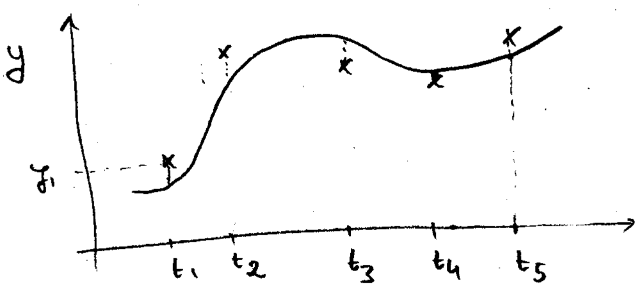
§10 Non linear Least Squares problems

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) $i = 1, .. m$

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:

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

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

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

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

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

$$\Leftrightarrow \quad \min_x z$$

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

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

$m \overset{n}{\square}$

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

$$R(x+p) = R(x) + DR(x)^T p + \underbrace{o(\|p\|)}_{\text{meaning this terms:}}$$

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

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

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

<u>Gauss-Newton method</u> for solving NL. Least Squares Problems.

<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 <u>provided</u> $R(x_*)$ is small:

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

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

<u>Gauss-Newton algorithm:</u>

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

$\rightsquigarrow$ can expect q-linear convergence

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

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

$\Leftrightarrow$ $\operatorname{rank}(DR_k) = n$, <u>full rank problem</u>

When $\operatorname{rank}(DR_k) \leq n$: <u>rank deficient problem</u>

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

$\rightsquigarrow$ pick solution with minimal norm:

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

← 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:

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

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

Example: $A = \sum_{i=1}^{m} \lambda_i u_i u_i^*$ $\qquad \lambda_i \geq 0$ (A pos semi def)

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

"Levenberg-Marquardt" method is this + choice of μ.

Solving Linear Least Squares problems

$$\min \frac{1}{2} \| Ax - b \|_2^2$$

First order necessary optimality cdts.

$$A^T (Ax - b) = 0$$

$$\Leftrightarrow \boxed{(A^T A) x = A^T b} \quad \text{"normal equations"}$$

Can be solved with:

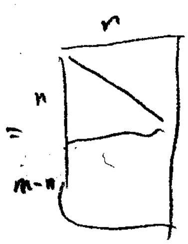
- *iterative method*: CG
- *direct method*: SVD or QR

Singular Value Decomposition:

$A \in \mathbb{R}^{m \times n}$ $\quad A = U \Sigma V^T,$

where $U \in \mathbb{R}^{m \times m}, \quad U^T U = I$

$V \in \mathbb{R}^{n \times n}, \quad V^T V = I$

$\Sigma \in \mathbb{R}^{m \times n} = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_{\min(n,m)}) =$ 

σ_i are *singular values* of A and

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(n,m)}$

$\sigma_1 = \|A\|_2$ (matrix norm induced by $\| \cdot \|_2$)

$$U = \begin{array}{cc} \overset{n}{\longleftrightarrow} & \overset{m-n}{\longleftrightarrow} \\ \end{array}$$
$$U = [U_1 \; U_2]$$

$$\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix} \begin{array}{l} \updownarrow n \\ \updownarrow m-n \end{array}$$

$$\|Ax - b\|^2 = \| U(\Sigma V^T x - U^T b) \|^2$$
$$= \| \Sigma V^T x - U^T b \|^2$$
$$= \left\| \begin{bmatrix} S \\ 0 \end{bmatrix} V^T x - \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} b \right\|^2$$
$$= \| S V^T x - U_1^T b \|^2 + \| U_2^T b \|^2$$

$$\Rightarrow \boxed{x^* = V S^{-1} U_1^T b} \quad \text{provided } S \text{ is invertible.}$$

