6.854 Advanced Algorithms October 21, 2011

Optional Lecture 3: Interior Point Method

Lecturer: Zeyuan Allen Zhu *Scribe: Zeyuan Allen Zhu*

Disclaimer: These lecture notes have not been fully checked by Zeyuan.

1 References

This file of notes serves as a reference for Zeyuan himself about the materials to be delivered in class. It copies a lot of materials from Prof Michel X. Goemans' lecture notes on 6.854 in 1994, (see `http://www-math.mit.edu/~goemans/notes-lp.ps`), and Prof Sven O. Krumke's report on interior point methods (see `http://optimierung.mathematik.uni-kl.de/~krumke/Notes/interior-lecture.pdf`).

Ye's interior point algorithm achieves the best known asymptotic running time in the literature, and this presentation incorporates some simplifications made by Freund.

2 Introduction

- $O(n^6 L)$ complexity for ellipsoid, $O(\sqrt{n}L) \times O(n^3)$ complexity of interior point method.
- Only describe for LP today, and can be generalized to SDP and other convex programming.
- Primal / Dual:

$$(P) \begin{cases} \min & c^T x \\ s.t. & Ax = b, \\ & x \geq 0 \end{cases} \qquad\qquad (D) \begin{cases} \max & b^T y \\ s.t. & A^T y + s = c, \\ & s \geq 0 \end{cases}$$

- Assume here that A is of full row-rank, A is $m \times n$, with $m \leq n$.
- A primal-dual algorithm, keeping track of $(\bar{x}, \bar{s})$ such that $\bar{x} > 0$ and $\bar{s} > 0$ (i.e., $\exists \bar{y} : A^T \bar{y} = c - \bar{s}$).
- Stay away from the boundary of the polytope, and thus making the duality gap $c^T \bar{x} - b^T \bar{y} = \bar{x}^T \bar{s} > 0$.
- Define a potential function $G(x, s) := q \log(x^T s) - \sum_{j=1}^n \log(x_j s_j)$.
- Choose $q = n + \sqrt{n}$ in order to guarantee $O(\sqrt{n}L)$ iterations.

3 The high-level description

- Start from some feasible solution
  - The primal/dual may not always be feasible!
  - It may be hard to find a feasible solution!
  - The starting potential might be huge!
  - Please refer to `http://www-math.mit.edu/~goemans/18415/18415-FALL01/init-lp.ps`.

- Claim: can always turn the problem into a feasible primal/dual pair, with a starting potential of $O(\sqrt{n}L)$.

- At each iteration,
  - Scale from $(\bar{x}, \bar{s})$ to (e, s') using affine transformation (Q3).
  - Compute the gradient $g = \nabla_x G(x,s)|_{(e,s')}$ (which is actually the "same" as $\nabla_s G(x,s)|_{(e,s')}$). (Q4)
  - If the projection of g on the primal feasible subspace ($Ax = b$) is large, make a primal gradient descent step. (Q4)
  - Otherwise, make a dual gradient descent step. (Q4)
- Stop until the potential is smaller than $-k\sqrt{n}L$ for some constant k. (Q2)
- Only get an approximate solution so far, turn it into an exact one! (Q1)

4 Q1: What accuracy do we need?

Lemma 1. *Let x_1, x_2 be vertices of $Ax = b$ and $x \geq 0$. If $c^T x_1 \neq c^T x_2$, then $|c^T x_1 - c^T x_2| > 2^{-2L}$.*

Proof. Exists integers q_1, q_2 such that $1 \leq q_1, q_2 < 2^L$ and $q_1 x_1, q_2, x_2 \in \mathbb{N}$. Therefore:

$$|c^T x_1 - c^T x_2| = \left| \frac{q_1 q_2 (c^T x_1 - c^T x_2)}{q_1 q_2} \right| \geq \frac{1}{q_1 q_2} \quad .$$

□

Corollary 2. *Let OPT be the optimal answer. If x is a feasible solution to P with $c^T x \leq OPT + 2^{-2L}$, then any vertex x' such that $c^T x' \leq c^T x$ is an optimal solution of P.*

Proof.
- Suppose x' is not optimal, but some other vertex x^* is.
- We have $c^T x' - c^T x^* > 2^{-2L}$ according to the lemma.
- It gives $c^T x' > OPT + 2^{-2L} \geq c^T x \geq c^T x'$.

