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 6 月 10 日

目录

1 绪论

1.1 最优化简介

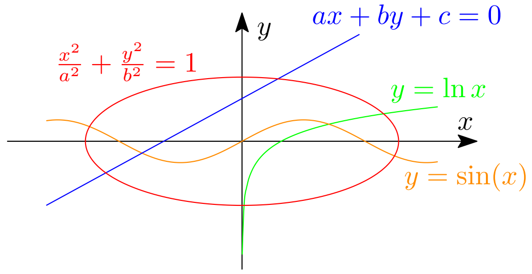
优化问题可以表示成

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

可行域有多种表达形式，一般用等式和不等式定义，即

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

注 线性问题：



- 线性规划

$$\begin{aligned} &\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ &\mathrm{s.t.}\ \boldsymbol{A}\boldsymbol{x} = 0 \\ &\quad\ \boldsymbol{x} \geqslant 0 \end{aligned}$$

- 非线性规划

$$\begin{aligned} &\min\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} + \boldsymbol{g}^{\mathrm{T}}\boldsymbol{x} \\ &\mathrm{s.t.}\ \boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{b}_1 \\ &\quad\ \boldsymbol{A}_2\boldsymbol{x} \geqslant \boldsymbol{b}_2 \end{aligned}$$

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。

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

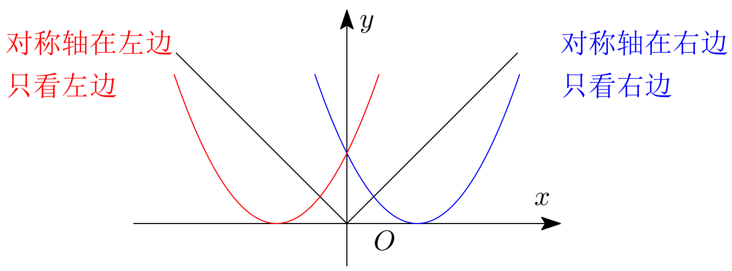
如二元函数 $f(\boldsymbol{x}) = x_1^2 + (1 - x_1x_2)^2$ 在 $\mathbb{R}^2$ 上的最优值为 0，但只有在 $x_1 = \frac{1}{x_2}$ 且 $x_2 \to \infty$ 是才能达到。

1.2 最优化方法概述

1.2.1 常见的优化方法

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

1. 解析法：给出问题最优解的解析式
   - 二次优化问题
     $$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$$
     若 Hesse 阵正定，最优解：$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$
   - 可分离的优化问题，其中 $\boldsymbol{a}\in\mathbb{R}^n, \lambda > 0$ ★★★★★
     $$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{a}\|^2 + \lambda\|\boldsymbol{x}\|_1$$
     求解子问题
     $$\sum_{i=1}^{n}\left(\min_{x_i\in\mathbb{R}} \frac{1}{2}(x_i-a_i)^2 + \lambda|x_i|\right) = \begin{cases} \min\limits_{x_i\in\mathbb{R}} \dfrac{1}{2}(x_i-a_i)^2 + \lambda x_i, & a_i \geqslant 0,\ \text{对称轴在右边} \\ \min\limits_{x_i\in\mathbb{R}} \dfrac{1}{2}(x_i-a_i)^2 - \lambda x_i, & a_i < 0,\ \text{对称轴在左边} \end{cases}$$
     最优解



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

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 $\boldsymbol{x}_0 \in \mathbb{R}^n$，及有关参数. 令 $k=0$
2. 验证停机准则.
3. 计算当前迭代点 $\boldsymbol{x}_k$ 的搜索方向 $\boldsymbol{d}_k \in \mathbb{R}^n$
4. 计算步长 $\alpha_k > 0$ 使满足
$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) < f(\boldsymbol{x}_k)$$
$$\alpha_k = \arg\min_{\alpha_k > 0} f(\boldsymbol{x}_k + \alpha_k + \boldsymbol{d}_k)$$
5. 令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$，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。

$$\begin{aligned} f(\boldsymbol{x}_k + \boldsymbol{d}) &= f(x_k) + \boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}_k) + \frac{1}{2} \boldsymbol{d}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \\ &\approx f(\boldsymbol{x}_k) + \boldsymbol{d}^{\mathrm{T}} \nabla f(\boldsymbol{x}_k) + \frac{1}{2} \boldsymbol{d}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{d} \end{aligned}$$

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

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

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

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

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

$$\limsup_{k\to\infty} = \frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|} \leqslant q$$

  Q-线性收敛：$0<q<1$

  Q-超线性收敛：$q=0$

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

$$\limsup_{k\to\infty} = \frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|^r} \leqslant q$$

  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

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

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

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

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 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$ 等价于下述约束优化问题

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

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

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

的最优值为 $f(\boldsymbol{x}^*)$。证毕！

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

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

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

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

从而由三角不等式

$$\begin{aligned}\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|} &\leqslant \frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|+\|\boldsymbol{x}^*-\boldsymbol{x}_k\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|} \\ &= \lim_{k\to\infty} \frac{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|} = 1\end{aligned}$$

另一方面，

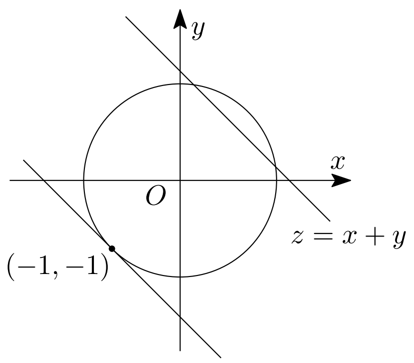
$$\begin{aligned}\lim_{k\to\infty}\frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|} &\geqslant \frac{\|\boldsymbol{x}^*-\boldsymbol{x}_k\|-\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}\\ &=\lim_{k\to\infty}\frac{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}=1\end{aligned}$$

