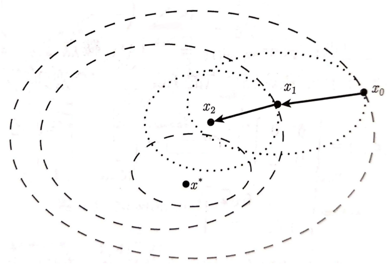
4-2 牛顿法

牛顿法

基本思想: 利用目标函数$f(\boldsymbol{x})$在迭代点$\boldsymbol{x}_k$处的二次Taylor展开作为模型函数，每一步迭代求得二次模型函数的极小点，产生极小点序列去逼近目标函数的极小点。



$$q_0(\boldsymbol{d}) \to q_1(\boldsymbol{d}) \to \cdots \to q_k(\boldsymbol{d})$$

牛顿法

设$f(\boldsymbol{x})$二次连续可微，$\boldsymbol{x}_k \in \mathbb{R}^n$，Hesse矩阵$\nabla^2 f(\boldsymbol{x}_k)$正定。在$\boldsymbol{x}_k$附近用二次Taylor展开近似$f$，

$$f(\boldsymbol{x}_k + \boldsymbol{d}) \approx q_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T G_k \boldsymbol{d}, \tag{4.2.1}$$

其中$\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}_k$，$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k)$, $G_k = \nabla^2 f(\boldsymbol{x}_k)$, $q_k(\boldsymbol{d})$为$f(\boldsymbol{x})$的二次近似。

将极小化近似二次函数$q_k(\boldsymbol{d})$，即得到如下**牛顿方程**：

$$\nabla q_k(\boldsymbol{d}) = \boldsymbol{g}_k + G_k \boldsymbol{d} = 0, \tag{4.2.2}$$

通过上式求得$\boldsymbol{d}_k = -G_k^{-1}\boldsymbol{g}_k$, 称该方向为**牛顿方向**。

牛顿法

迭代公式:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - G_k^{-1}\boldsymbol{g}_k, \tag{4.2.3}$$

使用牛顿方向作为迭代方向的最优化方法称为**牛顿法**。因为G_k正定，故牛顿方向总能保证函数值下降，即$\boldsymbol{g}_k^T\boldsymbol{d}_k = -\boldsymbol{g}_k^T G_k \boldsymbol{g}_k^T < 0$。

算法4.2.1 – 基本牛顿法

步1 给出$\boldsymbol{x}_0 \in \mathbb{R}^n$，$\varepsilon > 0$，$k := 0$.

步2 如果终止准则满足，输出近似最优解，停止迭代。

步3 由公式(4.2.2)计算搜索步长$\boldsymbol{d}_k$;

步4 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k, k := k+1$，转步2.

基本牛顿法中步长因子$\alpha_k = 1$。

牛顿法算例

例4.2.1 考虑问题：

$$\min f(\boldsymbol{x}) = 3x_1^2 + 3x_2^2 - x_1^2 x_2$$

从不同的初始点出发，用牛顿法求解该问题，终止准则为$\|\boldsymbol{g}_k\|_\infty < 10^{-6}$。

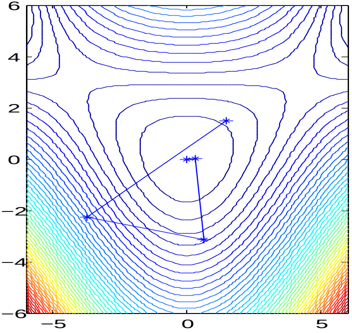
解: $f(\boldsymbol{x})$的一二阶导数信息为：

$$\boldsymbol{g}(\boldsymbol{x}) = (6x_1 - 2x_1x_2, 6x_2 - x_1^2)^T, G(\boldsymbol{x}) = \begin{bmatrix} 6-2x_2 & -2x_1 \\ -2x_1 & 6 \end{bmatrix}$$

由一阶最优性必要条件可求得三个稳定点：$\boldsymbol{x}_{(1)} = (0,0)^T$ – 极小点，$\boldsymbol{x}_{(2)} = (3\sqrt{2}, 3)^T$ – 鞍点，和$\boldsymbol{x}_{(3)} = (-3\sqrt{2}, 3)^T$.

