

最优化笔记

Notes of Optimization

Zebediah

参考刘浩洋、户将、李勇锋、文再文编著的[《最优化计算方法》](#)

2024 年 7 月 23 日

目录

1	基础知识	1
1.1	范数	1
1.2	导数	2
1.2.1	梯度	2
1.2.2	矩阵变量函数的导数	3
1.3	广义实值函数	4
1.3.1	适当函数	4
1.3.2	闭函数	4
1.4	凸集	5
1.4.1	凸集的相关定义	5
1.4.2	分离超平面定理	7
1.5	凸函数	7
1.5.1	凸函数的定义	7
1.5.2	凸函数判定定理	8
1.5.3	凸函数的性质	9
1.6	共轭函数	9
1.7	次梯度	10
1.7.1	次梯度的定义和性质	10
1.7.2	次梯度的计算规则	11
2	典型优化问题	12
2.1	线性规划	12
2.2	最小二乘问题	12
2.3	复合优化问题	13
2.4	随机优化问题	13
2.5	半定规划	14
2.6	矩阵优化	14
2.7	优化模型语言	15
3	最优性理论	16
3.1	最优化问题解的存在性	16
3.2	无约束可微问题的最优性理论	16

3.3	无约束不可微问题的最优性理论	17
3.4	对偶理论	18
3.4.1	拉格朗日函数与对偶问题	18
3.4.2	带广义不等式约束优化问题的对偶	19
3.5	一般约束优化问题的最优性理论	20
3.5.1	一阶最优性条件	20
3.5.2	二阶最优性条件	23
3.6	带约束凸优化问题的最优性理论	25
3.6.1	Slater 约束品性与强对偶原理	25
3.6.2	一阶充要条件	25
4	无约束优化算法	27
4.1	线搜索方法	27
4.1.1	线搜索准则	27
4.1.2	线搜索算法	30
4.1.3	收敛性	31
4.2	梯度类算法	31
4.2.1	梯度下降法	31
4.2.2	Barzilai-Borwein 方法	32
4.2.3	Tikhonov 正则化模型求解	33
4.3	次梯度算法	34
4.3.1	次梯度算法结构	34
4.3.2	收敛性	34
4.4	牛顿类算法	36
4.4.1	经典牛顿法	36
4.4.2	收敛性	36
4.4.3	修正牛顿法	37
4.4.4	非精确牛顿法	37
4.5	拟牛顿类算法	38
4.5.1	割线方程	38
4.5.2	拟牛顿矩阵更新方式	39
4.5.3	拟牛顿法的全局收敛性	40
4.5.4	有限内存 BFGS 方法	40
4.6	信赖域算法	42
4.6.1	信赖域算法框架	42
4.6.2	信赖域子问题求解	44

Chapter 1

基础知识

1.1 范数

定义 1.1 (ℓ_p 范数) 向量范数为 ℓ_p 范数 ($p \geq 1$) :

$$\|v\|_p = (|v_1|^p + |v_2|^p + \cdots + |v_n|^p)^{\frac{1}{p}}.$$

当 $p = \infty$ 时, ℓ_∞ 范数定义为

$$\|v\|_\infty = \max_i |v_i|.$$

命题 1.1 (Cauchy 不等式) 设 $a, b \in \mathbb{R}^n$, 则

$$|a^T b| \leq \|a\|_2 \|b\|_2,$$

等号成立当且仅当 a 与 b 线性相关.

定义 1.2 (矩阵的 ℓ_p 范数) 这里介绍常用的 ℓ_1 和 ℓ_2 范数. 矩阵 $A \in \mathbb{R}^{m \times n}$ 的 ℓ_1 范数定义为

$$\|A\|_1 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|,$$

ℓ_2 范数也称为 Frobenius 范数 (F 范数), 定义为

$$\|A\|_F = \sqrt{\text{Tr}(AA^T)} = \sqrt{\sum_{i,j} a_{ij}^2}. \quad (1.1)$$

注 矩阵的 F 范数具有正交不变性, 即对于任意的正交矩阵 $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$, 我们有

$$\begin{aligned} \|UAV\|_F^2 &= \text{Tr}(UAVV^T A^T U^T) = \text{Tr}(UAA^T U^T) \\ &= \text{Tr}(AA^T U^T U) = \text{Tr}(AA^T) = \|A\|_F^2, \end{aligned}$$

定义 1.3 (算子范数) 给定矩阵 $A \in \mathbb{R}^{m \times n}$, 以及 m 维和 n 维空间的向量范数 $\|\cdot\|_{(m)}$ 和 $\|\cdot\|_{(n)}$, 其诱

导的矩阵范数定义如下:

$$\|A\|_{(m,n)} = \max_{x \in \mathbb{R}^n, \|x\|_{(n)}=1} \|Ax\|_{(m)},$$

如果将 $\|\cdot\|_{(m)}$ 和 $\|\cdot\|_{(n)}$ 都取为相应向量空间的 ℓ_p 范数, 我们可以得到矩阵的 p 范数. 常用的是矩阵的 2 范数, 即

$$\|A\|_2 = \max_{x \in \mathbb{R}^n, \|x\|_2=1} \|Ax\|_2. \quad (1.2)$$

容易验证, 矩阵的 2 范数是该矩阵的最大奇异值 (利用 SVD 和正交矩阵的性质即可).

性质 1.1 (矩阵范数的相容性) 记号同定义 1.3, 所有算子范数都满足如下性质:

$$\|Ax\|_{(m)} \leq \|A\|_{(m,n)} \|x\|_{(n)}.$$

定义 1.4 (核范数) 给定矩阵 $A \in \mathbb{R}^{m \times n}$, 其核范数定义为

$$\|A\|_* = \sum_{i=1}^r \sigma_i,$$

其中 $\sigma_i, i = 1, 2, \dots, r$ 为 A 的所有非零奇异值, $r = \text{rank}(A)$.

定义 1.5 (Frobenius 内积) 对于矩阵空间 $\mathbb{R}^{m \times n}$ 的两个矩阵 A 和 B , 它们的 Frobenius 内积定义为

$$\langle A, B \rangle \stackrel{\text{def}}{=} \text{Tr}(AB^T) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}.$$

命题 1.2 (矩阵范数的 Cauchy 不等式) 设 $A, B \in \mathbb{R}^{m \times n}$, 则

$$|\langle A, B \rangle| \leq \|A\|_F \|B\|_F,$$

等号成立当且仅当 A 和 B 线性相关.

1.2 导数

1.2.1 梯度

定义 1.6 (梯度) 给定函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 且 f 在点 x 的一个邻域内有意义, 若存在向量 $g \in \mathbb{R}^n$ 满足

$$\lim_{p \rightarrow 0} \frac{f(x+p) - f(x) - g^T p}{\|p\|} = 0, \quad (1.3)$$

就称 f 在点 x 处可微 (或 Fréchet 可微). 此时 g 称为 f 在点 x 处的梯度, 记作 $\nabla f(x)$. 如果对区域 D 上的每一个点 x 都有 $\nabla f(x)$ 存在, 则称 f 在 D 上可微, 且

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

Hessian 矩阵和 Jacobi 矩阵在数学分析中已经定义过. 不难发现, 梯度 $\nabla f(x)$ 的 Jacobi 矩阵就是 $f(x)$ 的 Hessian 矩阵.

定理 1.1 (Taylor 展开) 设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是连续可微的, $p \in \mathbb{R}^n$ 为向量, 那么

$$f(x+p) = f(x) + \nabla f(x)^T p,$$

其中 $0 < t < 1$. 进一步地, 如果 f 是二阶连续可微的, 则

$$\begin{aligned} \nabla f(x+p) &= \nabla f(x) + \int_0^1 \nabla^2 f(x+tp) p \, dt, \\ f(x+p) &= f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p, \end{aligned}$$

其中 $0 < t < 1$.

定义 1.7 (梯度利普希茨连续) 给定可微函数 f , 若存在 $L > 0$, 对任意的 $x, y \in \text{dom } f$ 有

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

则称 f 是梯度利普希茨连续的, 相应利普希茨常数为 L . 有时也简记为梯度 L -利普希茨连续或 L -光滑.

引理 1.1 (二次上界) 设可微函数 $f(x)$ 的定义域 $\text{dom } f = \mathbb{R}^n$, 且为梯度 L -利普希茨连续的, 则函数 $f(x)$ 有二次上界:

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|^2, \quad \forall x, y \in \text{dom } f. \quad (1.4)$$

证明思路. 对任意的 $x, y \in \mathbb{R}^n$, 构造辅助函数 $g(t) = f(x + t(y - x))$, $t \in [0, 1]$. □

注 引理 1.1 指出 $f(x)$ 可被一个二次函数上界所控制, 即要求 $f(x)$ 的增长速度不超过二次. 实际上, 该引理对 $f(x)$ 定义域的要求可减弱为 $\text{dom } f$ 是凸集, 此条件的作用是保证证明中的 $g(t)$ 当 $t \in [0, 1]$ 时是有定义的.

推论 1.1 设可微函数 $f(x)$ 的定义域为 \mathbb{R}^n 且存在一个全局极小点 x^* , 若 $f(x)$ 为梯度 L -利普希茨连续的, 则对任意的 x 有

$$\frac{1}{2L} \|\nabla f(x)\|^2 \leq f(x) - f(x^*). \quad (1.5)$$

证明思路. 利用引理 1.1 和二次函数的性质. □

注 推论 1.1 说明, 若 f 是梯度利普希茨连续的, 且有一个全局极小点 x^* , 我们能够利用二次上界 (1.4) 来估计 $f(x) - f(x^*)$ 的大小.

1.2.2 矩阵变量函数的导数

定义 1.8 (Gâteaux 可微) 设 $f(X)$ 为矩阵变量函数, 如果存在矩阵 $G \in \mathbb{R}^{m \times n}$, 对任意方向 $V \in \mathbb{R}^{m \times n}$ 满足

$$\lim_{t \rightarrow 0} \frac{f(X + tV) - f(X) - t\langle G, V \rangle}{t} = 0, \quad (1.6)$$

则称 f 关于 X 是 Gâteaux 可微的. 满足 (1.6) 式的 G 称为 f 在 X 处在 Gâteaux 可微意义下的梯度.

可以证明, 当 f 是 Fréchet 可微函数时, f 也是 Gâteaux 可微的, 且这两种意义下的梯度相等. 在实际中, 由于 Gâteaux 可微定义式更容易操作, 因此通常是利用 (1.6) 式进行矩阵变量函数 $f(X)$ 的求导运算.

例 1.1 (1) 考虑线性函数: $f(X) = \text{Tr}(AX^T B)$, 其中 $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{m \times p}$, $X \in \mathbb{R}^{m \times n}$, 对任意方向 $V \in \mathbb{R}^{m \times n}$ 以及 $t \in \mathbb{R}$, 有

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{f(X + tV) - f(X)}{t} &= \lim_{t \rightarrow 0} \frac{\text{Tr}(A(X + tV)^T B) - \text{Tr}(AX^T B)}{t} \\ &= \text{Tr}(AV^T B) = \langle BA, V \rangle, \end{aligned}$$

其中最后一个等号用到了 Frobenius 内积的定义. 因此, $\nabla f(X) = BA$.

(2) 考虑二次函数: $f(X, Y) = \frac{1}{2} \|XY - A\|_F^2$, 其中 $(X, Y) \in \mathbb{R}^{m \times p} \times \mathbb{R}^{p \times n}$, 对变量 Y , 其中 $X \in \mathbb{R}^{m \times p}$, $Y \in \mathbb{R}^{p \times n}$, 取任意方向 V 以及充分小的 $t \in \mathbb{R}$, 有

$$\begin{aligned} f(X, Y + tV) - f(X, Y) &= \frac{1}{2} \|X(Y + tV) - A\|_F^2 - \frac{1}{2} \|XY - A\|_F^2 \\ &= \langle tXV, XY - A \rangle + \frac{1}{2} t^2 \|XV\|_F^2 \\ &= t \langle V, X^T(XY - A) \rangle + \mathcal{O}(t^2). \end{aligned}$$

由定义可知 $\frac{\partial f}{\partial Y} = X^T(XY - A)$. 同理也有 $\frac{\partial f}{\partial X} = (XY - A)Y^T$.

注 在对函数求导的过程中, 应当注意函数的自变量和相应的导数应该有相同的维数.

1.3 广义实值函数

定义 1.9 (广义实值函数) 令 $\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间, 则映射 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 称为广义实值函数.

1.3.1 适当函数

定义 1.10 (适当函数) 给定广义实值函数 f 和非空集合 \mathcal{X} . 如果存在 $x \in \mathcal{X}$ 使得 $f(x) < +\infty$, 并且对任意的 $x \in \mathcal{X}$, 都有 $f(x) > -\infty$, 那么称函数 f 关于集合 \mathcal{X} 是适当的.

注 适当函数 f 的特点是“至少有一处取值不为正无穷”, 以及“处处取值不为负无穷”. 对最优化问题 $\min_x f(x)$, 适当函数可以帮助去掉一些我们不感兴趣的函数, 从而在一个比较合理的函数类中考虑最优化问题. 今后若无特殊说明, 所讨论的函数均为适当函数. 此外, 对于适当函数 f , 规定其定义域

$$\text{dom } f = \{x \mid f(x) < +\infty\}.$$

1.3.2 闭函数

定义 1.11 (α -下水平集) 对于广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$,

$$C_\alpha = \{x \mid f(x) \leq \alpha\}$$

称为 f 的 α -下水平集.

定义 1.12 (上方图) 对于广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$,

$$\text{epi } f = \{(x, t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$$

称为 f 的上方图.

上方图的一个直观的例子如图 1.1 所示.

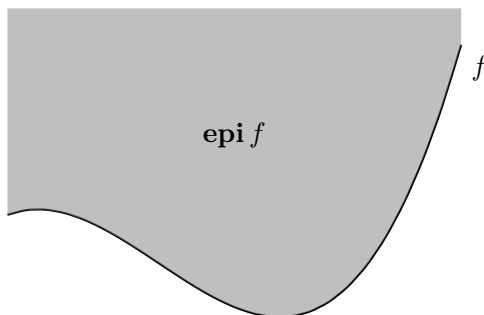


图 1.1: 函数 f 和其上方图 $\text{epi } f$

定义 1.13 (闭函数) 设 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 为广义实值函数, 若 $\text{epi } f$ 为闭集, 则称 f 为闭函数.

定义 1.14 (下半连续函数) 设广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 若对任意的 $x \in \mathbb{R}^n$, 有

$$\liminf_{y \rightarrow x} f(y) \geq f(x),$$

则 $f(x)$ 为下半连续函数.

定理 1.2 (闭函数和下半连续函数的等价性) 设广义实值函数 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 则以下命题等价:

- (1) $f(x)$ 的任意 α -下水平集都是闭集;
- (2) $f(x)$ 是下半连续的;
- (3) $f(x)$ 是闭函数.

1.4 凸集

1.4.1 凸集的相关定义

定义 1.15 (仿射集) 如果过集合 C 中任意两点的直线都在 C 内, 则称 C 为仿射集, 即

$$x_1, x_2 \in C \implies \theta x_1 + (1 - \theta)x_2 \in C, \forall \theta \in \mathbb{R}.$$

线性方程组 $Ax = b$ 的解集是仿射集. 反之, 任何仿射集都可以表示成一个线性方程组的解集.

定义 1.16 (凸集) 如果连接集合 C 中任意两点的线段都在 C 内, 则称 C 为凸集, 即

$$x_1, x_2 \in C \implies \theta x_1 + (1 - \theta)x_2 \in C, \forall 0 \leq \theta \leq 1.$$

定义 1.17 (凸组合和凸包) 形如

$$\begin{aligned} x &= \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k \\ 1 &= \theta_1 + \theta_2 + \cdots + \theta_k, \quad \theta_i \geq 0, i = 1, 2, \cdots, k \end{aligned}$$

的点称为 x_1, x_2, \cdots, x_k 的组合. 集合 S 中点所有可能的凸组合构成的集合称作 S 的凸包, 记作 $\text{conv } S$. $\text{conv } S$ 是包含 S 的最小的凸集.

定义 1.18 (仿射包) 设 S 为 \mathbb{R}^n 的子集, 称如下集合为 S 的仿射包:

$$\{x \mid x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k, \quad x_1, x_2, \cdots, x_k \in S, \quad \theta_1 + \theta_2 + \cdots + \theta_k = 1\},$$

记为 $\text{affine } S$. $\text{affine } S$ 是包含 S 的最小的仿射集.

