

MA 5037: Optimization Methods and Applications

Unconstrained Optimization



Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: May 17, 2018/Last updated: August 31, 2026

Global minimum and global maximum

Definition: Let $f : S \rightarrow \mathbb{R}$ be a real-valued function defined on a nonempty set $S \subseteq \mathbb{R}^n$.

- (1) $\mathbf{x}^* \in S$ is called a *global minimum point (minimizer)* of f over S if $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in S$.
- (2) $\mathbf{x}^* \in S$ is called a *strict global minimum point (minimizer)* of f over S if $f(\mathbf{x}^*) < f(\mathbf{x})$, $\forall \mathbf{x} \in S$ and $\mathbf{x} \neq \mathbf{x}^*$. (for short, $\forall \mathbf{x}^* \neq \mathbf{x} \in S$)
- (3) $\mathbf{x}^* \in S$ is called a *global maximum point (maximizer)* of f over S if $f(\mathbf{x}) \leq f(\mathbf{x}^*)$, $\forall \mathbf{x} \in S$.
- (4) $\mathbf{x}^* \in S$ is called a *strict global maximum point (maximizer)* of f over S if $f(\mathbf{x}) < f(\mathbf{x}^*)$, $\forall \mathbf{x}^* \neq \mathbf{x} \in S$.
- (5) *The set S on which the optimization of f is performed is called the feasible set, and any point $\mathbf{x} \in S$ is called a feasible solution.*

We will frequently omit the adjective “global”.

Minimal value and maximal value of f over S

Definition: Let $f : S \rightarrow \mathbb{R}$ be a real-valued function defined on a nonempty set $S \subseteq \mathbb{R}^n$.

- (1) $\mathbf{x}^* \in S$ is called a **global optimum** of f over S if it is either a global minimizer or a global maximizer.
- (2) **The minimal value of f over $S := \inf\{f(\mathbf{x}) : \mathbf{x} \in S\}$. If $\mathbf{x}^* \in S$ is a global minimum of f over S , then $\inf\{f(\mathbf{x}) : \mathbf{x} \in S\} = f(\mathbf{x}^*)$.**
- (3) **The maximal value of f over $S := \sup\{f(\mathbf{x}) : \mathbf{x} \in S\}$. If $\mathbf{x}^* \in S$ is a global maximum of f over S , then $\sup\{f(\mathbf{x}) : \mathbf{x} \in S\} = f(\mathbf{x}^*)$.**
- (4) The set of all global minimizers of f over S is denoted by

$$\operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

The set of all global maximizers of f over S is denoted by

$$\operatorname{argmax}\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

Example 1

Find the global minimum and maximum points of $f(x, y) = x + y$ over $S = B[\mathbf{0}, 1] = \{(x, y)^\top : x^2 + y^2 \leq 1\}$.

Solution:

- By the Cauchy-Schwarz inequality, for any $(x, y)^\top \in S$, we have

$$x + y \leq |x + y| = \left| (x, y) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right| \leq \sqrt{x^2 + y^2} \sqrt{1^2 + 1^2} \leq \sqrt{2}. \quad (\star)$$

Therefore, the maximal value of f over S is upper bounded by $\sqrt{2}$. Note that $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \in S$ and $f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = \sqrt{2}$ and this is the *only* point that attains this value. Thus, $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is the *strict* global maximum point of f over S , and the maximal value is $\sqrt{2}$.

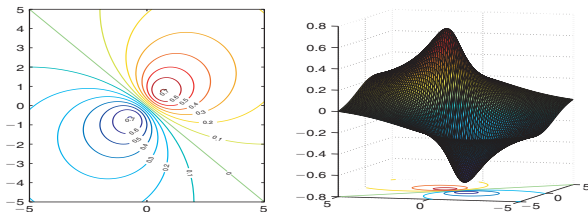
- On the other hand, by $(\star)$, we have $x + y \geq -\sqrt{2}$ and $(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ is the only point that attains this value. Therefore, $(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ is the strict global minimum point of f over S , and the minimal value is $-\sqrt{2}$.

Example 2

Consider the following 2-D function defined over the entire space $\mathbb{R}^2$:

$$f(x, y) = \frac{x + y}{x^2 + y^2 + 1}.$$

The contour and surface plots of the function are given below:



- The global maximizer = $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, the maximal value = $\frac{1}{\sqrt{2}}$.
- The global minimizer = $(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$, the minimal value = $\frac{-1}{\sqrt{2}}$.

The proof of these facts will be given later on page 26.

Local minimum and local maximum

Definition: Let $f : S \rightarrow \mathbb{R}$ be a real-valued function defined on a nonempty set $S \subseteq \mathbb{R}^n$.

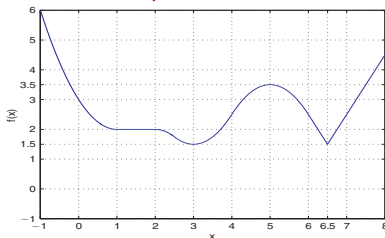
- (1) $\mathbf{x}^* \in S$ is called a **local minimum point** of f over S if $\exists r > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.
- (2) $\mathbf{x}^* \in S$ is called a **strict local minimum point** of f over S if $\exists r > 0$ such that $f(\mathbf{x}^*) < f(\mathbf{x}), \forall \mathbf{x}^* \neq \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.
- (3) $\mathbf{x}^* \in S$ is called a **local maximum point** of f over S if $\exists r > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{x}^*), \forall \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.
- (4) $\mathbf{x}^* \in S$ is called a **strict local maximum point** of f over S if $\exists r > 0$ such that $f(\mathbf{x}) < f(\mathbf{x}^*), \forall \mathbf{x}^* \neq \mathbf{x} \in S \cap B(\mathbf{x}^*, r)$.

