DDA6010

by Luo Yidong

2025 年 3 月 12 日

1 Mathematical Basis

1.1 Definition (Mathematical Formulation)

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & && \mathbf{x} \in C \end{aligned} \tag{1}$$

Vector $\mathbf{x} = (x_1, \ldots, x_n)^\top$ represents **optimization (decision) variables**.

Function $f : \mathbb{R}^n \to \mathbb{R}$ is an **objective function**.

Functions $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$ are **constraint functions** (representing inequality constraints).

Set $C \subseteq \mathbb{R}^n$ is a **constraint set**.

1.2 Open and Close Line

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the closed line segment between $\mathbf{x}$ and $\mathbf{y}$ is given by:

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

The open line segment $(\mathbf{x}, \mathbf{y})$ is similarly defined as:

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

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

1.3 Unit-Simplex

$$\Delta_n = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq 0, \mathbf{e}^\top \mathbf{x} = 1, \text{where} \quad \mathbf{e} = [1, 1, \cdots]^T\right\}.$$

1.4 Norms

The ℓ_p-norm ($p \geq 1$) is defined by:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

The ℓ_1-norm (Manhattan norm or Taxicab Norm) is:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|.$$

The ℓ_2-norm (Euclidean norm) is :

$$\|\mathbf{x}\| = \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The ℓ_∞-norm (infinity norm) is:

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

1.5 Cauchy-Schwarz Inequality

For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$|\mathbf{x}^\top \mathbf{y}| \le \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2.$$

1.6 Matrix Norm

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

- (Nonnegativity) $\|\mathbf{A}\| \ge 0$ for any $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$.
- (Positive homogeneity) $\|\lambda\mathbf{A}\| = |\lambda|\|\mathbf{A}\|$ for any $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\lambda \in \mathbb{R}$.
- (Triangle inequality) $\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$ for any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$.
- (Submultiplicativity) $\|\mathbf{AB}\| \le \|\mathbf{A}\|\|\mathbf{B}\|$ for any compatible $\mathbf{A}, \mathbf{B}$.

1.6.1 Frobenius Norm

The Frobenius norm of a matrix $\mathbf{A}$ is given by:

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

The Frobenius norm is not an induced norm.

1.6.2 Spectral Norm

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

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

- ℓ_1-norm: When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, the induced $(1,1)$-matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by
$$\|\mathbf{A}\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{m} |A_{i,j}|.$$
- ℓ_∞-norm: When $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, the induced (∞,∞)-matrix norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by
$$\|\mathbf{A}\|_\infty = \max_{i=1,2,\ldots,m} \sum_{j=1}^{n} |A_{i,j}|.$$

1.6.3 Induced Matrix Norm

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

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

Conclusion:

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

1.6.4 Basic Properties

1. The Frobenius norm is related to the trace of the matrix:

$$\|\mathbf{A}\|_F^2 = \operatorname{trace}(\mathbf{A}^\top \mathbf{A}).$$

2. The Frobenius norm is related to the singular values σ_i of the matrix:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{\min(m,n)} \sigma_i^2,$$

where σ_i are the singular values of the matrix.

3. If the matrix is symmetric positive definite, the Frobenius norm is related to the eigenvalues λ_i of the matrix:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n} \lambda_i^2,$$

where λ_i are the eigenvalues of the matrix.

1.6.5 Calculation Techniques

Matrix Decompositions

- **Singular Value Decomposition (SVD):** If the SVD of $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top$, where Σ is diagonal, then:

$$\|\mathbf{A}\|_F^2 = \operatorname{trace}(\Sigma^2) = \sum_i \sigma_i^2.$$

This is efficient when computing singular values is faster than element-wise computation.

- **QR Decomposition:** If $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is orthogonal, then:

$$\|\mathbf{A}\|_F^2 = \|\mathbf{R}\|_F^2.$$

This is suitable for sparse matrices.

- **Eigenvalue Decomposition (for Symmetric Matrices):** For symmetric matrices:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{n} \lambda_i^2,$$

where λ_i are the eigenvalues of $\mathbf{A}$.

Trace Property

Using the trace property of the Frobenius norm:

$$\|\mathbf{A}\|_F^2 = \text{trace}(\mathbf{A}^\top \mathbf{A}).$$

Only the diagonal elements of $\mathbf{A}^\top \mathbf{A}$ need to be computed:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2.$$

Block Matrices

For a block matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_3 & \mathbf{A}_4 \end{bmatrix},$$

the Frobenius norm is:

$$\|\mathbf{A}\|_F^2 = \|\mathbf{A}_1\|_F^2 + \|\mathbf{A}_2\|_F^2 + \|\mathbf{A}_3\|_F^2 + \|\mathbf{A}_4\|_F^2.$$

Vectorization

Using matrix vectorization:

$$\|\mathbf{A}\|_F^2 = \|\text{vec}(\mathbf{A})\|_2^2,$$

where $\text{vec}(\mathbf{A})$ stacks the columns of $\mathbf{A}$ into a vector. In matrix form:

$$\|\mathbf{A}\|_F^2 = \mathbf{A}(:)^\top \mathbf{A}(:),$$

where $\mathbf{A}(:)$ denotes the vectorized form of $\mathbf{A}$.

Low-Rank Matrices

If $\mathbf{A}$ is a low-rank matrix and can be factorized as $\mathbf{A} = \mathbf{U}\mathbf{V}^\top$ (where $\mathbf{U}$ and $\mathbf{V}$ are small or column matrices), then:

$$\|\mathbf{A}\|_F^2 = \|\mathbf{U}\|_F^2 \cdot \|\mathbf{V}\|_F^2.$$

This greatly reduces computational complexity.

1.7 Eigenvalues and Eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. Then a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ is called an eigenvector of $\mathbf{A}$ if there exists a $\lambda \in \mathbb{C}$ for which

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

The scalar λ is the eigenvalue corresponding to the eigenvector $\mathbf{v}$.

In general, real-valued matrices can have complex eigenvalues, but when the matrix is symmetric the eigenvalues are necessarily real.

1.8 Positive Semi-Definite Check

1.8.1 Definition-Based Approach

A **symmetric** matrix A is positive semi-definite if, for any vector $\mathbf{x} \in \mathbb{R}^n$, the following holds:

$$\mathbf{x}^T A\mathbf{x} \geq 0$$

Pros: This is a strict and general standard for checking positive semi-definite.

Cons: It is often impractical to verify for all possible vectors $\mathbf{x}$.

1.8.2 Eigenvalue Test

A matrix A is positive semi-definite if all of its eigenvalues are non-negative:

$$\lambda_i \geq 0 \quad \forall i$$

Steps:

- 1. Solve the characteristic equation $\det(A - \lambda I) = 0$;
- 2. Check if all eigenvalues λ_i are non-negative.

1.8.3 Principal Minor Test

For a symmetric matrix A, it is positive semi-definite if all of its **leading principal minors** (determinants of upper-left submatrices) are non-negative.

Example: Determine whether the following 3×3 matrix is positive semi-definite.

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Step 1: Confirm matrix symmetry

Clearly, A is symmetric since $A = A^\top$.

Step 2: Calculate the principal minors

1st principal minor: The diagonal elements of A.

$$\det(A_1^{(1)}) = 2 \geq 0 \quad \det(A_1^{(2)}) = 2 \geq 0 \quad \det(A_1^{(3)}) = 2 \geq 0$$

2nd principal minor: Select submatrices corresponding to any two rows and columns.

1. Choose the first two rows and columns:

$$A_2^{(1)} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\det(A_2^{(1)}) = (2)(2) - (-1)(-1) = 4 - 1 = 3 \geq 0$$

2. Choose the 1st and 3rd rows and columns:

$$A_2^{(2)} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\det(A_2^{(2)}) = (2)(2) - (0)(0) = 4 \geq 0$$

3. Choose the last two rows and columns:

$$A_2^{(3)} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\det(A_2^{(3)}) = (2)(2) - (-1)(-1) = 4 - 1 = 3 \geq 0$$

3rd principal minor: The determinant of the entire matrix.

$$\det(A) = 2 \cdot 2 - (-1)(-1) - (1)(-2 \cdot 0 \cdot (-1)) + 0 = 2(4-1) - (-1)(2-0) + 0 = 2 \times 3 + 2 = 8 \geq 0$$

1.8.4 Cholesky Decomposition

A matrix A is positive semi-definite if and only if it can be decomposed as:

$$A = LL^T$$

where L is a lower triangular matrix with non-negative diagonal entries.

Steps:

- 1. Attempt to perform Cholesky decomposition;
- 2. If successful, ensure that the diagonal elements of L are non-negative.

Example:

The matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ has the following Cholesky decomposition:

$$L = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A = LL^T$$

Thus, A is positive semi-definite.

Criterion	Positive Definite	Positive Semi-Definite
Definition	$\mathbf{x}^T A\mathbf{x} > 0$ for all non-zero $\mathbf{x}$	$\mathbf{x}^T A\mathbf{x} \geq 0$ for all $\mathbf{x}$
Eigenvalues	All $\lambda_i > 0$	All $\lambda_i \geq 0$
Principal minors	All positive	All non-negative
Cholesky decomposition	$L_{ii} > 0$	$L_{ii} \geq 0$

表 1: Comparison of Positive Definite and Positive Semi-Definite Matrices

1.9 The Spectral Factorization Theorem

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ ($\mathbf{U}^\top \mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}$) and a diagonal matrix $\mathbf{D} = \operatorname{diag}(d_1, d_2, \dots, d_n)$ for which

$$\mathbf{U}^\top \mathbf{A} \mathbf{U} = \mathbf{D}.$$

A direct result is that $\operatorname{Tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i(\mathbf{A})$ and

$$\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i(\mathbf{A}).$$

1.10 Properties (Orthogonal Matrices)

An orthogonal matrix $\mathbf{Q}$ has the following key properties:

1. **Inverse equals transpose:**

   $$\mathbf{Q}^\top = \mathbf{Q}^{-1}$$

   This means that multiplying $\mathbf{Q}$ by its transpose results in the identity matrix:

   $$\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}.$$

2. **Preserves vector norms:** For any vector $\mathbf{x}$,

   $$\|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|.$$

   This indicates that orthogonal matrices preserve the length (norm) of vectors.

3. **Preserves dot products:** For any vectors $\mathbf{x}$ and $\mathbf{y}$,

   $$(\mathbf{Q}\mathbf{x})^\top (\mathbf{Q}\mathbf{y}) = \mathbf{x}^\top \mathbf{y}.$$

   This means orthogonal matrices preserve angles between vectors.

4. **Determinant:** The determinant of an orthogonal matrix is either $+1$ or -1:

   $$\det(\mathbf{Q}) = \pm 1.$$

   A determinant of $+1$ represents a proper rotation, and -1 represents a reflection.

5. **Eigenvalues:** If λ is an eigenvalue of $\mathbf{Q}$, then $|\lambda| = 1$, meaning all eigenvalues lie on the unit circle in the complex plane.

6. **Orthogonal matrix columns:** The columns (and rows) of an orthogonal matrix form an orthonormal set, meaning they are orthogonal to each other and have unit length.

1.11 Definition (Open and Close Ball)

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

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

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

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

1.12 Definition (Ellipsoids)

An *ellipsoid* is a set of the form

$$E = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \leq 0\},$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

1.13 Definition (Interior points)

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

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

1.14 Example

$$\text{int}(\mathbb{R}^n_+) = \mathbb{R}^n_{++}, \quad \text{int}(B[\mathbf{c}, r]) = B(\mathbf{c}, r), \quad \text{int}([\mathbf{x}, \mathbf{y}]) = ?$$

1.15 Proposition

A union of any number of open sets is an open set and the intersection of a nite number of open sets is open.

1.16 Definition (Closeness)

A set $U \subseteq \mathbb{R}^n$ is closed if it contains all the limits of convergent sequences of vectors in U, that is, if $\{\mathbf{x}_i\}_{i=1}^{\infty} \subseteq U$ satisfies $\mathbf{x}_i \to \mathbf{x}^*$ as $i \to \infty$, then $\mathbf{x}^* \in U$.

Remark. A known result states that U is closed if and only if its complement U^c is open.

1.17 Example

The closed ball $B[\mathbf{c}, r]$, closed line segments, the nonnegative orthant $\mathbb{R}^n_+$, and the unit simplex Δ_n.

1.18 Definition (Boundary Points)

Given a set $U \subseteq \mathbb{R}^n$, a boundary point of U is a vector $\mathbf{x} \in \mathbb{R}^n$ satisfying the following: any neighborhood of $\mathbf{x}$ 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 $\mathrm{bd}(U)$.

1.19 Example

$$\mathrm{bd}(B(\mathbf{c}, r)) =$$

$$\mathrm{bd}(B[\mathbf{c}, r]) =$$

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

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

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

$$\mathrm{bd}(\Delta_n) =$$

1.20 Definition(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:

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

Another equivalent definition of $\mathrm{cl}(U)$ is:

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

1.21 Example

$$\mathrm{cl}(\mathbb{R}^n_{++}) =$$

$$\mathrm{cl}(B(\mathbf{c}, r)) =$$

$$(\mathbf{x} \neq \mathbf{y}), \mathrm{cl}((\mathbf{x}, \mathbf{y})) =$$

1.22 Definition (Boundedness)

A set $U \subseteq \mathbb{R}^n$ is called bounded if there exists $M > 0$ for which $U \subseteq B(0, M)$.

1.23 Definition (Compactness)

A set $U \subseteq \mathbb{R}^n$ is called compact if it is closed and bounded.

1.24 Definition (Directional Derivatives and Gradients)

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

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

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

1.25 Remark

1. For any $i = 1, 2, \dots, n$, if the limit

   $$\lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t}$$

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

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

   $$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)^\top.$$

1.26 Definition (Continuous Dierentiability)

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,

$$f'(\mathbf{x}; \mathbf{d}) = \nabla f(\mathbf{x})^\top \mathbf{d}, \quad \mathbf{x} \in U, \mathbf{d} \in \mathbb{R}^n.$$

1.27 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

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

Remark. Another way to write the above result is as follows:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + o(\|\mathbf{y} - \mathbf{x}\|),$$

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

1.28 Definition (Twice Dierentiability)

The partial derivatives ∂f are themselves real-valued functions that can be partially differentiated. The (i, j)-partial derivatives of f at $\mathbf{x} \in U$ (if they exist) are defined by

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

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 $\mathbf{x} \in U$:

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

1.29 Definition (The Hessian)

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

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

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

1.30 Theorem (Linear Approximation 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 $\mathbf{x} \in U$ and $r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$ there exists $\xi \in [\mathbf{x}, \mathbf{y}]$ such that:

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

1.31 Theorem (Quadratic Approximation 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 $\mathbf{x} \in U$ and $r > 0$ satisfy $B(\mathbf{x}, r) \subseteq U$. Then for any $\mathbf{y} \in B(\mathbf{x}, r)$:

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

2 Unconstrained Optimization

2.1 Definition (Stationary Point)

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

2.2 Example

$$\min \left\{ f(x,y) = \frac{x+y}{x^2+y^2+1} : x, y \in \mathbb{R} \right\}$$

$$\nabla f(x,y) = \frac{1}{(x^2+y^2+1)^2} \begin{pmatrix} (x^2+y^2+1) - 2(x+y)x \\ (x^2+y^2+1) - 2(x+y)y \end{pmatrix}.$$

Stationary points are those satisfying:

$$\begin{aligned} -x^2 - 2xy + y^2 &= -1, \\ x^2 - 2xy - y^2 &= -1. \end{aligned}$$

Hence, the stationary points are:

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) - \text{global maximizer}, \qquad \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) - \text{global minimizer}.$$

2.3 Theorem (Second order differentiable)

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 $\mathbf{x}^*$ is a stationary point. Then

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

2. If $\mathbf{x}^*$ is a local maximum point, then $\nabla^2 f(\mathbf{x}^*) \preceq 0$.

2.4 Proof

1. **Stationary Point Condition:**

   Since $\mathbf{x}^*$ is a stationary point, the first derivative (gradient) of f at $\mathbf{x}^*$ is zero:

   $$\nabla f(\mathbf{x}^*) = 0$$

2. **Second-Order Taylor Expansion:**

   For a twice continuously differentiable function f, the second-order Taylor expansion around $\mathbf{x}^*$ is:

   $$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top \mathbf{h} + \frac{1}{2}\mathbf{h}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{h} + o(\|\mathbf{h}\|^2)$$

Given that $\nabla f(\mathbf{x}^*) = 0$, this simplifies to:

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \frac{1}{2}\mathbf{h}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{h} + o(\|\mathbf{h}\|^2)$$

3. **Local Minimum Condition:**

   Since $\mathbf{x}^*$ is a local minimum, for sufficiently small $\mathbf{h}$, the function satisfies:

$$f(\mathbf{x}^* + \mathbf{h}) \geq f(\mathbf{x}^*)$$

   Substituting the Taylor expansion:

$$\frac{1}{2}\mathbf{h}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{h} + o(\|\mathbf{h}\|^2) \geq 0$$

   Dividing both sides by $\|\mathbf{h}\|^2$ (assuming $\mathbf{h} \neq 0$) and taking the limit as $\mathbf{h} \to 0$, the higher-order term $o(\|\mathbf{h}\|^2)$ becomes negligible:

$$\frac{1}{2}\mathbf{h}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{h} \geq 0$$

   This implies that:

$$\mathbf{h}^\top \nabla^2 f(\mathbf{x}^*)\mathbf{h} \geq 0 \quad \text{for all} \quad \mathbf{h} \in \mathbb{R}^n$$

   By definition, this means that the Hessian matrix $\nabla^2 f(\mathbf{x}^*)$ is positive semidefinite, denoted as:

$$\nabla^2 f(\mathbf{x}^*) \succeq 0$$

2.5 Definition (Saddle Point)

Let $f : U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. A stationary point $\mathbf{x}^* \in U$ 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.

2.6 Theorem (Saddle Point Condition)

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 $\mathbf{x}^*$ 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.

2.7 Theorem (Weierstrass Theorem)

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.

2.8 Definition (Coercive)

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

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

2.9 Theorem (Coercivity of Quadratic Functions)

Suppose that

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

where $A \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then f is coercive if and only if $A \succ 0$.

2.10 Proof

Sufficiency: Assume $A \succ 0$, then using eigenvalue properties, we have

$$\lambda_{\min}(A) \leq \frac{\mathbf{x}^\top A\mathbf{x}}{\|\mathbf{x}\|^2} \leq \lambda_{\max}(A),$$

which implies

$$\mathbf{x}^\top A\mathbf{x} \geq \alpha\|\mathbf{x}\|^2,$$

where $\alpha = \lambda_{\min}(A) > 0$. Therefore,

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

Using the Cauchy-Schwarz inequality and taking the limit as $\|\mathbf{x}\| \to \infty$, it follows that $f(\mathbf{x}) \to \infty$, showing that f is coercive.

Necessity: Assume f is coercive. We will prove that A is positive definite. Suppose for contradiction that A is not positive definite, then there exists an eigenvalue $\lambda \leq 0$ and an associated eigenvector $\mathbf{v} \neq 0 \in \mathbb{R}^n$ such that $A\mathbf{v} = \lambda\mathbf{v}$. Consider the function

$$f(a\mathbf{v}) = a^2\lambda\|\mathbf{v}\|^2 + 2a(\mathbf{b}^\top \mathbf{v}) + c.$$

If $\mathbf{b}^\top\mathbf{v} = 0$, then $f(a\mathbf{v}) \to c$, and if $\mathbf{b}^\top\mathbf{v} \neq 0$, $f(a\mathbf{v}) \to -\infty$ or ∞, contradicting the assumption that f is coercive. Thus, A must be positive definite.

If f is not a quadratic function but is strongly convex, then it is also coercive.

2.11 Theorem (Attainment of Global Optima Points for Coercive Functions)

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

2.12 Theorem (Global Optimality Condition)

Let f be a twice continuously differentiable function defined over $\mathbb{R}^n$. Suppose that $\nabla^2 f(\mathbf{x}) \succeq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f. Then $\mathbf{x}^*$ is a global minimum point of f. Proof (using Taylor expansion)

2.13 Definition (Quadratic Function)

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

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

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

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

2.14 Lemma

Let $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \quad (A \in \mathbb{R}^{n\times n}$ sym., $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R})$.

