非线性最小二乘问题

第一章：

局部最优的必要条件是一阶导为 0，驻点可以按照如下方式去区分为极大极小值点：

1. 若$F''(x_s)$正定，则 $\boldsymbol{x}_s$ 点为极小值点
2. 若 $F''(x_s)$负定，则$\boldsymbol{x}_s$点为极大值点

其中，$F''(x_s)$为对应点的海塞矩阵。对于非线性的目标函数，没办法求其导数为 0 的解析解，所以一般上通过迭代下降的方法，使得$F(x)$的值不断下降，也即下降条件：$F(\mathbf{x}_{k+1}) < F(\mathbf{x}_k)$

最小二乘问题的定义：

Definition 1.1. Least Squares Problem

Find $\mathbf{x}^*$, a local minimizer for[1)]

$$F(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m}(f_i(\mathbf{x}))^2 ,$$

where $f_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i=1,\ldots,m$ are given functions, and $m \geq n$.

全局最小和局部最小的定义：

Definition 1.2. Global Minimizer
Given $F : \mathbb{R}^n \mapsto \mathbb{R}$. Find

$$\mathbf{x}^+ = \operatorname{argmin}_{\mathbf{x}}\{F(\mathbf{x})\} .$$

Definition 1.3. Local Minimizer
Given $F : \mathbb{R}^n \mapsto \mathbb{R}$. Find $\mathbf{x}^*$ so that

$$F(\mathbf{x}^*) \leq F(\mathbf{x}) \quad \text{for} \quad \|\mathbf{x}-\mathbf{x}^*\| < \delta .$$

关于局部最优的一个局部最优导数为 0 定理和驻点的定义：

Theorem 1.5. Necessary condition for a local minimizer.
If $\mathbf{x}^*$ is a local minimizer, then

$$\mathbf{g}^* \equiv \mathbf{F}'(\mathbf{x}^*) = \mathbf{0} .$$

Definition 1.6. Stationary point. If

$$\mathbf{g}_s \equiv \mathbf{F}'(\mathbf{x}_s) = \mathbf{0} ,$$

then $\mathbf{x}_s$ is said to be a *stationary point* for F.

第二章：

下降方法的定义：

Algorithm 2.4. Descent method

```
begin
  k := 0;  x := x_0;  found := false                  {Starting point}
  while (not found) and (k < k_max)
    h_d := search_direction(x)                        {From x and downhill}
    if (no such h exists)
      found := true                                   {x is stationary}
    else
      α := step_length(x, h_d)                         {from x in direction h_d}
      x := x + α h_d;   k := k+1                       {next iterate}
end
```

泰勒公式：

$$
\begin{aligned}
F(\mathbf{x}+\alpha\mathbf{h}) &= F(\mathbf{x}) + \alpha\mathbf{h}^\top\mathbf{F}'(\mathbf{x}) + O(\alpha^2) \\
&\simeq F(\mathbf{x}) + \alpha\mathbf{h}^\top\mathbf{F}'(\mathbf{x}) \quad \text{for } \alpha \text{ sufficiently small.}
\end{aligned}
$$

下降方向 h 满足 $\mathbf{h}^\top\mathbf{F}'(\mathbf{x}) < 0$

Definition 2.6. Descent direction.

$\mathbf{h}$ is a descent direction for F at $\mathbf{x}$ if $\quad \mathbf{h}^\top\mathbf{F}'(\mathbf{x}) < 0$.

