一般的约束优化问题

$c_1(x)=0$
$c_2(x)=0$
$\vdots$
$c_m(x)=0$

$$\min\ f(x)$$
$$\text{s.t.}\quad c_i(x)=0 \quad i\in\mathcal{E} \quad |\mathcal{E}|=m$$
$$c_{2j}(x)\le 0 \quad j\in\mathcal{I} \quad |\mathcal{I}|=p$$

$c_{1i}:\mathbb{R}^n\to\mathbb{R}^m$
$c_{2j}:\mathbb{R}^n\to\mathbb{R}^p$

最优值: $p^*=f(x^*)$

可行域 $\Omega=\{x\in\mathbb{R}^n \mid c_{1i}(x)=0,\ c_{2j}(x)\le 0,\ \forall i\in\mathcal{E},\ j\in\mathcal{I}\}$

Lagrangian 函数: $\mathcal{L}:\mathbb{R}^n\to\mathbb{R}$

$$\mathcal{L}(x;\lambda,\mu)=f(x)+\langle\lambda,c_1(x)\rangle+\langle\mu,c_2(x)\rangle$$

$\min f(x)=0$

$c_1(x)=0$

$c_2(x)=0$

$c_3(x)\le 0$

$c_4(x)\le 0$

$\bar{x}$: $\begin{cases} c_1(\bar{x})=0 \\ c_2(\bar{x})=0 \\ c_3(\bar{x})<0 \\ c_4(\bar{x})=0 \end{cases}$

$\Rightarrow \mathcal{A}(\bar{x})=\{1,2,4\}$

Defn: 积极集 (active set) 对于可行点 $x\in\Omega$, 该点处的积极集 $\mathcal{A}(x)$ 定义为

$$\mathcal{A}(x)=\mathcal{E}\cup\{j\in\mathcal{I}: c_{2j}(x)=0\}$$

Defn: 线性无关约束品性 (LICQ. Linear Independent Constraint Qualification)

给定 $x\in\Omega$, 及相对应的积极集 $\mathcal{A}(x)$. 若 $\nabla c_i(x)$, $i\in\mathcal{A}(x)$ 是线性无关的, 则称 LICQ 在 x 点成立.

$c_1 = 2x_1+3x_2+4x_3=0$ $\quad \nabla c_1=\begin{bmatrix}2\\3\\4\end{bmatrix}$

$c_2 = 4x_1+6x_2+8x_3=0$ $\quad \nabla c_2=\begin{bmatrix}4\\6\\8\end{bmatrix}$

$c_3 = 3x_1+7x_2+x_3=0$ $\quad \nabla c_3=\begin{bmatrix}3\\7\\1\end{bmatrix}$

$$\min f(x)$$
$$\text{s.t.} \quad c_i(x) = 0 \quad i \in \mathcal{E} \quad |\mathcal{E}| = m$$
$$c_j(x) \le 0 \quad j \in \mathcal{I} \quad |\mathcal{I}| = p$$

$C_1(x): \mathbb{R}^n \to \mathbb{R}^m$

$\nabla C_1 \in \mathbb{R}^{m \times n} \to$ Jacobian 矩阵

$C_1 \to [\, c_{11} \quad c_{12} \quad c_{13} \cdots c_{1m} \,]$

$$\nabla C_1 \to \begin{bmatrix} - & \nabla c_{11} & - \\ - & \nabla c_{12} & - \\ & \vdots & \\ - & \nabla c_{1m} & - \end{bmatrix}$$

Lagrangian 函数: $\mathcal{L}: \mathbb{R}^n \to \mathbb{R}$

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \langle \lambda, C_1(x) \rangle + \langle \mu, C_2(x) \rangle$$

Karush-Kuhn-Tucker (KKT 条件)

Thm 设 $f(x)$, $C_1(x)$, $C_2(x)$ 可微, 若 x^* 是原问题的一个局部最优点且 x 点处 LICQ 成立, 那么存在 (λ^*, μ^*) 使得

KKT 条件

稳定性条件 $\nabla \mathcal{L}(x^*; \lambda^*, \mu^*) = \nabla f(x^*) + \nabla C_1^T(x)\lambda + \nabla C_2^T(x)\mu = 0$

原问题可行性 $c_1(x^*) = 0$

$c_2(x^*) \le 0 \qquad \mu^* \ge 0$ 对偶可行性

互补松弛条件 $\langle (\mu^*), c_2(x^*) \rangle = 0$ → Variational Inequality

$$c_2(x^*) = \begin{bmatrix} c_{21}(x^*) \\ c_{22}(x^*) \\ c_{23}(x^*) \end{bmatrix} \quad \mu^* = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

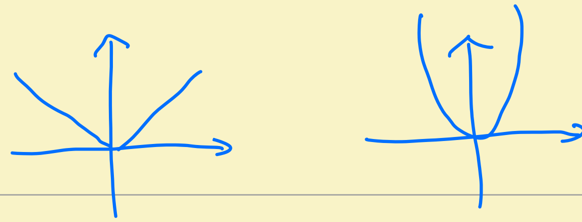
$$c_{21}^* \mu_1^* + c_{22}^* \mu_2^* + c_{23}^* \mu_3^* = 0$$

$$\Rightarrow \quad c_{21}^* \mu_1^* = 0, \quad c_{22}^* \mu_2^* = 0, \quad c_{23}^* \mu_3^* = 0$$

