简约梯度法

处理的问题

简约梯度法（**reduced gradient method**）处理带有线性约束的非线性规划问题。其标准形式为：

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad \textbf{(LC-MP)}$$

其中，$x \in \mathrm{R}^n$，$f: \mathrm{R}^n \mapsto \mathrm{R}$，$r\left(A_{m\times n}\right) = m$，$b \in \mathrm{R}^m$。

LC-MP 的可行域记为：$X_l = \left\{x \in \mathrm{R}^n \mid Ax = b,\ x \geq 0\right\}$。

简约梯度法的基本思想

- 将一个可行解 $\boldsymbol{x}^k$ 的前 $\boldsymbol{m}$ 个最大的正分量定义为基变量，其余的 $\boldsymbol{n-m}$ 个变量定义为非基变量。（仿照 **LP** 的单纯形算法——但不完全相同）。
- 为此，假设（1）$\boldsymbol{X_l}$ 中的每一个可行解至少有 $\boldsymbol{m}$ 个大于零的分量；（2）$\boldsymbol{A}$ 的任意 $\boldsymbol{m}$ 列线性无关。——非退化假设。
- 因为可行解 $\boldsymbol{x}^k$ 满足约束 $A\begin{pmatrix} x_B^k \\ x_N^k \end{pmatrix} = b$，$\boldsymbol{x}^k$ 中的基变量 x_B^k 可用非基变量 x_N^k 表示。从而，目标函数 $\boldsymbol{f}$ 可写为非基变量 x_N^k 的函数。
- 基本思想：根据这样的目标函数的负梯度方向构造可行下降方向 $\boldsymbol{p}^k$，沿 $\boldsymbol{p}^k$ 搜索下一个可行解 $\boldsymbol{x}^{k+1}$。

计算目标函数的梯度

- 按基变量的定义，可行解 $x=\begin{pmatrix} x_B \\ x_N \end{pmatrix}$，约束矩阵 ***A*** 也相应划分为(***B***, ***N***)。不失一般性，假设 ***B*** 恰好由 ***A*** 的第 **1** 列～第 ***m*** 列组成，***N*** 因此由第 ***m*** **+ 1** 列～第 ***n*** 列组成。
- $Ax=b \Rightarrow Bx_B+Nx_N=b \Rightarrow x_B=B^{-1}b-B^{-1}Nx_N$。
- 因此目标函数 $f(x)=f(x_B,x_N)=f\left(B^{-1}b-B^{-1}Nx_N,x_N\right)$，记为 $\boldsymbol{F(x_N)}$。
- 下面计算 $\nabla F(x_N)=\left(\dfrac{\partial F}{\partial x_{m+1}}(x_N),\dfrac{\partial F}{\partial x_{m+2}}(x_N),\cdots,\dfrac{\partial F}{\partial x_n}(x_N)\right)^{\mathrm{T}}$。

计算目标函数的梯度

- 为此，将 $F(x_N)$ 重写为 $F(x_N)=f(u(x_N),v(x_N))$，其中 $u(x_N)=B^{-1}b-B^{-1}Nx_N$，$v(x_N)=x_N$。
- 应用复合函数求导法则，有：

$$
\begin{aligned}
&\frac{\partial F}{\partial x_{m+1}}(x_N)\\
&=\frac{\partial f(u,v)}{\partial u_1}\frac{\partial u_1(x_N)}{\partial x_{m+1}}+\frac{\partial f(u,v)}{\partial u_2}\frac{\partial u_2(x_N)}{\partial x_{m+1}}+\cdots+\frac{\partial f(u,v)}{\partial u_m}\frac{\partial u_m(x_N)}{\partial x_{m+1}}\\
&\quad+\frac{\partial f(u,v)}{\partial v_1}\frac{\partial v_1(x_N)}{\partial x_{m+1}}+\frac{\partial f(u,v)}{\partial v_2}\frac{\partial v_2(x_N)}{\partial x_{m+1}}+\cdots+\frac{\partial f(u,v)}{\partial v_{n-m}}\frac{\partial v_{n-m}(x_N)}{\partial x_{m+1}}
\end{aligned}
$$

其中 $\frac{\partial v_1(x_N)}{\partial x_{m+1}}=1$， $\frac{\partial v_2(x_N)}{\partial x_{m+1}}=0$， …， $\frac{\partial v_{n-m}(x_N)}{\partial x_{m+1}}=0$。