选取步长：

$$\alpha_e = \operatorname{argmin}_{\alpha>0}\{F(\mathbf{x}+\alpha\mathbf{h})\} .$$

最速下降法

■令 $\boldsymbol{x}^{(0)}$ 作为初始搜索点，并沿着梯度负方向构造一个新点 $\boldsymbol{x}^{(0)}-\alpha\nabla f(\boldsymbol{x}^{(0)})$，由泰勒定理可得

$$f(\boldsymbol{x}^{(0)}-\alpha\nabla f(\boldsymbol{x}^{(0)}))=f(\boldsymbol{x}^{(0)})-\alpha\|\nabla f(\boldsymbol{x}^{(0)})\|^2+o(\alpha)$$

■因此，如果 $\nabla f(\boldsymbol{x}^{(0)})\neq\boldsymbol{0}$，那么当 $\alpha>0$ 足够小时，有

$$f(\boldsymbol{x}^{(0)}-\alpha\nabla f(\boldsymbol{x}^{(0)}))<f(\boldsymbol{x}^{(0)})$$

■可以设计一种方法实现以上理念。给定一个搜索点 $\boldsymbol{x}^{(k)}$ 由此点出发，根据向量 $-\alpha_k\nabla f(\boldsymbol{x}^{(k)})$ 指定的方向和幅值运动，构造新点 $\boldsymbol{x}^{(k+1)}$，其中，α_k 是一个正实数，称为步长。这样，可得到迭代公式:

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}-\alpha_k\nabla f(\boldsymbol{x}^{(k)})$$

■最速下降法

■最速下降法是梯度方法的一种具体实现，其理念为在每次迭代中选择合适的步长 α_k，使得目标函数值能够最大程度减小

■α_k 满足:

$$\alpha_k=\arg\min_{\alpha\geqslant 0} f(\boldsymbol{x}^{(k)}-\alpha\nabla f(\boldsymbol{x}^{(k)}))$$

每一次迭代的相对增益

$$\lim_{\alpha\to 0}\frac{F(\mathbf{x})-F(\mathbf{x}+\alpha\mathbf{h})}{\alpha\|\mathbf{h}\|}=-\frac{1}{\|\mathbf{h}\|}\mathbf{h}^\top\mathbf{F}'(\mathbf{x})=-\|\mathbf{F}'(\mathbf{x})\|\cos\theta$$

选择梯度的负方向：

$$h_{sd}=-F'(x)$$

- 如果有 $\nabla f(\boldsymbol{x}^{(k)})=\mathbf{0}$ ，说明 $\boldsymbol{x}^{(k)}$ 满足局部极小点一阶必要条件，这可以作为设计迭代停止规则的基础。
- 在实际应用中，采用数值计算方法很难恰好得到梯度为零的结果，因此，$\nabla f(\boldsymbol{x}^{(k)})=\mathbf{0}$ 并不适合直接作为停止规则。
- 一种实用的停止规则是计算梯度的范数 $\|\nabla f(\boldsymbol{x}^{(k)})\|$ ，如果小于某个预设的阈值，则迭代停止。

牛顿法

牛顿法

- 求解一元单值函数 f 在区间上求极小点的问题
- 假设函数连续二阶可微，即 $x^{(k)}$ 处的 $f'(x^{(k)})$ 和 $f''(x^{(k)})$ 均可求得。
- 构造一个经过点 $(x^{(k)}, f(x^{(k)}))$ 处的二次函数，该函数在的一阶和二阶导数分别为 $f'(x^{(k)})$ 和 $f''(x^{(k)})$ ，如下所示：

$$q(x)=f(x^{(k)})+f'(x^{(k)})(x-x^{(k)})+\frac{1}{2}f''(x^{(k)})(x-x^{(k)})^2$$

- $q(x^{(k)})$ 为 $f(x^{(k)})$ 在 $x^{(k)}$ 的二阶近似。

牛顿法

- 求函数 f 的极小点可近似于求解 q 的极小点
- 函数 q 的极小点应满足一阶必要条件：

$$0=q'(x)=f'(x^{(k)})+f''(x^{(k)})(x-x^{(k)})$$

- 令 $x=x^{(k+1)}$ ，可得：

$$x^{(k+1)}=x^{(k)}-\frac{f'(x^{(k)})}{f''(x^{(k)})}$$

从驻点出发，由于驻点的一阶导数为 0，结合泰勒展开：

$$\begin{aligned}\mathbf{F}'(\mathbf{x}+\mathbf{h}) &= \mathbf{F}'(\mathbf{x}) + \mathbf{F}''(\mathbf{x})\mathbf{h} + O(\|\mathbf{h}\|^2) \\ &\simeq \mathbf{F}'(\mathbf{x}) + \mathbf{F}''(\mathbf{x})\mathbf{h} \quad \text{for } \|\mathbf{h}\| \text{ sufficiently small}\end{aligned}$$

