MAT 3007 Optimization Homework 6
Due: 11:59 pm on July 20, 2025
Solution

1. Consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^2} f(x) = x_1^2 + \frac{1}{2}x_1^3 - x_1(2 + x_2^2) + \frac{1}{8}x_2^4.$$

(a) Compute the gradient and Hessian of f and calculate all critical points.
The gradient and Hessian of f are given by

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2x_1 + \frac{3}{2}x_1^2 - 2 - x_2^2 \\ -2x_1x_2 + \frac{1}{2}x_2^3 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 3x_1 + 2 & -2x_2 \\ -2x_2 & -2x_1 + \frac{3}{2}x_2^2 \end{pmatrix}.$$

Moreover, it holds that $\nabla f(\boldsymbol{x}) = 0$ if and only if $x_2(\frac{1}{2}x_2^2 - 2x_1) = 0$. We first consider the case $x_2 = 0$. Then, it follows $3x_1^2 + 4x_1 - 4 = 0$, i.e., $x_1 = \frac{1}{6}(-4 \pm \sqrt{16 + 16 \cdot 3}) = \frac{2}{3}$ or -2. Otherwise, we have $x_2^2 = 4x_1$ which implies $3x_1^2 - 4x_1 - 4 = 0$, i.e.,

$$x_1 = \frac{4 \pm \sqrt{16 + 16 \cdot 3}}{6} = \frac{4 \pm 8}{6} = 2 \text{ or } -\frac{2}{3}.$$

In total, f has the following four stationary points:

$$\bar{\boldsymbol{x}}_1 = [\tfrac{2}{3}\,,\, 0], \quad \bar{\boldsymbol{x}}_2 = [-2\,,\, 0], \quad \bar{\boldsymbol{x}}_3 = [2\,,\, 2\sqrt{2}], \quad \bar{\boldsymbol{x}}_4 = [2\,,\, -2\sqrt{2}].$$

(b) For each critical point x^* found in part (a), investigate whether x^* is a local maximizer, local minimizer, or saddle point and explain your answer.
We have

$$\nabla^2 f(\bar{\boldsymbol{x}}_1) = \begin{pmatrix} 4 & 0 \\ 0 & -\frac{4}{3} \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\bar{\boldsymbol{x}}_2) = \begin{pmatrix} -4 & 0 \\ 0 & 4 \end{pmatrix}.$$

Both Hessians are diagonal matrices with eigenvalues 4, $-\frac{4}{3}$ and -4, 4, respectively and hence $\nabla^2 f(\bar{\boldsymbol{x}}_1)$ and $\nabla^2 f(\bar{\boldsymbol{x}}_2)$ are indefinite and the stationary points $\bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_2$ are saddle points. Furthermore, it holds that

$$\nabla^2 f(\bar{\boldsymbol{x}}_3) = \begin{pmatrix} 8 & -4\sqrt{2} \\ -4\sqrt{2} & 8 \end{pmatrix} \quad \text{and} \quad \nabla^2 f(\bar{\boldsymbol{x}}_4) = \begin{pmatrix} 8 & 4\sqrt{2} \\ 4\sqrt{2} & 8 \end{pmatrix}$$

and $\det(\nabla^2 f(\bar{\boldsymbol{x}}_3)) = \det(\nabla^2 f(\bar{\boldsymbol{x}}_4)) = 64 - 32 > 0$. This shows that $\nabla^2 f(\bar{\boldsymbol{x}}_3)$ and $\nabla^2 f(\bar{\boldsymbol{x}}_4)$ are positive definite. Thus, by the second order sufficient conditions, $\bar{\boldsymbol{x}}_3$ and $\bar{\boldsymbol{x}}_4$ are local minimizers.

(c) Does the mapping f possess any global minimizer?
The restricted function $f(x_1, 0) = x_1^2 + \frac{1}{2}x_1^3 - 2x_1$ has "x_1^3" as leading term which diverges to $-\infty$ as $x_1 \to -\infty$. Hence, the function is not bounded from below and does not possess a global minimizer.

2. Consider the function $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$,

$$f_\alpha(x) := \alpha x_1^2 + x_2^2 - 2x_1x_2 - 2x_2,$$

where $\alpha \in \mathbb{R}$ is a scalar.