Example

Consider the following 1-D function defined over $[-1, 8]$:

$$f(x) = \begin{cases} (x-1)^2 + 2, & -1 \leq x \leq 1, \\ 2, & 1 \leq x \leq 2, \\ -(x-2)^2 + 2, & 2 \leq x \leq 2.5, \\ (x-3)^2 + 1.5, & 2.5 \leq x \leq 4, \\ -(x-5)^2 + 3.5, & 4 \leq x \leq 6, \\ -2x + 14.5, & 6 \leq x \leq 6.5, \\ 2x - 11.5, & 6.5 \leq x \leq 8. \end{cases}$$

Classify each of the points $x = -1, 1, 2, 3, 5, 6.5, 8$ as strict/nonstrict, global/local, minimum/maximum points.



First order optimality condition for local optimum points

Theorem: Let $f : U \rightarrow \mathbb{R}$ be a function defined on a set $\emptyset \neq U \subseteq \mathbb{R}^n$. Assume that $\mathbf{x}^* \in \text{int}(U)$ is a local optimum point and that all the partial derivatives of f exist at $\mathbf{x}^*$. Then $\nabla f(\mathbf{x}^*) = \mathbf{0}$. (Fermat's theorem in 1D)

Proof: Given $1 \leq i \leq n$, we define the function $g_i(t) := f(\mathbf{x}^* + te_i)$. Then g_i is differentiable at $t = 0$ and $g'_i(0) = \nabla f(\mathbf{x}^*) \cdot \mathbf{e}_i = \frac{\partial f}{\partial x_i}(\mathbf{x}^*)$. Since $\mathbf{x}^*$ is a local optimum point of f , it follows that $t = 0$ is a local optimum of g_i . By Fermat's theorem, we have $g'_i(0) = 0$, which implies that $\nabla f(\mathbf{x}^*) = \mathbf{0}$. $\square$

Note: First order optimality condition is only *a necessary condition*.

Definition: Let $f : U \rightarrow \mathbb{R}$ be a function defined on a set $\emptyset \neq U \subseteq \mathbb{R}^n$. Assume that $\mathbf{x}^* \in \text{int}(U)$ and all the partial derivatives of f exist over some neighborhood of $\mathbf{x}^*$. If $\nabla f(\mathbf{x}^*) = \mathbf{0}$, then $\mathbf{x}^*$ is called a stationary point of f .

Recall: a point either $\nabla f(\mathbf{x}^*) = \mathbf{0}$ or undefined (i.e., does not exist) is called a *critical point*.

Positive definiteness (review)

- Let $A \in \mathbb{C}^{n \times n}$ and $x, y \in \mathbb{C}^n$. Define $x^* := \bar{x}^\top$, $(x, y) := y^*x \in \mathbb{C}$. Then $(Ax, x) = x^*Ax$ is called a *quadratic form*.

- Definition:** Let $A \in \mathbb{C}^{n \times n}$.

A is positive definite $\iff (Ax, x) > 0, \forall 0 \neq x \in \mathbb{C}^n$.

- Note 1:** $A = A^* (:= \bar{A}^\top) \iff (Ax, x) \in \mathbb{R}, \forall x \in \mathbb{C}^n$. (see page 10)
- Note 2:** If $A \in \mathbb{C}^{n \times n}$ is positive definite, then $A = A^*$. (by Note 1)
- Note 3:** $A \in \mathbb{R}^{n \times n}$ is positive definite

$$\iff A = A^\top \text{ and } (Ax, x) > 0, \forall 0 \neq x \in \mathbb{R}^n.$$

Proof: ($\Rightarrow$) Trivial!

($\Leftarrow$) Let $0 \neq x := x_1 + ix_2 \in \mathbb{C}^n$. Then $x_1 \neq 0$ or $x_2 \neq 0$.

$$\therefore (A(x_1 + ix_2), (x_1 + ix_2)) = (Ax_1, x_1) - i(Ax_1, x_2) + i(Ax_2, x_1) + (Ax_2, x_2)$$

$$\because -i(Ax_1, x_2) = -i(x_1, A^*x_2) = -i(x_1, A^\top x_2) = -i(x_1, Ax_2) = -i(Ax_2, x_1)$$

$$\therefore (A(x_1 + ix_2), (x_1 + ix_2)) = (Ax_1, x_1) + (Ax_2, x_2) > 0$$

- Note 4:** Let $A \in \mathbb{C}^{n \times n}$ and $A = A^*$. Then A is positive definite $\iff$ all of its eigenvalues are real and positive.

Proof of Note 1

$$\begin{aligned}(\Rightarrow) \because (Ax, x) &= x^*Ax = (Ax)^*x = (x, Ax) = \overline{(Ax, x)}, \forall x \in \mathbb{C}^n \\ \therefore (Ax, x) &\in \mathbb{R}, \forall x \in \mathbb{C}^n\end{aligned}$$

$$\begin{aligned}(\Leftarrow) \forall x, y \in \mathbb{C}^n, \text{ we have} \\ \mathbb{R} \ni (x+y)^*A(x+y) &= x^*Ax + y^*Ay + x^*Ay + y^*Ax. \\ \therefore x^*Ay + y^*Ax &\in \mathbb{R}\end{aligned}$$