计算目标函数的梯度

- 因此，梯度向量$\nabla F\left(x_N\right)$可表示为：

$$\begin{pmatrix} \dfrac{\partial F\left(x_N\right)}{\partial x_{m+1}} \\ \dfrac{\partial F\left(x_N\right)}{\partial x_{m+2}} \\ \vdots \\ \dfrac{\partial F\left(x_N\right)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial u_1\left(x_N\right)}{\partial x_{m+1}} & \dfrac{\partial u_2\left(x_N\right)}{\partial x_{m+1}} & \cdots & \dfrac{\partial u_m\left(x_N\right)}{\partial x_{m+1}} \\ \dfrac{\partial u_1\left(x_N\right)}{\partial x_{m+2}} & \dfrac{\partial u_2\left(x_N\right)}{\partial x_{m+2}} & \cdots & \dfrac{\partial u_m\left(x_N\right)}{\partial x_{m+2}} \\ \vdots & & & \\ \times & \times & \cdots & \times \end{pmatrix} \begin{pmatrix} \dfrac{\partial f\left(u,v\right)}{\partial u_1} \\ \dfrac{\partial f\left(u,v\right)}{\partial u_2} \\ \vdots \\ \dfrac{\partial f\left(u,v\right)}{\partial u_m} \end{pmatrix} + \begin{pmatrix} \dfrac{\partial f\left(u,v\right)}{\partial v_1} \\ \dfrac{\partial f\left(u,v\right)}{\partial v_2} \\ \vdots \\ \dfrac{\partial f\left(u,v\right)}{\partial v_{n-m}} \end{pmatrix}$$

- 下面计算$\dfrac{\partial u_1}{\partial x_{m+1}}\left(x_N\right)$，$\dfrac{\partial u_2}{\partial x_{m+1}}\left(x_N\right)$，等项。

计算目标函数的梯度

- $u(x_N) = B^{-1}b - B^{-1}Nx_N$：

$$\begin{pmatrix} u_1(x_N) \\ u_2(x_N) \\ \vdots \\ u_m(x_N) \end{pmatrix} = \begin{pmatrix} \times \\ \times \\ \vdots \\ \times \end{pmatrix} - \begin{pmatrix} \times & \times & \cdots & \times \\ \times & \times & \cdots & \times \\ \vdots & & & \\ \times & \times & \cdots & \times \end{pmatrix} \begin{pmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{pmatrix}$$

- $\Rightarrow \dfrac{\partial u_1}{\partial x_{m+1}}(x_N) = -(B^{-1}N)_{11}$，$\dfrac{\partial u_2}{\partial x_{m+1}}(x_N) = -(B^{-1}N)_{21}$，…

$\dfrac{\partial u_1}{\partial x_{m+2}}(x_N) = -(B^{-1}N)_{12}$，…

计算目标函数的梯度

- 因此，$\nabla F(x_N) = -\left(B^{-1}N\right)^{\mathrm{T}} \nabla_u f(u,v) + \nabla_v f(u,v)$

$$= -\left(B^{-1}N\right)^{\mathrm{T}} \nabla_B f(x) + \nabla_N f(x),$$

记为r_N，称为简约梯度，

其中$\nabla_B f(x)$表示 ***f*** 对各基变量的偏导数组成的向量，$\nabla_N f(x)$表示 ***f*** 对各非基变量的偏导数组成的向量。

- 可行解 $\boldsymbol{x}^{\boldsymbol{k}}$ 的简约梯度为$r_N^k = -\left(\left(B^k\right)^{-1} N^k\right)^{\mathrm{T}} \nabla_B f(x^k) + \nabla_N f(x^k)$。

下面确定搜索方向 $p^k = \begin{pmatrix} p_B^k \\ p_N^k \end{pmatrix}$ 以构造下一个可行解

$x^{k+1} = x^k + t_k p^k$。先给出构造方法，再证明其可行、下降。

构造搜索方向

- 为了保持新的解值 $f\left(x^{k+1}\right)$ 下降， p_N^k 可取为 $-r_N^k$ 。但同时要保证 x^{k+1} 可行。因此， p_N^k 构造为：

$$p_i^k = \begin{cases} -r_i^k, & r_i^k \le 0 \\ -x_i^k r_i^k, & r_i^k > 0 \end{cases}，\quad i \in \operatorname{index}\left(N^k\right)。\qquad \textbf{(18a, 18b)}$$

- 为了保证下一个解 x^{k+1} 可行，必须有

$Ax^{k+1} = Ax^k + t_k Ap^k = b$， 因此 $Ap^k = 0$ 。

- $\Rightarrow B^k p_B^k + N^k p_N^k = 0$， $\Rightarrow p_B^k = -\left(B^k\right)^{-1} N^k p_N^k$ 。**(19)**

