**CS270 Algorithms & Data Structures** **Spring 2008**

# Lecture 14 Ye's Interior Point Algorithm. : 3.06.08

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

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

Note: this is blatant borrowing (cribbing) from Goeman's notes linked on the website. (As were parts of previous lectures on linear programming.)

**Question: Can you give a summary of the interior point algorithm?**

Sure. We wish to solve the following primal-dual pair of linear programs.

$$minz = cx$$

$$AX = b$$
$$x \geq 0$$

and its dual

$$maxw = by$$
$$A^T y + s = c$$
$$s \geq 0$$

We will simultaneously solve both. That is, it is a primal-dual algorithm. We will do so by maintaining a pair of solutions $(x, s)$ that are interior to the polytope, and trying to drive it towards optimal. The duality gap is $x^T s$. A primal-dual optimal pair has duality gap 0.

**Question: What is the potential function that we wish to drive down?**

$$G(x, s) = q \ln(x^T s) - \sum_{j=1}^{n} \ln(x_j s_j)$$

The first goes to $-\infty$ as the duality gap goes to 0, the second term keeps each value $x_i, s_j$ greater than 0.

We neglected to prove the following statement about this potential function. (But for an exercise.)

Claim:

$$n \ln(x^T s) - \sum_{j=1}^{n} \ln(x_j s_j) \geq n \ln n$$

**Question: What's the proof?**

The geometric mean is less than the arithmetic mean,

$$(\pi_i^n t_j)^{1/n} \leq \frac{1}{n}\left(\sum_i^n t_j\right).$$

Taking logs, and (re)arranging we get the Claim.

**Question: Recall a small potential function gives a small duality gap. How small?**

**Lemma 14.1** *When $G(x,s) < k\sqrt{N}L$ for feasible $x, s$, then*

$$x^T s < e^{-kL}$$

Proof:

$$-k\sqrt{n}L = (n+\sqrt{n})\ln x^t s = \sum_I^n \ln x_j s_j$$

From the claim above, we get that this is larger than $\sqrt{n}\ln x^T s + n\ln n$. Then, we get

$$\ln x^T s \leq -kL - \sqrt{n}\ln n.$$

**Question: Recall the algorithm?**

Sure, at iteration $i$, we use the affine transformation on $x$ to compute the gradient with respect to $x$, $g$, We then either move $x$ in direction $d$: which is $g$ projected onto the nullspace of $A$ to maintain primal feasibility, since any such vector does not change $Ad$.

If this leads to a small vector, we move the dual slack variables in the direction $g - d$, which is orthogonal to the nullspace of $A$, which means that there is a value of $y'$ such $A(y - y') = \delta s$. (This is slightly odd since this is not necessarily in the gradient direction of $s$, but it is related up to a pointwise division by $s_j$.)

In short, we get the following rules.

Let

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

And

$$d = (I - \bar{A}(\bar{A}\bar{A}^T)^{-1}\bar{A})g,$$

the projection onto the nullspace.

If $||d|| > .4$. then

$$\tilde{x} = e - \frac{1}{4||d||}d.$$

Otherwise,

$$\tilde{s} = \frac{e^T s'}{q}(d+e).$$

## Exercise 1: Show that $\tilde{x}, \tilde{s} > 0$.

**Question: Before continuing shall we get some intuition?**

Ok, I guess.

**Question: For the primal step, what is some intution?**

The derivative is basically suggesting we decrease some $x_i$'s to drive down the $\ln x^T s$ term and decrease some others to keep from driving up the $-\ln x_i s_i$ terms.

For some intuition, we think of the example that all the $s_i$ values are the same. Since $q > n$, the decrease will tend to be more important than the increase. (We wish to go in the $-g$ direction.)

Moreover, you can go in that direction and remain feasible. So, we fix that as necessary. If we remain good, then we go in that direction so that we don't violate any $x > 0$ constraint.

The proof that you improve in this case is that you should get at least the projection along the gradient per unit travel. So, we need to show that the gradient is large.

But, then too, the gradient changes as you go. So, we examine the log function and see that it does not change too much if we do not move too far. And thus, we get some improvement.

**Question: Alas, shouldnt we do a bit of math?**

Ok. (From Goemans...)

$$\begin{aligned}
G(\tilde{x},\tilde{s}) - G(e,s') &= G(e - \frac{1}{4||d||}d, \tilde{s}) - G(e,s') \\
&= qln\left(e^T - \frac{d^T s'}{4||d||}\right) - \sum_i^n \ln\left(1 - \frac{d_j}{4||d||}\right) - \sum_j^n \ln s'_j \\
&\quad -q\ln(e^T s') + \sum_j^n \ln 1 + \sum_j^n \ln s'_j \\
&= -q\ln\left(1 - \frac{d^T s'}{4||d|| e^T s'}\right) - \sum_j^n \ln\left(1 - \frac{d_j}{4||d||}\right)
\end{aligned}$$

Notice that all the arguments of the logs are pretty close to 1 at this point. This, due to the limited moves we make (i.e., only 1/4 of the direction.) This ensures that the benefit at the beginning doesn't change too drastically.

