10-725: Optimization **Fall 2012**

Lecture 22: Introduction to Interior Point Methods

Lecturer: Geoff Gordon/Ryan Tibshirani *Scribes: Abhinav Shrivastava*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

22.1 Interior-point methods

Interior-point methods are important because they are by far the most efficient methods to solve optimization problems with relatively small numbers of variables and potentially very large numbers of linear constraints (specially if we need very accurate solutions.) Also, they are based on geometric intuition and deriving the interior-point method for a problem can give interesting insights.

Historical Interest: (1) First practical polynomial time algorithm for solving an LP. (It is still an open question whether they are strongly polynomial time or not.) (2) First proposed method in literature: Ellipsoid Method. (Though it is not practical for a lot of problems, it is a great theoretical tool.)

Combinatorial or real analysis of LPs: We can either think of solving for Linear Program (LP) as being a combinatorial algorithm (for example, in case of the simplex algorithm, searching amongst discrete sets of vertices for solution) or a continuous problem (for example, maximizing a continuous function over a convex set). Interior-point methods can act as a bridge between both forms of reasoning (as we will see further in lectures.)

22.1.1 Ball Center (aka Chebyshev center)

Suppose for a problem, the feasible region (figure 22.2) is given by

$$X = \{\ x \mid Ax + b \geq 0\}. \tag{22.1}$$

In general, this is too big of a description (e.g. intersection of too many linear inequalities), and we would like to come up with a shorter description of this feasible region. One try might be to find the ball center, which is the center of the largest sphere that can be inscribed inside this feasible region (dotted blue line in figure 22.2.) As an optimization problem, we can write it as:

$$\max_x \min_i \ \texttt{dist}(x,\, a_i^T x + b_i = 0) \tag{22.2}$$

where $a_i^T x + b_i = 0$ is the constraint line for the i^{th} constraint, `dist` is euclidean distance of x to that constraint line, $\min_i$ finds closest constraint to the point x and $\max_x$ maximizes the distance of x to that closest constraint. Thus, (eq. 22.2) will maximize the radius of the ball that we can inscribe centered at x. We can find this ball center using a LP as follows:

1. If $\|a_i\| = 1$ (i.e. a_i is normalized), then $\texttt{dist}(x, a_i^T x + b_i = 0)$ is just $a_i^T x + b_i$ (a linear function of x).

So we can re-write (eq. 22.2) as:

$$\max_{x,t} \; t \quad \text{subject to } Ax + b \geq t\,\mathbf{1} \tag{22.3}$$

$$\max_{x,t} \; t \quad \text{subject to } \mathbf{s} \geq t\,\mathbf{1} \text{ where } \mathbf{s} = Ax + b \tag{22.4}$$

2. In general, we can write the constraint as $\dfrac{s_i}{\|a_i\|} \geq t \; \forall \, i$. Since $\|a_i\|$ is constant, the constraint is linear in s_i and t, therefore we have a LP.

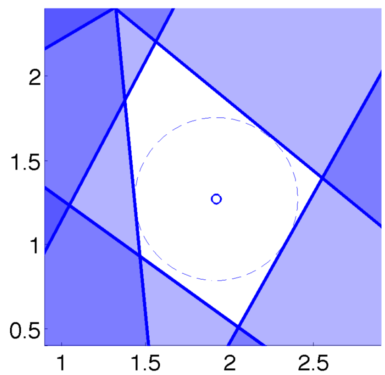
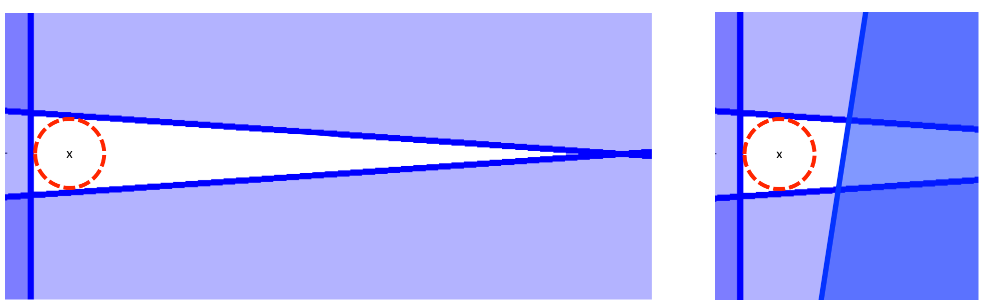


Figure 22.1: Ball Center (solid blue circle) for the feasible region (white polyhedron) (corresponding ball center in dotted blue)



(a) Feasible region with bad conditioning (b)

Figure 22.2: Ball center (dashed red circle) for a feasible region (white polyhedron) (corresponding center marked by "x".) For two very varied feasible regions given in (a) and (b), we have the same ball center.

