可行方向法

1. *Frank – Wolfe* 方法

2. 可行方向法

3. 既约梯度法

一. *Frank – Wolfe* 方法

1. 问题

给定非线性规划问题 $\min \quad f(x)$

$$s.t. \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases} \qquad (1)$$

其中 A 是 $m \times n$ 矩阵，秩为 m, b 是 m 维列向量, $f(x)$ 是可微函数，$x \in R^n$。

令 $D = \{x \mid Ax = b, x \geq 0\}$, 称 D 为可行域。

2. 算法分析

算法思想：在每次迭代中，将目标 函数 $f(x)$线性化，再利用线性规划方法求解。

算法分析：

设已知可行点 x^k,则有

$$
\begin{aligned}
f(x) &\approx f(x^k) + \nabla f(x^k)^T (x - x^k) \\
&= \nabla f(x^k)^T x + [f(x^k) - \nabla f(x^k)^T x^k]
\end{aligned}
$$

求解线性规划

$$
\begin{aligned}
\min \quad & \nabla f(x^k)^T x \\
s.t. \quad & x \in D
\end{aligned}
\qquad (2)
$$

定理 设 $f(x)$可微，$x^k \in D$，如果 y^k 是线性规划（2）的最优解，则有

（i）当$\nabla f(x^k)^T(y^k - x^k) = 0$时，则 x^k 是（1）的 $K-T$ 点。

（ii）当$\nabla f(x^k)^T(y^k - x^k) \neq 0$时，则向量 $d^k = y^k - x^k$ 是 $f(x)$在点 x^k 处关于 D的可行下降方向。

证明：(i) $\because y^k$ 是线性规划的最优解，

而 $\nabla f(x^k)^T y^k = \nabla f(x^k)^T x^k$.

$\therefore x^k$ 也是线性规划问题的最优解。

$\therefore x^k$ 也是线性规划问题(2)的$K-T$点。

$$\therefore \begin{cases} \nabla f(x^k) - \lambda - A^T u = 0 \\ \lambda^T x^k = 0 \\ \lambda \geq 0 \end{cases} \qquad (*)$$

其中 $\lambda \in R^n, u \in R^m$.

而 (*) 式也是非线性规划问题 (1) 的 $K-T$ 条件。

$\therefore x^k$ 也是非线性规划问题 (1)的 $K-T$ 点。

(ii) $\because\ \forall x\in D, \nabla f(x^k)^T y^k \le \nabla f(x^k)^T x$

$\therefore\ \nabla f(x^k)^T y^k \le \nabla f(x^k)^T x^k$

$\because\ \nabla f(x^k)^T (y^k - x^k) \ne 0$

$\therefore\ \nabla f(x^k)^T (y^k - x^k) < 0$，即 $y^k - x^k$ 为下降方向。

$\because D$ 是凸集。 $\therefore y^k - x^k$ 为可行方向。

$\therefore y^k - x^k$ 为可行下降方向。

当$\nabla f(x^k)^T(y^k - x^k) \neq 0$时，求解下列一维搜索问题：

$$\begin{aligned} \min \quad & f(x^k + t(y^k - x^k)) \\ s.t. \quad & 0 \le t \le 1 \end{aligned}$$

设极小点为 t_k，则可取 $x^{k+1} = x^k + t_k(y^k - x^k)$。因为$D$为凸集，则$x^{k+1} \in D$。再对点 x^{k+1}重复上述过程。

3. 算法步骤

Frank－Wolfe 算法:

1. 给定初始可行点 x^0,允许误差 $\varepsilon > 0$，令 $k = 0$。

2. 求解线性规划

$$\begin{aligned} \min \quad & \nabla f(x^k)^T x \\ s.t. \quad & x \in D \end{aligned}$$

解得极小点 y^k。

3. 若 $|\nabla f(x^k)^T (y^k - x^k)| \le \varepsilon$,则停止计算，得到点 x^k；否则转步4。

4. 从 x^k 出发，沿方向 $y^k - x^k$求解下列一维搜索问题 :

$$\begin{aligned} \min \quad & f(x^k + t(y^k - x^k)) \\ s.t. \quad & 0 \le t \le 1 \end{aligned}$$

