**高级工程数学个性化教材**

|  |  |  |  |
| --- | --- | --- | --- |
| **教师** | **序号** | **姓名** | **学号** |
|  |  |  |  |
|  |  |  |  |

**目录**

[序章 数学史 4](#_Toc186278832)

[数学的词源： 4](#_Toc186278833)

[关键数学家与贡献： 4](#_Toc186278834)

[数学史上的里程碑： 4](#_Toc186278835)

[数学发展的视角： 4](#_Toc186278836)

[名言引用： 4](#_Toc186278837)

[第一章 预备知识：数学证明方法与基本记号 4](#_Toc186278838)

[本章导言: 4](#_Toc186278839)

[1.1 数学证明方法 5](#_Toc186278840)

[1.2 基本记号 6](#_Toc186278841)

[本章小结 6](#_Toc186278842)

[第二章 向量空间与矩阵 6](#_Toc186278843)

[本章导言: 6](#_Toc186278844)

[2.1 向量与矩阵 6](#_Toc186278845)

[2.2 矩阵的秩 8](#_Toc186278846)

[2.3 线性方程组 9](#_Toc186278847)

[2.4 内积与范数 9](#_Toc186278848)

[本章小结 10](#_Toc186278849)

[第三章 线性变换、特征值与特征向量及正交投影 11](#_Toc186278850)

[本章导言: 11](#_Toc186278851)

[3.1 线性变换 11](#_Toc186278852)

[3.2 特征值与特征向量 12](#_Toc186278853)

[3.3 正交投影 12](#_Toc186278854)

[本章小结 13](#_Toc186278855)

[第四章 几何概念：线段、超平面与凸集 13](#_Toc186278856)

[本章导言: 13](#_Toc186278857)

[4.1 线段 13](#_Toc186278858)

[4.2 超平面与线性簇 13](#_Toc186278859)

[4.3 凸集 14](#_Toc186278860)

[4.4 邻域 15](#_Toc186278861)

[4.5 多面体与多胞体 15](#_Toc186278862)

[本章小结 16](#_Toc186278863)

[第五章 微积分基础：序列、极限与微分 16](#_Toc186278864)

[本章导言: 16](#_Toc186278865)

[5.1 序列与极限 16](#_Toc186278866)

[5.2 可微性 17](#_Toc186278867)

[5.3 导数矩阵 17](#_Toc186278868)

[5.4 求导法则 18](#_Toc186278869)

[5.5 水平集与梯度 18](#_Toc186278870)

[5.6 泰勒级数 19](#_Toc186278871)

[本章小结 19](#_Toc186278872)

[第六章 约束优化问题的最优性条件 19](#_Toc186278873)

[本章导言: 19](#_Toc186278874)

[6.1 约束优化问题概述 19](#_Toc186278875)

[6.2 最优性条件 20](#_Toc186278876)

[6.3 拉格朗日乘子法 (Lagrange Multipliers) 21](#_Toc186278877)

[本章小结 21](#_Toc186278878)

[第七章 不等式约束优化问题的最优性条件：KKT条件 22](#_Toc186278879)

[本章导言: 22](#_Toc186278880)

[7.1 对偶性质与 KKT 条件的引入 22](#_Toc186278881)

[7.2 KKT 条件 22](#_Toc186278882)

[7.3 KKT 条件的应用 23](#_Toc186278883)

[7.4 KKT 条件的局限性与实际应用 24](#_Toc186278884)

[本章小结 25](#_Toc186278885)

[第八章 牛顿法 (Newton's Method) 25](#_Toc186278886)

[本章导言: 25](#_Toc186278887)

[8.1 牛顿法的基本思想与迭代公式 25](#_Toc186278888)

[8.2 牛顿法的收敛性分析 26](#_Toc186278889)

[8.3 牛顿法的优缺点 26](#_Toc186278890)

[8.4 牛顿法的改进 26](#_Toc186278891)

[8.5 牛顿法与梯度下降法的比较 27](#_Toc186278892)

[8.6 应用举例 (结合前面章节) 27](#_Toc186278893)

[8.7 关于牛顿法中步长的一种理解方式 27](#_Toc186278894)

[习题 27](#_Toc186278895)

[本章小结 28](#_Toc186278896)

[第九章 次梯度 (Sub-gradient) 28](#_Toc186278897)

[本章导言: 28](#_Toc186278898)

[9.1 次梯度与次微分 28](#_Toc186278899)

[9.2 次梯度的例子 29](#_Toc186278900)

[9.3 次梯度运算法则 30](#_Toc186278901)

[9.4 次梯度方法 30](#_Toc186278902)

[9.5 次梯度方法的应用：Lasso 问题 30](#_Toc186278903)

[9.6 次梯度方法的优缺点 31](#_Toc186278904)

[习题 31](#_Toc186278905)

[本章小结 31](#_Toc186278906)

[第十章 迭代法与临近点算法 (Iterative Methods and Proximal Algorithms) 32](#_Toc186278907)

[本章导言: 32](#_Toc186278908)

[10.1 求解线性方程组的迭代方法 32](#_Toc186278909)

[10.2 临近点算法 (Proximal Algorithm) 33](#_Toc186278910)

[10.3 临近点梯度法 (Proximal Gradient Method) 34](#_Toc186278911)

[10.4 应用：Lasso 问题 34](#_Toc186278912)

[习题 35](#_Toc186278913)

[本章小结 35](#_Toc186278914)

[第十一章 练习题 35](#_Toc186278915)

[本章导言: 35](#_Toc186278916)

[练习题 1：向量空间与线性方程组 35](#_Toc186278917)

[练习题 2：线性方程组求解 36](#_Toc186278918)

[练习题 3：二次型与曲线拟合 36](#_Toc186278919)

[练习题 4：概率图模型 37](#_Toc186278920)

[练习题 5：最优性条件 39](#_Toc186278921)

[练习题 6：拉格朗日乘子法与 KKT 条件 39](#_Toc186278922)

[本章小结 40](#_Toc186278923)

[结束语 41](#_Toc186278924)

## 序章 数学史

### 数学的词源：

* 数学一词起源于希腊语“μαθηματικός (Mathematikós)”，意为“学问的基础”，其更早的词根“μάθημα (máthema)”意为“科学、知识或学问”。
* 数学的专业化使用可以追溯到毕达哥拉斯学派，这一学派首次明确将“数学”与数的研究及逻辑推理联系起来。

### 关键数学家与贡献：

* **泰勒斯 (Thales)**：提出了著名的“半圆内切角是直角”的几何理论，为后来的几何学奠定了基础。
* **毕达哥拉斯 (Pythagoras)**：在数学领域具有开创性贡献，特别是对数论和比例的研究。
* **欧几里得 (Euclid)**：以《几何原本》奠定了公理化方法，将几何系统化为一个逻辑体系。
* **阿基米德 (Archimedes)**：以数学的方式研究物理学问题，被称为古代数学的巅峰人物，对微积分的萌芽发展起到了重要作用。
* 微积分的创立者是**牛顿 (Isaac Newton)和莱布尼茨 (Gottfried Wilhelm Leibniz)**，他们独立地提出了这一伟大的数学工具。

### 数学史上的里程碑：

* 微积分的萌芽可以追溯到阿基米德的研究，他通过分割法接近曲线下的面积，间接为微积分的诞生奠定了思想基础。
* 在此后的1500多年间，数学从古希腊的几何学逐步扩展到代数和解析几何，为近代数学的快速发展打下了基础。
* 牛顿和莱布尼茨的微积分推动了数学的现代化发展，使数学在物理、天文等领域的应用取得重大突破。

### 数学发展的视角：

* 学生高中毕业时的数学水平相当于400年前的水平。
* 学完高等数学的学生可以达到150年前的数学水平。
* 学完本课程后，数学水平提升至约50年前的水平，同时掌握部分前沿内容。

### 名言引用：

* 普希金认为，数学是“跟随伟大人物的思想，是一门最引人入胜的科学”。

## 第一章 预备知识：数学证明方法与基本记号

### 本章导言:

本章作为教材的开篇，旨在为读者建立坚实的数学基础。我们将从最基本的数学证明方法入手，介绍常用的逻辑运算符，并详细讲解直接证明法、反证法、逆否命题证明法和数学归纳法等重要的证明技巧。此外，本章还将规范化数学符号的使用，包括标量、向量、矩阵和集合的表示方法，这些都是后续章节学习的基石。通过本章的学习，读者将掌握严谨的数学思维方式，并具备使用规范数学语言进行表达的能力。

### 1.1 数学证明方法

#### 1.1.1 命题与逻辑运算

* **命题**: 明确陈述的、可以判断真假的陈述句，例如 "2是偶数" (真命题) 或 "3是偶数" (假命题)。
* **逻辑运算符:**
  + **与 (and)**: 命题 A and B，当 A 和 B 都为真时，结果为真；否则为假。
  + **或 (or)**: 命题 A or B，当 A 或 B 至少有一个为真时，结果为真；当 A 和 B 都为假时，结果为假。
  + **非 (not)**: 命题 not A，当 A 为真时，结果为假；当 A 为假时，结果为真。
* **真值表**: 使用表格形式清晰地展示逻辑运算的结果。

|  |  |  |  |
| --- | --- | --- | --- |
| A | B | A and B | A or B |
| True | True | True | True |
| True | False | False | True |
| False | True | False | True |
| False | False | False | False |

|  |  |
| --- | --- |
| A | not A |
| True | False |
| False | True |

* **德摩根定律 (DeMorgan's Law):** not (A and B) 等价于 (not A) or (not B)

#### 1.1.2 蕴含关系与等价关系

* **蕴含 (implies)**： 命题 A implies B，当A为真时，B必然为真。常用符号： A ⇒ B
  + 等价表达: (not A) or B
  + 理解方式：将 A 分为 “成立” 和 “不成立” 两种情况进行讨论
  + 蕴含的几种常用表达方式：
    - If A then B
    - A only if B (B 成立时 A 才成立，反之 A 不成立则 B 也不成立)
    - A is sufficient for B （A 是 B 的充分条件）
    - B is necessary for A (B 是 A 的必要条件)
* **等价 (equivalent)**: 命题 A is equivalent to B，当A为真时，B也为真；当A为假时，B也为假。 常用符号：A ⇔ B
  + 等价表达： (A ⇒ B) and (B ⇒ A)
  + 也可以表达为 A if and only if B
  + 等价关系的证明通常需要证明双向的蕴含关系。

#### 1.1.3 数学证明的基本方法

* **直接证明法 (The direct method)**：从已知条件出发，通过逻辑推理，逐步推导出结论。
* **逆否命题证明法 (Proof by contraposition)**：要证明 A ⇒ B，可以转化为证明 (not B) ⇒ (not A)。
* **反证法 (Proof by contradiction)**：先假设结论不成立，然后通过逻辑推理，导出与已知条件或公理矛盾的结果，从而证明结论是成立的。
* **数学归纳法 (Principle of induction)**: 用于证明与自然数相关的命题。包含以下两个步骤：
  1. **基本情况 (Base Case)**：证明当 n = 1 时，命题成立。
  2. **归纳步骤 (Inductive Step)**：假设当 n = k 时，命题成立；然后证明当 n = k + 1 时，命题也成立。
  3. **注意事项**: 在使用数学归纳法时，一定要注意基本情况的验证，否则可能导致错误的结论。

### 1.2 基本记号

#### 1.2.1 标量、向量与矩阵

* **标量 (Scalar):** 用小写字母表示，例如：x, y, a, b。
* **向量 (Vector):** 用小写粗体字母表示，例如：**x**, **y**, **a**, **b**。一个列向量 x 可以表示为 x = (x1, x2, …, xn)T，其中 x1, x2, …, xn 是向量的分量。
* **矩阵 (Matrix):** 用大写粗体字母表示，例如：**A**, **B**。一个 m × n 的矩阵 **A** 可以表示为：

A = | a₁₁ a₁₂ ... a₁n |

| a₂₁ a₂₂ ... a₂n |

| ... ... ... ... |

| aₘ₁ aₘ₂ ... aₘn |

* **转置 (Transpose):** 向量或矩阵的转置用上标 T 表示。例如，列向量 **a** 的转置是行向量 **a**T = [a1, a2, ..., an]。矩阵 **A** 的转置是将矩阵的行和列互换。

#### 1.2.2 集合及其表示

* **集合 (Set):** 用大写花体字母表示，例如：X, Y。
* **集合元素 (Element):** 用小写字母表示，例如：x, y。
* **集合表示方法：**
  + 列举法: X = {x1, x2, …, xn}
  + 描述法: X = {x | x 满足某种性质}

### 本章小结

本章介绍了数学证明的基础知识和基本记号，这是进行高级工程数学学习的必要准备。通过对逻辑运算符、证明方法和数学符号的理解和掌握，读者将为后续章节的学习打下坚实的基础。

## 第二章 向量空间与矩阵

### 本章导言:

本章将深入探讨向量空间与矩阵的概念，这是高级工程数学中极为重要的基石。我们将介绍向量的定义、向量的基本运算，以及向量空间的概念和性质。随后，我们将详细讲解矩阵的定义、矩阵运算以及矩阵的秩等重要概念，并介绍线性方程组及其解的存在性和求解方法。此外，本章还将引入内积、范数等概念，它们是衡量向量和矩阵大小的重要工具，为后续的分析和计算打下基础。通过本章的学习，读者将掌握线性代数的基本理论和方法，为解决实际工程问题提供数学工具。

### 2.1 向量与矩阵

#### 2.1.1 向量的定义与表示

* **列向量 (Column vector):** n 个数的有序排列，表示为一个列的形式。
  + 例如： **a** = [a₁, a₂, …, aₙ]T
  + aᵢ 表示向量 **a** 的第 i 个分量。
* **行向量 (Row vector):** n 个数的有序排列，表示为一个行的形式。
  + 例如：[a₁, a₂, …, aₙ]
* **向量的相等:** 两个向量的对应元素均相等时，两向量相等。
* **转置 (Transpose):** 将列向量转化为行向量，或将行向量转化为列向量。 例如： **a**T = [a₁, a₂, …, aₙ]

#### 2.1.2 向量的基本运算

* **向量加法:** 两个向量对应分量相加。
  + **交换律 (Commutative law):** **a** + **b** = **b** + **a**
  + **结合律 (Associative law):** (**a** + **b**) + **c** = **a** + (**b** + **c**)
  + **零向量 (Zero vector):** 存在一个零向量 **0** = [0, 0, …, 0]T ，使得 **a** + **0** = **0** + **a** = **a**
* **向量的减法:** **a** - **b** = **a** + (-**b**)，其中 -**b** = **0** - **b** 。
  + **向量的差 (Difference):** 向量 **a** 与 **b** 的差是 [a₁ - b₁, a₂ - b₂, …, aₙ - bₙ]T 。
  + **性质：**
    - (- **b**) = **b**
    - -(**a** - **b**) = **b** - **a**
* **数乘 (Scalar multiplication):** 一个向量乘以一个标量。
  + 若 **a** ∈ Rⁿ，标量 α ∈ R，则 α**a** = [αa₁, αa₂, …, αaₙ]T
  + **分配律 (Distributive law):**
    - α(**a** + **b**) = α**a** + α**b**
    - (α + β)**a** = α**a** + β**a**
  + **结合律 (Associative law):** α(β**a**) = (αβ)**a**
  + **单位标量:** 存在单位标量 1 使得 1**a** = **a**
  + **向量 0:** 对于任意标量 α，有 α**0** = **0**
  + **零标量:** 对于任意向量 **a**，有 0**a** = **0**
* **数乘性质**:
  + α**a** = **0** ⇔ α = 0 或 **a** = **0**
  + 此性质可以推导出： 如果 α**a** = **0** 且 **a** ≠ **0**, 则 α = 0.