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

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

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

2.15 Lemma (coerciveness of quadratic functions)

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

2.16 Theorem (characterization of the nonnegativity of quadratic functions)

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

(i) $f(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

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

3 Least Squares

3.1 The least squares

To find the solution $\mathbf{x}_{LS}$ that minimizes the quadratic function for the least squares (LS) problem, we need to solve the following optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{b}^\top \mathbf{A}\mathbf{x} + \|\mathbf{b}\|^2$$

Step 1: Take the gradient of $f(\mathbf{x})$

The gradient of the function $f(\mathbf{x})$ with respect to $\mathbf{x}$ is computed as:

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{A}^\top \mathbf{b}$$

Step 2: Set the gradient to zero

To find the critical point, we set the gradient equal to zero:

$$2\mathbf{A}^\top \mathbf{A}\mathbf{x} - 2\mathbf{A}^\top \mathbf{b} = 0$$

Step 3: Solve for x

Simplifying the above equation:

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}$$

This is a system of linear equations. Assuming $\mathbf{A}^\top \mathbf{A}$ is invertible, we can solve for $\mathbf{x}$ by multiplying both sides by the inverse of $\mathbf{A}^\top \mathbf{A}$:

$$\mathbf{x}_{LS} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$$

Thus, the least-squares solution is:

$$\boxed{\mathbf{x}_{LS} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}}$$

3.2 Definition (Regularized Least Squares)

There are several situations in which the least squares solution does not give rise to a good estimate of the "true" vector $\mathbf{x}$. In these cases, a regularized problem (called regularized least squares (RLS)) is often solved:

$$\text{(RLS)} \quad \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda R(\mathbf{x}).$$

Here $\lambda > 0$ is the regularization parameter and $R(\cdot)$ is the regularization function (also called a penalty function). Quadratic regularization is a specific choice of regularization function:

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{D}\mathbf{x}\|^2.$$

The optimal solution of the above problem is

$$\mathbf{x}_{\mathrm{RLS}} = \left(\mathbf{A}^\top\mathbf{A} + \lambda\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{A}^\top\mathbf{b}.$$

How to assure that $\mathbf{A}^\top\mathbf{A} + \lambda\mathbf{D}^\top\mathbf{D}$ is invertible? (answer: $\mathrm{Null}(\mathbf{A}) \cap \mathrm{Null}(\mathbf{D}) = \{0\}$)

3.3 Definition (Nonlinear Least Squares)

Given m points $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$, the goal is to find a circle $C(\mathbf{x}, r)$ defined by a center $\mathbf{x} \in \mathbb{R}^n$ and radius $r \in \mathbb{R}_+$ that best fits these points. The equation of the circle is:

$$C(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}\| = r\}$$

To approximate the circle fitting, we start by minimizing the distance from each point $\mathbf{a}_i$ to the circle, giving:

$$\|\mathbf{x} - \mathbf{a}_i\| \approx r, \quad i = 1, 2, \dots, m$$

Squaring both sides to avoid nondifferentiability:

$$\|\mathbf{x} - \mathbf{a}_i\|^2 \approx r^2, \quad i = 1, 2, \dots, m$$

The problem can now be formulated as a nonlinear least squares minimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n, r\in\mathbb{R}_+} \sum_{i=1}^{m} \left(\|\mathbf{x} - \mathbf{a}_i\|^2 - r^2\right)^2$$

Expanding the terms gives:

$$\min_{\mathbf{x}, r} \sum_{i=1}^{m} \left(-2\mathbf{a}_i^\top\mathbf{x} + \|\mathbf{x}\|^2 - r^2 + \|\mathbf{a}_i\|^2\right)^2$$

Introducing the new variable $R = \|\mathbf{x}\|^2 - r^2$, the problem becomes:

$$\min_{\mathbf{x}, R} \sum_{i=1}^{m} \left(-2\mathbf{a}_i^\top\mathbf{x} + R + \|\mathbf{a}_i\|^2\right)^2$$

The constraint $\|\mathbf{x}\|^2 \geq R$ can be dropped because any optimal solution $(\hat{\mathbf{x}}, \hat{R})$ automatically satisfies it.

Thus, the final circle fitting least squares (CF-LS) formulation is:

$$\min_{\mathbf{x}, R} \sum_{i=1}^{m} \left(-2\mathbf{a}_i^\top\mathbf{x} + R + \|\mathbf{a}_i\|^2\right)^2$$

4 Gradient

4.1 Definition (The Gradient Method)

Objective: find an optimal solution of the problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$$

The iterative algorithms we consider are of the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k, \quad k = 0, 1, ...$$

- $\mathbf{d}_k$ - direction.
- t_k - stepsize.

4.2 Stepsize Selection Rules

- **Constant stepsize** - $t_k = \bar{t}$ for any k.

- **Exact stepsize** - t_k is a minimizer of f along the ray $\mathbf{x}_k + t\mathbf{d}_k$:

  $$t_k \in \arg\min_{t \geq 0} f(\mathbf{x}_k + t\mathbf{d}_k)$$

  The second strategy is completely theoretical. It is never used in practice since even in one dimensional case we cannot find the exact minimum in finite time.

- **Backtracking** - requires three parameters:
  $s > 0, \alpha \in (0,1), \beta \in (0,1)$. Here start with an initial stepsize $t_k = s$.
  While

  $$f(\mathbf{x}_k) - f(\mathbf{x}_k + t_k\mathbf{d}_k) < -\alpha t_k \nabla f(\mathbf{x}_k)^\top \mathbf{d}_k,$$

  set $t_k := \beta t_k$.
  Sufficient Decrease Property:

  $$f(\mathbf{x}_k) - f(\mathbf{x}_k + t_k\mathbf{d}_k) \geq -\alpha t_k \nabla f(\mathbf{x}_k)^\top \mathbf{d}_k.$$

4.3 Direction

$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k).$$

4.3.1 Lemma

Let f be a continuously differentiable function and let $\mathbf{x} \in \mathbb{R}^n$ be a non-stationary point ($\nabla f(\mathbf{x}) \neq 0$). Then an optimal solution of

$$\min_{\mathbf{d}} \{f'(\mathbf{x};\mathbf{d}) : \|\mathbf{d}\| = 1\}$$

is $\mathbf{d} = -\frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$.

4.4 Exact Line Search

4.4.1 Algorithm

Algorithm 1 The Gradient Method

```
0: Input: $\epsilon > 0$ - tolerance parameter.
0: Initialization: pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily.
0: for $k = 0, 1, 2, ...$ do
0:   pick a stepsize $t_k$ by a line search procedure on the function
         $g(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k))$.
0:   set $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k)$.
0:   if $\|\nabla f(\mathbf{x}_{k+1})\| \leq \epsilon$ then
0:     then STOP and $\mathbf{x}_{k+1}$ is the output.
0:   end if
0: end for=0
```

4.4.2 Example

$$\min x^2 + 2y^2$$

$$\mathbf{x}_0 = (2, 1),\ \epsilon = 10^{-5},\ \text{exact line search.}$$

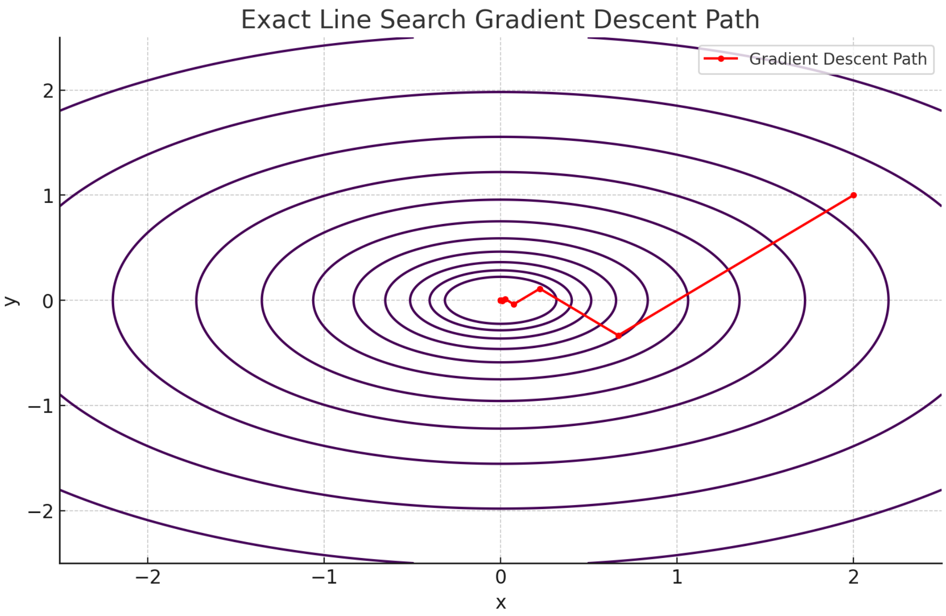
4.4.3 Lemma

Let $\{\mathbf{x}_k\}_{k\geq 0}$ be the sequence generated by the gradient method with exact line search for solving a problem of minimizing a continuously differentiable function f. Then for any $k = 0, 1, 2, ...$

$$(\mathbf{x}_{k+2} - \mathbf{x}_{k+1})^\top (\mathbf{x}_{k+1} - \mathbf{x}_k) = 0.$$

4.4.4 Proof

$$\mathbf{x}_{k+1} - \mathbf{x}_k = -t_k \nabla f(\mathbf{x}_k),\ \mathbf{x}_{k+2} - \mathbf{x}_{k+1} = -t_{k+1} \nabla f(\mathbf{x}_{k+1})$$



Therefore, need to prove

$$\nabla f(\mathbf{x}_k)^\top \nabla f(\mathbf{x}_{k+1}) = 0.$$

$$t_k \in \arg\min_{t \geq 0} \{ g(t) \equiv f(\mathbf{x}_k - t \nabla f(\mathbf{x}_k)) \}$$

Hence, $g'(t_k) = 0$

$$-\nabla f(\mathbf{x}_k)^\top \nabla f(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k)) = 0$$

$$\nabla f(\mathbf{x}_k)^\top \nabla f(\mathbf{x}_{k+1}) = 0$$

4.5 Lipschitz Continuity of the Gradient

4.5.1 Definition

Let f be a continuously differentiable function over $\mathbb{R}^n$. Say that f has a Lipschitz gradient if there exists $L \geq 0$ for which

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|$$

for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

L is called the Lipschitz constant.

- If ∇f is Lipschitz with constant L, then it is also Lipschitz with constant $\tilde{L}$ for all $\tilde{L} \geq L$.
- The class of functions with Lipschitz gradient with constant L is denoted by $C_L^{1,1}(\mathbb{R}^n)$ or just $C_L^{1,1}$.

4.5.2 Remark

Here the first "1" means that f is first-order differentiable or has first-order smoothness. Lipschitz continuity can be viewed as "1-Half Order" smoothness between C^1 and C^2.

$C^{k,\alpha}$ is a widely used notation to describe finer levels of smoothness:

- C^k: The function has k-th order continuous derivatives.

- $C^{k,\alpha}$: The k-th derivative satisfies Hölder continuity with an exponent $\alpha \in (0, 1]$.
  - Hölder continuity is defined as:

$$\|f^{(k)}(x) - f^{(k)}(y)\| \leq C\|x - y\|^\alpha,$$

    where $\alpha = 1$ corresponds to Lipschitz continuity.

4.5.3 Example

- **Linear functions** - Given $\mathbf{a} \in \mathbb{R}^n$, the function $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x}$ is in $C_0^{1,1}$.
- **Quadratic functions** - Let $\mathbf{A}$ be a symmetric $n \times n$ matrix, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then the function

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

  is a $C^{1,1}$ function. The smallest Lipschitz constant of ∇f is $2\|\mathbf{A}\|_2$.

4.5.4 Theorem (Equivalence to Boundedness of the Hessian)

Let f be a twice continuously differentiable function over $\mathbb{R}^n$. Then the following two claims are equivalent:

1. $f \in C_L^{1,1}(\mathbb{R}^n)$
2. $\|\nabla^2 f(\mathbf{x})\| \leq L$ for any $\mathbf{x} \in \mathbb{R}^n$

4.6 Convergence

4.6.1 Lemma (Descent Lemma)

Let $D \subseteq \mathbb{R}^n$ and $f \in C_L^{1,1}(D)$ for some $L > 0$. Then for any $\mathbf{x}, \mathbf{y} \in D$ satisfying $[\mathbf{x}, \mathbf{y}] \subseteq D$ it holds that

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2.$$

4.6.2 Lemma (Sufficient Decrease Lemma)

Suppose that $f \in C_L^{1,1}(D)$ for some $L > 0$. Then for any $\mathbf{x} \in \mathbb{R}^n$ and $t > 0$

$$f(\mathbf{x}) - f(\mathbf{x} - t\nabla f(\mathbf{x})) \geq t\left(1 - \frac{Lt}{2}\right)\|\nabla f(\mathbf{x})\|^2.$$

4.6.3 Lemma (Sufficient Decrease of the Gradient Method)

Let $f \in C_L^{1,1}(D)$. Let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by GM for solving $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ with one of the following stepsize strategies:

- constant stepsize $\bar{t} \in \left(0, \frac{2}{L}\right)$
- exact line search
- backtracking procedure with parameters $s > 0$ and $\alpha, \beta \in (0, 1)$

Then

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq M\|\nabla f(\mathbf{x}_k)\|^2,$$

where

$$M = \begin{cases} \bar{t}\left(1 - \frac{\bar{t}L}{2}\right) & \text{constant stepsize} \\ \frac{1}{2L} & \text{exact line search} \\ \alpha \min\left\{s, \frac{2(1-\alpha\beta)}{L}\right\} & \text{backtracking} \end{cases}$$

4.6.4 Theorem (Rate of Convergence of Gradient Norms)

Under the setting of Theorem 4.25, let f^* be the limit of the convergent sequence $\{f(\mathbf{x}_k)\}_{k\geq 0}$. Then for any $n = 0, 1, 2, \dots,$

$$\min_{k=0,1,\dots,n} \|\nabla f(\mathbf{x}_k)\| \leq \sqrt{\frac{f(\mathbf{x}_0) - f^*}{M(n+1)}}.$$

4.6.5 Theorem (The Effect of Range of Eigenvalue on Convergence Rate)

Let $\{\mathbf{x}_k\}_{k\geq 0}$ be the sequence generated by the gradient method with exact line search for solving the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^\top A\mathbf{x} \quad (A \succeq 0)$$

Then for any $k = 0, 1, \dots,$

$$f(\mathbf{x}_{k+1}) \leq \left(\frac{M-m}{M+m}\right)^2 f(\mathbf{x}_k)$$

where $M = \lambda_{\max}(A)$, $m = \lambda_{\min}(A)$.

4.6.6 Lemma (Kantorovich Inequality)

Let $\mathbf{A}$ be a positive definite $n \times n$ matrix. Then for any $0 \neq \mathbf{x} \in \mathbb{R}^n$, the inequality

$$\frac{(\mathbf{x}^\top\mathbf{x})^2}{(\mathbf{x}^\top\mathbf{A}\mathbf{x})(\mathbf{x}^\top\mathbf{A}^{-1}\mathbf{x})} \geq \frac{4\lambda_{\max}(\mathbf{A})\lambda_{\min}(\mathbf{A})}{(\lambda_{\max}(\mathbf{A}) + \lambda_{\min}(\mathbf{A}))^2}$$

holds.

4.6.7 Definition (The Condition Number)

Let $\mathbf{A}$ be a positive definite $n \times n$ matrix. Then the condition number of $\mathbf{A}$ is defined by

$$\kappa(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}.$$

- Matrices (or quadratic functions) with large condition number are called ill-conditioned.
- Matrices with small condition number are called well-conditioned.
- Large condition number implies large number of iterations of the gradient method.
- Small condition number implies small number of iterations of the gradient method.

- For a non-quadratic function, the asymptotic rate of convergence of $\mathbf{x}_k$ to a stationary point $\mathbf{x}^*$ is usually determined by the condition number of $\nabla^2 f(\mathbf{x}^*)$.

4.6.8 Proposition (Perturbation and condition number)

Suppose that we are given the linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A} \succ 0$ and assume that $\mathbf{x}$ is indeed the solution of the system ($\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$).

Suppose that the right-hand side is perturbed $\mathbf{b} + \Delta\mathbf{b}$. What can be said on the solution of the new system $\mathbf{x} + \Delta\mathbf{x}$?

$$\Delta\mathbf{x} = \mathbf{A}^{-1}\Delta\mathbf{b}.$$

Result:

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \kappa(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}.$$

4.7 Scaled Gradient Method

Consider the minimization problem

$$(P) \quad \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}.$$

For a given nonsingular matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$, we make the linear change of variables $\mathbf{x} = \mathbf{S}\mathbf{y}$, and obtain the equivalent problem

$$(P') \quad \min\{g(\mathbf{y}) \equiv f(\mathbf{S}\mathbf{y}) : \mathbf{y} \in \mathbb{R}^n\}.$$

Since $\nabla g(\mathbf{y}) = \mathbf{S}^\top \nabla f(\mathbf{S}\mathbf{y}) = \mathbf{S}^\top \nabla f(\mathbf{x})$, the gradient method for (P') is

$$\mathbf{y}_{k+1} = \mathbf{y}_k - t_k \mathbf{S}^\top \nabla f(\mathbf{S}\mathbf{y}_k).$$

Multiplying the latter equality by $\mathbf{S}$ from the left, and using the notation $\mathbf{x}_k = \mathbf{S}\mathbf{y}_k$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \mathbf{S}\mathbf{S}^\top \nabla f(\mathbf{x}_k).$$

Defining $\mathbf{D} = \mathbf{S}\mathbf{S}^\top$, we obtain the scaled gradient method:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \mathbf{D} \nabla f(\mathbf{x}_k).$$

$\mathbf{D} \succ 0$, so the direction $-\mathbf{D}\nabla f(\mathbf{x}_k)$ is a descent direction:

$$f'(\mathbf{x}_k; -\mathbf{D}\nabla f(\mathbf{x}_k)) = -\nabla f(\mathbf{x}_k)^\top \mathbf{D} \nabla f(\mathbf{x}_k) < 0.$$

The scaled gradient method with scaling matrix $\mathbf{D}$ is equivalent to the gradient method employed on the function $g(\mathbf{y}) = f(\mathbf{D}^{1/2}\mathbf{y})$. Note that the gradient and Hessian of g are given by

$$\nabla g(\mathbf{y}) = \mathbf{D}^{1/2}\nabla f(\mathbf{D}^{1/2}\mathbf{y}) = \mathbf{D}^{1/2}\nabla f(\mathbf{x})$$

$$\nabla^2 g(\mathbf{y}) = \mathbf{D}^{1/2}\nabla^2 f(\mathbf{D}^{1/2}\mathbf{y})\mathbf{D}^{1/2} = \mathbf{D}^{1/2}\nabla^2 f(\mathbf{x})\mathbf{D}^{1/2}$$

The objective is usually to pick $\mathbf{D}_k$ so as to make $\mathbf{D}_k^{1/2}\nabla^2 f(\mathbf{x}_k)\mathbf{D}_k^{1/2}$ as well-conditioned as possible.

4.7.1 Newton's method

$$\mathbf{D}_k = (\nabla^2 f(\mathbf{x}_k))^{-1}$$

4.7.2 Diagonal scaling

$\mathbf{D}_k$ is picked to be diagonal. For example,

$$(\mathbf{D}_k)_{ii} = \left(\frac{\partial^2 f(\mathbf{x}_k)}{\partial x_i^2}\right)^{-1}$$

Diagonal scaling can be very effective when the decision variables are of different magnitudes.

4.7.3 The Gauss-Newton Method

Nonlinear least squares problem:

$$(\text{NLS}): \quad \min_{\mathbf{x}\in\mathbb{R}^n} \left\{ g(\mathbf{x}) \equiv \sum_{i=1}^{m} (f_i(\mathbf{x}) - c_i)^2 \right\}.$$

$f_1, \dots, f_m$ are continuously differentiable over $\mathbb{R}^n$ and $c_1, \dots, c_m \in \mathbb{R}$. Denote:

$$F(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) - c_1 \\ f_2(\mathbf{x}) - c_2 \\ \vdots \\ f_m(\mathbf{x}) - c_m \end{pmatrix}$$

Then the problem becomes:

$$\min \|F(\mathbf{x})\|^2$$

Given the k-th iterate $\mathbf{x}_k$, the next iterate is chosen to minimize the sum of squares of the linearized terms, that is,

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \sum_{i=1}^{m} \left[f_i(\mathbf{x}_k) + \nabla f_i(\mathbf{x}_k)^\top (\mathbf{x} - \mathbf{x}_k) - c_i\right]^2 \right\}.$$

The general step actually consists of solving the linear LS problem:

$$\min \|\mathbf{A}_k\mathbf{x} - \mathbf{b}_k\|^2$$

where

$$\mathbf{A}_k = \begin{pmatrix} \nabla f_1(\mathbf{x}_k)^\top \\ \nabla f_2(\mathbf{x}_k)^\top \\ \vdots \\ \nabla f_m(\mathbf{x}_k)^\top \end{pmatrix} = J(\mathbf{x}_k)$$

is the so-called Jacobian matrix, assumed to have full column rank, and

$$\mathbf{b}_k = \begin{pmatrix} \nabla f_1(\mathbf{x}_k)^\top \mathbf{x}_k - f_1(\mathbf{x}_k) + c_1 \\ \nabla f_2(\mathbf{x}_k)^\top \mathbf{x}_k - f_2(\mathbf{x}_k) + c_2 \\ \vdots \\ \nabla f_m(\mathbf{x}_k)^\top \mathbf{x}_k - f_m(\mathbf{x}_k) + c_m \end{pmatrix} = J(\mathbf{x}_k)\mathbf{x}_k - F(\mathbf{x}_k).$$

The Gauss-Newton method can thus be written as below according to LSE normal equation:

$$\mathbf{x}_{k+1} = (J(\mathbf{x}_k)^\top J(\mathbf{x}_k))^{-1} J(\mathbf{x}_k)^\top \mathbf{b}_k$$

The GN method can be rewritten as follows:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - (J(\mathbf{x}_k)^\top J(\mathbf{x}_k))^{-1} J(\mathbf{x}_k)^\top F(\mathbf{x}_k) \\ &= \mathbf{x}_k - \frac{1}{2}(J(\mathbf{x}_k)^\top J(\mathbf{x}_k))^{-1} \nabla g(\mathbf{x}_k)\end{aligned}$$

Where the gradient of the objective function $g(\mathbf{x}) = \|F(\mathbf{x})\|^2$ is

$$\nabla g(\mathbf{x}) = 2J(\mathbf{x})^\top F(\mathbf{x})$$

That is, it is a scaled gradient method with a special choice of scaling matrix:

$$\mathbf{D}_k = \frac{1}{2}(J(\mathbf{x}_k)^\top J(\mathbf{x}_k))^{-1}$$

The Gauss-Newton method does not incorporate a stepsize, which might cause it to diverge. A well-known variation of the method incorporating stepsizes is the damped Gauss-Newton Method.

