CS 798 - Convexity and Optimization, Winter 2017, Waterloo

Lecture 13: Linear programming

We present a self-contained analysis of the interior point method in solving linear programs.

Setup

We consider linear programming of the form: $\min \langle c, x\rangle$

$$Ax \geq b.$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$.

To solve it using interior point method, we reduce it to an unconstrained optimization problem using the log-barrier function: $\min f_t(x) := t\langle c,x\rangle - \sum_{i=1}^{m} \log(a_i^T x - b_i)$

We denote $\phi(x) = -\sum_{i=1}^{m} \log(a_i^T x - b_i)$ as the "penalty" function, so that $f_t(x) := t\langle c,x\rangle + \phi(x)$.

To solve the unconstrained problem using Newton's method, we need to compute the gradient and Hessian of f_t.

Note that $\nabla \phi(x) = \sum_{i=1}^{m} \frac{-a_i}{a_i^T x - b_i}$ and $\nabla^2 \phi(x) = \sum_{i=1}^{m} \frac{a_i a_i^T}{(a_i^T x - b_i)^2}$,

thus $\nabla f_t(x) = tc - \sum_{i=1}^{m} \frac{a_i}{a_i^T x - b_i}$ and $\nabla^2 f_t(x) = \nabla^2 \phi(x)$.

Let x^* be an optimal solution to the original linear program, and let x_t^* be the optimal solution to $f_t(x)$.

Recall that for $t \geq \frac{m}{\varepsilon}$, then we proved that $\langle c, x_t^* \rangle \leq \langle c, x\rangle + \varepsilon$, so our goal is to solve $f_{\frac{m}{\varepsilon}}$.

The interior point method is to gradually increase t, and use the previous iterate as the initial solution of the current iterate and use Newton's method.

Algorithm

- Start with an initial solution x which is an approximately optimal solution of $f_{t_0}(x)$ for some $t_0 \geq 0$.
- While $t < \frac{m}{\varepsilon}$ do
  - (inner iteration) $x \leftarrow x - (\nabla^2 f_t(x))^{-1} \nabla f_t(x)$. This is one Newton step
  - (outer iteration) $t \leftarrow t + \Delta_t$.
- Return x.

In the outer iteration, last time we wrote $t \leftarrow (1+\alpha)t$ for some $\alpha > 0$, while this time we write an additive increase. It will be clear that $\Delta_t \geq \alpha t$ for some $\alpha > 0$, and we write the additive term to give a better lower bound for the first iteration when t is close to zero.

Invariant: At each step, we maintain the invariant that x is an approximately optimal solution to f_t.

The precise measure is to use the Newton decrement: $\| \nabla f_t(x) \|_{\nabla^2 f_t(x)^{-1}} \leq \delta$ for some $\delta \leq 0.01$

To simplify notation, we use x as the current iterate, $\bar{x} = x - (\nabla^2 f_t(x)^{-1}) \nabla f_t(x)$ as the next iterate.

Also, we use H to denote $\nabla^2 f_t(x)$, and $\bar{H}$ to denote $\nabla^2 f_t(\bar{x})$. Also $\bar{t} = t + \Delta_t$.

Using this notation, we want to prove that if $\| \nabla f_t(x) \|_{H^{-1}} \leq \delta$, then $\| \nabla f_{\bar{t}}(x) \|_{\bar{H}^{-1}} \leq \delta$.

Plan

The analysis has roughly four parts. The constants are rather arbitrary.

1. (inner iteration): We will show that $\| \nabla f_t(\bar{x}) \|_{\bar{H}^{-1}} \leq \frac{1}{\sqrt{2}} \| \nabla f_t(x) \|_{H^{-1}}$, i.e. the Newton decrement decreases by a constant factor after the Newton step.

   To prove this, we will use the property of the log-barrier function to show that the Hessian doesn't change quickly when we move from x to $\bar{x}$.

