SLP & SQP Methods

LI-PING ZHANG

Department of Mathematical Sciences

Tsinghua University, Beijing 100084

Office: New Science Building #A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Successive Linear Programming Method
- Frank-Wolfe Method
- Successive Quadratic Programming Method

序列线性规划法

基本思想：将非线性规划线性化，通过解线性规划问题来求解原问题的近似解。

$$
(NP)\quad \begin{cases} \min\ f(x) \\ s.t.\quad g_j(x) \geq 0 \quad j=1,\cdots,m \\ \qquad\quad h_j(x)=0 \quad j=1,\cdots,l \end{cases}
$$

其中$x \in R^n, f, g_i, h_j$均存在一阶连续偏导数。

清華大學

基本思想：

将(NP)中的目标函数$f(x)$和约束函数$g_i(x)$, $h_j(x)$线性化，并对变量的取值范围加以限制，从而得到线性近似规划，用单纯形方法求解此线性规划问题，把其最优解作为(NP)的解的近似。

设$x^{(k)}$是原问题的可行解，将$f(x), g_i(x)$, $h_j(x)$在$x^{(k)}$点$Taylor$展开。

$$\begin{cases} \min\ f(x^{(k)}) + \nabla f(x^{(k)})^T(x - x^{(k)}) \\ s.t.\quad g_i(x^{(k)}) + \nabla g_i(x^{(k)})^T(x - x^{(k)}) \geq 0, i = 1, \cdots m \\ \qquad\ \ h_j(x^{(k)}) + \nabla h_j(x^{(k)})^T(x - x^{(k)}) = 0, j = 1, \cdots l \end{cases}$$

$$
\begin{cases}
\min\ f(x^{(k)})+\nabla f(x^{(k)})^T(x-x^{(k)}) \\
s.t.\quad g_i(x^{(k)})+\nabla g_i(x^{(k)})^T(x-x^{(k)})\geq 0, i=1,\cdots m \\
\qquad\quad h_j(x^{(k)})+\nabla h_j(x^{(k)})^T(x-x^{(k)})=0, j=1,\cdots l \\
\qquad\quad |x_j-x_j^{(k)}|\leq \delta_j^{(k)},\quad j=1,2,\cdots,n
\end{cases}
$$

为了保证有最优解，限制在紧集上

步骤

1. 给定初始可行解$x^{(1)}$，$\delta_j^{(1)}, j=1,2,\cdots,n$，缩小误差$\beta\in(0,1)$，允许误差$\varepsilon_1,\varepsilon_2$，置$k:=1$。

2. 求解线性规划问题：

$$\begin{cases}\min f(x^{(k)})+\nabla f(x^{(k)})^T(x-x^{(k)})\\ s.t.\quad g_i(x^{(k)})+\nabla g_i(x^{(k)})^T(x-x^{(k)})\ge 0, i=1,\cdots m\\ \qquad h_j(x^{(k)})+\nabla h_j(x^{(k)})^T(x-x^{(k)})=0, j=1,\cdots l\\ \qquad |x_j-x_j^{(k)}|\le\delta_j^{(k)},\quad j=1,2,\cdots,n\end{cases}$$

得最优解$\bar{x}$.

3. 若$\bar{x}$是(NP)的可行解，则令$x^{(k+1)}=\bar{x}$，转4；否则，置$\delta_j^{(k)}:=\beta\delta_j^{(k)}, j=1,2,\cdots,n$，返回2。

4. 若$|f(x^{(k+1)})-f(x^{(k)})|<\varepsilon_1$，且$\left\|x^{(k+1)}-x^{(k)}\right\|<\varepsilon_2$或$|\delta_j^{(k)}|<\varepsilon_2, j=1,2,\cdots,n$，则点$x^{(k+1)}$为近似最优解；否则，令$\delta_j^{(k+1)}:=\delta_j^{(k)}$，$j=1,2,\cdots,n$，置$k:=k+1$，返回2。