4.7.4 Damped Gauss-Newton Method

1. Compute direction:
$$\mathbf{d}_k = -(J(\mathbf{x}_k)^\top J(\mathbf{x}_k))^{-1} \nabla g(\mathbf{x})$$

2. Determine step size t_k: Find the optimal step size t_k along direction $\mathbf{d}_k$ using a line search method, minimize:
$$h(t) = g(\mathbf{x}_k + t\mathbf{d}_k)$$

3. Update:
$$\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$$

4.7.5 Fermat-Weber Problem

The Fermat-Weber problem is formulated as follows: Given m points in $\mathbb{R}^n$: $\mathbf{a}_1, \dots, \mathbf{a}_m$, and weights $\omega_1, \dots, \omega_m > 0$, the goal is to find a point $\mathbf{x} \in \mathbb{R}^n$ that minimizes the weighted sum of distances to these points:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) = \sum_{i=1}^{m} \omega_i \|\mathbf{x} - \mathbf{a}_i\| \right\}$$

4.7.6 Weiszfeld's Method

Weiszfeld's method is used to solve the Fermat-Weber problem iteratively. Starting with an initial guess $\mathbf{x}_0$, the update at each iteration k is given by:

$$\mathbf{x}_{k+1} = T(\mathbf{x}_k) = \frac{\sum_{i=1}^{m} \frac{\omega_i \mathbf{a}_i}{\|\mathbf{x}_k - \mathbf{a}_i\|}}{\sum_{i=1}^{m} \frac{\omega_i}{\|\mathbf{x}_k - \mathbf{a}_i\|}}$$

This method is a fixed-point iteration and can be seen as a gradient method with a specific step size determined by the weights and distances.

5 Newton's Method

5.1 Former Newton's Method

Newton's method is widely known as a technique for finding a root of a univariate function. Let $\phi(\cdot) : \mathbb{R} \to \mathbb{R}$. Consider the equation

$$\phi(t^*) = 0$$

Assume that we know some $t \in \mathbb{R}$ which is close to t^*. Note that

$$\phi(t + \Delta t) = \phi(t) + \phi'(t)\Delta t + o(|\Delta t|).$$

Therefore, the solution of the equation $\phi(t + \Delta t) = 0$ can be approximated by the solution of the following linear equation:

$$\phi(t) + \phi'(t)\Delta t = 0$$

$$\Delta = -\frac{\phi(t)}{\phi'(t)}$$

We expect Δt to be a good approximation to $\Delta t^* = t^* - t$

$$t_{k+1} = t_k - \frac{\phi(t_k)}{\phi'(t_k)}$$

5.2 Definition

In the unconstrained minimization problem, we want to find a root of the nonlinear system

$$\nabla f(x) = 0$$

Thus replace $f(x)$ with $\nabla f(x)$, the Newton system is as followes:

$$\nabla f(x) + \nabla^2 f(x)\Delta x = 0$$

The objective is to find an optimal solution of the problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\},$$

If $\nabla^2 f(\mathbf{x}_k) \succ 0$,

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\nabla^2 f(\mathbf{x}_k)\right)^{-1} \nabla f(\mathbf{x}_k)$$

The vector $-\left(\nabla^2 f(\mathbf{x}_k)\right)^{-1} \nabla f(\mathbf{x}_k)$ is called **Newton's direction**.

5.3 Convergence of Newton' s method

Let us derive the local rate of convergence of Newton's Method. Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

under the following assumptions:

1. $f \in C_M^{2,2}(\mathbb{R}^n)$.
2. There exists a local minimum of the function f with positive definite Hessian:

$$\nabla^2 f(x^*) \succeq \mu I_n, \quad \mu > 0. \tag{(1)}$$

3. Our starting point x_0 is close enough to x^*.

Consider the process $x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$. Then, using the same reasoning as for the Gradient Method, we obtain the following representation:

$$x_{k+1} - x^* = x_k - x^* - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

$$= x_k - x^* - [\nabla^2 f(x_k)]^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x_k - x^*))(x_k - x^*) d\tau$$

$$= [\nabla^2 f(x_k)]^{-1} G_k (x_k - x^*),$$

where $G_k = \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))] d\tau$.
Let $r_k = \|x_k - x^*\|$. Then

$$\|G_k\| = \left\| \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))] d\tau \right\|$$

$$\leq \int_0^1 \|\nabla^2 f(x_k) - \nabla^2 f(x^* + \tau(x_k - x^*))\| d\tau$$

$$\leq \int_0^1 M(1-\tau) r_k d\tau = \frac{r_k M}{2}.$$

In view of relation (1), we have

$$\nabla^2 f(x_k) \succeq \nabla^2 f(x^*) - M r_k I_n \succeq (\mu - M r_k) I_n.$$

Therefore, if $r_k < \frac{\mu}{M}$, then $\nabla^2 f(x_k)$ is positive definite and

$$\|[\nabla^2 f(x_k)]^{-1}\| \leq (\mu - M r_k)^{-1}.$$

Hence, for r_k small enough ($r_k \leq \frac{2\mu}{3M}$), we have

$$r_{k+1} \leq \frac{M r_k^2}{2(\mu - M r_k)} \quad (\leq r_k).$$

The rate of convergence of this type is called *quadratic*.

5.3.1 Theorem

Let the function $f(\cdot)$ satisfy our assumptions. Suppose that the initial starting point x_0 is close enough to x^*:

$$\|x_0 - x^*\| \leq \bar{r} = \frac{2\mu}{3M}.$$

Then $\|x_k - x^*\| \leq \bar{r}$ for all k and Newton's Method converges quadratically:

$$\|x_{k+1} - x^*\| \leq \frac{M\|x_k - x^*\|^2}{2(\mu - M\|x_k - x^*\|)}.$$

5.4 Drawbacks of Newton's Method

- It can break down if $\nabla^2 f(x_k)$ is degenerate.
- Newton's process can diverge

5.4.1 Example

Let us apply Newton's Method for finding a root of the following univariate function:

$$\phi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Clearly, $t^* = 0$. Note that

$$\phi'(t) = \frac{1}{(1+t^2)^{3/2}}.$$

Therefore, Newton's process is as follows:

$$t_{k+1} = t_k - \frac{\phi(t_k)}{\phi'(t_k)} = t_k - \frac{t_k}{\sqrt{1+t_k^2}} \cdot (1+t_k^2)^{3/2} = -t_k^3.$$

Thus, if $|t_0| < 1$, then this method converges, and the convergence is extremely fast. The points ±1 are oscillation points of this scheme. If $|t_0| > 1$, then the method diverges.

5.5 Damped Newton's Method

In order to avoid a possible divergence, in practice we can apply the *damped Newton's method*:

$$\boxed{x_{k+1} = x_k - h_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k),}$$

where $h_k > 0$ is a step size parameter. At the initial stage of the method, we can use the same step size strategies as for the gradient scheme. At the final stage, it is reasonable to choose $h_k = 1$.

6 Convex Sets

6.1 Definition

A set $C \subseteq \mathbb{R}^n$ is called convex if for any $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0,1]$, the point $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ belongs to C.

6.2 Theorem (Convexity Preservation)

1. Let $C_1, C_2, \dots, C_k \subseteq \mathbb{R}^n$ be convex sets and let $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}$. Then the set $\mu_1 C_1 + \mu_2 C_2 + \cdots + \mu_k C_k$ is convex.

2. Let $C_i \subseteq \mathbb{R}^{k_i}$, $i = 1, 2, \dots, m$ be convex sets. Then the Cartesian product

$$C_1 \times C_2 \times \cdots \times C_m = \{(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) : \mathbf{x}_i \in C_i,\ i = 1, 2, \dots, m\}$$

   is convex.

3. Let $M \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then the set

$$\mathbf{A}(M) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in M\}$$

   is convex.

4. Let $D \subseteq \mathbb{R}^m$ be a convex set and let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then the set

$$\mathbf{A}^{-1}(D) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \in D\}$$

   is convex.

6.3 Theorem (Convex Combinations)

Let $C \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in C$. Then for any $\lambda \in \Delta_m$, the relation $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i \in C$ holds.

6.4 Definition (Convex Hull, 凸包)

Let $S \subseteq \mathbb{R}^n$. The convex hull of S, denoted by $\operatorname{conv}(S)$, is the set comprising all the convex combinations of vectors from S:

$$\operatorname{conv}(S) \equiv \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in S, \lambda \in \Delta_k, k \in \mathbb{N} \right\}.$$

6.5 Lemma

Let $S \subseteq \mathbb{R}^n$. If $S \subseteq T$ for some convex set T, then $\operatorname{conv}(S) \subseteq T$.

(The convex hull $\operatorname{conv}(S)$ is the “smallest” convex set containing S.)

6.6 Theorem (Caratheodory Theorem)

Let $S \subseteq \mathbb{R}^n$ and let $\mathbf{x} \in \operatorname{conv}(S)$. Then there exist $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n+1} \in S$ such that $\mathbf{x} \in \operatorname{conv}(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n+1}\})$, that is, there exist $\lambda \in \Delta_{n+1}$ such that

$$\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{x}_i.$$

Carathéodory 定理的核心思想：在 n-维空间中，任意一个在凸包中的点 x，都可以由最多 $n+1$ 个点的凸组合来表示。即使原集合 S 非常大，描述 x 只需要最多 $n+1$ 个点。

为什么是 n+1 个点？凸包在形成 d 维体时需要牺牲一个自由度，因此需要 $d+1$ 个点.

6.7 Definition (Cones, 锥)

A set S is called a *cone* if it satisfies the following property: For any $\mathbf{x} \in S$ and $\lambda \geq 0$, the inclusion $\lambda \mathbf{x} \in S$ is satisfied.

不难发现，锥一定包括原点。

6.8 Lemma

A set S is a convex cone if and only if the following properties hold:

1. $\mathbf{x}, \mathbf{y} \in S \implies \mathbf{x} + \mathbf{y} \in S$.
2. $\mathbf{x} \in S, \lambda \geq 0 \implies \lambda \mathbf{x} \in S$.

6.9 Examples

- The **convex polyhedron**

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq 0\}$$

  where $\mathbf{A} \in \mathbb{R}^{m \times n}$.

- The **Lorenz cone**, or **ice cream cone**, is given by

$$L^n = \left\{ \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \|\mathbf{x}\| \leq t, \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R} \right\}.$$

- **Nonnegative polynomials**: set consisting of all possible coefficients of polynomials of degree $n-1$ which are nonnegative over $\mathbb{R}$:

$$K^n = \left\{ \mathbf{x} \in \mathbb{R}^n : x_1 t^{n-1} + x_2 t^{n-2} + \cdots + x_n \geq 0, \forall t \in \mathbb{R} \right\}.$$

6.10 Definition (Conic Combination)

Given m points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n$, a *conic combination* of these m points is a vector of the form

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \lambda_m \mathbf{x}_m,$$

where $\lambda \in \mathbb{R}_+^m$.

6.11 Three Kinds of Combination

- Affine: affine set: $\theta x + (1-\theta)x$
- Convex: convex set: $\theta \in [0,1]$
- Conic: convex cone: $\theta_1 x + \theta_2 y$

6.12 Definition (The Conic Hull, 锥包)

Let $S \subseteq \mathbb{R}^n$. Then the **conic hull** of S, denoted by cone(S), is the set comprising all the conic combinations of vectors from S:

$$\text{cone}(S) \equiv \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in S, \lambda \in \mathbb{R}_+^k \right\}.$$

注意锥包和凸包的区别。

6.13 Lemma

Let $S \subseteq \mathbb{R}^n$. If $S \subseteq T$ for some convex cone T, then cone(S) $\subseteq T$.

(The conic hull of a set S is the smallest convex cone containing S.)

6.14 Theorem (Conic Representation Theorem)

Let $S \subseteq \mathbb{R}^n$ and let $\mathbf{x} \in \text{cone}(S)$. Then there exist k linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in S$ such that $\mathbf{x} \in \text{cone}(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\})$, that is, there exist $\lambda \in \mathbb{R}_+^k$ such that

$$\mathbf{x} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i.$$

In particular, $k \leq n$.

6.15 Definition (Basic Feasible Solutions of LP)

Consider the convex polyhedron.

$$P = \{x \in \mathbb{R}^n : Ax = b,\, x \geq 0\}, \quad (A \in \mathbb{R}^{m\times n},\, b \in \mathbb{R}^m)$$

The rows of $\mathbf{A}$ are assumed to be linearly independent. The above is a standard formulation of the constraints of a linear programming problem.

$\bar{\mathbf{x}}$ is a basic feasible solution (bfs) of P if the columns of $\mathbf{A}$ corresponding to the indices of the positive values of $\bar{\mathbf{x}}$ are linearly independent. 线性无关性保证了这个解是一个极端点（extreme point）

6.16 Proof

If $P \neq \emptyset$, then there exists $\mathbf{x} \in P$, meaning that $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq 0$. This implies that $\mathbf{b} \in \text{cone}(\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\})$, where $\mathbf{a}_i$ denotes the i-th column of $\mathbf{A}$.

By the conic representation theorem, there exist indices $i_1 < i_2 < \cdots < i_k$ and k numbers $y_{i_1}, y_{i_2}, \dots, y_{i_k} \geq 0$ such that

$$\mathbf{b} = \sum_{j=1}^{k} y_{i_j} \mathbf{a}_{i_j} \quad \text{and} \quad \mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \dots, \mathbf{a}_{i_k} \text{ are linearly independent.}$$

Now, define $\bar{\mathbf{x}} = \sum_{j=1}^{k} y_j \mathbf{e}_{i_j}$, where $\mathbf{e}_{i_j}$ is the i_j-th standard unit vector in $\mathbb{R}^n$. Clearly, $\bar{\mathbf{x}} \geq 0$ and

$$\mathbf{A}\bar{\mathbf{x}} = \sum_{j=1}^{k} y_j \mathbf{A}\mathbf{e}_{i_j} = \sum_{j=1}^{k} y_j \mathbf{a}_{i_j} = \mathbf{b}.$$

Therefore, $\bar{\mathbf{x}}$ is contained in P, and since the columns of $\mathbf{A}$ corresponding to the positive components of $\bar{\mathbf{x}}$ are linearly independent, it follows that $\bar{\mathbf{x}}$ is a basic feasible solution.

6.17 Theorem

Let $C \subseteq \mathbb{R}^n$ be a convex set. Then $\text{cl}(C)$ is a convex set.

6.18 Proof

Let $\mathbf{x}, \mathbf{y} \in \text{cl}(C)$ and let $\lambda \in [0, 1]$.

Since $\mathbf{x}, \mathbf{y} \in \text{cl}(C)$, there exist sequences $\{\mathbf{x}_k\}_{k \geq 0} \subseteq C$ and $\{\mathbf{y}_k\}_{k \geq 0} \subseteq C$ such that $\mathbf{x}_k \to \mathbf{x}$ and $\mathbf{y}_k \to \mathbf{y}$ as $k \to \infty$.

Define $\mathbf{z}_k = \lambda \mathbf{x}_k + (1 - \lambda)\mathbf{y}_k$. Since C is convex, we know that:

$$\mathbf{z}_k \in C \quad \forall k.$$

Next, we compute the limit of $\mathbf{z}_k$:

$$\lim_{k \to \infty} \mathbf{z}_k = \lambda \mathbf{x} + (1 - \lambda)\mathbf{y}.$$

Since $C \subseteq \text{cl}(C)$ and $\text{cl}(C)$ is closed, the limit of the sequence $\mathbf{z}_k \in C$ also lies in $\text{cl}(C)$. Therefore:

$$\lambda \mathbf{x} + (1 - \lambda)\mathbf{y} \in \text{cl}(C).$$

Thus, $\text{cl}(C)$ is convex.

6.19 Theorem (The Line Segment Principle)

Let C be a convex set and assume that $\text{int}(C) \neq \emptyset$. Suppose that $\mathbf{x} \in \text{int}(C)$ and $\mathbf{y} \in \text{cl}(C)$. Then $(1 - \lambda)\mathbf{x} + \lambda \mathbf{y} \in \text{int}(C)$ for any $\lambda \in [0, 1)$.

6.20 Theorem (Convexity of the Interior)

Let $C \subseteq \mathbb{R}^n$ be a convex set. Then $\text{int}(C)$ is convex.

6.21 Lemma (Combination of Closure and Interior)

Let C be a convex set with a nonempty interior. Then

1. $\text{cl}(\text{int}(C)) = \text{cl}(C)$.
2. $\text{int}(\text{cl}(C)) = \text{int}(C)$.

6.22 Theorem (Compactness of the Convex Hull of a Compact Set)

Let $S \subseteq \mathbb{R}^n$ be a compact set. Then $\text{conv}(S)$ is compact.

6.23 Definition (Extreme Points)

Let $S \subseteq \mathbb{R}^n$ be a convex set. A point $\mathbf{x} \in S$ is called an extreme point of S if there do not exist $\mathbf{x}_1, \mathbf{x}_2 \in S$ ($\mathbf{x}_1 \neq \mathbf{x}_2$) and $\lambda \in (0,1)$, such that

$$\mathbf{x} = \lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2.$$

The set of extreme points is denoted by $\text{ext}(S)$.

For example, the set of extreme points of a convex polytope (bounded polyhedron) consists of all its vertices.

6.24 Theorem (Equivalence Between bfs’ s and Extreme Points)

Let $P = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ has linearly independent rows and $\mathbf{b} \in \mathbb{R}^m$. Then $\bar{\mathbf{x}}$ is a basic feasible solution of P if and only if it is an extreme point of P.

6.25 Theorem (Krein-Milman Theorem)

Let $S \subseteq \mathbb{R}^n$ be a compact convex set. Then

$$S = \text{conv}(\text{ext}(S)).$$

7 Convex Functions

7.1 Definition

A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is called *convex* (or *convex over* C) if

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

for any $\mathbf{x}, \mathbf{y} \in C$, $\lambda \in [0,1]$.

7.2 Examples

Affine Functions. $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + \mathbf{b}$, where $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}$.

Norms. $g(\mathbf{x}) = \|\mathbf{x}\|$.

7.3 Theorem (Jensen's Inequality)

Let $f : C \to \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^n$ is a convex set. Then for any $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in C$ and $\lambda \in \Delta_k$, the following inequality holds:

$$f\left(\sum_{i=1}^{k} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{k} \lambda_i f(\mathbf{x}_i).$$

想象一个两端含有极端值的凸函数。

7.4 Theorem (The Gradient Inequality)

Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \quad \text{for any} \quad \mathbf{x}, \mathbf{y} \in C. \tag{1}$$

7.4.1 Proof

Step 1: Using the definition of convex functions.

Let $t \in (0,1)$, and set $z = x + t(y-x)$. By the convexity of f, we have:

$$f(z) = f(x + t(y-x)) \leq t f(y) + (1-t) f(x).$$

Step 2: Using the differentiability of f at x.

Since f is differentiable at x, $f(z)$ can be expanded as:

$$f(z) = f(x + t(y-x)) = f(x) + \nabla f(x)^\top (t(y-x)) + o(t).$$

Thus:

$$f(x + t(y-x)) = f(x) + t \nabla f(x)^\top (y-x) + o(t).$$

Step 3: Combining Step 1 and Step 2.

