   (c) With \( Q \) and \( L \) as defined in the part (b), show that
   \[
   f(x) = \frac{1}{2} \| L^T x + L^{-1} c \|_2^2 - \frac{1}{2} c^T Q^{-1} c.
   \]
   (d) If \( f \) is convex, under what conditions is \( \min_{x \in \mathbb{R}^n} f(x) = -\infty \)?
   (e) Assume that \( Q \) is symmetric and positive definite and let \( S \) be a subspace of \( \mathbb{R}^n \). Show that \( \bar{x} \) solves the problem
   \[
   \min_{x \in S} f(x)
   \]
   if and only if \( \nabla f(\bar{x}) \perp S \).
3. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ and consider the function

$$f(x) = \frac{1}{2} \|Ax - b\|_2^2.$$  

(a) Show that $\text{Ran}(A^T A) = \text{Ran}(A^T)$.  

(b) Show that for every $b \in \mathbb{R}^m$ there exists a solution to the equation $A^T Ax = A^T b$.  

(c) Show that there always exists a solution to the problem

$$\mathcal{LS} \min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2,$$

and describe the set of solutions to this problem.  

(d) Under what conditions on $A$ is the solution to the problem $\mathcal{LS}$ unique.  

(e) Using the condition from part (d) derive a formula for the unique solution to $\mathcal{LS}$.  

(f) If $A^T A$ is invertible, show that the matrix $P = A(A^T A)^{-1} A^T$ is the orthogonal projection onto $\text{Ran}(A)$, that is, show that

i. $P^2 = P$  

ii. $P^T = P$  

iii. $\text{Ran}(P) = \text{Ran}(A)$.  

4. (a) Matrix secant questions.  

i. Let $u, v \in \mathbb{R}^n$ with $u \neq 0 \neq v$. Show that the matrix $uv^T \in \mathbb{R}^{n \times n}$ is rank 1.  

ii. Let $B \in \mathbb{R}^{n \times n}$ and $s, y \in \mathbb{R}^n$. For what values of $u, v \in \mathbb{R}^n$ is it true that the matrix $B + uv^T$ satisfies $B + s = y$? Justify your answer.  

iii. Let $H \in \mathbb{R}^{n \times n}$ be symmetric and $s, y \in \mathbb{R}^n$. Under what conditions do there exist vectors $u, v \in \mathbb{R}^n$ such that the matrix $H + uv^T$ is also symmetric and satisfies $H + s = y$? Justify your answer.  

iv. Let $M \in \mathbb{R}^{n \times n}$ be symmetric and positive definite and let $s, y \in \mathbb{R}^n$ be such that $s^T y > 0$. The BFGS applied to $M$ is

$$M_+ = M + \frac{yy^T}{s^T y} - \frac{Mss^T M}{s^T Ms}.$$  

Show that $M_+$ can be rewritten in the form

$$M_+ = M + UD^{-1}U^T,$$

where $U \in \mathbb{R}^{n \times 2}$ and $D \in \mathbb{R}^{2 \times 2}$ is an invertible diagonal matrix, by deriving formulas for both $U$ and $D$.  

(b) (This problem is over the top. But parts might be trimmed down for a suitable final exam question.) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable, and let $\| \cdot \|$ be any norm on $\mathbb{R}^m$. In this problem we consider the function $f(x) = \|F(x)\|$ and properties of the the associated Gauss-Newton direction for minimizing $f$. Recall that $\bar{x} \in \mathbb{R}^n$ is a first-order stationary point for $f$ if $0 \leq f'(\bar{x}; d)$ for all $d \in \mathbb{R}^n$.  

i. Given $x, d \in \mathbb{R}^n$ and $t > 0$ show that

$$\|F(x + td)\| - \|F(x) + tF'(x)d\| \leq \|F(x + td) - (F(x) + tF'(x)d)\|.$$  

ii. Why is it true that

$$\lim_{t \to 0} \frac{\|F(x + td) - (F(x) + tF'(x)d)\|}{t} = 0 ?$$  

Hint: What is the definition of $F'(x)$?
iii. Use parts i and ii to show that
\[ f'(x; d) = \lim_{t \to 0} \frac{\|F(x) + tF'(x)d\| - \|F(x)\|}{t} . \]