Bad News: Consider the feasible region shown in figure 2(a) (long skinny triangle). The selected ball center is not a very good summary of the feasible region because we will have the same ball center if we had the feasible region from figure 2(b), which is indeed very different from figure 2(a). A symptom of this problem is that the ball may contain an arbitrarily small fraction of the volume of the feasible region.

22.1.2 Ellipsoid Center (aka max-volume inscribed ellipsoid)

It is the center of the inscribed ellipsoid with the largest possible volume. The ellipse centered at d is given by

$$E = \{Bu + d \mid \|u\|_2 \leq 1\} \tag{22.5}$$

i.e. a scaled and shifted version of the unit sphere. The volume of E is proportional to $\det B$, therefore 1/volume is proportional to $\det B^{-1}$. It turns out that we can find the ellipse centered at d as a solution of the semi-definite program given by

$$\begin{aligned} &\min \ \log\det B^{-1} \\ &\text{subject to: } a_i^T(Bu+d)+b_i \geq 0 \quad \forall i \quad \forall u \text{ with } \|u\| \leq 1 \\ &\qquad B \succeq 0 \end{aligned}$$

The above constraints imply that the entire ellipse has to be contained in the feasible region. For each i and each u, we have a linear constraint in B and d (actually, there are infinitely many such constraints.) The constraints are linear, and therefore we have a convex optimization problem.

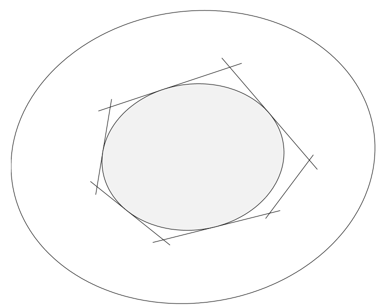


Figure 22.3: Maximum volume ellipsoid (shaded) inscribed in the feasible region (polyhedron)

As can be seen in example shown in figure 22.3, the resulting ellipsoid fills up the feasible region quite nicely. We can show that in general the $\texttt{Volume}(E) \geq \left(\frac{1}{n^n}\right) \texttt{Volume}(X)$ (E from eq. 22.5 and X from eq. 22.1) i.e. at least a polynomial fraction of the feasible region is covered by the ellipsoid.

In conclusion, we have a convex optimization problem to find a good summary of the feasible region. But the problem is very expensive, with infinitely many constraints in an SDP (but we can still do it.)

So, the next obvious question: Is there some summary of the feasible region which can be computed cheaply, and which can be said to be a good summary? (e.g. if it contains at least a polynomial fraction of the volume of the feasible region, like ellipsoid center discussed above.) Find the answer in the next section!

22.1.3 Analytic Center

Take the slack $\mathbf{s} = Ax + b \Leftrightarrow s_i = a_i^T x + b_i$ from eq. 22.4. We can define analytic center as:

$$\max_{x,s>=0} \prod_i \frac{s_i}{\|a_i\|} \tag{22.6}$$

This is similar to maximizing the minimum distance (from eq. 22.2, see general formulation) as if we have a small distance, then the above product is going to be small. On the other hand, the only way to maximize the product is to make all the slacks (s_i) approximately equal or at least some subset of slacks approximately equal and others bigger than that (can be seen as a soft version of the ball center.) Since $-\log$ is antimonotone function, therefore we can rewrite eq. 22.6 as

$$\max_{x,s>=0} \prod_i \frac{s_i}{\|a_i\|} \tag{22.7}$$

$$\Leftrightarrow \max_{x,s>=0} -\sum_i \log \frac{s_i}{\|a_i\|} \tag{22.8}$$

$$\Leftrightarrow \max_{x,s>=0} -\sum_i \log s_i + \sum_i log\|a_i\| \tag{22.9}$$

$$\Leftrightarrow \max_{x,s>=0} -\sum_i \log s_i \tag{22.10}$$

since $\|a_i\|$ is constant. Thus, minimizing the sum of negative logs of the slacks is same as maximizing the product in eq. 22.6. Therefore, the definition analytic center of a feasible region is given by eq. 22.10 (x in figure 22.4 is the analytic center.)

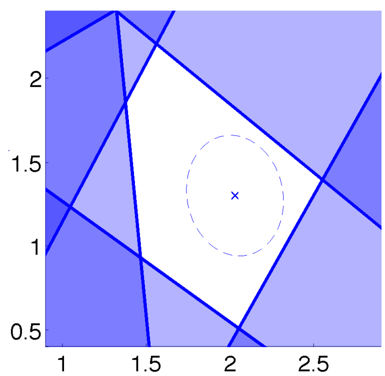


Figure 22.4: Analytic Center (blue "x") for the feasible region (white polyhedron) (corresponding Dikin Ellipsoid in dotted blue)

