

数值最优化算法及应用

张鑫航

国防科技大学

版本：1.0

更新：2024 年 7 月 28 日



目录

1	绪论	3
1.1	最优化简介	3
1.2	最优化方法概述	3
1.2.1	常见的优化方法	4
1.2.2	算法评价体系	5
2	无约束优化最优性问题	8
2.1	无约束优化最优性条件	8
2.2	无约束优化线搜索方法	11
2.2.1	精确线搜索方法	11
2.2.2	非精确线搜索方法	13
2.3	最速算法与牛顿算法	14
2.3.1	最速下降算法	14
2.3.2	牛顿算法	15
2.4	共轭梯度法	16
2.4.1	线性共轭方向法	17
2.5	拟牛顿算法	18
2.5.1	对称秩-1 校正公式	19
2.5.2	DFP 校正公式	19
2.6	信赖域方法	20
2.7	最小二乘问题	21
2.7.1	线性最小二乘	22
2.7.2	非线性最小二乘	22
3	凸优化	23

3.1	凸分析基础	23
3.1.1	临近点算子	25
3.1.2	和式优化最优性条件	27
3.2	线性化临近点方法	27
3.3	交替极小化方法	28
3.3.1	和式凸优化交替极小化方法	30
3.3.2	二分块交替极小化方法	31
3.4	凸分析基础 II	31
3.4.1	凸集与凸集分离定理	31
4	约束优化最优性条件	33
4.1	等式优化最优化条件	34
4.1.1	不等式优化约束条件	35
4.2	鞍点与凸规划	36
4.3	约束优化对偶	36
4.3.1	Lagrange 对偶	36
4.3.2	对偶定理	37
4.3.3	混合约束优化的对偶	38
4.4	二次规划	40
4.4.1	模型与基本性质	40
4.4.2	对偶理论	41
4.4.3	等式约束二次规划	41
4.5	有效集方法	43
4.6	约束优化可行方向法	45
4.6.1	Zoutendijk 可行方向法	46
4.7	投影算子	46

4.7.1	矩阵的广义逆	48
4.8	约束优化投影算法	49
5	其他方法	49
5.1	罚函数方法	49
5.1.1	外点罚函数	49
5.2	乘子罚函数	49
6	张量优化	50
6.1	基本概念	50

1 绪论

1.1 最优化简介

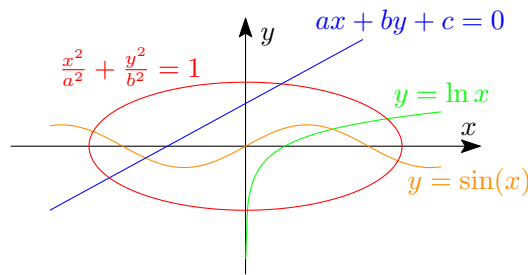
优化问题可以表示成

$$\begin{aligned} \min f(\mathbf{x}) \\ \text{s. t. } \mathbf{x} \in \Omega \end{aligned}$$

可行域有多种表达形式，一般用等式和不等式定义，即

$$\Omega = \{\mathbf{x} \in \mathbb{R}^n | c_i(\mathbf{x}) = 0, i \in \mathcal{E}; c_i(\mathbf{x}) \geq 0, i \in \mathcal{I}\}$$

注 线性问题：



- 线性规划

$$\begin{aligned} \min \mathbf{c}^T \mathbf{x} \\ \text{s. t. } \mathbf{Ax} = \mathbf{0} \\ \mathbf{x} \geq 0 \end{aligned}$$

- 非线性规划

$$\begin{aligned} \min \frac{1}{2} \mathbf{x}^T \mathbf{Gx} + \mathbf{g}^T \mathbf{x} \\ \text{s. t. } \mathbf{A}_1 \mathbf{x} = \mathbf{b}_1 \\ \mathbf{A}_2 \mathbf{x} \geq \mathbf{b}_2 \end{aligned}$$

例 1.1 若目标函数在可行域上有下界，但无最优解，则目标函数在可行域上的下确界成为优化问题的最优值。

$$\inf \{f(\mathbf{x}) | \mathbf{x} \in \Omega\}$$

如二元函数 $f(\mathbf{x}) = x_1^2 + (1 - x_1 x_2)^2$ 在 \mathbb{R}^2 上的最优值为 0，但只有在 $x_1 = \frac{1}{x_2}$ 且 $x_2 \rightarrow \infty$ 是才能达到。

1.2 最优化方法概述

1.2.1 常见的优化方法

$$\mathbf{x}^* \stackrel{?}{=} \arg \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$$

1. 解析法：给出问题最优解的解析式

- 二次优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

若 Hesse 阵正定，最优解： $\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$

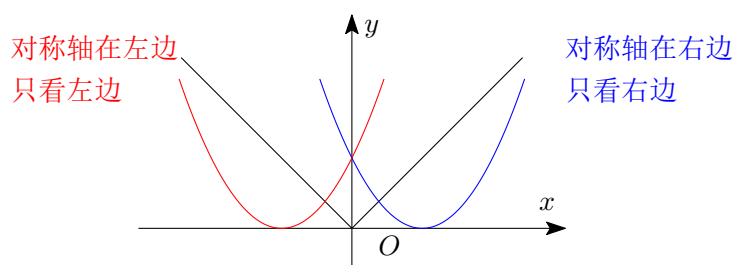
- 可分离的优化问题，其中 $\mathbf{a} \in \mathbb{R}^n, \lambda > 0$ ★★★★★

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{a}\|^2 + \lambda \|\mathbf{x}\|_1$$

求解子问题

$$\begin{aligned} & \sum_{i=1}^n \left(\min_{x_i \in \mathbb{R}} \frac{1}{2} (x_i - a_i)^2 + \lambda |x_i| \right) \\ &= \begin{cases} \min_{x_i \in \mathbb{R}} \frac{1}{2} (x_i - a_i)^2 + \lambda x_i, & a_i \geq 0, \text{ 对称轴在右边} \\ \min_{x_i \in \mathbb{R}} \frac{1}{2} (x_i - a_i)^2 - \lambda x_i, & a_i < 0, \text{ 对称轴在左边} \end{cases} \end{aligned}$$

最优解



$$x_i^* = \begin{cases} 0 & \text{若 } |a_i| \leq \lambda \\ a_i - \text{sgn}(a_i) \lambda & \text{若 } |a_i| > \lambda \end{cases}$$

2. 图解法与实验法（手工作坊）

- 图解法：根据目标函数的图像求解
- 实验法：依据一定规则对变量取不同的值，通过实验观察目标函数值的变化规律，找出问题的最优解。

3. 形式转化法：利用问题结构或最优性条件将其转化为有容易求解的另一类数学问题，然后对后者套用现有的方法求解。

4. 数值迭代法：利用函数值或梯度信息按一定规则从一个点产生一个新的点，直到不能改进为止。得到的是数值解，是近似解。

- 模式搜索法：根据函数值变化规律探测函数的下降方向并沿着该方向寻求更优的点。
- 梯度方法：最优解、最优值未知。

注 线搜索方法：在每一步迭代，首先基于目标函数梯度信息产生搜索方向，然后沿该方向寻求目标函数的（近似）最小值点。重复上述过程，直到不能改进。

1. 取初始点 $\mathbf{x}_0 \in \mathbb{R}^n$ ，及有关参数。令 $k = 0$
2. 验证停机准则。
3. 计算当前迭代点 \mathbf{x}_k 的搜索方向 $\mathbf{d}_k \in \mathbb{R}^n$
4. 计算步长 $\alpha_k > 0$ 使满足

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k)$$

$$\alpha_k = \arg \min_{\alpha_k > 0} f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)$$

5. 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ ，转到第 2 步

注 信赖域方法：基于当前迭代点信息建立目标函数在该点邻域内的一个近似二次模型，然后将该模型小邻域内的最小值点作为新的迭代点。

$$f(\mathbf{x}_k + \mathbf{d}) = f(\mathbf{x}_k) + \mathbf{d}^T \nabla f(\mathbf{x}_k) + \frac{1}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}_k) \mathbf{d} + o(\|\mathbf{d}\|^2)$$

$$\approx f(\mathbf{x}_k) + \mathbf{d}^T \nabla f(\mathbf{x}_k) + \frac{1}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}_k) \mathbf{d}$$

$$m_k(\mathbf{d}) = f(\mathbf{x}_k) + \mathbf{d}^T \nabla f(\mathbf{x}_k) + \frac{1}{2} \mathbf{d}^T \mathbf{B}_k \mathbf{d}$$

其中， \mathbf{B}_k 为 $\nabla^2 f(\mathbf{x}_k)$ 或近似

$$\min_{\mathbf{d} \in \mathbb{R}^n} f(\mathbf{x}_k + \mathbf{d}) \Rightarrow \min\{m_k(\mathbf{d}) \mid \mathbf{d} \in \mathbb{R}^n, \|\mathbf{d}\| \leq \Delta_k\}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$$

1.2.2 算法评价体系

注 收敛性

- 全局收敛：从任意的初始点出发，算法产生的迭代点列都收敛到问题的最优值点。
- 局部收敛：只有在初始点和最优值点具有某种程度的靠近时才能保证迭代点列收敛到最优值点。
- 弱收敛：迭代点列的某一聚点为优化问题的最优值点。

注 收敛速度

- Q-收敛：前后两迭代点靠近最优值点的距离之比

$$\limsup_{k \rightarrow \infty} = \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \leq q$$

Q-线性收敛： $0 < q < 1$

Q-超线性收敛： $q = 0$

Q-r 阶收敛：若存在 $0 \leq p < \infty$ 和 $r \geq 1$

$$\limsup_{k \rightarrow \infty} = \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|^r} \leq q$$

r 越大，收敛速度越快。若 $r > 1$ ，可推出其超线性收敛。

- R-线性收敛：借助一趋于零的等比数列度量迭代点收敛的快慢程度

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq \kappa q^k, \kappa > 0, q \in (0, 1)$$

R-超线性收敛：若存在 $\kappa > 0$ 和收敛于 0 的正数列 $\{q_k\}$

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq \kappa \prod_{i=0}^k q_i$$

Q-（超）线性收敛 \Rightarrow R-（超）线性收敛

例 1.2 证明无约束优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ 等价于下述约束优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}} \{t | t - f(\mathbf{x}) \geq 0\}$$

证明. \mathbf{x}^* 为无约束优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ 的最优解，当且仅当对任意的 $\mathbf{x} \in \mathbb{R}^n, f(\mathbf{x}) - f(\mathbf{x}^*) \geq 0$ 。所以优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}} \{t | t - f(\mathbf{x}) \geq 0\}$$

的最优值为 $f(\mathbf{x}^*)$ 。证毕！

例 1.3 设点列 $\{\mathbf{x}_k\}$ Q-超线性收敛到 \mathbf{x}^* ，则

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

证明. 设点列 $\{\mathbf{x}_k\}$ Q-超线性收敛到 \mathbf{x}^* ，则

$$\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 0$$

从而由三角不等式

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} &\leq \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\| + \|\mathbf{x}^* - \mathbf{x}_k\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \\ &= \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 1 \end{aligned}$$

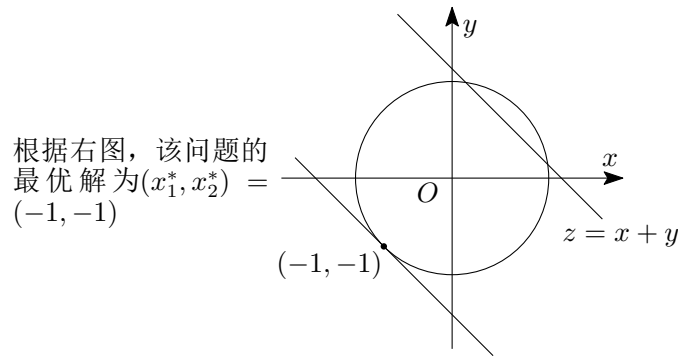
另一方面,

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} &\geq \frac{\|\mathbf{x}^* - \mathbf{x}_k\| - \|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \\ &= \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 1\end{aligned}$$

两式结合得命题结论。证毕！

例 1.4 用图解法求解下述优化问题的最优解

$$\min \{x_1 + x_2 | x_1^2 + x_2^2 \leq 2\}$$



例 1.5 设 $\mathbf{a} \in \mathbb{R}^n, \kappa > 0, \|\mathbf{a}\| > 1$, 试求解优化问题

$$\begin{aligned}\min \quad & \|\mathbf{x}\| \\ \text{s. t.} \quad & \|\mathbf{x}\| - \mathbf{a}^T \mathbf{x} + \kappa \leq 0\end{aligned}$$

解：由 Cauchy-Schwarz 不等式，对同样长度的 \mathbf{x} , \mathbf{x} 与 \mathbf{a} 同向时达到最小， \mathbf{x} 的取值范围达到最大。故该优化问题的最优解满足 $\mathbf{x} = t\mathbf{a}$, 其中 $t > 0$ 待定。故上述优化问题化为

$$\begin{aligned}\min \quad & t\|\mathbf{a}\| \\ \text{s. t.} \quad & t\|\mathbf{a}\| - t\|\mathbf{a}\|^2 + \kappa \leq 0\end{aligned}$$

即

$$\begin{aligned}\min \quad & t\|\mathbf{a}\| \\ \text{s. t.} \quad & t \geq \frac{\kappa}{\|\mathbf{a}\|^2 - \|\mathbf{a}\|}\end{aligned}$$

显然，其最优解为 $t^* = \frac{\kappa}{\|\mathbf{a}\|^2 - \|\mathbf{a}\|}$ 。这样，原问题的最优解为

$$\mathbf{x}^* = t^* \mathbf{a} = \frac{\kappa \mathbf{a}}{\|\mathbf{a}\|^2 - \|\mathbf{a}\|}$$

例 1.6 设 $\mathbf{a} \in \mathbb{R}^n, \rho > 0$, 矩阵 $\Sigma_0 \succ 0$ (对称正定)。试用解析法求解如下优化问题。

$$\begin{aligned}\max_{\Sigma \succ 0} \quad & \mathbf{a}^T \Sigma \mathbf{a} \\ \text{s. t.} \quad & \|\Sigma - \Sigma_0\|_F \leq \rho\end{aligned}$$

解：令 $\Delta \Sigma = \Sigma - \Sigma_0$ ，则原优化问题转化为

$$\begin{aligned} \max \quad & \mathbf{a}^T \Delta \Sigma \mathbf{a} + \mathbf{a}^T \Sigma_0 \mathbf{a} \\ \text{s. t.} \quad & \|\Delta \Sigma\|_F \leq \rho \end{aligned}$$

由 Cauchy-Schwarz 不等式：

$$\mathbf{a}^T \Delta \Sigma \mathbf{a} = \langle \Delta \Sigma, \mathbf{a} \mathbf{a}^T \rangle \leq \|\Delta \Sigma\|_F \|\mathbf{a} \mathbf{a}^T\|_F \leq \rho \|\mathbf{a} \mathbf{a}^T\|_F$$

当且仅当 $\Delta \Sigma$ 与 $\mathbf{a} \mathbf{a}^T$ 方向一致时取等号，当 $\|\Delta \Sigma\|_F = \rho$ 是取最优值。

原问题的最优解为

$$\Sigma^* = \Sigma_0 + \rho \frac{\mathbf{a} \mathbf{a}^T}{\|\mathbf{a} \mathbf{a}^T\|_F}$$

上述问题的最优解为

$$\mathbf{a}^T \Sigma^* \mathbf{a} = \mathbf{a}^T \Sigma_0 \mathbf{a} + \rho \mathbf{a}^T \mathbf{a}$$

2 无约束优化最优性问题

2.1 无约束优化最优性条件

数学模型 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ ，其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$

注 最优解判定：

- 直观地，验证最优解，需将其和附近点的目标函数值逐一比较，工作量太大。
- 利用目标函数的连续可微性质，可建立一个简易的判断方法，这就是光滑优化问题的最优性条件。

