Interior Point:

$$\min\ c^Tx \qquad \Big| \qquad \max\ b^Ty$$
$$Ax = b \qquad \Big| \qquad \text{st } A^Ty + s = c$$
$$x \geq 0. \qquad \Big| \qquad s \geq 0.$$

$\mathcal{P} = \{x \mid Ax = b,\ x \geq 0\}.$

$\mathcal{D} = \{s \mid \exists y \text{ st } A^Ty + s = c,\ s \geq 0\}.$

Fact: $c^Tx - b^Ty = s^Tx \geq 0.$

So. weak duality. And. given any x, s know that $c^Tx = b^Ty + s^Tx \leq c^Tx^* + s^Tx.$

$\Rightarrow$ loss $\leq s^Tx.$

Dont worry about the y just the slack!

Central Path:

View 1;

pair $x, s \in \mathcal{P}^o \times \mathcal{D}^o$

st $\bar{x}\bar{s} = t\mathbb{1}$

paramelized by $t > 0$

View 2: solve $\min\ c^Tx - \sum_i \mu_i \log x_i$ st $Ax = b.$ ↑ barrier to x being negative

KKT conditions:

∇ of the Lagrangian $\bar{c} - \bar{\mu}/\bar{x} = A^Ty$

$\Leftrightarrow -A^Ty + c = \bar{\mu}/\bar{x}$

$\Leftrightarrow \bar{s} = \bar{\mu}/\bar{x}$

$\Leftrightarrow \bar{x} \cdot \bar{s} = \bar{\mu}$

So if we solve

$$\min\ c^Tx - t\,\text{Log Barrier}(x)$$
$$Ax = b.$$

the optimal point gives x_t. (and dual gives s_t)

②

OK: in fact have bijection between $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ \leftrightarrow \mu \in \mathbb{R}_{>0}$.

using this idea.

IPM (path following):

how to do this? discuss.

Start with t v. large so that essentially solve only for the barriers minimization

Now:

given $\bar{\mu}_{old} = t\mathbb{1}$

sps have $\bar{x}, \bar{s}$ for it

Move to nearby $\bar{\mu}$

Find $\bar{x}_\mu \bar{s}_\mu$ for it.

just solve a linear system.

May not get exactly $(t-\epsilon)\mathbb{1}$ but something close enough to it good enough to continue the process.

Goal: $\boxed{\bar{\mu}^{old} \text{ close enough to } t \cdot \mathbb{1} \longrightarrow \bar{\mu}^{new} \text{ close enough to } t(1-\tfrac{1}{\sqrt{n}})\mathbb{1}}$

OK: sps $\bar{x}, \bar{s}$ sat μ^{old} so $\bar{x}, \bar{s} \in \mathcal{P}^\circ \times \mathcal{D}^\circ$. Now for new $\bar{\mu}$

Want $(\bar{x}+\bar{\delta}_x)(\bar{s}+\bar{\delta}_s) = \bar{\mu}$ ← drop product terms. $\bar{\delta}_x \bar{\delta}_s$

$A(x+\delta x) = b$

$(s+\delta s) + A(y+\delta y) = c$

$(\bar{x}+\bar{\delta}_x), (\bar{s}+\bar{\delta}_s) \in \mathbb{R}_{>0}$ ← drop for the time being

↑ strict

So "relax" the problem (drop the quadratic term $\bar{\delta}_s \delta_x$ and the $R_{\geq 0}$) ③

Solve it using $\boxed{\text{Linear System Solving}}$.

Then see how badly we messed up the "close to $t\mathbb{1}$" constraint.

Go from there

How? ~~[illegible]~~ $\underset{= \text{old } \bar{\mu}}{\bar{x}\cdot\bar{s}} + \bar{\delta}_x \cdot s + \bar{x}\cdot\delta_s = \underset{\text{new } \bar{\mu} \text{ vector}}{\bar{\mu}}$

$$A\bar{x} + A\bar{\delta}_x = b$$

$$A^T(\bar{y} + \bar{\delta_y}) + \bar{s} + \delta_s = c.$$

but $A^T\bar{y} + \bar{s} = c$ and $A\bar{x} = b$ already (we started with a sol^n)!!

So.

$$\boxed{\begin{aligned} \bar{x}\bar{s} + \bar{\delta}_x\bar{s} + \bar{x}\bar{\delta}_s &= \bar{\mu} \\ A\bar{\delta}_x &= 0. \\ A^T\delta_y + \delta_s &= 0. \end{aligned}}$$

$\to$ $\bar{\delta}_x \cdot \bar{s} + \bar{x}\cdot\bar{\delta}_s = \bar{\delta}_\mu = (\bar{\mu} - \bar{x}\bar{s})$

Linear system.

