Convex Optimization

Prof. Nati Srebro

Lecture 11:

Equality Constrained Optimization
Implicit and Explicit Constraints

Read: Boyd and Vandenberghe Sections 10.1-10.4,11.1-11.2

Methods for Constraint Optimization

Methods for Constrained Optimization

$$\begin{array}{lll} \min_{x \in \mathbb{R}^n} & f_0(x) & \\ s.t. & f_i(x) \leq 0 & i = 1..m \\ & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{array}$$

Using access to:

- 2nd order oracle for f_0, f_i

$$x \mapsto f_0(x), \nabla f_0(x), \nabla^2 f_0(x)$$
$$x \mapsto f_i(x), \nabla f_i(x), \nabla^2 f_i(x)$$

- Explicit access to A, b

Can obtain with 1st order access to $h_i(x) = \langle a_i, x \rangle - b_i$

$$b_i = h_i(0) \qquad\qquad a_i = \nabla h_i(0)$$

Equality Constrained Optimization

$$\boxed{\begin{array}{lll} \min\limits_{x\in\mathbb{R}^n} & f(x) & \\ s.t. & Ax = b & \qquad A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p \end{array}}$$

Using access to:

- 2nd order oracle for f
$$x \mapsto f(x), \nabla f(x), \nabla^2 f(x)$$

- Explicit access to A, b

Change of variables

$$\boxed{\begin{array}{lll} \min_{x \in \mathbb{R}^n} & f(x) & \\ s.t. & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{array}}$$

$$\{Ax = b | x \in \mathbb{R}^n\} = \left\{Fz + x^{(0)} \middle| z \in \mathbb{R}^{n-p}\right\}$$

Where $Ax^{(0)} = b$ and $null(A) = image(F)$

i.e. $AF = 0$ and $rank(F) = n - rank(A)$

$$\boxed{\min_{z \in \mathbb{R}^{n-p}} f(Fz + x^{(0)})}$$

Input: feasible $x^{(0)}$

Do: Use Newton to miminize $\tilde{f}(z) = f\left(Fz + x^{(0)}\right)$ starting from $z^{(0)} = 0$

Return $Fz^{(k)} + x^{(0)}$

Many F are possible. Does it matter which we use?

No—Newton is affine invariant

f self-concordant ➔ $\tilde{f}$ self-concordant ➔ $k = O\left(\left(f\left(x^{(0)}\right) - f(x^*)\right) + \log\log\frac{1}{\epsilon}\right)$

Optimization: Solving KKT

$$\begin{array}{lll} \min\limits_{x \in \mathbb{R}^n} & f(x) & \\ s.t. & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{array}$$

- KKT conditions: $Ax^* = b \qquad \nabla f(x^*) + A^\top \nu^* = 0$
- For $f(x) = \frac{1}{2} x^\top H x + c^\top x$:

$$Ax^* = b \text{ and } Hx^* + c + A^\top \nu^* = 0$$

i.e.

$$\begin{bmatrix} H & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

- If no solution?
  - No solution to $Ax = b$ ➔ primal infeasible
  - Solution to $Ax = b$, but not to KKT ➔ no dual feasible

    ➔ $p^* = d^* = -\infty$ ➔ primal unbounded

Newton on KKT

$$\begin{array}{lll} \min_{x \in \mathbb{R}^n} & f(x) & \\ s.t. & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{array}$$

- KKT conditions: $Ax^* = b \qquad \nabla f(x^*) + A^\top \nu^* = 0$
- For non-quadratic $f(x)$:

$$f\left(x^{(k)} + \Delta x\right) \approx \frac{1}{2} \Delta x^\top \nabla^2 f\left(x^{(k)}\right) \Delta x + \langle \nabla f\left(x^{(k)}\right), \Delta x \rangle + f\left(x^{(k)}\right)$$

- Linearized KKT:

$$A\left(x^{(k)} + \Delta x\right) = b \qquad \nabla^2 f\left(x^{(k)}\right) \Delta x + \nabla f\left(x^{(k)}\right) + A^\top \nu = 0$$

$$\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) \\ b - Ax^{(k)} \end{bmatrix}$$

Equality Constrained Newton

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad Ax = b \qquad A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$$

Init: Feasible $x^{(0)}$ ($x^{(0)} \in dom(f)$ and $Ax^{(0)} = b$)

Iterate: Solve: $\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{(k)} \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) \\ b - Ax^{(k)} \end{bmatrix}$

Stop if $\lambda\left(x^{(k)}\right)^2 = \Delta x^{(k)} \nabla^2 f\left(x^{(k)}\right) \Delta x^{(k)} \leq \epsilon$

Set $t^{(k)}$ by backtracking linesearch

$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}$

- Equivalent to unconstrained Newton on $\tilde{f}(z) = f\left(Fz + x^{(0)}\right)$
- $\lambda(z) = \sqrt{\nabla \tilde{f}(z)^\top \nabla^2 \tilde{f}(z)^{-1} \nabla \tilde{f}(x)} = \sqrt{\Delta z \nabla^2 \tilde{f}(z) \Delta z}$
- With equality constraints: $\lambda\left(x^{(k)}\right) = \sqrt{\Delta x^{(k)} \nabla^2 f\left(x^{(k)}\right) \Delta x^{(k)}}$

Equality Constrained Newton

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad Ax = b \qquad A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$$

Init — Feasible $x^{(0)}$ ($x^{(0)} \in dom(f)$ and $Ax^{(0)} = b$)

Iterate — Solve: $\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{(k)} \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) \\ b - Ax^{(k)} \end{bmatrix}$

Stop if $\lambda\left(x^{(k)}\right)^2 = \Delta x^{(k)} \nabla^2 f\left(x^{(k)}\right) \Delta x^{(k)} \leq \epsilon$

Set $t^{(k)}$ by backtracking linesearch

$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}$

Is $x^{(k+1)}$ feasible?

$$Ax^{(k+1)} = Ax^{(k)} + t^{(k)} \underbrace{A\Delta x^{(k)}}_{A\Delta x^{(k)} = b - Ax^{(k)} = 0} = b$$

If $x^{(k)}$ feasible

Equality Constrained Newton

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$s.t. \quad Ax = b \qquad A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$$

Init Feasible $x^{(0)}$ ($x^{(0)} \in dom(f)$ and $Ax^{(0)} = b$)

Iterate Solve: $\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{(k)} \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) \\ b - Ax^{(k)} \end{bmatrix}$

Stop if $\lambda\left(x^{(k)}\right)^2 = \Delta x^{(k)} \nabla^2 f\left(x^{(k)}\right) \Delta x^{(k)} \leq \epsilon$

Set $t^{(k)}$ by backtracking linesearch

$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}$

Analysis?

Equivalent to unconstrained Newton after change of variable

➔ If $f(\cdot)$ convex and self concordant,

$$k = O\left(\left(f_0\left(x^{(0)}\right) - f_0(x^*)\right) + \log\log\frac{1}{\epsilon}\right)$$

Starting at a Feasible Point

- Is it always easy to find feasible $x^{(0)}$?
  (i.e. $f\left(x^{(0)}\right) < \infty$, $Ax^{(0)} = b$)
- E.g.:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & -\sum_{i=1}^n \log(x_i) \\ \text{s.t.} & Ax = b \end{array}$$

$$\begin{array}{ll} \text{find} & x \in \mathbb{R}^n \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}$$

- Requires finding a feasible point for an LP
- As hard as LP optimization:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & c^\top x \\ \text{s.t.} & Ax = b \\ & Gx \leq h \end{array} \quad \xrightarrow[\text{over } \theta]{\text{bisection search}} \quad \begin{array}{ll} \min\limits_{\substack{x \in \mathbb{R}^n, z \in \mathbb{R}^m \\ t \in \mathbb{R}}} & -\sum_{i=1}^m \log(z_i) - \log(t) \\ \text{s.t.} & Ax = b \\ & Gx + z = h \\ & t = c^\top x - \theta \end{array}$$