In particular, we proceed from the inequality on ln.

$$-x - \frac{x^2}{2(1-a)} \leq \ln(1-x) \leq -x, \text{ when } |x| \leq a < 1 \tag{14.1}$$

$$\begin{aligned}
\Delta G &\leq -\frac{qd^Ts'}{4||d||e^Ts'} + \sum_j^n \frac{d_j}{4||d||} + \sum_j^n \frac{d_j^2}{16||d||^2 2(3/4)} \text{ for } a = 1/4 \\
&= -\frac{qd^Ts'}{4||d||e^Ts'} + \frac{e^Td}{4||d||} + \frac{1}{24} \\
&= \frac{1}{4||d||}(e - \frac{q}{e^Tx'}s')^Td + \frac{1}{24} \\
&= \frac{1}{4||d||}(-g)^Td + \frac{1}{24} \\
&= -\frac{||d||^2}{4||d||} + \frac{1}{24} \text{ projection of } g \text{ onto } d \\
&= -\frac{||d||}{4} + \frac{1}{24} \\
&= -\frac{1}{10} + \frac{1}{24} = \frac{-7}{120}.
\end{aligned}$$

Ok..onto the dual.

**Question: For the dual move, what is the intuition?**

Let's look at the pieces of the proof, and discuss.

**Lemma 14.2**

$$\sum_j^n \ln(\tilde{s_j}) - n\ln(\frac{e^T\tilde{s}}{n}) \geq \frac{-2}{15}$$

Proof: $\tilde{s} = \frac{\Delta}{q}(e+d)$ and equation 14.1 about logs.

$$\begin{aligned}
\sum_j^n \ln(\tilde{s_j}) - n\ln(\frac{e^T\tilde{s}}{n}) &= \sum_j^n \ln(\frac{\Delta}{q}(1+d_j)) - n\ln(\frac{\Delta}{q}(1+\frac{e^Td}{n}) \\
&\geq \sum_j^n (d_j - \frac{d_j^2}{2(3/5)}) - n\frac{e^Td}{n} \\
&\geq -\frac{||d||^2}{6/5} \\
&\geq -\frac{2}{15}
\end{aligned}$$

I guess this says that the new $\tilde{s}$ does not have some coordinate that gets too too small relative to the average smallness of a coordinate.

**Question: Can we finish please?**

We have the following inequality from the concavity of the log function,

$$\sum_{j}^{n} \ln(s_j) \leq n \ln(\frac{e^T s}{n})).$$

Now, we start bounding...

$$\begin{aligned}
G(e, \tilde{s}) - G(e, s') &= q \ln(\frac{e^T \tilde{s}}{e^T s'}) - \sum_{j}^{n} \ln(\tilde{s}_j) + \sum_{j}^{n} \ln(s'_j) \\
&= q \ln(\frac{e^T \tilde{s}}{e^T s'}) + \frac{2}{15} - n \ln(\frac{e^T \tilde{s}}{n} + n \ln(\frac{e^T s'}{n}) \\
&- \frac{2}{15} + \sqrt{n} \ln(\frac{e^T \tilde{s}}{e^T s'}
\end{aligned}$$

This states that the potential function reduces by $\sqrt{n}$ times the reduction in norm of the slack variables (in the transformed space.)

We will proceed to show that the reduction in norm is $\Omega(1/\sqrt{n})$.

So,

$$e^T \tilde{s} = \frac{\Delta}{q}(n + e^T d)$$

and $\Delta = e^T s'$, so

$$\frac{e^T \tilde{s}}{e^T s'} = \frac{1}{q}(n + e^T d) \leq \frac{1}{n + \sqrt{n}}(n + .4\sqrt{n}).$$

The last follows since by Cauchy-Schwartz $|e^T d| \leq ||e|| ||d|| = \sqrt{n}||d||$.

Thus, we have

$$\begin{aligned}
G(e, \tilde{s}) - G(e, s') &\leq \frac{2}{15} + \sqrt{n} \ln(1 + \frac{0.6\sqrt{n}}{n + \sqrt{n}}) \\
&\leq \frac{2}{15} - \frac{.6\sqrt{n}}{n + \sqrt{n}}. \\
&\leq \frac{2}{15} - \frac{3}{10} = -1/6
\end{aligned}$$

The last follows since $n + \sqrt{n} < 2n$.

**Question: What is the intuition again?**

Hmmm. The progress a primal step is essentially that dictated by the gradient (i.e., the gradient value projected onto the feasible subspace.) One needs to move carefully to ensure the inital gradient direction is not misinforming you as you move away from the initial point.

Progress in a dual step really starts by using duality. That is, that the gradient direction that is not in the feasible primal space, is in the feasible dual direction. This follows from the definition of the dual. Now, the gradient before was for changes in the $x$ variables, rather than the $s$ variables. But the change in the $x_j$ variable is proportional to a change in the $s_j$ variable with constant $1/s_j$. Since we could (conceivably) reduce the $s_j$ variable by as much as $O(s_j)$. The changes in potential could be related to the change in $x_j$'s which is ruled by the value of $g - d$.

The idea that $g$ is reasonably good perhaps follows from the notion that one should be able to reduce an average $s_j x_j$ value by a good deal (say $x$) which than makes the first term go down by $qx$ and the second go up by $nx$ (with the same scaling).

The analysis gets a bit more, that is obtaining benefit from $1/\sqrt{n}$ of the total potential instead of perhaps only one coordinate.

## Exercise 2: If $q$ is set to $n + n^{3/4}$ what changes in the analysis? Hint: the dual step does better, but what must the total potential function get down to?