最优化方法：救命笔记

献给还没复习的朋友们

作者：Youwei Zhang

组织：天之智慧研究会

时间：May 18,2022

版本：1.0

各人自扫门前雪，休管他人瓦上霜。——我不同意

1. 牛顿法

算法

给定控制误差 $\varepsilon > 0$.

Step 1 取初始点 $\boldsymbol{x}_0$, 令 $k = 0$.

Step 2 计算 $\boldsymbol{g}_{\boldsymbol{k}}$.

Step 3 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 则 $\boldsymbol{x}^* = \boldsymbol{x}_k$, 停; 否则计算 $\boldsymbol{G}_k$, 并由 $G_k p_k = -g_k$ 解出 $\boldsymbol{p}_k$.

Step 4 令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{p}_k, k = k+1$, 转 Step 2 .

♣

练习 试用牛顿 (Newton) 法求下列问题的最优解:

$$\min f(x) = \frac{1}{2}x_1^2 + \frac{3}{2}x_2^2 - x_1x_2 - 2x_2,$$

初始点 $x^{(0)} = (-1,-1)^T$;

解

$g(x) = \begin{pmatrix} x_1 - x_2 \\ 3x_2 - x_1 - 2 \end{pmatrix}, G(x) = \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}$ 正定

$x^{(0)} = (-1,-1)^T, g_0 = (0,-4)^T, G_0 = \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}$

$\Rightarrow X^{(1)} = X^{(0)} - G^{-1}g_0 = (1,1)^T$,

此时,$g_1 = (0,0)^T$, 故$X^* = X^{(1)} = (1,1)^T, f^* = -1$.

练习 试用牛顿 (Newton) 法求下列问题的最优解:

$$\min f(X) = x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1,$$

取初始点$X^{(0)} = (1,1)^T$.

解

$g(x) = \begin{pmatrix} 2x_1 - 2x_2 - 4 \\ 4x_2 - 2x_1 \end{pmatrix}, G(x) = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$ 正定

$x^{(0)} = (1,1)^T, g_0 = (-4,2)^T, G_0 = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$

$\Rightarrow X^{(1)} = X^{(0)} - G^{-1}g_0 = (4,2)^T$,

此时,$g_1 = (0,0)^T$, 故$X^* = X^{(1)} = (4,2)^T, f^* = -8$.

笔记 对于二次型函数，由于牛顿法的方向直指中心，且一步到位，能够一步走到极值点，所以求 X^* 的时候，直接令 $g(x) = 0$ 即可。

2. 拟牛顿法

算法

给定控制误差 ε,

Step 1 给定初始点 $\boldsymbol{x}_0$, 初始矩阵 $\boldsymbol{H}_0$ (通常取单位阵), 计算 g_0 , 令 $k=\mathbf{0}$.

Step 2 令 $\boldsymbol{p}_k=-\boldsymbol{H}_k\boldsymbol{g}_k$.

Step 3 由精确一维搜索确定步长 α_k ,

$f\left(x_k+\alpha_k p_k\right)=\min\limits_{\alpha\geq 0} f\left(x_k+\alpha p_k\right)$.

Step 4 令 $x_{k+1}=x_k+\alpha_k p_k$.

Step 5 若 $\left\|\boldsymbol{g}_{k+1}\right\|\leqslant\varepsilon$, 则 $\boldsymbol{x}^*=\boldsymbol{x}_{k+1}$ 停; 否则令

$\boldsymbol{s}_k=\boldsymbol{x}_{k+1}-\boldsymbol{x}_k,\quad \boldsymbol{y}_k=\boldsymbol{g}_{k+1}-\boldsymbol{g}_k$.

Step 6 由 DFP 修正公式 $\boldsymbol{H}_{k+1}=\boldsymbol{H}_k-\frac{\boldsymbol{H}_k\boldsymbol{y}_k{\boldsymbol{y}_k}^{\mathrm{T}}\boldsymbol{H}_k}{{\boldsymbol{y}_k}^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{y}_k}+\frac{\boldsymbol{s}_k{\boldsymbol{s}_k}^{\mathrm{T}}}{\boldsymbol{y}_k^{\mathrm{T}}\boldsymbol{s}_k}$ 得 $\boldsymbol{H}_{k+1}$. 令 $k=k+1$, 转 Step 2 .

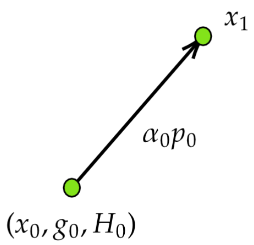
笔记

1. 确定初始点, 行走方向 (求这点的梯度 g_k 和 H_k, 方向 $\boldsymbol{p}_k=-\boldsymbol{H}_k\boldsymbol{g}_k$)
2. 确定步长 ($\min\limits_{\alpha\geq 0} f\left(x_k+\alpha p_k\right)$)
3. 走到下一点 ($x_{k+1}=x_k+\alpha_k p_k$), 再重复上述操作

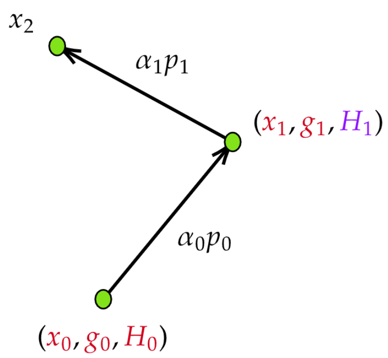
终止条件: 这一点的梯度 $g_k=0$ 或很小的时候, 就不要往下一个点走了, 这点就为最优点.