   We can interpret this as saying that we are in the quadratic convergence zone of the Newton method, as we had shown last time.

2. (outer iteration): We will then increase t as much as possible while $\| \nabla f_{\bar{t}}(\bar{x}) \|_{\bar{H}^{-1}} \leq \delta$.

   We will use the property of the log-barrier function to show that $\bar{t} \geq (1 + \frac{1}{\sqrt{m}}) t$.

   This will eventually imply that there are $\tilde{O}(\sqrt{m})$ iterations for most combinatorial problems, after we establish a good initial lower bound on Δ_0.

3. (ending): We will argue that when $\| \nabla f_t(x) \|_{H^{-1}}$ is small, then x is an approximately optimal solution to f_t. So, in particular, when we stop, $\langle c, x \rangle \leq \langle c, x^* \rangle + 2\varepsilon$.

   As we mentioned last time, this is an analog of the following statement in gradient descent:

   if f is strongly convex and $\|\nabla f(x)\|_2$ is small, then x is close to the optimal solution.

4. (starting): We need to find an initial point to start the central path algorithm, so that the number of iterations can be bounded.

   In general, the number of iterations depends on the bit-complexity of the problem, but for combinatorial problems we can find initial solutions that guarantee $\tilde{O}(\sqrt{m})$ iterations to converge.

Inner iteration

Inner iteration

The proof has two steps: The first step is to show that the log-barrier function satisfies a Hessian smoothness property, and the second step uses this property to show the decrease in Newton decrement.

Property: If $\|\bar{x}-x\|_H = \|(\nabla^2 f_t(x))^{-1}\nabla f_t(x)\|_H = \|\nabla f_t(x)\|_{H^{-1}} \le \delta$, then all the Hessians $H_y := \nabla^2 f_t(y)$ for y between x and $\bar{x}$ satisfy $(1-\delta)^2 H_y \precsim H \precsim (1+\delta)^2 H_y$.

proof: Recall that $H = \nabla^2 f_t(x) = \nabla^2 \phi(x) = \sum_{i=1}^m \frac{a_i a_i^T}{(a_i^T x - b_i)^2}$. Similarly, $H_y = \nabla^2 f_t(y) = \sum_{i=1}^m \frac{a_i a_i^T}{(a_i^T y - b_i)^2}$.

To show $(1-\delta)^2 H_y \precsim H \precsim (1+\delta)^2 H_y$, we show $(1-\delta)^2 v^T H_y v \le v^T H v \le (1+\delta)^2 v^T H_y v$ when $\|y-x\|_H \le \delta$.

To do this, it is enough to show that $(a_i^T x - b_i)^2 \approx (a_i^T y - b_i)^2$ when $\|y-x\|_H \le \delta$.

The condition $\|y-x\|_H \le \delta$ means $\delta^2 \ge (y-x)^T\left(\sum_{i=1}^m \frac{a_i a_i^T}{(a_i^T x-b_i)^2}\right)(y-x) = \sum_{i=1}^m \frac{\langle a_i, y-x\rangle^2}{(a_i^T x - b_i)^2}$.

This implies that $\left|\frac{\langle a_i, y-x\rangle}{a_i^T x - b_i}\right| \le \delta$ for all $1 \le i \le m$, and thus $|\langle a_i, y-x\rangle| \le \delta\, |a_i^T x - b_i|$,

hence $\big|\, |a_i^T y - b_i| - |a_i^T x - b_i| \,\big| \le \delta\, |a_i^T x - b_i|$ and so $(1-\delta)^2 (a_i^T x - b_i)^2 \le (a_i^T y - b_i)^2 \le (1+\delta)^2 (a_i^T x - b_i)^2$.

Therefore, $v^T H v = \sum_{i=1}^m \frac{\langle a_i, v\rangle^2}{(a_i^T x - b_i)^2} \le \sum_{i=1}^m (1+\delta)^2 \frac{\langle a_i, v\rangle^2}{(a_i^T y - b_i)^2} = (1+\delta)^2 v^T H_y v$.