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

$$\min\left\{x_1+x_2|x_1^2+x_2^2\leqslant 2\right\}$$

根据右图，该问题的最优解为$(x_1^*,x_2^*)=(-1,-1)$



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

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

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

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

即

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

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

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

例 1.6 设 $\boldsymbol{a}\in\mathbb{R}^n,\rho>0$，矩阵 $\boldsymbol{\Sigma}_0\succ 0$（对称正定）。试用解析法求解如下优化问题。

$$\begin{aligned}\max_{\boldsymbol{\Sigma}\succ 0}\quad&\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{a}\\ \text{s.t.}\quad&\|\boldsymbol{\Sigma}-\boldsymbol{\Sigma}\|_{\mathrm{F}}\leqslant\rho\end{aligned}$$

解：令 $\Delta\boldsymbol{\Sigma}=\boldsymbol{\Sigma}-\boldsymbol{\Sigma}_0$，则原优化问题转化为

$$\begin{array}{ll}\max & \boldsymbol{a}^{\mathrm{T}}\Delta\boldsymbol{\Sigma}\boldsymbol{a}+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{a} \\ \text{s.t.} & \|\Delta\boldsymbol{\Sigma}\|_{\mathrm{F}}\leqslant\rho\end{array}$$

由 Cauchy-Scwarzy 不等式：

$$\boldsymbol{a}^{\mathrm{T}}\Delta\boldsymbol{\Sigma}\boldsymbol{a}=\left\langle\Delta\boldsymbol{\Sigma},\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\right\rangle\leqslant\|\Delta\boldsymbol{\Sigma}\|_{\mathrm{F}}\|\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\|_{\mathrm{F}}\leqslant\rho\|\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\|_{\mathrm{F}}$$

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

原问题的最优解为

$$\boldsymbol{\Sigma}^*=\boldsymbol{\Sigma}_0+\rho\frac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\|\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\|}$$

上述问题的最优解为

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}^*\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{a}+\rho\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$$

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 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n}f(\boldsymbol{x})$，其中 $f:\mathbb{R}^n\to\mathbb{R}$

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) $\boldsymbol{x}^*$ 是 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n}f(\boldsymbol{x})$(局部) 最优解，则 $\nabla f(\boldsymbol{x}^*)=\mathbf{0}$

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

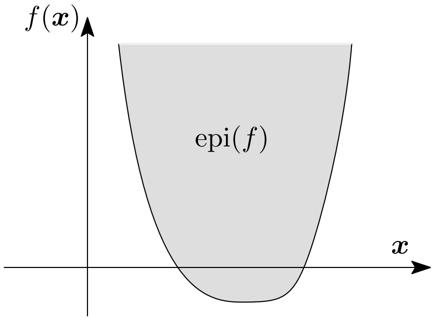
$$\begin{aligned}f(\boldsymbol{x}^*+\alpha\boldsymbol{d}) &=f(\boldsymbol{x}^*)+\alpha\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d}+o(\alpha)\\ &=f(\boldsymbol{x}^*)-\alpha\|\nabla f(\boldsymbol{x}^*)\|^2+o(\alpha)\\ &<f(\boldsymbol{x}^*)\end{aligned}$$

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

定义 2.1 (凸函数) $f(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y})\leqslant\lambda f(\boldsymbol{x})+(1-\lambda)f(\boldsymbol{y}),\forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n,\lambda\in[0,1]$ 或者 $f\left(\sum\limits_{i=1}^m\lambda_i\boldsymbol{x}_i\right)\leqslant\sum\limits_{i=1}^m\lambda_i f(\boldsymbol{x}_i),\forall\boldsymbol{x}_i\in\mathbb{R}^n,\lambda_i\geqslant0,i=1,2,\cdots,m,\sum\limits_{i=1}^m\lambda_i=1$

定义 2.2 (凸组合) 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m \in \mathbb{R}^n$, $\lambda_1+\lambda_2+\cdots+\lambda_m = 1, \lambda_1, \lambda_2, \cdots, \lambda_m \in (0,1)$，则称 $\lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2 + \cdots + \lambda_m\boldsymbol{x}_m$ 为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 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 $\mathrm{epi}(f)$ 是凸集。$\mathrm{epi}(f) := \{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \alpha \geqslant f(\boldsymbol{x})\}$ 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。$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geqslant \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

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

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

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

则称其为一致凸函数，又称强凸函数。

$$\begin{gathered}
f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leqslant \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}) - \lambda(1-\lambda)\eta\|\boldsymbol{x}-\boldsymbol{y}\|^2 \\
\Big\downarrow \text{连续可微} \\
f(\boldsymbol{y}) - f(\boldsymbol{x}) \geqslant \nabla f(\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) + \tfrac{1}{2}\eta\|\boldsymbol{y}-\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \\
\Big\updownarrow \text{或者} \\
(\boldsymbol{y}-\boldsymbol{x})^{\mathrm{T}}(\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})) \geqslant \eta\|\boldsymbol{y}-\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \\
\Big\downarrow \text{二阶连续可微} \\
\boldsymbol{h}^{\mathrm{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{h} \geqslant \eta\|\boldsymbol{h}\|^2, \quad \forall \boldsymbol{x}, \boldsymbol{h} \in \mathbb{R}^n
\end{gathered}$$

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

定理 2.2 (凸优化) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$，$f: \mathbb{R}^n \to \mathbb{R}$ 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

$$f(\boldsymbol{x})-f(\boldsymbol{x}^*)\geqslant\langle\nabla f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle=0$$
$$f(\boldsymbol{x})>f(\boldsymbol{x}^*)$$

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

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

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

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

取 $\alpha>0$ 充分小。由泰勒展开式

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

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 $\nabla f(\boldsymbol{x}^*)=\mathbf{0}$，$\nabla^2 f(\boldsymbol{x})$ 半正定，但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 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$，设 $\nabla f(\boldsymbol{x})=\mathbf{0}$，$\nabla^2\boldsymbol{f}(\boldsymbol{x}^*)$ 正定，则 $\boldsymbol{x}^*$ 为该优化问题的严格最优解

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

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

其中 $\alpha>0$ 充分小。则

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

例 2.1 $\min\limits_{x>0,y\geqslant 0} f(x,y)=\dfrac{10}{x}+\dfrac{(x-y)^2}{2x}+\dfrac{3y^2}{2x}$ ★★★★★