解得t_k。

5. 令 $x^{k+1} = x^k + t_k(y^k - x^k)$,令$k := k+1$,返回2。

算法收敛性：

定理 设 $f(x)$ 可微, D 有界, $x^0 \in D$，如果 $\{x^k\}$ 是由 $Frank-wolfe$ 算法求解（1）产生的点列，则

（1）如果 $\{x^k\}$ 是有穷点列，则其最后一个点 x^k 是（1）的 $K-T$ 点。

（2）如果 $\{x^k\}$ 是无穷点列，则必有极限点，且其任意一个极限点 $x^* \in D$ 是（1）的 $K-T$ 点。

二. *Zoutendijk* 可行方向法

1. 问题

给定非线性规划问题

$$\min \quad f(x)$$
$$s.t. \quad \begin{cases} Ax \geq b \\ Cx = e \end{cases} \qquad (3)$$

其中$f(x)$是可微函数。A是$m \times n$矩阵，秩为m。C是$l \times n$矩阵，秩为l。$x \in R^n$,b和e分别是m维和l维列向量。

令 $D = \{x \mid Ax \geq b, Cx = e\}$,称$D$为可行域。

算法思想：在每次迭代中沿迭代点 处的可行下降方向进行搜索。

如何确定可行下降方向 ？

2. 算法分析

⑴ 利用迭代点的积极约 束确定可行方向。

定理1 设 $\bar{x} \in D$，在点 $\bar{x}$ 处有 $A_1 \bar{x} = b_1, A_2 \bar{x} > b_2$,其中

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

则非零向量 d 是 $\bar{x}$ 处的可行方向的充分必 要条件是 $A_1 d \geq 0$, $Cd = 0$。

证明：必要性。

设非零向量 d 是 $\bar{x}$ 处的可行方向。则存在 $\delta > 0$，使得对任意的 $t \in (0, \delta)$, $\bar{x} + td$ 仍是可行解。即有

$$A(\bar{x} + td) \geq b$$

$$C(\bar{x} + td) = e$$

$$A(\bar{x}+td)=\begin{bmatrix}A_1\\A_2\end{bmatrix}(\bar{x}+td)$$

$$=\begin{bmatrix}b_1+tA_1d\\A_2\bar{x}+tA_2d\end{bmatrix}\geq\begin{bmatrix}b_1\\b_2\end{bmatrix}$$

因为 $t>0$，所以 $A_1d\geq 0$。

$$C(\bar{x}+td)=e+tCd=e$$

所以 $Cd=0$。

充分性同理可证。

⑵ 可行下降方向的确定

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ A_1 d \geq 0 \\ Cd = 0 \end{cases} \Rightarrow \quad d \text{ 是点 } \bar{x} \text{ 处的可行下降方向。}$$

如何求出满足上述条件 的可行下降方向？

求解下列线性规划问题

$$\min \quad \nabla f(\bar{x})^T d$$

$$s.t. \quad \begin{cases} A_1 d \geq 0 \\ Cd = 0 \\ |d_j| \leq 1, \forall j \end{cases} \qquad (4)$$

结果:

（1）$d=0$是可行解，因此最优目 标函数值不大于 0。

（2）如果线性规划的最优 值小于0，则得到可行下降方向 。

（3）如果线性规划的最优 值等于0，则$\bar{x}$是$K-T$点。

定理 2 设$\bar{x}\in D$，在点$\bar{x}$处有$A_1\bar{x}=b_1, A_2\bar{x}>b_2$,其中

$$A=\begin{bmatrix}A_1\\A_2\end{bmatrix}, b=\begin{bmatrix}b_1\\b_2\end{bmatrix}$$

则$\bar{x}$是$K-T$点的充要条件是（4）的最优目标函数值为 0 。

(3) 搜索步长的确定

已知迭代点 x^k 和该点的可行下降方向 d^k,则可令 $x^{k+1}=x^k+t_k d^k$。

其中 t_k 应满足:

(*i*) $x^k+t_k d^k$ 仍为可行解;(*ii*) 使目标函数值下降。

利用下列辅助线性规划 问题求 t_k:

$$
\begin{aligned}
\min\quad & f(x^k+td^k) \\
s.t.\quad & \begin{cases} A(x^k+td^k)\geq b \\ C(x^k+td^k)=e \\ t\geq 0 \end{cases}
\end{aligned}
\qquad (5)
$$

化简线性规划问题（5）

因为 d^k 是可行下降方向，所以 $Cd^k=0$，即约束条件

$$C(x^k+td^k)=Cx^k+tCd^k=Cx^k=e$$

总是成立。

设在点 x^k 处有 $A_1x^k=b_1, A_2x^k>b_2$，其中

$$A=\begin{bmatrix}A_1\\A_2\end{bmatrix}, b=\begin{bmatrix}b_1\\b_2\end{bmatrix}$$