定理 2.1 (无约束优化问题) \mathbf{x}^* 是 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ (局部) 最优解，则 $\nabla f(\mathbf{x}^*) = 0$

证明. 反证法。取 $\mathbf{d} = -\nabla f(\mathbf{x}^*)$ ， $\alpha > 0$ 充分小。则

$$\begin{aligned} f(\mathbf{x}^* + \alpha \mathbf{d}) &= f(\mathbf{x}^*) + \alpha \nabla f(\mathbf{x}^*)^T \mathbf{d} + o(\alpha) \\ &= f(\mathbf{x}^*) - \alpha \|\nabla f(\mathbf{x}^*)\|^2 + o(\alpha) \\ &< f(\mathbf{x}^*) \end{aligned}$$

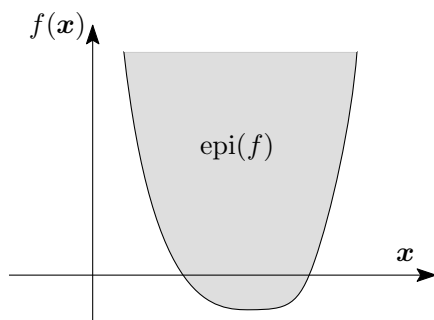
和 \mathbf{x}^* 是 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ (局部) 最优解矛盾。证毕！

定义 2.1 (凸函数) $f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y})$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \lambda \in [0, 1]$ 或者

$$f\left(\sum_{i=1}^m \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^m \lambda_i f(\mathbf{x}_i), \forall \mathbf{x}_i \in \mathbb{R}^n, \lambda_i \geq 0, i = 1, 2, \dots, m, \sum_{i=1}^m \lambda_i = 1$$

定义 2.2 (凸组合) 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n$, $\lambda_1 + \lambda_2 + \dots + \lambda_m = 1$, $\lambda_1, \lambda_2, \dots, \lambda_m \in (0, 1)$, 则称 $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \dots + \lambda_m \mathbf{x}_m$ 为 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 的一个凸组合。

定义 2.3 (凸函数的等价定义 1 (上图)) 函数 f 的上图 $\text{epi}(f)$ 是凸集。 $\text{epi}(f) := \{(\mathbf{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \geq f(\mathbf{x})\}$ 是凸集。



定义 2.4 (凸函数的等价定义 2 (连续可微)) 切平面在曲面下方。 $f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

定义 2.5 (凸函数的等价定义 3 (二阶连续可微)) $\mathbf{h}^T \nabla^2 f(\mathbf{x}) \mathbf{h} \geq 0$, $\forall \mathbf{x}, \mathbf{h} \in \mathbb{R}^n$

定义 2.6 (一致凸函数 (凸函数的加强版)) 对函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 若存在 $\eta > 0$, 使对任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ 和 $\lambda \in (0, 1)$ 成立

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \lambda(1 - \lambda) \eta \|\mathbf{x} - \mathbf{y}\|^2$$

则称其为一一致凸函数, 又称强凸函数。

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \lambda(1 - \lambda) \eta \|\mathbf{x} - \mathbf{y}\|^2$$

↓ 连续可微

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{1}{2} \eta \|\mathbf{y} - \mathbf{x}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

↕ 或者

$$(\mathbf{y} - \mathbf{x})^T (\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})) \geq \eta \|\mathbf{y} - \mathbf{x}\|^2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

↓ 二阶连续可微

$$\mathbf{h}^T \nabla^2 f(\mathbf{x}) \mathbf{h} \geq \eta \|\mathbf{h}\|^2, \quad \forall \mathbf{x}, \mathbf{h} \in \mathbb{R}^n$$

注 线性函数是凸函数但不一致凸。对二次函数, 一致凸与严格凸等价。

注 $f(\mathbf{x})$ 一致凸等价于 $f(\mathbf{x}) - \eta \|\mathbf{x}\|^2$ 为凸函数, 存在 $\eta > 0$

定理 2.2 (凸优化) $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为连续可微的凸函数。凸优化问题的稳定点为其全局最优解。

证明. 利用凸函数的性质

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}^*) &\geq \langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle = 0 \\ f(\mathbf{x}) &> f(\mathbf{x}^*) \end{aligned}$$

定理 2.3 (无约束优化二阶最优性条件) 设 \mathbf{x}^* 为优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ 的最优解, 则 $\nabla f(\mathbf{x}^*) = \mathbf{0}, \nabla^2 f(\mathbf{x}^*)$ 半正定。

证明. 结论 $\nabla f(\mathbf{x}^*) = \mathbf{0}$ 已证。

反设 $\nabla^2 f(\mathbf{x}^*)$ 非半正定, 则存在向量 \mathbf{d} 使得

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} < 0.$$

取 $\alpha > 0$ 充分小。由泰勒展开式

$$\begin{aligned} f(\mathbf{x}^* + \alpha \mathbf{d}) &= f(\mathbf{x}^*) + \alpha \nabla f(\mathbf{x}^*)^T \mathbf{d} + \frac{1}{2} \alpha^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\alpha^2) \\ &< f(\mathbf{x}^*). \end{aligned}$$

得矛盾。证毕！

注 二阶必要条件的非充分性：

函数 $f(x)$ 在零点满足 $\nabla f(\mathbf{x}^*) = \mathbf{0}, \nabla^2 f(\mathbf{x}^*)$ 半正定, 但零点不是其最优值点。

定理 2.4 (二阶充分条件) 对优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$, 设 $\nabla f(\mathbf{x}^*) = \mathbf{0}, \nabla^2 f(\mathbf{x}^*)$ 正定, 则 \mathbf{x}^* 为该优化问题的严格最优解

证明. 对任意充分靠近 \mathbf{x}^* 的 \mathbf{x} , 存在单位向量 \mathbf{d} , 使得

$$\mathbf{x} = \mathbf{x}^* + \alpha \mathbf{d}$$

其中 $\alpha > 0$ 充分小。则

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2} \alpha^2 \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} + o(\alpha^2) > f(\mathbf{x}^*).$$

例 2.1 $\min_{x>0, y \geq 0} f(x, y) = \frac{10}{x} + \frac{(x-y)^2}{2x} + \frac{3y^2}{2x}$ ★★★★★

先忽略约束：利用 $\min_{x,y} f(x, y) = \min_x \min_y f(x, y)$, 先固定 x , 关于 y 做内层优化, 再求解关于 x 的外层优化

$$\min_y f(x, y) = \frac{10}{x} + \frac{(x-y)^2}{2x} + \frac{3y^2}{2x}$$

目标函数关于 y 为凸函数，利用最优性条件得最优解

$$\begin{aligned} f(x, y) &= \frac{1}{2x} (20 + x^2 - 2xy + 4y^2) \\ y &= \frac{1}{4}x \end{aligned}$$

将上述最优解代入目标函数得外层优化问题

$$\min f(x) = \frac{10}{x} + \frac{3}{8}x$$

目标函数关于 x 为凸函数（二阶导数大于 0）。再利用最优性条件得 $x = \frac{4}{3}\sqrt{15}, y = \frac{1}{4}x = \frac{1}{3}\sqrt{15}$

它们满足约束条件，自然为原问题的最优解。

2.2 无约束优化线搜索方法

2.2.1 精确线搜索方法

注 优化模型：

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

其中， $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微。

注 有关记号：

$$\begin{aligned} \mathbf{g}_k &= \nabla f(\mathbf{x}_k), & \mathbf{g}(\mathbf{x}) &= \nabla f(\mathbf{x}) \\ \mathbf{G}_k &= \nabla^2 f(\mathbf{x}_k), & \mathbf{G}(\mathbf{x}) &= \nabla^2 f(\mathbf{x}). \end{aligned}$$

我们先回顾作为最速下降方向的负梯度方向。假设 $\nabla f(\mathbf{x}) \neq \mathbf{0}$ 。固定方向 $\mathbf{d} \in \mathbb{R}^n$ 且 $\|\mathbf{d}\| = 1$ 。考虑沿此方向函数值的变化。由可微的定义可知

$$f(\mathbf{x} + \tau \mathbf{d}) = f(\mathbf{x}) + \tau \nabla f(\mathbf{x})^T \mathbf{d} + o(\tau), \quad \tau > 0$$

当 $\nabla f(\mathbf{x})^T \mathbf{d} < 0$ 时，由于 $\frac{o(\tau)}{\tau} \rightarrow 0, \tau \rightarrow 0_+$ 可知，存在 $\varepsilon > 0$ ，使得当 $0 < \tau < \varepsilon$ 时，

$$\frac{o(\tau)}{\tau} \leq -\frac{\nabla f(\mathbf{x})^T \mathbf{d}}{2}$$

此时

$$f(\mathbf{x} + \tau \mathbf{d}) - f(\mathbf{x}) \leq \frac{\tau}{2} \cdot \nabla f(\mathbf{x})^T \mathbf{d} < 0$$

上式表明，从 \mathbf{x} 点沿 \mathbf{d} 方向出发，当步长 $\tau < \varepsilon$ 时，

$$f(\mathbf{x} + \tau \mathbf{d}) < f(\mathbf{x})$$

因此, 称满足条件 $\nabla f(\mathbf{x})^T \mathbf{d} < 0$ 的方向为下降方向。而称下降最快或 $\nabla f(\mathbf{x})^T \mathbf{d}$ 最小的方向为最速下降方向, 也即

$$\begin{aligned}\hat{\mathbf{d}} &= \arg \min \{ \nabla f(\mathbf{x})^T \mathbf{d} \} \\ \text{s.t. } &\|\mathbf{d}\| = 1\end{aligned}$$

则 $\mathbf{d} = -\frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ 。事实上, 有 Cauchy-Schwarz 不等式

$$|\nabla f(\mathbf{x})^T \mathbf{d}| \leq \|\nabla f(\mathbf{x})\| \cdot \|\mathbf{d}\| \leq \|\nabla f(\mathbf{x})\|$$

可知

$$-\|\nabla f(\mathbf{x})\| \leq \nabla f(\mathbf{x})^T \mathbf{d} \leq \|\nabla f(\mathbf{x})\|$$

从而 $\hat{\mathbf{d}} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ 可取得下界 $-\|\nabla f(\mathbf{x})\|$ 。

注 算法分析:

- 迭代过程: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$
- 精确线搜索 (最优步长): $\alpha = \arg \min_{\alpha \geq 0} f(\mathbf{x}_k + \alpha \mathbf{d}_k)$
- $\{\mathbf{x}_n\}$ 满足 $\{f(\mathbf{x})\}$ 递减, $\nabla f(\mathbf{x}) \rightarrow 0$
- 沿下降方向寻求目标函数的最大下降点
- 基本性质: $f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) \leq f(\mathbf{x}_k + \alpha \mathbf{d}_k), \quad \forall \alpha > 0$

$$f'_\alpha(\mathbf{x}_k + \alpha_k \mathbf{d}_k) = \nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^T \mathbf{d}_k = 0$$

- 最优步长的计算: 一元函数的极值问题, 全局最优步长难求. 通常只能求得局部最优或近似最优步长。
- 常用的方法: 黄金分割法、多项式插值

精确线搜索下降算法:

1. 取初始点 $\mathbf{x}_0 \in \mathbb{R}^n$ 和参数 $\varepsilon \geq 0$, 令 $k = 0$
2. 若 $\|\mathbf{g}_k\| \leq \varepsilon$, 算法终止; 否则进入下一步
3. 计算下降方向 \mathbf{d}_k 满足 $\mathbf{g}_k^T \mathbf{d}_k < 0$
4. 计算步长

$$\alpha_k = \arg \min \{ f(\mathbf{x}_k + \alpha \mathbf{d}_k) \mid \alpha \geq 0 \}$$

5. 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, k = k + 1$, 转步 2

注 算法收敛性: 要保证算法收敛, 要对搜索方向、目标函数做些假设。

定理 2.5 设目标函数 $f \in C^2$ (二阶连续) 有下界

下降方向 \mathbf{d}_k 与负梯度方向 $-\mathbf{g}_k$ 夹角满足

$$\theta_k \leq \frac{\pi}{2} - \theta, 0 < \theta \leq \frac{\pi}{2}$$

若精确线搜索方法产生无穷迭代点列, 且满足 $\|\nabla^2 f(\mathbf{x}_k + \alpha \mathbf{d}_k)\| \leq M, \forall \alpha > 0$, 其中, $M > 0$ 为常数, 则算法收敛, 即 $\lim_{k \rightarrow \infty} \mathbf{g}_k = \mathbf{0}$

2.2.2 非精确线搜索方法

注 最优步长的缺陷:

- 最优步长的初衷是在每一迭代步使目标函数下降量达到最大. 但计算量太大, 而且不好求.
- 最优化问题关注全局最优值点。集中于某个方向上的线搜索似乎没有必要。

设 \mathbf{d}_k 为 \mathbf{x}_k 的下降方向, 即 $\mathbf{d}_k^T \mathbf{g}_k < 0$, 对 $\sigma \in (0, 1)$, 只要 $\alpha > 0$ 充分小, 有

$$f(\mathbf{x}_k + \alpha \mathbf{d}_k) = f(\mathbf{x}_k) + \alpha \mathbf{g}_k^T \mathbf{d}_k + o(\alpha) < 0$$

步长选取: 先取一个较大的步长, 看目标函数是否有满意的下降量。否则依次按比例压缩, 直至满足要求。又称进退试探法。

定义 2.7 (Armijo 步长规则) 取常数 $\beta > 0, \sigma, \gamma \in (0, 1)$ 。步长 $\alpha_k = \beta \gamma^{m_k}$, 其中 m_k 为 $0, 1, \dots$ 中满足下式的最小值

$$f(\mathbf{x}_k + \beta \gamma^{m_k} \mathbf{d}_k) \leq f_k + \sigma \beta \gamma^{m_k} \mathbf{g}_k^T \mathbf{d}_k$$

如 $\alpha_k < \beta$, 则

$$\begin{aligned} f(\mathbf{x}_k + \beta \gamma^{m_k} \mathbf{d}_k) &\leq f(\mathbf{x}_k) + \sigma \beta \gamma^{m_k} \mathbf{g}_k^T \mathbf{d}_k \\ f(\mathbf{x}_k + \beta \gamma^{m_k - 1} \mathbf{d}_k) &> f(\mathbf{x}_k) + \sigma \beta \gamma^{m_k - 1} \mathbf{g}_k^T \mathbf{d}_k \end{aligned}$$

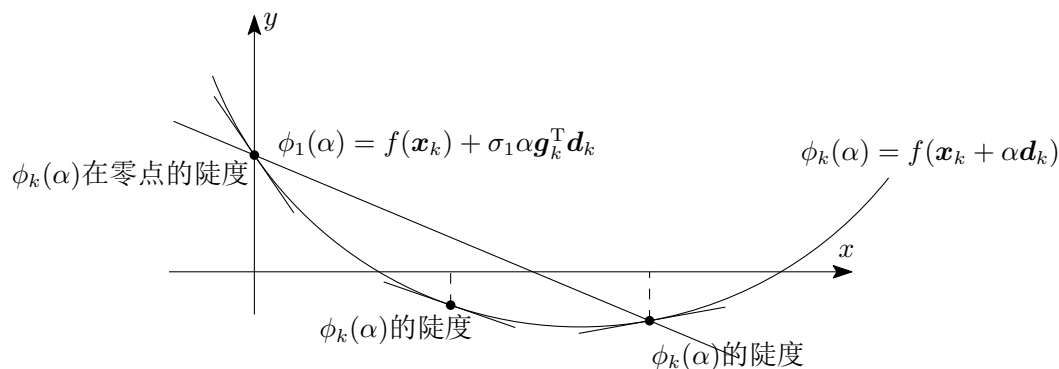
定义 2.8 (Wolfe 步长) 常数 $0 < \sigma_1 < \sigma_2 < 1$ 。步长 α_k 同时满足

$$f(\mathbf{x}_k + \alpha \mathbf{d}_k) \leq f(\mathbf{x}_k) + \sigma_1 \alpha \mathbf{g}_k^T \mathbf{d}_k \quad \text{下降性条件}$$