(a) Find the stationary points (in case they exist) of f_α for each value of α.
The gradient and Hessian of f_α are given by

$$\nabla f_\alpha(x) = \begin{pmatrix} 2\alpha x_1 - 2x_2 \\ 2x_2 - 2x_1 - 2 \end{pmatrix}, \quad \nabla^2 f_\alpha(x) = \begin{pmatrix} 2\alpha & -2 \\ -2 & 2 \end{pmatrix}$$

and it holds that

$$\nabla f_\alpha(x) = 0 \quad \iff \quad x_2 = x_1 + 1 \quad \text{and} \quad 2\alpha x_1 - 2x_1 = 2$$

which implies $(\alpha - 1)x_1 = 1$. This equation only has a solution if $\alpha \neq 1$. In this case, we obtain

$$x_1^* = \frac{1}{\alpha - 1}, \quad x_2^* = 1 + \frac{1}{\alpha - 1} = \frac{\alpha}{\alpha - 1}.$$

This is also the unique stationary point of f_α.

(b) For each stationary point x^* in part (a), determine whether x^* is a local maximizer or a local minimizer or a saddle point of f_α.
We have

$$\det(\nabla^2 f_\alpha(x)) = 4\alpha - 4 = 4(\alpha - 1) \quad (\forall\ x).$$

Consequently, if $\alpha < 1$, the Hessian is indefinite and x^* is a saddle point. If $\alpha > 1$, then $\nabla^2 f_\alpha(x)$ is positive definite and x^* is a local minimizer of f_α.

(c) For which values of α can f_α have a global minimizer?
The function f_α can only have a global minimizer if $\alpha > 1$. In the case $\alpha \leq 1$, f_α is unbounded and it does not possess a global minimizer (all stationary points are saddle points)

3. Consider the following inequality-constrained optimization problem:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Prove: If x^* is a local minimum of the above problem, then there exists some $y \in \mathbb{R}^m$, with $y \geq 0$, such that

$$\nabla f(x^*) = A^\top y,$$
$$y_i \cdot (a_i^\top x^* - b_i) = 0, \quad \forall i,$$

where $a_i^\top$ is the ith row of matrix A.
We consider the descent directions and the feasible directions at x^*. It is easy to see that the descent directions are:

$$S_D(x^*) = \{d : \nabla f(x^*)^\top d < 0\}$$

For the feasible directions, it is

$$S_F(x^*) = \{d : a_i^\top d \geq 0, \text{ if } a_i^\top x^* = b_i\}$$

Local optimality requires that $S_D(x^*) \cap S_F(x^*) = \emptyset$. We define

$$A(x) = \{i : a_i^\top x = b_i\}$$

to be the *active constraints* at x, then the necessary condition should be: there does not exist d such that 1)$\nabla f(x^*)^\top d < 0$ and 2) $a_i^\top d \geq 0$ for $i \in A(x^*)$. The nonexistence of d is equivalent to the existence of $y \geq 0$, such that

$$\nabla f(x) = \sum_{i\in A(x)} a_i y_i$$

This can be further written as there exists $y \geq 0$ such that

$$\nabla f(x) = A^\top y$$
$$y_i \cdot (a_i^\top x - b_i) = 0, \quad \forall i.$$

4. Answer the following questions:

   (a) If $Y \subseteq \mathbb{R}^n$ is a convex set, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$, prove $S = \{x \in \mathbb{R}^n : Ax + b \in Y\}$ is a convex set.
   Consider $x_1, x_2 \in S$ and $\lambda \in [0, 1]$, we have $Ax_1 + b \in Y$ and $Ax_2 + b \in Y$. Since Y is a convex set, we have $\lambda(Ax_1 + b) + (1 - \lambda)(Ax_2 + b) \in Y$. Since $A(\lambda x_1 + (1 - \lambda)x_2) + b = \lambda(Ax_1 + b) + (1 - \lambda)(Ax_2 + b) \in Y$, so $\lambda x_1 + (1 - \lambda)x_2 \in S$. That completes the proof.