先忽略约束：利用 $\min\limits_{x,y} f(x,y)=\min\limits_{x}\min\limits_{y} f(x,y)$，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

$$f(x,y)=\frac{1}{2x}\left(20+x^2-2xy+4y^2\right)$$
$$y=\frac{1}{4}x$$

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：

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

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

注 有关记号：

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

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

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

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

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

此时

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

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

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

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

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

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

$$|\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}|\le\|\nabla f(\boldsymbol{x})\|\cdot\|\boldsymbol{d}\|\le\|\nabla f(\boldsymbol{x})\|$$

可知

$$-\|\nabla f(\boldsymbol{x})\|\le\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}\le\|\nabla f(\boldsymbol{x})\|$$

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

注 算法分析：

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

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

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

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

5. 令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k$, $k=k+1$，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 $\boldsymbol{d}_k$ 与负梯度方向 $-\boldsymbol{g}_k$ 夹角满足

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

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

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

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

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 $0,1,\cdots$ 中满足下式的最小值

$$f(\boldsymbol{x}_k+\beta\gamma^m \boldsymbol{d}_k)\leqslant f_k+\sigma\beta\gamma^m \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k$$

如 $\alpha_k<\beta$，则

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

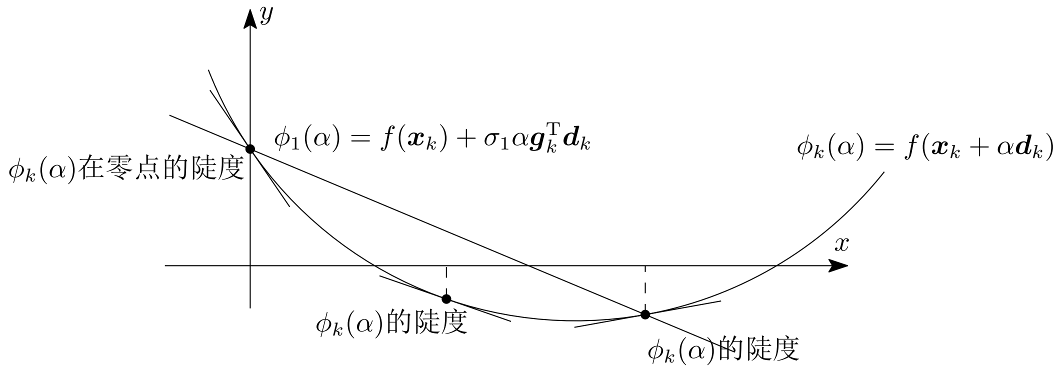
定义 2.8 (Wolfe 步长) 常数 $0<\sigma_1<\sigma_2<1$。步长 α_k 同时满足

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

和

$$\nabla f(\boldsymbol{x}_k+\alpha \boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k\geqslant \sigma_2 \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k \Longleftrightarrow \boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\geqslant\sigma_2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k$$

曲线 $\varphi_k(\alpha)=f(\boldsymbol{x}_k+\alpha \boldsymbol{d}_k)$ 的陡度 $\nabla f(\boldsymbol{x}_k+\alpha \boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k$ 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 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$

迭代算法

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

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 $\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{d}_{k+1}=0$

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

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

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

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

- 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 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^2} f(x_1,x_2)=\dfrac{1}{3}x_1^2+\dfrac{1}{2}x_2^2$。★★★★★

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

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

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

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

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

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

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

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

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

利用最优性条件

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

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

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

定理 2.7 设目标函数二阶连续可微，Hesse 阵在最优值点 $\boldsymbol{x}^*$ 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，$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^*$
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

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

左乘 $\boldsymbol{G}_k^{-1}$

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

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 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x}+\boldsymbol{g}^{\mathrm{T}}\boldsymbol{x}+c\Rightarrow\boldsymbol{x}^*=-\boldsymbol{G}^{-1}\boldsymbol{g}$
  任意初始点 $\boldsymbol{x}_0\in\mathbb{R}^n$

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

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

0 为最优解

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

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 $x_0\geqslant 1$，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（$\boldsymbol{A}$ 对称正定）：

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

- 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

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

2.4.1 线性共轭方向法

定义 2.9 (共轭方向) 对对称正定阵 $\boldsymbol{A}$，若 $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^n$，满足 $\boldsymbol{d}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_2 = 0$，则称 $\boldsymbol{d}_1, \boldsymbol{d}_2$ 关于矩阵 $\boldsymbol{A}$ 共轭，并称其为 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 $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n$ 关于对称正定阵 $\boldsymbol{A}$ 两两共轭，即满足

$$\boldsymbol{d}_i\boldsymbol{A}\boldsymbol{d}_j = 0, 1 \leqslant i \neq j \leqslant k$$

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

注 线性共轭方向法：

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

1. 初始点 $\boldsymbol{x}_0$，搜索方向 $\boldsymbol{d}_0$ 满足 $\langle \boldsymbol{d}_0, \boldsymbol{g}_0\rangle < 0$，终止参数 $\varepsilon \geqslant 0$，令 $k = 0$
2. 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$，算法终止；否则，进入下一步
3. 计算最优步长 $\alpha = \arg\min\limits_{\alpha\geqslant 0}\{f(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k)\}$，令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\boldsymbol{d}_k$
4. 构造 $\boldsymbol{d}_{k+1}$ 使其与 $\boldsymbol{d}_0, \cdots, \boldsymbol{d}_k$ 关于矩阵 $\boldsymbol{A}$ 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 $\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$,向量组 $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}$ 关于 $\boldsymbol{A}$ 共轭。共轭方向发产生点列 $\{\boldsymbol{x}_k\}$。则对任意 $0 \leqslant k \leqslant n-1$，$\boldsymbol{x}_{k+1}$ 是目标函数在仿射集 $\boldsymbol{x}_0 + \mathrm{span}\,[\boldsymbol{d}_0, \cdots, \boldsymbol{d}_k]$ 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 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 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，

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

解：

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

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

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

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

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

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

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

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

定理 2.9 (收敛速度) 对严格凸二次函数 $f(\boldsymbol{x})=\dfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$. 若系数矩阵 $\boldsymbol{A}$ 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。