定义 1.19 (锥组合和凸锥) 形如

$$x = \theta_1 x_1 + \theta_2 x_2, \quad \theta_1 > 0, \theta_2 > 0$$

的点称为点 x_1, x_2 的锥组合. 若集合 S 中任意点的锥组合都在 S 中, 则称 S 为凸锥.

下面是一些重要的凸集:

1. 超平面和半空间

任取非零向量 a , 形如 $\{x \mid a^T x = b\}$ 的集合称为超平面, 形如 $\{x \mid a^T x \leq b\}$ 的集合称为半空间. 超平面是仿射集和凸集, 半空间是凸集但不是仿射集.

2. 球和锥

令 $\|\cdot\|$ 是任意一个范数, $\{x \mid \|x - x_c\| \leq r\}$ 称为中心为 x_c , 半径为 r 的范数球. $\{(x, t) \mid \|x\| \leq t\}$ 称为范数锥. 欧几里得范数锥也称为二次锥. 范数球和范数锥都是凸集.

3. 多面体

满足线性等式和不等式组的点的集合称为多面体, 即 $\{x \mid Ax \leq b, Cx = d\}$, 其中 $A \in \mathbb{R}^{m \times n}, C \in \mathbb{R}^{p \times n}$, $x \leq y$ 表示向量 x 的每个分量均小于等于 y 的对应分量. 多面体是有限个半空间和超平面的交集, 因此是凸集.

4. (半) 正定锥

记 \mathcal{S}^n 为 $n \times n$ 对称矩阵的集合, $\mathcal{S}_+^n = \{X \in \mathcal{S}^n \mid X \succeq 0\}$ 为 $n \times n$ 半正定矩阵的集合, $\mathcal{S}_{++}^n = \{X \in \mathcal{S}^n \mid X \succ 0\}$ 为 $n \times n$ 正定矩阵的集合. 容易证明 \mathcal{S}_+^n 是凸锥, 因此 \mathcal{S}_+^n 又称为半正定锥.

下面介绍证明一个集合 (设为 C) 为凸集的两种方式. 第一种是利用定义

$$x_1, x_2 \in C, 0 \leq \theta \leq 1 \implies \theta x_1 + (1 - \theta)x_2 \in C$$

来证明集合 C 是凸集. 第二种方法是说明集合 C 可由简单的凸集 (超平面、半空间、范数球等) 经过保凸的运算后得到. 为此, 我们需要掌握一些常见的保凸运算, 如缩放、平移和投影变换 (它们都是仿射变换). 此外, 任意多个凸集的交以及仿射变换都是保凸的.

1.4.2 分离超平面定理

凸集一个重要的性质是可以用超平面分离不相交的凸集.

定理 1.3 (分离超平面定理) 如果 C 和 D 是不相交的两个凸集, 则存在非零向量 a 和常数 b , 使得

$$a^T x \leq b, \quad \forall x \in C \quad \text{且} \quad a^T x \geq b, \quad \forall x \in D,$$

即超平面 $\{x \mid a^T x = b\}$ 分离了 C 和 D (如图 1.2).

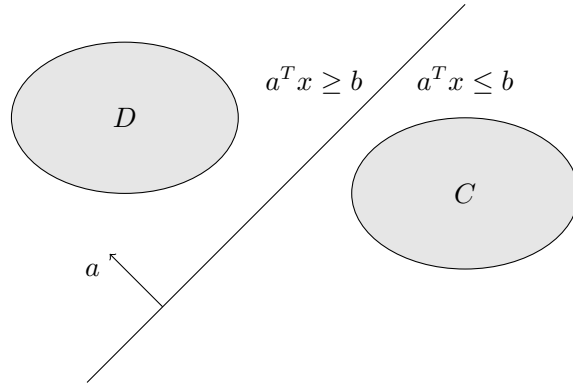


图 1.2: 分离超平面

定理 1.4 (严格分离定理) 设 C 是闭凸集, 点 $x_0 \notin C$, 则存在非零向量 a 和常数 b , 使得

$$a^T x < b, \quad \forall x \in C, \quad \text{且} \quad a^T x_0 > b.$$

定义 1.20 (支撑超平面) 给定集合 C 及其边界上一点 x_0 , 如果 $a \neq 0$ 满足 $a^T x \leq a^T x_0, \forall x \in C$, 那么称集合

$$\{x \mid a^T x = a^T x_0\}$$

为 C 在边界点 x_0 处的支撑超平面.

定理 1.5 (支撑超平面定理) 如果 C 是凸集, 则在 C 的任意边界点处都存在支撑超平面.

注 支撑超平面定理有非常强的几何直观: 给定一个平面后, 可把凸集边界上的任意一点当成支撑点将凸集放置在该平面上. 其他形状的集合一般没有这个性质.

1.5 凸函数

1.5.1 凸函数的定义

定义 1.21 (凸函数) 设函数 f 为适当函数, 如果 $\text{dom } f$ 是凸集, 且

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

对所有 $x, y \in \text{dom } f, 0 \leq \theta \leq 1$ 都成立, 则称 f 是凸函数. 类似地也可以定义凹函数.

定义 1.22 (强凸函数) 若存在常数 $m > 0$, 使得

$$g(x) = f(x) - \frac{m}{2}\|x\|^2$$

为凸函数, 则称 $f(x)$ 为强凸函数, 其中 m 为强凸参数. 为了方便我们也称 $f(x)$ 为 m -强凸函数.

定义 1.23 (强凸函数的等价定义) 若存在常数 $m > 0$, 使得对任意 $x, y \in \mathbf{dom} f$ 以及 $\theta \in (0, 1)$, 有

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) - \frac{m}{2}\theta(1 - \theta)\|x - y\|^2,$$

则称 $f(x)$ 为强凸函数, 其中 m 为强凸参数.

命题 1.3 设 f 为强凸函数且存在最小值, 则 f 的最小值点唯一.

注 f 存在最小值是前提. 强凸函数 f 的全局极小点不一定存在, 例如 $f(x) = x^2, \mathbf{dom} f = (1, 2)$.

1.5.2 凸函数判定定理

一个函数是凸函数当且仅当将函数限制在任意直线在定义域内的部分上时仍是凸的.

定理 1.6 $f(x)$ 是凸函数当且仅当对任意的 $x \in \mathbf{dom} f, v \in \mathbb{R}^n, g: \mathbb{R} \rightarrow \mathbb{R}$,

$$g(t) = f(x + tv), \quad \mathbf{dom} g = \{t \mid x + tv \in \mathbf{dom} f\}$$

是凸函数.

定理 1.7 (一阶条件) 对于定义在凸集上的可微函数 f , f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^T(y - x), \quad \forall x, y \in \mathbf{dom} f.$$

定理 1.8 (梯度单调性) 设 f 可微, 则 f 为凸函数当且仅当 $\mathbf{dom} f$ 为凸集且 ∇f 为单调映射, 即

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

定理 1.9 (二阶条件) 设 f 为定义在凸集上的二阶连续可微函数, 则 f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbf{dom} f.$$

如果 $\nabla^2 f(x) \succ 0, \forall x \in \mathbf{dom} f$, 则 f 是严格凸函数.

定理 1.10 函数 $f(x)$ 为凸函数当且仅当其上方图 $\mathbf{epi} f$ 是凸集.

定理 1.11 (1) 若 f 是凸函数, 则 αf 是凸函数, 其中 $\alpha \geq 0$.

(2) 若 f_1, f_2 是凸函数, 则 $f_1 + f_2$ 是凸函数.

(3) 若 f 是凸函数, 则 $f(Ax + b)$ 是凸函数.

(4) 若 f_1, f_2, \dots, f_m 是凸函数, 则 $f(x) = \max \{f_1(x), f_2(x), \dots, f_m(x)\}$ 是凸函数.

(5) 若对每个 $y \in \mathcal{A}, f(x, y)$ 关于 x 是凸函数, 则 $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ 是凸函数.

(6) 给定函数 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 和 $h: \mathbb{R} \rightarrow \mathbb{R}$, 令 $f(x) = h(g(x))$. 若 g 是凸函数, h 是凸函数且单调不减, 那么 f 是凸函数; 若 g 是凹函数, h 是凸函数且单调不增, 那么 f 是凸函数.

(7) 若 $f(x, y)$ 关于 (x, y) 整体是凸函数, C 是凸集, 则 $g(x) = \inf_{y \in C} f(x, y)$ 是凸函数.

1.5.3 凸函数的性质

命题 1.4 设 $f(x)$ 是凸函数, 则 $f(x)$ 所有的 α -下水平集 C_α 为凸集.

引理 1.2 (二次下界) 设 $f(x)$ 是参数为 m 的可微强凸函数, 则如下不等式成立:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{m}{2} \|y - x\|^2, \quad \forall x, y \in \text{dom } f. \quad (1.7)$$

推论 1.2 设 f 为可微强凸函数, 则 f 的所有 α -下水平集有界.

1.6 共轭函数

定义 1.24 (共轭函数) 任一适当函数 f 的共轭函数定义为

$$f^*(y) = \sup_{x \in \text{dom } f} \{y^T x - f(x)\}.$$

如图 1.3 所示, 共轭函数 $f^*(y)$ 是线性函数 yx 和 $f(x)$ 之间的最大差值,

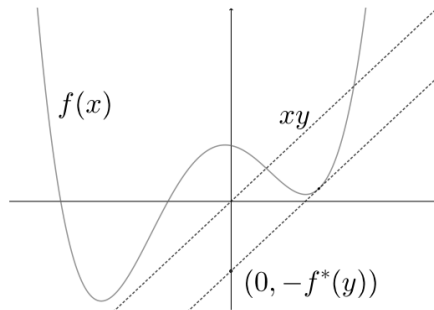


图 1.3: 共轭函数

命题 1.5 (Fenchel 不等式) $f(x) + f^*(y) \geq x^T y$, 这由共轭函数的定义可以直接得出.

定义 1.25 (二次共轭函数) 任一函数 f 的二次共轭函数定义为

$$f^{**}(x) = \sup_{y \in \text{dom } f^*} \{x^T y - f^*(y)\}.$$

显然 f^{**} 恒为闭凸函数, 且由 Fenchel 不等式可知

$$f^{**}(x) \leq f(x), \quad \forall x,$$

或等价地, $\text{epi } f \subseteq \text{epi } f^{**}$. 对于凸函数 f , 下面的定理描述了 f 的二次共轭函数与其自身的关系.

定理 1.12 若 f 为闭凸函数, 则

$$f^{**}(x) = f(x), \quad \forall x,$$

或等价地, $\text{epi } f = \text{epi } f^{**}$.

1.7 次梯度

1.7.1 次梯度的定义和性质

定义 1.26 (次梯度和次微分) 设 f 为适当凸函数, x 为定义域 $\text{dom } f$ 中的一点. 若向量 $g \in \mathbb{R}^n$ 满足

$$f(y) \geq f(x) + g^T(y - x), \quad \forall y \in \text{dom } f, \quad (1.8)$$

则称 g 为函数 f 在点 x 处的一个次梯度. 进一步地, 称集合

$$\partial f(x) = \{g \mid g \in \mathbb{R}^n, f(y) \geq f(x) + g^T(y - x), \forall y \in \text{dom } f\} \quad (1.9)$$

为 f 在点 x 处的次微分.

定义次梯度的初衷之一是希望它具有类似于梯度的一些性质. 我们在此指出, 无论是梯度法还是次梯度法, 本质上都在使用一阶泰勒展开式的原理去逼近在某点的原函数.

如图 1.4 所示, 对适当凸函数 $f(x)$, g_1 为点 x_1 处的唯一次梯度, 而 g_2, g_3 为点 x_2 处的两个不同的次梯度.

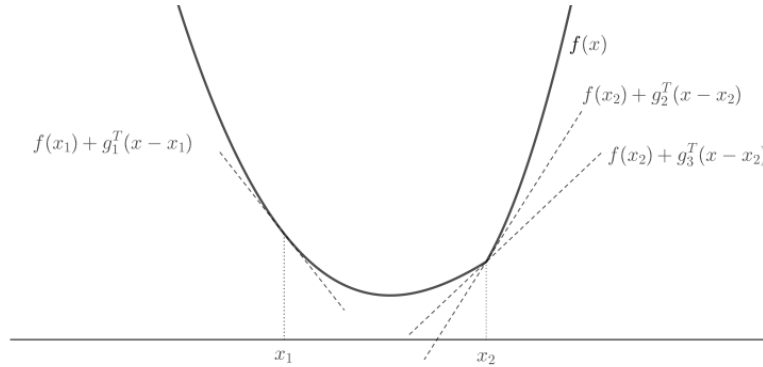


图 1.4: 函数 $f(x)$ 的次梯度

定理 1.13 设 f 是凸函数, 则次微分 $\partial f(x)$ 有如下性质:

- (1) 对任何 $x \in \text{dom } f$, $\partial f(x)$ 是一个闭凸集 (可能为空集);
- (2) 如果 $x \in \text{int dom } f$, 则 $\partial f(x)$ 非空有界集.

命题 1.6 设凸函数 $f(x)$ 在 $x_0 \in \text{int dom } f$ 处可微, 则 $\partial f(x_0) = \{\nabla f(x_0)\}$, 即当凸函数 $f(x)$ 在某点处可微时, $\nabla f(x)$ 就是 $f(x)$ 在该点处唯一的次梯度.

定理 1.14 (次梯度的单调性) 设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为凸函数, $x, y \in \mathbf{dom} f$, 则

$$(u - v)^T(x - y) \geq 0,$$

其中 $u \in \partial f(x), v \in \partial f(y)$.

1.7.2 次梯度的计算规则

(1) 可微凸函数: 设 f 为凸函数, 若 f 在点 x 处可微, 则 $\partial f(x) = \{\nabla f(x)\}$.

(2) 凸函数的非负线性组合: 设 f_1, f_2 为凸函数, 且满足 $\mathbf{int} \mathbf{dom} f_1 \cap \mathbf{dom} f_2 \neq \emptyset$, 而 $x \in \mathbf{dom} f_1 \cap \mathbf{dom} f_2$. 若 $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x), \alpha_1, \alpha_2 \geq 0$, 则 $f(x)$ 的次微分 $\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$.

(3) 线性变量替换: 设 h 为适当凸函数, 并且函数 f 满足

$$f(x) = h(Ax + b), \quad \forall x \in \mathbb{R}^m,$$

其中 $A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$. 若存在 $x^\# \in \mathbb{R}^m$, 使得 $Ax^\# + b \in \mathbf{int} \mathbf{dom} h$, 则

$$\partial f(x) = A^T \partial h(Ax + b), \quad \forall x \in \mathbf{int} \mathbf{dom} f.$$

定理 1.15 (Moreau-Rockafellar) 设 $f_1, f_2: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 是两个凸函数, 则对任意的 $x_0 \in \mathbb{R}^n$,

$$\partial f_1(x_0) + \partial f_2(x_0) \subseteq \partial (f_1 + f_2)(x_0). \quad (1.10)$$

进一步地, 若 $\mathbf{int} \mathbf{dom} f_1 \cap \mathbf{dom} f_2 \neq \emptyset$, 则对任意的 $x_0 \in \mathbb{R}^n$,

$$\partial (f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0). \quad (1.11)$$

定理 1.16 (Dubovitskii-Milyutin) 设 $f_1, f_2, \dots, f_m: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 均为凸函数, 令

$$f(x) = \max \{f_1(x), f_2(x), \dots, f_m(x)\}, \quad \forall x \in \mathbb{R}^n.$$

对 $x_0 \in \bigcap_{i=1}^m \mathbf{int} \mathbf{dom} f_i$, 定义 $I(x_0) = \{i \mid f_i(x_0) = f(x_0)\}$, 则

$$\partial f(x_0) = \mathbf{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0). \quad (1.12)$$

例 1.2 (ℓ_1 范数) 定义 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为 ℓ_1 范数, 则对 $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, 有

$$f(x) = \|x\|_1 = |x_1| + |x_2| + \dots + |x_n| = \max_{s \in \{-1, 1\}^n} s^T x.$$

于是

$$\partial f(x) = J_1 \times J_2 \times \dots \times J_n, \quad J_k = \begin{cases} [-1, 1], & x_k = 0, \\ \{1\}, & x_k > 0, \\ \{-1\}, & x_k < 0. \end{cases}$$

Chapter 2

典型优化问题

2.1 线性规划

线性规划问题的一般形式如下:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x, \\ \text{s.t.} \quad & Ax = b, Gx \leq e, \end{aligned} \tag{2.1}$$

其中 $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{p \times n}$ 和 $e \in \mathbb{R}^p$ 是给定的矩阵和向量, $x \in \mathbb{R}^n$ 是决策变量. 在实际中, 我们考虑两种特殊形式 (其他形式都可以转化成这两种形式): 标准形 (等式约束和决策变量非负)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x, \\ \text{s.t.} \quad & Ax = b, x \geq 0, \end{aligned} \tag{2.2}$$

以及不等式形 (没有等式约束)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^T x, \\ \text{s.t.} \quad & Ax \leq b. \end{aligned} \tag{2.3}$$

2.2 最小二乘问题

最小二乘问题的一般形式如下:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m r_i^2(x), \tag{2.4}$$

其中 $r_i : \mathbb{R}^n \rightarrow \mathbb{R}$ 为实值函数. 如果所有的 r_i 都是线性函数, 我们称问题 (2.4) 为线性最小二乘问题, 否则称其为非线性最小二乘问题.

