Newton - CG (see p 169)

Idea: ① find descent direction p_k by approx. solving Newton system: $\nabla^2 f_k p_k = -\nabla f_k$ with CG.

what tolerance?

require
$$\| \underbrace{\nabla^2 f_k p_k + \nabla f_k}_{r_k} \| \leq \overset{\text{forcing seq}}{\eta_k} \| \nabla f_k \| , \quad 1 < \eta_k < 0$$

what to do when CG breaks down?

This can potentially happen as $\nabla^2 f_k$ is not necessarily pos def so CG can encounter "negative curvature". directions d_j s.t. $d_j^T \nabla^2 f_k d_j < 0$ (we called them p_i in CG). $\Rightarrow$ keep last updated version of p_k.

② Use line search to determine step length.

Convergence result for inexact Newton methods

What happens when gradient or Hessian are not available exactly:

$$\nabla f_k + \delta_k \,, \qquad \delta_k \in \mathbb{R}^n = \text{error in gradient comp}$$
$$\nabla^2 f_k + \Delta_k \qquad \Delta_k \in \mathbb{R}^{n\times n} = \text{error in Hessian comp}$$

and we compute p_k as sol to:

$$(\nabla^2 f_k + \Delta_k)\, p_k = -(\nabla f_k + \delta_k)$$

Thm (Convergence of inexact Newton methods) (see thm 7.2 p168)

Assume that:

- x^* is a point where 2nd order suff opt. cdt° hold
- $\nabla^2 f(x)$ is Lipschitz cont close to x^*
- x_0 is sufficiently close to x^*.

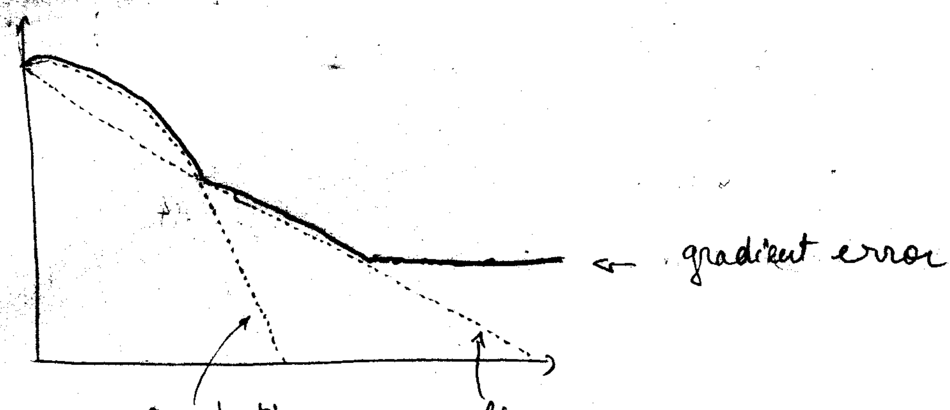
Then $\lim_{k\to\infty} x_k = x^*$ (if $\delta_k = 0$)

and

$$\| x_k - x^* \|_2 \leq C_1 \| x_k - x^* \|_2^2 + C_2 \eta_k \| x_k - x^* \| + C_3 \delta_k$$

$C_1, C_2, C_3 > 0$.

Typical convergence history:



We can use η_k (precision of Newton's system solves) to control convergence rate.

If $\eta_k \leq \| x_k - x^* \|$ ⇝ we would get quadratic convergence!

However $\| x_k - x^* \|$ is not known.

Use $\eta_k \leq c_k \| \nabla f_k \|^\alpha$

then

- $\| x_{k+1} - x^* \| \leq c \left(\| x_k - x^* \|^2 + \| x_k - x^* \|^{1+\alpha} \right)$

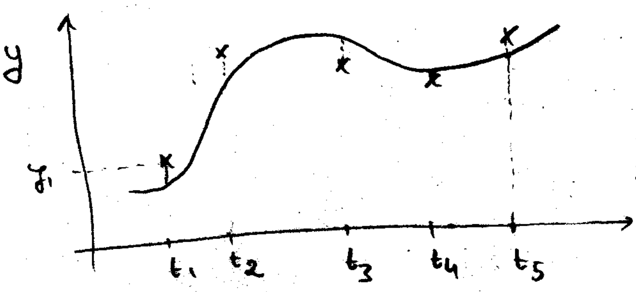
§10 Non linear Least Squares problems

let $R : \mathbb{R}^n \to \mathbb{R}^m$ be the "residual function".

Find $x^* \in \mathbb{R}^n$ solving $\min_x \frac{1}{2} \|R(x)\|_2^2$ (1)

$$\frac{1}{2}\|R(x)\|_2^2 = \frac{1}{2}\sum_{i=1}^{m} R_i(x)^2$$

Archetypical example: curve fitting



Given points (t_i, y_i) $i = 1, \ldots m$ find curve $\varphi(t; x_*)$ (model parameters) s.t. x_* solves (1) with $R_i(x) = \varphi(t_j; x) - y$

Example: line regression $\varphi(t; x) = x_1 + x_2 t$

model space : $\mathbb{R}^2$.

