5.6 共轭梯度法

SMaLL

[1] 中国石油大学（华东）
SMaLL 课题组
small.sem.upc.edu.cn
liangxijunsd@163.com

2023

5.6 共轭梯度法　(Conjugate Gradient Method)

1. 共轭方向法

2. 生成 Q 共轭方向

3. 共轭梯度法

5.6 共轭梯度法 (Conjugate Gradient Method)

1. 共轭方向法
2. 生成 Q 共轭方向
3. 共轭梯度法

Sec. 1 共轭方向法

定义 给定一个 $n \times n$ 的正定矩阵 Q, 一个非零向量的集合 $d^1, \ldots, d^k$ 是 Q 共轭向量, 如果

$$d^{i\top} Q d^j = 0, \tag{1}$$

对于所有 i 和 j 都有 $i \neq j$.

- 如果 $d^l, \ldots, d^k$ 是 Q 共轭向量, 那么它们是线性无关的.

证明 假设 d^k 可以表示为其他向量的线性组合,

$$d^k = \alpha^1 d^1 + \cdots + \alpha^{k-1} d^{k-1},$$

然后左乘 $d^{k\top} Q$,

$$d^{k\top} Q d^k = \alpha^1 d^{k\top} Q d^1 + \cdots + \alpha^{k-1} d^{k\top} Q \alpha^{k-1} = 0,$$

上式不可能成立，因为 $d^k \neq 0$ 且 Q 是正定的.

Sec. 1 共轭方向法

- Task. 二次函数的无约束最小化

$$\min f(x) = \frac{1}{2} x^\top Q x - b^\top x. \tag{2}$$

- Inputs.
  - 给定 n 个 Q 共轭方向 $d^0, \ldots, d^{n-1}$
  - x^0 为任意初始化向量
- 共轭方向迭代
  - $$\begin{aligned} \alpha_k &= \operatorname{argmin}_\alpha f(x^k + \alpha d^k) \\ &= \frac{{d^k}^\top (b - Q x^k)}{{d^k}^\top Q d^k} \end{aligned} \tag{3}$$
  - $x^{k+1} = x^k + \alpha^k d^k, \qquad k = 0, \ldots, n-1$

解释

Sec. 1 共轭方向法

定义 线性空间 V 的线性流形 (linear manifold):

$$P = r_0 + V_1 = \{r_0 + \alpha \mid \alpha \in V_1\},$$

其中 V_1 是 V 的子空间，$r_0 \in V$，V_1 的维数称为线性流形 P 的维数，线性流形是直线、二维平面、...、$n-1$ 维平面的总称。

定理 对于 Q 共轭方向 $d^0, \cdots, d^{n-1}$,

$\Rightarrow$ 共轭方向法迭代解 $\{x_k\}$, 满足

$$x^{k+1} = \arg\min_{x \in M^k} f(x), \quad k = 0, \cdots, n-1 \tag{4}$$

其中 $M^k = \{x \mid x = x^0 + v\}, \quad v \in \operatorname{span}\{d^0, \ldots, d^k\}.$

解释

证明

共轭方向. $\Rightarrow \forall\, i$,

$$\frac{df(x^i+\alpha d^i)}{d\alpha}\Big|_{\alpha=\alpha^i} = \nabla f(x^{i+1})^\top d^i = 0.$$

我们需要展示:

$$\begin{aligned}
& x^{k+1} = \operatorname{argmin}_{x\in M^k} f(x) \\
\Leftrightarrow\quad & (\alpha^0,\cdots,\alpha^k) = \operatorname{argmin}_{\gamma^0,\cdots,\gamma^k} f(x^0+\gamma^0 d^0+\gamma^k d^k) \\
\Leftrightarrow\quad & \frac{\partial f(x^0+\gamma^0 d^0+\cdots+\gamma^k d^k)}{\partial \gamma^i}\Big|_{\gamma^j=\alpha^j, j=0,\ldots,k} = \langle \nabla f(x^{k+1}), d^k\rangle = 0, \qquad i=0,\ldots,k.
\end{aligned}$$