H_k 求解时的两个关键向量:分别是两点的坐标差和梯度差

拟牛顿法步骤示意图



1.标注初始点的信息:坐标,梯度,H_0
2.三个信息已知,可立刻求出方向p_0
3.此时可将$\alpha_0 p_0$代入f,求出步长,使得f在p_0这个方向的值最小
4.步长和方向都确定后,就能让x_0走到x_1了



1.标注x_1点的信息:坐标,梯度,H_1
其中H_1由红色信息求出
2.三个信息已知,可立刻求出方向p_1
3.此时可将$\alpha_1 p_1$代入f,求出步长
4.步长和方向都确定后,就能让x_1走到x_2了

练习 用 DFP(变尺度) 法求解下列无约束最优化问题:

$$\min f(x)=2x_1^2+x_2^2+2x_1x_2+x_1-x_2;$$

取初始点 $x^{(0)}=(0,0)^T, H_0=E;$

解

$\nabla f(x)=(4x_1+2x_2+1, 2x_2+2x_1-1)$,

初始点信息: $x^{(0)}=(0,0)^T, g_0=(1,-1), H_0=\begin{pmatrix}1 & \\ & 1\end{pmatrix}$,

确定方向:$p_0=-H_0g_0=(-1,1)^T$,

确定步长:$\min\limits_{\lambda\geq 0} f(x^{(0)}+\lambda p_0)=\min\limits_{\lambda\geq 0}\lambda^2-2\lambda, \varphi'(\lambda)=2\lambda-2=0$, 得$\lambda_0=1$,

走向下一点 $x^{(1)}$: $x^{(1)}=x^{(0)}+\lambda p_0=(-1,1)^T$,

新点 $x^{(1)}$ 的信息:$x^{(1)}=(-1,1)^T, g_1=(-1,-1)^T, H_1=$ 待求

求解 H_1:

$s_0=x^{(1)}-x^{(0)}=(-1,1)^T, y_0=g_1-g_0=(-2,0)^T, H_1=H_0+\frac{s_0s_0^T}{y_0^Ts_0}-\frac{H_0y_0y_0^TH_0}{y_0^TH_0y_0}$

$=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\frac{1}{2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}-\frac{1}{4}\begin{pmatrix}4 & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}1/2 & -1/2\\ -1/2 & 3/2\end{pmatrix}$,

确定 $x^{(1)}$ 的行走方向: $p_1=-H_1g_1=\begin{pmatrix}0\\ -1\end{pmatrix}$,

确定步长: $\min\limits_{\lambda\geq 0} f(x^{(1)}+\lambda p_1)=\min\limits_{\lambda\geq 0}\lambda^2-\lambda-1$, 得$\lambda_1=\dfrac{1}{2}$,

走向下一点 $x^{(2)}$: $x^{(2)}=x^{(1)}+\lambda p_1=\left(-1,\frac{3}{2}\right)^T$,

$x^{(2)}$ 的信息: $x^{(2)}=\left(-1,\frac{3}{2}\right)^T, g_2=(0,0)^T$, 梯度为 0, 终止迭代.

所以极小点 $x^{(2)}=(-1,3/2), f^*=-\frac{5}{4}$.

练习 试用 DFP(变尺度) 法, 求出 $X^{(1)}$ 和变尺度矩阵 H_1 以及搜索方向 $\boldsymbol{p}_1$,

$$\min f(X)=x_1^2+2x_2^2-2x_1x_2-4x_1,$$

取初始点$X^{(0)}=(1,1)^T, H_0=E$.

解 过程如上, 太麻烦了, 自己写.

3. 共轭梯度法

算法

给定控制误差 ε .

Step 1 给定初始点 $x_1, k=1$.

Step 2 计算 $\boldsymbol{g}_k=\boldsymbol{g}\left(\boldsymbol{x}_{\mathbf{k}}\right)$.

