This exam will consist of three parts: (I) Linear Least Squares, (II) Quadratic Optimization, and (III) Optimality Conditions and Lagrangian Duality. The first two parts ((I) Linear Least Squares and (II) Quadratic Optimization) will have 2 multipart questions, and the third part ((III) Optimality Conditions and Lagrangian Duality) will have 3 multipart questions. This give a total of 7 questions is worth 50 points for a total of 350 points. The first two parts ((I) Linear Least Squares and (II) Quadratic Optimization) are identical to the two parts of the midterm exam, however, on the final, the questions will be taken from only two of the three question types ((i) theory, (ii) linear algebra, and (iii) computations). Please use the midterm exam study guide to prepare for these questions. A more detailed description of the third part of the final exam is given below.

III Optimality Conditions and Lagrangian Duality

1 Theory Question: For this question you will need to review all of the vocabulary words as well as the theorems from the notes on *Elements of Multivariable Calculus, Optimality Conditions for Unconstrained Problems*, and *Optimality Conditions for Constrained Optimization*. You may be asked to provide statements of first- and second-order optimality conditions for both constrained and unconstrained problems. In addition, you may be asked about the role of convexity in optimization, how it is detected, as well as first- and second-order conditions under which it is satisfied.

2 Computation: In this question you will be asked to compute gradients and Hessians, located and classify stationary points for specific optimizations problems, as well as test for the convexity of a problem. You may be asked to verify whether a function or set is convex.

3 Lagrangian Duality: In this problem you will be given a primal formulation of a convex optimization problem and then asked to compute its dual.

Sample Questions

(III) Optimality Conditions and Lagrangian Duality

Question 1: Theory Question

1. State the first- and second-order conditions for optimality for the following two problems:

   (a) Linear least squares: $\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

   Solution
   Let $f(x) := \frac{1}{2}\|Ax-b\|_2^2$.
   First order: If $\bar{x}$ is a local solution, then $A^T(A\bar{x}-b) = \nabla f(\bar{x}) = 0$.
   Second order: Since $\nabla^2 f(x) = A^T A$ for all x, f is convex. Hence the first-order optimality condition is both necessary and sufficient for optimality.

   (b) Quadratic Optimization: $\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Qx + g^T x$, where $Q \in \mathbb{R}^{n\times n}$ is symmetric and $g \in \mathbb{R}^n$.

   Solution
   Let $f(x) := \frac{1}{2}x^T Qx + g^T x$.
   First order: If $\bar{x}$ is a local solution, then $Q\bar{x} + g = \nabla f(\bar{x}) = 0$.
   Second order: (Necessary) If $\bar{x}$ is a local solution, then $Q\bar{x} + g = \nabla f(\bar{x}) = 0$ and $Q = \nabla^2 f(\bar{x})$ is PSD.
   (Sufficient) Since $Q = \nabla^2 f(x)$ for all x, f s convex if and only if Q is PSD. Hence the second-order necessary conditions for optimality are also sufficient for any point $\bar{x}$ satisfying $\nabla f(\bar{x}) = 0$ to a global optimal solution to $\min_{x\in\mathbb{R}^n} f(x)$.

2. Provide necessary and sufficient conditions under which a quadratic optimization problem be written as a linear least squares problem.

Solution Consider $\min_{x\in\mathbb{R}^n} \frac{1}{2}x^TQx + g^Tx$, then the necessary and sufficient condition is Q is PSD and $g \in \mathrm{Ran}(Q)$. Indeed, if Q is PSD, then Q has a Cholesky factorization $Q = LL^T$ where $L \in \mathbb{R}^{n\times k}$ with $k = \mathrm{rank}(Q)$. Since $g \in \mathrm{Ran}(Q) = \mathrm{Ran}(L)$, there is a vector $b \in \mathbb{R}^k$ such that $-g = Lb$. Then

$$\begin{aligned}
\frac{1}{2}x^TQx + g^Tx &= \frac{1}{2}x^TLL^Tx - (Lb)^Tx \\
&= \left[\frac{1}{2}(L^Tx)^T(L^Tx) - b^T(L^Tx) + \frac{1}{2}b^Tb\right] - \frac{1}{2}b^Tb \\
&= \frac{1}{2}\|L^Tx - b\|_2^T - \frac{1}{2}b^Tb\ .
\end{aligned}$$

3. State the second-order necessary and sufficient optimality conditions for the problem $\min_{x\in\mathbb{R}^n} f(x)$, where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

Solution

Theorem 1.6 from Chapter 6.