#### 2.1.3 向量的线性相关性与基

* **线性组合 (Linear combination):** 向量的线性组合是由向量乘以标量并相加得到的。 例如： α₁**a**₁ + α₂**a**₂ + ... + αₙ**a**ₙ
* **线性无关 (Linearly independent):** 一组向量，如果它们的任何非零线性组合都不等于零向量，则称这些向量线性无关。
  + 换句话说，如果 α₁**a**₁ + α₂**a**₂ + ... + αₙ**a**ₙ = **0** 仅当 α₁ = α₂ = ... = αₙ = 0 成立，则向量 **a**₁, **a**₂, …, **a**ₙ 线性无关。
* **线性相关 (Linearly dependent):** 如果一组向量不是线性无关的，则称它们是线性相关的，即存在一组不全为零的标量 α₁，α₂，…，αₙ，使得 α₁**a**₁ + α₂**a**₂ + ... + αₙ**a**ₙ = **0**
* **向量空间的基 (Basis):** 对于一个向量空间 V 的子空间，如果一组线性无关的向量 {**a**₁, …, **a**ₖ} 能够张成这个子空间，那么称这组向量为该子空间的一个基。也就是说，子空间内的任意向量都能用该基的线性组合表示。
* **向量空间的维数 (Dimension):** 一个向量空间的所有基包含的向量个数都相同，这个数就称为向量空间的维数，记为 dim V.
* **基的唯一性表示**: 对于 V 中的任何向量 a, 都可以被唯一地表示为 a = a₁a₁ + ... + aₖaₖ
* **坐标 (Coordinates):** 在一组基下，向量 a 的表示式 a = a₁a₁ + ... + aₖaₖ 中，系数 a₁, ..., aₖ被称为向量 a 在该基下的坐标。
* **标准基 (Natural basis):** Rⁿ 的标准基是一组特殊的基，由向量 e₁ = [1, 0, ..., 0]T，e₂ = [0, 1, ..., 0]T, ..., eₙ = [0, 0, ..., 1]T构成。

#### 2.1.4 矩阵的定义与表示

* **矩阵 (Matrix):** 由 m 行 n 列的数字组成的一个矩形阵列。一个 m × n 的矩阵可以表示为：
  + **A** = [aᵢⱼ]，其中 i = 1, 2, ..., m， j = 1, 2, ..., n
* **矩阵的转置 (Transpose):** 将矩阵的行和列互换。 例如， A = [aᵢⱼ]，则 AT = [aⱼᵢ]
* **矩阵的相等**: 两个矩阵的行数和列数都相等，且对应元素都相等时，两矩阵相等。

### 2.2 矩阵的秩

#### 2.2.1 矩阵的秩的定义

* 对于矩阵 A，其列向量构成的集合中，线性无关的列向量的最大数目称为矩阵 A 的秩，记为 rank A。
* 矩阵 A 的秩也是矩阵 A 列向量所张成的子空间的维度。
* 矩阵 A 的第 k 列记为 aₖ，则 A = [a₁, …, aₙ]

#### 2.2.2 矩阵秩的性质

* **秩的不变性 (Invariance of rank)**: 矩阵 A 的秩在下列操作下保持不变：
  1. 将矩阵 A 的列向量乘以非零标量。
  2. 交换矩阵 A 的列向量。
  3. 将矩阵 A 的一列加上其他列的线性组合。
* 矩阵 A 中，线性无关向量的个数与排列顺序无关。

#### 2.2.3 行列式 (Determinant)

* **行列式的定义**: 对于方阵A，可以定义一个标量，称为行列式，记为 det A 或 |A|。
* **行列式的性质**:
  + 矩阵的行列式是矩阵每列的线性函数
    - det[a₁, …, aₖ₋₁, α**a**ₖ(1) + β**a**ₖ(2), aₖ₊₁, ..., aₙ] = αdet[a₁, …, **a**ₖ(1), aₖ₊₁, ..., aₙ] + βdet[a₁, …, **a**ₖ(2), aₖ₊₁, ..., aₙ]
  + 如果矩阵 A 中存在两列向量相等，即 **a**ₖ = **a**ₖ₊₁ ， 则 det A = 0
  + 单位矩阵 Iₙ 的行列式等于 1，即 det Iₙ = 1
* **其他性质:**
  + det[a₁, ..., aₖ₋₁, aₖ + αaⱼ, aₖ₊₁, ..., aⱼ, ..., aₙ] = det[a₁, ..., aₙ]
* **p阶子式**: 一个 m x n 的矩阵 A 的 p 阶子式，其中 p<min{m,n}，指通过删除矩阵 m-p 行和 n-p 列后，得到的 p x p 方阵的行列式值
* **定理**: 如果一个 m x n (m > n) 的矩阵 A 有一个非零的 n 阶子式，则矩阵A 的列是线性无关的，即 rank A = n。

#### 2.2.4 非奇异矩阵

* **非奇异矩阵 (Nonsingular matrix):** 对于一个 n × n 的方阵 A，如果存在一个 n × n 的矩阵 B，使得 AB = BA = In，则称 A 是非奇异的，B 为 A 的逆矩阵，记为 B = A⁻¹

### 2.3 线性方程组

#### 2.3.1 线性方程组的表示

* **线性方程组:** 包含若干个线性方程的方程组。
  + 例如：

a₁₁x₁ + a₁₂x₂ + ... + a₁ₙxₙ = b₁

a₂₁x₁ + a₂₂x₂ + ... + a₂ₙxₙ = b₂

...

aₘ₁x₁ + aₘ₂x₂ + ... + aₘₙxₙ = bₘ

* + 可以表示为：**Ax = b**，其中 A 是系数矩阵， **x** 是未知向量，**b** 是常数向量。
  + 也可以表示为： x₁**a**₁ + x₂**a**₂ + ... + xₙ**a**ₙ = **b**，其中 **a**₁, …, **a**ₙ 是矩阵 A 的列向量。
* **增广矩阵 (Augmented matrix):** [**A**, **b**] = [**a**₁, …, **a**ₙ, **b**]。

#### 2.3.2 线性方程组解的存在性

* **定理 2.1:** 线性方程组 **Ax = b** 有解的充分必要条件是 rank **A** = rank [**A**, **b**]。
  + **证明:**
    - **充分性 (=>)**: 若 Ax = b 有解，则 **b** 可以表示为 **A** 的列向量的线性组合，从而 **b** ∈ span( **a**₁, …, **a**ₙ).因此， rank **A** = dim span( **a**₁, …, **a**ₙ) = dim span( **a**₁, …, **a**ₙ, **b**) = rank [**A**, **b**]
    - **必要性 (<=>)**: 若 rank **A** = rank [**A**, **b**] = r，则 **A** 的前 r 列线性无关。 由于 rank [**A**, **b**] = r，则 **b** 可以由 **A** 的前 r 列表示，因此， 存在向量 **x**，使得 **Ax** = **b**

#### 2.3.3 线性方程组的求解

* **定理 2.2:** 对于方程组 **Ax = b**，其中 A ∈ Rᵐˣⁿ，rank A = m，则方程组的解可以通过给定 n - m 个变量的任意值，然后求解剩余的 m 个变量得到。
  + 证明：
    - 可将方程组表示为 x₁**a**₁ + x₂**a**₂ + ... + xₘ**a**ₘ = **b** - xₘ₊₁**a**ₘ₊₁ - ... - xₙ**a**ₙ
    - 将 xₘ₊₁, ... , xₙ 设为任意值 dₘ₊₁, ... , dₙ
    - 构造矩阵 B = [**a**₁, ..., **a**ₘ]， 由于 rank A = m，因此 B 是可逆矩阵，此时方程组为 Bx = **b** - dₘ₊₁**a**ₘ₊₁ - ... - dₙ**a**ₙ
    - 解得 x = B⁻¹( **b** - dₘ₊₁**a**ₘ₊₁ - ... - dₙ**a**ₙ)

### 2.4 内积与范数

#### 2.4.1 绝对值

* **定义:** |a| 表示实数 a 的绝对值
* **性质:**
  1. |a| = |-a|
  2. -|a| ≤ a ≤ |a|
  3. |a + b| ≤ |a| + |b|
  4. ||a| - |b|| ≤ |a - b| ≤ |a| + |b|
  5. |ab| = |a||b|
  6. |a| < c and |b| < d imply that |a + b| < c + d
  7. |a| < b is equivalent to -b < a < b (i.e., a < b and -a < b)。此不等式对于 "<=" 也适用

#### 2.4.2 欧几里得内积

* **欧几里得内积 (Euclidean inner product)** 对于两个 Rⁿ 中的向量 x 和 y 定义为：
  + (x, y) = ∑ᵢ xᵢyᵢ = **x**T**y**
* **内积的性质:**
  + **正定性 (Positivity):** (x, x) > 0, 当且仅当 x = 0 时 (x, x) = 0
  + **对称性 (Symmetry):** (x, y) = (y, x)
  + **加法性 (Additivity):** (x + y, z) = (x, z) + (y, z)
  + **齐性 (Homogeneity):** (rx, y) = r(x, y)，其中 r ∈ R
  + 第二向量满足加法性和齐性:
  + (x, y + z) = (x, y) + (x, z)
  + (x, ry) = r(x, y), 其中r∈R
* **正交 (Orthogonal):** 如果 (x, y) = 0，则称向量 x 和 y 正交。

#### 2.4.3 欧几里得范数

* **欧几里得范数 (Euclidean norm)** 向量 x 的欧几里得范数定义为:
  + ||x|| = √<x, x> = √(xTx)
* **范数的性质：**
  + **正定性 (Positivity):** ||x|| > 0, 当且仅当 x = 0 时 ||x|| = 0
  + **齐性 (Homogeneity):** ||rx|| = |r|· ||x||，其中 r ∈ R
  + **三角不等式 (Triangle inequality):** ||x + y|| ≤ ||x|| + ||y||

#### 2.4.4 柯西-施瓦茨不等式 (Cauchy-Schwarz Inequality)

* 对于 Rⁿ 中的任意两个向量 x 和 y，有： |(x, y)| ≤ ||x|| · ||y||。等号成立当且仅当 x = αy, 其中 α∈R。

#### 2.4.5 p-范数 (p-norm)

* 对于向量 x ∈ Rⁿ，定义 p-范数为：
* ||x||ₚ = (|x₁|ᵖ + ... + |xₙ|ᵖ)¹/ᵖ, 当 1 ≤ p < ∞
* ||x||ₚ = max{|x₁|, ..., |xₙ|}, 当 p = ∞
* **p 范数可用于描述连续函数**

#### 2.4.6 复向量空间的内积

* 对于复数向量空间 Cⁿ， 内积定义为 (x, y) = ∑ᵢ xᵢy̅ᵢ
* 复数空间的内积性质：
  1. **正定性**: (x, x) > 0, 当且仅当 x = 0 时 (x, x) = 0
  2. **对称性**: (x, y) = (y, x) 的共轭
  3. **加法性**: (x + y, z) = (x, z) + (y, z)
  4. **齐性 (齐性):** (rx, y) = r(x, y) for every r∈ C
  5. 对于第二个变量满足: (x, r₁y + r₂z) = r̅₁(x, y) + r̅₂(x, z)

### 本章小结

本章深入探讨了向量空间和矩阵的基础知识，包括向量的运算、矩阵的表示和秩、线性方程组的解、以及内积和范数等重要概念。这些概念和方法为进一步学习高级工程数学奠定了基础，并为解决实际工程问题提供了数学工具。

## 第三章 线性变换、特征值与特征向量及正交投影

### 本章导言:

本章将继续深入探讨线性代数的核心概念，主要围绕线性变换、特征值与特征向量以及正交投影展开讨论。线性变换是向量空间之间保持线性关系的映射，它可以通过矩阵来表示。特征值和特征向量则揭示了线性变换的内在结构，它们描述了在变换下保持方向不变的向量及其对应的缩放因子。最后，正交投影则是一种特殊的线性变换，它将向量投影到子空间上，保持垂直性，是解决线性方程组、优化问题和信号处理等领域的重要工具。通过本章的学习，读者将对线性变换的几何意义、特征分解的概念以及投影变换的性质有深刻的理解，并能够灵活运用这些工具解决实际问题。

### 3.1 线性变换

#### 3.1.1 线性变换的定义

* **线性变换 (Linear transformation):** 从向量空间 Rⁿ 到向量空间 Rᵐ 的一个映射 L: Rⁿ → Rᵐ，如果满足以下两个条件，则称为线性变换：
  1. **齐次性 (Homogeneity):** L(α**x**) = αL(**x**)，对于任意 **x** ∈ Rⁿ，任意标量 α ∈ R 成立。
  2. **可加性 (Additivity):** L(**x**₁ + **x**₂) = L(**x**₁) + L(**x**₂)，对于任意 **x**₁, **x**₂ ∈ Rⁿ 成立。

#### 3.1.2 线性变换的矩阵表示

* 如果确定了 Rⁿ 和 Rᵐ 的基，那么线性变换 L 可以用一个矩阵表示。
* 若 **x** ∈ Rⁿ， **x'** 为 **x** 在 Rⁿ 的给定基下的表示； **y** = L(**x**)， **y'**为 **y** 在 Rᵐ 的给定基下的表示。则存在一个矩阵 A ∈ Rᵐˣⁿ，使得 y' = Ax'。此时称 A 为线性变换 L 的矩阵表示。
* 特别地，当假设 Rⁿ 和 Rᵐ 的基均为标准基时， 则有 L(**x**) = **Ax**

