NP-hard problems → optimal solutions in expon. time (so far)

NP-hard problems → <u>suboptimal</u> solutions in polyn. time (now)

↓

performance analysis:
- empirical
- average case
- <u>worst case</u>

Approximation algorithms

Def. A <u>factor c approximation Algorithm</u> for a minimization problem is a polyn. time alg. that computes for any feasible input instance $I \in \mathcal{I}$ a feasible solution with cost

$$Alg(I) \leq c \cdot OPT(I),$$

where $OPT(I)$ is the cost of an optimal solution to I.

[A c-approximative alg. for a maximization problem computes a solution with cost

$$Alg(I) \geq \frac{1}{c} OPT(I), \quad \forall I \in \mathcal{I}.$$]

We call c <u>the approximation factor</u>.

<u>Dilemma</u>: How to compare against an <u>unknown</u> opt. solution?

⟶ use lower bounds instead (LP relax., Lagrange ...)

<u>General approach</u>: (min. problem)

Find $LB(I) \leq OPT(I)$, $\forall I \in \mathcal{I}$, and

show $Alg(I) \leq c \cdot LB(I)$, $\forall I \in \mathcal{I}$.

Then Alg is a c-approx. since $Alg(I) \leq c \cdot OPT(I)$, $\forall I \in$

<u>Outlook</u>:
- combinatorial alg. analyzed using LP relaxation
- LP based algorithms + techniques, how to turn a fractional solution into <u>good</u> integral one.

But first: introductory example: TSP (purely combinatorial)

Metric TSP

Problem: Given is a complete graph with non-negative edge cost, that satisfy the triangle inequality, i.e., for all $u, v, w \in V$:

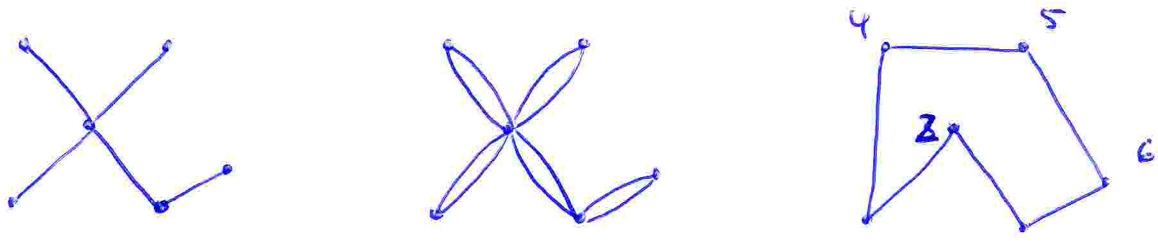
$$c(u,v) \leq c(u,w) + c(w,v).$$

Find a minimum cost cycle (tour) visiting each node exactly once.

Algorithm [Christofides '76]

1. Find a MST T in G
2. Double every edge of T to obtain a Eulerian graph
   [Recall: G' is Eulerian $\Leftrightarrow$ every node in G' has node degree 2.]
3. Find a Eulerian Tour $\mathcal{T}$ on the doubled MST.
4. Output the tour that visits vertices of G in the order of their first appearance in $\mathcal{T}$. (= take shortcuts)

Ex.



MST T — doubled MST, eulerian tour $\mathcal{T}$ — Final tour.

Thm. The algorithm is a factor 2 approximation for Metric TSP.

Pf. Let $OPT(I)$ be the length of an optimal tour for inst. I.
Removing one edge gives a spanning tree for I.

$$\Rightarrow OPT(I) \geq MST(I) = cost(T).$$

Obviously, $cost(\mathcal{T}) = 2 \cdot cost(T)$.
With the triangle ineq., the algorithms final tour has length

$$Alg(I) \leq cost(\mathcal{T}) \leq 2\, OPT(I).$$

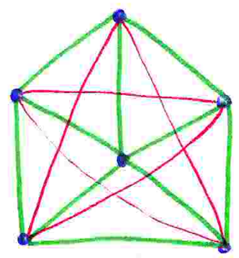
$\square$

→ Analysis of this algorithm is tight (asymptotically).

Example:

- complete graph K_n with edge cost 1 and 2.

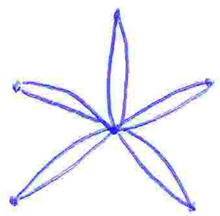
Pic. for $n=6$:



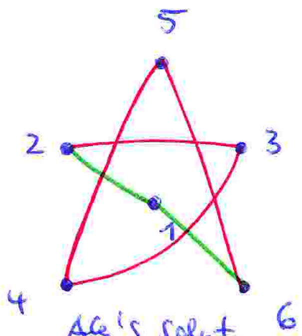
— cost = 1

— cost = 2

- alg.:

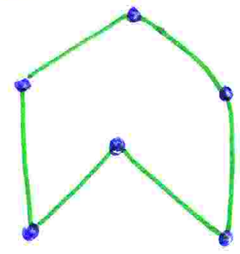


doubled MST



Alg's solut.

- opt.:



$Alg = (n-2)\cdot 2 + 2\cdot 1 = 2n-2$

$OPT = n \cdot 1$

$$\frac{Alg(I)}{OPT(I)} \xrightarrow[n\to\infty]{} 2$$

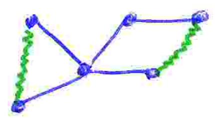
Question: Is there a better alg. for metric TSP? YES.

- the above alg. finds Eulerian tour by doubling
- can we find a cheaper way from MST → Eulerian graph?
  - → all nodes need even degree!