Step 3 若 $\left\|\boldsymbol{g}_k\right\| \leqslant \varepsilon$, 则 $\boldsymbol{x}^*=\boldsymbol{x}_k$ 停; 否则令

$$
\begin{gathered}
\boldsymbol{p}_k=-\boldsymbol{g}_k+\beta_{k-1} \boldsymbol{p}_{k-1}, \\
\beta_{k-1}= \begin{cases}\frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}}, & \text {当} k>1 \text {时}, \\ 0, & \text {当} k=1 \text {时}.\end{cases}
\end{gathered}
$$

Step 4 由精确一维搜索确定步长 α_k, 满足

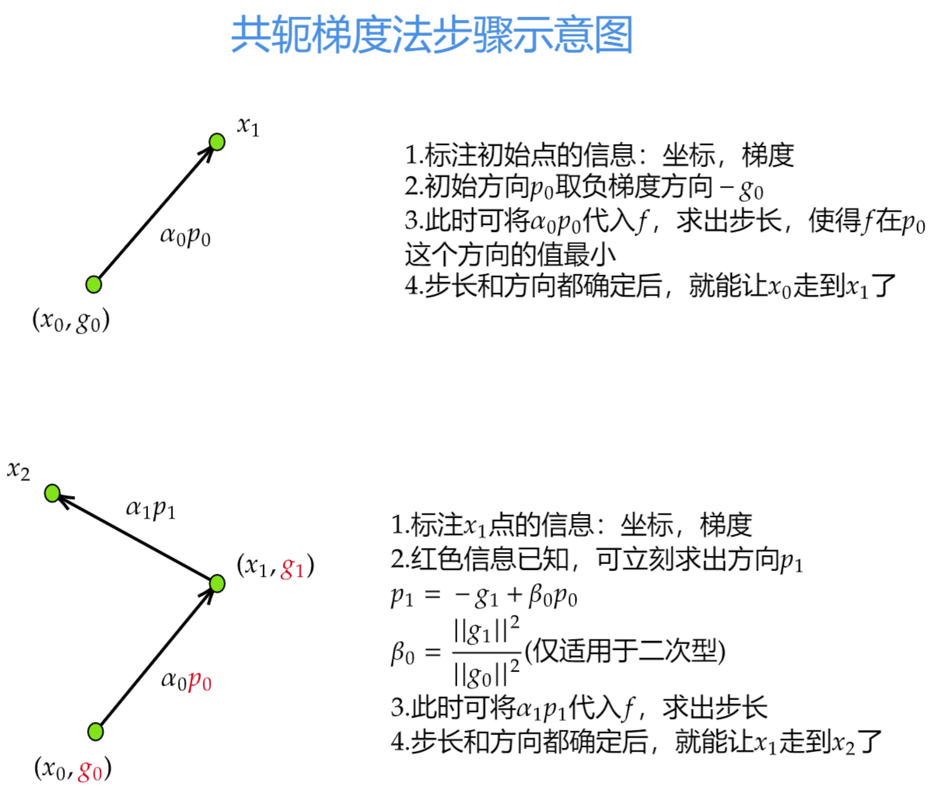
$$
f\left(x_k+\alpha_k p_k\right)=\min _{a \geq 0} f\left(x_k+\alpha p_k\right) .
$$

Step 5 令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k \boldsymbol{p}_k, k=k+1$, 转 Step 2 .

笔记

下面 3 句话不清楚什么意思不用管，直接看图即可.

1. 共轭梯度法核心是求一组共轭向量，也是 G 的特征向量，组成过渡矩阵 P；在坐标变换 $x=Py$ 下就是正交向量了，二次型的变量也解耦了。看二次型的等高线（椭圆）有，沿着共轭方向走，其实就是沿着变换后坐标系下的坐标轴方向走，由于变换后二次函数变量之间互不耦合，所以一维搜索下分别取极值后，就是最终的极值.
2. 共轭向量可以由已知点的梯度用类似" 施密特共轭化" 的方法求得，对于二次型函数，这个公式还可以简化
3. 共轭梯度法也可以用来解实对称矩阵特征值，也就是解方程，参数量巨大也无所谓



✍ **练习** 试用共轭梯度 (FR) 法求下列问题的最优解:

$$min f(x) = x_1^2 + 4x_2^2,$$

取初始点为 $x^{(0)} = (4,1)^T$;

解

$\nabla f(x) = g(x) = (2x_1, 8x_2)^T$

初始点信息: $x^{(0)} = (4,1)^T, g_0 = (8,8)^T$.

确定方向 p_0: $p_0 = -g_0 = (-8,-8)^T$.

确定步长 α_0: $f(x^{(0)} + \alpha_0 p_0) = (4-8\alpha_0)^2 + 4(1-8\alpha_0)^2$, 令 $f'(\alpha_0) = 0$, 得 $\alpha_0 = \frac{1}{5}$.

走向下一点 $x^{(1)}$: $x^{(1)} = x^{(0)} + \alpha_0 p_0 = (\frac{12}{5}, -\frac{3}{5})^T$.

新点 $x^{(1)}$ 的信息:$x^{(1)} = (\frac{12}{5}, -\frac{3}{5})^T, g_1 = (\frac{24}{5}, -\frac{24}{5})^T$

确定方向 p_1: $p_1 = -g_1 + \beta_0 p_0 = -g_1 + \frac{||g_1||^2}{||g_0||^2} p_0 = \frac{48}{5}(-4,1)^T$, 因为就是一个方向, 所以取 $p_1 = (-4,1)^T$.