和

$$\nabla f(\mathbf{x}_k + \alpha \mathbf{d}_k)^T \mathbf{d}_k \geq \sigma_2 \mathbf{g}_k^T \mathbf{d}_k \iff \mathbf{g}_{k+1}^T \mathbf{d}_k \geq \sigma_2 \mathbf{g}_k^T \mathbf{d}_k$$

曲线 $\varphi_k(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$ 的陡度 $\nabla f(\mathbf{x}_k + \alpha \mathbf{d}_k)^T \mathbf{d}_k$ 比 $\alpha = 0$ 点有所减缓从而远离当前迭代点



2.3 最速算法与牛顿算法

2.3.1 最速下降算法

优化模型 $\min_{x \in \mathbb{R}^n} f(x)$

迭代算法

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \alpha_k \mathbf{d}_k \\ \alpha_k &= \arg \min_{\alpha \in \mathbb{R}_+} f(\mathbf{x}_k + \alpha \mathbf{d}_k) \\ \mathbf{d}_k &= -\nabla f(\mathbf{x}_k) \end{aligned}$$

推论 (最优步长规则下的收敛性质) 最优步长规则下的收敛性质 $\mathbf{d}_k^T \mathbf{d}_{k+1} = 0$

证明. 记 $\varphi_k(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{d}_k)$, 利用最优步长规则

$$\begin{aligned} 0 &= \varphi'_k(\alpha_k) = \langle \nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k), \mathbf{d}_k \rangle \\ &= \langle \nabla f(\mathbf{x}_{k+1}), \mathbf{d}_k \rangle \\ &= -\langle \mathbf{d}_{k+1}, \mathbf{d}_k \rangle. \end{aligned}$$

定理 2.6 (收敛速度估计) 对严格凸二次函数 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0$ 为矩阵 \mathbf{G} 的特征根。则最优步长最速下降算法线性收敛, 即

$$\frac{f_{k+1} - f(\mathbf{x}^*)}{f_k - f(\mathbf{x}^*)} \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right)^2, \quad \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \leq \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \sqrt{\frac{\lambda_1}{\lambda_n}}.$$

- 很强条件下 (二次严格凸) 最速下降算法的收敛速度估计;
- 线性收敛是一个很慢的收敛速度;
- 有例为证: 该理论估计是准确的, 没有提升的空间。

例 2.2 $\min_{x \in \mathbb{R}^2} f(x_1, x_2) = \frac{1}{3} x_1^2 + \frac{1}{2} x_2^2$. ★★★★★

显然，唯一最优解 $\mathbf{x}^* = (0, 0)$ ，取初始点 $\mathbf{x}_0 = (3, 2)$ ，那么最速下降方法产生的迭代点列 \mathbf{d}_0

$$\mathbf{x}_k = \left(\frac{3}{5^k}; (-1)^k \frac{2}{5^k} \right) \xrightarrow{k \rightarrow \infty} (0; 0)$$

全局收敛，收敛速度线性！有 $\|\mathbf{x}_k - \mathbf{x}^*\| = \sqrt{13} \left(\frac{1}{5}\right)^k$ 。那么

$$\frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = \frac{1}{5} < \frac{1}{5} \sqrt{\frac{3}{2}} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \sqrt{\frac{\lambda_1}{\lambda_2}}$$

$$\frac{f(\mathbf{x}_{k+1}) - f(\mathbf{x}^*)}{f(\mathbf{x}_k) - f(\mathbf{x}^*)} = \frac{1}{25} = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2$$

注 优点：

- 迭代过程简单，计算量与存储量小；
- 初始步函数值下降较快，可以快速靠近最优解。

缺点：

- 相邻两搜索方向正交，导致锯齿现象：越靠近最优值点，靠近速度越慢。
- 不适于算法收局。

2.3.2 牛顿算法

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

利用目标函数 $f(\mathbf{x}_k + \mathbf{d})$ 在 \mathbf{x}_k 的二阶近似求新的迭代点

$$m_k(\mathbf{d}) \triangleq f(\mathbf{x}_k) + \mathbf{d}^T \mathbf{g}_k + \frac{1}{2} \mathbf{d}^T \mathbf{G}_k \mathbf{d}.$$

利用最优性条件

$$\nabla_{\mathbf{d}} m_k(\mathbf{d}) = \mathbf{g}_k + \mathbf{G}_k \mathbf{d} = \mathbf{0} \longrightarrow \mathbf{d}_k^N = -\mathbf{G}_k^{-1} \mathbf{g}_k$$

令 $\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{G}_k^{-1} \mathbf{g}_k$ ，即得牛顿算法。牛顿步

$$\mathbf{d}_k^N = -\mathbf{G}_k^{-1} \mathbf{g}_k$$

定理 2.7 设目标函数二阶连续可微，Hesse 阵在最优值点 \mathbf{x}^* 非奇异。则

1. 若初始点充分靠近最优值点，则算法超线性收敛， $\lim_{k \rightarrow \infty} \mathbf{x}_k = \mathbf{x}^*$
2. 若 Hesse 阵在最优值点附近 Lipschitz 连续，则二阶收敛。

证明. 利用目标函数二阶展式

$$\mathbf{0} = \mathbf{g}(\mathbf{x}^*) = \mathbf{g}_k + \mathbf{G}_k(\mathbf{x}^* - \mathbf{x}_k) + o(\|\mathbf{x}_k - \mathbf{x}^*\|).$$

左乘 G_k^{-1}

$$\begin{aligned} \mathbf{x}_k - \mathbf{x}^* - G_k^{-1} \mathbf{g}_k &= o(\|\mathbf{x}_k - \mathbf{x}^*\|) \\ \Rightarrow \mathbf{x}_{k+1} - \mathbf{x}^* &= o(\|\mathbf{x}_k - \mathbf{x}^*\|). \end{aligned}$$

证毕！

注 算法特点：

- 无全局收敛性。仅局部收敛，即要求初始点靠近最优值点；
- 二阶收敛：迭代点越靠近最优值点，收敛速度越快！
- 具有二次终止性：对严格凸二次函数，一步即得最优解
- 严格二次函数 $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} + c \Rightarrow \mathbf{x}^* = -\mathbf{G}^{-1} \mathbf{g}$
任意初始点 $\mathbf{x}_0 \in \mathbb{R}^n$

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_0 - G^{-1} \nabla f(\mathbf{x}_0) \\ &= \mathbf{x}_0 - G^{-1} (\mathbf{G} \mathbf{x}_0 + \mathbf{g}) \\ &= -G^{-1} \mathbf{g} = \mathbf{x}^*. \end{aligned}$$

例 2.3 用牛顿法求解 $\min f(x) = \sqrt{1+x^2}$ ★★★★★

0 为最优解

$$\text{目标函数导数 } f'(x) = \frac{x}{\sqrt{1+x^2}} \quad f''(x) = \frac{1}{(1+x^2)^{3/2}}$$

迭代过程

$$\begin{aligned} x_{k+1} &= x_k - \frac{f'(x_k)}{f''(x_k)} \\ &= x_k - x_k(1+x_k^2) \\ &= -x_k^3 \end{aligned}$$

- 当 $x_0 < 1$ ，算法快速收敛到最优解；
- 当 $x_0 \geq 1$ ，算法不收敛

注 牛顿方法的缺陷

- 初始点远离最优解时，算法可能不收敛；
- Hesse 矩阵奇异时，算法不可行；
- 计算量和存储量大：需计算目标函数的梯度和 Hesse 阵

2.4 共轭梯度法

线性方程组 (\mathbf{A} 对称正定)：

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

- 经典算法：Gauss 消元法、系数矩阵三角分解法

- 算法缺陷：计算时间随问题规模急速增长。

将问题转化为

$$\min f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

2.4.1 线性共轭方向法

定义 2.9 (共轭方向) 对对称正定阵 \mathbf{A} , 若 $\mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^n$, 满足 $\mathbf{d}_1^T \mathbf{A} \mathbf{d}_2 = 0$, 则称 $\mathbf{d}_1, \mathbf{d}_2$ 关于矩阵 \mathbf{A} 共轭, 并称其为 \mathbf{A} 的共轭方向。

共轭是正交的推广。

定义 2.10 (线性共轭方向的推广) 若向量组 $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$ 关于对称正定阵 \mathbf{A} 两两共轭, 即满足

$$\mathbf{d}_i^T \mathbf{A} \mathbf{d}_j = 0, 1 \leq i \neq j \leq n$$

推论 若向量组 $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$, 关于矩阵 \mathbf{A} 共轭, 则它们线性无关。

注 线性共轭方向法:

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

1. 初始点 \mathbf{x}_0 , 搜索方向 \mathbf{d}_0 满足 $\langle \mathbf{d}_0, \mathbf{g}_0 \rangle < 0$, 终止参数 $\varepsilon \geq 0$, 令 $k = 0$
2. 若 $\|\mathbf{g}_k\| \leq \varepsilon$, 算法终止; 否则, 进入下一步
3. 计算最优步长 $\alpha = \arg \min_{\alpha \geq 0} \{f(\mathbf{x}_k + \alpha \mathbf{d}_k)\}$, 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{d}_k$
4. 构造 \mathbf{d}_{k+1} 使其与 $\mathbf{d}_0, \dots, \mathbf{d}_k$ 关于矩阵 \mathbf{A} 共轭, 令 $k \leftarrow k + 1$, 返回步 2

定理 2.8 (二次终止性) 对严格凸二次函数 $\min f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$, 向量组 $\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{n-1}$ 关于 \mathbf{A} 共轭。共轭方向法产生点列 $\{\mathbf{x}_k\}$ 。则对任意 $0 \leq k \leq n-1$, \mathbf{x}_{k+1} 是目标函数在仿射集 $\mathbf{x}_0 + \text{span}[\mathbf{d}_0, \dots, \mathbf{d}_k]$ 上的最小值点, 算法至多 n 步迭代后终止。

例 2.4 利用共轭梯度法求 $\mathbf{A} \mathbf{x} = \mathbf{b}$ 或者说, 利用共轭梯度法求 $\min x_1^2 + \frac{1}{2} x_2^2 + \frac{1}{2} x_3^2$

★★★★★

其中,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

解:

取初始点 $\mathbf{x}_0 = (1, 1, 1)^T$, 迭代过程:

$$1. \mathbf{x}_0 = (1, 1, 1)^T, \mathbf{g}_0 = \mathbf{A}\mathbf{x}_0 - \mathbf{0} = (2, 1, 1)^T, \beta_{-1} = 0, \mathbf{d}_0 = -\mathbf{g}_0.$$

$$\begin{aligned} \alpha &= \arg \min f(\mathbf{x}_0 + \alpha \mathbf{d}_0) \\ &= (1 - 2\alpha)^2 + \frac{1}{2}(1 - \alpha)^2 + \frac{1}{2}(1 - \alpha)^2 = \frac{3}{5} \end{aligned}$$

$$2. \mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{d}_0 = \frac{1}{5}(-1, 2, 2)^T, \mathbf{g}_1 = \mathbf{A}\mathbf{x}_1 - \mathbf{0} = \frac{1}{5}(-1, 2, 2)^T, \beta_0 = \frac{\mathbf{g}_1^T \mathbf{g}_1}{\mathbf{g}_0^T \mathbf{g}_0} = \frac{2}{25}$$

$$\mathbf{d}_1 = -\mathbf{g}_1 + \beta_0 \mathbf{g}_0 = -\frac{6}{25}(1, -2, -2)^T$$

$$\begin{aligned} \alpha &= \arg \min f(\mathbf{x}_1 + \alpha \mathbf{d}_1) \\ &= (1 - 2\alpha)^2 + \frac{1}{2}(1 - \alpha)^2 + \frac{1}{2}(1 - \alpha)^2 = \frac{3}{5} \end{aligned}$$

$$3. \mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = \mathbf{0}, \|\mathbf{g}_2\| = 0, \text{ 终止。}$$

k	\mathbf{x}_k	\mathbf{g}_k	β_{k-1}	\mathbf{d}_k	α_k
0	$(1, 1, 1)^T$	$(2, 1, 1)^T$	0	$-(2, 1, 1)^T$	$\frac{3}{5}$
1	$\frac{1}{5}(-1, 2, 2)^T$	$\frac{1}{5}(-2, 2, 2)^T$	$\frac{2}{25}$	$-\frac{6}{25}(1, -2, -2)^T$	$\frac{5}{6}$
2	$(0, 0, 0)^T$	$(0, 0, 0)^T$			

定理 2.9 (收敛速度) 对严格凸二次函数 $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} - \mathbf{b}^T \mathbf{x}$. 若系数矩阵 \mathbf{A} 有 r 个相异特征根, 则最优步长规则下的共轭梯度法至多 r 步迭代后终止。

2.5 拟牛顿算法

注 基本思想: 利用已产生迭代点的梯度信息构造目标函数 Hesse 矩阵的一个近似, 进而建立牛顿方向的一个近似, 再通过线搜索产生新的迭代点。

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \underbrace{\mathbf{G}_k^{-1} \mathbf{g}_k}_{\text{牛顿方法}} \longrightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - \underbrace{\alpha \mathbf{B}_k^{-1} \mathbf{g}_k}_{\text{拟牛顿方法}}$$

牛顿方法

拟牛顿方法

考察目标函数梯度 $\mathbf{g}(\mathbf{x})$ 在 \mathbf{x}_{k+1} 点的线性近似

$$\mathbf{g}(\mathbf{x}_k) \approx \mathbf{g}_{k+1} + \mathbf{G}_{k+1}(\mathbf{x}_k - \mathbf{x}_{k+1})$$

\implies

$$\mathbf{g}_{k+1} - \mathbf{g}_{x_k} \approx \mathbf{G}_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

$$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$

\implies

$$\mathbf{y}_k \approx \mathbf{G}_{k+1} \mathbf{s}_k$$

$$\begin{aligned} \mathbf{H}_{k+1} &= \mathbf{B}_{k+1}^{-1} \\ \implies \end{aligned}$$

$$\mathbf{H}_{k+1} \mathbf{y}_k \approx \mathbf{s}_k$$

得到重要结论

$$\mathbf{H}_{k+1} \mathbf{y}_k \approx \mathbf{s}_k$$

2.5.1 对称秩-1 校正公式

注 对称秩-1 校正公式:

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{v}_k \mathbf{v}_k^T$$

使用 $\mathbf{H}_{k+1} \mathbf{y}_k \approx \mathbf{s}_k$ 两边同时乘以 \mathbf{y}_k 得到

$$\mathbf{H}_k \mathbf{y}_k + \mathbf{v}_k \mathbf{v}_k^T \mathbf{y}_k = \mathbf{s}_k$$

得到

$$\mathbf{v}_k = \frac{\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k}{\mathbf{v}_k^T \mathbf{y}_k}$$

所以对于 $\mathbf{v}_k \mathbf{v}_k^T$ 有

$$\begin{aligned} \mathbf{v}_k \mathbf{v}_k^T &= \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{(\mathbf{v}_k^T \mathbf{y}_k)(\mathbf{y}_k^T \mathbf{v}_k)} \\ &= \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{\mathbf{y}_k^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{y}_k} \\ &= \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T \mathbf{y}_k} \end{aligned}$$

故而,

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T \mathbf{y}_k} \quad (1)$$

1. 取 $\mathbf{x}_0 \in \mathbb{R}^n$, 矩阵 \mathbf{H}_0 (一般取 $\mathbf{H}_0 = \mathbf{I}$) 和计算精度 ε , 令 $k = 0$
2. 计算搜索方向 $\mathbf{d}_k = -\mathbf{H}_k \mathbf{g}_k$. 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$. 其中, 步长由线搜索产生. 若 $\|\mathbf{g}_{k+1}\| \leq \varepsilon$, 算法停止; 否则转下一步
3. 利用

