Numerical Optimization: Trust Region Methods

Qiang Zhu

University of Nevada Las Vegas

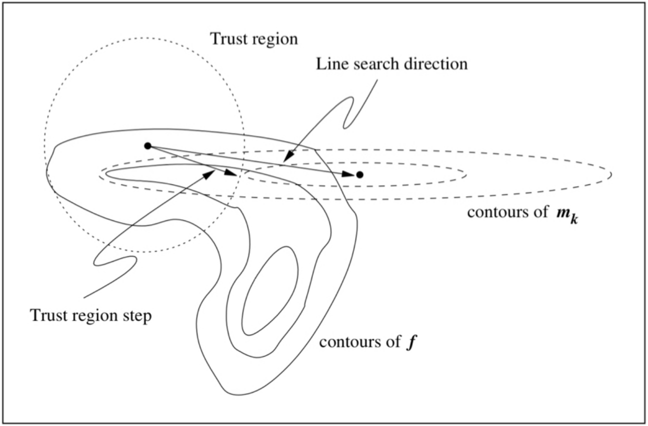
May 14, 2020

Overview

1. Trust Region Model
2. The outline of trust region approach
3. Summary

The problem of line search

In the line search method, one usually use the direction based on the first- or second-order derivative, and then do an approximate 1D search. If the derivative is far from the local minimum, such search may result in excessively large steps or premature convergence.



Line search .v.s trust region

Line search

- Find the direction of improvement
- Select a step length

Trust region

- Select a trust region (within a hypersphere)
- Find a point of improvement

Quadratic approximation

In this chapter, we will assume that the model function m_k that is used at each iterate x_k is quadratic. m_k is based on the Taylor-series expansion of f.

$$f(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp)p,$$

where t is some scalar in the interval (0,1).
By using an approximation B_k to the Hessian in the second-order term, m_k is defined as follows:

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

The difference between $m_k(p)$ and $f(x_k + p)$ is $O(p^2)$,which is small when p is small.

Trust region step

The trust-region method steps to the minimizer of m_k within the dotted circle, yielding a more significant reduction in f and better progress toward the solution.

To obtain each step, we seek a solution of the subproblem

$$\min \ m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k f(x_k + tp) p,$$
$$\text{s.t.} \ \ ||p||_2 \leq \Delta_k,$$

where Δ_k is the trust-region radius.

Thus, the trust-region approach requires us to solve a sequence of subproblems in which the objective function and constraint (which can be written as $p^T p \leq \Delta_k$) are both quadratic, which is easy to solve if it is convex.

How to adjust the Δ_k?

For a given step, we define

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

The numerator is called the actual reduction.
The denominator is the predicted reduction, which is non-negative.

- if $\rho_k < 0$, the new objective value $f(x_k + p_k)$ is greater than $f(x_k)$, reject.
- if $\rho_k \approx 1$, there is good agreement between the model m_k and the function f, expand the trust region
- if $0 < \rho_k \ll 1$, shrink the trust region by reducing δ_k

Algorithm

```
function trust_region_descent(f, G, H, x, k_max;
    eta1=0.25, eta2=0.5, gamma1=0.5, gamma2=2.0, delta=1.0)

    y = f(x)
    for k in 1 : k_max
        x1, y1 = solve_subproblem(G, H, x, delta)
        r = (y - f(x1)) / (y - y1)
        if r < eta1
            delta *= gamma1
        else
            x, y = x1, y1
            if r > eta2
                delta *= gamma2
            end
        end
    end
    return x
end
```

Summary

- Trust region method may perform better when the initial point is far from the local minimum
- The correctness of trust region method relies on the accuracy of the model function
- The step size is controled by the trust-region radius which is updated each step
- Quadratic approximation needs the information of hessian
- The subproblem optimization may be tricky when hessian is not positive definite