#### 3.1.3 变换矩阵与基的关系

* 设 {**e**₁, **e**₂, ..., **e**ₙ} 和 {**e'**₁, **e'**₂, ..., **e'**ₙ} 是 Rⁿ 的两组基，定义矩阵：
  + **T** = [**e'**₁, **e'**₂, ..., **e'**ₙ]⁻¹ [**e**₁, **e**₂, ..., **e**ₙ]
* 称 **T** 为从基 {**e**₁, **e**₂, ..., **e**ₙ} 到基 {**e'**₁, **e'**₂, ..., **e'**ₙ} 的变换矩阵。
* 显然有：[**e'**₁, **e'**₂, ..., **e'**ₙ] **T** = [**e**₁, **e**₂, ..., **e**ₙ]
* **T** 的第 i 列就是向量 eᵢ 在基 {**e'**₁, **e'**₂, ..., **e'**ₙ} 下的坐标。
* 对于任意向量 **v**，设 **x** 为其在 {**e**₁, ..., **e**ₙ} 下的坐标， **x'** 为其在 {**e'**₁, ..., **e'**ₙ} 下的坐标，则有： **x** = **Tx'**。
* **过渡矩阵 (Transition matrix):** 若 ( **η**₁, **η**₂, …, **η**ₙ) = (**ε**₁, **ε**₂, …, **ε**ₙ) C， 则称矩阵 C 为由基 **ε**₁, **ε**₂, …, **ε**ₙ 到基 **η**₁, **η**₂, …, **η**ₙ 的过渡矩阵。

### 3.2 特征值与特征向量

#### 3.2.1 特征值与特征向量的定义

* **特征值 (Eigenvalue):** 对于 n × n 的实数方阵 A，如果存在一个标量 λ (可以是复数) 和一个非零向量 **v**，使得 **Av** = λ**v**， 则称 λ 为矩阵 A 的一个特征值。
* **特征向量 (Eigenvector):** 满足 **Av** = λ**v** 的非零向量 **v** 为矩阵 A 的属于特征值 λ 的一个特征向量。

#### 3.2.2 特征方程

* 要使 λ 为 A 的特征值， 必须并且只须矩阵 λI - A 为奇异矩阵， 即 det[λI - A] = 0, 其中 I 是 n x n 的单位矩阵。
* det[λI - A] = 0 是关于 λ 的一个 n 次多项式方程，称为 **特征方程 (characteristic equation)**。det[λI - A]称为矩阵 A 的 **特征多项式 (characteristic polynomial)** 。
* n 阶特征方程有 n 个复数根（可能相同）。
* 若有 n 个不同的特征根，则有 n 个线性无关的特征向量。

#### 3.2.3 实对称矩阵的特征值与特征向量

* **定理 3.2：** 实对称矩阵(A=Aᵀ)的所有特征值都是实的。
  + **证明:** 若 **Ax** = λ**x**, **x** ≠ 0，则有 ( **Ax**, **x** ) = λ<**x**, **x** >。另一方面, ( **Ax**, **x** ) = (**x**, **A**T**x**) = (**x**, **Ax**) = λ( **x**, **x** ) 。 由于 ( **x**, **x** ) 是实数并且>0，因此 λ = λ̄，即 λ 是实的。
* **定理 3.3：** 任意 n × n 实对称矩阵具有 n 个相互正交的特征向量.
* **证明**: 此处只证 n 个特征值不同的情形。
  + 设 **Av**₁ = λ₁**v**₁, **Av**₂ = λ₂**v**₂, 其中 λ₁ ≠ λ₂，则有：< **Av**₁, **v**₂> = λ₁<**v**₁, **v**₂>。
  + 由于 A = Aᵀ，则 < **Av**₁, **v**₂> = <**v**₁, Aᵀ**v**₂> = <**v**₁, **Av**₂> = λ₂<**v**₁, **v**₂>
  + 因此， λ₁<**v**₁, **v**₂> = λ₂<**v**₁, **v**₂> ，由于 λ₁ ≠ λ₂，则 <**v**₁, **v**₂> = 0。 \* 当特征值 λ₁ ≠ λ₂时， 对应的特征向量是正交的。

### 3.3 正交投影

#### 3.3.1 子空间和正交补

* **子空间 (Subspace):** 向量空间 Rⁿ 的一个子集 V，如果满足以下条件，则称为 Rⁿ 的一个子空间：
  + 若 **x**₁, **x**₂ ∈ V， 则 α**x**₁ + β**x**₂ ∈ V, 对于任意 α, β ∈ R 成立
* **子空间的维度 (Dimension):** 一个子空间 V 的维度等于子空间中线性无关向量的最大个数。
* **正交补 (Orthogonal complement):** 对于 Rⁿ 的一个子空间 V，其正交补记为 V⊥，V⊥ 由所有与 V 中所有向量正交的向量构成。即 V⊥ = {**x**: **v**T**x** = 0 for all **v** ∈ V}
  + V 和 V⊥ 张成 Rⁿ， 即 ∀**x** ∈ Rⁿ 可唯一表示为 **x** = **x**₁ + **x**₂，其中 **x**₁ ∈ V, **x**₂ ∈ V⊥。称为正交分解。
* **正交投影 (Orthogonal projection):** **x**₁ 和 **x**₂ 分别是向量 **x** 在子空间 V 和 V⊥ 上的正交投影。
* Rⁿ = V ⊕ V⊥ (直和)

#### 3.3.2 正交投影的定义

* **正交投影矩阵 (Orthogonal projector):** 线性变换 P 是一个到子空间 V = R(P) 的正交投影矩阵 如果对于任意向量 **x** ∈ Rⁿ, 有 P**x** ∈ V。
* **定理 3.5：** 矩阵 P 是正交投影矩阵 当且仅当 P² = P = Pᵀ
  + **证明:**
    - 利用 x = Px + (x - Px)
    - R(P) ⊥ = N(Pᵀ)
    - => 如果 P 是正交投影， 则 R(I-P) ⊆ R(P)⊥ = N(Pᵀ). 所以 Pᵀ (I-P) = O, Pᵀ = PᵀP。即 P=Pᵀ=P² \* <= 如果 P=Pᵀ=P²，对于任意 **x**，(P**y**)ᵀ(I − P)**x** = **y**ᵀ Pᵀ (I − P) **x** = **y**ᵀ P(I − P)**x** = 0 。 因此，(I-P)**x** ∈ R(P)⊥.

### 本章小结

本章介绍了线性变换、特征值与特征向量以及正交投影等概念，这些概念不仅是线性代数的核心内容，也是其他数学领域和工程应用的重要基础。理解这些概念将有助于读者更好地分析和解决实际问题。

## 第四章 几何概念：线段、超平面与凸集

### 本章导言:

本章将介绍一些基本的几何概念，这些概念在优化问题、机器学习等领域有着广泛的应用。我们将首先定义线段的概念，这是构建更复杂几何结构的基石。然后，我们将深入探讨超平面和线性簇的概念，这它们是高维空间中线性关系的重要体现，也是许多优化问题的约束条件。最后，我们将介绍凸集及其相关性质，凸集是一类重要的几何对象，在优化理论中具有重要的地位，因为局部最优解通常也是全局最优解。理解这些几何概念，将有助于读者从几何角度认识和分析实际问题，从而更好地理解和应用相关的数学工具。

### 4.1 线段

* **线段的定义:** 在 Rⁿ 空间中，连接两个点 **x** 和 **y** 的线段，是位于连接这两点的直线上的点的集合。
  + 如果点 **z** 在 **x** 和 **y** 之间的线段上，则 **z** - **y** = α( **x** - **y**), 其中 α ∈ [0, 1]。
  + 也可以写为 **z** = α**x** + (1 - α)**y**, 其中 α ∈ [0, 1]。
  + 线段可以用集合表示为：{ α**x** + (1 - α)**y** : α ∈ [0,1] }

### 4.2 超平面与线性簇

#### 4.2.1 超平面的定义

* **超平面 (Hyperplane):** 在 Rⁿ 空间中，超平面是由满足线性方程的所有点组成的集合。
  + 方程形式： u₁x₁ + u₂x₂ + ... + uₙxₙ = v, 其中 u₁, u₂,..., uₙ 为实数，至少一个非零，**x** = [x₁, x₂, ..., xₙ]ᵀ ∈ Rⁿ，v 为实数。
  + 向量形式： { **x** ∈ Rⁿ: **u**ᵀ**x** = v}
* **法向量 (Normal):** 向量 **u** 称为超平面的法向量， 它与超平面内任意两个向量的差，也即超平面内任意向量正交。
* 如果 **a** 是超平面上的任意一点，则 **u**ᵀ( **x** - **a** ) = 0，也可以理解为 **u** 和 **x** - **a** 是相互正交的。
  + **超平面的法向量**： 对于超平面 **u**ᵀ **x** = v, **u** 是法向量，代表了超平面方向，而 v 控制了超平面与原点的距离。
* **半空间 (Half-space):** 超平面将空间分为两个半空间。
  + **正半空间 (Positive half-space):** 由满足不等式 u₁x₁ + ... + uₙxₙ ≥ v 的点组成，表示为 H₊ = {**x** ∈ Rⁿ: **u**ᵀ**x** ≥ v}
  + **负半空间 (Negative half-space):** 由满足不等式 u₁x₁ + ... + uₙxₙ < v 的点组成，表示为 H₋ = {**x** ∈ Rⁿ: **u**ᵀ**x** < v}

#### 4.2.2 线性簇

* **线性簇 (Linear variety):** 由满足线性方程组 **Ax = b** 的所有点构成的集合。表示为：
  + { **x** ∈ Rⁿ: **Ax = b** }
  + 其中 A ∈ Rᵐˣⁿ，b∈ Rᵐ。
* **线性簇的维度 (Dimension):** 如果 dim N(A) = r, 我们称该线性簇的维度为 r。
* 当 **b** = 0 时，线性簇是一个 **子空间**
* 当 A = O 时, 线性簇是 Rⁿ。
* 维度小于n的线性簇可以被看作有限个超平面的交集。

### 4.3 凸集

#### 4.3.1 凸集的定义

* **凸组合 (Convex combination):** 两个点 **u** 和 **v** 的凸组合是 w = α**u** + (1 − α)**v**, 其中 α ∈ [0,1]。
* **凸集 (Convex set):** 如果对于任意 **u**, **v** ∈ Θ，连接 **u** 和 **v** 的线段上的所有点都属于 Θ，则称集合 Θ 是凸集。即若 α**u** + (1-α)**v** ∈ Θ 对于 ∀**u**, **v** ∈ Θ， ∀α ∈ [0,1]都成立，则称 Θ 为凸集。

#### 4.3.2 凸集的性质

* **定理 4.1：** 凸集具有以下性质：
  1. 如果 Θ 是凸集，且 β 是实数，则 βΘ = { **x**: **x** = β**v**, **v** ∈ Θ }也是凸集。
  2. 如果 Θ₁ 和 Θ₂ 是凸集，则 Θ₁ + Θ₂ = { **x**: **x** = **v**₁ + **v**₂, **v**₁ ∈ Θ₁, **v**₂ ∈ Θ₂ } 也是凸集。
  3. 任何凸集的集合的交集仍然是凸集。

#### 4.3.3 凸集的例子

* 空集是凸集。
* 只包含一个点的集合是凸集。
* 直线或线段是凸集。
* 子空间是凸集。
* 超平面是凸集。
* 线性簇是凸集。
* 半空间是凸集。
* Rⁿ 是凸集。

#### 4.3.4 凸集的极点

* **极点 (Extreme point):** 凸集 Θ 中的一个点 **x** 如果无法被表示为 Θ 中其他两个不同点的凸组合，则称 **x** 为极点。
  + 例如：圆的边界上的点是极点， 凸多边形的顶点是极点。

### 4.4 邻域

* **邻域 (Neighborhood):** 点 **x** ∈ Rⁿ 的邻域是指以 **x** 为中心，半径为 ε 的球形区域。
  + 数学表示： {**y** ∈ Rⁿ : ||**y** - **x**|| < ε}, 其中 ε 为正实数。
  + 在二维平面 R² 上，一个邻域是以 **x** 为圆心的圆盘。
  + 在三维空间 R³ 中，一个邻域是以 **x** 为球心的球体。
* **内点 (Interior point):** 如果一个点 **x** ∈ S ， 并且集合 S 包含 **x** 的一个邻域，则称 **x** 为 S 的一个内点。换句话说，S 中的每一个点都是其内点，就称 S 是开集(open set)。
  + 开集不包含任何边界点。
* **边界点 (Boundary point):** 一个点 **x** 是集合 S 的边界点，如果 **x** 的每个邻域都包含 S 中的点，也包含 S 外的点。
  + 边界点可能属于 S, 也可能不属于 S。
  + 所有边界点的集合构成了集合 S 的边界。
* **闭集 (Closed set):** 如果一个集合包含其所有边界点，则称该集合是闭集。
  + 一个集合是闭集当且仅当其补集是开集。
* **有界集 (Bounded set):** 如果一个集合能够包含在一个有限半径的球内，则称该集合是有界的。
* **紧集 (Compact set):** 如果一个集合是闭集且是有界的，则该集合称为紧集。
  + 紧集在优化问题中非常重要，因为可以保证最值点的存在，例如： 连续函数在紧集上必定有最大值和最小值 (Weierstrass 定理)。
    - **定理 4.2 维尔斯特拉斯定理 (Theorem of Weierstrass):** 对于一个连续函数 f: Ω → R，其中 Ω ⊂ Rⁿ 是一个紧集，则必定存在一个点 x₀ ∈ Ω，使得 f(x₀) ≤ f(x) 对所有 x ∈ Ω 成立，即函数 f 在 Ω 上可以取得最小值。

### 4.5 多面体与多胞体

* **凸多胞体 (Convex polytope):** 可以表示成有限个半空间的交集，称该集合为凸多胞体。
* **多面体 (Polyhedron):** 有界的凸多胞体称为多面体。

### 本章小结

本章介绍了线段、超平面、凸集以及邻域等基本的几何概念。这些概念为后续讨论优化问题和机器学习算法奠定了基础。理解这些几何结构能够帮助读者从几何的角度理解数学，从而为解决复杂问题提供新的视角。

## 第五章 微积分基础：序列、极限与微分

### 本章导言:

本章将介绍微积分的基本概念，为后续的优化算法和理论提供必要的工具。我们将从序列及其极限的概念入手，定义单调序列、有界序列和收敛序列。然后，我们将详细讲解函数的可微性，包括导数矩阵的概念和求导法则。最后，我们将引入水平集和梯度，以及泰勒级数等重要概念，这些工具在优化问题的分析和求解中扮演着关键角色。通过本章的学习，读者将掌握微积分的基本原理和运算方法，为深入学习优化理论奠定坚实的数学基础。

### 5.1 序列与极限

#### 5.1.1 实数序列的定义与分类

* **实数序列 (Sequence of real numbers):** 定义域为自然数集 N (1, 2, 3, …)，值域包含在实数集 R 的函数。可表示为 {x₁, x₂, x₃, ...} 或者 {xₖ} 或 {xₖ}k=1∞
* **递增序列 (Increasing sequence):** 对于所有 k，有 xₖ < xₖ₊₁。
* **非递减序列 (Nondecreasing sequence):** 对于所有 k，有 xₖ ≤ xₖ₊₁。 \* **递减序列 (Decreasing sequence):** 对于所有 k，有 xₖ > xₖ₊₁ \* **非递增序列 (Nonincreasing sequence):** 对于所有 k，有 xₖ ≥ xₖ₊₁ \* **单调序列 (Monotone sequences):** 非递增或非递减的序列统称为单调序列。

