Notes on Nocedal and Wright's "Numerical Optimization"
Chapter 4 – "Trust-Region Methods"

Lucas N. Ribeiro

Introduction I

- 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.
- Example: quadratic model (2nd order Taylor expansion around x_k)
$$m_k(p) = f_k - \nabla_k^{\mathsf{T}} p + 0.5 p^{\mathsf{T}} B_k p, \tag{1}$$
where $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$ and B_k is some symmetric matrix. If $B_k = \nabla^2 f(x_k)$ then we call trust-region Newton method.
- To obtain each step, we seek a solution of the subproblem:
$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + \nabla f_k^{\mathsf{T}} p + 0.5 p^{\mathsf{T}} B_k p, \quad \text{s.t.}\ \|p\| \le \Delta_k, \tag{2}$$
where Δ_k is the trust-region radius.
- *full step*: when B_k is positive definite and $\|B_k^{-1} \nabla f_k\| \le \Delta_k$, the solution of (2) is $p_k^B = -B_k^{-1} \nabla f_k$.
- Full step might be expensive. Consider *approximations*.

Algorithm outline I

- First issue: how to choose the trust-region radius Δ_k at each iteration?
- Define the metric

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} \tag{3}$$

  The numerator is the actual reduction and the denominator the predicted reduction.
- Decisions:
  - $\rho_k < 0$ step rejected (as the new obj. value is larger than the current);
  - $\rho_k \approx 1$ good agreement, it's safe to expand the radius at the next it;
  - $\rho_k > 1$ keep it;
  - $0 \leq \rho_k \leq 1$ shrink it.

Algorithm outline II

Algorithm outline (more details in the book)

1. Initialize parameters;
2. Obtain p_k by (approximately) solving (2);
3. Evaluate ρ_k
4. Update $x_k = x_k + p_k$ (or not if decided based on ρ_k)

Solving (2)

Strategies for approximate solutions

- dogleg method – when B_k is positive definite;
- 2D subspace minimization – applied to indefinite B_k;
- Steihaug's method – B_k is the exact Hessian.

The Cauchy point I

Although we are seeking the optimal solution for (2), it is enough for global convergence purposes to find an approx. solution p_k that lies within the trust region and gives **sufficient reduction** in the model.
The sufficient reduction can be quantified in terms of the Cauchy point.
To find the Cauchy point, we use the following procedure:

1. Find $p_k^s = \arg\min_p f_k + \nabla f_k^\mathsf{T} p = -\frac{\Delta_k}{\|\nabla f_k\|}\nabla f_k$
2. Calculate the scalar satisfying the trust-region bound:

$$\tau_k = \arg\min_{\tau > 0} m_k(\tau p_k^s)$$

3. The Cauchy point is then given by $p_k^C = \tau_k p_k^s$.

The Cauchy point II

The Cauchy point is inexpensive to calculate and is important in deciding if an approx solution of (2) is acceptable. Specifically, a trust region method will be globally convergent if its steps attain a sufficient reduction in m_k, i.e., they give a reduction on m_k that is at least some fixed multiple of the decrease attained by the Cauchy step!

Improving on the Cauchy step:

- Dogleg method
- 2D subspace min
- Steihaug