Getting back to GN method, why do we take steps

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

when DR_k is not full rank?

First order necessary optimality cd^{ns} states:

x^* minimizer of $\frac{1}{2}\|R(x)\|^2 = f(x) \Rightarrow \nabla f(x^*) = DR^T(x^*) R(x^*) =$

Thus if our current iterate happens to be x^*,

$$s_* = -DR_*^{+} R_* = -(DR_*^T DR_*)^{+} \underbrace{DR_*^T R_*}_{=0}$$

$$= 0$$

so we will not move from optimum if we find it.
(regardless of having small / large residual R_*)

Yet another way of seen this:

s_* is the minimal norm sol to $\min_s \frac{1}{2}\| R_* + DR_* s\|_2^2$

$$\Rightarrow \quad DR_*^T DR_* s = -DR_*^T R_* = 0$$

And the smallest such s is clearly $s = 0$.

<u>Large residual problems</u>

Gauss Newton relies on approximating the Hessian by parts involving only first derivative information:

if $f(x) = \frac{1}{2}\|R(x)\|^2$ then $\nabla^2 f(x) = DR^T(x) DR(x) + \sum_{i=1}^{m} R_i(x) \nabla^2 R_i(x)$.

$$\approx DR^T(x) DR(x),$$

This approximation is not very good if $\|R(x_*)\|$ is large.

~ there are secant type methods (~ BFGS) that capture this missing second order information. (pp 262-265)

§11 Non linear Equations

$F: \mathbb{R}^n \to \mathbb{R}^n$, want to find $x*$ s.t.

$F(x_*) = 0$.

x_* is a root or zero of F.

Newton's method:

$F(x_k+s) \approx m(x_k+s) = F(x_k) + DF(x_k)s$ (lin model)

where $DF(x_k) =$ Jacobian of F at x_k

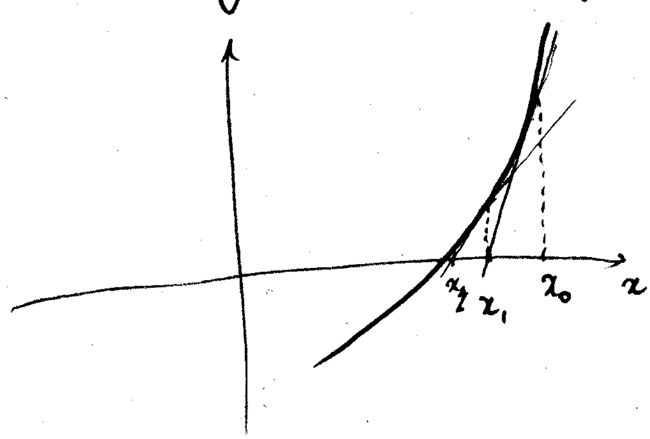
$$= \begin{bmatrix} \nabla F_1(x_k)^T \\ \nabla F_2(x_k)^T \\ \vdots \\ \nabla F_n(x_k)^T \end{bmatrix}$$

compute s for which $m(x_k+s) = 0$ (we want a root!)

$\Rightarrow DF_k s = -F_k$ (2)

Newton's method for NL eq is : $x_{k+1} = x_k + s_k$, where s_k solves (2

Note that:
$T(x) = F(x_k) + F'(x_k)(x - x_k)$ is tangent to $F(x)$ at x_k,
so by solving (2) we are finding root of tangent. For 1D function



Connection with unconstrained optimization

necessary optimality condition:

solving $\underbrace{Df(x)}_{F(x)} = 0$ is solving a NL sys. of eq.

$DF(x) = \nabla^2 f(x)$ (guaranteed to be symmetric, but in general $DF(x)$ needs not be symm)

Non linear eq as an optimization problem

$$F(x)=0 \iff \text{find } x \text{ that solves } \min_x \tfrac{1}{2}\|F(x)\|_2^2$$

If we solve this problem using Gauss-Newton we find step as solution to:

$\min_s \frac{1}{2}\| DF_k s + F_k \|_2^2$ then assuming $DF_k \in \mathbb{R}^{n\times n}$, non singular

$$\Rightarrow \quad s_k = -(DF_k^T DF_k)^{-1} DF_k^T F_k$$

$= -DF_k^{-1} F_k$ which is equivalent to what we derived before.

Convergence result for Newton's method for non-linear equations:

Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be continuously diff'ble with $DF(x)$ Lipschitz continuous.
Let x_* be a point at which $DF(x_*)$ is invertible.
Then for x_0 close enough to x_*,

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2 \quad \text{(q-quadratic convergence)}$$

__proof__: Banach fixed pt theorem + Banach Lemma

It is possible to come up with Quasi Newton methods that do not require the computation of $DF(x)$. Their derivation is similar to ~~that~~ of DFP, BFGS, and the most popular of such methods is "Broyden's method" (P 279-283) which can achieve q-superlinear convergence.

Broyden's method:

choose initial point x_0
choose initial approx B_0 to $DR(x_0)$ (usually I)
B_0 non-singular.

for $k = 0, 1, \dots$

Solve $B_k p_k = -R(x_k)$
choose step size t_k (line search ...)
$x_{k+1} = x_k + t_k p_k$
$s_k = x_{k+1} - x_k$
$y_k = R(x_{k+1}) - R(x_k)$
$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}$$

can also do update on approx to $DR(x_k)^{-1}$.
(as BFGS does)

Update is s.t. it satisfies secant equation:

$$y_k = B_{k+1} s_k$$

and

B_{k+1} solves $\min_{B} \|B - B_k\|$ s.t. $y_k = B s_k$

proof:

$$\|B_{k+1} - B_k\| = \left\| \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k} \right\|$$
$$= \left\| \frac{(B - B_k) s_k s_k^T}{s_k^T s_k} \right\| \leq \|B - B_k\| \left\| \frac{s_k s_k^T}{s_k^T s_k} \right\| = \|B - B_k\|$$

SVD gives 1 as largest singular value