#### 5.1.2 实数序列的极限

* **极限 (Limit):** 实数序列 {xₖ} 的极限是一个数 x*， 如果对于任意 ε > 0，存在一个整数 K (可能取决于 ε)，使得当 k > K 时，有 |xₖ - x*| < ε。 \* 可以表示为：x\* = limₖ→∞ xₖ 或者 xₖ → x\*

#### 5.1.3 高维空间序列及其极限

* **高维空间序列**: Rⁿ 中的序列可以用 {**x**⁽¹⁾, **x**⁽²⁾, ... } 或者 {**x**⁽ᵏ⁾} 表示，其中 **x**⁽ᵏ⁾ ∈ Rⁿ。
* **高维空间序列的极限:** 对于 Rⁿ 中的序列 {**x**⁽ᵏ⁾}，如果对于任意 ε > 0，存在一个整数 K (可能取决于 ε)，使得当 k > K 时，有 ||**x**⁽ᵏ⁾ - **x\***|| < ε， 则称 **x\*** 为序列 {**x**⁽ᵏ⁾} 的极限。
* 可以表示为: **x\*** = limₖ→∞ **x**⁽ᵏ⁾ 或者 **x**⁽ᵏ⁾ → **x\***

#### 5.1.4 收敛序列的性质

* **唯一性定理 (Theorem 5.1):** 如果一个序列收敛，它的极限是唯一的。
* **有界性定理 (Theorem 5.2):** 每一个收敛序列都是有界的。
* **子序列收敛性定理 (Theorem 5.4):** 如果一个序列收敛到某个极限，那么该序列的任何子序列也都收敛到相同的极限。
* **单调有界收敛定理 (Theorem 5.3):** R 中的每个单调有界序列都是收敛的。

#### 5.1.5 连续函数与极限

* **函数在一点的连续性:** 如果对于任意收敛到 x₀ 的序列{x⁽ᵏ⁾} ，都有 limₖ→∞ f(x⁽ᵏ⁾) = f(x₀), 則稱函数 f 在 x₀ 处是连续的。
* 如果函数 f 在 x₀ 处连续，我们可以使用 limx→x₀ f(x) = f(x₀)来表示。

#### 5.1.6 矩阵序列的极限

* **矩阵序列的极限:** 如果一个 m × n 矩阵的序列 {**A**ₖ} 的所有元素组成的实数序列都收敛到对应位置的矩阵 **A** 的元素， 则称 矩阵序列 {**A**ₖ} 收敛到矩阵 **A**。 \* 数学表示为： limₖ→∞ ||**A** - **A**ₖ|| = 0
* **引理 5.1:** 对于任意 A ∈ Rⁿˣⁿ, limₖ→∞ **A**ᵏ = 0 当且仅当 A 的特征值的绝对值都小于 1。

### 5.2 可微性

#### 5.2.1 仿射函数

* **仿射函数 (Affine function):** 从 Rⁿ 到 Rᵐ 的一个函数 **A**: Rⁿ → Rᵐ，如果存在一个线性变换 L: Rⁿ → Rᵐ 以及一个向量 **y** ∈ Rᵐ，使得对于所有 **x** ∈ Rⁿ, 有 **A**(**x**) = L(**x**) + **y**， 则称 **A** 为仿射函数。

#### 5.2.2 可微性的定义

* 若要用仿射函数在 x₀ 附近逼近函数 f: Rⁿ → Rᵐ，则需要保证：
  1. **A**(**x₀**) = f(**x₀**)
  2. 当 **x** 接近 **x₀** 时， **A**(**x**) 趋近 f(**x**) 的速度要快于 **x** 趋近 **x₀** 的速度。 \* 数学上，就是让误差 || f(**x**) - **A**(**x**)|| 与 ||**x**-**x₀**|| 的比值在 **x** 趋近 **x₀** 时趋于 0 。
  3. 即 limx→x₀ || f(x) - A(x)||/||x - x₀|| = 0

### 5.3 导数矩阵

#### 5.3.1 导数矩阵的定义

* 对于一个可微函数 f: Rⁿ → Rᵐ，在给定点 x₀，其导数可以用一个 m × n 的矩阵来表示，记作 Df(x₀)，称为**导数矩阵 (Derivative matrix)**。
* 将 f : Rⁿ → Rᵐ 看作 f = [f₁, f₂, ... , fm]ᵀ , 则导数矩阵 Df(x₀) 的第 j 列是 **L** **e**ᵢ , 其中 **e**ᵢ 为 Rⁿ 的标准基的第 i 列。 也就是函数在 x₀ 处沿着第 j 个坐标轴方向的偏导数向量。
  + 数学表示：
    - Df(x₀) 的第 j 列 = limₜ→₀ [f(x₀ + t**e**ⱼ) - f(x₀)]/t。
    - Df(x₀) = [∂f/∂x₁, ∂f/∂x₂,..., ∂f/∂xₙ]
* **梯度的定义:** 当 f: Rⁿ → R 时，梯度记为 ∇f(x)，是导数矩阵 Df(x)的转置。 \* ∇f(x) = [∂f/∂x₁, ∂f/∂x₂,..., ∂f/∂xₙ] ᵀ

#### 5.3.2 导数矩阵与函数的关系

* 一个可微函数 f : Rⁿ → Rᵐ 在 x₀ 点的最佳仿射逼近为： A(x) = f(x₀) + Df(x₀)(x − x₀)
* 一个函数在某点可微，该点的导数矩阵是唯一的。
* 导数矩阵 Df(x₀) 的列是向量偏导数。 \* 向量 ∂f/∂xⱼ(x₀) 是在点 x₀ 处，沿着第 j 个坐标轴方向的切向量。

#### 5.3.3 海森矩阵

* **二阶可微函数：** 如果函数 f: Rⁿ → R 的梯度 ∇f 是可微的，则称 f 为二阶可微的。
* **海森矩阵 (Hessian Matrix):** 二阶可微函数 f 在 x 处的二阶导数用一个 n × n 的矩阵表示，称为海森矩阵，记为 D²f(x) 或 F(x) \* D² f(x) = [ ∂²f/(∂xᵢ∂xⱼ)]。 \* (∂²f/∂xᵢ∂xⱼ 是 ∂f/∂xⱼ 关于 xᵢ 的偏导数)
* 如果函数 f 在 x 处二阶连续可微，则海森矩阵是对称的，满足 Clairaut 定理（也叫Schwarz定理），即 ∂²f/∂xᵢ∂xⱼ = ∂²f/∂xⱼ∂xᵢ
* 当二阶偏导数不连续时，海森矩阵不一定是对称的。

### 5.4 求导法则

#### 5.4.1 链式法则 (Chain rule)

* 对于函数 f: R → Rⁿ 和 g: Rⁿ → R，复合函数 h(t) = g(f(t)) 的导数 h'(t) = ∇g(f(t))·f'(t)
* 更一般地，h'(t) = Dg(f(t)) Df(t)

#### 5.4.2 乘积法则

* 对于可微函数 f : Rⁿ → Rᵐ 和 g: Rⁿ → Rᵐ， 定义 h : Rⁿ → R 为 h(**x**) = f(**x**)ᵀ g(**x**) , 则 Dh(x) = f(x)ᵀDg(x) + g(x)ᵀDf(x)。

#### 5.4.3 常用导数公式

* D(yᵀ Ax) = yᵀ A , 其中 y ∈ Rᵐ ， A ∈ Rᵐˣⁿ
* D(xᵀ Ax) = xᵀ (A + Aᵀ) , 如果m = n
* D(xᵀ y) = yᵀ ,其中 y ∈ Rⁿ， 如果 y 与 x 无关
* D(xᵀ Qx) = 2xᵀ Q ,如果 Q 是对称矩阵
* D(xᵀx) = 2xᵀ

### 5.5 水平集与梯度

#### 5.5.1 水平集

* **水平集 (Level set):** 函数 f: Rⁿ → R 的水平集是指函数值等于某个常数的点的集合。
  + 即 S = { **x** : f(**x**) = c}。
  + 当 f : R² → R 时，水平集 S 通常是一条曲线。
  + 当 f : R³ → R 时，水平集 S 通常是一个曲面。

#### 5.5.2 梯度与水平集的关系

* 如果存在一条曲线 y 位于 水平集S 上, 且参数化表示为函数 g : R → Rⁿ， 那么 f(g(t))=c。若 g(t₀) = x₀, 且g' (t₀) = v， 则切向量 v 应该与梯度 ∇f(x₀) 正交。
* **定理：** 梯度 ∇f(x₀) 正交于过点 x₀ 的水平集的切线方向。 \* 梯度 ∇f(**x**) 指向函数值增长最快的方向。
* 梯度 -∇f(**x**) 指向函数值下降最快的方向，即最速下降方向。

### 5.6 泰勒级数

* **泰勒定理 (Taylor's Theorem):** 如果函数 f: R → R 在区间 [a, b] 上 m 次连续可微，则： \* f(b) = f(a) + f'(a)(b-a)/1! + f''(a)(b-a)²/2! + ... + f⁽ᵐ⁻¹⁾(a)(b-a)ᵐ⁻¹/(m-1)! + Rm \* 余项 Rm = f⁽ᵐ⁾(a+θh) hᵐ (1-θ)ᵐ⁻¹ / (m - 1)!, 或 Rm = f⁽ᵐ⁾(a+θ’h) hᵐ / m! \* 其中 h = b - a, θ, θ' ∈ (0,1) 。
* 泰勒定理也可以推广到多元函数 \* 对于多元函数 f，其在 x₀ 处的泰勒展开式为:
  + f(x) = f(x₀) + ∇f(x₀)ᵀ (x-x₀) + (x-x₀)ᵀ D²f(x₀) (x-x₀) /2! + o(||x-x₀||²)
* **中值定理:** 如果 f: Rⁿ → Rᵐ 在开集 Ω 内可微，那么对于 Ω 中任意两点 x 和 y，存在矩阵 **M**， 使得 f(x) - f(y) = **M**(x-y)
* **M** 的每一行，均为 Df 在联结 x 和 y 路径上的点的导数。

### 本章小结

本章介绍了微积分的基本概念，包括序列及其极限，函数的可微性、导数矩阵、梯度、水平集和泰勒展开等。这些概念是优化算法的基础，也是理解许多高级数学概念的基石。

## 第六章 约束优化问题的最优性条件

### 本章导言:

本章将深入探讨约束优化问题的最优性条件，这是优化理论的核心内容。我们将首先介绍约束优化问题的基本概念和分类，然后重点讨论一阶必要条件（FONC）和二阶必要条件（SONC），以及二阶充分条件（SOSC）。这些条件为判断一个点是否为局部最优解提供了理论依据。在此基础上，我们将详细介绍拉格朗日乘子法，这是一种将约束优化问题转化为无约束优化问题的重要方法。通过本章的学习，读者将掌握判断约束优化问题最优解的条件和方法，并能够运用拉格朗日乘子法解决实际的优化问题。

### 6.1 约束优化问题概述

#### 6.1.1 约束优化问题的定义

* **约束优化问题 (Constrained optimization problem):** 在一个给定的集合 Ω (称为可行域) 中找到一个点 **x\***，使得目标函数 f(**x**) 取得最小值或最大值。
  + 数学表示：
  + minimize f(x)
  + subject to x ∈ Ω

其中 f: Rⁿ → R 是目标函数，Ω ⊆ Rⁿ 是可行域。

* **可行域 (Feasible region):** 满足所有约束条件的点的集合。
* **局部极小点 (Local minimizer):** 如果存在 **x\*** 的一个邻域 N，使得对于所有 **x** ∈ N ∩ Ω 且 **x** ≠ **x\***，都有 f(**x**) ≥ f(**x\***)，则称 **x\*** 为一个局部极小点。
* **严格局部极小点 (Strict local minimizer):** 如果存在 **x\*** 的一个邻域 N，使得对于所有 **x** ∈ N ∩ Ω 且 **x** ≠ **x\***，都有 f(**x**) > f(**x\***)，则称 **x\*** 为一个严格局部极小点。
* **全局极小点 (Global minimizer):** 如果对于所有 **x** ∈ Ω，都有 f(**x**) ≥ f(**x\***)，则称 **x\*** 为一个全局极小点。

#### 6.1.2 约束优化问题的分类

* **等式约束 (Equality constraints):** 约束条件为等式形式，例如 hᵢ(**x**) = 0。
* **不等式约束 (Inequality constraints):** 约束条件为不等式形式，例如 gᵢ(**x**) ≤ 0。
* **线性约束 (Linear constraints):** 约束函数为线性函数。
* **非线性约束 (Nonlinear constraints):** 约束函数为非线性函数。

### 6.2 最优性条件

#### 6.2.1 一阶必要条件 (First-Order Necessary Condition, FONC)

* **可行方向 (Feasible direction):** 在可行域 Ω 中的点 **x** 处，如果存在一个向量 **d** 和一个正数 α₀，使得对于所有 α ∈ [0, α₀]，都有 **x** + α**d** ∈ Ω，则称 **d** 为在 **x** 处的一个可行方向。
* **定理 6.1 (FONC):** 如果 **x\*** 是约束优化问题的一个局部极小点，且 **d** 是 **x\*** 处的一个可行方向，则 ∇f(**x\***)ᵀ**d** ≥ 0。
  + **证明:** 类似于无约束情况，利用泰勒展开和反证法。
    - 定义 φ(α) = f(**x\*** + α**d**)，其中 **d** 是可行方向。
    - 将 φ(α) 在 α = 0 处进行泰勒展开：f(**x\*** + α**d**) = f(**x\***) + α∇f(**x\***)ᵀ**d** + o(α)
    - 如果 ∇f(**x\***)ᵀ**d** < 0，则对于足够小的 α > 0，有 f(**x\*** + α**d**) < f(**x\***)，与 **x\*** 是局部极小点矛盾。
* **推论:** 如果 **x\*** 是可行域内部的一个局部极小点，则 ∇f(**x\***) = 0。
  + 因为此时任意方向都是可行方向，可以取 **d** = -∇f(**x\***)。

#### 6.2.2 二阶必要条件 (Second-Order Necessary Condition, SONC)

* **定理 6.2 (SONC):** 如果 **x\*** 是约束优化问题的一个局部极小点，**d** 是 **x\*** 处的一个可行方向，且 ∇f(**x\***)ᵀ**d** = 0，则 **d**ᵀ∇²f(**x\***)**d** ≥ 0。
  + **证明:** 同样利用泰勒展开和反证法。
    - 将 φ(α) 在 α = 0 处进行二阶泰勒展开：f(**x\*** + α**d**) = f(**x\***) + α∇f(**x\***)ᵀ**d** + (α²/2)**d**ᵀ∇²f(**x\***)**d** + o(α²)
    - 由于 ∇f(**x\***)ᵀ**d** = 0，如果 **d**ᵀ∇²f(**x\***)**d** < 0，则对于足够小的 α > 0，有 f(**x\*** + α**d**) < f(**x\***)，与 **x\*** 是局部极小点矛盾。
* **推论:** 如果 **x\*** 是可行域内部的一个局部极小点，则 ∇f(**x\***) = 0 且 ∇²f(**x\***) 是半正定矩阵。

