优化笔记

Leoeon

2015.11.11

Contents

Chapter 1

凸优化

1.1 凸集

1.1.1 集合

1.1.1.1 仿射

多点$\{x_k\}$的仿射组合：$\{\sum\limits_k \theta_k x_k | \sum\limits_k \theta_k = 1\}$

集合C的仿射包：$\mathrm{aff}[C] = \{\sum\limits_k \theta_k x_k | \forall x_k \in C, \sum\limits_k \theta_k = 1\}$

集合C是仿射集合：$C \supseteq \mathrm{aff}[C]$

1.1.1.2 凸

多点$\{x_k\}$的凸组合：$\{\sum\limits_k \theta_k x_k | \sum\limits_k \theta_k = 1, \theta_k \geq 0\}$

集合C的凸包：$\mathrm{conv}[C] = \{\sum\limits_k \theta_k x_k | \forall x_k \in C, \sum\limits_k \theta_k = 1, \theta_k \geq 0\}$

集合C是凸集：$C \supseteq \mathrm{conv}[C]$

一般化：集合C是凸集，随机变量x，$P(x \in C) = 1$，则$E[x] = \int\limits_{x \in C} p(x) x dx \in C$

1.1.1.3 锥

多点$\{x_k\}$的锥组合：$\{\sum\limits_k \theta_k x_k | \theta_k \geq 0\}$

集合C的锥包：$\mathrm{cone}[C] = \{\sum\limits_k \theta_k x_k | \forall x_k \in C, \theta_k \geq 0\}$

集合C是凸锥：$C \supseteq \mathrm{cone}[C]$

。

1.1.2 正常锥

正常锥：锥K为凸、闭、实（有非空内部）、尖（不包含直线）

广义不等式：$x_1 \succeq_K x_2 \iff x_2 - x_1 \in K$

1.1.3 超平面

超平面分离定理：凸集$C \bigcap$凸集$D = \emptyset$，则$\exists a \neq 0, b$，对$\forall x \in C$ 有$a^T x \leq b$，对$\forall x \in D$ 有$a^T x \geq b$

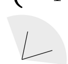
支撑超平面：取$a \neq 0$，$x_0 \in$ 边界$[C]$，对$\forall x \in C$，有$a^T(x - x_0) \leq 0$，则$\{x | a^T(x - x_0) = 0\}$为C的支撑超平面

支撑超平面定理：对$\forall$非空凸集C，$\forall x_0 \in$ 边界$[C]$，在x_0处$\exists C$的支撑超平面

1.1.4 对偶锥

锥K的对偶锥$K^* = \{z | x^T z \geq 0, \forall x \in K\}$ 是闭凸锥

$$\begin{cases} x_1 \succeq_K x_2 \quad \iff \quad \text{对}\forall z \succeq_{K^*} 0\text{，有}z^T x_1 \leq z^T x_2 \quad (z \neq 0\text{时，}z^T x_1 < z^T x_2) \\ z_1 \succeq_{K^*} z_2 \quad \iff \quad \text{对}\forall x \succeq_K 0\text{，有}z_1^T x \leq z_2^T x \quad (z \neq 0\text{时，}z_1^T x < z_2^T x) \end{cases}$$



K^{**}是K的凸包的闭包

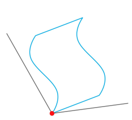
1.1.5 最值与极值

1.1.5.1 最小元

$$x_0 \in S\text{，对}\forall x \in S \text{ 均有}x_0 \preceq_K x$$

$$\iff \text{对}\forall \lambda \succeq_{K^*} 0\text{，}\forall x \in S\text{，}x_0 = \arg\text{最小}_x \lambda^T x$$

$$\iff \text{对}\forall \lambda \succeq_{K^*} 0\text{，超平面}\{x | \lambda^T(x - x_0) = 0\}\text{是}x_0\text{处对S的严格支撑超平面}$$



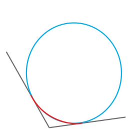
1.1.5.2 极小元

$$x_0 \in S\text{，对}\forall x \in S \text{ 均有}(x \preceq_K x_0 \Longrightarrow x = x_0)$$

$$\Longleftarrow \exists \lambda_0 \succeq_{K^*} 0\text{，}\forall x \in S\text{，}x_0 = \arg\text{极小}\lambda_0^T x$$

$$x_0 \in S\text{，对}\forall x \in \text{凸集}S \text{ 均有}(x \preceq_K x_0 \Longrightarrow x = x_0)$$

$$\iff \exists \lambda_0 \succeq_{K^*} 0\text{，}\forall x \in \text{凸集}S\text{，}x_0 = \arg\text{极小}\lambda_0^T x$$



