Solving (cont.) Convex Problems

Optimization - 10725

Carlos Guestrin
Carnegie Mellon University

April ~~9th~~ 14th, 2008

Solving problems with equality constraints

- Equality constraints:

$$\min_x f(x) \quad \text{s.t. } {}_m\hat{A}x = b$$

$$m < n$$

- Seems very hard

Null space

- Equality constraints: $Ax = b$
- Given one solution: $\hat{x}$: $\underline{A\hat{x} = b}$
  $\bar{x} = \hat{x} + y$
- Find other solutions: $A\bar{x} = b$
  $A(\hat{x}+y) = b \Rightarrow Ay = 0$ null space
  $y \in N(A)$
- Since Null Space is a linear subspace: columns of F are basis of N(A)
  $x \in \mathbb{R}^n$ $Ay = 0$ linear subspace
  $y = F z$ $z \in \mathbb{R}^d$ $d \leq n$
  $\{x \mid Ax = b\} = \{\hat{x} + Fz \mid z \in \mathbb{R}^d\}$ $d = \dim(N(A))$

Eliminating linear equalities

- Equivalent optimization problems:
  $\min_x f(x)$, $Ax = b$, $x \in \mathbb{R}^n$ $\Leftrightarrow$ $\min_z f(\hat{x} + Fz)$, $z \in \mathbb{R}^d$
  $\hat{x} \leftarrow$ *any* point in $Ax = b$
- Find basis for null space of A (linear algebra) $\leftarrow \hat{x}, F$
  - Solve unconstrained problem
- A concern… e.g., A has structure
  F may lose structure

Solving quadratic problems with equality constraints

- Quadratic problem with equality constraints:

$$\min_x \frac{1}{2} x^T P x + \xi^T x + r = f(x)$$

$$m < n \qquad \underset{m}{A} x = b$$

$v \leftarrow$ unconstrained

$$L(x, v) = \frac{1}{2} x^T P x + \xi^T x + r + v^T (Ax - b)$$

- KKT condition x* solution iff

feasibility

$$Ax = b$$

$$\nabla f(x^*) + A^T v^* = 0 \; ; \quad Px + \xi + A^T v = 0$$

- Rewriting: (x^*, v^*)

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -\xi \\ b \end{bmatrix}$$

- Solve linear system:
  - Any solution is OPT (x^*, v^*)
  - If no solution, unbounded

Newton's method with equality constraints

- Quadratic approximation:

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

$$A(x + \Delta x) = b$$

- Start feasible, stay feasible: $x^{(0)} \leftarrow$ feasible $\quad Ax^{(0)} = b$

$$x^{(1)} = x^0 + t \Delta x \Rightarrow Ax^{(1)} = b \; , \quad A \Delta x = 0$$

- KKT:
- Solve linear system:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

Lagrange multipliers for $A \Delta x = 0$

- Move accordingly:

line search

$$x \leftarrow x + t \, \Delta x_{nt}$$

General convex problem

- General (differentiable) convex problem:

$$\min_x f_0(x) \qquad \forall i = 1..k \qquad f_i \in \text{convex}$$
$$f_i(x) \le 0$$
$$Ax = b$$

- Equivalent problem with only equality constraints:

$$I_-(u) = \begin{cases} 0, & u \le 0 \\ \infty & u > 0 \end{cases}$$

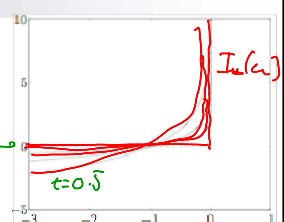
$$\min_x f_0(x) + \sum_i I_-(f_i(x))$$
$$Ax = b$$

not differentiable !!