使一阶导为 0：

$$\mathbf{H}\,\mathbf{h}_{\mathrm{n}} = -\mathbf{F}'(\mathbf{x}) \quad \text{with } \mathbf{H} = \mathbf{F}''(\mathbf{x})$$

下一次的迭代表示为： $\mathbf{x} := \mathbf{x} + \mathbf{h}_{\mathrm{n}}$ 。但是这个思路目的是为了求可能的驻点，在下降的方向上，会不会出现问题呢，因此这里还需要验证一下：

假设这个海塞矩阵 H 是正定的，也就是说这个驻点是极小值点。我们对如下式子两边同乘一个 h_n^T 。

$\mathbf{H}\,\mathbf{h}_{\mathrm{n}} = -\mathbf{F}'(\mathbf{x})$ 就变成了 $0 < \mathbf{h}_{\mathrm{n}}^{\top}\mathbf{H}\,\mathbf{h}_{\mathrm{n}} = -\mathbf{h}_{\mathrm{n}}^{\top}\mathbf{F}'(\mathbf{x})$ ，$\mathbf{h}_{\mathrm{n}}^{\top}\mathbf{H}\,\mathbf{h}_{\mathrm{n}}$ 是一个二次型，一定正定，满足定义 2.6，使得我们所求的方向一定是一个下降方向。

Definition 2.6. Descent direction.

$\mathbf{h}$ is a descent direction for F at $\mathbf{x}$ if $\mathbf{h}^{\top}\mathbf{F}'(\mathbf{x}) < 0$.

基于最速下降和牛顿法的混合模式：

if $\mathbf{F}''(\mathbf{x})$ is positive definite
  $\mathbf{h} := \mathbf{h}_{\mathrm{n}}$
else
  $\mathbf{h} := \mathbf{h}_{\mathrm{sd}}$
$\mathbf{x} := \mathbf{x} + \alpha\mathbf{h}$

这里的 $\mathbf{h}_{\mathrm{sd}}$ 是由最速下降法得到的下降方向，而 α 是由线性搜索得到的。至于如何去检查这个 h 是否正定，作者这里也提了一嘴 Cholesky 方法。

线性搜索的核心选取步长 α ：

$$\alpha_{\mathrm{e}} = \operatorname{argmin}_{\alpha>0}\{F(\mathbf{x}+\alpha\mathbf{h})\}$$

满足如下条件时会停下来：

$|\varphi'(\alpha_s)| \leq \tau\, |\varphi'(0)|$ 而 t 在这里是一个比较小的正值。但是准确的线性搜索可能是非常浪费时间的，而我们实际上不是真的很需要一个非常精准的φ，因此更多的我们会使用软线性搜索：

$$\varphi(\alpha_s) \leq \varphi(0) + \gamma_1 \cdot \varphi'(0) \cdot \alpha \quad \text{with } 0 < \gamma_1 < 1$$

但是仅有这一条还不够，这样可能导致这个新的 a 值太过接近初始值，因此还需要加上以下这条限制：

$$\varphi'(\alpha_s) \geq \gamma_2 \cdot \varphi'(0) \quad \text{with } \gamma_1 < \gamma_2 < 1$$

在满足以上两条 条件时，这个 a 就被我们接受为需要的步长。

信任区域法和阻尼法：

我们首先对当前估计状态附近的原代价函数进行一个近似：

$F(\mathbf{x}+\mathbf{h}) \simeq L(\mathbf{h}) \equiv F(\mathbf{x}) + \mathbf{h}^\top \mathbf{c} + \frac{1}{2}\mathbf{h}^\top \mathbf{B}\,\mathbf{h}$。这里的 c 和 B 都是对称阵。