#### 6.2.3 二阶充分条件 (Second-Order Sufficient Condition, SOSC)

* **定理 6.3 (SOSC):** 设 **x\*** 是可行域 Ω 的一个内点，如果 ∇f(**x\***) = 0 且 ∇²f(**x\***) 是正定矩阵，则 **x\*** 是一个严格局部极小点。
  + **证明:** 利用泰勒展开和正定矩阵的性质。
  + 将 f(x) 在 x\* 处进行泰勒展开： f(x) = f(x*) + ∇f(x*)ᵀ(x-x*) + (1/2)(x-x*)ᵀ∇²f(x*) (x-x*) + o(||x-x\*||²)
  + 由于∇f(x*) = 0， ∇²f(x*) 正定，则存在 λₘᵢₙ > 0 使得 (x-x*)ᵀ∇²f(x*) (x-x*) > λₘᵢₙ||x-x*||²
  + 于是 f(x) - f(x*) = (1/2)(x-x*)ᵀ∇²f(x*) (x-x*) + o(||x-x*||²) ≥ (λₘᵢₙ/2)||x-x*||² + o(||x-x\*||²) > 0
  + 因此 f(x) > f(x\*)
* **注意:** 对于边界点，SOSC 需要更复杂的形式。

### 6.3 拉格朗日乘子法 (Lagrange Multipliers)

#### 6.3.1 等式约束优化问题

* 考虑如下等式约束优化问题：
* minimize f(x)
* subject to hᵢ(x) = 0, i = 1, 2, ..., m

其中 f: Rⁿ → R，hᵢ: Rⁿ → R。

* **拉格朗日函数 (Lagrangian):** L(**x**, **λ**) = f(**x**) + Σᵢ λᵢhᵢ(**x**)，其中 **λ** = [λ₁, λ₂, ..., λₘ]ᵀ 称为拉格朗日乘子。
* **定理 6.4 (拉格朗日乘子法):** 如果 **x\*** 是上述等式约束优化问题的一个局部极小点，并且 hᵢ(**x**) 在 **x\*** 处线性无关 (即 ∇hᵢ(**x\***) 线性无关)，则存在一组拉格朗日乘子 **λ\***，使得：
  + ∇xL(**x\***, **λ\***) = ∇f(**x\***) + Σᵢ λᵢ*∇hᵢ(****x***) = 0
  + ∇λL(**x\***, **λ\***) = hᵢ(**x\***) = 0, i = 1, 2, ..., m
  + **几何解释:** 在最优点 **x\*** 处，目标函数的梯度 ∇f(**x\***) 可以表示为约束函数梯度 ∇hᵢ(**x\***) 的线性组合，即目标函数的等值线与约束曲面相切。

#### 6.3.2 拉格朗日乘子的意义

* 拉格朗日乘子 λᵢ\* 表示约束条件 hᵢ(**x**) = 0 发生微小变化时，目标函数最优值的变化率。
  + 假设 hᵢ(**x**) = 0 变为 hᵢ(**x**) = εᵢ，则最优值的变化近似为 Σᵢ λᵢ\*εᵢ。

#### 6.3.3 不等式约束优化问题 (简要介绍)

* 对于不等式约束 gᵢ(**x**) ≤ 0，可以引入松弛变量将其转化为等式约束，然后应用拉格朗日乘子法。
* **KKT 条件 (Karush-Kuhn-Tucker conditions):** 推广的拉格朗日乘子法，用于处理不等式约束。

### 本章小结

本章介绍了约束优化问题的最优性条件，包括一阶必要条件、二阶必要条件和二阶充分条件，以及将约束优化问题转化为无约束优化问题的拉格朗日乘子法。这些理论和方法为解决实际的约束优化问题提供了有力的工具。

## 第七章 不等式约束优化问题的最优性条件：KKT条件

### 本章导言:

本章将重点介绍处理不等式约束优化问题的关键工具——卡罗需-库恩-塔克条件 (Karush-Kuhn-Tucker Conditions)，简称 KKT 条件。KKT 条件是一阶必要条件的推广，它不仅适用于等式约束，也适用于不等式约束。我们将首先回顾对偶性质引出KKT条件的重要性，然后详细介绍 KKT 条件的具体内容及其推导过程，包括互补松弛条件、原始可行性和对偶可行性等关键概念。最后，我们将通过例题讲解 KKT 条件的具体应用，并简要讨论 KKT 条件的局限性和实际应用中的求解方法。通过本章的学习，读者将能够理解并掌握 KKT 条件，并将其应用于解决实际的不等式约束优化问题。

### 7.1 对偶性质与 KKT 条件的引入

#### 7.1.1 原问题与对偶问题

* **原问题 (Primal problem):**
* minimize f(x)
* subject to: hᵢ(x) = 0, i = 1, ..., m
* gᵢ(x) ≤ 0, i = 1, ..., p
* x ∈ Rⁿ
* **拉格朗日函数 (Lagrangian function):**
  + L(x, λ, μ) = f(x) + Σᵢ λᵢhᵢ(x) + Σⱼ μⱼgⱼ(x)
  + 其中 λᵢ 为等式约束的拉格朗日乘子，μⱼ 为不等式约束的拉格朗日乘子。
* **对偶函数 (Dual function):**
  + d(λ, μ) = infx L(x, λ, μ)
* **对偶问题 (Dual problem):**
* maximize d(λ, μ)
* subject to: μ ≥ 0

#### 7.1.2 弱对偶性与强对偶性

* **弱对偶性 (Weak duality):** 对于任意可行解 x 和对偶可行解 (λ, μ)，都有 d(λ, μ) ≤ f(x)。即对偶问题的最优解是对偶问题的最优解是对原问题最优解的下界。
  + 证明:
  + 令 A(x) = maxλ,μ:μ≥0 L(x, λ, μ)
  + A(x) = maxλ,μ:μ≥0 L(x, λ, μ) ≥ L(x, λ, μ) ≥ minx L(x, λ, μ) = d(λ, μ)
  + A(x) ≥ minx A(x) ≥ maxλ,μ:μ≥0 d(λ, μ) ≥ d(λ, μ)
* **强对偶性 (Strong duality):** 在满足一定条件下 (例如 Slater 条件)，原问题的最优解等于对偶问题的最优解，即 p\* = d\*。
* **Slater 条件:** 存在一个 x 使得所有不等式约束严格成立，即 gᵢ(x) < 0。
* 通过引入对偶问题，我们可以得到关于原问题最优解的一个下界。这促使我们思考：在什么条件下，对偶问题的最优解能够精确地给出原问题的最优解呢？

### 7.2 KKT 条件

#### 7.2.1 KKT 条件的陈述

* 对于一般约束优化问题：
* min f(x)
* subject to hᵢ(x) = 0, i = 1, ...m
* gᵢ(x) ≤ 0, i = 1, ...p
* **定理 7.1 (KKT 条件):** 如果 x\* 是上述约束优化问题的一个局部最优解，且在 x\* 处满足一定的正则性条件（例如线性无关约束规范 LICQ），则存在拉格朗日乘子 λ\* 和 μ\*，使得以下条件成立：
  1. **稳定性条件 (Stationarity):** ∇f(x*) + Σᵢ λᵢ*∇hᵢ(x*) + Σⱼ μⱼ*∇gⱼ(x\*) = 0
  2. **互补松弛条件 (Complementary slackness):** μⱼ*gⱼ(x*) = 0, j = 1, ..., p
  3. **原问题可行性 (Primal feasibility):**
     + hᵢ(x\*) = 0, i = 1, ..., m
     + gⱼ(x\*) ≤ 0, j = 1, ..., p
  4. **对偶可行性 (Dual feasibility):** μⱼ\* ≥ 0, j = 1, ..., p

#### 7.2.2 KKT 条件的解释

* **稳定性条件:** 在最优点处，目标函数的负梯度可以表示为所有起作用约束 (包括等式约束和不等式约束) 的梯度的线性组合，且线性组合系数非负。
* **互补松弛条件:** 对于每个不等式约束，要么拉格朗日乘子 μⱼ\* 为 0 (约束不起作用)，要么约束条件取等号 gⱼ(x\*) = 0 (约束起作用)。
* **原问题可行性:** 解必须满足所有约束条件。
* **对偶可行性:** 不等式约束对应的拉格朗日乘子非负。

#### 7.2.3 KKT 条件的推导 (简要说明)

* KKT 条件可以看作是拉格朗日乘子法在不等式约束情况下的推广。其推导过程较为复杂，主要思想是通过构造一个辅助函数，利用可行方向和泰勒展开等工具，推导出最优解必须满足的条件。

### 7.3 KKT 条件的应用

#### 7.3.1 求解步骤

1. 写出约束优化问题的拉格朗日函数。
2. 列出 KKT 条件 (稳定性条件、互补松弛条件、原始可行性、对偶可行性)。
3. 求解 KKT 条件得到的方程组和不等式组，得到候选的最优解和对应的拉格朗日乘子。
4. 验证正则性条件 (例如 LICQ)，并根据二阶条件或其他方法进一步判断候选解是否为局部最优解。

#### 7.3.2 例题讲解

* **例题 1 (参考 PDF 第 9-10 页):**
* min x²
* subject to 1 ≤ x ≤ 2
  + 解题过程：
    1. 将约束条件改写为标准形式： -x + 1 ≤ 0 和 x - 2 ≤ 0
    2. 构造拉格朗日函数： L(x, λ₁, λ₂) = x² + λ₁(-x + 1) + λ₂(x - 2)
    3. 列出 KKT 条件：
       - 2x - λ₁ + λ₂ = 0
       - λ₁(-x + 1) = 0
       - λ₂(x - 2) = 0
       - -x + 1 ≤ 0
       - x - 2 ≤ 0
       - λ₁ ≥ 0
       - λ₂ ≥ 0
    4. 求解 KKT 条件：
       - 分析 λ₁ 和 λ₂ 的取值情况，得到 x\* = 1, λ₁\* = 2, λ₂\* = 0
    5. 验证解的有效性： x\* = 1 是该问题的全局最优解。
* **例题 2 (参考 PDF 第 11-16 页):**
* min f(x₁, x₂) = x₁² + 2x₂² - 4x₁ - 4x₂
* subject to: g₁(x₁, x₂) = x₁ + x₂ - 3 ≤ 0
* g₂(x₁, x₂) = 5 - x₁ - 2x₂ ≤ 0
  + 解题过程：
    1. 构造拉格朗日函数：L(x₁, x₂, λ₁, λ₂) = x₁² + 2x₂² - 4x₁ - 4x₂ + λ₁(x₁ + x₂ - 3) + λ₂(5 - x₁ - 2x₂)
    2. 列出 KKT 条件：
       - 2x₁ - 4 + λ₁ - λ₂ = 0
       - 4x₂ - 4 + λ₁ - 2λ₂ = 0
       - λ₁g₁(x₁, x₂) = 0
       - λ₂g₂(x₁, x₂) = 0
       - g₁(x₁, x₂) ≤ 0
       - g₂(x₁, x₂) ≤ 0
       - λ₁ ≥ 0
       - λ₂ ≥ 0
    3. 求解 KKT 条件：
       - 分别讨论 λ₁ 和 λ₂ 的四种取值组合 (0, 0), (0, +), (+, 0), (+, +)，并求解相应的方程组。
       - 对于每种组合，检查求得的解是否满足 KKT 条件。
       - 最终得到满足所有 KKT 条件的解：x₁\* = 1, x₂\* = 2, λ₁\* = 8, λ₂\* = 6。
    4. 验证解的有效性： 可以验证该解满足二阶充分条件，因此是局部最优解 (由于目标函数是凸函数，也是全局最优解)。

### 7.4 KKT 条件的局限性与实际应用

#### 7.4.1 KKT 条件的局限性

* KKT 条件是必要条件，而不是充分条件。满足 KKT 条件的点可能是局部最优解、鞍点或全局最优解，需要进一步判断。
* KKT 条件的求解通常需要解一个非线性方程组和不等式组，计算复杂度较高。
* KKT 条件要求目标函数和约束函数可微。
* 对于某些非凸优化问题，即使满足强对偶性，KKT 条件也不一定能找到全局最优解。

#### 7.4.2 实际应用

* KKT 条件是许多优化算法的基础，例如内点法、序列二次规划 (SQP) 等。
* 在实际应用中，通常需要结合数值方法和启发式方法来求解 KKT 条件。
* 可以使用一些现成的优化软件 (例如 MATLAB 的 fmincon 函数) 来求解 KKT 条件。
* 对于大规模问题，可以考虑使用分解协调等方法来降低计算复杂度。
* 一些计算工具和程序可以帮助求解KKT条件，比如PDF文件中提到的利用SymPy库进行符号求解(参考PDF文件第19页) 和利用梯度下降法进行数值求解(参考PDF文件第20页)

### 本章小结

本章详细介绍了 KKT 条件，这是解决不等式约束优化问题的重要工具。我们从对偶性质出发，引出了 KKT 条件的重要性，并详细阐述了 KKT 条件的内容、解释和应用。通过例题讲解，读者可以更好地理解 KKT 条件的具体应用步骤。最后，我们也讨论了 KKT 条件的局限性和实际应用中的一些处理方法。

## 第八章 牛顿法 (Newton's Method)

### 本章导言:

本章将介绍一种经典的求解无约束优化问题的迭代算法——牛顿法。牛顿法利用目标函数的二阶导数信息，通过构造二次近似模型来逼近最优点，具有较快的收敛速度，特别是对于二次函数，牛顿法可以一步到位找到最优解。我们将首先介绍牛顿法的基本思想和迭代公式，然后分析其收敛性质，包括收敛速度和收敛条件。接着，我们将讨论牛顿法的优缺点，并介绍一些改进的牛顿法，例如阻尼牛顿法和 Levenberg-Marquardt 方法。此外, 我们还将比较牛顿法和梯度下降法, 尤其会通过具体的例子以及收敛速度的对比, 来体现牛顿法的优势。最后，我们将通过习题来巩固所学内容。通过本章的学习，读者将能够深入理解牛顿法的原理，掌握其应用方法，并了解其优缺点和改进方向。

### 8.1 牛顿法的基本思想与迭代公式

#### 8.1.1 牛顿法的基本思想

* 牛顿法的核心思想是在当前迭代点 xₖ 处，用一个二次函数 q(x) 来近似目标函数 f(x)，然后求解二次函数的极小点作为下一个迭代点 xₖ₊₁。这个二次函数是通过目标函数 f(x) 在 xₖ 处的二阶泰勒展开得到的。
* 从几何上看，牛顿法是用一个抛物面来逼近目标函数的等值线，并用抛物面的顶点作为下一个迭代点。