1.1.6 单调性

对可微函数f，$dom[f]$ 为凸集

f是k-非减的 $\iff$ 对$\forall x \in dom[f]$ 有$\nabla f(x) \succeq_{K^*} 0$

f是k-增的 $\Longleftarrow$ 对$\forall x \in dom[f]$ 有$\nabla f(x) \succ_{K^*} 0$

1.2 凸函数

1.2.1 凸函数

$f : \mathbb{R}^n \to \mathbb{R}$

若定义域$dom[f]$是凸集，且对$\forall x_1, x_2 \in dom[f]$、$\forall\theta \in [0,1]$，有$f(\theta x_1+(1-\theta)x_2) \leq \theta f(x_1)+(1-\theta)f(x_2)$

一般化：$f(E[x]) \leq E[f(x)]$ 或$f(x) \leq E[f(x+z)]$(z是均值为0随机变量)

$\iff$ f在与定义域相交的任何直线上都是凸的

即对$\forall x \in dom[f]$ 与$\forall v$，$g(t) = f(x+tv)$ 是凸的

$\iff$ $dom[f]$是凸集，且对$\forall x_1, x_2 \in dom[f]$ 有$f(x_2) \geq f(x_1) + \nabla f(x_1)^T(x_2 - x_1)$

$\iff$ $dom[f]$是凸集，且对$\forall x \in dom[f]$ 有$\nabla^2 f(x) \succeq 0$

（可将$f(x)$延拓为$\tilde{f}(x) = \begin{cases} f(x) & , \quad x \in dom[f] \\ +\infty & , \quad x \notin dom[f] \end{cases}$）

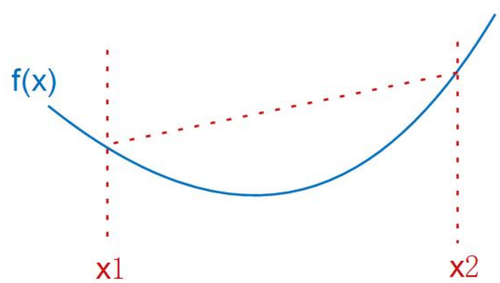


Figure 1.1: 凸函数

1.2.2 上境图

函数f的上境图$epi[f] \overset{def}{=} \{(x,t)|x \in dom[f], t \geq f(x)\}$

f是凸函数 $\iff$ epi[f]是凸函数

1.2.3 共轭函数

$f^* : \mathbb{R}^n \to \mathbb{R}$，$f^*(z) \overset{def}{=} \sup\limits_{x \in dom[f]} (z^T x - f(x)) = \nabla f(x)^T x - f(x)|_{z=\nabla f(x)}$

$f^{**} : \mathbb{R}^n \to \mathbb{R}$，$f^{**}(x) \overset{def}{=} \sup\limits_{z \in dom[f^*]} (x^T z - f^*(z))$

无论f是否为凸函数，共轭函数f^*必为闭凸函数

恒有$f(x) \geq f^{**}(x)$

若f为凸且闭，则$f^{**}(x) = f(x)$

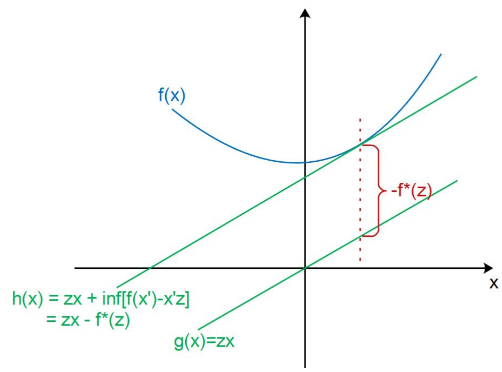


Figure 1.2: f^*

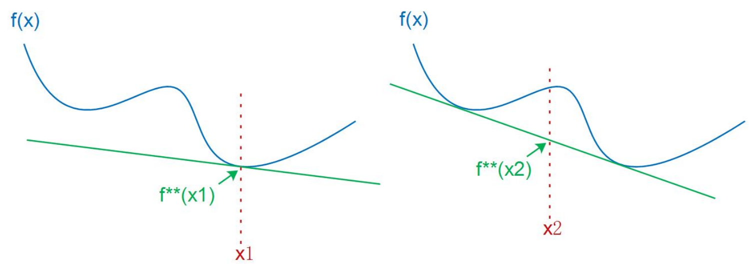


Figure 1.3: f^{**}

1.2.4 拟凸函数

$f : \mathbb{R}^n \to \mathbb{R}$

若对$\forall a \in \mathbb{R}$，$\{x \in dom[f] | f(x) \leq a\}$ 都是凸集

$\iff$ $dom[f]$是凸集，且对$\forall x_1, x_2 \in dom[f]$、$\forall \theta \in [0,1]$ 有$f(\theta x_1 + (1-\theta)x_2) \leq \max\{f(x_1), f(x_2)\}$

