Lecture 14 Solving LPs using the interior point method

---

original LP

$$\min \ \langle c, x\rangle$$

$$\text{s.t} \quad Ax \geq b$$

$$, x \in \mathbb{R}^n \quad A = \begin{pmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_m^T - \end{pmatrix} \in \mathbb{R}^{m \times n} \quad , b \in \mathbb{R}^m$$

$\downarrow$ **unconstrained problem**

$$\min f_t(x) = t\langle c, x\rangle - \underbrace{\sum_{i=1}^{m} \log(\langle a_i, x\rangle - b_i)}_{= \varphi(x)}$$

$$\nabla f_t(x) = t \cdot c - \sum_{i=1}^{m} \frac{a_i}{\langle a_i, x\rangle - b_i}$$

$$\nabla^2 f_t(x) = \sum_{i=1}^{m} \frac{a_i^T a_i}{(\langle a_i, x\rangle - b_i)^2}$$

Remember from last lecture that in order to find a solution within an error of ε from the optimal solution, then we need to solve $f_{\frac{m}{\varepsilon}}$.

Algorithm

- start with an initial solution which is approximately optimal for $f_{t_0}(x)$, $t_0 > 0$
- repeat until $t > m/\varepsilon$
  - inner iteration $x \leftarrow x - (\nabla^2 f_t(x))^{-1} \nabla f_t(x)$
  - outer iteration $t \leftarrow t + \Delta t$
- return x

Let's introduce some notation to simplify the proof

x : current iterate

$\bar{x}$: next iterate $\left(\bar{x} = x - (\nabla^2 f_t(x))^{-1} \nabla f_t(x)\right)$

x_t^* : optimal point of $f_t(\cdot)$

$H = \nabla^2 f_t(x)$, Hessian of f_t at the current iterate

$\bar{H} = \nabla^2 f_t(\bar{x})$, Hessian of f_t at the next iterate

(the Hessian does not depend on t !!!)

The invariant that we will maintain during the analysis is that x and x_t^* are close. As we mentioned last time, in this case it makes a lot of sense to do a Newton step as the Newton method guarantees quadratic convergence speed when the starting point (in this case x, the current iterate) is "close" to the optimum x_t^*. What we will additionally prove is that $\bar{x}$ will be close to $x_{t+\Delta t}^*$. So,

in every iteration:

(1) we are close to the optimum of the current function f_t

(2) doing the Newton step makes us ready for the next iteration!!!

To argue that x and x_t^* are close we will set Δt so as to maintain the invariant:

$\|\nabla f_t(x)\|_{H^{-1}} \leq \delta$ for some small δ (e.g. $\delta = 0{,}01$)

Similarly to strongly convex functions

(3) a small gradient norm means that we are close to the optimum.

(In the case of strongly convex functions the norm was the ℓ_2-norm and now it is the quadratic norm w.r.t inverse of the Hessian H^{-1}.)

Let's start by proving that .

As an intermediary lemma we will prove that the Hessian does not change fastly!!!

Hessian is "smooth" lemma

Let $H^y = \nabla^2 f_t(y)$ be the hessian at a point y and H be the hessian at x.
(we defined before $H = \nabla^2 f_t(x)$ to be the hessian at the current iterate).
If $\|y-x\|_H \leq \delta$ then $(1-\delta)^2 H^y \preceq H \preceq (1+\delta)^2 H^y$

proof

It is enough to prove that
$(1-\delta)^2 v^T H^y v \leq v^T H v \leq (1+\delta)^2 v^T H^y v \quad \forall v$
when $\|y-x\|_H \leq \delta$

$$H = \sum_{i=1}^{m} \frac{a_i^T a_i}{(a_i^T x - b_i)^2} \qquad H^y = \sum_{i=1}^{m} \frac{a_i a_i^T}{(a_i^T y - b_i)^2}$$

$$\|y-x\|_H \leq \delta \Rightarrow \sum_{i=1}^{m} \frac{(a_i^T (y-x))^2}{(a_i^T x - b_i)^2} \leq \delta^2 \Rightarrow$$

