Numerical Optimisation:

Trust region methods

f.rullan@cs.uc

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Problems
University College London

Lecture 4

- Choose a region around the current iterate $f(x_k)$ in which we trust a model.
- Choose a relatively easy solvable model which we trust is an adequate representation of f in this region.

-
-

  the region is too large, the minimiser of the mo
  away from the minimiser of f.

- If the model is consistently reliable, the trust r
  increased.
- If the step length is not acceptable, reduce the size of the trust region and find a new minimiser. In general both the direction and step length change when the trust region changes.

Trustregion: model

Here we assume a quadratic model based on Taylor expansion of f at x_k

$$m_k(p) = f(x_k) + g_k^T p + \frac{1}{2} p^T B_k p,$$

wher

Hess

In gen

B_k. The difference between $\nabla^2 f(x_k +$) is $\mathcal{O}(\|p\|^2)$.

The choice of $B_k = \nabla^2 f(x_k)$ leads to

methods and the model accuracy is $\mathcal{O}(\|p\|^3)$.

In each step we solve

$$\min_{p \in \mathbb{R}^n} m_k(p) = f(x_k) + g_k^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} B_k p, \quad \text{s.t. } \|p\| \leq \Delta_k, \quad \text{(CM)}$$

and $\Delta_k > 0$ is the radius of the trust region. The constraint can be equivalently written $p^{\mathrm{T}} p \quad \Delta^2$.

If B_k

quad
$\|p^B\| = \|B_k^{-1} g_k\| \leq \Delta_k$ this is also the solution to the constrained problem and we call p^B a full step.

Solution in other cases is less straight forward but c
obtained at moderate computational cost. In particular, only approximate solution is necessary to obtain convergence and good practical behaviour.

Trust region vs line search

Figure: Nocedal Wright Fig 4.1

Compare the actual reduction in objective function to the predicted reduction i.e. reduction in the model m_k.

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

-
- $\rho_k > 0$, small – accept step, shrink trust regi iteration

- $\rho_k > 0$, but significantly smaller than 1 – ac do not alter trust region
- $\rho \approx 1$: good agreement between f and m_k – accept step and expand trust region for next iteration.

Algorithm: Trust region

Given $\hat{\Delta} > 0, \Delta_0 \in (0, \hat{\Delta})$ and $\eta \in [0, \frac{1}{4})$
for $k = 1, 2, 3, \ldots$ **do**
  Obtain p_k by (approximatively) solving (CM)
  Evaluate $\rho_k = \frac{f(x_k) - f(x_k + p)}{m_k(0) - m_k(p)}$
  if $\rho_k < \frac{1}{4}$ **then**
    $\Delta_{k+1} = \frac{1}{4}\Delta_k$

    else
      $\Delta_{k+1} = \Delta_k$
    end if
  end if
  if $\rho_k > \eta$ **then**
    $x_{k+1} = x_k + p_k$
  else
    $x_{k+1} = x_k$
  end if
end for

$p^\star$ is a global solution of the trust region problem (CM)

$$\min_{p \in \mathbb{R}^n} m_k(p) = f(x_k) + g_k^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} B_k p, \quad \text{s.t. } \|p\| \le \Delta_k$$

if and o ... at
the fol

$$(B + \lambda I) p^\star = -g_k, \tag{1a}$$

$$\lambda(\Delta - \|p^\star\|) = 0, \tag{1b}$$

$$B + \lambda I \ \text{ is positive se} \quad ($$

Solution of (CM) for different radii

Figure: Nocedal Wright Fig 4.2 (note that $p_3^\star$ and $p_1^\star$ should be swapped)

For Δ_1, $\|p^\star\| < \Delta$ hence $\lambda = 0$ and so

$$Bp^\star = -g_k$$

with

For Δ

trust region, hence $\|p^\star\| = \Delta$ and $\lambda \geq 0$. From (1)(a) we have

$$\lambda p^\star = -Bp^\star - g_k =$$

Thus if $\lambda > 0$, $p^\star$ is collinear with the negative gradient of m_k and normal to its contours.