4. State the first-order optimally conditions for the problem

$$\min_{x\in\Omega} f_0(x), \tag{$\spadesuit$}$$

where

$$\Omega := \{x : f_i(x) \le 0, i = 1, \dots, s, f_i(x) = 0, i = s+1, \dots, m\}. \tag{$\clubsuit$}$$

Solution

Theorem 1.3 from Chapter 7.

5. State the second-order sufficient conditions for optimality for the problem ($\spadesuit$) where Ω is given by ($\clubsuit$).

Solution Theorem 3.2 from Chapter 7.

6. State first- and second-order necessary and sufficient conditions for a function $f : \mathbb{R}^n \to \mathbb{R}$ to be convex.

Solution

Theorem 1.14 from Chapter 6.

7. Use a first-order necessary and sufficient condition for convexity to show that if $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function and $C \subset \mathbb{R}^n$ is a convex set, then $\bar{x}$ solves $\min_{x\in C} f(x)$ if and only if $\nabla f(\bar{x})^T(x - \bar{x}) \ge 0$ for all $x \in C$.

Solution If $\bar{x}$ solves $\min_{x\in C} f(x)$, then $f((1-t)\bar{x}+tx) \ge f(\bar{x})\ \forall\, x \in C$ and $0 \le t \le 1$ since $(1-t)\bar{x}+tx \in C$ by the convexity of C. Hence, for all $x \in C$,

$$\nabla f(\bar{x})^T(x-\bar{x}) = f'(\bar{x}; x-\bar{x}) = \lim_{t\downarrow 0} \frac{f(\bar{x}+t(x-\bar{x})) - f(\bar{x})}{t} = \lim_{t\downarrow 0} \frac{f((1-t)\bar{x}+tx) - f(\bar{x})}{t} \ge 0,$$

since $(1-t)\bar{x}+tx \in C$ for all $0 \le t \le 1$.

On the other hand, since f is convex

$$f(y) \ge f(x) + \nabla f(x)^T(y-x) \quad \forall\, y \in \mathbb{R}^n \ \text{ and } \ x \in \mathrm{dom}\,(f).$$

Therefore, if $\nabla f(\bar{x})^T(x-\bar{x}) \geq 0$ for all $x \in C$, then

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) \geq f(\bar{x}) \quad \forall\, x \in C,$$

or equivalently, $\bar{x}$ solves $\min_{x\in C} f(x)$.

Question 2: Computation

1. If f_1 and f_2 are convex functions mapping $\mathbb{R}^n$ into $\mathbb{R}$, show that $f(x) := \max\{f_1(x),\ f_2(x)\}$ is also a convex function.

   Solution

$$\begin{aligned}
f((1-\lambda)x_1+\lambda x_2) &= \max\{f_1((1-\lambda)x_1+\lambda x_2),\ f_2((1-\lambda)x_1+\lambda x_2)\} \\
&\leq \max\{(1-\lambda)f_1(x_1)+\lambda f_1(x_2),\ (1-\lambda)f_2(x_1)+\lambda f_2(x_2)\} \\
&\leq \max\{(1-\lambda)\max\{f_1(x_1),\ f_2(x_1)\}+\lambda\max\{f_1(x_2),\ f_2(x_2)\}, \\
&\qquad\qquad (1-\lambda)\max\{f_1(x_1),\ f_2(x_1)\}+\lambda\max\{f_1(x_2),\ f_2(x_2)\}\} \\
&= (1-\lambda)\max\{f_1(x_1),\ f_2(x_1)\}+\lambda\max\{f_1(x_2),\ f_2(x_2)\} \\
&= (1-\lambda)f(x_1)+\lambda f(x_2)
\end{aligned}$$

2. Let $F:\mathbb{R}^n \to \mathbb{R}^m$. Use the delta method to show that the gradient of the function $f(x) := \frac{1}{2}\|F(x)\|_2^2$ is

$$\nabla f(x) = \nabla F(x)^T F(x)\ .$$

   Solution

   Section 4 from Chapter 5.