对于 $i=0,\ldots,k$,

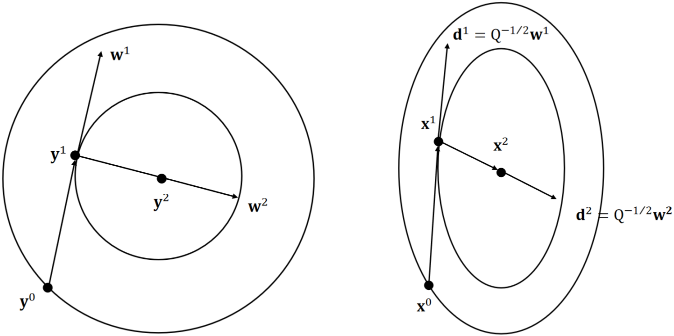
$$\begin{aligned}
\nabla f(x^{k+1})^\top d^i &= (Qx^{k+1}-b)^\top d^i \\
&= \langle x^{i+1} + \sum_{j=i+1}^{k} \alpha^j d^j, Qd^i\rangle - b^\top d^i \\
&= x^{i+1\top} Q d^i - b^\top d^i = \nabla f(x^{i+1})^\top d^i = 0.
\end{aligned}$$

SMaLL
Statistical Machine Learning Lyceum

解释

具体的 当 $b=0$， $Q=I$ (单位矩阵) 时

- f 的等值曲面是同心球;
- Q 共轭 $\longrightarrow$ 通常的正交;
- n 个正交方向的最小化 $\Rightarrow x^*$ (球的中心).



解释

一般的

$I \to$ 一般正定矩阵 Q (缩放 + 旋转)

$y = Q^{\frac{1}{2}}x$: $\frac{1}{2}x^\top Qx = \frac{1}{2}\| y \|^2$.

- 如果$w^0, \ldots, w^{n-1}$ 是 R^n 中任意正交非零向量, 那么会有
  $y^{k+1} = y^k + \alpha^k w^k, \quad k = 0, \ldots, n-1$
- 上述方程左乘 $Q^{-\frac{1}{2}}$
  $\to x^{k+1} = x^k + \alpha^k d^k$, where $d^k = Q^{-\frac{1}{2}}w^k$.
- 方向 $d^0, \ldots, d^{n-1}$ 是 Q 共轭的.

1. 共轭方向法
2. 生成 Q 共轭方向
3. 共轭梯度法

Sec. 2 生成 Q 共轭方向

- Task. 给定任意线性无关向量 $\xi^0, \ldots, \xi^k$
  构造 Q 共轭方向 $d^0, \ldots, d^k$ 对于所有的 $i = 0, \ldots, k$:

$$\text{span}\{d^0, \ldots, d^i\} = \text{span}\{\xi^0, \ldots, \xi^i\}. \tag{5}$$

- Procedure (*Gram-Schmidt procedure*)
  (1) $d^0 = \xi^0$.
  (2) 假设: 对于 $i < k$, 所选择的 Q 共轭方向 $d^0, \ldots, d^i$ 可以使上述性质成立.
  将 d^{i+1} 表示为

$$d^{i+1} = \xi^{i+1} + c^{(i+1),0} d^0 + \cdots + c^{(i+1),i} d^i \tag{6}$$

Sec. 2 生成 Q 共轭方向

选择 $c^{(i+1),j}$: d^{i+1} 是 $d^0,\ldots,d^i$ 的 Q 共轭的向量:

$$d^{i+1\top}Qd^j = \xi^{i+1\top}Qd^j + (c^{(i+1),0}d^0 + \cdots + c^{(i+1),i}d^i)^\top Qd^j = 0. \qquad (7)$$

因为 $d^0,\ldots,d^i$ 是 Q 共轭的:

$$c^{(i+1),j} = -\frac{\xi^{i+1\top}Qd^j}{d^{j\top}Qd^j}, \quad j = 0,\ldots,i. \qquad (8)$$