From convexity:

$$f(x) + t\nabla f(x)^\top (y - x) + o(t) \leq tf(y) + (1 - t)f(x).$$

Rearranging terms:

$$t\nabla f(x)^\top (y - x) + o(t) \leq tf(y) - tf(x).$$

That is:

$$t\nabla f(x)^\top (y - x) + o(t) \leq t(f(y) - f(x)).$$

Step 4: Eliminating t and taking the limit.

Divide through by t (noting $t > 0$) to get:

$$\nabla f(x)^\top (y - x) + \frac{o(t)}{t} \leq f(y) - f(x).$$

As $t \to 0^+$, $\frac{o(t)}{t} \to 0$. Therefore:

$$\nabla f(x)^\top (y - x) \leq f(y) - f(x).$$

Reorganizing, we obtain:

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

7.5 Proposition

Let f be a continuously differentiable function which is convex over a convex set $C \subseteq \mathbb{R}^n$. Suppose that $\nabla f(\mathbf{x}^*) = 0$ for some $\mathbf{x}^* \in C$. Then $\mathbf{x}^*$ is the global minimizer of f over C.

7.6 Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function given by $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then f is (strictly) convex if and only if $\mathbf{A} \succeq 0$ ($\mathbf{A} \succ 0$).

7.7 Theorem (Monotonicity of the Gradient)

Suppose that f is a continuously differentiable function over a convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \geq 0 \quad \text{for any} \quad \mathbf{x}, \mathbf{y} \in C.$$

一个凸函数的梯度总是“指向更高的位置”，或者说梯度的变化不会与路径的方向相反。

7.8 Theorem (Second-Order Characterization of Convexity)

Suppose that f is a twice continuously differentiable function over an open convex set $C \subseteq \mathbb{R}^n$. Then f is convex over C if and only if

$$\nabla^2 f(\mathbf{x}) \succeq 0 \quad \text{for any} \quad \mathbf{x} \in C.$$

7.9 Theorem (Operations Preserving Convexity)

- Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$ and let $\alpha \geq 0$. Then αf is a convex function over C.
- Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $C \subseteq \mathbb{R}^n$. Then the sum function $f_1 + f_2 + \cdots + f_p$ is convex over C.
- Let f be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the function g defined by
$$g(\mathbf{y}) = f(\mathbf{A}\mathbf{y} + \mathbf{b})$$
is convex over the convex set $D = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y} + \mathbf{b} \in C\}$.

7.10 Definition (Perspective Transformation)

If $f : \mathbb{R}^n \to \mathbb{R}$, then the **perspective** of f is the function $g : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$g(x, t) = t f(x/t),$$

with domain

$$\mathbf{dom}\ g = \{(x, t) \mid x/t \in \mathbf{dom}\ f,\ t > 0\}.$$

“Perspective” 一词来源于几何直观：想象 $\mathbf{x}$ 在一个“空间”中，t 相当于焦距或相机的视角距离；当我们固定一个正的 t 时，就相当于在“深度” t 的平面上看 $\mathbf{x}$ 的坐标，从而得到“透视”后的值。

7.10.1 透视变换可以把非凸函数转换成凸函数

考虑以下非凸约束：

$$\|x\|^2 \leq z^2, \quad z > 0.$$

这个约束是非凸的，因为 $\|x\|^2 = z^2$ 是一个锥体的边界，而 $z > 0$ 的条件使其不对称。

通过透视变换凸化：

引入一个新变量 $t > 0$，定义：

$$\|x\|^2 \leq t^2 z, \quad t > 0.$$

转换为：

$$\|x\| \leq \sqrt{tz}, \quad t > 0,$$

该约束可以转化为 second-order cone 形式（SOC），从而成为凸约束。

7.11 Theorem (Convexity of Perspective Functions)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function such that $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Then, its perspective transformation

$$h(\mathbf{x}, t) = \frac{f(\mathbf{x})}{t}, \quad \text{for} \quad t > 0$$

is a convex function over the set $\{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{>0}\}$.

Furthermore, if $t(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + d$ is an affine function that is strictly positive over a domain $D \subset \mathbb{R}^n$, then the composite function

$$g(\mathbf{x}) = \frac{f(\mathbf{x})}{t(\mathbf{x})} = \frac{f(\mathbf{x})}{\mathbf{c}^\top \mathbf{x} + d}$$

is convex over the domain D.

7.12 Theorem (Preservation of Convexity under Composition)

Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$.

Let $g : I \to \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $I \subseteq \mathbb{R}$. Assume that the image of C under f is contained in $I : f(C) \subseteq I$.

Then the composition of g with f defined by $h(\mathbf{x}) \equiv g(f(\mathbf{x}))$ is convex over C.

7.13 Theorem (Point-Wise Maximum of Convex Functions)

Let $f_1, f_2, \dots, f_p : C \to \mathbb{R}$ be p convex functions over the convex set $C \subseteq \mathbb{R}^n$. Then the maximum function

$$f(\mathbf{x}) \equiv \max_{i=1,2,\dots,p} \{f_i(\mathbf{x})\}$$

is convex over C.

7.14 Theorem (Preservation of Convexity Under Partial Minimization)

Let $f : C \times D \to \mathbb{R}$ be a convex function defined over the set $C \times D$ where $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$ are convex sets.

Let

$$g(\mathbf{x}) = \min_{\mathbf{y} \in D} f(\mathbf{x}, \mathbf{y}), \quad \mathbf{x} \in C$$

where we assume that the minimum is finite. Then g is convex over C.

7.15 Definition (Level Sets)

Let $f : S \to \mathbb{R}$ be a function defined over a set $S \subseteq \mathbb{R}^n$.

Then the level set of f with level α is given by

$$\operatorname{Lev}(f, \alpha) = \{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}.$$

7.16 Theorem (Level Sets)

Let $f : C \to \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the level set $\mathrm{Lev}(f, \alpha)$ is convex.

f convex $\implies C_\alpha$ convex.

反过来说并不成立，可以想像一个向下的丁丁状函数。

7.17 Definition (Quasi-Convex Functions)

A function $f : C \to \mathbb{R}$ defined over the convex set $C \subseteq \mathbb{R}^n$ is called quasi-convex if for any $\alpha \in \mathbb{R}$ the set $\mathrm{Lev}(f, \alpha)$ is convex.

Or

A function $f : C \to \mathbb{R}$ is called a quasi-convex function on the convex set $C \subseteq \mathbb{R}^n$ if, for any $\mathbf{x}, \mathbf{y} \in C$ and $t \in [0, 1]$, the following inequality holds:

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}.$$

7.18 Theorem (Lipschitz continuous on interior points)

Let $f : C \to \mathbb{R}$ be a convex function defined over a convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \mathrm{int}(C)$. Then there exist $\epsilon > 0$ and $L > 0$ such that $B[\mathbf{x}_0, \epsilon] \subseteq C$ and

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \leq L\|\mathbf{x} - \mathbf{x}_0\| \quad \text{for any} \quad \mathbf{x} \in B[\mathbf{x}_0, \epsilon].$$

7.19 Theorem (Existence of Directional Derivatives of Convex Functions)

Let $f : C \to \mathbb{R}$ be a convex function over the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \mathrm{int}(C)$. Then for any $\mathbf{d} \neq \mathbf{0}$, the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists.

7.20 Properties of Extended-Valued Functions

The effective domain of an extended real-valued function is the set of vectors for which the function takes a real value:

$$\mathrm{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < \infty\}.$$

An extended real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is called proper if it is not always equal to infinity, meaning that there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $f(\mathbf{x}_0) < \infty$.

An extended real-valued function is convex if and only if $\mathrm{dom}(f)$ is a convex set and the restriction of f to its effective domain is a convex real-valued function over $\mathrm{dom}(f)$.

7.21 Definition (The Epigraph)

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$. Then its epigraph $\mathrm{epi}(f) \subseteq \mathbb{R}^{n+1}$ is defined to be the set

$$\mathrm{epi}(f) = \{(\mathbf{x}, t) : f(\mathbf{x}) \leq t\}.$$

An extended real-valued function f is convex if and only if its epigraph set $\mathrm{epi}(f)$ is convex.

7.22 Proposition

If $\operatorname{epi} f$ is a convex set, then at some point $(x_0, f(x_0))$, we can find a supporting hyperplane:

$$a^T(x - x_0) + b[t - f(x_0)] \leq 0,$$

which holds for all (x, t) satisfying $t \geq f(x)$. Here, $(a, b) \in \mathbb{R}^n \times \mathbb{R}$ is the normal vector of the hyperplane (nonzero).

7.23 Theorem (Preservation of Convexity Under Supremum)

Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be an extended real-valued convex function for any $i \in I$ (I being an arbitrary index set). Then the function $f(\mathbf{x}) = \sup_{i \in I} f_i(\mathbf{x})$ is an extended real-valued convex function.

7.24 Theorem (Maximum of a Convex Function over a Compact Convex Set)

Let $f : C \to \mathbb{R}$ be convex and continuous over the nonempty convex and compact set $C \subseteq \mathbb{R}^n$. Then there exists at least one maximizer of f over C that is an extreme point of C.

7.25 Conclusion of Convexity Check

- 定义法 or $epi(f)$ or slices
- 二阶导法
- Convexity preservation (f is convex)
  - 非负加权 $\sum \alpha_i f_j$
  - Composition $f \circ g$ with non-decreasing f and convex g or g is affine
  - 最大值 $\max f_i$
  - 固定参数最大化 $g(x) = \sup_{y \in C} f(x, y)$, C is arbitary
  - 固定参数最小化 when f is jointly convex in both x and y, and C is convex. $g(x) = \inf_{y \in C} f(x, y)$

7.26 Conjugate Function

Let $f : \mathbb{R}^n \to \mathbf{R}$. The function $f^* : \mathbb{R}^n \to \mathbf{R}$, defined as

$$f^*(y) = \sup_{x \in \operatorname{dom} f} \left(y^T x - f(x)\right),$$

is called the *conjugate* of the function f. The domain of the conjugate function consists of $y \in \mathbb{R}^n$ for which the supremum is finite, *i.e.*, for which the difference $y^T x - f(x)$ is bounded above on $\operatorname{dom} f$. This definition is illustrated in figure.

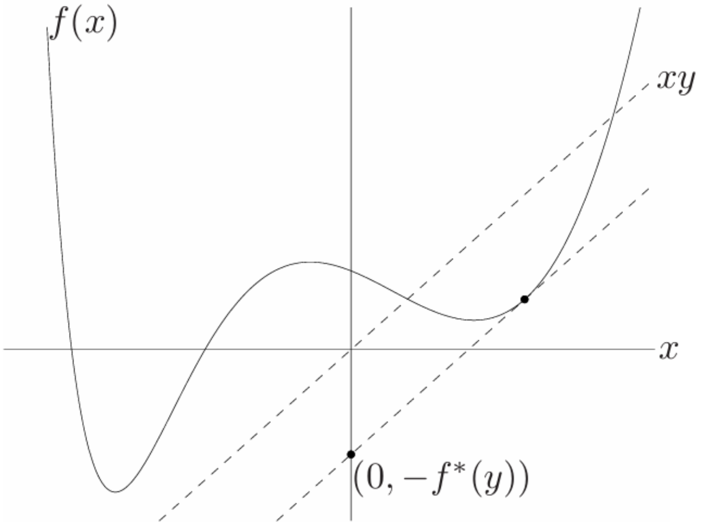


Figure 3.8 A function $f : \mathbf{R} \to \mathbf{R}$, and a value $y \in \mathbf{R}$. The conjugate function $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line in the figure. If f is differentiable, this occurs at a point x where $f'(x) = y$.

8 Convex Optimization

8.1 Definition

A convex optimization problem (or just a convex problem) is a problem consisting of minimizing a convex function over a convex set:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in C \end{aligned} \tag{1}$$

- C: convex set.
- f: convex function over C.

A functional form of a convex problem can be written as

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, p. \end{aligned}$$

$$f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R} \quad \text{are convex functions and} \quad h_1, h_2, \dots, h_p : \mathbb{R}^m \to \mathbb{R} \quad \text{are affine functions.}$$

8.2 Theorem (Global Minimum of Convex Function)

Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set $C \subseteq \mathbb{R}^n$. Let $\mathbf{x}^* \in C$ be a local minimum of f over C. Then $\mathbf{x}^*$ is a global minimum of f over C.

8.3 Theorem

Let $f : C \to \mathbb{R}$ be a strictly convex function defined on the convex set C. Let $\mathbf{x}^* \in C$ be a local minimum of f over C. Then $\mathbf{x}^*$ is a strict global minimum of f over C.

8.4 Theorem (Optimal Solution)

Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set $C \subseteq \mathbb{R}^n$. Then the set of optimal solutions of the problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

is convex. If, in addition, f is strictly convex over C, then there exists at most one optimal solution of the problem.

Proof

第一部分：最优解集是凸集

目标： 证明最优解集 $S = \{\mathbf{x} \in C : f(\mathbf{x}) = f^*\}$ 是凸集，其中 $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in C\}$。

证明：

1. **定义最优解集：**
   设 $S = \{\mathbf{x} \in C : f(\mathbf{x}) = f^*\}$，其中 $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in C\}$。

2. **取任意两点：**
   取 $\mathbf{x}_1, \mathbf{x}_2 \in S$。

3. **考虑任意组合：**
   对于任意 $\lambda \in [0, 1]$，设
   $$\mathbf{x}_\lambda = \lambda \mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2.$$
   由于 C 是凸集，故 $\mathbf{x}_\lambda \in C$。

4. **利用凸函数的定义：**
   由于 f 是凸函数，
   $$f(\mathbf{x}_\lambda) \leq \lambda f(\mathbf{x}_1) + (1 - \lambda) f(\mathbf{x}_2).$$
   因为 $\mathbf{x}_1, \mathbf{x}_2 \in S$，有 $f(\mathbf{x}_1) = f(\mathbf{x}_2) = f^*$，故
   $$f(\mathbf{x}_\lambda) \leq \lambda f^* + (1 - \lambda) f^* = f^*.$$

5. **确定最优性：**
   因为 f^* 是最小值，有
   $$f(\mathbf{x}_\lambda) \geq f^*.$$
   结合上述式子，有
   $$f(\mathbf{x}_\lambda) = f^*.$$
   因此，$\mathbf{x}_\lambda \in S$。

6. **结论：**
   任意两个最优解的凸组合仍然属于最优解集 S，因此 S 是凸集。

第二部分：严格凸时最优解唯一

目标： 如果 f 在 C 上严格凸，则优化问题

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

至多有一个最优解。

证明：

1. **假设存在两个不同的最优解：**
   假设存在 $\mathbf{x}_1, \mathbf{x}_2 \in S$ 且 $\mathbf{x}_1 \neq \mathbf{x}_2$。

2. **考虑中点：**
   设 $\lambda = \frac{1}{2}$，则
   $$\mathbf{x}_\lambda = \frac{1}{2}\mathbf{x}_1 + \frac{1}{2}\mathbf{x}_2.$$
   由于 C 是凸集，故 $\mathbf{x}_\lambda \in C$。

3. **利用严格凸的定义：**
   因为 f 是严格凸的，
   $$f(\mathbf{x}_\lambda) < \frac{1}{2}f(\mathbf{x}_1) + \frac{1}{2}f(\mathbf{x}_2).$$
   由于 $\mathbf{x}_1, \mathbf{x}_2 \in S$，有 $f(\mathbf{x}_1) = f(\mathbf{x}_2) = f^*$，故
   $$f(\mathbf{x}_\lambda) < \frac{1}{2}f^* + \frac{1}{2}f^* = f^*.$$

4. **矛盾：**
   但这与 f^* 是最小值相矛盾，因为 $\mathbf{x}_\lambda \in C$ 且
   $$f(\mathbf{x}_\lambda) < f^*,$$
   这意味着 $\mathbf{x}_\lambda$ 也是可行解且具有更小的函数值，与 f^* 为最小值的定义矛盾。

5. **结论：**
   因此，假设存在两个不同的最优解是不成立的，即在严格凸的情况下，最优解唯一。

8.5 Linear Programming

$$\begin{aligned} & \min \quad \mathbf{c}^\top \mathbf{x} \\ \text{(LP):} \quad & \text{s.t.} \quad \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \qquad\quad \mathbf{B}\mathbf{x} = \mathbf{g} \end{aligned}$$

When the feasible set C is compact (i.e., closed and bounded) and nonempty, **the Weierstrass theorem** guarantees that the convex optimization problem has at least one optimal solution.

8.6 Convex quadratic problems

$$\begin{aligned} \min \quad & \mathbf{x}^\top \mathbf{Q}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \le \mathbf{c} \end{aligned}$$

$$\mathbf{Q} \in \mathbb{R}^{n\times n} \quad \text{is positive semidefinite}, \quad \mathbf{b} \in \mathbb{R}^n, \quad \mathbf{A} \in \mathbb{R}^{m\times n}, \quad \mathbf{c} \in \mathbb{R}^m.$$

8.7 Conic Programming

Conic programming is a special type of convex optimization problem, whose standard form is usually written as

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^T x \\ \text{s.t.} \quad & Ax = b, \\ & x \in \mathcal{K}, \end{aligned}$$

where:

- $x \in \mathbb{R}^n$ is the decision variable;

- $c \in \mathbb{R}^n$ is the coefficient vector of the objective function;
- $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ define the linear equality constraints;
- $\mathcal{K}$ is a convex cone, meaning that if $x \in \mathcal{K}$ and $\alpha \geq 0$, then $\alpha x \in \mathcal{K}$, and for any $x, y \in \mathcal{K}$, we have $x + y \in \mathcal{K}$.

8.7.1 Uniformity and Generalization:

The form of conic programming problems covers many common convex optimization problems. For example:

- When $\mathcal{K} = \mathbb{R}^n_+$, the problem reduces to a standard **linear programming problem**.
- When $\mathcal{K}$ is a second-order cone, i.e.,

  $$\mathcal{Q} = \{(t, x) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x\|_2 \leq t\},$$

  the problem becomes a **second-order cone programming (SOCP) problem**.
- When $\mathcal{K}$ is a semidefinite cone, i.e.,

  $$\mathcal{S}^n_+ = \{X \in \mathbb{R}^{n \times n} \mid X = X^T, X \succeq 0\},$$

  the problem becomes a **semidefinite programming (SDP) problem**.

8.8 Quadratically Constrained Quadratic Problems:

$$\begin{aligned}
\text{(QCQP)} \quad \min \quad & \mathbf{x}^\top \mathbf{A}_0 \mathbf{x} + 2\mathbf{b}_0^\top \mathbf{x} + c_0 \\
\text{s.t.} \quad & \mathbf{x}^\top \mathbf{A}_i \mathbf{x} + 2\mathbf{b}_i^\top \mathbf{x} + c_i \leq 0, \quad i = 1, 2, \dots, m, \\
& \mathbf{x}^\top \mathbf{A}_j \mathbf{x} + 2\mathbf{b}_j^\top \mathbf{x} + c_j = 0, \quad j = m+1, m+2, \dots, m+p. \\
\mathbf{A}_0, \dots, \mathbf{A}_{m+p} - n \times n\, \text{symmetric}, \quad & \mathbf{b}_0, \dots, \mathbf{b}_{m+p} \in \mathbb{R}^n, \quad c_0, \dots, c_{m+p} \in \mathbb{R}.
\end{aligned}$$

QCQPs are not necessarily convex problems. When there are no equality constraints ($p = 0$) and all the matrices are positive semidefinite:

$\mathbf{A}_i \succeq 0, \quad i = 0, 1, \dots, m$, the problem is convex, and is therefore called a convex QCQP.

8.9 Relaxation of QCQP

8.10 Robust LP

Robust Linear Programming (Robust LP) mainly deals with situations where parameters in linear programming (LP) problems are uncertain. The general form is:

$$\min_x \quad c^T x$$

$$\text{s.t.} \quad a_i^T x \le b_i, \quad i = 1, \ldots, m$$

where a_i are **uncertain parameters**.

To ensure that the decision variable x satisfies the constraints under all possible scenarios, we typically consider the **worst-case** scenario for each constraint:

$$\max_{a_i \in \mathcal{E}_i} a_i^T x \le b_i,$$

which means we need to find the most difficult-to-satisfy coefficient a_i in the given set $\mathcal{E}_i$.

8.10.1 Example 1 When the uncertainty set is an ellipsoid

When the uncertainty set is an ellipsoid, it is typically defined as:

$$\mathcal{E}_i = \{a_i \mid a_i = \bar{a}_i + P_i u, \ \|u\|_2 \le 1\}$$

where:

- $\bar{a}_i$ is the nominal value of the parameter;
- P_i is the matrix defining the uncertainty ellipsoid;
- u is an arbitrary unit vector within a unit ball, describing the range of parameter uncertainty.