线性最小二乘问题是回归分析中的一个基本模型, 它可以表示为

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (a_i^T x - b_i)^2,$$

记 $A = [a_1, a_2, \dots, a_m]^T$, 那么线性最小二乘问题可以等价地写成如下形式:

$$\min_{x \in \mathbb{R}^n} f(x) \stackrel{\text{def}}{=} \frac{1}{2} \|Ax - b\|_2^2,$$

这是一个无约束二次目标函数的优化问题. 因为二次函数 f 是凸的, 故 $x \in \mathbb{R}^n$ 为其全局极小解当且仅当 x 满足方程

$$\nabla f(x) = A^T(Ax - b) = 0.$$

2.3 复合优化问题

复合优化问题一般可以表示为如下形式:

$$\min_{x \in \mathbb{R}^n} \psi(x) = f(x) + h(x), \quad (2.5)$$

其中 $f(x)$ 是光滑函数 (比如数据拟合项), $h(x)$ 可能是非光滑的 (比如 ℓ_1 范数正则项, 约束集合的示性函数, 或它们的线性组合).

下面是一个应用举例: 图像去噪问题. 记带噪声的图像为 y , 噪声为 ε , 那么

$$y = x + \varepsilon,$$

其中 x 为要恢复的真实图像. 图像去噪的一个常用的模型是全变差模型. 考虑图像中相邻位置的像素值存在阶跃的地方是稀疏的, 定义全变差算子

$$\|x\|_{TV} = \sum_{i,j} |(\nabla x)_{i,j}|$$

其中 $(\nabla x)_{i,j}$ 取为 $\sqrt{(x_{i+1,j} - x_{ij})^2 + (x_{i,j+1} - x_{ij})^2}$ 或 $|x_{i+1,j} - x_{ij}| + |x_{i,j+1} - x_{ij}|$, 对应的全变差称为是各项同性或者各项异性的. 借助于全变差算子, 去噪模型可以表示为

$$\min_{x \in \mathbb{R}^{n_1 \times n_2}} \|x - y\|_F^2 + \lambda \|x\|_{TV}$$

其中目标函数中的第一项是数据保真项, 即重构出的图片要与已有的采集信息相容. 第二项是正则项, 用来保证重构出的图像的阶跃是稀疏的.

2.4 随机优化问题

随机优化问题可以表示成以下形式:

$$\min_{x \in \mathcal{X}} \mathbb{E}_\xi [F(x, \xi)] + h(x),$$

其中 $\mathcal{X} \subseteq \mathbb{R}^n$ 表示决策变量 x 的可行域, ξ 是一个随机变量 (分布一般是未知的). 对于每个固定的 ξ , $F(x, \xi)$ 表示样本 ξ 上的损失或者奖励. 正则项 $h(x)$ 用来保证解的某种性质. 由于变量 ξ 分布的未知性, 其数学期望 $\mathbb{E}_\xi [F(x, \xi)]$ 一般是不可计算的. 为了得到目标函数值的一个比较好的估计, 在实际问题中往往利用 ξ 的

经验分布来代替其真实分布. 具体地, 假设有 N 个样本 $\xi_1, \xi_2, \dots, \xi_N$, 令 $f_i(x) = F(x, \xi_i)$, 我们得到优化问题

$$\min_{x \in \mathcal{X}} f(x) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N f_i(x) + h(x), \quad (2.6)$$

并称其为经验风险极小化问题或者采样平均极小化问题. 这个问题通常是难以求解的, 一方面是因为样本数 N 比较多 (因此函数值、梯度计算代价比较高), 另一方面是因为优化问题的可行域所在空间维数 n 比较大.

2.5 半定规划

半定规划问题的一般形式如下:

$$\begin{aligned} \min \quad & c^T x, \\ \text{s.t.} \quad & x_1 A_1 + x_2 A_2 + \dots + x_n A_n + B \preceq 0, Gx = h, \end{aligned} \quad (2.7)$$

其中 $c \in \mathbb{R}^n, A_i \in \mathcal{S}^m, i = 1, 2, \dots, n, B \in \mathcal{S}^m, G \in \mathbb{R}^{p \times n}, h \in \mathbb{R}^p$ 为已知的向量和矩阵, $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ 是自变量. 如果矩阵 A_i, B 是对角的, 那么问题 (2.7) 退化为线性规划问题. 类似于线性规划问题, 我们考虑半定规划的标准形式

$$\begin{aligned} \min \quad & \langle C, X \rangle, \\ \text{s.t.} \quad & \langle A_1, X \rangle = b_1, \dots, \langle A_n, X \rangle = b_n, X \succeq 0, \end{aligned} \quad (2.8)$$

和不等式形式

$$\begin{aligned} \min \quad & c^T x, \\ \text{s.t.} \quad & x_1 A_1 + x_2 A_2 + \dots + x_n A_n + B \preceq 0. \end{aligned} \quad (2.9)$$

形如 (2.7) 的优化问题都可以转化成 (2.8) 或者 (2.9) 的形式.

2.6 矩阵优化

矩阵优化问题具有如下形式:

$$\min_{X \in \mathcal{X}} \psi(X),$$

其中 \mathcal{X} 为特定的矩阵空间, $\psi(X) : \mathcal{X} \rightarrow \mathbb{R}$ 为给定的函数, 可能是非光滑的. 对于矩阵优化问题, 如果决策变量为一个 $n \times n$ 矩阵, 那么我们可能需要确定 n^2 个元素. 因此, 决策变量的维数过大往往是矩阵优化问题难以快速求解的一个重要原因.

2.7 优化模型语言

考虑如下简单的凸优化问题:

$$\begin{aligned} \min \quad & \|Ax - b\|_2, \\ \text{s.t.} \quad & Cx = d, \|x\|_\infty \leq e. \end{aligned} \tag{2.10}$$

这里主要介绍 cvx 这一工具. Matlab 中使用的是 cvx 工具箱: <https://cvxr.com/cvx/>. 下面给出问题 (2.10) 的代码示例:

```

1 m = 20; n = 10; p = 4;
2 A = randn(m, n); b = randn(m, 1);
3 C = randn(p, n); d = randn(p, 1); e = rand;
4 cvx_begin
5     variable x(n)
6     minimize(norm(A * x - b, 2))
7     subject to
8         C * x == d
9         norm(x, Inf) <= e
10 cvx_end

```

Python 中使用的是 cvxpy 库, 可直接通过 pip install cvxpy 进行安装.

```

1 import cvxpy as cp
2 import numpy as np
3
4 m, n, p = 20, 10, 4
5 np.random.seed(1)
6 A = np.random.randn(m, n)
7 b = np.random.randn(m)
8 C = np.random.randn(p, n)
9 d = np.random.randn(p)
10 e = np.random.rand()
11
12 x = cp.Variable(n)
13 objective = cp.Minimize(cp.norm(A @ x - b, 2))
14 constraints = [
15     C @ x == d,
16     cp.norm_inf(x) <= e
17 ]
18
19 problem = cp.Problem(objective, constraints)
20 problem.solve()
21
22 print("Optimal value:", problem.value)
23 print("Optimal x:", x.value)

```

Chapter 3

最优性理论

3.1 最优化问题解的存在性

考虑下面优化问题最优解的存在性:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned} \tag{3.1}$$

其中 $\mathcal{X} \subseteq \mathbb{R}^n$ 为可行域.

定理 3.1 (Weierstrass 定理) 考虑一个适当且闭的函数 $f: \mathcal{X} \rightarrow (-\infty, +\infty]$, 假设下面三个条件中任意一个成立:

- (1) $\text{dom } f \stackrel{\text{def}}{=} \{x \in \mathcal{X} : f(x) < +\infty\}$ 是有界的;
- (2) 存在一个常数 $\bar{\gamma}$ 使得下水平集

$$C_{\bar{\gamma}} \stackrel{\text{def}}{=} \{x \in \mathcal{X} : f(x) \leq \bar{\gamma}\}$$

是非空且有界的;

- (3) f 是强制的, 即对于任一满足 $\|x^k\| \rightarrow +\infty$ 的点列 $\{x^k\} \subset \mathcal{X}$, 都有

$$\lim_{k \rightarrow \infty} f(x^k) = +\infty,$$

那么, 问题 (3.1) 的最小值点集 $\{x \in \mathcal{X} \mid f(x) \leq f(y), \forall y \in \mathcal{X}\}$ 是非空且紧的.

注 定理 3.1 的三个条件在本质上都是保证 $f(x)$ 的最小值不能在无穷远处取到, 因此我们可以仅在一个有界的下水平集中考虑 $f(x)$ 的最小值.

3.2 无约束可微问题的最优性理论

无约束可微优化问题通常表示为如下形式:

$$\min_{x \in \mathbb{R}^n} f(x), \tag{3.2}$$

其中假设 f 是连续可微函数.

定义 3.1 (下降方向) 对于可微函数 f 和点 $x \in \mathbb{R}^n$, 如果存在向量 d 满足

$$\nabla f(x)^T d < 0,$$

那么称 d 为 f 在点 x 处的一个下降方向.

定理 3.2 (一阶必要条件) 假设 f 在全空间 \mathbb{R}^n 可微. 如果 x^* 是一个局部极小点, 那么

$$\nabla f(x^*) = 0.$$

定理 3.3 (二阶最优性条件) 假设 f 在点 x^* 的一个开邻域内是二阶连续可微的, 则成立:

(1) 如果 x^* 是 f 的一个局部极小点, 那么

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succeq 0;$$

(2) 如果在点 x^* 处有

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succ 0$$

成立, 那么 x^* 为 f 的一个局部极小点.

3.3 无约束不可微问题的最优性理论

定理 3.4 (凸优化问题一阶充要条件) 假设 f 是适当且凸的函数, 则 x^* 为问题 (3.2) 的一个全局极小点当且仅当 $0 \in \partial f(x^*)$.

定理 3.5 (复合优化问题一阶必要条件) 考虑一般的复合优化问题

$$\min_{x \in \mathbb{R}^n} \psi(x) = f(x) + h(x), \quad (3.3)$$

令 x^* 为问题 (3.3) 的一个局部极小点, 那么

$$-\nabla f(x^*) \in \partial h(x^*),$$

其中 $\partial h(x^*)$ 为凸函数 h 在点 x^* 处的次梯度集合.

下面给出一个实例, 考虑如下问题:

$$\min_{x \in \mathbb{R}^n} \psi(x) \stackrel{\text{def}}{=} f(x) + \mu \|x\|_1, \quad (3.4)$$

其中 $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ 为光滑函数, 正则系数 $\mu > 0$ 用来调节解的稀疏度. 尽管 $\|x\|_1$ 不是可微的, 但我们可以计算其次微分

$$\partial_i \|x\|_1 = \begin{cases} \{1\}, & x_i > 0, \\ [-1, 1], & x_i = 0, \\ \{-1\}, & x_i < 0. \end{cases}$$

因此, 如果 x^* 是问题 (3.4) 的一个局部最优解, 那么其满足

$$-\nabla f(x^*) \in \mu \partial \|x^*\|_1,$$

即

$$\nabla_i f(x^*) = \begin{cases} -\mu, & x_i^* > 0, \\ a \in [-\mu, \mu], & x_i^* = 0, \\ \mu, & x_i^* < 0. \end{cases}$$

进一步地, 如果 $f(x)$ 是凸的, 那么满足上式的 x^* 就是问题 (3.4) 的全局最优解.

3.4 对偶理论

这一节以及本章之后的章节考虑一般的约束优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leq 0, i \in \mathcal{I}, \\ & c_i(x) = 0, i \in \mathcal{E}, \end{aligned} \tag{3.5}$$

其中 c_i 为定义在 \mathbb{R}^n 或其子集上的实值函数, \mathcal{I} 和 \mathcal{E} 分别表示不等式约束和等式约束对应的下标集合且各下标互不相同. 这个问题的可行域定义为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \leq 0, i \in \mathcal{I} \text{ 且 } c_i(x) = 0, i \in \mathcal{E}\}.$$

3.4.1 拉格朗日函数与对偶问题

定义 3.2 (拉格朗日函数和拉格朗日对偶函数) 设 $\lambda \in \mathbb{R}_+^m$, 记 $m = |\mathcal{I}|, p = |\mathcal{E}|$, 则拉格朗日函数的具体形式 $L: \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ 定义为

$$L(x, \lambda, v) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} v_i c_i(x). \tag{3.6}$$

函数 (3.6) 中的加号也可以修改为减号, 同时调整相应乘子的约束条件使得上述下界原则满足即可.

拉格朗日对偶函数 $g: \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow [-\infty, +\infty)$ 是拉格朗日函数 $L(x, \lambda, v)$ 对于 $\lambda \in \mathbb{R}_+^m, v \in \mathbb{R}^p$ 关于 x 取的下确界:

$$g(\lambda, v) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, v). \tag{3.7}$$

易知其为凹函数 (无论原始问题是否为凸问题).

引理 3.1 (弱对偶原理) 对于任意的 $\lambda \geq 0$ 和 v , 拉格朗日对偶函数给出了优化问题 (3.5) 最优值的一个下界, 即

$$g(\lambda, v) \leq p^*, \quad \lambda \geq 0, \tag{3.8}$$

其中 p^* 依赖于参数 λ 和 v 的选取.

一个自然的问题是, 从拉格朗日对偶函数获得的下界中, 哪个是最优的呢? 为了求解该最优的下界, 便有如下拉格朗日对偶问题:

$$\max_{\lambda \geq 0, v} g(\lambda, v) = \max_{\lambda \geq 0, v} \inf_{x \in \mathbb{R}^n} L(x, \lambda, v). \quad (3.9)$$

向量 λ 和 v 也称为问题 (3.5) 的**对偶变量**或者**拉格朗日乘子向量**. 由于其目标函数的凹性和约束集合的凸性, 拉格朗日对偶问题是一个凸优化问题. 当 $g(\lambda, v) = -\infty$ 时, 对偶函数提供的 p^* 的下界变得没有实际意义. 因此我们规定拉格朗日对偶函数的定义域

$$\text{dom } g = \{(\lambda, v) \mid \lambda \geq 0, g(\lambda, v) > -\infty\}.$$

当 $(\lambda, v) \in \text{dom } g$ 时, 称其为**对偶可行解**. 记对偶问题的最优值为 q^* . 称 $p^* - q^* (\geq 0)$ 为**对偶间隙**. 如果对偶间隙为 0 ($p^* = q^*$), 称**强对偶原理**成立. 假设 (λ^*, v^*) 是使得对偶问题取得最优值的解, 称其为**对偶最优解**或者**最优拉格朗日乘子**.

推导拉格朗日对偶问题最重要的是能把拉格朗日对偶函数的具体形式方便地写出来. 拉格朗日对偶问题的写法并不唯一. 如果问题 (3.5) 中有些约束, 比如对应于下标集 $\mathcal{I}_1 = \{i_1, i_2, \dots, i_q\}$ 的不等式约束, 则可以不把这些约束松弛到拉格朗日函数里. 此时拉格朗日函数为

$$L(x, s, v) = f(x) + \sum_{i \in \mathcal{I} \setminus \mathcal{I}_1} s_i c_i(x) + \sum_{i \in \mathcal{E}} v_i c_i(x), s \geq 0, c_i(x) \leq 0, i \in \mathcal{I}_1,$$

相应地, 对偶问题为

$$\max_{s \geq 0, v} \left\{ \inf_{x \in \mathbb{R}^n} L(x, s, v), \text{ s.t. } c_i(x) \leq 0, i \in \mathcal{I}_1 \right\}.$$

对于强对偶原理满足的凸问题, 不同写法的拉格朗日对偶问题是等价的.

3.4.2 带广义不等式约束优化问题的对偶

定义 3.3 (适当锥) 称满足如下条件的锥 K 为适当锥 (proper cone):

- (1) K 是凸锥;
- (2) K 是闭集;
- (3) K 是实心的 (solid), 即 $\text{int } K \neq \emptyset$;
- (4) K 是尖的 (pointed), 即对任意非零向量 x , 若 $x \in K$, 则 $-x \notin K$, 也即 K 中无法容纳直线.