Infeasible Start Newton

- KKT conditions:

$$r_P(x) = Ax - b = 0 \qquad r_D(x, \nu) = \nabla f(x) + A^\top \nu = 0$$

- Linearized about $x = x^{(k)} + \Delta x$ and $\nu = \nu^{(k)} + \Delta \nu$

$$\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) - A^\top \nu^{(k)} \\ b - Ax^{(k)} \end{bmatrix}$$

- Compare with linearizing about $x = x^{(k)} + \Delta x$ and solving for Δx,ν:

$$A\left(x^{(k)} + \Delta x\right) = b \qquad \nabla^2 f\left(x^{(k)}\right)\Delta x + \nabla f\left(x^{(k)}\right) + A^\top \nu = 0$$

$$\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) \\ b - Ax^{(k)} \end{bmatrix}$$

- Same equations, but now we use $\Delta \nu$ instead of ν.

Infeasible Start Newton

- KKT conditions:

$$r_P(x) = Ax - b = 0 \quad r_D(x, \nu) = \nabla f(x) + A^\top \nu = 0$$

- Linearized about $x = x^{(k)} + \Delta x$ and $\nu = \nu^{(k)} - \Delta\nu$

$$\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) - A^\top \nu^{(k)} \\ b - Ax^{(k)} \end{bmatrix}$$

Init	$x^{(0)} \in dom(f)$ and any $\nu^{(0)}$
Iterate	Solve: $\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) - A^\top \nu^{(k)} \\ b - Ax^{(k)} \end{bmatrix}$
	Set $t^{(k)}$ by backtracking linesearch on $\lVert r(x, \nu) \rVert_2$
	(while $\left\lVert r\left(x^{(k)} + t\Delta x^{(k)}, \nu^{(k)} + t\Delta \nu^{(k)}\right) \right\rVert > (1 - \alpha t) \lVert r\left(x^{(k)}, \nu^{(k)}\right) \rVert$, set $t \leftarrow \beta t$)
	$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}$, $\nu^{(k+1)} \leftarrow \nu^{(k)} + t^{(k)} \Delta \nu^{(k)}$

Use $\lVert r(x, \nu) \rVert_2 = \lVert [r_P(x) \;\; r_D(x, \nu)] \rVert_2$ as measure of progress

(See Boyd and Vandenberghe p. 533)

Infeasible Start Newton

$$r_P(x) = Ax - b = 0 \quad r_D(x, \nu) = \nabla f(x) + A^\top \nu^* = 0$$
$$r(x, \nu) = [r_P \;\; r_D]$$

Init	$x^{(0)} \in dom(f)$ and any $\nu^{(0)}$
Iterate	Stop if $Ax^{(k)} = b$ and $r\left(x^{(k)}, \nu^{(k)}\right) \leq \epsilon$
	Solve: $\begin{bmatrix} \nabla^2 f\left(x^{(k)}\right) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \nu \end{bmatrix} = \begin{bmatrix} -\nabla f\left(x^{(k)}\right) - A^\top \nu^{(k)} \\ b - Ax^{(k)} \end{bmatrix}$
	Set $t^{(k)}$ by backtracking linesearch on $\|r(x)\|_2$
	$x^{(k+1)} \leftarrow x^{(k)} + t^{(k)} \Delta x^{(k)}, \nu^{(k+1)} \leftarrow \nu^{(k)} + t^{(k)} \Delta \nu^{(k)}$

Will $x^{(k)}$ be feasible?

