Noname manuscript No.
(will be inserted by the editor)

Preemptive and Non-Preemptive Generalized Min Sum Set Cover

Sungjin Im · Maxim Sviridenko · Ruben van der Zwaan

Received: date / Accepted: date

Abstract In the (non-preemptive) Generalized Min Sum Set Cover Problem, we are given n ground elements and a collection of sets $\mathcal{S} = \{S_1, S_2, ..., S_m\}$ where each set $S_i \in 2^{[n]}$ has a positive requirement $\kappa(S_i)$ that has to be fulfilled. We would like to order all elements to minimize the total (weighted) cover time of all sets. The cover time of a set S_i is defined as the first index j in the ordering such that the first j elements in the ordering contain $\kappa(S_i)$ elements in S_i. This problem was introduced in [1] with interesting motivations in web page ranking and broadcast scheduling. For this problem, constant approximations are known [2, 16].

We study the version where preemption is allowed. The difference is that elements can be fractionally scheduled and a set S is covered in the moment when $\kappa(S)$ amount of elements in S are scheduled. We give a 2-approximation for this preemptive problem. Our linear programming relaxation and analysis are completely different from [2, 16]. We also show that any preemptive solution can be transformed into a non-preemptive one by losing a factor of 6.2 in the objective function. As a byproduct, we obtain an improved 12.4-approximation for the non-preemptive problem.

This work was partially supported by NSF grant CCF-1016684.

Sungjin Im
Department of Computer Science, University of Illinois, 201 N. Goodwin Ave., Urbana, IL 61801, USA.
Tel.: +1-217-2446433 Fax: +1-217-2654035
E-mail: im3@illinois.edu

Maxim Sviridenko
Department of Computer Science, University of Warwick, Coventry CV4 7AL,UK.
Tel.:+44-24-76573792
E-mail: sviri@dcs.warwick.ac.uk

Ruben van der Zwaan
Department of Quantitative Economics, Maastricht University, The Netherlands.
E-mail: r.vanderzwaan@maastrichtuniversity.nl

Keywords Set cover · Latency · Preemption · Average cover time · Approximation

1 Introduction

The Min Sum Set Cover problem is a minimum latency version of the hitting set problem. We are given as input n elements, $\{1, 2, \ldots, n\} = [n]$ and a collection of sets $\mathcal{S} = \{S_1, S_2, ..., S_m\}$ where each set $S_i \in 2^{[n]}$. The goal is to find a permutation of the elements such that the total sum of (or equivalently average) cover/hitting times of all sets is minimized. For simplicity, we will say that an element e is covered at time slot t or it has cover time $\texttt{cov}(e) = t$ if it is placed in the t-th position in the permutation. Equivalently, we may say that the element e is scheduled at time t. The cover time $\texttt{cov}(S_i)$ of a set S_i is defined as $\min_{e \in S_i} \texttt{cov}(e)$ and the goal is to minimize $\sum_{S_i \in \mathcal{S}} \texttt{cov}(S_i)$. For this problem, a simple greedy algorithm is known to achieve an approximation factor 4 [4,8]. The greedy algorithm iteratively picks the element that hits the most sets that are not yet hit. Also it is known that the problem cannot be approximated within a factor of $4 - \epsilon$ for any $\epsilon > 0$ unless $P = NP$ [8]. A closely related problem known as Min Sum Coloring was studied before in [4, 5] with applications in scheduling. Also the special case of the Min Sum Vertex Cover was used in [6] as a heuristic for speeding up a solver for semidefinite programs.

The Min Latency Set Cover problem is a variant where the cover time is defined as the time where all elements in the set are covered e.g. $\texttt{cov}(S_i) = \max_{e \in S_i} \texttt{cov}(e)$. This problem is in fact equivalent to the precedence-constrained scheduling on a single machine [17], for which various 2-approximation algorithms are known [10,7,11]. It was shown that, assuming a variant of the Unique Games Conjecture, unless P=NP there is no $2 - \epsilon$ approximation for any $\epsilon > 0$ [3].

A generalization of the aforementioned problems was introduced by Azar, Gamzu and Yin [1] to provide a better framework for ranking web pages in response to queries that could have multiple intentions. This generalized problem was later named Generalized Min Sum Set Cover [2], and can be stated as follows. Every set S_i has a requirement $\kappa(S_i) \in \{1, 2, \ldots, |S_i|\} = [|S_i|]$. For a permutation of the ground set we define $\texttt{cov}(e)$ as before and S_i is covered at time t if t is the earliest time such that $|\{e \in S_i : \texttt{cov}(e) \leq t\}| \geq \kappa(S_i)$. Again, the goal is to find a permutation of the elements in $[n]$ minimizing $\sum_{S_i \in \mathcal{S}} \texttt{cov}(S_i)$. Azar et al. [1] give a modified greedy algorithm that has a performance guarantee of $O(\ln(\max_{S_i \in \mathcal{S}} \kappa(S_i)))$. The question whether there exists an $O(1)$-approximation was answered affirmatively by Bansal, Gupta and Krishnaswamy [2]. In order to obtain an $O(1)$-approximation, they used a time indexed linear program together with knapsack cover inequalities and gave a clever randomized rounding scheme. Very recently, their approximation ratio of 485 was improved by Skutella and Williamson to 28 via the same LP but a different rounding scheme [16].

In this paper we study the Preemptive Generalized Min Sum Set Cover. Like the Generalized Min Sum Set Cover problem, when $\kappa(S) = |S|$ for all $S \in \mathcal{S}$ it is a special case of (and in fact is a equivalent to) the single machine scheduling problem with precedence constraints and preemptions: $1|prec, pmtn|\sum w_j C_j$. It is known that preemption does not improve the solution quality for this problem (shown by a simple exchange argument), i.e. the optimal preemptive and non-preemptive schedules have the same optimal value. Hence it follows that there is no $2-\epsilon$ approximation for any $\epsilon > 0$ assuming a variant of the Unique Games Conjecture and $P \neq NP$ [3].

Preemptive Generalized Min Sum Set Cover is formally defined as follows. Given the ground set of elements $[n]$, sets $\mathcal{S} = \{S_1, S_2, ..., S_m\}$ and requirement $\kappa(S) \in [|S|]$ for each set $S \in \mathcal{S}$, we should fractionally assign elements of the ground set to the interval $[0, n]$. Formally, we define functions $x_e(t) : [0, n] \to \{0, 1\}$ where $x_e(t)$ is the indicator function that denotes whether element e is scheduled at time t such that $\int_{t=0}^{n} x_e(t)\, \mathtt{d}t = 1$ for all $e \in [n]$ and $\sum_{e\in[n]} x_e(t) = 1$ for any time $t \in [0, n]$. Then, the cover (or completion) time $\mathtt{cov}(S)$ of the set S is defined as the earliest time t such that $\int_{\tau=0}^{t} \sum_{e\in S} x_e(\tau)\, \mathtt{d}\tau \geq \kappa(S)$ and the goal is to minimize the sum of cover times over all sets. Note that the cover time $\mathtt{cov}(S)$ is not necessarily an integer unlike in the non-preemptive problem.

Our main motivation to study Preemptive Generalized Min Sum Set Cover is the fact that it provides a lower bound for the optimal value of the Generalized Min Sum Set Cover. We decouple finding an approximate solution to the relaxed problem (see Section 2) and the question of the lower bound quality (see Section 3 and Conjecture 1).

1.1 Our Results

Our main result is a polynomial time approximation algorithm with performance guarantee of 2 for the Preemptive Generalized Min Sum Set Cover. As we noticed before this result is tight modulo some complexity assumptions [3]. We note that one can easily show that the linear program used in [2,16] is a valid relaxation for the preemptive problem (see Proposition 4), thus the best known approximation for the non-preemptive problem also carries for the preemptive problem as well.

We introduce a *configuration linear program* which completely differs from the linear programming relaxation used in [2,16]. Interestingly, it is not obvious that our new linear program is a valid relaxation for the preemptive problem, unlike the previous linear program in [2,16] which can be easily shown to be a valid relaxation for the preemptive (and non-preemptive) problem. Our new LP is provably stronger than the previous LP, for both the preemptive and non-preemptive problems.

Further, we study the "gap" between the preemptive and non-preemptive solutions of the Generalized Min Sum Set Cover Problem, which is of independent interest. With some modifications of the rounding scheme in [16], we show that one can transform any α-approximate preemptive schedule into 6.2α-approximate non-preemptive one. With this transformation, we obtain an 12.4-approximation for the non-preemptive Generalized Min Sum Set Cover Problem, improving upon the previous best 28-approximation by Skutella and Williamson[16]. We conjecture that the gap between optimal preemptive and non-preemptive solutions is precisely two.

All our proofs easily extend to the case where every set S_i has a non-negative weight $w_i \geq 0$ and the objective is to minimize $\sum_{S_i \in \mathcal{S}} w_i \cdot \mathsf{cov}(S_i)$.

1.2 Organization

The remainder of this paper is organized as follows. In Section 2 we introduce the configuration linear program $\mathsf{LP}_{\mathsf{primal}}$. First, we prove that our configuration linear program is a valid relaxation for Preemptive Generalized Min Sum Set Cover and that this linear program can be solved in polynomial time. Finally, we design a rounding procedure that results in a randomized 2-approximation (Section 2.4) that can be derandomized. In Section 3 we obtain a transformation from a preemptive schedule to a non-preemptive schedule with a loss of factor 6.2, which immediately implies a 12.4-approximation in expectation to Generalized Min Sum Set Cover. In Section 4 we compare the time indexed linear programming relaxation used in [2, 16] to our configuration linear programming relaxation and show our relaxation is stronger.

2 2-Approximation for Preemptive Generalized Min Sum Set Cover

This section is devoted to proving the following theorem.

Theorem 1 *There is a randomized polynomial time 2-approximation algorithm for Preemptive Generalized Min Sum Set Cover.*

Throughout this section, for any integer $t \in [n]$, the t-th time slot will be equivalent to the time interval $(t-1, t]$.