$$\Rightarrow \frac{(a_i^T (y-x))^2}{(a_i^T x - b_i)^2} \leq \delta^2 \quad \forall i \in [m] \Rightarrow$$

$$\Rightarrow \left(a_i^T y - b_i - (a_i^T x - b_i)\right)^2 \le \delta^2 (a_i^T x - b_i)^2 \quad \forall i \in [m]$$

$$\Rightarrow (1-\delta)^2 (a_i^T x - b_i)^2 \le (a_i^T y - b_i)^2 \le (1+\delta)^2 (a_i^T x - b_i)^2$$

Thus $v^T H v = \sum_{i=1}^{m} \frac{(a_i^T v)^2}{(a_i^T x - b_i)^2} \le \sum_{i=1}^{m} \frac{(1+\delta)^2 (a_i^T v)^2}{(a_i^T y - b_i)^2} =$

$$= (1+\delta)^2 v^T H^y v$$

(and in the same way $v^T H v \geq (1-\delta)^2 v^T H^y v$

$\square$

observation

$(1-\delta)^2 H^y \preceq H \preceq (1+\delta)^2 H^y \Rightarrow$

all matrices are positive definite $\Longrightarrow$ $(1+\delta)^{-2} (H^y)^{-1} \preceq H^{-1} \preceq (1-\delta)^{-2} (H^y)^{-1}$

$\square$

Now we are ready to prove that if $\|\nabla f_t(x)\|_{H^{-1}}$ is small then x and x_t^* are close.

small gradient => close to OPT lemma

If $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0{,}001$ then

$$\|x - x_t^*\|_H \leq 3\delta$$

proof sketch

From Taylor theorem we have that:

$$f_t(x+v) = f_t(x) + \langle \nabla f_t(x), v \rangle + \frac{1}{2} v^T H^y v$$

for y between x and $x+v \Rightarrow \|y-x\|_H \leq \|v\|_H$

by the smooth Hessian lemma we have that

$$f_t(x+v) \geq f_t(x) + \langle \nabla f_t(x), v \rangle + \frac{v^T H v}{2(1+\|v\|_H^2)} \geq$$

$$\geq f_t(x) - \|\nabla f_t(x)\|_{H^{-1}} \|v\|_H + \frac{\|v\|_H^2}{2(1+\|v\|_H)^2} \Rightarrow$$