适当锥 K 可以诱导出广义不等式, 它定义了全空间上的偏序关系:

$$x \preceq_K y \iff y - x \in K.$$

类似地, 可以定义严格广义不等式:

$$x \prec_K y \iff y - x \in \text{int } K.$$

定义 3.4 (对偶锥) 令 K 为全空间 Ω 的子集, 称集合

$$K^* = \{y \in \Omega \mid \langle x, y \rangle \geq 0, \forall x \in K\}$$

为其对偶锥, 其中 $\langle \cdot, \cdot \rangle$ 是 Ω 上的一个内积. 此外, 称满足 $K = K^*$ 的锥 K 为**自对偶锥**, 非负锥和半正定锥

都是自对偶锥.

如果将不等式约束函数换成向量函数, 并且推广定义相应的广义不等式约束, 我们可以得到如下形式的优化问题:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \preceq_{K_i} 0, i \in \mathcal{I}, \\ & c_i(x) = 0, i \in \mathcal{E}, \end{aligned} \quad (3.10)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}, c_i: \mathbb{R}^n \rightarrow \mathbb{R}, i \in \mathcal{E}$ 为实值函数, $c_i: \mathbb{R}^n \rightarrow \mathbb{R}^{k_i}, k_i \in \mathbb{N}_+, i \in \mathcal{I}$ 为向量值函数, K_i 为某种适当锥且 \preceq_{K_i} 表示由锥 K_i 定义的广义不等式. 因此, 问题 (3.5) 是在问题 (3.10) 中取 $k_i = 1, K_i = \mathbb{R}_+, \forall i \in \mathcal{I}$ 时的特殊情形.

根据 $K_i, i \in \mathcal{I}$ 的对偶锥 K_i^* , 我们对广义不等式约束分别引入乘子 $\lambda_i \in K_i^*, i \in \mathcal{I}$, 对等式约束引入乘子 $v_i \in \mathbb{R}, i \in \mathcal{E}$, 构造如下拉格朗日函数:

$$L(x, \lambda, v) = f(x) + \sum_{i \in \mathcal{I}} \langle c_i(x), \lambda_i \rangle + \sum_{i \in \mathcal{E}} v_i c_i(x), \quad \lambda_i \in K_i^*, v_i \in \mathbb{R}.$$

容易验证 $L(x, \lambda, v) \leq f(x), \forall x \in \mathcal{X}, \lambda_i \in K_i^*, v_i \in \mathbb{R}$. 式我们定义拉格朗日对偶函数

$$g(\lambda, v) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, v).$$

因此, 对偶问题为

$$\max_{\lambda_i \in K_i^*, v_i \in \mathbb{R}} g(\lambda, v).$$

每个优化问题都对应一个对偶问题. 相比原始问题, 对偶问题总是凸的, 其最优值给出了原始问题最优值 (极小化问题) 的一个下界. 如果原始问题满足一定的条件, 我们可以从理论上证明原始问题和对偶问题的最优值是相等的. 当原始问题的约束个数比决策变量维数更小时, 对偶问题的决策变量维数会比原始问题的小, 从而可能在相对较小的决策空间中求解. 因此, 对于对偶问题的研究非常必要.

3.5 一般约束优化问题的最优性理论

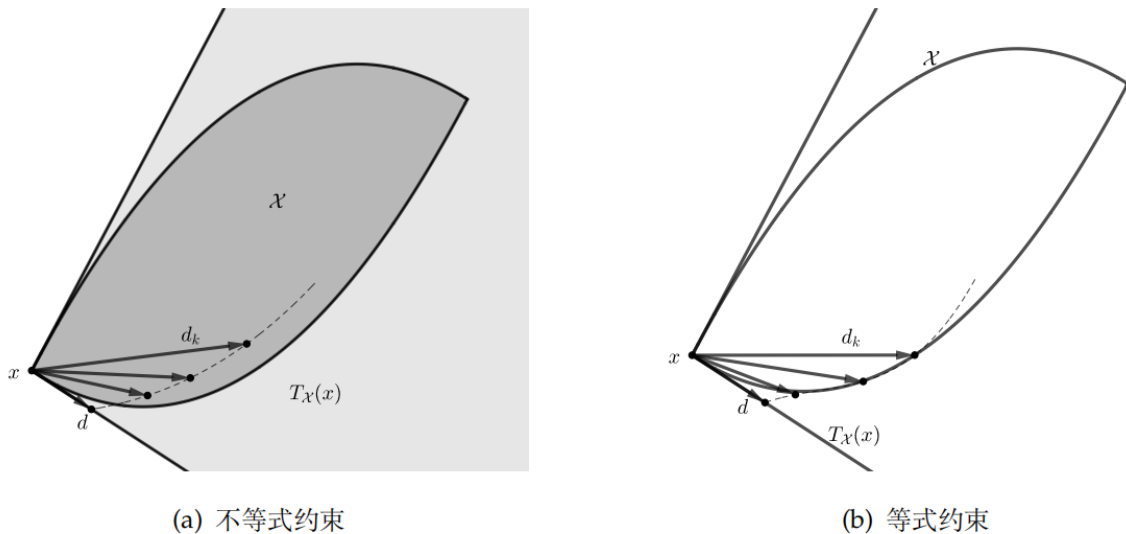
3.5.1 一阶最优性条件

定义 3.5 (切锥) 给定可行域 \mathcal{X} 及其内一点 x , 若存在可行序列 $\{z_k\}_{k=1}^\infty \subset \mathcal{X}$ 逼近 x (即 $\lim_{k \rightarrow \infty} z_k = x$) 以及正标量序列 $\{t_k\}_{k=1}^\infty, t_k \rightarrow 0$ 满足

$$\lim_{k \rightarrow \infty} \frac{z_k - x}{t_k} = d,$$

则称向量 d 为 \mathcal{X} 在点 x 处的一个切向量. 点 x 处的所有切向量构成的集合称为切锥, 用 $T_{\mathcal{X}}(x)$ 表示.

如图 3.1 所示, 左图中深色区域 \mathcal{X} 表示两个不等式约束, 其在点 x 处的切锥 $T_{\mathcal{X}}(x)$ 为图中浅色区域. 右图中则是考虑等式约束, 这里可行域 \mathcal{X} 是图中可行域的边界, 此时 $T_{\mathcal{X}}(x)$ 对应点 x 处与 \mathcal{X} 相切的两条射线.

图 3.1: \mathbb{R}^2 上的约束和切锥

定理 3.6 (几何最优性条件) 假设可行点 x^* 是问题 (3.5) 的一个局部极小点. 如果 $f(x)$ 和 $c_i(x), i \in \mathcal{I} \cup \mathcal{E}$ 在点 x^* 处是可微的, 那么

$$d^T \nabla f(x^*) \geq 0, \quad \forall d \in T_{\mathcal{X}}(x^*),$$

这也等价于 $T_{\mathcal{X}}(x^*) \cap \{d \mid \nabla f(x^*)^T d < 0\} = \emptyset$.

定义 3.6 (线性化可行方向锥) 对于可行点 $x \in \mathcal{X}$, 该点处的积极集 (active set) $\mathcal{A}(x)$ 定义为两部分下标的集合, 一部分是等式约束对应的下标, 另外一部分是不等式约束中等号成立的约束对应的下标, 即

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(x) = 0\}.$$

进一步地, 点 x 处的线性化可行方向锥定义为

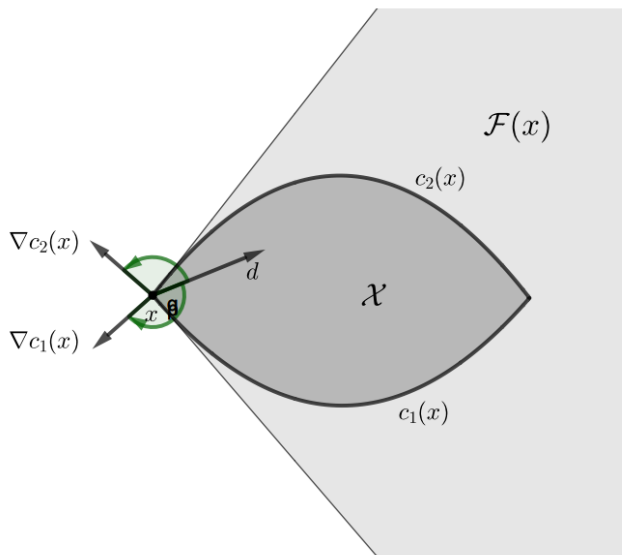
$$\mathcal{F}(x) = \left\{ d \mid \begin{array}{l} d^T \nabla c_i(x) = 0, \forall i \in \mathcal{E}, \\ d^T \nabla c_i(x) \leq 0, \forall i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}.$$

如图 3.2 所示, 在点 x 处, 两个不等式约束的等号均成立. 而 $\mathcal{F}(x)$ 中的向量应保证和 $\nabla c_i(x), i = 1, 2$ 的夹角为钝角或直角.

命题 3.1 设 $c_i(x), i \in \mathcal{E} \cup \mathcal{I}$ 一阶连续可微, 则对任意可行点 x 有 $T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$, 反之未必成立.

定义 3.7 (线性无关约束品性) 给定可行点 x 及相应的积极集 $\mathcal{A}(x)$. 如果积极集对应的约束函数的梯度, 即 $\nabla c_i(x), i \in \mathcal{A}(x)$, 是线性无关的, 则称线性无关约束品性 (LICQ) 在点 x 处成立.

引理 3.2 给定任意可行点 $x \in \mathcal{X}$, 如果在该点处 LICQ 成立, 则有 $T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

图 3.2: \mathbb{R}^2 上的约束集合和线性化可行方向锥

定义 3.8 (MFCQ) 给定可行点 x 及相应的积极集 $\mathcal{A}(x)$. 如果存在一个向量 $w \in \mathbb{R}^n$ 使得

$$\begin{aligned} \nabla c_i(x)^T w &< 0, \quad \forall i \in \mathcal{A}(x) \cap \mathcal{I}, \\ \nabla c_i(x)^T w &= 0, \quad \forall i \in \mathcal{E}, \end{aligned}$$

并且等式约束对应的梯度集 $\{\nabla c_i(x), i \in \mathcal{E}\}$ 是线性无关的, 则称 MFCQ 在点 x 处成立.

注 由 LICQ 可以推出 MFCQ, 反之未必成立.

定义 3.9 (线性约束品性) 如果所有的约束函数 $c_i(x), i \in \mathcal{I} \cup \mathcal{E}$ 都是线性的, 则称线性约束品性成立, 此时也成立 $T_{\mathcal{X}}(x) = \mathcal{F}(x)$. 一般来说, 线性约束品性和 LICQ 之间没有相互包含关系.

下面介绍 KKT 条件, 首先需要如下引理:

引理 3.3 (Farkas 引理) 设 p 和 q 为两个非负整数, 给定

$$\{a_i \in \mathbb{R}^n, i = 1, 2, \dots, p\}, \{b_i \in \mathbb{R}^n, i = 1, 2, \dots, q\}, c \in \mathbb{R}^n.$$

则满足以下条件

$$\begin{aligned} d^T a_i &= 0, \quad i = 1, 2, \dots, p, \\ d^T b_i &\geq 0, \quad i = 1, 2, \dots, q, \\ d^T c &< 0 \end{aligned}$$

的 d 不存在当且仅当存在 $\lambda_i, i = 1, 2, \dots, p$ 和 $\mu_i \geq 0, i = 1, 2, \dots, q$, 使得

$$c = \sum_{i=1}^p \lambda_i a_i + \sum_{i=1}^q \mu_i b_i.$$

由 Farkas 引理, 可以得到如下一阶必要条件, 也称为 KKT 条件, 并称满足 (3.11) 的变量对 (x^*, λ^*) 为 KKT 对, x^* 为 KKT 点.

定理 3.7 (KKT 条件) 假设 x^* 是问题 (3.5) 的一个局部最优点. 如果 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 成立, 那么存在拉格朗日乘子 λ_i^* 使得如下条件成立:

$$\begin{aligned} \text{稳定性条件} \quad & \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0, \\ \text{原始可行性条件} \quad & c_i(x^*) = 0, \forall i \in \mathcal{E}, \\ \text{原始可行性条件} \quad & c_i(x^*) \leq 0, \forall i \in \mathcal{I}, \\ \text{对偶可行性条件} \quad & \lambda_i^* \geq 0, \forall i \in \mathcal{I}, \\ \text{互补松弛条件} \quad & \lambda_i^* c_i(x^*) = 0, \forall i \in \mathcal{I}. \end{aligned} \tag{3.11}$$

注 上面的定理只给出了切锥与线性化可行方向锥相同时的最优性条件. 也就是说, 如果在局部最优点 x^* 处 $T_{\mathcal{X}}(x^*) \neq \mathcal{F}(x^*)$, 那么 x^* 不一定是 KKT 点. 同样地, 因为 KKT 条件只是必要的, 所以 KKT 点不一定是局部最优点.

3.5.2 二阶最优性条件

定义 3.10 (临界锥) 设 (x^*, λ^*) 满足 KKT 条件 (3.11), 定义临界锥为

$$\mathcal{C}(x^*, \lambda^*) = \left\{ d \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T d = 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ 且 } \lambda_i^* > 0 \right\},$$

其中 $\mathcal{F}(x^*)$ 为点 x^* 处的线性化可行方向锥.

利用上述定义, 可得如下结论:

$$d \in \mathcal{C}(x^*, \lambda^*) \Rightarrow \lambda_i^* \nabla c_i(x^*)^T d = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I}.$$

更进一步地,

$$d \in \mathcal{C}(x^*, \lambda^*) \Rightarrow d^T \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^T \nabla c_i(x^*) = 0.$$

也就是说, 临界锥定义了依据一阶导数不能判断是否为下降或上升方向的线性化可行方向, 必须使用高阶导数信息加以判断.

定理 3.8 (二阶必要条件) 假设 x^* 是问题 (3.5) 的一个局部最优解, 并且 $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 成立. 令 λ^* 为相应的拉格朗日乘子, 即 (x^*, λ^*) 满足 KKT 条件, 那么

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*).$$

定理 3.9 (二阶充分条件) 假设在可行点 x^* 处, 存在一个拉格朗日乘子 λ^* , 使得 (x^*, λ^*) 满足 KKT 条件. 如果

$$d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*), d \neq 0,$$

那么 x^* 为问题 (3.5) 的一个严格局部极小解.

为了更深刻地理解约束优化的最优性理论, 我们考虑一个具体的例子.

例 3.1 给定如下约束优化问题:

$$\min x_1^2 + x_2^2, \quad \text{s.t.} \quad \frac{x_1^2}{4} + x_2^2 - 1 = 0,$$

其拉格朗日函数为

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda \left(\frac{x_1^2}{4} + x_2^2 - 1 \right).$$

该问题可行域在任意一点 $x = (x_1, x_2)^T$ 处的线性化可行方向锥为

$$\mathcal{F}(x) = \left\{ (d_1, d_2) \mid \frac{x_1}{4} d_1 + x_2 d_2 = 0 \right\}.$$

因为只有一个等式约束且其对应函数的梯度非零, 故有 LICQ 成立, 且在 KKT 对 (x, λ) 处有 $\mathcal{C}(x, \lambda) = \mathcal{F}(x)$. 可以计算出其 4 个 KKT 对

$$(x^T, \lambda) = (2, 0, -4), \quad (-2, 0, -4), \quad (0, 1, -1) \quad \text{和} \quad (0, -1, -1).$$

我们考虑第一个 KKT 对 $y = (2, 0, -4)^T$ 和第三个 KKT 对 $z = (0, 1, -1)^T$. 计算可得,

$$\nabla_{xx}^2 L(y) = \begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix}, \quad \mathcal{C}(y) = \{(d_1, d_2) \mid d_1 = 0\}.$$

取 $d = (0, 1)$, 则

$$d^T \nabla_{xx}^2 L(y) d = -6 < 0,$$

因此 y 不是局部最优点. 类似地, 对于 KKT 对 $z = (0, 1, -1)$,

$$\nabla_{xx}^2 L(z) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathcal{C}(z) = \{(d_1, d_2) \mid d_2 = 0\}.$$

对任意的 $d = (d_1, 0)$ 且 $d_1 \neq 0$,

$$d^T \nabla_{xx}^2 L(z) d = \frac{3}{2} d_1^2 > 0.$$

因此, z 为一个严格局部最优点.