Through computation, we obtain:

$$\max_{\|u\|_2 \le 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2.$$

Thus, the original problem can be reformulated as a **second-order cone** constraint:

$$\bar{a}_i^T x + \|P_i^T x\|_2 \le b_i.$$

8.10.2 Example 2 Probability-based constraints

Consider the following probabilistic problem:

$$\min_x \quad c^T x$$

$$\text{s.t.} \quad \mathbb{P}(a_i^T x \le b_i) \ge \eta.$$

Assume that the random variable a_i follows a normal distribution:

$$a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i).$$

This means

$$\mathbb{P}\left(\frac{b_i - \bar{a}_i^T x}{\sqrt{x^T \Sigma_i x}} \ge z \right) = \Phi(z) \ge \eta.$$

Rearranging, we obtain the deterministic equivalent form of the probabilistic constraint:

$$\frac{b_i - \bar{a}_i^T x}{\sqrt{x^T \Sigma_i x}} \geq \Phi^{-1}(\eta),$$

which leads to:

$$\bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{\frac{1}{2}} x\|_2 \leq b_i.$$

Again, we obtain a **second-order cone** constraint.

8.11 Chebyshev Center Problem

Given m points $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ in $\mathbb{R}^n$. The objective is to find the center of the minimum radius closed ball containing all the points.

$$\begin{aligned} \min_{x,r} \quad & r \\ \text{s.t.} \quad & \mathbf{a}_i \in B[\mathbf{x}, r], \quad i = 1, 2, \dots, n \end{aligned}$$

8.12 Portfolio Selection Problem

We are given n assets numbered as $1, 2, \dots, n$. Let $Y_j\,(j = 1, 2, \dots, n)$ be the random variable representing the return from asset j.

Assume the expected returns are known:

$$\mu_j = E(Y_j), \quad j = 1, 2, \dots, n$$

and that the covariances of all the pairs of variables are also known:

$$\sigma_{i,j} = \mathrm{COV}(Y_i, Y_j), \quad i, j = 1, 2, \dots, n.$$

Let $x_j\,(j = 1, 2, \dots, n)$ be the proportion of the budget invested in asset j.

The decision variables are constrained to satisfy $\mathbf{x} \in \Delta_n$.

The overall return is the random variable:

$$R = \sum_{j=1}^{n} x_j Y_j,$$

whose expectation and variance are given by:

$$E(R) = \mu^\top \mathbf{x}, \quad V(R) = \mathbf{x}^\top \mathbf{C} \mathbf{x},$$

where

$$\mu = (\mu_1, \mu_2, \dots, \mu_n)^\top \quad \text{and} \quad \mathbf{C} \text{ is the covariance matrix:} \quad C_{i,j} = \sigma_{i,j}.$$

8.12.1 The Markowitz Model

Minimizing the risk under a minimal return level:

$$\begin{aligned} \min \quad & \mathbf{x}^\top \mathbf{C}\mathbf{x} \\ \text{s.t.} \quad & \mu^\top \mathbf{x} \geq \alpha, \quad \mathbf{e}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq 0 \end{aligned}$$

Maximize the expected return subject to a bounded risk constraint:

$$\begin{aligned} \max \quad & \mu^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^\top \mathbf{C}\mathbf{x} \leq \beta, \quad \mathbf{e}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq 0 \end{aligned}$$

A penalty approach:

$$\begin{aligned} \min \quad & -\mu^\top \mathbf{x} + \gamma \left(\mathbf{x}^\top \mathbf{C}\mathbf{x}\right) \\ \text{s.t.} \quad & \mathbf{e}^\top \mathbf{x} = 1, \quad \mathbf{x} \geq 0 \end{aligned}$$

8.13 The Orthogonal Projection Operator

Given a nonempty closed convex set C, the **orthogonal projection** operator $P_C : \mathbb{R}^n \to C$ is defined by

$$P_C(\mathbf{x}) = \arg\min\{\|\mathbf{y} - \mathbf{x}\|^2 : \mathbf{y} \in C\}.$$

8.14 Theorem (The First Projection Theorem)

Let $C \subseteq \mathbb{R}^n$ be a nonempty closed and convex set.
Then for any $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection $P_C(\mathbf{x})$ exists and is unique.

8.15 Linear Classification

Suppose that we are given two types of points in $\mathbb{R}^n$: type A and type B points.

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n \quad \text{- type A.}$$

$$\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \dots, \mathbf{x}_{m+p} \in \mathbb{R}^n \quad \text{- type B.}$$

The objective is to find a **linear separator**, which is a hyperplane of the form

$$H(\mathbf{w}, \beta) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} + \beta = 0\}$$

for which the type A and type B points are in its opposite sides:

$$\mathbf{w}^\top \mathbf{x}_i + \beta < 0, \quad i = 1, 2, \dots, m$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta > 0, \quad i = m+1, m+2, \dots, m+p$$

Underlying Assumption: The two sets of points are **linearly separable**, meaning that the set of inequalities has a solution.

可以通过原点来判断 Half-Space 方向，或者看 Half-Space 方向是和法向量 $\mathbf{a}$ 相反一侧的。

8.15.1 Lemma (Margin)

Let $H(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Let $\mathbf{y} \in \mathbb{R}^n$. Then the distance between $\mathbf{y}$ and the set H is given by

$$d(\mathbf{y}, H(\mathbf{a}, b)) = \frac{\left|\mathbf{a}^\top \mathbf{y} - b\right|}{\|\mathbf{a}\|}.$$

8.15.2 Mathematical Formulation

$$\max \left\{ \min_{i=1,2,\ldots,m+p} \frac{\left|\mathbf{w}^\top \mathbf{x}_i + \beta\right|}{\|\mathbf{w}\|} \right\}$$

$$\text{s.t.} \quad \mathbf{w}^\top \mathbf{x}_i + \beta < 0, \quad i = 1, 2, \ldots, m$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta > 0, \quad i = m+1, m+2, \ldots, m+p$$

The problem has a degree of freedom in the sense that if $(\mathbf{w}, \beta)$ is an optimal solution, then so is any nonzero multiplier of it, that is, $(\alpha\mathbf{w}, \alpha\beta)$ for $\alpha \neq 0$. We can therefore decide that

$$\min_{i=1,2,\ldots,m+p} \left|\mathbf{w}^\top \mathbf{x}_i + \beta\right| = 1$$

Problem Reformulation:

Thus, the problem can be written as:

$$\max \frac{1}{\|\mathbf{w}\|} \quad \text{s.t.} \quad \min_{i=1,2,\ldots,m+p} \left|\mathbf{w}^\top \mathbf{x}_i + \beta\right| = 1$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta < 0, \quad i = 1, 2, \ldots, m$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta > 0, \quad i = m+1, m+2, \ldots, m+p$$

Equivalent Quadratic Problem:

This can also be written as:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{s.t.} \quad \min_{i=1,2,\ldots,m+p} \left|\mathbf{w}^\top \mathbf{x}_i + \beta\right| = 1$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta \leq -1, \quad i = 1, 2, \ldots, m$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta \geq 1, \quad i = m+1, m+2, \ldots, m+p$$

Dropping Redundant Constraints:

Since the first constraint can be dropped, the equivalent form becomes:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{s.t.} \quad \mathbf{w}^\top \mathbf{x}_i + \beta \leq -1, \quad i = 1, 2, \ldots, m$$

$$\mathbf{w}^\top \mathbf{x}_i + \beta \geq 1, \quad i = m+1, m+2, \ldots, m+p$$

8.16 Trust Region Subproblem (TRS) Reformulation as a Convex Optimization Problem

8.16.1 Problem Definition: Trust Region Subproblem (TRS)

Problem (TRS):

$$\begin{aligned} \min \quad & \mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\mathbf{b}^\top \mathbf{x} + c \\ \text{subject to} \quad & \|\mathbf{x}\|^2 \le 1 \end{aligned}$$

where:

- $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a symmetric matrix.
- $\mathbf{b} \in \mathbb{R}^n$ is a vector.
- $c \in \mathbb{R}$ is a constant.

Nonconvexity: This problem is generally nonconvex unless $\mathbf{A}$ is positive semidefinite.

8.16.2 Application of the Spectral Decomposition Theorem

Spectral Decomposition: Any symmetric matrix $\mathbf{A}$ can be decomposed as:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top,$$

where:

- $\mathbf{U} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix.
- $\mathbf{D} = \operatorname{diag}(d_1, d_2, \dots, d_n)$ is a diagonal matrix containing the eigenvalues of $\mathbf{A}$.

Applying this decomposition to the TRS, the problem becomes:

$$\min \left\{ \mathbf{x}^\top \mathbf{U}\mathbf{D}\mathbf{U}^\top \mathbf{x} + 2\mathbf{b}^\top \mathbf{U}\mathbf{U}^\top \mathbf{x} + c : \|\mathbf{U}^\top \mathbf{x}\|^2 \le 1 \right\}.$$

8.16.3 Linear Change of Variables

Linear Transformation: Let $\mathbf{y} = \mathbf{U}^\top \mathbf{x}$. Then:

$$\|\mathbf{y}\|^2 = \|\mathbf{U}^\top \mathbf{x}\|^2 = \|\mathbf{x}\|^2 \le 1.$$

The problem in terms of $\mathbf{y}$ becomes:

$$\min \left\{ \mathbf{y}^\top \mathbf{D}\mathbf{y} + 2\mathbf{b}^\top \mathbf{U}\mathbf{y} + c : \|\mathbf{y}\|^2 \le 1 \right\}.$$

Simplification: Let $\mathbf{f} = \mathbf{U}^\top \mathbf{b}$. The problem simplifies to:

$$\min \sum_{i=1}^n d_i y_i^2 + 2\sum_{i=1}^n f_i y_i + c, \quad \text{s.t.} \quad \sum_{i=1}^n y_i^2 \le 1.$$

8.16.4 Normalization and Reformulation

Lemma: If $\mathbf{y}^*$ is an optimal solution, then $f_i y_i^* \leq 0$ for all $i = 1, 2, \ldots, n$.

Variable Transformation:

$$y_i = -\operatorname{sgn}(f_i)\sqrt{z_i}, \quad z_i \geq 0.$$

The optimization problem becomes:

$$\min \sum_{i=1}^{n} d_i z_i - 2\sum_{i=1}^{n} |f_i|\sqrt{z_i} + c$$

$$\text{s.t.} \quad \sum_{i=1}^{n} z_i \leq 1, \quad z_i \geq 0.$$

8.16.5 Convexity Analysis of the Reformulated Problem

Objective Function:

- The term $\sum_{i=1}^{n} d_i z_i$ is linear, hence convex.
- The term $-2\sum_{i=1}^{n} |f_i|\sqrt{z_i}$ is convex because $\sqrt{z_i}$ is concave, and the negative of a concave function is convex.

Constraints:

- The constraint $\sum_{i=1}^{n} z_i \leq 1$ is linear, hence defines a convex set.
- The nonnegativity constraints $z_i \geq 0$ are also linear and convex.

9 Optimization over a Convex Set

$$\begin{aligned} (P) \quad \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in C \end{aligned}$$

- C: closed convex subset of $\mathbb{R}^n$.
- f: continuously differentiable over C. Not necessarily convex.

9.1 Theorem (Stationarity or Optimal Condition)

Let f be a continuously differentiable function over a closed and convex set C. Then $\mathbf{x}^*$ is called a **stationary point** or **optimal point** of (P) if

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \text{for all} \quad \mathbf{x} \in C$$

Note that it's different from unconstrained optimization since constraints should be met. **The inequality ensures that the directional derivative is non-negative in all possible directions.**

9.2 Lemma

Let f be a continuously differentiable function over a nonempty closed convex set C, and let $\mathbf{x}^*$ be a **local minimum** of (P). Then $\mathbf{x}^*$ is a **stationary point** of (P).

9.3 Explicit Stationarity Condition

feasible set	explicit stationarity condition
$\mathbb{R}^n$	$\nabla f(\mathbf{x}^*) = 0$
$\mathbb{R}^n_+$	$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0 & \text{if } x_i^* > 0 \\ \geq 0 & \text{if } x_i^* = 0 \end{cases}$
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^\top \mathbf{x} = 1\}$	$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*)$
$B[0,1]$	$\nabla f(\mathbf{x}^*) = 0 \quad \text{or} \quad \|\mathbf{x}^*\| = 1$ and $\exists \lambda \leq 0 : \nabla f(\mathbf{x}^*) = \lambda \mathbf{x}^*$

9.4 Theorem (The Second Projection Theorem)

Let C be a nonempty closed convex set and let $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{z} = P_C(\mathbf{x})$ if and only if

$$(\mathbf{x} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \leq 0 \quad \text{for any} \quad \mathbf{y} \in C. \tag{1}$$

9.5 Theorem (Nonexpansivness)

Let C be a nonempty closed and convex set.

1. For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:

$$(P_C(\mathbf{v}) - P_C(\mathbf{w}))^\top (\mathbf{v} - \mathbf{w}) \geq \|P_C(\mathbf{v}) - P_C(\mathbf{w})\|^2 . \tag{2}$$

   Projections at different points do not deviate too much from the relative positions of their original points

2. **(non-expansiveness)** For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$:

$$\|P_C(\mathbf{v}) - P_C(\mathbf{w})\| \leq \|\mathbf{v} - \mathbf{w}\|. \tag{3}$$

   The projection operation does not ”magnify” the distance between two points

9.6 Theorem (Representation of Stationarity)

Let f be a continuously differentiable function over the nonempty closed convex set C, and let $s > 0$. Then $\mathbf{x}^*$ is a stationary point of

$$\begin{aligned} (P) \quad \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in C \end{aligned}$$

if and only if

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right).$$

9.7 Proof

Direction 1: If $\mathbf{x}^*$ is a stationary point, then $\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right)$.

A point $\mathbf{x}^*$ is a stationary point if:

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in C. \tag{1}$$

We want to show that:

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right).$$

The **Projection Theorem** states that for any point $\mathbf{v} \in \mathbb{R}^n$, the projection $\mathbf{z} = P_C(\mathbf{v})$ onto a closed convex set C satisfies:

$$(\mathbf{v} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \leq 0, \quad \forall \mathbf{y} \in C. \tag{2}$$

Assuming $\mathbf{x}^*$ is a stationary point, let:

$$\mathbf{v} = \mathbf{x}^* - s\nabla f(\mathbf{x}^*), \quad \mathbf{z} = \mathbf{x}^*.$$

We need to verify that $\mathbf{x}^*$ satisfies the projection condition. Substituting into (2), we get:

$$\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*) - \mathbf{x}^*\right)^\top (\mathbf{y} - \mathbf{x}^*) \leq 0, \quad \forall \mathbf{y} \in C.$$

The left-hand side simplifies to:

$$(-s\nabla f(\mathbf{x}^*))^\top(\mathbf{y}-\mathbf{x}^*).$$

Thus, the inequality becomes:

$$(-s\nabla f(\mathbf{x}^*))^\top(\mathbf{y}-\mathbf{x}^*) \leq 0.$$

Dividing both sides by $-s$ (note that $s > 0$), we obtain:

$$\nabla f(\mathbf{x}^*)^\top(\mathbf{y}-\mathbf{x}^*) \geq 0, \quad \forall \mathbf{y} \in C.$$

This is exactly the definition of a stationary point in (1). Therefore, $\mathbf{x}^*$ satisfies the projection condition, and we conclude that:

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right).$$

Direction 2: If $\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right)$, then $\mathbf{x}^*$ is a stationary point.

Assume:

$$\mathbf{x}^* = P_C\left(\mathbf{x}^* - s\nabla f(\mathbf{x}^*)\right).$$

Since $\mathbf{x}^*$ is the projection of $\mathbf{x}^* - s\nabla f(\mathbf{x}^*)$ onto C, the projection theorem implies:

$$(\mathbf{x}^* - s\nabla f(\mathbf{x}^*) - \mathbf{x}^*)^\top(\mathbf{y}-\mathbf{x}^*) \leq 0, \quad \forall \mathbf{y} \in C.$$

The left-hand side simplifies to:

$$(-s\nabla f(\mathbf{x}^*))^\top(\mathbf{y}-\mathbf{x}^*).$$

Thus, the inequality becomes:

$$(-s\nabla f(\mathbf{x}^*))^\top(\mathbf{y}-\mathbf{x}^*) \leq 0.$$

Dividing both sides by $-s$ (since $s > 0$), we get:

$$\nabla f(\mathbf{x}^*)^\top(\mathbf{y}-\mathbf{x}^*) \geq 0, \quad \forall \mathbf{y} \in C.$$

This is exactly the definition of a stationary point in (1). Therefore, $\mathbf{x}^*$ is a stationary point.

9.8 The Gradient Mapping

The gradient mapping is defined as

$$G_L(\mathbf{x}) = L\left[\mathbf{x} - P_C\left(\mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})\right)\right]$$

where $L > 0$.

In the unconstrained case $G_L(\mathbf{x}) = \nabla f(\mathbf{x})$.

$G_L(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x}$ is a stationary point of (P). This means that we can consider $\|G_L(\mathbf{x})\|^2$ to be an optimality measure.

9.9 Lemma (Sufficient decrease lemma for constrained problems)

Suppose that $f \in C_L^{1,1}(C)$ for some $L > 0$, where C is a closed convex set. Then for any $\mathbf{x} \in C$ and $t \in \left(0, \frac{2}{L}\right)$ the following inequality holds:

$$f(\mathbf{x}) - f\left(P_C\left(\mathbf{x} - t\nabla f(\mathbf{x})\right)\right) \geq t\left(1 - \frac{Lt}{2}\right)\left\|\frac{1}{t}\left(\mathbf{x} - P_C\left(\mathbf{x} - t\nabla f(\mathbf{x})\right)\right)\right\|^2.$$

9.9.1 Recall (Sufficient Decrease Lemma for unconstrained problems)

Suppose that $f \in C_L^{1,1}(D)$ for some $L > 0$. Then for any $\mathbf{x} \in \mathbb{R}^n$ and $t > 0$

$$f(\mathbf{x}) - f(\mathbf{x} - t\nabla f(\mathbf{x})) \geq t\left(1 - \frac{Lt}{2}\right)\|\nabla f(\mathbf{x})\|^2.$$

9.10 Convergence of the Gradient Projection Method

Let $\{\mathbf{x}_k\}$ be the sequence generated by the gradient projection method for solving problem (P) with either a constant stepsize $\bar{t} \in \left(0, \frac{2}{L}\right)$, where L is a Lipschitz constant of ∇f or a backtracking stepsize strategy. Assume that f is bounded below. Then

1. The sequence $\{f(\mathbf{x}_k)\}$ is nonincreasing.
2. $G_d(\mathbf{x}_k) \to 0$ as $k \to \infty$, where

$$d = \begin{cases} \frac{1}{t} & \text{constant stepsize,} \\ \frac{1}{s} & \text{backtracking.} \end{cases}$$

9.11 Theorem (Rate of convergence of the sequence of function values)

Consider the problem

$$(P) \quad \min \quad f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in C,$$

where C is a nonempty closed and convex set, and $f \in C_L^{1,1}(C)$ is convex over C. Let $\{\mathbf{x}_k\}_{k\geq 0}$ be generated by GPM for solving (P) with a constant stepsize $t_k = \bar{t} \in \left(0, \frac{1}{L}\right]$. Assume the set of optimal solutions X^* is nonempty, and let f^* be the optimal value of (P). Then,

1. For any $k \geq 0$ and $\mathbf{x}^* \in X^*$,

$$2\bar{t}\left(f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)\right) \leq \|\mathbf{x}_k - \mathbf{x}^*\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}^*\|^2,$$

2. For any $n \geq 1$:

$$f(\mathbf{x}_n) - f^* \leq \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|^2}{2\bar{t}n}.$$

9.12 Theorem (Convergence of the sequence generated by the gradient projection method)

Under the same setting of the previous theorem, the sequence $\{\mathbf{x}_k\}_{k\geq 0}$ generated by the gradient projection method with a constant stepsize $t_k = \bar{t} \in \left(0, \frac{1}{L}\right]$ converges to an optimal solution.

9.13 Sparsity Constrained Problems

The sparsity constrained problem is given by

$$(S) \quad \min \quad f(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq s,$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is a lower-bounded continuously differentiable function.
- $s > 0$ is an integer smaller than n.
- $\|\mathbf{x}\|_0$ is the ℓ_0 norm of $\mathbf{x}$, which counts the number of nonzero components in $\mathbf{x}$.
- We do not assume that f is a convex function. The constraint set is of course not convex.

9.13.1 Notation

- $\mathcal{I}_1(\mathbf{x}) \equiv \{i : x_i \neq 0\}$ - the support set.
- $\mathcal{I}_0(\mathbf{x}) \equiv \{i : x_i = 0\}$ - the off-support set.
- $C_s = \{\mathbf{x} : \|\mathbf{x}\|_0 \leq s\}$.
- For a vector $\mathbf{x} \in \mathbb{R}^n$ and $i \in \{1, 2, \dots, n\}$, the i-th largest absolute value component in $\mathbf{x}$ is denoted by $M_i(\mathbf{x})$.

