Second Order Methods for Solving Convex Problems (cont.)

Optimization - 10725
Carlos Guestrin
Carnegie Mellon University
April 7th, 2010

Backtracking line search alg.

- **Given**
  - Point `x`
  - Descent direction `Δx`
  - `α ∈ (0, 0.5)`
  - `β ∈ (0,1)`
- `t=1`
- While `f(x + tΔx) > f(x) + αt∇f(x)Δx`
  - `t := βt`

Boyd & Vandenberghe: pick
- `α` in `[0.01, 0.3]`
- `β` in `[0.1, 0.8]`
Newton’s method – alg.

- Start from some \( x \) in \( \text{dom} \ f \)
- Repeat
  - Determine descent direction \( \Delta x_{nt} \)
    - Solving a linear system \( \nabla^2 f(x) \Delta x = -\nabla f(x) \)
  - Line search to choose step size \( t \)
  - Update: \( x \leftarrow x + t \Delta x_{nt} \)
- Until stopping criterion

Good stopping criterion:
\[
\frac{1}{2} \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \leq \epsilon \]

Convergence analysis for Newton’s

- (Really see book for details.)

Two phases:
- Gradient is large
  - Damped Newton Phase
    - Step size \( t<1 \)
    - Linear convergence
  - Gradient is small
    - Pure Newton Phase
      - Step size \( t=1 \)
      - Quadratic convergence
        - \( c^2 \) by exp
      - Only lasts 6 steps
Proof intuition

- Assumptions:
  - Strong convexity:
  - Upper bound on Hessian:
    - Implied by strong convexity
  - Hessian of f is Lipschitz continuous: (bound on third derivative)

- Proof shows Newton follows two phases (details in book):
  - Constants:
  - Damped Newton Phase:
  - Pure Newton Phase:

Complexity of Damped Newton Phase

- When gradient is large:
  - Relate to suboptimality:
    - Total number of iterations in Damped Newton Phase
Pure Newton Phase

- Proof guarantees that
  - When gradient is small:
    - Gradient decreases quadratically:
  - From convexity, can relate gradient to error:
  - Applying bound recursively:

The power of quadratic convergence

- Linear convergence (e.g., at rate ½)
  - Bits of precision improve linearly

- Quadratic convergence (e.g., at rate ½)
  - Bits of precision improve exponentially:
    - So, done after 6 iterations!
Putting it all together: Complexity of Newton’s method – Simple Analysis

- Damped Newton Phase:

- Pure Newton Phase:

- Total:
  - Bound depends on unknown quantities! 😊

Self-Concordance:
A smoothness assumption

- Self-concordant function in 1D:
  - A bound on the third derivative in terms of the second, rather than unknown constants
  - Example: $f(x) = -\log x$
When function is Self-Concordant,
Analysis of Newton’s method is very clean!

- For self-concordant functions, no constants in bound:

- For example:

In Practice, bounds are loose,
but general behavior captured

- Bound on #iterations from previous slide:

- Example from book:
  - Self concordant function
  - Empirical bound:
Solving problems with equality constraints

- Equality constraints:
  - Seems very hard

Null space

- Equality constraints:
  - Given one solution:
    - Find other solutions:
      - Since Null Space is a linear subspace:
Eliminating linear equalities

- Equivalent optimization problems:
  - Find basis for null space of A (linear algebra)
    - Solve unconstrained problem
  - A concern…

Solving quadratic problems with equality constraints

- Quadratic problem with equality constraints:
  - KKT condition \( x^* \) solution iff

  - Rewriting:

  - Solve linear system:
    - Any solution is OPT
    - If no solution, unbounded
Newton’s method with equality constraints

- Quadratic approximation:

  - Start feasible, stay feasible:
  - KKT:

    - Solve linear system:

    - Move accordingly:

General convex problem

- General (differentiable) convex problem:

  - Equivalent problem with only equality constraints:
Approximating the indicator

- Approximate indicator:
  - Correct as t
  - Differentiable

- Approximate optimization problem:

- Convex, if $f_i$ are convex, because

Log-barrier function

- Solve log-barrier problem with parameter $t$:

  - Nice property:
    - Gradient:
    - Hessian:
Force field interpretation

- Log-barrier function:
  - Descending gradient of log barrier

- Each term:
  - Want $f_i(x) \leq 0$
  - As we approach 0:

Central path

- For each $t$, solve:

- As $t$ goes to infinity, approach solution of original problem

- Problem becomes badly conditioned for very large $t$, so want to stay close to path and make small steps on $t$
Barrier method

- **Given:**
  - Feasible \(x^{(0)}\)
  - Initial \(t > 0\)
  - \(\mu > 1\)

- **Repeat**
  - **Centering:**
    - Starting from \(x\), compute:
  - **Update:** \(x := \ldots\)
  - **Stopping criterion:** When \(t\) is “large enough”
  - **Increase barrier param:** \(t := \ldots\)

---

When is \(t\) large enough???

- **Solve centering step:**

  - There exists values for dual vars (See book), such that duality gap \(\leq k/t\)

  - **Thus:**

  - **Stopping criterion** \(k/t \leq \varepsilon\)
Centering step not (necessarily) exact

- Finding exact point on central path can take a while…

- Usually:
  - Run a few steps of Newton to recenter
  - Then increase t
  - (problem: duality gap result no longer holds!!)

- Most often use primal-dual method
  - Equivalent to Newton’s method on Lagrangian

  - See book for details

What about feasible starting point???

- Phase I: Solve feasibility problem, e.g.,
  - Starting from feasible point:
    - (don’t solve to optimality!!! Stop when s<0)
    - When feasible region “not too small”, find point very quickly
  - Phase II: use feasible point from Phase I as starting point for Newton’s or other method

- Also possible:
  - Change Phase I to guarantee starting point (near) central path
  - Combine Phase I and Phase II