
Chapter 1. 基础知识

21307099

liyj323@mail2.sysu.edu.cn

1 绪论

1.1 最优化

一般描述为:

$$\text{mimize } f(x) \tag{1}$$

$$\text{subject to } x \in \Omega \tag{2}$$

1.2 课程内容

- 凸集、凸函数、凸优化问题
- 对偶理论
- 无约束优化: 算法结构、一阶方法、二阶方法
- 有约束优化: 罚方法、增广拉格朗日乘子法、交替方向乘子法
- 现代优化算法

2 基础知识

2.1 向量

Definition 1. 范数 称一个从向量空间 \mathbb{R}^n 到实数域 \mathbb{R} 的非负函数 $\|\cdot\|$ 为范数, 如果

-
- 正定性 $\|v\| \geq 0, \forall v \in \mathbb{R}^n$ and $\|v\| = 0 \Leftrightarrow v = 0$
- 齐次性 $\|\alpha v\| = |\alpha| \|v\|, \forall v \in \mathbb{R}^n, \alpha \in \mathbb{R}$
- 三角不等式 $\|v + w\| \leq \|v\| + \|w\|, \forall v, w \in \mathbb{R}^n$

Definition 2. 相容范数

满足相容性 (次可乘性) 的范数, 即 $\|AB\| \leq \|A\| \cdot \|B\|$

注意到 AB, A, B 可能在不同的空间上, 相容范数应在所有空间上都满足该性质.

对于给定 v, w :

- 内积: $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$
- $\ell_\infty = \max_{1 \leq i \leq n} |x_i|$
- $\ell_p (p \geq 1) : \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$
- ℓ_0 半范: $\|x\| = (x \text{ 的所有分量中非 } 0 \text{ 元素的个数})$

Proposition 1.

$$\forall x \in \mathbb{R} : \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n \|x\|_\infty$$

Cauchy-Schwarz 不等式:

$$\forall x, y \in \mathbb{R}^n : -\|x\|_2 \|y\|_2 \leq \langle x, y \rangle \leq \|x\|_2 \|y\|_2$$

Hölder 不等式:

$$|\langle x, y \rangle| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

Minkowski 不等式:

$$\forall x, y \in \mathbb{R}^n, p \in [1, \infty)$$

$$\|x + y\|_p \leq \|x\|_p + \|y\|_q$$

2.2 矩阵

$$A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times n}$$

Definition 3. 矩阵

- 内积: $\langle A, B \rangle = \text{tr}(AB^T)$ 即对应位置元素积的和
- Frobenius 范数: $\|A\| = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\text{tr}(A^T A)}$
- 诱导范数
我们希望 $\|AX\| \leq \|A\| \|x\|$, 故定义

$$\|A\| \triangleq \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}$$

- 诱导 ℓ_1 范数 (最大列和范数)

$$\|A\|_1 = \max_j \sum_i |a_{ij}|$$

- 诱导 ℓ_2 范数 (谱范数)

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

特别地, 当 A 是 Hermi

- 诱导 ℓ_∞ 范数 (最大行和范数)

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$

- 核范数

$$\|A\|_* = \sum_{i=1}^r \sigma_i, r = \text{rank} A, \sigma \text{ 为非零奇异值}$$

Definition 4. 正定矩阵

如果对于任意 $x \neq 0, x^T A x > 0$, 则称 A 是正定的, 记为 $A \succ 0$

Definition 5. 半正定矩阵

如果对于任意 $x \neq 0, x^T A x \geq 0$, 则称 A 是半正定的, 记为 $A \succeq 0$

$$\forall x \neq 0, x^T A x \geq 0 \quad (3)$$

$$\Rightarrow \forall A \vec{v} = \lambda \vec{v}, v^T A v \geq 0 \quad (4)$$

$$\Rightarrow \lambda v^T v \geq 0 \quad (5)$$

$$\text{Since } v^T v \geq 0 \quad (6)$$

$$\Rightarrow \lambda \geq 0 \quad (7)$$

2.3 Basic Concepts In Optimization

Definition 6. 最优化问题的数学模型

$$\text{mimize } f(x) \quad (8)$$

$$\text{subject to } x \in \Omega \quad (9)$$

- 决策变量 $x = (x_1, \dots, x_n)^T$

- $f : \Omega \rightarrow \mathbb{R}$

- 可行域 (决策集) $\Omega \subseteq \mathbb{R}^n$ $\begin{cases} \text{无约束优化问题, } \Omega = \mathbb{R}^n, \\ \text{约束优化问题, } \text{Otherwise.} \end{cases}$