搜索方向可行下降

定理 4.5.3 对于 **LC-MP** 问题 $\min\{f(x) | Ax=b, x\geq 0\}$，设目标函数 $\boldsymbol{f}$ 可微，基变量、非基变量、非退化假设、搜索方向按如上约定。则：

（**1**） $p^k \neq 0 \Rightarrow \boldsymbol{p}^{\boldsymbol{k}}$ 是 $\boldsymbol{f}$ 在点 $\mathbf{x}^{\boldsymbol{k}}$ 处的可行下降方向；

（2） $p^k = 0 \Rightarrow \mathbf{x}^{\boldsymbol{k}}$ 是 **MP** 问题的 **K-T** 点。（（**2**）的逆也成立。）

- 证。（**1**）下一个探索点 $x^{k+1} = x^k + t_k p^k$，其中 $t_k > 0$。由 $\boldsymbol{p}^{\boldsymbol{k}}$ 的构造，已有 $Ax^{k+1} = b$。下面证 $x^{k+1} \geq 0$。
- $x_B^{k+1} = x_B^k + t_k p_B^k$。由非退化假设，$x_B^k > 0$。因此，只要 $\boldsymbol{t_k}$ 取得适当小，即有 $x_B^{k+1} > 0$。

搜索方向可行

- $x_N^{k+1}=x_N^k+t_k p_N^k$。由于 x_N^k 是非基变量，不能保证 $x_N^k>0$

  （但必有 $x_N^k\geq 0$☺）。令 $i\in \mathrm{index}\left(N^k\right)$，分两种情况讨论：

- 若 $x_i^k=0$，则

$$x_i^{k+1}=x_i^k+t_k p_i^k=t_k p_i^k=\begin{cases}-t_k r_i^k\geq 0, & r_i^k\leq 0\\ -t_k x_i^k r_i^k=0, & r_i^k>0\end{cases},$$

  因此总有 $x_i^{k+1}\geq 0$。

- 若 $x_i^k>0$，则可类比 x_B^k 的情形处理。
- 综上，对于适当选取的 $\boldsymbol{t_k}$，$\boldsymbol{x^{k+1}}$ 是可行解。

搜索方向下降

- 由定理 **4.4.1**，只要 $\nabla f\left(x^k\right)^{\mathrm{T}} p^k < 0$，则 $\boldsymbol{p}^{\boldsymbol{k}}$ 即是函数 $\boldsymbol{f}$ 在 $\boldsymbol{x}^{\boldsymbol{k}}$ 处的下降方向。

- $$\begin{aligned}\nabla f\left(x^k\right)^{\mathrm{T}} p^k &= \nabla_B f\left(x^k\right)^{\mathrm{T}} p_B^k + \nabla_N f\left(x^k\right)^{\mathrm{T}} p_N^k \\ &= \nabla_B f\left(x^k\right)^{\mathrm{T}}\left(-\left(B^k\right)^{-1} N^k p_N^k\right) + \nabla_N f\left(x^k\right)^{\mathrm{T}} p_N^k \\ &= \left(-\nabla_B f\left(x^k\right)^{\mathrm{T}}\left(B^k\right)^{-1} N^k + \nabla_N f\left(x^k\right)^{\mathrm{T}}\right) p_N^k \\ &= \left(r_N^k\right)^{\mathrm{T}} p_N^k = \sum_{i \in \mathrm{index}\left(N^k\right)} r_i^k p_i^k\end{aligned}$$ 。

- 而 $r_i^k p_i^k = \begin{cases} -\left(r_i^k\right)^2 \le 0, & r_i^k \le 0 \\ -x_i^k\left(r_i^k\right)^2 \le 0, & r_i^k > 0 \end{cases}$，因为 $x_i^k \ge 0$ 。

搜索方向下降

- 因为 $p^k = \left(-\left(B^k\right)^{-1} N^k p_N^k \quad p_N^k\right) \neq 0$，因此 $p_N^k \neq 0$。
- 由 p_N^k 的定义，可知 p_N^k 的各项（r_i^k **(18a)**，及 $-x_i^k r_i^k$ **(18b)**）不全为 **0**。（若**(18a)**的各项全为 **0**，则必有**(18b)**的某项不为 **0**。若**(18b)**的各项全为 **0**，则必有**(18a)**的某项不为 **0**。）
- 因此 $\nabla f\left(x^k\right)^{\mathrm{T}} p^k < 0$。
- 至此，已经证明了 $\boldsymbol{p}^k$ 既是可行方向，又是下降方向。因此（**1**）得证。