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T \mathbf{y}_k}$$

计算 \mathbf{H}_{k+1} , 令 $k = k + 1$, 转步 2

2.5.2 DFP 校正公式

注 DFP 校正公式:

基于对称秩-1 校正公式中的校正项

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T}{(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^T \mathbf{y}_k}$$

对 \mathbf{H}_k 做如下对称秩-2 校正

$$\mathbf{H}_{k+1} = \mathbf{H}_k + a \mathbf{s}_k \mathbf{s}_k^T + b \mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^T \mathbf{H}_k$$

利用 $\mathbf{H}_{k+1} \mathbf{y}_k = \mathbf{s}_k$ 得到

$$\mathbf{H}_k \mathbf{y}_k + a \mathbf{s}_k \mathbf{s}_k^T \mathbf{y}_k + b \mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^T \mathbf{H}_k \mathbf{y}_k = \mathbf{s}_k$$

两边相等，故而

$$a = \frac{1}{\mathbf{s}_k^T \mathbf{y}_k}, \quad b = -\frac{1}{\mathbf{y}_k^T \mathbf{H}_k \mathbf{y}_k}$$

故而得到 DFP 校正公式

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} - \frac{\mathbf{H}_k \mathbf{y}_k \mathbf{y}_k^T \mathbf{H}_k}{\mathbf{y}_k^T \mathbf{H}_k \mathbf{y}_k}$$

例 2.5 用 DFP 方法求解优化问题★★★★★

$$\min f(\mathbf{x}) = 2x_1^2 + x_2^2 - 4x_1 + 2$$

解：容易计算 $\mathbf{x}^* = (1, 0)^T$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 4x_1 - 4 \\ 2x_2 \end{pmatrix}$$

1. 取初始值 $\mathbf{x}_0 = (2, 1)^T, \mathbf{H}_0 = \mathbf{I}$ ，则 $\mathbf{g}_0 = (4, 2)^T$ ，沿着 $\mathbf{d}_0 = -\mathbf{H}_0 \mathbf{g}_0 = (-4, -2)^T$ 精确线搜索得到 $\alpha = \frac{5}{18}$

新迭代点 $\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathbf{d}_0 = (\frac{8}{9}, \frac{4}{9})^T$ 进入下一次迭代

2. 计算 $\mathbf{x}_1 = \mathbf{x}_0 + \alpha \mathbf{d}_0 = (\frac{8}{9}, \frac{4}{9})^T$ ，梯度 $\mathbf{g}_0 = (-\frac{4}{9}, \frac{8}{9})^T$ ，进而计算

$$\mathbf{s}_0 = \mathbf{x}_1 - \mathbf{x}_0 = \alpha_0 \mathbf{d}_0 = (-\frac{10}{9}, -\frac{5}{9})^T$$

$$\mathbf{y}_0 = \mathbf{g}_1 - \mathbf{g}_0 = (-\frac{40}{9}, -\frac{40}{9})^T$$

得校正矩阵

$$\mathbf{H}_1 = \mathbf{H}_0 + \frac{\mathbf{s}_0 \mathbf{s}_0^T}{\mathbf{s}_0^T \mathbf{y}_0} - \frac{\mathbf{H}_0 \mathbf{y}_0 \mathbf{y}_0^T \mathbf{H}_0}{\mathbf{y}_0^T \mathbf{H}_0 \mathbf{y}_0} = \frac{1}{360} \begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix}$$

沿着 $\mathbf{d}_1 = -\mathbf{H}_1 \mathbf{g}_1 = \frac{12}{51}(1, -4)^T$ 精确线搜索得到 $\alpha_1 = \frac{17}{36}$

新迭代点 $\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{d}_1 = (1, 0)^T = \mathbf{x}^*$ ，算法终止。

2.6 信赖域方法

优化模型 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$

信赖域方法： $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k$ 使得 $f(\mathbf{x}_{k+1}) < f(\mathbf{x})$

算法思想：构造目标函数的二阶近似，将其在当前点邻域内的最小值点作为新的迭代点。

构造 $f(\mathbf{x}_k + \mathbf{d})$ 的二阶近似

$$m_k(\mathbf{d}) = f(\mathbf{x}_k) + \mathbf{g}_k^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T \mathbf{B}_k \mathbf{d}$$

其中 $\mathbf{B}_k = \nabla^2 f(\mathbf{x}_k)$ 或者 $\mathbf{B}_k \approx \nabla^2 f(\mathbf{x}_k)$ 。

注 信赖域子问题

$$\min\{m_k(\mathbf{d}) \mid \mathbf{d} \in \mathbb{R}^n, \|\mathbf{d}\| \leq \Delta_k\}.$$

其中, $\Delta_k > 0$ 为信赖域半径。

设 \mathbf{d}_k 为其最优解。则目标函数与二次模型的近似度为

$$f(\mathbf{x}_k + \mathbf{d}_k) - m_k(\mathbf{d}_k) = O(\|\mathbf{d}_k\|^2)$$

若 $\mathbf{B}_k = \nabla^2 f(\mathbf{x}_k)$, 则

$$f(\mathbf{x}_k + \mathbf{d}_k) - m_k(\mathbf{d}_k) = O(\|\mathbf{d}_k\|^2)$$

注 调整依据: $m_k(\mathbf{d}_k)$ 与 $f(\mathbf{x}_k + \mathbf{d}_k)$ 的近似度

- 预下降量: $\text{Pred}_k = m_k(\mathbf{0}) + m_k(\mathbf{d}_k)$,
- 实下降量: $\text{Ared}_k = f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{d}_k)$
- 下降比: $r_k = \frac{\text{Ared}_k}{\text{Pred}_k}$

注 信赖域算法:

1. 取 $\hat{\Delta} > 0, \mathbf{x}_0 \in \mathbb{R}^n, \Delta_0^i \in (0, \hat{\Delta}], \eta \in [0, \frac{1}{4}), \varepsilon \geq 0$. 令 $k = 0$
2. 若 $\|g_k\| \leq \varepsilon$, 算法终止; 否则进入下一步
3. 解子问题 $\mathbf{d}_k = \arg \min\{m_k(\mathbf{d}) \mid \mathbf{d} \in \mathbb{R}^n, \|\mathbf{d}\| \leq \Delta_k\}$
计算 $r_k = \frac{\text{Ared}_k}{\text{Pred}_k}$
若 $r_k < \frac{1}{4}$. 令 $\Delta_{k+1} = \frac{1}{4}\Delta_k$;
若 $r_k > \frac{3}{4}$, 且 $\|\mathbf{d}_k\| = \Delta_k$, 令 $\Delta_{k+1} = \min\{2\Delta_k, \hat{\Delta}\}$;
否则, 令 $\Delta_{k+1} = \Delta_k$
4. 若 $r_k > \eta$, 令 $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{d}_k$
否则, 令 $\mathbf{x}_{k+1} = \mathbf{x}_k, k = k + 1$, 转 2

2.7 最小二乘问题

最小二乘问题数学模型

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^n (y_i - \phi(\mathbf{x}, t_i))^2$$

注 模型特点:

- $m \gg n$
- 目标函数为平方和形式, 函数值非负
- 目标函数选取值恰当时, 目标函数的最优值靠近零

设 $r_i(\mathbf{x}) = y_i - \phi(\mathbf{x}, t_i)$ 模型可以写为

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x})$$

其中 $\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), r_2(\mathbf{x}), \dots, r_m(\mathbf{x}))^T$ 。

2.7.1 线性最小二乘

数学模型

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x})$$

其中 $\mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$ 问题转换为

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

定理 2.10 线性最小二乘问题存在全局最优解。它有唯一最优解的充分必要条件是矩阵 \mathbf{A} 列满秩。

2.7.2 非线性最小二乘

数学模型

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) := \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2$$

注 梯度

$$\nabla f(\mathbf{x}) = \mathbf{g}(\mathbf{x}) = \nabla \left(\frac{1}{2} \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) \right) = \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) = \sum_{i=1}^m r_i(\mathbf{x}) \nabla r_i(\mathbf{x})$$

数值乘以向量求和

其中

$$\mathbf{J}(\mathbf{x}) = D_{\mathbf{x}} \mathbf{r}(\mathbf{x}) = \begin{pmatrix} \frac{\partial r_1(\mathbf{x})}{\partial x_1} & \frac{\partial r_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial r_2(\mathbf{x})}{\partial x_1} & \frac{\partial r_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m(\mathbf{x})}{\partial x_1} & \frac{\partial r_m(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial r_m(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

注 Hesse 阵

$$\begin{aligned} \nabla^2 f(\mathbf{x}) = \mathbf{G}(\mathbf{x}) &= \sum_{i=1}^m \nabla r_i(\mathbf{x}) \nabla^T r_i(\mathbf{x}) + \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}) \\ &= \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x}) \\ &\triangleq \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \mathbf{S}(\mathbf{x}). \end{aligned}$$

其中

$$\mathbf{S}(\mathbf{x}) = \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x})$$

注 牛顿算法

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{J}_k^T \mathbf{J}_k + \mathbf{S}_k)^{-1} \mathbf{J}_k^T \mathbf{r}(\mathbf{x}_k)$$

注 Gauss-Newton 方法: 当残量函数 $\mathbf{r}(x)$ 很小时, $\mathbf{S}(\mathbf{x}) = \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x})$ 很小

$$\mathbf{d}_k^{\text{GN}} = -(\mathbf{J}_k^T \mathbf{J}_k)^{-1} \mathbf{J}_k^T \mathbf{r}(\mathbf{x}_k)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k^{\text{GN}}$$

3 凸优化

3.1 凸分析基础

定义 3.1 (Lipschitz 连续)

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

定义 3.2 (局部 Lipschitz 连续)

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L_{x_0} \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in N(\mathbf{x}_0, \delta)$$

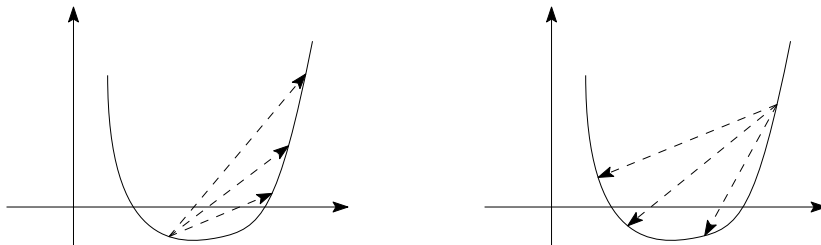
定义 3.3 (正常函数) $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 满足 $f(\mathbf{x}) > -\infty$, $\forall \mathbf{x} \in \mathbb{R}^n$ 且存在 $\bar{\mathbf{x}} \in \mathbb{R}^n$ 使得 $f(\bar{\mathbf{x}}) < \infty$. ($f(\mathbf{x})$ 有下确界, 且不恒取无穷大)

定理 3.1 正常凸函数在其定义域内部局部 Lipschitz 连续.

定理 3.2 设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为正常凸函数, 则对任意 $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n$,

$$\frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \text{ 关于 } t \text{ 单调递增}$$

或者说: 对凸函数, 沿任一方向, 若函数值增长则逐渐加快, 否则逐渐变慢. 方向导数 $\lim_{t \rightarrow 0+} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t}$ 存在



定义 3.4 (次梯度) 非光滑凸函数: 存在 $\xi \in \mathbb{R}^n$ 使得

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \xi, \mathbf{y} - \mathbf{x} \rangle, \quad \forall \mathbf{y} \in \mathbb{R}^n,$$

ξ 称为凸函数 $f(\mathbf{x})$ 在 \mathbf{x} 点的次梯度.

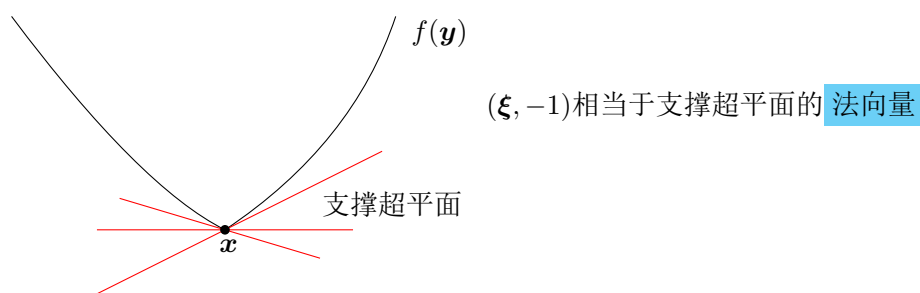
定义 3.5 (次微分) 满足

$$\partial f(x) = \{\xi \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle \xi, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

称

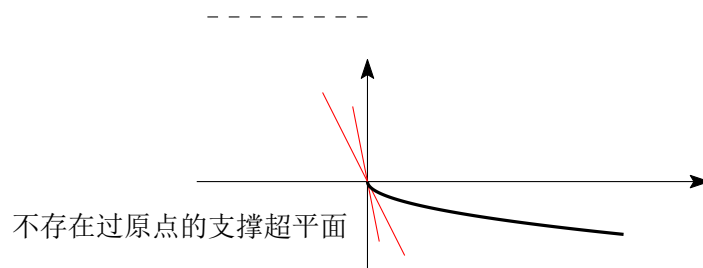
$$h(y) \triangleq f(x) + \langle \xi, y - x \rangle$$

为函数 $f(x)$ 在 x 点的支撑超平面。



注 次梯度可能存在, 可能不存在, 也可能存在不唯一。

例 3.1 凸函数 $f(x) = \begin{cases} -\sqrt{x}, & x \geq 0 \\ \infty, & \text{else} \end{cases}$ 在零点没有次微分



定理 3.3 对凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 有

$$\xi \in \partial f(x) \iff f'(x; d) \geq \xi^T d, \quad \forall d \in \mathbb{R}^n.$$

定理 3.4 $x^* \in \mathbb{R}^n$ 为凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的最优值点当且仅当 $0 \in \partial f(x)^*$

证明.

$$\begin{aligned} 0 &\in \partial f(x^*) \\ f(x) &\geq f(x^*) + \langle 0, x - x^* \rangle, \quad \forall x \in \mathbb{R}^n \\ f(x) &\geq f(x^*), \quad \forall x \in \mathbb{R}^n \end{aligned}$$

证毕!

注 次梯度是凸函数梯度的推广. 如果凸函数连续可微, 则 梯度 = 次梯度.

定理 3.5 若凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 \mathbf{x} 点可微, 则

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$$

定理 3.6 凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 和其定义域内任一内点 \mathbf{x} , $\partial f(\mathbf{x})$ 非空有界.

注 凸函数一定次可微! 可行域边界点除外。

定理 3.7 若 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 在其定义域内部次可微, 则它为凸函数。

注 次梯度的推广:

- 非凸 Lipschitz 连续

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

- 非凸局部 Lipschitz 连续

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L_{x_0} \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in N(\mathbf{x}_0, \delta)$$

3.1.1 临近点算子

定义 3.6 (临近点算子) 对函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 定义其在 $\mathbf{x} \in \mathbb{R}^n$ 点的临近点算子

$$\text{prox}_f(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathbb{R}^n} \left\{ f(\mathbf{y}) + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 \right\}$$

注 临近点算子使子问题的性能得到加强, 如非凸函数变为凸函数, 凸函数变为强凸函数, 便于在当前迭代点临近找到比该点更好的点.

