第4章 非线性规划

4.5 约束最优化方法

约束最优化方法

一般的数学规划问题（带约束的非线性规划问题）：

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad \begin{cases} g_i(x) \le 0, & i=1,\cdots,p \\ h_j(x) = 0, & j=1,\cdots,q \end{cases} \end{aligned} \qquad \textbf{(MP)}$$

其中，$x \in \mathrm{R}^n$，$f, g_i, h_j : \mathrm{R}^n \mapsto \mathrm{R}$。

- 约束最优化问题的最优性条件
- 简约梯度法
- 惩罚函数法

1. Kuhn-Tucker条件

MP最优解的必要条件

- 记 X 为 MP 的可行域，指标集 $I = \{1, \ldots, p\}$, $J = \{1, \ldots, q\}$。
- 任给一个可行点 $x \in X$，记 $I(x) = \{i \mid g_i(x) = 0, i \in I\}$。即，$I(x)$和 J 均为取得等式约束的下标的集合。
- 称以等式（$g_i(x) = 0$）成立的约束 $g_i(x) \le 0$ 为积极约束。

K-T条件

定理 4.5.1 （**Kuhn**（**库恩**）**, Tucker**（**塔克**）**, 1951**）

(1) x^*是可行域 X 中的一个点。

(2) 设函数 f 和 g_i（$i \in I(x^*)$）在点 x^*处可微，

(3) g_i（$i \in I \setminus I(x^*)$）在点 x^*处连续，

(4) h_j（$j \in J$）在点 x^*处连续可微，

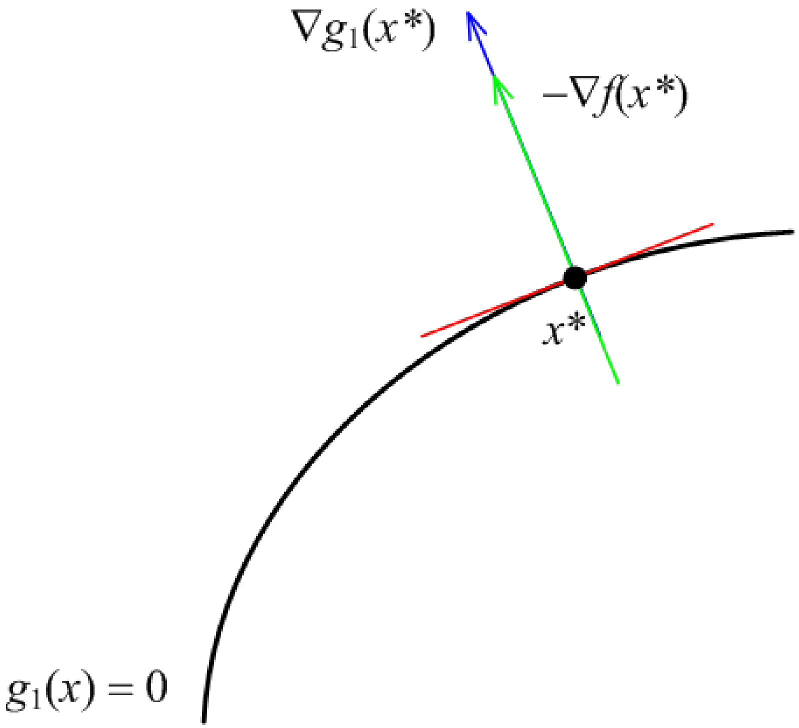
(5) 并且各 $\nabla g_i(x^*),\ i \in I(x^*)$，$\nabla h_j(x^*),\ j \in J$ 线性无关。

