Lecture 1: Mathematical Preliminaries and Optimality Conditions for Unconstrained Optimization

Cunlu Zhou

Center for Quantum Information and Control
University of New Mexico

CQuIC Summer Course
June 5, 2024

Mathematical Preliminaries

In this course, we will focus on the real n-dimensional vector space $\mathbb{R}^n$ and the space of real valued $m \times n$ matrices $\mathbb{R}^{m \times n}$.

Important Subsets of $\mathbb{R}^n$

- $\rightarrow$ Nonnegative orthant: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0,\ i = 1, 2, \ldots, n\}$.
- $\rightarrow$ Positive orthant: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i > 0,\ i = 1, 2, \ldots, n\}$.
- $\rightarrow$ The closed line segment between $x, y \in \mathbb{R}^n$:

$$[x, y] = \{x + \alpha(y - x) : \alpha \in [0, 1]\}.$$

- $\rightarrow$ The open line segment between $x, y \in \mathbb{R}^n$:

$$[x, y] = \{x + \alpha(y - x) : \alpha \in (0, 1)\}$$

for $x \neq y$ and $(x, x) = \emptyset$.

- $\rightarrow$ Unit simplex: $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0, e^T x = 1\}$.

Inner Products

An inner product on $\mathbb{R}^n$ is a map $\langle \cdot , \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ with the following properties:

1. (symmetry) $\langle x , y \rangle = \langle y , x \rangle, \ \forall x, y \in \mathbb{R}^n$.
2. (additivity) $\langle x , y + z \rangle = \langle x , y \rangle + \langle x , z \rangle, \ \forall x, y, z \in \mathbb{R}^n$.
3. (homogeneity) $\langle \lambda x , y \rangle = \lambda \langle x , y \rangle, \ \forall \lambda \in \mathbb{R}^n$ and $x, y \in \mathbb{R}^n$.
4. (positive definiteness) $\langle x , x \rangle \geq 0, \ \forall x \in \mathbb{R}^n$ and $\langle x , x \rangle = 0$ if and only if $x = 0$.

Example: the "dot product"

$$\langle x , y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i, \ \forall x, y \in \mathbb{R}^n.$$

Vector Norms

A norm $\|\cdot\|$ on $\mathbb{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ satisfying:

1. (Nonnegativity) $\|x\| \geq 0, \ \forall x \in \mathbb{R}^n$ and $\|x\| = 0$ if and only if $x = 0$.
2. (positive homogeneity) $\|\lambda x\| = |\lambda| \|x\|, \ \forall x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.
3. (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|, \ \forall x, y \in \mathbb{R}^n$.

→ One natural way to generate a norm on $\mathbb{R}^n$ is to take any inner product $\langle \cdot , \cdot \rangle$ defined on $\mathbb{R}^n$, and define the associated norm

$$\|x\| = \sqrt{\langle x , x \rangle}, \ \forall x \in \mathbb{R}^n.$$

→ For example, the Euclidean norm or l_2-norm:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \ \forall x \in \mathbb{R}^n.$$

l_p-norms

→ The l_p-norm ($p \geq 1$) is defined by $\|x\|_p \equiv \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$.

→ The l_∞-norm is

$$\|x\|_\infty \equiv \max_{i=1,2,\ldots,n} |x_i|.$$

→ It can be shown that

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p.$$

The Cauchy-Schwartz Inequality

For all $x, y \in \mathbb{R}^n$,

$$\left| x^T y \right| \leq \|x\| \cdot \|y\|.$$

Matrix Norms

Definition. A norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$ is a function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}_+$ satisfying

1. (nonnegativity) $\|A\| \geq 0$ for any $A \in \mathbb{R}^{m\times n}$ and $\|A\| = 0$ if and only if $A = 0$.
2. (positive homogeneity) $\|\lambda A\| = |\lambda|\|A\|$ for any $A \in \mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$.
3. (triangle inequality) $\|A + B\| \leq \|A\| + \|B\|$ for any $A, B \in \mathbb{R}^{m\times n}$.

Induced Norms

→ Given a matrix $A \in \mathbb{R}^{m \times n}$ and two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the induced matrix norm $\|A\|_{a,b}$ (called (a, b)-norm) is defined by

$$\|A\|_{a,b} = \max_x \{\|Ax\|_b : \|x\|_a \leq 1\}.$$

→ By definition, we have

$$\|Ax\|_b \leq \|A\|_{a,b}\|x\|_a.$$

→ An induced norm is a norm.

Matrix Norms Contd

→ *spectral norm*: If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$, the induced $(2,2)$-norm of a matrix $A \in \mathbb{R}^{m\times n}$ is the maximum singular value of A:

$$\|A\|_2 = \|A\|_{2,2} = \sqrt{\lambda_{\max}(A^T A)} \equiv \sigma_{\max}(A).$$

→ 1-*norm*: when $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, the induced $(1,1)$-norm of a matrix $A \in \mathbb{R}^{m\times n}$ is given by (*maximum absolute column sum*)

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} \left|A_{i,j}\right|.$$

→ ∞-*norm*: when $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, the induced (∞,∞)-norm of a matrix $A \in \mathbb{R}^{m\times n}$ is given by (*maximum absolute row sum*)

$$\|A\|_\infty = \max_{i=1,2,\ldots,n} \sum_{j=1}^{m} \left|A_{i,j}\right|.$$

The Frobenius Norm

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}, \; A \in \mathbb{R}^{m\times n},$$

which is not an induced norm (why?).

Basic Topological Concepts

→ The open ball with center $c \in \mathbb{R}^n$ and radius r:

$$B(c, r) = \{x : \|x - c\| < r\}.$$

→ The closed ball with center $c \in \mathbb{R}^n$ and radius r:

$$B[c, r] = \{x : \|x - c\| \leq r\}.$$

Definition. Given a set $U \subseteq \mathbb{R}^n$, a point $c \in U$ is called an interior point of U if there exists $r > 0$ for which $B(c, r) \subseteq U$.

→ The set of all interior points of a given set U is called the interior of the set and is denoted by $\text{int}(U)$:

$$\text{int}(U) = \{x \in U : B(x, r) \subseteq U \text{ for some } r > 0\}.$$

→ Example: $\text{int}(B[c, r]) = B(c, r)$

Open and Closed Sets

- $\rightarrow$ An open set is a set that contains only interior points, meaning that $U = \text{int}(U)$. For example, open balls and the positive orthant $\mathbb{R}^n_{++}$.
- $\rightarrow$ A union of any number of open sets is an open set and the intersection of a finite number of open sets is open.
- $\rightarrow$ A set $U \subseteq \mathbb{R}^n$ is closed if it contains all the limits of convergent sequences of vectors in U, i.e., if $\{x_i\}_{i=1}^{\infty} \subseteq U$ satisfies $x_i \rightarrow x^*$ as $i \rightarrow \infty$, then $x^* \in U$.
- $\rightarrow$ U is closed iff its complement U^c is open.
- $\rightarrow$ Examples of closed sets: the closed ball $B[c, r]$, closed line segments, the nonnegative orthant $\mathbb{R}^n_+$ and the unit simplex $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0, e^T x = 1\}$.

Boundary Points

Definition. Given a set $U \subseteq \mathbb{R}^n$, a boundary point of U is a vector $x \in \mathbb{R}^n$ satisfying the following: any neighborhood of x contains at least one point in U and at least one point in its complement U^c.

→ The set of all boundary points of a set U is denoted by $\text{bd}(U)$.

→ Examples:

$$\text{bd}(B(c,r)) = \text{bd}(B[c,r]) = \{x \in \mathbb{R}^n : \|x - c\| = r\}$$

$$\text{bd}(\mathbb{R}^n_{++}) = \text{bd}(\mathbb{R}^n_{+}) = ?$$

$$\text{bd}(\mathbb{R}^n) = ?$$

$$\text{bd}(\Delta_n) = ?$$

Closure

→ The closure of a set $U \subseteq \mathbb{R}^n$ is denoted by $\mathrm{cl}(U)$ and is defined to be the smallest closed set containing U:

$$\mathrm{cl}(U) = \bigcap \{T : U \subseteq T,\ T \text{ is closed}\}.$$

→ Equivalently,

$$\mathrm{cl}(U) = U \cup \mathrm{bd}(U).$$

Boundedness and Compactness

- → A set $U \subseteq \mathbb{R}^n$ is called bounded if there exists $M > 0$ such that $U \subseteq B(0, M)$.
- → A set $U \subseteq \mathbb{R}^n$ is called compact if it is closed and bounded.
- → Examples of compact sets: closed balls, unit simplex, closed line segments.