定义 3.7 (临近点算子的扩展)

$$\begin{aligned} \text{prox}_{\frac{1}{L}f}(\mathbf{x}) &= \arg \min_{\mathbf{y} \in \mathbb{R}^n} \left\{ f(\mathbf{y}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2 \right\} \\ &= \arg \min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \frac{1}{L} f(\mathbf{y}) + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 \right\} \end{aligned}$$

对于 $f(\mathbf{x}) + g(\mathbf{x})$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续

$$\begin{aligned} &\arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_0\|^2 + g(\mathbf{x}) \right\} \\ &= \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{1}{L} g(\mathbf{x}) + \frac{1}{2} \left\| \mathbf{x} - \left(\mathbf{x}_0 - \frac{1}{L} \nabla f(\mathbf{x}_0) \right) \right\|^2 \right\} \\ &\triangleq \text{prox}_{\frac{1}{L}g} \left(\mathbf{x}_0 - \frac{1}{L} \nabla f(\mathbf{x}_0) \right) \end{aligned}$$

例 3.2 求 $f_2(x) = \begin{cases} 0, & x \neq 0 \\ -\lambda, & x = 0 \end{cases}$ 的临近点算子 ($\lambda > 0$)。★★★★★

$$\begin{aligned}
\text{prox}_{f_2}(x) &= \arg \min_{y \in \mathbb{R}} \left\{ f_2(y) + \frac{1}{2}(y-x)^2 \right\} \\
&= \arg \min_{y \in \mathbb{R}} \left\{ \min_{y=0} \left\{ -\lambda + \frac{x^2}{2} \right\}, \min_{y \neq 0} \left\{ \frac{1}{2}(y-x)^2 \right\} \right\} \\
&= \arg \min_{y \in \mathbb{R}} \left\{ -\lambda + \frac{x^2}{2}, 0 \right\} \\
&= \begin{cases} \{0\}, & |x| < \sqrt{2\lambda} \\ \{x\}, & |x| > \sqrt{2\lambda} \\ \{0, x\}, & |x| = \sqrt{2\lambda} \end{cases}
\end{aligned}$$

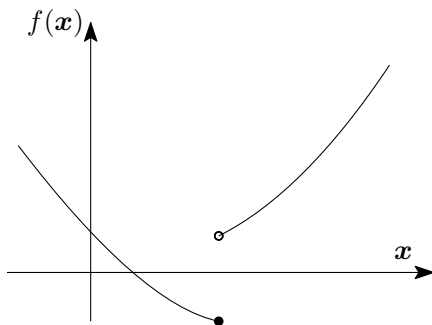
例 3.3 求 $f_3(x) = \begin{cases} 0, & x \neq 0 \\ \lambda, & x = 0 \end{cases}$ 的临近点算子 ($\lambda > 0$)。★★★★★

解:

$$\text{prox}_{f_3}(x) = \begin{cases} \{x\}, & x \neq 0 \\ \emptyset, & x = 0 \end{cases}$$

定义 3.8 (下半连续)

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0)$$



定理 3.8 对下半连续的凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 其临近点算子 $\text{prox}_f(\cdot)$ 为单一映射。

定理 3.9 设凸函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 下半连续. 则对任意的 $x, y \in \mathbb{R}^n$

$$y = \text{prox}_f(x)$$

$$x - y \in \partial f(y)$$

$$\langle x - y, z - y \rangle \leq f(z) - f(y), \forall z \in \mathbb{R}^n$$

证明. 由临近点算子的定义, 对任意的 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbf{y} = \text{prox}_f(\mathbf{x}) \iff \mathbf{y} = \arg \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ f(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|^2 \right\}.$$

后面等价于 $\mathbf{0} \in \partial f(\mathbf{y}) + \mathbf{y} - \mathbf{x}$. 由次梯度的定义 $\mathbf{x} - \mathbf{y} \in \partial f(\mathbf{y})$ 等价于,

$$\langle \mathbf{x} - \mathbf{y}, \mathbf{z} - \mathbf{y} \rangle \leq f(\mathbf{z}) - f(\mathbf{y}), \forall \mathbf{z} \in \mathbb{R}^n$$

证毕!

定理 3.10 设 f 为下半连续的凸函数. 则对任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\langle \mathbf{x} - \mathbf{y}, \text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y}) \rangle \geq \|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\|^2$$

$$\|\text{prox}_f(\mathbf{x}) - \text{prox}_f(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

3.1.2 和式优化最优性条件

为充分利用函数的解析性质, 将目标函数中的项根据其解析性质分开, 可得如下和式优化问题

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

用梯度和次梯度建立优化问题的最优性条件

定理 3.11 $\mathbf{x}^* \in \mathbb{R}^n$ 为优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) + g(\mathbf{x})$ 的最优解, 则 $-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*)$

3.2 线性化临近点方法

用梯度和函数值建立优化问题的最优性条件

$$\min f(\mathbf{x}) + g(\mathbf{x})$$

其中 $\Omega \subset \mathbb{R}^n$, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微的凸函数, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 凸函数

定理 3.12 为上述优化问题的最优解当且仅当

$$g(\mathbf{x}) - g(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

注 临近点算法

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}$$

注 算法收敛性

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) + \frac{L}{2} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}$$

由

$$\begin{aligned} \mathbf{x}^* &= \arg \min f(\mathbf{x}) + g(\mathbf{x}) \\ g(\mathbf{x}) - g(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^\top (\mathbf{x} - \mathbf{x}^*) &\geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

得

$$f(\mathbf{x}) - f(\mathbf{x}_{k+1}) + L(\mathbf{x}_{k+1} - \mathbf{x}_k)^\top (\mathbf{x} - \mathbf{x}_{k+1}) \geq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

令 $\mathbf{x} = \mathbf{x}_k$

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq L \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2.$$

$$f(\mathbf{x}_0) - \lim_{k \rightarrow \infty} f(\mathbf{x}_k) \geq L \sum_{k=0}^{\infty} \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 \longrightarrow \lim_{k \rightarrow \infty} \|\mathbf{x}_k - \mathbf{x}_{k+1}\| = 0.$$

算法产生的函数值序列 $\{f(\mathbf{x}_k)\}$ 单调不减. 如果目标函数 $f(\mathbf{x})$ 在 \mathbb{R}^n 上有下界, 则数列 $\{f(\mathbf{x}_k)\}$ 收敛.

定理 3.13 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 下半连续的凸函数, 有下界; 临近点算法 $\mathbf{x}_{k+1} = \text{prox}_{\frac{1}{L}f}(\mathbf{x}_k)$ 产生迭代点列的任一聚点为凸优化问题 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ 的最优值点.

3.3 交替极小化方法

$$\min_{\mathbf{x}_1 \in \mathbb{R}^{n_1}, \mathbf{x}_2 \in \mathbb{R}^{n_2}, \dots, \mathbf{x}_s \in \mathbb{R}^{n_s}} \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s)$$

以循环方式交替对 s 个块变量求最小. 对迭代点 $\mathbf{x}^k = (\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}_s^k)$, 分别基于块变量依次对 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s$ 求极小, 得到新的迭代点

$$\begin{aligned} \mathbf{x}^{k,1} &= (\mathbf{x}_1^{k+1}, \mathbf{x}_2^k, \dots, \mathbf{x}_s^k), \\ \mathbf{x}^{k,2} &= (\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \mathbf{x}_3^k, \dots, \mathbf{x}_s^k), \\ \mathbf{x}^{k,i} &= (\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \dots, \mathbf{x}_i^{k+1}, \mathbf{x}_{i+1}^k, \dots, \mathbf{x}_s^k), \\ \mathbf{x}^{k,s} &= \mathbf{x}^{k+1} = (\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \dots, \mathbf{x}_s^{k+1}). \end{aligned}$$

定理 3.14 设目标函数 $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, 目标函数水平集有界且对任意的 $\mathbf{x} \in \mathbb{R}^n$ 和 $i \in \{1, 2, \dots, s\}$

$$\min_{\mathbf{y} \in \mathbb{R}^{n_i}} \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1}, \mathbf{y}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_s)$$

有唯一最优解. 则算法产生迭代点列的任一聚点为优化问题的稳定点.

例 3.4 反例: 连续不可微. $\min \Psi(x_1, x_2) = |3x_1 + 4x_2| + |x_1 - 2x_2|$ ★★★★★

解 连续凸函数, 水平集有界, 且对任一分量有唯一最优解. 对任意 $\alpha > 0$,

$$\Psi(-4\alpha, t) = |4t - 12\alpha| + |2t + 4\alpha| = \begin{cases} -6t + 8\alpha, & t < -2\alpha, \\ -2t + 16\alpha, & -2\alpha \leq t \leq 3\alpha \\ 6t - 8\alpha, & t > 3\alpha, \end{cases}$$

$$\Psi(t, 3\alpha) = |3t + 12\alpha| + |t - 6\alpha| = \begin{cases} -4t - 6\alpha, & t < -4\alpha, \\ 2t + 18\alpha, & -4\alpha \leq t \leq 6\alpha, \\ 4t + 6\alpha, & t > 6\alpha, \end{cases}$$

对任意 $\alpha \leq 0$

$$\begin{aligned} -4\alpha &= \arg \min_{x_1 \in \mathbb{R}} \Psi(x_1, 3\alpha), \\ 3\alpha &= \arg \min_{x_2 \in \mathbb{R}} \Psi(-4\alpha, x_2) \end{aligned}$$

交替极小化方法

$$\min \Psi(x_1, x_2) = |3x_1 + 4x_2| + |x_1 - 2x_2|$$

若 x_1 非零, 则在首次迭代后, 算法滞留在 $(-4\alpha, 3\alpha)$ 点 (聚点)。

$\mathbf{x} = \mathbf{0}$ 为函数的唯一最小值点, $(-4\alpha, 3\alpha)$ 既不是该函数的最小值点, 也不是其稳定点。

定义 3.9 (坐标轮换最小值点) 若 $\mathbf{x}^* \in \mathbb{R}^n$ 满足

$$\Psi(\mathbf{x}^*) \leq \Psi(\mathbf{x}_1^*, \dots, \mathbf{x}_{i-1}^*, \mathbf{x}_i, \mathbf{x}_{i+1}^*, \dots, \mathbf{x}_s^*),$$

$\forall i = 1, 2, \dots, s, \mathbf{x}_i \in \mathbb{R}^{n_i}$, 则称 \mathbf{x}^* 为函数 $\Psi(\mathbf{x})$ 的坐标轮换最小值点。

定理 3.15 设优化问题

$$\min_{\mathbf{x}_1 \in \mathbb{R}^{n_1}, \mathbf{x}_2 \in \mathbb{R}^{n_2}, \dots, \mathbf{x}_s \in \mathbb{R}^{n_s}} \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s)$$

目标函数下半连续, 水平集有界, 子问题

$$\min_{\mathbf{y} \in \mathbb{R}^{n_i}} \Psi(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{y}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_s)$$

有唯一最优解。则迭代点列的任一聚点为坐标轮换最小值点。

注 连续不可微函数的坐标轮换极小值点未必是其最小值点或稳定点. 那么, 在什么情况下是呢?

定理 3.16 对优化问题

$$\min_{\mathbf{x} \in \mathbb{R}^n} \Psi(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$$

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, $g(\mathbf{x}) = \sum_{i=1}^s g_i(\mathbf{x}_i), g_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ 连续凸函数。

坐标轮换极小值 \iff 稳定点

对分块和式优化问题，若子问题最优解不唯一，算法收敛性不能保证。

例 3.5 反例：子问题最优解不唯一★★★★★

解

$$f(x, y, z) = -xy - yz - zx + [x - 1]_+^2 + [-x - 1]_+^2 + [y - 1]_+^2 + [-y - 1]_+^2 + [z - 1]_+^2 + [-z - 1]_+^2.$$

依次两两固定 y, z, x (注: $[x]_+ = \frac{x+|x|}{2}$)

$$\begin{aligned} f(x; y, z) &= -x(y+z) + \left(\frac{(x-1) + |x-1|}{2} \right)^2 + \left(\frac{-(x+1) + |x+1|}{2} \right)^2 + a \\ &= \begin{cases} -x(y+z) + (x+1)^2 + a & x < -1 \\ -x(y+z) + 0 + a & -1 < x < 1 \\ -x(y+z) + (x-1)^2 + a & x > 1 \end{cases} \\ f'_x(x; y, z) &= \begin{cases} 2(x+1) - (y+z) & x < -1 \\ -(y+z) & -1 < x < 1 \\ 2(x-1) - (y+z) & x > 1 \end{cases} \end{aligned}$$

根据 $f'_x(x; y, z) = 0$, 得到

$x < -1$ 时, $x = -1 + \frac{y+z}{2}$, ($y+z < 0$ 时取到)

$x > 1$ 时, $x = 1 + \frac{y+z}{2}$, ($y+z > 0$ 时取到) 故而, 有

$$\begin{aligned} \arg \min_x f(x, y, z) &= \begin{cases} \operatorname{sgn}(y+z)(1 + \frac{1}{2}|y+z|), & y+z \neq 0, \\ [-1, 1], & y+z = 0. \end{cases} \\ \arg \min_y f(x, y, z) &= \begin{cases} \operatorname{sgn}(x+z)(1 + \frac{1}{2}|x+z|), & x+z \neq 0, \\ [-1, 1], & x+z = 0, \end{cases} \\ \arg \min_z f(x, y, z) &= \begin{cases} \operatorname{sgn}(x+y)(1 + \frac{1}{2}|x+y|), & x+y \neq 0, \\ [-1, 1], & x+y = 0. \end{cases} \end{aligned}$$

3.3.1 和式凸优化交替极小化方法

$$\min_{\mathbf{x} \in \mathbb{R}^n} \Psi(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$$

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微的凸函数

$g(\mathbf{x}) = \sum_{i=1}^s g_i(\mathbf{x}_i)$, $g_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ 连续凸函数

对该优化问题，无需子问题最优解唯一，就能建立算法的全局收敛性。

定理 3.17 对上述优化问题, 若目标函数水平集有界, 则交替极小化方法产生迭代点列的任一聚点为问题的最优解.

3.3.2 二分块交替极小化方法

$$\min_{\mathbf{x}_1 \in \mathbb{R}^{n_1}, \mathbf{x}_2 \in \mathbb{R}^{n_2}} \{\Psi(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1, \mathbf{x}_2) + g_1(\mathbf{x}_1) + g_2(\mathbf{x}_2)\}$$

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微的凸函数

$\nabla_i f(x)$ Lipschitz 连续, 常数为 $L_i, i = 1, 2$

$g_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ 下半连续的凸函数

注 二分块交替极小化算法

- 初始步: 取 $\mathbf{x}_1^0 \in \mathbb{R}^{n_1}$, 计算

$$\mathbf{x}_2^0 = \arg \min_{\mathbf{x}_2} f(\mathbf{x}_1^0, \mathbf{x}_2) + g_2(\mathbf{x}_2)$$

- 迭代步: 依次计算

$$\begin{aligned} \mathbf{x}_1^{k+1} &= \arg \min_{\mathbf{x}_1} f(\mathbf{x}_1, \mathbf{x}_2^k) + g_1(\mathbf{x}_1), \\ \mathbf{x}_2^{k+1} &= \arg \min_{\mathbf{x}_2} f(\mathbf{x}_1^{k+1}, \mathbf{x}_2) + g_2(\mathbf{x}_2). \end{aligned}$$

3.4 凸分析基础 II

3.4.1 凸集与凸集分离定理

定义 3.10 (锥) $\mathcal{K} \subset \mathbb{R}^n$ 对任意的 $\mathbf{x} \in \mathcal{K}$ 和 $\lambda \geq 0$, 都有 $\lambda \mathbf{x} \in \mathcal{K}$ 。

并不是所有的锥都是凸的

定义 3.11 (极锥) 与原锥对立的锥。

$$\mathcal{K}^\circ = \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{y} \rangle \leq 0, \forall \mathbf{x} \in \mathcal{K}\}.$$

子空间的极锥为其正交补空间!

