2. Convex Sets and Convexity

2.1. Convex Sets

A set S is *convex* if any line segment joining two elements of S is a subset of S, i.e., for any $\mathbf{x}_1, \mathbf{x}_2 \in S$ and $\alpha \in [0, 1]$, $\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2 \in S$. By definition, the empty set ϕ is convex. There are multiple reasons to define convex sets. One practical reason is to have a connected domain.

Exercise 2.1.1. Let S_i be a convex set for all i. Show that $\bigcap_i S_i$ is also a convex set. ✧✧

Exercise 2.1.2. Let $g_1(x_1, x_2, x_3) = 2x_1 - 5x_2 + 3x_3$ and $g_2(x_1, x_2, x_3) = x_1 + 2x_2 + x_3$. Show that the $G = \{\mathbf{x}|\ g_1(\mathbf{x}) \le 8 \text{ and } g_2(\mathbf{x}) \le 8\}$ is a convex set. *Remark*. By a similar argument, the feasible set of a linear program is a convex set. ✧✧

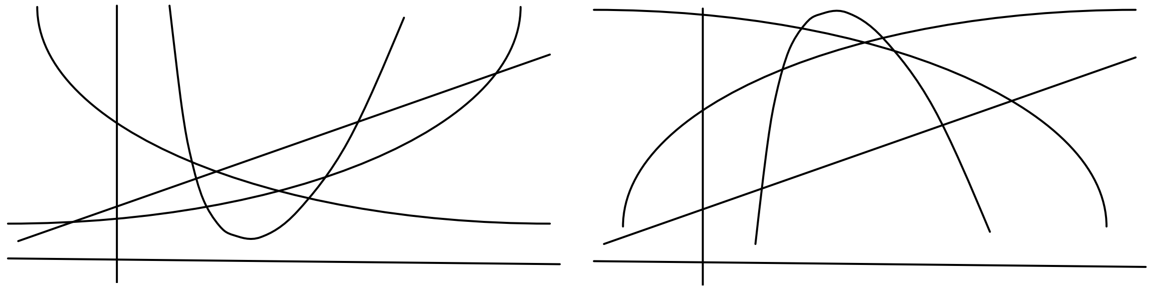
2.2. Convex Functions and Concave Functions

Many results in non-linear optimization involve convex (and concave) functions. In fact, most of those nice, beautiful results rely on convexity (and concavity).

Definition 2.2.1. A function $f(\mathbf{x})$ is (*strictly*) *convex* over a convex $S \subseteq \Re^n$ iff

$$f(\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2)\ (<) \le \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \text{ for all } \mathbf{x}_1, \mathbf{x}_2 \in S \text{ and } 0 \le \alpha \le 1.$$ ✧✧

Definition 2.2.1 has a nice graphical representation, i.e., function f is a convex function in a convex set S if the line segment joining $(\mathbf{x}_1, f(\mathbf{x}_1))$ and $(\mathbf{x}_2, f(\mathbf{x}_2))$ is not less than the functional value of any point $\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2$ between $\mathbf{x}_1$ and $\mathbf{x}_2$.



Examples of Convex Functions　　Examples of Concave Functions

Definition 2.2.2. A function $f(\mathbf{x})$ is (*strictly*) *concave* over a convex $S \subseteq \Re^n$ iff

$$f(\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2)\ (>) \geq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \text{ for all } \mathbf{x}_1, \mathbf{x}_2 \in S \text{ and } 0 \leq \alpha \leq 1.$$ ✧✧

Properties of Convex and Concave Functions

- Let $f(\mathbf{x})$ be a convex (resp. concave) function. Then $-f(\mathbf{x})$ is a concave (resp. convex) function.
- Let $f_i(\mathbf{x})$ be a convex (resp. concave) function and $c_i > 0$. Then $\Sigma_i c_i f_i(\mathbf{x})$ is a convex (resp. concave) function.
- Let $f_i(\mathbf{x})$ be a convex function and $b_i \in \Re$ for $i = 1, \ldots, m$. The set $S = \{\mathbf{x} \in \Re^n | f_i(\mathbf{x}) \leq b_i, i = 1, \ldots, m\}$ is a convex set.
- Let $f_i(\mathbf{x})$ be a convex (resp. concave) function. Then $\max_i\{f_i\}$ (resp. $\min_i\{f_i\}$) is also a convex (resp. concave) function.
- A convex (resp. concave) function is continuous except possibly at the boundary.
- **Theorem 1** (pp 325 of JB). Let S be the solution set defined by a set of linear constraints $\mathbf{a}_i^T\mathbf{x} \leq b_i$, $i = 1, \ldots, m$. If the maximization problem $\max\{f(\mathbf{x})| \mathbf{x} \in S\}$ is feasible for a convex function f, then a global maximum exists at a corner point of S.