□

This corollary tells us that as long as our value is 2^{-2L} close to OPT, we are pretty safe: any vertex better than this would be a solution. But how to actually find one? One can adopt the following procedure:

- Let $P(x) := \{j : x_j > 0\}$.
- If the columns of A on indices $P(x)$ are linearly independent, then can fill it up to m columns (by adding $m - |P(x)|$ ones). This can be done since A has full row-rank, and will imply that x is already a vertex.
- Otherwise, the columns $\{a_i : i \in P(x)\}$ are linearly dependent: $\sum_{i \in P(x)} \lambda_i a_i = 0$.
- I claim that $x + \delta \cdot \lambda$ is still feasible for small enough $\delta > 0$.
- One of the two directions will not increase the objective of (P), and at the same time destroy some coordinate.
- Repeat.

A naive implementation takes time $O(n \times n^3)$ because we might iterate $O(n)$ times and each time do a Gaussian elimination. One can actually be clever and do the Gaussian elimination only once.

5 Q2: How is the potential related to the duality gap

For a generic choice of q, we are uncertain about its behavior when we are getting close to the optimal solution.

Lemma 3. *Let $x, s > 0$ be vectors in $\mathbb{R}^{n\times 1}$, then*

$$n \log x^T s - \sum_{j=1}^{n} \log x_j s_j \geq n \log n \ .$$

Proof. $\left(\prod_{j=1}^{n} x_j s_j\right)^{1/n} \leq \frac{1}{n}\left(\sum_{j=1}^{n} x_j s_j\right) \Rightarrow \frac{1}{n}\left(\sum_{j=1}^{n} \log x_j s_j\right) \leq \log\left(\sum_{j=1}^{n} x_j s_j\right) - \log n$ $\square$

- This means, we should choose some $q > n$ and this ensures $G \to -\infty$ as $x^T s \to 0$.
- $q = n + 1$ will result in $O(nL)$ number of iterations.
- $q = n + \sqrt{n}$ will result in $O(\sqrt{n}L)$ number of iterations.
- But when can we stop?

Lemma 4. *If $G(x, s) \leq -k\sqrt{n}L$ for some constant k, then $x^T s < e^{-kL}$.*

Proof. By the previous lemma, $-k\sqrt{n}L \geq \sqrt{n} \log x^T s + n \log n \Rightarrow \log x^T s \leq -kL - \sqrt{n} \log n < -kL$. $\square$

6 Q3: What is an affine transformation?

- Given $\bar{x} > 0$, we consider the affine scaling $x = (x_1, \ldots, x_n) \to x' = (\frac{x_1}{\bar{x}_1}, \ldots, \frac{x_n}{\bar{x}_n})$.
- By defining $X = \text{diag}\{\bar{x}_1, \ldots, \bar{x}_n\}$:

$$(P)\begin{cases} \min & c^T X x' \\ s.t. & AXx' = b, \\ & x' \geq 0 \end{cases} \qquad (D)\begin{cases} \max & b^T y \\ s.t. & (AX)^T y + Xs = Xc, \\ & Xs \geq 0 \end{cases}$$

- Also written as (if we define $c' = Xc, s' = Xs, A' = AX$):

$$(P)\begin{cases} \min & c'^T x' \\ s.t. & A'x' = b, \\ & x' \geq 0 \end{cases} \qquad (D)\begin{cases} \max & b^T y \\ s.t. & (A')^T y + s' = c', \\ & s' \geq 0 \end{cases}$$

- The above analysis implies that, starting from a feasible solution $(\bar{x}, \bar{s})$, one can change the (P) and (D) program, making (e, s') a feasible solution to the new program.
- Since $e = \bar{x}X^{-1}$ and $s' = \bar{s}X$, we have $\bar{x}_j \bar{s}_j = s'_j$, the duality gap $G(\bar{x}, \bar{s}) = G(e, s')$.
- In this transformed space, all points are far from the boundaries.