- Let $x = e_j \in \mathbb{R}^n, y = e_k \in \mathbb{R}^n$. Then $\mathbb{R} \ni x^*Ay + y^*Ax = a_{jk} + a_{kj}$

$$\therefore \operatorname{Im}(a_{jk}) = -\operatorname{Im}(a_{kj})$$

$$\therefore a_{jk} := a + bi \text{ and } a_{kj} := c - bi \text{ for some } a, b, c \in \mathbb{R}$$

- Let $x = ie_j \in \mathbb{C}^n, y = e_k \in \mathbb{R}^n$. Then

$$\mathbb{R} \ni x^*Ay + y^*Ax = -ia_{jk} + ia_{kj} = (-ia + b) + (ci + b) = (c - a)i + 2b.$$

$$\therefore c = a. \text{ Then } a_{jk} := a + bi = \overline{a - bi} = \overline{a_{kj}}$$

$$\therefore A = \overline{A}^\top = A^*$$

Positive definiteness

- **Definition:** A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *positive semidefinite*, denoted by $A \succeq \mathbf{0}$, if $\mathbf{x}^\top A \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$.
- **Definition:** A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called *positive definite*, denoted by $A \succ \mathbf{0}$, if $\mathbf{x}^\top A \mathbf{x} > 0, \forall \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$.
- **Example:** Let $A := \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$. $\forall \mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$, we have

$$\begin{aligned} \mathbf{x}^\top A \mathbf{x} &= [x_1, x_2] \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^2 - 2x_1x_2 + x_2^2 = x_1^2 + (x_1 - x_2)^2 \geq 0. \end{aligned}$$

Since $x_1^2 + (x_1 - x_2)^2 = 0$ iff $x_1 = x_2 = 0$, we have $A \succ \mathbf{0}$.

- **Example:** Let $A := \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. One can show that A is not positive definite by considering $\mathbf{x} = (1, -1)^\top$.

The diagonal elements of a positive definite matrix

- **Note:** Let $A \in \mathbb{R}^{n \times n}$ and $A \succ 0$. Then the diagonal elements of A are positive. Proof: $A_{ii} = e_i^\top A e_i > 0, \forall i$. $\square$

Let $A \in \mathbb{R}^{n \times n}$ and $A \succeq 0$. Then the diagonal elements of A are nonnegative.

- **Definition:** $A \preceq 0$ (negative semidefinite) iff $-A \succeq 0$.

$A \prec 0$ (negative definite) iff $-A \succ 0$.

- **Note:** Let $A \in \mathbb{R}^{n \times n}$ and $A \prec 0$. Then the diagonal elements of A are negative.

Let $A \in \mathbb{R}^{n \times n}$ and $A \preceq 0$. Then the diagonal elements of A are nonpositive.

- **Definition:** A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called indefinite if $\exists x, y \in \mathbb{R}^n$ such that $x^\top A x > 0$ and $y^\top A y < 0$.

- **Note:** Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. If there exist positive and negative elements in the diagonal of A , then A is indefinite.

Proof: Let i and j be the indices such that $A_{ii} > 0$ and $A_{jj} < 0$. Then $e_i^\top A e_i = A_{ii} > 0$ and $e_j^\top A e_j = A_{jj} < 0$. $\square$

Eigenvalue characterization theorem

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then

- $A \succ 0$ if and only if all its eigenvalues are positive.

Proof: By the spectral decomposition theorem, there exist an orthogonal matrix U and a diagonal matrix $D = \text{diag}(d_1, \dots, d_n)$ whose diagonal elements are the eigenvalues of A , for which $U^T A U = D$. For any $0 \neq x \in \mathbb{R}^n$, let $y = U^{-1}x$. Then

$$x^T A x = y^T U^T A U y = y^T D y = \sum_{i=1}^n d_i y_i^2.$$

Therefore, $x^T A x > 0$ for any $x \neq 0$ if and only if $\sum_{i=1}^n d_i y_i^2 > 0$ for any $y \neq 0$.

(1) For any given i , let $y = e_i$, we have $d_i > 0$, i.e., all eigenvalues are positive.

(2) If $d_i > 0 \forall i$, then $\sum_{i=1}^n d_i y_i^2 > 0$ for any $y \neq 0$, i.e., $x^T A x > 0$ for any $x \neq 0$. $\square$

- $A \succeq 0$ if and only if all its eigenvalues are nonnegative.
- $A \prec 0$ if and only if all its eigenvalues are negative.
- $A \preceq 0$ if and only if all its eigenvalues are nonpositive.
- A is indefinite if and only if it has at least one positive eigenvalue and at least one negative eigenvalue.

Trace and determinant

- If $A \succ \mathbf{0} (\succeq \mathbf{0})$, then $\text{Tr}(A) > (\geq) 0$ and $\det(A) > (\geq) 0$.

Idea of the proof: The trace and determinant of a symmetric matrix are the sum and product of its eigenvalues respectively. $\square$

- Above two conditions are necessary and sufficient for 2×2 matrix A .

Idea of the proof: For any two real number $a, b \in \mathbb{R}$, one has $a, b > (\geq) 0$ if and only if $a + b > (\geq) 0$ and $ab > (\geq) 0$. $\square$

- **Example:** Consider the matrices

$$A := \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}, \quad B := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0.1 \end{bmatrix}.$$

(1) $A \succ \mathbf{0}$ since $\text{Tr}(A) = 7 > 0$ and $\det(A) = 11 > 0$.