Cauchy point

Cauchy point p^C is the minimiser of m_k along the steepest descent direction $-g_k$ subject to the trust region bound.

Find p^s:

$$p^s = \arg\min_{n} f(x_k) + g_k^{\mathrm{T}} p, \quad \text{s.t. } \|p\| \le \Delta_k$$

Calc

$$\tau \geq 0$$

Set $p^C = \tau_k p^s$

The solution to the first problem can be written down explicitly, simply by going as far as allowed in the steepest descent direction

$$p^s = -\frac{\Delta_k}{\|g\|} g.$$

To obtain τ_k we substitute $p^s = -\frac{\Delta_k}{\|g_k\|} g_k$ into the second problem we obtain

$$\arg\min_{\tau} m_k(\tau p^s) = f(x_k) - \tau \underbrace{\frac{\Delta_k}{\|g_k\|} g_k^{\mathrm{T}} g_k}_{>0} + \frac{1}{2}\tau^2 \frac{\Delta_k^2}{\|g_k\|^2} g_k^{\mathrm{T}} B_k g_k$$

subject to $\tau g_k \frac{\Delta_k}{g}$ Δ_k τ $[\ 1, 1]$.

We c

- $g_k \neq 0$. Hence, the minimum is attained for largest $\tau \in [-1, 1]$ i.e. $\tau = 1$.
- $g_k^{\mathrm{T}} B_k g_k > 0$: $m_k(\tau p^s)$ is strictly conv τ, thus the minimum is either the unconstrai whenever in $[-1, 1]$ or otherwise 1 ($\arg\min_{\tau=\{-1,1\}} m_k(\tau p^s)$)

$$\tau_k = \begin{cases} 1 & g_k^{\mathrm{T}} B_k g_k \le 0 \\ \min\left(\|g_k\|^3/(\Delta_k g_k^{\mathrm{T}} B_k g_k), 1\right) & g_k^{\mathrm{T}} B_k g_k > 0. \end{cases}$$

Cauchy point for positive definite B_k

Sufficient reduction in the model is reduction of at least a positive fraction of that achieved by the Cauchy point p^C.

Figure: Nocedal Wright Fig 4.3

- Cauchy points p^C provides sufficient reduction to yield global convergence.
- Cauchy points is cheap to compute.
- Cauchy point essentially corresponds to the steepest descent method with a particular choice of step length. Steepest

- te

  when B is used to compute both the desce

  the step length.

- A number of trust region methods compute and then attempt to improve on it. Often, the full step i.e. $p^B = -B^{-1} g_k$ is chosen whenever B is positive definite and $\|p^B\| \leq \Delta_k$. When $B = \nabla^2 f(x_k)$ or a quasi-Newton approximation, this strategy can be expected to yield superlinear convergence.

The dogleg method

Assumption: B positive definite.

If $p^B = -B^{-1} g_k$ with $\|p^B\| \le \Delta_k$ it is just the unconstrained minimum

$$p^\star = p^B, \quad \|p^B\| \le \Delta_k$$

On th

$\|p^B\| \le \Delta$ ensures that the quadratic term in m_k has little effect on the solution of (CM) and it could be omitted i.e.

$$p^\star \approx -\frac{\Delta_k}{\|g_k\|} g_k, \quad \Delta$$

For intermediate values of Δ_k, the solution $p^\star(\Delta_k)$ typically follows a curved trajectory (Fig. 4.4 Nocedal, Wright).

Figure: Nocedal Wright Fig 4.4

The dogleg method replaces the curved trajectory with path consisting of two line segments.

The first line segment runs from the origin to the minimiser of m_k along the steepest descent direction

$${}^{U} \qquad \frac{g_k^{\mathrm{T}} g_k}{}$$

The second line segment runs from p^U B minimum or full step).

Formally, the trajectory can be written as

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1 \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases}$$

The dogleg method chooses p to minimise the model m_k along this path subject the trust region bound.