7 Q4: How to make a gradient descent step

7.1 Primal descent

- Starting from (e, s'), and recall $G(x,s) = q\log(x^T s) - \sum_{j=1}^n \log(x_j s_j)$.
-
$$g = \nabla_x G(s,x)|_{(e,s')} = \left(\frac{q}{x^T s}s - (1/x_1, \ldots, 1/x_n)\right)\Big|_{(e,s')} = \frac{q}{e^T s'}s' - e$$
- To maximize the change in G, we want to move in the direction of $-g$.
- However, we need to insure the new point is still feasible (i.e. $Ax = b$) still holds.
- Let d be the projection of g onto the null space $\{x : Ax = 0\}$.
-

  Claim 5. $d = (I - A^T(AA^T)^{-1}A)g$

  Proof. $g - d$ must be orthogonal to the null space of A, and thus being some combination of row vectors of A.
$$\begin{cases} Ad = 0, \\ \exists w \text{ s.t. } A^T w = g - d \end{cases} \Rightarrow \quad (AA^T)w = Ag$$
  Solving the equation we get $w = (AA^T)^{-1}Ag$ and thus $d = g - A^T(AA^T)^{-1}Ag = (I - A^T(AA^T)^{-1}A)g$. □

- If $\|d\|_2 = \sqrt{d^T d} \geq 0.4$, we make a primal step:
$$\begin{cases} \tilde{x} &= e - \frac{1}{4\|d\|}d \\ \tilde{s} &= s' \end{cases}$$
- Observe that $\tilde{x} > 0$, because $\tilde{x}_j = 1 - \frac{1}{4}\frac{d_j}{\|d\|} \geq 3/4 > 0$.
-

  Lemma 6. *When making a primal step,* $G(\tilde{x}, \tilde{s}) - G(e, s') \leq -\frac{7}{120}$.

7.2 Dual descent

- If $\|d\|_2 < 0.4$, we make a dual step. But let us compute the gradient first:
$$h = \nabla_s G(x,s)|_{(e,s')} = \frac{q}{e^T s'}e - (1/s'_1, 1/s'_2, \ldots, 1/s'_n)$$
- Notice that $h_j = g_j/s_j$, and thus h and g can be "seen" to be approximately in the same direction.
- Need to maintain the feasibility of (D): $A^T y + s = c$. We want:
$$\begin{cases} A^T\tilde{y} + \tilde{s} &= c \\ \Rightarrow \tilde{s} - s' &= A^T(\tilde{y} - y') \end{cases}$$
- This means, the change $\tilde{s} - s'$ should be in the image space of A^T (and thus perpendicular to the null space). Let us move in the direction of $-(g - d)$.

- Recall that $(g-d) = A^T w = A^T(AA^T)^{-1}Ag$, we choose $\tilde{y} = y' + \mu w$ and $\tilde{s} = s' - (g-d)\mu$.
- We choose some magic number $\mu = \frac{e^T s'}{q}$, and:

$$\begin{cases} \tilde{s} &= s' - \frac{e^T s'}{q}(g-d) \\ &= s' - \frac{e^T s'}{q}(q\frac{s'}{e^T s'} - e - d) \\ &= \frac{e^T s'}{q}(d+e) \\ \tilde{x} &= e \end{cases}$$

- Similarly, we have $\tilde{s} > 0$ (using $\|d\| < 0.4$).
-

  Lemma 7. *When making a dual step, $G(\tilde{x}, \tilde{s}) - G(e, s') \le -\frac{1}{6}$.*

7.3 Missing proofs

Fact 8.

$$\forall |x| \le a < 1, \qquad -x - \frac{x^2}{2(1-a)} \le \log(1-x) \le -x$$

Proof of Lemma 6.

