Solving Convex Problems

Optimization - 10725
Carlos Guestrin
Carnegie Mellon University

April 9th, 2008

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 $\min_x f_0(x)$
  - Equality constraints $Ax = b$
  - General convex constraints $f_i(x) \leq 0$
- Good luck!! :)

Solving unconstrained problems

- Unconstrained problem $\min_{x} f(x)$
- Sequence of points: $x^{(0)}, x^{(1)}, x^{(2)}, \ldots \xrightarrow[k\to\infty]{} f(x^{(k)}) = p^*$
- Exactly: Stop when $f(x^{(k)}) = p^*$
- Approximately: Stop when $f(x^{(k)}) - p^* \le \varepsilon$

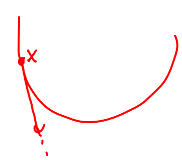
Descent methods

previous value

step size

- $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$ ← descent direction
  - Want: $f(x^{(k+1)}) < f(x^{(k)})$ unless $f(x^{(k)}) = p^*$
- From convexity:

$$f(y) \geq f(x^{(k)}) + \nabla f(x^{(k)})^T (y - x^{(k)})$$



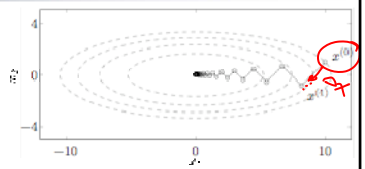
- Thus $\nabla f(x^{(k)})^T (y - x^{(k)}) \geq 0 \Rightarrow f(y) \geq f(x^{(k)})$:(
- Therefore, pick Δx such that:

$\Delta x = y - x^{(k)}$ $\quad \nabla f(x^{(k)})^T \Delta x < 0$

Generic descent algorithm

$\alpha^T \beta \le \|\alpha\| \|\beta\|$

- 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: $\|\nabla f(x)\|_2 \le \eta$ is really really small

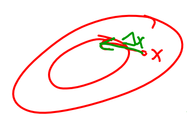
  $\forall y \quad f(y) \geq f(x) + \nabla f(x)^T (y - x)$

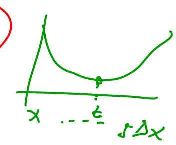
- In gradient descent, Δx = $-\nabla f(x)$

Exact line search

- Find best step size t: at x, pick direction Δx

  $t = \arg\min_s f(x + s\Delta x)$



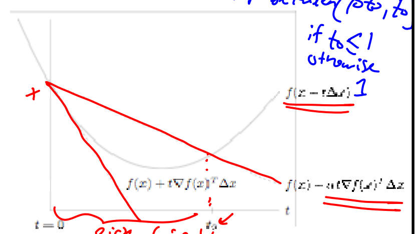


- Problem is
  - $g(s) = f(x + s\Delta x)$ is convex !!
  - Sometimes easy to solve in closed form
  - Other times can take a long time…

Backtracking line search

- From convexity,
  lower bound on $f(x+t\Delta x)$:

  $f(x+t\Delta x) \geq f(x) + t\,\nabla f(x)^T \Delta x$



stop between $(\beta t_0, t_0)$ if $t_0 \leq 1$, otherwise 1

  - Can't really hope to achieve ideal decrease of

find t, such that $f(x+t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$

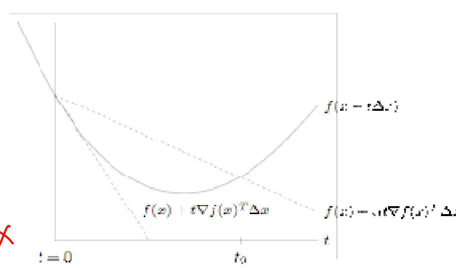
- Instead pick some $\alpha \in (0, 0.5)$
  - And achieve: $\exists t_0 \quad \forall t \in [0, t_0] \quad f(x+t\Delta x) \leq f(x) + \alpha t \nabla f(x)^T \Delta x$

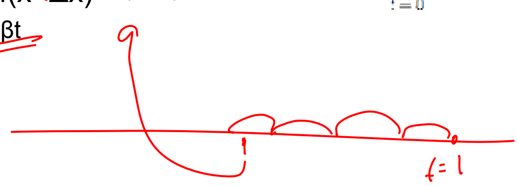
- Choosing t: start with $t=1$, if we are lucky, we are done

  otherwise $t := \beta t$, $\beta \in (0,1)$ Guarantee progress of at least $\beta\, \alpha\, t_0\, \nabla f(x)^T \Delta x$

Backtracking line search alg.

- Given
  - Point x
  - Descent direction Δx
  - α $\in (0, 0.5)$
  - β $\in (0, 1)$
- t=1
- While $f(x+t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$
  - $t := \beta t$





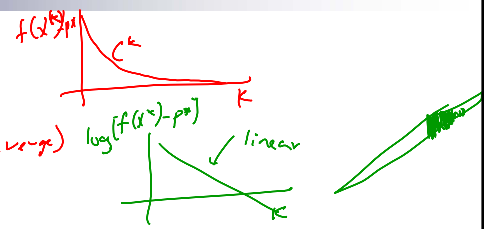
- Boyd & Vandenberghe: pick
  - α in [0.01,0.3]
  - β in [0.1,0.8]

Analysis of gradient descent

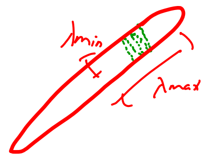
- (details in book…)
- Linear convergence rate:
  - $f(x^{(k)}) - p^* \leq c^k \, (f(x^{(0)}) - p^*)$

    how far opt — iterations — initial

    (linear converge)
  - Geometrically decreasing
    - $c<1$
    - In log plot, error decreases below a line…