(b) In what situation the quadratic-over-linear function, $f(x,y) = \frac{x^2}{y}$, is convex?

The Hessian matrix is $\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^\intercal$. If we want $\nabla^2 f(x,y) \succeq 0$, we need $y > 0$. So, the function is convex when $y > 0$.

(c) If function $g(x)$ (or $g_i(x)$ for $i = 1, .., m)$) is convex, Are the following functions convex?

(i) $\exp g(x)$
Convex since $\exp x$ is convex and non-decreasing and $g(x)$ is convex.

(ii) $\frac{1}{g(x)}$
Nonconvex since $\frac{1}{g(x)}$ is convex when $g(x)$ is concave and positive.

(iii) $\sum_{i=1}^m \exp g_i(x)$
Convex since $\sum_{i=1}^m \exp x_i$ is convex.

(d) Verify whether the following set is convex or not:

$$X = \{x \in \mathbb{R} : \alpha \leq \sqrt{x} \leq \beta\}, \quad \alpha \in \mathbb{R}, \quad \beta \geq 0, \quad \alpha \leq \beta.$$

The set X is convex. It can be alternatively represented as follows $X = [\max\{0, \alpha\}^2, \beta^2]$. This is just a closed convex interval.

(e) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as $f(x) = ||x||_0$, where $||x||_0$ is the number of nonzero elements in x. Is f convex?
f is not convex. Try $x = [0,1]^T$, $y = [1,0]^T$, and $\lambda = 0.5$.

(f) Prove the loss function in logistic regression $f(\mathbf{w}) = \frac{1}{m}\sum_{i=1}^m \ln\left(1 + \exp(-y_i \cdot \mathbf{w}^\top x_i)\right)$ is a convex function.
The function $f(w)$ is a nonnegative weighted sum of functions, each has the form $\log(1 + \exp(-y_i \cdot w^\top x_i))$. Consider a fixed term $\log(1 + \exp(-y_i \cdot w^\top x_i))$, this is in the format of $g(z) = \log(1+\exp(z))$ and note that $g(z)$ is composed with a linear function of w, namely $z = -y_i \cdot w^\top x_i$. Since z is linear in w, and since composition of a convex function with a linear function preserves convexity, the convexity of $f(w)$ will follow once we verify that $g(z)$ is convex for all $z \in \mathbb{R}$.
To check that $g(z)$ is convex, we compute its second derivative:

$$g''(z) = \frac{\exp(z)}{(1 + \exp(z))^2}.$$

This expression is nonnegative for all $z \in \mathbb{R}$, which implies that $g(z)$ is convex on $\mathbb{R}$. Therefore, each term $\log(1 + \exp(-y_i \cdot w^\top x_i))$ is convex in w, and hence $f(w)$, being a nonnegative average of convex functions, is also convex.

5. For each of the following statements, state whether it is true or false. If true, provide a proof, and if false provide a counter-example.

(a) The union of a finite number of convex sets is always convex.
False. Consider two convex sets: $[0,1]$ and $[2,3]$. Their union is not convex, because the midpoint between 1 and 2 is not in the union.

(b) A convex optimization problem can have at most one global optimal solution.
False. The problem $\min\{x+y: \; x+y \geq 1, \; x \geq 0, \; y \geq 0\}$ has an infinite number of global optimal solutions.

(c) A convex optimization problem must have an optimal solution.
False. The convex problem $\min\{x: \; x \in (0,1)\}$ does not have an optimal solution.

(d) A convex optimization problem that has an optimal solution can have either exactly one optimal solution or an infinite number of optimal solutions.
True. Consider the convex optimization problem $(P): \min\{f(x): x \in X\}$. Suppose (P) has two optimal solutions x', x'' with the objective value $f(x') = f(x'') = v^*$. Let $x(\lambda) = \lambda x' + (1-\lambda)x''$ for $\lambda \in [0,1]$. Note that $x(\lambda) \in X$ and $f(x(\lambda)) \leq \lambda f(x') + (1-\lambda) f(x'') = v^*$ by convexity. Thus the infinite collection of points $\{x(\lambda) : \lambda \in [0,1]\}$ must all be optimal solutions.