(2) As for the matrix B , $\text{Tr}(B) = 2.1 > 0$ and $\det(B) = 0$. Even so, we cannot conclude that the matrix B is positive semidefinite. *In fact, B is indefinite since*

$$e_1^\top B e_1 = 1 > 0, \quad (e_2 - e_3)^\top B (e_2 - e_3) = -0.9 < 0.$$

Positive semidefinite square root

Given $A \succeq \mathbf{0}$, let $A = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ be the spectral decomposition, where $\mathbf{U}$ is an orthogonal matrix, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_n)$ is a diagonal matrix whose diagonal elements are the eigenvalues of A . Since $A \succeq \mathbf{0}$, we have $d_1, d_2, \dots, d_n \geq 0$. We define

$$A^{\frac{1}{2}} := \mathbf{U}\mathbf{E}\mathbf{U}^\top, \quad \mathbf{E} = \text{diag}(\sqrt{d_1}, \sqrt{d_2}, \dots, \sqrt{d_n}).$$

Obviously,

$$A^{\frac{1}{2}}A^{\frac{1}{2}} = \mathbf{U}\mathbf{E}\mathbf{U}^\top\mathbf{U}\mathbf{E}\mathbf{U}^\top = \mathbf{U}\mathbf{E}\mathbf{E}\mathbf{U}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top = A.$$

The matrix $A^{\frac{1}{2}}$ is called the positive semidefinite square root.

Principal minors criterion

- **Definition:** Given an $n \times n$ matrix, the determinant of the upper left $k \times k$ submatrix is called the k th (leading) principal minor, denoted by $D_k(A)$.
- **Example:** The leading principal minors of the 3×3 matrix

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

are $D_1(A) = a_{11}$,

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

- **Leading principal minors criterion:** Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then $A \succ \mathbf{0}$ iff $D_1(A) > 0, D_2(A) > 0, \dots, D_n(A) > 0$.

Note: It cannot be used for detecting positive semidefiniteness (unless all principal minors are nonnegative). e.g., consider $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$.

Diagonally dominant matrices

- **Definition:** Let $A \in \mathbb{R}^{n \times n}$. Then

- (1) A is called diagonally dominant if $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|, \forall i$.
- (2) A is called strictly diagonally dominant if $|A_{ii}| > \sum_{j \neq i} |A_{ij}|, \forall i$.

- **Theorem:** Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix.

- (1) If A is a diagonally dominant matrix whose diagonal elements are nonnegative. Then $A \succeq \mathbf{0}$.
- (2) If A is a strictly diagonally dominant matrix whose diagonal elements are positive. Then $A \succ \mathbf{0}$.

Proof: (1). Suppose $\exists \lambda < 0$ an eigenvalue of A . Let $\mathbf{u} = (u_1, \dots, u_n)^\top$ be a corresponding eigenvector. Let $|u_i| = \max\{|u_1|, \dots, |u_n|\}$. Then by $A\mathbf{u} = \lambda\mathbf{u}$,

$$|A_{ii} - \lambda||u_i| = \left| \sum_{j \neq i} A_{ij}u_j \right| \leq \left(\sum_{j \neq i} |A_{ij}| \right) |u_i| \leq |A_{ii}| |u_i|,$$

implying $|A_{ii} - \lambda| \leq |A_{ii}|$. This is a contradiction.

(2). From (1), we know that $A \succeq \mathbf{0}$. Thus, all we need to show is that A has no zero eigenvalues. Suppose $\exists$ eigenvalue $\lambda = 0, \mathbf{u} \neq \mathbf{0}$ such that $A\mathbf{u} = \mathbf{0}$. Similar to part (1), we obtain

$$|A_{ii}||u_i| = \left| \sum_{j \neq i} A_{ij}u_j \right| \leq \left(\sum_{j \neq i} |A_{ij}| \right) |u_i| < |A_{ii}| |u_i|.$$

This is obviously a contradiction. $\square$

Necessary second order optimality condition

Theorem: Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Assume that f is twice continuously differentiable over U , i.e., $f \in C^2(U)$, and that $\mathbf{x}^*$ is a stationary point. Then

- (1) If $\mathbf{x}^*$ is a local minimum point of f over U , then $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$.
- (2) If $\mathbf{x}^*$ is a local maximum point of f over U , then $\nabla^2 f(\mathbf{x}^*) \preceq \mathbf{0}$.

Proof: (1). Since $\mathbf{x}^*$ is a local minimum point, $\exists B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}^*) \leq f(\mathbf{x})$, $\forall \mathbf{x} \in B(\mathbf{x}^*, r)$. Let $\mathbf{0} \neq \mathbf{d} \in \mathbb{R}^n$. For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* := \mathbf{x}^* + \alpha \mathbf{d} \in B(\mathbf{x}^*, r)$ and $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$. By the linear approximation theorem, $\exists \mathbf{z}_\alpha \in (\mathbf{x}^*, \mathbf{x}_\alpha^*)$ such that

$$f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) = \underbrace{\nabla f(\mathbf{x}^*)^\top}_{\mathbf{0}} (\mathbf{x}_\alpha^* - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x}_\alpha^* - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{z}_\alpha) (\mathbf{x}_\alpha^* - \mathbf{x}^*) = \frac{\alpha^2}{2} \mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d}.$$

Thus, $\mathbf{d}^\top \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d} \geq 0$, $\forall \alpha \in (0, \frac{r}{\|\mathbf{d}\|})$. Using the fact that $\mathbf{z}_\alpha \rightarrow \mathbf{x}^*$ as $\alpha \rightarrow 0^+$, and the continuity of the Hessian, we obtain $\mathbf{d}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$. We conclude that $\nabla^2 f(\mathbf{x}^*) \succeq \mathbf{0}$.