Frank-Wolfe方 法

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

线性约束

$A_{m\times n}, r(A)=m, E_{l\times n}, f(x)$连续可微

令 $S=\{x \mid Ax\geq b, Ex=e, x\in R^n\}$

基本思想：在每次迭代中，将目标函数$f(x)$线性化，通过解线性规划求得下降可行方向，进而沿此方向在可行域内做一维搜索。

任取$x^{(k)} \in S$，在$x^{(k)}$处以$f(x)$的一阶$Taylor$展开式作为$f(x)$的线性逼近函数：

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

求解线性规划问题

$$
\begin{array}{ll} \min & f_L(x) \\ s.t. & x \in S \end{array}
\quad \Longleftrightarrow \quad
(2) \quad \begin{cases} \min \nabla f(x^{(k)})^T x \\ s.t. \quad x \in S \end{cases}
$$

设问题(2)存在有限最优解$y^{(k)}$.

清華大學

求解(2)有下列两种情形之一：

1. 若$\nabla f(x^{(k)})^T(y^{(k)}-x^{(k)})=0$，则停止迭代，$x^{(k)}$是(1)的$KKT$点。 则由yk最优知xk为最优解

2. 若$\nabla f(x^{(k)})^T(y^{(k)}-x^{(k)})\neq 0$，则必有

$$\nabla f(x^{(k)})^T(y^{(k)}-x^{(k)})<0$$

$$\boxed{\begin{cases}\min\ \nabla f(x^{(k)})^T x\\ s.t.\quad x\in S\end{cases}}$$

$\Rightarrow y^{(k)}-x^{(k)}$为$x^{(k)}$处的下降方向。

$\because S$是凸集，$\therefore$ 对$\forall\lambda\in(0,1)$，有

$$\lambda y^{(k)}+(1-\lambda)x^{(k)}=x^{(k)}+\lambda(y^{(k)}-x^{(k)})\in S$$

$\Rightarrow y^{(k)}-x^{(k)}$为$x^{(k)}$处的可行方向。

$\Rightarrow y^{(k)}-x^{(k)}$为$x^{(k)}$处的下降可行方向。

清華大學

定理： 设$y^{(k)}$是(2)的最优解，且满足

$\nabla f(x^{(k)})^T (y^{(k)} - x^{(k)}) = 0$，则$x^{(k)}$是(1)的$KKT$点。

证明：由$\nabla f(x^{(k)})^T (y^{(k)} - x^{(k)}) = 0$，得

$$\nabla f(x^{(k)})^T y^{(k)} = \nabla f(x^{(k)})^T x^{(k)}$$

$\because y^{(k)}$是(2)的最优解，且$x^{(k)} \in S$

$\therefore x^{(k)}$是(2)的KKT点$\Rightarrow \exists w \ge 0 (w \in E^m), v \in E^n$使得

$$\begin{cases} \nabla f(x^{(k)}) - A^T w - E^T v = 0 \\ w^T \left(A x^{(k)} - b \right) = 0 \\ E x^{(k)} = e \end{cases}$$

$$\boxed{\begin{cases} \min \nabla f(x^{(k)})^T x \\ s.t. \quad x \in S \end{cases}}$$

这也是(1)的KKT条件$\therefore x^{(k)}$是(1)的KKT点。

清華大學

步骤

1.给定初始可行点 $x^{(1)}$，允许误差 $\varepsilon > 0$，置 $k = 1$。

2. 求解下面的线性规划问题得到最优解 $y^{(k)}$。

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

3. 若 $\left|\nabla f(x^{(k)})^T (y^{(k)} - x^{(k)})\right| \le \varepsilon$,则停止，得 $x^{(k)}$；否则转 4。

4. 从 $x^{(k)}$ 出发，沿方向 $y^{(k)} - x^{(k)}$ 进行一维搜索：

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

得最优解 λ_k.