$\iff$ $dom[f]$是凸集，且对$\forall x_1, x_2 \in dom[f]$ 有($f(x_2) \leq f(x_1) \Rightarrow \nabla f(x_1)^T(x_2 - x_1) \leq 0$)

$\iff$ 对$\forall x \in dom[f]$、$\forall u \in \mathbb{R}^n$ 有($u^T \nabla f(x) = 0 \Rightarrow u^T \nabla^2 f(x) u \geq 0$)

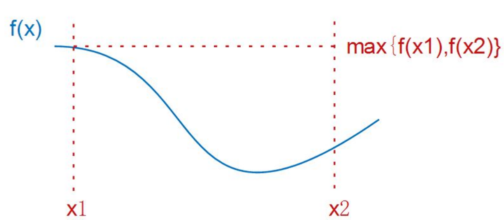


Figure 1.4: 拟凸函数

1.2.5 对数凸函数

$f : \mathbb{R}^n \to \mathbb{R}$

若对$\forall x \in dom[f]$ 有$f(x) > 0$ 且$\log f(x)$ 是凸函数

$\Longleftrightarrow$ $dom[f]$是凸集，且对$\forall x \in dom[f]$ 有$f(x) > 0$，且对$\forall x_1, x_2 \in dom[f]$、$\forall \theta \in [0,1]$ 有$f(\theta x_1 + (1-\theta)x_2) \le f(x_1)^{\theta} f(x_2)^{1-\theta}$

$\Longleftrightarrow$ $dom[f]$ 是凸集，且对$\forall x_1, x_2 \in dom[f]$ 有$\log f(x_2) \ge \log f(x_1) \frac{\nabla f(x_1)^T}{f(x_1)}(x_2 - x_1)$

$\Longleftrightarrow$ $dom[f]$ 是凸集，且对$\forall x \in dom[f]$ 有$f(x)\nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T$

$\Rightarrow$ f是凸函数，f是拟凸函数

1.2.6 K凸函数

$f : \mathbb{R}^n \to \mathbb{R}$，$K \subseteq \mathbb{R}^m$是一正常锥

若对$\forall x_1, x_2$、$\forall \theta \in [0,1]$ 有$f(\theta x_1 + (1-\theta)x_2) \preceq_K \theta f(x_1) + (1-\theta) f(x_2)$

$\Longleftrightarrow$ 对$\forall w \succeq_{K^*} 0$，$w^T f$是凸的

$\Longleftrightarrow$ $dom[f]$ 是凸集，且对$\forall x_1, x_2 \in dom[f]$ 有$f(x_2) \succeq_K f(x_1) + Df(x_1)(x_2 - x_1)$（D是Jacobian矩阵）

1.3 凸优化问题

1.3.1 凸优化问题

$$
\begin{aligned}
& \min_x f(x) && \text{（}f(x)\text{为凸函数）} \\
s.t.\quad & g_i(x) \le 0 && \text{（}g_i(x)\text{为凸函数）} \\
& h_j(x) = a_j^T x - b_j = 0 &&
\end{aligned}
$$

$$
\begin{aligned}
& x^*\text{是全局最优解，即}x^* = \mathrm{arginf}_x\{f(x) | g_i(x) \le 0, h_j(x) = 0\} \\
\Longleftrightarrow\quad & x^*\text{是任意局部最优解，即}x^* = \mathrm{arginf}_x\{f(x) | g_i(x) \le 0, h_j(x) = 0, ||x - x^*|| \le \varepsilon(\varepsilon > 0)\} \\
\Longleftrightarrow\quad & x^* \in \text{约束，对}\forall x \in \text{约束}(x \ne x^*)\text{，有}\nabla f^T(x^*)(x - x^*) \ge 0
\end{aligned}
$$

1.3.2 拟凸优化问题

$$
\begin{aligned}
& \min_x f(x) && \text{（}f(x)\text{为拟凸函数）} \\
s.t.\quad & g_i(x) \le 0 && \text{（}g_i(x)\text{为凸函数）} \\
& h_j(x) = a_j^T x - b_j = 0 &&
\end{aligned}
$$

$$
\begin{aligned}
& x^*\text{是全局最优解，即}x^* = \mathrm{arginf}_x\{f(x) | g_i(x) \le 0, h_j(x) = 0\} \\
(\not\Longleftrightarrow\quad & x^*\text{是任意局部最优解}) \\
\Longleftrightarrow\quad & x^* \in \text{约束，对}\forall x \in \text{约束}(x \ne x^*)\text{，有}\nabla f^T(x^*)(x - x^*) \ge 0
\end{aligned}
$$