9.14 Definition (Basic Feasibility)

A vector $\mathbf{x}^* \in C_s$ is called a basic feasible (BF) vector of (P) if:

1. when $\|\mathbf{x}^*\|_0 < s$, $\nabla f(\mathbf{x}^*) = 0$;
   $\mathbf{x}^*$ is a unconstrained critical point
2. when $\|\mathbf{x}^*\|_0 = s$, $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$ for all $i \in \mathcal{I}_1(\mathbf{x}^*)$.
   Only the gradients in the non-zero directions are required to be 0, ensuring that there is no room for further optimization in those directions

9.15 Theorem (BF is a necessary optimality condition)

Let $\mathbf{x}^*$ be an optimal solution of (P). Then $\mathbf{x}^*$ is a BF vector.

9.16 Definition (L-stationarity)

A vector $\mathbf{x}^* \in C_s$ is called an L-stationary point of (S) if it satisfies the relation

$$[\text{NC}_L] \quad \mathbf{x}^* \in P_{C_s}\left(\mathbf{x}^* - \frac{1}{L}\nabla f(\mathbf{x}^*)\right).$$

9.17 Lemma (Explicit Reformulation of L-stationarity)

For any $L > 0$, $\mathbf{x}^*$ satisfies [NC$_L$] if and only if $\|\mathbf{x}^*\|_0 \leq s$ and

$$\left|\frac{\partial f}{\partial x_i}(\mathbf{x}^*)\right| \begin{cases} \leq LM_s(\mathbf{x}^*) & \text{if } i \in \mathcal{I}_0(\mathbf{x}^*), \\ = 0 & \text{if } i \in \mathcal{I}_1(\mathbf{x}^*), \end{cases} \tag{6}$$

9.18 Theorem

Suppose that $f \in C^{1,1}_{L_f} \subseteq \mathbb{R}^n$, and that $L > L_f$. Let $\mathbf{x}^*$ be an optimal solution of (S). Then $\mathbf{x}^*$ is an L-stationary point.

9.19 The Iterative Hard-Thresholding (IHT) Method

Algorithm 2 The IHT method

```
Input: a constant $L \geq L_f$.
Initialization: Choose $\mathbf{x}_0 \in C_s$.
General step: $\mathbf{x}^{k+1} \in P_{C_s}\left(\mathbf{x}^k - \frac{1}{L}\nabla f(\mathbf{x}^k)\right), \quad (k = 0, 1, 2, \dots)$ =0
```

9.19.1 Theorem (convergence of IHT)

Suppose that $f \in C^{1,1}_{L_f}$ and let $\{\mathbf{x}^k\}_{k\geq 0}$ be the sequence generated by the IHT method with stepsize $\frac{1}{L}$ where $L > L_f$. Then any accumulation point of $\{\mathbf{x}^k\}_{k\geq 0}$ is an L-stationary point.

10 Optimality Conditions for Linearly Constrained Problems

10.1 Theorem (Separation of a point from a closed and convex set)

Let $C \subset \mathbb{R}^n$ be a nonempty closed and convex set, and let $\mathbf{y} \notin C$. Then there exists $\mathbf{p} \in \mathbb{R}^n \backslash \{0\}$ and $\alpha \in \mathbb{R}$ such that

$$\mathbf{p}^\top \mathbf{y} > \alpha \quad \text{and} \quad \mathbf{p}^\top \mathbf{x} \le \alpha \quad \text{for all } \mathbf{x} \in C.$$

10.1.1 Proof

By the second orthogonal projection theorem, the vector $\bar{\mathbf{x}} = P_C(\mathbf{y}) \in C$ satisfies

$$(\mathbf{y} - \bar{\mathbf{x}})^\top (\mathbf{x} - \bar{\mathbf{x}}) \le 0 \quad \text{for all } \mathbf{x} \in C$$

which is the same as

$$(\mathbf{y} - \bar{\mathbf{x}})^\top \mathbf{x} \le (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}} \quad \text{for all } \mathbf{x} \in C.$$

Denote $\mathbf{p} = \mathbf{y} - \bar{\mathbf{x}} \neq 0$ and $\alpha = (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}}$. Then

$$\mathbf{p}^\top \mathbf{x} \le \alpha \quad \text{for all } \mathbf{x} \in C.$$

On the other hand,

$$\mathbf{p}^\top \mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^\top \mathbf{y} = (\mathbf{y} - \bar{\mathbf{x}})^\top (\mathbf{y} - \bar{\mathbf{x}}) + (\mathbf{y} - \bar{\mathbf{x}})^\top \bar{\mathbf{x}} = \|\mathbf{y} - \bar{\mathbf{x}}\|^2 + \alpha > \alpha.$$

10.2 Lemma (Farkas Lemma)

Let $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then exactly one of the following systems has a solution:

(i) $\mathbf{A}\mathbf{x} \le 0, \quad \mathbf{c}^\top \mathbf{x} > 0$

(ii) $\mathbf{A}^\top \mathbf{y} = \mathbf{c}, \quad \mathbf{y} \ge 0$

Or

Let $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then the following claims are equivalent:

(i) The implication $\mathbf{A}\mathbf{x} \le 0 \Rightarrow \mathbf{c}^\top \mathbf{x} \le 0$ holds true.

(ii) There exists $\mathbf{y} \in \mathbb{R}^m_+$ such that $\mathbf{A}^\top \mathbf{y} = \mathbf{c}$.

10.3 Theorem (Gordans Alternative Theorem)

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then exactly one of the following systems has a solution:

(i) $\mathbf{A}\mathbf{x} < 0$

(ii) $\mathbf{p} \neq 0, \quad \mathbf{A}^\top \mathbf{p} = 0, \quad \mathbf{p} \ge 0$

10.4 Theorem (KKT conditions for linearly constrained problems - necessary optimality conditions)

Consider the minimization problem

$$\begin{aligned} (P) \quad & \min f(\mathbf{x}) \\ & \text{s.t. } \mathbf{a}_i^\top \mathbf{x} \le b_i, \quad i = 1, 2, \dots, m \end{aligned}$$

where f is continuously differentiable over $\mathbb{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$, $b_1, b_2, \dots, b_m \in \mathbb{R}$ and let $\mathbf{x}^*$ be a local minimum point of (P). Then there exist $\lambda_1, \lambda_2, \dots, \lambda_m \ge 0$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0 \tag{2}$$

and

$$\lambda_i (\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \dots, m \tag{3}$$

10.4.1 Proof

Step 1: Establish Directional Derivative Inequalities

Since $\mathbf{x}^*$ is a local minimum, for any feasible direction that satisfies the constraints, we have:

$$\nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \ge 0 \quad \text{for all } \mathbf{x} \text{ satisfying } \mathbf{a}_i^\top \mathbf{x} \le b_i.$$

This shows that in the feasible region, the directional derivative of the objective function at $\mathbf{x}^*$ is non-negative.

Step 2: Define the Set of Active Constraints

Define the set of active constraints at $\mathbf{x}^*$ as:

$$I(\mathbf{x}^*) = \{ i \mid \mathbf{a}_i^\top \mathbf{x}^* = b_i \}.$$

These are the constraints that are equalities at $\mathbf{x}^*$.

Step 3: Variable Transformation

Make the change of variables:

$$\mathbf{y} = \mathbf{x} - \mathbf{x}^*.$$

Thus, $\mathbf{y}$ represents a direction starting from $\mathbf{x}^*$.

Step 4: Rewrite the Inequalities

Since $\mathbf{x} = \mathbf{x}^* + \mathbf{y}$, substitute into the constraints:

- For $i \in I(\mathbf{x}^*)$:
  $\mathbf{a}_i^\top (\mathbf{x}^* + \mathbf{y}) \le b_i \Rightarrow \mathbf{a}_i^\top \mathbf{x}^* + \mathbf{a}_i^\top \mathbf{y} \le b_i \Rightarrow b_i + \mathbf{a}_i^\top \mathbf{y} \le b_i \Rightarrow \mathbf{a}_i^\top \mathbf{y} \le 0.$
- For $i \notin I(\mathbf{x}^*)$:
  $\mathbf{a}_i^\top (\mathbf{x}^* + \mathbf{y}) \le b_i \Rightarrow \mathbf{a}_i^\top \mathbf{x}^* + \mathbf{a}_i^\top \mathbf{y} \le b_i \Rightarrow \mathbf{a}_i^\top \mathbf{y} \le b_i - \mathbf{a}_i^\top \mathbf{x}^*.$

Since $\mathbf{a}_i^\top \mathbf{x}^* < b_i$ for $i \notin I(\mathbf{x}^*)$, it follows that $b_i - \mathbf{a}_i^\top \mathbf{x}^* > 0$.

Step 5: Simplify the Conditions

Thus, the inequalities can be simplified as:

$$\text{For all } \mathbf{y} \text{ satisfying } \mathbf{a}_i^\top \mathbf{y} \leq 0 \quad (i \in I(\mathbf{x}^*)), \quad \nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0.$$

Step 6: Remove the Influence of Inactive Constraints

For simplicity, we will consider only those directions $\mathbf{y}$ that satisfy $\mathbf{a}_i^\top \mathbf{y} \leq 0$ for $i \in I(\mathbf{x}^*)$ and ignore those with $i \notin I(\mathbf{x}^*)$.

Proof of Validity:

For $i \notin I(\mathbf{x}^*)$, $b_i - \mathbf{a}_i^\top \mathbf{x}^* > 0$. Therefore, we can find a small enough $\alpha > 0$ such that $\mathbf{a}_i^\top (\alpha \mathbf{y}) \leq b_i - \mathbf{a}_i^\top \mathbf{x}^*$.

Thus, as long as α is sufficiently small, $\alpha \mathbf{y}$ will not violate the constraint.

Therefore, we can focus on directions $\mathbf{y}$ that satisfy $\mathbf{a}_i^\top \mathbf{y} \leq 0$ for $i \in I(\mathbf{x}^*)$.

Step 7: Establish an Inequality for y

Now, we have:

$$\text{If } \mathbf{a}_i^\top \mathbf{y} \leq 0 \text{ for all } i \in I(\mathbf{x}^*), \text{ then } \nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0.$$

This means that for directions $\mathbf{y}$ satisfying $\mathbf{a}_i^\top \mathbf{y} \leq 0$ (for $i \in I(\mathbf{x}^*)$), the gradient $\nabla f(\mathbf{x}^*)^\top \mathbf{y}$ is non-negative.

Step 8: Apply Farkas' Lemma

Farkas' Lemma: For given $\mathbf{c} \in \mathbb{R}^n$ and matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, exactly one of the following two systems has a solution:

1. There exists $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{y} \leq 0$ and $\mathbf{c}^\top \mathbf{y} < 0$.
2. There exists $\lambda \geq 0$ such that $\mathbf{A}^\top \lambda = \mathbf{c}$.

In our case, let:

- $\mathbf{A}$ be a matrix with rows $\mathbf{a}_i^\top$ (for $i \in I(\mathbf{x}^*)$).
- $\mathbf{c} = -\nabla f(\mathbf{x}^*)$.

From the above, we have:

- For all $\mathbf{y}$ satisfying $\mathbf{a}_i^\top \mathbf{y} \leq 0$, $\nabla f(\mathbf{x}^*)^\top \mathbf{y} \geq 0$.

Applying Farkas' Lemma:

Since there does not exist a $\mathbf{y}$ such that $\mathbf{A}\mathbf{y} \leq 0$ and $\nabla f(\mathbf{x}^*)^\top \mathbf{y} < 0$, by Farkas' Lemma, there must exist $\lambda \geq 0$ (for $i \in I(\mathbf{x}^*)$) such that:

$$\mathbf{A}^\top \lambda = -\nabla f(\mathbf{x}^*).$$

This implies:

$$-\nabla f(\mathbf{x}^*) = \sum_{i \in I(\mathbf{x}^*)} \lambda_i \mathbf{a}_i.$$

Step 9: Define the Lagrange Multipliers

Extend λ for all $i = 1, 2, \ldots, m$ as follows:

- For $i \in I(\mathbf{x}^*)$, $\lambda_i \geq 0$ as derived.
- For $i \notin I(\mathbf{x}^*)$, set $\lambda_i = 0$.

Then, in total, we have:

$$-\nabla f(\mathbf{x}^*) = \sum_{i=1}^{m} \lambda_i \mathbf{a}_i,$$

or equivalently:

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0.$$

Step 10: Verify Complementary Slackness

For $i \notin I(\mathbf{x}^*)$:

$$\lambda_i = 0 \quad \text{and} \quad \mathbf{a}_i^\top \mathbf{x}^* < b_i.$$

Thus:

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0.$$

For $i \in I(\mathbf{x}^*)$:

$$\mathbf{a}_i^\top \mathbf{x}^* = b_i.$$

Therefore:

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = \lambda_i \times 0 = 0.$$

Step 11: Conclusion

In summary, we have proven the existence of $\lambda_i \geq 0$ (for $i = 1, 2, \dots, m$) such that:

- Stationarity condition:

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0$$

- Complementary slackness condition:

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \dots, m$$

This is the specific form of the KKT conditions in this problem.

10.5 Theorem (KKT conditions for convex linearly constrained problems - necessary and sufficient optimality conditions)

Consider the minimization problem

$$\begin{aligned} (P) \quad & \min f(\mathbf{x}) \\ & \text{s.t. } \mathbf{a}_i^\top \mathbf{x} \leq b_i, \quad i = 1, 2, \dots, m \end{aligned}$$

where f is a convex continuously differentiable function over $\mathbb{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$, $b_1, b_2, \dots, b_m \in \mathbb{R}$ and let $\mathbf{x}^*$ be a feasible solution of (P). Then $\mathbf{x}^*$ is an optimal solution if and only if there exist $\lambda_1, \lambda_2, \dots, \lambda_m \geq 0$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0 \tag{4}$$

and

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \dots, m \tag{5}$$

10.5.1 Proof

Necessity was proven.

Step 1: Define the Auxiliary Function $h(\mathbf{x})$

We define a new function $h : \mathbb{R}^n \to \mathbb{R}$ as:

$$h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i)$$

Explanation:

- This is similar to the structure of a Lagrangian function, but here, we add the constraint terms directly to the objective function rather than forming a Lagrangian (typically written as $L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i)$).
- Since $\lambda_i \geq 0$, we can directly add the constraint terms to the objective function f for subsequent analysis.

Step 2: Compute the Gradient of h at $\mathbf{x}^*$

We calculate the gradient of h at $\mathbf{x}^*$:

$$\nabla h(\mathbf{x}^*) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i$$

Using the stationarity condition (4):

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0,$$

we have:

$$\nabla h(\mathbf{x}^*) = 0$$

Conclusion:

- $\mathbf{x}^*$ is a stationary point of h (gradient is zero).
- Since f is convex and the constraints are linear (linear functions are also convex), h is a convex function.

Step 3: $\mathbf{x}^*$ is a Global Minimum of h

Reasoning:

- For convex functions, any stationary point is a global minimum.
- Since h is convex and $\nabla h(\mathbf{x}^*) = 0$, we have:

$$h(\mathbf{x}^*) \leq h(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$$

Step 4: Compare $h(\mathbf{x}^*)$ and $h(\mathbf{x})$

For any $\mathbf{x} \in \mathbb{R}^n$, we have:

$$h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i)$$

$$h(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i)$$

Step 5: Using the Complementary Slackness Condition

From the complementary slackness condition (5):

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0$$

Therefore, each constraint term in $h(\mathbf{x}^*)$ vanishes, so:

$$h(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = f(\mathbf{x}^*) + 0 = f(\mathbf{x}^*)$$

Step 6: Since $\mathbf{x}$ is Feasible

Because $\mathbf{x}$ satisfies the constraint conditions, we have $\mathbf{a}_i^\top \mathbf{x} - b_i \leq 0$, and since $\lambda_i \geq 0$, it follows that:

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i) \leq 0$$

Thus, for any feasible $\mathbf{x}$:

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i) \leq f(\mathbf{x})$$

Step 7: Combine the Above Results

From Steps 4 through 6, we have:

$$h(\mathbf{x}^*) = f(\mathbf{x}^*) \leq h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i(\mathbf{a}_i^\top \mathbf{x} - b_i) \leq f(\mathbf{x})$$

Therefore:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \quad \forall \mathbf{x} \text{ satisfying } \mathbf{a}_i^\top \mathbf{x} \leq b_i$$

10.6 Theorem (KKT conditions for linearly constrained problems)

Consider the minimization problem

$$\begin{aligned} (Q) \quad & \min f(\mathbf{x}) \\ & \text{s.t. } \mathbf{a}_i^\top \mathbf{x} \leq b_i, \quad i = 1, 2, \dots, m \\ & \qquad \mathbf{c}_j^\top \mathbf{x} = d_j, \quad j = 1, 2, \dots, p \end{aligned}$$

where f is continuously differentiable, $\mathbf{a}_i, \mathbf{c}_j \in \mathbb{R}^n$, $b_i, d_j \in \mathbb{R}$.

(i) (Necessity of the KKT conditions) If $\mathbf{x}^*$ is a local minimum of (Q), then there exist $\lambda_1, \lambda_2, \dots, \lambda_m \geq 0$ and $\mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i + \sum_{j=1}^{p} \mu_j \mathbf{c}_j = 0 \tag{6}$$

and

$$\lambda_i(\mathbf{a}_i^\top \mathbf{x}^* - b_i) = 0, \quad i = 1, 2, \ldots, m \tag{7}$$

(ii) (Sufficiency in the convex case) If f is convex over $\mathbb{R}^n$ and $\mathbf{x}^*$ is a feasible solution of (Q) for which there exist $\lambda_1, \ldots, \lambda_m \geq 0$ and $\mu_1, \ldots, \mu_p \in \mathbb{R}$ such that (6) and (7) are satisfied, then $\mathbf{x}^*$ is an optimal solution of (Q).

10.7 Lemma

Let C be the affine space

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Assume the rows of $\mathbf{A}$ are linearly independent. Then

$$P_C(\mathbf{y}) = \mathbf{y} - \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{b})$$

Consider the hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top \mathbf{x} = b\} \quad (0 \neq \mathbf{a} \in \mathbb{R}^n,\ b \in \mathbb{R})$$

Then by the previous slide:

$$P_H(\mathbf{y}) = \mathbf{y} - \mathbf{a}(\mathbf{a}^\top\mathbf{a})^{-1}(\mathbf{a}^\top\mathbf{y} - b) = \mathbf{y} - \frac{\mathbf{a}^\top\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}$$

10.8 Lemma (Distance of a point from a hyperplane)

Let $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^\top\mathbf{x} = b\}$, where $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

$$d(\mathbf{y}, H) = \frac{|\mathbf{a}^\top\mathbf{y} - b|}{\|\mathbf{a}\|}$$

10.9 Orthogonal Regression

Let $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.

For a given $0 \neq \mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$, we define the hyperplane:

$$H_{\mathbf{x},y} := \{\mathbf{a} \in \mathbb{R}^n : \mathbf{x}^\top\mathbf{a} = y\}$$

In the orthogonal regression problem, we seek to find a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ to $H_{\mathbf{x},y}$ is minimal:

$$\min_{\mathbf{x},y} \left\{ \sum_{i=1}^{m} d(\mathbf{a}_i, H_{\mathbf{x},y})^2 : 0 \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}$$

where

$$d(\mathbf{a}_i, H_{\mathbf{x},y})^2 = \frac{(\mathbf{a}_i^\top \mathbf{x} - y)^2}{\|\mathbf{x}\|^2}, \quad i = 1, \dots, m.$$

The Orthogonal Regression problem is the same as

$$\min_{\mathbf{x},y} \left\{ \sum_{i=1}^{m} \frac{(\mathbf{a}_i^\top \mathbf{x} - y)^2}{\|\mathbf{x}\|^2} : 0 \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R} \right\}$$

Fixing $\mathbf{x}$ and minimizing first with respect to y, we obtain that the optimal y is given by $y = \frac{1}{m} \sum_{i=1}^{m} \mathbf{a}_i^\top \mathbf{x} = \frac{1}{m} \mathbf{e}^\top \mathbf{A}\mathbf{x}$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]^\top$.