2.1 Configuration LP

We write a configuration linear program. For a set $S \in \mathcal{S}$, a valid configuration is an (integral) assignment of elements in S to time slots. More formally, such a map can be described as an injective function $f_S : S \rightarrow [n]$. For notational simplicity, we may represent the mapping via a relation (configuration) $F =_{def} \{(e, f_S(e)) \mid e \in S\}$. Let $\mathcal{F}(S)$ denote the collection of all possible configurations for set S. Let C_S^F denote the completion time t of set S under the configuration F, i.e. the first time t' such that $|f_S^{-1}([t'])| \geq \kappa(S)$. Let $x_{e,t}$ denote the fraction of element e we schedule in the t-th time slot. The variable y_S^F is used to indicate which configurations S adheres to. For example, if

$y_S^F = 1$, it means all elements in S are scheduled following the configuration F.

Our integer program is formulated as follows.

$$\min \sum_{S \in \mathcal{S}} \sum_{F \in \mathcal{F}(S)} C_S^F y_S^F \tag{ILP}$$

$$s.t. \quad \sum_{e} x_{e,t} = 1 \qquad \forall\, t \in [n] \tag{1}$$

$$\sum_{t} x_{e,t} = 1 \qquad \forall\, e \in [n] \tag{2}$$

$$\sum_{F \in \mathcal{F}(S)} y_S^F = 1 \qquad \forall\, S \in \mathcal{S} \tag{3}$$

$$\sum_{F \in \mathcal{F}(S), (e,t) \in F} y_S^F = x_{e,t} \qquad \forall\, e, t \in [n], S : e \in S \tag{4}$$

$$x_{e,t} \in \{0,1\} \qquad \forall\, e, t \in [n]$$

$$y_S^F \in \{0,1\} \qquad \forall\, S \in \mathcal{S}, F \in \mathcal{F}(S)$$

The constraints (1) and (2) enforce that exactly one element is scheduled at any time slot and that an element can be scheduled only once over all times. The constraint (3) states that each set S has a unique configuration. Finally, (4) says that if an element e is scheduled at time t, then it must align with the configuration of S.

The relaxation $\mathsf{LP}_{\mathsf{primal}}$ of ILP is then defined as follows.

$$\min \quad \sum_{S \in \mathcal{S}} \sum_{F \in \mathcal{F}(S)} C_S^F y_S^F \tag{$\mathsf{LP}_{\mathsf{primal}}$}$$

$$s.t. \quad \text{Constraints (1),(2),(3) and (4) hold}$$

$$x_{e,t} \geq 0 \qquad \forall\, e, t \in [n]$$

$$y_S^F \geq 0 \qquad \forall\, S \in \mathcal{S}, F \in \mathcal{F}(S)$$

2.2 Validity of the LP

It is easy to verify that $\mathsf{LP}_{\mathsf{primal}}$ is a valid linear programming relaxation for Generalized Min Sum Set Cover. However, it is not obvious that the $\mathsf{LP}_{\mathsf{primal}}$ is indeed a valid relaxation for the preemptive problem. Since we will use two different types of fractional schedules throughout the analysis, we first clearly define/remind those schedules. The first one is a *continuous* schedule that is defined by indicator functions $x_e(t) : [0, n] \to \{0, 1\}, e \in [n]$ such that (1) for any $t \in [0, n], \sum_{e \in [n]} x_e(t) = 1$ and (2) for any $e \in [n]$, $\int_{\tau=0}^{n} x_e(\tau) \, \mathsf{d}\tau = 1$. We say that $x_e(t), e \in [n]$ is a feasible schedule if all these conditions are satisfied. Recall that the completion (or cover) time C_S of each set S is defined by a continuous schedule as the earliest time t such that $\int_{\tau=0}^{t} \sum_{e \in S} x_e(\tau) \, \mathsf{d}\tau \geq$

$\kappa(S)$. The other version of schedule, which is somewhat *discretized*, is defined by $x_{e,t}, e, t \in [n]$ that satisfy (1) $\sum_{e\in[n]} x_{e,t} = 1$, (2) $\sum_{t\in[n]} x_{e,t} = 1$ and (3) $0 \leq x_{e,t} \leq 1$ for any $e, t \in [n]$. When these conditions are satisfied, we will say $x_{e,t}, e, t \in [n]$ is feasible. Note that this discretized version of schedule does not immediately define the completion time of sets since it does not specify how the fractions of various elements are ordered within one time step. Rather, it is used in $\mathsf{LP}_{\mathsf{primal}}$ as a relaxation of continuous schedules. We show the following theorem.

Theorem 2 *Consider any feasible continuous schedule $x_e(t), e \in [n]$. Let C_S denote the completion time of set S in this schedule. For any $e, t \in [n]$, let $x_{e,t} =_{def} \int_{\tau=t-1}^{t} x_e(\tau)\, \mathtt{d}\tau$. Then $x_{e,t}$ satisfy constraints (1) and (2). Also there exists y-values that satisfy the other constraints (3) and (4) as well and further satisfy*

$$\sum_{S\in\mathcal{S}} \sum_{F\in\mathcal{F}(S)} C_S^F y_S^F \leq \sum_{S\in\mathcal{S}} C_S. \tag{5}$$

The first claim in Theorem 2 that $x_{e,t}$ satisfy constraints (1) and (2) easily follows from the properties of continuous schedules and from how $x_{e,t}$ are defined:

$$\sum_{e\in[n]} x_{e,t} = \sum_{e\in[n]} \int_{\tau=t-1}^{t} x_e(\tau)\, \mathtt{d}\tau = \int_{\tau=t-1}^{t} \sum_{e\in[n]} x_e(\tau)\, \mathtt{d}\tau = \int_{\tau=t-1}^{t} 1\, \mathtt{d}\tau = 1,$$

and

$$\sum_{t\in[n]} x_{e,t} = \sum_{t\in[n]} \int_{\tau=t-1}^{t} x_e(\tau)\, \mathtt{d}\tau = \int_{\tau=0}^{n} x_e(\tau)\, \mathtt{d}\tau = 1.$$

In fact, it is not difficult to see that there exist y-values that satisfy all constraints (1)-(4). The following proposition however shows that not all such y-values serve our purpose.

Proposition 1 *There exist y_S^F-values that satisfy (1)-(4), but not (5) within any constant factor.*

Proof Consider the following simple example. The inputs are $\mathcal{S} = \{S_1 = \{e_1, e_2, e_3\}, S_2 = \{e_4, e_5, e_6\}, S_3 = \{e_7, e_8, e_9\}\}$ with $K(S_1) = K(S_2) = K(S_3) = 1$. The given schedule is as follows.

- $x_{e_1 t} = x_{e_2 t+1} = x_{e_3 t+2} = 1/3$ for all $t = 1, 4, 7$.
- $x_{e_4 t} = x_{e_5 t+1} = x_{e_6 t+2} = 1/3$ for all $t = 1, 4, 7$.
- $x_{e_7 t} = x_{e_8 t+1} = x_{e_9 t+2} = 1/3$ for all $t = 1, 4, 7$.

Note that all sets S_1, S_2 and S_3 are completed at time 3 in the above schedule, i.e. $C_{S_1} = C_{S_2} = C_{S_3} = 3$. Consider the following configurations and y-values that satisfy all constraints (1)-(4). See Figure 2.2.

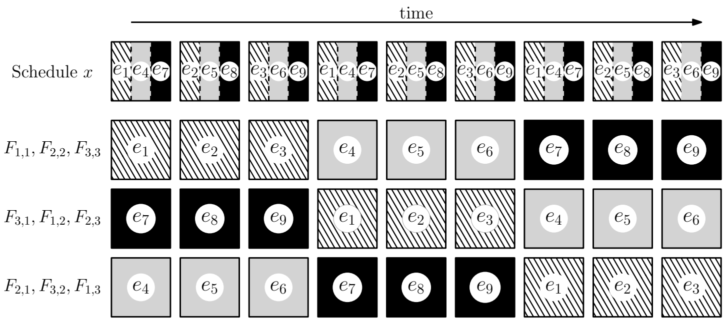

Fig. 1 Elements from the sets S_1, S_2, S_3 are striped, gray and black. The top row depicts the fractional schedule, and the three lower rows depict the decomposition into configurations.

- $F_{1,1} = \{(e_1, 1), (e_2, 2), (e_3, 3)\}$, $F_{1,2} = \{(e_1, 4), (e_2, 5), (e_3, 6)\}$, $F_{1,3} = \{(e_1, 7), (e_2, 8), (e_3, 9)\}$; $y_{S_1}^{F_{1,1}} = y_{S_1}^{F_{1,2}} = y_{S_1}^{F_{1,3}} = 1/3$.
- $F_{2,1} = \{(e_4, 1), (e_5, 2), (e_6, 3)\}$, $F_{2,2} = \{(e_4, 4), (e_5, 5), (e_6, 6)\}$, $F_{2,3} = \{(e_4, 7), (e_5, 8), (e_6, 9)\}$; $y_{S_2}^{F_{2,1}} = y_{S_2}^{F_{2,2}} = y_{S_2}^{F_{2,3}} = 1/3$.
- $F_{3,1} = \{(e_7, 1), (e_8, 2), (e_9, 3)\}$, $F_{3,2} = \{(e_7, 4), (e_8, 5), (e_9, 6)\}$, $F_{3,3} = \{(e_7, 7), (e_8, 8), (e_9, 9)\}$; $y_{S_3}^{F_{3,1}} = y_{S_3}^{F_{3,2}} = y_{S_3}^{F_{3,3}} = 1/3$.

The above configurations and y variables give a LHS for (5) of at least 1 + 4 + 7 = 12. One can easily adapt this instance to make the LHS arbitrarily greater than the RHS. $\square$

Henceforth, we focus on showing that there exist "good" y-values that also satisfy (5). We will show how to construct a feasible solution y such that the inequality

$$\sum_{F \in \mathcal{F}(S)} C_S^F y_S^F \le C_S \tag{6}$$

holds for any set $S \in \mathcal{S}$ which will imply the inequality (5). Since setting y_S-values for a specific S does not affect other y-values, we can focus on each $S \in \mathcal{S}$ separately. We will find "good" y_S^F-values that satisfy constraints (3) and (4), and further (6).