Similarly, we have $v^T H v \ge (1-\delta^2) v^T H_y v$ and we are done. □

Note that the property also implies that $(1+\delta)^{-2} v^T H_y^{-1} v \le v^T H^{-1} v \le (1-\delta)^{-2} v^T H_y^{-1} v$ for all $v \in \mathbb{R}^n$.

For $\delta \le 0.01$, we have $e^{-0.1} H_y \precsim H \precsim e^{0.1} H_y$ and $e^{-0.1} H_y^{-1} \precsim H^{-1} \precsim e^{0.1} H_y^{-1}$.

Now, we use the property to show the decrease of the Newton decrement, which is a straightforward calculation.

Lemma: If $\|\bar{x}-x\|_H = \|\nabla f_t(x)\|_{H^{-1}} \le \delta \le 0.01$, then $\|\nabla f_t(\bar{x})\|_{\bar{H}^{-1}} \le \frac{1}{\sqrt{2}} \|\nabla f_t(x)\|_{H^{-1}}$.

Proof: By the property, we have $\|\nabla f_t(\bar{x})\|^2_{\bar{H}^{-1}} \le e^{0.1} \|\nabla f_t(\bar{x})\|^2_{H^{-1}}$

$= e^{0.1} \|\nabla f_t(x - H^{-1}\nabla f_t(x))\|^2_{H^{-1}}$ by Newton step.

We use Taylor's theorem to write $\nabla f_t(x - H^{-1}\nabla f_t(x)) = \nabla f_t(x) - \nabla^2 f_t(y) H^{-1} \nabla f_t(x)$, where y is some point between x and $\bar{x}$.

So, $\|\nabla f_t(x - H^{-1}\nabla f_t(x))\|^2_{H^{-1}} = \|(I - H_y H^{-1})\nabla f_t(x)\|^2_{H^{-1}}$.

We would like to bound the RHS by $\|\nabla f_t(x)\|^2_{H^{-1}}$, and we are just going to expand it and show that $\|(I - H_y H^{-1}) v\|^2_{H^{-1}} \le 0.42 \|v\|^2_{H^{-1}}$ $\forall v \in \mathbb{R}^n$.

$$\|(I - H_y H^{-1})v\|_{H^{-1}}^2 = v^T(I - H^{-1}H_y)H^{-1}(I - H_y H^{-1})v$$
$$= v^T H^{-1} v - 2v^T H^{-1} H_y H^{-1} v + v^T H^{-1} H_y H^{-1} H_y H^{-1} v$$
$$\leq v^T H^{-1} v - 2e^{-0.1} v^T H^{-1} H H^{-1} v + v^T H^{-1} H_y H^{-1} H_y H^{-1} v \quad (\text{using } H_y \succeq e^{-0.1} H)$$
$$\leq (1 - 2e^{-0.1}) v^T H^{-1} v + e^{0.1} v^T H^{-1} H_y H_y^{-1} H_y H^{-1} v \quad (\text{using } H^{-1} \preceq e^{0.1} H_y^{-1})$$
$$\leq (1 - 2e^{-0.1}) v^T H^{-1} v + e^{0.2} v^T H^{-1} v \quad (\text{using } H_y^{-1} \preceq e^{0.1} H)$$

Therefore, $\|(I - H_y H^{-1})v\|_{H^{-1}}^2 \leq (1 - 2e^{-0.1} + e^{0.2})\|v\|_{H^{-1}}^2 \leq 0.42\|v\|_{H^{-1}}^2$.

Hence, $\|\nabla f_t(\bar{x})\|_{\bar{H}^{-1}}^2 \leq \frac{1}{2}\|\nabla f_t(x)\|_{H^{-1}}^2$, and the lemma follows. □

To conclude, if we maintain the invariant $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0.01$, then $\|\nabla f_t(\bar{x})\|_{\bar{H}^{-1}} \leq \frac{1}{\sqrt{2}}\|\nabla f_t(x)\|_{H^{-1}}$.