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

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

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

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

$$\boxed{\begin{aligned}\boldsymbol{y}_k&=\boldsymbol{g}_{k+1}-\boldsymbol{g}_k\\ \boldsymbol{s}_k&=\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\end{aligned}}$$

$$\Longrightarrow\quad \boldsymbol{y}_k\approx\boldsymbol{G}_{k+1}\boldsymbol{s}_k$$

$$\overset{\boldsymbol{H}_{k+1}=\boldsymbol{B}_{k+1}^{-1}}{\Longrightarrow}\quad \boldsymbol{H}_{k+1}\boldsymbol{y}_k\approx\boldsymbol{s}_k$$

得到重要结论

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

2.5.1 对称秩-1 校正公式

注 对称秩-1 校正公式：

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

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

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

得到

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

所以对于 $\boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{T}}$ 有

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

故而，

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

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

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

   计算 $\boldsymbol{H}_{k+1}$，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

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

对 $\boldsymbol{H}_k$ 做如下对称秩-2 校正

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

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

$$\cancel{\boldsymbol{H}_k \boldsymbol{y}_k} + a \underline{\boldsymbol{s}_k} \boldsymbol{s}_k^{\mathrm{T}} \boldsymbol{y}_k + b \cancel{\boldsymbol{H}_k \boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{T}} \boldsymbol{H}_k} \boldsymbol{y}_k = \underline{\boldsymbol{s}_k}$$

两边相等，故而

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

故而得到 DFP 校正公式

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

例 2.5 用 DFP 方法求解优化问题★★★★★

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

解：容易计算 $\boldsymbol{x}^* = (1,0)^{\mathrm{T}}$

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

1. 取初始值 $\boldsymbol{x}_0 = (2,1)^{\mathrm{T}}, \boldsymbol{H}_0 = \boldsymbol{I}$，则 $\boldsymbol{g}_0 = (4,2)^{\mathrm{T}}$，沿着 $\boldsymbol{d}_0 = -\boldsymbol{H}_0 \boldsymbol{g}_0 = (-4,-2)^{\mathrm{T}}$
   精确线搜索得到 $\alpha = \dfrac{5}{18}$
   新迭代点 $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha \boldsymbol{d}_0 = (\frac{8}{9}, \frac{4}{9})^{\mathrm{T}}$ 进入下一次迭代
2. 计算 $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha \boldsymbol{d}_0 = (\frac{8}{9}, \frac{4}{9})^{\mathrm{T}}$，梯度 $\boldsymbol{g}_0 = (-\frac{4}{9}, \frac{8}{9})^{\mathrm{T}}$，进而计算

$$\begin{aligned} \boldsymbol{s}_0 &= \boldsymbol{x}_1 - \boldsymbol{x}_0 = \alpha_0 \boldsymbol{d}_0 = (-\frac{10}{9}, -\frac{5}{9})^{\mathrm{T}} \\ \boldsymbol{y}_0 &= \boldsymbol{g}_1 - \boldsymbol{g}_0 = (-\frac{40}{9}, -\frac{40}{9})^{\mathrm{T}} \end{aligned}$$

   得校正矩阵

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

   沿着 $\boldsymbol{d}_1 = -\boldsymbol{H}_1 \boldsymbol{g}_1 = \frac{12}{51}(1,-4)^{\mathrm{T}}$ 精确线搜索得到 $\alpha_1 = \dfrac{17}{36}$
   新迭代点 $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 \boldsymbol{d}_1 = (1,0)^{\mathrm{T}} = \boldsymbol{x}^*$，算法终止。

2.6 信赖域方法

优化模型 $\min\limits_{\boldsymbol{x} \in \mathbb{R}^n} f(x)$

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

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

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

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

注 信赖域子问题

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

其中，$\Delta_k > 0$ 为信赖域半径。

设 $\boldsymbol{d}_k$ 为其最优解。则目标函数与二次模型的近似度为

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

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

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

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

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

注 信赖域算法：

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

2.7 最小二乘问题

最小二乘问题数学模型

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

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

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

其中 $\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \cdots, r_m(\boldsymbol{x}))^{\mathrm{T}}$。

2.7.1 线性最小二乘

数学模型

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

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

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

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 $\boldsymbol{A}$ 列满秩。

2.7.2 非线性最小二乘

数学模型

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

注 梯度

$$\nabla f(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x}) = \nabla\left(\frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})\right) = \boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}) = \sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla r_i(\boldsymbol{x})$$ 数值乘以向量求和

其中

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

注 Hesse 阵

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

其中

$$\boldsymbol{S}(x) = \sum_{i=1}^{m} \boldsymbol{r}_i(\boldsymbol{x})\nabla^2 \boldsymbol{r}_i(\boldsymbol{x})$$

注 牛顿算法

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\left(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k+\boldsymbol{S}_k\right)^{-1}\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}_k)$$

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

$$\boldsymbol{d}_k^{\mathrm{GN}}=-\left(\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{J}_k\right)^{-1}\boldsymbol{J}_k^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}_k)$$

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k^{\mathrm{GN}}$$

3 凸优化

3.1 凸分析基础

定义 3.1 (Lipschitz 连续)

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

定义 3.2 (局部 Lipschitz 连续)

$$|f(\boldsymbol{x})-f(\boldsymbol{y})|\leqslant L_{\boldsymbol{x}_0}\|\boldsymbol{x}-\boldsymbol{y}\|,\quad\forall\boldsymbol{x},\boldsymbol{y}\in N(\boldsymbol{x}_0,\delta)$$

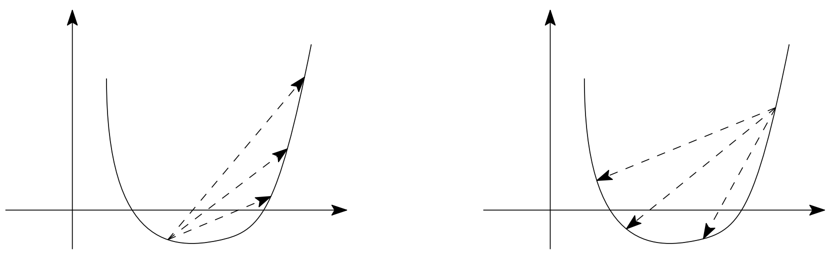
定义 3.3 (正常函数) $f:\mathbb{R}^n\to\mathbb{R}$ 满足 $f(x)>-\infty,\quad\forall x\in\mathbb{R}^n$ 且存在 $\bar{\boldsymbol{x}}\in\mathbb{R}^n$ 使得 $f(\bar{\boldsymbol{x}})<\infty$。（$f(\boldsymbol{x})$ 有下确界，且不恒取无穷大）