- We will always have $f\left(x^{(k)}\right) < \infty$ (otherwise we backtrack)
- After a "full" step, with $t = 1$:
  $$Ax^{(k+1)} = A\left(x^{(k)} + 1 \cdot \Delta x^{(k)}\right) = Ax^{(k)} + \left(b - Ax^{(k)}\right) = b$$
- If $Ax^{(k)} = b$ then:
  $$Ax^{(k+1)} = A\left(x^{(k)} + t\Delta x^{(k)}\right) = Ax^{(k)} + t\left(b - Ax^{(k)}\right) = b$$
- Conclusion: once we enter quadratic phase, $x^{(k)}$ feasible

Solving an LP

$$\min_{x \in \mathbb{R}^n} \; c^\top x \quad \text{s.t.} \quad Ax = b, \quad Gx \le h$$

$$\xrightarrow{\text{bisection search over } \theta}$$

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^m, t \in \mathbb{R}} \; -\sum_{i=1}^{m} \log(z_i) - \log(t) \quad \text{s.t.} \quad Ax = b, \quad Gx + z = h, \quad t = c^\top x - \theta$$

For each θ,

Solve using infeasible start equality constraint Newton,

initializing to: $z^{(0)} = (1,1,\dots,1), t^{(0)} = 1$

$x^{(0)} = 0$ (or some solution to $Ax = b$)

- Runtime analysis? How many Newton iterations?

$$O\left(\left(f_0\left(x^{(0)}\right) - f_0(x^*)\right) + \log\log\frac{1}{\epsilon}\right) ???$$

- If $\theta < p^*$, no feasible solution.
  How do we know? When do we stop Newton?

Equality Constrained Optimization

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f_0(x) \\ s.t. \quad & Ax = b \qquad\qquad A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p \end{aligned}$$

Using access to:

- 2^{nd} order oracle for f_0

$$x \mapsto f_0(x), \nabla f_0(x), \nabla^2 f_0(x)$$

- Explicit access to A, b
- $x^{(0)} \in dom(f_0)$

- Analysis:
  - If f_0 convex and self-concordant
  - $Ax^{(0)} = b$

  Then, feasible ϵ-suboptimal solution at:

$$k = O\left(\left(f_0\left(x^{(0)}\right) - f_0(x^*)\right) + \log\log\frac{1}{\epsilon}\right)$$

Adding Inequality Constraints

$$
\begin{array}{lll}
\min_{x \in \mathbb{R}^n} & f_0(x) & \\
s.t. & f_i(x) \leq 0 & i = 1..m \\
 & Ax = b & A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p
\end{array}
$$

Using access to:

- 2nd order oracle for f_0, f_i

$$x \mapsto f_0(x), \nabla f_0(x), \nabla^2 f_0(x)$$
$$x \mapsto f_i(x), \nabla f_i(x), \nabla^2 f_i(x)$$

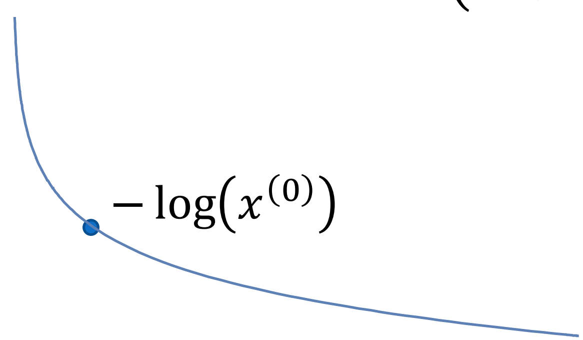
- Explicit access to A, b
- $x^{(0)} \in dom\big(f_0(x)\big), x_i^{(0)} \in dom\big(f_i(x)\big)$

but for now, assume access to feasible $x^{(0)}$

Optimization with Implicit Constraints

$$\boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & -\sum_{i=1}^{n}\log(x_i) \\ \text{s.t.} & Ax = b \end{array}}$$