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

清華大學

例：
$$\begin{cases} \min\ f(x) = 4x_1^2 + (x_2 - 2)^2 \\ s.t.\quad -2 \le x_1 \le 2 \\ \qquad\quad -1 \le x_2 \le 1 \end{cases}$$

解：取初始点$x^{(1)} = (-2, -1)^T$

第一次迭代 $\nabla f(x^{(1)}) = (-16, -6)^T$

$$\text{求}\begin{cases} \min\ \nabla f(x^{(1)})^T x = -16x_1 - 6x_2 \\ s.t.\quad -2 \le x_1 \le 2 \\ \qquad\quad -1 \le x_2 \le 1 \end{cases}$$

得$y^{(1)} = (2, 1)^T$.

从$x^{(1)}$出发，沿$y^{(1)}-x^{(1)}$作一维搜索

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

得$\lambda_1 = 0.56$

$$\therefore x^{(2)} = x^{(1)} + \lambda_1(y^{(1)} - x^{(1)}) = (0.24, 0.12)^T$$

第二次迭 代

$$求\begin{cases} \min \nabla f(x^{(2)})^T x = 1.92x_1 - 3.76x_2 \\ s.t. \quad -2 \le x_1 \le 2 \\ \qquad\quad -1 \le x_2 \le 1 \end{cases}$$

得$y^{(2)} = (-2, 1)^T$.

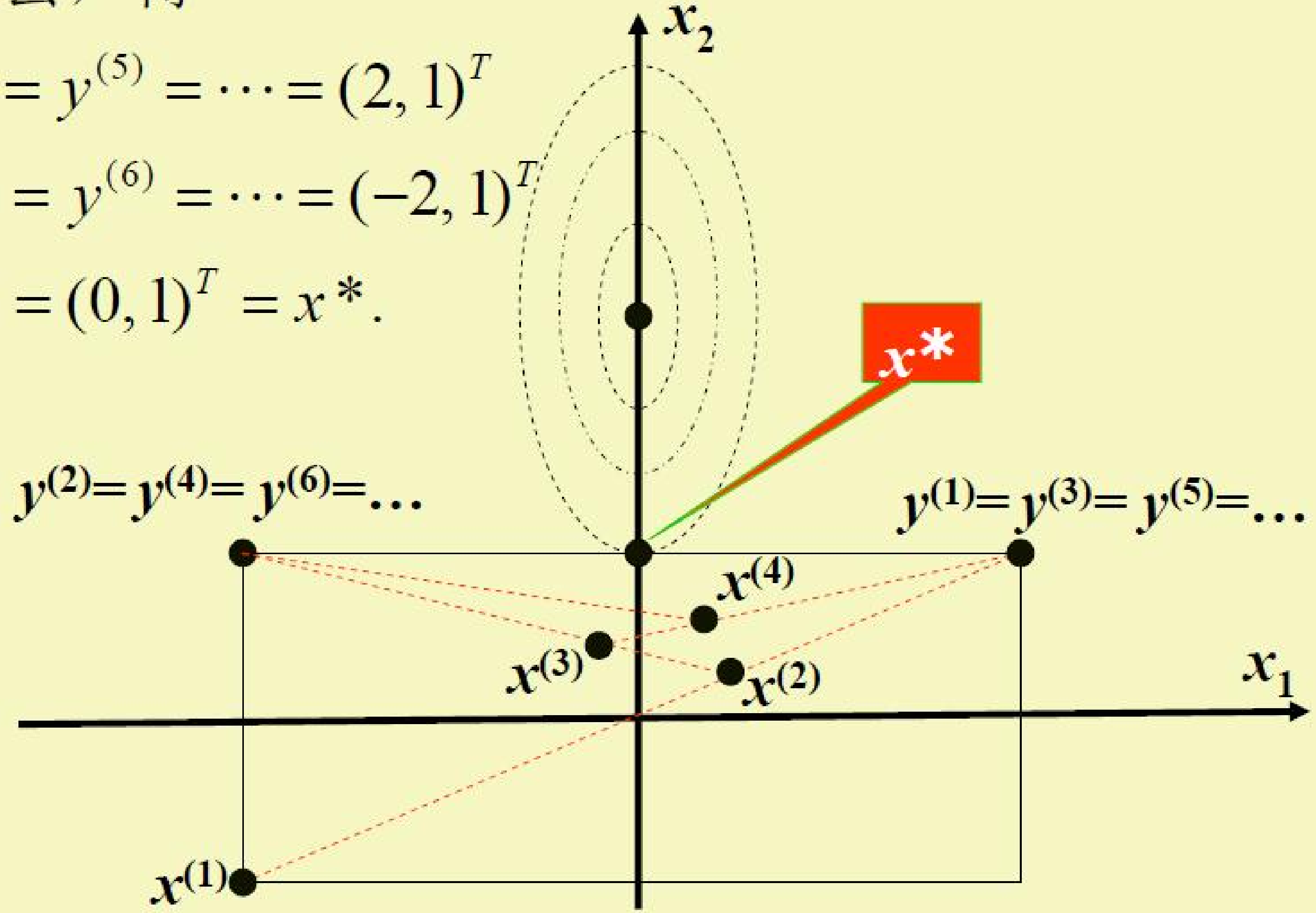
从$x^{(2)}$出发，沿$y^{(2)}-x^{(2)}$作一维搜索，得$x^{(3)}$，