则约束条件 $A(x^k+td^k)\ge b$ 可以改写为

$$\begin{bmatrix}A_1(x^k+td^k)\\A_2(x^k+td^k)\end{bmatrix}\ge\begin{bmatrix}b_1\\b_2\end{bmatrix}$$

因为 $A_1x^k=b_1, A_1d^k\ge 0, t\ge 0$，所以不等式约束

$A_1(x^k+td^k)\ge b_1$ 自然成立。

所以问题（5）简化为

$$\min \quad f(x^k+td^k)$$

$$s.t. \quad \begin{cases} A_2x^k+t\,A_2d^k\ge b_2 \\ t\ge 0 \end{cases}$$

令 $\bar{b}=b_2-A_2x^k, \bar{d}=A_2d^k$，则 $\bar{b}<0$。所以约束条件可改写　　为

$$\begin{cases} t\bar{d}\ge \bar{b} \\ t\ge 0 \end{cases}$$

则可得 t 的上界为

$$t_{\max}=\begin{cases} \min\{\dfrac{\bar{b}_i}{\bar{d}_i}\,|\,\bar{d}_i<0\} & \text{当存在 } \bar{d}_i<0 \\ +\infty & \text{当 } \bar{d}\ge 0 \end{cases} \qquad (*)$$

所以问题可最终简化为

$$\begin{aligned} \min \quad & f(x^k + td^k) \\ s.t. \quad & 0 \le t \le t_{\max} \end{aligned} \qquad (6)$$

3. 算法步骤

(1) 给定初始点 x^0，令 $k=0$。

(2) 在点 x^k 处将 A 和 b 分解成 $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ 和 $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$，使得 $A_1 x^k = b_1$，

$A_2 x^k > b_2$。计算 $\nabla f(x^k)$。

(3) 求解线性规划问题

$$\begin{aligned} \min \quad & \nabla f(x^k)^T d \\ s.t. \quad & \begin{cases} A_1 d \ge 0 \\ Cd = 0 \\ |d_j| \le 1, \forall j \end{cases} \end{aligned}$$

求得最优解 d^k。

(4) 如果 $\nabla f(x^k)^T d^k = 0$,则算法结束，x^k 是 $K-T$ 点；否则转（5）。

(5) 利用（*）式计算 $t_{\max}$，求解线性规划问题

$$\begin{array}{ll} \min & f(x^k + td^k) \\ s.t. & 0 \le t \le t_{\max} \end{array}$$

解得极小值点 t_k，令 $x^{k+1} = x^k + t_k d^k$。令 $k := k+1$,返回（2）。

例 用 *Zoutendijk* 可行方向法解下列问题

$$\min \quad x_1^2 + x_2^2 - 2x_1 - 4x_2 + 6$$

$$s.t. \begin{cases} -2x_1 + x_2 + 1 \geq 0 \\ -x_1 - x_2 + 2 \geq 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

初始点取为 $x^1 = (0,0)^T$ 。

解：

$$A = \begin{pmatrix} -2 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \nabla f(x^1) = [-2,-4]^T.$$

在 x^1 处：$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b_1 = [0,0]^T.$

$$A_2=\begin{pmatrix}-2 & 1\\ -1 & -1\end{pmatrix},\ b_2=[-1,-2]^T.$$

先求 x^1 处的下降方向： $\min \quad \nabla f(x^1)^T d$

$$s.t.\ \begin{cases} A_1 d \ge 0 \\ |d_j| \le 1,\ j=1,2 \end{cases}$$

即 $\min \quad -2d_1-4d_2$

$$s.t.\ \begin{cases} d_1 \ge 0 \\ d_2 \ge 0 \\ -1 \le d_1 \le 1 \\ -1 \le d_2 \le 1 \end{cases}$$

$\therefore$ 最优解为 $d^1=[1,1]^T$.