可行域: 等式约束, 不等式约束

$$\Omega = \{x \in \mathbb{R}^n : h(x) = 0, c(x) \leq 0\}$$

Definition 7.

- 可行域中的点, 或满足约束条件的点称为可行点;
- 对于 $x^* \in \Omega$, 若对任意的 $x \in \Omega$, 都有 $f(x^*) \leq f(x)$, 则称 x^* 是问题的一个全局最优解。对应的目标函数值, 即 $f(x^*)$, 为全局最优值。记为 $x^* \in \arg \min_{x \in \Omega} f(x)$ 。
- 对于局部最优, 若 x^* 还满足对任意的 $x \in \Omega \setminus \{x^*\}$, 都有 $f(x) > f(x^*)$, 则称 x^* 为严格全局最优解;
- 对于 $x^* \in \Omega$, 若存在 x^* 的邻域 $B(x^*, \delta) = \{x : \|x - x^*\| \leq \delta\}$ 使得对于任意的 $x \in \Omega \cap B(x^*, \delta)$, 有 $f(x^*) \leq f(x)$, 则称 x^* 为局部最优解;
- 对于局部最优, 若 x^* 满足对任意的 $x \in \Omega \cap B(x^*, \delta), x \neq x^*$, 有 $f(x) > f(x^*)$, 则称 x^* 为严格局部最优解。

2.4 迭代算法

基本框架

- 取初始点 $x^0 \in \mathbb{R}^n$ 及其他有关参数, $k = 0$.

- 验证停机准则.

- 给出搜索方向 $d^k \in \mathbb{R}^n$, 通常要求 d^k 是下降方向, 即 $(\nabla f_0(x^k))^T d^k < 0$.

- 计算迭代步长 $\alpha_k > 0$ 使得 $f_0(x^k + \alpha_k d^k) < f_0(x^k)$.

- $x^{k+1} := x^k + \alpha_k d^k, k := k + 1$.

2.5 最优化问题

2.5.1 分类

- 约束优化 & 非约束优化
- 连续优化 & 非连续优化

$$\Omega = \{x \in \mathbb{R}^n : h(x) = 0, c(x) \leq 0\}$$

- 线性规划: f, h, c 都是线性的
- 非线性规划: f, h, c 至少有一个非线性的
- 二次规划: f 是二次函数, h, c 是线性
- 凸优化: 目标函数为凸函数, 可行域为凸集

2.5.2 算法评价标准

全局收敛与局部收敛

- 若 $\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$, 则称 x^k 收敛于 x^*
- 子列收敛于 x^* , 称聚点
- 全局收敛性、局部收敛性

收敛速度

1 设算法产生的迭代点列 x^k 收敛于 x^* , 即 $x^k \rightarrow x^*$

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = \mu$$

- $\mu \in (0, 1)$ 称 Q-线性收敛 (Q 表示 quotient (分式), 线性表示误差取对数后, 随迭代步数 k 显线性)
- $\mu = 0$: Q-超线性收敛, $\mu = 1$: Q-次线性收敛

2

$$\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|^p} = \mu, p > 1 \quad \text{称为 Q-p 阶收敛到 } x^*$$

3 R-线性收敛, R-超线性收敛

3 凸集 仿射集 凸锥

Definition 8. 凸集 Convex Set

过集合 C 内任意两点的线段都在 C 内, 则称 C 为凸集:

$$x_1, x_2 \in C \Rightarrow \theta_1 x_1 + \theta_2 x_2 \in C, \forall \theta_1 + \theta_2 = 1, \theta_1, \theta_2 \geq 0$$