The minimum along the dogleg can be found easily because
(i) $\|\tilde{p}(\tau)\|$ is an increasing function function of τ
(ii) $m(\tilde{p}(\tau))$ is a decreasing function of τ

Pro $\in [1, 2]$
can be s
nonn

Intuition:
(i) The length of $\tilde{p}$ could only decrease wit
at $\tau = 1$ i.e. the vector $p^B - p^U$ make 2
with p^U which is not possible for the steepest descent solution.
(ii) $m(\tilde{p}(2))$ is the minimum of a strictly convex function, hence $m(\tilde{p}(1)) > m(\tilde{p}(2))$ and the function decreases for $\tau \in [1, 2]$.

As a consequence the path $\tilde{p}(\tau)$ intersects the trust region boundary at exactly one point if $\|p^B\| \geq \Delta$ and the intersection point can be computed solving the quadratic equation

$$\|p^U + (\tau - 1)(p^B - p^U)\|^2 = \Delta^2.$$

In cas
defin

Newton dogleg method. If $\nabla^2 f(x_k)$ is no
could use one of the modified Hessians and close to th
we will recover the Newton step. However, the som
perturbation introduced by the modification ca
benefits of the trust region methods. In fact, the trust region introduces its own modification (1)(a,c) thus the dogleg method is most appropriate when B is positive definite.

An extension of the dogleg method

$$\min_p m_k(p) = f(x_k) + g_k^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} B p \quad \text{s.t. } p \in \text{span}[g, B^{-1} g].$$

The o
as $\tilde{p}$ m_k
is ofte
on the subspace). The subspace span[
looking for a minimiser for a quadratic model (Tay

This subspace minimisation strategy can be modified for indefinite B.

The Cauchy point p^C satisfies the sufficient reduction condition

$$m_k(0) - m_k(p) \geq c_1 \|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right) \qquad \text{(SR)}$$

with $c_1 = \frac{1}{2}$.

Pro ... the

ineq ...

If a vect ... $\| \ \| \leq \ _k$

$$m_k(0) - m_k(p) \geq c_2(m_k$$

then it satisfies (SR) with $c_1 = c_2/2$

$m_k(0) - m_k(p) \geq c_2(m_k(0) - m_k(p^C)) \geq \frac{1}{2} c_2 \|g_k\| \min \ \Delta_k, \frac{\ _k\|}{\|B_k\|}\Big).$

In particular, if p is the exact solution $p^\star$ of (CM), then it satisfies (SR) with $c_1 = \frac{1}{2}$. Note that both the dogleg and 2d-subspace minimisation algorithms satisfy (SR) with $c_1 = \frac{1}{2}$ because the both produce approximate solutions p for which $m_k(p) \leq m_k(p^C)$.

Global convergence

Let $\|B_k\| \leq \beta$ for some constant $\beta > 0$ and f be bounded below on the level set $S = \{x \ : \ f(x) \leq f(x_0)\}$ and Lipschitz continuously differentiable in the neighbourhood of S, $\mathcal{N}(S, R_0)$, $R_0 > 0$ and all the approximate solutions p_k of (CM) satisfy the inequalities (SR) for so t regio

-

$$\liminf_{k\to\infty} \|g_k$$

- $\eta \in (0, \frac{1}{4})$ in Algorithm: Trust region

$$\lim_{k\to\infty} g_k = 0.$$

Let f be twice Lipschitz continuously differentiable in the neighbourhood of a point $x^\star$ at which the second-order sufficient conditions are satisfied. Suppose that the sequence $\{x_k\}$ converges to $x^\star$ and that for all k sufficiently large, the trust region algorithm base tisfy the Ca asym $\frac{1}{2}\Delta_k$ i.e.

$$\|p_k - p_k^N\| = o($$

Then the trust region bound Δ_k beco sufficiently large and the sequence $\{x_k\}$ converges superlinearly to $x^\star$.