在信任区域法中，我们首先假设一致一个正数 δ 且在这个半径内我们的模型足够精确。那么步长可以按照以下方式来确定：

$$\mathbf{h} = \mathbf{h}_{\mathrm{tr}} \equiv \operatorname{argmin}_{\|\mathbf{h}\| \leq \Delta}\{L(\mathbf{h})\}$$

而在阻尼法中，步长按以下方法来确定：

$\mathbf{h} = \mathbf{h}_{\mathrm{dm}} \equiv \operatorname{argmin}_{\mathbf{h}}\{L(\mathbf{h}) + \frac{1}{2}\mu\, \mathbf{h}^\top \mathbf{h}\}$后面加的这一项相当于对大步长的惩罚。

介绍这两个东西有什么用呢？这两种方法在实现上更加简单一点。那么相对于最速下降法和牛顿法，如何验证这两种估计方法的精度呢？给出如下方法：

定义：$\varrho = \dfrac{F(\mathbf{x}) - F(\mathbf{x}+\mathbf{h})}{L(\mathbf{0}) - L(\mathbf{h})}$，用来表示实际与预测下降之间的比率。如果这个比率太小，说明说明实际下降误差过小于模型下降误差，模型计算的下降

步长过大，因此对于信任区域法来说，要减小 δ 的值，对于阻尼法来说，则要增大阻尼参数 μ 。反之亦然。

第三章

非线性最小二乘问题：

对于给定的向量函数$\mathbb{R}^n \mapsto \mathbb{R}^m$，且 m≥n 我们想最小化这个向量函数的二范数 $\|\mathbf{f}(\mathbf{x})\|$，也即相当于去找$\mathbf{x}^* = \mathrm{argmin}_{\mathbf{x}}\{F(\mathbf{x})\}$，这里的：

$$F(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m}(f_i(\mathbf{x}))^2 = \frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2 = \frac{1}{2}\mathbf{f}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$$

假定 f 的泰勒展开如下：

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h} + O(\|\mathbf{h}\|^2)$$

这里的 J 是雅克比矩阵 $(\mathbf{J}(\mathbf{x}))_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x})$ 。

由于 $\frac{\partial F}{\partial x_j}(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})\frac{\partial f_i}{\partial x_j}(\mathbf{x})$. ，因此大 F 的导数可以表示为：

$$\mathbf{F}'(\mathbf{x}) = \mathbf{J}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$$

对于大 F 的二阶导，我们需要首先引入海塞矩阵：

$$\frac{\partial^2 F}{\partial x_j \partial x_k}(\mathbf{x}) = \sum_{i=1}^{m}\left(\frac{\partial f_i}{\partial x_j}(\mathbf{x})\frac{\partial f_i}{\partial x_k}(\mathbf{x}) + f_i(\mathbf{x})\frac{\partial^2 f_i}{\partial x_j \partial x_k}(\mathbf{x})\right)$$

因此 $\mathbf{F}''(\mathbf{x}) = \mathbf{J}(\mathbf{x})^\top \mathbf{J}(\mathbf{x}) + \sum_{i=1}^{m} f_i(x)\mathbf{f}_i''(\mathbf{x})$

第一节:高斯牛顿法

高斯-牛顿迭代法的基本思想是，使用泰勒级数展开式去近似地代替非线性回归模型，然后通过多次迭代，多次修正回归系数，使回归系数不断通过逼近非线性回归模型的最佳回归系数，最后使原模型的残差平方和达到最小。

首先在估计点附近展开小 f：$\mathbf{f}(\mathbf{x}+\mathbf{h}) \simeq \boldsymbol{\ell}(\mathbf{h}) \equiv \mathbf{f}(\mathbf{x}) + \mathbf{J}(\mathbf{x})\mathbf{h}$，带入进大 F，有：

$$
\begin{aligned}
F(\mathbf{x}+\mathbf{h}) \simeq L(\mathbf{h}) &\equiv \tfrac{1}{2}\boldsymbol{\ell}(\mathbf{h})^\top \boldsymbol{\ell}(\mathbf{h}) \\
&= \tfrac{1}{2}\mathbf{f}^\top\mathbf{f} + \mathbf{h}^\top\mathbf{J}^\top\mathbf{f} + \tfrac{1}{2}\mathbf{h}^\top\mathbf{J}^\top\mathbf{J}\mathbf{h} \\
&= F(\mathbf{x}) + \mathbf{h}^\top\mathbf{J}^\top\mathbf{f} + \tfrac{1}{2}\mathbf{h}^\top\mathbf{J}^\top\mathbf{J}\mathbf{h}
\end{aligned}
$$