    Worry only about {nodes with odd degree in MST} =: V'.

    Observe: $|V'|$ is even since the sum of all node degrees in MST is even. (handshaking argument)

    ⇒ MST + perfect matching on V' give Eulerian graph



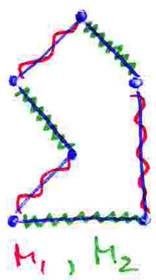
Algorithm II [Christofides 76]

1. Find an MST T on G
2. Compute a minimum cost perfect matching M on the set of odd degree nodes of T. Add M to T. $\to$ Eulerian
3. Find a Eulerian tour $\mathcal{T}$ on this graph.
4. Output the tour that visits vertices of G in the order of their first appearance in $\mathcal{T}$.

Thm. The alg. is a factor 3/2 approx. for the metric TSP.

Lemma (other lower bound)
Let $V' \subseteq V$ s.t. $|V'|$ is even and let M be a min cost perfect matching on V'. Then $\text{cost}(M) \leq OPT/2$.

Pf. Let $OPT' \leq OPT$ be the subtour of OPT on V' obtained by shortcutting (Δ-ineq.).



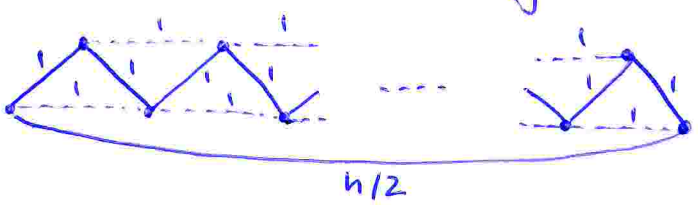
OPT' is the union of 2 disjoint perfect matchings M_1 and M_2. Then $\min\{\text{cost}(M_1), \text{cost}(M_2)\} \leq \frac{1}{2} OPT'$!

$$\text{cost}(M) \leq \min\{\text{cost}(M_1), \text{cost}(M_2)\} \leq \frac{1}{2} OPT. \quad \square$$

Pf of Thm.

$$\text{cost}(\mathcal{T}) \leq \text{cost}(T) + \text{cost}(M) \leq \frac{3}{2} OPT. \quad \square$$

Tight example: n vertices, with n being odd



— MST + expensive unique matching $\Rightarrow$ $Alg = n - 1 + \frac{n}{2}$

$OPT = n$

$\to$ This is the currently best known approx. alg. for metric TSP. It is conjectured – but open – if a 4/3 exists.

Question: How "good" can any polyn. time alg. be?
$\hat{=}$ lower bound on the approx. factor

Inapproximability

▷ example gap reduction (general) TSP.

Thm. [Gonzalez, Sahni '76]
For any polynomial time computable function $\alpha(n)$, there is no $\alpha(n)$-factor approx. alg., unless $P = NP$.

Pf. by contradiction: Suppose there is an $\alpha(n)$-approx. $\mathcal{A}$. We show that $\mathcal{A}$ can be used to decide the Hamiltonian Circuit problem, which is NP-hard. $\Rightarrow$ This would imply $P = NP$.

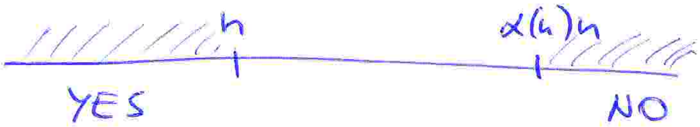
↳ Given a graph, does there exist a tour that visits each node exactly once? [NP-hard, Karp 72]

Ham. circuit problem $G = (V,E)$ $\xrightarrow{\text{construct}}$ TSP
G' is $G(V,E)$ with $w_e = 1, \forall e \in E$ plus edges E' s.t. G is complete and $w_e = \alpha(n) \cdot n$, $\forall e \in E'$.

Obs. If G has HC, then $\exists$ tour with cost $= n$.
If G has no HC, then cost of any tour $> \alpha(n) \cdot n$.

$\Rightarrow$ An $\alpha(n)$-approx. alg. $\mathcal{A}$ could decide HC problem in polyn. time. Contradiction.

□



$\leadsto$ Analysis based on LP relaxation:

<u>The knapsack problem</u>

$$\max \sum_{j=1}^{n} c_j x_j$$
$$\sum_{j=1}^{n} w_j x_j \le K$$
$$x_j \in \{0,1\}$$

<u>Alg</u>:

0.) Remove all items with $w_j > K$.

1.) Sort and reindex items s.t. $\frac{c_1}{w_1} \ge \frac{c_2}{w_2} \ge \dots \ge \frac{c_n}{w_n}$.

2.) Let $k = \max\{ j \in \{1,\dots,n\} \mid \sum_{\ell=1}^{j} w_j \le K \}$.

3.) Choose the best of the two solutions

$$\{1,\dots,k\} \quad \text{or} \quad \{k+1\}.$$

<u>Thm.</u> Alg. is a factor 2-approximation for knapsack.

<u>Pf.</u> Let OPT be cost of an optimal solution and z^{LP} the cost of the optimum to the LP-relaxation.

$$OPT \le z^{LP}.$$

<u>Prop.</u> $z^{LP} \le \sum_{j=1}^{k+1} c_j$. (Proved in homework.)

Then, our solution $\max\{\sum_{j=1}^{k} c_j, c_{k+1}\} \ge \frac{1}{2} z^{LP} \ge \frac{1}{2} OPT$.

$\square$