Definition 9. 凸组合 Convex Combination

$\theta_i \geq 0$ and $\sum_{i=1}^m \theta_i = 1$ 则称 $\sum_{i=1}^m \theta_i x_i$ 为 x_1, \dots, x_m 的一个凸组合

Definition 10. 仿射集 Affine Set

过集合 C 内任意两点的直线都在 C 内, 则称 C 为仿射集:

$$x_1, x_2 \in C \Rightarrow \theta_1 x_1 + \theta_2 x_2 \in C, \forall \theta_1 + \theta_2 = 1$$

显然仿射集的要求更高, 因此任何仿射集都是凸集, 但凸集未必是仿射集

Definition 11. 仿射组合 Affine Combination

$\sum_{i=1}^m \theta_i = 1$ 则称 $\sum_{i=1}^m \theta_i x_i$ 为 x_1, \dots, x_m 的一个仿射组合

3.1 仿射维数和相对内部

Definition 12. 仿射维数

集合 C 的仿射维数为其仿射包的维数。

Definition 13. 相对内部

定义集合 C 的相对内部为 $\text{aff } C$ 的内部, 记为 $\text{relint } C$ 即

$$\text{relint } C = \{x \in C | B(x, r) \cap \text{aff } C \subseteq C, \exists r > 0\}$$

Definition 14. 相对边界

定义集合 C 的相对边界为 $\text{cl}C \setminus \text{relint } C$

Definition 15. 凸锥 Convex Cone

$\forall x \in C, \theta \geq 0$, 都有 $\theta x \in C$, 则称 C 为锥

若 C 为凸集, 则称 C 为凸锥, 即

$$x_1, x_2 \in C \Rightarrow \theta_1 x_1 + \theta_2 x_2 \in C, \forall \theta_1, \theta_2 \geq 0$$

Definition 16. 凸锥组合 Conic Combination

$$\theta_1 \dots \theta_m \geq 0$$

称

$$\sum_{i=1}^m \theta_i x_i$$

为 $x_1 \dots x_m \geq 0$ 的一个凸锥组合

Definition 17. 凸包 仿射包 凸锥包

集合 C 中任意元素的凸组合、仿射组合、凸锥组合称为 C 的凸包、仿射包、凸锥包

3.2 超平面与半空间

超平面:

$$\{x \in \mathbb{R}^n | a^T x = b\}$$

或者

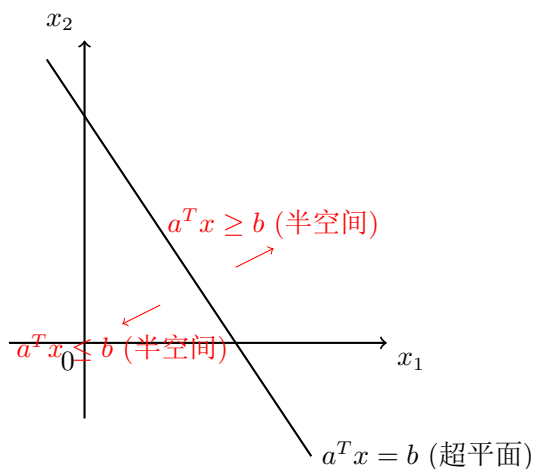
$$\{x \in \mathbb{R}^n | a^T (x - x_0) = 0\}$$

其中 x_0 是超平面上任意一点。超平面是凸集、仿射集，过原点时为凸锥

半空间:

$$\{x \in \mathbb{R}^n | a^T x \leq b\}, a \neq 0$$

显然，半空间是凸集，但不是仿射集。过原点时为凸锥



3.3 范数球

Definition 18. Norm Ball

$$\{x \in \mathbb{R}^n \mid \|x - x_c\|_p \leq r\}$$

显然, 范数球是凸集且当 $r \neq 0$ 时不可能是仿射集和凸锥

3.4 球和椭球

Definition 19. 球

$$b(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} \quad (10)$$

$$= \{x \mid \sqrt{(x - x_c)^T (x - x_c)} \leq r\} \quad (11)$$

Definition 20. 椭球

$$\mathcal{E} = \{x \in \mathbb{R}^n \mid (x - x_c)^T P (x - x_c) \leq 1\}$$