- Generally not a problem for descent methods: if $x^{(k)} \in interior\big(dom(f_0)\big)$, linesearch will find $x^{(k+1)} = x^{(k)} + t^{(\mathrm{k})}\Delta x^{(k)} \in dom(f_0)$
- In fact, never leave sublevel set $\left\{x \middle| f_0(x) \le f_0\big(x^{(0)}\big)\right\} \subset dom(f_0)$
- Runtime:
  - With Gradient Descent $k \propto \dfrac{\max\limits_{x\in S} f_0''(x)}{\min\limits_{\mathrm{x}\in S} f_0''(x)} \propto \left(\dfrac{x_{\max}}{x_{\min}}\right)^2$
  - But with Newton, for self concordant: $k \propto \Big(f_0\big(x^{(0)}\big) - f_0(x^*)\Big) \propto \log\left(\dfrac{x_{\max}}{x_{\min}}\right)$



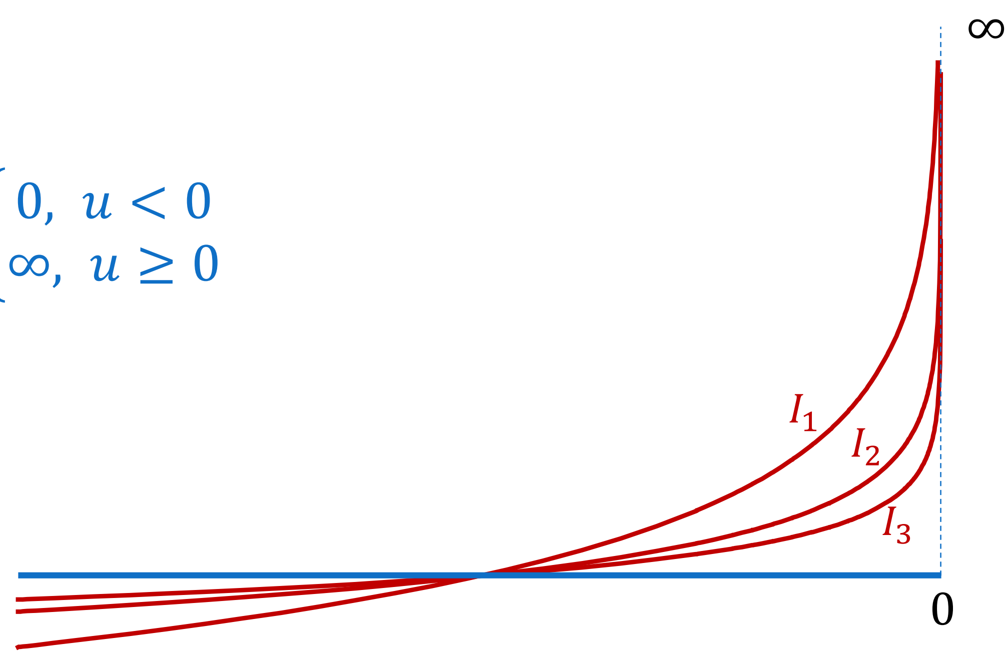
Making Constraints Implicit

$$\boxed{\begin{array}{ll} \min_{x\in\mathbb{R}^n} & f_0(x) \\ s.t. & f_i(x) \le 0 \\ & Ax = b \end{array}}$$

$$|||$$

$$\boxed{\begin{array}{ll} \min_{x\in\mathbb{R}^n} & f_0(x) + \sum_{i=1}^m I\big(f_i(x)\big) \\ s.t. & Ax = b \end{array}} \approx \boxed{\begin{array}{ll} \min_{x\in\mathbb{R}^n} & f_0(x) + \sum_{i=1}^m I_t\big(f_i(x)\big) \\ s.t. & Ax = b \end{array}}$$

