Algorithms For Unconstrained Minimisation
An Overview

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

February 2, 2024

Overview

- All algorithms for unconstrained minimization require the user to supply a starting point(an initial guess denoted by x_0).
- x_0 is chosen using some insight about the problem at hand.
- Beginning at x_0, optimization algorithms generate a sequence of iterates $\{x_k\}_{k=0}^{\infty}$ that terminate when either no more progress could be made or when the solution is approximated to some desired accuracy.

Stopping Criteria

can't we use $||x_K - x^*||$ or $||f(x_k) - f(x^*)||$??
In practice given a small $\epsilon > 0$ (Tolerance) s.t.

- $||\nabla f(x_k)|| < \epsilon$.
- $||x_k - x_{k-1}|| < \epsilon$ or $||x_k - x_{k-1}|| < \epsilon ||x_{k-1}||$.
- $|f(x_k) - f(x_{k-1})| < \epsilon$ or $|f(x_k) - f(x_{k-1})| < \epsilon |f(x_{k-1})|$.

Overview

- In order to move from one iterate x_k to the next i.e. x_{k+1}, the algorithms may use just the information about the function f at x_k or it may use some or all information at previous iterates $(x_0, x_1, \ldots, x_{k-1})$.
- At the new iterate x_{k+1} desirably the function value is lesser than that at x_k (monotone algorithms).
- There do exist non-monotone algorithms, but even for them at some $m > 0$, $f(x_{k+m}) < f(x_k)$.
- There are two fundamental strategies (families of procedures) to move from the point x_k to x_{k+1}:

1. Line Search Methods

2. Trust Region Methods

Line Search Strategy

- The algorithm chooses a direction p_k and searches along this direction from the current iterate x_k for a new iterate with a lower function value.
- The distance to move along p_k can be found by approximately solving the following one-dimensional minimization problem to find a step length α:

$$\min_{\alpha>0} f(x_k + \alpha p_k)$$

- Exact solution would give the maximum benefit of moving along p_k.
- But, an exact minimisation may be expensive and is usually unnecessary.
- Instead, a limited number of trial step lengths are generated by the algorithm until it finds an approximation of the minimum (loosely).
- At the new point a new step-direction and step length are calculated.

Line Search Strategy

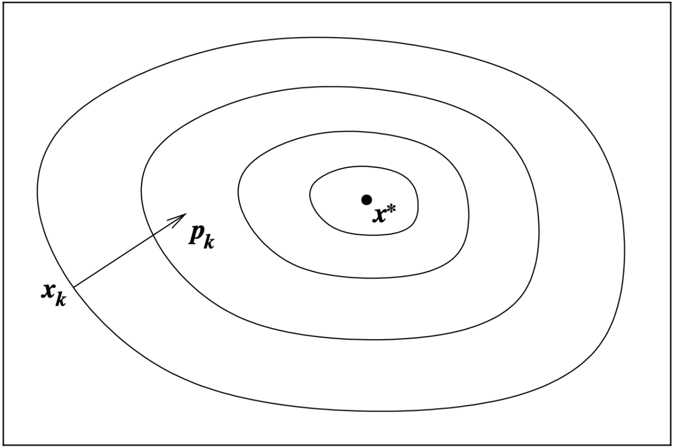

Figure: Line Search Search Direction.

Trust Region Strategy

- The information gathered about f is used to construct a model function m_k.
- The behaviour near the current point x_k is similar to that of the actual objective function f near x_k.
- As the model m_k may not be a good approximation of f when x is far from x_k, the search for a minimizer of m_k is restricted to a small region (trust region) around x_k .
- The candidate step p is found by approximately solving the following sub-problem

$$\min_{p} m_k(x_k + p)$$

  where $x_k + p$ lies inside the trust region.
- If the candidate step doesn't procedure sufficient reduction then the trust region radius is deemed to be too large.

Trust Region Strategy

- The trust region is shrunk and the process is resumed.
- The region usually is a ball defined by

$$||p||_2 \leq \Delta,$$

  where $\Delta > 0$ is called the trust-region radius.
- The model m_k is usually defined by a quadratic function of the form:

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

  f_k and ∇f_k are the functional values and gradient of f at x_k.
- m_k is in agreement with f at x_k upto first order.
- The matrix B_k is either the Hessian or some approximation of it.

Trust Region Strategy

Example

Suppose that the objective is given by

$$f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

At the point $x_k = (0, 1)$ its gradient and Hessian are:

$$\nabla f_k = \begin{bmatrix} -2 \\ 20 \end{bmatrix}, \qquad \nabla^2 f_k = \begin{bmatrix} -38 & 0 \\ 0 & 20 \end{bmatrix}$$

Trust Region Strategy

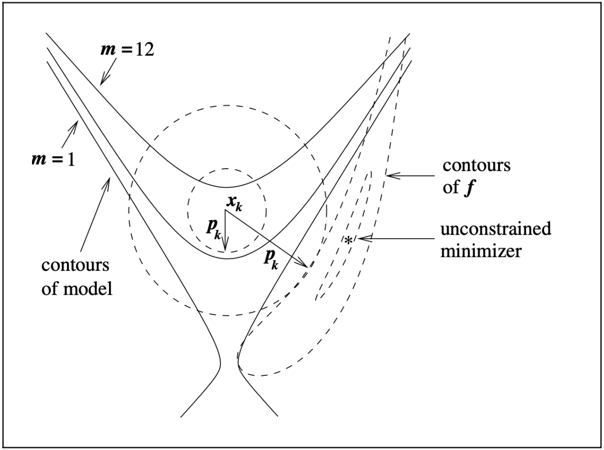

Figure: Two possible trust regions (circles) and their corresponding steps p_k. The solid lines are contours of the model function m_k.

Trust Region Strategy

- Each time the size of the trust region is decreased after failure of a candidate iterate, the step from x_k to the new candidate will be shorter, and it usually points in a different direction from the previous candidate.
- The trust-region strategy differs in this respect from line search, which stays with a single search direction.
- The line search and trust-region approaches differ in the order in which they choose the direction and distance of the move to the next iterate.
- Line search starts by fixing the direction p_k and then identifying an appropriate distance, namely the step length α_k.
- In trust region, a maximum distance the trust-region radius Δ_k is chosen and then a direction and step that attain the best improvement possible subject to this distance constraint is seeked.
- If this step proves to be unsatisfactory, we reduce the distance measure Δ_k and try again.