Algorithms Based On The Cauchy Point

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

February 20, 2024

The Cauchy Point

Perspective From Line Search

- Even when optimal step lengths are not used methods could be globally convergent.
- The step length α_k needs to only satisfy fairly loose criteria.

Perspective For Trust-Region Method

- A similar imposition rather relation applies to trust-region methods as well.
- Even though the optimal solution to the sub-problem is seeked, it is enough to find an approximate solution p_k within the trust region which gives some sufficient reduction to obtain global convergence.
- The sufficient reduction could be quantified in terms of Cauchy Point, which is denoted by p_k^C and defined as follows:

Cauchy Point (Algorithm)

Cauchy Point Calculation

Step-I:

Find the vector p_k^S that solves a linear version of the trust region sub-problem i.e.

Find p_k^s s.t. $||p_k^s|| \leq \Delta_k$ and

$$p_k^s = \arg\left\{\min_{p\in\mathbb{R}^n} f_k + g_k^T p\right\}$$

Step-II:

Calculate the scalar $\tau_k > 0$ that minimizes $m_k(\tau p_k^s)$ subject to satisfying the trust-region bound i.e.

$$\tau_k = \arg\left\{\min_{\tau\geq 0} m_k(\tau p_k^s)\right\} \qquad \text{s.t. } ||\tau p_k^s|| \leq \Delta_k$$

Step-III:

Set $p_k^C = \tau_k p_k^s$.

Explicit Computation of The Cauchy Point

- Note that the problem in Step-I is a linear function.
- As a consequence p_k^s chosen in the direction of the negative gradient should keep yielding reduction in the function value.
- That is to reduce the function one can move along the direction:
$$-\frac{g_k}{||g_k||}$$
- As the function is linear the minimiser will lie at the boundary of the trust-region, giving
$$p_k^s = -\frac{\Delta_k}{||g_k||} g_k.$$
- Now in Step-II we calculate τ_k, we search for the minimiser of the model m in the direction of p_k^s (along the ray).

Explicit Computation of The Cauchy Point

Towards this end we consider two cases

Case-I: $g_k^T B_k g_k \leq 0$
Case-II: $g_k^T B_k g_k > 0$

Case-I:

- The function $m_k(\tau p_k^s)$ decreases monotonically with increasing τ whenever $g_k \neq 0$.
- $f_k + \tau {p_k^s}^T g_k$ is decreasing as a consequence of the choice of ${p_k^s}^T$.
- Now since $g_k^T B_k g_k \leq 0$ we have

$$f_k + \tau {p_k^s}^T g_k + \frac{1}{2} {p_k^s}^T B_k p_k^s$$

also decreases as ${p_k^s}^T B_k p_k^s = \tau^2 \Delta_k^2 \frac{g_k^T B_k g_k}{||g_k||^2} \leq 0$, when τ increases.

Explicit Computation of The Cauchy Point

- Therefore, the minimum is attained at simply the largest value that satisfies the trust-region bound for τ_k, i.e. $\tau_k = 1$.

Case-II:

As $g_k^T B_k g_k > 0$, $m_k(\tau p_k^s)$ is a convex quadratic in τ, so

- τ_k is either the unconstrained minimiser of this quadratic i.e.

$$\tau_k = \frac{||g_k||^3}{\Delta_k g_k^T B_k g_k}$$

- or, the boundary value 1.

which ever comes first.

Explicit Computation of The Cauchy Point

Summary

$$p_k^C = -\tau \frac{\Delta_k}{||g_k||} g_k,$$

where

$$\tau_k = \begin{cases} 1, & \text{if } g_k^T B_k g_k \leq 0 \\ \min\left(\frac{||g_k||^3}{\Delta_k g_k^T B_k g_k}, 1\right) & \text{otherwise.} \end{cases}$$

- The Cauchy step is inexpensive to calculate, no matrix factorisation are required.
- It is of crucial importance in deciding if an approximate solution of the trust-region sub-problem is acceptable.
- Specifically, a trust-region method will be globally convergent if its steps p_k give a reduction in the model m_k that is at least some fixed positive multiple of the decrease attained by Cauchy step.

The Cauchy Point

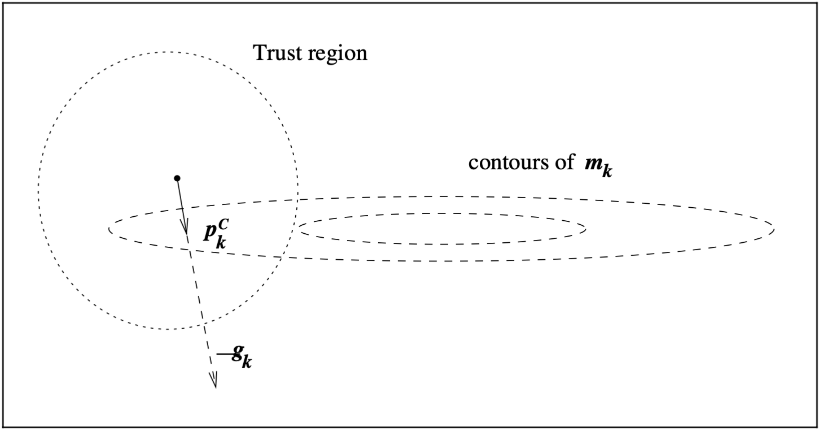

Figure: The Cauchy point for a subproblem in which B_k is positive definite. In this example, p_k^C lies strictly inside the trust region

Improving On The Cauchy Point

- Why look any further, if p_k^C provides sufficient reduction for convergence and the cost of calculating it is so small.
- By always taking the Cauchy point as our step, we are simply implementing the steepest descent method with a particular choice of step length.
- Anyways, steepest descent performs poorly even for an optimal step length choice at each iteration.
- Similar issues one can raise as were done against steepest descent,

  The Cauchy point doesn't depend very strongly on the matrix B_k, as it is used only for step length calculation.

Improving On The Cauchy Point

- For rapid convergence B_k must be involved in the choice of direction as well. (if B_k contains valid information about the curvature of the function)
- Many trust-region algorithms compute the Cauchy point and then try to improve on it.
- The improvement strategy is often designed so that the full step length
$$p_k^B = -B_k^{-1} g_k$$
is chosen whenever B_k is positive definite and $||p_k^B|| \leq \Delta_k$.
- When B_k is the exact Hessian $\nabla^2 f(x_k)$ or a quasi-Newton approximation, it can be expected to yield superlinear convergence.