#### 8.1.2 牛顿法的迭代公式

* 考虑无约束优化问题：min f(x), x ∈ Rⁿ
* 将 f(x) 在当前迭代点 xₖ 处进行二阶泰勒展开，得到二次近似函数：
  + q(x) = f(xₖ) + ∇f(xₖ)ᵀ(x - xₖ) + (1/2)(x - xₖ)ᵀ∇²f(xₖ)(x - xₖ)
* 求解二次函数的极小点，即令 ∇q(x) = 0，得到：
  + ∇f(xₖ) + ∇²f(xₖ)(x - xₖ) = 0
* 如果 ∇²f(xₖ) 可逆，则可以解得：
  + x = xₖ - [∇²f(xₖ)]⁻¹∇f(xₖ)
* 因此，牛顿法的迭代公式为：
  + xₖ₊₁ = xₖ - [∇²f(xₖ)]⁻¹∇f(xₖ)
* 另一种表示形式：
  + 求解线性方程组： ∇²f(xₖ)dₖ = -∇f(xₖ) 得到搜索方向 dₖ
  + 更新迭代点： xₖ₊₁ = xₖ + dₖ
* 当 n = 1 时，牛顿法的迭代公式简化为：
  + xₖ₊₁ = xₖ - f'(xₖ) / f''(xₖ)
  + 几何解释：在一维情况下，牛顿法是用 f(x) 在 xₖ 处的切线的零点作为下一个迭代点。

### 8.2 牛顿法的收敛性分析

#### 8.2.1 收敛速度

* **定理 8.1:** 设 f ∈ C³，x\* 是一个局部极小点，且 ∇f(x*) = 0，∇²f(x*) 可逆。如果初始点 x₀ 充分靠近 x*，则牛顿法产生的序列 {xₖ} 收敛到 x*，且收敛速度是二阶的，即：
  + ||xₖ₊₁ - x*|| ≤ C||xₖ - x*||², 其中 C 是一个常数。
* **推导过程：** (参考 PDF 第29-31页)
  + 令 F(x) = ∇²f(x), 根据泰勒展开、海森矩阵的性质以及 F(x\*) 可逆性进行推导。
  + 关键步骤：
  + xₖ₊₁ - x\* = xₖ - x\* - [∇²f(xₖ)]⁻¹∇f(xₖ) = [∇²f(xₖ)]⁻¹ [∇²f(xₖ)(xₖ - x\*) - ∇f(xₖ)]
  + ∇f(x*) = ∇f(xₖ) + ∇²f(xₖ)(x* - xₖ) + O(||xₖ - x\*||²) = 0
  + 利用 ∇f(x*) = 0, 进行符号变换，得到：∇²f(xₖ)(xₖ - x*) - ∇f(xₖ) = O(||xₖ - x\*||²)
  + 最终得到 ||xₖ₊₁ - x*|| ≤ C₁||xₖ - x*||²

#### 8.2.2 收敛条件

* 初始点 x₀ 必须充分靠近局部极小点 x\*。
* 目标函数 f(x) 必须三阶连续可微。
* 在迭代过程中，海森矩阵 ∇²f(xₖ) 必须可逆。

### 8.3 牛顿法的优缺点

#### 8.3.1 优点

* **收敛速度快:** 在满足条件下，牛顿法具有二阶收敛速度，比梯度下降法 (一阶收敛) 快得多。尤其对于二次函数，牛顿法可以一步收敛到最优解。(参考 PDF 第25页图)
* **仿射不变性 (Affine invariant):** 牛顿法的收敛性不依赖于坐标系的选择。

#### 8.3.2 缺点

* **局部收敛性:** 牛顿法只有当初始点充分靠近最优解时才能保证收敛。如果初始点远离最优解，牛顿法可能不收敛，甚至可能收敛到鞍点或极大点。(参考 PDF 第26页图)
* **计算量大:** 每次迭代都需要计算海森矩阵 ∇²f(xₖ) 及其逆矩阵，对于大规模问题，计算量很大。
  + 计算海森矩阵的时间复杂度为 O(n²)，求逆的复杂度为 O(n³)。
* **海森矩阵奇异性问题:** 如果海森矩阵 ∇²f(xₖ) 不可逆 (奇异)，则牛顿法无法进行。

### 8.4 牛顿法的改进

#### 8.4.1 阻尼牛顿法 (Damped Newton's Method)

* 为了改善牛顿法的全局收敛性，可以在迭代公式中引入一个步长因子 αₖ，即：
  + xₖ₊₁ = xₖ - αₖ[∇²f(xₖ)]⁻¹∇f(xₖ)
* 步长因子 αₖ 可以通过线性搜索 (line search) 来确定，例如：
  + αₖ = argminα≥0 f(xₖ - α[∇²f(xₖ)]⁻¹∇f(xₖ))
* 阻尼牛顿法可以保证每次迭代都使目标函数值下降，从而提高算法的全局收敛性。

#### 8.4.2 Levenberg-Marquardt 方法

* 针对海森矩阵奇异或不正定的情况，Levenberg-Marquardt 方法对牛顿法的迭代公式进行了修改：
  + xₖ₊₁ = xₖ - (∇²f(xₖ) + μₖI)⁻¹∇f(xₖ)
  + 其中 μₖ 是一个正数，I 是单位矩阵。
* 当 μₖ = 0 时，Levenberg-Marquardt 方法退化为牛顿法。
* 当 μₖ → +∞ 时，Levenberg-Marquardt 方法趋近于梯度下降法。
* 通过调整 μₖ 的大小，Levenberg-Marquardt 方法可以在牛顿法和梯度下降法之间进行切换，从而兼顾二者的优点。
* μₖ 的选择策略：
  + 选择 μₖ > 0 使得 ∇²f(xₖ) + μₖI 正定。
  + 如果目标函数值下降，则减小 μₖ。
  + 如果目标函数值上升，则增大 μₖ。

### 8.5 牛顿法与梯度下降法的比较

|  |  |  |
| --- | --- | --- |
| 特性 | 牛顿法 | 梯度下降法 |
| 收敛速度 | 二阶收敛 | 一阶收敛 |
| 计算量 | 较大 (计算海森矩阵及其逆矩阵) | 较小 (仅计算梯度) |
| 收敛性 | 局部收敛 | 全局收敛 (步长选择合适的情况下) |
| 海森矩阵要求 | 可逆 | 无 |
| 步长 | 通常为 1 (经典牛顿法) 或通过线搜索确定 | 需要选择合适的步长 (例如，通过线搜索) |
| 适用范围 | 初始点靠近最优解，且海森矩阵可逆的情况 | 更广泛，可用于海森矩阵不可逆或难以计算的情况 |

### 8.6 应用举例 (结合前面章节)

* 利用牛顿法求解方程组：
* f(x) = 0, 其中 f: Rⁿ → Rⁿ
* 迭代公式： xₖ₊₁ = xₖ - J(xₖ)⁻¹f(xₖ), 其中 J(xₖ) 为雅可比矩阵。
* 求解逻辑回归 (Logistic Regression) 的参数：
* 目标函数为对数似然函数的负数。
* 利用牛顿法迭代求解参数。

### 8.7 关于牛顿法中步长的一种理解方式

* 将 f(xₖ₊₁) 在 xₖ 附近进行泰勒展开： f(xₖ₊₁) ≈ f(xₖ) + f'(xₖ)(xₖ₊₁ - xₖ) + (1/2)f''(xₖ)(xₖ₊₁ - xₖ)²
* 为了简化计算，将最后一项的 f''(xₖ) 替换为 1/t，得到：f(xₖ₊₁) ≈ f(xₖ) + f'(xₖ)(xₖ₊₁ - xₖ) + (1/(2t))(xₖ₊₁ - xₖ)²
* 此时，可以推导出 xₖ₊₁ = xₖ - tf'(xₖ)

### 习题

1. 证明对于二次函数，牛顿法可以一步到位找到最优解。
2. 使用牛顿法求解函数 f(x) = x³ - 2x + 2 的极小点，初始点分别为 x₀ = 0 和 x₀ = 1.5，并观察迭代过程。(可参考教材第26页图)
3. 使用牛顿法求解方程组：
4. x₁² + x₂² - 1 = 0
5. x₁ - x₂ = 0

初始点为 x₀ = (1, 0)ᵀ。

1. 考虑逻辑回归模型，目标函数为对数似然函数的负数，推导参数的牛顿法迭代公式。
2. (参考教材第22页) 考虑函数 f(x₁, x₂) = x₁⁴ + x₂², 从初始点 x₀ = (1, 1)ᵀ 开始，分别使用梯度下降法和牛顿法进行迭代，比较二者的收敛速度。
3. (参考教材第36页) 考虑函数 y = x4/3，如何修改牛顿法使其能够有效求解该函数的极小点？

### 本章小结

本章详细介绍了牛顿法的原理、迭代公式、收敛性、优缺点以及改进方法。牛顿法是一种高效的求解无约束优化问题的迭代算法，特别适用于目标函数具有良好二次性质的情况。然而，牛顿法也存在局部收敛性和计算量大的问题，需要根据具体情况选择合适的算法或改进策略。

## 第九章 次梯度 (Sub-gradient)

### 本章导言:

在前面的章节中，我们学习的优化算法，如梯度下降法和牛顿法，都依赖于目标函数的可微性。然而，在实际问题中，我们经常会遇到不可微的函数，例如带有绝对值项的函数 (如 ℓ₁ 范数)。为了处理这类不可微的凸优化问题，本章将引入次梯度的概念，它是梯度概念的推广。我们将首先定义次梯度和次微分，然后探讨次梯度的性质和计算规则。接着，我们将介绍基于次梯度的优化方法——次梯度方法，并分析其收敛性。最后，我们将通过一些例子，例如 Lasso 问题，来展示次梯度方法在求解实际问题中的应用。通过本章的学习，读者将能够理解次梯度的概念，掌握次梯度的计算方法，并能够运用次梯度方法解决实际的不可微凸优化问题。

### 9.1 次梯度与次微分

#### 9.1.1 次梯度的定义

* **回顾梯度的性质:** 对于可微的凸函数 f，其在任意一点 x 处的梯度 ∇f(x) 满足以下不等式：
  + f(y) ≥ f(x) + ∇f(x)ᵀ(y - x), ∀y
  + 几何意义：函数 f(x) 的图像始终位于其在点 x 处切线的上方，即线性近似总是低估了函数值。
* **次梯度 (Sub-gradient):** 对于凸函数 f (不一定可微)，其在点 x 处的次梯度是一个向量 g，满足以下不等式：
  + f(y) ≥ f(x) + gᵀ(y - x), ∀y
  + 几何意义：函数 f(x) 的图像始终位于过点 (x, f(x)) 且斜率为 g 的超平面的上方。
* **注:** 即使对于非凸函数，也可以定义次梯度，但次梯度不一定存在。

#### 9.1.2 次微分 (Sub-differential)

* **次微分的定义:** 函数 f 在点 x 处的所有次梯度的集合称为 f 在 x 处的次微分，记作 ∂f(x)。
* **性质:**
  + ∂f(x) 是一个闭凸集 (即使 f 是非凸函数)。
  + 对于凸函数，∂f(x) 非空。
  + 如果 f 在 x 处可微，则 ∂f(x) = {∇f(x)}，即次微分只包含梯度。
  + 如果 ∂f(x) = {g}，即次微分只包含一个元素，则 f 在 x 处可微，且 ∇f(x) = g。

### 9.2 次梯度的例子

#### 9.2.1 一维绝对值函数

* f(x) = |x|
* 当 x > 0 时，∂f(x) = {1}
* 当 x < 0 时，∂f(x) = {-1}
* 当 x = 0 时，∂f(x) = [-1, 1]

#### 9.2.2 ℓ₂ 范数

* f(x) = ||x||₂, x ∈ Rⁿ
* 当 x ≠ 0 时，∂f(x) = {x / ||x||₂}
* 当 x = 0 时，∂f(x) = {z : ||z||₂ ≤ 1}
* **说明**: 当 x=0 时，函数 f(x) = ||x||₂ 不可微。此时，其在原点处的次微分是单位球内的所有向量。

#### 9.2.3 ℓ₁ 范数

* f(x) = ||x||₁, x ∈ Rⁿ
* ∂f(x) = {g : gᵢ ∈ sign(xᵢ) if xᵢ ≠ 0, gᵢ ∈ [-1, 1] if xᵢ = 0}

#### 9.2.4 两个可微凸函数的最大值

* f(x) = max{f₁(x), f₂(x)}，其中 f₁(x) 和 f₂(x) 是可微的凸函数。
* 当 f₁(x) > f₂(x) 时，∂f(x) = {∇f₁(x)}
* 当 f₁(x) < f₂(x) 时，∂f(x) = {∇f₂(x)}
* 当 f₁(x) = f₂(x) 时，∂f(x) = conv{∇f₁(x), ∇f₂(x)}，即 ∇f₁(x) 和 ∇f₂(x) 的凸包。

### 9.3 次梯度运算法则

#### 9.3.1 数乘

* ∂(af)(x) = a∂f(x)，其中 a > 0

#### 9.3.2 加法

* ∂(f₁ + f₂)(x) = ∂f₁(x) + ∂f₂(x)

#### 9.3.3 仿射变换

* 如果 g(x) = f(Ax + b)，则 ∂g(x) = Aᵀ∂f(Ax + b)

#### 9.3.4 有限逐点最大值

* 如果 f(x) = maxi=1,...,m fᵢ(x)，则 ∂f(x) = conv(∪i:fᵢ(x)=f(x) ∂fᵢ(x))，即在 x 处取到最大值的所有 fᵢ(x) 的次微分的并集的凸包。

### 9.4 次梯度方法

#### 9.4.1 次梯度方法的迭代公式

* 考虑无约束凸优化问题：min f(x)
* 次梯度方法的迭代公式为：
  + x⁽ᵏ⁺¹⁾ = x⁽ᵏ⁾ - tₖg⁽ᵏ⁾
  + 其中 x⁽ᵏ⁾ 是第 k 次迭代的解，tₖ 是步长，g⁽ᵏ⁾ 是 f(x) 在 x⁽ᵏ⁾ 处的任意一个次梯度，即 g⁽ᵏ⁾ ∈ ∂f(x⁽ᵏ⁾)。
* **注意:** 次梯度方法与梯度下降法的区别在于，次梯度方法使用的是次梯度，而不是梯度。

#### 9.4.2 步长选择

* **固定步长:** tₖ = t
* **逐渐减小步长:** tₖ = α / √k，其中 α 是一个常数。
* **其他更复杂的步长选择策略。**

#### 9.4.3 收敛性分析

* **定理 9.1:** 如果 f(x) 是凸函数且满足 Lipschitz 连续条件 (即存在常数 L > 0，使得 ||g|| ≤ L, ∀g ∈ ∂f(x))，并且步长选择满足一定条件 (例如 tₖ = α / √k)，则次梯度方法收敛，即：
  + limk→∞ f(x⁽ᵏ⁾) = f*，其中 f* 是 f(x) 的最优值。
* **收敛速度:** 次梯度方法的收敛速度通常为 O(1/√k)，比梯度下降法的收敛速度慢。