<u>Exercise 2.2.1</u>. Let S be defined by $-2 \leq x \leq 5$, $-2 \leq y \leq 4$, and $f(x, y) = x^2+y^2$. Identify the local and global maxima of f in S. ✧✧

- **Theorem 2** (pp 326 of JB). Let $f(\mathbf{x})$ be a convex function and the solution set S be convex. If the minimization problem $\min\{f(\mathbf{x})| \mathbf{x} \in S\}$ is feasible, then all local minima are also global minima. The minimum is unique if f is a strictly convex function.

<u>Exercise 2.2.2</u>. Give the intuition of Theorem 2. ✧✧

2.3. Definiteness of Square Matrices

In this subsection we divert to discuss the *definiteness* of a square symmetric matrix, because the convexity of function f depends on the definiteness of its Hessian.

Let $\mathbf{A}$ be an $n\times n$ *symmetric* matrix and $\mathbf{x}$ an n-dimensional vector. A quadratic function is formed by $\mathbf{x}^T\mathbf{A}\mathbf{x}$. Then

$$\mathbf{A} \text{ is } \begin{array}{c}\text{positive definite}\\ \text{negative definite}\\ \text{positive semi-definite}\\ \text{negative semi-definite}\end{array} \text{ if } \mathbf{x}^T\mathbf{A}\mathbf{x} \begin{array}{c}>\\ <\\ \geq\\ \leq\end{array} 0 \text{ for any } \mathbf{x} \neq 0.$$

A is *indefinite* if it does not belong to the above four types.

Example 2.3.1. Check the definiteness of these matrices: $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} -1 & 1 \\ 1 & -4 \end{bmatrix}$, $\begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix}$, $\begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}$.

Proof. (a). $(x \quad y)\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (2x+y, x+2y)\begin{pmatrix} x \\ y \end{pmatrix} = 2x^2+xy+xy+2y^2 = x^2+ y^2+(x+y)^2 > 0$ for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ is positive definite.

(b). $(x \quad y)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x+y)^2 \geq 0$ for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is positive semi-definite.

(c). $(x \quad y)\begin{bmatrix} -1 & 1 \\ 1 & -4 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (-x+y,x-4y)\begin{pmatrix} x \\ y \end{pmatrix} = -x^2+xy+xy-4y^2 = -(x-y)^2-3y^2 < 0$ for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} -1 & 1 \\ 1 & -4 \end{bmatrix}$ is negative definite.

(d). $(x \quad y)\begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = -9x^2+12xy-4y^2 = -(3x-2y)^2 \leq 0$ for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix}$ is negative semi-definite.

(e). $(x \quad y)\begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 3x^2+2xy = x^2 - y^2 + (x+y)^2$, which can be positive and negative for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}$ is indefinite.

(f). $(x \quad y)\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = -2x^2+2xy+ y^2 = (x+y)^2 - 3x^2$, which can be positive and negative for $(x, y) \neq (0, 0)$. Therefore, $\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}$ is indefinite. ✧✧

Remarks. The definiteness of a matrix is also related to the optimization result.

Minimizing $(x \quad y)\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ gives (0, 0) as the unique global minimum; minimizing $(x \quad y)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ gives multiple minimum when $x = -y$; maximizing $(x \quad y)\begin{bmatrix} -1 & 1 \\ 1 & -4 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ gives (0, 0) as the unique global maximum; maximizing $(x \quad y)\begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ gives multiple maximum for 3x = 2y; there is neither minimum nor maximum for the two functions $(x \quad y)\begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ and $(x \quad y)\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$. ✧✧

In general, the definiteness of a *symmetric* matrix **A** can be found from the determinants of **A**[1]:

- **A** is *positive definite* iff the determinants of *all* the *leading principal* sub-matrices[#2] are positive for $i = 1, \ldots, n$.
- **A** is *negative definite* iff $a_{11} < 0$ and the determinants of remaining leading principal sub-matrices alternate in sign, i.e., the determinant of the second principal sub-matrix > 0, the third < 0, and so on.
- **A** is *positive semi-definite* iff the determinants of *all principal* sub-matrices, *leading or not*, are non-negative.
- **A** is *negative semi-definite* iff the determinants of *all* principal sub-matrices of *odd order*, leading or not, are non-positive, and the determinants of *all* principal sub-matrices of *even order*, leading or not, are non-negative.
- **A** is *indefinite* iff it is none of the above four types.

It is also possible to check the (semi)definiteness of a square symmetric matrix by diagonalizing it. We skip this method.

<u>Example 2.3.2</u>. There are simple formulas to get the determinants For 2×2 and 3×3 matrices.

1

#2 The leading principal sub-matrices of $\begin{bmatrix} 4 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 6 \end{bmatrix}$ are [4], $\begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$, and $\begin{bmatrix} 4 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 6 \end{bmatrix}$. However, $\begin{bmatrix} 4 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 6 \end{bmatrix}$ does have principal sub-matrices (that are not leading), including [2] and [6] of order 1, $\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$ and $\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}$ of order 2. Naturally, for an $n \times n$ matrix, it has C_k^n principal sub-matrices of order k.

