ORIE 6300 Mathematical Programming I	November 3, 2016

Lecture 20

Lecturer: Damek Davis *Scribe: Pamela Badian-Pessot*

1 Recap

- We can view the simplex method as a nonsmooth equation solver.

2 Primal-dual Interior Point Method (IPM)

Reference Today's lecture is based on Jim Renegar's excellent textbook [**?**].
History:

- 1984 Karmarkar developed new *polynomial time algorithm* for linear programming
- First polynomial time algorithm called Ellipsoid method, developed in 1972. Proved to have polynomial complexity by Khachiyan in 1979.
- Ellipsoid method is very slow in practice. Much slower than simplex.
- Throughout the 1980s-1990s IPMs actively researched.
- We will study a simple primal-dual IPMs that often performs well in practice.

Idea:

- Given primal dual pair

$$\min\{c^T x | Ax = b,\ x \geq 0\}, \qquad \max\{b^t y | A^T y + s = c,\ s \geq 0\}$$

  form primal dual system

$$C_1 = \left\{ \begin{bmatrix} x \\ y \\ s \end{bmatrix} \middle| \begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \end{bmatrix} \begin{bmatrix} x \\ y \\ s \end{bmatrix} \right\}, \qquad C_2 = \left\{ \begin{bmatrix} x \\ y \\ s \end{bmatrix} \middle| \begin{bmatrix} x \\ s \end{bmatrix} \geq 0 \right\}$$

  together with the complementary slackness condition

$$x^T(c - A^T y) = c^T - b^T y = 0.$$

- Then realize that $x^T s = x^T(c - A^T y)$.

- IPMs solve a series of relaxed problems

$$(P_v) = \left\{ \begin{bmatrix} x \\ y \\ s \end{bmatrix} \in C_1 \cap C_2^\circ,\ x \odot s = v,\ v > 0 \right\}.$$

  depending on vectors $v \in \mathbb{R}^n_{>0}$ which tend to zero. Where $C_2^\circ = \text{int}(C_2)$ and $x \odot s := (x_i s_i)_{i=1}^n$, i.e., the componentwise product.

- In the limit, we get a solution.

Three Questions

1. When is there a solution to P_v?
2. How do we choose initial v and solve P_v?
3. Given v and a solution to P_v, how should we choose v_+ (the next v)? and can we easily update the solution of P_v to a solution of P_{v_+}?

2.1 Question 1

The answer to question 1 is *always.*

Define:

$$C = \left\{ (x, s) \middle| \exists y \text{ with } \begin{bmatrix} x \\ y \\ s \end{bmatrix} \in C_1 \cap C_2^\circ \right\}$$

Theorem 1 *The mapping*

$$F : C \to \mathbb{R}^m_{>0}$$
$$(x, s) \mapsto x \odot s$$

is a bijection.

The proof of this theorem relies on basic techniques in convex optimization, so we omit it.

Why does a solution always exist?

$$\text{Given } v, \text{ set } (x, s) = F^{-1}(v).$$

2.2 Question 2

- We choose

$$v = \mu e$$

  where $e = (1, \dots, 1)$ and $\mu > 0$. Then by the theorem, $\exists x(\mu), s(\mu)$,

$$x(\mu) \odot s(\mu) = \mu e.$$

Definition 1 (Central Path) *We call* $\{(x(\mu), s(\mu)) \mid \mu > 0\}$ *the central path.*

- It is typical to initialize IPMs on the central path.
- Why do this?
  - To get best computational complexity.
  - To only have one algorithm parameter μ.
  - To keep variables "balanced:" we want all variables to violate optimality conditions by the same amount.
- How do we find initial $(x(\mu), s(\mu))$?

  In practice, we can't find the points exactly, but we can assume we satisfy

$$||x \odot s - \mu e|| < const \cdot \mu.$$

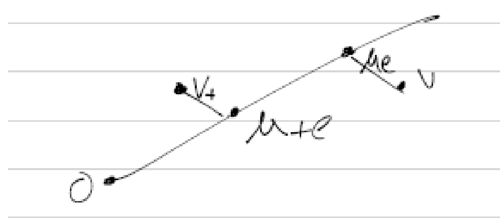
- This is typically achieved by inexactly solving another related optimization problem, which we won't dwell on here.
- This is similar to how simplex method requires solving an auxiliary LP to get an initial BFS.

2.3 Question 3

- Suppose have a solution to P_v such that $\|v - \mu e\| < const\mu$.
- We want to easily find a point v_+ so that

$$\|v_+ - \mu_+ e\| < const\mu_+$$

  where $\mu_+ < \mu$.



  and a solution to P_{v_+}.

