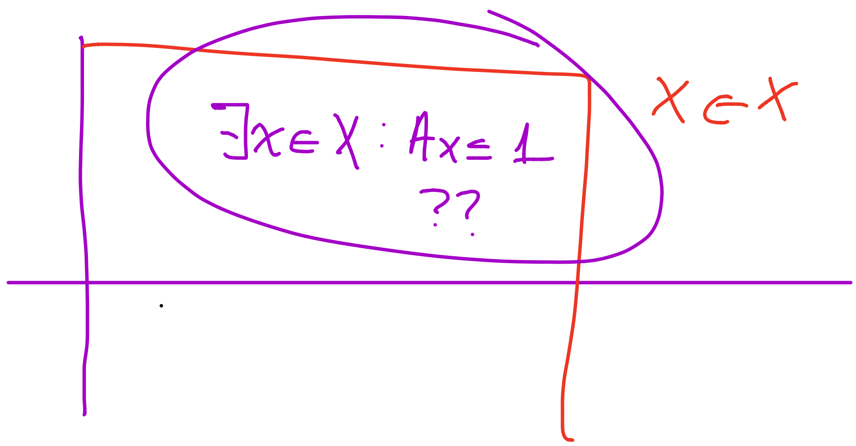
Last time:

Optimize



Came up with a robust way to answer with

$$f_\mu(x) = \mathrm{smax}_\mu (Ax)$$ softmax violation

$$f(x) = \max_{1 \le i \le m} (Ax)_i - 1$$ max violation

Alg: $\bar{x}_0 = 0$

$$x_{t+1} = \arg\min_{x \in X} \langle \nabla f_\mu(\bar{x}_t), x \rangle$$

It turns out we can use

$$x_{t+1} = \underline{\underline{\text{any}}} \left\{ x \in X : \langle f_\mu(\bar{x}_t), x \rangle \leq 1 \right\}$$

So either x_{t+1} does not exists (no feasible sol.) or ge for $T = O\left(\frac{\rho^2 \log n}{\varepsilon^2}\right)$ iterations

then

$$A \overline{x_T} \leq 1 + \varepsilon$$

$$\overline{x_T} \in X$$

Relation to the saddle point problem.

$$\underbrace{\min_{x \in X} \underbrace{\max_{p \in \Delta_m} p^T A x - 1}_{f(x)} + \mu H(p)}$$

$f'_\mu(z)$

Could have also argued through the dual/primal player relationship.

$$\max_{p \in \Delta_m} \left(\min_{x \in X} p^T A x - 1 \right)$$

Less illustrative of the regularization method.

<u>Rmk:</u> This method extends $\|\cdot\|_\infty$ with any convex constraints.

for approximate solutions in quadratic programming.

$$f_\mu(z) = \mu \log \sum \exp\left(z_i/\mu \right)$$

$$\nabla_i f_\mu(z) = \frac{e^{z_i}}{\sum_{i=1}^{n} e^{z_i/\mu}}$$

$$\nabla^2_{ii} f_\mu(z) = \left(\frac{e^{z_i/\mu}}{\sum e^{z_i/\mu}} - \frac{e^{z_i/\mu}\, e^{z_i/\mu}}{\left(\sum e^{z_i/\mu}\right)^2} \right) \frac{1}{\mu}$$

$$= \frac{1}{\mu}\left(p_i - p_i^2\right)$$

$$\nabla^2_{ij} f_\mu(z) = \frac{-1}{\mu} \frac{e^{z_i/\mu}\, e^{z_j/\mu}}{\left(\sum e^{z_i/\mu}\right)^2}$$

$$= -\frac{1}{\mu} p_i p_j$$

$$\nabla^2 f_\mu(z) = \frac{1}{\mu}\left(\mathrm{diag}(p) - pp^T\right)$$

* Cumulant generating function.

$$f_\mu(z+h) = f_\mu(z) + \langle \nabla f_\mu(z), h \rangle + \underbrace{\frac{h^T \nabla^2 f(\tilde{z}) h}{2}}_{\le \frac{1}{2\mu}[\mathbb{E}(h)]^2 \le \frac{1}{2\mu}\|h\|_\infty^2} + O(h^3)$$

So

$$f_\mu(z+h) \le f_\mu(z) + \langle - \rangle + \frac{1}{2\mu}\|h\|_\infty^2$$

So we have $\|\cdot\|_\infty$ smoothness.

<u>How does this work for SDPs?</u>

$$\sum y_i A^{(i)} \preceq I \qquad y \in \mathbb{R}^m$$

$$f_\mu(y) = \lambda_{max}\left(\sum y_i A_i\right) - 1$$

$$f_\mu(y) = \max_{\substack{I \cdot X = 1 \\ X \succeq 0}} X \cdot \left(\sum y_i A_i\right) - 1 + \mu H(X)$$

$X \succeq 0$

$$H(X) = \mathrm{Tr}(X \log X) = H(\lambda_i(X))$$

$$\nabla f_\mu(x) = \frac{e^{\sum y_i A_i / \mu}}{I \cdot e^{\sum y_i A_i / \mu}}$$