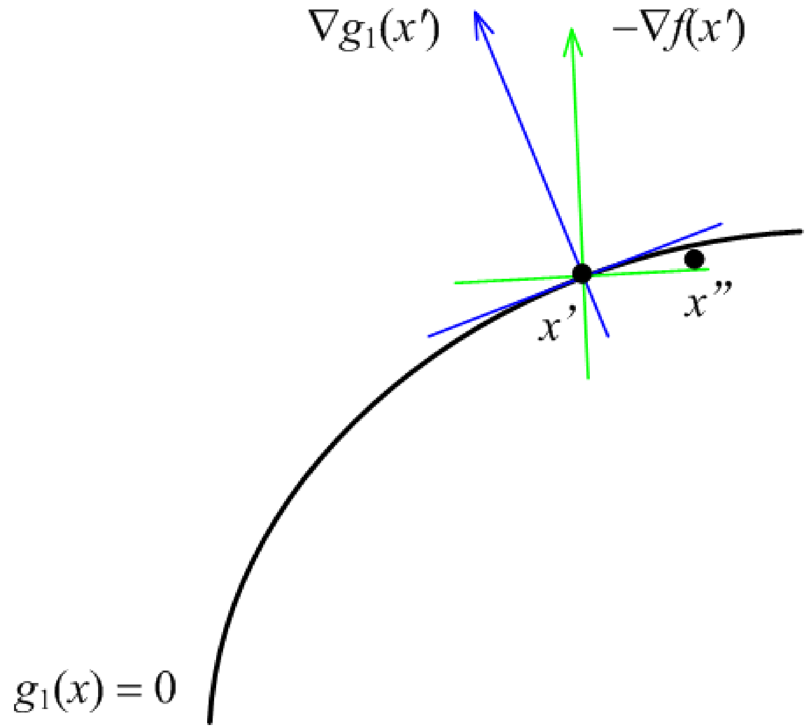
若 x^*是 MP 的局部最优解，则存在两组实数 $\lambda_i^*, i \in I(x^*)$ 和 $\mu_j^*, j \in J$，使得

$$\begin{cases} \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J} \mu_j^* \nabla h_j(x^*) = 0 \\ \lambda_i^* \geq 0, \qquad i \in I(x^*) \end{cases}$$