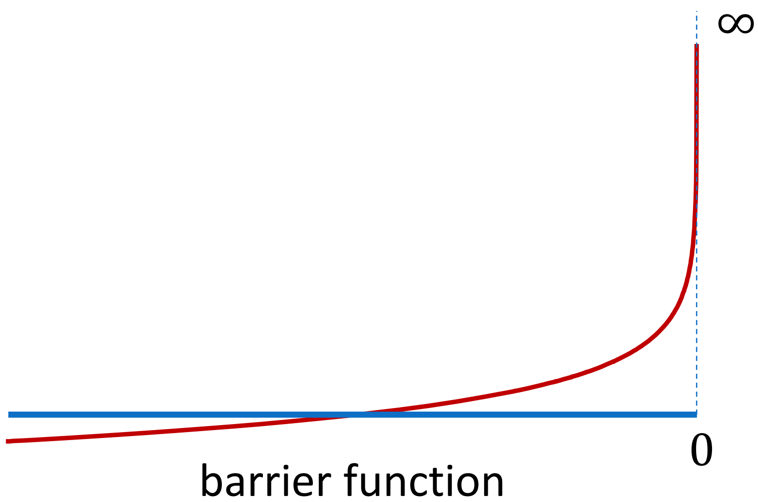
- $I(u) = \begin{cases} 0, & u \le 0 \\ \infty, & u > 0 \end{cases}$
- $I_t(u) = -\frac{1}{t}\log(-u) \xrightarrow{t\to\infty} \begin{cases} 0, & u < 0 \\ \infty, & u \ge 0 \end{cases}$

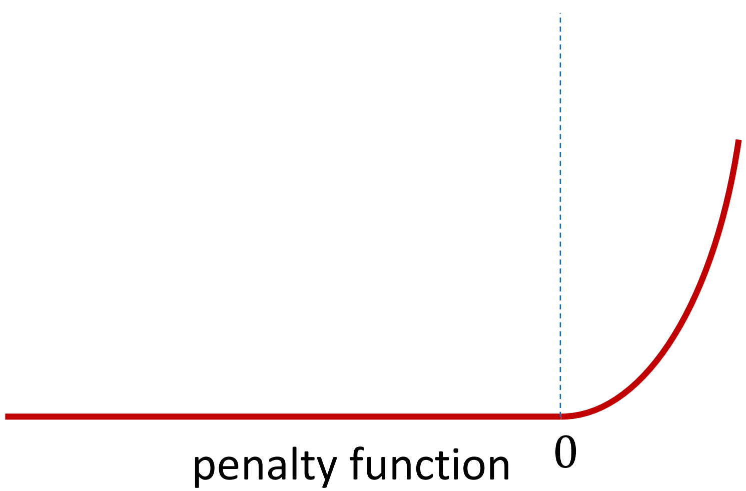


Barrier Methods

$$\boxed{\begin{array}{ll} \min_{x \in \mathbb{R}^n} & f_0(x) + \sum_{i=1}^{m} I_t\big(f_i(x)\big) \\ s.t. & Ax = b \end{array}}$$

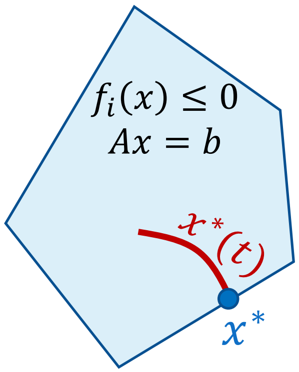
- Could use other barriers or penalties
- Log barriers $I_t(u) = -\log(-u)$:
    - Clean analysis with good guarantees
    - Connection with dual (as we will see)
    - Works well in practice





Log Barrier

$$\boxed{\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f_0(x) - \frac{1}{t}\sum_{i=1}^m \log\big(-f_i(x)\big) \\ s.t. & Ax = b \end{array}}$$



- Central Path: $\{x^*(t) | t > 0\}$

- $x^*(t)$ strictly feasible for original (constrained) problem
- We will show: $x^*(t) \to x^*$

- How suboptimal (for original problem) is $x^*(t)$?
- How do we set t?
- What's the complexity of optimizing with that t?