确定步长 α_1: $f(x^{(1)} + \alpha_1 p_1) = (2.4 - 4\alpha_1)^2 + 4\left(-0.6 + \alpha_1\right)^2$, 令 $f'(\alpha_1) = 0$, 得 $\alpha_1 = 0.6$.

走向下一点 $x^{(2)}$: $x^{(2)} = x^{(1)} + \alpha_1 p_1 = (0,0)^T$, 顺便求出梯度: $g_2 = (0,0)^T$

所以, $X^* = (0,0), \min f(X^*) = 0$.

✍ **练习** 试用共轭梯度 (FR) 法求下列问题的最优解:

$$\min f(X) = 2x_1^2 + x_2^2 + 2x_1x_2 + x_1 - x_2,$$

取初始点为$X^{(0)} = (0,0)^T$.

解 步骤同上.

4.KT 条件

定理

对于一般约束最优化问题:

$$\begin{aligned}
&\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbf{R}^n, \\
&\text{s.t. } c_i(\boldsymbol{x}) = 0, \quad i \in E = \{1, \cdots, l\}, \\
&\qquad c_i(\boldsymbol{x}) \geqslant 0, \quad i \in I = \{l+1, \cdots, m\}
\end{aligned}$$

若

(i) x^* 为局部最优解, 其有效集 $I^* = \{i \mid c_i(x^*) = 0, i \in I\}$;

(ii) $f(\boldsymbol{x}), c_i(\boldsymbol{x})(i = 1, \cdots, m)$ 在点 $\boldsymbol{x}^*$ 可微;

(iii) 对所有 $i \in E \cup I^*, \nabla c_i(x^*)$ 线性无关.

则存在向量 $\lambda^* = (\lambda_1^*, \cdots, \lambda_m^*)^{\mathrm{T}}$ 使得

$$\begin{aligned}
&\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, \\
&\lambda_i^* c_i(x^*) = 0, \lambda_i^* \geqslant 0, \quad i \in I, \\
&c_i(x^*) = 0, \quad i \in E.
\end{aligned}$$

♡

✍ **练习** 求解下列约束问题:

$$\min f(x) = (x_1 - x_2 + x_3)^2$$
$$\text{st. } \begin{cases} c_1(x) = x_1 + 2x_2 - x_3 = 5 \\ c_2(x) = x_1 - x_2 - x_3 = -1 \end{cases}$$

(1) 写出 K – T 条件;

(2) 求出 K – T 点, 最优解和最优值.

解 ...

✍ **练习** 求解下列约束最优化问题:

$$\min f(X) = x_1^2 + x_2^2 + x_3^2$$
$$\text{s.t. } \begin{cases} c_1(X) = x_1 + 2x_2 - x_3 - 4 = 0, \\ c_2(X) = x_1 - x_2 - x_3 + 2 = 0; \end{cases}$$

(1) 写出 K – T 条件;

(2) 求出 K – T 点, 最优解和最优值.

解 ...

5. 罚函数法

算法

外罚函数法

对于约束问题:

$$\begin{aligned} &\min f(\boldsymbol{x}), \boldsymbol{x} \in \mathbf{R}^n, \\ &\text{s.t.} \qquad c_i(\boldsymbol{x}) = 0, \quad i \in E = \{1, 2, \cdots, l\} \\ &\qquad\qquad c_i(\boldsymbol{x}) \geqslant 0, \quad i \in I = \{l+1, \cdots, m\} \end{aligned}$$

取控制误差 $\varepsilon > 0$ 和罚因子的放大系数 $c > 1$ (可取 $\varepsilon = 10^{-4}, c = 10$).

Step 1 给定初始点 x_0 (可以不是可行点) 和初始罚因子 σ_1 (可取 $\sigma_1 = 1$), 令 $k = 1$.

Step 2 以 x_{k-1} 为初始点求无约束问题:

$$\min P(x, \sigma_k) = f(x) + \sigma_k \tilde{P}(x),$$

其中

$$\tilde{\boldsymbol{P}}(\boldsymbol{x}) = \sum_{i=1}^{l} |c_i(\boldsymbol{x})|^\beta + \sum_{j=l+1}^{m} |\min(0, c_j(\boldsymbol{x}))|^\alpha, \quad \alpha \geqslant 1, \beta \geqslant 1,$$

得最优解 $\boldsymbol{x}_k = \boldsymbol{x}(\sigma_k)$.

Step 3 若 $\sigma_k \tilde{P}(x_k) < \varepsilon$, 则以 x_k 为近似最优解, 停止. 否则令 $\sigma_{k+1} = c\sigma_k, k = k+1$, 转 Step 2 .

♣

笔记

1. 这种方法就是给目标函数 $f(x)$ 加一个惩罚项 $\tilde{P}(x)$，变成增广目标函数 $P(x)$. 使原目标函数在可行域内不受惩罚，在违反约束的地方受到惩罚（理想情况为 ∞）, 其几何解释之前在群聊天里发过，不记得可以翻记录, 那里讲的非常清楚.
2. 直接求 $P(x)$ 的无约束极值即可.