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 $f:\mathbb{R}^n\to\mathbb{R}$ 为正常凸函数，则对任意 $\boldsymbol{x},\boldsymbol{d}\in\mathbb{R}^n$，

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

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 $\lim\limits_{t\to0+}\dfrac{f(\boldsymbol{x}+t\boldsymbol{d})-f(\boldsymbol{x})}{t}$ 存在



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

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

$\boldsymbol{\xi}$ 称为凸函数 $f(\boldsymbol{x})$ 在 x 点的次梯度。

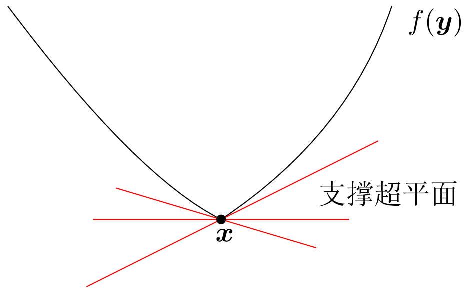
定义 3.5 (次微分) 满足

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

称

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

为函数 $f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 点的支撑超平面。

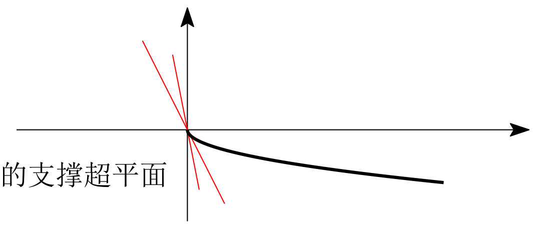


$(\boldsymbol{\xi},-1)$相当于支撑超平面的法向量

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

- - - - - - - -



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

$$\boldsymbol{\xi}\in\partial f(\boldsymbol{x})\Longleftrightarrow f'(\boldsymbol{x};\boldsymbol{d})\geqslant\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{d},\quad\forall\boldsymbol{d}\in\mathbb{R}^n.$$

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

证明.

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

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 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\boldsymbol{x}$ 点可微，则

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

定理 3.6 凸函数 $f:\mathbb{R}^n\to\mathbb{R}$ 和其定义域内任一内点 $\boldsymbol{x}$，$\partial f(x)$ 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 $f:\mathbb{R}^n\to\mathbb{R}$ 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

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

- 非凸局部 Lipschitz 连续

$$|f(\boldsymbol{x})-f(\boldsymbol{y})|\leqslant L_{\boldsymbol{x}_0}\|\boldsymbol{x}-\boldsymbol{y}\|,\quad \forall \boldsymbol{x},\boldsymbol{y}\in N(\boldsymbol{x}_0,\delta)$$

3.1.1 临近点算子

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

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

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)

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

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

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

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

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

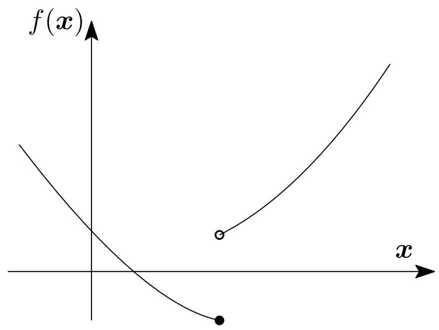
例 3.3 求 $f_3(x)=\begin{cases} 0, & x\neq 0 \\ \lambda, & x=0 \end{cases}$ 的临近点算子 $(\lambda>0)$。⭐⭐⭐⭐⭐

解：

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

定义 3.8 (下半连续)

$$
\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}\inf f(\boldsymbol{x})\geqslant f(\boldsymbol{x}_0)
$$



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

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

$$
\boldsymbol{y}=\operatorname{prox}_f(\boldsymbol{x})
$$

$$
\boldsymbol{x}-\boldsymbol{y}\in\partial f(\boldsymbol{y})
$$

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

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

$$\boldsymbol{y} = \operatorname{prox}_f(\boldsymbol{x}) \Longleftrightarrow \boldsymbol{y} = \arg\min_{\boldsymbol{z}\in\mathbb{R}^n}\{f(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2\}.$$

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

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

证毕！

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

$$\langle \boldsymbol{x}-\boldsymbol{y}, \operatorname{prox}_f(\boldsymbol{x}) - \operatorname{prox}_f(\boldsymbol{y})\rangle \geqslant \|\operatorname{prox}_f(\boldsymbol{x}) - \operatorname{prox}_f(\boldsymbol{y})\|^2$$

$$\|\operatorname{prox}_f(\boldsymbol{x}) - \operatorname{prox}_f(\boldsymbol{y})\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|.$$

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

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

用梯度和次梯度建立优化问题的最优性条件

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

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

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

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

定理 3.12 为上述优化问题的最优解当且仅当

$$g(\boldsymbol{x}) - g(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

注 临近点算法

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

注 算法收敛性

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

由

$$\boxed{\begin{gathered}\boldsymbol{x}^* = \arg\min f(\boldsymbol{x}) + g(\boldsymbol{x}) \\ g(\boldsymbol{x}) - g(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\Gamma(\boldsymbol{x} - \boldsymbol{x}^*) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^n\end{gathered}}$$

得

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_{k+1}) + L(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}_{k+1}) \geqslant 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$$

令 $\boldsymbol{x} = \boldsymbol{x}_k$

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

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

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

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

3.3 交替极小化方法

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

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

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

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

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

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 $\alpha > 0$,

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

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

对任意 $\alpha \leqslant 0$

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

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）。

$\boldsymbol{x} = \boldsymbol{0}$ 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 $\boldsymbol{x}^* \in \mathbb{R}^n$ 满足

$$\Psi(\boldsymbol{x}^*) \leqslant \Psi(\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_{i-1}^*, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}^*, \cdots, \boldsymbol{x}_s^*),$$