。

K-T条件的解释

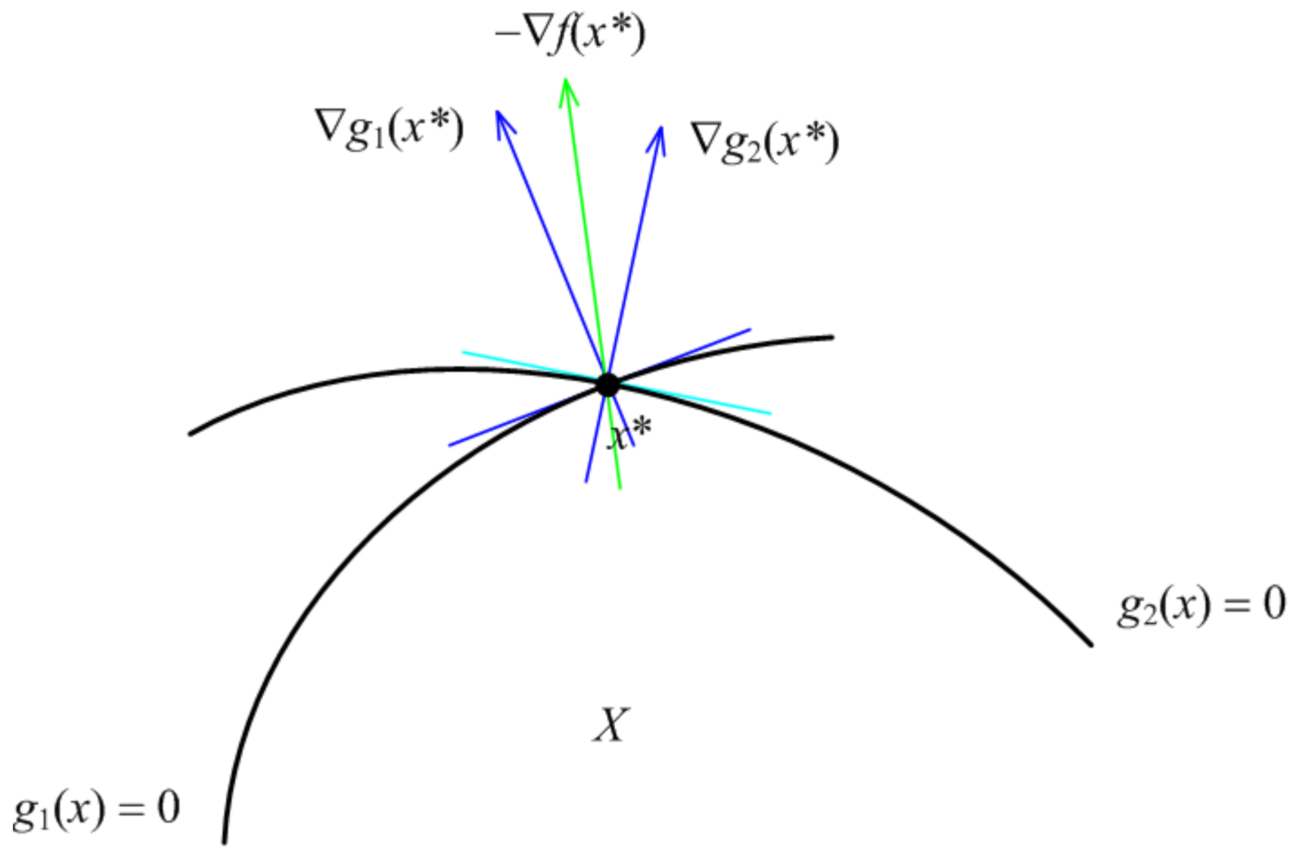




左图：$-\nabla f(x^*) = \lambda_1 \nabla g_1(x^*)$，$x^*$为局部最优解。

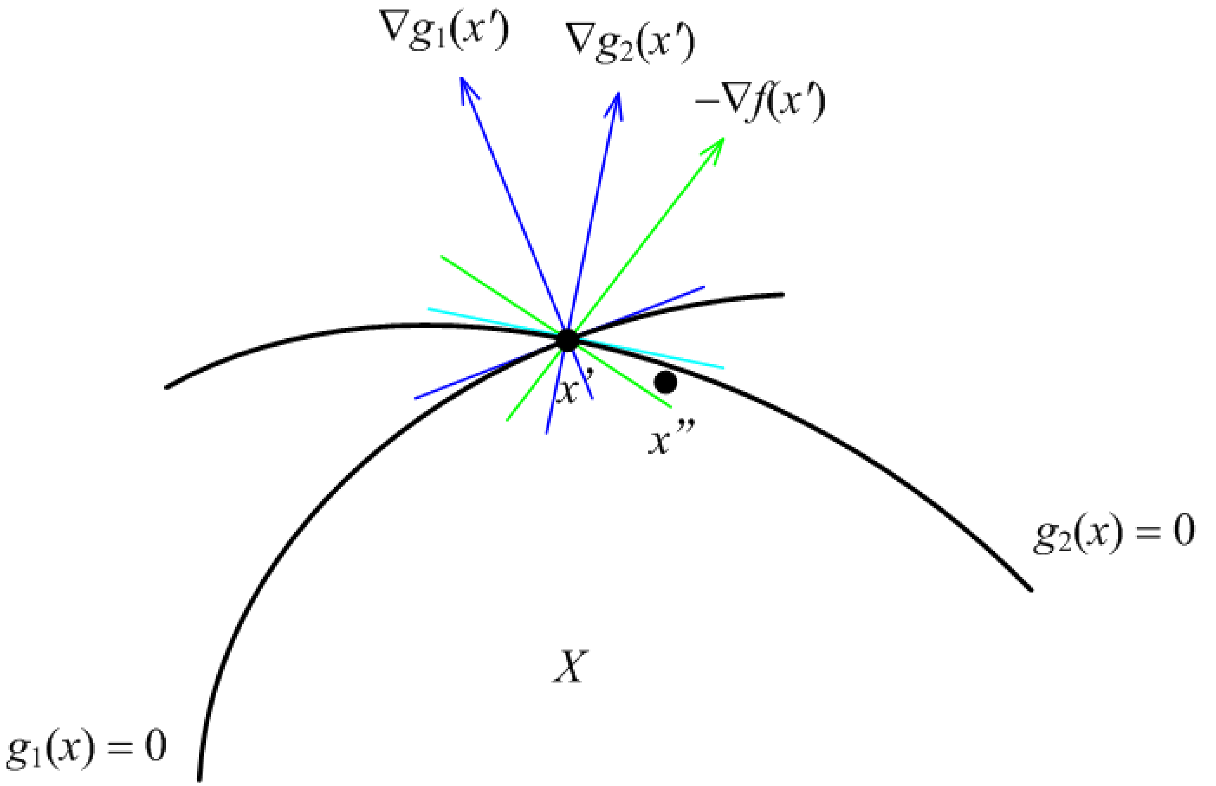
右图：$-\nabla f(x')$与$\nabla g_1(x')$方向不重合，x'还可更优。