高斯牛顿法的步长

$\mathbf{h}_{\mathrm{gn}}$ 会试图最小化 $L(\mathbf{h})$：$\mathbf{h}_{\mathrm{gn}} = \mathrm{argmin}_{\mathbf{h}}\{L(\mathbf{h})\}$

$$\mathbf{L}'(\mathbf{h}) = \mathbf{J}^\top\mathbf{f} + \mathbf{J}^\top\mathbf{J}\mathbf{h}, \qquad \mathbf{L}''(\mathbf{h}) = \mathbf{J}^\top\mathbf{J}$$，

类似于牛顿法，还是找一阶导数为 0 的点，因此有

$$(\mathbf{J}^\top\mathbf{J})\mathbf{h}_{\mathrm{gn}} = -\mathbf{J}^\top\mathbf{f}$$

综上：

$$
\begin{aligned}
&\text{Solve } (\mathbf{J}^\top\mathbf{J})\mathbf{h}_{\mathrm{gn}} = -\mathbf{J}^\top\mathbf{f} \\
&\mathbf{x} := \mathbf{x} + \alpha\mathbf{h}_{\mathrm{gn}}
\end{aligned}
$$

这里的步长仍然需要用线性搜索去实现。

传统的高斯牛顿法中 α 是一个定值为 1。

第二节：The Levenberg–Marquardt Method

Algorithm 3.16. Levenberg–Marquardt method

begin
  $k := 0; \quad \nu := 2; \quad \mathbf{x} := \mathbf{x}_0$
  $\mathbf{A} := \mathbf{J}(\mathbf{x})^\top \mathbf{J}(\mathbf{x}); \quad \mathbf{g} := \mathbf{J}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$
  $found := (\|\mathbf{g}\|_\infty \le \varepsilon_1); \quad \mu := \tau * \max\{a_{ii}\}$
  while (**not** *found*) **and** $(k < k_{\max})$
    $k := k+1;$ Solve $(\mathbf{A} + \mu\mathbf{I})\mathbf{h}_{\mathrm{lm}} = -\mathbf{g}$
    if $\|\mathbf{h}_{\mathrm{lm}}\| \le \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2)$
      found := **true**
    else
      $\mathbf{x}_{\mathrm{new}} := \mathbf{x} + \mathbf{h}_{\mathrm{lm}}$
      $\varrho := (F(\mathbf{x}) - F(\mathbf{x}_{\mathrm{new}}))/(L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{lm}}))$
      if $\varrho > 0$ {step acceptable}
        $\mathbf{x} := \mathbf{x}_{\mathrm{new}}$
        $\mathbf{A} := \mathbf{J}(\mathbf{x})^\top \mathbf{J}(\mathbf{x}); \quad \mathbf{g} := \mathbf{J}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$
        $found := (\|\mathbf{g}\|_\infty \le \varepsilon_1)$
        $\mu := \mu * \max\{\frac{1}{3}, 1 - (2\varrho - 1)^3\}; \quad \nu := 2$
      else
        $\mu := \mu * \nu; \quad \nu := 2 * \nu$
end

LM 算法是一种阻塞式高斯牛顿法。这也是使用最广泛的非线性最小二乘算法。其区别在于其步长$\mathbf{h}_{\mathrm{lm}}$的计算按照如下方法实现：

$$(\mathbf{J}^\top \mathbf{J} + \mu\mathbf{I})\mathbf{h}_{\mathrm{lm}} = -\mathbf{g} \quad \text{with } \mathbf{g} = \mathbf{J}^\top \mathbf{f} \text{ and } \mu \ge 0$$