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

定理 3.15 设优化问题

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

目标函数下半连续，水平集有界，子问题

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

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

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

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

$$\text{坐标轮换极小值} \Longleftrightarrow \text{稳定点}$$

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

$$\begin{aligned} f(x,y,z) = &-xy-yz-zx+[x-1]_+^2+[-x-1]_+^2+[y-1]_+^2 \\ &+[-y-1]_+^2+[z-1]_+^2+[-z-1]_+^2. \end{aligned}$$

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

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

$$f_x'(x;y,z)=\begin{cases} 2(x+1)-(y+z) & x<-1 \\ -(y+z) & -1<x<1 \\ 2(x-1)-(y+z) & x>1 \end{cases}$$

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

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

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

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

3.3.1 和式凸优化交替极小化方法

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

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

$g(\boldsymbol{x})=\sum_{i=1}^s g_i(\boldsymbol{x}_i)$, $\quad g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 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

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

$f:\mathbb{R}^n\to\mathbb{R}$ 连续可微的凸函数
$\nabla_i f(x)$ Lipschitz 连续，常数为 $L_i, i=1,2$
$g_i:\mathbb{R}^{n_i}\to\mathbb{R}$ 下半连续的凸函数

注 二分块交替极小化算法

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

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

- 迭代步：依次计算

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

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 $\boldsymbol{x}\in\mathcal{K}$ 和 $\lambda\leqslant 0$，都有 $\lambda\boldsymbol{x}\in\mathcal{K}$。

并不是所有的锥都是凸的

定义 3.11 (极锥) 与原锥对立的锥。

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

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，$\boldsymbol{x}\in\Omega$。

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

定义为凸集 Ω 在该点的切锥.

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

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

$$\mathcal{K} = \left\{ \sum_{i=1}^{m} \mu_i \boldsymbol{b}_i \middle| \mu_i \geqslant 0, i = 1, 2, \cdots, m \right\},$$

$\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_m$ 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

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

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

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

为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 的一个凸组合.

定义 3.16 (凸包) $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 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.

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

有界的凸多面体就是多面胞

定义 3.18 (回收方向) $\boldsymbol{d}$ 满足

$$\forall \boldsymbol{x} \in \mathcal{S}, \alpha \geqslant 0 \Longrightarrow x + \alpha \boldsymbol{d} \in \mathcal{S}$$

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

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

$$\mathcal{P} = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = b\}$$

$$\mathcal{K} = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{0}\}$$

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

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

为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m$ 的一个仿射组合.

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

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

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

$$\boldsymbol{p}^{\mathrm{T}}\boldsymbol{y} > \alpha, \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \leqslant \alpha, \forall \boldsymbol{x} \in \mathcal{S}$$

直观意义：若点 $\boldsymbol{y}$ 不属于闭凸集 $\mathcal{S}$，则存在超平面

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

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 $\boldsymbol{b}$ 可以表示为向量组 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m$ 的非负线性组合的充要条件是对任意满足 $\boldsymbol{a}_i^{\mathrm{T}} \geqslant 0$ 的 $\boldsymbol{x}$, 都有 $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} \geqslant 0$

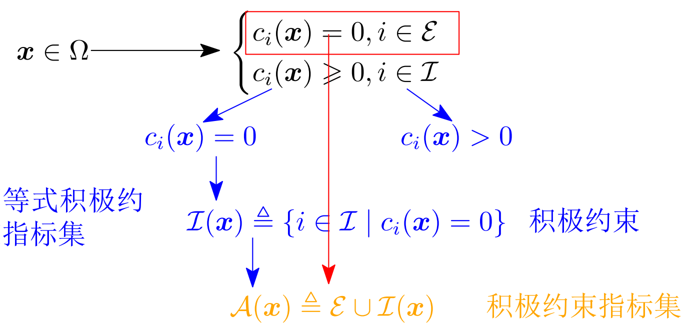
定理 3.21 (Farkas 引理') 对于 $\boldsymbol{A} \in \mathbb{R}^{n\times m}, \boldsymbol{b} \in \mathbb{R}^n$ 下述两系统恰有一个解

1. $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \geqslant \mathbf{0}, \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} < 0$;
2. $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}, \boldsymbol{y} \geqslant 0$.

4 约束优化最优性条件

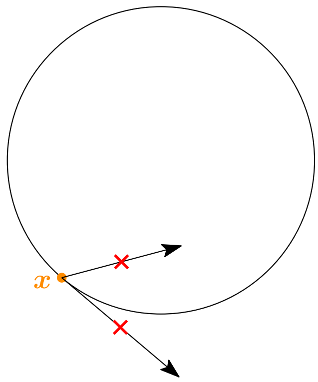
$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{x} \in \Omega \end{aligned}$$

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



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

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



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

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

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

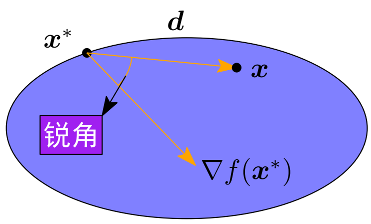
特别提醒：上述条件不是充分的，除非约束函数是线性的。

定理 4.1 (可行下降方向) 若 $\boldsymbol{d}$ 为 $\boldsymbol{x}\in\Omega$ 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 $\boldsymbol{x}^*$ 为约束优化问题的稳定点，即满足

$$\langle\nabla f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle\geqslant 0,\quad \forall\boldsymbol{x}\in\Omega.$$



4.1 等式优化最优化条件

例 4.1 设 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\in\mathbb{R}^n$ 满足 $\sum\limits_{i=1}^m\boldsymbol{a}_i\neq\boldsymbol{0}$ 求解优化问题★★★★★

$$\min\{\sum_{i=1}^m\|\boldsymbol{x}-\boldsymbol{a}_i\|^2\mid\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=1\}$$

解 令

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

对 x_k 求偏导，得到

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

故而有

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

所以，有

$$
\begin{cases}
\sum\limits_{i=1}^{m} 2(\boldsymbol{x}-\boldsymbol{a}_i) = 2\lambda\boldsymbol{x} \\
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1
\end{cases}
$$

