Newton-CG (see p169)

Idea: find descent direction of by appear.
     solving Newton system: $D^2 f_k p_k = -Df_k$ with CG.
     what tolerance?
     require
     \[
     \frac{\|D^2 f_k p_k + Df_k\|}{\|Df_k\|} \leq \eta_k \|Df_k\|, \quad 0 < \eta_k < 1
     \]

what to do when CG breaks down?

This can potentially happen as $D^2 f_k$ is not necessarily
     pos def so CG can encounter negative curvature.
     directions $d_j$ s.t.
     \[
     d_j^T D^2 f_k d_j < 0
     \]
     (we called these $p_j$ in CG). => keep last updated version of $p_k$

2. Use line search to determine step length.

Convergence result for inexact Newton method

What happens when gradient or Hessian are not
     available exactly:

\[
\nabla f_k + s_k, \quad s_k \in \mathbb{R}^n = \text{error in gradient comp}
\]
\[
D^2 f_k + \Delta k, \quad \Delta k \in \mathbb{R}^{n \times n} = \text{error in Hessian comp}
\]

and we compute $p_k$ as sol to

\[
(\nabla^2 f_k + \Delta k) p_k = -(\nabla f_k + s_k)
\]
Assume that:
- \( x^* \) is a point where 2nd order suff opt. cond. hold
- \( \nabla^2 f(x) \) is Lipschitz cont. close to \( x^* \)
- \( z_0 \) is sufficiently close to \( x^* \)

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

\[
\| x_k - x^* \|_2 \leq C_1 \| x_k - x^* \|_2^2 + C_2 \eta k \| x_k - x^* \| + C_3 s_k
\]

\( C_1, (C_2, C_3) > 0 \)

Typical convergence history:

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

If \( \eta_k \leq \| x_k - x^* \| \sim \eta k \) we would get quadratic convergence.

However, \( \| x_k - x^* \| \) is not known.

Use \( \eta_k \leq c_k \| \nabla f_k \| \eta \)

then

\[
\| x_{k+1} - x^* \| \leq c (\| x_k - x^* \|_2^2 + \| x_k - x^* \|_1 + \eta^d)
\]
min \| R(x) \|_{\infty} = \max_{i=1, \ldots, m} |R_i(x)|

(3)

\begin{align*}
\text{min} & \quad \sum_{i=1}^{m} |R_i(x)| \\
\text{s.t.} & \quad |x_1(2)| < 1 \quad \text{and} \quad |x_2(2)| < 1
\end{align*}

\begin{align*}
\text{(2)} & \quad \text{min} \quad \sum_{i=1}^{m} |R_i(x)| \\
\text{s.t.} & \quad |x_1(2)| < 1 \quad \text{and} \quad |x_2(2)| < 1
\end{align*}

\text{Note: a different norm can be used to measure misfit.}

\begin{align*}
\text{Example: line regression} \quad y(x) = x_1 + x_2 t \\
\text{model space: } R^2
\end{align*}

Find \( z \in \mathbb{R}^m \) s.t. \( \min \| R(x) \|_2^2 \)

\begin{align*}
\text{Typical example: curve fitting} \quad y(x) = \frac{1}{\| R(x) \|_2^2} \\
\text{model space: } R^2
\end{align*}

\begin{align*}
\text{Find } & \quad z \in \mathbb{R}^m \quad \text{to the } \quad \text{"virtual function"} \\
\text{given point} \quad & \quad (x_1(2), x_2(2)) \cdot \mathbf{z} \\
\text{with function} \quad & \quad y(x) = x_1 + x_2 t \\
\text{and point} \quad & \quad (x_1(1), x_2(1)) \cdot \mathbf{z} \\
\text{subject to} \quad & \quad |x_1(2)| < 1 \quad \text{and} \quad |x_2(2)| < 1
\end{align*}
\( n = \# \text{ of model parameters} \) \hspace{1cm} m > n \text{ over determined} \\
\( m = n \) \text{ non linear system} \\
\( m < n \) \text{ under determined} \\

\[
f(x) = \frac{1}{2} \| R(x) \|^2
\]

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

where

\[
DR(x) = J(x) = \begin{bmatrix}
\nabla R_1(x)^T \\
\nabla R_2(x)^T \\
\vdots \\
\nabla R_m(x)^T
\end{bmatrix}
\]

Jacobian as "best" linear \( R^n \to R^m \) approx to \( R(x) \),
can be seen from Taylor's theorem.

\[
R(x+p) = R(x) + DR(x)^T p + o(\| p \|)
\]

meaning this terms:

\[
\lim_{p \to 0} \frac{\| R(x+p) \|}{\| p \|} = 0
\]

thus:

\[
DR(x) : R^n \to R^m = m \times n
\]

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

**Gauss-Newton method** for solving **NL Least Square Problem**

**Idea:** If \( x_c \) = current iterate:

instead of solving for updates:

\[
\min_{s} \frac{1}{2} \| R(x_c+s) \|^2
\]

\[
\min_{s} \frac{1}{2} \| f(x_c+s) + DR(x_c)s \|^2
\]
This model agrees up to first order and gives a reasonable approx of second order terms provided $R(x^*)$ is small:

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

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

\textit{Gauss-Newton algorithm:}

Choose $x_0$

for $k = 0, 1, \ldots$:

$$x_k = -(D R_k^T D R_k)^{-1} D R_k^T R_k$$

$$x_{k+1} = x_k + s_k$$

(+ line search or other globalization)

GN can be viewed as an inexact Newton method

$\Rightarrow$ can expect $q$-linear convergence.

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

$$D R_k^T D R_k = n \begin{bmatrix} I \\ 0 \end{bmatrix}$$

is invertible.

$\Rightarrow$ Rank $(D R_k) = n$, full rank problem.

When rank $(D R_k) \leq n$ : rank deficient problem

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

$\Rightarrow$ pick solution with minimal norm:

$$s_k = -(D R_k)^+ R_k$$

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: \[ \min \frac{1}{2} \| DR(x^*) b + R(x^*) \|_2^2 \]

\[ \Rightarrow \quad DR(x^*)^T DR(x^*) \lambda = - DR(x^*)^T R(x^*) = 0 \]

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

Can be hard to tell if a matrix is full rank numerically.

\[ (DR(x^*)^T DR(x^*) + \mu I) \lambda = - DR(x^*)^T R(x) \]

shifts all eigenvalues so that they are all pos.

\[ (\mu > 0) \]

Example: \[ A = \sum_{i=1}^{\infty} \lambda_i \xi_i \xi_i^* \]

\[ \lambda_i \geq 0 \quad (A \text{ pos. semidef}) \]

\[ A + \mu I = \sum_{i=1}^{\infty} \left( \lambda_i + \mu \right) \xi_i \xi_i^* \]

\[ \geq 0 \quad \Rightarrow \quad A + \mu I \text{ is invertible} \]