The Trust Region Method

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

February 16, 2024

Similarities and Contrasts With Line Search Methods

- Line search methods generates a search direction based on a (linear or quadratic) model, and then focus their efforts on finding a suitable step length α along this direction.
- Trust-region methods define a region around the current iterate within which they trust the model to be an adequate representation of the objective function, and then choose the step to be the approximate minimizer of the model in this region.
- In effect, baring the choice of the trust region, they choose the direction and length of the step simultaneously.
- If a step is not acceptable, they reduce the size of the region and find a new minimizer.
- In general the direction of the step may change whenever the size of the trust region is altered (quite unlike the line-search methods).

Trust-Region Methods

- The size of the trust region is critical to the effectiveness of each step.
- If the region is too small, the algorithm misses an opportunity to take a substantial step that would have moved it very close to the minimiser.
- If the trust region is too large, then the minimiser of the model might be far away from the minimiser of the objective function.
- In practice the choice of a larger or smaller trust region depends on the performance of the model in the previous iteration.
- Trust region is expanded if the model consistently keeps producing good steps and is a accurate approximation to the objective function.
- A failed step hampers the confidence of the method leading to a shrinkage of the trust region

Line Search Vs Trust-Region Methods

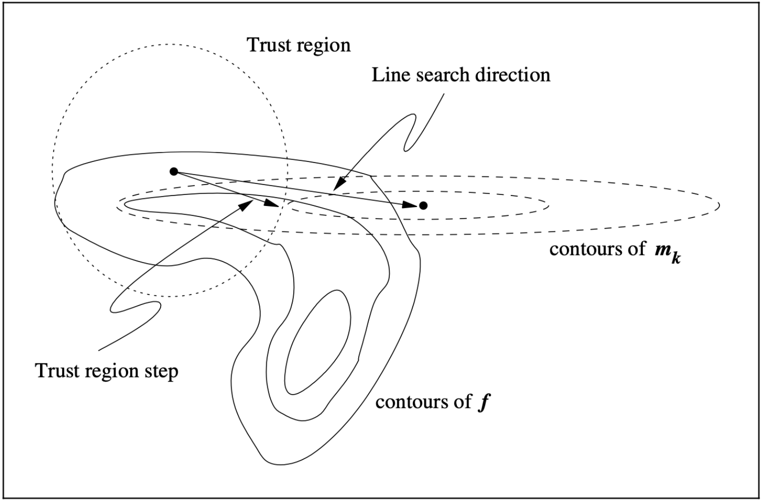

Figure: Trust-region and line search steps.

Line Search Vs Trust-Region Methods

- Figure shows the contrast between trust-region and line-search approaches on a function f of two variables in which the current point x_k and the minimiser x^* lie at opposite ends of a curved valley.
- The quadratic model m_k is constructed from the function and derivative information at x_k and also may possibly also have considered information accumulated from previous iterations and steps.
- A line search method based on this model searched along to the minimiser to of m_k, but this direction will yield atmost a small reduction in f, even if the optimal step length is used.
- Whereas the trust-region methods step to the minimiser of m_k within the dotted circle (trust region), yielding a more significant reduction in f and better progress toward the solution.

Model Derivation

Consider the Taylor series expansion of f around x_k, which is

$$f(x_k + p) = f_k + g_k^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p. \tag{1}$$

where $f_k = f(x_k)$, $g_k = \nabla f(x_k)$ and $t \in (0, 1)$.

- Using an approximation B_k to the Hessian in the second-order term in (1), m_k (the model function) is defined as follows:

  $$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p. \tag{2}$$

  where B_k is some symmetric matrix.
- Note that the difference between m_k and $f(x_k + p)$ is $\mathcal{O}(||p||^2)$, which is small when p is small.
- For $B_k = \nabla^2 f_k$ we get the trust-region Newton method and the error in the model is further brought down to $\mathcal{O}(||p||^3)$ for $||p||$ small.

The Trust Region Subproblem

- The general trust region step is found by seeking a solution to the subproblem
$$\min_{p \in \mathbb{R}^n} m_k(p) \tag{3}$$
where $\Delta_k > 0$ is called the trust-region radius.
- Typically we choose $||.||$ to be the Euclidean norm effecting the trust region to be a ball of radius Δ_k around p.
- The trust-region approach requires us to solve a sequence of subproblems of the type (3) in which the objective function and constraint $(p^T p \leq \Delta_k^2)$ are both quadratic.
- When B_k is positive definite and $||B_k^{-1} g_k|| \leq k$,the solution of (3) is easy to identify, as it is simply the unconstrained minimum $p_k^B = -B^{-1} g_k$ of the quadratic $m_k(p)$).
- In this case, p_k^B is called the full-step.
- The solution of (3) is not so obvious in other cases, but it can usually be found without too much computational expense.