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

若 $\sum\limits_{i=1}^{m}\boldsymbol{a}_i \neq \boldsymbol{0}$，则

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

相应地，最优值为

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

4.1.1 不等式优化约束条件

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

定理 4.3 (完整 KKT 条件)

$$\begin{cases} \nabla f(\boldsymbol{x}^*) = \sum\limits_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i^* \nabla c_i(\boldsymbol{x}^*), \\ \lambda_i^* \geqslant 0,\ c_i(\boldsymbol{x}^*) \geqslant 0,\ \lambda_i^* c_i(\boldsymbol{x}^*) = 0,\ i\in\mathcal{I}, \\ c_i\big(\boldsymbol{x}^*\big) = 0,\ i\in\mathcal{E}. \end{cases}$$

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：

$$\begin{aligned} &\min f(\boldsymbol{x}) && \Longrightarrow && \min f(\boldsymbol{x}) \\ &\mathrm{s.t.}\ c_i(\boldsymbol{x}) = 0,\ i\in\mathcal{E} && \Longrightarrow && \boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{0} \\ &\mathrm{s.t.}\ c_i(\boldsymbol{x}) \geqslant 0,\ i\in\mathcal{I} && \Longrightarrow && \boldsymbol{G}(\boldsymbol{x}) \geqslant \boldsymbol{0} \end{aligned}$$

Lagrange 函数

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

对偶问题：

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

例 4.2 线性规划的对偶⭐⭐⭐⭐⭐

$$\begin{aligned} \min\quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.}\quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geqslant \boldsymbol{0} \end{aligned}$$

对偶

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

$$\begin{aligned} L(\boldsymbol{x};\boldsymbol{u},\boldsymbol{v}) &= \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= \left(\boldsymbol{c} - \boldsymbol{u} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\right)^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{b}, \quad \boldsymbol{u} \geqslant \boldsymbol{0}. \end{aligned}$$

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x};\boldsymbol{u},\boldsymbol{v}) &= \min_{\boldsymbol{x}\in\mathbb{R}^n} \left(\boldsymbol{c} - \boldsymbol{u} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v}\right)^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{b} \\ &= \begin{cases} -\infty, & \boldsymbol{c} - \boldsymbol{u} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} \neq \boldsymbol{0}, \\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{b}, & \boldsymbol{c} - \boldsymbol{u} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{0}, \end{cases} \end{aligned}$$

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

化简得到

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

4.3.2 对偶定理

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

$$
f(\boldsymbol{x}_0) \geqslant \theta(\boldsymbol{u_0}, \boldsymbol{v}_0)
$$

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

$$
f(\boldsymbol{x}_0) = \theta(\boldsymbol{u_0}, \boldsymbol{v}_0)
$$

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：

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

- 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★☆☆☆☆

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

其中 $\boldsymbol{G}$ 非半正定。

解 Lagrange 函数

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

Lagrange 对偶

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

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

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

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

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

内层优化最优解，原问题可行解

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

得到 $\boldsymbol{x}^*$，之后有

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

解得

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

$(\boldsymbol{x}^*, \lambda)$ 为鞍点，强对偶定理成立

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

4.3.3 混合约束优化的对偶

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

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 $\max\limits_{\lambda \geqslant 0} \min\limits_{x \geqslant 0} L(\lambda, \boldsymbol{x})$

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

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⭐⭐⭐⭐⭐

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

解 令 $\boldsymbol{y} = \boldsymbol{x}$ 得

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

对偶

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

变量可分离

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

对于 $\min\limits_{\boldsymbol{y} \in \mathbb{R}^n} \left(\lambda \|\boldsymbol{y}\|_1 - \boldsymbol{z}^{\mathrm{T}} \boldsymbol{y} \right)$ 有

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

综上

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

- 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

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

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

目标函数二次, 二阶连续可微

$$\nabla\boldsymbol{Q}(\boldsymbol{x})=\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{g}\quad\nabla^2\boldsymbol{Q}(\boldsymbol{x})=\boldsymbol{Q}$$

约束线性

$$\Omega=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}=b_i,i\in\mathcal{E};\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}\geqslant b_i,i\in\mathcal{I}\}$$

定理 4.5 (Frank-Wolfe 定理,1956)

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

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

$$\begin{array}{ll} \min & Q(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x}+\boldsymbol{g}^{\mathrm{T}}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}=b_i,\ i\in\mathcal{E} \\ & \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}\geqslant b_i,\ i\in\mathcal{I} \end{array}$$

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

Lagrange 函数

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

Lagrange 对偶

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

由于矩阵 $\boldsymbol{G}$ 对称正定，那么

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

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

有

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

代入得到

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

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

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

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

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

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

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

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

即

$$\operatorname{rank}(\boldsymbol{G}) = \operatorname{rank}(\boldsymbol{G}, \boldsymbol{g})$$

证明.