To this end, we define two matroids M_1 and M_2 that enforce that any independent set in the intersection of M_1 and M_2 which in addition is a base in M_1 corresponds to a feasible configuration $F \in \mathcal{F}(S)$. Then we show that the vector $x_{e,t}, e \in S, t \in [n]$ lies in the intersection of the polytopes of the two matroids. Using the fact that such an intersection polytope is integral, we will be able to decompose x into a convex combination of integer points that lie in the intersection of the polytopes of M_1 and M_2. As already mentioned, due to the structure of the matroids, each integer point will correspond to a configuration $F \in \mathcal{F}(S)$. By setting y-values as suggested by the decomposition, we will guarantee that y satisfy constraints (3) and (4). Finally, we will complete the analysis by showing that such y-values satisfy (6) as well. This is enabled by some additional constraints we impose on the matroids. We refer the reader to Chapters 39-41 in [14] for an extensive overview of algorithmic matroid theory.

We begin with defining each of the two matroids M_1 and M_2 which have the same common ground set, $U = \{(e,t) \mid e \in S, t \in [n]\}$ (Recall that we are focusing on each fixed $S \in \mathcal{S}$ separately). We will call (e,t) a pair in order to distinguish it from elements, $[n]$. The first matroid $M_1 = (U, \mathcal{I}(M_1))$ enforces that each element in S can be scheduled in at most one time slot. Formally, the collection $\mathcal{I}(M_1)$ of independent sets of M_1 is defined as follows: $A \in \mathcal{I}(M_1)$ if and only if for any $e \in S$, $|A \cap \{(e,t) \mid t \in [n]\}| \leq 1$. Observe that M_1 is a partition matroid since pairs in U are partitioned based on each common element, and any independent set collects at most one pair from each group. Hence the polytope $P(M_1)$ of M_1 (polymatroid) is defined as follows.

$$\begin{aligned} \sum_{t \in [n]} x_{e,t} &\leq 1 && \forall e \in S && (P(M_1)) \\ x_{e,t} &\geq 0 && \forall e \in S, t \in [n] \end{aligned}$$

Proposition 2 *The vector $x = (x_{e,t})$, $e \in S$, $t \in [n]$ is in the polytope $P(M_1)$. Moreover, $\sum_{e \in S, t \in [n]} x_{e,t} = |S|$, i.e. x belongs to the base polymatroid of M_1.*

Proof For any $e \in S$, from the definition of $x_{e,t}$, we know that $\sum_{t \in [n]} x_{e,t} = \sum_{t=1}^{n} \int_{\tau=t-1}^{t} x_e(\tau)\, \mathtt{d}\tau = \int_{\tau=0}^{n} x_e(\tau)\, \mathtt{d}\tau = 1$. □

The second matroid $M_2 = (U, \mathcal{I}(M_2))$ has a more involved structure. It enforces that in each time slot, at most one element in S can be scheduled. Additionally, it enforces that at most $\kappa(S)$ elements can be scheduled during the first $C-1$ time slots and at most $|S| - \kappa(S)$ elements can be scheduled during the time slots, $C+1, C+2, ..., n$, where C is an integer such that $C - 1 < C_S \leq C$. These additional constraints will be crucial in finding "good" y-values. Formally, $A \in \mathcal{I}(M_2)$ if and only if A satisfies

- For each integer time $t \in [n]$, $|A \cap \{(e,t) \mid e \in S\}| \leq 1$.
- $|A \cap \{(e,t) \mid e \in S, 1 \leq t \leq C-1\}| \leq \kappa(S)$.
- $|A \cap \{(e,t) \mid e \in S, C+1 \leq t \leq n\}| \leq |S| - \kappa(S)$.

We observe that $\mathcal{I}(M_2)$ is a laminar matroid: All pairs in U are partitioned into groups with the same time t, and at most one pair can be chosen from each group to be in an independent set. Further, the second and third constraints put a limit on the number of pairs that can be chosen from the groups of time slots $t = 1, 2, ..., C-1$ and from the groups of time slots $t = C+1, C+2, ..., n$, respectively. We define the polymatroid $P(M_2)$ as follows.

$$\begin{aligned}
\sum_{e\in S} x_{e,t} &\leq 1 && \forall t \in [n] && (P(M_2))\\
\sum_{t=1}^{C-1}\sum_{e\in S} x_{e,t} &\leq \kappa(S) \\
\sum_{t=C+1}^{n}\sum_{e\in S} x_{e,t} &\leq |S| - \kappa(S) \\
x_{e,t} &\geq 0 && \forall e \in S, t \in [n]
\end{aligned}$$

Proposition 3 *The vector $x = (x_{e,t})$ lies in the polymatroid $P(M_2)$.*

Proof We begin with proving that $x_{e,t}$ satisfies the first constraint. From the definition of $x_{e,t}$, we have that

$$\sum_{e\in S} x_{e,t} = \sum_{e\in S}\int_{\tau=t-1}^{t} x_e(\tau)\ \mathtt{d}\tau = \int_{\tau=t-1}^{t}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau \leq \int_{\tau=t-1}^{t} 1\ \mathtt{d}\tau = 1.$$

Now we consider the second constraint. Recall that $C - 1 < C_S \leq C$.

$$\sum_{t=1}^{C-1}\sum_{e\in S} x_{e,t} = \int_{\tau=0}^{C-1}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau \leq \int_{\tau=0}^{C_S}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau = \kappa(S)$$

The last inequality is due to the definition of C_S. Finally,

$$\begin{aligned}
\sum_{t=C+1}^{n}\sum_{e\in S} x_{e,t} &= \int_{\tau=C}^{n}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau \leq \int_{\tau=C_S}^{n}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau \\
&= |S| - \int_{\tau=0}^{C_S}\sum_{e\in S} x_e(\tau)\ \mathtt{d}\tau = |S| - \kappa(S).
\end{aligned}$$

□

It is well known the intersection of two polymatroids is an integral polytope (see e.g. [14]), i.e. any vertex point is integral. Hence since $(x_{e,t})$ lies in the intersection of two polytopes $P(M_1)$ and $P(M_2)$, it can be decomposed into a linear combination of vertex (hence integer) points in $P(M_1) \cap P(M_2)$. Note that each of such integer points corresponds to an independent set in $\mathcal{I}(M_1) \cap \mathcal{I}(M_2)$, which is of size at most $|S|$ due to the constraints of M_1. In fact, the size must be exactly $|S|$, since $\sum_{e\in S}\sum_{t\in[n]} x_{e,t} = |S|$. By the constraints of M_1 and the first constraints of M_2, we conclude that each of such integer points corresponds to a configuration $F \in \mathcal{F}(S)$. Hence we have shown the following lemma.

Lemma 1 *There exist $\mathcal{F}'(S) \subseteq \mathcal{F}(S)$ and positive constants $\theta_S^F, F \in \mathcal{F}'(S)$ that satisfy*

– $\sum_{F\in\mathcal{F}'(S)} \theta_S^F = 1$.
– *For any* $e \in S$, $t \in [n]$, $x_{e,t} = \sum_{F\in\mathcal{F}'(S)} \theta_S^F \cdot \mathbf{1}[(e,t) \in F]$.

where an indicator variable $\mathbf{1}[(e,t) \in F] = 1$ *if and only if* $(e,t) \in F$.

We let $y_S^F = \theta_S^F$ for all $F \in \mathcal{F}'(S)$ and $y_S^F = 0$ for all $F \in \mathcal{F}(S) \setminus \mathcal{F}'(S)$. Note that x and y satisfy constraints (3) and (4).

It remains to show that y satisfy (6). Now the second and third constraints of M_2 play a crucial role. We make the following observation.

Lemma 2 *For any* $F \in \mathcal{F}'(S)$ *exactly one of the following holds.*

– $|F \cap \{(e,t) \mid e \in S, 1 \le t \le C-1\}| = \kappa(S)$.
– $|F \cap \{(e,t) \mid e \in S, 1 \le t \le C-1\}| = \kappa(S) - 1$ *and* $(e, C) \in F$ *for some* $e \in S$.

Proof Recall that $|F| = |S|$. By the third constraints of M_2, we know that $N_{\ge C+1} =_{def} |F \cap \{(e,t) \mid e \in S, C+1 \le t \le n\}| \le |S| - \kappa(S)$, hence that $N_{\ge C} =_{def} |F \cap \{(e,t) \mid e \in S, C \le t \le n\}| \le |S| - \kappa(S) + 1$. Therefore, we have $N_{\le C-1} =_{def} |F \cap \{(e,t) \mid e \in S, 1 \le t \le C-1\}| \ge \kappa(S) - 1$. Further, we know $N_{\le C-1} \le \kappa(S)$ from the second constraint of M_2. Thus unless $N_{\le C-1} = \kappa(S)$, it must be the case that $N_{\le C-1} = \kappa(S) - 1$. In that case, since $N_{\ge C+1} \le |S| - \kappa(S)$, we conclude that $(e, C) \in F$ for some $e \in S$. □

Motivated by the above lemma, we can now prove that our linear program is a valid relaxation for the preemptive version of the problem.

Proof (Proof of Theorem 2) Partition $\mathcal{F}'(S)$ into $\mathcal{F}'_1(S)$ and $\mathcal{F}'_2(S)$ by letting $\mathcal{F}'_1(S)$ to denote all $F \in \mathcal{F}'(S)$ that fall in the first case in the Lemma 2 and letting $\mathcal{F}'_2(S) = \mathcal{F}'(S) \setminus \mathcal{F}'_1(S)$. Let $\theta' = \sum_{F\in\mathcal{F}'_2(S)} \theta_S^F$. Note that for any $F \in \mathcal{F}'_1(S)$, $C_S^F \le C - 1$ and for any $F \in \mathcal{F}'_2(S)$, $C_S^F = C$. In words, the set S is completed no later than time $C-1$ for $(1-\theta')$ fraction of configurations in $\mathcal{F}'(S)$ and exactly at time C for θ' fraction of configurations in $\mathcal{F}'(S)$. Hence we have that

$$\sum_{F\in\mathcal{F}(S)} C_S^F y_S^F = \sum_{F\in\mathcal{F}'(S)} C_S^F \theta_S^F = \sum_{F\in\mathcal{F}'_1(S)} C_S^F \theta_S^F + \sum_{F\in\mathcal{F}'_2(S)} C_S^F \theta_S^F$$
$$\le (1-\theta')(C-1) + \theta' C = C - 1 + \theta'. \tag{7}$$

Now we focus on upper-bounding θ'. From the definition of C_S and the fact that $\sum_{e\in S} x_e(\tau) \le 1$ for any τ, we know that

$$\int_{\tau=0}^{C-1} \sum_{e\in S} x_e(\tau)\,\mathtt{d}\tau = \int_{\tau=0}^{C_S} \sum_{e\in S} x_e(\tau)\,\mathtt{d}\tau - \int_{\tau=C-1}^{C_S} \sum_{e\in S} x_e(\tau)\,\mathtt{d}\tau$$
$$\ge \kappa(S) - (C_S - (C-1)). \tag{8}$$

