ME598/494 Homework 2 Solution

1. (20 points) Show that the stationary point of the function

$$f = 2x_1^2 - 4x_1x_2 + 1.5x_2^2 + x_2$$

is a saddle. Find the directions of downslopes away from the saddle. Hint: Use Taylor's expansion at the saddle point. Find directions that reduce f.

Solution

(a) Find a stationary point

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 4x_1 - 4x_2 = 0 \\ \frac{\partial f}{\partial x_2} &= -4x_1 + 3x_2 + 1 = 0 \end{aligned} \tag{1}$$

Solve these to get the stationary point $x^* = [1, 1]^T$.

(b) Calculate the Hessian to get $H = [4, -4; -4, 3]$. It is indefinite since one eigenvalue is positive and the other is negative. So the stationary point is a saddle point.

(c) To find the direction of downslope, denote $\partial \mathbf{x}^* = \mathbf{x} - \mathbf{x}^*$ and $\partial f^* = f(\mathbf{x}) - f(\mathbf{x}^*)$:

$$\begin{aligned} \partial f^* &= \nabla f^* \partial \mathbf{x}^* + \frac{1}{2} \partial \mathbf{x}^{*T} \mathbf{H} \partial \mathbf{x}^* \\ &= \frac{1}{2}(2\partial x_1 - \partial x_2)(2\partial x_1 - 3\partial x_2) \end{aligned} \tag{2}$$

Set $\partial f^* < 0$ to get the downslopes $2\partial x_1 - \partial x_2 < 0$ and $2\partial x_1 - 3\partial x_2 > 0$ or $2\partial x_1 - \partial x_2 > 0$ and $2\partial x_1 - 3\partial x_2 < 0$.

2. (a) (10 points) Find the point in the plane $x_1 + 2x_2 + 3x_3 = 1$ in $\mathbb{R}^3$ that is nearest to the point $(-1, 0, 1)^T$. Is this a convex problem? Hint: Convert the problem into an unconstrained problem using $x_1 + 2x_2 + 3x_3 = 1$.

(b) (40 points) Implement the gradient descent and Newton's algorithm for solving the problem. Attach your codes in the report, along with a short summary of your findings. The summary should include: (1) The initial points tested; (2) corresponding solutions; (3) A log-linear convergence plot. Based on your results, which algorithm do you think is better? Why? Hint: A template can be found here.

Solution

(a) Solve the following problem

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & (x_1+1)^2 + x_2^2 + (x_3-1)^2 \\ \text{subject to:} \quad & x_1 + 2x_2 + 3x_3 = 1 \end{aligned} \tag{3}$$

Substituting $x_1 = 1 - 2x_2 - 3x_3$, the problem reduces to an unconstrained optimization problem. The solution is $x_1 = -15/14, x_2 = -1/7, x_3 = 11/14$. The unconstrained problem has positive definite Hessian everywhere. The problem is thus convex. You can also show that the original problem has a convex objective function and a convex feasible solution set.

(b) See code.

3. (5 points) Prove that a hyperplane is a convex set. Hint: A hyperplane in $\mathbb{R}^n$ can be expressed as: $\mathbf{a}^T\mathbf{x} = c$ for $\mathbf{x} \in \mathbb{R}^n$, where $\mathbf{a}$ is the normal direction of the hyperplane and c is some constant.

   Solution: Let $\mathcal{H} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, \mathbf{a}^T\mathbf{x} = c\}$ be the hyperplane. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be any two points in $\mathcal{H}$. Then $\mathbf{a}^T\mathbf{x}_1 = c$ and $\mathbf{a}^T\mathbf{x}_2 = c$. Then $\mathbf{a}^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) = \lambda\mathbf{a}^T\mathbf{x}_1 + (1-\lambda)\mathbf{a}^T\mathbf{x}_2 = c$. Therefore $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$ belongs to $\mathcal{H}$. So $\mathcal{H}$ is convex.

4. Let $f(x)$ and $g(x)$ be two convex functions defined on the convex set $\mathcal{X}$.

   (a) (5 points) Prove that $af(x) + bg(x)$ is convex for $a > 0$ and $b > 0$.

   (b) (5 points) In what conditions will $f(g(x))$ be convex?

   Solution

(a) For any two points x_1 and x_2, we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \tag{4}$$

and

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2). \tag{5}$$

Then

$$\begin{aligned} &af(\lambda x_1 + (1-\lambda)x_2) + bg(\lambda x_1 + (1-\lambda)x_2) \\ &\leq a(\lambda f(x_1) + (1-\lambda)f(x_2)) + b(\lambda g(x_1) + (1-\lambda)g(x_2)) \\ &= \lambda(af(x_1) + bg(x_1)) + (1-\lambda)(af(x_2) + bg(x_2)). \end{aligned} \tag{6}$$

Therefore by definition $af(x) + bg(x)$ is convex.

(b) The second-order derivative is $f''(g')^2 + g''f'$. If $f''(g')^2 + g''f' > 0$ then $f(g)$ is convex. One special case will be when f is monotonically increasing, i.e., $f' > 0$.

5. (15 points, optional for MAE494) Show that $f(\mathbf{x}_1) \geq f(\mathbf{x}_0) + \mathbf{g}_{\mathbf{x}_0}^T(\mathbf{x}_1 - \mathbf{x}_0)$ for a convex function $f(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$ and for $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathcal{X}$.

Solution: The following is a proof for a 1D case. Necessity: If f is convex, we have

$$f(x + \lambda(y-x)) \leq (1-\lambda)f(x) + (y), \tag{7}$$

for $\lambda \in [0,1]$. Divide both sides by λ to have:

$$f(y) \geq f(x) + \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}. \tag{8}$$

Take $t \rightarrow 0$ to get $f(y) \geq f(x) + f'(x)(y-x)$.

Sufficiency: Let $z = \lambda x + (1-\lambda)y$ for $\lambda \in [0,1]$. We have $f(x) \geq f(z) + f'(z)(x-z)$, $f(y) \geq f(z) + f'(z)(y-z)$. Multiplying the first inequality by λ, the second by $1-\lambda$, add the two together to have $\lambda f(x) + (1-\lambda)f(y) \geq f(z)$. Thus f is convex.

For a general case in $\mathbb{R}^n$, please see Page 70 "Proof of first-order convexity condition" in *Convex Optimization* by Stephen Boyd.