Directional Derivatives and Gradients

Definition. Let f be a function defined on a set $S \subseteq \mathbb{R}^n$. Let $x \in \text{int}(S)$ and let $d \in \mathbb{R}^n$. If the limit

$$\lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t}$$

exists, then it is called the directional derivative of f at x along the direction d and is denoted by $f'(x; d)$.

Directional Derivatives and Gradients

→ For any $i = 1, 2, \ldots, n$, if the limit

$$\lim_{t \to 0^+} \frac{f(x + te_i) - f(x)}{t}$$

exists, then its value is called the i-th partial derivative and is denoted by $\frac{\partial f}{\partial x_i}(x)$.

→ If all the partial derivatives of a function f exist at a point $x \in \mathbb{R}^n$, then the gradient of f at x is

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Continuous Differentiability

A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called continuously differentiable over U if all the partial derivatives exist and are continuous on U. In that case,

$$f'(x; d) = \nabla f(x)^T d, \ \forall x \in U, \ d \in \mathbb{R}^n.$$

Proposition. Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is continuously differentiable over U. Then

$$\lim_{d \to 0} \frac{f(x+d) - f(x) - \nabla f(x)^T d}{\|d\|} = 0, \ \forall x \in U.$$

Equivalently, we can write the above result as follows:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + o(\|y - x\|),$$

where $o(\cdot) : \mathbb{R}^n_+ \to \mathbb{R}$ is a 1-D function satisfying $\frac{o(t)}{t} \to 0$ as $t \to 0^+$.

Twice Differentiability

→ The partial derivatives $\frac{\partial f}{\partial x_i}$ are themselves real-valued functions that can be partially differentiated. The (i, j)-partial derivatives of f at $x \in U$ (if exists) is defined by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial \left(\frac{\partial f}{\partial x_j}\right)}{\partial x_i}(x).$$

→ A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called twice continuously differentiable over U if all the second-order partial derivatives exist and are continuous over U. In that case, for any $i \neq j$ and any $x \in U$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

The Hessian

The Hessian of f at a point $x \in U$ is the $n \times n$ matrix:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

For twice continuously differentiable functions, the Hessian is a symmetric matrix.

Linear Approximation Theorem

Theorem. Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U. Let $x \in U$ and $r > 0$ satisfy $B(x, r) \subseteq U$. Then $\forall y \in B(x, r)$ there exists $\xi \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(\xi)(y - x).$$

Quadratic Approximation Theorem

Theorem. Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U. Let $x \in U$ and $r > 0$ satisfy $B(x, r) \subseteq U$. Then $\forall y \in B(x, r)$,

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + o(\|y - x\|^2).$$

Optimality Conditions for Unconstrained Optimization

Global Minima

Definition. Let $f : S \to \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $x^* \in S$ is a global minimum point of f over S if $f(x^*) \leq f(x),\ \forall x \in S$.
2. $x^* \in S$ is a strict global minimum point of f over S if $f(x^*) < f(x),\ \forall x^* \neq x \in S$.

Definition. The minimum value of f over S is defined as

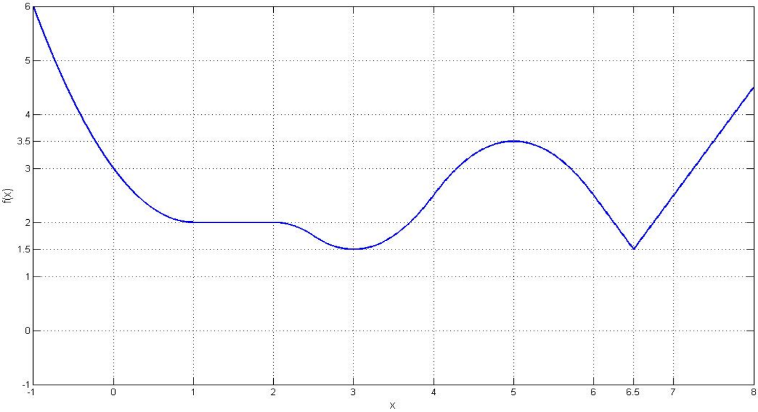
$$\inf\{f(x) : x \in S\}.$$

Local Minima

Definition. Let $f : S \to \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then

1. $x^* \in S$ is a local minimum of f over S if there exists $r > 0$ for which $f(x^*) \le f(x), \ \forall x \in S \bigcap B(x^*, r)$.
2. $x^* \in S$ is a strict local minimum of f over S if there exists $r > 0$ for which $f(x^*) < f(x), \ \forall x^* \ne x \in S \bigcap B(x^*, r)$.