$$f_t(x+v) \geq f_t(x) - \delta \|v\|_H + \frac{\|v\|_H^2}{2(1+\|v\|_H)^2}$$

setting $v = x^* - x$ we get that

$$-\delta \|v\|_H + \frac{\|v\|_H^2}{2(1+\|v\|_H)^2} < 0 \Rightarrow$$

$\Rightarrow$ $\boxed{\|v\|_H \leq 3\delta}$

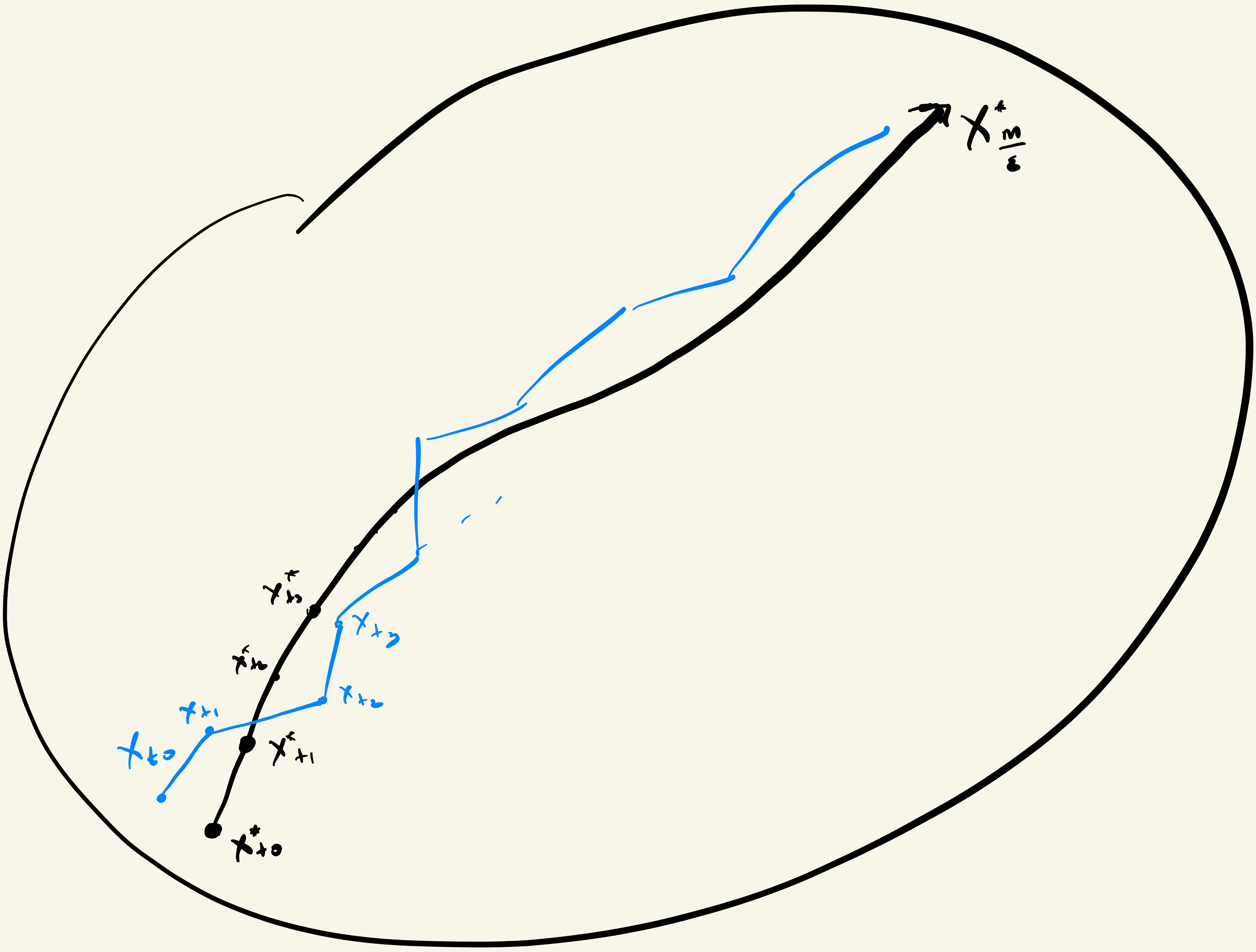
$\|v\|_H$ very big. And we need to exclude this case by arguing that the steps along the algorithm are never too big!!

Takeaway point.

If $\|\nabla f_t(x)\|_{H^{-1}}$ is small then

x and x_t^* are close!!!

Pictorial representation



proof sketch at any point in time
If we start from a "good" point
(i.e. x_{t_0} is close to $x^*_{t_0}$) then

x_t is close to x^*_t $\longrightarrow$ $x_{t+\Delta t}$ is close to $x^*_{t+\Delta t}$

$\|\nabla f_t(x_t)\|_{(\nabla^2 f_t(x_t))^{-1}}$ is small $\longrightarrow$ $\|\nabla f_{t+\Delta t}(x_{t+\Delta t})\|_{(\nabla^2 f_{t+\Delta t}(x_{t+\Delta t}))^{-1}}$ is small