Outer iteration

After one Newton step, we have $\|\nabla f_t(\bar{x})\|_{\bar{H}^{-1}} \leq \frac{1}{\sqrt{2}}\delta$, and we would like to increase t as much as possible while maintaining $\|\nabla f_{\bar{t}}(\bar{x})\|_{\bar{H}^{-1}} \leq \delta$.

First, we lower bound the additive increase of t to maintain the invariant.

Lemma We can set $t' = t + \frac{1}{200\|c\|_{H^{-1}}}$ and maintain $\|\nabla f_{\bar{t}}(\bar{x})\|_{\bar{H}^{-1}} \leq \delta$, given $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta$ and $\delta \leq 0.01$.

proof $\|\nabla f_{\bar{t}}(\bar{x})\|_{\bar{H}^{-1}} = \|\bar{t}c + \nabla\phi(\bar{x})\|_{\bar{H}^{-1}} = \|(\bar{t} - t)c + tc + \nabla\phi(\bar{x})\|_{\bar{H}^{-1}}$

$$\leq \|(\bar{t} - t)c\|_{\bar{H}^{-1}} + \|tc + \nabla\phi(\bar{x})\|_{\bar{H}^{-1}} \quad (\text{triangle inequality})$$
$$= (\bar{t} - t)\|c\|_{\bar{H}^{-1}} + \|\nabla f_t(\bar{x})\|_{\bar{H}^{-1}}$$
$$\leq (\bar{t} - t)\|c\|_{\bar{H}^{-1}} + \frac{1}{\sqrt{2}}\delta. \quad (\text{inner iteration lemma})$$

By setting $\bar{t} = t + \frac{\delta}{\sqrt{2}\|c\|_{\bar{H}^{-1}}} \geq t + \frac{1}{200\|c\|_{H^{-1}}}$ as $\|c\|_{\bar{H}^{-1}} \leq e^{0.05}\|c\|_{H^{-1}}$, we have $\|\nabla f_{\bar{t}}(\bar{x})\|_{\bar{H}^{-1}} \leq \delta$. □

We would like to lower bound the additive term as a multiplicative term of t.

Before that, we first compute an important parameter of the log-barrier function, which is the reason of the $\sqrt{m}$ term in the number of iterations.

Property $\|\nabla\phi(x)\|_{H^{-1}} \leq \sqrt{m}$.

proof Recall that $\nabla\phi(x) = \sum_{i=1}^m \frac{-a_i}{a_i^T x - b_i}$ and $H = \nabla^2 f_t(x) = \nabla^2\phi(x) = \sum_{i=1}^m \frac{a_i a_i^T}{(a_i^T x - b_i)^2}$.

We need to compute $\sqrt{\nabla\phi(x)^T H^{-1} \nabla\phi(x)}$ and it is best to write expressions using matrix form.

Let A be the $m \times n$ matrix where the i-th row is a_i.

Let $\vec{s} \in \mathbb{R}^m$ be the vector with the i-th entry is $a_i^T x - b_i$ and $S := \mathrm{diag}(\vec{s})$ be the $m \times m$ diagonal matrix where the (i,i)-th entry being $a_i^T x - b_i$.

Then $\nabla \phi(x) = A^T S \vec{1}$ and $H = A^T S^2 A$

So, $\|\nabla \phi(x)\|_{H^{-1}}^2 = \nabla \phi(x)^T H^{-1} \nabla \phi(x) = \vec{1}^T (SA)(A^T S S A)^{-1} (A^T S) \vec{1}$.

Note that $\Pi = (SA)(A^T S S A)^{-1}(A^T S)$ is a projection matrix as $\Pi^2 = \Pi$.

