CSE 599: Interplay between Convex Optimization and Geometry **Winter 2018**

Lecture 17: Lee-Sidford Barrier

Lecturer: Yin Tat Lee

Disclaimer: *Please tell me any mistake you noticed.*

In this lecture, we talk about the Lee-Sidford barrier (apparently, I have no talent in naming). My original motive to study linear programming and convex optimization is to solve the maximum flow problem faster. When we use the standard logarithmic barrier function to solve the maximum flow problem, we get a $O^*(\sqrt{m})$ iterations algorithm where m is the number of edges. Each iteration involves solving a Laplacian system and hence takes nearly linear time. Therefore, we get a $O^*(m^{1.5})$ time algorithm, matching the previous best algorithm by Goldberg and Rao [91].

Problem 17.0.2. What is the relation between interior point methods and Goldberg and Rao algorithm?

Last lecture, we showed that in general one can get an $O^*(\sqrt{n})$ iterations algorithm for linear programs and hence naturally one conjectured that a $O^*(m\sqrt{n})$ time algorithm for the maximum flow problem is possible where n is the number of vertices. To get the $O^*(m\sqrt{n})$ time algorithm, we need to overcome the following hurdles:

1. Design a $O^*(n)$-self concordant barrier that can be computed very efficiently.
2. Modify the algorithm to work on the linear program of the form
$$\min_{B^\top f=d, 0\leq f\leq u} c^\top x.$$

Note that this linear program does not have $O^*(n)$-self concordant barrier. However, this is not a proof that we cannot design an $O^*(\sqrt{n})$ iterations algorithm. In particular, if there is no the upper constraint $f \leq u$, the dual linear program has n variables and hence we can run the interior point method on the dual. With the upper constraint, what we did is simply move the dual algorithm to primal and modified it such that it works with the upper constraint.

For simplicity, we will only focus on getting a $O^*(n)$-self concordant barrier that can be computed pretty efficiently.

17.1 Problem of Logarithmic barrier

Before we talk about our barrier, let us understand the problem of the logarithm barrier. It is known that interior point methods on logarithmic barrier can take $O(\sqrt{m})$ even if m is exponential to n. (See for example [92]). One reason is that we can perturb the barrier arbitrary by repeating constraints. Suppose we repeat the i^{th} constraint w_i many times, essentially our barrier function becomes

$$-\sum w_i \ln(a_i^\top x - b_i).$$

By choosing w_i appropriately, one can make the central path nearly every vertices of cube like the following

For these polytopes, the easiest way to fix it is to remove redundant constraints. However, when there are many constraints with similar but not identical direction, then it is less clear how to handle.

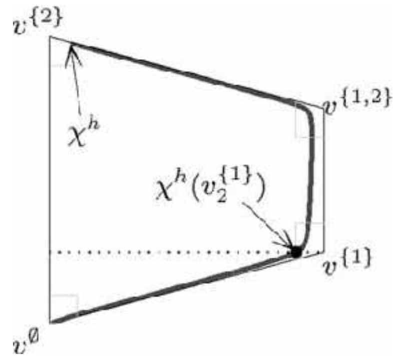

Figure 17.1: Copied from [92]. Please ignore the labels.

17.2 Volumetric Barrier

Vaidya proposed the volumetric barrier [94]. The self-concordance of this barrier has a better dependence on m compared to the log-barrier. The volumetric barrier is

$$\phi_2(x) = \ln\det(A^\top S^{-2} A)$$

where $s = a_i^\top x - b_i$ and $S = \mathrm{diag}(s)$. By simple calculations, we have the following:

Lemma 17.2.1. *We have that*

$$\nabla \phi_2(x) = A_x^\top \sigma_x,$$
$$\nabla^2 \phi_2(x) = A_x^\top (6\Sigma_x - 4P_x^{(2)}))A_x.$$

where $A_x = S^{-1}A$, $P_x = A_x^\top (A_x^\top A_x)^{-1} A_x$, $P_x^{(2)}$ is the Schur product of P, σ_x is the diagonal of P_x as a vector, and Σ_x is the diagonal of P_x as a diagonal matrix. Furthermore, we have that

$$2A_x^\top \Sigma_x A_x \preceq \nabla^2 \phi_2(x) \preceq 6A_x^\top \Sigma_x A_x.$$

Intuitively, one can think

$$\phi_2(x) \sim -\sum_i \sigma_{x,i} \ln(a_i^\top x - b_i).$$

Instead of a formal proof, we show how to estimate the self-concordance of ϕ_2:

Lemma 17.2.2. *Consider $\phi(x) = -\sum_i w_i \ln(a_i^\top x - b_i)$. Then,*

$$\nabla \phi(x)^\top (\nabla^2 \phi(x))^{-1} \nabla \phi(x) \leq \sum_i w_i$$

and that

$$D^3\phi(x)[h,h,h] \leq 2 \max_i \left| \frac{\sigma_i(\sqrt{W}A_x)}{w_i} \right| \left(D^2\phi(x)[h,h]\right)^{3/2}.$$

where $\sigma_i(\sqrt{W}A_x)$ is the leverage score of $A_x\sqrt{W}$, namely $\sigma_i(\sqrt{W}A_x) = (\sqrt{W}A_x(A_x^\top W A_x)^{-1} A_x^\top \sqrt{W})_{ii}$.