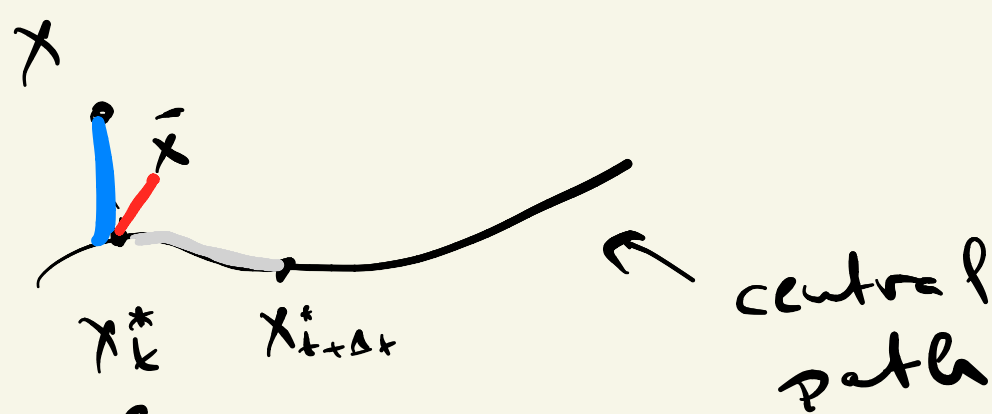
x_t is close to x_t^* $\longrightarrow$ $x_{t+\Delta t}$ is close to $x^*_{t+\Delta t}$

$\|\nabla f_t(x_t)\|_{(\nabla^2 f_t(x_t))^{-1}}$ is small $\longrightarrow$ $\|\nabla f_{t+\Delta t}(x_{t+\Delta t})\|_{(\nabla^2 f_{t+\Delta t}(x_{t+\Delta t}))^{-1}}$ is small

This is proven in two steps:

(remember that $x_t = x$ $\quad H = \nabla^2 f_t(x)$

$x_{t+\Delta t} = \bar{x}$ $\quad \bar{H} = \nabla^2 f_{t+\Delta t}(\bar{x})$



Lemma 1st step

If is small then is small

If $\|\nabla f_t(x)\|_{H^{-1}} \leq \delta \leq 0.001$ then

$$\|\nabla f_t(\bar{x})\|_{(\bar{H})^{-1}} \leq \frac{1}{\sqrt{2}} \|\nabla f_t(x)\|_{H^{-1}} \leq \frac{\delta}{\sqrt{2}}$$

Lemma 2nd step

If is small then is small

we want to increase t as much as possible

so that $\|\nabla f_t(\bar{x})\|_{(\bar{H})^{-1}} \leq \frac{\delta}{\sqrt{2}} \Rightarrow$

$$\|\nabla f_{t+\Delta t}(\bar{x})\|_{(\bar{H})^{-1}} \leq \delta$$

let's calculate Δt $\Longrightarrow$

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

$$= \|\Delta t c + tc + \nabla\varphi(\bar{x})\|_{(\bar{H})^{-1}} \leq$$

$$\leq \Delta t \|c\|_{(\bar{H})^{-1}} + \|tc + \nabla\varphi(\bar{x})\|_{(\bar{H})^{-1}} =$$

$$= \Delta t \|c\|_{(\bar{H})^{-1}} + \|\nabla f_t(\bar{x})\|_{(\bar{H})^{-1}} \leq$$

$$\leq \Delta t \|c\|_{(\bar{H})^{-1}} + \frac{\delta}{\sqrt{2}}$$

$$\Rightarrow \Delta t = \frac{\delta}{\sqrt{2}\|c\|_{(\bar{H})^{-1}}} \geq \frac{1}{200\|c\|_{(\bar{H})^{-1}}}$$

for small enough δ

Thus, now to bound the number of iterations before arriving close to $x^*_{m/\varepsilon}$ we need to upper bound $\|c\|_{(\bar{H})^{-1}}$ in terms of t

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

$$\leq \frac{1}{t}\left(\| \nabla f_t(x) \|_{H^{-1}} + \| \nabla \varphi(x) \|_{H^{-1}} \right) \leq$$

by lemma 1 — this needs a proof

$$\leq \frac{1}{t}\left(\delta + \sqrt{m} \right)$$

Now we are ready to bound the number of iterations.