### 9.5 次梯度方法的应用：Lasso 问题

#### 9.5.1 Lasso 问题的定义

* Lasso 问题可以表示为以下优化问题：
  + minβ (1/2)||y - Xβ||₂² + λ||β||₁
  + 其中 y ∈ Rⁿ 是观测向量，X ∈ Rⁿˣᵖ 是设计矩阵，β ∈ Rᵖ 是待估计的参数向量，λ 是正则化参数。

#### 9.5.2 Lasso 问题的最优性条件

* 利用次梯度 optimality condition，Lasso 问题的最优解 β\* 满足以下条件：
  + 0 ∈ -Xᵀ(y - Xβ*) + λ∂||β*||₁
  + 即 Xᵀ(y - Xβ*) = λv，其中 v ∈ ∂||β*||₁
* 根据 ℓ₁ 范数的次微分性质，可以得到：
  + Xᵢᵀ(y - Xβ*) = λ⋅sign(βᵢ*) if βᵢ\* ≠ 0
  + |Xᵢᵀ(y - Xβ*)| ≤ λ if βᵢ* = 0

#### 9.5.3 软阈值算子 (Soft-thresholding operator)

* 当 X = I 时，Lasso 问题可以得到显式解：
  + β\* = Sλ(y)
  + 其中 Sλ(y) 是软阈值算子，其定义为：
    - [Sλ(y)]ᵢ = (yᵢ - λ)₊ if yᵢ > λ
    - [Sλ(y)]ᵢ = 0 if -λ ≤ yᵢ ≤ λ
    - [Sλ(y)]ᵢ = (yᵢ + λ)₋ if yᵢ < -λ
  + (y)₊ 表示取 y 的正部，(y)₋ 表示取 y 的负部。

### 9.6 次梯度方法的优缺点

#### 9.6.1 优点

* 可以处理不可微的凸优化问题。
* 算法简单，易于实现。
* 对目标函数的性质要求较低，只需要凸性和 Lipschitz 连续性。

#### 9.6.2 缺点

* 收敛速度较慢，通常为 O(1/√k)。
* 步长选择对算法性能影响较大。
* 次梯度方法不一定是下降方法，目标函数值在迭代过程中可能会出现波动。

### 习题

1. 计算函数 f(x) = max{x², 2x + 3} 的次微分。
2. 证明次微分的加法法则。
3. 使用次梯度方法求解以下优化问题：
   * min |x - 1| + |x - 2|
4. 考虑 Lasso 问题，证明软阈值算子 Sλ(y) 满足 Lasso 问题的最优性条件。
5. (参考教材第33页) 编写程序实现软阈值算子，并求解简化的 Lasso 问题。

### 本章小结

本章介绍了次梯度的概念、性质和计算规则，以及基于次梯度的优化方法——次梯度方法。次梯度方法可以用于求解不可微的凸优化问题，具有广泛的应用。我们通过 Lasso 问题展示了次梯度方法在实际问题中的应用。

## 第十章 迭代法与临近点算法 (Iterative Methods and Proximal Algorithms)

### 本章导言:

本章将介绍求解线性方程组的迭代方法和一类重要的优化算法——临近点算法 (Proximal Algorithm)。我们将首先回顾求解线性方程组 Ax = b 的经典迭代方法，例如 Jacobi 方法和 Gauss-Seidel 方法，并讨论它们的收敛性。随后，我们将引入临近点算法的概念，并详细讲解其与梯度下降法的联系和区别。临近点算法通过求解一个更简单的子问题来逼近原问题的解，特别适合于处理目标函数中包含不可微项的情况，例如包含 ℓ₁ 范数的优化问题。我们将通过推导软阈值算子来具体说明临近点算法的应用，并介绍其在求解 Lasso 问题中的应用。最后, 我们将介绍临近点梯度法 (Proximal Gradient Method) 及其快速版本 (FISTA)，并简要讨论其收敛性。通过本章的学习，读者将掌握求解线性方程组的迭代方法和临近点算法的基本原理，并能够运用这些方法解决实际的优化问题。

### 10.1 求解线性方程组的迭代方法

#### 10.1.1 Jacobi 迭代法

* 考虑线性方程组 Ax = b，其中 A ∈ Rⁿˣⁿ 是非奇异矩阵，b ∈ Rⁿ。
* 将 A 分解为 A = D - L - U，其中 D 是 A 的对角部分，-L 是 A 的严格下三角部分，-U 是 A 的严格上三角部分。
* Jacobi 迭代法的迭代公式为：
  + x⁽ᵏ⁺¹⁾ = D⁻¹(L + U)x⁽ᵏ⁾ + D⁻¹b
  + 分量形式：
    - xᵢ⁽ᵏ⁺¹⁾ = (1/aᵢᵢ)(bᵢ - Σj≠i aᵢⱼxⱼ⁽ᵏ⁾), i = 1, 2, ..., n
* **基本思想:** 每次迭代，用其他分量的当前值来更新一个分量。
* **例子:** (参考 PDF 第 23 页)
* 5x₁ - x₂ + 2x₃ = 12
* 3x₁ + 8x₂ - 2x₃ = -25
* x₁ + x₂ + 4x₃ = 6

可以改写为：

x₁ = (12 + x₂ - 2x₃) / 5

x₂ = (-25 - 3x₁ + 2x₃) / 8

x₃ = (6 - x₁ - x₂) / 4

然后进行迭代求解。

* **局限性:** Jacobi 迭代法不总是收敛的。例如, 对于以下方程组 (参考 PDF 第 25 页):
* x₁ + 7x₂ = 0
* 3x₁ + x₂ = 1

使用 Jacobi 迭代法会发散。

#### 10.1.2 Gauss-Seidel 迭代法

* Gauss-Seidel 迭代法的迭代公式为：
  + x⁽ᵏ⁺¹⁾ = (D - L)⁻¹Ux⁽ᵏ⁾ + (D - L)⁻¹b
  + 分量形式：
    - xᵢ⁽ᵏ⁺¹⁾ = (1/aᵢᵢ)(bᵢ - Σj<i aᵢⱼxⱼ⁽ᵏ⁺¹⁾ - Σj>i aᵢⱼxⱼ⁽ᵏ⁾), i = 1, 2, ..., n
* **基本思想:** 每次迭代，用已经更新过的分量来更新其他分量。
* **例子:** (参考 PDF 第 24 页) 使用与 Jacobi 迭代法相同的例子，Gauss-Seidel 迭代公式为：
* x₁⁽ᵏ⁺¹⁾ = (12 + x₂⁽ᵏ⁾ - 2x₃⁽ᵏ⁾) / 5
* x₂⁽ᵏ⁺¹⁾ = (-25 - 3x₁⁽ᵏ⁺¹⁾ + 2x₃⁽ᵏ⁾) / 8
* x₃⁽ᵏ⁺¹⁾ = (6 - x₁⁽ᵏ⁺¹⁾ - x₂⁽ᵏ⁺¹⁾) / 4
* **与 Jacobi 迭代法的比较:** Gauss-Seidel 迭代法通常比 Jacobi 迭代法收敛更快，但仍然不能保证收敛。

#### 10.1.3 收敛性分析

* **定理 10.1:** 如果 A 是严格对角占优矩阵，则 Jacobi 迭代法和 Gauss-Seidel 迭代法都收敛。
* **严格对角占优矩阵:** 对于矩阵 A，如果满足 |aᵢᵢ| > Σj≠i |aᵢⱼ|, ∀i，则称 A 是严格对角占优矩阵。

### 10.2 临近点算法 (Proximal Algorithm)

#### 10.2.1 临近点算子 (Proximal Operator)

* 对于凸函数 f: Rⁿ → R，其临近点算子定义为：
  + proxλf(v) = argminx (f(x) + (1/(2λ))||x - v||₂²)，其中 λ > 0 是一个参数。
* **直观理解:** proxλf(v) 是在函数 f(x) 的值和与点 v 的距离之间进行权衡后得到的点。参数 λ 控制权衡的力度。
* **几何解释:** (参考 PDF 第 26 页) proxλf(v) 可以看作是在 v 附近寻找一个使 f(x) 尽可能小的点，同时又不会离 v 太远。

#### 10.2.2 临近点算法的迭代公式

* 考虑无约束优化问题：min f(x) + g(x)，其中 f(x) 是光滑凸函数，g(x) 是凸函数 (不一定光滑)。
* 临近点算法的迭代公式为：
  + x⁽ᵏ⁺¹⁾ = proxλg(x⁽ᵏ⁾ - λ∇f(x⁽ᵏ⁾))，其中 λ > 0 是步长。
* **基本思想:** 每次迭代，先沿着 f(x) 的负梯度方向走一步，然后用 g(x) 的临近点算子将迭代点拉回到一个使 g(x) 较小的区域。

#### 10.2.3 临近点算子的性质

* **投影算子:** 当 g(x) 是一个闭凸集 C 的示性函数 (indicator function) 时，即：
  + g(x) = IC(x) = { 0, if x ∈ C; +∞, if x ∉ C }
  + 临近点算子 proxλg(v) 退化为投影算子，即 proxλg(v) = PC(v)，其中 PC(v) 表示 v 在集合 C 上的投影 (参考 PDF 第 27 页)。
* **更一般的形式:** (参考 PDF 第 28 页)
  + min f(x) + λg(x)
  + 其中 f(x) 光滑，而 g(x) 不一定光滑，但其临近点算子 proxλg(x) 容易计算。
  + 迭代公式：x⁽ᵏ⁺¹⁾ = proxλₖg(x⁽ᵏ⁾ - λₖ∇f(x⁽ᵏ⁾))，其中 λₖ > 0 是步长。

#### 10.2.4 临近点算子的计算

* 对于一些特殊的函数 g(x)，其临近点算子 proxλg(x) 可以显式地计算出来。
* **例 1:** g(x) = |x|, x ∈ R \* proxλg(b) = { b - λ, if b ≥ λ; 0, if |b| ≤ λ; b + λ, if b ≤ -λ } (参考 PDF 第 29 页) \* 证明过程：
  + 当 b ≥ 0 时，目标函数为 (1/(2λ))(x - b)² + x，其极小点为 x̃ = b - λ。
  + 如果 x̃ ≥ 0，即 b ≥ λ，则 proxλg(b) = b - λ。
  + 如果 x̃ < 0，即 b < λ，则 proxλg(b) = 0。
  + 当 b < 0 时，同理可得 proxλg(b) = b + λ (如果 b ≤ -λ) 或 0 (如果 |b| < λ)。 \* 可以合并写成：
  + proxλg(b) = { sgn(b)(|b| - λ), if |b| > λ; 0, otherwise }
* **例 2:** g(x) = ||x||₁, x ∈ Rⁿ \* proxλg(b) = Sλ(b)，其中 Sλ(b) 是软阈值算子 (soft-thresholding operator)，其定义为：
  + [Sλ(b)]ᵢ = { bᵢ - λ, if bᵢ > λ; 0, if |bᵢ| ≤ λ; bᵢ + λ, if bᵢ < -λ } (参考 PDF 第 30 页)
  + 即逐分量应用一维情况下的软阈值算子。

### 10.3 临近点梯度法 (Proximal Gradient Method)

#### 10.3.1 临近点梯度法的迭代公式

* 考虑优化问题：min F(x) = f(x) + g(x)，其中 f(x) 是光滑凸函数，g(x) 是凸函数 (不一定光滑)。
* 临近点梯度法的迭代公式为：
* x⁽ᵏ⁺¹⁾ = proxtₖg(x⁽ᵏ⁾ - tₖ∇f(x⁽ᵏ⁾))，其中 tₖ > 0 是步长。
* **与梯度下降法的关系:** 当 g(x) = 0 时，临近点梯度法退化为梯度下降法。
* **与迭代软阈值算法 (ISTA) 的关系:** 当 g(x) = λ||x||₁ 时，临近点梯度法退化为 ISTA 算法。

#### 10.3.2 快速临近点梯度法 (FISTA - Fast Iterative Shrinkage-Thresholding Algorithm)

* FISTA 是对临近点梯度法的一种加速方法，其迭代公式为：
  + x⁽ᵏ⁺¹⁾ = proxtₖg(y⁽ᵏ⁾ - tₖ∇f(y⁽ᵏ⁾))
  + y⁽ᵏ⁺¹⁾ = x⁽ᵏ⁺¹⁾ + ((k - 1)/(k + 2))(x⁽ᵏ⁺¹⁾ - x⁽ᵏ⁾)
  + 其中 y⁽ᵏ⁾ 是一个辅助变量。
* **与临近点梯度法的比较:** FISTA 通过引入一个额外的动量项 (momentum term) 来加速收敛。

#### 10.3.3 收敛性分析

* **定理 10.2:** 如果 f(x) 是光滑凸函数，∇f(x) 是 Lipschitz 连续的，g(x) 是凸函数，则临近点梯度法和 FISTA 都收敛。
* **收敛速度:** 临近点梯度法的收敛速度为 O(1/k)，FISTA 的收敛速度为 O(1/k²)。

### 10.4 应用：Lasso 问题

* Lasso 问题的目标函数可以写成 f(x) + g(x) 的形式，其中：
  + f(x) = (1/2)||Ax - b||₂² (光滑凸函数)
  + g(x) = λ||x||₁ (凸函数，不可微)
* ∇f(x) = Aᵀ(Ax - b)
* proxλg(x) = Sλ(x) (软阈值算子)
* 因此，Lasso 问题可以用临近点梯度法 (ISTA) 或 FISTA 求解 (参考 PDF 第 31 页)。

### 习题

1. 证明 Gauss-Seidel 迭代法的收敛性定理 (当 A 是严格对角占优矩阵时)。
2. 推导 ℓ₀ 范数的临近点算子。
3. 使用 FISTA 算法求解 Lasso 问题，并与 ISTA 算法的收敛速度进行比较。
4. (参考教材第 2 页) 实现求解 Lasso 问题的程序，并输出类似第 3 页图的结果。
5. (参考教材第 2 页) 使用不同的初始值，测试牛顿法求解教材第 26 页给出的例子，并观察迭代结果。

### 本章小结

本章介绍了求解线性方程组的迭代方法和临近点算法。迭代方法为求解大型线性方程组提供了有效的途径。临近点算法是一类重要的优化算法，特别适合于处理目标函数中包含不可微项的情况。我们通过 Lasso 问题展示了临近点算法在实际问题中的应用，并介绍了 FISTA 算法来加速收敛。