Example: classify all the global and local optima



Fermat's Theorem - First-Order Optimality Condition

Theorem. Let $f : U \to \mathbb{R}$ be a function defined on a set $U \subset \mathbb{R}^n$. Suppose that $x^* \in \text{int}(U)$ is a local optimum point and that all the partial derivatives of f exist at x^*. Then $\nabla f(x^*) = 0$.

Proof. Consider the 1-D function $g(t) = f(x^* + te_i)$. □

Stationary Points

Definition. Let $f : U \longrightarrow \mathbb{R}$ be a function defined on a set $U \subset \mathbb{R}^n$. Suppose that $x^* \in \mathrm{int}(U)$ and that all the partial derivatives of f are defined at x^*. Then x^* is called a stationary point of f if $\nabla f(x^*) = 0$.

Classification of Matrices - Positive Definiteness

- → A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called positive semidefinite, denoted by $A \succeq 0$, if $x^T A x \geq 0, \ \forall x \in \mathbb{R}^n$.
- → A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called positive definite, denoted by $A \succ 0$, if $x^T A x > 0, \ \forall 0 \neq x \in \mathbb{R}^n$.
- → A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called indefinite, if there exists $x, y \in \mathbb{R}^n$ such that $x^T A x > 0, \ y^T A y < 0$.

The Principal Minors Criteria

Definition. Given an $n \times n$ matrix, the determinant of the upper left $k \times k$ submatrix is called the k-th principal minor and is denoted by $D_k(A)$. For example,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$D_1(A) = a_{11},\ D_2(A) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},\ D_3(A) = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Theorem (principal minors criteria). Let A be an $n \times n$ symmetric matrix. Then A is positive definite if and only if $D_1(A) > 0, D_2(A) > 0, \ldots, D_n(A) > 0.$

Diagonal Dominance

Definition. Let A be a symmetric $n \times n$ matrix.

(a). A is called diagonally dominant if

$$|A_{ii}| \geq \sum_{j \neq i} \left|A_{ij}\right|, \ \forall i = 1, 2, \ldots, n$$

(b). A is called strictly diagonally dominant if

$$|A_{ii}| > \sum_{j \neq i} \left|A_{ij}\right|, \ \forall i = 1, 2, \ldots, n$$

Theorem (positive (semi)definiteness of diagonally dominant matrices).

(a). If A is symmetric, diagonally dominant with nonnegative diagonal elements, then A is positive semidefinite.

(b). If A is symmetric, **strictly** diagonally dominant with **positive** diagonal elements, then A is positive definite.

Necessary Second-Order Optimality Conditions

Theorem. Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that x^* is a stationary point. Then if x^* is a local minimum point, then $\nabla^2 f(x^*) \succeq 0$.

Proof. Use the Quadratic Approximation Theorem. □

Sufficient Second-Order Optimality Conditions

Theorem. Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that x^* is a stationary point. Then if $\nabla^2 f(x^*) \succ 0$, x^* is a strict local minimum point of f over U.

Proof. Since Hessian is continuous, there exists a ball $B(x^*, r) \subseteq U$ for which $\nabla^2 f(x) \succ 0, \ \forall x \in B(x^*, r)$. By the Linear Approximation Theorem, there exists a vector $z_x \in [x^*, x]$ (and hence $z_x \in B(x^*, r)$) for which

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T \nabla^2 f(z_x)(x - x^*).$$

$\nabla^2 f(z_x) \succ 0 \Rightarrow$ for any $x \in B(x^*, r)$ such that $x \neq x^*$, the inequality $f(x) > f(x^*)$ holds, implying that x^* is a strict local minimum point of f over U. □

Saddle Points

Definition. Let $f : U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. A stationary point $x^* \in U$ is called a saddle point of f over U if it is neither a local mimimum point nor a local maximum point of f over U.

Theorem (sufficient condition for saddle points). Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that x^* is a stationary point. if $\nabla^2 f(x^*)$ is an indefinite matrix, then x^* is a saddle point of f over U.

