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
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 $f(x + \Delta x)$:
  - 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 $f(x + \Delta x) >$
    - $t := \beta t$

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

- (details in book...)
- 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
  - $c = 1 - 1/\text{condition number}$

- 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 $\Delta x_{nt}$
  - **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 $t \leq 1$
      - Linear convergence
  - Gradient is small
    - 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

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 \( x \)
  - Initial \( t > 0 \)
  - \( \mu > 1 \)
- 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 \( \leq \frac{k}{t} \)
  - Thus:
    - Stopping criterion \( \frac{k}{t} \leq \varepsilon \)

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