On the other hand, it follows that

$$\int_{\tau=0}^{C-1} \sum_{e \in S} x_e(\tau)\,\mathsf{d}\tau = \sum_{t=1}^{C-1} \sum_{e \in S} x_{e,t} \qquad \text{[By the definition of } x_{e,t}\text{]}$$

$$= \sum_{t=1}^{C-1} \sum_{e \in S} \sum_{F \in \mathcal{F}'(S):(e,t) \in F} y_S^F \qquad \text{[From the decomposition of } x \text{ into } y_S^F\text{]}$$

$$= \sum_{F \in \mathcal{F}_1'(S)} y_S^F \sum_{e \in S} \sum_{t=1}^{C-1} \mathbf{1}[(e,t) \in F] + \sum_{F \in \mathcal{F}_2'(S)} y_S^F \sum_{e \in S} \sum_{t=1}^{C-1} \mathbf{1}[(e,t) \in F]$$

$$= \sum_{F \in \mathcal{F}_1'(S)} \theta_S^F \cdot \kappa(S) + \sum_{F \in \mathcal{F}_2'(S)} \theta_S^F \cdot (\kappa(S) - 1)$$

$$= (1 - \theta') \cdot \kappa(S) + \theta' \cdot (\kappa(S) - 1) = \kappa(S) - \theta'. \tag{9}$$

From (8) and (9), we have $\theta' \leq C_S - (C-1)$. By combining this with (7), we complete the proof of Theorem 2. □

2.3 Solving the LP

The linear programming relaxation $\mathsf{LP}_{\mathsf{primal}}$ has exponentially many variables. Hence, we solve the dual LP and show there are only polynomially many non-zero variables in the primal LP that achieve the optimal LP value. The dual LP is as follows.

$$\max \sum_{t \in [n]} \alpha_t + \sum_{e \in [n]} \beta_e + \sum_{S \in \mathcal{S}} \gamma_S \tag{$\mathsf{LP}_{\mathsf{dual}}$}$$

$$s.t. \quad \alpha_t + \beta_e - \sum_{S: e \in S} \delta_{etS} \leq 0 \qquad \forall e, t \tag{10}$$

$$\gamma_S + \sum_{(e,t) \in F} \delta_{etS} \leq C_S^F \qquad \forall S \in \mathcal{S}, F \in \mathcal{F}(S) \tag{11}$$

To solve $\mathsf{LP}_{\mathsf{dual}}$ with the ellipsoid algorithm, we need a *separation oracle* for finding a violated constraint (see [9]). Since constraints (10) are easy to verify (there are only n^2 of them), we focus on constraints (11). We need a polynomial time algorithm that given γ_S and δ_{etS}-values, finds (if any) $S \in \mathcal{S}$ and $F \in \mathcal{F}(S)$ that violate constraints (11).

We model this problem as a classical minimum cost s-t flow problem. In this problem, we are given a digraph $G = (V, A)$, a capacity function $c : A \to \mathbb{Q}_+$, a cost function $k : A \to \mathbb{Q}$ and the volume $\phi \in \mathbb{Q}_+$. The goal is to send ϕ amount of flow from the source s to the sink t, i.e. to find an s-t flow f of volume ϕ, subject to capacity constraints $0 \leq f(e) \leq c(e)$ for all $e \in A$ and the standard flow conservation constraints, minimizing the costs $\sum_{e \in A} f(e)k(e)$.

It is known that if the volume ϕ and capacities $c_e, e \in E$ are integral then we can test in polynomial time if there is an s-t flow of volume ϕ. Moreover, if

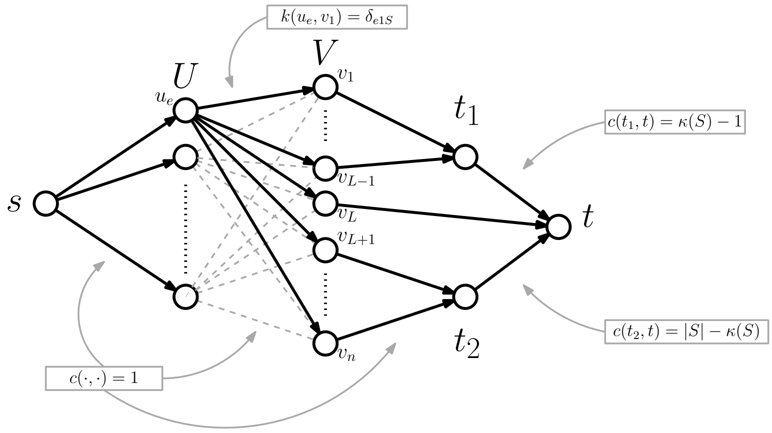

Fig. 2 An illustration of the construction of the graph G, in which we want to find a maximum-value flow.

there is such a flow (i.e. there is a feasible solution to the problem) then there is an integral minimum-cost s-t flow, and it can be found in polynomial time (see Chapter 12 in [14]).

We now show how to reduce our separation problem for constraints (11) to the minimum cost s-t flow problem. It will be convenient for us to consider an equivalent maximum cost s-t flow problem where the goal is to maximize the value of the objective function $\sum_{e \in A} f(e)k(e)$.

Fix a set S and an integer $L \in [n]$. We will try to find a violated constraint for the constraints (11) corresponding to the set S and configurations $F \in \mathcal{F}(S)$ with $C_S^F = L$. Create a directed complete bipartite graph $G_L = (U, V, A)$ where part U has vertex u_e for each $e \in S$, part V has vertex v_i for each time slot $i \in [n]$. Arc $a = (u_e, v_i) \in A$ has cost $k(e) = \delta_{eiS}$ and capacity $c(e) = 1$. We augment G_L as follows. We add a source vertex s and connect it to all vertices in U. There are two "intermediate" sinks t_1 and t_2, both connected to the "final" sink t. The vertices $v_1, v_2, ..., v_{L-1}$ in V are connected to t_1 and the vertices $v_{L+1}, v_{L+2}, ..., v_n$ in V are connected to the other intermediate sink t_2. The arcs a between the source s and part U have cost $k(a) = 0$ and capacity $c(a) = 1$. Analogously, all arcs a between part V and intermediate sinks t_1 and t_2 have cost $k(a) = 0$ and capacity $c(a) = 1$. Arcs $a' = (t_1, t)$ and $a'' = (t_2, t)$ have capacities $c(a') = \kappa(S) - 1$ and $c(a'') = |S| - \kappa(S)$ respectively, and all of them have zero costs. The vertex v_L is special and is directly connected to t. The arc (v_L, t) has a unit capacity and zero cost. The goal is to find the s-t flow of volume $\phi = |S|$ of maximum cost. See Figure 2 for an illustration of this construction.

Note that any integral s-t flow f of value $|S|$ in digraph G_L corresponds to a valid configuration F for set S such that $C_S^F = L$, and vice versa. Hence, if the maximum-cost s-t flow in G_L has cost more than $L - \gamma_S$, the constraint (11) is violated for S and $F \in \mathcal{F}(S)$ that corresponds to the flow. The converse also holds: if the maximum-cost s-t flow has cost less than or equal to $L - \gamma_S$

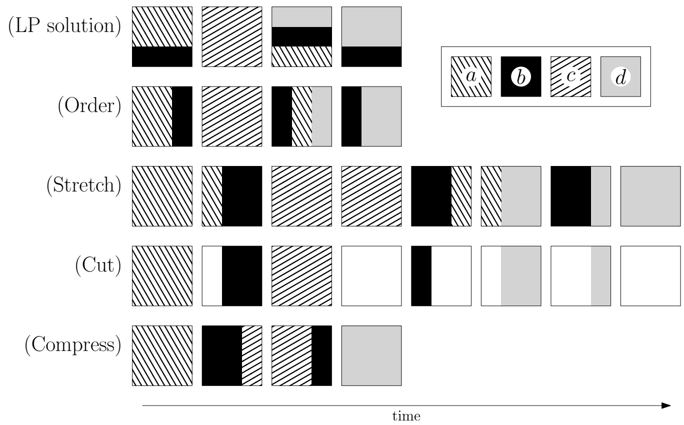

Fig. 3 In this example the schedule is stretched by a factor of two e.g. $\lambda = \frac{1}{2}$.

there is *no* configuration $F \in \mathcal{F}(S)$ with $C_S^F = L$ that violates (11). With the help of this separation oracle and classical connection between separation and optimization [9], we can solve $\mathsf{LP}_{\mathsf{dual}}$ in polynomial time.

Then we can optimally solve $\mathsf{LP}_{\mathsf{primal}}$ by focusing only on y_S^F variables that correspond to the constraints that were considered by the ellipsoid method in solving $\mathsf{LP}_{\mathsf{dual}}$. A more formal (and well-known) argument is that the $\mathsf{LP}_{\mathsf{dual}}$ with the subset of constraints considered by the ellipsoid method is a relaxation of the original problem but it has the same optimal solution. The dual of the relaxed problem is $\mathsf{LP}_{\mathsf{primal}}$ restricted to the subset of corresponding variables which by the strong duality theorem has the same optimal value.

2.4 Rounding procedure

Let $x_{e,t}$ and y_S^F be a basic optimal solution of the linear programming relaxation $\mathsf{LP}_{\mathsf{primal}}$. In particular we know that there are at most $2n + m + n^2m$ non-zero variables (this is the number of constraints (1)-(4)). Let C_S^{LP} denote the completion time of set S in the LP. That is, $C_S^{LP} = \sum_{F\in\mathcal{F}(S)} C_S^F y_S^F$. We create a schedule parameterized by $\lambda \in (0, 1]$, where λ is randomly drawn from $(0, 1]$ according to the density function $f(v) = 2v$.