## 第十一章 练习题

### 本章导言:

本章包含六道练习题，涵盖了前面章节的主要内容，包括：

* 向量空间的性质
* 线性方程组的求解
* 二次型的应用
* 概率图模型 (PGM)
* 最优性条件
* 拉格朗日乘子法与 KKT 条件

每道题都提供了详细的解答，希望读者能够通过这些练习题巩固所学知识，并提高解决实际问题的能力。

### 练习题 1：向量空间与线性方程组

设 f: Rⁿ → R 为连续可微函数，写出相应的梯度和 Hessian 矩阵，并写出 Taylor 公式 (写至 2 阶)。

#### 解答:

* **梯度 (Gradient):** ∇f(x) 是一个 n 维向量，其第 i 个分量为 ∂f(x)/∂xᵢ。

∇f(x) = [ ∂f(x)/∂x₁ ]

[ ∂f(x)/∂x₂ ]

[ ... ]

[ ∂f(x)/∂xₙ ]

* **海森矩阵 (Hessian Matrix):** ∇²f(x) 是一个 n × n 的矩阵，其第 i 行第 j 列的元素为 ∂²f(x)/∂xᵢ∂xⱼ。

∇²f(x) = [ ∂²f(x)/∂x₁² ∂²f(x)/∂x₁∂x₂ ... ∂²f(x)/∂x₁∂xₙ ]

[ ∂²f(x)/∂x₂∂x₁ ∂²f(x)/∂x₂² ... ∂²f(x)/∂x₂∂xₙ ]

[ ... ... ... ... ]

[ ∂²f(x)/∂xₙ∂x₁ ∂²f(x)/∂xₙ∂x₂ ... ∂²f(x)/∂xₙ² ]

* **二阶泰勒公式 (Taylor Expansion to the second order):** f(x) ≈ f(x₀) + ∇f(x₀)ᵀ(x - x₀) + (1/2)(x - x₀)ᵀ∇²f(x₀)(x - x₀)

### 练习题 2：线性方程组求解

设 A 为 m × n 矩阵：

(1) 假设 m > n，且 AᵀA 正定，给出 minx ||Ax - b||₂² 的公式。

(2) 假设 m < n，且 AAᵀ 可逆，给出以下问题的解的公式：

min. ||x||₂² s.t. Ax = b

#### 解答:

(1) 当 m > n 且 AᵀA 正定 (此时 A 列满秩) 时，问题 minx ||Ax - b||₂² 是一个最小二乘问题，其解可以通过求解正规方程得到。

\* 目标函数可以写成： ||Ax - b||₂² = (Ax - b)ᵀ(Ax - b) = xᵀAᵀAx - 2bᵀAx + bᵀb

\* 对 x 求导并令其等于 0，得到正规方程： AᵀAx = Aᵀb

\* 由于 AᵀA 正定，因此 AᵀA 可逆，解得： x = (AᵀA)⁻¹Aᵀb

(2) 当 m < n 且 AAᵀ 可逆 (此时 A 行满秩) 时，问题是一个带有等式约束的优化问题。

\* 构造拉格朗日函数： L(x, λ) = ||x||₂² + λᵀ(Ax - b) = xᵀx + λᵀ(Ax - b)

\* 对 x 求导并令其等于 0，得到： 2x + Aᵀλ = 0，即 x = -(1/2)Aᵀλ

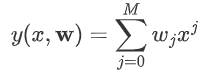
\* 将 x = -(1/2)Aᵀλ 代入约束条件 Ax = b，得到：-(1/2)AAᵀλ = b

\* 由于 AAᵀ 可逆，解得：λ = -2(AAᵀ)⁻¹b

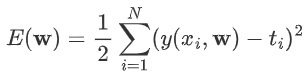
\* 将 λ 代回 x 的表达式，得到：x = Aᵀ(AAᵀ)⁻¹b

### 练习题 3：二次型与曲线拟合

对曲线拟合问题，设有 N 个观测点 x₁, ..., xₙ，其观测值为 t₁, ..., tₙ。现用 M 次多项式拟合，记多项式的形式为：



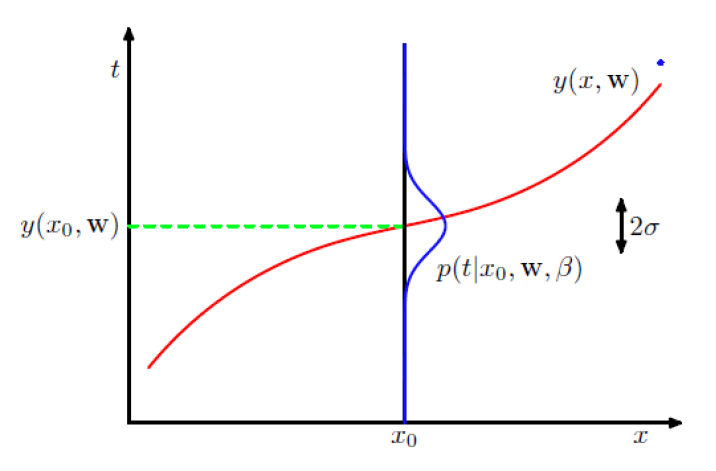
定义



则多项式的系数为



现请用 MLE 方法推导出以上情形。附：用 Bayes 观点看待曲线拟合的示意图。

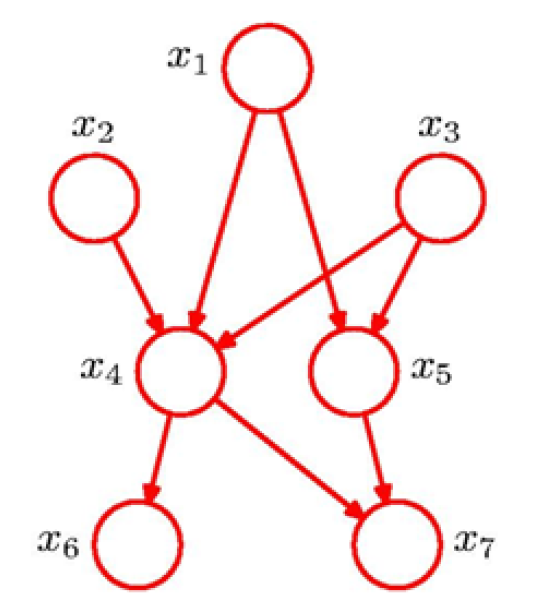


#### 解答:

* **最大似然估计 (MLE):** 假设观测值 tᵢ 是由真实函数值 y(xᵢ, w) 加上一个服从高斯分布的噪声 εᵢ 得到的，即：
  + tᵢ = y(xᵢ, w) + εᵢ
  + 其中 εᵢ ~ N(0, σ²)
* 因此，给定 xᵢ 和 w，tᵢ 服从均值为 y(xᵢ, w)，方差为 σ² 的高斯分布：
  + p(tᵢ|xᵢ, w, σ²) = (1/(√(2π)σ))exp(-(tᵢ - y(xᵢ, w))² / (2σ²))
* 假设观测值 t₁, ..., tₙ 相互独立，则似然函数为：
  + p(t₁, ..., tₙ|x₁, ..., xₙ, w, σ²) = Πᵢ p(tᵢ|xᵢ, w, σ²)
* 取对数似然函数：
  + ln p(t₁, ..., tₙ|x₁, ..., xₙ, w, σ²) = Σᵢ ln p(tᵢ|xᵢ, w, σ²) = - (N/2)ln(2π) - Nln(σ) - (1/(2σ²))Σᵢ (tᵢ - y(xᵢ, w))²
* 最大化对数似然函数等价于最小化 (1/(2σ²))Σᵢ (tᵢ - y(xᵢ, w))²，由于 σ² 是常数，因此等价于最小化：
  + E(w) = (1/2)Σᵢ (y(xᵢ, w) - tᵢ)²
* 因此，通过 MLE 方法，我们得到了与最小化 E(w) 相同的结果。

### 练习题 4：概率图模型

对如下的 PGM 图：



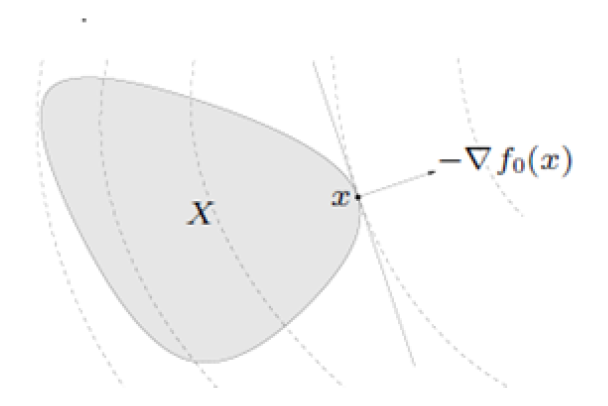
写出 p(x₁, ..., x₇) 的公式。

#### 解答:

根据图中的条件独立性假设，可以将联合概率分解为： p(x₁, ..., x₇) = p(x₁)p(x₂|x₁)p(x₃|x₁)p(x₄|x₂, x₆)p(x₅|x₂, x₃)p(x₆|x₄, x₅)p(x₇|x₃, x₅, x₆)

### 练习题 5：最优性条件

试根据下图，写出 optimality condition (即 x 使 f(x) 达到极值，x 应满足的条件)。



#### 解答:

根据图示，x 位于可行域 X 的边界上，且 -∇f(x) 指向可行域外部。在最优点 x 处，目标函数 f(x) 的负梯度 -∇f(x) 必须与指向可行域内部的法向量方向一致 (或负梯度与该点处的可行方向均呈钝角)，或者 -∇f(x) 为零向量。

因此，最优性条件可以描述为：

* 对于任意可行方向 d (即从 x 出发指向可行域 X 内部的方向)，都有 ∇f(x)ᵀd ≥ 0，或者 ∇f(x) = 0。

### 练习题 6：拉格朗日乘子法与 KKT 条件

1. 写出 f: Rⁿ → R 为凸函数的定义。
2. 写出凸优化问题的标准形式。
3. 对以下问题：
4. min f(x) = x₁² + x₂ + 4
5. s.t. -x₁² - (x₂ + 4)² + 16 ≥ 0
6. x₁ - x₂ - 6 ≥ 0

写出其相应的 Lagrange 函数，对偶函数和 KKT 条件，并求出相应的解。

#### 解答:

(1) 凸函数的定义：对于函数 f: Rⁿ → R，如果对于任意 x, y ∈ Rⁿ 和任意 α ∈ [0, 1]，都有： f(αx + (1 - α)y) ≤ αf(x) + (1 - α)f(y) 则称 f(x) 为凸函数。

(2) 凸优化问题的标准形式： min f(x) s.t. gᵢ(x) ≤ 0, i = 1, ..., m hᵢ(x) = 0, i = 1, ..., p 其中 f(x) 是凸函数，gᵢ(x) 是凸函数，hᵢ(x) 是仿射函数。

(3) 针对给定的优化问题：

* **Lagrange 函数:** L(x, λ) = x₁² + x₂ + 4 + λ₁(-x₁² - (x₂ + 4)² + 16) + λ₂(x₁ - x₂ - 6)
* **对偶函数:** g(λ) = infx L(x, λ)
* **KKT 条件:**
  1. ∇xL(x*, λ*) = 0:
     + 2x₁\* - 2λ₁*x₁* + λ₂\* = 0
     + 1 - 2λ₁*(x₂* + 4) - λ₂\* = 0
  2. λ₁\* ≥ 0, λ₂\* ≥ 0 (对偶可行性)
  3. -x₁*² - (x₂* + 4)² + 16 ≥ 0, x₁\* - x₂\* - 6 ≥ 0 (原始可行性)
  4. λ₁*(-x₁*² - (x₂\* + 4)² + 16) = 0, λ₂*(x₁* - x₂\* - 6) = 0 (互补松弛性)
* **求解:**
  1. 分析 λ₁ 和 λ₂ 的取值情况：
     + **情况 1:** λ₁\* = 0, λ₂\* = 0: 此时 x₁\* = 0, 但无法满足第一个等式。
     + **情况 2:** λ₁\* = 0, λ₂\* > 0: 此时 x₁\* - x₂\* - 6 = 0, x₁\* = 0, λ₂\* = -1, 不满足 λ₂\* > 0。
     + **情况 3:** λ₁\* > 0, λ₂\* = 0: 此时 -x₁*² - (x₂* + 4)² + 16 = 0, 且 x₁\* = 0, x₂\* = -3, λ₁\* = 1/2。验证得到: -x₁*² - (x₂* + 4)² + 16 = -1 + 16 > 0, x₁\* - x₂\* - 6 = -3 < 0， 不满足原始可行性。
     + **情况 4:** λ₁\* > 0, λ₂\* > 0: 此时 -x₁*² - (x₂* + 4)² + 16 = 0 且 x₁\* - x₂\* - 6 = 0。联立方程求解得到 x₁\* = (1 + √33)/2, x₂\* = (-11 + √33)/2, λ₁\* = (1 + √33)/(1 + √33 + 8), λ₂\* = 16 + 2√33 - (1 + √33)(1 + √33)/(1 + √33 + 8)。
  2. 最终解需要验证情况 4 中 λ₁\* > 0, λ₂\* > 0 是否成立。经过计算, 可以验证成立。

因此，最优解为 x₁\* = (1 + √33)/2, x₂\* = (-11 + √33)/2, λ₁\* = (1 + √33)/(1 + √33 + 8), λ₂\* = 16 + 2√33 - (1 + √33)(1 + √33)/(1 + √33 + 8)。

### 本章小结

本章通过六道练习题，复习了向量空间、线性方程组、二次型、概率图模型、最优性条件以及拉格朗日乘子法与 KKT 条件等重要知识点。希望读者能够认真完成这些练习题，加深对相关概念和方法的理解，提升解决实际问题的能力。

## 结束语

本学期我完成了高级工程数学的学习！从线性代数的基石和微积分的精妙，到优化理论的探索和迭代方法的实践，我走过了一段充满挑战与发现的数学旅程。我从向量空间和矩阵的构建开始，逐步深入到线性变换、特征值与特征向量的奥秘之中。领略了微积分的魅力，掌握了序列、极限、导数矩阵以及泰勒级数等重要概念。在优化理论的殿堂里，探索了无约束优化和约束优化的奥秘，学习了梯度下降法、牛顿法、次梯度方法以及临近点算法等强大的工具，更领悟了拉格朗日乘子法和 KKT 条件在解决复杂问题中的精妙运用。

这趟旅程不仅仅是知识的积累，更是思维的锤炼。我学会了如何用严谨的数学语言描述问题，如何运用抽象的数学工具分析问题，更重要的是，学会了如何用数学的思维解决问题。这些知识和能力，将成为未来学习和工作中宝贵的财富。数学的海洋浩瀚无垠，本书仅仅撷取了其中的一小部分。我希望这本教材能够成为我继续探索数学世界的引航灯，激发我对数学的热情，引导我在未来的道路上不断前行。数学不仅仅是一门学科，更是一种思维方式，一种探索未知的工具。在未来的学习和工作中，我将灵活运用所学知识，不断创新，勇攀高峰！