Numerical Optimization, 2021 Fall
Homework 6 Solution

1 Minimum Point

(1) Using the first-order necessary conditions, find a minimum point of the function [5pts]

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9. \tag{1}$$

(2) Verify that the point is a relative minimum point by verifying that the second-order sufficiency conditions hold. [5pts]

(3) Prove that the point is a global minimum point. [5pts]

(1) Note that x, y, z are unconstrained, so a minimum point must be an "interior" point.
First-order necessary conditions: $\nabla f(x^*, y^*, z^*) = \mathbf{0}$, that is,

$$\begin{aligned} \frac{\partial f}{\partial x} &= 4x + y - 6 = 0 \\ \frac{\partial f}{\partial y} &= x + 2y + z - 7 = 0 \\ \frac{\partial f}{\partial z} &= y + 2z - 8 = 0 \end{aligned} \tag{2}$$

$\Rightarrow x^* = \frac{6}{5}, y^* = \frac{6}{5}, z^* = \frac{17}{5}$ satisfies *F.O.N.C.*

(2) $f \in C^2$ and (x^*, y^*, z^*) is an interior point with $\nabla f(x^*, y^*, z^*) = \mathbf{0}$. Compute the Hessian:

$$F(x, y, z) = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \text{ independent of } (x, y, z). \tag{3}$$

Determinants of the principal minors of F:

$$|4| = 4, \quad \begin{vmatrix} 4 & 1 \\ 1 & 2 \end{vmatrix} = 7, \quad \begin{vmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix} = 10 \tag{4}$$

are all positive, so $F(x, y, z)$, and in particular $F(x^*, y^*, z^*)$, are positive definite.
Therefore, (x^*, y^*, z^*) is a strict relative minimum point.

(3) The Hessian $F(x, y, z)$ is positive definite throughout the entire region, so f is a convex function. Therefore, (x^*, y^*, z^*) satisfying *F.O.N.C.* is a global minimum point.

2 Cholesky Factorization

We know that an $n \times n$ symmetric matrix $\boldsymbol{A}$ is positive definite if and only if it has an $\boldsymbol{LU}$ decomposition (without interchange of rows) and the diagonal elements of $\boldsymbol{U}$ are all positive. Now, show that an $n \times n$ matrix $\boldsymbol{A}$ is symmetric and positive definite if and only if it can be written as $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^T$ where $\boldsymbol{G}$ is a lower triangular matrix with positive diagonal elements. This representation is known as the *Cholesky factorization* of $\boldsymbol{A}$. [25pts]

- The 'IF' part:
  Assume that $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{G}^T$, where $\boldsymbol{G}$ is a lower triangular matrix with positive diagonal elements. We only have to show that $\boldsymbol{A}$ has an $\boldsymbol{LU}$ decomposition. Let $\boldsymbol{g}_i = (g_{1i}, g_{2i}, \ldots, g_{ni})^T$ be the ith column vector of $\boldsymbol{G}$, we have

$$\boldsymbol{A} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_n] \begin{bmatrix} \boldsymbol{g}_1^T \\ \boldsymbol{g}_2^T \\ \cdots \\ \boldsymbol{g}_n^T \end{bmatrix} = \left[g_{11}^{-1}\boldsymbol{g}_1, g_{22}^{-1}\boldsymbol{g}_2, \cdots, g_{nn}^{-1}\boldsymbol{g}_n\right] \begin{bmatrix} g_{11}\boldsymbol{g}_1^T \\ g_{22}\boldsymbol{g}_2^T \\ \cdots \\ g_{nn}\boldsymbol{g}_n^T \end{bmatrix} \tag{5}$$

  Let

$$\boldsymbol{L} = \left[g_{11}^{-1}\boldsymbol{g}_1, g_{22}^{-1}\boldsymbol{g}_2, \cdots, g_{nn}^{-1}\boldsymbol{g}_n\right], \quad \boldsymbol{U} = \begin{bmatrix} g_{11}\boldsymbol{g}_1^T \\ g_{22}\boldsymbol{g}_2^T \\ \cdots \\ g_{nn}\boldsymbol{g}_n^T \end{bmatrix} \tag{6}$$

  Then we have $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$ and $\boldsymbol{L}$ is a lower triangular matrix with unit diagonals and $\boldsymbol{U}$ is an upper triangular matrix with positive diagonal elements ($u_{ii} = g_{ii}^2$).

- The 'ONLY IF' part:
  We can assume that $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$. Since $u_{ii} > 0$, we can define

$$\boldsymbol{S} = \begin{bmatrix} \sqrt{u_{11}} & & \\ & \ddots & \\ & & \sqrt{u_{nn}} \end{bmatrix} \tag{7}$$

  Let $\tilde{\boldsymbol{U}} = \left[u_{11}^{-1}\boldsymbol{u}_1, u_{22}^{-1}\boldsymbol{u}_2, \cdots, u_{nn}^{-1}\boldsymbol{u}_n\right]$ where $\boldsymbol{u}_i$ denotes the ith column vector of $\boldsymbol{U}$. Then we have

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{S}\boldsymbol{S}\tilde{\boldsymbol{U}} \tag{8}$$