K-T条件的解释



$-\nabla f(x^*)$ 位于 $\nabla g_1(x^*)$ 和 $\nabla g_1(x^*)$ 的夹角内（位于 $\nabla g_1(x^*)$ 和 $\nabla g_1(x^*)$ 所构成的锥内），$-\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$，$x^*$ 为局部最优解。

K-T条件的解释



$-\nabla f(x')$不位于$\nabla g_1(x')$和$\nabla g_1(x')$所构成的锥内，x'还可以更优。

两个概念

- **MP 问题若无等式约束，则 K-T 条件简化为**

  $$-\nabla f\left(x^*\right)=\sum_{i\in I\left(x^*\right)}\lambda_i^*\nabla g_i\left(x^*\right),\ \text{其中}\lambda_i^*\geq 0,\ \forall i\in I\left(x^*\right)。$$

  此时称$-\nabla f(x^*)$位于$\nabla g_i(x^*), i\in I(x^*)$所张成的锥内。

- **MP 问题若无不等式约束，则 K-T 条件简化为**

  $$-\nabla f\left(x^*\right)=\sum_{j\in J}\mu_j^*\nabla h_j\left(x^*\right),$$

  此时称$-\nabla f(x^*)$位于$\nabla h_j(x^*), j\in J$ 所生成的子空间内。

K-T条件的变化

定理 4.5.1’

(1) x^*是可行域 X 中的一个点。

(2) 设函数 f, g_i（$i \in I$），h_j（$j \in J$）在点 x^*处可微，

(3) 并且各 $\nabla g_i(x^*), i \in I(x^*)$，$\nabla h_j(x^*), j \in J$ 线性无关。

若 x^*是 MP 的局部最优解，则存在两组实数 $\lambda_i^*, i \in I$ 和 $\mu_j^*, j \in J$，使得

$$\begin{cases} \nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J} \mu_j^* \nabla h_j(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0, \quad i \in I \\ \lambda_i^* \geq 0, \quad i \in I \end{cases},$$

其中 $\lambda_i^* g_i(x^*) = 0$，$i \in I$ 称为互补松紧条件。

K-T条件的简写

记 $L(x,\lambda,\mu)=f(x)+\sum_{i\in I}\lambda_i g_i(x)+\sum_{j\in J}\mu_j h_j(x)$，则 K-T 条件可简写为：

$$\begin{cases} \nabla_x L(x^*,\lambda^*,\mu^*)=0 \\ \lambda_i^* g_i(x^*)=0, \quad i\in I \\ \lambda_i^* \geq 0, \quad i\in I \end{cases},$$

其中 $\nabla_x L(x^*,\lambda^*,\mu^*)$ 表示函数 $L(x,\lambda,\mu)$ 对 x 的梯度向量在 (x^*,λ^*,μ^*) 处的值。

$L(x,\lambda,\mu)$ 称为 Lagrange 函数，λ，μ称为 Lagrange 乘子。

K-T条件的有限充分性

定理 4.5.2

1. 设 x^*为 MP 问题的可行点，

2. 函数 f, g_i, h_j 在 x^*处连续可微，$\forall i \in I$，$\forall j \in J$，

3. f, g_i 是凸函数，h_j 是线性函数，$\forall i \in I(x^*)$，$\forall j \in J$。

若点 x^*满足 MP 的 K-T 条件，则 x^*是 MP 的整体最优解。

2. 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， $x \in \mathrm{R}^n$**，** $f : \mathrm{R}^n \mapsto \mathrm{R}$**，** $r(A_{m\times n}) = m$**，** $b \in \mathrm{R}^m$**。**