So, $\|\nabla \phi(x)\|_{H^{-1}}^2 = \vec{1} \Pi \vec{1} = \vec{1} \Pi^2 \vec{1} = \|\Pi \vec{1}\|_2^2 \leq \|\vec{1}\|_2^2 = m$. □

Lemma $1/\|c\|_{H^{-1}} \geq t/(\delta + \sqrt{m}) \geq t/2\sqrt{m}$.

proof $\|c\|_{H^{-1}} = \frac{1}{t} \| \nabla f_t(x) - \nabla \phi(x)\|_{H^{-1}} \leq \frac{1}{t} \|\nabla f_t(x)\|_{H^{-1}} + \frac{1}{t} \|\nabla \phi(x)\|_{H^{-1}} \leq \frac{1}{t}\|\nabla f_t(x)\|_{H^{-1}} + \frac{1}{t}\|\nabla \phi(x)\|_{H^{-1}}$.

Using the invariant $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta$ and $\|\nabla \phi(x)\|_{H^{-1}} = m$, we have $\|c\|_{H^{-1}} \leq \frac{1}{t}(\delta + \sqrt{m})$. □

Theorem It takes at most $O\left(\sqrt{m} \log\left(\frac{m \cdot \|c\|_{H_0^{-1}}}{\varepsilon}\right)\right)$ iterations for $\bar{t} \geq \frac{m}{\varepsilon}$, where $H_0 = \nabla^2 f_{t_0}(x_0)$ and x_0 is the initial solution and t_0 is such that $\|\nabla f_{t_0}(x_0)\|_{\nabla^2 f_{t_0}(x_0)^{-1}} \leq \delta$.

proof In every outer iteration, we can increase t by $\frac{1}{200\|c\|_{H^{-1}}}$ by the first lemma.

This is at least $\frac{t}{400\sqrt{m}}$ by the second lemma, and so $\bar{t} \geq t(1 + \frac{1}{400\sqrt{m}})$.

If we do the outer iteration $k+1$ times, then $\bar{t} \geq \frac{1}{200\|c\|_{H_0^{-1}}}(1 + \frac{1}{400\sqrt{m}})^k$, by using the additive bound for the first iteration and then use the multiplicative bound k times.

So, to get $\bar{t} \geq \frac{m}{\varepsilon}$, we set k so that $\frac{1}{200\|c\|_{H_0^{-1}}}(1+\frac{1}{400\sqrt{m}})^k \geq \frac{m}{\varepsilon}$.

It takes at most $\log\left(\frac{200\|c\|_{H_0^{-1}} m}{\varepsilon}\right) / \log\left(1 + \frac{1}{400\sqrt{m}}\right) \leq 800\sqrt{m} \log\left(\frac{200\|c\|_{H_0^{-1}} \cdot m}{\varepsilon}\right) = O\left(\sqrt{m} \log\left(\frac{m \cdot \|c\|_{H_0^{-1}}}{\varepsilon}\right)\right)$,

where the inequality follows from $\log(1+x) \geq \frac{x}{2}$ for $0 \leq x \leq 1$.

Note that $\|c\|_{H_0^{-1}} = \|c\|_{\nabla^2 f_{t_0}(x_0)^{-1}}$ where x_0 is the initial point and t_0 is such that the invariant holds. □

Ending

We are almost there: We need to find a good starting point, and we need to show that we are close to an optimal solution when we stop. Here we address the latter issue.

There are two steps. Let x_t^* be the optimal solution to f_t.

First, we show that if $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta$, then x is close to x_t^*.

Lemma If $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0.01$, then $\|x - x_t^*\|_H \leq 3\delta$.

proof We will prove that if $\|\nabla f_t(x)\|_{H^{-1}}$ is small, then $f_t(x+v) \geq f_t(x)$ for $\|v\|_H$ large, and thus $\|x - x_t^*\|_H$ must be small.