这里的 μ 是阻塞系数，类似于阻塞牛顿法，其也有如下几种可能的影响：

如果 μ 全为正数，那么这就能保证$\mathbf{h}_{\mathrm{lm}}$是下降方向的

如果 μ 特别大，我们可以忽略掉其他的东西，解出 h 有：

$$\mathbf{h}_{\mathrm{lm}} \simeq -\frac{1}{\mu}\mathbf{g} = -\frac{1}{\mu}\mathbf{F}'(\mathbf{x})$$

反之，如果 μ 特别小，那么这个算法其实就和高斯牛顿法差不多了。可以看出这个 μ 对于下降方向以及步长的大小都有着比较重要的影响。

增益率：$\varrho = \dfrac{F(\mathbf{x}) - F(\mathbf{x}+\mathbf{h}_{\mathrm{lm}})}{L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{lm}})}$，分母和是线性模型预测的结果的增益

$$\begin{aligned} L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{lm}}) &= -\mathbf{h}_{\mathrm{lm}}{}^{\top}\mathbf{J}^{\top}\mathbf{f} - \tfrac{1}{2}\mathbf{h}_{\mathrm{lm}}{}^{\top}\mathbf{J}^{\top}\mathbf{J}\mathbf{h}_{\mathrm{lm}} \\ &= -\tfrac{1}{2}\mathbf{h}_{\mathrm{lm}}{}^{\top}\big(\mathbf{2g} + (\mathbf{J}^{\top}\mathbf{J} + \mu\mathbf{I} - \mu\mathbf{I})\mathbf{h}_{\mathrm{lm}}\big) \\ &= \tfrac{1}{2}\mathbf{h}_{\mathrm{lm}}{}^{\top}(\mu\mathbf{h}_{\mathrm{lm}} - \mathbf{g}) \ . \end{aligned}$$

停止准则，我们希望能够反应一个最小值点而并非一个局部的极小值点，也即 $\mathbf{F}'(\mathbf{x}^*) = \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$，因此我们可以用 $\|\mathbf{g}\|_\infty \le \varepsilon_1$，这里 e 是一个非常小的值。或者使用：$\|\mathbf{x}_{\mathrm{new}} - \mathbf{x}\| \le \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2)$。

第三节：**Dog Leg Method**

Algorithm 3.21. Dog Leg Method

begin
  $k := 0;\quad \mathbf{x} := \mathbf{x}_0;\quad \Delta := \Delta_0;\quad \mathbf{g} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{f}(\mathbf{x})$ {1°}
  found $:= (\|\mathbf{f}(\mathbf{x})\|_\infty \le \varepsilon_3)$ **or** $(\|\mathbf{g}\|_\infty \le \varepsilon_1)$ {2°}
  while (**not** *found*) **and** $(k < k_{\max})$
    $k := k+1;$ Compute α by (3.19)
    $\mathbf{h}_{\mathrm{sd}} := -\alpha\mathbf{g};$ Solve $\mathbf{J}(\mathbf{x})\mathbf{h}_{\mathrm{gn}} \simeq -\mathbf{f}(\mathbf{x})$ {3°}
    Compute $\mathbf{h}_{\mathrm{dl}}$ by (3.20)
    if $\|\mathbf{h}_{\mathrm{dl}}\| \le \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2)$
      found := **true**
    else
      $\mathbf{x}_{\mathrm{new}} := \mathbf{x} + \mathbf{h}_{\mathrm{dl}}$
      $\varrho := (F(\mathbf{x}) - F(\mathbf{x}_{\mathrm{new}}))/(L(\mathbf{0}) - L(\mathbf{h}_{\mathrm{dl}}))$ {4°}
      if $\varrho > 0$
        $\mathbf{x} := \mathbf{x}_{\mathrm{new}};\quad \mathbf{g} := \mathbf{J}(\mathbf{x})^{\top}\mathbf{f}(\mathbf{x})$
        found $:= (\|\mathbf{f}(\mathbf{x})\|_\infty \le \varepsilon_3)$ **or** $(\|\mathbf{g}\|_\infty \le \varepsilon_1)$
      if $\varrho > 0.75$ {5°}
        $\Delta := \max\{\Delta, 3*\|\mathbf{h}_{\mathrm{dl}}\|\}$
      elseif $\varrho < 0.25$
        $\Delta := \Delta/2;$ *found* $:= (\Delta \le \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2))$ {6°}
end