- $d^{i\top}Qd^i \neq 0$
- $d^{i+1} \neq 0$
- $i \to i+1$, 格拉姆-施密特过程的性质也显然成立

步骤

- 给定一组向量 $\xi^0, \ldots, \xi^k$ (不一定线性无关)
- 初始 $d^0 = \xi^0$
- 使 $d^{i+1} = \xi^{i+1} + \sum_{m=0}^{i} c^{(i+1)m} d^m$, $c^{(i+1)j} = -\frac{{\xi^{i+1}}^\top Q d^j}{{d^j}^\top Q d^j}$
- 如果 $d^{i+1} = 0 \to$ 没有将其置于 Q 共轭方向上
- 如果 $d^{i+1} \neq 0 \to$ 将其置于 Q 共轭方向上

1. 共轭方向法
2. 生成 Q 共轭方向
3. 共轭梯度法

Sec. 3 共轭梯度法

- 对梯度向量应用格拉姆-施密特过程，得到了共轭梯度法 $\xi^0 = -g^0, \ldots, \xi^{n-1} = -g^{n-1}$.
- $d^0 = -g^0$;
- $d^k = -g^k + \sum_{j=0}^{k-1} d^j,\ c^{k,j} = \frac{{g^k}^\top Q d^j}{{d^j}^\top Q d^j}$;
- $x^{k+1} = x^k + \alpha^k d^k$;
- 当 $g^k = 0$ 时，该方法找到最优解.

命题 1

对于二次凸优化，CG 方法在最多 n 步后得到最优解.

证明

首先用归纳法证明：迭代停止前生成的所有梯度 g^k 是线性无关的.

- g^0 本身是线性无关的, 否则 $g^0 = 0$
- 假设在 k 步后该方法还没有停止, 那么 $g^0, \ldots, g^{k-1}$ 是线性无关的.
- $g^k = 0$, 在这种情况下迭代终止.
- $g^k \neq 0$, g^k 与 $d^0, \ldots, d^{k-1}$ 正交等价于 g^k 与 $g^0, \ldots, g^{k-1}$ 正交.
- g^k 与 $g^0 \quad, \ldots, g^{k-1}$, 这就完成了归纳法.

简化更新方程

命题 1 共轭梯度法中的方向可由以下算法生成:

$d^0 = -g^0$;

$d^k = -g^k + \beta^k d^{k-1}$, 其中 $\beta^k = \frac{{g^k}^\top g^k}{{g^{k-1}}^\top g^{k-1}}$.

证明

- $g^{j+1} - g^j = Q(x^{j+1} - x^j) = \alpha_j Q d^j$, 其中 $\alpha^j \neq 0$, 否则 $g^{j+1} = g^j$, 这意味着 $d^j = 0$.
- $g^{i\top} Q d^j = \frac{1}{\alpha_j} g^{i\top}(g^{j+1} - g^j) = \begin{cases} 0, & j = 0, \ldots, i-2, \\ \frac{1}{\alpha^j} g^{i\top} g^i, & j = i-1, \end{cases}$
- $d^{j\top} Q d^j = \frac{1}{\alpha_j} d^{j\top}(g^{j+1} - g^j)$.
- $\Rightarrow$ $d^k = -g^k + \beta_k d^{k-1}, \beta_k = \frac{g^{k\top} g^k}{d^{k-1\top}(g^k - g^{k-1})}$,
- 因为 $d^{k-1} = -g^{k-1} + \beta_{k-1} d^{k-2} \Rightarrow \beta_k = \frac{g^{k\top} g^k}{g^{k-1\top} g^{k-1}}$
- 根据 $\langle g^k, g^{k-1} \rangle = 0 \Rightarrow \beta_k = \frac{g^{k\top}(g^k - g^{k-1})}{g^{k-1\top} g^{k-1}}$.

作业

1. Python 编程：用共轭梯度法求解 $f(x_1, x_2) = 2{x_1}^2 + 8{x_2}^2 + 8x_1x_2$ 的最小值和最小值点。