称满足 KKT 条件的点 (x^*, λ^*, μ^*) 为稳定点.

$$\min f(x)$$
$$\text{s.t.} \quad c(x) = 0$$

① 罚方法
② 增广拉格朗日方法
③* 投影梯度法
④* 流形优化法



min $f(x)$
s.t $c(x)=0$

罚方法

Defn：对于等式约束优化问题，其二次罚函数 $g(x)=\|x\|_2^2$

$$P_\sigma(x)=f(x)+\boxed{\frac{1}{2}\sigma\|c(x)\|_2^2}$$

罚项

罚参数（罚因子）

Note：a. 罚函数对不满足约束的点进行惩罚，但在迭代过程中，点列一般处于可行域外。

b. 增大 σ 使得罚项在罚函数中权重加大，迫使迭代向可行域靠近。

c. 在可行域中，$P_\sigma(x)$ 的全局极小点与原问题的最优解相同

d. 当罚参数过小时，不可行点处的函数下降可能会抵消罚项对约束违反的惩罚。

E.g. $\min\ -x^2+2y^2 \quad s.t\ x=1 \quad \Rightarrow (x^*,y^*)=(1,0)$

$$P_\sigma(x,y)=-x^2+2y^2+\frac{\sigma}{2}(x-1)^2 \qquad \sigma\le 2$$

$$=2y^2-2x+1 \quad \rightarrow \text{无下界.}$$

Q：为什么不能给解到大的罚参数？

E.g. $\min\ x+\sqrt{3}y$
s.t. $x^2+y^2=1 \qquad (x^*,y^*)=\left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)$

$$P_\sigma(x,y)=x+\sqrt{3}y+\frac{\sigma}{2}(x^2+y^2-1)^2$$

Algo: 二次罚函数法

给定 $\sigma_0 > 0$, x^0, $\rho > 1$　　ρ: 罚因子增长系数

while 未达到收敛准则, do

　　以 x^k 为初始点, 求解 $x^{k+1} = \arg\min_x P_{\sigma_k}(x)$

　　$\sigma_{k+1} = \sigma_k$

　　$k = k+1$

end while

收敛性分析

假设有解

Thm: 设 x^{k+1} 是 $P_{\sigma_k}(x)$ 的全局极小解, σ_k 单调上升, $\sigma_k \to +\infty$ $k \to \infty$

则 $\{x^k\}$ 的每一个极限点 x^* 都是原问题的全局极小解.

Proof: 设 $\bar{x}$ 是原问题的全局极小解, 则

$$f(\bar{x}) \le f(x) \quad \forall x \text{ satisfies } c(x) = 0$$

由于 x^{k+1} 是 $P_{\sigma_k}(x)$ 的全局极小值, 有 $P_{\sigma_k}(x^{k+1}) \le P_{\sigma_k}(\bar{x})$

$$\Rightarrow f(x^{k+1}) + \frac{\sigma_k}{2}\|c(x^{k+1})\|^2 \le f(\bar{x}) + \frac{\sigma_k}{2}\|c(\bar{x})\|^2 = f(\bar{x})$$

$$\Rightarrow \|c(x^{k+1})\|^2 \le \frac{2}{\sigma_k}\left(f(\bar{x}) - f(x^{k+1})\right)$$

设 $x^k \to x^*$,

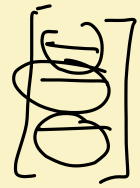
$$\|c(x^*)\|^2 = \lim_{k\to\infty} \|c(x^{k+1})\|^2 \le \lim_{k\to\infty} \frac{2}{\sigma_k}\left(f(\bar{x}) - f(x^{k+1})\right) = 0$$

$$\Rightarrow \|c(x^*)\|^2 = 0$$

$$\Rightarrow f(x^{k+1}) \le f(\bar{x}) \quad \Rightarrow \quad f(x^*) \le f(\bar{x}) \quad \text{as } k \to \infty,$$

由于 $\bar{x}$ 为全局极小解, 有 $f(x^*) = f(\bar{x})$

$$\begin{bmatrix} x_1 + 2x_2 + 3x_3 + 4x_4 \\ x_1 + 3x_2 + 2x_3 + 5x_4 \end{bmatrix}$$



终止准则.

LICQ: $\nabla c(x^*)$ 线性无关

$c(x): \mathbb{R}^n \to \mathbb{R}^m$

$\nabla c(x) \in \mathbb{R}^{m\times n}$

$n \geq m$

<u>Thm</u>: 设 $f(x)$, $c(x)$ 可微, 若 x^{k+1} 满足 $\|\nabla_x P_{\sigma_k}(x^{k+1})\| \leq \varepsilon_k$, 且 $\varepsilon_k \to 0$. 对 $\{x^k\}$ 的任何极限点 x^*, 满足 LICQ, 则 x^* 是原问题的 KKT 点, 且

$$\lim_{k\to\infty} (-\sigma c_i(x^{k+1})) = \lambda_i^*$$

其中, λ_i^* 是 $c_i(x^*)=0$ 的 Lagrangian 乘子.