牛顿法算例

从$(1.5,1.5)^T$以及$(-2,4)^T$出发, 牛顿法产生的迭代点列作图如下：



从$(0,3)^T$出发, 该点处海森矩阵奇异, 牛顿法无法迭代。

牛顿法算例

初始点1: $(1.5, 1.5)^T$

k	x_k^T	f_k	$\|g_k\|$
0	$(1.5000, 1.5000)$	10.1250	8.1125
1	$(-3.7500, -2.2500)$	89.0156	48.0633
2	$(0.6250, -3.1250)$	31.6895	20.6151
3	$(0.3190, 0.0014)$	0.3052	1.9155
4	$(-0.0020, -0.0172)$	0.0009	0.1037
5	$(-0.0000, -0.0000)$	0.0000	0.0000
6	$(-0.0000, -0.0000)$	0.0000	0.0000

收敛到极小点$\boldsymbol{x}_{(1)}$

初始点2: $(-2, 4)^T$

k	x_k^T	f_k	$\|g_k\|$
0	$(-2.0000, 4.0000)$	44.0000	20.3961
1	$(4.0000, 2.0000)$	28.0000	8.9443
2	$(4.3077, 3.0769)$	26.9750	0.6695
3	$(4.2439, 3.0011)$	27.0000	0.0105
4	$(4.2426, 3.0000)$	27.0000	0.0000
5	$(4.2426, 3.0000)$	27.0000	0.0000

收敛到鞍点$\boldsymbol{x}_{(2)}$

定理4.2.1 牛顿法局部收敛性与二阶收敛速度.

设$f(\boldsymbol{x})$二阶连续可微，$\boldsymbol{x}^*$是$f(\boldsymbol{x})$的局部极小点，$\nabla^2 f(\boldsymbol{x}^*)$正定。假定$f(\boldsymbol{x})$的Hesse矩阵$G_k$满足Lipschitz条件，即存在$\beta > 0$，以下不等式成立:

$$\|G(\boldsymbol{x}) - G(\boldsymbol{y})\| \leq \beta \|\boldsymbol{x} - \boldsymbol{y}\|, \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \tag{4.2.4}$$

则当初始点$\boldsymbol{x}_0$充分靠近$\boldsymbol{x}^*$时，对于一切k，牛顿迭代(4.2.3)有意义，迭代序列$\{\boldsymbol{x}_k\}$收敛到$\boldsymbol{x}^*$，并且具有二阶收敛速度。

证明: 对梯度向量函数$\boldsymbol{g}(\boldsymbol{x})$做Taylor展开得：

$$\boldsymbol{g}(\boldsymbol{x}_k + \boldsymbol{d}) = \boldsymbol{g}_k + G_k \boldsymbol{d} + O(\|\boldsymbol{d}\|^2).$$

其中：$\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}_k$。

定理4.2.1证明

证明续: 令$\boldsymbol{d} = -\boldsymbol{h}_k = \boldsymbol{x}^* - \boldsymbol{x}_k$，梯度向量展开为：

$$0 = \boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{g}_k - G_k \boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2).$$

由于$f(\boldsymbol{x})$二次连续可微，$G^* = \nabla^2 f(\boldsymbol{x}^*)$正定，故当$\boldsymbol{x}_k$充分靠近$\boldsymbol{x}^*$时，$G_k$也正定且$\|G_k^{-1}\|$有界, 故第$k$次迭代存在。用$G_k^{-1}$乘以上式两边得：

$$\begin{aligned} 0 &= G_k^{-1}\boldsymbol{g}_k - \boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2) = -\boldsymbol{d}_k - \boldsymbol{h}_k + O(\|\boldsymbol{h}_k\|^2) \\ &= -\boldsymbol{h}_{k+1} + O(\|\boldsymbol{h}_k\|^2). \end{aligned}$$

由$O(\cdot)$的定义，存在常数C，使得

$$\|\boldsymbol{h}_{k+1}\| \le C\|\boldsymbol{h}_k\|^2. \tag{a}$$

证明续: 若$\boldsymbol{x}_k$充分靠近$\boldsymbol{x}^*$使得:

$$\boldsymbol{x}_k \in \Omega = \{\boldsymbol{x} | \|\boldsymbol{x}_k - \boldsymbol{x}^*\| \leq \frac{\gamma}{C}, \ \gamma \in (0,1)\}$$

由(a)有

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| = \|\boldsymbol{h}_{k+1}\| \leq \gamma \|\boldsymbol{h}_k\| \leq \frac{\gamma^2}{C} < \frac{\gamma}{C},$$

这表明$\boldsymbol{x}_{k+1}$也在这个邻域Ω中。由归纳法，迭代对所有k有定义。由于

$$\|\boldsymbol{h}_k\| \leq \gamma \|\boldsymbol{h}_{k-1}\| \leq \cdots \leq \gamma^k \|\boldsymbol{h}_0\|,$$

令$k \to \infty$，得$\|\boldsymbol{h}_k\| \to 0$.因此迭代序列$\{\boldsymbol{x}_k\}$收敛。(a)表明收敛速度是二阶的。 ■