再求步长 λ_1：

$$\bar{d}=A_2 d^1=\begin{pmatrix}-2 & 1\\ -1 & -1\end{pmatrix}\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}-1\\-2\end{bmatrix}$$

$$\bar{b} = b_2 - A_2 x^1 = \begin{bmatrix} -1 \\ -2 \end{bmatrix} - \begin{pmatrix} -2 & 1 \\ -1 & -1 \end{pmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$\therefore \lambda_{\max} = \min\{\frac{-1}{-1}, \frac{-2}{-2}\} = 1.$$

求解 $\min \quad f(x^1 + \lambda d^1) = 2\lambda^2 - 6\lambda + 6$

$s.t. \quad 0 \le \lambda \le 1$

解得$:\lambda_1 = 1$。

$$\therefore x^2 = x^1 + \lambda_1 d^1 = [1,1]^T \ .$$

三. 既约梯度法

1. 问题 $\min \quad f(x)$

$$s.t. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

其中 A 是 $m \times n$ 矩阵，秩为 m，$f(x)$ 是 R^n 上的连续可微函数。

假设：A 的任意 m 列均线性无关，并且每个基本可行解都有 m 个正分量。则每个可行解至少有 m 个正分量。

算法思想：将变量分为基变量和非 基变量，并利用约束条件 $Ax = b$，将基变量由非基变量 表示，得到简化的目标函数，并利用此函数的负梯度 构造下降可行方向。

2. 算法分析

不妨设 $A=[B, N]$, 其中 B 是 $m\times m$ 的可逆矩阵，

x_B, x_N 分别为相应的基变量和 非基变量。

$$\therefore Ax=b \Rightarrow [B, N]\begin{bmatrix} x_B \\ x_N \end{bmatrix}=b$$

即 $\quad Bx_B+Nx_N=b$

$$\therefore\ x_B=B^{-1}b-B^{-1}Nx_N \qquad (1)$$

将 (1) 式代入目标函数得:

$$f(x)=f(x_B, x_N) \underline{\underline{\Delta}} F(x_N)$$

则原问题化为

$$\min \quad F(x_N)$$
$$s.t. \quad x_B, x_N \geq 0$$

记 $r(x_N)=\nabla F(x_N)$

$$=\nabla_{x_N} f(x_B,x_N)-(B^{-1}N)^T\nabla_{x_B} f(x_B,x_N) \quad (2)$$

称 $r(x_N)$ 为目标函数的既约梯度，且 $-r(x_N)$ 为下降方向。

如何确定可行下降方向？

设已知迭代点 x^k，记第 k 个搜索方向 $d^k=[d_B^k,d_N^k]^T$，

则 $x^{k+1}=x^k+\lambda_k d^k$.

确定 d_N^K： 令 $d_{N_j}^k=\begin{cases}-x_{N_j}^k r_j(x_N^k) & r_j(x_{N_j}^k)>0\\ -r_j(x_N^k) & r_j(x_{N_j}^k)\le 0\end{cases}$ (3)

$$令\ d_{N_j}^k=\begin{cases}-x_{N_j}^k r_j(x_N^k) & r_j(x_{N_j}^k)>0\\ -r_j(x_N^k) & r_j(x_{N_j}^k)\le 0\end{cases} \qquad (3)$$

$\because$ 为使目标函数值下降，则 d_N^k 可取 $-r(x_N)$ 方向。

但是当 $x_{N_j}=0, r_j(x_N^k)>0$ 时，

$$x_{N_j}-\lambda_k r_j(x_N^k)<0 \Rightarrow 不满足可行条件\ x\ge 0.$$

确定 d_B^K ：

由上一节的定理 1, 如果 d^k 是可行方向，则

$$Ad^k=0$$

即 $$Bd_B^k+Nd_N^k=0$$

$\therefore$ 取 $d_B^k=-B^{-1}Nd_N^k \qquad (4)$.

$\therefore d^k=[-B^{-1}Nd_N^k, d_N^k]^T$ 为可行下降方向 .

如何确定搜索步长 ？

要使 $x_j^{k+1} = x_j^k + \lambda d_j^k \geq 0$, $j = 1,2,\cdots,n$. 则：

(1) 当 $d_j^k \geq 0$ 时, $\lambda \geq 0$ 即可；

(2) 当 $d_j^k < 0$ 时,

$$x_j^{k+1} = x_j^k + \lambda d_j^k \geq 0 \Rightarrow \lambda \leq \frac{-x_j^k}{d_j^k}$$

$$\therefore \quad \lambda_{\max} = \begin{cases} +\infty & d^k \geq 0 \\ \min\{\dfrac{-x_j^k}{d_j^k} \mid d_j^k < 0\} & \text{其它情形} \end{cases} \qquad (5)$$

定理3 设 x 是原问题的可行解，A,B 和 $f(x)$ 如上假设。设 d 是由(3)、(4)式确定的方向。则有：

(1) 如果 $d \neq 0$，则 d 是可行下降方向；

(2) $d = 0$ 的充要条件是 x 为 $K-T$ 点。

证明：(1) $\because Ad = Bd_B + Nd_N$

$$= B(-B^{-1}Nd_N) + Nd_N = 0$$

且当 $x_{N_j} = 0$ 时，$d_{N_j} \geq 0$. 而 $x_B > 0$.

