Advanced Algorithmics

Strategies for Tackling Hard Problems

Sebastian Wild

Markus Nebel

Lecture 19

2017-06-22

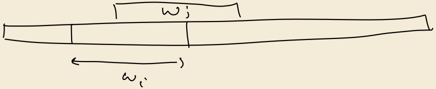
Definition 5.16 (Shortest-Superstring)

Given: alphabet Σ, set of strings $W = \{w_1, \ldots, w_n\} \subseteq \Sigma^+$

Feasible Instances: *superstrings* s of S, i.e., s contains w_i as substring for $1 \leq i \leq n$.

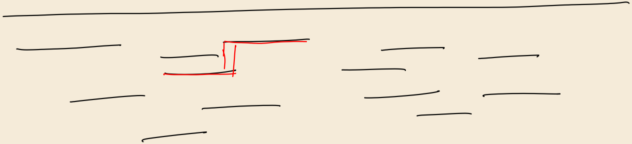
Cost: $|s|$

Goal: min



Remark 5.17

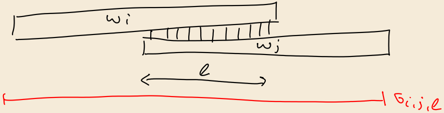
Without-loss-of-generality assumption: no string is a substring of another.



NP-hard

Shortest Superstring by Set Cover

Construct *all* pairwise superstrings:



$\sigma_{i,j,\ell} = (w_i)_{1,|w_i|-\ell}(w_j)$ valid iff $(w_j)_{1,\ell} = (w_i)_{|w_i|-\ell+1,|w_i|}$

$M = \big\{\sigma_{i,j,\ell} : i, j, \in [u], \ell \in \big[0..\min\{|w_i|, |w_j|\}\big]\big\}$

$\rightsquigarrow$ **Set Cover instance:**

Universe: $[n]$ $\rightsquigarrow$ try to *cover* all words in W with superstring ...

Subsets: $S = \{S_\pi : \pi \in W \cup M\}$ ... by combining pairwise superstrings.

where $S_\pi = \{k \in [n] \ : \ \exists i, j : w_k = \pi_{i,j}\}$

Cost function: $c(S_\pi) = |\pi|$

Given set-cover solution $\{S_{\pi_1}, \ldots, S_{\pi_k}\}$

$\rightsquigarrow$ superstring $\pi_1 \ldots \pi_k$ $\leftarrow$ in any order

Lemma 5.18 (Pairwise superstrings yield 2-SC-approx)

Let W be an instance for SHORTEST-SUPERSTRING and (n, S, c) the corresponding SET-COVER instance. Let further OPT resp. OPT_{sc} be the optimal objective value of W resp. (n, S, c). Then holds $OPT \leq OPT_{sc} \leq 2 \cdot OPT$.

Proof: • $OPT \leq OPT_{sc}$

$OPT = |s^*|$ $\quad$ s^* shortest superstring for W

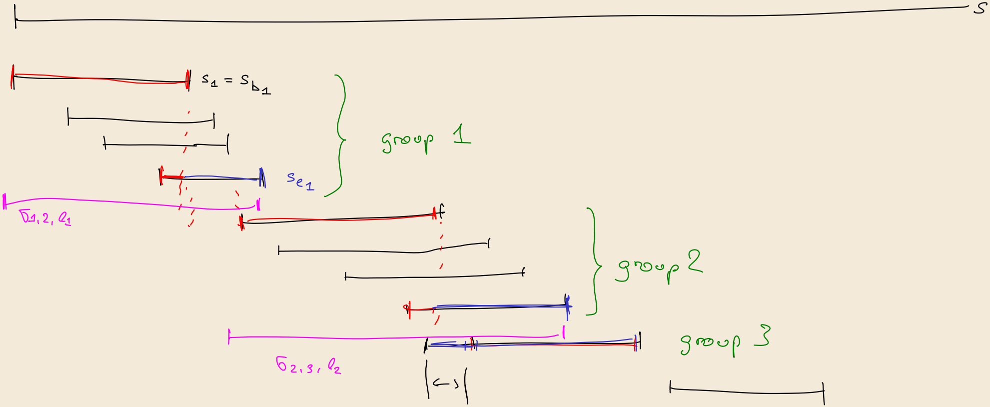
$OPT_{sc} = |s|$ $\quad$ where $s = \pi_1 \cdots \pi_k$ is from set cover solution

