Trust region methods

Mark S. Gockenbach

1 Introduction

I have presented a single strategy for globalizing Newton's method and quasi-Newton methods, namely, the use of a line search. The reader will recall that, in the line search approach, the local model

$$f(x+p) \doteq q_k(x) = f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p \cdot H_k p \tag{1}$$

is used to generate a descent direction $p^{(k)}$. A line search is then used to choose the step length α_k and then $x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$.

Trust region methods use the quadratic model q_k in an essentially different fashion. Since (1) is a local model, it is expected to accurately represent f only in a region near $x^{(k)}$. Using heuristics, it is possible to maintain an estimate Δ of the radius of a ball on which the quadratic model is an accurate representation of f. The step from $x^{(k)}$ is then determined by minimizing q_k over this ball, which is called the *trust region* (it is the region in which the quadratic model is trusted). The trust region subproblem is

$$\begin{array}{ll} \min & f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p \cdot H_k p \\ s.t. & \|p\| \leq \Delta. \end{array}$$

This should be compared with the line search subproblem, which is the one-dimensional minimization

$$\min_{\alpha > 0} f(x^{(k)} + \alpha p^{(k)}).$$

The advantage of the trust region approach is partially due to the fact that H_k is not required to be positive definite (so, in particular, $H_k = \nabla^2 f(x^{(k)})$ can be used even if it is singular or indefinite). If H_k has directions of negative curvature, then the trust region algorithm can take advantage of them directly.

As a result of this better use of the curvature information, it is possible to prove stronger convergence theorems. In particular, when $H_k = \nabla^2 f(x^{(k)})$, then it is possible (under some assumptions) to prove convergence to a stationary point x^* that also satisfies the second-order necessary condition that $\nabla^2 f(x^*)$ be positive semidefinite.

2 The trust region subproblem

The trust region subproblem is interesting in its own right. It is an inequality-constrained minimization problem, and therefore falls into a class of problems that I will discuss later in the course. However, the trust region subproblem is special in that it is possible to write down necessary and sufficient conditions for p to be a global minimizer. For general inequality-constrained optimization problems, as for general unconstrained optimization problems, only necessary conditions are available for global solutions.[1]

[1] As I will discuss later, there is a class of inequality-constrained nonlinear programs (called *convex* programs) for which the necessary conditions are necessary and sufficient. The trust region subproblem has (almost) the same property even though it is *not* a convex problem.

Theorem 2.1 *Suppose $f \in \mathrm{R}$, $g \in \mathrm{R}^n$, and $H \in \mathrm{R}^{n \times n}$. The vector p^* is a global solution of*

$$\begin{aligned} \min \quad & f + g \cdot p + \frac{1}{2} p \cdot Hp \\ s.t. \quad & \|p\| \le \Delta \end{aligned}$$

if and only if $\|p^*\| \le \Delta$ and there is a scalar $\lambda \ge 0$ such that:

1. $(H + \lambda I)p^* = -g$;
2. $\lambda(\Delta - \|p^*\|) = 0$;
3. *and $H + \lambda I$ is positive semidefinite.*

The scalar λ is the *Lagrange multiplier* for the trust region subproblem (I will define Lagrange multipliers later), and it can be shown that λ increase as Δ decreases, with $\lambda \to \infty$ as $\Delta \to 0^+$. If H is nonsingular and the quasi-Newton step $p^N = -H^{-1}g$ lies inside the trust region, then p^N is the solution of the trust region subproblem and $\lambda = 0$. On the other hand, as Δ decreases to zero, λ increases to infinity, which means that $H + \lambda I$ is increasingly dominated the λI term. The effect is that $p^* = -(H + \lambda I)^{-1} g$ lies more and more in the steepest descent direction as $\Delta \to 0$. Thus the solution to the trust region subproblem follows a path from the quasi-Newton direction to the steepest descent direction as Δ decreases.

I will not discuss trust region methods further. There are a variety of algorithms for solving the trust region subproblem and these have been incorporated into effectives optimization codes. The reader is referred to Nocedal and Wright [1], Chapter 4, for more details.

References

[1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer-Verlag, New York, 1999.