By Taylor's theorem, $f_t(x+v) = f_t(x) + \langle \nabla f_t(x), v\rangle + \frac{1}{2} v^T \nabla^2 f_t(y) v$ where y is between x and $x+v$.

$$\geq f_t(x) - \|\nabla f_t(x)\|_{H^{-1}} \|v\|_H + \frac{1}{2} v^T \nabla^2 f_t(y) v$$ using $\langle a,b\rangle \geq -\|a\|_{H^{-1}} \|b\|_{H^{-1}_*} = -\|a\|_{H^{-1}} \|b\|_H$

Since $\|y-x\|_H \leq \|v\|_H$, apply the property in the inner iteration section. we have $(1-\|v\|_H)^2 H_y \preceq H \preceq (1+\|v\|_H)^2 H_y$.

Therefore, $f_t(x+v) \geq f_t(x) - \delta \|v\|_H + \frac{1}{2(1+\|v\|_H)^2} v^T H v = f_t(x) - \delta \|v\|_H + \frac{\|v\|_H^2}{2(1+\|v\|_H)}$.

For v with $\|v\|_H \geq 3\delta$, we have $f_t(x+v) > f_t(x)$.

Therefore, since $f_t(x_t^*) \leq f_t(x)$, we must have $\|x_t^* - x\|_H \leq 3\delta$. □

Next, we show that the objective value of x is close to the objective value of x_t^*.

<u>Lemma</u> If $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0.01$, then $c^T x - c^T x^* \leq \frac{1}{t}(\delta + \sqrt{m}) \cdot 3\delta$.

<u>proof</u> $c^T x - c^T x_t^* = \langle c, x - x_t^* \rangle \leq \|c\|_{H^{-1}} \|x - x_t^*\|_H$ using generalized Cauchy-Schwarz and $\|a\|_{H*} = \|a\|_{H^{-1}}$.

By the second lemma in the outer iteration section, $\|c\|_{H^{-1}} \leq \frac{1}{t}(\delta + \sqrt{m})$, and thus using the previous lemma

$$c^T x - c^T x_t^* \leq \frac{1}{t}(\delta + \sqrt{m}) \cdot 3\delta. \quad \square$$

Finally. we can bound the distance to optimality when we stop.

<u>Theorem</u> If $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0.01$ for $t \geq \frac{m}{\varepsilon}$, then $c^T x - c^T x^* \leq 2\varepsilon$.

<u>proof</u> $c^T x - c^T x^* = c^T x - c^T x_t^* + c^T x_t^* - c^T x^*$

$\leq c^T x - c^T x_t^* + \frac{m}{t}$ by the claim in central path in L12.

$\leq \frac{1}{t}(m + (\delta + \sqrt{m}) \cdot 3\delta)$ by the previous lemma

$\leq 2m/t \leq 2\varepsilon$ using $t \geq \frac{m}{\varepsilon}$. □

<u>Starting</u>

We need to find an initial point x_0 that is close to the central path to start the algorithm.

One natural such point is the optimal solution x_{ac} when $t=0$, i.e. x_{ac} is the minimizer of $f_0(x) = \phi(x)$.

This point is called the analytical center of the polytope, as we have seen in L06 and HW1.

For some combinatorial problems, it is easy to find an analytical center.

One nice example is the <u>undirected</u> maxflow problem, which can be written as:

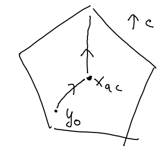
$$\max \ x_{ts} \quad \text{s.t.} \quad Bx = 0, \ -\vec{1} \leq x \leq \vec{1}.$$

In this case, the analytic center is just the original, which is also feasible.

One can check that $\|c\|_{\nabla^2 f(x_{ac})^{-1}} = O(\text{poly}(m))$, and by the theorem in the outer iteration section,

the number of iterations to compute an ε-approximate max-flow is $O(\sqrt{m}\log(\frac{m\cdot\|c\|_{H^{-1}}}{\varepsilon})) = \tilde{O}(\sqrt{m}\log\frac{1}{\varepsilon})$.

