OPER 627: Nonlinear Optimization
Lecture 9: Trust-region methods

Department of Statistical Sciences and Operations Research
Virginia Commonwealth University

Sept 25, 2013

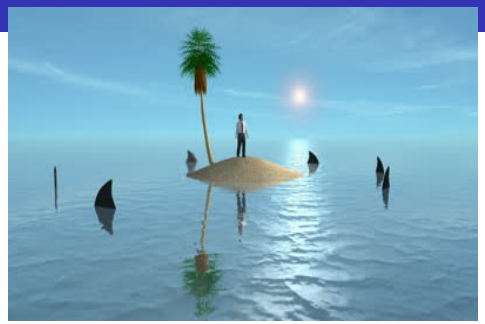
Quiz

- True or False:
  - From any arbitrary initial point, BFGS algorithm guarantees convergence to a stationary point
  - For a general nonlinear smooth function, BFGS algorithm has superlinear local convergence under some mild conditions
  - Quasi-Newton methods are good at sparse large-scale optimization problems
- What is the idea of LBFGS method?

Today's Outline

- Introduction to trust region methods
- Trust region algorithm
- Trust region subproblems

An alternative perspective



Model function to approximate $f(x_k + p)$:

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

This model is only "locally" reliable: p cannot be too large

Let's go local

1. Step 1: Choose a trust region (a compact convex neighborhood around the current point)
2. Step 2: Find a minimizer of some model function in the trust region

Q: Why do we need the trust region?

Trust region subproblem

1. Model function: $m_k(p) = f(x_k) + g_k^\top p + \frac{1}{2} p^\top B_k p$
2. Trust region radius: Δ, $\|p\| \leq \Delta$ (2-norm by default)

- B_k could be the true Hessian
- B_k could be a quasi-Newton approximate matrix

Trust region subproblem (TRP):

$$\min_p \; m(p)$$
$$\text{s.t.} \;\; \|p\| \leq \Delta$$

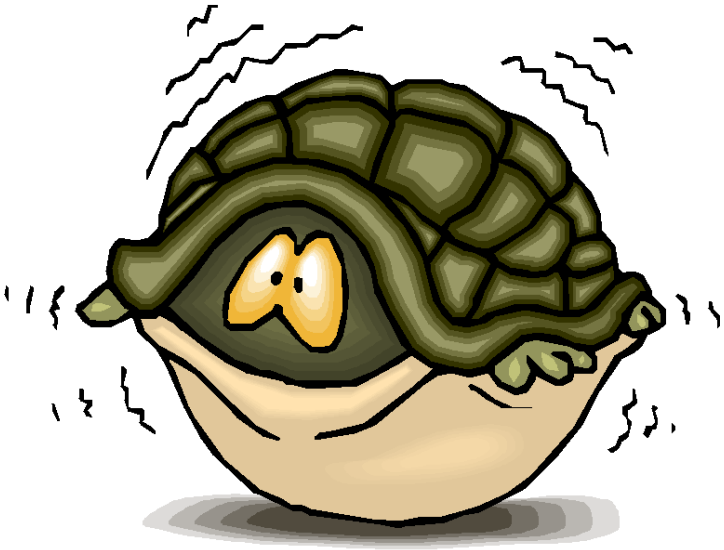
Make p_k as large as possible (Why?), as long as $m(p)$ agrees "closely" with $f(x+p)$

How close?

- A-Red_k: actual reduction in f, $\Delta f_k := f(x_k) - f(x_k + p_k)$
- P-Red_k: predicted reduction in f, $\Delta m_k := f(x_k) - m_k(p_k)$
- Ratio: $\rho_k = \frac{\Delta f_k}{\Delta m_k}$, if close to 1, then we have a good agreement

Trust region: simple idea

If ρ is far from 1:



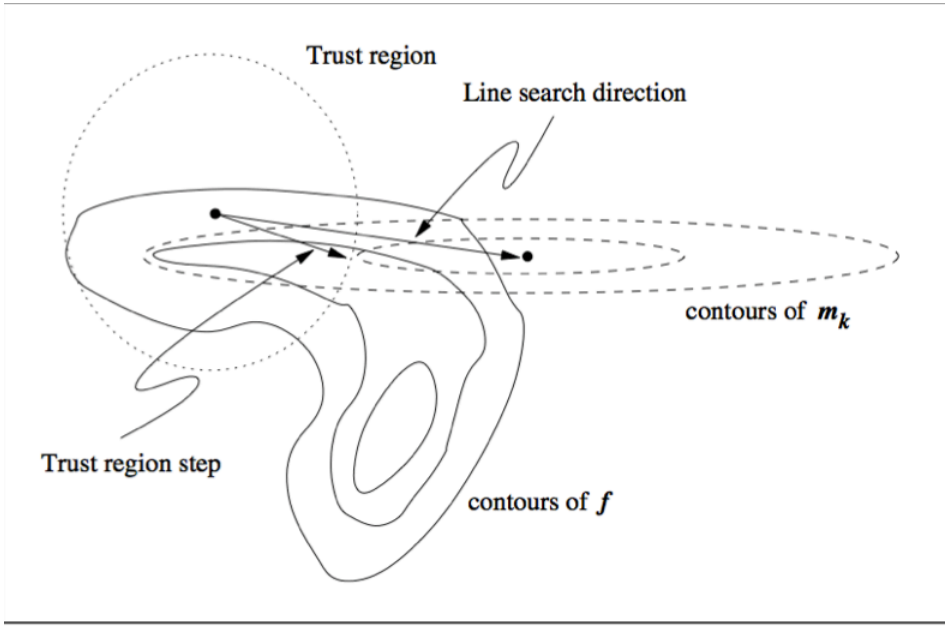
If ρ is close to 1, and p hit the boundary: $\|p\| = \delta$:

Trust region algorithm

Algorithm 1 Trust region ($\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta}), \eta \in [0, \frac{1}{4})$)

Given x_k, $\Delta_k > 0$, calculate $g_k = \nabla f(x_k)$ and B_k
(Key) Solve TRP_k for p_k
Evaluate $f(x_k + p_k)$, and calculate Δf_k, ρ_k
if $\rho_k < \frac{1}{4}$ (not a good model, reduce the trust region radius) **then**
    $\Delta_{k+1} = \frac{\Delta_k}{4}$
else if $\|p_k\| = \Delta_k$ (hit the boundary) **then**
    $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$
else
    $\Delta_{k+1} = \Delta_k$
end if
if $\rho_k > \eta$ (so there are some reduction in the model) **then**
    $x_{k+1} = x_k + p_k$
else
    $x_{k+1} = x_k$ (null step)
end if

Trust region and line search



Example: $\min_x (x^2 - 1)^2$

1. Line search at 0: cannot get over saddle point
2. Trust region:

$$\min_d -2d^2 \text{ s.t. } |d| \leq \Delta$$

Escape the saddle point!

Solution of TRP

Trust region subproblem (TRP):

$$\min_p \; m(p)$$
$$\text{s.t.} \quad \|p\| \leq \Delta$$

Theorem (More and Sorensen 1983)

Vector p^* is a global minimizer of TRP if and only if $\|p^*\| \leq \Delta$ and $\exists \lambda \geq 0$ that:

1. $(B + \lambda I)p^* = -g$
2. $\lambda(\Delta - \|p^*\|) = 0$
3. *$B + \lambda I$ is PSD*

Two cases

1. $\lambda = 0$, this happens only if B is PSD, and $p^* = -B^{-1}g$ satisfies $\|p^*\| \leq \Delta$
2. $\lambda > 0$, then we look for a number λ that is big enough so that $B + \lambda I$ is PSD, and $p^* = -(B + \lambda I)^{-1}g$ satisfies $\|p^*\| = \Delta$

Let $p(\lambda) = -(B + \lambda I)^{-1}g$, we "just" need to solve $\|p(\lambda)\| = \Delta$, which is a single-variable root finding problem!

How to solve? No explicit form $p(\lambda)$!

Let's do some derivation!

Algorithm for solving TRP

Algorithm 2 Trust region subproblem

Given $\lambda_0, \Delta > 0$
for $l = 0, 1, \cdots$ **do**
    Factorize $B + \lambda_l I = R^\top R$ (safeguard here to ensure Cholesky factorization)
    Solve $R^\top R p_l = -g$
    Solve $R^\top q_l = p_l$
    Set $\lambda_{l+1} = \lambda_l + \frac{\|p_l\|^2}{\|q_l\|^2} \frac{\|p_l\| - \Delta}{\Delta}$
end for

Good news: typically just 2 or 3 steps of Newton iteration!