  Since $\boldsymbol{A}$ is symmetric, we have

$$\boldsymbol{A} = \boldsymbol{A}^T = \tilde{\boldsymbol{U}}^T\boldsymbol{S}\boldsymbol{S}\boldsymbol{L}^T = \boldsymbol{L}'\boldsymbol{U}' \tag{9}$$

  where $\boldsymbol{L}' = \tilde{\boldsymbol{U}}^T$ and $\boldsymbol{U}' = \boldsymbol{S}\boldsymbol{S}\boldsymbol{L}^T$. Since the $\boldsymbol{LU}$ decomposition is unique, we know that $\boldsymbol{L}' = \boldsymbol{L}$ and $\boldsymbol{U}' = \boldsymbol{U}$. So we have

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{S}\boldsymbol{S}\tilde{\boldsymbol{U}} = \boldsymbol{L}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{L}^T = \boldsymbol{G}\boldsymbol{G}^T \tag{10}$$

  where $\boldsymbol{G} = \boldsymbol{L}\boldsymbol{S}$ is lower triangular matrix with positive diagonal elements.

3 Convex Function

Let γ be a monotone non-decreasing function of a single variable (that is, $\gamma(r) \leq \gamma(r')$ for $r' > r$) which is also convex; and let f be a convex function defined on a convex set Ω. Show that the function $\gamma(f)$ defined by $\gamma(f)(\boldsymbol{x}) =$

$\gamma[f(\boldsymbol{x})]$ is convex on Ω. [20pts]
We have a convex function $f : R^n \to R$ and a nondecreasing convex function $\gamma : R \to R$. We have to show that the function $h(\boldsymbol{x}) = \gamma[f(\boldsymbol{x})]$ is convex on R^n. Let $\boldsymbol{x}, \boldsymbol{y}$ be any two points in R^n and let $\alpha \in [0, 1]$ be arbitrary. Then

$$f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \tag{11}$$

Since γ is nondecreasing we have

$$\gamma\left[f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right)\right] \leq \gamma\left[\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})\right] \tag{12}$$

Since γ is convex on R, we have

$$\gamma\left[\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})\right] \leq \alpha\gamma\left[f(\boldsymbol{x})\right] + (1-\alpha)\gamma\left[f(\boldsymbol{y})\right] \tag{13}$$

Combining (12) and (13), we have

$$\gamma\left[f\left(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}\right)\right] \leq \gamma\left[\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})\right] \leq \alpha\gamma\left[f(\boldsymbol{x})\right] + (1-\alpha)\gamma\left[f(\boldsymbol{y})\right] \tag{14}$$

This shows that $h = \gamma(f)$ is convex.

4 Sufficient Condition

Let f be twice continuously differentiable on a region $\Omega \subset E^n$. Show that a sufficient condition for a point $\boldsymbol{x}^*$ in the interior of Ω to be a relative minimum point of f is that $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ and that f be locally convex at $\boldsymbol{x}^*$. [15pts]
f locally convex at $\boldsymbol{x}^*$ means that there is an $\epsilon > 0$ such that for all $\boldsymbol{y}$ satisfying $\|\boldsymbol{y} - \boldsymbol{x}^*\| < \epsilon$ we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)(\boldsymbol{y} - \boldsymbol{x}^*) \tag{15}$$

But if $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$, then $f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*)$ for $\|\boldsymbol{y} - \boldsymbol{x}^*\| < \epsilon$.
Therefore, $\boldsymbol{x}^*$ is a relative minimum.

5 Order of Convergence

Prove a proposition, similar to the one in textbook Section 7.8, showing that the order of convergence is insensitive to the error function. [25pts]
Proposition. Let f and g be two error functions satisfying $f(\boldsymbol{x}^*) = g(\boldsymbol{x}^*) = 0$ and for all $\boldsymbol{x}$, a relation of the form

$$0 \leq a_1 g(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq a_2 g(\boldsymbol{x}) \tag{16}$$

for some fixed $a_1 > 0, a_2 > 0$. If the sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty} = 0$ converges to $\boldsymbol{x}^*$ with order of convergence p with respect to one of these functions, it also does so with respect to the other.

Proof. The statement is easily seen to be symmetric in f and g. Thus we assume $\{\boldsymbol{x}_k\}$ is convergent with order p with respect to f and will prove that the same is true with respect to g. We have

$$\overline{\lim_{k\to\infty}} \frac{a_1 g(\boldsymbol{x}_{k+1})}{(a_2 g(\boldsymbol{x}_k))^p} \leq \overline{\lim_{k\to\infty}} \frac{f(\boldsymbol{x}_{k+1})}{f(\boldsymbol{x}_k)^p} = \beta \leq \overline{\lim_{k\to\infty}} \frac{a_2 g(\boldsymbol{x}_{k+1})}{(a_1 g(\boldsymbol{x}_k))^p} \tag{17}$$

Rearranging:

$$\frac{a_1^p}{a_2}\beta \leq \overline{\lim_{k\to\infty}} \frac{g(\boldsymbol{x}_{k+1})}{(g(\boldsymbol{x}_k))^p} \leq \frac{a_2^p}{a_1}\beta \tag{18}$$

Since p is the supremum of the nonnegative powers for which β is finite, then p is also the supremum of the numbers for which the above limit is finite. Therefore, the order of convergence of $\{\boldsymbol{x}_k\}$ with respect to g is also p.