LM 算法需要对每一个待估参数求偏导，所以如果你的拟合函数 f 非常复杂，或者待估参数相当地多，那么可能不适合使用 LM 算法，而可以选择 **Dog Leg Method** 算法。

最速下降法的方向向量可以按以下方法给出：$\mathbf{h}_{\mathrm{sd}} = -\mathbf{g} = -\mathbf{J}(\mathbf{x})^{\top}\mathbf{f}(\mathbf{x})$

$$
\begin{aligned}
\mathbf{f}(\mathbf{x}+\alpha\mathbf{h}_{\mathrm{sd}}) &\simeq \mathbf{f}(\mathbf{x}) + \alpha\mathbf{J}(\mathbf{x})\mathbf{h}_{\mathrm{sd}} \\
\Downarrow \qquad & \\
F(\mathbf{x}+\alpha\mathbf{h}_{\mathrm{sd}}) &\simeq \tfrac{1}{2}\|\mathbf{f}(\mathbf{x}) + \alpha\mathbf{J}(\mathbf{x})\mathbf{h}_{\mathrm{sd}}\|^2 \\
&= F(\mathbf{x}) + \alpha\mathbf{h}_{\mathrm{sd}}{}^{\top}\mathbf{J}(\mathbf{x})^{\top}\mathbf{f}(\mathbf{x}) + \tfrac{1}{2}\alpha^2\|\mathbf{J}(\mathbf{x})\mathbf{h}_{\mathrm{sd}}\|^2
\end{aligned}
$$

α 的最小值可以由以下算式得到：

$$
\alpha = -\frac{\mathbf{h}_{\mathrm{sd}}{}^{\top}\mathbf{J}(\mathbf{x})^{\top}\mathbf{f}(\mathbf{x})}{\|\mathbf{J}(\mathbf{x})\mathbf{h}_{\mathrm{sd}}\|^2} = \frac{\|\mathbf{g}\|^2}{\|\mathbf{J}(\mathbf{x})\mathbf{g}\|^2}
$$

使用如下方法来选择最终步长，其中置信区域半径为 δ

if $\|\mathbf{h}_{\mathrm{gn}}\| \le \Delta$
    $\mathbf{h}_{\mathrm{dl}} := \mathbf{h}_{\mathrm{gn}}$
elseif $\|\alpha\mathbf{h}_{\mathrm{sd}}\| \ge \Delta$
    $\mathbf{h}_{\mathrm{dl}} := (\Delta/\|\mathbf{h}_{\mathrm{sd}}\|)\mathbf{h}_{\mathrm{sd}}$
else
    $\mathbf{h}_{\mathrm{dl}} := \alpha\mathbf{h}_{\mathrm{sd}} + \beta(\mathbf{h}_{\mathrm{gn}} - \alpha\mathbf{h}_{\mathrm{sd}})$
        with β chosen so that $\|\mathbf{h}_{\mathrm{dl}}\| = \Delta$.

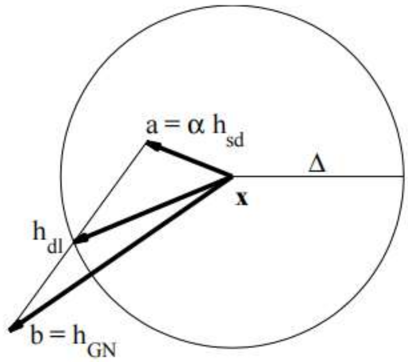

Figure 3.4. *Trust region and Dog Leg step.*[4)]