Theorem

It takes at most $O\left(\sqrt{m} \log\left(\frac{m \cdot \|c\|_{H_0^{-1}}}{\varepsilon}\right)\right)$ iterations for $t \geq \frac{m}{\varepsilon}$. Where

$H_0 = \nabla^2 f_{t_0}(x_0)$, x_0 an initial solution such that $\| \nabla f_{t_0}(x_0) \|_{H_0^{-1}} \leq \delta$.

Proof

$$\Delta t \geq \frac{1}{200 \|c\|_{H^{-1}}} \geq \frac{t}{200(\delta + \sqrt{m})} \geq \frac{t}{400\sqrt{m}}$$

$\Rightarrow \bar{t} \approx \left(1 + \frac{1}{\sqrt{m}}\right) t \Rightarrow$ after k iterations

$$t^k \approx \left(1 + \frac{1}{\sqrt{m}}\right)^k \cdot \frac{1}{\|c\|_{H_0^{-1}}}$$

Thus by solving $t^k \geq m/\varepsilon$ we get the desired bound on k.

Let's say that we arrived at the last iteration. We know that

$$x_{m/\varepsilon} \text{ --- } x^*_{m/\varepsilon}$$ are close

$$c^T \cdot x^*_{m/\varepsilon} - c^T x^* \leq \varepsilon$$

solution of the unconstrained problem

Thus in order to bound the suboptimality of $x_{m/\varepsilon}$. We need to bound

$$c^T x_{m/\varepsilon} - c^T x^*_{m/\varepsilon}$$

Lemma

If $\|\nabla f_t(x_t)\|_{H^{-1}} \le \delta \le 0{,}001$,

then $c^T x_t - c^T x_t^* \le \frac{1}{t}(\delta + \sqrt{m}) \cdot 3\delta$

proof

$$c^T x_t - c^T x_t^* = c^T (x_t - x_t^*) \le$$

$$\le \|c\|_{H^{-1}} \cdot \|x_t - x_t^*\|_H \le$$

previous lemma $\quad \|\nabla f_t(x_t)\|_{H^{-1}} \le \delta$ + previous lemma

$$\le \frac{1}{t}(\delta + \sqrt{m}) \cdot 3\delta$$

If $t \geq m/\varepsilon \Rightarrow \frac{(\delta + \sqrt{m})3\delta}{t} \le \frac{(\delta + \sqrt{m})3\delta \cdot \varepsilon}{m}$

$$\le \varepsilon \quad \blacksquare$$

Thus, overall

optimal point of unconstrained problem

$$c^T x_{m/\varepsilon} - c^T x^* \le \varepsilon + \varepsilon = 2\varepsilon$$

Observations

- a crucial thing in our analysis is that we need to find a good starting point x_{t_0} which is very close to $x^*_{t_0}$ in such a way that $\|c\|_{H_0^{-1}}$ is small and the invariant $\|\nabla f_{t_0}(x_{t_0})\|_{H_0^{-1}} \leq \delta$ holds in the begining. This is in general not so easy to do and requires to run the Barrier method backwards to find, e.g a starting point close to the analytic center x_{ac} which solves $f_0(x)$. We refer to the excellent Lap Chi Lau's notes in Lecture 13 for that.

- If the analytic center can be computed and $\|c\|_{H_0^{-1}}$ can be bounded for example by poly(m) in most combinatorial problems. Then the Barrier method needs $O(\sqrt{m}\log {}^{m}/_{\varepsilon})$ iterations

<u>example</u> max-flow (undirected)

$$\max f_{st}$$
$$\text{s.t } Bf = 0$$
$$0 \le f_e \le 1 \ \forall e \in E$$

then the analytic center can be computed and $\|c\|_{H_0^{-1}} = O(\text{poly}(m))$

$\Rightarrow$ $\sqrt{m}\log {}^{m}/_{\varepsilon}$ iterations to solve it.

$O(m)$ in each iteration thanks to Laplacian solvers $\Rightarrow$ $\tilde{O}(m^{3/2}\log {}^{1}/_{\varepsilon})$