$\therefore$ 由定理1可知 d 为可行方向。

$$\because \nabla f(x)^T d = \nabla_{x_B} f(x)^T d_B + \nabla_{x_N} f(x)^T d_N$$

$$= \nabla_{x_B} f(x)^T (-B^{-1}Nd_N) + \nabla_{x_N} f(x)^T d_N$$

$$= r(x_N)^T d_N$$

$$\because\ d_{N_j}^k = \begin{cases} -x_{N_j}^k r_j(x_N^k) & r_j(x_{N_j}^k) > 0 \\ -r_j(x_N^k) & r_j(x_{N_j}^k) \le 0 \end{cases}$$

$$\therefore r(x_N)^T d_N = -\sum_{r_j(x_N) \le 0} [r_j(x_N)]^2 - \sum_{r_j(x_N) > 0} [r_j(x_N)]^2 x_{N_j} < 0$$

$\therefore d$ 是可行下降方向。

3. 算法步骤

(1) 给定初始可行解 x^1, 误差 $\varepsilon > 0$, 令 $k = 1$。

(2) 记 A 的第 j 列为 p_j。从 x^k 中选择前 m 个最大的分量, 其下标集合记为 J_K。令

B 是由$\{p_j \mid j \in J_k\}$组成的 m 阶矩阵;

N 是由$\{p_j \mid j \notin J_k\}$组成的 $m \times (n-m)$ 阶矩阵.

由 (2) 式求出 $r(x_N^k)$, 并由 (3) 和 (4) 式 求出 d_N^k 和 d_B^k, 得到搜索方向 d^k 。

⑶ 若 $\|d^k\| \le \varepsilon$, 算法结束，得到点 x^k; 否则, 转 (4)。

(4) 由 (5) 式求出 $\lambda_{\max}$, 利用一维搜索

$$\begin{aligned} \min \quad & f(x^k + \lambda d^k) \\ s.t. \quad & 0 \le \lambda \le \lambda_{\max} \end{aligned}$$

求出 λ_k 。

(5) 令 $x^{k+1}=x^k+\lambda_k d^k$.$k:=k+1$, 转 (2)。

4. 例题

例 用既约梯度法求解下列 问题

$$\min \quad 2x_1^2+x_2^2$$

$$s.t. \quad \begin{cases} x_1-x_2+x_3=2 \\ -2x_1+x_2+x_4=1 \\ x_j\geq 0, j=1,2,3,4 \end{cases}$$

初始点为 $x^1=(1,3,4,0)^T$.

解: $$A=\begin{pmatrix} 1 & -1 & 1 & 0 \\ -2 & 1 & 0 & 1 \end{pmatrix},$$

$$\nabla f(x) = [4x_1, 2x_2, 0, 0]^T.$$

第1次迭代： $\nabla f(x^1) = [4,6,0,0]^T$.

$\because J_1 = \{2,3\}$. $\therefore x_B = [x_2, x_3]^T = [3,4]^T$， $x_N = [1,0]^T$.

且 $B = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}, N = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$。 $\therefore B^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$.

$$\therefore r(x_N^1) = \begin{bmatrix} 4 \\ 0 \end{bmatrix} - \left[\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \right]^T \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 16 \\ -6 \end{bmatrix}$$

$$\therefore d_N^1 = [d_1^1, d_4^1]^T = [-16, 6]^T$$

$$d_B^1 = -\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -16 \\ 6 \end{bmatrix} = \begin{bmatrix} -38 \\ -22 \end{bmatrix}$$

$$\therefore \lambda_{\max} = \min\{\frac{-1}{-16}, \frac{-3}{-38}, \frac{-4}{-22}\} = \frac{1}{16}$$

$$\therefore f(x^1 + \lambda d^1) = 2(1-16\lambda)^2 + (3-38\lambda)^2$$

求解一维搜索问题：

$$\begin{aligned} &\min \quad f(x^1 + \lambda d^1) = 2(1-16\lambda)^2 + (3-38\lambda)^2 \\ &s.t. \quad 0 \le \lambda \le \frac{1}{16} \end{aligned}$$

解得：$\lambda_1 = \frac{1}{16}$。

$$\therefore x^2 = x^1 + \lambda_1 d^1 = (0, \frac{5}{8}, \frac{21}{8}, \frac{3}{8})^T.$$