Backward to analytical center



In general, it is not easy to find a point close to the analytical center.

The idea is to use the interior point method itself to find the analytical center.

Suppose we have a strictly feasible solution y_0 (which may be far away from the central path for f).

The observation is that y_0 is on the central path for the objective function $-\nabla\phi(y_0)$ for $t=1$.

That is, consider the problem $g_t(y) := -t\nabla\phi(y_0)y + \phi(y)$, then y_0 is the optimizer of g_1, because

it satisfies the optimality condition $0 = \nabla g_1(y) = -\nabla\phi(y_0) + \nabla\phi(y)$.

Now, the idea is to follow the central path for g backward to reach the analytical center.

In every step, we update $t \leftarrow t(1+\frac{1}{400\sqrt{m}})^{-1} \approx t(1-\frac{1}{400\sqrt{m}})$ to decrease t to approach zero, but

otherwise using exactly the same interior point algorithm.

Number of iterations in backward central path

The plan is to compute y so that $\|\nabla g_t(y)\|_{H_y^{-1}} \le \frac{\delta}{\sqrt{2}}$ for a small enough t, and then we

set the initial point x_0 for the central path of f to be y.

The question is how small we need to set t to be, as it affects the number of iterations of

the backward interior point algorithm.

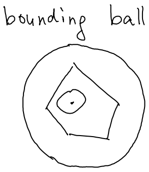
Suppose we set $x_0 = y$ when t becomes very small, and use x_0 as an approximately optimal solution for f_0.

Then
$$\|\nabla f_0(x_0)\|_{H_{x_0}^{-1}} = \|\nabla\phi(x_0)\|_{H_{x_0}^{-1}} \le \|-t\nabla\phi(y_0) + \nabla\phi(x_0)\|_{H_{x_0}^{-1}} + t\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}$$
$$= \|\nabla g_t(y)\|_{H_{x_0}^{-1}} + t\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}$$
$$\le \frac{\delta}{\sqrt{2}} + t\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}.$$

We need to set $t = \frac{\delta}{\sqrt{2}}\cdot\frac{1}{\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}}$ so that $\|\nabla f_0(x_0)\|_{H_{x_0}^{-1}} \le \delta$ and we can use it as the starting point.

Generally speaking, if the polytope is contained in a ball of reasonable radius, then we know that

the maximum eigenvalue of $H_{x_0}^{-1}$ is bounded, as the Dikin ellipsoid is contained in the bounding ball.



Also, if we can find a point y_0 such that it is not too close to the boundaries,

then $\|\nabla\phi(y_0)\|_2$ would not be too big.

then $\|\nabla\phi(y_0)\|_2$ would not be too big.

And then $\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}$ is reasonably upper bound and we don't need to do so many iterations in the backward central path algorithm.

This is satisfied in all natural problems in combinatorial optimization, with $\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}} = \mathrm{poly}(n,m)$.

In general, however, $\|\nabla\phi(y_0)\|_{H_{x_0}^{-1}}$ could be as large as $\mathrm{poly}(n) \cdot 2^{O(L)} \cdot m$ where L is the bit complexity of the problem. That would take $\tilde{O}(\sqrt{m}L)$ iterations.

The worst case is when the polytope is very wide in some direction and very narrow in some other's.

Finally, if it is not easy to find a strictly feasible point, one can add an auxiliary variable s and solve the problem: $\min s$ s.t. $Ax + s\vec{1} \geq b$

We can set $x=0$ and s large enough to obtain a strictly feasible solution, and then we can use the above method to solve this LP until we find the objective to be negative.

References

- Yin Tat Lee, personal communications.
- Nesterov, Introductory lectures on convex optimization, chapter 4.
- Vishnoi, A mini-course on convex optimization, chapter 3.