Note: a different norm can be used to measure misfit:

$$\min_x \|R(x)\|_1 = \sum_{i=1}^{m} |R_i(x)| \qquad (2)$$

$$\Leftrightarrow \quad \min_x \sum_{i=1}^{m} z_i \qquad (2^*)$$

s.t. $|R_i(x)| \leq z_i$, $i = 1..m$

$$\min_x \|R(x)\|_\infty = \max_{i=1..m} |R_i(x)| \qquad (3)$$

$$\Leftrightarrow \quad \min_x z$$

s.t. $|R_i(x)| \leq z$, $i = 1 \ldots m$ $\qquad (3^*)$

$(2*)$ & $(3*)$ are constrained optimization problems that we will learn how to solve.

$n = \#$ of model params
$m = \#$ of measurements

$m > n$ over determined
$m = n$ non linear system
$m < n$ under determined

$$f(x) = \frac{1}{2} \| R(x) \|_2^2$$

$$\nabla f(x) = DR(x)^T R(x) = \sum_{j=1}^{m} R_j(x) \nabla R_j(x)$$

where

$$DR(x) = J(x) = \begin{bmatrix} \nabla R_1(x)^T \\ \nabla R_2(x)^T \\ \vdots \\ \nabla R_m(x)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial R_1}{\partial x_1} & \frac{\partial R_1}{\partial x_2} & \cdots & \frac{\partial R_1}{\partial x_n} \\ \vdots & & & \\ \frac{\partial R_m}{\partial x_1} & \frac{\partial R_m}{\partial x_2} & \cdots & \frac{\partial R_m}{\partial x_n} \end{bmatrix}$$

= Jacobian ($m \times n$)

Jacobian is "best" linear $\mathbb{R}^n \to \mathbb{R}^m$ approx to $R(x)$. can be seen from Taylor's theorem.

$$R(x+p) = R(x) + DR(x)^T p + \underbrace{o(\|p\|)}$$

meaning this terms $\lim_{p \to 0} \frac{\|g(p;x)\|}{\|p\|} = 0$

thus: $DR(x) : \mathbb{R}^n \to \mathbb{R}^m = m \times n$

$$\nabla^2 f(x) = DR(x)^T DR(x) + \sum_{j=1}^{m} R_j(x) \nabla^2 R_j(x)$$

<u>Gauss-Newton method</u> for solving NL. Least Square Problem.

<u>Idea</u>: If x_c = current iterate:

instead of solving for update s:

$$\min_s \underbrace{\frac{1}{2} \| R(x_c + s) \|_2^2}_{f(x_c+s)}$$

We use Taylor's theorem to solve "nearby" problem

$$\min_s \frac{1}{2} \| R(x_c) + DR(x_c) s \|_2^2$$

This model agrees up to first order and gives a reasonable approx of second order terms provided $R(x*)$ is small:

$$\frac{1}{2}\|R(x_c)+DR(x_c)s\|_2^2 = \frac{1}{2}\|R(x_c)\|_2^2 + R(x_c)^T DR(x_c)s + \frac{1}{2}s^T DR_c^T DR_c s$$

$$= f(x_c) + \nabla f(x_c)^T s + \frac{1}{2}s^T\left(\nabla^2 f(x_c) - \sum_{i=1}^{m} R_i(x)\nabla^2 R_i(x)\right)$$

Gauss-Newton algorithm:

Choose x_0

for $k = 0, 1, \ldots$

$$s_k = -\left(DR_k^T DR_k\right)^{-1} DR_k^T R_k$$

$$x_{k+1} = x_k + s_k \qquad \text{(+ line search or other globalization)}$$

GN can be viewed as an inexact Newton method

⇝ can expect q-linear convergence

Here we have assumed that close enough to x^*,

$DR_k^T DR_k = {}_n\overset{n}{\square}$ is invertible

$\Leftrightarrow$ $\text{rank}(DR_k) = n$, full rank problem

When $\text{rank}(DR_k) \leq n$: rank deficient problem

means $\min \frac{1}{2}\|DR_k s + R_k\|_2^2$ has more than one sol.

⇝ pick solution with minimal norm:

$$s_k = -(DR_k)^{+} R_k$$

← Moore Penrose Pseudoinverse

computed using QR or SVD

see later for review of these methods

If we did not take min norm solution we could miss x_*!!

since: $\min \frac{1}{2} \| DR(x_*) s + R(x_*) \|_2^2$

$$\Leftrightarrow \quad DR(x_*)^T DR(x_*) s = - DR(x_*)^T R(x_*) = 0$$

maybe many solutions
but $s=0$ is the only valid
one here (min. norm solution).

Can be hard to tell if a matrix is full rank numerically.
-> make system we need to solve invertible: i.e. solve:

$$\left(DR(x)^T DR(x) + \underset{\text{identity}}{\mu I} \right) s = - DR(x)^T R(x)$$

shifts all eigenvalues so that they are all pos.
$(\mu > 0)$

Example: $A = \sum_{i=1}^{n} \lambda_i u_i u_i^*$ $\quad \lambda_i \geq 0$ (A pos semidef)

$$A + \mu I = \sum_{i=1}^{n} \underbrace{(\lambda_i + \mu)}_{>0} u_i u_i^* \Rightarrow A + \mu I \text{ is invertible}$$