定义 3.12 (闭凸集的切锥) Ω 为 \mathbb{R}^n 中的非空闭凸集, $\mathbf{x} \in \Omega$ 。

$$\mathcal{T}_\Omega(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \text{存在 } \mathbf{d}_k \rightarrow \mathbf{d}, t_k \downarrow 0, \text{ 使得 } \mathbf{x} + t_k \mathbf{d}_k \in \Omega\}$$

定义为凸集 Ω 在该点的切锥。

切锥 $\mathcal{T}_\Omega(\mathbf{x})$ 的极锥 $\mathcal{T}_\Omega^\circ(\mathbf{x})$ 称为集合 Ω 在 \mathbf{x} 点的法锥或正则锥, 记为 $\mathcal{N}_\Omega(\mathbf{x})$

定义 3.13 (有限生成锥) 锥中任一元素都可以表示有限个元素的非负组合, 即

$$\mathcal{K} = \left\{ \sum_{i=1}^m \mu_i \mathbf{b}_i \mid \mu_i \geq 0, i = 1, 2, \dots, m \right\},$$

$\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ 称为生成元。

有限生成锥是凸锥, 它等价于多面锥。

定义 3.14 (多面锥) 可以表示成如下形式的锥

$$\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{0}\}$$

定义 3.15 (凸组合) 设 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n, \lambda_i \geq 0, i = 1, 2, \dots, m$ 满足 $\lambda_1 + \lambda_2 + \dots + \lambda_m = 1$, 称

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

为 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 的一个凸组合。

定义 3.16 (凸包) $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 的所有凸组合构成的集合称为由其生成的凸包。

定义 3.17 (多面胞) 含有限个顶点的闭凸集, 又称多面胞。

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$$

有界的凸多面体就是多面胞

定义 3.18 (回收方向) \mathbf{d} 满足

$$\forall \mathbf{x} \in \mathcal{S}, \alpha \geq 0 \implies \mathbf{x} + \alpha \mathbf{d} \in \mathcal{S}$$

定理 3.18 若多面体 $\mathcal{S} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$ 无界, 则存在多面胞 \mathcal{P} 和多面锥 \mathcal{K} 使得

$$\mathcal{S} = \mathcal{P} + \mathcal{K}.$$

$$\mathcal{P} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$$

$$\mathcal{K} = \{\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{0}\}$$

定义 3.19 (仿射集) $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m \in \mathbb{R}^n, \lambda_i \in \mathbb{R}, i = 1, 2, \dots, m$ 满足 $\lambda_1 + \lambda_2 + \dots + \lambda_m = 1$, 称

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

为 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 的一个仿射组合。

定义 3.20 (内点) 存在该点的邻域 $N(\mathbf{x}, \delta)$, 使得 $N(\mathbf{x}, \delta) \subseteq \mathcal{S}$

定义 3.21 (相对内点) 存在该点的邻域 $N(\mathbf{x}, \delta)$, 使得 $N(\mathbf{x}, \delta) \cap \text{Aff}(\mathcal{S}) \subseteq \mathcal{S}$ 。即该点在仿射子空间 $\text{Aff}(\mathcal{S})$ 中是集合的内点

定理 3.19 (凸集分离定理) 设 $S \subset \mathbb{R}^n$ 是非空闭凸集, $y \notin S$ 。则存在非零向量 p 和常数 α 使得

$$p^T y > \alpha, p^T x \leq \alpha, \forall x \in S$$

直观意义: 若点 y 不属于闭凸集 S , 则存在超平面

$$H = \{x \mid p^T x = \alpha\}$$

将该点与闭凸集严格分离. 这种分离不是唯一的!

定理 3.20 (Farkas 引理) 向量 b 可以表示为向量组 a_1, a_2, \dots, a_m 的非负线性组合的充要条件是对任意满足 $a_i^T \geq 0$ 的 x , 都有 $b^T x \geq 0$

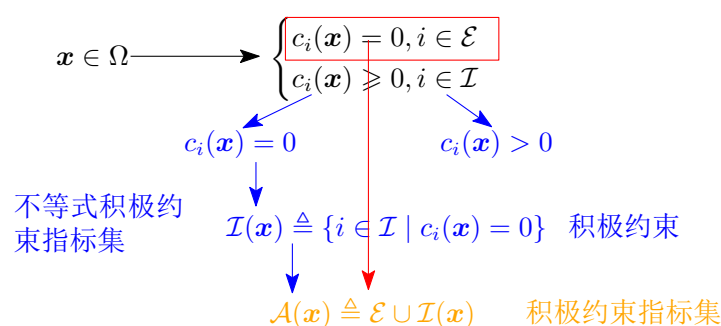
定理 3.21 (Farkas 引理') 对于 $A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$ 下述两系统恰有一个解

1. $A^T x \geq 0, b^T x < 0$;
2. $Ay = b, y \geq 0$.

4 约束优化最优性条件

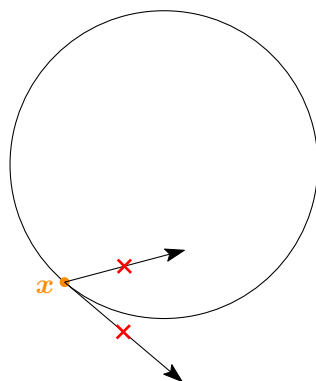
$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega \end{aligned}$$

$$\Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\},$$



定义 4.1 (可行方向) 设 $x \in \Omega$ 对 $d \in \mathbb{R}^n$ 若存在 $\delta > 0$, 使对任意的 $\alpha \in [0, \delta]$ 都有 $x + \alpha d \in \Omega$ 则称 d 为约束优化问题在 x 点的可行方向.

注 有的可行域没有可行方向, 如 $\Omega = \{x \in \mathbb{R}^n \mid x^T x = 1\}$



推论 若 d 为约束优化问题在 $x \in \Omega$ 点的可行方向, 则 (必要条件)

$$\nabla c_i(x)^T d = 0, \quad \forall i \in \mathcal{E}; \quad \nabla c_i(x)^T d \geq 0, \quad \forall i \in \mathcal{I}(x).$$

第一个条件是确保垂直; 第二个条件是确保边界点的下一步在可行域内

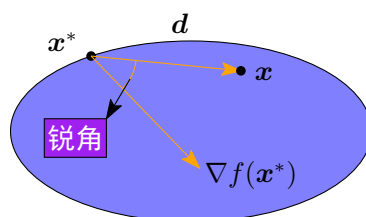
特别提醒: 上述条件不是充分的, 除非约束函数是线性的。

定理 4.1 (可行下降方向) 若 d 为 $x \in \Omega$ 点的可行方向, 同时还是目标函数在该点的下降方向, 则称其为目标函数在该点的可行下降方向。

最优值点不存在可行下降方向

定理 4.2 若约束优化问题的可行域为非空闭凸集, 则最优解 x^* 为约束优化问题的稳定点, 即满足

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega.$$



4.1 等式优化最优化条件

例 4.1 设 $a_1, a_2, \dots, a_m \in \mathbb{R}^n$ 满足 $\sum_{i=1}^m a_i \neq 0$ 求解优化问题★★★★★

$$\min \left\{ \sum_{i=1}^m \|x - a_i\|^2 \mid x^T x = 1 \right\}$$

解 令

$$\begin{aligned} L(\mathbf{x}, \lambda) &= \|\mathbf{x} - \mathbf{a}_i\|^2 - \lambda(\mathbf{x}^T \mathbf{x} - 1) \\ &= \sum_{i=1}^m \sum_{j=1}^n (x_j - (\mathbf{a}_i)_j)^2 - \lambda \left(\sum_{j=1}^n x_j^2 - 1 \right) \end{aligned}$$

对 x_k 求偏导, 得到

$$\frac{\partial L}{\partial x_k} = \sum_{i=1}^m 2(x_k - (\mathbf{a}_i)_k) - \lambda \cdot 2x_k$$

故而有

$$\nabla_{\mathbf{x}} L = \begin{pmatrix} \sum_{i=1}^m 2(x_1 - (\mathbf{a}_i)_1) - \lambda 2x_1 \\ \sum_{i=1}^m 2(x_2 - (\mathbf{a}_i)_2) - \lambda 2x_2 \\ \vdots \\ \sum_{i=1}^m 2(x_n - (\mathbf{a}_i)_n) - \lambda 2x_n \end{pmatrix}$$

所以, 有

$$\begin{cases} \sum_{i=1}^m 2(\mathbf{x} - \mathbf{a}_i) = 2\lambda \mathbf{x} \\ \mathbf{x}^T \mathbf{x} = 1 \end{cases}$$

设方程的解为 \mathbf{x}^* , 则 \mathbf{x}^* 与 $\sum_{i=1}^m \mathbf{a}_i$ 同向或者反向。

若 $\sum_{i=1}^m \mathbf{a}_i \neq \mathbf{0}$, 则

$$\mathbf{x}^* = \frac{\sum_{i=1}^m \mathbf{a}_i}{\|\sum_{i=1}^m \mathbf{a}_i\|}$$

相应地, 最优值为

$$\begin{aligned} f(\mathbf{x}^*) &= \sum_{i=1}^m \|\mathbf{x}^* - \mathbf{a}_i\|^2 \\ &= \sum_{i=1}^m \langle \mathbf{x}^* - \mathbf{a}_i, \mathbf{x}^* - \mathbf{a}_i \rangle \\ &= \sum_{i=1}^m \|\mathbf{x}^*\|^2 - 2 \left\langle \frac{\sum_{i=1}^m \mathbf{a}_i}{\|\sum_{i=1}^m \mathbf{a}_i\|}, \sum_{i=1}^m \mathbf{a}_i \right\rangle + \sum_{i=1}^m \|\mathbf{a}_i\|^2 \\ &= m - 2 \left\| \sum_{i=1}^m \mathbf{a}_i \right\| + \sum_{i=1}^m \|\mathbf{a}_i\|^2 \end{aligned}$$

4.1.1 不等式优化约束条件

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} + b_i = 0, i \in \mathcal{E} \\ & \mathbf{a}_i^T \mathbf{x} + b_i \geq 0, i \in \mathcal{I} \end{aligned}$$

定理 4.3 (完整 KKT 条件)

$$\begin{cases} \nabla f(\mathbf{x}^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(\mathbf{x}^*), \\ \lambda_i^* \geq 0, c_i(\mathbf{x}^*) \geq 0, \lambda_i^* c_i(\mathbf{x}^*) = 0, i \in \mathcal{I}, \\ c_i(\mathbf{x}^*) = 0, i \in \mathcal{E}. \end{cases}$$

4.2 鞍点与凸规划

定义 4.2 (函数的鞍点) 梯度零点, 但既不是最大值点、也不是最小值点

4.3 约束优化对偶

4.3.1 Lagrange 对偶

原问题:

$$\begin{aligned} \min f(\mathbf{x}) &\implies \min f(\mathbf{x}) \\ \text{s.t. } c_i(\mathbf{x}) = 0, i \in \mathcal{E} &\implies \mathbf{H}(\mathbf{x}) = \mathbf{0} \\ \text{s.t. } c_i(\mathbf{x}) \geq 0, i \in \mathcal{I} &\implies \mathbf{G}(\mathbf{x}) \geq \mathbf{0} \end{aligned}$$

Lagrange 函数

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\mathbf{u} \geq \mathbf{0}, \mathbf{v}} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \mathbf{G}(\mathbf{x})^T \mathbf{u} - \mathbf{H}(\mathbf{x})^T \mathbf{v}, \mathbf{x} \in \mathbb{R}^n, \mathbf{u} \in \mathbb{R}_+^{|\mathcal{I}|}, \mathbf{v} \in \mathbb{R}^{|\mathcal{E}|}$$

对偶问题:

$$\max_{\mathbf{u} \geq \mathbf{0}, \mathbf{v}} \min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \max_{\mathbf{u} \geq \mathbf{0}, \mathbf{v}} \theta(\mathbf{u}, \mathbf{v})$$

例 4.2 线性规划的对偶★★★★★

$$\begin{aligned} \min \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{Ax} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

对偶

$$\begin{aligned} \max_{\mathbf{u} \geq \mathbf{0}, \mathbf{v}} \min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) \\ L(\mathbf{x}; \mathbf{u}, \mathbf{v}) &= \mathbf{c}^T \mathbf{x} - \mathbf{u}^T \mathbf{x} - \mathbf{v}^T (\mathbf{Ax} - \mathbf{b}) \\ &= (\mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v})^T \mathbf{x} + \mathbf{v}^T \mathbf{b}, \quad \mathbf{u} \geq \mathbf{0}. \\ \min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}; \mathbf{u}, \mathbf{v}) &= \min_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v})^T \mathbf{x} + \mathbf{v}^T \mathbf{b} \\ &= \begin{cases} -\infty, & \mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v} \neq \mathbf{0}, \\ \mathbf{v}^T \mathbf{b}, & \mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v} = \mathbf{0}, \end{cases} \end{aligned}$$

$$\begin{aligned} \max \quad & (\mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v})^T \mathbf{x} + \mathbf{b}^T \mathbf{v} \\ \text{s.t.} \quad & \nabla_x L(\mathbf{x}; \mathbf{u}, \mathbf{v}) = \mathbf{c} - \mathbf{u} - \mathbf{A}^T \mathbf{v} = \mathbf{0} \\ & \mathbf{u} \geqslant \mathbf{0} \end{aligned}$$

化简得到

$$\begin{aligned} \max \quad & \mathbf{v}^T \mathbf{b} \\ \text{s.t.} \quad & \mathbf{A}^T \mathbf{v} \leqslant \mathbf{c} \end{aligned}$$

4.3.2 对偶定理

定理 4.4 (对偶定理) 设 \mathbf{x}_0 是原问题的可行解, $(\mathbf{u}_0, \mathbf{v}_0)$ 是对偶问题的可行解, 则 $f(\mathbf{x}_0) \geqslant \theta(\mathbf{u}_0, \mathbf{v}_0)$, 则

$$f(\mathbf{x}_0) \geqslant \theta(\mathbf{u}_0, \mathbf{v}_0)$$

有时, 原问题的最优解 \mathbf{x}^* , 对偶规划问题的最优解 $(\mathbf{u}^*, \mathbf{v}^*)$ 满足

$$f(\mathbf{x}_0) = \theta(\mathbf{u}_0, \mathbf{v}_0)$$

也可以表述为: 原规划问题的任一可行解对应的目标函数值都不小于对偶规划问题的任一可行解对应的目标函数值!

定义 4.3 (对偶间隙、完全对偶、强对偶) 关于对偶有以下定义:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\mathbf{u} \geqslant \mathbf{0}, \mathbf{v}} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) - \max_{\mathbf{u} \geqslant \mathbf{0}, \mathbf{v}} \min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \mathbf{u}, \mathbf{v})$$

- 对偶间隙: 原规划与对偶规划问题的最优值之间的差
- 完全对偶: 对偶间隙为零.
- 强对偶: 对偶间隙为零, 原问题和对偶问题都存在最优解

注 强对偶定理对哪些优化问题成立?

- 对凸规划问题, 若弱 Slater 约束规格成立, 且存在最优解, 则对偶规划也有最优解, 强对偶定理成立
- 对线性规划问题及线性约束的凸规划问题, 弱 Slater 约束规格自然成立, 故强对偶定理成立.
- 强对偶定理并非只对凸规划问题成立.