3.6 带约束凸优化问题的最优性理论

考虑如下形式的凸优化问题:

$$\begin{aligned} \min_{x \in \mathcal{D}} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leq 0, \quad i = 1, 2, \dots, m, \\ & Ax = b, \end{aligned} \tag{3.12}$$

其中 $f(x)$ 为适当的凸函数, $c_i(x), i = 1, 2, \dots, m$ 是凸函数且 $\text{dom } c_i = \mathbb{R}^n$, 以及 $A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$ 是已知的. 我们用集合 \mathcal{D} 表示自变量 x 的自然定义域, 即

$$\mathcal{D} = \text{dom } f = \{x \mid f(x) < +\infty\}.$$

自变量 x 除了受到自然定义域的限制以外, 还需要受到约束的限制. 我们定义可行域

$$\mathcal{X} = \{x \in \mathcal{D} : c_i(x) \leq 0, i = 1, 2, \dots, m; Ax = b\}.$$

3.6.1 Slater 约束品性与强对偶原理

定义 3.11 (相对内点集) 给定集合 \mathcal{D} , 记其仿射包为 $\text{affine } \mathcal{D}$. 集合 \mathcal{D} 的相对内点集定义为

$$\text{relint } \mathcal{D} = \{x \in \mathcal{D} \mid \exists r > 0, \text{ s.t. } B(x, r) \cap \text{affine } \mathcal{D} \subseteq \mathcal{D}\}.$$

相对内点是内点的推广.

定义 3.12 (Slater 约束品性) 若对凸优化问题 (3.12), 存在 $x \in \text{relint } \mathcal{D}$ 满足

$$c_i(x) < 0, \quad i = 1, 2, \dots, m, \quad Ax = b,$$

则称对此问题 Slater 约束品性满足. 有时也称该约束品性为 Slater 条件.

注 当一些不等式约束是仿射函数时, Slater 条件可以适当放宽. 不妨假设前 k 个不等式约束是仿射函数, 此时 Slater 约束品性可变为: 存在 $x \in \text{relint } \mathcal{D}$ 满足

$$c_i(x) \leq 0, \quad i = 1, 2, \dots, k; \quad c_i(x) < 0, \quad i = k+1, k+2, \dots, m; \quad Ax = b,$$

即对线性不等式约束无要求其存在严格可行点.

定理 3.10 (强对偶原理) 如果凸优化问题 (3.12) 满足 Slater 条件, 则强对偶原理成立.

3.6.2 一阶充要条件

对于一般的约束优化问题, 当问题满足特定约束品性时, 我们知道 KKT 条件是局部最优解处的必要条件. 而对于凸优化问题, 当 Slater 条件满足时, KKT 条件则变为局部最优解的充要条件 (根据凸性, 局部最优解也是全局最优解).

定理 3.11 (凸问题 KKT 条件) 对于凸优化问题 (3.12), 如果 Slater 条件成立, 那么 x^*, λ^* 分别是原始、对偶全局最优解当且仅当

$$\begin{aligned}
 &\text{稳定性条件} \quad 0 \in \partial f(x^*) + \sum_{i \in \mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* a_i, \\
 &\text{原始可行性条件} \quad Ax^* = b, \\
 &\text{原始可行性条件} \quad c_i(x^*) \leq 0, \forall i \in \mathcal{I}, \\
 &\text{对偶可行性条件} \quad \lambda_i^* \geq 0, \forall i \in \mathcal{I}, \\
 &\text{互补松弛条件} \quad \lambda_i^* c_i(x^*) = 0, \forall i \in \mathcal{I}.
 \end{aligned} \tag{3.13}$$

其中 a_i 是矩阵 A^T 的第 i 列.

最后, 我们将原始问题和对偶问题解的关系总结在表 3.1 中. 可以看到, 针对线性规划问题及其对偶问题, 解的情况只有四种可能的组合, 即

- 如果原始问题有界, 则对偶问题也有界;
- 如果原始问题无界, 则对偶问题不可行;
- 如果原始问题不可行, 则对偶问题要么无界, 要么不可行.

表 3.1: 线性规划原始问题和对偶问题的对应关系

原始问题	对偶问题		
	有界	无界	不可行
有界	✓	×	×
无界	×	×	✓
不可行	×	✓	✓

Chapter 4

无约束优化算法

本章考虑如下无约束优化问题:

$$\min_{x \in \mathbb{R}^n} f(x), \quad (4.1)$$

其中 $f(x)$ 是 $\mathbb{R}^n \rightarrow \mathbb{R}$ 的函数.

4.1 线搜索方法

线搜索类算法的数学表述为: 给定当前迭代点 x^k , 首先通过某种算法选取向量 d^k , 之后确定正数 α_k , 则下一步的迭代点可写作

$$x^{k+1} = x^k + \alpha_k d^k. \quad (4.2)$$

我们称 d^k 为迭代点 x^k 处的搜索方向, α_k 为相应的步长. 这里要求 d^k 是一个下降方向, 即 $(d^k)^T \nabla f(x^k) < 0$. 这个下降性质保证了沿着此方向搜索函数 f 的值会减小. 线搜索类算法的关键是如何选取一个好的方向 $d^k \in \mathbb{R}^n$ 以及合适的步长 α_k .

首先构造辅助函数

$$\phi(\alpha) = f(x^k + \alpha d^k),$$

其中 d^k 是给定的下降方向, $\alpha > 0$ 是该辅助函数的自变量. 函数 $\phi(\alpha)$ 的几何含义非常直观: 它是目标函数 $f(x)$ 在射线 $\{x^k + \alpha d^k : \alpha > 0\}$ 上的限制.

线搜索的目标是选取合适的 α_k 使得 $\phi(\alpha_k)$ 尽可能减小. 一个自然的想法是寻找 α_k 使得

$$\alpha_k = \arg \min_{\alpha > 0} \phi(\alpha),$$

即 α_k 为最佳步长. 这种线搜索算法被称为精确线搜索算法. 但该方法选取 α_k 通常需要很大计算量, 在实际应用中较少使用. 另一个想法不要求 α_k 是 $\phi(\alpha)$ 的最小值点, 而是仅仅要求 $\phi(\alpha_k)$ 满足某些不等式性质. 这种线搜索方法被称为非精确线搜索算法, 此方法在实际应用中较为常见.

4.1.1 线搜索准则

线搜索准则的合适与否直接决定了算法的收敛性, 若选取不合适的线搜索准则将会导致算法无法收敛.

1. Armijo 准则

定义 4.1 (Armijo 准则) 设 d^k 是点 x^k 处的下降方向, 若

$$f(x^k + \alpha d^k) \leq f(x^k) + c_1 \alpha \nabla f(x^k)^\top d^k, \quad (4.3)$$

则称步长 α 满足 Armijo 准则, 其中 $c_1 \in (0, 1)$ 是一个常数.

Armijo 准则有非常直观的几何含义, 它指的是点 $(\alpha, \phi(\alpha))$ 必须在直线

$$l_1(\alpha) = \phi(0) + c\alpha \nabla f(x^k)^\top d^k$$

的下方, 如图 4.1 所示:

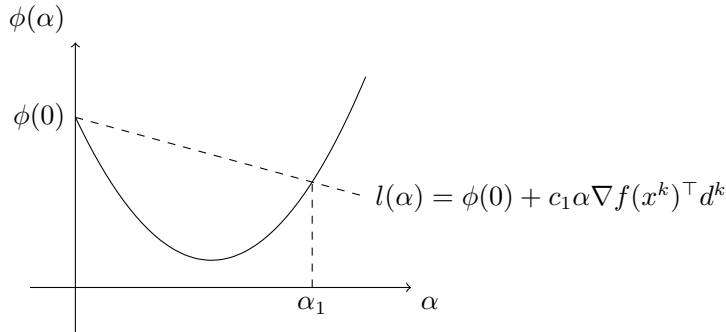


图 4.1: Armijo 准则

区间 $[0, \alpha_1]$ 中的点均满足 Armijo 准则. 我们注意到 d_k 为下降方向, 这说明 $l(\alpha)$ 的斜率为负, 选取符合条件 (4.3) 的 α 确实会使得函数值下降.

下面介绍一个寻找满足 Armijo 准则的步长的常用算法——回退法. 给定初值 $\hat{\alpha}$, 回退法通过不断以指数方式缩小试探步长, 找到第一个满足 Armijo 准则 (4.3) 的点. 具体来说, 回退法选取

$$\alpha_k = \gamma^{j_0} \hat{\alpha},$$

其中

$$j_0 = \min \left\{ j = 0, 1, \dots \mid f(x^k + \gamma^j \hat{\alpha} d^k) \leq f(x^k) + c_1 \gamma^j \hat{\alpha} \nabla f(x^k)^\top d^k \right\},$$

参数 $\gamma \in (0, 1)$ 为一个给定的实数. 回退法的基本过程如算法 1 所示.

Algorithm 1 线搜索回退法

- 1: 选择初始步长 $\hat{\alpha}$, 参数 $\gamma, c \in (0, 1)$. 初始化 $\alpha \leftarrow \hat{\alpha}$.
 - 2: **while** $f(x^k + \alpha d^k) > f(x^k) + c\alpha \nabla f(x^k)^\top d^k$ **do**
 - 3: 令 $\alpha \leftarrow \gamma\alpha$.
 - 4: **end while**
 - 5: 输出 $\alpha_k = \alpha$.
-

注 当 α 充分小时, Armijo 准则总是成立的. 在实际应用中我们通常也会给 α 设置一个下界, 防止步长过小.

2. Goldstein 准则

为了克服 Armijo 准则的缺陷, 我们需要引入其他准则来保证每一步的 α_k 不会太小.

定义 4.2 (Goldstein 准则) 设 d^k 是点 x^k 处的下降方向, 若

$$\begin{aligned} f(x^k + \alpha d^k) &\leq f(x^k) + c\alpha \nabla f(x^k)^\top d^k, \\ f(x^k + \alpha d^k) &\geq f(x^k) + (1-c)\alpha \nabla f(x^k)^\top d^k, \end{aligned} \quad (4.4)$$

则称步长 α 满足 Goldstein 准则, 其中 $c \in (0, \frac{1}{2})$.

同样, Goldstein 准则 (4.4) 也有非常直观的几何含义, 它指的是点 $(\alpha, \phi(\alpha))$ 必须在两条直线

$$\begin{aligned} l_1(\alpha) &= \phi(0) + c\alpha \nabla f(x^k)^\top d^k, \\ l_2(\alpha) &= \phi(0) + (1-c)\alpha \nabla f(x^k)^\top d^k \end{aligned}$$

之间. 如图 4.2 所示, 区间 $[\alpha_1, \alpha_2]$ 中的点均满足 Goldstein 准则, Goldstein 准则也确实去掉了过小的 α .

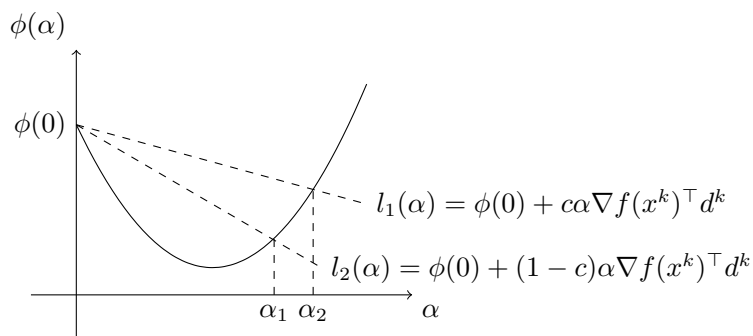


图 4.2: Goldstein 准则

3. Wolfe 准则

Goldstein 准则能够使得函数值充分下降, 但是它可能避开了最优的函数值. 为此我们引入 Wolfe 准则.

定义 4.3 (Wolfe 准则) 设 d^k 是点 x^k 处的下降方向, 若

$$\begin{aligned} f(x^k + \alpha d^k) &\leq f(x^k) + c_1 \alpha \nabla f(x^k)^\top d^k, \\ \nabla f(x^k + \alpha d^k)^\top d^k &\geq c_2 \nabla f(x^k)^\top d^k, \end{aligned} \quad (4.5)$$

则称步长 α 满足 Wolfe 准则, 其中 $c_1, c_2 \in (0, 1)$ 为给定的常数且 $c_1 < c_2$.

在准则 (4.5) 中, 第一个不等式即是 Armijo 准则, 而第二个不等式则是 Wolfe 准则的本质要求. 注意到 $\nabla f(x^k + \alpha d^k)^\top d^k$ 恰好就是 $\phi(\alpha)$ 的导数, Wolfe 准则实际要求 $\phi(\alpha)$ 在点 α 处切线的斜率不能小于 $\phi'(0)$ 的 c_2 倍.

如图 4.3 所示, 在区间 $[\alpha_1, \alpha_2]$ 中的点均满足 Wolfe 准则. 注意到在 $\phi(\alpha)$ 的极小值点 α^* 处有 $\phi'(\alpha^*) = \nabla f(x^k + \alpha^* d^k)^\top d^k = 0$, 因此 α^* 永远满足第二个等式. 而选择较小的 c_1 可使得 α^* 同时满足第一个等式, 即 Wolfe 准则在绝大多数情况下会包含线搜索子问题的精确解. 在实际应用中, 参数 c_2 通常取为 0.9.

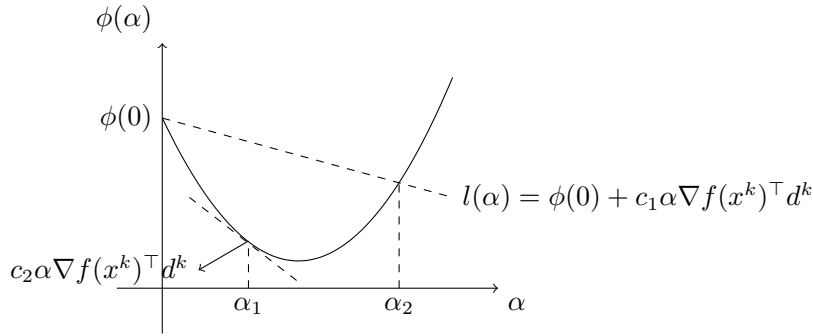


图 4.3: Wolfe 准则

以上介绍的三种准则都有一个共同点: 使用这些准则产生的迭代点列都是单调的. 在实际应用中, 非单调算法有时会有更好的效果. 这就需要我们应用非单调线搜索准则, 这里介绍其中两种.

4. 非单调线搜索准则

定义 4.4 (Grippo) 设 d^k 是点 x^k 处的下降方向, $M > 0$ 为给定的正整数. 以下不等式可作为一种线搜索准则:

$$f(x^k + \alpha d^k) \leq \max_{0 \leq j \leq \min\{k, M\}} f(x^{k-j}) + c_1 \alpha \nabla f(x^k)^T d^k, \quad (4.6)$$

其中 $c_1 \in (0, 1)$ 为给定的常数.

准则 (4.6) 和 Armijo 准则非常相似, 区别在于 Armijo 准则要求下一次迭代的函数值 $f(x^{k+1})$ 相对于本次迭代的函数值 $f(x^k)$ 有充分下降, 而准则 (4.6) 只需要下一步函数值相比前面至多 M 步以内迭代的函数值有下降就可以了. 显然这一准则的要求比 Armijo 准则更宽, 它也不要求 $f(x^k)$ 的单调性.

定义 4.5 (Zhang, Hager) 设 d^k 是点 x^k 处的下降方向, 以下不等式可作为一种线搜索准则:

$$f(x^k + \alpha d^k) \leq C^k + c_1 \alpha \nabla f(x^k)^T d^k, \quad (4.7)$$

其中 C^k 满足递推式 $C^0 = f(x^0)$, $C^{k+1} = \frac{1}{Q^{k+1}} (\eta Q^k C^k + f(x^{k+1}))$, 序列 $\{Q^k\}$ 满足 $Q^0 = 1$, $Q^{k+1} = \eta Q^k + 1$, 参数 $\eta, c_1 \in (0, 1)$.

我们可以用以下的方式理解这个准则: 变量 C^k 实际上是本次搜索准则的参照函数值, 即充分下降性质的起始标准; 而下一步的标准 C^{k+1} 则是函数值 $f(x^{k+1})$ 和 C^k 的凸组合, 并非仅仅依赖于 $f(x^{k+1})$, 而凸组合的两个系数由参数 η 决定. 可以看到当 $\eta = 0$ 时, 此准则就是 Armijo 准则.