Create an arbitrary continuous schedule $x_e(t)$, $e \in [n]$, $t \in [0, n]$ from $x_{e,t}, e, t \in [n]$ such that for any $e, t \in [n]$, $\int_{\tau=t-1}^{t} x_e(\tau)\,\mathsf{d}\tau = x_{e,t}$. For example, this can be done by processing each element e for the amount $x_{e,t}$ during the time step t in an arbitrary order between the elements. For notational convenience, let σ denote the continuous schedule $x_e(t)$. We will also use the standard machine scheduling terminology. The new schedule $\sigma(\lambda)$ is defined as follows. Stretch out the schedule σ by a factor of $\frac{1}{\lambda}$. In other words, map every point τ in time onto τ/λ. For each element e define $\tau_e \in [1, n/\lambda]$ to be the earliest point in time when the element has been processed for one time unit (out of total $1/\lambda$). Leave the machine idle whenever it processes

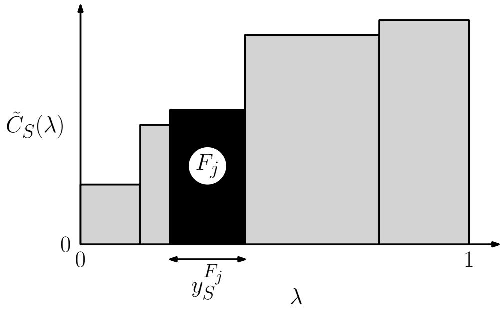

Fig. 4 An illustration of $\tilde{C}_S(\lambda)$. The jth leftmost rectangle, which corresponds to F_j, has width $y_S^{F_j}$ and height $C_S^{F_j}$.

the element e after time τ_e. After repeating this procedure for all elements $e \in [n]$, we shift the whole schedule to the left to eliminate all idle times. The final schedule $\sigma(\lambda)$ has total length n. Let $x_e^{(\lambda)}(t), e \in [n], t \in [0, n]$ be the resulting continuous schedule $\sigma(\lambda)$. Note that similar algorithms were used in scheduling before to design approximation algorithms for various preemptive scheduling problems with total completion time objective [15, 13].

Example 1 See Figure 3 for an illustration. Consider an instance with 4 elements $\{a, b, c, d\}$, with the LP solution $x_{a,1} = 2/3$, $x_{b,1} = 1/3$, $x_{c,2} = 1$, $x_{d,3} = 1/3$, $x_{b,3} = 1/3$, $x_{a,3} = 1/3$, $x_{d,4} = 2/3$, $x_{b,4} = 1/3$. Construct a continuous schedule by randomly *ordering* the elements in each time step. For example in time step 3, three elements, a, b, d are scheduled seamlessly, each for 1/3 time steps. Then *stretch* the whole schedule by a factor two ($\lambda = 1/2$), and *cut* out each element after being scheduled by a unit amount. Finally, *compress* the schedule, by shifting everything to the left removing the idle times.

Let $C_S(\lambda)$ denote the completion time of S in the new schedule $\sigma(\lambda)$. Order all configurations $F \in \mathcal{F}(S)$ for $y_S^F > 0$ in non-decreasing order of C_S^F. Let $F_1, F_2, ..., F_k$ be such an ordering. Define $\tilde{C}_S(\lambda) =_{def} C_S^{F_j}$ where $\sum_{i=1}^{j-1} y_S^{F_i} < \lambda$ and $\sum_{i=1}^{j} y_S^{F_i} \geq \lambda$. See Figure 4 for an illustration. Let $\mathbf{1}[\phi]$ be an indicator function such that $\mathbf{1}[\phi] = 1$ if and only if ϕ is true and zero otherwise.

Lemma 3 *For any $S \in \mathcal{S}$ and $0 < \lambda \leq 1$, $C_S(\lambda) \leq \frac{1}{\lambda} \cdot \tilde{C}_S(\lambda)$.*

Proof To simplify the proof we assume that there exists j such that $\sum_{1 \leq l \leq j} y_S^{F_l} = \lambda$. Otherwise, let j be the lowest index such that $\sum_{1 \leq l \leq j} y_S^{F_l} > \lambda$, then we define two copies F_j' and F_j'' of configuration F_j, with $y_S^{F_j'} = \lambda - \sum_{1 \leq l \leq j-1} y_S^{F_l}$ and $y_S^{F_j''} = \sum_{1 \leq l \leq j} y_S^{F_l} - \lambda$. Here F_j' and F_j'' are the same configurations as F_j. Now, $\sum_{1 \leq l \leq j-1}(y_S^{F_l}) + y_S^{F_j'} = \lambda$.

We will show the following inequality:

$$\int_{\tau=0}^{\tilde{C}_S(\lambda)/\lambda} \sum_{e\in S} x_e^{(\lambda)}(\tau)\,\mathsf{d}\tau \geq \kappa(S) \tag{12}$$

that would imply that the completion time $C_S(\lambda)$ of the set S in the schedule $\sigma(\lambda)$ must be no later than $\tilde{C}_S(\lambda)/\lambda$. Since for every $e \in S$ we have $\int_{\tau=0}^{\tilde{C}_S(\lambda)/\lambda} x_e^{(\lambda)}(\tau)\,\mathsf{d}\tau \geq \min\left\{1, \frac{1}{\lambda}\int_{\tau=0}^{\tilde{C}_S(\lambda)} x_e(\tau)\,\mathsf{d}\tau\right\} \geq \min\left\{1, \frac{1}{\lambda}\sum_{t\leq\lfloor\tilde{C}_S(\lambda)\rfloor} x_{e,t}\right\}$, and $\tilde{C}_S(\lambda)$ is integral by definition for any $\lambda \in (0,1]$, it is sufficient to show the inequality

$$\sum_{e\in S} \min\left\{\lambda, \sum_{t\leq\tilde{C}_S(\lambda)} x_{e,t}\right\} \geq \lambda\kappa(S) \tag{13}$$

to derive (12). We now derive the inequality (13).

$$\begin{aligned}
\sum_{e\in S} \min\left\{\lambda, \sum_{t\leq\tilde{C}_S(\lambda)} x_{e,t}\right\} &\geq \sum_{e\in S} \min\left\{\lambda, \sum_{t\leq\tilde{C}(\lambda)} \sum_{l=1}^{j} y_S^{F_l}\cdot\mathbf{1}[(e,t)\in F_l]\right\} \\
&= \sum_{e\in S} \min\left\{\lambda, \sum_{l=1}^{j} y_S^{F_l}\cdot\mathbf{1}[(e,t)\in F_l \;\text{ for some }\; t\leq\tilde{C}(\lambda)]\right\} \\
&= \sum_{e\in S}\sum_{l=1}^{j} y_S^{F_l}\cdot\mathbf{1}[(e,t)\in F_l \;\text{ for some } t\leq\tilde{C}(\lambda)] \\
&= \sum_{l=1}^{j} y_S^{F_l}\sum_{e\in S}\mathbf{1}[(e,t)\in F_l \;\text{ for some } t\leq\tilde{C}(\lambda)] \\
&\geq \sum_{l=1}^{j} y_S^{F_l}\kappa(S) = \lambda\kappa(S)
\end{aligned}$$

The first inequality follows from constraints (4). The second equality holds because $\sum_{l=1}^{j} y_S^{F_l} = \lambda$. The last inequality holds because for any F_l, $l \leq j$, $C_S^{F_l} \leq \tilde{C}_S(\lambda)$. □

The following lemma can be easily shown from the definition of $\tilde{C}_S(\lambda)$ (see also Figure 4).

Lemma 4 *For any $S \in \mathcal{S}$, $\int_{\lambda=0}^{1} \tilde{C}_S(\lambda)\,d\lambda = C_S^{LP}$.*

Proof (Proof of Theorem 1) By Theorem 2, $\mathsf{LP}_{\mathsf{primal}}$ is a valid relaxation, and we now estimate the expected set cover time in the approximate solution.

$$\begin{aligned}
\mathbb{E}[C_S(\lambda)] &= \int_{\lambda=0}^{1} C_S(\lambda)\cdot 2\lambda\;\mathsf{d}\lambda \qquad \text{[By definition]} \\
&\leq \int_{\lambda=0}^{1} \frac{1}{\lambda}\cdot\tilde{C}_S(\lambda)\cdot 2\lambda\;\mathsf{d}\lambda \qquad \text{[By Lemma 3]} \\
&= 2\int_{\lambda=0}^{1} \tilde{C}_S(\lambda)\;\mathsf{d}\lambda \quad = 2C_S^{LP} \qquad \text{[By Lemma 4]}
\end{aligned}$$

□

We shortly indicate how our approximation algorithm can be derandomized. The function $\tilde{C}_S(\lambda)$ is a piecewise constant function, with at most a polynomial number of pieces since there are at most polynomially many non-zero variables y_S^F for each S. This implies that there are at most polynomially many "interesting" λ-values that we need to consider, among which at least one gives the desired approximation ratio.

3 Gap between Preemptive and Non-preemptive Schedules

In this section, we study the lower bound quality of the preemptive problem for the non-preemptive problem. Note that if we show a way to convert any given preemptive schedule into a non-preemptive one losing a factor of η, we would immediately obtain a 2η-approximation algorithm for the non-preemptive Generalized Min Sum Set Cover.

Our scheme for transforming a preemptive schedule into a non-preemptive one is similar to the one used by Skutella and Williamson [16]. We obtain a better gap by utilizing several additional tricks *and* starting from a preemptive schedule. Formally we will prove the following theorem.

Theorem 3 *Given a preemptive schedule with cost C, there exists a non-preemptive schedule with expected cost at most $6.2C$. Furthermore, this non-preemptive schedule can be found in polynomial time.*

Combining Theorem 1 and Theorem 3 we derive

Theorem 4 *There exists a polynomial time 12.4-approximation algorithm for Generalized Min Sum Set Cover.*

Theorem 3 implies an upper bound on the gap of 6.2, and any gap lower than 2 would result in an approximation factor strictly less than 4 for the non-preemptive problem, which is impossible unless P=NP [8]. We believe that our gap is not tight. In fact, we make the following bold conjecture:

Conjecture 1 Given a preemptive schedule with cost C then there is a non-preemptive schedule with cost at most $2C$. Further, such a non-preemptive schedule can be found in polynomial time.

It would be also interesting to show if the optimal gap between values of preemptive and non-preemptive schedules depends on parameter $\xi = \min_S\{\kappa(S)/|S|\}$. For example, we know if $\xi = 1$ then there is no advantage for preemptive schedules, i.e. $\eta = 1$ in this case.