(2). Employing the result of part (1) on the function $-f$, we obtain (2). $\square$

Sufficient second order optimality condition

Theorem: Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Assume that $f \in C^2(U)$ and that $\mathbf{x}^*$ is a stationary point. Then

- (1) If $\nabla^2 f(\mathbf{x}^*) \succ \mathbf{0}$, then $\mathbf{x}^*$ is a strict local minimum point of f over U .
- (2) If $\nabla^2 f(\mathbf{x}^*) \prec \mathbf{0}$, then $\mathbf{x}^*$ is a strict local maximum point of f over U .

Proof: (1) Since the Hessian is continuous, it follows that there exists a ball $B(\mathbf{x}^*, r) \subseteq U$ s.t. $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}, \forall \mathbf{x} \in B(\mathbf{x}^*, r)$ (using the leading principal minors criterion on page 16). By the linear approximation theorem, it follows that for any $\mathbf{x} \in B(\mathbf{x}^*, r)$, $\exists \mathbf{z}_x \in (\mathbf{x}^*, \mathbf{x}) \subset B(\mathbf{x}^*, r)$ such that

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{z}_x)(\mathbf{x} - \mathbf{x}^*).$$

Since $\nabla^2 f(\mathbf{z}_x) \succ \mathbf{0}$, it follows that

$$f(\mathbf{x}) - f(\mathbf{x}^*) > 0, \quad \text{for } \mathbf{x} \neq \mathbf{x}^*.$$

That is, $\mathbf{x}^*$ is a strict local minimum point of f over U .

(2) This part follows from part (1) by considering the function $-f$. $\square$

Sufficient condition for a saddle point

- **Definition:** Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Assume that $f \in C^1(U)$. A stationary point $\mathbf{x}^*$ is called a saddle point of f over U if it is neither a local minimum point nor a local maximum point of f over U .
- **Sufficient condition for a saddle point:** Let $f : U \rightarrow \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Assume that $f \in C^2(U)$ and that $\mathbf{x}^*$ is a stationary point. If $\nabla^2 f(\mathbf{x}^*)$ is an indefinite matrix, then $\mathbf{x}^*$ is a saddle point of f over U .

Proof: Let $\lambda > 0$ be an eigenvalue of $\nabla^2 f(\mathbf{x}^*)$ with a normalized eigenvector $\mathbf{v}$. Since U is open, $\exists r > 0$ such that $\mathbf{x}^* + \alpha \mathbf{v} \in U, \forall \alpha \in (0, r)$. By the quadratic approximation theorem and $\nabla f(\mathbf{x}^*) = \mathbf{0}$, we have

$$\begin{aligned} f(\mathbf{x}^* + \alpha \mathbf{v}) &= f(\mathbf{x}^*) + \frac{\alpha^2}{2} \mathbf{v}^\top \nabla^2 f(\mathbf{x}^*) \mathbf{v} + o(\alpha^2 \|\mathbf{v}\|^2) \\ &= f(\mathbf{x}^*) + \frac{\lambda \alpha^2}{2} \|\mathbf{v}\|^2 + o(\alpha^2 \|\mathbf{v}\|^2) = f(\mathbf{x}^*) + \frac{\lambda \alpha^2}{2} + o(\alpha^2). \end{aligned}$$

Since $\frac{o(\alpha^2)}{\alpha^2} \rightarrow 0$ as $\alpha \rightarrow 0^+$, $\exists \varepsilon_1 \in (0, r)$ such that $o(\alpha^2) > -\frac{\lambda}{2} \alpha^2, \forall \alpha \in (0, \varepsilon_1)$.

Hence, $f(\mathbf{x}^* + \alpha \mathbf{v}) > f(\mathbf{x}^*)$. This shows that $\mathbf{x}^*$ cannot be a local maximum point of f over U . Similarly, by considering a negative eigenvalue, $\mathbf{x}^*$ cannot be a local minimum point of f over U . Therefore, $\mathbf{x}^*$ is a saddle point of f over U . $\square$

Weierstrass theorem

- **Weierstrass Theorem:** *Let $f : \emptyset \neq C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function and C is a compact set. Then there exist a global minimum point of f over C and a global maximum point of f over C .*
- **Definition:** *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous function defined over $\mathbb{R}^n$. The function f is called coercive if $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$.*
- **Attainment under coerciveness:** *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f has a global minimum point over S .*

Proof:

- (1) Let $x_0 \in S$. Since f is coercive, $\exists M > 0$ such that $f(x) > f(x_0)$, $\forall x \in \mathbb{R}^n$ and $\|x\| > M$.
- (2) Since any global minimizer x^* of f over S satisfies $f(x^*) \leq f(x_0)$, it follows that the set of global minimizers of f over S is the same as the set of global minimizers of f over $S \cap B[0, M]$.
- (3) The set $S \cap B[0, M]$ is compact and nonempty, by the Weierstrass theorem, there exists a global minimizer of f over $S \cap B[0, M]$ and hence also over S .

□

Example 1

Consider $f(x_1, x_2) = x_1^2 + x_2^2$ over the set $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \leq -1\}$.

- Since C is not bounded, the Weierstrass theorem does not guarantee the existence of global minimizer and maximizer of f over C . Obviously, f has no global maximizer over C .
- f is coercive and C is closed, f has a global minimizer over C .
- In the interior of C : $\nabla f(x_1, x_2) = \mathbf{0} \Rightarrow (x_1, x_2) = (0, 0) \notin C$.