4.1.2 线搜索算法

之前的讨论已经初步介绍了回退法 (算法 1), 并指出该算法可以用于寻找 Armijo 准则 (4.3) 的步长. 实际上只要修改一下算法的终止条件, 回退法就可以被用在其他线搜索准则之上. 回退法的实现简单、原理直观, 所以它是最常用的线搜索算法之一. 然而, 回退法的缺点也很明显: 第一, 它无法保证找到满足 Wolfe 准则的步长, 即 (4.5) 的第二个条件不一定成立; 第二, 回退法以指数的方式缩小步长, 因此对初值 $\hat{\alpha}$ 和参数 γ 的选取比较敏感. 下面简单介绍其他类型的线搜索算法.

为了提高回退法的效率, 我们有基于多项式插值的线搜索算法. 假设初始步长 $\hat{\alpha}_0$ 已给定, 如果经过验证, $\hat{\alpha}_0$ 不满足 Armijo 准则, 下一步就需要减小试探步长. 和回退法不同, 我们不直接将 $\hat{\alpha}_0$ 缩小常数倍, 而

是基于 $\phi(0), \phi'(0), \phi(\hat{\alpha}_0)$ 这三个信息构造一个二次插值函数 $p_2(\alpha)$, 即寻找二次函数 $p_2(\alpha)$ 满足

$$p_2(0) = \phi(0), \quad p_2'(0) = \phi'(0), \quad p_2(\hat{\alpha}_0) = \phi(\hat{\alpha}_0).$$

由于二次函数只有三个参数, 以上三个条件可以唯一决定 $p_2(\alpha)$, 而且不难验证 $p_2(\alpha)$ 的最小值点恰好位于 $(0, \hat{\alpha}_0)$ 内. 此时取 $p_2(\alpha)$ 的最小值点 $\hat{\alpha}_1$ 作为下一个试探点, 利用同样的方式不断递归下去直至找到满足 Armijo 准则的点.

基于插值的线搜索算法可以有效减少试探次数, 但仍然不能保证找到的步长满足 Wolfe 准则. Fletcher 提出了一个用于寻找满足 Wolfe 准则的算法, 这里不展开叙述.

4.1.3 收敛性

定理 4.1 (Zoutendijk) 考虑一般的迭代格式 (4.2), 其中 d^k 是搜索方向, α_k 是步长, 且在迭代过程中 Wolfe 准则 (4.5) 满足. 假设目标函数 f 下有界、连续可微且梯度 L -利普希茨连续, 即

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

那么

$$\sum_{k=0}^{\infty} \cos^2 \theta_k \|\nabla f(x^k)\|^2 < +\infty, \quad (4.8)$$

其中 $\cos \theta_k$ 为负梯度 $-\nabla f(x^k)$ 和下降方向 d^k 夹角的余弦, 即

$$\cos \theta_k = \frac{-\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \|d^k\|}.$$

不等式 (4.8) 也被称为 Zoutendijk 条件.

定理 4.1 指出, 只要迭代点满足 Wolfe 准则, 对梯度利普希茨连续且下有界函数总能推出 (4.8) 式成立.

命题 4.1 (线搜索算法的收敛性) 对于迭代法 (4.2), 设 θ_k 为每一步负梯度 $-\nabla f(x^k)$ 与下降方向 d^k 的夹角, 并假设对任意的 k , 存在常数 $\gamma > 0$, 使得

$$\theta_k < \frac{\pi}{2} - \gamma, \quad (4.9)$$

则在定理 4.1 成立的条件下, 有

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0.$$

4.2 梯度类算法

4.2.1 梯度下降法

对于光滑函数 $f(x)$, 在迭代点 x^k 处, 我们需要选择一个较为合理的 d^k 作为下降方向. 注意到 $\phi(\alpha) = f(x^k + \alpha d^k)$ 有泰勒展开

$$\phi(\alpha) = f(x^k) + \alpha \nabla f(x^k)^T d^k + \mathcal{O}(\alpha^2 \|d^k\|^2),$$

根据柯西不等式, 当 α 足够小时, 取 $d^k = -\nabla f(x^k)$ 会使得函数下降最快. 因此梯度法就是选取 $d^k = -\nabla f(x^k)$ 的算法, 它的迭代格式为

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k). \quad (4.10)$$

步长 α_k 的选取可依赖于上一节的线搜索算法, 也可直接选取固定的 α_k .

定理 4.2 考虑正定二次函数

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

其最优值点为 x^* . 若使用梯度法 (4.10) 并选取 α_k 为精确线搜索步长, 即

$$\alpha_k = \frac{\|\nabla f(x^k)\|^2}{\nabla f(x^k)^T A \nabla f(x^k)}, \quad (4.11)$$

则梯度法关于迭代点列 $\{x^k\}$ 是 Q-线性收敛的, 即

$$\|x^{k+1} - x^*\|_A^2 \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n} \right)^2 \|x^k - x^*\|_A^2,$$

其中 λ_1, λ_n 分别为 A 的最大、最小特征值, $\|x\|_A \stackrel{\text{def}}{=} \sqrt{x^T A x}$ 为由正定矩阵 A 诱导的范数.

注 这个结果说明了梯度法的一个很重大的缺陷: 当目标函数的 Hessian 矩阵条件数较大时, 它的收敛速度会非常缓慢.

定理 4.3 (梯度法在凸函数上的收敛性) 设函数 $f(x)$ 为凸的梯度 L -利普希茨连续函数, $f^* = f(x^*) = \inf_x f(x)$ 存在且可达. 如果步长 α_k 取为常数 α 且满足 $0 < \alpha \leq \frac{1}{L}$, 那么由迭代 (4.10) 得到的点列 $\{x^k\}$ 的函数值收敛到最优值, 且在函数值的意义下收敛速度为 $\mathcal{O}(\frac{1}{k})$.

引理 4.1 设函数 $f(x)$ 是 \mathbb{R}^n 上的凸可微函数, 则以下结论等价:

- (1) f 的梯度为 L -利普希茨连续的;
- (2) 函数 $g(x) \stackrel{\text{def}}{=} \frac{L}{2}x^T x - f(x)$ 是凸函数;
- (3) $\nabla f(x)$ 有余强制性, 即对任意的 $x, y \in \mathbb{R}^n$, 有

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2. \quad (4.12)$$

定理 4.4 (梯度法在强凸函数上的收敛性) 设函数 $f(x)$ 为 m -强凸的梯度 L -利普希茨连续函数, $f^* = f(x^*) = \inf_x f(x)$ 存在且可达. 如果步长 α 满足 $0 < \alpha \leq \frac{2}{m+L}$, 那么由梯度下降法 (4.10) 迭代得到的点列 $\{x^k\}$ 收敛到 x^* , 且为 Q-线性收敛.

4.2.2 Barzilai-Borwein 方法

当问题的条件数很大, 也即问题比较病态时, 梯度下降法的收敛性质会受到很大影响. Barzilai-Borwein (BB) 方法是一种特殊的梯度法, 经常比一般的梯度法有着更好的效果. 从形式上来看, BB 方法的下降方向仍是点 x^k 处的负梯度方向 $-\nabla f(x^k)$, 但步长 α_k 并不是直接由线搜索算法给出的. 考虑梯度下降法的格式:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k),$$

这种格式也可以写成

$$x^{k+1} = x^k - D^k \nabla f(x^k),$$

其中 $D^k = \alpha_k I$. BB 方法选取的 α_k 是如下两个最优问题之一的解:

$$\min_{\alpha} \|\alpha y^{k-1} - s^{k-1}\|^2, \quad (4.13)$$

$$\min_{\alpha} \|y^{k-1} - \alpha^{-1} s^{k-1}\|^2, \quad (4.14)$$

其中 $s^{k-1} \stackrel{\text{def}}{=} x^k - x^{k-1}$ 以及 $y^{k-1} \stackrel{\text{def}}{=} \nabla f(x^k) - \nabla f(x^{k-1})$. 容易验证问题 (4.13) 和 (4.14) 的解分别为

$$\alpha_{\text{BB1}}^k \stackrel{\text{def}}{=} \frac{(s^{k-1})^T y^{k-1}}{(y^{k-1})^T y^{k-1}}, \quad \alpha_{\text{BB2}}^k \stackrel{\text{def}}{=} \frac{(s^{k-1})^T s^{k-1}}{(s^{k-1})^T y^{k-1}}, \quad (4.15)$$

因此可以得到 BB 方法的两种迭代格式:

$$x^{k+1} = x^k - \alpha_{\text{BB1}}^k \nabla f(x^k), \quad (4.16)$$

$$x^{k+1} = x^k - \alpha_{\text{BB2}}^k \nabla f(x^k). \quad (4.17)$$

从表达式 (4.15) 注意到, 计算两种 BB 步长的任何一种仅仅需要函数相邻两步的梯度信息和迭代点信息, 不需要任何线搜索算法即可选取算法步长. 对于一般的问题, 通过 (4.15) 式计算出的步长可能过大或过小, 因此我们还需要将步长做上界和下界的截断, 即选取 $0 < \alpha_m < \alpha_M$ 使得 $\alpha_m \leq \alpha_k \leq \alpha_M$.

还需注意的是, BB 方法本身是非单调方法, 有时也配合非单调收敛准则使用以获得更好的实际效果. 算法 2 给出了一种 BB 方法的框架.

Algorithm 2 非单调线搜索的 BB 方法

- 1: 给定 x^0 , 选取初值 $\alpha > 0$, 整数 $M \geq 0, c_1, \beta, \varepsilon \in (0, 1), k = 0$.
 - 2: **while** $\|\nabla f(x^k)\| > \varepsilon$ **do**
 - 3: **while** $f(x^k - \alpha \nabla f(x^k)) \geq \max_{0 \leq j \leq \min(k, M)} f(x^{k-j}) - c_1 \alpha \|\nabla f(x^k)\|^2$ **do**
 - 4: 令 $\alpha \leftarrow \beta \alpha$.
 - 5: **end while**
 - 6: 令 $x^{k+1} = x^k - \alpha \nabla f(x^k)$. 根据公式 (4.15) 之一计算 α , 并做截断使得 $\alpha \in [\alpha_m, \alpha_M]$.
 - 7: $k \leftarrow k + 1$.
 - 8: **end while**
-

4.2.3 Tikhonov 正则化模型求解

设 x 为真实图像, 其大小为 $m \times n$, y 为带噪声的图像, 即 $y = x + e$, 其中 e 是高斯白噪声. 为了从有噪声的图像 y 中还原出原始图像 x , 利用 Tikhonov 正则化的思想可以建立如下模型:

$$\min_x f(x) = \frac{1}{2} \|x - y\|_F^2 + \lambda \left(\|D_1 x\|_F^2 + \|D_2 x\|_F^2 \right), \quad (4.18)$$

其中 D_1x, D_2x 分别表示对 x 在水平方向和坚直方向上做向前差分, 即

$$(D_1x)_{ij} = \frac{1}{h}(x_{i+1,j} - x_{ij}), \quad (D_2x)_{ij} = \frac{1}{h}(x_{i,j+1} - x_{ij}),$$

其中 h 为给定的离散间隔. 模型 (4.18) 由两项组成: 第一项为保真项; 第二项为 Tikhonov 正则项, 该项期望原始图像 x 各个部分的变化不要太剧烈, 这种正则项会使得恢复出的 x 有比较好的光滑性.

模型 (4.18) 的目标函数是光滑的, 因此可以利用梯度法来求解. 容易求出 $f(x)$ 的梯度

$$\nabla f(x) = x - y - 2\lambda\Delta x,$$

其中 Δ 是图像 x 的离散拉普拉斯算子, 即

$$(\Delta x)_{ij} = \frac{x_{i+1,j} + x_{i-1,j} + x_{i,j+1} + x_{i,j-1} - 4x_{ij}}{h^2},$$

因此梯度法的迭代格式为

$$x^{k+1} = x^k - t(I - 2\lambda\Delta)x^k + ty^k.$$

我们注意到离散拉普拉斯算子是一个线性算子, 因此求解模型 (4.18) 本质上是求解一个线性方程组. 由于该线性方程组的系数矩阵不方便写出, 所以可以使用梯度法求解. 实际上该梯度法和 Landweber 迭代是等价的.

4.3 次梯度算法

4.3.1 次梯度算法结构

现在我们在问题 (4.1) 中假设 $f(x)$ 为凸函数, 但不一定可微. 对凸函数可以在定义域的内点处定义次梯度 $g \in \partial f(x)$. 类比梯度法的构造, 我们有如下次梯度算法的迭代格式:

$$x^{k+1} = x^k - \alpha_k g^k, \quad g^k \in \partial f(x^k), \quad (4.19)$$

其中 $\alpha_k > 0$ 为步长. 它通常有如下四种选择:

- (1) 固定步长 $\alpha_k = \alpha$;
- (2) 固定 $\|x^{k+1} - x^k\|$, 即 $\alpha_k \|g^k\|$ 为常数;
- (3) 消失步长 $\alpha_k \rightarrow 0$ 且 $\sum_{k=0}^{\infty} \alpha_k = +\infty$;
- (4) 选取 α_k 使其满足某种线搜索准则.

4.3.2 收敛性

首先我们列出 $f(x)$ 所要满足的基本假设. 对无约束优化问题(4.1), 目标函数 $f(x)$ 满足:

- (1) f 为凸函数;
- (2) f 至少存在一个有限的极小值点 x^* , 且 $f(x^*) > -\infty$;
- (3) f 的次梯度是有界的, 即

$$\|g\| \leq G, \quad \forall g \in \partial f(x), x \in \mathbb{R}_{m_i}^n.$$

1. 不同步长下的收敛性

定理 4.5 (次梯度算法的收敛性) 在假设的条件下, 设 $\{\alpha_k > 0\}$ 为任意步长序列, $\{x^k\}$ 是由算法产生的迭代序列, 则对任意的 $k \geq 0$, 有

$$2 \left(\sum_{i=0}^k \alpha_i \right) (\hat{f}^k - f^*) \leq \|x^0 - x^*\|^2 + \sum_{i=0}^k \alpha_i^2 G^2, \quad (4.20)$$

其中 x^* 是 $f(x)$ 的一个全局极小值点, $f^* = f(x^*)$, \hat{f}^k 为前 k 次迭代 $f(x)$ 的最小值, 即

$$\hat{f}^k = \min_{0 \leq i \leq k} f(x^i).$$

定理 4.5 揭示了次梯度算法的一些关键性质: 次梯度算法的收敛性非常依赖于步长的选取; 次梯度算法是非单调算法, 可以配套非单调线搜索准则 (4.6) 和 (4.7) 一起使用. 根据定理 4.5 可以直接得到不同步长取法下次梯度算法的收敛性.

推论 4.1 在假设的条件下, 次梯度算法的收敛性满足 (\hat{f}^k 的定义和定理 4.5 中的定义相同):

(1) 取 $\alpha_i = t$ 为固定步长, 则

$$\hat{f}^k - f^* \leq \frac{\|x^0 - x^*\|^2}{2kt} + \frac{G^2 t}{2};$$

(2) 取 α_i 使得 $\|x^{i+1} - x^i\|$ 固定, 即 $\alpha_i \|g^i\| = s$ 为常数, 则

$$\hat{f}^k - f^* \leq \frac{G \|x^0 - x^*\|^2}{2ks} + \frac{Gs}{2};$$

(3) 取 α_i 为消失步长, 即 $\alpha_i \rightarrow 0$ 且 $\sum_{i=0}^{\infty} \alpha_i = +\infty$, 则

$$\hat{f}^k - f^* \leq \frac{\|x^0 - x^*\|^2 + G^2 \sum_{i=0}^k \alpha_i^2}{2 \sum_{i=0}^k \alpha_i};$$

进一步可得 \hat{f}^k 收敛到 f^* .

从推论 4.1 可以看到, 无论是固定步长还是固定 $\|x^{k+1} - x^k\|$, 次梯度算法均没有收敛性, 只能收敛到一个次优的解, 这和梯度法的结论有很大的不同; 只有当 α_k 取消失步长时 \hat{f}^k 才具有收敛性. 一个常用的取法是 $\alpha_k = \frac{1}{k}$, 这样不但可以保证其为消失步长, 还可以保证 $\sum_{i=0}^{\infty} \alpha_i^2$ 有界.