例 4.3 非凸优化 ★☆☆☆☆

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^T \mathbf{x} \leqslant 1 \end{aligned}$$

其中 \mathbf{G} 非半正定。

解 Lagrange 函数

$$\begin{aligned} L(\mathbf{x}, \lambda) &= \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} - \lambda(1 - \mathbf{x}^T \mathbf{x}) \\ &= \mathbf{x}^T \left(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \right) \mathbf{x} + \mathbf{g}^T \mathbf{x} + \lambda \end{aligned}$$

Lagrange 对偶

$$\max_{\lambda \geq 0} \theta(\lambda) = \max_{\lambda \geq 0} \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^T \left(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \right) \mathbf{x} + \mathbf{g}^T \mathbf{x} - \lambda$$

若 $(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I})$ 非半正定

$$\theta(\lambda) = \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^T \left(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \right) \mathbf{x} + \mathbf{g}^T \mathbf{x} - \lambda = -\infty$$

那么有 $(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I}) \succcurlyeq \mathbf{0}$

$$\max_{\substack{\lambda \geq 0 \\ \frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \succcurlyeq \mathbf{0}}} \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^T \left(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \right) \mathbf{x} + \mathbf{g}^T \mathbf{x} - \lambda$$

内层优化最优解，原问题可行解

$$\begin{aligned} (\mathbf{G} + 2\lambda \mathbf{I}) \mathbf{x} + \mathbf{g} &= \mathbf{0} \\ \mathbf{x}^T \mathbf{x} &= 1 \end{aligned}$$

得到 \mathbf{x}^* ，之后有

$$\begin{aligned} \max \quad & -(\mathbf{x}^*)^T \left(\frac{1}{2} \mathbf{G} + \lambda \mathbf{I} \right) \mathbf{x}^* - \lambda \\ \text{s.t.} \quad & \lambda \geq 0, (\frac{1}{2} \mathbf{G} + \lambda \mathbf{I}) \succcurlyeq \mathbf{0} \end{aligned}$$

解得

$$\lambda^* = -\lambda_{\min} \left(\frac{1}{2} \mathbf{G} \right) > 0$$

(\mathbf{x}^*, λ) 为鞍点，强对偶定理成立

$$\begin{aligned} \theta(\lambda^*) &= L(\mathbf{x}^*, \lambda^*) \\ &= \frac{1}{2} (\mathbf{x}^*)^T \mathbf{G} \mathbf{x}^* + \mathbf{g}^T \mathbf{x}^* - \lambda^* (1 - (\mathbf{x}^*)^T \mathbf{x}^*) \\ &= \frac{1}{2} (\mathbf{x}^*)^T \mathbf{G} \mathbf{x}^* + \mathbf{g}^T \mathbf{x}^* \\ &= f(\mathbf{x}^*) \end{aligned}$$

4.3.3 混合约束优化的对偶

$$\begin{aligned} \min \quad & f(\mathbf{x}) && \text{连续可微凸函数} \\ \text{s.t.} \quad & \mathbf{H}(\mathbf{x}) = \mathbf{0} && \text{连续可微凹函数} \\ & \mathbf{G}(\mathbf{x}) \geq \mathbf{0} && \text{约束条件} \\ & \mathbf{x} \in \Theta && \text{约束条件} \end{aligned}$$

例 4.4 求下述优化问题的对偶★★★★★

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 + x_2 - 4 \geq 0 \\ & x \geq 0 \end{aligned}$$

解 Lagrange 函数

$$L(x, \lambda) = x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 4), \quad x \geq 0, \quad \lambda \geq 0$$

Lagrange 对偶 $\max_{\lambda \geq 0} \min_{x \geq 0} L(\lambda, \mathbf{x})$

$$\begin{aligned} \theta(\lambda) &= \min_{x \geq 0} L(x, \lambda) \\ &= \min_{x \geq 0} \{x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 4)\} \\ &= \left(\min_{x_1 \geq 0} \{x_1^2 - \lambda x_1\} + \min_{x_2 \geq 0} \{x_2^2 - \lambda x_2\} + 4\lambda \right) \\ &= -\frac{1}{4}\lambda^2 - \frac{1}{4}\lambda^2 + 4\lambda \\ &= -\frac{1}{2}\lambda^2 + 4\lambda \end{aligned}$$

其中 $x_1^* = x_2^* = \frac{1}{2}\lambda$

$$\max_{\lambda \geq 0} \theta(\lambda) = -\frac{1}{2}\lambda^2 + 4\lambda = 8$$

例 4.5 求解无约束优化问题★★★★★

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|_1$$

解 令 $\mathbf{y} = \mathbf{x}$ 得

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 + \lambda \|\mathbf{y}\|_1 \\ \text{s.t.} \quad & \mathbf{y} - \mathbf{x} = 0 \end{aligned}$$

对偶

$$\max_{\mathbf{z} \in \mathbb{R}^n} \min_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^n} = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 + \lambda \|\mathbf{y}\|_1 + \mathbf{z}^T(\mathbf{x} - \mathbf{y}).$$

变量可分离

$$\boxed{\min_{\mathbf{x} \in \mathbb{R}^n} \left(\frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 + \mathbf{z}^T \mathbf{x} \right)} + \boxed{\min_{\mathbf{y} \in \mathbb{R}^n} (\lambda \|\mathbf{y}\|_1 - \mathbf{z}^T \mathbf{y})}$$

对于 $\min_{\mathbf{y} \in \mathbb{R}^n} (\lambda \|\mathbf{y}\|_1 - \mathbf{z}^T \mathbf{y})$ 有

$$\min_{\mathbf{y} \in \mathbb{R}^n} \lambda \|\mathbf{y}\|_1 - \mathbf{z}^T \mathbf{y} = \begin{cases} 0, & \|\mathbf{z}\|_\infty \leq \lambda, \\ -\infty, & \|\mathbf{z}\|_\infty > \lambda. \end{cases}$$

综上

$$\begin{aligned} & \max_{z \in \mathbb{R}^n} \min_{x, y \in \mathbb{R}^n} L(x, y, z) \\ &= \max_{\|z\|_\infty \leq \lambda} \frac{1}{2} \|A(A^T A)^{-1}(A^T b - z) - b\|^2 + z^T (A^T A)^{-1}(A^T b - z) \\ &= \max_{\|z\|_\infty \leq \lambda} -\frac{1}{2} z^T (A^T A)^{-1} z + z^T (A^T A)^{-1} A^T b \\ &= \min_{\|z\|_\infty \leq \lambda} \frac{1}{2} z^T (A^T A)^{-1} z - z^T (A^T A)^{-1} A^T b \end{aligned}$$

- 原问题为线性约束的凸优化问题, 强对偶定理成立.
- 从而, 原问题等价地化为一带简单约束的光滑优化问题.

4.4 二次规划

4.4.1 模型与基本性质

优化模型

$$\begin{aligned} \min \quad & Q(x) = \frac{1}{2} x^T G x + g^T x \\ \text{s.t.} \quad & a_i^T x = b_i, i \in \mathcal{E} \\ & a_i^T x \geq b_i, i \in \mathcal{I} \end{aligned}$$

其中 $G \in \mathbb{R}^{n \times n}, g \in \mathbb{R}^n$

目标函数二次, 二阶连续可微

$$\nabla Q(x) = Gx + g \quad \nabla^2 Q(x) = G$$

约束线性

$$\Omega = \{x \in \mathbb{R}^n \mid a_i^T x = b_i, i \in \mathcal{E}; a_i^T x \geq b_i, i \in \mathcal{I}\}$$

定理 4.5 (Frank-Wolfe 定理, 1956)

$$\begin{aligned} \min \quad & Q(x) = \frac{1}{2} x^T G x + g^T x \\ \text{s.t.} \quad & a_i^T x = b_i, i \in \mathcal{E} \\ & a_i^T x \geq b_i, i \in \mathcal{I} \end{aligned}$$

若目标函数在可行域上有下界, 则有全局最优解.

推论 若严格凸二次规划可行域非空, 则有唯一全局最优解.

推论 更深的结论: 严格凸二次函数在任意非空闭凸集上都有唯一全局最优解.

4.4.2 对偶理论

$$\begin{aligned} \min \quad & Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^T \mathbf{x} = b_i, i \in \mathcal{E} \\ & \mathbf{a}_i^T \mathbf{x} \geq b_i, i \in \mathcal{I} \end{aligned}$$

矩阵 \mathbf{G} 对称正定, 记 $\mathbf{A}^T = (\mathbf{a}_i)_{\mathcal{I} \cup \mathcal{E}}, \mathbf{b} = (b_i)_{\mathcal{I} \cup \mathcal{E}}$.

Lagrange 函数

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} - \boldsymbol{\lambda}^T (\mathbf{A} \mathbf{x} - \mathbf{b}), \quad \lambda_i \geq 0, i \in \mathcal{I},$$

Lagrange 对偶

$$\max_{\boldsymbol{\lambda}} \min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda})$$

由于矩阵 \mathbf{G} 对称正定, 那么

$$\max_{\boldsymbol{\lambda}} \left[\min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x} - \boldsymbol{\lambda}^T (\mathbf{A} \mathbf{x} - \mathbf{b}) \right]$$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{G} \mathbf{x} + \mathbf{g} - \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{0}$$

有

$$\mathbf{x} = \mathbf{G}^{-1}(\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}).$$

代入得到

$$\begin{aligned} & \frac{1}{2}(\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g})^T (\mathbf{G}^{-1})^T \mathbf{G} \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) + \mathbf{g}^T \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) - \boldsymbol{\lambda}^T [\mathbf{A} \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) - \mathbf{b}] \\ \Rightarrow & \frac{1}{2}(\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g})^T \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) + \mathbf{g}^T \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) - \boldsymbol{\lambda}^T [\mathbf{A} \mathbf{G}^{-1} (\mathbf{A}^T \boldsymbol{\lambda} - \mathbf{g}) - \mathbf{b}] \end{aligned}$$

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{\lambda}^T (\mathbf{A} \mathbf{G}^{-1} \mathbf{A}^T) \boldsymbol{\lambda} - (\mathbf{b} + \mathbf{A} \mathbf{G}^{-1} \mathbf{g})^T \boldsymbol{\lambda} \\ \text{s.t.} \quad & \lambda_i \geq 0, \quad i \in \mathcal{I} \end{aligned}$$

定理 4.6 设 $\mathbf{x}^*, \boldsymbol{\lambda}^*$ 分别为严格凸二次规划及对偶规划的最优解. 则

$$\mathbf{x}^* = \mathbf{G}^{-1}(\mathbf{A}^T \boldsymbol{\lambda}^* - \mathbf{g})$$

对严格凸二次规划, 利用其对偶解可得原规划最优解。

4.4.3 等式约束二次规划

注 无约束二次规划

$$\min Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{g}^T \mathbf{x}$$

定理 4.7 凸二次函数 $\min Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x}$ 有最优解等价于

$$\mathbf{g} \in \mathcal{R}(\mathbf{G})$$

即

$$\text{rank}(\mathbf{G}) = \text{rank}(\mathbf{G}, \mathbf{g})$$

证明.

$$\begin{aligned}\nabla Q(\mathbf{x}) &= \mathbf{G}\mathbf{x} + \mathbf{g} = \mathbf{0} \\ \Rightarrow \mathbf{G}\mathbf{x} &= -\mathbf{g}\end{aligned}$$

证毕!

例 4.6 求解以下 无约束二次规划 ★★★★★☆

$$\begin{aligned}\min \quad & Q(\mathbf{x}) = x_1^2 - x_2^2 - x_3^2 \\ \text{s. t.} \quad & x_1 + x_2 + x_3 = 1 \\ & x_2 - x_3 = 1\end{aligned}$$

解

$$\mathbf{x}_B = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mathbf{x}_N = x_3$$

则 $x_2 = x_3 + 1, x_1 = -2x_3$ 问题转化为

$$\begin{aligned}\min Q(x_3) &= 4x_3^2 - (x_3 + 1)^2 - x_3^2 \\ &= 2x_3^2 - x_3 - 1\end{aligned}$$

得到

$$x_3^* = \frac{1}{2} \quad \mathbf{x}^* = \begin{pmatrix} -1 \\ 3/2 \\ 1/2 \end{pmatrix}$$

注 等式约束二次规划的 Lagrange 方法

$$\begin{aligned}\min \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & c_i(\mathbf{x}) = 0, i \in \mathcal{E}\end{aligned}$$

原理：通过 KKT 条件获取最优解

$$\begin{cases} \nabla f(\mathbf{x}^*) = \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(\mathbf{x}^*), \\ \mathbf{c}(\mathbf{x}^*) = \mathbf{0}. \end{cases}$$

二次规划问题

$$\begin{aligned}\min \quad & Q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{G}\mathbf{x} + \mathbf{g}^T \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}\end{aligned}$$

K-T 条件

$$\begin{cases} \mathbf{G}\mathbf{x} + \mathbf{g} = \mathbf{A}^T\boldsymbol{\lambda} \\ \mathbf{A}\mathbf{x} = \mathbf{b} \end{cases}$$

转化为

$$\begin{pmatrix} \mathbf{G} & -\mathbf{A}^T \\ -\mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\mathbf{g} \\ -\mathbf{b} \end{pmatrix}$$

定理 4.8 设矩阵 \mathbf{A} 行满秩, 并对任意满足 $\mathbf{A}\mathbf{x} = \mathbf{0}$ 的非零向量 \mathbf{x} , 均有 $\mathbf{x}^T\mathbf{G}\mathbf{x} > 0$ 则矩阵

$$\begin{pmatrix} \mathbf{G} & -\mathbf{A}^T \\ -\mathbf{A} & \mathbf{0} \end{pmatrix}$$

非奇异。

例 4.7 用 Lagrange 法求解如下问题★★★★★

$$\begin{aligned} \min \quad & Q(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + x_1x_3 + 2.5x_2^2 + 2x_2x_3 + 2x_3^2 - 8x_1 - 3x_2 - 3x_3 \\ \text{s.t.} \quad & x_1 + x_3 = 3 \\ & x_2 + x_3 = 0 \end{aligned}$$

解 由目标函数及约束条件

$$\mathbf{G} = \begin{pmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{pmatrix}, \mathbf{g} = \begin{pmatrix} -8 \\ -3 \\ -3 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

KKT 系统

$$\begin{pmatrix} 6 & 2 & 1 & -1 & 0 \\ 2 & 5 & 2 & 0 & -1 \\ 1 & 2 & 4 & -1 & -1 \\ -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \\ 3 \\ -3 \\ 0 \end{pmatrix}$$

解得

$$\mathbf{x}^* = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \quad \boldsymbol{\lambda}^* = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

4.5 有效集方法

例 4.8 求解二次规划问题★★★★★

$$\begin{aligned} \min \quad & Q(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1 - 4x_2 \\ \text{s.t.} \quad & -x_1 - x_2 + 1 \geq 0 \\ & x_1, x_2 \geq 0 \end{aligned}$$

解 第一步迭代: 取初始点

$$\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, S_0 = \mathcal{A}(\mathbf{x}^{(0)}) = \{2, 3\}$$

求解等式约束优化子问题

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 4d_2 \\ \text{s.t.} \quad & d_1 = 0 \\ & d_2 = 0 \end{aligned}$$

得最优解和相应的 Lagrange 乘子

$$\mathbf{d}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\lambda}^{(0)} = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$$

因为 $\mathbf{d}^{(0)} = \mathbf{0}$, 新的迭代点为

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

因为 $\lambda_i < 0$, 修正指标集

$$S_1 = S_0 / \{3\} = \mathcal{A}_0 / \{3\} = \{2\}$$

进入第二次迭代:

求解子问题

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 4d_2 \\ \text{s.t.} \quad & d_1 = 0 \end{aligned}$$

得最优解

$$\mathbf{d}^{(1)} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

因为 $\mathbf{d}^{(1)} \neq \mathbf{0}$, 计算步长

$$\begin{aligned} \alpha_1 &= \min \left\{ 1, \frac{b_i - \mathbf{a}_i^T \mathbf{x}^{(1)}}{\mathbf{a}_i^T \mathbf{d}^{(1)}} \mid i = 1, 3, \mathbf{a}_i^T \mathbf{d}^{(1)} < 0 \right\} \\ &= \frac{b_1 - \mathbf{a}_1^T \mathbf{x}^{(1)}}{\mathbf{a}_1^T \mathbf{d}^{(1)}} = \frac{1}{2} \end{aligned}$$

从 $\mathcal{I}/S_1 = \{1, 3\}$ 中取

$$\begin{aligned} 0 + \frac{1}{2} [(-1 \times 0) + (-1 \times 2)] &= -1 \\ 0 + \frac{1}{2} [(0 \times 0) + (1 \times 2)] &\neq 0 \end{aligned}$$