牛顿法优缺点

优点：

- 当初始点$\boldsymbol{x}_0$充分接近$\boldsymbol{x}^*$时，方法以二阶收敛速度收敛。
- 方法具有二次终止性。

缺点：

- 当初始点$\boldsymbol{x}_0$没有充分接近$\boldsymbol{x}^*$时，G_k出现不正定或者奇异情形，迭代无法进行。即使非奇异也不能保证f_k单调下降。
- 每步迭代需要计算海森矩阵，涉及到$\frac{n\times(n+1)}{2}$个二阶偏导数。
- 每步迭代需要求解一个线性方程组，计算量为$O(n^3)$。

阻尼牛顿法

考虑采用带一维搜索的牛顿法，即：

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k,$$

其中:

- $\boldsymbol{d}_k = -G_k^{-1}\boldsymbol{g}_k$;
- α_k由线性搜索策略确定。

以上带一维步长因子的牛顿法称为**阻尼牛顿法**或**带步长因子牛顿法**。在G_k正定前提下，保证$\{f_k\}$单调下降，且当$\boldsymbol{x}_k$离$\boldsymbol{x}^*$稍远，产生点列$\{\boldsymbol{x}_k\}$仍能收敛。

阻尼牛顿法

算法4.2.2 阻尼牛顿法

步1 取初始点$\boldsymbol{x}_0$，终止误差$\varepsilon > 0$，令$k := 0$;

步2 若$\|\boldsymbol{g}_k\| \leq \varepsilon$，停止迭代，输出$\boldsymbol{x}_k$;

步3 解方程组构造牛顿方向，即解：$G_k \boldsymbol{d} = -\boldsymbol{g}_k$得$\boldsymbol{d}_k$;

步4 进行线性搜索求α_k，使得

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) = \min_{\alpha \geq 0} f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k).$$

步5 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, k := k + 1$，转步2。

阻尼牛顿法全局收敛性

对于严格凸函数，采用Wolfe准则的阻尼Newton方法具有全局收敛性。

定理4.2.2 阻尼牛顿法/Wolfe全局收敛性.

设$f(\boldsymbol{x}) \in C^2$，且对任给$\boldsymbol{x}_0 \in \mathbb{R}^n$，存在常数$\beta > 0$，使得$f(\boldsymbol{x})$在水平集$L(\boldsymbol{x}_0) = \{\boldsymbol{x} | f(\boldsymbol{x}) \le f(\boldsymbol{x}_0)\}$上满足:

$$\boldsymbol{u}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{u} \ge \beta \|\boldsymbol{u}\|^2, \ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{x} \in L(\boldsymbol{x}_0). \tag{4.2.5}$$

则在采用Wolfe准则的阻尼牛顿法产生的点列$\{\boldsymbol{x}_k\}$满足:

- $\{\boldsymbol{x}_k\}$为有穷点列，即存在N，使得$\boldsymbol{g}_N = 0$;
- $\{\boldsymbol{x}_k\}$为无穷点列，$\{\boldsymbol{x}_k\}$收敛到$f(\boldsymbol{x})$的唯一极小点。

定理4.2.2证明

证明： 首先证明水平集$L(\boldsymbol{x}_0)$为有界闭凸集。(详见教材53页，分别证明凸性、闭性、有界性)。由(4.2.5)与凸函数判定定理知，$f(\boldsymbol{x})$为$L(\boldsymbol{x}_0)$上严格凸函数，从而其稳定点为唯一全局极小点。

若$\{\boldsymbol{x}_k\}$为无穷点列，由于$\{\boldsymbol{x}_k\} \subset L(\boldsymbol{x}_0)$，存在极限点$\tilde{\boldsymbol{x}} \in L(\boldsymbol{x}_0)$及子列$\{\boldsymbol{x}_k\}_{\mathcal{K}}, \boldsymbol{x}_k \to \tilde{\boldsymbol{x}}, k \in \mathcal{K}$. 因阻尼牛顿法产生的$\{f_k\}$单调下降、有下界，故必有：

$$f_k \to f(\tilde{\boldsymbol{x}})$$

根据非精确线性搜索方法的收敛性定理(定理3.2.3)，只要证明如下夹角条件满足：

$$\theta_k \leq \frac{\pi}{2} - \mu(\mu > 0)$$

即可得：$\boldsymbol{g}_k \to \tilde{\boldsymbol{g}} = 0$. 由稳定点的唯一性知$\boldsymbol{x}^* = \tilde{\boldsymbol{x}}$。

定理4.2.2证明

证明续： 由于$G(\boldsymbol{x})$连续和$L(\boldsymbol{x}_0)$为有界闭凸集知，存在$\gamma > 0$, 对任意给的$\boldsymbol{x} \in L(\boldsymbol{x}_0)$, 有$\|G(\boldsymbol{x})\| \leq \gamma$, 则

$$\|\boldsymbol{g}_k\| = \|G_k \boldsymbol{d}_k\| \leq \gamma \|\boldsymbol{d}_k\|$$