2. 收敛速度和步长的关系

在推论 4.1 中, 通过适当选取步长 α_i 可以获得对应次梯度算法的收敛速度. 假设 $\|x^0 - x^*\| \leq R$, 即初值和最优解之间的距离有上界. 假设总迭代步数 k 是给定的, 根据推论 4.1 的第一个结论,

$$\hat{f}^k - f^* \leq \frac{\|x^0 - x^*\|^2}{2kt} + \frac{G^2 t}{2} \leq \frac{R^2}{2kt} + \frac{G^2 t}{2}.$$

在固定步长下, 由平均值不等式得知当 t 满足 $\frac{R^2}{2kt} = \frac{G^2 t}{2}$, 即 $t = \frac{R}{G\sqrt{k}}$ 时, 我们有估计

$$\hat{f}^k - f^* \leq \frac{GR}{\sqrt{k}}.$$

以上分析表明要使得目标函数值达到 ε 的精度, 即 $\hat{f}^k - f^* \leq \varepsilon$, 必须取迭代步数 $k = \mathcal{O}(\frac{1}{\varepsilon^2})$ 且固定步长 α_k 要满足 $t = \mathcal{O}(\frac{1}{\sqrt{k}})$. 注意这里的固定步长依赖于最大迭代步数, 这和之前构造梯度法的步长是不太一样的. 从上面的取法中还可以看出对于满足假设的函数 f , 最大迭代步数可以作为判定迭代点是否最优的一个终止准则.

类似地, 根据推论 4.1 的第二个结论以及平均值不等式, 在固定 $\|x^{i+1} - x^i\|$ 的条件下可以取 $s = \frac{R}{\sqrt{k}}$, 同样会得到估计

$$\hat{f}^k - f^* \leq \frac{GR}{\sqrt{k}}.$$

如果我们知道 $f(x)$ 的更多信息, 则可以利用这些信息来选取步长. 例如在某些应用中可预先知道 f^* 的值 (但不知道最小值点), 当 $\alpha_i = \frac{f(x^i) - f^*}{\|g^i\|^2}$ 时, 不等式右侧达到极小, 这等价于

$$\frac{(f(x^i) - f^*)^2}{\|g^i\|^2} \leq \|x^i - x^*\|^2 - \|x^{i+1} - x^*\|^2.$$

递归地利用上式并结合 $\|x^0 - x^*\| \leq R$ 和 $\|g^i\| \leq G$, 可以得到

$$\hat{f}^k - f^* \leq \frac{GR}{\sqrt{k}}.$$

注意, 此时步长的选取已经和最大迭代数无关, 它仅仅依赖于当前点处的函数值与最优值的差和次梯度模长.

4.4 牛顿类算法

4.4.1 经典牛顿法

对二次连续可微函数 $f(x)$, 考虑 $f(x)$ 在迭代点 x^k 处的二阶泰勒展开

$$f(x^k + d^k) = f(x^k) + \nabla f(x^k)^T d^k + \frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k + o(\|d^k\|^2). \quad (4.21)$$

我们的目的是根据这个二阶近似来选取合适的下降方向 d^k . 如果忽略 (4.21) 式中的高阶项, 并将等式右边看成关于 d^k 的函数求其稳定点, 可以得到

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k). \quad (4.22)$$

方程 (4.22) 也被称为牛顿方程, 容易得出当 $\nabla^2 f(x^k)$ 非奇异时, 更新方向 $d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$. 一般称满足方程 (4.22) 的 d^k 为牛顿方向. 因此经典牛顿法的更新格式为

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k). \quad (4.23)$$

4.4.2 收敛性

定理 4.6 (经典牛顿法的收敛性) 假设目标函数 f 是二阶连续可微的函数, 且 Hessian 矩阵在最优值点 x^* 的一个邻域 $N_\delta(x^*)$ 内是利普希茨连续的, 即存在常数 $L > 0$ 使得

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \quad \forall x, y \in N_\delta(x^*).$$

如果函数 $f(x)$ 在点 x^* 处满足 $\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则对于迭代法 (4.23) 有如下结论:

- (1) 如果初始点离 x^* 足够近, 则牛顿法产生的迭代点列 $\{x^k\}$ 收敛到 x^* ;
- (2) $\{x^k\}$ 收敛到 x^* 的速度是 Q -二次的;
- (3) $\{\|\nabla f(x^k)\|\}$ Q -二次收敛到 0.

注 牛顿法适用于优化问题的高精度求解, 但它没有全局收敛性质. 因此在实际应用中, 通常会使用梯度类算法先求得较低精度的解, 而后调用牛顿法来获得高精度的解.

4.4.3 修正牛顿法

经典牛顿法具有以下缺陷:

- (1) 每一步迭代需要求解一个 n 维线性方程组, 这导致在高维问题中计算量很大. Hessian 矩阵 $\nabla^2 f(x^k)$ 既不容易计算又不容易储存.
- (2) 当 $\nabla^2 f(x^k)$ 不正定时, 由牛顿方程 (4.22) 给出的解 d^k 的性质通常比较差. 事实上, 当 Hessian 矩阵正定时, d^k 是一个下降方向, 而在其他情况下不一定.
- (3) 当迭代点距最优值较远时, 直接选取步长 $\alpha_k = 1$ 会使得迭代不稳定甚至发散.

下面介绍带线搜索的修正牛顿法, 其基本思想是对牛顿方程中的 Hessian 矩阵 $\nabla^2 f(x^k)$ 进行修正, 使其变成正定矩阵; 同时引入线搜索以改善算法稳定性. 它的一般框架见算法 3. 该算法的关键在于修正矩阵 E^k 如何选取. 一个最直接的取法是取 $E^k = \tau_k I$, 即取 E^k 为单位矩阵的常数倍. 当 τ_k 充分大时, 总可以保证 B^k 是正定矩阵. 然而 τ_k 不宜取得过大, 这是因为当 τ_k 趋于无穷时, d^k 的方向会接近负梯度方向. 比较合适的取法是先估计 $\nabla^2 f(x^k)$ 的最小特征值, 再适当选择 τ_k . 除此之外, E^k 还可以采用修正 Cholesky 分解进行隐式更新.

Algorithm 3 带线搜索的修正牛顿法

- 1: 给定初始点 x^0 .
 - 2: **for** $k = 0, 1, 2, \dots$ **do**
 - 3: 确定矩阵 E^k 使得矩阵 $B^k \stackrel{\text{def}}{=} \nabla^2 f(x^k) + E^k$ 正定且条件数较小.
 - 4: 求解修正的牛顿方程 $B^k d^k = -\nabla f(x^k)$ 得方向 d^k .
 - 5: 使用任意一种线搜索准则确定步长 α_k .
 - 6: 更新 $x^{k+1} = x^k + \alpha_k d^k$.
 - 7: **end for**
-

4.4.4 非精确牛顿法

在经典牛顿法中, 计算牛顿方向 d^k 依赖于求解线性方程组. 当 n 较大但 $\nabla^2 f(x^k)$ 有稀疏结构时, 需要通过迭代法来求解牛顿方程 (4.22). 事实上, 考虑牛顿方程 (4.22) 的非精确解 d^k , 我们引入向量 r^k 来表示残差, 则非精确牛顿方向满足

$$\nabla^2 f(x^k) d^k = -\nabla f(x^k) + r^k.$$

这里假设相对误差 η_k 满足

$$\|r^k\| \leq \eta_k \|\nabla f(x^k)\|.$$

显然, 牛顿法的收敛性依赖于相对误差 η_k 的选取, 直观上牛顿方程求解得越精确, 非精确牛顿法的收敛性就越好.

定理 4.7 (非精确牛顿法的收敛性) 设函数 $f(x)$ 二阶连续可微, 且在最小值点 x^* 处的 Hessian 矩阵正定, 则在非精确牛顿法中,

- (1) 若存在常数 $t < 1$ 使得 η_k 满足 $0 < \eta_k < t, k = 1, 2, \dots$, 则该算法收敛速度是 Q-线性的;
- (2) 若 $\lim_{k \rightarrow \infty} \eta_k = 0$, 则该算法收敛速度是 Q-超线性的;
- (3) 若 $\eta_k = \mathcal{O}(\|\nabla f(x^k)\|)$, 则该算法收敛速度是 Q-二次的.

4.5 拟牛顿类算法

拟牛顿方法不计算 Hessian 矩阵 $\nabla^2 f(x)$, 而是构造其近似矩阵 B^k 或其逆的近似矩阵 H^k . 同时我们希望 B^k 或 H^k 仍然保留 Hessian 矩阵的部分性质.

4.5.1 割线方程

设 $f(x)$ 是二次连续可微函数, 根据泰勒展开, 向量值函数 $\nabla f(x)$ 在点 x^{k+1} 处的近似为

$$\nabla f(x) = \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x - x^{k+1}) + \mathcal{O}(\|x - x^{k+1}\|^2). \quad (4.24)$$

令 $x = x^k$, $s^k = x^{k+1} - x^k$ 及 $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$, 得到

$$\nabla^2 f(x^{k+1}) s^k + \mathcal{O}(\|s^k\|^2) = y^k. \quad (4.25)$$

忽略高阶项 $\|s^k\|^2$, 我们希望 Hessian 矩阵的近似矩阵 B^{k+1} 满足方程

$$y^k = B^{k+1} s^k, \quad (4.26)$$

或者其逆的近似矩阵 H^{k+1} 满足方程

$$s^k = H^{k+1} y^k, \quad (4.27)$$

并称 (4.26) 式和 (4.27) 式为割线方程.

还可以从另一个角度理解割线方程 (4.26). 考虑在点 x^{k+1} 处的二阶近似

$$m_{k+1}(d) = f(x^{k+1}) + \nabla f(x^{k+1})^T d + \frac{1}{2} d^T B^{k+1} d,$$

我们要求 $m_{k+1}(d)$ 在 $d = -s^k$ 和 $d = 0$ 处的梯度与 $f(x)$ 在 $x = x^k$ 和 $x = x^{k+1}$ 处的梯度分别保持一致. 注意到 $\nabla m_{k+1}(0) = \nabla f(x^{k+1})$ 是自然满足的, 为了使得第一个条件满足只需

$$\nabla m_{k+1}(-s^k) = \nabla f(x^{k+1}) - B^{k+1} s^k = \nabla f(x^k),$$

整理即可得到 (4.26) 式.

另外, 注意到近似矩阵 B^k 的正定性, 在 (4.26) 式两边同时左乘 $(s^k)^T$ 可得 $(s^k)^T B^{k+1} s^k = (s^k)^T y^k$, 因此条件

$$(s^k)^T y^k > 0 \quad (4.28)$$

为 B^{k+1} 正定的一个必要条件. 我们额外要求条件 (4.28) 在迭代过程中始终满足, 该条件也称为曲率条件.

对于一般的目标函数 $f(x)$, 需要使用 Wolfe 准则线搜索来保证曲率条件 (4.28) 成立. 实际上, 由 Wolfe 准则的第二个条件有 $\nabla f(x^{k+1})^T s^k \geq c_2 \nabla f(x^k)^T s^k$, 两边同时减去 $\nabla f(x^k)^T s^k$,

$$(y^k)^T s^k \geq (c_2 - 1) \nabla f(x^k)^T s^k > 0,$$

这是因为 $c_2 < 1$ 以及 $s^k = \alpha_k d^k$ 是下降方向. 仅仅使用 Armijo 准则不能保证曲率条件成立.

在实际应用中基于 H^k 的拟牛顿法更加实用, 这是因为根据 H^k 计算下降方向 d^k 不需要求解线性方程组. 但基于 B^k 的拟牛顿法有比较好的理论性质, 产生的迭代序列比较稳定. 拟牛顿算法框架见算法 4.

Algorithm 4 拟牛顿算法

- 1: 给定 $x^0 \in \mathbb{R}^n$, 初始矩阵 $B^0 \in \mathbb{R}^{n \times n}$ (或 H^0), 令 $k = 0$.
 - 2: **while** 未达到停机准则 **do**
 - 3: 计算方向 $d^k = -(B^k)^{-1} \nabla f(x^k)$ 或 $d^k = -H^k \nabla f(x^k)$.
 - 4: 通过线搜索找到合适的步长 $\alpha_k > 0$, 令 $x^{k+1} = x^k + \alpha_k d^k$.
 - 5: 更新 Hessian 矩阵的近似矩阵 B^{k+1} 或其逆矩阵的近似 H^{k+1} .
 - 6: $k \leftarrow k + 1$.
 - 7: **end while**
-

4.5.2 拟牛顿矩阵更新方式

1. 秩一更新 (SR1)

$$B^{k+1} = B^k + \frac{(y^k - B^k s^k)(y^k - B^k s^k)^T}{(y^k - B^k s^k)^T s^k}, \quad (4.29)$$

$$H^{k+1} = H^k + \frac{(s^k - H^k y^k)(s^k - H^k y^k)^T}{(s^k - H^k y^k)^T y^k}. \quad (4.30)$$

SR1 公式虽然结构简单, 但它不能保证矩阵在迭代过程中保持正定. 容易验证 $(y^k - B^k s^k)^T s^k > 0$ 是 B^{k+1} 正定的一个充分条件, 但这个条件在迭代过程中未必得到满足. 因此在实际中较少使用 SR1 公式.

2. BFGS 公式

$$B^{k+1} = B^k + \frac{y^k (y^k)^T}{(s^k)^T y^k} - \frac{B^k s^k (B^k s^k)^T}{(s^k)^T B^k s^k}, \quad (4.31)$$

$$H^{k+1} = \left(I - \rho_k y^k (s^k)^T \right)^T H^k \left(I - \rho_k y^k (s^k)^T \right) + \rho_k s^k (s^k)^T, \quad (4.32)$$

其中 $\rho_k = \frac{1}{(s^k)^T y^k}$. 这里 H^{k+1} 公式的推导需要用到 SMW 公式. BFGS 公式是目前最有效的拟牛顿更新格式之一, 对 (4.32) 进行改动可得到有限内存 BFGS 格式 (L-BFGS), 它是常用的处理大规模优化问题的拟牛顿类算法.

3. DFP 公式

$$B^{k+1} = \left(I - \rho_k s^k (y^k)^T \right)^T B^k \left(I - \rho_k s^k (y^k)^T \right) + \rho_k y^k (y^k)^T, \quad (4.33)$$

$$H^{k+1} = H^k - \frac{H^k y^k (H^k y^k)^T}{(y^k)^T H^k y^k} + \frac{s^k (s^k)^T}{(y^k)^T s^k}. \quad (4.34)$$

其中 $\rho_k = \frac{1}{(s^k)^T y^k}$. 尽管 DFP 格式在很多方面和 BFGS 格式存在对偶关系, 但从实际效果来看, DFP 格式整体上不如 BFGS 格式. 因此在实际使用中人们更多使用 BFGS 格式.

4.5.3 拟牛顿法的全局收敛性

定理 4.8 (BFGS 全局收敛性) 假设初始矩阵 B^0 是对称正定矩阵, 目标函数 $f(x)$ 是二阶连续可微函数, 且下水平集

$$\mathcal{L} = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$$

是凸的, 并且存在正数 m 以及 M 使得对于任意的 $z \in \mathbb{R}^n$ 以及任意的 $x \in \mathcal{L}$ 有

$$m\|z\|^2 \leq z^T \nabla^2 f(x) z \leq M\|z\|^2,$$

则采用 BFGS 格式 (4.31) 并结合 Wolfe 线搜索的拟牛顿算法全局收敛到 $f(x)$ 的极小值点 x^* .

定理 4.9 (BFGS 收敛速度) 设 $f(x)$ 二阶连续可微, 在最优点 x^* 的一个邻域内 Hessian 矩阵利普希茨连续, 且使用 BFGS 迭代格式收敛到 f 的最优值点 x^* . 若迭代点列 $\{x^k\}$ 满足

$$\sum_{k=1}^{\infty} \|x^k - x^*\| < +\infty$$

则 $\{x^k\}$ 以 Q-超线性收敛到 x^* .

4.5.4 有限内存 BFGS 方法

一般来说, 拟牛顿矩阵 B^k 或 H^k 是稠密矩阵, 而存储稠密矩阵要消耗 $\mathcal{O}(n^2)$ 的内存. 有限内存 BFGS 方法 (L-BFGS) 解决了这一存储问题.

L-BFGS 方法是根据 BFGS 公式 (4.31), (4.32) 变形而来的. 为了推导方便, 我们以 H^k 的更新公式为基础来推导相应的 L-BFGS 公式. 首先引入新的记号重写:

$$H^{k+1} = (V^k)^T H^k V^k + \rho_k s^k (s^k)^T, \quad (4.35)$$

其中

$$\rho_k = \frac{1}{(y^k)^T s^k}, \quad V^k = I - \rho_k y^k (s^k)^T.$$

观察到 (4.35) 式有类似递推的性质, 为此我们可将 (4.35) 式递归地展开 m 次, 其中 m 是一个给定的整数:

$$\begin{aligned} H^k &= (V^{k-m} \dots V^{k-1})^T H^{k-m} (V^{k-m} \dots V^{k-1}) \\ &\quad + \rho_{k-m} (V^{k-m+1} \dots V^{k-1})^T s^{k-m} (s^{k-m})^T (V^{k-m+1} \dots V^{k-1}) \\ &\quad + \rho_{k-m+1} (V^{k-m+2} \dots V^{k-1})^T s^{k-m+1} (s^{k-m+1})^T (V^{k-m+2} \dots V^{k-1}) \\ &\quad + \dots + \rho_{k-1} s^{k-1} (s^{k-1})^T. \end{aligned} \quad (4.36)$$

为了达到节省内存的目的, (4.35) 式不能无限展开下去, 但这会产生一个问题: H^{k-m} 还是无法显式求出. 一个想法就是用 H^{k-m} 的近似矩阵来代替 H^{k-m} 进行计算, 近似矩阵的选取方式非常多, 但基本原则是要保

证近似矩阵具有非常简单的结构. 假定我们给出了 H^{k-m} 的一个近似矩阵 \hat{H}^{k-m} , (4.36) 式便可以用于计算拟牛顿迭代.