不等式积极约束 $i = 1$ 令

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \alpha_1 \mathbf{d}^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, S_2 = S_1 \cup \{1\} = \{1, 2\}$$

进入第三次迭代:

求解子问题

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 2d_2 \\ \text{s.t.} \quad & d_1 + d_2 = 0 \\ & d_1 = 0 \end{aligned}$$

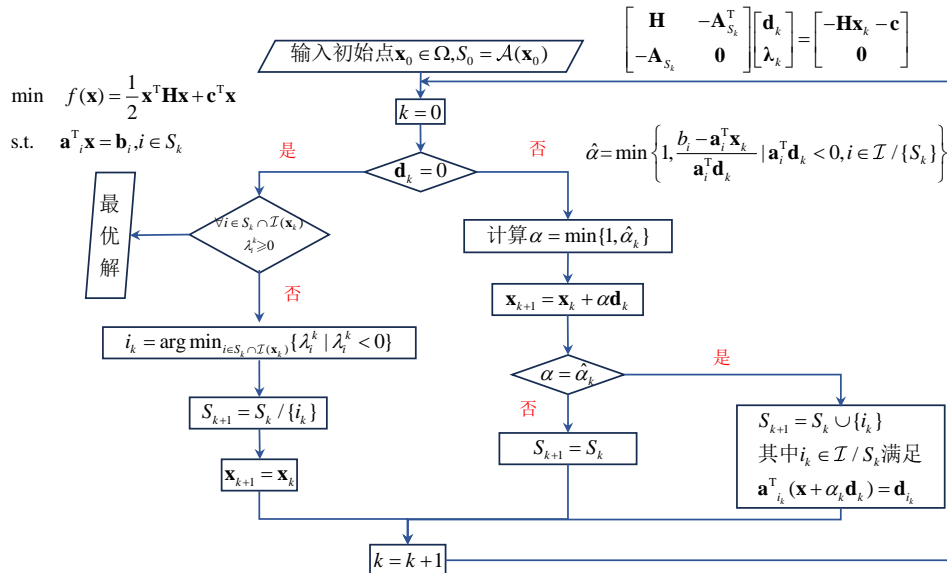
得最优解

$$\mathbf{d}^{(2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\lambda}^{(2)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

由 $\mathbf{d}^{(2)} = 0$, 且对于 $\forall i \in S_k \cap \mathcal{I}(\mathbf{x}^{(2)}), \lambda \geq 0$ 算法停止。原问题最优解和最优 Lagrange 乘子分别为:

$$\mathbf{x}^* = \mathbf{x}^{(2)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \boldsymbol{\lambda}^* = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

上述过程的流程见下图



4.6 约束优化可行方向法

4.6.1 Zoutendijk 可行方向法

策略: 在可行条件下, 让下降性达到最大

Frank-Wolfe 方法:

$$\begin{aligned} \min & \nabla f(\mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) \\ \text{s.t. } & \mathbf{x} \in \Omega \end{aligned}$$

缺陷: 会产生模很大, 下降性很差的搜索方向

Zoutendijk 可行方向法: 在可行方向锥中寻求最靠近负梯度方向的搜索方向

$$\mathbf{x}_k \in \Omega \Rightarrow \begin{cases} \mathbf{a}_i^T \mathbf{x}_k = b_i, & i \in \mathcal{E} \\ \mathbf{a}_i^T \mathbf{x}_k = b_i, & i \in \mathcal{I}(\mathbf{x}_k) \\ \mathbf{a}_i^T \mathbf{x}_k > b_i, & i \in \mathcal{I}/\mathcal{I}(\mathbf{x}_k) \end{cases}$$

$$\mathbf{d}_k \text{可行} \Rightarrow \begin{cases} \mathbf{a}_i^T (\mathbf{x}_k + \alpha \mathbf{d}_k) = b_i, & i \in \mathcal{E} \\ \mathbf{a}_i^T (\mathbf{x}_k + \alpha \mathbf{d}_k) \geq b_i, & i \in \mathcal{I} \\ \mathbf{a}_i^T \mathbf{d}_k = 0, & i \in \mathcal{E} \\ \mathbf{a}_i^T \mathbf{d}_k \geq 0, & i \in \mathcal{I}(\mathbf{x}_k) \end{cases}$$

$$\begin{aligned} \mathbf{d}_k &= \arg \min \mathbf{d}^T \nabla f(\mathbf{x}_k) \\ \text{s.t. } & \mathbf{a}_i^T \mathbf{d} \geq 0, \quad i \in \mathcal{I}(\mathbf{x}_k) \\ & \mathbf{a}_i^T \mathbf{d} = 0, \quad i \in \mathcal{E} \\ & -1 \leq d_i \leq 1, \quad 1 \leq i \leq n \end{aligned}$$

计算最大可行步长

$$\begin{aligned} \mathbf{a}_i^T (\mathbf{x}_k + \alpha \mathbf{d}_k) &\geq b_i, \quad \alpha \geq 0, i \in \mathcal{I}/\mathcal{I}(\mathbf{x}_k^*) \\ \hat{\alpha}_k &= \min_{i \in \mathcal{I}/\mathcal{I}(\mathbf{x}_k)} \left\{ \frac{b_i - \mathbf{a}_i^T \mathbf{x}_k}{\mathbf{a}_i^T \mathbf{d}_k} \mid \mathbf{a}_i^T \mathbf{d}_k < 0 \right\} \end{aligned}$$

最优步长

$$\alpha_k = \arg \min \{f(\mathbf{x}_k + \alpha \mathbf{d}_k) \mid 0 \leq \alpha \leq \hat{\alpha}_k\}$$

4.7 投影算子

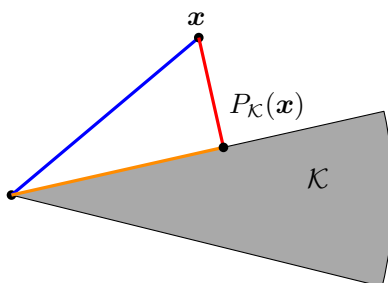
定义 4.4 (投影算子) 设 Ω 为非空闭凸集. 对任意 $\mathbf{x} \in \mathbb{R}^n$, 定义

$$P_\Omega(\mathbf{x}) = \arg \min \{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{y} \in \Omega\},$$

并称其为 \mathbf{x} 到 Ω 上的投影。称 $P_\Omega(\cdot)$ 为从 \mathbb{R}^n 到 Ω 上的投影算子

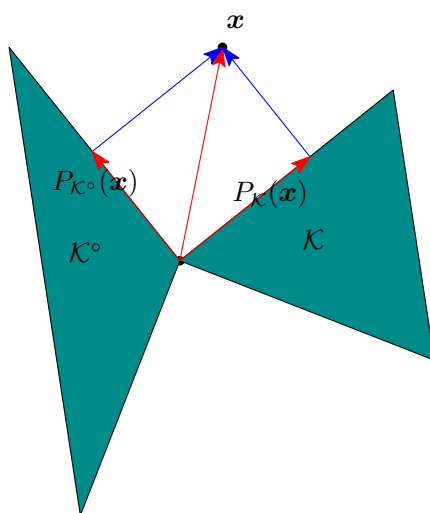
定理 4.9 对闭凸锥 \mathcal{K} 和任意 $\mathbf{x} \in \mathbb{R}^n$

1. $\langle \mathbf{x} - P_{\mathcal{K}}(\mathbf{x}), P_{\mathcal{K}}(\mathbf{x}) \rangle = 0$, 即 $\langle \mathbf{x}, P_{\mathcal{K}}(\mathbf{x}) \rangle = \|P_{\mathcal{K}}(\mathbf{x})\|^2$
2. $\|\mathbf{x}\| \geq \|P_{\mathcal{K}}(\mathbf{x})\|$



定理 4.10 (Moreau 分解定理) 设 \mathcal{K} 为非空闭凸锥, $\mathcal{K}^0 = \{\mathbf{y} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{y} \rangle \leq 0, \forall \mathbf{x} \in \mathcal{K}\}$, 对任意的 $\mathbf{x} \in \mathbb{R}^n$, 令 $\mathbf{y} = P_{\mathcal{K}}(\mathbf{x})$, $\mathbf{z} = P_{\mathcal{K}^0}(\mathbf{x})$, 则 $\mathbf{x} = \mathbf{y} + \mathbf{z}$ 且 $\langle \mathbf{y}, \mathbf{z} \rangle = 0$.

几何解释：任一向量可分解成闭凸锥与其极锥上的两个投影的直和。



定理 4.11 $\mathbf{x} \in \Omega$ 为凸约束优化问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in \Omega\}$ 的稳定点, 则对任意的 $\alpha > 0$, 均有 $\mathbf{x} = P_{\Omega}(\mathbf{x} - \alpha \nabla f(\mathbf{x}))$

残量函数 $r(\mathbf{x}) = \mathbf{x} - P_{\Omega}(\mathbf{x} - \nabla f(\mathbf{x}))$ 可作为价值函数度量当前点与最优解（稳定点）之间的近似程度

易操作的最优解判定方式。可作为算法终止规则。

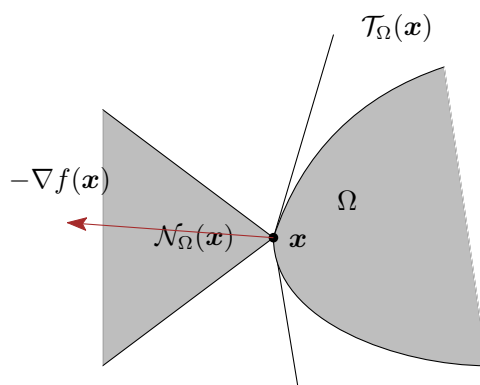
定义 4.5 $\mathbf{x} \in \Omega$. 称 $\mathcal{T}_{\Omega}(\mathbf{x}) = \{\mathbf{d} \in \mathbb{R}^n \mid \text{存在 } \mathbf{d}_k \rightarrow \mathbf{d}, t_k \rightarrow 0^+, \text{ 使得 } \mathbf{x} + t_k \mathbf{d}_k \in \Omega\}$ 为闭凸集 Ω 在 \mathbf{x} 点的切锥。

定义 4.6 定义 $-\nabla f(\mathbf{x})$ 在切锥 $\mathcal{T}_{\Omega}(\mathbf{x})$ 上的投影定义为约束优化问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in \Omega\}$ 在 \mathbf{x} 点的投影梯度, 即

$$\nabla_{\Omega} f(\mathbf{x}) \triangleq \arg \min\{\|\mathbf{d} + \nabla f(\mathbf{x})\| \mid \mathbf{d} \in \mathcal{T}_{\Omega}(\mathbf{x})\}$$

定理 4.12 $\mathbf{x} \in \Omega$ 为约束优化问题 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in \Omega\}$ 稳定点当且仅当 $\nabla_{\Omega} f(\mathbf{x}) = \mathbf{0}$.

几何意义：在最优值点（稳定点），负梯度属于该点的法锥。



4.7.1 矩阵的广义逆

若矩阵 \mathbf{A} 奇异，或者非方阵，是否存在与逆矩阵类似的矩阵？

定义 4.7 矩阵的广义逆： $\mathbf{A} \in \mathbb{R}^{m \times n}$ 。若存在矩阵 $\mathbf{G} \in \mathbb{R}^{n \times m}$ 使得

$$\begin{aligned} \mathbf{AGA} &= \mathbf{A}, \quad \mathbf{GAG} = \mathbf{G}, \\ (\mathbf{AG})^T &= \mathbf{AG}, \quad (\mathbf{GA})^T = \mathbf{GA} \end{aligned}$$

则称矩阵 \mathbf{G} 为矩阵 \mathbf{A} 的广义逆，又称 Moore-Penrose 广义逆。记为 \mathbf{A}^+

- 矩阵非奇异 $\mathbf{A}^+ = \mathbf{A}^{-1}$
- 矩阵行满秩 $\mathbf{A}^+ = \mathbf{A}^T(\mathbf{AA}^T)^{-1}$

$$\mathbf{A} \mathbf{A}^T (\mathbf{AA}^T)^{-1} \mathbf{A} = \mathbf{A}$$

- 矩阵列满秩 $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$

$$\mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} = \mathbf{A}$$

- 矩阵行（或列）单位正交，即满足矩阵 $\mathbf{AA}^T = \mathbf{I}_m$ 或者 $(\mathbf{A}^T \mathbf{A} = \mathbf{I}_n)$ ，则 $\mathbf{A}^+ = \mathbf{A}^T$

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}_1^T & \mathbf{v}_2^T & \cdots & \mathbf{v}_m^T \end{bmatrix}^T \text{ 且 } \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$$

$$\mathbf{AA}^T = \begin{bmatrix} \mathbf{v}_i^T \mathbf{v}_j \end{bmatrix}_{m \times m} = \mathbf{I}_{m \times m}$$

有

$$\mathbf{A} \mathbf{A}^T \mathbf{A} = \mathbf{I}_{m \times m} \mathbf{A} = \mathbf{A}$$

4.8 约束优化投影算法

凸约束优化 $\min\{f(\mathbf{x}) \mid \mathbf{x} \in \Omega\}$ 可行域非空闭凸

基本思想：沿负梯度方向行进，借助投影算子拉回可行域

迭代格式 $\mathbf{x}_{k+1} = P_{\Omega}(\mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k))$.

5 其他方法

5.1 罚函数方法

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & c_i(\mathbf{x}) = 0, i \in \mathcal{E} \\ & c_i(\mathbf{x}) \geq 0, i \in \mathcal{I} \end{aligned}$$

定义 5.1 (罚函数方法) 罚函数方法：借助正则化技术将约束函数揉进目标函数，得到一无约束优化问题，然后极小化后者得约束函数违反度和目标函数值都尽可能小的解。

5.1.1 外点罚函数

$$\begin{aligned} \min \{f(\mathbf{x}) \mid c_i(\mathbf{x}) = 0, i \in \mathcal{E}\} \\ \min_{\mathbf{x} \in \mathbb{R}^n} P(\mathbf{x}, \pi) = f(\mathbf{x}) + \pi \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x}) \end{aligned}$$

5.2 乘子罚函数

等式约束优化	$\min \{f(\mathbf{x}) \mid c_i(\mathbf{x}) = 0, i \in \mathcal{E}\}$
	↓
外点罚函数	$P(\mathbf{x}, \pi) = f(\mathbf{x}) + \pi \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x})$
	↓
乘子罚函数	$P(\mathbf{x}, \boldsymbol{\lambda}, \pi) = f(\mathbf{x}) - \sum_{i \in \mathcal{E}} \lambda_i c_i(\mathbf{x}) + \pi \sum_{i \in \mathcal{E}} c_i^2(\mathbf{x})$

定理 5.1 设 $\mathbf{x}^* = \arg \min \{f(\mathbf{x}) \mid c_i(\mathbf{x}) = 0, i \in \mathcal{E}\}$, $\boldsymbol{\lambda}^*$ 为最优 Lagrange 乘子; $\nabla c_i(\mathbf{x}^*), i \in \mathcal{E}$ 线性无关。

原规划问题在最优值点 \mathbf{x}^* 满足二阶最优性条件。

$$\text{乘子罚函数 } P(\mathbf{x}, \boldsymbol{\lambda}, \pi) = f(\mathbf{x}) - \sum_{i \in \mathcal{S}} \lambda_i c_i(\mathbf{x}) + \pi \sum_{i \in \mathcal{S}} c_i^2(\mathbf{x})$$

则存在 $\pi^* > 0$, 使对任意的 $\pi \geq \pi^*$, \mathbf{x}^* 为罚函数 $P(\mathbf{x}, \boldsymbol{\lambda}^*, \pi)$ 严格最小值点.

6 张量优化

6.1 基本概念

定义 6.1 (d 阶张量) 又称超矩阵, d 维数组

$$\mathbf{T} = (t_{i_1 i_2 \dots i_d})$$