要满足的K-T条件

- **(2)，(⇒)。将 MP 问题重写为：**

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) = -x_i \le 0, \quad i = 1..n \\ & h_j(x) = a_j^{\mathrm{T}} x = 0, \quad j = 1..m \end{aligned}$$

，其中 a_j^{T} 为 $\boldsymbol{A}$ 的第 $\boldsymbol{j}$ 行。

- 要证 $\boldsymbol{x}^k$ 为 **K-T** 点，即要证明：

$$\begin{cases} \nabla f(x^k) + \sum \lambda_i^* \nabla g_i(x^k) + \sum \mu_j^* \nabla h_j(x^k) = 0 \\ \lambda_i^* g_i(x^k) = 0, \quad i = 1..n \\ \lambda_i^* \ge 0, \quad i = 1..n \end{cases}$$

，代入 **MP** 各项，

$$\begin{cases} \nabla f(x^k) - \lambda^* + A^{\mathrm{T}} \mu^* = 0 \\ (\lambda^*)^{\mathrm{T}} x^k = 0 \\ \lambda^* \ge 0 \end{cases}$$

。

对K-T条件各项的解释

- $g_i(x^k) = -x_i \Rightarrow \nabla g_i(x^k) = -e_i$，其中 $\boldsymbol{e_i} = (\mathbf{0\ 1} \cdots \mathbf{0})^{\mathrm{T}}$ 为第 ***i*** 项为 **1**，其余各项均为 **0** 的列向量。因此，

$$\sum_i \lambda_i^* \nabla g_i(x^k) = -\sum_i e_i \lambda_i^* = -I\lambda^* = -\lambda^*$$ 。

- $g_i(x^k) = -x_i \Rightarrow \lambda_i^* g_i(x^k) = -\lambda_i^* x_i$ 。
- $h_j(x^k) = a_j^{\mathrm{T}} x^k \Rightarrow \nabla h_j(x^k) = a_j$ 。于是，

$$\sum_j \mu_j^* \nabla h_j(x^k) = \sum_j a_j \mu_j^* = (a_1 \quad a_2 \quad \cdots \quad a_m) \begin{pmatrix} \mu_1^* \\ \mu_2^* \\ \vdots \\ \mu_m^* \end{pmatrix} = A^{\mathrm{T}} \mu^*$$ 。

要满足的K-T条件

- 按照 $x^k=\begin{pmatrix} x_B^k \\ x_N^k \end{pmatrix}$ 的划分，重新表示 $\lambda^*=\begin{pmatrix} \lambda_B^* \\ \lambda_N^* \end{pmatrix}$， $\mu^*=\begin{pmatrix} \mu_B^* \\ \mu_N^* \end{pmatrix}$。
- 于是 $\boldsymbol{x}^k$ 要满足的 **K-T** 条件可重写为：

$$\nabla_B f\left(x^k\right)-\lambda_B^*+\left(B^k\right)^{\mathrm{T}}\mu^*=0\,, \qquad \textbf{(1)}$$

$$\nabla_N f\left(x^k\right)-\lambda_N^*+\left(N^k\right)^{\mathrm{T}}\mu^*=0\,, \qquad \textbf{(2)}$$