At the boundary of C : $\{(x_1, x_2) : x_1 + x_2 = -1\} \Rightarrow x_1 = -x_2 - 1$.

$$g(x_2) := f(-x_2 - 1, x_2) = (-x_2 - 1)^2 + x_2^2$$

$$g'(x_2) = 2(1 + x_2) + 2x_2 \Rightarrow g'(x_2) = 0 \Rightarrow x_2 = -\frac{1}{2} \Rightarrow x_1 = -\frac{1}{2}.$$

Thus, $(x_1, x_2) = (-\frac{1}{2}, -\frac{1}{2})$ is the only candidate for a global minimum point.

Therefore, $(x_1, x_2) = (-\frac{1}{2}, -\frac{1}{2})$ is the global minimum point of f over C .

Example 2

Consider the function $f(x_1, x_2) = 2x_1^3 + 3x_2^2 + 3x_1^2x_2 - 24x_2$ over $\mathbb{R}^2$.

- $\nabla f(x_1, x_2) = \begin{bmatrix} 6x_1^2 + 6x_1x_2 \\ 6x_2 + 3x_1^2 - 24 \end{bmatrix} := \mathbf{0}$. Then the stationary points of the function f are $(x_1, x_2) = (0, 4), (4, -4), (-2, 2)$.
- The Hessian of f is given by $\nabla^2 f(x_1, x_2) = 6 \begin{bmatrix} 2x_1 + x_2 & x_1 \\ x_1 & 1 \end{bmatrix}$.
- $\nabla^2 f(0, 4) = 6 \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \succ \mathbf{0} \Rightarrow (0, 4) \text{ is a strict local minimum point. It is not a global minimum point, since } f(x_1, 0) = 2x_1^3 \rightarrow -\infty \text{ as } x_1 \rightarrow -\infty.$
 $\nabla^2 f(4, -4) = 6 \begin{bmatrix} 4 & 4 \\ 4 & 1 \end{bmatrix}$, $\text{tr}(A) > 0$ but $\det(A) < 0$, an indefinite matrix.
 $\therefore (4, -4) \text{ is a saddle point}$
 $\nabla^2 f(-2, 2) = 6 \begin{bmatrix} -2 & -2 \\ -2 & 1 \end{bmatrix}$ is indefinite, since it has both positive and negative elements on its diagonal (cf. page 10).
 $\therefore (-2, 2) \text{ is a saddle point}$

Example 3

Consider the function $f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$ over $\mathbb{R}^2$.

- $\nabla f(x_1, x_2) = 4 \begin{bmatrix} (x_1^2 + x_2^2 - 1)x_1 \\ (x_1^2 + x_2^2 - 1)x_2 + (x_2^2 - 1)x_2 \end{bmatrix} := \mathbf{0}.$

Then the stationary points are $(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1)$.

- The Hessian of the function is

$$\nabla^2 f(x_1, x_2) = 4 \begin{bmatrix} 3x_1^2 + x_2^2 - 1 & 2x_1x_2 \\ 2x_1x_2 & x_1^2 + 6x_2^2 - 2 \end{bmatrix}.$$

- $\nabla^2 f(0, 0) = 4 \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \prec \mathbf{0}.$

$\therefore (0, 0)$ is a strict local maximum point

(not global, $\because f(x_1, 0) = (x_1^2 - 1)^2 + 1 \rightarrow \infty$)

$$\nabla^2 f(1, 0) = \nabla^2 f(-1, 0) = 4 \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}, \text{ indefinite matrix.}$$

$\therefore (1, 0), (-1, 0)$ saddle points

$$\nabla^2 f(0, 1) = \nabla^2 f(0, -1) = 4 \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \succeq \mathbf{0}, \text{ no conclusion!}$$

$\because f(0, 1) = f(0, -1) = 0$ and f is bounded below by 0

$\therefore (0, 1), (0, -1)$ are global minimum points

Contour and surface plots of Example 3

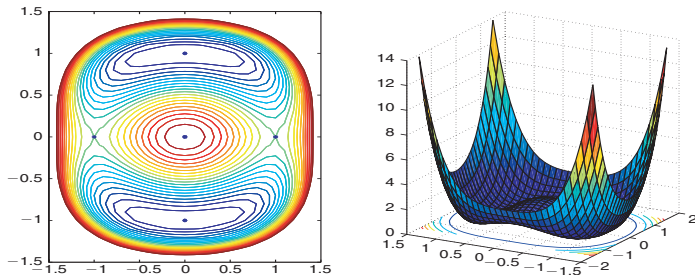


Figure 2.3. Contour and surface plots of $f(x_1, x_2) = (x_1^2 + x_2^2 - 1)^2 + (x_2^2 - 1)^2$. The five stationary points $(0,0), (0,1), (0,-1), (1,0), (-1,0)$ are denoted by asterisks. The points $(0,-1), (0,1)$ are strict local minimum points as well as global minimum points, $(0,0)$ is a local maximum point, and $(-1,0), (1,0)$ are saddle points.

```
ezsurf('x^2 + y^2 - 1)^2 + (y^2 - 1)^2', [-2 2 -1.5 1.5])  
colorbar  
view(-30, 30)
```

Example 4

Consider the function $f(x, y) = \frac{x + y}{x^2 + y^2 + 1}$ over $\mathbb{R}^2$.