We can write:

$$x^* = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i$$

where:
$U = [u_1 \cdots u_n]$
$V = [v_1 \cdots v_n]$

When S is singular: $\exists r \ \forall i > r \ \sigma_i = 0$

$r = \text{rank}(A)$

$r < \min(m,n) \Rightarrow$ rank deficient problem

any x of the form:

$$x = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i + \sum_{i=r+1}^{n} z_i v_i$$

solves optim problem.

$$\|x\|^2 = \sum_{i=1}^{r} \frac{(u_i^T b)^2}{\sigma_i^2} + \sum_{i=r+1}^{n} z_i^2$$

choose minimal norm one: $(z_i = 0)$.

$$x^* = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

$$= A^+ b$$

, $A^+ =$ Moore-Penrose pseudo inverse of A. (pinv in Matlab

If $A = U_r \Sigma_r V_r^T$ where $U_r \in \mathbb{R}^{m \times r}$, $U_r^T U_r = I$
$V_r \in \mathbb{R}^{n \times r}$, $V_r^T V_r = I$
$\Sigma_r \in \mathbb{R}^{r \times r}$ non singular

$$\Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

$$A^+ = V_r \Sigma_r^{-1} U_r$$

[we just keep singular values (and corresp singular vector) that are non zero]

Solving Linear Least Squares problem using QR facto:

Theorem: Let $A \in \mathbb{R}^{m\times n}$, $m \geq n$ (similar result holds for $m < n$)

Then there exists a permutation matrix $P \in \mathbb{R}^{n\times n}$

an orthonormal matrix $Q \in \mathbb{R}^{m\times m}$

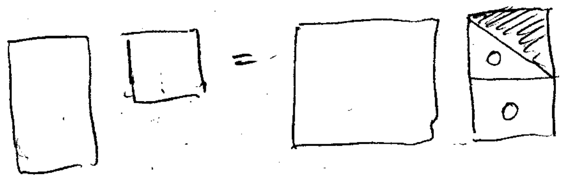
an upper triangular matrix $R \in \mathbb{R}^{m\times n}$ s.t.

$$AP = QR$$

$P^TP = PP^T = I$

$Q^TQ = QQ^T = I$

P has only one nonzero coeff per column



in general:

$$R = \begin{array}{c} \overset{r}{\longleftrightarrow}\overset{n-r}{\longleftrightarrow} \\ \begin{bmatrix} R_1 & R_2 \\ 0 & 0 \\ \multicolumn{2}{c}{0} \end{bmatrix} \end{array}$$

$R_1 \in \mathbb{R}^{r\times r}$, $r = \operatorname{rank}(A)$

$R_2 \in \mathbb{R}^{r\times n-r}$

First the full rank problem ($r = n$)

$$\|Ax - b\|^2 = \|APP^Tx - b\|^2 = \|QRP^Tx - QQ^Tb\|^2$$

$$= \|RP^Tx - Q^Tb\|^2$$

$$\Rightarrow x_* = P R^{-1} Q^T b$$

cheap inverse of ∇ matrix

Now the rank deficient problem ($r < n$)

For convenience, let

$$P^Tx = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} \quad \text{and} \quad Q^Tb = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix}$$

then:

$$\|Ax - b\|_2^2 = \|RP^T x - Q^T b\|_2^2$$

$$= \left\| \begin{bmatrix} R_1 & R_2 \\ \hline \multicolumn{2}{c}{0} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \right\|_2^2$$

$$= \|R_1 y_1 - (d_1 - R_2 y_2)\|_2^2 + \|d_2\|_2^2$$

if x solves least squares problem, we must have

$$P^T x = \begin{bmatrix} R_1^{-1}(d_1 - R_2 y_2) \\ y_2 \end{bmatrix}$$

letting $y_2 = 0$ we get the "basic solution"

$$x_B = P \begin{bmatrix} R_1^{-1} d_1 \\ 0 \end{bmatrix}$$ however $x_B \neq A^+ b$ (min norm sol)

need more work to get min norm sol if needed.

Both QR & SVD are similar in computing time (probably QR is cheaper)