Sample Questions

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

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

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

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

   **Solution**
   
   Let \( f(x) := \frac{1}{2} x^TQx + g^Tx \).
   
   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 \( x \), \( f \) is 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 be a global optimal solution to \( \min_{x \in \mathbb{R}^n} f(x) \).

2. What is an H-orthogonal vector pair, and why must the vectors in this pair be linearly independent?

   **Solution**
   
   Given a positive definite matrix \( H \), two vectors \( v \) and \( w \) are H-orthogonal if \( v^THw = 0 \). To see that nonzero H-orthogonal vectors are linearly independent, see Proposition 7.1.

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

   **Solution**
   
   Consider \( \min_{x \in \mathbb{R}^n} \frac{1}{2} x^THx + v^Tx \). Then the necessary and sufficient condition is \( H \) is PSD and \( v \in \text{Ran}(H) \).

4. 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 3.1 from Chapter 6.

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

   **Solution**
   
   Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) is smooth. Then the following are equivalent:
   
   (a) \( f \) is convex
   
   (b) \( f(y) \geq f(x) + \nabla f(x)^T(y - x) \) for all \( y, x \in \mathbb{R}^n \)
   
   (c) \( \nabla^2 f(x) \) is PSD for all \( x \in \mathbb{R}^n \)

6. What is the relation between local minimizers, global minimizers, and critical points of differentiable convex functions?

   **Solution**
   
   All these notions coincide for convex functions.
7. What is a line search methods?

Solution Section 1 of Chapter 8.


Solution A direction of descent for a function $f$ at a point $x$ is any vector $v \in \mathbb{R}^n$ satisfying $f(x + tv) < f(x)$ for some $t > 0$. If $f$ is differentiable at $x$ with $\nabla f(x) \neq 0$, then the steepest descent direction of $f$ at $x$ is the vector $v = -\nabla f(x)$. If $f$ is twice differentiable at $x$ with $\nabla f(x) \neq 0$ and $\nabla^2 f(x)$ nonsingular, then the Newton direction of $f$ at $x$ is the vector $v = -[\nabla^2 f(x)]^{-1}\nabla f(x)$.

9. State what is the backtracking line search.

Solution Section 1 of Chapter 8.

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

\[
\quad f((1 - \lambda)x_1 + \lambda x_2) = \max\{f_1((1 - \lambda)x_1 + \lambda x_2), f_2((1 - \lambda)x_1 + \lambda x_2)\} \\
\leq \max\{(1 - \lambda)f_1(x_1) + \lambda f_1(x_2), (1 - \lambda)f_2(x_1) + \lambda f_2(x_2)\} \\
\leq (1 - \lambda) \max\{f_1(x_1), f_2(x_1)\} + \lambda \max\{f_1(x_2), f_2(x_2)\}, \\
= (1 - \lambda)f(x_1) + \lambda f(x_2)
\]

2. Use the delta method to compute the gradient of the function $f(x) = x^T H x + v^T x$.

Solution

Section 3 from Chapter 5.

3. Let $Q \in \mathbb{R}^{n \times m}$ and $v \in \mathbb{R}^m$. Compute the gradient of the function

\[f(x) := ||Qx||^2 + v^T x.\]

Solution Notice that $f$ is a quadratic function $f(x) = x^T (Q^T Q)x + v^T x$. Hence $\nabla f(x) = 2Q^T Q x + v$.

4. 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) $f(x) = x_1^2 - 2x_1 x_2 + \frac{1}{3} x_2^3 - 8x_2$
(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 $x = (0.5, 1, 1)$.
5. 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**

The origin is the unique critical point for both functions, and we have
\[
\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 6x_2 \end{bmatrix}, \quad \nabla^2 g(x_1, x_2) = \begin{bmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{bmatrix}.
\]
So the Hessians are PSD at the origin. Moreover, \(\nabla^2 g(x_1, x_2)\) is always PSD and hence \(g\) is convex. It follows that its critical point (the origin) is the unique global minimizer of \(g\). On the other hand, \(0\) is not a local minimizer of the function \(t \mapsto f(0, t)\) and hence the origin is not a local minimizer of \(f\).

6. 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) minimizer;
(c) \((x_1, x_2) = (0, 0)\) is local (global, since \(f\) is convex) maximizer;
(d) \((x_1, x_2, x_3) = \left(\frac{1}{2}, 1, 1\right)\) is local (global, since \(f\) is convex) minimizer.

7. Determine which function in question (6) are convex and what that means for your answer in question (6).

**Solution** See the previous answer.

8. Compute the gradients and Hessians of the functions

(a) \(f(x) = \ln \left(e^{x_1^2} + e^{x_2^2} + e^{x_3^2}\right)\)

(b) \(f(x) = e^{x_1} \sin(x_1) \sin(x_2)\)

(c) \(f(x) = \ln \left(2 + \sin(x_1) + \sin(x_2)\right)\)

**Solution** Use the chain rule many times as in Homework set 4.

9. State which of the following are coercive.

(a) \(f(x) = x_1^2 + x_2^2\)

(b) \(f(x) = \sqrt{|x_1 \sin(x_2)|}\)

(c) \(f(x) = (x_1^2 + x_2^2) \sin(x_1 + x_2)\)

(d) \(f(x) = \frac{1}{x_1 + x_2}e^{x_1^2 + x_2^2}\)
Solution (a) is clearly coercive. (d) is coercive because $\frac{1}{t}e^t \to \infty$ as $t \to +\infty$. (c) is not coercive because $f(t, -t) = 0$ for all $t$. (b) is not coercive because $f(t, k\pi) = 0$ for each integer $k = 1, 2, 3, \ldots$.

10. Run the gradient descent line search method on the function $f(x, y) = \frac{1}{2}x^2 + y^2$ starting at the point $x_1 = (2, 2)$ and using only step sizes $t_k = \frac{1}{2}$. Output the points $x_1, x_2, x_3$.

Solution Observe $\nabla f(x, y) = [x, 2y]^T$. Hence $x_2 = x_1 - t_1 \nabla f(x_1) = (1, 0)$ and $x_3 = x_2 - t_2 \nabla f(x_2) = \left(\frac{1}{2}, 0\right)$.