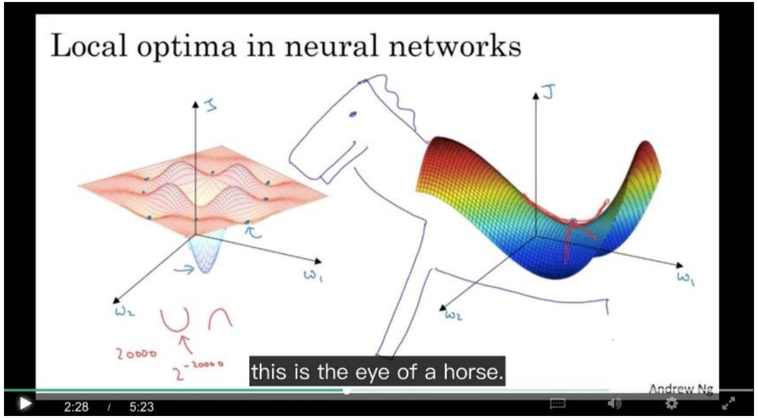
4. 非线性最小二乘

4. 非线性最小二乘

- 先考虑简单的问题： $\min_{x} \frac{1}{2}\|f(x)\|_2^2.$ 这里 $x \in \mathbb{R}^n$ ，f 为任意函数
- 当 f 很简单时：
  - 解： $\frac{\mathrm{d}f}{\mathrm{d}x} = 0.$ 将得到极值点或鞍点，比较这些解即可
- 当 f 复杂时：
  - df/dx难求，或df/dx=0很难解
  - 使用迭代方式求解



4. 非线性最小二乘

- 迭代方式：

问题：如何确定这个增量？

1. 给定某个初始值 $\boldsymbol{x}_0$。
2. 对于第 k 次迭代，寻找一个增量 $\Delta\boldsymbol{x}_k$，使得 $\|f(\boldsymbol{x}_k+\Delta\boldsymbol{x}_k)\|_2^2$ 达到极小值。
3. 若 $\Delta\boldsymbol{x}_k$ 足够小，则停止。
4. 否则，令 $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\Delta\boldsymbol{x}_k$，返回 2.

4. 非线性最小二乘

- 确定增量的方法（即梯度下降策略）：一阶的或二阶的
- 泰勒展开：

$$\|f(x+\Delta x)\|_2^2 \approx \|f(x)\|_2^2 + J(x)\Delta x + \frac{1}{2}\Delta x^T H \Delta x.$$

- 若只保留一阶梯度：

$$\min_{\Delta x} \|f(x)\|_2^2 + J\Delta x$$

增量的方向： $\Delta x^* = -J^T(x).$ （通常还需要计算步长）

- 称为最速下降法（Steepest Method）

4. 非线性最小二乘

- 若保留二阶梯度：

$$\|f(x+\Delta x)\|_2^2 \approx \|f(x)\|_2^2 + \boldsymbol{J}(x)\Delta x + \frac{1}{2}\Delta x^T \boldsymbol{H}\Delta x.$$

$$\Delta x^* = \arg\min \|f(x)\|_2^2 + \boldsymbol{J}(x)\Delta x + \frac{1}{2}\Delta x^T \boldsymbol{H}\Delta x.$$

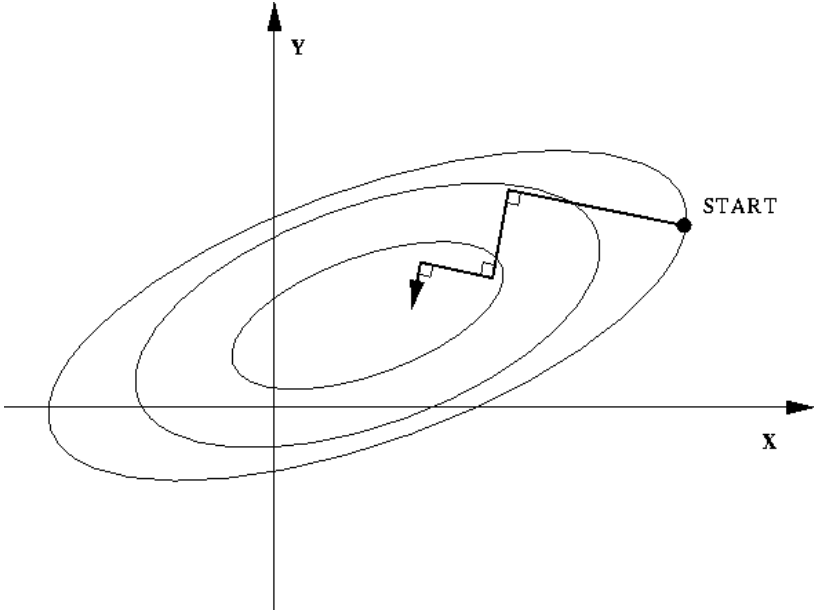
- 则得到（令上式关于 Δx 的导数为零）：

$$\boldsymbol{H}\Delta x = -\boldsymbol{J}^T.$$

- 该方法又称为牛顿法

4. 非线性最小二乘

- 最速下降法和牛顿法虽然直观，但实用当中存在一些缺点
  - 最速下降法会碰到zigzag问题（过于贪婪）
  - 牛顿法迭代次数少，但需要计算复杂的Hessian矩阵
- 能否回避Hessian的计算？
  - Gauss-Newton
  - Levenberg-Marquadt



4. 非线性最小二乘

- Gauss-Newton
  - 一阶近似 f(x)： $$f\left(x+\Delta x\right)\approx f\left(x\right)+J\left(x\right)\Delta x.$$
  - 平方误差变为：

$$\begin{aligned}\frac{1}{2}\|f\left(x\right)+J\left(x\right)\Delta x\|^2&=\frac{1}{2}\left(f\left(x\right)+J\left(x\right)\Delta x\right)^T\left(f\left(x\right)+J\left(x\right)\Delta x\right)\\&=\frac{1}{2}\left(\|f(x)\|_2^2+2f\left(x\right)^TJ(x)\Delta x+\Delta x^TJ(x)^TJ(x)\Delta x\right).\end{aligned}$$

  - 令关于 Δx 导数为零：

$$2J(x)^Tf\left(x\right)+2J(x)^TJ\left(x\right)\Delta x=\mathbf{0}.$$

$$J(x)^TJ\left(x\right)\Delta x=-J(x)^Tf\left(x\right).$$

$$\Rightarrow \quad H\Delta x=g.$$

4. 非线性最小二乘

$$\boldsymbol{J}(\boldsymbol{x})^T \boldsymbol{J}(\boldsymbol{x}) \Delta \boldsymbol{x} = -\boldsymbol{J}(\boldsymbol{x})^T f(\boldsymbol{x}).$$

$$\Rightarrow$$

$$\boldsymbol{H} \Delta \boldsymbol{x} = \boldsymbol{g}.$$

- G-N用J的表达式近似了H
- 步骤：

1. 给定初始值 $\boldsymbol{x}_0$。
2. 对于第 k 次迭代，求出当前的雅可比矩阵 $\boldsymbol{J}(\boldsymbol{x}_k)$ 和误差 $f(\boldsymbol{x}_k)$。
3. 求解增量方程：$\boldsymbol{H} \Delta \boldsymbol{x}_k = \boldsymbol{g}$.
4. 若 $\Delta \boldsymbol{x}_k$ 足够小，则停止。否则，令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \Delta \boldsymbol{x}_k$，返回 2.

4. 非线性最小二乘

在求解 ICP 问题中，我们已知损失函数如式 3.1 所示。因此，在进行优化求解时，已知误差函数：

$$err = (Rp_i + t) - q_i \tag{3.14}$$

此时要求得优化时的雅可比矩阵J，需对 R 和 t 分别求导：

$$\frac{\partial err}{\partial t} = I, \frac{\partial err}{\partial R} = -Rp_i^{\wedge} \tag{3.15}$$

其中I表示单位矩阵，$p_i^{\wedge}$表示点p_i的反对称矩阵；如式 2.13 所示，采用右扰动求导方式求得。

此时得到的雅可比矩阵J的形式如下：

$$J = \left[\frac{\partial err}{\partial t}, \frac{\partial err}{\partial R}\right] = [\mathrm{I}, -Rp_i^{\wedge}] \tag{3.16}$$

求得雅可比矩阵J后，按照式 3.13 的形式分别求得每一个点对所对应的矩阵 Hessian 和矩阵 B：

$$Hessian_i = J^T J\, , B_i = -J^T err \tag{3.17}$$

4. 非线性最小二乘

最终当遍历完所有的点对后，累加所有的矩阵 Hessian 和矩阵 B，最后计算增量Δx：

$$\Delta x = \left(\sum_{i=1}^{n} Hessian_i\right)^{-1} \cdot \left(\sum_{i=1}^{n} B_i\right) \tag{3.18}$$

此时的增量Δx为一个 6 维向量，前三维为平移向量 t，后三维为旋转向量。再使用式 2.7 的李代数的指数映射将旋转向量转换为旋转矩阵即得到了需要求得的旋转矩阵 R。

以上为一次迭代，若变换后不满足精度需求，则进行迭代直至满足匹配精度。