3. A critical point of a function $f:\mathbb{R}^n \to \mathbb{R}$ is any point x at which $\nabla f(x) = 0$. Compute all of the critical points of the following functions. If no critical point exists, explain why.

   (a) $f(x) = x_1^2 - 4x_1 + 2x_2^2 + 7$
   (b) $f(x) = e^{-\|x\|^2}$
   (c) $f(x) = x_1^2 - 2x_1x_2 + \frac{1}{3}x_2^3 - 8x_2$
   (d) $f(x) = (2x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - 1)^2$

   Solution

   (a) $\nabla f(x) = [2x_1 - 4, 4x_2]^T = 0$, then $x = (2, 0)$.
   (b) $\nabla f(x) = -2e^{-\|x\|^2}x = 0$, then $x = 0$.
   (c) $\nabla f(x) = [2x_1 - 2x_2, -2x_1 + x_2^2 - 8]^T = 0$, then $x = (-2, -2)$ or $x = (4, 4)$.
   (d) $\nabla f(x) = [4(2x_1 - x_2), -2(2x_1 - x_2) + 2(x_2 - x_3), -2(x_2 - x_3) + 2(x_3 - 1)]^T = 0$, then $x = (0.5, 1, 1)$.

4. Show that the functions

$$f(x_1, x_2) = x_1^2 + x_2^3, \quad \text{and} \quad g(x_1, x_2) = x_1^2 + x_2^4$$

   both have a critical point at $(x_1, x_2) = (0, 0)$ and that their associated Hessians are positive semi-definite. Then show that $(0, 0)$ is a local (global) minimizer for g but is not a local minimizer for f.

   Solution

Both f and g are completely separable. The origin is the unique critical point for both functions. However,

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 6x_2 \end{bmatrix} \qquad \nabla^2 g(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{bmatrix}.$$

Clearly, $\nabla^2 f$ is not positive semi-definite for $x_2 < 0$, so f is not convex, while $\nabla^2 g$ is everywhere positive semi-definite and so is convex. Thus, f has no local (global) optima, while the origin is a global minimizer of g.

5. Find the local minimizers and maximizers for the following functions if they exist:
   (a) $f(x) = x^2 + \cos x$
   (b) $f(x_1, x_2) = x_1^2 - 4x_1 + 2x_2^2 + 7$
   (c) $f(x_1, x_2) = e^{-(x_1^2+x_2^2)}$
   (d) $f(x_1, x_2, x_3) = (2x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - 1)^2$

   Solution

   (a) $x = 0$ is local (global, since f is convex) minimizer;
   (b) $(x_1, x_2)^T = (2, 0)^T$ is local (global, since f is convex) minizer;
   (c) $(x_1, x_2) = (0, 0)$ is local (global, since f is convex) maximizer;
   (d) $(x_1, x_2, x_3) = (\frac{1}{2}, 1, 1)$ is local (global, since f is convex) minimizer.

6. Locate all of the KKT points for the following problems. Can you show that these points are local solutions? Global solutions?

   (a)
$$\begin{array}{ll} \text{minimize} & e^{(x_1 - x_2)} \\ \text{subject to} & e^{x_1} + e^{x_2} \leq 20 \\ & 0 \leq x_1 \end{array}$$

   Solution
   This is convex problem so any local solution is a global solution. Obviously, we wish to make x_1 as small as possible and x_2 as big as possible. Hence, we must have $x_1 = 0$ which gives the solution $(x_1, x_2) = (0, \ln(19))$. By plugging this solution into the KKT conditions, we obtain the multipliers $(y_1, y_2) = (1/19)^2(1, 20)$.

   (b)
$$\begin{array}{ll} \text{minimize} & e^{(-x_1 + x_2)} \\ \text{subject to} & e^{x_1} + e^{x_2} \leq 20 \\ & 0 \leq x_1 \end{array}$$

   Solution: Here we want to make x_1 as big as possible and x_2 as small as possible. By fixing x_1 at zero and sending x_2 to $-\infty$, the constraints are satisfied and the objective goes to zero. Hence, no solution exists and the optimal value is 0.