Proof: 由终止准则可得,

$$\|\nabla P_{\sigma_k}(x^{k+1})\| = \|\nabla f(x^{k+1}) + \sigma_k [\nabla c(x^{k+1})]^T c(x^{k+1})\| \leq \varepsilon_k$$

利用三角不等式 $\|a\| - \|b\| \leq \|a+b\|$

$$\|[\nabla c(x^{k+1})]^T c(x^{k+1})\| \leq \frac{1}{\sigma_k}\left(\varepsilon_k + \|\nabla f(x^{k+1})\|\right)$$

令 $k\to\infty$, 得

$$\|[\nabla c(x^*)]^T c(x^*)\| = 0$$

由于 $\nabla c_i(x^*)$ 线性无关, 可得 $c(x^*)=0$. $\therefore$ x^* 为可行点,

构造 Lagrangian 乘子 $\lambda_i^k = -\sigma_k c_i(x^{k+1})$, 则

$$\nabla P_{\sigma_k}(x^{k+1}) = \nabla f(x^{k+1}) + [\nabla c(x^{k+1})]^T \lambda^k$$

线性无关
$\therefore \nabla c(x^{k+1})$ 行满秩

$$\Rightarrow [\nabla c(x^{k+1})]^T \lambda^k = \nabla f(x^{k+1}) - \nabla P_{\sigma_k}(x^{k+1})$$

$$\Rightarrow \lambda^k = [\nabla c(x^{k+1}) \nabla c(x^{k+1})^T]^{-1} \nabla c(x^{k+1}) \left(\nabla f(x^{k+1}) - \nabla P_{\sigma_k}(x^{k+1})\right)$$

由于 $\|\nabla_x P_{\sigma_k}(x^{k+1})\| \leq \varepsilon_k \to 0$, 有

$$\lambda^* = \lim_{k\to\infty} \lambda^k = [\nabla c(x^*) \nabla c(x^*)^T]^{-1} \nabla c(x^*) \nabla f(x^*)$$

且

$$\nabla P_{\sigma_k}(x^*) = \nabla f(x^*) - [\nabla c(x^*)]^T \lambda^* = 0$$

$\min f(x)$ $\Rightarrow$ $-\sigma_k c_i(x^*) = \lambda_i^*$ $\lambda_i^* \neq 0$

s.t. $c_i(x) \geq 0 \quad i \in \mathcal{E}$ $\Rightarrow$ $c_i(x^*) = -\frac{\lambda_i^*}{\sigma_k}$

$c_i(x^*) \to 0$ as $\sigma_k \to +\infty$ $\to$ 条件数爆炸

精确罚方法： $g(x) = \|x\|_1$ 是精确罚.

增广拉格朗日方法

增广 Lagrangian 函数

$$L_\sigma(x;\lambda) = f(x) + \langle \lambda, c(x) \rangle + \frac{1}{2}\sigma \| c(x) \|_2^2$$

增广 Lagrangian 函数 FOC

$$\nabla_x L_\sigma(x^{k+1}; \lambda^k) = \nabla f(x^{k+1}) + [\nabla c(x^{k+1})]^T \lambda^k + \sigma [\nabla c(x^{k+1})]^T c(x^{k+1}) = 0$$

原问题稳定点条件

$$\nabla_x L(x^{k+1}; \lambda^*) = \nabla f(x^{k+1}) + [\nabla c(x^{k+1})]^T \lambda^* = 0$$

$$[\nabla c(x^{k+1})]^T \lambda^k + \sigma [\nabla c(x^{k+1})]^T c(x^{k+1}) = [\nabla c(x^{k+1})]^T \lambda^* = 0$$

LICQ

$$\lambda^k + \sigma_k c(x^{k+1}) = \lambda^*$$

$$\Rightarrow c(x^{k+1}) = \frac{1}{\sigma_k}(\lambda^* - \lambda^k)$$

$c(x^{k+1}) \to 0$ $\sigma_k \to +\infty$ 或 $\lambda^* - \lambda^k \to 0$

一般 $\lambda^0=0$

Algo. 增广拉格朗日函数法 (ALM)

1. 选取初始点 x^0, λ^0, 罚参数 $\sigma_0>0$, 罚参数更新 $\rho>0$,
2. while $\|\nabla_x L_\sigma(x^k,\lambda^k)\| \ge \eta_k$ do
3. 以 x^k 为初始点, 求解 $L_{\sigma_k}(x,\lambda^k)$, $x^{k+1}=\arg\min L_{\sigma_k}(x,\lambda^k)$
4. 更新乘子. $\lambda^{k+1}=\lambda^k+\sigma_k c(x^{k+1})$
5. if $\|c(x^{k+1})\|\le\varepsilon$ then
6. break.
7. end if
8. 更新罚参数: $\sigma_{k+1}=\rho\sigma_k$
9. end for

E.g.

$$\min \; x+\sqrt{3}y$$
$$\text{s.t.} \; x^2+y^2=1$$