Approximating the indicator

$$\frac{d \log(f(u))}{du} = \frac{f'}{f}$$

- Approximate indicator:
  - $I_-(u) \approx -\frac{1}{t}\log(-u) \qquad t > 0$
  - Correct as t $\to \infty$
  - Differentiable $\frac{\partial\left(-\frac{1}{t}\log(-u)\right)}{du} = \frac{-1}{tu}$



- Approximate optimization problem:

$$\min_x f_0(x) \qquad \approx \qquad \min_x f_0(x) - \frac{1}{t}\sum_i \log(-f_i(x))$$
$$f_i(x) \le 0 \qquad\qquad Ax = b$$
$$Ax = b$$

- Convex, if f_i are convex, because
  - $-\frac{1}{t}\log(-u)$ ← convex and increasing function of u

Log-barrier function

- Solve log-barrier problem with parameter t:

$$\min_x \; t f_0(x) + \phi(x) \quad ; \quad \phi(x) = -\sum_i \log(-f_i(x))$$
$$Ax = b$$

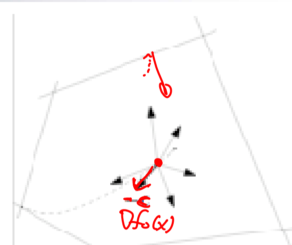
- Nice property:
  - Gradient: $\nabla(t f_0(x) + \phi(x)) = t \nabla f_0(x) + \nabla \phi(x)$
  - Hessian: $t \nabla^2 f(x) + \nabla^2 \phi(x)$

very simple

Force field interpretation

- Log-barrier function: $t f_0(x) + \phi(x)$
- Descending gradient of log barrier

$$-\left[t \nabla f_0(x) + \nabla \phi(x)\right]$$



- Each term: of $-\nabla \phi(x)$ $\quad \frac{\nabla f_i(x)}{f_i(x)}$
  - Want $f_i(x) \leq 0$
  - As we approach 0 :

as $f_i(x)$ tries to become ≥ 0

$\frac{\nabla f_i(x)}{f_i(x)}$ becomes a

$\frac{1}{f_i(x)} \to -\infty$; $-\infty \, \nabla f_i(x)$ make $f_i(x)$ smaller (more negative)

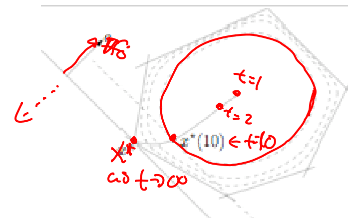
$$\phi(x) = -\sum_i \log(-f_i(x))$$

$$\nabla \phi(x) = \sum_i - \frac{\nabla f_i(x)}{f_i(x)}$$

Central path

- For each t, solve:

  $x^*(t) \in \arg\min_x t f_0(x) + \phi(x)$
  $Ax = b$



- 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
  - μ>1 update parameter
- Repeat
  - *Centering*: start from solution for previous t; don't solve to completion ε
    - Starting from x, compute: $x^*(t) = \arg\min_x t f_0(x) + \phi(x)$, $Ax = b$
  - *Update*: x:= $x^*(t)$
  - *Stopping criterion*: When t is “large enough”
  - *Increase barrier param*: t:= μt

When is t large enough???

- Solve centering step: $x^*(t) = \arg\min_x t f_0(x) + \phi(x)$, $Ax = b$
- There exists values for dual vars (See book), $(\lambda^*(t), \nu^*(t))$ such that duality gap ≤ k/t
  - change for diff t's
  - for original problem
  - k ← # of inequality constraints
- Thus: $f_0(x^*(t)) - p^* \le k/t$ (OPT)
- Stopping criterion k/t ≤ ε

$(\lambda^*(t), \nu^*(t))$ feasible in dual of; $x^*(t)$ feasible

$$\min_x f_0(x) \quad f_i(x) \le 0 \quad Ax = b$$

Centering step not (necessarily) exact

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

$$x^*(t) = \arg\min_x t f_0(x) + \phi(x), \quad Ax = b$$

- 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
    - $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T(Ax - b)$, $\lambda_i \ge 0$
    - See book for details

What about feasible starting point???

- Phase I: Solve feasibility problem, e.g.,

  $\min_{s,x} s$ $\quad f_i(x) \le s \;\; \forall i$
  $Ax = b$

  - Starting from feasible point:

    pick $\hat{x}$ $A\hat{x} = b$, set $\hat{s} = \max_i f_i(\hat{x})$

  - (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