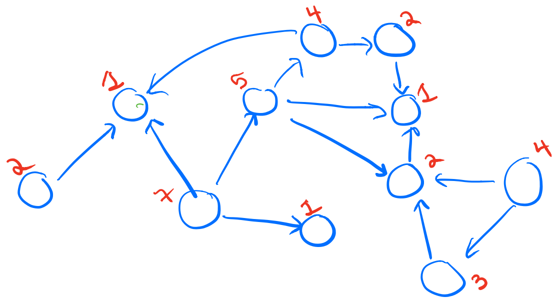
A cool application: reachability

Input: directed graph $G=(V,E)$

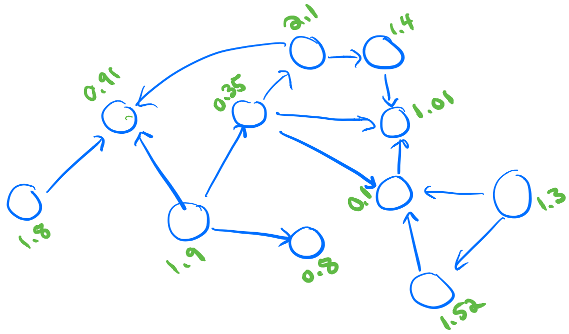
Output: $\forall v \in V$, $|S(v)|$ where $S(v)$: set of nodes reachable from v

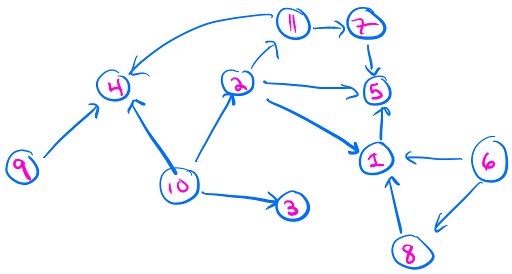


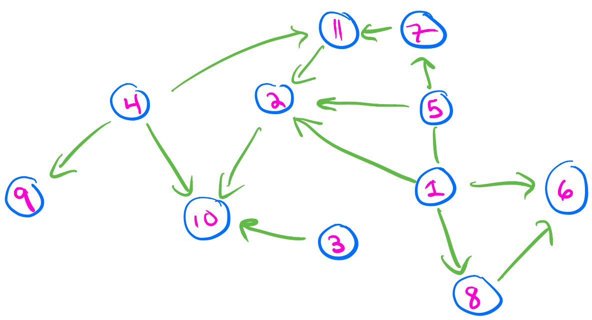
BFS/DFS $O(mn)$

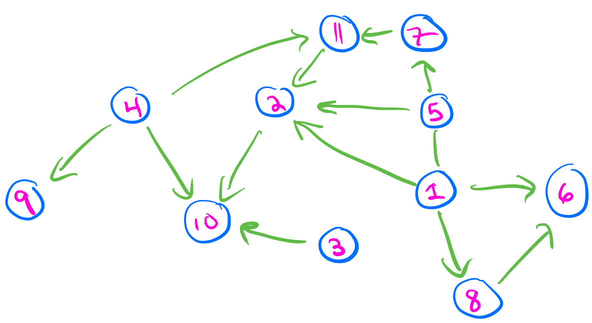
Randomized alg $O(m \log n)$ approximately correct

Today
- Announcements
  - Stirling's Approx
  - hw 2 - google form
- Reachability
- intro to LP
- randomized rounding of LPs









Reverse edge directions.

While $V \neq \emptyset$

let $i :=$ min rank of nodes in V

Perform search to find V_i, set of nodes reachable from node of rank i

$\forall v \in V_i$, let $s(v) := R(v_i)$

$V := V \setminus V_i$

Remove from E all edges incident to nodes in V_i

$$\boxed{\begin{aligned}
&\Pr\left(X \le (1-\delta)\mu\right) \le e^{-\frac{\delta^2\mu}{2}} && 0 \le \delta \le 1 \\
&\Pr\left(X \ge (1+\delta)\mu\right) \le e^{-\mu\left[(1+\delta)\ln(1+\delta)-\delta\right]} \le \begin{cases} e^{-\frac{\delta^2\mu}{2+\delta}} & \delta > 0 \\ e^{-\frac{\delta^2\mu}{3}} & 0 < \delta \le 1 \end{cases}
\end{aligned}}$$

Similarly, we want to show $\exists$ constant c_2 s.t.

$$\frac{k}{1.69} < \boxed{\frac{1}{\hat{s}(v)} \leq c_2 k} \quad \text{w.h.p.}$$

consider single $s(v)$

$$\Pr\left(s(v) \geq \frac{1}{c_2 k}\right) = e^{-k \cdot \frac{1}{c_2 k}} = e^{-\frac{1}{c_2}}$$

$$N_i' = \mathbb{1}_{s_i(v) \geq \frac{1}{c_2 k}} \qquad E\left(\sum_{i=1}^{t} N_i'\right) = t e^{-\frac{1}{c_2}}$$

choose c_2 so $\Pr\left(\sum_{i=1}^{t} N_i' < \frac{t}{e}\right)$ small

$\underbrace{s_1(v) \geq \dots s_{\frac{t}{e}}(v) \geq}_{\text{want all } \geq \frac{1}{c_2 k}}$

want this to be $(1-\delta) t e^{-\frac{1}{c_2}}$

For example, if we take $c_2 = 2 \Rightarrow$

$$e^{-\frac{1}{c_2}} = e^{-1/2} \approx 0.606$$

whereas $\frac{1}{e} \approx 0.3678$

$$\Rightarrow \quad \frac{t}{e} \simeq 0.6 \, t e^{-1/2}$$

$$\Rightarrow \delta \approx 0.4$$

$0 \leq \delta \leq 1$

$\Pr\left(X \leq (1-\delta)\mu\right) \leq e^{-\frac{\delta^2 \mu}{2}}$

$$\Rightarrow \Pr\left(\sum_{i=1}^{t} N_i' < \frac{t}{e}\right) \leq e^{-\frac{0.4^2}{2} \cdot 0.6 t}$$

So again if $t = \Omega(\log n)$

this is $o\left(\frac{1}{n}\right)$

$\Rightarrow$ by a union bound with prob $1 - o(1)$

$$\forall v \quad \frac{1}{\hat{s}(v)} \leq 2k$$

Optimization

min/max $f(x_1, \ldots, x_n)$
subject to constraints

When can we solve efficiently?

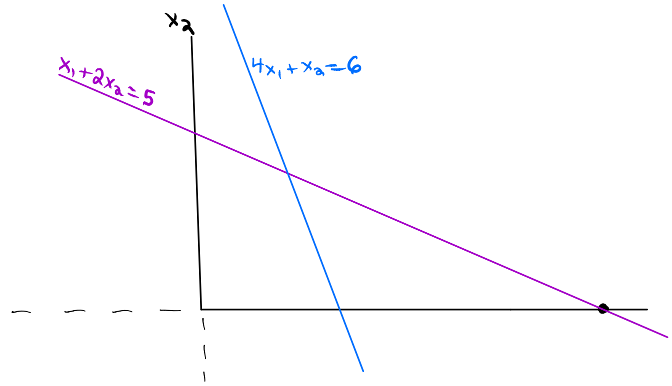
Linear programming

L

$f(x_1, \ldots, x_n)$ linear fn.
constraints linear

Example: Diet Problem
Athlete wants to max protein consumption
subject to ≤ 5 units of fat/day
$\le \$6$/day

	protein/lb	fat/lb	\$/lb
steak	2	1	4
peanut butter (PB)	1	2	1



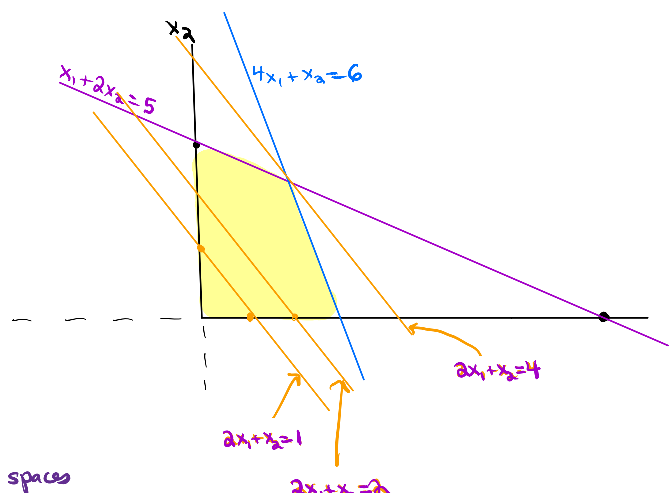
x_1 #lbs of steak/day
x_2 #lbs of PB/day

max $\boxed{2x_1 + x_2}$ ← objective fn
subject to

$$\boxed{\begin{array}{l} 4x_1 + x_2 \le 6 \\ x_1 + 2x_2 \le 5 \\ x_1, x_2 \ge 0 \end{array}}$$

← feasible set
feasible region

feasible pt with max objective fn value is "optimal solution"



feasible set is "polyhedron": intersection of half spaces
if also bounded & nonempty ⇒ "polytope"

feasible set convex

linear cost fns define family of parallel hyperplanes
optimal feasible pt must occur at corner, a.k.a. vertex
[can't be expressed as convex comb of feasible pts]

Unfortunately too many vertices to enumerate $\approx m^{n/2}$
m constraints, n variables

Ex: $0 \le x_i \le 1$ $1 \le i \le n$
n dimensional hypercube

max $\boxed{2x_1 + x_2}$

subject to
$$\boxed{\begin{aligned} 4x_1 + x_2 &\le 6 \\ x_1 + 2x_2 &\le 5 \\ x_1, x_2 &\ge 0 \end{aligned}}$$

$c = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$A = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$

$b = \begin{pmatrix} 6 \\ 5 \end{pmatrix}$

Input to LP problem

$$\boxed{\begin{aligned} &\vec{c} \in \mathbb{R}^n \\ &A \in \mathbb{R}^{m \times n} \\ &\vec{b} \in \mathbb{R}^m \end{aligned}}$$

find $x \in \mathbb{R}^n$ to

max $c \cdot x$

subject to $Ax \le b$

$x_i \ge 0 \ \forall i$

$$\equiv \max \ c_1x_1 + c_2x_2 + \dots + c_nx_n$$

$$\begin{aligned} \text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \le b_1 \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \le b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \le b_m \\ & x_1, x_2, \dots, x_n \ge 0 \end{aligned}$$

There are efficient algs for LP.

- simplex } exp worst case, "smoothed" poly time
- ellipsoid
- interior pt methods } polynomial time

smoothed analysis of algs:

measures perf of alg under slight random perturbation of input

ellipsoid: sometimes even exponential-sized LPs can be solved in poly time

poly time w/ "separation oracle" - poly time alg

given x $\to$ yes x feasible

given x $\to$ no separating hyperplane

Bottom line: If you can formulate your problem as LP, it can be solved in time $\text{poly}(m, n, \log(\max_{i,j}(a_{ij}, b_j, c_i)))$

Using "linear-programming relaxations" to get approx algs for NP-hard problems & then apply randomized rounding

Example: MAX 2-SAT

Given a Boolean formula in CNF (and of ors) where each clause has 2 literals, find T/F assignment to variables that maximizes # satisfied clauses

max-version NP-hard

Ex: $x_1 \vee x_3) \wedge (x_1 \vee \bar{x}_3) \wedge (\bar{x}_1 \vee x_4) \wedge (\bar{x}_4 \vee x_1) \wedge (x_1 \vee x_4) \wedge (\bar{x}_1 \vee \bar{x}_4)$

Step 1:

Example.

$$(x_1 \vee x_2) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_2) \wedge (x_1 \vee \bar{x}_3)$$

$$\max \quad z_1 + z_2 + z_3 + z_4 + z_5$$

s.t.

$$x_1 + x_2 \geq z_1$$

$$x_1 + (1 - x_2) \geq z_2$$

$$(1 - x_1) + x_2 \geq z_3$$

$$(1 - x_1) + (1 - x_2) \geq z_4$$

$$x_1 + (1 - x_3) \geq z_5$$

$$x_1, x_2, x_3 \in \{0, 1\}$$

$$z_1, z_2, z_3, z_4, z_5 \in \{0, 1\}$$

Randomized rounding summary

important technique for obtaining approx algs for NP-hard problems

Recipe

1. Set problem up as integer linear program. vars $\in \{0,1\}$
2. Relax ILP $\to$ LP vars $\in [0,1]$
3. Solve LP to optimality $\to$ x_i^* $i=1,..,n$
4. Construct solution by randomly rounding vars $\to \{0,1\}$
   treat x_i^* as probability
5. bound quality of soln by comparing to LP OPT

Another example: $\boxed{\text{congestion minimization}}$

Input: directed graph $G=(V,E)$
set of pairs (s_i,t_i) $i=1..k$

Output: path P_i from $s_i \rightsquigarrow t_i$ $\forall i=1..k$
s.t. $\underbrace{\text{congestion}}_{\text{max \# paths that intersect any edge}}$ is minimized

NP-hard

Approx alg via randomized rounding

1. Set up ILP [multi commodity flow]
   vars $f_i(e) \in \{0,1\}$ flow from $s_i \rightsquigarrow t_i$ on edge e
   C congestion

$$\min C$$

subject to

$$\sum_{\substack{e \text{ s.t.} \\ e=(u\to v) \\ \text{for some } u}} f_i(e) = \sum_{\substack{e \text{ s.t.} \\ e=(v\to w) \\ \text{for some } w}} f_i(e) \quad \forall v \neq s_i,t_i \ \forall i \qquad \text{conservation of flow}$$

$$\sum_{\substack{e \text{ s.t.} \\ e=(s_i\to u) \\ \text{for some } u}} f_i(e) = 1 \quad \forall i \qquad \text{route 1 unit of flow from } s_i \to t_i$$

$$\sum_i f_i(e) \le C \qquad \text{congestion bound}$$

$$f_i(e) \in \{0,1\} \quad \forall i,e$$

Theorem $\Pr\left(\text{any edge has congestion} \geq \underbrace{\boxed{\frac{6 \ln n}{\ln\ln n}}}_{\alpha} C^*\right) \leq \frac{1}{n}$

gives approx ratio of α

Proof Fix $e = (u \to v)$

Let $X_i(e) = \begin{cases} 1 & \text{if } e \in P_i \\ 0 & \text{o.w.} \end{cases}$

$$E(X_i(e)) = \sum_{\substack{P \in \mathcal{P}_i \text{ s.t.} \\ e \in P}} f_P^i = f_i^*(e)$$

Let $X(e) = \sum_{i=1}^{k} X_i(e) \Rightarrow \forall e \quad E(X(e)) = \sum_{i=1}^{k} f_i^*(e) \leq C^*$

$$\Pr\left(X(e) \geq \underbrace{(1+\delta)}_{\alpha} C^*\right) \leq e^{-C^*[(1+\delta)\ln(1+\delta) - \delta]} \qquad \text{[Chernoff bound]}$$

$$\Pr(X(e) \geq \alpha C^*) \leq e^{-C^*[\alpha \ln \alpha + 1 - \alpha]} \underset{C^* \geq 1}{\leq} e^{-\alpha(\ln \alpha - 1)}$$

$$\leq e^{-3 \ln n} = \frac{1}{n^3}$$

$\alpha(\ln \alpha - 1)$

$= \frac{6 \ln n}{\ln\ln n}\left[\ln\ln n + \ln 6 - \ln\ln\ln n - 1\right]$

$\geq 3 \ln n$

$$\Rightarrow \Pr(\text{any edge has congestion} \geq \alpha C^*) \underset{\text{union bound}}{\leq} \sum_e \Pr(X(e) \geq \alpha C^*)$$

$$\leq n^2 \cdot \frac{1}{n^3} = \frac{1}{n}$$

Optimal!

$\exists$ graphs with $\frac{OPT}{C^*} = \Omega\left(\frac{\ln n}{\ln\ln n}\right)$ integrality gap

so can't do better with such an approach

Hardness (directed graphs)

Every poly time alg has $\Omega\left(\frac{\log n}{\log\log n}\right)$ approx ratio

unless $NP \subseteq BPTIME\left(n^{O(\log\log n)}\right)$

Final note: If $C^* \geq c \ln n$ then the randomized rounding approach gives max congestion $O(C^*)$ w.h.p.