s is a superstring for W

require $\{S_{\pi_1}, \ldots, S_{\pi_k}\}$ cover all w_i

• $OPT_{sc} \leq 2 \cdot OPT$

Let s is a shortest superstring $\overset{\text{wlog.}}{\rightsquigarrow}$ contains s contains $w_1, \ldots, w_n$ in this order



$\Rightarrow$ no overlaps between group i and group $i+2$

$\Rightarrow$ every string w_i is a substring of at most 2 groups $\sigma_{i,i+1,\ell}$

$\Rightarrow$ every string covered at most twice

$|\pi_1, \dots, \pi_k| \leq 2 \cdot |S| = 2\,\mathrm{OPT}$

$\square$.

Corollary 5.19 ($2H_n$ approximation for superstring)

By solving the transformed set cover instance with `greedySetCover`, we obtain a $2H_n$-approximation for the shortest superstring problem.

Shortest superstring can be approximated better

$\rightsquigarrow$ 3-approximation

5.5 (Integer) Linear Optimization

Definition 5.20 (LP)
A linear program (LP) in *standard form* with n *variables* and m *constraints* is characterized by a matrix $A \in \mathbb{Z}^{m \times n}$, a vector $b \in \mathbb{Z}^m$, and a vector $c \in \mathbb{Z}^n$ and is written as

$x = (x_1, \ldots, x_n)$

$$\boxed{\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b \\ & x \geq 0 \end{aligned}} \qquad \begin{aligned} \min \quad & \textstyle\sum_{j=1}^{n} c_j \cdot x_j \\ \text{s.t.} \quad & \textstyle\sum_{j=1}^{n} a_{ij} \cdot x_j \geq b_i \quad \text{for all } i \in [m] \\ & x_j \geq 0 \quad \text{for all } j \in [n] \end{aligned}$$

(Comparisons on vectors are meant componentwise.)
Any vector $x \in \mathbb{R}^n$ with $Ax \geq b$ and $x \geq 0$ is called a *feasible solution* for the LP, and $c^T x$ is its objective value. An *optimal solution* is a feasible vector x^* with **min**imal objective value. ◀

Remark 5.21 (Rational coefficients)
We can in general allow $A \in \mathbb{Q}^{m \times x}$, $b \in \mathbb{Q}^m$ and $c \in \mathbb{Q}^n$; by multiplying constraints and scaling objective function with the common denominator we obtain an equivalent LP. ◀

Example LP

$$\begin{aligned} \min\ & 7x_1 + x_2 + 5x_3 \\ \text{s.t.}\ & x_1 - x_2 + 3x_3 \geq 10 \\ & 5x_1 + 2x_2 - x_3 \geq 6 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

$c = (7, 1, 5)$

$A = \begin{pmatrix} 1 & -1 & 3 \\ 5 & 2 & -1 \end{pmatrix} \quad b = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$

$\rightsquigarrow$ Optimal solution $x^* = (1.75, 0, 2.75)$ with $c^T x^* = 26$.