- $\nabla f(x, y) = \frac{1}{(x^2 + y^2 + 1)^2} \begin{bmatrix} (x^2 + y^2 + 1) - 2(x + y)x \\ (x^2 + y^2 + 1) - 2(x + y)y \end{bmatrix} := \mathbf{0} \Rightarrow$
 $-x^2 - 2xy + y^2 = -1, x^2 - 2xy - y^2 = -1$
 $\Rightarrow xy = 1/2$ (adding), $x^2 = y^2$ (subtracting)
 $\Rightarrow$ stationary points are $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$
- For any $(x, y)^\top \in \mathbb{R}^2$, from the Cauchy-Schwarz inequality,

$$f(x, y) = \frac{(x, y)^\top \cdot (1, 1)^\top}{x^2 + y^2 + 1} \leq \sqrt{2} \frac{\sqrt{x^2 + y^2}}{x^2 + y^2 + 1} \leq \sqrt{2} \max_{t \geq 0} \frac{t}{t^2 + 1} \leq \frac{\sqrt{2}}{2}.$$

$$\because (t - 1)^2 \geq 0 \Rightarrow t^2 + 1 \geq 2t$$

- $\because f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = \frac{\sqrt{2}}{2} \quad \therefore (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \text{ is the global maximum point}$

$$\text{Similarly, } \because \frac{(-x, -y)^\top \cdot (1, 1)^\top}{x^2 + y^2 + 1} \leq \frac{\sqrt{2}}{2} \quad \therefore f(x, y) \geq \frac{-\sqrt{2}}{2}$$

$$\because f(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}) = \frac{-\sqrt{2}}{2} \quad \therefore (\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}) \text{ is the global minimum point}$$

Example 5

Consider the function $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4$ over $\mathbb{R}^2$.

- $\nabla f(x_1, x_2) = \begin{bmatrix} -4x_1 + x_2^2 + 16x_1^3 \\ 2x_1x_2 \end{bmatrix} := \mathbf{0}$.
 $\Rightarrow$ stationary points are $(0,0), (1/2,0), (-1/2,0)$.
- The Hessian of the function is $\nabla^2 f(x_1, x_2) = \begin{bmatrix} -4 + 48x_1^2 & 2x_2 \\ 2x_2 & 2x_1 \end{bmatrix}$.
- $\nabla^2 f(1/2, 0) = \begin{bmatrix} 8 & 0 \\ 0 & 1 \end{bmatrix} \succ \mathbf{0}$. $\therefore (1/2, 0)$ is a strict local minimum point
(not global, $f(-1, x_2) = 2 - x_2^2 \rightarrow -\infty, x_2 \rightarrow \infty$)
 $\nabla^2 f(-1/2, 0) = \begin{bmatrix} 8 & 0 \\ 0 & -1 \end{bmatrix}$, indefinite. $\therefore (-1/2, 0)$ saddle point
 $\nabla^2 f(0, 0) = \begin{bmatrix} -4 & 0 \\ 0 & 0 \end{bmatrix}$, a negative semidefinite matrix.
 $\therefore f(\alpha^4, \alpha) = \alpha^6(-2\alpha^2 + 1 + 4\alpha^{10}) > 0$
 $f(-\alpha^4, \alpha) = \alpha^6(-2\alpha^2 - 1 + 4\alpha^{10}) < 0$ for $0 < \alpha \ll 1$
 $\therefore (0, 0)$ is a saddle point of f

Contour and surface plots of Example 5

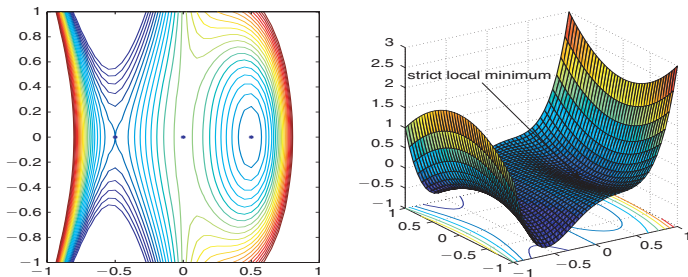


Figure 2.4. Contour and surface plots of $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4$. The three stationary point $(0, 0)$, $(0.5, 0)$, $(-0.5, 0)$ are denoted by asterisks. The point $(0.5, 0)$ is a strict local minimum, while $(0, 0)$ and $(-0.5, 0)$ are saddle points.

```
ezsurf('-2*x^2 + x*y^2 + 4*x^4', [-1 1 -1 1])  
colorbar  
view(-45, 30)
```

Global optimality conditions

- **Theorem:** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice continuously differentiable function. Assume that $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}, \forall \mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f . Then $\mathbf{x}^*$ is a global minimum point of f .

Proof: By the linear approximation theorem, $\forall \mathbf{x} \in \mathbb{R}^n, \exists \mathbf{z}_x \in (\mathbf{x}^*, \mathbf{x})$ such that

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\mathbf{z}_x)(\mathbf{x} - \mathbf{x}^*).$$

Since $\nabla^2 f(\mathbf{z}_x) \succeq \mathbf{0}$, we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$. $\therefore \mathbf{x}^*$ is a global minimum point of f $\square$

- **Example:**

$$f(\mathbf{x}) := x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + (x_1^2 + x_2^2 + x_3^2)^2.$$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 + x_2 + x_3 + 4x_1(x_1^2 + x_2^2 + x_3^2) \\ 2x_2 + x_1 + x_3 + 4x_2(x_1^2 + x_2^2 + x_3^2) \\ 2x_3 + x_1 + x_2 + 4x_3(x_1^2 + x_2^2 + x_3^2) \end{bmatrix}.$$

Obviously, $(x_1, x_2, x_3) = (0, 0, 0)$ is a stationary point.