这样继续下去，得

$$y^{(1)}=y^{(3)}=y^{(5)}=\cdots=(2,1)^T$$

$$y^{(2)}=y^{(4)}=y^{(6)}=\cdots=(-2,1)^T$$

显然$\lim\limits_{k\to\infty}x^{(k)}=(0,1)^T=x^*.$



Main idea of SQP

Successive quadratic programming (SQP) method was first proposed by Wilson (1963) and developed by Han (1976) and Powell (1977). Hence, it is also called the Wilson-Han-Powell method. This method is based on directly solving the Lagrange first-order necessary conditions of constrained optimization problems. The main idea is to solve the KKT system by using quasi-Newton method at each iteration and to reformulate the KKT system into a quadratic programming subproblem.

Lagrange-Newton method

Consider the following equality-constrained nonlinear programming problem

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & h_i(x) = 0, \quad i \in \mathcal{E}. \end{aligned} \tag{1}$$

The KKT system of (1) is written as

$$\begin{aligned} & \nabla f(x) - \textstyle\sum_{i \in \mathcal{E}} \lambda_i^T \nabla h_i(x) = 0 \\ & -h_i(x) = 0, \ i \in \mathcal{E} \end{aligned} \tag{2}$$

and its Lagrange function is

$$L(x, \lambda) = f(x) - \lambda^T h(x).$$

We use Newton method to solve (2) and the current iterate is (x^k, λ^k). Let $h_i(x), \ i \in \mathcal{E}$ group into a vector $h(x)$ and let $\nabla h_i(x), \ i \in \mathcal{E}$ group into a matrix $H(x)$. Then the Newton-step procedure is

$$\nabla_x^2 L(x^k, \lambda^k)(x - x^k) - H(x^k)(\lambda - \lambda^k) = -(\nabla f(x^k) - H(x^k)\lambda^k)$$
$$H(x^k)(x - x^k) = -h(x^k),$$

which can be rewritten as

$$\begin{aligned} &\nabla_x^2 L(x^k, \lambda^k)d - H(x^k)\lambda = -\nabla f(x^k) \\ &H(x^k)d = -h(x^k). \end{aligned} \tag{3}$$

Solving the above system of linear equations, we obtain (d^k, λ^{k+1}) and then derive a new iterate $(x^{k+1} = x^k + d^k, \lambda^{k+1})$. This is called Lagrange-Newton method.