The remaining section is devoted to proving Theorem 3. We start with defining several schedules that will be used throughout the analysis. Let $x_e(t), e \in [n], t \in [0, n]$ be a given preemptive solution where $x_e(t) \in \{0, 1\}$ is the indicator function if an element e is scheduled at time t. It would be convenient to extend the domain of $x_e(t)$ to $[0, \infty)$ by setting $x_e(t) = 0$ for $t > n$. Let

C_S^P be the cover time of set S in the preemptive schedule corresponding to the solution $x_e(t)$; the superscript P stands for the preemptive schedule. Our goal is to randomly construct a non-preemptive schedule that completes set S at time C_S^R such that $\mathbb{E}[C_S^R] = O(1)C_S^P$. In the following we will use the notion of a solution (feasible or infeasible) interchangeably with the notion of a schedule.

We will define a new fractional solution (and a schedule) $\tilde{x}$ from x. Let $\sigma =_{def} \sigma_0$ denote the schedule defined by x. For each integer $i \geq 1$, stretch out σ horizontally by a factor of r^i and let σ_i be the resulting schedule. Here $r \geq 1$ is a constant to be fixed later. More formally, $x_e^{(i)}(r^i t) = \frac{1}{r^i} x_e(t)$ for all $e \in [n]$, defines the schedule σ_i. Note that now we allow $x_e^{(i)}(\tau)$ to be non-boolean: it denotes the rate at which we process element e. Note that σ_i schedules element e during $[r^i t, r^i(t+\mathtt{d}t)]$ by the same amount as σ does during $[t, t+\mathtt{d}t]$. For two parameters $Q \geq 0$ and $\rho \geq 0$, which we will fix later, define $\tilde{x}_e(t)$ as follows:

$$\tilde{x}_e(t) =_{def} Q \left(x_e(t) + \rho \sum_{i=1}^{\infty} x_e^{(i)}(t) \right).$$

Note that $\tilde{x}_e(t)$ may not yield a feasible preemptive schedule because it may schedule elements at a rate of more than one at an instantaneous time. Let $\tilde{\sigma}$ denote the (infeasible) fractional schedule defined by $\tilde{x}_e(t)$. Via a randomized rounding, we will first obtain an intermediate infeasible integral schedule σ^I and then the final feasible integral schedule σ^R. Throughout the analysis, we will be mostly concerned with the intermediate schedules $\tilde{\sigma}$ and σ^I. In these schedules, we are allowed to schedule more than one elements at some times, and will define the cover time of sets in a natural way; the formal definitions will be given later when necessary.

Example 2 See Figure 5 for an illustration. We set $Q = 2$, $\rho = 1$ and $r = 2$. Consider an instance with 4 elements $\{a, b, c, d\}$, with a preemptive schedule $(x_e(t))_{e \in [n]} =_{def} \frac{3}{4}a, \frac{1}{4}b, 1c, \frac{1}{2}b, \frac{1}{4}a, \frac{1}{4}d, \frac{1}{4}b, \frac{3}{4}d$; for example, $\frac{3}{4}a$ implies that element a is schedule for $\frac{3}{4}$ unit times. Then we create schedules $x^{(i)}$ from x by stretching out x horizontally by a factor r^i. By adding Q copies of $x^{(0)}$ and $Q\rho$ copies of each $x^{(i)}$, $i \geq 1$, we obtain the "thick" schedule $\tilde{x}$.

Suppose $\alpha_a = 1/2$, $\alpha_b = 1/4$, $\alpha_c = 1$ and $\alpha_d = 1/2$. Then, $t_{a,\alpha_a} = 1/8$, $t_{b,\alpha_b} = 7/8$, $t_{c,\alpha_c} = 3/2$ and $t_{d,\alpha_d} = 3$.

The following lemma easily follows from the definition of $\tilde{x}_e(t)$.

Lemma 5 *For any time* $t \in [0, \infty)$,

$$\int_{\tau=0}^{t} \sum_{e \in [n]} \tilde{x}_e(\tau)\, \mathtt{d}\tau \leq Q \left(1 + \frac{\rho}{r-1}\right) t.$$

Proof The desired bound easily follows by the definition of $\tilde{x}$ and the fact that σ_i, $i \geq 0$ schedules elements by an amount of at most t/r^i until time t. $\square$

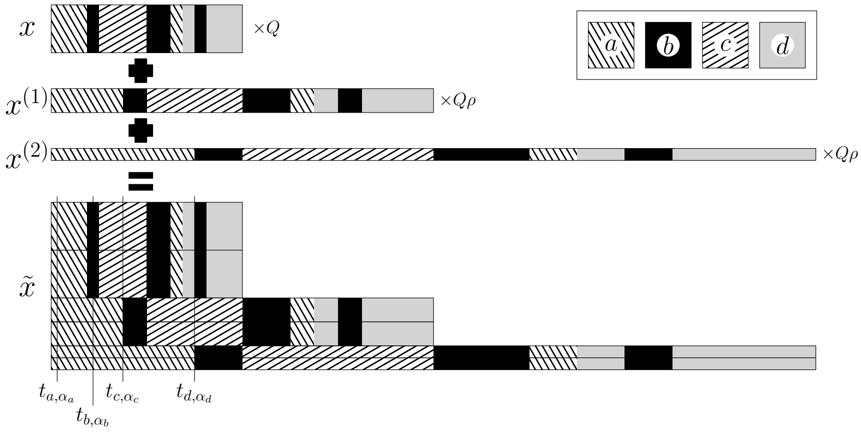

Fig. 5 An illustration of the construction of schedule $\tilde{x}$ from x, with parameters $r = 2$, $\rho = 1$ and $Q = 2$ and alpha values $\alpha_a = 1/2$, $\alpha_b = 1/4$, $\alpha_c = 1$ and $\alpha_d = 1/2$.

We now give our rounding procedure. For each $e \in [n]$, choose $\alpha_e \in [0, 1]$ uniformly at random. Let t_{e,α_e} be the first time t such that $\int_0^t \tilde{x}_e(\tau)\mathtt{d}\tau \geq \alpha_e$. Let σ^I denote the resulting (infeasible) schedule where e is scheduled at time t_{e,α_e}. Here when we say that element e is scheduled at time t_{e,α_e} in σ^I, we ignore that element e takes a unit amount of time to be completely scheduled; this will be taken care of in the final schedule. Rather, we think of element e as being tied to the instantaneous time t_{e,α_e}. In the same spirit, we define the cover time C^I_S of S in σ^I as the first time t such that $|\{e \in S \mid t_{e,\alpha_e} \leq t\}| \geq \kappa(S)$. We will schedule the elements in non-decreasing order of t_{e,α_e} as our final schedule σ^R, breaking ties arbitrarily. The algorithm is based on a popular scheduling concept of α-points and similar to the one in [16].

Let $p_{S,i}$ denote the probability that S is not satisfied until time $r^i \cdot C^P_S$ in σ^I, i.e.

$$p_{S,i} =_{def} \Pr[|\{e \in S \mid t_{e,\alpha_e} \leq r^i \cdot C^P_S\}| < \kappa(S)]$$

In the following lemma we upper bound $p_i =_{def} \max_{S \in \mathcal{S}} p_{S,i}$.

Lemma 6 *Consider any $Q \geq 1$, $r > 1$ and $\rho \leq 1$. Then we have $\max_S p_{S,i} = p_i \leq \max\{K_{1i}, K_{2i}, K_{3i}\}$ where*

$$\begin{aligned}
K_{1i} &= \exp(-Q(1 + \rho i)), \\
K_{2i} &= \exp(-2Q(1 + \rho i)) + 2Q(1 + \rho i)\exp(-2Q(1 + \rho i) + 1), \\
K_{3i} &= \exp\left(-1.5Q\left(\left(1 - \frac{1}{(1 + \rho i)Q}\right)^2 (1 + \rho i)\right)\right).
\end{aligned}$$

Proof Consider any S and fixed $i \geq 0$. Let $A =_{def} \left\{e \in S \ : \ \int_0^{r^i \cdot C^P_S} \tilde{x}_e(\tau)\mathtt{d}\tau \geq 1\right\}$. Observe that any element $e \in A$ is scheduled no later than time $r^i \cdot C^P_S$ in the schedule σ^I. Note by definition of C^P_S that $\sum_{e \in S} \int_{\tau=0}^{C^P_S} x_e(\tau)\, \mathtt{d}\tau \geq \kappa(S)$. Hence it follows (for all $A \subseteq S$) that

$$\sum_{e \in S \setminus A} \int_{\tau=0}^{C_S^P} x_e(\tau)\, \mathtt{d}\tau \geq \kappa(S) - |A|.$$

Then from the definition of $\tilde{x}$ and by observing that for every $i \geq 0$ each σ_i schedules elements in $S \setminus A$ by an amount of at least $\kappa(S) - |A|$ until time $r^i \cdot C_S^P$, for any $i \geq 0$,

$$\sum_{e \in S \setminus A} \int_{\tau=0}^{r^i \cdot C_S^P} \tilde{x}_e(\tau)\, \mathtt{d}\tau \geq Q(\kappa(S) - |A|)(1 + \rho i).$$

For any $e \in S \setminus A$, let X_e denote an indicator random variable such that $X_e = 1$ if and only if $t_{e,\alpha_e} \leq r^i \cdot C_S^P$. Observe that X_e are independent of each other, since the value of X_e is determined by α_e, which is randomly chosen independent of the other elements. Let $X =_{def} \sum_{e \in S \setminus A} X_e$ and $\mu =_{def} \mathbb{E}[X]$. Then, by observing that $\Pr[X_e = 1] = \int_{\tau=0}^{r^i \cdot C_S^P} \tilde{x}_e(\tau)\ \mathtt{d}\tau$, it follows that

$$\mu = \mathbb{E}[X] = \sum_{e \in S \setminus A} \int_{\tau=0}^{r^i \cdot C_S^P} \tilde{x}_e(\tau)\ \mathtt{d}\tau \geq Q \cdot (\kappa(S) - |A|) \cdot (1 + \rho i).$$

Note that $p_{S,i} \leq \Pr[X < \kappa(S) - |A|]$. We consider a few cases depending on the value of $\kappa(S) - |A|$. Assume that $\kappa(S) > |A|$, since otherwise $p_{S,i} = 0$.