(e) If the feasible region of an optimization problem is non-empty, closed, and convex, suppose that the optimization problem has the property that every local optimal solution is also globally optimal then the objective function must be a convex function.
False. Consider $\min\{-x^2 : x \in [-1.1]\}$.

6. Consider the problem

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & 4x_1^2 + x_2^2 - x_1 - 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 1, \quad x_1^2 \leq 1. \end{aligned}$$

(a) Show the above problem is a convex optimization problem.
The constraint functions $g_1(x) := 2x_1 + x_2 - 1$ and $g_2(x) := x_1^2 - 1$ are obviously convex. The objective function $f(x) := 4x_1^2 + x_2^2 - x_1 - 2x_2$ satisfies

$$\nabla f(x) = \begin{pmatrix} 8x_1 - 1 \\ 2x_2 - 2 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}.$$

The Hessian of f is a diagonal matrix with eigenvalues 8 and 2 and thus, f and the optimization problem are convex.

(b) Show that Slater's condition is satisfied for the above problem.
We need to find a point $\hat{x} \in \mathbb{R}^2$ such that $g_1(\hat{x}), g_2(\hat{x}) < 0$. This holds, e.g., for $\hat{x} = (0,0)^\top$.

(c) Derive the KKT conditions for the above problem and find all KKT points.
The KKT conditions are given by

$$\begin{aligned}\nabla f(x) + \nabla g_1(x)\lambda_1 + \nabla g_2(x)\lambda_2 &= \begin{pmatrix} 8x_1 - 1 \\ 2x_2 - 2 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \lambda_1 + \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix} \lambda_2 \\ &= \begin{pmatrix} 8x_1 - 1 + 2\lambda_1 + 2x_1\lambda_2 \\ 2x_2 - 2 + \lambda_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\end{aligned}$$

$$\lambda_1, \lambda_2 \geq 0, \quad g_1(x), g_2(x) \leq 0, \quad \lambda_1 g_1(x) = 0, \quad \lambda_2 g_2(x) = 0.$$

We first consider the case $\lambda_1 = \lambda_2 = 0$. Then, we have $x_1 = \frac{1}{8}$ and $x_2 = 1$. This point is not feasible. Let us continue with the case $\lambda_1 > 0$ and $\lambda_2 = 0$. Then, it follows

$$\lambda_1 = 2 - 2x_2, \quad 8x_1 - 4x_2 + 3 = 0, \quad g_1(x) = 2x_1 + x_2 - 1 = 0.$$

This yields $2x_1 = 1 - x_2$ and $4 - 8x_2 + 3 = 0$, i.e., $x_2 = \frac{7}{8}$, $x_1 = \frac{1}{16}$, and $\lambda_1 = \frac{1}{4}$. Since this point is feasible, this is a KKT point of the problem.

We now consider $\lambda_1 = 0$ and $\lambda_2 > 0$. Then, by the complementarity conditions, we obtain $x_1 = \pm 1$ and $x_2 = 1$ (from the second main condition). Only the choice $x_1 = -1$ and $x_2 = 1$ is feasible. However, it holds that

$$-8 - 1 - 2\lambda_2 < 0 \quad \forall \lambda_2 > 0.$$

Finally, let us discuss the case $\lambda_1, \lambda_2 > 0$. Then, we have $x_1 = \pm 1$ and $x_2 = 1 - 2x_1 = -1$ or 3. In the case $x_1 = 1$ and $x_2 = -1$, we obtain $\lambda_1 = 4$ and $8 - 1 + 8 + 2\lambda_2 = 0$ if and only if $\lambda_2 < 0$. In the case $x_1 = -1$ and $x_2 = 3$, we have $\lambda_2 = -4 < 0$. Thus, this is not a KKT point.

Hence, $x^* = \left(\frac{1}{16}, \frac{7}{8}\right)^\top$ is the single KKT point of the problem.

(d) Does this problem have a unique global solution? Briefly explain your answer!
Yes, x^* is the unique global solution of this problem. First of all, since the problem is convex, the KKT point x^* is a global solution. In addition, since Slater's condition is satisfied, every global solution of the problem has to satisfy the KKT conditions.