Trust region methods

- previously:
  - use a dirn to choose length of step
- Now:
  - use a local model of fn to choose dir, size of step within trust region

Local model:

$$f(x) = f(x_k) + p'\nabla f_k + \frac{1}{2} p' B_k p$$

this could be Hessian or other matrix

Trust region:

$$p'p \leq d^2$$

Issues

- how to choose p?
- how to choose d?

Choosing d:

write $m_k(p)$ for model

consider

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

$$= \frac{\text{actual reduction}}{\text{predicted reduction}}$$

$$> 0$$

Cases:

ρ_k close to 1:

→ model is good,

trust region can be expanded

ρ_k smallish

→ do not alter tr

ρ_k 0, negative (step rejected)

→ reduce trust region

Notice: no need to increase trust region if model is good and step is inside

Solving the subproblem:

$$\min_{p} \; m(p) = f + p'g + \frac{1}{2} p' B p$$

subject to $\quad p'p \leq d^2$

global soln if:

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

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

$$(B + \lambda I) \succeq 0$$

positive semi-de

(significant optimization problem on its own)

<u>But</u> approximation may be OK.

<u>Cauchy point.</u>

1) solve $p_k^s = \arg\min_p \left(f + p' g \right.$

linear version of m_k

2) compute $\tau_k > 0$ to minimize

$$m_k(\tau_k p_k^s)$$

such that

$$\|\tau_k p_k^s\| \leq d_k$$

$$p_k^c = \tau_k p_k^s$$

Fact: a tr method will be globally convergent if steps give a reduction in m_k that is at least a fixed positive multiple of Cauchy point

Why not just use CP?

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

The Dogleg method:

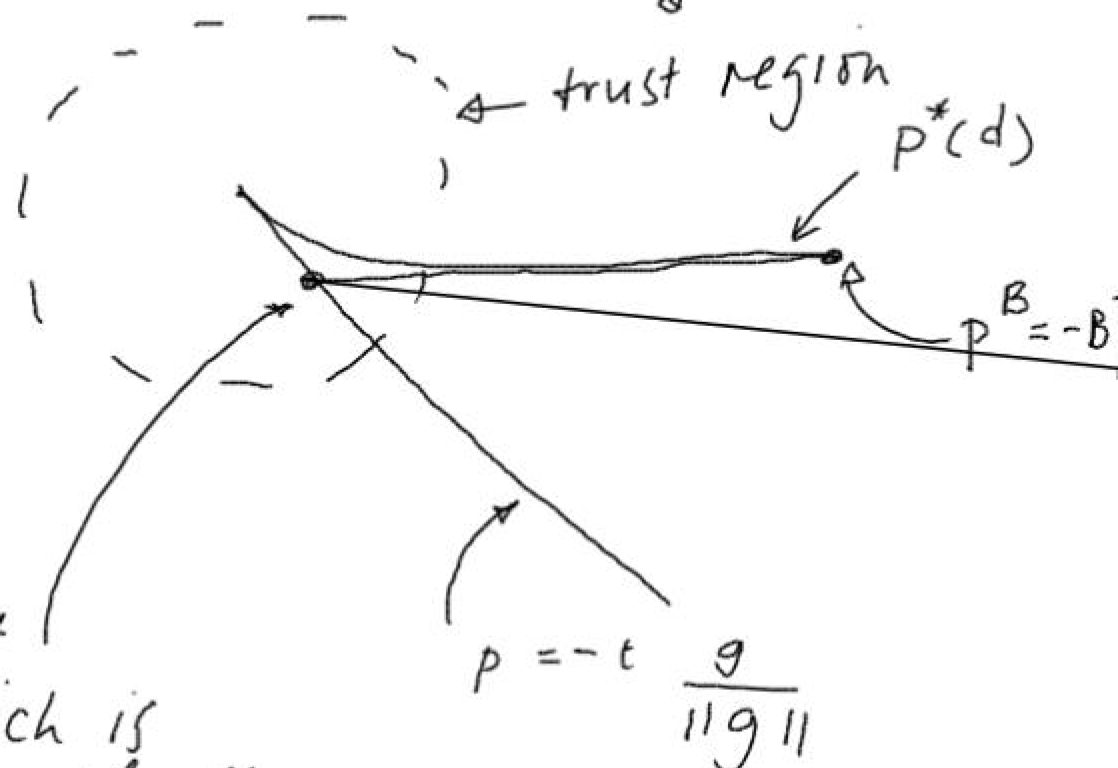
- notice that if B_k is pd and

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

$-B_k^{-1} g$ is a soln (the full step).

now for small p ($\|p\| < d$)
the quadratic term isn't important

$\rightarrow \quad p^*(d) \approx -d \dfrac{g}{\|g\|}$ (for small d



p^u
which is
min of m
along descent
dir

$$= -\frac{g^T g}{g^T B g} \cdot g$$

dog-leg path is

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

Notice:

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

2) $m(\hat{p}(\tau))$ is an ~~in~~ de-creasing fn of τ

So: if $\|p^B\| > d$, path intersects tr at one point

$\Rightarrow$ 2 possibilities

1) p^B

2) int of $\hat{p}$ and TR.

strategy is most useful for convex functions

<u>2D subspace minimization</u>:

- consider subspace spanned by ~~pu~~ p^u and p^B
- confine m to this subspace
- Solve here

  $\equiv$ root finding for degree 4 polynomial

Now if B is <u>not</u> p.d.

note span $[p^u \; p^B]$

$= \text{Span}[g, -B^{-1}g]$

→ est λ_1 = smallest eval of B

→ use $-\lambda_1 < \alpha < -2\lambda_1$

→ subspace is

span $[g, (B+\alpha I)^{-1}g]$

→ if $\|(B+\alpha I)^{-1}g\| \leq d$

adjust step so that

$\|p\| \geq \|(B+\alpha I)^{-1}g\|$

→ no neg evals → use Cauchy point

. This is convergent.

General notes

- one can work with

$$p^T D p \leq d^2$$

scaling matrix gives elliptical T.R.

- one can work with 1-norm, ∞ norm trust region