SQP method

直接求(3)不方便因此将其转化为别的问题的KKT条件

Obviously, (3) is equivalent to the following equality-constrained quadratic programming problem

$$\begin{array}{ll} \min & f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla_x^2 L(x^k, \lambda^k) d \\ s.t. & h_i(x^k) + \nabla h_i(x^k)^T d = 0, \quad i \in \mathcal{E} \end{array}$$

Thus, we can extend the Lagrange-Newton method to solve the mixed constrained nonlinear programming:

$$\begin{array}{ll} \min & f(x) \\ s.t. & h_i(x) = 0, \ \ i \in \mathcal{E} \\ & h_i(x) \geq 0, \ \ i \in \mathcal{I}. \end{array} \tag{4}$$

Here, the direction d^k is obtained by solving the quadratic program

$$\begin{aligned} \min \quad & f(x^k) + \nabla f(x^k)^T d + \tfrac{1}{2} d^T \nabla_x^2 L(x^k, \lambda^k) d \\ s.t. \quad & h_i(x^k) + \nabla h_i(x^k)^T d = 0, \quad i \in \mathcal{E} \\ & h_i(x^k) + \nabla h_i(x^k)^T d \geq 0, \quad i \in \mathcal{I}, \end{aligned}$$

where

$$L(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x).$$

If $d^k = 0$, then x^k is a KKT point of (4).

There are two disadvantages: to compute the Hessian matrix of the Lagrange function $L(x, \lambda)$; to guarantee the positive definiteness of the Hessian matrix. Hence, we use the approximation of $\nabla_x^2 L(x^k, \lambda^k)$ based on the quasi-Newton method. Thus, we describe the SQP method as follows.

0. Given a starting point $x^0 \in \mathcal{R}^n$ and an initial positive definite matrix B_0. Set $k = 0$.

1. Compute d^k by solving the quadratic program

$$
\begin{array}{rll}
 & \min & f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d \\
Q(x^k, B_k) & s.t. & h_i(x^k) + \nabla h_i(x^k)^T d = 0, \quad i \in \mathcal{E} \\
 & & h_i(x^k) + \nabla h_i(x^k)^T d \geq 0, \quad i \in \mathcal{I}.
\end{array}
$$

2. If $d^k = 0$, stop; x^k is a KKT point of (4). Otherwise, let $x^{k+1} = x^k + \alpha_k d^k$ where α_k is determined by a certain line search.

3. Update B_k to form B_{k+1} such that B_{k+1} is positive definite. Set $k = k + 1$ and go to Step 1.

这里的精确搜索可以由加上罚函数的目标函数(此时相当于无约束)来确定

$$d^k = \left[I - A^T(AA^T)^{-1}A\right]\nabla f(x)$$

可以利用这个投影矩阵，考虑有约束的规划的搜索

Update B_k

Let

$$L(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i h_i(x),$$

where $\lambda_i,\ i \in \mathcal{E}\cup\mathcal{I}$ are optimal multipliers of $Q(x,B)$. Generally, we use the BFGS formulation to update B_k

$$B_{k+1} = B_k + \frac{\gamma_k\gamma_k^T}{\gamma_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k},$$

where

$$s_k = x^{k+1} - x^k, \quad \gamma_k = \nabla_x L(x^{k+1},\lambda^{k+1}) - \nabla_x L(x^k,\lambda^k).$$

理论上需要 $\gamma_k^T s_k > 0$ 但是数值上需要大于某个定值

To guarantee the positive definiteness, Powell (1978) introduced the following formulation:

$$B_{k+1} = B_k + \frac{\eta_k \eta_k^T}{\eta_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k},$$

where $\eta_k = \theta_k \gamma_k + (1 - \theta_k) B_k s_k$, and

$$\theta_k = \begin{cases} 1, & \textbf{if } s_k^T \gamma_k \geq 0.2 s_k^T B_k s_k \\ \dfrac{0.8 s_k^T B_k s_k}{s_k^T B_k s_k - s_k^T \gamma_k}, & \textbf{otherwise} \end{cases}$$

Clearly, $s_k^T \eta_k \geq 0.2 s_k^T B_k s_k$. When B_k is positive definite, so is B_{k+1} from Theorem 6 in Lecture Note #9 (Page 67).

Remarks

- SQP method is globally convergent and has the rate of superlinear convergence under some suitable assumptions.
- Maratos observed that the unit step-size is not reached even at a superlinearly convergent step. Thus, the rate of convergence of the SQP method is only linear. This phenomenon is called “Maratos effect”.