Today…

- Thus far, focused on formulating convex problems and gradient methods
  - Now: second order and interior point methods
  - Plan: 200 pages of book (Part III) in two lectures

- Focus:
  - Convex functions
  - Twice differentiable

- Overview
  - Unconstrained
  - Equality constraints
  - General convex constraints
Solving unconstrained problems

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

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 $\Delta x$ such that:
Generic descent algorithm

- Start from some \( x \) in \( \text{dom} \ f \)
- Repeat
  - Determine descent direction \( \Delta x \)
  - **Line search** to choose step size \( t \)
  - Update: \( x \leftarrow x + t \Delta x \)
- Until stopping criterion

Good stopping criterion:

- In gradient descent, \( \Delta x = \)

**Exact line search**

- Find best step size \( t \):

Problem is

- Sometimes easy to solve in closed form
- Other times can take a long time…
Backtracking line search

- From convexity, lower bound on $f(x+t\Delta x)$:
  - Can’t really hope to achieve ideal decrease of
  - Instead pick some $\alpha$
    - And achieve:

Choosing $t$:

---

Backtracking line search alg.

- Given
  - Point $x$
  - Descent direction $\Delta x$
  - $\alpha$
  - $\beta$
- $t=1$
- While $f(x+t\Delta x)>$
  - $t := \beta t$

- Boyd & Vandenberghe: pick
  - $\alpha$ in [0.01,0.3]
  - $\beta$ in [0.1,0.8]
Analysis of gradient descent

- (details in book and Geoff’s lecture earlier in semester…)
- 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
  - \( c \equiv 1 - 1/\text{condition number} \)
- For quadratic problem:
  - Condition number is \( \lambda_{\text{max}}/\lambda_{\text{min}} \)
- 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:
  - 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…

Newton’s method

- Second order Taylor expansion:
  - Descent direction, solution to linear system

- Nice property:
  - We wanted:
    - We get:
Newton’s method – alg.

- Start from some \( x \) in \( \text{dom } f \)
- Repeat
  - Determine descent direction \( \Delta x_{nt} \)
  - \textbf{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
      - \( c^n(2^n) \)
    - 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

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: