ORF 523
Convex and Conic Optimization

Liza Rebrova

Liza

U Michigan (PhD Math, Random matrix theory, high-dimensional probability) →

UCLA (Postdoc Applied Math, Randomized algorithms, numerical linear algebra, tensor methods, math of data) →

Princeton ORFE

Sherrerd 324, elre@princeton.edu, erebrova.github.io

OH: **Monday 9:30-11:30am** + one-on-one meetings scheduling: **https://calendly.com/liza-rebrova**

AI: **Abraar Chaudhry**

Graduate student at ORFE

Office hours: **Wednesday: 5-7PM (tentative) Sherrerd 123**

Email: **azc@princeton.edu**

Solving optimization problems

An optimization problem:

- **Maximize/minimize something/make a decision**
- **Subject to some constraints**

Everything we do in life is solving optimization problems :)

However, in math optimization, main questions are:

1) **Which optimization problems can we solve efficiently?**
2) How to solve them?

In ORF 524 (last semester):

Optimization problem: min f(x) subject to x in Omega

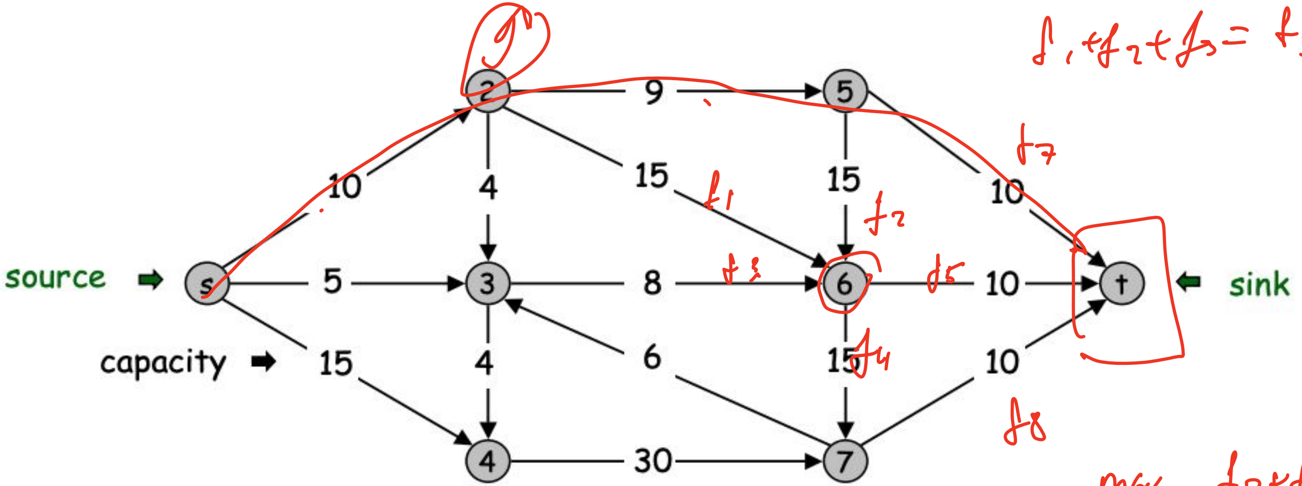
Algorithms for solving “nice” optimization problems:

linear problems (simplex method, interior point method);

non-linear problems (first-order methods like gradient descent, ADMM…)

Another example: maximum flow problem

$f_1 + f_2 + f_3 = f_5 + f_4$



max $f_7 + f_5 + f_8$

From https://www.cs.princeton.edu/courses/archive/spr04/cos226/lectures/maxflow.4up.pdf

Problem set-up

Given: *directed graph* (G = (V, E)) with *capacities* on edges (c_e) and fixed *source* (S) and *target* (T) nodes.

Flow: nonnegative number (f_e) assigned to every edge, at most the capacity of this edge, satisfying *flow conservation property*. Namely, induced flow at each node, defined as (a sum of all incoming edges - a sum of all outcoming edges), must be 0, except from the S and T nodes.

Max flow problem: find flow on G that maximizes induced flow in T.

Solution 1:

Ford-Fulkerson algorithm

While possible:

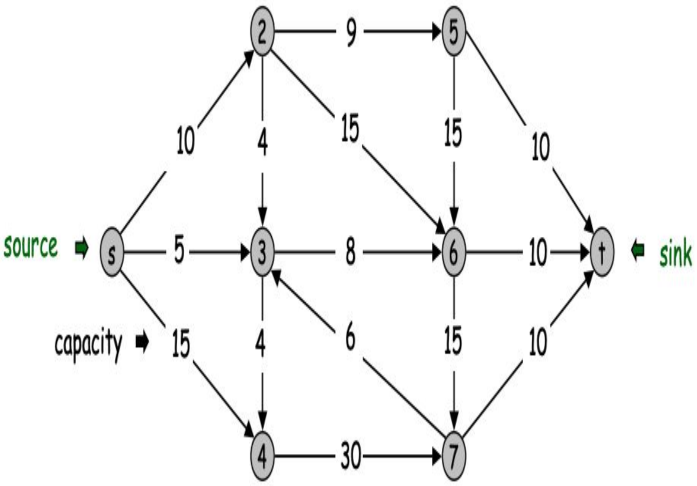
1. Find an *acceptable* flow path from S to T
2. *"Augment"* the graph to get a new one (that shows if we can improve current flow)

Acceptable flow: find a directed path from S to T and take flow value as its minimum edge capacity

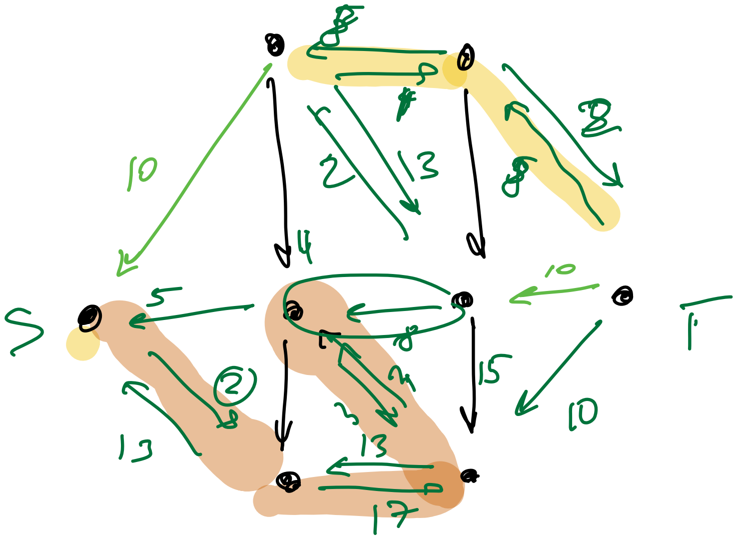
Graph is *augmented* along the last flow path: (a) old edges are removed (b) residual flow edges are added (c) reverse flow edges are added

Algorithm

10 5



30 - 2 = 28 max flow



Solution 2:

f_e $\quad \xrightarrow[e]{} T$

edges that go in target node

Linear problem?

$$\max \sum_{\overrightarrow{e}T} f_e$$

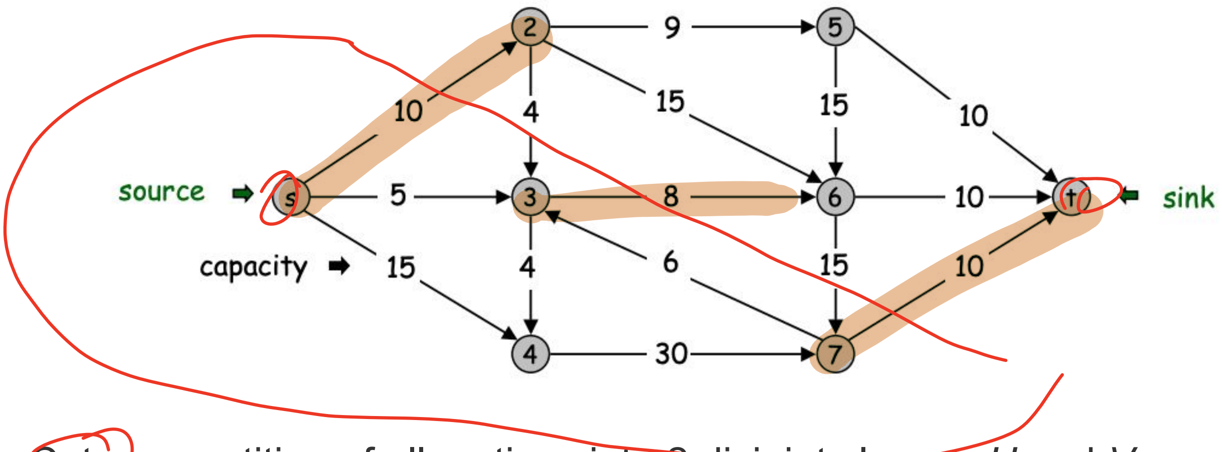
s.t. $0 \le f_e \le c_e$

$$\sum_{\overrightarrow{e}v} f_e = \sum_{v\overrightarrow{e}} f_e$$

Certificates for solutions

28

How can we check that a given solution cannot be improved?



4

Cut – a partition of all vertices into 2 disjoint classes $\underline{U}$ and $\underline{V}$, such that S in $\underline{U}$, T in $\underline{V}$. Value of cut – sum of capacities of the edges directed from $\underline{U}$ to $\underline{V}$.

Minimal cut problem

- Duality principles give us lower/upper bound for the solutions (in linear cases they are tight bounds!)
- Min cut problem is important by itself (and typically solved via conversion to max flow problem)
- E.g., image segmentation

With how many people (out of 20) you can meet? 19

Doodle ★ Features Pricing Create account Sign in

February 2016

20 participants	Mon 1 3:00 PM - 3:15 PM	Mon 1 3:15 PM - 3:30 PM	Mon 1 3:30 PM - 3:45 PM	Mon 1 3:45 PM - 4:00 PM	Tue 2 3:00 PM - 3:15 PM	Tue 2 3:15 PM - 3:30 PM	Tue 2 3:30 PM - 3:45 PM	Tue 2 3:45 PM - 4:00 PM	Wed 3 3:00 PM - 3:15 PM	Wed 3 3:15 PM - 3:30 PM	Wed 3 3:30 PM - 3:45 PM	Wed 3 3:45 PM - 4:00 PM	Thu 4 3:00 PM - 3:15 PM	Thu 4 3:15 PM - 3:30 PM	Thu 4 3:30 PM - 3:45 PM	Thu 4 3:45 PM - 4:00 PM	Fri 5 3:00 PM - 3:15 PM	Fri 5 3:15 PM - 3:30 PM	Fri 5 3:30 PM - 3:45 PM	Fri 5 3:45 PM - 4:00 PM
Woodrow Wilson	✓	✓															✓			
F. Scott Fitzgerald						✓				✓								✓		
Richard Feynman															✓			✓		
Michelle Obama		✓										✓								
Paul Volcker								✓		✓										
John Nash				✓															✓	
Terence Tao													✓							
Ben Bernanke					✓					✓										
Paul Krugman							✓				✓									
Andrew Wiles	✓																			
Steve Forbes					✓			✓			✓									
John Milnor	✓													✓						
Kazuyo Sejima									✓											
Albert Einstein			✓										✓							
George A. Miller			✓													✓				
Alan Turing		✓		✓																
Jeff Bezos					✓															
Meg Whitman												✓								
Donald Rumsfeld												✓					✓			✓
Eugene O'Neill								✓									✓		✓	

Credits: Amir Ali Ahmadi (ORF 523 2021)

How do we know this is best possible?

Doodle

Features Pricing Create account Sign in

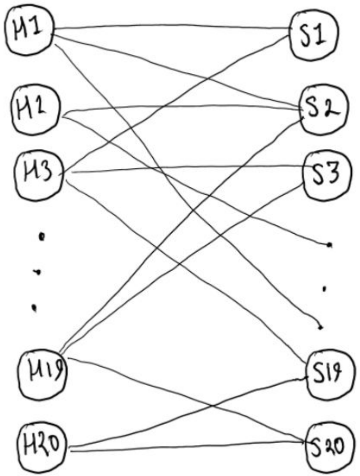
February 2016	Mon 1				Tue 2				Wed 3				Thu 4				Fri 5			
20 participants	3:00 PM - 3:15 PM	3:15 PM - 3:30 PM	3:30 PM - 3:45 PM	3:45 PM - 4:00 PM	3:00 PM - 3:15 PM	3:15 PM - 3:30 PM	3:30 PM - 3:45 PM	3:45 PM - 4:00 PM	3:00 PM - 3:15 PM	3:15 PM - 3:30 PM	3:30 PM - 3:45 PM	3:45 PM - 4:00 PM	3:00 PM - 3:15 PM	3:15 PM - 3:30 PM	3:30 PM - 3:45 PM	3:45 PM - 4:00 PM	3:00 PM - 3:15 PM	3:15 PM - 3:30 PM	3:30 PM - 3:45 PM	3:45 PM - 4:00 PM
Woodrow Wilson	✓	✓															✓			
F. Scott Fitzgerald						✓				✓								✓		
Richard Feynman															✓			✓		
Michelle Obama		✓										✓								
Paul Volcker								✓		✓										
John Nash				✓															✓	
Terence Tao													✓							
Ben Bernanke					✓					✓										
Paul Krugman							✓				✓									
Andrew Wiles	✓																			
Steve Forbes					✓			✓			✓									
John Milnor	✓													✓						
Kazuyo Sejima									✓											
Albert Einstein			✓										✓							
George A. Miller			✓													✓				
Alan Turing		✓		✓																
Jeff Bezos					✓															
Meg Whitman												✓								
Donald Rumsfeld												✓					✓			✓
Eugene O'Neill								✓									✓		✓	