$$\left(\lambda_B^*\right)^{\mathrm{T}} x_B^k=0\,, \qquad \textbf{(3)}$$

$$\left(\lambda_N^*\right)^{\mathrm{T}} x_N^k=0\,, \qquad \textbf{(4)}$$

$$\lambda_B^*\geq 0\,, \qquad \textbf{(5)}$$

$$\lambda_N^*\geq 0\,。 \qquad \textbf{(6)}$$

x^k满足K-T条件

- 由已知，$\boldsymbol{p}^{\boldsymbol{k}} = \mathbf{0}$。于是 $p_N^k = 0$ 。
- 由**(18a)**， $\forall i \in \operatorname{index}\left(N^k\right)$，当 $r_i^k \le 0$ 时 $p_i^k = -r_i^k$ 。 $p_N^k = 0$ 表明 r_i^k 不能严格小于 **0**。因此有 $r_N^k \ge 0$ 。 **(7)**
- 由 p_N^k 的定义（**18a, 18b**）， $p_N^k = 0$ 表明 $\forall i \in \operatorname{index}\left(N^k\right)$，或者 $r_i^k = 0$ ，或者 $x_i^k r_i^k = 0$，二者必居其一。因此，
$\left(r_N^k\right)^{\mathrm{T}} x_N^k = 0$ 。 **(8)**
- 定义 $\lambda_B^* = 0$ ， $\lambda_N^* = r_N^k$ ， $\mu^* = -B^{k,\mathrm{T},-1} \nabla_B f\left(x^k\right)$， **(9a, 9b, 9c)**
其中 $B^{k,\mathrm{T},-1} = \left(\left(B^k\right)^{\mathrm{T}}\right)^{-1}$ 。
- **(9a)** $\Rightarrow$ **(3)**，**(9b), (8)** $\Rightarrow$ **(4)**，**(9a)** $\Rightarrow$ **(5)**，**(9b), (7)** $\Rightarrow$ **(6)**。

x^k满足K-T条件

- 由**(9a)**和**(9c)**，可知

$$\nabla_B f\left(x^k\right)-\lambda_B^* +\left(B^k\right)^{\mathrm{T}}\mu^* =\nabla_B f\left(x^k\right)+\left(B^k\right)^{\mathrm{T}}\left(-B^{k,\mathrm{T},-1}\nabla_B f\left(x^k\right)\right)$$

$$=\nabla_B f\left(x^k\right)-\nabla_B f\left(x^k\right)=0\text{，}\textbf{(1)}\text{成立。}$$

- 由**(9a)**，**(9c)**，以及 r_N^k 的定义，

$$\nabla_N f\left(x^k\right)-\lambda_N^* +\left(N^k\right)^{\mathrm{T}}\mu^*$$

$$=\nabla_N f\left(x^k\right)-\left[-\left(\left(B^k\right)^{-1}N^k\right)^{\mathrm{T}}\nabla_B f\left(x^k\right)+\nabla_N f\left(x^k\right)\right]+\left(N^k\right)^{\mathrm{T}}\left(-B^{k,\mathrm{T},-1}\nabla_B f\left(x^k\right)\right)$$

$$=\nabla_N f\left(x^k\right)+\left(N^k\right)^{\mathrm{T}}B^{k,-1,\mathrm{T}}\nabla_B f\left(x^k\right)-\nabla_N f\left(x^k\right)-\left(N^k\right)^{\mathrm{T}}B^{k,\mathrm{T},-1}\nabla_B f\left(x^k\right)$$

$$=0\text{，}\textbf{(2)}\text{成立。}$$

- 综上，可知 $\boldsymbol{x^k}$ 满足 **K-T** 条件。
- 类似的证明，可知**(2)**的**(⇐)**成立。□

确定t_k的上界$t_k^{\max}$

- 由定理 **4.5.3** 的证明，$t_k > 0$ 要取为适当小的一个数。问题：t_k 需要小到什么界以内？
- 由新的探索点 x^{k+1} 的定义，有 $x_i^{k+1} = x_i^k + t_k p_i^k$， $i = 1..n$ 。
- t_k 的选择要保证 $x_i^{k+1} \geq 0$ 。
- 对于 $p_i^k \geq 0$ 的 i，则由 $x_i^k \geq 0$， $t_k \geq 0$， 已有 $x_i^{k+1} \geq 0$ 。
- 对于 $p_i^k < 0$ 的 i，应要求 $t_k \leq -\dfrac{x_i^k}{p_i^k}$ 。
- 综上，

$$t_k^{\max} = \begin{cases} +\infty, & p^k \geq 0 \\ \min\left\{-\dfrac{x_i^k}{p_i^k} \mid p_i^k < 0, i = 1..n\right\} & \text{存在} p_i^k < 0 \end{cases}。$$

Wolfe简约梯度法

（参数$\varepsilon>0$为终止误差。）

选取初始可行点 $\boldsymbol{x}^0$，$\boldsymbol{k} \leftarrow \mathbf{0}$。

while true do

    定义 $\boldsymbol{x}^k$的前 $\boldsymbol{m}$ 个最大分量的基变量。

    计算简约梯度r_N^k。

    按**(18)(19)**式构造可行下降方向 $\boldsymbol{p}^k$。

    if $|\boldsymbol{p}^k| \le \varepsilon$ **then** 停止迭代，退出循环。

    （一维搜索）求解$\min_{0\le t\le t_k^{\max}} f(x^k+tp^k)$，得到 $\boldsymbol{t_k}$。

    $\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \boldsymbol{t_k p}^k$，$\boldsymbol{k} \leftarrow \boldsymbol{k}+\mathbf{1}$。

endwhile

return $\boldsymbol{x}^k$。