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

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）X_l 中的每一个可行解至少有 $\boldsymbol{m}$ 个大于零的分量；（2）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 ***F***(***x***$_N$)。
- 下面计算 $\nabla F(x_N)=\left(\frac{\partial F}{\partial x_{m+1}}(x_N),\frac{\partial F}{\partial x_{m+2}}(x_N),\cdots,\frac{\partial F}{\partial x_n}(x_N)\right)^{\mathrm{T}}$ 。

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：

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

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$\nabla F(x_N)$可表示为：

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

- 下面计算$\frac{\partial u_1}{\partial x_{m+1}}(x_N)$，$\frac{\partial u_2}{\partial x_{m+1}}(x_N)$，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x^k的简约梯度为$r_N^k=-\left(\left(B^k\right)^{-1}N^k\right)^{\mathrm{T}}\nabla_B f\left(x^k\right)+\nabla_N f\left(x^k\right)$。

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：

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

- 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 $\min\{f(x) \mid Ax=b, x\geq 0\}$，设目标函数 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） $p^k \neq 0 \Rightarrow \boldsymbol{p}^k$ 是 f 在点 $\boldsymbol{x}^k$ 处的可行下降方向；

（2） $p^k = 0 \Rightarrow \boldsymbol{x}^k$ 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 $\boldsymbol{p}^k$ 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 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 $\boldsymbol{p}^k$ 即是函数 $\boldsymbol{f}$ 在 $\boldsymbol{x}^k$ 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 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 A 的第 j 行。

- 要证 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 $\boldsymbol{e_i} = (\boldsymbol{0\ 1\ \cdots\ 0})^{\mathrm{T}}$ 为第 $\boldsymbol{i}$ 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，

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

。

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 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，$\boldsymbol{p}^k = \mathbf{0}$。于是 $p_N^k = 0$。
- 由(18a)，$\forall i \in \mathrm{index}\left(N^k\right)$，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 $\forall i \in \mathrm{index}\left(N^k\right)$，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) ⇒ (3)**，**(9b), (8) ⇒ (4)**，**(9a) ⇒ (5)**，**(9b), (7) ⇒ (6)**。

x^k满足K-T条件

- 由(9a)和(9c)，可知

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

$$= \nabla_B f(x^k) - \nabla_B f(x^k) = 0$$，(1)成立。

- 由(9a)，(9c)，以及 r_N^k 的定义，

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

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

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

$= 0$，(2)成立。

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

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

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

Wolfe简约梯度法

- 例4.5.2 用Wolfe法求解约束极小化问题

$$
\begin{cases}
\min & x_1^2 + x_2^2 + 2x_1x_2 + 2x_1 + 6x_2 \\
s.t. & x_1 + x_2 \le 4 \\
 & -x_1 + x_2 \le 2 \\
 & x_1, x_2 \ge 0
\end{cases}
$$

取 $x^0 = (1,1)^T, \varepsilon = 10^{-6}$

Wolfe简约梯度法

解 首先将问题化为如下的问题

$$\begin{cases} \min & x_1^2+x_2^2+2x_1x_2+2x_1+6x_2, \\ s.t. & x_1+x_2+x_3=4, \\ & -x_1+x_2+x_4=2, \\ & x_j \geq 0, \quad j=1,2,3,4 \end{cases}$$

取 $x^0=(1,1,2,2)^{\mathrm{T}}, \varepsilon=10^{-6}$.

Wolfe简约梯度法

第 1 轮迭代. 对于$x^0=(1,1,2,2)^T$, 有$\mathrm{I}_B^0=\{3,4\}$, 故矩阵

$$A=\begin{pmatrix}1 & 1 & 1 & 0\\ -1 & 1 & 0 & 1\end{pmatrix}$$

相应分解为

$$B_0=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad N_0=\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$$

因为$\nabla\mathrm{f}(\mathrm{x})=\ (2x_1+2x_2+2,2x_2+2x_1+6,0,0)^T$

所以$\nabla\mathrm{f}(\mathrm{x}^0)=\ (6,10,0,0)^T$.据此可求得$\mathrm{x}^0$点的简约梯度

$$\mathrm{r}_N^0=-(\mathrm{B}_0^{-1}N_0)^T\nabla_B\mathrm{f}(\mathrm{x}^0)+\nabla_N\mathrm{f}(\mathrm{x}^0)$$

$$=-\left[\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}\right]^T\begin{pmatrix}0\\ 0\end{pmatrix}+\begin{pmatrix}6\\ 10\end{pmatrix}=\begin{pmatrix}6\\ 10\end{pmatrix}$$