   (c)
$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 - 4x_1 - 4x_2 \\ \text{subject to} & x_1^2 \leq x_2 \\ & x_1 + x_2 \leq 2 \end{array}$$

   Solution: This is a convex optimization problem where the objective $x_1^2 + x_2^2 - 4x_1 - 4x_2 = (x_1 - 2)^2 + (x_2 - 2)^2 - 8$ is strictly convex. Hence any KKT point will give a unique global optimal solution. Check that $(x_1, x_2) = (1, 1)$ and $(y_1, y_2) = (0, 2)$ is a KKT pair for this problem.

(d)

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|x\|^2 \\ \text{subject to} & Ax = b \end{array}$$

where $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ satisfies $\mathrm{Nul}\,(A^T) = \{0\}$.

Solution: This is a convex problem so $\bar{x}$ is a solution if and only if there is a $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a KKT pair for this problem. The Lagrangian is $L(x, y) = \frac{1}{2}\|x\|_2^2 + y^T(b - Ax)$. The KKT conditions are $A\bar{x} = b$ and $\bar{x} = A^T\bar{y}$. Hence $b = A\bar{x} = AA^T\bar{y}$. Since $\mathrm{Nul}\,(A^T) = \{0\}$, $\mathrm{Nul}\,(AA^T) = \{0\}$ so that the matrix AA^T is invertible. Consequently, $\bar{y} = (AA^T)^{-1}b$ and $\bar{x} = A^T\bar{y} = A^T(AA^T)^{-1}b$.

Question 3: Lagrangian Duality

1. Let $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$ and compute the Lagrangian dual to the problem

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & c^T x \\ & \text{subject to} & Ax \le b,\ 0 \le x\ . \end{array}$$

Solution:

$$\begin{aligned} g(y, z) &= \min_{x\in\mathbb{R}^n} L(x, y, z) := c^T x + y^T(Ax - b) - z^T x \\ 0 &= \nabla_x L(x, y, z) = c + A^T y - z \\ g(y, z) &= -y^T b \end{aligned}$$

Then the dual problem is

$$\begin{aligned} &\max_{y,z}\ g(y, z) \\ &\text{s.t.}\ \ c + A^T y - z = 0 \\ &\qquad y \ge 0, z \ge 0, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\max_{y}\ -y^T b \\ &\text{s.t.}\ \ c + A^T y \ge 0 \\ &\qquad y \ge 0. \end{aligned}$$

2. Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$ and compute the Lagrangian dual to the problem

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ & \text{subject to} & Ax \le b,\ 0 \le x\ . \end{array}$$

Solution:

Section 5.2 from Chapter 7.

3. Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. Consider the optimization problem

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ & \text{subject to} & \|x\|_\infty \le 1\ . \end{array}$$

(a) Show that this problem is equivalent to the problem

$$\hat{\mathcal{P}} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ \text{subject to} & -e \leq x \leq e \ , \end{array}$$

where e is the vector of all ones.

Solution:

$\{x | \|x\|_\infty \leq 1\} = \{x | \ |x_i| \leq 1 \quad \forall i = 1, \cdots, n\} = \{x | -e \leq x \leq e\}$.

(b) What is the Lagrangian for $\hat{\mathcal{P}}$?

Solution:

$$L(x, u, v) = \frac{1}{2}x^T Qx + c^T x - u^T(x + e) + v^T(x - e)$$

(c) Show that the Lagrangian dual for $\hat{\mathcal{P}}$ is the problem

$$\mathcal{D} \qquad \max -\frac{1}{2}(y - c)^T Q^{-1}(y - c) - \|y\|_1 \qquad = \qquad -\min \frac{1}{2}(y - c)^T Q^{-1}(y - c) + \|y\|_1 \ .$$

This is also the Lagrangian dual for $\mathcal{P}$.