练习 求解约束问题

$$\begin{aligned} &\min f(x) = x_1^2 + x_2^2 \\ &\text{s.t. } x_1 + 1 \leq 0 \end{aligned}$$

解 令

$$P(x, \sigma) = \begin{cases} x_1^2 + x_2^2, & x_1 + 1 \leq 0 \\ x_1^2 + x_2^2 + \sigma(x_1 + 1)^2, & x_1 + 1 > 0 \end{cases}$$

令 $\begin{cases} \frac{\partial P}{\partial x_1} = 0 \\ \frac{\partial P}{\partial x_2} = 0 \end{cases}$

当 $x_1 \leq -1$ 时, $\begin{cases} \frac{\partial P}{\partial x_1} = 2x_1, \\ \frac{\partial P}{\partial x_2} = 2x_2, \end{cases}$, 解得 $(x_1, x_2) = (0, 0)$, 舍去;

当 $x_1 > -1$ 时, $\begin{cases} \frac{\partial P}{\partial x_1} = 2x_1 + 2\sigma(x_1 + 1) \\ \frac{\partial P}{\partial x_2} = 2x_2, \end{cases}$, 解得 $(x_1, x_2) = (-\frac{\sigma}{\sigma+1}, 0)$,

它是 $P(x, \sigma)$ 的最优点，最优值为 $P(x, \sigma) = \frac{\sigma}{\sigma+1}$.

当 $\sigma \to +\infty$ 时, $x_1(\sigma) \to -1, x_2(\sigma) \to 0$. 因此 $x(\sigma) \to x^*, P(x, \sigma) \to f(x^*) = 1$.

注 一般题目约束是起作用的，所以不必求没有惩罚的那一段函数的极值.

练习 求解约束问题

$$\min f(x) = x_1^2 + x_2^2$$
$$\text{s.t. } x_1 + 1 = 0$$

解 令

$$P(x,\sigma) = \begin{cases} x_1^2 + x_2^2, & x_1 + 1 = 0, \\ x_1^2 + x_2^2 + \sigma (x_1+1)^2, & x_1 + 1 \neq 0, \end{cases} = x_1^2 + x_2^2 + \sigma (x_1+1)^2$$

令 $\begin{cases} \frac{\partial P}{\partial x_1} = 0 \\ \frac{\partial P}{\partial x_2} = 0 \end{cases}$, $\begin{cases} \frac{\partial P}{\partial x_1} = 2x_1 + 2\sigma(x_1+1) \\ \frac{\partial P}{\partial x_2} = 2x_2, \end{cases}$, 解得 $(x_1, x_2) = (-\frac{\sigma}{\sigma+1}, 0)$,

它是 $P(x,\sigma)$ 的最优点，最优值为 $P(x,\sigma) = \frac{\sigma}{\sigma+1}$.

当 $\sigma \to +\infty$ 时, $x_1(\sigma) \to -1, x_2(\sigma) \to 0$. 因此 $x(\sigma) \to x^*, P(x,\sigma) \to f(x^*) = 1$.

练习 求解约束问题:

$$\min f(x) = x_1^2 + x_1x_2 + x_2^2,$$
$$\text{s.t. } c(x) = x_1 + x_2 - 2 = 0$$

解 这道题就是后面的乘子法的试题，这里先用外罚函数法写一写，后面的乘子法求解步骤类似. 考试的话只考其一，最近两次考的都是乘子法.

笔记

计算机求解思路

构造一堆增广函数，惩罚项系数按一定的倍数增加，分别用无约束算法求解每个增广函数的极值点

以上上题为例，大致步骤如下（惩罚系数 σ 的增大倍数设为 10）:

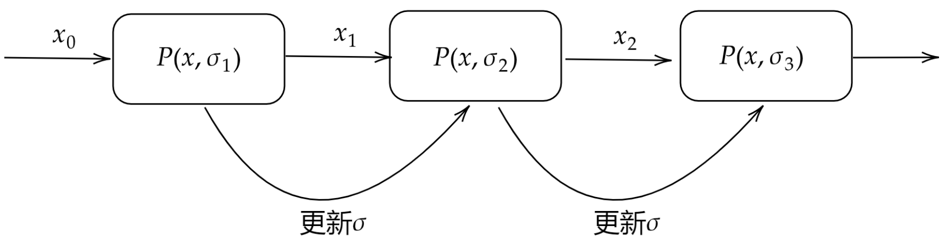
1. 给定初始点 x_0，增广目标函数 $P(x,\sigma_1) = P(x, 0.1) = x_1^2 + x_2^2 + 0.1(x_1+1)^2$, 用牛顿法或者别的无约束算法求得 $P(x,\sigma_1)$ 的最优点，记为 x_1.

2. 将 x_1 视为新的初始点，增广目标函数 $P(x,\sigma_2) = P(x,1) = x_1^2 + x_2^2 + (x_1+1)^2$, 用牛顿法或者别的无约束算法求得 $P(x,\sigma_2)$ 的最优点，记为 x_2.