由(4.2.5)知：

$$\frac{\pi}{2} - \theta_k \geq \sin\left(\frac{\pi}{2} - \theta_k\right) = \cos\theta_k = \frac{-\boldsymbol{g}_k^T \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|\|\boldsymbol{d}_k\|} = \frac{\boldsymbol{d}_k^T G \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|\|\boldsymbol{d}_k\|} \geq \frac{\beta}{\gamma} \quad (4.2.6)$$

即夹角条件满足：$\theta_k \leq \frac{\pi}{2} - \frac{\beta}{\gamma}$。 ■

修正牛顿法– 混合方法

阻尼牛顿法迭代过程中，可能会出现海森矩阵可逆，不正定情形，导致牛顿方向与梯度方向正交。采用修正方法可以克服这些问题。针对海森矩阵始终可逆，但可能非正定的情形，混合修正方法如下(设$\boldsymbol{g}_k \neq 0$)：

- 若G_k正定，则:

  $$\boldsymbol{d}_k = -G_k^{-1}\boldsymbol{g}_k$$

  是下降方向。
- 若G_k不正定且可逆，且$\boldsymbol{g}_k^T G_k^{-1}\boldsymbol{g}_k < 0$，则

  $$\boldsymbol{d}_k = G_k^{-1}\boldsymbol{g}_k$$

  是下降方向。
- 若G_k不正定，且奇异或$G_k^{-1}\boldsymbol{g}_k = 0$，则取:

  $$\boldsymbol{d}_k = -\boldsymbol{g}_k.$$

Levenberg-Marquardt, LM方法

LM方法是处理G_k奇异、不正定等情形的一个最简单方法。该方法通过求解以下方程组：

$$(G_k + \nu_k I)\boldsymbol{d} = -\boldsymbol{g}_k$$

来确定迭代方向的牛顿型方法，这里$\nu_k > 0$, I是单位阵。

- 当ν_k足够大时，可以保证$(G_k + \nu_k I)$正定；
- 当ν_k很小时，方程组的解偏向于牛顿方向，随着ν_k的增大，方程组的解向负梯度方向偏移。

这种修正牛顿方法的思想来自解最小二乘问题的LM方法，具体修正ν_k的方法将在最小二乘部分介绍。这里当$G_k + \nu_k I$不正定时，简单地取$\nu_k := 2\nu_k$。

LM方法算例

例4.2.2 从不同初始点出发，用LM方法求解4.2.1问题。方法采用Wolfe线性搜索准则。

初始点为 $(1.5, 1.5)^T$ 时，LM 方法每一迭代步的结果

迭代次数 k	$x^{(k)T}$	f_k	$\|\|g_k\|\|_\infty$
0	(1.5000, 1.5000)	10.1250	8.1125
1	(−0.2121, 0.2771)	0.3528	1.9875
2	(−0.0481, 0.0768)	0.0244	0.5376
3	(−0.0022, 0.0051)	0.0001	0.0332
4	(−0.0000, 0.0000)	0.0000	0.0001
5	(0.0000, 0.0000)	0.0000	0.0000

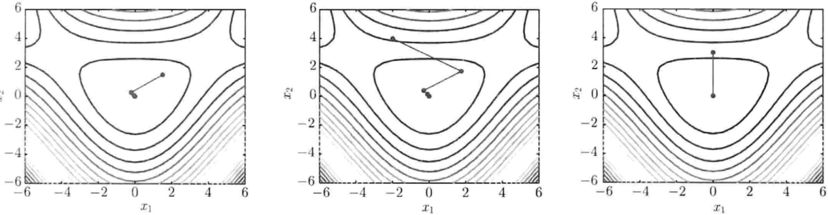
初始点为 $(-2, 4)^T$ 时，LM 方法每一迭代步的结果

迭代次数 k	$x^{(k)T}$	f_k	$\|\|g_k\|\|_\infty$
0	(−2.0000, 4.0000)	44.0000	20.3961
1	(1.7599, 1.7441)	13.0152	8.5918
2	(−0.2963, 0.4088)	0.7288	2.8197
3	(−0.0886, 0.1511)	0.0908	1.0306
4	(−0.0070, 0.0178)	0.0011	0.1145
5	(−0.0000, 0.0001)	0.0000	0.0008
6	(0.0000, 0.0000)	0.0000	0.0000

LM方法算例

初始点为 $(0,3)^T$ 时，LM 方法每一迭代步的结果

迭代次数 k	$x^{(k)T}$	f_k	$\|g_k\|_\infty$
0	(0.0000, 3.0000)	27.0000	18.0000
1	(0.0000, −0.0000)	0.0000	0.0000



(a) 初始点为$(1.5,1.5)^T$　(b) 初始点为$(-2,4)^T$　(c) 初始点为$(0,3)^T$

$(-2,4)^T$接近鞍点，$(0,3)^T$处Hesse矩阵不正定，但LM方法均能够产生收敛迭代点列。