Solving Convex Problems

Optimization - 10725
Carlos Guestrin
Carnegie Mellon University

April 9th, 2008

©2008 Carlos Guestrin

Today…

- Thus far, focused on formulating convex problems
  - Today: How do we solve them!
  - Plan: 200 pages of book (Part III) in one lecture
- Focus:
  - Convex functions
  - Twice differentiable
- Overview
  - Unconstrained
  - Equality constraints
  - General convex constraints
-

©2008 Carlos Guestrin

Solving unconstrained problems

- Unconstrained problem
- Sequence of points:
- Exactly: Stop when
- Approximately: Stop when

©2008 Carlos Guestrin

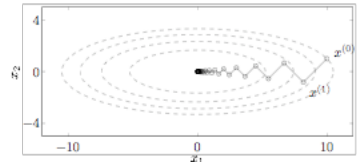
Descent methods

- $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$
  - Want:
- From convexity:
- Thus $\nabla f(x^{(k)})^T (y - x^{(k)}) \geq 0$
- Therefore, pick Δx such that:

©2008 Carlos Guestrin

Generic descent algorithm

- Start from some x in **dom** f
- Repeat
  - Determine descent direction Δx
  - **Line search** to choose step size t
  - Update: x ← x + t Δx
- Until stopping criterion



- Good stopping criterion:
- In gradient descent, Δx =

©2008 Carlos Guestrin

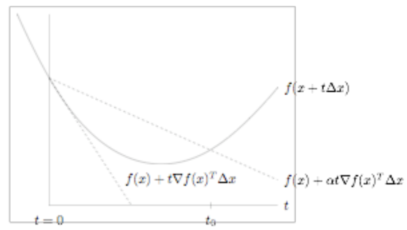
Exact line search

- Find best step size t:
- Problem is
  -
  - Sometimes easy to solve in closed form
  - Other times can take a long time…

©2008 Carlos Guestrin

Backtracking line search

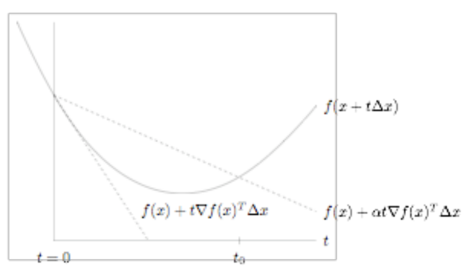
- From convexity,
  lower bound on f(x+Δx):
  - Can't really hope to achieve ideal decrease of
- Instead pick some α
  - And achieve:
- Choosing t:



©2008 Carlos Guestrin

Backtracking line search alg.

- Given
  - Point x
  - Descent direction Δx
  - α
  - β
- t=1
- While f(x+Δx)>
  - t := βt
- Boyd & Vandenberghe: pick
  - α in [0.01,0.3]
  - β in [0.1,0.8]



©2008 Carlos Guestrin

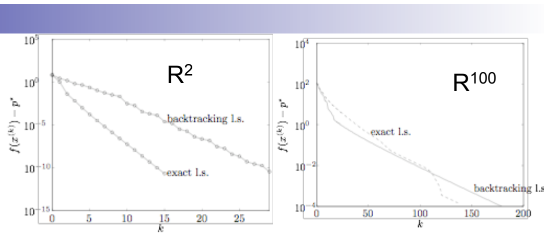
Analysis of gradient descent

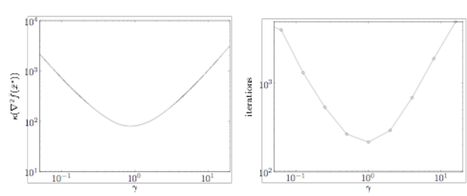
- (details in book…)
- Linear convergence rate:
  - $f(x^{(k)}) - p^* \leq c^k (f(x^{(0)}) - p^*)$
  - Geometrically decreasing
    - c<1
    - In log plot, error decreases below a line…
- Rate c related to "condition number" of Hessian
  - $c \cong 1 - 1/$"condition number"
- For quadratic problem:
  - Condition number is $\lambda_{max}/\lambda_{min}$
- Gradient descent bad when condition number is large

©2008 Carlos Guestrin

Observations about descent algorithms

- Observe linear convergence in practice
- Boyd & Vandenberghe: difference often not significant in large dimensional problems
  - May not be worth implementing exact LS when complex
- Condition number can greatly affect convergence





©2008 Carlos Guestrin

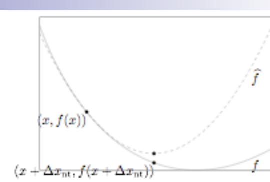
Solving quadratic problems is easy

- Quadratic problem:
- Solving equivalent to solving linear system:
  - If system has at least one solution: done!
  - If system has no solutions: problem is unbounded
- Usually don't have simple quadratic problems, but…

©2008 Carlos Guestrin

Newton's method

- Second order Taylor expansion:



- Descent direction, solution to linear system
- Nice property:
  - We wanted:
  - We get:

©2008 Carlos Guestrin

Newton's method – alg.