在拟牛顿迭代中, 实际上并不需要计算 H^k 的显式形式, 只需要利用 $H^k \nabla f(x^k)$ 来计算迭代方向 d^k . 一个利用展开式 (4.36) 直接求解 $H^k \nabla f(x^k)$ 的算法如算法 5 所示.

Algorithm 5 L-BFGS 双循环递归算法

```

1: 初始化  $q \leftarrow \nabla f(x^k)$ .
2: for  $i = k-1, k-2, \dots, k-m$  do
3:   计算并保存  $\alpha_i \leftarrow \rho_i (s^i)^\top q$ .
4:   更新  $q \leftarrow q - \alpha_i y^i$ .
5: end for
6: 初始化  $r \leftarrow \hat{H}^{k-m} q$ , 其中  $\hat{H}^{k-m}$  是  $H^{k-m}$  的近似矩阵.
7: for  $i = k-m, k-m+1, \dots, k-1$  do
8:   计算  $\beta \leftarrow \rho_i (y^i)^\top r$ .
9:   更新  $r \leftarrow r + (\alpha_i - \beta) s^i$ .
10: end for
11: 输出  $r$ , 即  $H^k \nabla f(x^k)$ .
```

我们现在给出算法 5 直观的执行过程. 对 (4.36) 的左右两边同时右乘 $\nabla f(x^k)$, 则需要计算

$$V^{k-1} \nabla f(x^k), V^{k-2} V^{k-1} \nabla f(x^k), \dots, V^{k-m} \dots V^{k-2} V^{k-1} \nabla f(x^k).$$

这些结果可以递推地进行计算. 另外, 在计算 $V^{k-l} \dots V^{k-1} \nabla f(x^k)$ 的过程中恰好同时计算了上一步的 $\rho_{k-l} (s^{k-l})^\top [V^{k-l+1} \dots V^{k-1} \nabla f(x^k)]$, 这是一个标量, 对应着算法 5 的 α_i . 因此执行完第一个循环后, 我们得到了 α_i, q , 公式 (4.36) 变成了如下形式:

$$\begin{aligned} H^k \nabla f(x^k) &= (V^{k-m} \dots V^{k-1})^\top H^{k-m} q + (V^{k-m+1} \dots V^{k-1})^\top s^{k-m} \alpha_{k-m} \\ &\quad + (V^{k-m+2} \dots V^{k-1})^\top s^{k-m+1} \alpha_{k-m+1} + \dots + s^{k-1} \alpha_{k-1}, \end{aligned} \quad (4.37)$$

这已经简化了不少, 接下来算法 5 的第二个循环就是自上而下合并每一项. 以合并前两项为例, 它们有公共的因子 $(V^{k-m+1} \dots V^{k-1})^\top$, 提取出来之后前两项的和可以写为

$$(V^{k-m+1} \dots V^{k-1})^\top \left((V^{k-m})^\top r + \alpha_{k-m} s^{k-m} \right) = (V^{k-m+1} \dots V^{k-1})^\top (r + (\alpha_{k-m} - \beta) s^{k-m}),$$

这正是第二个循环的迭代格式. 注意合并后 (4.37) 式的结构仍不变, 因此可递归地计算下去, 最后变量 r 就是我们期望的结果 $H^k \nabla f(x^k)$.

算法 5 双循环约需要 $4mn$ 次乘法运算与 $2mn$ 次加法运算, 若近似矩阵 \hat{H}^{k-m} 是对角矩阵, 则额外需要 n 次乘法运算. 由于 m 不会很大, 因此该算法的复杂度是 $\mathcal{O}(mn)$. 算法所需要的额外存储为临时变量 α_i , 它的大小是 $\mathcal{O}(m)$. 综上所述, L-BFGS 双循环算法是非常高效的. 近似矩阵 \hat{H}^{k-m} 的取法可以是对角矩阵 $\hat{H}^{k-m} = \gamma_k I$, 其中

$$\gamma_k = \frac{(s^{k-1})^\top y^{k-1}}{(y^{k-1})^\top y^{k-1}},$$

这恰好是 BB 方法的第一个步长.

Algorithm 6 L-BFGS 方法

```

1: 选择初始点  $x^0$ , 参数  $m > 0, k \leftarrow 0$ .
2: while 未达到收敛准则 do
3:   选取近似矩阵  $\hat{H}^{k-m}$ .
4:   使用算法 5 计算下降方向  $d^k = -H^k \nabla f(x^k)$ .
5:   使用线搜索算法计算满足 Wolfe 准则的步长  $\alpha_k$ .
6:   更新  $x^{k+1} = x^k + \alpha_k d^k$ .
7:   if  $k > m$  then
8:     从内存空间中删除  $s^{k-m}, y^{k-m}$ .
9:   end if
10:  计算并保存  $s^k = x^{k+1} - x^k, y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ .
11:   $k \leftarrow k + 1$ .
12: end while

```

4.6 信赖域算法

信赖域算法和线搜索算法都是借助泰勒展开来对目标函数进行局部近似, 但看待近似函数的方式不同. 在线搜索算法中, 我们先利用近似模型求出下降方向, 然后给定步长; 而在信赖域类算法中, 我们直接在一个有界区域内求解这个近似模型, 而后迭代到下一个点. 因此信赖域算法实际上是同时选择了方向和步长.

4.6.1 信赖域算法框架

由带拉格朗日余项的泰勒展开,

$$f(x^k + d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k + td) d,$$

其中 $t \in (0, 1)$ 为和 d 有关的正数. 和牛顿法相同, 我们利用 $f(x)$ 的一个二阶近似来刻画 $f(x)$ 在点 $x = x^k$ 处的性质:

$$m_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B^k d, \quad (4.38)$$

其中 B^k 是对称矩阵. 这里要求 B^k 是 Hessian 的近似矩阵, 如果 B^k 恰好是函数 $f(x)$ 在点 $x = x^k$ 处的 Hessian 矩阵, 那么当 $f(x)$ 充分光滑时, $m_k(d)$ 的逼近误差是 $O(\|d\|^3)$.

我们使用了二阶泰勒展开来近似目标函数 $f(x)$, 但还需要考虑到泰勒展开是函数的局部性质, 它仅仅对模长较小的 d 有意义. 当 d 过长时, 模型 (4.38) 便不再能刻画 $f(x)$ 的特征, 为此需要添加约束. 我们仅在如下球内考虑 $f(x)$ 的近似:

$$\Omega_k = \{x^k + d \mid \|d\| \leq \Delta_k\}, \quad (4.39)$$

其中 $\Delta_k > 0$ 是一个和迭代有关的参数. 我们称 Ω_k 为信赖域, Δ_k 为信赖域半径. 信赖域就是我们相信 $m_k(d)$ 能很好地近似 $f(x)$ 的区域, 而 Δ_k 表示了这个区域的大小. 因此信赖域算法每一步都需要求解如下子问题

$$\min_{d \in \mathbb{R}^n} m_k(d), \text{ s.t. } \|d\| \leq \Delta_k. \quad (4.40)$$

图 4.4 显示了子问题的求解过程. 图中实线表示 $f(x)$ 的等高线, 虚线表示 $m_k(d)$ 的等高线 ($d = x - x^k$), d_N^k 表示求解无约束问题得到的下降方向 (若 B^k 为 Hessian 矩阵则 d_N^k 为牛顿方向), d_{TR}^k 表示求解信赖域子问题 (4.40) 得到的下降方向. 信赖域算法正是限制了 d 的大小, 使得迭代更加保守, 因此可以在牛顿方向

很差时发挥作用.

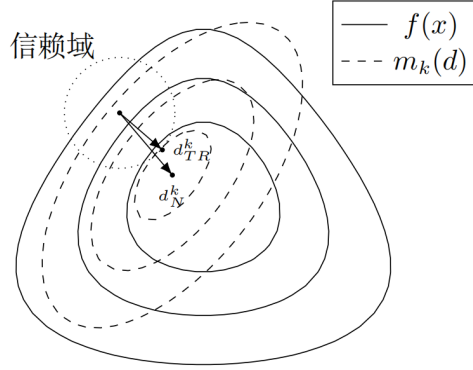


图 4.4: 信赖域算法一步迭代

此外, 选取信赖域半径 Δ_k 决定了算法的收敛性. 如果 $m_k(d)$ 对函数 $f(x)$ 近似较好, 就应该扩大信赖域半径, 反之则应该减小. 我们引入如下定义来衡量 $m_k(d)$ 近似程度的好坏:

$$\rho_k = \frac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)}, \quad (4.41)$$

其中 d^k 为求解子问题 (4.40) 得到的迭代方向. 如果 ρ_k 接近于 1, 说明用 $m_k(d)$ 来近似 $f(x)$ 是比较成功的, 我们应该扩大 Δ_k ; 如果 ρ_k 非常小甚至为负, 此时应该缩小 Δ_k .

算法 7 给出了完整的信赖域方法. 该算法对参数的取值并不敏感. 实际中可取 $\bar{\rho}_1 = 0.25, \bar{\rho}_2 = 0.75$ 以及 $\gamma_1 = 0.25, \gamma_2 = 2$. 注意, 信赖域半径 Δ_k 不会无限增长, 一是因为它有上界的控制, 二是如果信赖域约束不起作用 (即二次模型最优值处于信赖域内), 我们 also 无需增加信赖域半径. 只有当 $m_k(d)$ 近似足够好并且信赖域约束起作用时, 才需要增加 Δ_k .

Algorithm 7 信赖域算法

- 1: 给定最大半径 Δ_{\max} , 初始半径 Δ_0 , 初始点 $x^0, k \leftarrow 0$.
- 2: 给定参数 $0 \leq \eta < \bar{\rho}_1 < \bar{\rho}_2 < 1, \gamma_1 < 1 < \gamma_2$.
- 3: **while** 未达到收敛准则 **do**
- 4: 计算子问题 (4.40) 得到迭代方向 d^k .
- 5: 根据 (4.41) 式计算下降率 ρ_k .
- 6: 更新信赖域半径:

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k, & \rho_k < \bar{\rho}_1, \\ \min \{ \gamma_2 \Delta_k, \Delta_{\max} \}, & \rho_k > \bar{\rho}_2 \text{ 以及 } \|d^k\| = \Delta_k, \\ \Delta_k, & \text{其他.} \end{cases}$$

- 7: 更新自变量:

$$x^{k+1} = \begin{cases} x^k + d^k, & \rho_k > \eta, \quad / \star \text{ 只有下降比例足够大才更新 } \star / \\ x^k, & \text{其他.} \end{cases}$$

- 8: $k \leftarrow k + 1$.
 - 9: **end while**
-

4.6.2 信赖域子问题求解

本小节说明算法 7 剩下一个关键问题: 如何求解信赖域子问题 (4.40)?

4.6.2.1 迭代法

定理 4.10 d^* 是信赖域子问题

$$\min m(d) = f + g^T d + \frac{1}{2} d^T B d, \text{ s.t. } \|d\| \leq \Delta \quad (4.42)$$

的全局极小解当且仅当 d^* 是可行的且存在 $\lambda \geq 0$ 使得

$$(B + \lambda I) d^* = -g, \quad (4.43)$$

$$\lambda (\Delta - \|d^*\|) = 0, \quad (4.44)$$

$$(B + \lambda I) \succeq 0. \quad (4.45)$$

证明. 先证明必要性. (4.43) 和 (4.44) 由 KKT 条件可直接得到, 现说明 (4.45). 任取 d 满足 $\|d\| = \Delta$, 根据最优性有

$$m(d) \geq m(d^*) = m(d^*) + \frac{\lambda}{2} (\|d^*\|^2 - \|d\|^2).$$

利用 (4.43) 式消去 g , 代入上式整理得

$$(d - d^*)^T (B + \lambda I) (d - d^*) \geq 0,$$

由任意性可知 $B + \lambda I$ 半正定.

再证明充分性. 定义辅助函数

$$\hat{m}(d) = f + g^T d + \frac{1}{2} d^T (B + \lambda I) d = m(d) + \frac{\lambda}{2} d^T d,$$

由条件 (4.45) 可知 $\hat{m}(d)$ 关于 d 是凸函数. 根据条件 (4.43), d^* 满足凸函数一阶最优性条件, 从而 d^* 是 $\hat{m}(d)$ 的全局极小值点, 进而对任意可行解 d , 我们有

$$m(d) \geq m(d^*) + \frac{\lambda}{2} (\|d^*\|^2 - \|d\|^2).$$

由互补条件 (4.44) 可知 $\lambda (\Delta^2 - \|d^*\|^2) = 0$, 代入上式消去 $\|d^*\|^2$ 得

$$m(d) \geq m(d^*) + \frac{\lambda}{2} (\Delta^2 - \|d\|^2) \geq m(d^*).$$

定理 4.10 得证. □

定理 4.10 提供了问题维数 n 较小时寻找 d^* 的一个方法. 根据 (4.43) 式, 最优解是以 λ 为参数的一族向量. 我们定义

$$d(\lambda) = -(B + \lambda I)^{-1} g,$$

则只需要寻找合适的 λ 使得 (4.44) 和 (4.45) 成立即可. 根据互补条件 (4.44), 当 $\lambda > 0$ 时必有 $\|d(\lambda)\| = \Delta$;

根据半正定条件 (4.45), λ 须大于等于 B 的最小特征值的相反数.

现在研究 $\|d(\lambda)\|$ 随 λ 变化的性质. 设 B 有特征值分解 $B = Q\Lambda Q^T$, 其中 $Q = [q_1, q_2, \dots, q_n]$ 是正交矩阵, $\Lambda = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 是对角矩阵, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ 是 B 的特征值. 为了方便, 以下仅考虑 $\lambda_1 \leq 0$ 且 λ_1 是单特征根的情形. 其他情形可类似分析. 显然, $B + \lambda I$ 有特征值分解 $B + \lambda I = Q(\Lambda + \lambda I)Q^T$. 对 $\lambda > -\lambda_1 \geq 0$, 我们可直接写出 $d(\lambda)$ 的表达式:

$$d(\lambda) = -Q(\Lambda + \lambda I)^{-1}Q^T g = -\sum_{i=1}^n \frac{q_i^T g}{\lambda_i + \lambda} q_i.$$

这正是 $d(\lambda)$ 的正交分解, 由正交性可容易求出

$$\|d(\lambda)\|^2 = \sum_{i=1}^n \frac{(q_i^T g)^2}{(\lambda_i + \lambda)^2}. \quad (4.46)$$

根据 (4.46) 式可知当 $\lambda > -\lambda_1$ 且 $q_1^T g \neq 0$ 时, $\|d(\lambda)\|^2$ 是关于 λ 的严格减函数, 且有

$$\lim_{\lambda \rightarrow \infty} \|d(\lambda)\| = 0, \quad \lim_{\lambda \rightarrow -\lambda_1^+} \|d(\lambda)\| = +\infty.$$

根据连续函数介值定理, $\|d(\lambda)\| = \Delta$ 的解必存在且唯一.

$$\begin{aligned} \frac{1}{a^2} + \frac{4}{b} &= \frac{1}{a^2} + \frac{4}{b} + \lambda \left(\frac{a}{2} + \frac{a}{2} + b - 1 \right) \\ &\geq 3\sqrt[3]{\frac{1}{a^2} \cdot \frac{\lambda a}{2} \cdot \frac{\lambda a}{2}} + 2\sqrt{\frac{4}{b} \cdot \lambda b - \lambda} \end{aligned}$$

取等当且仅当 $\frac{1}{a^2} = \frac{\lambda a}{2}$, $\frac{4}{b} = \lambda b$, 即 $a = \sqrt[3]{\frac{2}{\lambda}}$, $b = -\frac{4}{\lambda}$ (b 取负号只是方便计算). 由 $a + b = 1$ 可观察出一个解 $\lambda = -2$.