<u>Claim</u>: Sps A has full row rank and $(x,s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ then

~~define~~ solve the Lin System.

no proof. See the ee-Vempala notes

if $\|\bar{\delta}_\mu / \bar{\mu}\|^2_{\bar{\mu}} < \min_i \mu_i$ then $\bar{x} + \bar{\delta}_x, \bar{s} + \bar{\delta}_s \in \mathcal{P}^\circ \times \mathcal{D}^\circ$.

if small change to μ then new solution is still feasible.

wierd norm. $\|z\|_v = \sqrt{z^T \mathrm{Diag}(v) z}$

so $\|\bar{\delta}_\mu/\bar{\mu}\|^2_{\bar{\mu}} = \sum_i \frac{(\delta_\mu)_i}{\mu_i} \cdot \mu_i \cdot \frac{(\delta_\mu)_i}{\mu_i} = \sum_i \frac{(\delta_\mu)_i^2}{\mu_i}$

Great! But how much ~~reduction~~ progress did we make?

sps. $\bar{\delta}_\mu = \alpha . \bar{\mu}$ (↑ scalar) so want $\sum_i \frac{\alpha^2 . \mu_i^2}{\mu_i} \overset{?}{\leq} \min_i \mu_i$

$$\Rightarrow \alpha \leq \sqrt{\frac{\min_i \mu_i}{\sum_i \mu_i}} = \sqrt{\frac{\min_i \mu_i}{\|\mu\|_1}}.$$

Great: So reduce the size of $\bar{\mu}$ ~~...~~ the most when $\mu = \mathbb{1}$.

reduce by $\frac{1}{\sqrt{n}}$.

But if μ very imbalanced, then not much progress. :(

So: new goal: stay close to $t\mathbb{1}$ at all times.

Can we? why not by construction? Remember: actual

$$(\bar{x}+\bar{\delta}_x)(\bar{s}+\bar{\delta}_s) = \bar{\mu} + \bar{\delta}_x \cdot \bar{\delta}_s$$

So even if we were exactly at $\bar{\mu}_{old} = \bar{x} \cdot \bar{s}$

we may not be at ~~...~~ $\bar{\mu}$

Claim 2: suppose $(\bar{x}_0, \bar{s}_0) \in P^\circ \times D^\circ$ and $\|\bar{x}_0 \bar{s}_0 - t_{start}\mathbb{1}\|_2 \leq \frac{t_{start}}{4}$.

then can maintain

$$(x_t, s_t) \in P^\circ \times D^\circ \quad st \quad \|\bar{x}_t \bar{s}_t - t\mathbb{1}\|_2 \leq t/6 . \cancel{st/4}$$

Pf: Simple algebra, uses Cauchy-Schwarz.

~~But the~~ But not edifying, so skip here

Algo: Input $(x_0, s_0) \in \mathcal{P}^0_\circ \times \mathcal{D}^0$ $\quad \|\bar{x}_0 \bar{s}_0 - t_0 \mathbb{1}\| \le t_0/4$.

$\alpha = \frac{1}{16\sqrt{n}}$. $\quad n = \#$columns of A. (full row rank A).

repeat: $\quad$ stopping condition! how to choose?

$$t' = \max\left(\frac{t}{1+\alpha}, t_{end}\right)$$

$$\bar{\mu} = \bar{X}_t \bar{S}_t, \quad \delta_\mu = t'\mathbb{1} - \bar{\mu}$$

Solve ~~for~~

$$\begin{cases} \bar{X}\bar{\delta}_s + \bar{S}\bar{\delta}_x = \delta_\mu \\ A\bar{\delta}_x = 0 \\ A^T\delta_y + \delta_s = 0. \end{cases}$$

Set $x_{t'} \leftarrow x_t + \delta_x \quad s_{t'} \leftarrow s_t + \delta_s$.

$t \leftarrow t'$

until $t = t_{end}$.

———— x ————

Final solution (x_{end}, s_{end})

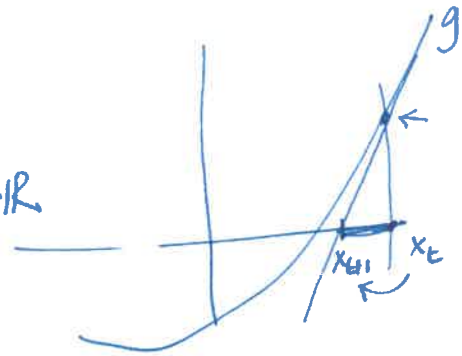
satisfies $\|\bar{X}_{end} \cdot \bar{S}_{end} - t\mathbb{1}\|_2 \le t/6$

$$\Rightarrow \bar{x}_{end}^T \bar{s}_{end} \le 2tn$$

$$\Rightarrow c^T x_{end} \le c^T x^* + 2tn$$

so set t to be small enough so that this loss is not large.

Newton's Method



- For finding a zero of $g: \mathbb{R} \to \mathbb{R}$

$$x_{t+1} = x_t - \frac{g(x_t)}{g'(x_t)}$$

- To find a minimizer of $f: \mathbb{R} \to \mathbb{R}$

apply to $g = f'$. $\Rightarrow$ $x_{t+1} = x_t - \dfrac{f'(x_t)}{f''(x_t)}$

Same intuition for higher dimensions: $f \in \mathbb{R}^n \to \mathbb{R}$, $\underset{=\nabla f}{g} \in \mathbb{R}^n \to \mathbb{R}^n$

$$x_{t+1} = x_t - H_{x_t}^{-1} \nabla f(x_t)$$

Another way of seeing this:-

take $f(x) \cong \underbrace{f(x_t) + \langle \nabla f(x_t), x - x_t \rangle + \frac{1}{2} \langle x - x_t, H_{x_t}(x - x_t) \rangle}_{S(x)}$

the 2nd order approximation,

and minimize it. So get $\nabla S(x_{t+1}) = 0$

i.e. $\nabla f(x_t) + H_{x_t}(x_{t+1} - x_t) = 0$ or $x_{t+1} - x_t = -H_{x_t}^{-1} \nabla f(x_t)$

which is the Newton update

N.b. Newton's method does not always converge — well known examples where it diverges, and also classes of problems where it does converge.

Quadratic Convergence of Newton Raphson

To be continued ...