Solution:

$$\begin{aligned} g(u, v) &= \min_x L(x, u, v) = \min_x \frac{1}{2}x^T Qx + c^T x - u^T(x + e) + v^T(x - e) \\ 0 &= \nabla_x L(x, u, v) = Qx + c - u + v \\ x &= Q^{-1}(u - v - c) \\ g(u, v) &= -\frac{1}{2}(u - v - c)^T Q^{-1}(u - v - c) - (u + v)^T e \end{aligned}$$

then the Lagrangian dual of $\hat{\mathcal{P}}$ is

$$\max_{u \geq 0, v \geq 0} g(u, v) = -\frac{1}{2}(u - v - c)^T Q^{-1}(u - v - c) - (u + v)^T e.$$

Set $y = u - v$, then $-v \leq y \leq u$ and so $-v_i \leq y_i \leq u_i, \ i = 1, \ldots, n$ which implies that $\|y\|_1 \leq e^T(u + v)$. Therefore, we can eliminate u and v to obtain the dual

$$\max_y g(y) = -\frac{1}{2}(y - c)^T Q^{-1}(y - c) - \|y\|_1.$$

4. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Consider the optimization problem

$$\mathcal{P} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|_2^2 \\ \text{subject to} & \|x\|_1 \leq 1 \ . \end{array}$$

(a) Show that this problem is equivalent to the problem

$$\hat{\mathcal{P}} \qquad \begin{array}{ll} \text{minimize}_{(x,z,w)} & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} & Ax - b = w, \\ & -z \leq x \leq z \ \text{ and } \ e^T z \leq 1, \end{array}$$

where e is the vector of all ones.

Solution:

Observe that $\{x | -z \leq x \leq z \ \text{ and } \ e^T z \leq 1\} = \{x | \|x\|_1 \leq 1\}$.

(b) What is the Lagrangian for $\hat{\mathcal{P}}$?

Solution:

$$L(w,x,z,y,u,v,\lambda) = \frac{1}{2}w^T w + y^T(Ax-b-w) - u^T(x+z) + v^T(x-z) + \lambda(e^T z - 1)$$

(c) Show that the Lagrangian dual for $\hat{\mathcal{P}}$ is the problem

$$\mathcal{D} \qquad \max -\frac{1}{2}\|y\|_2^2 - y^T b - \|A^T y\|_\infty \qquad = \qquad -\min \frac{1}{2}\|y-b\|_2^2 + \|A^T y\|_\infty - \frac{1}{2}\|b\|_2^2 .$$

This is also the Lagrangian dual for $\mathcal{P}$.

Solution:

$$\begin{aligned}
L(w,x,z,y,u,v,\lambda) &= \frac{1}{2}w^T w + y^T(Ax-b-w) - u^T(x+z) + v^T(x-z) + \lambda(e^T z - 1)\\
g(y,u,v,\lambda) &= \min_{w,x,z} L(w,x,z,y,u,v,\lambda)\\
0 &= \nabla_w L = w - y\\
0 &= \nabla_x L = A^T y - u + v\\
0 &= \nabla_z L = -u - v + \lambda e,
\end{aligned}$$

hence $g(y,u,v,\lambda) = -\frac{1}{2}y^T y - y^T b - \lambda$ and the Lagrangian dual is

$$\begin{aligned}
(\tilde{\mathcal{D}}) \quad & \max \ -\frac{1}{2}y^T y - y^T b - \lambda\\
& \text{s.t. } A^T y = u - v\\
& \qquad \lambda e = u + v\\
& \qquad \lambda \geq 0, u \geq 0, v \geq 0.
\end{aligned}$$

Observe that

$$\{(y,\lambda) | A^T y = u - v, \lambda e = u + v, u \geq 0, v \geq 0, \lambda \geq 0\} = \{(y,\lambda) | A^T y + \lambda e \geq 0, -A^T y + \lambda e \geq 0, \lambda \geq 0\},$$

hence $(\tilde{\mathcal{D}})$ is equivalent to

$$\begin{aligned}
(\tilde{\mathcal{D}}) \quad & \max \ -\frac{1}{2}y^T y - y^T b - \lambda\\
& \text{s.t. } -\lambda e \leq A^T y \leq \lambda e
\end{aligned}$$

which is equivalent to $\mathcal{D}$.