For $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, det(**A**) = *ab*–*cd*, and for $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, det(**A**) = *aei* + *cdh* + *bfg* – *ceg* – *afh* – *bdi*. ✧✧

In the context of Hessians of function *f*, **A** may be a function of variables rather than a set of pure numbers.

Example 2.3.3. The following parts are examples from the textbook JB. Check the (semi-)definiteness of the Hessians of the following functions. Note especially that in (d) the Hessian changes with the value of (x_1, x_2).

(a). $f(\mathbf{x}) = 3x_1x_2 + x_1^2 + 3x_2^2$.

(b). $f(\mathbf{x}) = 24x_1x_2 + 9x_1^2 + 16x_2^2$.

(c). $f(\mathbf{x}) = x_1x_2 + x_1x_3 + x_2x_3 + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$.

(d). $f(\mathbf{x}) = \left(x_2 - x_1^2\right)^2 + (1 - x_1)^2$. ✧✧

2.4. Differentiable Convex Functions

The Mean Value Theorem and the Taylor's Theorem can be regarded as the linear and the quadratic approximations of a function in a given direction.

Mean Value Theorem. For $f \in C^1$ in the linear segment container [$\mathbf{x}_1$, $\mathbf{x}_2$], there exists an α, $0 \le \alpha \le 1$, such that

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)\ (\mathbf{x}_2 - \mathbf{x}_1).$$ ✧✧

Exercise 2.4.1. Let $f(x, y) = x^2 + y^2$.

(a). Find α such that $f\begin{pmatrix}1\\1\end{pmatrix} = f\begin{pmatrix}0\\0\end{pmatrix} + \nabla^T f\left((1-\alpha)\begin{pmatrix}1\\1\end{pmatrix}\right)\begin{pmatrix}1\\1\end{pmatrix}$.

(b). For the α found in (a), $f\begin{pmatrix}0.99\\0.99\end{pmatrix} \approx f\begin{pmatrix}0\\0\end{pmatrix} + \nabla^T f\left((1-\alpha)\begin{pmatrix}1\\1\end{pmatrix}\right)\begin{pmatrix}0.99\\0.99\end{pmatrix}$ and $f\begin{pmatrix}1.01\\1.01\end{pmatrix} \approx$

$$f\begin{pmatrix}0\\0\end{pmatrix}+\nabla^T f\left((1-\alpha)\begin{pmatrix}1\\1\end{pmatrix}\right)\begin{pmatrix}1.01\\1.01\end{pmatrix}.$$ ✧✧

Taylor's Theorem. For $f \in C^2$ in the linear segment container [$\mathbf{x}_1$, $\mathbf{x}_2$], there exists an α, $0 \le \alpha \le 1$, such that

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2-\mathbf{x}_1) + \tfrac{1}{2}(\mathbf{x}_2-\mathbf{x}_1)^T\mathbf{H}(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)(\mathbf{x}_2 - \mathbf{x}_1).$$ ✧✧

Definition of a convex function as a function above tangent: Suppose that $f \in C^1$. Then f is convex (over convex set) S iff $f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$.

Proof. First consider a convex f. Take $\mathbf{x} = \mathbf{x}_1$ and $\mathbf{y} = \mathbf{x}_2$. From the convexity of f, $\frac{f(\mathbf{x}_1+\alpha(\mathbf{x}_2-\mathbf{x}_1))-f(\mathbf{x}_1)}{\alpha} \le f(\mathbf{x}_2) - f(\mathbf{x}_1)$. Passing to the limit as α goes to zero gives $\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \le f(\mathbf{x}_2) - f(\mathbf{x}_1)$. Now suppose that $f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$. Choosing $\mathbf{y} = \mathbf{x}_1$ and again $\mathbf{y} = \mathbf{x}_2$ gives two inequalities. Adding the two inequalities gives

$$\alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \ge f(\mathbf{x}) + \nabla f(\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2 - \mathbf{x}).$$

Further choosing $\mathbf{x} = \alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2$ completes the proof. ✧✧

Definition of a convex function by its Hessian. Suppose that $f \in C^2$. Then f is convex (over convex set) S iff the Hessian of f is positive semi-definite throughout S.

Proof. From Taylor's Theorem, for $f \in C^2$,

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2-\mathbf{x}_1) + \tfrac{1}{2}(\mathbf{x}_2-\mathbf{x}_1)^T\mathbf{H}(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)(\mathbf{x}_2 - \mathbf{x}_1).$$

$\mathbf{H}$ is positive semi-definite is equivalent to $f(\mathbf{x}_2) \ge f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2-\mathbf{x}_1)$. ✧✧

For a single-variable function $f \in C^2$, f is convex iff $\frac{d^2 f(x)}{d^2 x} \ge 0$. For a quadratic form $f(\mathbf{x}) = a + \mathbf{c}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} \in C^2$, its Hessian $\mathbf{H} = \mathbf{Q}$. Thus, f is f is convex if it is a positive semi-definite function.

There are different types of optimization problems. Please refer to Figure 9.22 **Categorization of Optimization Problems** of JB.