CS294 Algorithms Toolkit **Fall 2008**

Lecture 6 Linear programming (Interior Point.). : 2.19.08

Lecturer: Satish Rao *Scribe: Rao, for now*

Disclaimer: *These are rough notes, with some exercises.*

Question: Shall we review the previous day's algorithm?

We have the following linear program.

$$yA + s = c$$

$$\max by$$

and

$$\min cx$$

$$Ax = b$$

The idea was to iteratively reduce $\sum_i x_i s_i = cx - by$, while maintaining primal and dual feasible solutions.

Question: The first thing isa potential function.

This is the barrier function potential.

$$G(x, s) = q \ln x^T s - \sum_i \ln x_i s_i.$$

Goes to negative infinity as the complementary slackness goes to 0, as long as the $x_i s_i$ products go to 0 in a relatively uniform fashion.

Question: What is the error with respect to $G(x, s)$?

It is easy enough to verify that when $G(x, s) \leq k\sqrt{n}$ that $\sum_i x_i s_i \leq e^{-k}$, when $q = n + \sqrt{n}$.

Question: Does this have a history?

This has been called the central path.

$$F(x, s) = t \ln x^T s - \sum_i \ln x_i.$$

Stay away from borders of the primal polytope (the analytic center of the polytope.) In that case, t is generally increased as one gets closer and closer to a solution. In the algorithm presented last time, one doesn't have to do this.

Question: Other approaches?

Don't stay feasible, and introduce penalties for that. Actually that is what dual variables are.

Question: Another idea?

Affine transformation to make current iterate $x = e$. Does not affect potential function. Moreover, the LP's can easily be translated into new a pair of LPs.

Still, why? Perhaps we'll see.

Question: Note that primal and feasible updates are orthogonal!

That is, given feasible solution x, s. We get that any update d should have $Ad = 0$; since

$$A(x + d) = Ax + Ad = b$$

and $Ax = b$.

Moreover any update f to s should be of the form a linear combination of rows since

$$yA + s = c$$

and

$$y'A + s + f = c$$

and the difference is

$$(y - y')A = f.$$

Since f is a linear combination of the rows of A it must therefore be orthogonal to any d in the (right) nullspace of A.

Thus, if there is some direction that improves the potential function, then it is the sum of a possible primal update and a possible dual update.

We note that changing the primal variables will have the benefit indicated by the gradient but changing the duals will have a different benefit. Something to consider.

Question: Given a vector, how do we compute the primal and dual directions?

Given g. $g = d + f$. $Ad = 0$ and thereis w such that $wA = f$.

So, the $Ag = Af = AA^T w$. And $w = (AA^T)^{-1} Ag$. And

$$f = A^T (AA^T)^{-1} g.$$

And, the primal $d = g - f$, so we have

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

Question: What should the direction of improvement be?

Take the derivative of $G(e, s)$ (by the way $x = e$.)

$$g = q\frac{s}{e^t s} - e.$$

(Without the affine transformation, what happens? Well each direction is scaled by $\bar{x}_i$ and $1/x_i$ is subtracted. Not sure what this effect is? Definitely seems to change things.)

We can examine this and see that it is not a tiny number, when q is large enough. For example, if all the s_i's are equal, we get a vector with $(q-n)/n$ in each coordinate. For $q = \sqrt{n}$, this is of size 1.

Another example is when one s_i is small. In this case, we get a large (negative) value in this component. Which suggests that we should raise x_i in this component to get the primal-dual pair of solutions back onto the "central" path.

Question: If constant improvent in d direction?

We update in the direction $-d/4||d||$, which leaves all the x_i's positive.

The reduction in norm is the rate of reduction that you get in this direction (after subtracting the maximum change in the rate of the reduction) times the distance; since we don't move much the rate of reducton does not change significantly and we get a constant reduction in $G(x,s)$.

Question: Otherwise, we move s.

In this case, we move it by $\mu(g-d)$, which is

$$\tilde{s} = s' - \mu(g-d).$$

Where

$$\mu = \frac{e^T s'}{q}.$$

Here, we get

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

Now, the potential function decreases from here by two observations (made quantitative.) The penalty from the s_i's getting smaller is not too bad (since each entry in d is not too large, i.e., 2/5.)

Moreover, the duality gap decreases by some factor $(1+1/\sqrt{n})$. This more or less follows from the fact that new s' are the unit vector (since d is small) scaled by $e^T s'/q$ which has norm at most $||e^T s'||(n/q) \leq (n/(n+\sqrt{n})$.

Question: Now, the multiplicative updates algorithm "solves" linear programs (at least of a certain form. How does it compare?

The bounds one gets is as follows.

$$C^* \leq (1+\epsilon)C_{opt} + \log n/\epsilon T.$$

But, one does no linear algebra. What is the iterations to get delta of optimal?

$$(\epsilon)C_{opt} + \log n/\epsilon T \leq \delta.$$

Here, we get

$$\epsilon \leq \delta/C_{opt}$$

and also that,

$$\delta \leq \delta \log n/C_{opt}T$$

Thus, $T \geq 1/\delta^2$. Versus, $T = \log(1/\delta)$ for interior point.

Question: Why so different?