3. 将 x_2 视为新的初始点，增广目标函数 $P(x,\sigma_3) = P(x,10) = x_1^2 + x_2^2 + 10(x_1+1)^2$, 用牛顿法或者别的无约束算法求得 $P(x,\sigma_3)$ 的最优点，记为 x_3.

终止条件：$\sigma_k \tilde{P}(x_k) < \varepsilon$　　因为 KT 条件没法直接拿来用，这里是因为 $\{x_k\}$ 收敛时，$\sigma_k \tilde{P}(x_k) \to 0$

罚函数法步骤示意图



算法步骤：将x_0丢进$P(x,\sigma_1)$中做无约束优化，得到x_1;
再把x_1丢进$P(x,\sigma_2)$中优化，得到x_2; 以此类推。

算法

内罚函数法

求解约束问题

$$\min f(x), \quad x \in \mathbf{R}^n$$
$$s.t. \quad c_i(x) \geqslant 0, \quad i=1, \cdots, m$$

且其可行域的内点集 $D_0 \neq \varnothing$.
构造如下的增广目标函数:

$$B(x, r)=f(x)+r \tilde{B}(x),$$

其中令

$$\tilde{B}(\boldsymbol{x})=\sum_{i=1}^m \frac{1}{c_i(\boldsymbol{x})} \text{ 或} \tilde{B}(\boldsymbol{x})=-\sum_{i=1}^m \ln \left(c_i(\boldsymbol{x})\right),$$

取控制误差 $\varepsilon>0$ 和罚因子的缩小系数 $0<c<1$(可取 $\varepsilon=10^{-4}, c=0.1$).
Step 1 选定初始点 $x_0 \in D_0$, 给定 $r_1>0$ (取 $r_1=10$), 令 $k=1$.
Step 2 以 x_{k-1} 为初始点, 求解无约束问题

$$\min B\left(x, r_k\right)=f(x)+r_k \tilde{B}(x),$$

得最优解 $\boldsymbol{x}_k=\boldsymbol{x}\left(r_k\right)$.
Step 3 若 $r_k \tilde{B}\left(x_k\right)<\varepsilon$, 则 $\boldsymbol{x}_k$ 为近似最优解, 停. 否则, 令 $r_{k+1}=c r_k, k=k+1$ 转 Step 2 .

♣

练习 (非数学专业必答) 请用内点法求解下列问题:

$$\min f(X)=\frac{1}{6}\left(x_1+1\right)^3+x_2-1$$
$$\text{s.t.} \begin{cases} c_1(X)=x_1-1 \geq 0 \\ c_2(X)=x_2-1 \geq 0 \end{cases}$$

(1) 推导其最优解;
(2) 若取 $X_0=(3,5), r_1=2$, 公比 $c=\frac{1}{2}$ 时, 指出用何种方法可计算点列 $X_1, X_2, \cdots, X_k, \cdots$;

解 ...

6. 乘子法

算法

等式约束问题的乘子法 (PH 算法)

Step 1 选定初始点 x_0 , 初始乘子向量 λ_1 , 初始罚因子 σ_1 及其放大系数 $c>1$, 控制误差 $\varepsilon>0$ 与常数 $\theta\in(0,1)$, 令 $k=1$.

Step 2 以 x_{k-1} 为初始点求解无约束问题

$$\min M\left(\boldsymbol{x},\lambda_k,\sigma_k\right)=f(\boldsymbol{x})-{\lambda_k}^{\mathrm{T}}\boldsymbol{c}(\boldsymbol{x})+\frac{\sigma_k}{2}\boldsymbol{c}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{c}(\boldsymbol{x})$$

得最优解 $\boldsymbol{x}_k$.

Step 3 当 $\|\boldsymbol{c}\left(x_k\right)\|<\varepsilon$ 时, x_k 为所求最优解, 停. 否则转 Step 4 .

Step 4 当 $\|c\left(x_k\right)\|/\|c\left(x_{k-1}\right)\|\leqslant\theta$ 时, 转 Step 5 , 否则令 $\sigma_{k+1}=c\sigma_k$, 转 Step 5 .

Step 5 令 $\lambda_{k+1}=\lambda_k-\sigma_k c\left(x_k\right)$, $k=k+1$, 转 Step 2 .