Using the above expression for y, we obtain that

$$\sum_{i=1}^{m} (\mathbf{a}_i^\top \mathbf{x} - y)^2 = \sum_{i=1}^{m} \left(\mathbf{a}_i^\top \mathbf{x} - \frac{1}{m} \mathbf{e}^\top \mathbf{A}\mathbf{x} \right)^2$$

$$= \sum_{i=1}^{m} (\mathbf{a}_i^\top \mathbf{x})^2 - \frac{2}{m} \sum_{i=1}^{m} (\mathbf{e}^\top \mathbf{A}\mathbf{x})(\mathbf{a}_i^\top \mathbf{x}) + \frac{1}{m} (\mathbf{e}^\top \mathbf{A}\mathbf{x})^2$$

$$= \mathbf{x}^\top \mathbf{A}^\top \left(\mathbf{I}_m - \frac{1}{m} \mathbf{e}\mathbf{e}^\top \right) \mathbf{A}\mathbf{x}$$

Therefore, a reformulation of the problem is

$$\min_{\mathbf{x}} \left\{ \frac{\mathbf{x}^\top \left[\mathbf{A}^\top \left(\mathbf{I}_m - \frac{1}{m} \mathbf{e}\mathbf{e}^\top \right) \mathbf{A} \right] \mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \neq 0 \right\}$$

10.9.1 Proposition

An optimal solution of the orthogonal regression problem is $(\mathbf{x}, y)$, where $\mathbf{x}$ is an eigenvector of $\mathbf{A}^\top \left(\mathbf{I}_m - \frac{1}{m} \mathbf{e}\mathbf{e}^\top \right) \mathbf{A}$ associated with the minimum eigenvalue and

$$y = \frac{1}{m} \sum_{i=1}^{m} \mathbf{a}_i^\top \mathbf{x}.$$

The optimal function value of the problem is

$$\lambda_{\min} \left[\mathbf{A}^\top \left(\mathbf{I}_m - \frac{1}{m} \mathbf{e}\mathbf{e}^\top \right) \mathbf{A} \right].$$

11 KKT Conditions

11.1 Definition (Feasible Descent Direction)

Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in C \end{aligned}$$

where f is continuously differentiable over the set $C \subseteq \mathbb{R}^n$. Then a vector $\mathbf{d} \neq \mathbf{0}$ is called a **feasible descent direction** at $x \in C$ if $\nabla f(x)^\top \mathbf{d} < 0$ and there exists $\epsilon > 0$ such that $x + t\mathbf{d} \in C$ for all $t \in [0, \epsilon]$.

11.2 Lemma

Consider the problem

$$\begin{aligned} (\mathbf{P}) \quad \min \quad & f(x) \\ \text{s.t.} \quad & x \in C \end{aligned}$$

where f is continuously differentiable over the set C. If $\mathbf{x}^*$ is a local optimal solution of $(\mathbf{P})$, then there are no feasible descent directions at $\mathbf{x}^*$.

11.3 Lemma

Let $\mathbf{x}^*$ be a local minimum of the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, m \end{aligned}$$

where $f, g_1, \ldots, g_m$ are continuously differentiable over $\mathbb{R}^n$. Let $I(\mathbf{x}^*)$ be the set of active constraints at $\mathbf{x}^*$:

$$I(\mathbf{x}^*) = \{i : g_i(\mathbf{x}^*) = 0\}.$$

Then there does not exist a vector $\mathbf{d} \in \mathbb{R}^n$ such that

$$\nabla f(\mathbf{x}^*)^\top \mathbf{d} < 0$$

and

$$\nabla g_i(\mathbf{x}^*)^\top \mathbf{d} < 0, \quad i \in I(\mathbf{x}^*).$$

11.4 The Fritz-John Necessary Conditions

Let $\mathbf{x}^*$ be a local minimum of the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \ldots, m \end{aligned}$$

where $f, g_1, \dots, g_m$ are continuously differentiable functions over $\mathbb{R}^n$. Then there exist multipliers $\lambda_0, \lambda_1, \dots, \lambda_m \geq 0$, which are not all zeros, such that

$$\lambda_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) = \mathbf{0}$$

and

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m.$$

11.5 KKT Conditions for Inequality/Equality Constrained Problems

Introduce the equality constraints:

Let $\mathbf{x}^*$ be a local minimum of the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(x) = 0, \quad j = 1, 2, \dots, p \end{aligned} \tag{1}$$

where $f, g_1, \dots, g_m, h_1, \dots, h_p$ are continuously differentiable functions over $\mathbb{R}^n$. Suppose that the gradients of the active constraints and the equality constraints:

$$\{\nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*)\}, \quad i \in I(\mathbf{x}^*), \, j = 1, 2, \dots, p$$

are **linearly independent**. Then there exist multipliers $\lambda_1, \dots, \lambda_m \geq 0, \mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$, such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m.$$

11.6 Definition (KKT point)

Consider problem (1) where $g_1, \dots, g_m, h_1, h_2, \dots, h_p$ are continuously differentiable functions over $\mathbb{R}^n$. A feasible point $\mathbf{x}^*$ is called a **KKT point** if there exist $\lambda_1, \dots, \lambda_m \geq 0, \mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m.$$

11.7 Definition (Regularity)

A feasible point $\mathbf{x}^*$ is called **regular** if the set

$$\{\nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*) \mid i \in I(\mathbf{x}^*), j = 1, 2, \dots, p\}$$

is linearly independent.

11.8 Example

$$\begin{aligned} \min \quad & f(x) = x_1 + x_2 \\ \text{s.t.} \quad & h(x) = \left[x_1^2 + x_2^2 - 1\right]^2 = 0 \end{aligned}$$

This problem has a highly nonlinear constraint, which includes a squared form. The constraint is effectively represented as the definition of a unit circle $x_1^2+x_2^2=1$. Such constraints often involve squaring operations or introduce squared or equivalent forms due to modeling errors, but during optimization, they may lead to the following issues:

1. **Gradient Degeneration**: At feasible points, the gradient of the constraint $\nabla h(x)$ contains the factor $[x_1^2+x_2^2-1]$, which becomes zero when the constraint is satisfied, resulting in the complete loss of gradient information for the constraint.
2. **KKT Condition Failure**: Gradient degeneration may render the stationarity condition (a part of the KKT conditions) inoperative. The necessary conditions for optimization problems may degenerate into meaningless identities.
3. **Numerical Optimization Challenges**: Gradient degeneration causes optimization algorithms to fail to find valid update directions, potentially leading to stagnation or erroneous results.

11.9 KKT Conditions in the Convex Case

In the convex case the KKT conditions are always sufficient. If a point satisfies the KKT condition, then it must be a globally optimal solution.

Let $\mathbf{x}^*$ be a feasible solution of the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, \quad i = 1, 2, \dots, m, \\ & h_j(x) = 0, \quad j = 1, 2, \dots, p \end{aligned} \tag{2}$$

where $f, g_1, \dots, g_m$ are continuously differentiable **convex** functions over $\mathbb{R}^n$ and $h_1, h_2, \dots, h_p$ are affine functions. Suppose that there exist multipliers $\lambda_1, \dots, \lambda_m \ge 0, \mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$, such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) = \mathbf{0}$$

and

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m.$$

Then $\mathbf{x}^*$ is the optimal solution of (2).

11.10 Slaters condition

If constraints are convex, regularity can be replaced by Slaters condition.

For a convex optimization problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, p, \end{aligned}$$

if all $g_i(x)$ are convex functions and $h_j(x)$ are affine functions, then the Slater's condition requires:

- There exists a strictly feasible point $x_0 \in \mathbb{R}^n$ such that:

  $$g_i(x_0) < 0, \quad \forall i = 1, \dots, m,$$

  and

  $$h_j(x_0) = 0, \quad \forall j = 1, \dots, p.$$

This means that there exists a point x_0 that strictly satisfies all inequality constraints and satisfies the equality constraints.

11.11 Constrained Least Squares

$$\textbf{(CLS)} \quad \min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|^2 \leq \alpha$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ has full column rank, $\mathbf{b} \in \mathbb{R}^m$, $\alpha > 0$.

—

Analysis of Problem (CLS):

Problem (CLS) is a convex problem and satisfies Slater's condition. The Lagrangian is:

$$L(\mathbf{x}, \lambda) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda(\|\mathbf{x}\|^2 - \alpha), \quad (\lambda \geq 0)$$

The KKT conditions are:

$$\nabla_{\mathbf{x}} L = 2\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) + 2\lambda\mathbf{x} = 0$$

$$\lambda(\|\mathbf{x}\|^2 - \alpha) = 0$$

$$\|\mathbf{x}\|^2 \leq \alpha, \quad \lambda \geq 0$$

—

If $\lambda = 0$, then by the first equation:

$$\mathbf{x} = \mathbf{x}_{\text{LS}} \equiv (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$$

Optimal if and only if $\|\mathbf{x}_{\mathrm{LS}}\|^2 \leq \alpha$.

—

On the other hand, if $\|\mathbf{x}_{\mathrm{LS}}\|^2 > \alpha$, then necessarily $\lambda > 0$. By the complementary slackness condition, we have $\|\mathbf{x}\|^2 = \alpha$ and the first equation implies that:

$$\mathbf{x} = \mathbf{x}_\lambda \equiv (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{b}$$

The multiplier $\lambda > 0$ should be chosen to satisfy $\|\mathbf{x}_\lambda\|^2 = \alpha$, that is, λ is the solution of:

$$f(\lambda) = \|(\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{b}\|^2 - \alpha = 0$$

—

At $\lambda = 0$:

$$f(0) = \|(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}\|^2 - \alpha = \|\mathbf{x}_{\mathrm{LS}}\|^2 - \alpha > 0$$

and $f(\lambda) \to -\alpha$ as $\lambda \to \infty$. The function f is strictly decreasing.

Conclusion: The optimal solution of the CLS problem is given by:

$$\mathbf{x} = \begin{cases} \mathbf{x}_{\mathrm{LS}}, & \|\mathbf{x}_{\mathrm{LS}}\|^2 \leq \alpha \\ \mathbf{x}_\lambda, & \|\mathbf{x}_{\mathrm{LS}}\|^2 > \alpha \end{cases}$$

where λ is the unique root of $f(\lambda)$ over $(0, \infty)$.

11.12 Second Order Necessary Optimality Conditions for Inequality/Equality Constrained Problems

Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(x) = 0, \quad j = 1, 2, \dots, p \end{aligned}$$

where $f, g_1, \dots, g_m, h_1, \dots, h_p$ are twice continuously differentiable functions. Let $\mathbf{x}^*$ be a local minimum and suppose that $\mathbf{x}^*$ is regular, meaning that the set

$$\{\nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*) \mid i \in I(\mathbf{x}^*), j = 1, 2, \dots, p\}$$

is linearly independent. Then $\exists \lambda_1, \lambda_2, \dots, \lambda_m \geq 0$ and $\mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ such that

$$\nabla_x L(\mathbf{x}^*, \lambda, \mu) = 0,$$

$$\lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m,$$

and

$$\mathbf{d}^\top \nabla^2_{xx} L(\mathbf{x}^*, \lambda, \mu) \mathbf{d} \geq 0 \quad \text{for all } \mathbf{d} \in \Lambda(\mathbf{x}^*),$$

where

$$\Lambda(\mathbf{x}^*) \equiv \{\mathbf{d} \in \mathbb{R}^n \mid \nabla g_i(\mathbf{x}^*)^\top \mathbf{d} = 0,\ i \in I(\mathbf{x}^*),\ \nabla h_j(\mathbf{x}^*)^\top \mathbf{d} = 0,\ j = 1, 2, \dots, p\}.$$

11.13 Trust Region Methods

Objective Problem

Consider a general unconstrained optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f(x)$ is the objective function to be minimized.

Core Ideas of the Trust Region Method

1. **Local Quadratic Model**:

   At each iteration, construct a quadratic approximation $m_k(p)$ of the objective function near the current point x_k:
   $$m_k(p) = f(x_k) + \nabla f(x_k)^\top p + \frac{1}{2} p^\top H_k p,$$

   where:

   - $p = x - x_k$ is the search direction.
   - $\nabla f(x_k)$ is the gradient at the current point.
   - H_k is the Hessian matrix (or its approximation) at the current point.

2. **Trust Region Constraint**:

   Restrict the step length p such that the new step stays within a trusted region (called the trust region):
   $$\|p\| \le \Delta_k,$$

   where $\Delta_k > 0$ is the radius of the trust region.

3. **Step Adjustment**:

   Evaluate the quality of the approximation $m_k(p)$ to the actual objective function, and adaptively adjust the radius Δ_k of the trust region to control the step length.

11.14 Trust Region Subproblem

During each iteration, a trust region subproblem needs to be solved:

$$\min_p \quad m_k(p) = f(x_k) + \nabla f(x_k)^\top p + \frac{1}{2} p^\top H_k p,$$

$$\text{s.t.} \quad \|p\| \le \Delta_k.$$

Characteristics of the Trust Region Subproblem

1. **Quadratic Optimization Problem**:

   The objective function is quadratic.

2. **Step Length Constraint**:

   The constraint is a simple quadratic constraint $\|p\| \leq \Delta_k$.

3. **Analytical and Numerical Solutions**:

   - When H_k is positive definite, an analytical solution may exist.
   - When H_k is not positive definite, numerical methods are required to solve the problem.

Trust Region Method Workflow

The workflow of the trust region method can be divided into the following steps:

Step 1: Initialization

- Specify an initial point x_0 and set the initial trust region radius $\Delta_0 > 0$.
- Set the convergence tolerance $\epsilon > 0$ and step length adjustment parameters.

Step 2: Solve the Trust Region Subproblem

At the current point x_k, solve the trust region subproblem to obtain the step p_k:

$$\min_{p} \quad m_k(p) = \nabla f(x_k)^\top p + \frac{1}{2} p^\top H_k p,$$

$$\text{s.t.} \quad \|p\| \leq \Delta_k.$$

Step 3: Update

Compute the new point $x_{k+1} = x_k + p_k$ and evaluate the actual reduction in the objective function.

Step 4: Evaluate Model Accuracy

Compare the model-predicted reduction in the objective function with the actual reduction, and define the ratio:

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}.$$

- $\rho_k \approx 1$: The model $m_k(p)$ accurately predicts the behavior of $f(x)$, and the trust region radius Δ_k may remain unchanged.
- $\rho_k \ll 1$: The model prediction is inaccurate, and the trust region may need to be reduced.
- $\rho_k \gg 1$: The model prediction is conservative, and the trust region may need to be enlarged.

Step 5: Adjust Trust Region Radius

Adjust the trust region radius based on the value of ρ_k:

- If $\rho_k \geq \eta_1$ (successful step), increase the trust region radius: $\Delta_{k+1} = \gamma_1 \Delta_k$, where $\gamma_1 > 1$.
- If $\rho_k \leq \eta_2$ (unsuccessful step), decrease the trust region radius: $\Delta_{k+1} = \gamma_2 \Delta_k$, where $0 < \gamma_2 < 1$.
- Otherwise, keep $\Delta_{k+1} = \Delta_k$.

Step 6: Termination Condition

Terminate the iteration if:

- $\|\nabla f(x_k)\| < \epsilon$, or
- The trust region radius is smaller than a certain threshold, $\Delta_k < \Delta_{\min}$.

11.15 KKT conditions for the Trust Region Subproblem

The Trust Region Subproblem (TRS) is a problem of minimizing an indefinite quadratic function under an ℓ_2-norm constraint, given by:

$$\text{(TRS)}: \quad \min_{\mathbf{p} \in \mathbb{R}^n} m(\mathbf{p}) = f(x_k) + \nabla f(x_k)^\top \mathbf{p} + \frac{1}{2} \mathbf{p}^\top \mathbf{H} \mathbf{p}, \\ \text{s.t.} \quad \|\mathbf{p}\|^2 \leq \Delta^2,$$

where:

- $\mathbf{p} = \mathbf{x} - x_k$ represents the search direction relative to the current point x_k;
- $m(\mathbf{p})$ is a quadratic approximation of the objective function $f(x)$;
- $\nabla f(x_k)$ is the gradient of the objective function at the current point x_k;
- $\mathbf{H} \in \mathbb{R}^{n \times n}$ is the Hessian matrix of the objective function or its approximation;
- $\Delta > 0$ is the trust region radius, restricting the step length.

A vector $\mathbf{p}^*$ is the optimal solution of the Trust Region Subproblem (TRS) if and only if there exists $\lambda^* \geq 0$ such that the following conditions are satisfied:

1. **Stationarity**:
$$(\mathbf{H} + \lambda^* \mathbf{I}) \mathbf{p}^* = -\nabla f(\mathbf{x}_k) \tag{10}$$
where $\mathbf{H}$ is the Hessian matrix of the objective function and $\nabla f(\mathbf{x}_k)$ is the gradient.

2. **Feasibility**:
$$\|\mathbf{p}^*\|^2 \leq \Delta^2 \tag{11}$$
This ensures that the search direction $\mathbf{p}^*$ lies within the trust region.

3. **Complementary Slackness**:
$$\lambda^*(\|\mathbf{p}^*\|^2 - \Delta^2) = 0 \tag{12}$$
When $\|\mathbf{p}^*\|^2 < \Delta^2$, $\lambda^* = 0$; when $\|\mathbf{p}^*\|^2 = \Delta^2$, $\lambda^* \geq 0$.

4. **Positive Semidefiniteness**:
$$\mathbf{H} + \lambda^*\mathbf{I} \succeq 0 \tag{13}$$
This ensures that the matrix is positive semidefinite.

11.16 Least Squares (LS): Noise Only on the Right-Hand Side

In the least squares method, it is assumed that **only the right-hand side b** contains noise. The goal of the optimization is to find a vector $\mathbf{x}$ such that:

$$\mathbf{A}\mathbf{x} = \mathbf{b} + \mathbf{w},$$

where $\mathbf{w} \in \mathbb{R}^m$ is the noise vector.

Optimization Problem

$$\min_{\mathbf{w},\mathbf{x}} \|\mathbf{w}\|^2,$$

i.e., minimizing the norm of the noise vector $\mathbf{w}$.

11.17 Total Least Squares (TLS): Considering Noise in Both A and b

In the total least squares method, it is assumed that **both the coefficient matrix A and the right-hand side b** contain noise, and their errors are to be optimized.

Model Modification

The model becomes:
$$(\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{w},$$

where:

- $\mathbf{E} \in \mathbb{R}^{m\times n}$ is the error matrix for $\mathbf{A}$;
- $\mathbf{w} \in \mathbb{R}^m$ is the error vector for $\mathbf{b}$.

Optimization Problem

$$\min_{\mathbf{E},\mathbf{w},\mathbf{x}} \|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2,$$
$$\text{s.t.} \quad (\mathbf{A} + \mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{w}.$$

11.18 The New Formulation of (TLS)

To reduce the complexity of the problem, we take a decomposition approach:
Fix $\mathbf{x}$, solve the subproblem (P_x), and obtain the analytic expressions of $\mathbf{E}$ and $\mathbf{w}$.

$$(P_x) \quad \min_{\mathbf{E},\mathbf{w}} \|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2,$$
$$\text{s.t. } (\mathbf{A}+\mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{w}.$$

The Lagrangian function is constructed as:

$$L(\mathbf{E},\mathbf{w},\lambda) = \|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2 + 2\lambda^\top[(\mathbf{A}+\mathbf{E})\mathbf{x} - \mathbf{b} - \mathbf{w}].$$

To satisfy the KKT conditions, we need:

1. **Stationarity:**

$$\nabla_{\mathbf{E}} L = 0, \quad \nabla_{\mathbf{w}} L = 0.$$

For $\mathbf{E}$:

$$\nabla_{\mathbf{E}} L = 2\mathbf{E} + 2\lambda\mathbf{x}^\top = 0 \implies \mathbf{E} = -\lambda\mathbf{x}^\top. \tag{19}$$

For $\mathbf{w}$:

$$\nabla_{\mathbf{w}} L = 2\mathbf{w} - 2\lambda = 0 \implies \mathbf{w} = \lambda. \tag{20}$$

2. **Feasibility:**

$$(\mathbf{A}+\mathbf{E})\mathbf{x} = \mathbf{b} + \mathbf{w}.$$

Combining $\mathbf{E} = -\lambda\mathbf{x}^\top$ and $\mathbf{w} = \lambda$, we have:

$$(\mathbf{A} - \lambda\mathbf{x}^\top)\mathbf{x} = \mathbf{b} + \lambda.$$

Simplifying gives:

$$\lambda(\|\mathbf{x}\|^2 + 1) = \mathbf{A}\mathbf{x} - \mathbf{b}.$$

Thus:

$$\lambda = \frac{\mathbf{A}\mathbf{x} - \mathbf{b}}{\|\mathbf{x}\|^2 + 1}. \tag{21}$$

From this, we deduce:

$$\mathbf{E} = -\lambda\mathbf{x}^\top, \quad \mathbf{w} = \lambda, \quad \lambda = \frac{\mathbf{A}\mathbf{x} - \mathbf{b}}{\|\mathbf{x}\|^2 + 1}.$$

Substituting these expressions into the original objective function:

$$\|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2 = \|\lambda\mathbf{x}^\top\|_F^2 + \|\lambda\|^2.$$

Note that:

$$\|\lambda\mathbf{x}^\top\|_F^2 = \|\lambda\|^2\|\mathbf{x}\|^2,$$

thus:

$$\|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2 = \|\lambda\|^2(\|\mathbf{x}\|^2 + 1).$$

Substituting $\lambda = \frac{\mathbf{Ax}-\mathbf{b}}{\|\mathbf{x}\|^2+1}$, the objective function simplifies to:

$$\|\mathbf{E}\|_F^2 + \|\mathbf{w}\|^2 = \frac{\|\mathbf{Ax}-\mathbf{b}\|^2}{\|\mathbf{x}\|^2+1}.$$

By substituting the solution of the subproblem, the original problem is transformed into the following unconstrained optimization problem about $\mathbf{x}$:

$$(\text{TLS}') \quad \min_{\mathbf{x}\in\mathbb{R}^n} \frac{\|\mathbf{Ax}-\mathbf{b}\|^2}{\|\mathbf{x}\|^2+1}.$$

However, it's still a **nonconvex** problem. What's more, it resembles the problem of minimizing the **Rayleigh quotient**.

