Trust region methods

- previously:
  - use a hessian to choose length of step

- now:
  - use a local model of fun to choose dir, size of step within trust region

Local model:

\[ f(x) = f(x_k) + p^\top \nabla f_k + \frac{1}{2} p^\top B_k p \]

this could be Hessian or other matrix
Trust region:
\[ p'p < d^2 \]

Issues
- how to choose \( p \)?
- how to choose \( d \)?

Choosing \( d \):
write \( m_k(x_p) \) for model
consider

\[
C_k = \frac{f(x_k) - f(x_k + p_k)}{M_k(o) - M_k(p_k)}
\]

\[
= \frac{\text{actual reduction}}{\text{predicted reduction}} > 0
\]
Cases:

$P_k$ close to 1:
$\rightarrow$ model is good, trust region can be expanded

$P_k$ smallish
$\rightarrow$ do not alter trust region

$P_k$ 0, negative (step rejected)
$\rightarrow$ reduce trust region

Notice: no need to increase trust region if model is good and step is valid.
Solving the subproblem:

$$\min_p \ m(p) = f + p^T g + \frac{1}{2} p^T B p$$
subject to \quad \frac{p^T p}{d^2} \leq 1$$

\underline{global soln if}:

$$(B + \lambda I) p^* = -g$$

$$\lambda (d^2 - p^* p) = 0$$

$$\lambda (d^2 - p^* p) \geq 0$$

\[\text{positive semi-def}\]

(significant optimization problem on its own)
But approximation may be OK.

Cauchy point.

1) solve $p^s_k = \arg \min_p (f + p^l) \rightarrow $ linear w.r.t. $m_k$

2) compute $\gamma_k > 0$ to minimize $M_k(\gamma_k p_k)$ such that $\|\gamma_k p_k\| \leq d_k$

$p^c_k = \gamma_k p^s_k$
Fact: a TV method will be globally convergent if steps give a reduction in $m_k$ that is at least a fixed positive multiple of Cauchy.

Why not just use CP?

- It's gradient descent w/ a special step.

The Dogleg method:

- Notice that if $B_k$ is pd and

$$
\|B_k^{-1}g\| \leq d
$$

- $B_k^{-1}g$ is a step (the full step).
Now for small $p$ ($\|p\| < d$) the quadratic term isn't important.

\[ p^*(d) \approx -d \frac{g}{\|g\|} \]  
(for small $d$)

\[ p = -B^{-1}g \]

which is min of $m$ along descent dir

\[ p = -t \frac{g}{\|g\|} \]

\[ = -\frac{g^T g}{g^T B g} \cdot g \]
The dog-leg path is

$$\hat{p}(t) = \begin{cases} t \hat{p}^u & 0 \leq t \leq 1 \\ \hat{p}^u + (t-1)(\hat{p}^b - \hat{p}^u) & 1 \leq t \leq 2 \end{cases}$$

Notice:

1) $\|\hat{p}\|$ is an increasing fn of $u$

2) $m(\hat{p}(t))$ is an increasing fn of $u$

So: if $\|\hat{p}^b\| > d$, path intersects $tr$ at one point

$\Rightarrow$ 2 possibilities

1) $\hat{p}^b$

2) int of $\hat{p}$ and $tr$. 
strategy is most useful for convex functions

2D subspace minimization:

- consider subspace spanned by \( p \) and \( p^b \)

- confine \( m \) to this subspace

locate here

\[ = \text{root finding for degree 4 polynomial} \]
Now if 

\[ \begin{align*} \text{note: span} \left[ \rho^T \rho B \right] & = \text{span} \left[ g, -B' g \right] \\
\text{est} \lambda_{k+1} & = \text{smallest eval of } B \\
\text{use } -\lambda & < \alpha < -2\lambda, \\
\text{subspace is} \quad \text{span} \left[ g, (B + \alpha I)^{-1} g \right] \\
\text{if } \| (B + \alpha I)^{-1} g \| & \leq d \\
\text{adjust step so that} \quad \| p \| > \| (B + \alpha I)^{-1} g \| \\
\text{no neg evals } \rightarrow \text{use Cauchy point, } \\
\text{This is convergent.} \end{align*} \]
General notes

- One can work with

\[ p^T D p \leq d^2 \]

\[ L \text{ scaling matrix} \text{ elliptical TR.} \]

- One can work with

\[ 1 \text{-norm, } \infty \text{-norm trust region} \]