$$\begin{aligned} G(\tilde{x}, \tilde{s}) - G(e, s') &= G(e - \frac{1}{4\|d\|}d, s') - G(e, s') \\ &= q\log\left(e^T s' - \frac{d^T s'}{4\|d\|}\right) - \sum_{j=1}^{n}\log\left(1 - \frac{d_j}{4\|d\|}\right) - \sum_{j=1}^{n}\log s'_j \\ &\quad -q\log(e^T s') + \sum_{j=1}^{n}\log 1 + \sum_{j=1}^{n}\log s'_j \\ &= q\log\left(1 - \frac{d^T s'}{4\|d\| e^T s'}\right) - \sum_{j=1}^{n}\log\left(1 - \frac{d_j}{4\|d\|}\right) \qquad \text{(using the above fact for } a = 1/4) \\ &\le -\frac{q d^T s'}{4\|d\| e^T s'} + \sum_{j=1}^{n}\frac{d_j}{4\|d\|} + \sum_{j=1}^{n}\frac{d_j^2}{16\|d\|^2 2(3/4)} \\ &= -\frac{q d^T s'}{4\|d\| e^T s'} + \frac{e^T d}{4\|d\|} + \frac{1}{24} \\ &= \frac{1}{4\|d\|}\left(e - \frac{q}{e^T s'}s'\right)^T d + \frac{1}{24} \\ &= \frac{1}{4\|d\|}(-g)^T d + \frac{1}{24} = -\frac{\|d\|^2}{4\|d\|} + \frac{1}{24} \le -\frac{1}{10} + \frac{1}{24} = -\frac{7}{120} \end{aligned}$$

Notice that, this is nothing but saying that, the size of the descent should be proportional to the direction d dot product the gradient g, if we ignore the second order derivative. ☐

Proof of Lemma 7. Recall that $\tilde{s} = \frac{e^T s'}{q}(d+e)$. We first compute that

$$\begin{aligned}
\sum_{j=1}^{n} \log(\tilde{s}_j) - n\log\big(\frac{e^T\tilde{s}}{n}\big) &= \sum_{j=1}^{n}\log(1+d_j) - n\log(1+\frac{e^T d}{n}) \\
&\geq \sum_{j=1}^{n}(d_j - \frac{d_j^2}{2(3/5)}) - n\frac{e^T d}{n} \\
&= -\frac{\|d\|^2}{6/5} \geq -\frac{2}{15}
\end{aligned}$$

On the other hand, we have $\sum_{j=1}^{n}\log(\tilde{s}_j) - n\log\big(\frac{e^T\tilde{s}}{n}\big) \leq 0$ by Jensen's inequality. Now we can start to compute:

$$\begin{aligned}
G(e,\tilde{s}) - G(e,s') &= q\log\big(\frac{e^T\tilde{s}}{e^T s'}\big) - \sum_{j=1}^{n}\log(\tilde{s}_j) + \sum_{j=1}^{n}\log(s'_j) \\
&\leq q\log\big(\frac{e^T\tilde{s}}{e^T s'}\big) + \frac{2}{15} - n\log\big(\frac{e^T\tilde{s}}{n}\big) + n\log\big(\frac{e^T s'}{n}\big) \qquad (\text{recall } q = n+\sqrt{n}) \\
&= \frac{2}{15} + \sqrt{n}\log\big(\frac{e^T\tilde{s}}{e^T s'}\big) = \frac{2}{15} + \sqrt{n}\log\big(\frac{1}{q}(n+e^T d)\big) \\
&\leq \frac{2}{15} + \sqrt{n}\log\left(\frac{1}{n+\sqrt{n}}(n+0.4\sqrt{n})\right) \qquad (\text{using } |e^T d| \leq \|e\|\|d\| = \sqrt{n}\|d\|) \\
&\leq \frac{2}{15} + \sqrt{n}\log\left(1 - \frac{0.6\sqrt{n}}{n+\sqrt{n}}\right) \leq \frac{2}{15} - \frac{0.6n}{n+\sqrt{n}} \leq \frac{2}{15} - \frac{3}{10} = -\frac{1}{6}
\end{aligned}$$

□

8 Final analysis

- We have $O(\sqrt{n}L)$ iterations and each having a $O(n^3)$ time Gaussian elimination.
- Wait! Operations are not atomic. For instance,
- The two norm $\|d\|$ involves irrational operations
  - This can be taken care by the fact that, we can change from 0.4 to 0.399 and all the results still go through.
  - Therefore, we only need to compute d to the first a few bits.
- The complexity for Gaussian elimination requires $O(n^3)$ arithmetic operations (each arithmetic operation costs $O(L)$ in fact).
- Wait! Does the gradient descent steps increase the bit complexity?
- Yes, but we can round things down to 2^L, without affecting the duality gap too much.
- Yinyu Ye has an $O(n^3L)$-algorithm in his seminal paper `http://www.springerlink.com/content/n577730515780778/fulltext.pdf`.