Attainment of Minimal / Maximal Points

Theorem (Weierstrass). Let f be a continuous function defined over a nonempty compact set $C \subseteq \mathbb{R}^n$. Then there exists a global minimum point of f over C and a global maximum point of f over C.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function over $\mathbb{R}^n$. f is called coercive if

$$\lim_{\|x\| \to \infty} f(x) = \infty.$$

Theorem. Let Let $f : \mathbb{R}^n \to \mathbb{R}$ be a coercive and continuous function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f attains a global minimum point on S.

Proof. → Pick any $x_0 \in S$. f being coercive $\Rightarrow \exists M > 0$ such that

$$f(x) > f(x_0) \text{ for any } x \text{ such that } \|x\| > M.$$

→ Since any global minimizer x^* of f over S satisfies $f(x^*) \leq f(x_0)$, it follows that the set of global minimizer of f over S is the same as the set of global minimizers of f over $S \cap B[0, M]$.

→ The set $S \cap B[0, M]$ is compact and nonempty $\Rightarrow$ (by the Weierstrass theorem) $\exists$ a global minimizer of f over $S \cap B[0, M]$ and hence also over S.

□

Example

Classify the stationary points of the function $f(x_1, x_2) = -2x_1^2 + x_1 x_2^2 + 4x_1^4$.

$$\nabla f(x) = \begin{pmatrix} -4x_1 + x_2^2 + 16x_1^3 \\ 2x_1 x_2 \end{pmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{pmatrix} -4 + 48x_1^2 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}.$$

⇒ stationary points are solutions to

$$\begin{aligned} -4x_1 + x_2^2 + 16x_1^3 &= 0, \\ 2x_1 x_2 &= 0. \end{aligned}$$

⇒ stationary points are $(0, 0)$, $(0.5, 0)$, $(-0.5, 0)$.

$$\nabla^2 f(0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix}, \quad \nabla^2 f(-0.5, 0) = \begin{pmatrix} 8 & 0 \\ 0 & -1 \end{pmatrix}, \quad \nabla^2 f(0, 0) = \begin{pmatrix} -4 & 0 \\ 0 & 0 \end{pmatrix}$$

strict local minimum   saddle point   saddle point (why?)

Global Optimality Conditions

Theorem. Let f be a twice continuously defined over $\mathbb{R}^n$. Suppose that $\nabla^2 f(x) \succeq 0,\ \forall x \in \mathbb{R}^n$. Let $x^* \in \mathbb{R}^n$ be a stationary point of f. Then x^* is a global minimum point of f.

Proof. By the Linear Approximation Theorem, it follows that for any $x \in \mathbb{R}^n$, there exists a vector $z_x \in [x^*, x]$ for which

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T \nabla^2 f(z_x)(x - x^*).$$

Since $\nabla^2 f(z_x) \succeq 0$, we have that $f(x) \geq f(x^*)$, which implies that x^* is a global minimum point of f. $\square$

Quadratic Functions

A quadratic function over $\mathbb{R}^n$ is a function of the form

$$f(x) = x^T A x + 2b^T x + c,$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Its gradient and Hessian can be easily obtained (exercise):

$$\nabla f(x) = 2Ax + 2b,$$
$$\nabla^2 f(x) = 2A.$$

Lemma. Let $f(x)$ be a quadratic function. Then

1. x is a stationary point of f iff $Ax = -b$.
2. If $A \succeq 0$, then x is a global minimum point of f iff $Ax = -b$.
3. If $A \succ 0$, then $x = -A^{-1}b$ is a strict global minimum point of f.

Coerciveness of Quadratic Functions

Lemma. Let $f(x) = x^T Ax + 2b^T x + c$, where $A \in \mathbb{R}^{n\times n}$ is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then f is coercive if and only if $A \succ 0$.

Characterization of the Nonnegativity of Quadratic Functions

Theorem. Let $f(x) = x^T Ax + 2b^T x + c$, where $A \in \mathbb{R}^{n\times n}$ is symmetric, $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then the following two claims are equivalent:

(a). $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

(b). $\begin{pmatrix} A & b \\ b^T & c \end{pmatrix} \succeq 0$.

Exercises

Beck: 1.2, 1.14, 2.1, 2.2, 2.9, 2.14, 2.17.v