Extreme point: feasible point that is *not* a convex combination of two distinct feasible solutions. characterized by inequalities that are fulfilled with =

$x = \mu \cdot y + (1-\mu) \cdot z \qquad \mu \in (0,1)$

$\Rightarrow y = z = x$

Remark 5.22 (Facts on LPs)

1. More general versions of LP possible:
   = constraints, unrestricted variables, max instead of min ...
   $\rightsquigarrow$ can all be transformed into equivalent one in standard form.
2. LP can be *infeasible* (no solution), *unbounded* (no optimal solution) or *finite*.
3. If LP has optimal solution, there is an optimal extreme point $\rightsquigarrow$ finite problem!
4. $\boxed{\text{Optimal solutions can be computed in poly-time (ellipsoid method).}}$

Definition 5.23 (ILP)

An *integer linear program* in standard form is an LP with the additional integrality constraints $x_j \in \mathbb{N}_0$:

$$\begin{aligned} \min \quad & c^T x \\ \text{s. t.} \quad & Ax \geq b \\ & x \in \mathbb{N}_0^n \end{aligned}$$

◀

Remark 5.24 (Facts on ILPs)

1. Generalized versions can again be transformed into standard form.
2. Decision version of the problem $\mathcal{NP}$-complete.

5.6 Set Cover by LP Relaxation & Rounding

The Set Cover ILP

Idea $x_j = 1$ iff S_j in cover. ⇝ have to make sure $x_j > 1$ not optimal ✓
Notation: For $u \in U$ set $V(u) = \{j : u \in S_j\}$.

$$\min \quad \sum_{j=1}^{k} c(S_j) \cdot x_j$$

$$\text{s. t.} \quad \sum_{j \in V(u)} x_j \geq 1 \quad \forall u \in U = [n]$$

$$\cancel{x \in \mathbb{N}_0^k} \qquad\qquad x \in \{0,1\}^k$$

$$x \geq 0$$

Observation: Any optimal solution fulfills $x \in \{0, 1\}^k$

Problem: ILP not efficiently solvable ⇝ *relax* integrality constraints!
i.e., replace $x \in \mathbb{N}_0^k$ by $x \geq 0$.
⇝ efficiently solvable, but might get fractional solutions x^*.

Simple Rounding

procedure frequencyCutoffSetCover(n,S,c)
  f = global frequency of S // max. # sets any u appears in
  x^* = optimal solution of relaxed set cover LP.
  $\mathcal{C} = \emptyset$
  for $j = 1, \ldots, k$
    if $x^*_j \ge 1/f$ **then** add j to $\mathcal{C}$
  return $\mathcal{C}$

Theorem 5.25

`frequencyCutoffSetCover` is an f-approximation for SET-COVER.

Proofs set-cover x^* feasible $\sum_{j \in V(u)} x_j \ge 1 \quad \forall u \in U$

Since $|V(u)| \le f \Rightarrow$ one x_j with $j \in V(u)$: $x_j \ge \frac{1}{f}$

$\Rightarrow S_j \in \mathcal{C}$ ✓

f-approx; SC minimization problem

$$\begin{aligned} OPT_{frac} &\leq OPT \\ \| & \\ c^T x^* & \end{aligned}$$

since we only shrink the feasibility region

that's why LPs help to find approx. algorithms

$$c(\mathcal{C}) = \sum_{j \in \mathcal{C}} c(S_j) \leq \sum_{j \in \mathcal{C}} f \cdot x_j^* c(S_j)$$

$$= f \cdot \sum_{j \in \mathcal{C}} x_j^* c(S_j) \leq f \cdot \sum_{j=1}^{k} x_j^* c(S_j) = f \cdot OPT_{frac}$$

$$\leq f \cdot OPT$$

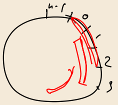
$\square$

Tight example $\quad U = \{0, \dots, n-1\} \quad S_i = \{i, i+1, \dots, i+f-1 \mod n\}$