Remark. Geometrically, $\frac{\sigma_i(\sqrt{W}A_x)}{w_i} = \max_{\|h\|_x \leq 1} \left| \frac{a_i^\top h}{s_i} \right|$, the distance from x to the boundary of i^{th} constraint.

Proof. Note that $\nabla \phi(x) = A_x^\top w$ and $\nabla^2 \phi(x) = A_x^\top W A_x$. Hence, we have

$$\nabla \phi(x)^\top (\nabla^2 \phi(x))^{-1} \nabla \phi(x) = w^\top A_x (A_x^\top W A_x)^{-1} A_x^\top w \leq \sqrt{w}^\top \sqrt{w} = \sum_i w_i$$

where we used that $\sqrt{W}A_x(A_x^\top W A_x)^{-1}A_x^\top\sqrt{W}$ is an orthogonal projection matrix.

For the second condition, we have

$$D^3\phi(x)[h,h,h] = -2\sum_i w_i(\frac{a_i^\top h}{s_i})^3 \le 2\cdot\sum_i w_i(\frac{a_i^\top h}{s_i})^2\cdot\max_i\left|\frac{a_i^\top h}{s_i}\right|.$$

Note that

$$\left|\frac{a_i^\top h}{s_i}\right| = \left|\frac{a_i^\top(A_x^\top W A_x)^{-\frac{1}{2}}(A_x^\top W A_x)^{\frac{1}{2}}h}{s_i}\right| \le \left(A_x(A_x^\top W A_x)^{-1}A_x^\top\right)_{ii}\|h\|_x.$$

Hence, we have

$$D^3\phi(x)[h,h,h] \le 2\max_i\left|\frac{\sigma_i(\sqrt{W}A_x)}{w_i}\right|\|h\|_x^3$$

$\square$

By rescaling the ϕ by constant, we obtain a barrier with the "self-concordance" at x

$$\sum_i w_i\cdot\max_i\left|\frac{\sigma_i(\sqrt{W}A_x)}{w_i}\right|.$$

Now, we analyze the case $w_i = \sigma_i(A_x)$.

Lemma 17.2.3. *We have that*

$$\sum_i\sigma_i(A_x)\cdot\max_i\left|\frac{\sigma_i(\sqrt{\Sigma_i(A_x)}A_x)}{\sigma_i(A_x)}\right| \le 2n\sqrt{m}.$$

Proof. By the property of leverage score, we have that $\sum_i\sigma_i(A_x)\le n$.

We bound the maximum term in two ways. First, we note that

$$\frac{\sigma_i(\sqrt{\Sigma_i(A_x)}A_x)}{\sigma_i(A_x)} \le \frac{1}{\sigma_i(A_x)}.$$

Second, we have that

$$\mathrm{tr}((A_x^\top A_x)^{-1}A_x^\top\mathrm{diag}(1_{\sigma\le\frac{1}{2m}})A_x) \le \sum_{\sigma_i\le\frac{1}{2m}}\sigma_i \le \frac{1}{2}$$

which implies that

$$A_x^\top\mathrm{diag}(1_{\sigma\le\frac{1}{2m}})A_x \preceq \frac{1}{2}A_x^\top A_x.$$

Therefore, we have

$$A_x^\top A_x \preceq 2A_x^\top\mathrm{diag}(1_{\sigma\ge\frac{1}{2m}})A_x \preceq 4m\cdot A_x^\top\Sigma A_x.$$

Hence, we have

$$\begin{aligned}\sigma_i(\sqrt{\Sigma_i}A_x) &\le (\sqrt{\Sigma}A_x(A_x^\top\Sigma A_x)^{-1}A_x^\top\sqrt{\Sigma})_{ii}\\ &\le 4m(\sqrt{\Sigma}A_x(A_x^\top A_x)^{-1}A_x^\top\sqrt{\Sigma})_{ii}\\ &= 4m\sigma_i(A_x)^2.\end{aligned}$$

Combining both cases, we have

$$\frac{\sigma_i(\sqrt{\Sigma_i(A_x)}A_x)}{\sigma_i(A_x)} \le \min(\frac{1}{\sigma_i(A_x)}, 4m\sigma_i(A_x)) \le 2\sqrt{m}.$$

$\square$

Formally, the weighing depends on x and hence one need to compute the derivatives formally. By some calculation, we indeed can prove that ϕ_2 has self concordance $O(n\sqrt{m})$. Note that this barrier has less dependence on m. In the same paper, Vaidya consider the barrier $\phi_2 - \frac{n}{m}\sum \ln s_i$ and proved that it has self-concordance $O(\sqrt{mn})$. You can again estimate the self-concordance of this hybrid barrier by the calculation above.