将拟凸优化问题转为凸优化问题：

1. 令$\phi_t : R^n \to R, t\ inR$ 为满足$f(x) \le t \iff \phi_t(x) \le 0$ 的一族函数
（如$\phi_t(x) = \begin{cases} 0 & , \quad f(x) \le t \\ +\infty & , \quad f(x) > t \end{cases}$ ）
2. 解凸可行性问题

$$\begin{aligned} & x \\ s.t. \quad & \phi_t(x) \le 0 \\ & g_i(x) \le 0 \\ & h_j(x) = 0 \end{aligned}$$

若有解，则$f(x^*) \le t$，否则$f(x^*) > t$
3. 由此用二分法划t限出$f(x^*)$ 所在

1.4 拉格朗日乘子法

原优化问题

$$\begin{aligned} & \min_x f(x) \\ s.t. \quad & g_i(x) \le 0 \\ & h_j(x) = 0 \end{aligned}$$

Lagrange函数

$$L(x, \lambda, \nu) = f(x) + \sum_i \lambda_i g_i(x) + \sum_j \nu_j h_j(x)$$

Lagrange对偶函数

$$F(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu) \qquad \left(D = \bigcap_i dom[g_i] \bigcap_j \bigcap_j dom[h_j] \right)$$

Lagrange对偶问题

$$\begin{aligned} & \max_{\lambda, \nu} F(\lambda, \nu) \quad (dom[F] = \{(\lambda, \nu) | F(\lambda, \nu) > -\infty\}) \\ s.t. \quad & \lambda \succeq 0 \end{aligned}$$

（无论原问题是否为凸优化问题，对偶问题必是凸优化问题）

对$\forall \lambda \succeq 0$ 与$\forall \nu$，有

$$F(\lambda, \nu) \le \text{最优值} f(x^*)$$

即

$$\text{最优值} F(\lambda^*, \nu^*) \le \text{最优值} f(x^*)$$

即

$$\sup_{\lambda \succeq 0} \inf_x L(x, \lambda, \nu) \le \inf_x \sup_{\lambda \succeq 0} L(x, \lambda, \nu)$$

当强对偶性成立，即最优值$F(\lambda^*, \nu^*) =$ 最优值$f(x^*)$ 时，有

$$\begin{cases} \nabla_x f(x^*) + \sum_i \lambda_i^* \nabla_x g_i(x^*) + \sum_j \nu_j^* \nabla_x h_j(x^*) = 0 \\ g_i(x^*) \le 0, \quad \lambda_i^* \ge 0, \quad \lambda_i^* g_i(x^*) = 0 \\ h_j(x^*) = 0 \end{cases}$$

$$\left(\begin{cases} g_i(x) \leq 0 \\ h_j(x) = 0 \end{cases} \text{与} \begin{cases} \lambda \succeq 0 \\ \inf\limits_{x \in D} [\sum\limits_i \lambda_i g_i(x) + \sum\limits_j \nu_j h_j(x)] > 0 \end{cases} \text{至多有一个可行}\right)$$

1.5 例子

1.5.1 不等式

1.5.1.1

算数几何平均不等式：凸函数$f(x) = -\log x$

$$\sum_{i=1}^{n} a_i x_i \geq \prod_{i=1}^{n} x_i^{a_i} \quad (x_i \geq 0)$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \quad (x_i \geq 0)$$

1.5.1.2

Holder不等式：凸函数$f(x) = x^p$

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} y_i^q\right)^{\frac{1}{q}} \quad (x_i, y_i > 0)(p, q > 1)(\frac{1}{p} + \frac{1}{q} = 1)$$

Minkowski和不等式：

$$\left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} y_i^p\right)^{\frac{1}{p}} \quad (x_i, y_i \geq 0)(p > 1)$$

Minkowski积不等式：

$$\left[\prod_{i=1}^{n} (x_i + y_i)\right]^{\frac{1}{n}} \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} + \left(\prod_{i=1}^{n} y_i\right)^{\frac{1}{n}} \quad (x_i, y_i \geq 0)$$

Jensen不等式：

$$\left(\sum_{i=1}^{n} x_i^{p_1}\right)^{\frac{1}{p_1}} > \left(\sum_{i=1}^{n} x_i^{p_2}\right)^{\frac{1}{p_2}} \quad (x_i > 0)(0 < p_1 < p_2)$$