$\Rightarrow$ every u appears in f sets

$\Rightarrow$ optimal: $\frac{n}{f}$ sets

by symmetry: $x^* = \left(\frac{1}{f}, \dots, \frac{1}{f}\right)$



f divides n

$c(S_j) = 1$

Randomized Rounding

$x_j^* \in [0,1] \mapsto$ interpret as probabilities

```
procedure randomizedRoundingSet(n,S,c)
    x* = optimal solution of relaxed set cover LP.
    C = ∅
    for j = 1,...,k
        for i = 1,...,⌈ln(4n)⌉
            b = coin flip with prob x*_j
            if b = 1 then add j to C
    return C
```

without repetitions

$\rightsquigarrow \mathbb{E}[c(\mathcal{C})] = \sum_j x_j^* \cdot c(S_j) = c^T x^* = OPT_{frac}$

Lemma 5.26 (Correct with prob 3/4)

randomizedRoundingSet computes with probability at least $\frac{3}{4}$ a feasible set-cover.

Proof: Consider $u \in \mathcal{U}$ $\quad |V(u)| = t \rightsquigarrow (p_1, \dots, p_t) = (x^*_{i_1}, \dots, x^*_{i_t})$ $\quad V(u) = \{i_1, \dots, i_t\}$

x^* feasible $\quad p_1 + \dots + p_t = c \geq 1$

$\Pr[u \text{ not covered by } \mathcal{C}] = \underbrace{(1-p_1)\cdots(1-p_t)}_{\text{concave}} \leftarrow$ decreases with c (falling)

in one iteration

concave $\Rightarrow$ maximal at $p_1 = p_2 = \dots = p_t = \frac{c}{t}$

~~minimal~~ maximal value attained at $p_1 = \dots = p_t = \frac{1}{t}$

$$\Pr[u \text{ not covered by } \mathcal{C} \text{ in one iteration}] \leq \left(1 - \frac{1}{t}\right)^t \leq \frac{1}{e}$$

$$\Pr[u \text{ not covered by } \mathcal{C}] \leq \left(\frac{1}{e}\right)^{\ln(4n)} = \frac{1}{4n}$$

$$\Rightarrow \Pr[\mathcal{C} \text{ is not a set cover}] = \Pr\left[\bigcup_{u \in \mathcal{U}} u \text{ not covered by } \mathcal{C}\right] \leq n \cdot \frac{1}{4n} = \frac{1}{4}$$

$\square$.

Lemma 5.27 (Expected quality)

The expected cost $\mathbb{E}[c(\mathcal{C})]$ of $\mathcal{C}$ computed by `randomizedRoundingSet` are bounded from above by $\ln(4n) \cdot OPT_{frac}$.

$\rightsquigarrow$ By Markov's inequality: $\Pr\left[c(\mathcal{C}) \geq 4 \ln(4n) \cdot OPT_{frac}\right] \leq \frac{1}{4}$

Proof: With repetitions $\mathcal{C} = \mathcal{C}_1 \cup \dots \cup \mathcal{C}_{\ln(4n)}$

$$\mathbb{E}[c(\mathcal{C})] \leq \sum_{i=1}^{\ln(4n)} \mathbb{E}[c(\mathcal{C}_i)] = \ln(4n) \cdot OPT_{frac}.$$

$\square$.

Randomized Rounding Approximation for Set Cover

```
procedure randomizedRoundingSetCover(n,S,c)
    𝒞 = randomizedRoundingSet(n,S,c)
    if c(𝒞) > 4 ln(4n) · OPT_frac ∨ 𝒞 not a set cover
        return S
    else
        return 𝒞
```

Theorem 5.28 (randomizedRoundingSetCover randomized approx)

`randomizedRoundingSetCover` is a randomized $4\ln(4n)$-approximation for SET-COVER. ◀

Proofs: if return $\mathcal{C}$ from line 6 ⇝ bound fulfilled by Lem 5.27.

$$\Pr[c(\mathcal{C}) > 4\ln(4n)\,OPT_{frac} \vee \mathcal{C} \text{ not cover}]$$

$$\le \Pr[\quad] + \Pr[\quad]$$

$$\underset{\text{Lem 5.26, 5.27}}{\le} \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \qquad \square$$