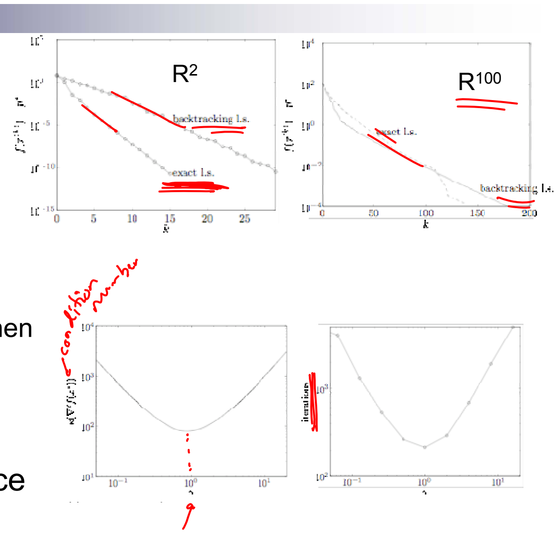
- Rate c related to "condition number" of Hessian
  - $c \cong 1 - 1/$"condition number"
- For quadratic problem: $\frac{1}{2} x^T P x + q^T x + r$
  - Condition number is $\lambda_{max}/\lambda_{min}$



- Gradient descent bad when condition number is large

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



Solving quadratic problems is easy

- Quadratic problem: $\min_x \frac{1}{2} x^T P x + q^T x + r$
- Solving equivalent to solving linear system:

$\nabla f = 0 \qquad \nabla f(x) = Px + q = 0 \quad \equiv \quad Px = -q$

  - If system has at least one solution: done!   many solutions
  - If system has no solutions: problem is unbounded

$Px = -q$ having no solutions

- Usually don't have simple quadratic problems, but…

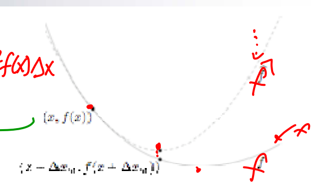
Newton's method

f

- Second order Taylor expansion:

$f(x+\Delta x) \approx \hat{f}(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x$

$\nabla^2 f(x) = \left[\frac{\partial f(x)}{\partial x_i \partial x_j} \right]$



$\hat{f}$ $\qquad f$

- Descent direction, solution to linear system

$\nabla \hat{f}(x+\Delta x) = 0 \quad \Rightarrow \quad \nabla^2 f(x) \Delta x = -\nabla f(x)$

solve for Δx_{nt}

- Nice property:
  - We wanted: $\nabla f(x)^T \Delta x < 0$ plug
  - We get: suppose $\nabla^2 f(x)$ invertible $\Rightarrow$ $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f$

$\Rightarrow \quad -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$ because $\nabla^2 f(x) \succ 0$ because convex

Newton's method – alg.

- Start from some x in **dom** f
- Repeat
  - Determine descent direction Δx_{nt}
    - solve system $\nabla^2 f(x) \Delta x = -\nabla f(x)$
  - **Line search** to choose step size t
  - Update: x ← x + t Δx_{nt}
- Until stopping criterion

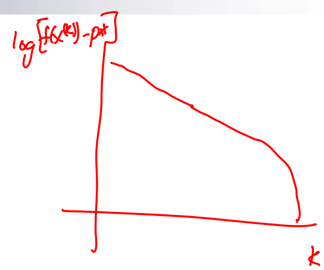
- Good stopping criterion:

$$\frac{1}{2}\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) \leq \epsilon \quad \text{“}\varepsilon\text{”}$$

$f(x+\Delta x) \geq f(x) + \nabla f(x) \Delta x$ by convexity

Convergence analysis for Newton's

- (Really see book for details.)



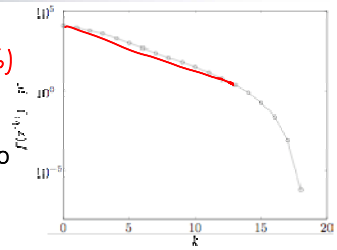
- Two phases:
  - Gradient is large $\|\nabla f(x)\|_2 \geq \eta$
    - Damped Newton Phase
      - Step size t<1
    - Linear convergence
  - Gradient is small $\|\nabla f(x)\|_2 < \eta$
    - Pure Newton Phase
      - Step size t=1
    - Quadratic convergence
      - c^(2^k)
    - Only lasts 6 steps

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

$g(x) = f(Ax+b)$

α, β



$n \; \hat{\nabla}^2 f(x)$

$\nabla^2 f(x)\, \Delta x = -\nabla f(x)$

Solving problems with equality constraints

- Equality constraints:

$\min_x f(x)$

$\text{s.t. } {}_m\hat{A}x = b$

$m < n$

- 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

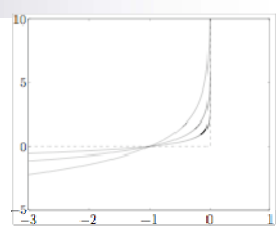
- Quadratic approximation:
- Start feasible, stay feasible:
- KKT:
- Solve linear system:
- Move accordingly:

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:

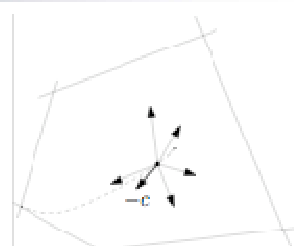
- Convex, if f_i are convex, because
  -

Log-barrier function

- Solve log-barrier problem with parameter t:

- Nice property:
  - Gradient:
  - Hessian:

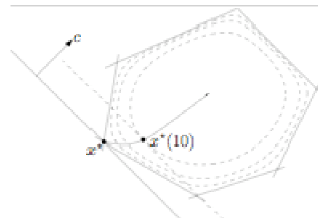
Force field interpretation



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

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

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