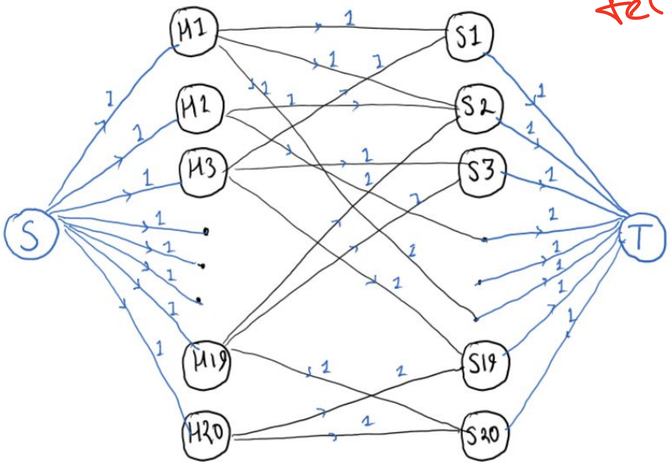
st. ...

$f_e \in \{0,1\}$

$f_e(1-f_e)=0$

Was it a really new problem?





Reduction is a powerful tool…but also note the difference!

Take-aways from max cut/min flow/scheduling game:

1. We will be interested in recognizing large classes of problems that we know how to solve (linear, convex, conic), rather than in devising special algorithms

2. Duality principles give us lower/upper bound for the solutions (in linear cases, they are tight bounds!)

3. If the problem is more complicated, there is a hope to reduce it to one of the "known" problems, or to otherwise simplify

Hard problem #1: maximum cut

Problem (max cut):

Find a partition of the graph's vertices into two complementary sets *U* and *V*, such that the number of edges between the set *U* and the set *V* is as large as possible.

- There is no known polynomial solution to this problem
- And not believed to be, unless P = NP [complexity theory!]
- Actually, no polynomial-time approximation scheme can get, arbitrarily close to the optimal solution, for it, unless P = NP (APX-hard problem). Thus, any approximation algorithm targets a ratio < 1.

We will study Goemans-Williamson algorithm that gives |cut| >= 0.878*|max cut| based on semidefinite (conic) relaxation and randomization

Approximate solution with randomization

$0.5\,|\text{max. cut}| \le |\text{cut}| \le |\text{max\_cut}|$

Idea: assign each vertex of the graph to one of two subsets U and V at random with probability ½.

Claim: In expectation, this cut will contain at least ½ of all the edges (thus, at least ½ of optimal cut size, approximation ratio 0.5)

$$\mathbb{E}(\text{number of edges in this random cut}) = \mathbb{E}\Big(\sum_e \mathbb{1}(e)\Big)$$

$\mathbb{1}(e) = 1$ if $e \in$ cut, $= 0$ if $e \notin$ cut

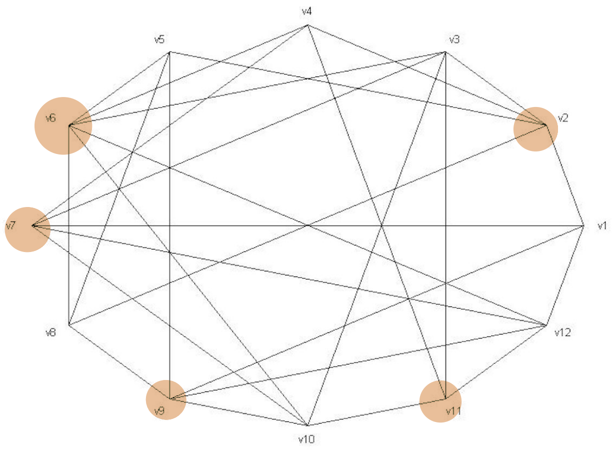
$v_1 \quad v_2$

$$= \sum_e \mathbb{E}(\mathbb{1}_e) = \sum_e P(\text{edge belongs to a cut})$$

$$= \sum_e P\begin{pmatrix} v_1 \in U \\ v_2 \in V \end{pmatrix} + P\begin{pmatrix} v_1 \in V \\ v_2 \in U \end{pmatrix} = \sum_e \Big(\frac{1}{4} + \frac{1}{4}\Big) = \frac{1}{2}|E|$$