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

证毕！

例 4.6 求解以下无约束二次规划★★★★☆

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

解

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

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

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

注 等式约束二次规划的 Lagrange 方法

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

原理：通过 KKT 条件获取最优解

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

二次规划问题

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

K-T 条件

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

转化为

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

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

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

非奇异。

例 4.7 用 Lagrange 法求解如下问题★★★★★

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

解 由目标函数及约束条件

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

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

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

4.5 有效集方法

例 4.8 求解二次规划问题★★★★★

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

解 **第一步迭代：** 取初始点

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

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

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

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

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

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

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

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

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

从 $\mathcal{I}/S_1 = \{1, 3\}$ 中取

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

不等式积极约束 $i=1$ 令

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

进入第三次迭代：

求解子问题

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

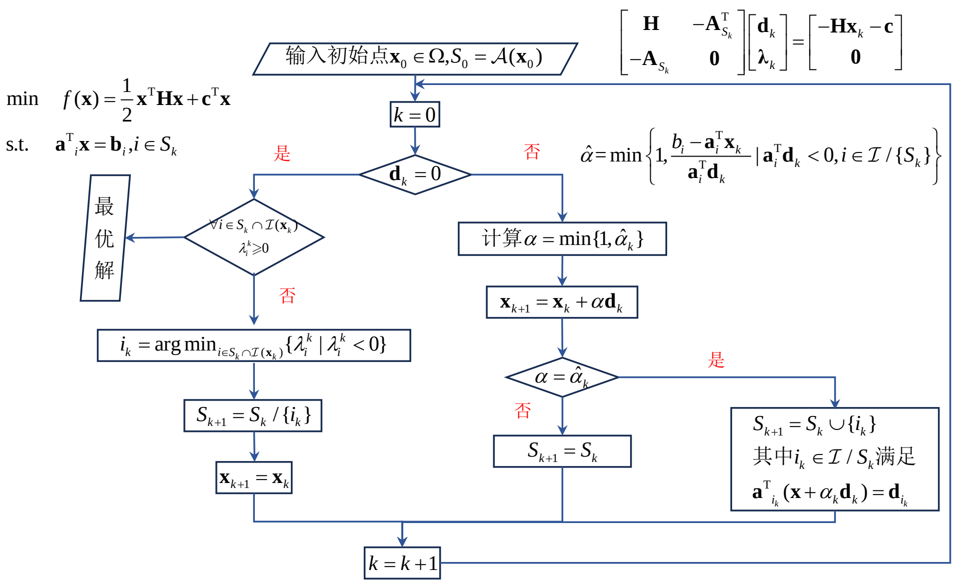
得最优解

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

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

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

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:

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

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

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

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

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

计算最大可行步长

$$\boldsymbol{a}_i^{\mathrm{T}}(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\geqslant b_i,\quad \alpha\geq 0, i\in\mathcal{I}/\mathcal{I}(\boldsymbol{x}_k^*)$$

$$\hat{\alpha}_k=\min_{i\in\mathcal{I}\backslash\mathcal{I}(\boldsymbol{x})}\left\{\frac{b_i-\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}_k}{\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}_k}|\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}_k<0\right\}$$

最优步长

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

4.7 投影算子

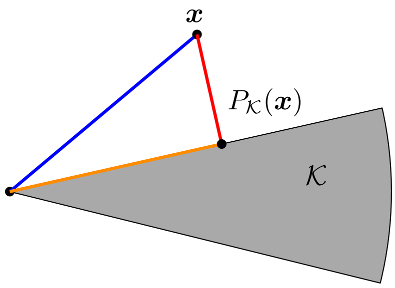
定义 4.4 (投影算子) 设 Ω 为非空闭凸集. 对任意 $\boldsymbol{x}\in\mathbb{R}^n$, 定义

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

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

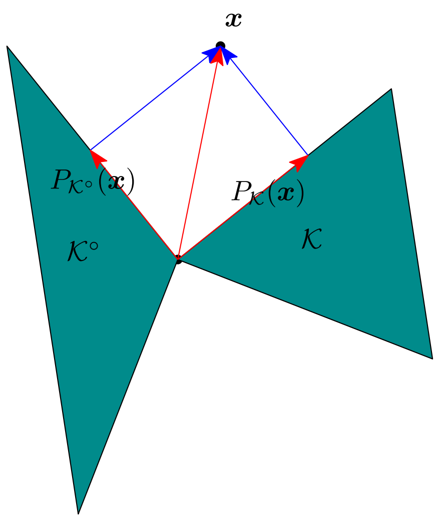
定理 4.9 对闭凸锥 $\mathcal{K}$ 和任意 $\boldsymbol{x} \in \mathbb{R}^n$

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



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

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

残量函数 $r(x) = \boldsymbol{x} - P_\Omega(\boldsymbol{x} - \nabla_f(\boldsymbol{x}))$ 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。

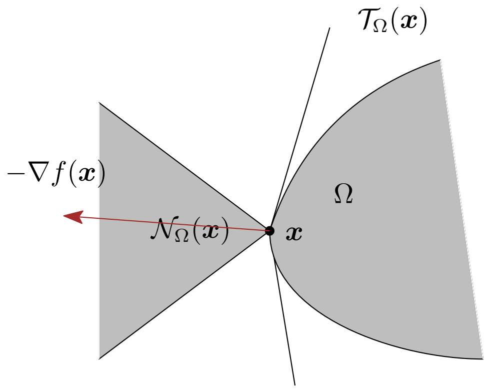
定义 4.5 $\boldsymbol{x} \in \Omega$. 称 $\mathcal{T}_\Omega(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \parallel$ 存在$\boldsymbol{d}_k \to \boldsymbol{d}, t_k \to 0^+$, 使得 $\boldsymbol{x} + t_k \boldsymbol{d}_k \in \Omega\}$ 为闭凸集 Ω 在 $\boldsymbol{x}$ 点的切锥。

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

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

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

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 $\boldsymbol{A}$ 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：$\boldsymbol{A} \in \mathbb{R}^{m\times n}$。若存在矩阵 $\boldsymbol{G} \in \mathbb{R}^{n\times m}$ 使得

$$\boldsymbol{AGA} = \boldsymbol{A}, \quad \boldsymbol{GAG} = \boldsymbol{G},$$
$$(\boldsymbol{AG})^{\mathrm{T}} = \boldsymbol{AG}, \quad (\boldsymbol{GA})^{\mathrm{T}} = \boldsymbol{GA}$$

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

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

$$\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\,\boldsymbol{A} = \boldsymbol{A}$$

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

$$\boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A} = \boldsymbol{A}$$

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

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

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j \end{bmatrix}_{m\times m} = \boldsymbol{I}_{m\times m}$$

有

$$\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A} = \boldsymbol{I}_{m\times m}\boldsymbol{A} = \boldsymbol{A}$$

4.8 约束优化投影算法

凸约束优化 $\min\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \Omega\}$ 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 $\boldsymbol{x}_{k+1} = P_\Omega(\boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k))$.

5 其他方法

5.1 罚函数方法

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

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

$$\min\{f(\boldsymbol{x}) \mid c_i(\boldsymbol{x}) = 0, i \in \mathcal{E}\}$$

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} P(\boldsymbol{x}, \pi) = f(\boldsymbol{x}) + \pi \sum_{i \in \mathcal{E}} c_i^2(\boldsymbol{x})$$

5.2 乘子罚函数

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

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

原规划问题在最优值点 $\boldsymbol{x}^*$ 满足二阶最优性条件。

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

则存在 $\pi^*>0$, 使对任意的 $\pi\geqslant\pi^*,\boldsymbol{x}^*$ 为罚函数 $P(\boldsymbol{x},\boldsymbol{\lambda}^*,\pi)$ 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

$$\boldsymbol{T}=(t_{i_1 i_2\cdots i_d})$$