17.3 Lee-Sidford Barrier

Intuitively, we can think the volumetric barrier reweigh the constraints by leverage score. Lemma 17.2.2 suggests that the best barrier should have $\frac{\sigma_i(\sqrt{W}A_x)}{w_i}$ all roughly the same. Ideally, one would set

$$w_i = \sigma_i(\sqrt{W}A_x).$$

This is exactly the condition of the ℓ_∞ Lewis weight. Recall from Definition 11.1.3 that ℓ_p Lewis weight $w_p(A)$ be the unique vector $w \in \mathbb{R}_+^m$ such that

$$w_i = \sigma\left(W^{\frac{1}{2}-\frac{1}{p}}A\right)_i.$$

Now, the problem is simply to find a barrier function ϕ such that at every x, we have

$$\phi_p(x) \sim -\sum w_p(A_x)\ln(a_i^\top x - b_i).$$

Note that if we indeed has such function for $p = \infty$, then this barrier should have self-concordant $O(n)$.

Recall from Lemma 11.1.5 that $w_p(A_x)$ is the maximizer of

$$\phi_p(x) = \max_{w_i \geq 0} \log\det\left(A_x^\top W^{1-\frac{2}{p}}A_x\right) - (1-\frac{2}{q})\sum_{i=1}^m w_i.$$

Note that this function looks very close to volumetric barrier, except that it optimizes over a family of volumetric barrier with different reweighing. Therefore, it is natural to ask if $\phi_p(x)$ is self-concordant.

Note that $\phi_2(x)$ is exactly the volumetric barrier. Also, one can prove that $w_0(A_x)$ is constant (if A is in general position) and hence $\phi_0(x)$ looks like the log barrier function. Therefore, ϕ_p can think as a family of barrier that includes log barrier and volumetric barrier.

By some calculations, it turns out ϕ_p is self-concordance for all $0 < p < \infty$.

Theorem 17.3.1. *Given a polytope $\Omega = \{Ax > b\}$. Suppose that Ω is bounded and non-empty and that every row of A is non-zero. For any $p > 0$, after rescaling, the $\phi_p(x)$ is a $O_p(n \cdot m^{\frac{1}{p+2}})$ self-concordant barrier.*

Take $p = \Theta(\log m)$, we have a barrier with self concordance is $O(n\log^{O(1)} m)$. Unfortunately, the proof of this is a little bit long. We left it as a 12-pages long calculus exercise.

17.4 Open Problems

Ideally, one would ask if the LS barrier is practical. Since the current best approach in practice is the primal dual central path, we have the following question:

Problem 17.4.1. What is the primal dual analogy of the LS result?

Beside linear programs, another important class of problems is semi-definite programming. Nesterov and Nemirovskii generalized the volumetric barrier to semidefinite programming setting [93]. Naturally, can we generalize the LS barrier to that setting? Note that the current fastest algorithm for semi-definite programming is cutting plane method instead of interior point method. This is very unusual for structured convex programming. Therefore, there are a lot of opportunities for improving algorithms on semi-definite programming.

Problem 17.4.2. Improve the convergence rate of interior point methods for semi-definite programming.

If we view the LS barrier as a weighted combination of log barrier, then it is natural to ask:

Problem 17.4.3. Given barriers ϕ_i for K_i, can we come with a barrier for $\bigcap K_i$ that is better than $\sum_i \phi_i$?

Finally, going back to my original purpose of solving linear programs:

Problem 17.4.4. Can you solve the maximum flow problem faster than $O^*(m\sqrt{n})$?

References

[91] Andrew V Goldberg and Satish Rao. Beyond the flow decomposition barrier. *Journal of the ACM (JACM)*, 45(5):783–797, 1998.

[92] Murat Mut and Tamás Terlaky. A tight iteration-complexity upper bound for the mty predictor-corrector algorithm via redundant klee-minty cubes. 2014.

[93] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming.* SIAM, 1994.

[94] Pravin M Vaidya. A new algorithm for minimizing convex functions over convex sets. In *Foundations of Computer Science, 1989., 30th Annual Symposium on*, pages 338–343. IEEE, 1989.