- Start from some x in **dom** f
- Repeat
  - Determine descent direction Δx_{nt}
    -
  - **Line search** to choose step size t
  - Update: x ← x + t Δx_{nt}
- Until stopping criterion

- Good stopping criterion:

$$\frac{1}{2}\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \leq \epsilon$$

©2008 Carlos Guestrin

Convergence analysis for Newton's

- (Really see book for details.)

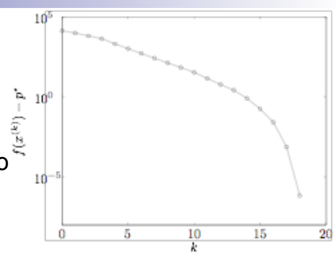
- Two phases:
  - Gradient is large
    - Damped Newton Phase
      - Step size t<1
    - Linear convergence

  - Gradient is small
    - Pure Newton Phase
      - Step size t=1
    - Quadratic convergence
      - c^(2^k)
    - Only lasts 6 steps

©2008 Carlos Guestrin

Summary on Newton's

- Converges in very few iterations, especially in quadratic phase
- Invariant to choice of coordinates or affine scaling
  - Very useful property!
- Performs well with problem size, not very sensitive to parameter choices
- Can prove even cooler things when function is smooth
  - E.g., "self-concordance," see book
- Many implementation tricks (see book)



- But…
  - Forming and storing Hessian is quadratic
    - Can be prohibitive
  - Solving linear system can be really expensive
  - Use *quasi-Newton* methods

©2008 Carlos Guestrin

Solving problems with equality constraints

- Equality constraints:

- Seems very hard

©2008 Carlos Guestrin

Null space

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

©2008 Carlos Guestrin

Eliminating linear equalities

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

©2008 Carlos Guestrin

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

©2008 Carlos Guestrin

Newton's method with equality constraints

- Quadratic approximation:
- Start feasible, stay feasible:
- KKT:
- Solve linear system:
- Move accordingly:

©2008 Carlos Guestrin

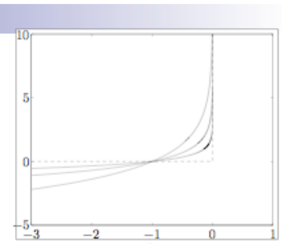
General convex problem

- General (differentiable) convex problem:

- Equivalent problem with only equality constraints:

©2008 Carlos Guestrin

Approximating the indicator



- Approximate indicator:
  -
  - Correct as t
  - Differentiable
- Approximate optimization problem:

- Convex, if f_i are convex, because
  -

©2008 Carlos Guestrin

Log-barrier function

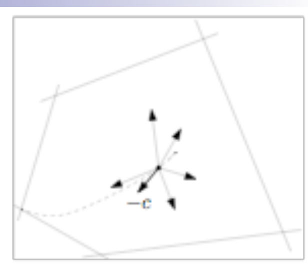
- Solve log-barrier problem with parameter t:

- Nice property:
  - Gradient:
  - Hessian:

©2008 Carlos Guestrin

Force field interpretation

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

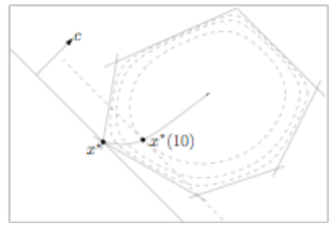


- Each term:
  - Want $f_i(x) \leq 0$
  - As we approach 0_- :

©2008 Carlos Guestrin

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

©2008 Carlos Guestrin

Barrier method

- Given:
  - Feasible x
  - Initial t>0
  - μ>1
- Repeat
  - *Centering*:
    - Starting from x, compute:
  - *Update*: x:=
  - *Stopping criterion*: When t is "large enough"
  - *Increase barrier param*: t:=

©2008 Carlos Guestrin

When is t large enough???

- Solve centering step:
- There exists values for dual vars (See book), such that duality gap ≤ k/t
- Thus:
- Stopping criterion k/t ≤ ε

©2008 Carlos Guestrin

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

©2008 Carlos Guestrin

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

©2008 Carlos Guestrin