The Hessian is $\nabla^2 f(\mathbf{x}) = \mathbf{A} + \mathbf{B}(\mathbf{x}) + \mathbf{C}(\mathbf{x})$, where

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \succeq \mathbf{0}, \text{ since it is diagonally dominant with positive diagonal}$$

elements, $\mathbf{B}(\mathbf{x}) = 4(x_1^2 + x_2^2 + x_3^2)\mathbf{I}_3 \succeq \mathbf{0}$, and $\mathbf{C}(\mathbf{x}) = 8\mathbf{x}\mathbf{x}^\top \succeq \mathbf{0}$.

$\therefore \nabla^2 f(\mathbf{x}) \succeq \mathbf{0} \quad \therefore \mathbf{x} = (0, 0, 0)^\top$ is a global minimum point of f over $\mathbb{R}^3$

Quadratic functions

Quadratic functions are an important class of functions that are useful in the modeling of many optimization problems.

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

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c,$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

- The gradient and Hessian of the above quadratic function $f(\mathbf{x})$:

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b}, \quad \nabla^2 f(\mathbf{x}) = 2\mathbf{A}.$$

- **Properties of quadratic functions:**

(1) $\mathbf{x}$ is a stationary point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.

(2) If $\mathbf{A} \succeq \mathbf{0}$, then $\mathbf{x}$ is a global minimum point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.

Proof: By Theorems on page 8 and page 29. $\square$

(3) If $\mathbf{A} \succ \mathbf{0}$, $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is a strict global minimum point of f .

Proof: If $\mathbf{A} \succ \mathbf{0}$, then $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is the unique solution to $\mathbf{A}\mathbf{x} = -\mathbf{b}$. Hence, it is the unique global minimum point of f . $\square$

Note: In (3), the minimal value of f is given by

$$f(\mathbf{x}) = (-\mathbf{A}^{-1}\mathbf{b})^\top \mathbf{A}(-\mathbf{A}^{-1}\mathbf{b}) - 2\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b} + c = c - \mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}.$$

Coerciveness of quadratic functions

Theorem: Let $f(x) = x^\top Ax + 2b^\top x + c$, 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 $A \succ 0$.

Proof:

($\Rightarrow$) Assume that $A \succ 0$. Then $x^\top Ax \geq \alpha \|x\|^2$ with $\alpha = \lambda_{\min}(A) > 0$. Thus,

$$f(x) \geq \alpha \|x\|^2 - 2\|b\|\|x\| + c = \alpha \|x\| \left(\|x\| - 2\frac{\|b\|}{\alpha} \right) + c \rightarrow \infty, \quad \text{as } \|x\| \rightarrow \infty.$$

Therefore, f is coercive.

($\Leftarrow$) Assume that f is coercive. We need to prove that $A \succ 0$. We first show that there does not exist a negative eigenvalue. Suppose $\exists \mathbf{0} \neq v \in \mathbb{R}^n$, $\lambda < 0$ s.t. $Av = \lambda v$. Then for any $\alpha \in \mathbb{R}$,

$$f(\alpha v) = \lambda \|\alpha v\|^2 + 2(b^\top v)\alpha + c \rightarrow -\infty \quad \text{as } \alpha \rightarrow \infty.$$

This is a contradiction. We now show that 0 cannot be an eigenvalue of A . Suppose $\exists \mathbf{0} \neq v \in \mathbb{R}^n$ s.t. $Av = \mathbf{0}$. Then for any $\alpha \in \mathbb{R}$,

$$f(\alpha v) = 2(b^\top v)\alpha + c.$$

If $b^\top v = 0$ then $f(\alpha v) \rightarrow c$ as $\alpha \rightarrow \infty$. If $b^\top v > 0$ then $f(\alpha v) \rightarrow -\infty$ as $\alpha \rightarrow -\infty$.

If $b^\top v < 0$ then $f(\alpha v) \rightarrow -\infty$ as $\alpha \rightarrow \infty$. All these contradict the coerciveness of f . $\square$

Characterization of the nonnegativity of quadratic functions

Theorem: Let $f(x) = x^\top Ax + 2b^\top x + c$, 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) $f(x) = x^\top Ax + 2b^\top x + c \geq 0, \forall x \in \mathbb{R}^n$.

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

Proof:

(b) $\Rightarrow$ (a): For any $x \in \mathbb{R}^n$, $0 \leq \begin{bmatrix} x \\ 1 \end{bmatrix}^\top \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} = x^\top Ax + 2b^\top x + c \Rightarrow$ (a).

(a) $\Rightarrow$ (b): We begin by showing that $A \succeq 0$.

Suppose not. $\exists v \neq 0 \in \mathbb{R}^n$ and $\lambda < 0$ s.t. $Av = \lambda v$. Thus, for any $\alpha \in \mathbb{R}$,

$$f(\alpha v) = \lambda \alpha^2 \|v\|^2 + 2(b^\top v)\alpha + c \rightarrow -\infty \quad \text{as } \alpha \rightarrow -\infty,$$

contradicting the nonnegativity of f . Our objective is to prove (b). We want to show

that for any $y \in \mathbb{R}^n$ and $t \in \mathbb{R}$, $\begin{bmatrix} y \\ t \end{bmatrix}^\top \begin{bmatrix} A & b \\ b^\top & c \end{bmatrix} \begin{bmatrix} y \\ t \end{bmatrix} \geq 0$, which is equivalent to

$$y^\top Ay + 2tb^\top y + ct^2 \geq 0. \quad (\star)$$

If $t = 0$ then $y^\top Ay + 2tb^\top y + ct^2 = y^\top Ay \geq 0$, since $A \succeq 0$. We obtain $(\star)$.

If $t \neq 0$ then $0 \leq t^2 f(y/t) = y^\top Ay + 2tb^\top y + ct^2$, we have $(\star)$. $\square$