`Case (a):` $\kappa(S) - |A| = 1$. Using the fact that X_e are independent of each other, we have

$$\begin{aligned}
\Pr[X < \kappa(S) - |A|] &= \Pr[X = 0] \\
&= \prod_{e \in S \setminus A} (1 - \Pr[X_e = 1]) \\
&\leq \exp(-\sum_{e \in S \setminus A} \Pr[X_e = 1]) = \exp(-\mu) \\
&\leq \exp(-Q(1 + \rho i)).
\end{aligned}$$

`Case (b):` $\kappa(S) - |A| = 2$.

$$\begin{aligned}
&\Pr[X < \kappa(S) - |A|] \\
&= \Pr[X = 0] + \Pr[X = 1] \\
&\leq \exp(-\mu) + \sum_{e' \in S \setminus A} \Pr[X_{e'} = 1] \prod_{e \in S \setminus (A \setminus \{e'\})} (1 - \Pr[X_e = 1]) \\
&\leq \exp(-2Q(1 + \rho i)) + \sum_{e' \in S \setminus A} \Pr[X_{e'} = 1] \exp(-\sum_{e \in S \setminus (A \setminus \{e'\})} \Pr[X_e = 1]) \\
&\leq \exp(-2Q(1 + \rho i)) + \mu \cdot \exp(-\mu + 1) \\
&\leq \exp(-2Q(1 + \rho i)) + 2Q(1 + \rho i) \exp(-2Q(1 + \rho i) + 1)
\end{aligned}$$

`Case (c):` $\kappa(S) - |A| \geq 3$. This case can be done similarly as is done in [16]. The main difference is that here we are using the fact that $K - |A| \geq 3$, which helps to obtain a tighter bound.

$$\begin{aligned}
\Pr[X < \kappa(S) - |A|] &\leq \Pr\Big[X < \frac{\mu}{Q(1+\rho i)}\Big] \\
&= \Pr\Big[X < \mu\Big(1 - \Big(1 - \frac{1}{Q(1+\rho i)}\Big)\Big)\Big] \\
&\leq \exp\Big(-\frac{1}{2}\Big(1 - \frac{1}{Q(1+\rho i)}\Big)^2 \cdot \mu\Big) \\
&\leq \exp\Big(-\frac{1}{2}\Big(1 - \frac{1}{Q(1+\rho i)}\Big)^2 \cdot 3Q(1+\rho i)\Big)
\end{aligned}$$

The second inequality comes from the Chernoff Bounds $\Pr[X < \mu(1-\delta)] \leq \exp(-\frac{1}{2}\delta^2\mu)$ [12]. Taking the maximum of the bounds in the above three cases completes the proof. □

In the following lemma, we bound the total expected cover time in the final schedule σ^R. Before giving a formal proof, we give a high-level explanation on how we obtain the upper bound in the lemma. First, the loss of factor $Q(1 + \frac{\rho}{r-1})$ comes from flatenning out the "thick" schedule $\tilde{\sigma}$ which was obtained by overlaying multiple schedules σ_i. The thickness is upper-bounded in Lemma 5. The other term $\Big(1 + (r-1)\sum_{i=0}^{\infty} r^i p_i\Big)$ follows from the definition of p_i: Set S is covered no later than time $r^i \cdot C_S^P$ with a probability of at least $1 - p_i$. The final term comes from the following. In obtaining the final schedule σ^R, we count the expected number of elements that are scheduled before C_S^I for each S. For a technical reason, those elements in S are separately counted, which results in the final term.

Lemma 7 *For any $\rho \leq 1$, $r > 1$ and $Q \geq 1$, we have*

$$\mathbb{E}\left[\sum_{S \in \mathcal{S}} C_S^R\right] \leq Q\left(1 + \frac{\rho}{r-1}\right)\Big(1 + (r-1)\sum_{i=0}^{\infty} r^i p_i\Big) \sum_{S \in \mathcal{S}} C_S^P + \sum_{S \in \mathcal{S}} \kappa(S).$$

Proof Recall that C_S^I is the earliest time t such that $|\{e \in S \ : \ t_{e,\alpha_e} \leq t\}| \geq \kappa(S)$. We first upper bound $\mathbb{E}[C_S^I]$. By definition of p_i, we have $\Pr[C_S^I > r^{i-1} \cdot C_S^P] \leq p_{i-1}$ for any $i \geq 1$. Thus it follows that

$$\begin{aligned}
\mathbb{E}[C_S^I] &\leq \Big((1 - p_0) + \sum_{i=1}^{\infty} r^i (p_{i-1} - p_i)\Big) C_S^P \\
&= \Big(1 + (r-1)\sum_{i=0}^{\infty} r^i p_i\Big) C_S^P.
\end{aligned}$$

As mentioned before, C_S^I is not the completion time C_S^R of S in the final schedule σ^R. This is because we might have scheduled more than C_S^I elements until time C_S^I. By counting the number of elements appearing no later than time C_S^I, we obtain the following relation between C_S^R and C_S^I. Such elements are counted separately depending on whether they are in S or not. We obtain

$$C_S^R \le |\{e \in [n] \setminus S \mid t_{e,\alpha_e} \le C_S^I\}| + \kappa(S)$$

Since for any fixed $e \notin S$ and time $t \in \mathbb{R}_+$, $\Pr[t_{e,\alpha_e} \le t] = \min\{1, \int_{\tau=0}^{t} \tilde{x}_e(\tau)\mathtt{d}\tau\}$, by Lemma 5, we have

$$\mathbb{E}[|\{e \in [n] \setminus S \ : \ t_{e,\alpha_e} \le t\}|] \le \sum_{e\in[n]} \int_{\tau=0}^{t} \tilde{x}_e(\tau)\mathtt{d}\tau \le Q\left(1+\frac{\rho}{r-1}\right)t. \quad (14)$$

Therefore we have

$$\begin{aligned}
&\mathbb{E}[|\{e \in [n] \setminus S \mid t_{e,\alpha_e} \le C_S^I\}|] \\
&= \int_{\tau=0}^{\infty} \left(\mathbb{E}[|\{e \in [n] \setminus S \ : \ t_{e,\alpha_e} \le \tau\}|] \cdot \Pr[C_S^I = \tau]\right) \mathtt{d}\tau \\
&\le \int_{\tau=0}^{\infty} \left(Q\left(1+\frac{\rho}{r-1}\right)\tau \cdot \Pr[C_S^I = \tau]\right) \mathtt{d}\tau \\
&\le Q\left(1+\frac{\rho}{r-1}\right) \int_{\tau=0}^{\infty} \left(\tau \cdot \Pr[C_S^I = \tau]\right) \mathtt{d}\tau \\
&= Q\left(1+\frac{\rho}{r-1}\right) \mathbb{E}[C_S^I].
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
\mathbb{E}[C_S^R] &\le Q\left(1+\frac{\rho}{r-1}\right)\mathbb{E}[C_S^I] + \kappa(S) \\
&\le Q\left(1+\frac{\rho}{r-1}\right)\left(1+(r-1)\sum_{i=1}^{\infty} r^i p_i\right) C_S^P + \kappa(S).
\end{aligned}$$

Summing over all sets S completes the proof. □

We are now ready to complete the proof of Theorem 3.

Proof (Proof of Theorem 3) Observe that $\sum_{S\in\mathcal{S}} \kappa(S)$ is a lower bound on the cost of any preemptive schedule. From Lemma 7, it suffices to show

$$Q\left(1+\frac{\rho}{r-1}\right)\left(1+(r-1)\sum_{i=0}^{\infty} r^i p_i\right) \le 5.2$$

We set $Q = 2.65$, $r = 1.40$ and $\rho = 0.22$. By applying the upper bound in Lemma 6, we obtain $Q(1+\frac{\rho}{r-1})\left(1+(r-1)\sum_{i=0}^{30} r^i p_i\right) \le 5.13$ (the computation was done numerically). For $i > 30$, we now prove that

$$\begin{aligned}
r^i p_i &\le r^i \max(K_{1i}, K_{2i}, K_{3i}) \\
&\le 2\exp(-Q(1+\rho i))r^i \\
&< 1.145 \cdot (e^{-Q\rho} r)^i \\
&< 1.145 \cdot (0.782)^i.
\end{aligned}$$

From Lemma 6, we derive

1. $K_{1i} =\leq 2\exp(-Q(1+\rho i))$;
2. Using the fact that $2e\frac{z}{\exp(z)} \leq 1$ for $z \geq 3$ and $Q(1+\rho i) \geq 2.65(1+30 \cdot 0.22) > 10$, it follows that

$$\begin{aligned}
K_{2i} &= \exp(-2Q(1+\rho i)) + 2Q(1+\rho i)\exp(-2Q(1+\rho i)+1) \\
&= \exp(-2Q(1+\rho i)) + 2e\frac{Q(1+\rho i)}{\exp(Q(1+\rho i))}\exp(-Q(1+\rho i)) \\
&\leq 2\exp(-Q(1+\rho i)).
\end{aligned}$$

3. Finally, from the fact that $Q(1+\rho i) > 10$, we know

$$\begin{aligned}
K_{3i} &= \exp\left(-1.5Q\left(\left(1-\frac{1}{(1+\rho i)Q}\right)^2(1+\rho i)\right)\right) \\
&\leq \exp(-1.5Q(1+\rho i)(1-1/10)^2) \\
&\leq \exp(-1.2Q(1+\rho i)) \\
&\leq 2\exp(-Q(1+\rho i)).
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
&Q(1+\frac{\rho}{r-1})\left((r-1)\sum_{i=31}^{\infty} r^i p_i\right) \\
&< Q(1+\frac{\rho}{r-1})\Big(1.145(r-1)\sum_{i=31}^{\infty}(0.782)^i\Big) \\
&= Q(1+\frac{\rho}{r-1})\Big(1.145(r-1)(0.782)^{31}\sum_{i=0}^{\infty}(0.782)^i\Big) \\
&< Q(1+\frac{\rho}{r-1})\Big(1.145(r-1)(0.782)^{31}\cdot\frac{1}{1-0.782}\Big) < 0.02
\end{aligned}$$

We also need to add one due to the last term in the Lemma 7. This establishes that the gap between between preemptive and non-preemptive schedules for the min sum objective is at most 6.2. □

4 Comparison of our LP and the previous one in [2, 16]

In this section, we demonstrate that our configuration LP ($\mathsf{LP}_{\mathsf{Primal}}$) is *stronger* for both the non-preemptive and preemptive problems than the LP ($\mathsf{LP}_{\mathsf{BGK}}$) considered in [2,16], which is based on the knapsack cover inequalities. For completeness, we present $\mathsf{LP}_{\mathsf{BGK}}$ as follows.