- Let $v' = \mu_+ e$. Given a solution to P_v, called $[x, y, s]$, the best case is that we solve

$$v' = x' \odot s', \quad x' = x + \Delta x, \quad s' = s + \Delta s, \quad y' = y + \Delta y$$

$$x', s' \geq 0, \quad A\Delta x = 0, \quad A^T \Delta y + \Delta s = 0.$$

- The last two conditions guarantee that $Ax' = b$, $A^T y' + s' = c$.

- The first equation can be expanded

$$v' = (x + \Delta) \odot (s + \Delta s) = x \odot s + x \odot \Delta s + \Delta x \odot s + \Delta x \odot \Delta s.$$

I.e.

$$v' - v = x \odot \Delta s + \Delta x \odot s + \Delta x \odot \Delta s.$$

- Clearly x', y', s' solves $P_{v'}$ but this is too hard in general because of the quadratic coupling $\Delta x \odot \Delta s$.

- However, we CAN solve the first order approximation

$$\begin{aligned} x \odot \Delta s + \Delta x \odot s &= v' - v \\ A\Delta x &= 0 \\ A^T \Delta y + \Delta s &= 0 \end{aligned}$$

Then we set:

$$x_+ = x + \Delta x; \quad s_+ = s + \Delta s; \quad y_+ = y + \Delta y.$$

And let $v_+ = x_+ \odot s_+$.

- **Observation:** $v - v_+ = \Delta x \odot \Delta s$.

- Is (x_+, y_+, s_+) feasible? To answer this we need a function and a theorem.

Definition 2 *Define a function* $r : \mathbb{R}^n \to \mathbb{R}$ *by*

$$(\forall v \in \mathbb{R}^n) \qquad r(v) = \min\{v_1, \ldots, v_n\}.$$

Theorem 2 *If* $v' \in B(v, r(v))$*, then* (x_+, y_+, s_+) *is feasible and*

$$||v_+ - v'|| \leq \frac{\|v' - v\|^2}{2r(v)}.$$

Before we prove the theorem we indicate its use in algorithmic analysis.

Corollary 3 *If* $v' \in B(v, tr(v))$ *where* $t < 1$*, then* (x_+, y_+, s_+) *is feasible and*

$$v_+ \in B\left(v', \frac{1}{2}\frac{t^2}{1-t} r(v')\right).$$

In particular, if

(a) $||v - \mu e|| < \frac{1}{24}\mu$*; and*

(b) $\mu_+ = \left(1 - \frac{1}{8\sqrt{n}}\right)\mu$

then $||v_+ - \mu_+ e|| \leq \frac{1}{24}\mu_+$.

Proof: By the theorem

$$||v_+ - v'|| \leq \frac{||v'_+ - v||^2}{2r(v)} \leq \frac{t^2 r(v)}{2}.$$

On the other hand,

$$r(v') \geq r(v) - ||v - v'|| \geq (1-t)r(v)$$

by 1-Lipschitz continuity of r. So we can substitute $r(v) \leq \frac{r(v')}{1-t}$ above.

Exercise: Assume (a) and (b) and derive the bound $||v_+ - \mu_+ e||$. □

- Thus if v is near the central path, then v_+ is also near the central path! Therefore iterating and finding v_{++} makes sense.
- After k iterations, we have

$$||v_k - \mu_k e|| \leq \frac{1}{24}\mu_k \leq \frac{1}{24}\left(1 - \frac{1}{8\sqrt{n}}\right)\mu_{k-1} \leq \cdots \leq \frac{1}{24}\left(1 - \frac{1}{8\sqrt{n}}\right)^k \mu_0.$$

- Moreover,

$$\begin{aligned} c^T x - b^T y = x^T s &= \sum v_j \\ &= ||v||_1 \\ &\geq ||\mu e||_1 - ||v - \mu e||_1 \\ &\geq n\mu - \sqrt{n}||v - \mu e|| \\ &= n\left(1 - \frac{1}{24\sqrt{n}}\right)\mu \end{aligned}$$

  and similarly

$$c^T x_+ - b^T y_+ \leq n\left(1 + \frac{1}{24\sqrt{n}}\right)\mu_+.$$

Exercise: Show

$$\frac{c^T x_+ - b^Y +}{c^T x - b^T y} \leq 1 - \frac{1}{24\sqrt{n}}.$$

- Thus, the *primal-dual gap* is halved once ever $O(\sqrt{n})$ iterations.
- We will prove theorem 2 next time.
- Why is each iteration of this IPM more expensive than DRS and MAP?
  Because the linear system changes at every iteration of this method, while each iteration of DRS/MAP limited to matrix-vector multiplications if we precompute $D^\dagger$.
- Paper of possible interest [**?**]