It turns out that we can define an ellipse, centered at the analytic center, that contains at least the polynomial fraction of the volume of the feasible region (for e.g. dikin ellipsoid, more on this in next lecture.). More concretely, for n dimensional problem with m constraints, we at least cover a factor of $(1/m^n)$ of the feasible region $\big(\texttt{Volume}(\text{Dikin Ellipsoid}) \geq \big(\frac{1}{m^n}\big) \texttt{Volume}(X)\big)$(the proof is in book). Therefore, we have a similar polynomial fraction guarantee as in ellipsoid center (though weaker). But how much weaker? In n dimensions, we need at least $m = n + 1$ constraints to make a closed polyhedron. So, $m > n \Leftrightarrow (1/m^n) < (1/n^n)$. Recall from previous section that $\texttt{Volume}(\text{Ellipsoid Center}) \geq \big(\frac{1}{n^n}\big) \texttt{Volume}(X)$, therefore we have $\texttt{Volume}(\text{Dikin Ellipsoid}) \leq \texttt{Volume}(\text{Ellipsoid Center})$ i.e. the ellipsoid center's volume has to be at least as large as analytic center's ellipsoid volume (but both are polynomial fraction of the total volume of the feasible region)

Invariance: The analytic center is invariant to scaling of the rows of A (scaling will only change the constant that we dropped.) i.e. it is invariant to individual scaling of the constraints (a_i).

Comparison to Ball Center Unlike the ball center, analytic center will be effected by every constraint, even the inactive ones, as the product/sum has every constraint (though far away constraints will have less effect.) That means that the analytic center is not the analytic center of the feasible region, it is just the center of this particular representation of the feasible region, as we could move an inactive constraint back and forth (a bit) and it will change the analytic center.

So why analytic centers? Some cool results in next lecture!

Feasible region with bad conditioning: Analytic relieves the problem of bad conditioning by stretching

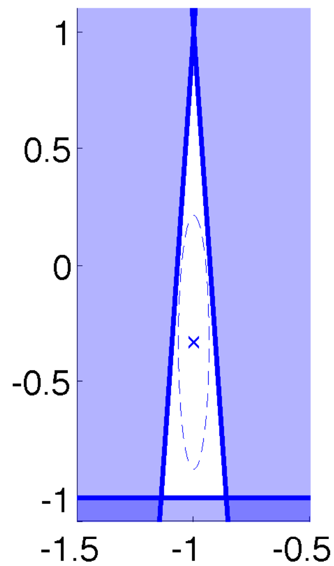


Figure 22.5: Example of badly conditioned feasible region (white polyhedron). Analytic Center (blue "x") for the feasible region (corresponding Dikin Ellipsoid in dotted blue)

out the ellipse a lot (as shown in figure 22.5) and we still get a decent volume of the feasible region. And in fact, we can show that not only analytic center is invariant to individual scaling of the constraints, but also any affine scaling of the feasible region. That is, we can shift and scale our feasible region, and the analytic center will shift and scale in exactly the same way, and the ellipsoid (dikin ellipsoid) will shift and scale the same way as well.

22.1.3.1 Newton's method for Analytic Center

The sum-of-logs in eq. 22.10 is very smooth and infinitely differentiable function, we know that Newton's method might be a good way to find the analytic center. From eq. 22.10 we have

Objective: $f(x) = -\sum_i \log(a_i^T x + b_i) = \sum_i \log(s_i)$ where $\mathbf{s} = Ax + b$

Gradient: $\dfrac{df(x)}{dx} = -\sum_i \dfrac{a_i}{a_i^T x + b_i} = -\sum_i a_i \dfrac{1}{s_i} = -A^T \left(\dfrac{1}{s}\right)$ where $(1/s)$ is component-wise.

Hessian: $\dfrac{d^2 f(x)}{d^2 x} = \sum_i \dfrac{a_i a_i^T}{(a_i^T x + b_i)^2} = A^T S^{-2} A$ where $S = \texttt{diag}(\mathbf{s})$

Using this gradient and hessian, therefore we can run Newton's method on $f(x)$. Because, our objective is strictly convex function, the hessian is strictly positive definite and hence the Newton's direction is actually a descent direction. This implies that the Newton's method converges from any feasible initializer (as long as we use line search).

But we still have to fine a strictly feasible initializer, where the objective is finite (see next lecture to get around this requirement). It turns out that the Newton's method will converge globally in a very small number of iterations.

Adding an Objective
Analytic center was for feasibility problem like finding x such that $Ax + b \geq 0$. Instead of this, now we want to find solution to $\min c^T x$ subject to $Ax + b \geq 0$.
We can use the same trick and solve

$$\min f_t(x) = c^T x - \left(\frac{1}{t}\right) \sum_i \log(a_i^T x + b_i)$$

where $t > 0$, is the trade-off between two terms. As $t \to 0$, $\left(\frac{1}{t}\right) \to \infty$ and the log part of $f_t(x)$ dominates, and we have the solution for analytic center. As $t \to \infty$, $\left(\frac{1}{t}\right) \to 0$ and the log part vanishes, thus we just get the LP optimal. The center path is smooth (infinitely differentiable for any finite t) and so the central path smoothly connects the analytic center to the LP optimal (more details on this in next lecture.)