♣

笔记

算法设计思路（不清楚在讲啥可跳过）

原始形式的外罚函数法存在一个严重的缺点，就是 σ 必须趋于无穷大，才能让点列收敛到最优点，但是这样的增广函数性质很垃圾，梯度差的一比，很难优化。而拉格朗日函数虽然可以直接使用无约束优化算法，但是有的情况下这个拉格朗日函数不存在极值点，其稳定点是个鞍点，无约束算法不一定能走到这个鞍点上来。

解决方案：对拉格朗日函数再进行约束，使它变成约束问题，并且这个约束问题与原始约束问题等价。对这个拉格朗日函数进行外罚函数法。

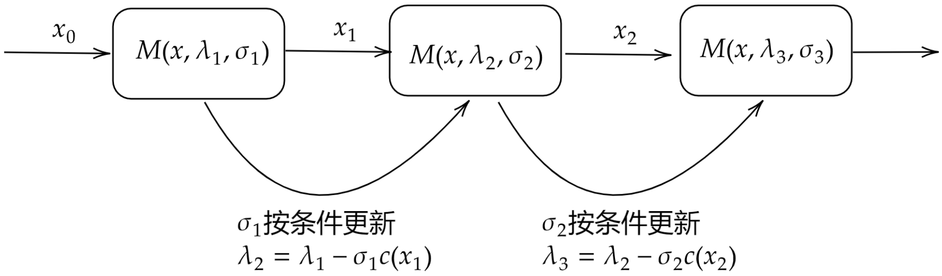
由于带约束的拉格朗日函数，在最优点处，其 Hessian 矩阵一定在约束可行域内正定，由 Finsler 定理（教材 P158）可以证明，一定存在一个数 $\sigma^*>0$, 使得当 $\sigma\geq\sigma^*$ 时，对于整个空间，都有 $\nabla_x^2 L\left(x^*,\lambda^*\right)+\sigma\nabla\boldsymbol{A}\left(x^*\right)\boldsymbol{A}\left(x^*\right)^{\mathrm{T}}$ 正定, 其中 $\nabla\boldsymbol{A}\left(x^*\right)\boldsymbol{A}\left(x^*\right)^{\mathrm{T}}$ 是罚函数 $\tilde{P}(x)=\frac{1}{2}\boldsymbol{c}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{c}(\boldsymbol{x})$ 的二阶导.

而 $\nabla_x^2 L\left(x^*,\lambda^*\right)+\sigma\nabla\boldsymbol{A}\left(x^*\right)\boldsymbol{A}\left(x^*\right)^{\mathrm{T}}$ 是如下函数（增广拉格朗日函数）的 Hessian 矩阵：

$$M(x,\lambda,\sigma)=L(x,\lambda)+\frac{\sigma}{2}c(x)^{\mathrm{T}}c(x)$$

说明了上述条件下，在最优点处，拉格朗日增广函数在全空间内都是正定的，就能用无约束优化算法，并且梯度性质较好.

乘子法步骤示意图



σ_1按条件更新：当$\dfrac{||c(x_2)||}{||c(x_1)||} > \theta$时，更新$\sigma_2 = c\sigma_1$；否则不更新$\sigma_1$

算法步骤：将x_0丢进$M(x, \lambda_1, \sigma_1$中做无约束优化，得到x_1；
再把x_1丢进$M(x, \lambda_2, \sigma_2)$中优化，得到$x_2$；以此类推。
终止条件：$||c(x_2)|| < \epsilon$

手算步骤：直接令$\nabla_x M(x, \lambda\ \ , \sigma\ \ ) = 0$，解出$x^*, \lambda^*, \sigma^*$即可

✍ **练习**　试用乘子法求下列问题的最优解:

$$\begin{aligned} &\min f(x) = x_1^2 + x_1 x_2 + x_2^2, \\ &\text{s.t. } c(x) = x_1 + x_2 - 2 = 0 \end{aligned}$$

解

增广 Lagrange 函数为:

$$M(\boldsymbol{x}, \lambda, \sigma) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{G}\boldsymbol{x} - \lambda(x_1 + x_2 - 2) + \frac{\sigma}{2}(x_1 + x_2 - 2)^2, \sigma > 0,$$

令

$$\nabla M(x, \lambda, \sigma) = \boldsymbol{G}\boldsymbol{x} - \lambda \nabla c(\boldsymbol{x}) + \sigma c(\boldsymbol{x}) \nabla c(\boldsymbol{x}) = 0$$

其中

$$\nabla c(\boldsymbol{x}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \lambda \nabla c(\boldsymbol{x}) = \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \sigma \nabla c(\boldsymbol{x}) \cdot c(\boldsymbol{x}) = \sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1, 1)\boldsymbol{x} - 2\sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \sigma \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \boldsymbol{x} - 2\sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

所以,

$$\nabla M(x, \lambda, \sigma) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} x - \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \sigma \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x - 2\sigma \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2+\sigma & 1+\sigma \\ 1+\sigma & 2+\sigma \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \lambda + 2\sigma \\ \lambda + 2\sigma \end{pmatrix} = 0,$$

解得,

$$x_1 = x_2 = \frac{\lambda + 2\sigma}{3 + 2\sigma} \to 1(\sigma \to \infty),$$

故存在 σ^*, 使得 $\frac{\lambda + 2\sigma^*}{3 + 2\sigma^*} = 1$, 得 $\lambda = 3$. 因此

$$x^* = (1, 1)^T, f* = 3.$$

✍ **练习** 请用乘子法求下列问题的最优解:

$$\min f(x) = x_1^2 - 3x_2 - x_2^2,$$
$$\text{s.t. } x_2 = 0.$$

解

增广 Lagrange 函数为:

$$M(x,\lambda,\sigma) = x_1^2 - (\lambda+3)x_2 + \frac{\sigma-2}{2}x_2^2$$

令

$$\frac{\partial M}{\partial x_1} = 2x_1 = 0,$$
$$\frac{\partial M}{\partial x_2} = (\sigma-2)x_2 - (\lambda+3) = 0,$$

得 $x_0 = \left(0, \frac{\lambda+3}{\sigma-2}\right)^{\mathrm{T}}$. 要求 x_0 满足的约束条件 $x_2 = 0$, 必须取 $\lambda = -3$, 从而 $x_0 = (0,0)^{\mathrm{T}} = x^*$, 即为原约束问题的最优解.