11.19 Theorem

$\mathbf{x}$ is an optimal solution of (TLS$'$) if and only if $(\mathbf{x}, \mathbf{E}, \mathbf{w})$ is an optimal solution of (TLS) where $\mathbf{E} = -\frac{(\mathbf{Ax}-\mathbf{b})\mathbf{x}^\top}{\|\mathbf{x}\|^2+1}$ and $\mathbf{w} = \frac{\mathbf{Ax}-\mathbf{b}}{\|\mathbf{x}\|^2+1}$.

11.20 Homogenization Argument and Relaxation of TLS

Introducing $t = 1$

To handle the fractional objective function $\frac{\|\mathbf{Ax}-\mathbf{b}\|^2}{\|\mathbf{x}\|^2+1}$, we can introduce an auxiliary variable t to homogenize the fraction. The new objective function becomes:

$$\min_{\mathbf{x}\in\mathbb{R}^n, t\in\mathbb{R}} \left\{ \frac{\|\mathbf{Ax}-t\mathbf{b}\|^2}{\|\mathbf{x}\|^2+t^2} : t = 1 \right\}.$$

This means replacing the constant 1 in the denominator with t^2, making the problem homogeneous with respect to the variables $(\mathbf{x}, t)$.

Reformulating into a Homogeneous Form

By defining a new variable $\mathbf{y} = \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix} \in \mathbb{R}^{n+1}$, the problem becomes:

$$f^* = \min_{\mathbf{y}\in\mathbb{R}^{n+1}} \left\{ \frac{\mathbf{y}^\top \mathbf{B} \mathbf{y}}{\|\mathbf{y}\|^2} : y_{n+1} = 1 \right\}, \tag{22}$$

where:

$$\mathbf{B} = \begin{pmatrix} \mathbf{A}^\top\mathbf{A} & -\mathbf{A}^\top\mathbf{b} \\ -\mathbf{b}^\top\mathbf{A} & \|\mathbf{b}\|^2 \end{pmatrix}.$$

Here:

- $\mathbf{B}$ is a symmetric matrix, which integrates information from $\mathbf{A}$ and $\mathbf{b}$.

- The key to homogenization is transforming the fractional objective into the quadratic form $\mathbf{y}^\top \mathbf{B}\mathbf{y}$ divided by $\|\mathbf{y}\|^2$.

Relaxing the Constraint $y_{n+1} = 1$

A key constraint in the above problem is $y_{n+1} = 1$, which fixes the last dimension of the homogeneous variable $\mathbf{y}$. However, this constraint may be unnecessary and increases the complexity of the solution.

Thus, we consider a relaxed version of the problem:

$$g^* = \min_{\mathbf{y} \in \mathbb{R}^{n+1}} \left\{ \frac{\mathbf{y}^\top \mathbf{B}\mathbf{y}}{\|\mathbf{y}\|^2} : \mathbf{y} \neq \mathbf{0} \right\}. \tag{23}$$

In the relaxed version, we drop the restriction $y_{n+1} = 1$, requiring only that $\mathbf{y} \neq \mathbf{0}$.

Advantages of the Relaxed Version

1. **Improved Generality:**

   - The relaxed version does not rely on specific coordinate constraints and is applicable to a broader range of problems.

2. **Simplified Solution:**

   - After relaxation, the problem reduces to finding the eigenvector of matrix $\mathbf{B}$ corresponding to its smallest eigenvalue.

11.21 Lemma

If the optimal solution $\mathbf{y}^*$ of the relaxed problem (23) satisfies $y^*_{n+1} \neq 0$, then through the normalization operation:

$$\tilde{\mathbf{y}} = \frac{\mathbf{y}^*}{y^*_{n+1}},$$

we can obtain a new vector $\tilde{\mathbf{y}}$, which satisfies:

1. $\tilde{\mathbf{y}}^\top \mathbf{B}\tilde{\mathbf{y}}/\|\tilde{\mathbf{y}}\|^2 = g^*$, meaning the objective value remains unchanged;

2. $\tilde{y}_{n+1} = 1$, meaning it satisfies the constraint of the original problem.

Thus, $\tilde{\mathbf{y}}$ is the optimal solution of the original problem (22).

12 Duality

12.1 Definition

$$
\begin{aligned}
f^* = \ & \min f(x) \\
\text{s.t. } & g_i(x) \le 0, \quad i = 1, 2, \dots, m, \\
& h_j(x) = 0, \quad j = 1, 2, \dots, p, \\
& x \in X
\end{aligned}
\tag{1}
$$

f, g_i, h_j $(i = 1, 2, \dots, m,\ j = 1, 2, \dots, p)$ are functions defined on the set $X \subseteq \mathbb{R}^n$.

Problem (1) will be referred to as the primal problem.

The dual objective function $q : \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ is defined to be

$$q(\lambda, \mu) = \min_{x \in X} L(x, \lambda, \mu). \tag{2}$$

The domain of the dual objective function is

$$\mathrm{dom}(q) = \{(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p : q(\lambda, \mu) > -\infty\}.$$

The dual problem is given by

$$q^* = \max q(\lambda, \mu) \quad \text{s.t.} \quad (\lambda, \mu) \in \mathrm{dom}(q). \tag{3}$$

12.2 Theorem

Consider problem (1) with f, g_i, h_j $(i = 1, 2, \dots, m,\ j = 1, 2, \dots, p)$ being functions defined on the set $X \subseteq \mathbb{R}^n$, and let q be the dual function defined in (2). Then

(a) $\mathrm{dom}(q)$ is a convex set.

(b) q is a concave function over $\mathrm{dom}(q)$.

12.3 The Weak Duality Theorem

Consider the primal problem (1) and its dual problem (3). Then

$$q^* \le f^*,$$

where f^*,q^* are the primal and dual optimal values respectively.

12.4 Theorem (Supporting Hyperplane Theorem)

Let $C \subseteq \mathbb{R}$ be a convex set and let $\mathbf{y} \notin C$. Then there exists $\mathbf{0} \ne \mathbf{p} \in \mathbb{R}^n$ such that

$$\mathbf{p}^T \mathbf{x} \le \mathbf{p}^T \mathbf{y} \quad \text{for any} \quad \mathbf{x} \in C.$$

12.5 Separation of Two Convex Sets

Let $C_1, C_2 \subseteq \mathbb{R}^n$ be two nonempty convex sets such that $C_1 \cap C_2 = \emptyset$. Then there exists $0 \neq \mathbf{p} \in \mathbb{R}^n$ for which

$$\mathbf{p}^\top \mathbf{x} \leq \mathbf{p}^\top \mathbf{y} \quad \text{for any } \mathbf{x} \in C_1,\ \mathbf{y} \in C_2.$$

12.6 The Nonlinear Farkas' Lemma

Let $X \subseteq \mathbb{R}^n$ be a convex set and let $f, g_1, g_2, \ldots, g_m$ be convex functions over X. Assume that there exists $\hat{x} \in X$ such that

$$g_1(\hat{x}) < 0,\ g_2(\hat{x}) < 0,\ \ldots,\ g_m(\hat{x}) < 0.$$

Let $c \in \mathbb{R}$. Then the following two claims are equivalent:

(a) the following implication holds:

$$\mathbf{x} \in X,\ g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m \implies f(\mathbf{x}) \geq c.$$

(b) there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that

$$\min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \right\} \geq c. \tag{5}$$

12.7 Strong Duality of Convex Problems with Inequality Constraints

Consider the optimization problem

$$\begin{aligned} f^* = \ & \min f(\mathbf{x}) \\ & \text{s.t.} \quad g_i(\mathbf{x}) \leq 0,\ i = 1, 2, \ldots, m, \\ & \mathbf{x} \in X \end{aligned} \tag{7}$$

where X is a convex set and $f, g_i,\ i = 1, 2, \ldots, m$ are convex functions over X. Suppose that there exists $\hat{\mathbf{x}} \in X$ for which $g_i(\hat{\mathbf{x}}) < 0,\ i = 1, 2, \ldots, m$. If problem (7) has a finite optimal value, then

(a) the optimal value of the dual problem is attained.

(b) $f^* = q^*$.

12.8 Complementary Slackness Conditions

Consider the optimization problem

$$f^* = \min\{f(\mathbf{x}) : g_i(\mathbf{x}) \leq 0,\, i = 1, 2, \ldots, m,\, \mathbf{x} \in X\}, \tag{8}$$

and assume that $f^* = q^*$ where q^* is the optimal value of the dual problem. Let $\mathbf{x}^*, \lambda^*$ be feasible solutions of the primal and dual problems, respectively. Then $\mathbf{x}^*, \lambda^*$ are **optimal** solutions of the primal and dual problems if and only if

$$\mathbf{x}^* \in \arg\min_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*), \tag{9}$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, 2, \dots, m. \tag{10}$$

12.9 A More General Strong Duality Theorem

Consider the optimization

$$\begin{aligned} f^* = & \min_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(\mathbf{x}) \leq 0, \quad j = 1, 2, \dots, p, \\ & s_k(\mathbf{x}) = 0, \quad k = 1, 2, \dots, q. \\ & \mathbf{x} \in \mathbf{X} \end{aligned} \tag{11}$$

where X is a convex set and f, g_i, $i = 1, 2, \dots, m$ are convex functions over X. The functions h_j, s_k are affine functions. Suppose that there exists $\hat{\mathbf{x}} \in \text{int}(X)$ for which $g_i(\hat{\mathbf{x}}) < 0, h_j(\hat{\mathbf{x}}) \leq 0, s_k(\hat{\mathbf{x}}) = 0$. Then if problem (11) has a finite optimal value, the optimal value of the dual problem

$$q^* = \max\{q(\lambda, \eta, \mu) : (\lambda, \eta, \mu) \in \text{dom}(q)\},$$

where

$$q(\lambda, \eta, \mu) = \min_{\mathbf{x}\in X} \left[f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \eta_j h_j(\mathbf{x}) + \sum_{k=1}^{q} \mu_k s_k(\mathbf{x}) \right]$$

is attained, and $f^* = q^*$.

13 Subgradient Method

13.1 Theorem (Informal)

Any Convex function is differentiable almost everywhere.

13.2 Introduction to Subgradient

For a differentiable convex function $f : \mathbb{R}^n \to (-\infty, \infty]$, its linearization at a vector $x \in \text{dom}(f)$ is given by:

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}),$$

And we have for all $\mathbf{y} \in \text{dom}(f)$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}).$$

It shows that the linearization of a convex function is always its lower bound.

13.3 Definition (Subgradient and Subdifferential)

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex function, i.e. f not always equal to ∞ . A vector $\mathbf{g} \in \mathbb{R}^n$ is a **subgradient** of f at a point $\mathbf{x} \in \text{dom}(f)$ if for all $\mathbf{y} \in \text{dom}(f)$, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^\top (\mathbf{y} - \mathbf{x}).$$

The **subdifferential** of f at $\mathbf{x}$ is the set of all subgradients of f at $\mathbf{x}$, denoted by $\partial f(\mathbf{x})$.

13.3.1 Subgradient of a Differentiable Point

For differentiable convex function, at every point $\mathbf{x} \in \text{dom}(f)$, the set of subgradients $\partial f(\mathbf{x})$ contains only the gradient $\nabla f(\mathbf{x})$.

13.3.2 Subgradient of a Nondifferentiable point

For a convex function that is not differentiable at a point, the subdifferential is a set of subgradients, rather than a single subgradient.

13.4 Epigraph

Epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set of points lying above the graph of f:

$$\text{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^n \times \mathbb{R} : f(\mathbf{x}) \leq t\}.$$

The subdifferential inequality can be combined with geometry, where the concept of subdifferential s can be described using a supporting hyperplane as follows:

For the subdifferential s at the point $(x, f(x))$, a **supporting hyperplane** H is defined as:

$$H = \{(y, \gamma) \in \mathbb{R}^{n+1} : (-s^\top, 1)^\top (y; \gamma) = (-s^\top, 1)^\top (x; f(x))\}.$$

The hyperplane passes through the point $(x, f(x))$ and supports the epigraph of f.

In other words, this hyperplane is ”tightly attached” to the boundary of the epigraph and satisfies the property that all points below the hyperplane are contained in epi(f). This is consistent with the subdifferential condition.

13.5 Theorem (Closedness and Convexity of the Subdierential Set)

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper function. Then the set $\partial f(\mathbf{x})$ is **closed and convex** for any $\mathbf{x} \in \text{dom}(f)$.

13.6 Definiiton (Subdifferentiability)

A proper function $f : \mathbb{R}^n \to (-\infty, \infty]$ is **subdifferentiable** at a point $\mathbf{x} \in \text{dom}(f)$ if its subdifferential $\partial f(\mathbf{x})$ is nonempty.

13.7 Lemma (Nonemptiness of Subdifferential Sets → Convexity)

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper function and assume dom(f) is convex. Suppose for any $\mathbf{x} \in \text{dom}(f)$, the subdifferential $\partial f(\mathbf{x})$ is nonempty. Then f is convex.

13.8 Theorem (Nonemptiness and boundedness of the subdifferential set at interior points of the domain)

Let f be a **proper convex** function, and assume $\mathbf{x} \in \text{int}(\text{dom}(f))$. Then, $\partial f(\mathbf{x})$ is **nonempty and bounded.**

13.8.1 Corollary

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then f is subdifferentiable over $\mathbb{R}^n$.

13.9 Max Formula

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper convex function. Then for any $\mathbf{x} \in \text{int}(\text{dom}(f))$ and $\mathbf{d} \in \mathbb{R}^n$,

$$f'(\mathbf{x}; \mathbf{d}) = \max\{\mathbf{s}^\top \mathbf{d} : \mathbf{s} \in \partial f(\mathbf{x})\}.$$

13.10 Theorem (The Subdierential at points of dierentiability)

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex function, and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. If f is differentiable at $\mathbf{x}$, then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

Conversely, if f has a unique subgradient at $\mathbf{x}$, then f is differentiable at $\mathbf{x}$, and $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

13.11 Theorem (Multiplication by a positive scalar)

In convex analysis, the **subgradient** is a generalized concept of the ”slope” of a convex function at a given point, especially useful for non-differentiable points. The computation of subgradients is particularly important in optimization problems, as it provides directional information for optimization algorithms.

The methods for computing subgradients are usually divided into two categories:

1. **Weak Results**: These are rules for computing some subgradients, i.e., in certain cases, specific subgradient vectors can be directly identified. However, they do not fully describe the entire subdifferential set.

2. **Strong Results**: These provide a full characterization or complete description of the subdifferential set, allowing us to completely understand and describe all possible subgradient vectors.

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex function, and let $\mathbf{x} \in \operatorname{dom}(f)$. Then for any $\mathbf{x}$ and any $\alpha > 0$, we have

$$\partial(\alpha f)(\mathbf{x}) = \alpha \partial f(\mathbf{x}).$$

13.12 Theorem (Summation)

Let $f_1, f_2 : \mathbb{R}^n \to (-\infty, \infty]$ be proper convex functions.

(a) If $\mathbf{x} \in \operatorname{dom}(f_1) \cap \operatorname{dom}(f_2)$, the following inclusion holds:

$$\partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}) \subseteq \partial(f_1 + f_2)(\mathbf{x}).$$

(b) If $\mathbf{x} \in \operatorname{int}(\operatorname{dom}(f_1)) \cap \operatorname{int}(\operatorname{dom}(f_2))$, then

$$\partial(f_1 + f_2)(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x}).$$

13.13 Theorem(Affine transformation)

Let $f : \mathbb{R}^m \to (-\infty, \infty]$ be a proper convex function and $A \in \mathbb{R}^{m \times n}$. Let

$$h(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b}) \quad \text{with } \mathbf{b} \in \mathbb{R}^m.$$

Assume that h is proper, meaning that

$$\operatorname{dom}(h) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} + \mathbf{b} \in \operatorname{dom}(f)\} \neq \emptyset.$$

(a) (weak affine transformation rule of subdifferential calculus) For any $\mathbf{x} \in \operatorname{dom}(h)$,

$$A^\top(\partial f(A\mathbf{x} + \mathbf{b})) \subseteq \partial h(\mathbf{x}).$$

- Left-hand explanation: $A^\top(\partial f(A\mathbf{x}+\mathbf{b}))$ represents the new set obtained by applying the linear transformation $A^\top$ to all subgradient vectors of f at the point $A\mathbf{x} + \mathbf{b}$.
- Right-hand explanation: $\partial h(\mathbf{x})$ represents the subdifferential set of the function h at the point $\mathbf{x}$.

Meaning: This part shows that the subdifferential set of the transformed function h includes at least all vectors in the subdifferential set of f at the corresponding point, after applying the linear transformation $A^\top$.

(b) (affine transformation rule of subdifferential calculus) If $\mathbf{x} \in \operatorname{int}(\operatorname{dom}(h))$, then

$$\partial h(\mathbf{x}) = A^\top(\partial f(A\mathbf{x} + \mathbf{b})).$$

In this case, the subdifferential set of h not only includes all vectors in $A^\top \partial f(A\mathbf{x} + \mathbf{b})$, but the two sets are actually equivalent.

13.14 Theorem (Max Rule of Subdifferential Calculus)

Let $f_1, f_2, \dots, f_m : \mathbb{R}^n \to (-\infty, \infty]$ be proper convex functions, and define

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}.$$

Let $\mathbf{x} \in \bigcap_{i=1}^m \operatorname{int}(\operatorname{dom}(f_i))$. Then

$$\partial f(\mathbf{x}) = \operatorname{conv}\left(\bigcup_{i \in I(\mathbf{x})} \partial f_i(\mathbf{x})\right),$$

where $I(\mathbf{x}) = \{i \in \{1, 2, \dots, m\} : f_i(\mathbf{x}) = f(\mathbf{x})\}$.

13.15 Theorem (Fermat's optimality condition)

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex function. Then

$$\mathbf{x}^* \in \operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$$

if and only if $\mathbf{0} \in \partial f(\mathbf{x}^*)$.

13.16 Theorem (Optimality Conditions for Convex Constrained Optimization)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper convex function, and let C be a convex set for which $\operatorname{int}(\operatorname{dom}(f)) \cap \operatorname{int}(C) \neq \emptyset$. Then

$$\mathbf{x}^* \in \operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

if and only if there exists $\mathbf{s} \in \partial f(\mathbf{x}^*)$ for which

$$\mathbf{s}^\top(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad \text{for any } \mathbf{x} \in C.$$

13.17 Theorem (Informal)

The direction of minus the subgradient is not necessarily a descent direction.

13.17.1 Example

consider $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $f(x_1, x_2) = |x_1| + 2|x_2|$. Then

$$\partial f(1,0) = \{(1,x) : |x| \leq 2\}.$$

In particular, $(1,2) \in \partial f(1,0)$. However, the direction $-(1,2)$ is not a descent direction.

13.18 Projected Subgradient Method

Algorithm 3 Projected Subgradient Method

```
0: Initialization: pick $\mathbf{x}_0 \in C$ arbitrarily.
0: General step:
0: for $k = 0, 1, 2, ...$ do
0:   pick a stepsize $t_k > 0$ and a subgradient $f'(\mathbf{x}_k) \in \partial f(\mathbf{x}_k)$
0:   set $\mathbf{x}_{k+1} = P_C(\mathbf{x}_k - t_k f'(\mathbf{x}_k))$
0: end for=0
```

13.19 Convergence of the Projected Subgradient Method

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ is a proper closed and convex function.
- Let $C \subseteq \text{int}(\text{dom}(f))$ be nonempty, closed, and convex.
- The optimal set X^* is nonempty, and the optimal value is f^*.
- There exists a constant $L_f > 0$ for which $\|\mathbf{s}\| \leq L_f$ for all $\mathbf{s} \in \partial f(\mathbf{x})$, $\mathbf{x} \in C$.

Under the above assumptions, let $\{\mathbf{x}_k\}_{k \geq 0}$ be the sequence generated by the projected subgradient method for solving

$$\min_{\mathbf{x} \in C} f(\mathbf{x})$$

with positive stepsizes $\{t_k\}_{k \geq 0}$. If

$$\frac{\sum_{n=0}^{k} t_n^2}{\sum_{n=0}^{k} t_n} \to 0 \quad \text{as } k \to \infty,$$

then

$$\hat{f}_k - f^* \to 0 \quad \text{as } k \to \infty,$$

where $\hat{f}_k \equiv \min_{n=0,1,\ldots,k} f(\mathbf{x}_n)$.