Wolfe简约梯度法

有

$$p_N^0=\begin{pmatrix}-6\\-10\end{pmatrix}, p_B^0=-B_0^{-1}N_0p_N^0=-\begin{pmatrix}1&0\\0&1\end{pmatrix}^{-1}\begin{pmatrix}1&1\\-1&1\end{pmatrix}\begin{pmatrix}-6\\-10\end{pmatrix}=\begin{pmatrix}16\\4\end{pmatrix}$$

因此得到可行下降方向 $p^0=(-6,-10,16,4)^T$

由于$\|p^0\|\nleq 10^{-6}$，要沿p^0进行有效一维搜索. 并确定

$$t_{max}^0=\min\{\frac{1}{6},\frac{1}{10}\}=\frac{1}{10}$$

求解 $\min\limits_{0\le t\le\frac{1}{10}} f(x^0+t_0p^0)=256t^2-136t+12$

得到最优解$t_0=\frac{1}{10}$, 于是得到下一个迭代点

$$x^1=x^0+t_0p^0=(1,1,2,2)^T+\frac{1}{10}(-6,-10,16,4)^T$$

$$=\left(\frac{2}{5},0,\frac{18}{5},\frac{12}{5}\right)^T$$

Wolfe简约梯度法

第 2 轮迭代. 对于x^1, 有$I_B^1 = \{3,4\}$. 故对矩阵 A 的分解仍有

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad N_1 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

由$\nabla f(x^1) = \left(\frac{14}{5}, \frac{34}{5}, 0, 0\right)^T$, 得到简约梯度

$$r_N^1 = -(B_1^{-1} N_1)^T \nabla_B f(x^1) + \nabla_N f(x^1)$$

$$= -\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\right]^T \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{14}{5} \\ \frac{34}{5} \end{pmatrix} = \begin{pmatrix} \frac{14}{5} \\ \frac{34}{5} \end{pmatrix}$$

又有

$$p_N^1 = \begin{pmatrix} -\frac{28}{25} \\ 0 \end{pmatrix}, p_B^1 = -B_1^{-1} N_1 p_N^1 = \begin{pmatrix} \frac{28}{25} \\ -\frac{28}{25} \end{pmatrix}$$

Wolfe简约梯度法

所以可行下降方向为$p^1=(-\frac{28}{25},0,\frac{28}{25},-\frac{28}{25})^T$

由于$\|p^1\|\not\leq 10^{-6}$, 为计算简便, 不妨重取p^1为

$$p^1=(-1,0,1,-1)^T$$

并有

$$t_{max}^1=\min\{\frac{\frac{2}{5}}{1},\frac{\frac{12}{5}}{1}\}=\frac{2}{5}$$

求解 $\min\limits_{0\leq t\leq\frac{2}{5}} f(x^1+tp^1)=(\frac{2}{5}-t)^2+2(\frac{2}{5}-t)$

求得$t_1=\frac{2}{5}$, 所以

$$x^2=x^1+t_1p^1=(\frac{2}{5},0,\frac{18}{5},\frac{12}{5})^T+\frac{2}{5}(-1,0,1,-1)^T$$

$$=(0,0,4,2)^T$$

Wolfe简约梯度法

第 3 轮迭代. 对于x^2, 有$I_B^2 = \{3,4\}$. 故对矩阵 A 的分解仍为

$$B_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

因为$\nabla f(x^2) = (2,6,0,0)^T$, 由公式求得$r_N^2 = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$，并求得

$p_N^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, P_B^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. 因而有

$$p^2 = (0,0,0,0)^T$$

$\|p^2\| = 0 < 10^{-6}$. 由定理知$x^2 = (0,0,4,2)^T$为初始问题的 K-T 点. 显然是一个凸规划，因此所求得的x^2为其整体最优解. 原问题的整体最优解为$x^* = (0,0)^T$

谢谢大家