iv. Use part iii and the convexity of the norm to show that
\[ f'(x; d) \leq \|F(x) + F'(x)d\| - \|F(x)\| . \]
Hint: \( F(x) + tF'(x)d = (1 - t)F(x) + t(F(x) + F'(x)d) \)
v. Use parts iii and iv to show that \( 0 \leq f'(x; d) \) for all \( d \) if and only if \( \|F(x)\| \leq \|F(x) + F'(x)d\| \) for all \( d \).
vi. If \( x \) is not a first-order stationary point for \( f \) and \( d \) solves
\[ \mathcal{GN} \min_{d \in \mathbb{R}^n} \|F(x) + F'(x)d\| , \]
show that \( d \) is a descent direction for \( f \) at \( x \) by showing that
\[ f'(x; d) \leq \|F(x) + F'(x)d\| - \|F(x)\| < 0 . \]
vii. Show that the problem
\[ \min_{d \in \mathbb{R}^n} \|F(x) + F'(x)d\|_1 \]
can be written as a linear program.
viii. Suppose at a given point \( x \in \mathbb{R}^n \) one solves the problem \( \mathcal{GN} \) in Part vi above to obtain a direction \( d \) for which \( \|F(x) + F'(x)d\| < \|F(x)\| \). Show that the following backtracking line search procedure is finitely terminating in the sense that the solution \( t \) is not zero.
Line Search: Let \( 0 < c < 1 \) and \( 0 < \gamma < 1 \) and set
\[ t := \min \gamma^s \]
s.t. \( s \in \{0, 1, 2, 3, \ldots\} \) and
\[ \|F(x + \gamma^s d)\| \leq (1 - \gamma^s c)\|F(x)\| + c\gamma^s\|F(x) + F'(x)d\|. \]
Hint: \( (1 - \gamma^s c)\|F(x)\| + c\gamma^s\|F(x) + F'(x)d\| = \|F(x)\| + c\gamma^s[\|F(x) + F'(x)d\| - \|F(x)\|]. \) Then use Part vi.

5. (a) Locate all of the KKT points for the following problem. Are these points local solutions? Are they global solutions?
\[
\begin{align*}
\text{minimize} & \quad x_1^2 + x_2^2 - 4x_1 - 4x_2 \\
\text{subject to} & \quad x_1^2 \leq x_2 \\
& \quad x_1 + x_2 \leq 2 \\
\end{align*}
\]
(b) Let \( Q \in \mathbb{R}^{m \times n} \) be symmetric and positive definite, \( g \in \mathbb{R}^n \), \( b \in \mathbb{R}^m \), and \( A \in \mathbb{R}^{m \times n} \) with \( \text{Nul}(A^T) = \{0\} \).
i. Show that the matrix \( AQ^{-1}A^T \) is nonsingular.
ii. Show that
\[ \bar{x} = Q^{-1}A^T(AQ^{-1}A^T)^{-1}b - [I - Q^{-1}A^T(AQ^{-1}A^T)^{-1}A]Q^{-1}c \]
solves the problem
\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2}x^TQx + c^Tx \\
\text{subject to} & \quad Ax = b \\
\end{align*}
\]
Hint: Use the Lagrangian \( L(x, y) = f(x) + y^T(b - Ax) \) and show that the Lagrange multiplier \( \bar{y} \) must satisfy \( \bar{x} = Q^{-1}(-c + A^T\bar{y}) \). Then multiply this expression through by \( A \) and solve for \( \bar{y} \).
(c) Let $Q \in \mathbb{R}^{n \times n}$ be symmetric and positive definite, and let $c \in \mathbb{R}^n$. Consider the optimization problem

$$\min_{0 \leq x} \frac{1}{2} x^T Q x + c^T x .$$

i. What is the Lagrangian function for this problem?

ii. Show that the Lagrangian dual is the problem

$$\max_{y \leq c} -\frac{1}{2} y^T Q^{-1} y = - \min_{y \leq c} \frac{1}{2} y^T Q^{-1} y .$$

iii. Show that if $\bar{x}, \bar{y} \in \mathbb{R}^n$ satisfy $\bar{y} = -Q \bar{x}$, then $\bar{x}$ solves the primal problem if and only if $\bar{y}$ solves the dual problem, and the optimal values in the primal and dual coincide.

(d) (A harder duality problem) Let $Q \in \mathbb{R}^{n \times n}$ be symmetric and positive definite. Consider the optimization problem

$$\begin{align*}
P & \quad \text{minimize} & & \frac{1}{2} x^T Q x + c^T x \\
& \quad \text{subject to} & & \|x\|_\infty \leq 1 .
\end{align*}$$

i. Show that this problem is equivalent to the problem

$$\begin{align*}
\hat{P} & \quad \text{minimize} & & \frac{1}{2} x^T Q x \\
& \quad \text{subject to} & & -e \leq x \leq e ,
\end{align*}$$

where $e$ is the vector of all ones.

ii. What is the Lagrangian for $\hat{P}$?

iii. Show that the Lagrangian dual for $\hat{P}$ is the problem

$$\begin{align*}
D & \quad \max -\frac{1}{2} (y - c)^T Q^{-1} (y - c) - \|y\|_1 = - \min \frac{1}{2} (y - c)^T Q^{-1} (y - c) + \|y\|_1 .
\end{align*}$$

This is also the Lagrangian dual for $cP$.

iv. Show that if $\bar{x}, \bar{y} \in \mathbb{R}^n$ satisfy $\bar{y} = Q \bar{x} + c$, then $\bar{x}$ solves $P$ if and only if $\bar{y}$ solves $D$, and the optimal values in $P$ and $D$ coincide.