Chapter 2

迭代法

求$\min_x f(x)$

2.1 梯度下降法

$$
\begin{aligned}
f(x) &\simeq f(x^{(t)}) \\
&+ \begin{bmatrix} \cdots & \frac{\partial f}{\partial x_i}(x^{(t)}) & \cdots \end{bmatrix} (\vec{x} - \vec{x}^{(t)})
\end{aligned}
$$

故Δx 模恒定的前提下，方向为$-\begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}$时，$f(x)$ 最小。

即

$$
\vec{x}^{(t+1)} = \vec{x}^{(t)} - \lambda \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}
$$

2.2 牛顿法

$$
\begin{aligned}
f(x) &\simeq f(x^{(t)}) \\
&+ \begin{bmatrix} \cdots & \frac{\partial f}{\partial x_i}(x^{(t)}) & \cdots \end{bmatrix} (\vec{x} - \vec{x}^{(t)}) \\
&+ \frac{1}{2}(\vec{x} - \vec{x}^{(t)}) \begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x^{(t)}) & \cdots \\ & \vdots & \end{bmatrix} (\vec{x} - \vec{x}^{(t)})
\end{aligned}
$$

为求$f(x)$最小，令

$$
\begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x) \\ \vdots \end{bmatrix} \simeq \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix} + \begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x^{(t)}) & \cdots \\ & \vdots & \end{bmatrix} (\vec{x} - \vec{x}^{(t)}) = 0
$$

即

$$\vec{x}^{(t+1)}=\vec{x}^{(t)}-\begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x^{(t)}) & \cdots \\ & \vdots & \end{bmatrix}^{-1}\begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}$$

2.2.1 拟牛顿法

令$\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)}) \overset{def}{=} \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t+1)}) \\ \vdots \end{bmatrix} - \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}$，$\Delta\vec{x}^{(t)}=\vec{x}^{(t+1)}-\vec{x}^{(t)}$

用$H^{(t)}$ 近似$\begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x^{(t)}) & \cdots \\ & \vdots & \end{bmatrix}$，或用$G^{(t)}$ 近似$\begin{bmatrix} & \vdots & \\ \cdots & \frac{\partial^2 f}{\partial x_i \partial x_j}(x^{(t)}) & \cdots \\ & \vdots & \end{bmatrix}^{-1}$，

则需满足拟牛顿条件

$$\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})=H^{(t)}\Delta\vec{x}^{(t)} \quad \text{或} \quad G^{(t)}\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})=\Delta\vec{x}^{(t)}$$

则令

$$\vec{x}^{(t+1)}=\vec{x}^{(t)}-\lambda^{(t)}G^{(t)}\begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}$$

其中

$$\lambda^{(t)}=\mathrm{argmin}_{\lambda\geq 0}\left(\vec{x}^{(t)}-\lambda G^{(t)}\begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_i}(x^{(t)}) \\ \vdots \end{bmatrix}\right)$$

2.2.1.1 BFGS

$$H^{(t+1)}=H^{(t)}+\frac{\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})^T}{\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})^T\Delta\vec{x}^{(t)}}+\frac{H^{(t)}\Delta\vec{x}^{(t)}\Delta\vec{x}^{(t)T}H^{(t)}}{\Delta\vec{x}^{(t)T}H^{(t)}\Delta\vec{x}^{(t)}}$$

2.2.1.2 DFP

$$G^{(t+1)}=G^{(t)}+\frac{\Delta\vec{x}^{(t)}\Delta\vec{x}^{(t)T}}{\Delta\vec{x}^{(t)T}\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})}+\frac{G^{(t)}\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})^TG^{(t)}}{\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})^TG^{(t)}\Delta\frac{\partial \vec{f}}{\partial x}(x^{(t)})}$$

2.3 有约束的迭代法

2.3.1 惩罚函数法

$$\begin{aligned} &\min f(x) \\ s.t. \quad &h_i(x)=0 \end{aligned}$$

近似于

$$\min f(x) + \sum_i \lambda_i h_i(x)$$

$\lambda_i \geq 0$作为惩罚力度

2.3.2 边界跟踪法

2.3.2.1 法一

若每一步不违反约束，则忽略约束按正常梯度方向$-\nabla f$走；若每一步违反约束，则按约束方向∇g_i走

2.3.2.2 法二

忽略约束走到终点。

若终点不违反约束，则完成；若终点违反约束，则按约束方向∇g_i方向走直至不违反