Notes on Nocedal and Wright's "Numerical Optimization"
Chapter 6 – "Practical Newton Methods"

Lucas N. Ribeiro

Introduction I

- Line search methods:

$$x_{k+1} = x_k + \alpha_k p_k$$

- p_k is the search direction (must point to descent direction $p_k^{\mathsf{T}} \nabla f_k < 0$ – leading to global convergence!)
- Gradient descent: $p_k = -\nabla f_k$.
- More general: $p_k = -B_k^{-1} \nabla f_k$, where B_k is a symmetric and non-singular matrix.
- Newton method: $B_k = \nabla^2 f_k$ (Hessian).
- Problems with standard Newton method: Hessian may not be positive definite all the time, and/or close to singular, possibly leading to ascent directions!
- The Newton step is obtained by solving the symmetric $n \times n$ linear system:

$$\nabla^2 f(x_k) p_k^N = -\nabla f(x_k)$$

Introduction II

- Approaches to ensure good Newton step:
  1. Solve the Newton problem by means of a conjugate gradient (Newton-CG)
  2. Modifying the Hessian matrix $\nabla^2 f(x_k)$ so it becomes sufficiently positive definite (modified Newton method)
- Computational complexity concerns:
  1. Newton-CG method – terminate CG iteration before an exact solution is found (inexact Newton approach)
  2. Computation of the Hessian matrix is quite complex

Inexact Newton steps I

- Standard Newton method has to calculate its step exactly, which can be expensive
- Classical linear system methods (Gaussian elimination etc) may be too complex when the number of variables is large $\rightarrow$ iterative solutions are appealing here
- An inexact solution may work
- Here we consider the CG residual $r_k = \nabla^2 f(x_k) p_k + \nabla f(x_k)$
- (Note that the standard CG residual is $r_k = Ax_k - b$!)
- A typical condition checked to terminate the iterative solver is

$$\|r_k\| \leq \eta_k \|\nabla f(x_k)\|$$

  where η_k is the forcing sequence. This factor is important to determine the rate of convergence of inexact Newton methods.

Inexact Newton steps II

Rate of convergence results

- Local convergence is obtained simply by ensuring that η_k is bounded away from 1.
- Rate of convergence is linear if $\eta_k \to 0$ and superlinear if $\eta_k = O(\|\nabla f(x_k)\|)$

Line search Newton methods I

- Each iteration has the form $x_{k+1} = x_k + \alpha_k p_k$
- $alpha_k$ is a step length that satisfied Wolfe-Goldstein conditions and can be obtained by backtracking (although $\alpha_k = 1$ is a good initial guess)
- p_k is either the exact Newton step or its approximation

Line search Newton CG

- Applies CG to calculate the Newton step.
- Recall that CG is designed to positive definite systems, but Hessian may admit negative eigenvalues, thus a test is introduced in the Newton CG to check for negative curvatures
- Preconditioning can be applied to correct the system's curvature.

Line search Newton methods II

Line Search Newton CG

- Given initial point x_0
- For ...
  1. Compute search direction applying CG method to $\nabla^2 f(x_k)p = -\nabla f(x_k)$. Terminate if $\|r_k\| \leq \min(0.5, \sqrt{\nabla f(x_k)})\|\nabla f(x_k)\|$ or if negative curvature is encountered.
  2. Find step size α_k
  3. Set $x_{k+1} = x_k + \alpha_k p_k$

Modified Newton's method I

- The search direction can be obtained by linear algebra factorization techniques.
- If the Hessian is not definite positive, we can modify this by adding a matrix E_k so that it becomes nice.
- Global convergence can be attained if E_k satisfies the bounded modified factorization property
- In this case $B_k = \nabla^2 f(x_k) + E_k$ has bounded condition number
- We discuss some Hessian modifications:
  - Eigenvalue modification: changes negative eigenvalues to non-negative
  - Adding multiple of identity
  - Modified Cholesky factorization, Gershgorin modification: adds elements to the Cholesky diagonal

Modified Newton's method II

Line Search Newton CG

- Given initial point x_0
- For ...
  1. Factorize $B_k = \nabla^2 f(x_k) + E_k$
  2. Solve $B_k p_k = -\nabla f(x_k)$
  3. Find step size α_k
  4. Set $x_{k+1} = x_k + \alpha_k p_k$

Trust-region Newton method

- Trust-region methods do not require the quadratic model to be positive definite, then we can use $B_k = \nabla^2 f(x_k)$ and get p_k by solving

$$\min_{p} m_k(p) = f_k + \nabla f_k^{\mathsf{T}} p + 0.5 p^{\mathsf{T}} B_k p$$

  subject to $\|p\| \leq \Delta_k$.
- It can be solved by many techniques, including:
  - Newton dogleg method
  - 2D subspace minimization
  - Accurate solution using iteration
  - CG method (CG-Steihaug!)
  - Lanczos method