Problem #2: max independent set

Problem: Find maximal subset of vertices not connected with any edges



5

Discussion

- The maximum independent set problem is also NP-hard. No certificate to show that a given/guessed solution is optimal

- Relaxation to a semidefinite program allows to give lower and upper bounds (that might not be tight as in the linear case, still can give a good idea about the size of the best independent set; we will show SDP relaxation is tighter than possible linear relaxation)

- We will study this relaxation and an interesting generalization to independent tuples of vertices, based on stability numbers of the strong graph products

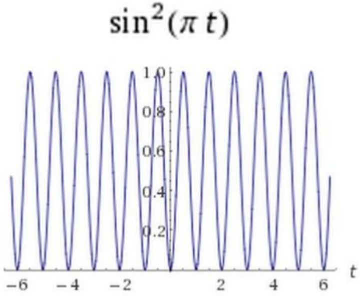
Problem #3

Is optimal solution of the following problem 0 or > 0?

$$\min_{x,y,z,n} \quad (x^n + y^n - z^n)^2$$

$$\text{s.t.} \quad x \geqslant 1,\ y \geqslant 1,\ z \geqslant 1,\ n \geqslant 3,$$

$$\sin^2 \pi n + \sin^2 \pi x + \sin^2 \pi y + \sin^2 \pi z = 0.$$



Discussion

- This is a formulation of Fermat's Great theorem

For $n \geq 3$, the equation $x^n + y^n = z^n$ has no solution over positive integers.

Posed in 17th century, solved by Andrew Wiles in 1994.

- In general, polynomial optimization quickly becomes hard, e.g.:

  (a) Given a smooth objective function f (even a degree-4 polynomial), and a point x0, it is NP-hard to decide if x0 is a local minimum or a strict local minimum of a problem [min f(x) for x in R^n]

  (b) Given a quadratic function f and a point x0, it is NP-hard to decide if x0 is a local minimum of [min f(x) for x in O], where O is defined as a set of linear inequalities

Take-aways from three Hard Problem examples:

1. An easy mathematical formulation can hinder complexity of the problem

2. The importance of understanding complexity theory

3. Reductions still help a lot, and randomization is a powerful tool too

Tentative list of topics

$\min f(x)$
s.t. $x \in \cancel{\Omega}$ $\mathbb{R}^n$

- Math review
- **Unconstrained nonlinear optimization**: first and second order optimality conditions
- **Convex analysis and convex optimization problems**
- **Duality and certificates of infeasibility**
- From linear programs to **positive semidefinite programs** (PSD)
- **Relaxations to linear and PSD problems**
- **Complexity theory**
- Applications to combinatorial optimization, data science etc
- Approximate solutions and randomization
- Convex feasibility problems and randomization
- Randomized sketching and dimensionality reduction for convex optimization

Grading

- **Homeworks** (50% of the grade; 4 to 5 total – biweekly, no extensions allowed) You can collaborate and encouraged to discuss the problems with other students, AI and with me. You should write up all the problems individually and turn in HW for yourself only (please write the names of all your collaborators, excluding instructors :))

- **Midterm** (20% of the grade, around spring break, around 2 hours, in class or timed submission in gradescope, no collaborations)

- **Final exam** (30% of the grade, take home, for a couple of days, no collaborations)

Some suggested materials

- **S. Boyd and L. Vandenberghe, Convex Optimization**
  https://stanford.edu/~boyd/cvxbook/
- A. Ben-Tal and A. Nemirovski, Lecture Notes on Modern Convex Optimization
  https://www2.isye.gatech.edu/~nemirovs/LMCO_LN.pdf
- M. Laurent and F. Vallentin, Semidefinite Optimization
  http://www.mi.uni-koeln.de/opt/wp-content/uploads/2015/10/laurent_vallentin_sdo_2012_05.pdf

Prof Amir Ali Ahmadi’s website includes relevant class notes http://aaa.princeton.edu/orf523

Other resources for the course

- A tab on my website erebrova.github.io/ORF523.html
- Canvas website (make sure you are subscribed for announcements)
- Gradescope (for all homeworks and exam submission, see on canvas)
- Ed discussions (see on canvas)
- Zoom link
  https://princeton.zoom.us/j/93766607587?pwd=QnBLVG1UOFpoRDZlYlNwRXdUeVdLZz09

Please email me every time before class if you are in isolation/quarantine!