Week 6, Lecture 1

22 February 2024 14:16

In this lecture we begin to look at **Trust-Region Methods**, taken from Chapters 7-8 of the typed notes

Rationale: LS methods — at each iteration, the OP is reduced to a 1D subproblem:

$$\alpha_k = \arg\min_{\alpha > 0} f(\underline{x}_k + \alpha p_k).$$

LS methods work well when the Hessian is positive-definite everywhere. Trust-region methods are more robust — they can be "tweaked" so that they converge even when the Hessian is not always positive-definite.

The idea (§ 7.2)

$$f(\underline{x}) = f(\underline{x}_k + p) \quad \dots \quad \underline{x} = \underline{x}_k + p$$

We approximate $f(\underline{x}_k + p)$ by a quadratic function:

$$f(\underline{x}_k + p) \simeq f_k + \langle g, p \rangle + \frac{1}{2}\langle p, Bp \rangle = m_k(p)$$

i.e. we are "approximating $f(\underline{x})$ by the model problem at each iteration".

Eq. (1)

Here, g can be the gradient ($g = \nabla f_k$) and B can be the Hessian $\left(B_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_{\underline{x}_k}\right)$

but this doesn't have to be the case.

Equation (1) is the quadratic approximation of the cost function. We introduce a trust region where $m_k(\underline{p})$ is a good approximation to $f(\underline{x}_k + \underline{p})$:

$$\|\underline{p}\|_2 \leq \Delta \qquad (2)$$

Here, Δ is the size of the trust region.

Once we have established the size of the trust region, we can solve the model problem for $\underline{p}$:

$$\underline{p}_k = \underset{\|\underline{p}\|_2 \leq \Delta}{\arg\min}\; m_k(\underline{p}) . \qquad (3)$$

Remark: Eqn (3) is a constrained minimization.

Size of trust region (§ 7.3)

- $f(\underline{x}_k) - f(\underline{x}_k + \underline{p}_k)$ Want this to be positive
- $m_k(0) - m_k(\underline{p}_k)$ GUARANTEED POSITIVE

Form the ratio:

$$\rho_k = \frac{f(\underline{x}_k) - f(\underline{x}_k + \underline{p}_k)}{m_k(0) - m_k(\underline{p}_k)} \qquad \frac{\text{Want this to be + ve}}{\text{GUARANTEED + ve}}$$

Now :

- If $\rho_k > 0$ then this is good, we are reducing the cost function at each iteration. Furthermore, if ρ_k is close to one, there is good agreement between the actual decrease in the cost function (numerator) and the model decrease (denominator), and we can expand the trust region at the next iteration.
- If $\rho_k > 0$ but significantly smaller than one, then we leave the trust region unchanged at the next iteration.
- If $\rho_k < 0$ or positive but much smaller than one we reduce the size of the trust region at the next iteration.

$\boxed{\text{ALGORITHM \#5}}$

Algorithm 5 Determining Size of Trust Region

Choose a maximum size of the trust region, $\widehat{\Delta}$ and an initial guess for the size of the trust region, Δ_0. Also, choose a criterion $\eta \in [0, 1/4)$ for a descent direction to be accepted.

for $k = 0, 1, 2, \cdots$ **do**

  Obtain p_k by (approximately) solving Equation (7.3). $\longrightarrow$ $p_k = \underset{\|p\|_2 \le \Delta}{\arg\min}\; m_k(p)$

  Evaluate ρ_k from Equation (7.4). $\longrightarrow$ $\rho_k = \dfrac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$

I { **if** some condition is true **then** e.g. $\rho_k < \frac{1}{4}$

    $\Delta_{k+1} = (1/4)\Delta_k$

  else

    if $\rho_k > 3/4$ and $\|p_k\|_2 = \Delta_k$ **then**

      $\Delta_{k+1} = \min(2\Delta_k, \widehat{\Delta})$

    else

      $\Delta_{k+1} = \Delta_k$;

    end if

  end if }

II { **if** $\rho_k > \eta$ **then**

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

  else

    $x_{k+1} = x_k$

  end if }

end for

I - expand / contract trust-region bdy

II - accept / reject proposed update

<u>The constrained model problem</u> (§ 7.4)

$$\left.\begin{aligned} m_k(p) &= f_k + \langle g, p\rangle + \tfrac{1}{2}\langle p, Bp\rangle \\ p_* &= \arg\min\; m_k(p) \\ \text{subject to}&: \|p\|_2 \le \Delta . \end{aligned}\right\} (4)$$

Theorem 7.1 : Let B be a symmetric matrix.

Then, p_* is a global solution of the trust-region model problem (Eq. (4)),

$$p_* = \arg\min_{\|p\|_2 \leq \Delta} m_k(p)$$

if and only if p_* is feasible and there exists a $\lambda \geq 0$ such that:

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

$$\lambda (\Delta - \|p_*\|_2) = 0$$

$B + \lambda I$ is positive semi-definite.

In practice, solving p_* in this way is "overkill", there are approximate solutions to the constrained model problem that will do.

The first approximate solution that we look at is the Cauchy Point (§ 7.6) EXAM

Idea : Suppose Δ is very small. Then,

$$m_k(p) \simeq f_k + \langle g, p \rangle, \quad \|p\|_2 \leq \Delta$$

We find the p that minimizes the cost function in this model: This is the steepest-descent direction, $p \propto -g$. A guess for p (magnitude and direction) puts p at the trust-region boundary:

$$p_{temp} = -\frac{\Delta}{\|g\|_2} g.$$

A refined guess is then:

$$p = \tau \, p_{temp} \quad , \quad 0 < \tau \le 1 .$$

We determine τ as follows:

$$\tau = \underset{\tau > 0}{\operatorname{argmin}} \; m_k \left(\tau \, p_{temp} \right) . \qquad (5)$$

i.e. the optimal value of τ is the one that minimizes the M.P. with the quadratic term included.

Eqn (5) can be re-written as:

$$m_k (\tau \, p_{temp}) = f_k - \frac{\tau \Delta \langle g, g \rangle}{\| g \|_2} + \frac{1}{2} \tau^2 \Delta^2 \frac{\langle g, B g \rangle}{\| g \|_2^2} .$$

The next step will be to solve for τ. TBC.