✍ **练习** (数学专业必答) 对上题, 已知初始点 X_0, 初始乘子向量 λ_1, 初始罚因子 σ_1, 放大系数 $c > 1$, 控制误差 $\varepsilon > 0$, 常数 $\theta \in (0,1)$, 如何求出迭代点列 $X_1, X_2, \cdots, X_k, \cdots$;

解

就是简要概述一下算法，结合上面给的算法步骤和图片，随便讲讲，注意 σ 和 λ 的更新，以及终止条件。

7. 简约梯度法

算法

需要用到的公式:

1. $r\left(x^N\right)=\nabla_N f\left(x^B\left(x^N\right), x^N\right)-\left(B^{-1}N\right)^{\mathrm{T}}\nabla_B f\left(x^B\left(x^N\right), x^N\right)$
2. $\left(p_k^N\right)_j=\begin{cases}-\left(x_k^N\right)_j r_j\left(x_k^N\right), & \text{当 } r_j\left(x_k^N\right)>0 \text{ 时},\\ -r_j\left(x_k^N\right), & \text{当 } r_j\left(x_k^N\right)\leqslant 0 \text{ 时}.\end{cases}$
3. $\boldsymbol{p}_k^B=-\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{p}_k^N$
4. $\alpha_{\max}=\begin{cases}\min\left\{\frac{(\boldsymbol{x}_k)_j}{-(\boldsymbol{p}_k)_j}\mid(\boldsymbol{p}_k)_j<0\right\}, & \text{当 } \boldsymbol{p}_k\neq\mathbf{0} \text{ 时},\\ +\infty, & \text{当 } \boldsymbol{p}_k\geqslant\mathbf{0} \text{ 时},\end{cases}$

Step 1 给定初始基可行解 $x_1=\begin{pmatrix}x_1^B\\ x_1^N\end{pmatrix}\geqslant 0$, 其中 x_1^B 为基向量, 令 $k=1$.

Step 2 对应于 $\boldsymbol{x}_k=\begin{pmatrix}\boldsymbol{x}_k^B\\ \boldsymbol{x}_k^N\end{pmatrix}$ 将 $\boldsymbol{A}$ 分解成 $\boldsymbol{A}=(\boldsymbol{B},\boldsymbol{N})$. 由公式 1, 2 和 3 分别计算 $r\left(x_k^N\right)$, $\boldsymbol{p}_k^N$ 和 $\boldsymbol{p}_k^B$, 令 $\boldsymbol{p}_k=\begin{pmatrix}\boldsymbol{p}_k^B\\ \boldsymbol{p}_k^N\end{pmatrix}$.

Step 3 若 $\boldsymbol{p}_k=\mathbf{0}$, 则 $\boldsymbol{x}_k$ 为 KT 点, 停; 否则, 由式 4 计算 $\alpha_{\max}$, 求 α_k 使

$$f\left(\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k\right)=\min_{0\leqslant\alpha\leqslant\alpha_{\max}} f\left(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k\right),$$

令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k$, 转 Step 4 .

Step 4 若 $\boldsymbol{x}_{k+1}^B>0$, 则基向量不变, 令 $k=k+1$, 转 Step 2; 若有某个 j 使 $\left(\boldsymbol{x}_{k+1}^B\right)_j=0$, 则将 $\left(\boldsymbol{x}_{k+1}^B\right)_j$ 换出基, 而以 $\boldsymbol{x}_{k+1}^N$ 中具有最大分量的变量换入基, 构成新的基向量 $\boldsymbol{x}_{k+1}^B$ 与非基向量 $\boldsymbol{x}_{k+1}^N$, 令 $k=k+1$, 转 Step 2 .

♣

✍ **练习** 试述简约梯度法的迭代步骤;

解 ...

✍ **练习** 试用简约梯度法求出初始点 $X^{(1)}$ 和搜索方向 $\boldsymbol{p}_1$, 以及步长的上限 $\alpha_{\max}$;

$$\min f(X)=x_1^2+x_1x_2+2x_2^2-6x_1-14x_2,$$
$$\text{s.t.}\begin{cases}c_1(X)=x_1+x_2+x_3=2,\\ c_2(X)=-x_1+2x_2\quad+x_4=3,\\ x_i>0, i=1,2,3,4;\end{cases}$$

解 ...