$$
\begin{aligned}
\min \quad & \sum_{t\in[n]}\sum_{S\in\mathcal{S}}(1-z_{S,t}) && && (\mathsf{LP_{BGK}})\\
s.t. \quad & \sum_{e\in[n]} x_{e,t} = 1 && \forall t\in[n] && (15)\\
& \sum_{t\in[n]} x_{e,t} = 1 && \forall e\in[n] && (16)\\
\sum_{e\in S\setminus A}\sum_{t'<t} & x_{e,t'} \geq (\kappa(S)-|A|)\cdot z_{S,t} && \forall S\in\mathcal{S},\forall A\subseteq S,\forall t\in[n] && (17)\\
& x_{e,t}\geq 0 && \forall e\in[n],\forall t\in[n] &&\\
& z_{S,t}\in[0,1] && \forall S\in\mathcal{S},\forall t\in[n] &&
\end{aligned}
$$

Note that the constraints (1) and (2) in $\mathsf{LP_{Primal}}$ are exactly as the same as (15) and (16) in $\mathsf{LP_{BGK}}$. In $\mathsf{LP_{BGK}}$, the objective is based on the variables $z_{S,t}$, the extent to which S is covered until time $t-1$. Each set S contributes to the objective by $(1-z_{S,t})$ at each time t. The variables $z_{S,t}$ are defined by the knapsack cover inequalities. We show that $\mathsf{LP_{BGK}}$ is also a valid relaxation for the preemptive problem as well as for the non-preemptive one, as is $\mathsf{LP_{Primal}}$.

Proposition 4 $\mathsf{LP_{BGK}}$ *is a valid linear programming relaxation for Preemptive Generalized Min Sum Set Cover.*

Proof Consider any preemptive schedule $x_e(t)$, $e\in[n]$, $t\in[0,n]$. Define the value of $x_{e,t} := \int_{\tau=t-1}^{t} x_e(\tau)\ \mathsf{d}\tau$ from $x_e(t)$. Recall in Theorem 2 that we have shown that $x_{e,t}$ satisfy constraints (15) and (16). For each set $S\in\mathcal{S}$ and $t\in[n]$, let $z_{S,t} = \min_{A\subseteq S, |A|<\kappa(S)} \frac{\sum_{e\in S\setminus A}\sum_{t'<t} x_{e,t'}}{\kappa(S)-|A|}$. Then, constraints (17) are clearly satisfied.

We now shift our attention to proving that for the $x_{e,t}, z_{S,t}$ values we defined above, $\mathsf{LP_{BGK}}$ has an objective smaller than the total preemptive cover time under the continuous schedule $x_e(t)$. To this end, it suffices to show that for each set $S\in\mathcal{S}$, $\sum_{t\in[n]}(1-z_{S,t})\leq C_S^P$, where C_S^P be the (preemptive) cover time of S. Consider any set $S\in\mathcal{S}$. First observe that for any time $t\geq\lfloor C_S^P\rfloor+2$, $z_{S,t}=0$, since at least $\kappa(S)$ elements are completely scheduled from S by time $\lfloor C_S^P\rfloor+1$. Now we consider the time $\lfloor C_S^P\rfloor+1$. Let $S'\subseteq S$ be a subset of $\kappa(S)$ elements that are completely scheduled by time C_S^P. Hence it must be the case that for each element $e\in S'$, $\int_{\tau=0}^{\lfloor C_S^P\rfloor} x_e(\tau)\ \mathsf{d}\tau \geq 1-(C_S^P-\lfloor C_S^P\rfloor)$. As such, $z_{S,\lfloor C_S^P\rfloor}\geq 1-(C_S^P-\lfloor C_S^P\rfloor)$. Finally, for each time $t\leq\lfloor C_S^P\rfloor$, S adds at most 1 to the $\mathsf{LP_{Primal}}$ objective. In sum, we obtain $\sum_{t\in[n]}(1-z_{S,t})\leq \lfloor C_S^P\rfloor + (1-z_{S,\lfloor C_S^P\rfloor})\leq C_S^P$. □

We now focus on showing that our relaxation $\mathsf{LP_{Primal}}$ gives a stronger lower bound than $\mathsf{LP_{BGK}}$. We first provide an instance for which $\mathsf{LP_{Primal}}$ has an objective value strictly larger than $\mathsf{LP_{BGK}}$.

Proposition 5 *For any $\epsilon > 0$, there exists an instance for which $\mathsf{LP}_{\mathsf{Primal}}$ has an objective larger than the objective of $\mathsf{LP}_{\mathsf{BGK}}$ by a factor of more than $2 - \epsilon$.*

Proof Consider the instance where there exists a single set $S = [n]$ with $\kappa(S) = n$. We first consider the objective of $\mathsf{LP}_{\mathsf{Primal}}$. Since all configurations $F \in \mathcal{F}(S)$ have $C_S^F \geq n$, therefore $\mathsf{LP}_{\mathsf{Primal}}$ gives a solution with cost no smaller than n. Now we turn our attention to $\mathsf{LP}_{\mathsf{BGK}}$. Consider the schedule where all elements are equally scheduled in each time slot, i.e. $x_{e,t} = 1/n$ for all $e, t \in [n]$. Since $z_{S,t} = \frac{t-1}{n}$ satisfies the constraints, $\mathsf{LP}_{\mathsf{BGK}}$ will have an objective value of $\sum_{t=1}^{n}(1 - \frac{t-1}{n}) = \frac{n+1}{2}$. The claim immediately follows. □

We now show the following lemma. Since both linear programs, $\mathsf{LP}_{\mathsf{Primal}}$ and $\mathsf{LP}_{\mathsf{BGK}}$ are valid relaxations, it, together with the above proposition, will establish that $\mathsf{LP}_{\mathsf{Primal}}$ is stronger than $\mathsf{LP}_{\mathsf{BGK}}$.

Lemma 8 *For any instance, $\mathsf{LP}_{\mathsf{Primal}}$ has an objective no smaller than the objective of $\mathsf{LP}_{\mathsf{BGK}}$.*

Proof Consider an arbitrary instance and let x^*, y^* be a fixed optimal solution to $\mathsf{LP}_{\mathsf{Primal}}$. Define z as follows:

$$z_{S,t} =_{def} \sum_{F \in \mathcal{F}(S) \text{ and } C_S^F < t} y_S^{*F}.$$

We first show that x^*, z satisfy all constraints in $\mathsf{LP}_{\mathsf{BGK}}$. We focus on showing constraints (17), since all the other constraints are trivially satisfied. Consider any S and $A \subseteq S$. Then we have

$$\begin{aligned}
\sum_{e \in S \setminus A} \sum_{t' < t} x^*_{e,t'} &= \sum_{e \in S \setminus A} \sum_{t' < t} \sum_{F \in \mathcal{F}(S), (e,t') \in F} y_S^{*F} \qquad \text{[From constraints (4)]} \\
&= \sum_{F \in \mathcal{F}(F)} y_S^{*F} \sum_{e \in S \setminus A} \sum_{t' < t} \mathbf{1}[(e,t') \in F] \\
&\geq \sum_{F \in \mathcal{F}(F) \text{ and } C_S^F < t} y_S^{*F} \sum_{e \in S \setminus A} \sum_{t' < t} \mathbf{1}[(e,t') \in F] \\
&\geq \sum_{F \in \mathcal{F}(F) \text{ and } C_S^F < t} y_S^{*F} (\kappa(S) - |A|) \\
&= (\kappa(S) - |A|) \cdot z_{S,t}
\end{aligned}$$

The last inequality holds because for any $t \in [n]$, $S \in \mathcal{S}$, $F \in \mathcal{F}(S)$ with $C_S^F < t$, at least $\kappa(S)$ elements in S are scheduled before time t under the configuration F.

Let $\mathsf{LP}_{\mathsf{Primal}}(x^*, y^*)$ denote the objective of $\mathsf{LP}_{\mathsf{Primal}}$ by the solution x^*, y^*. Likewise $\mathsf{LP}_{\mathsf{BGK}}(x^*, z)$ denotes the objective of $\mathsf{LP}_{\mathsf{BGK}}$ by the solution x^*, z. We now show that

$$\mathsf{LP}_{\mathsf{Primal}}(x^*, y^*) = \mathsf{LP}_{\mathsf{BGK}}(x^*, z)$$

Since x^*, y^* are an optimal solution to $\mathsf{LP}_{\mathsf{Primal}}$, and x^*, z are a feasible solution to $\mathsf{LP}_{\mathsf{BGK}}$, the claim will follow. Consider any S and $F \in \mathcal{F}(S)$. By viewing C_S^F as adding one to the cost of S at each time $t \le C_S^F$, we have

$$\begin{aligned}
\mathsf{LP}_{\mathsf{Primal}}(x^*, y^*) &= \sum_{S \in \mathcal{S}} \sum_{F \in \mathcal{F}(S)} C_S^F y_S^{*F} \\
&= \sum_{S \in \mathcal{S}} \sum_{1 \le t \le n} \sum_{F \in \mathcal{F}(S): t \le C_S^F} y_S^{*F} \\
&= \sum_{1 \le t \le n} \sum_{S \in \mathcal{S}} (1 - \sum_{F \in \mathcal{F}(S), t > C_S^F} y_S^{*F}) \quad \text{[From constraints (3)]} \\
&= \sum_{1 \le t \le n} \sum_{S \in \mathcal{S}} (1 - z_{S,t}) \\
&= \mathsf{LP}_{\mathsf{BGK}}(x^*, z)
\end{aligned}$$

□

References

1. Yossi Azar, Iftah Gamzu, and Xiaoxin Yin. Multiple intents re-ranking. In *ACM Symposium on Theory of Computing*, pages 669–678, 2009.
2. Nikhil Bansal, Anupam Gupta, and Ravishankar Krishnaswamy. A constant factor approximation algorithm for generalized min-sum set cover. In *ACM-SIAM symposium on Discrete algorithms*, pages 1539–1545, 2010.
3. Nikhil Bansal and Subhash Khot. Optimal long code test with one free bit. In *Symposium on Foundations of Computer Science*, pages 453–462, 2009.
4. Amotz Bar-Noy, Mihir Bellare, Magnús M. Halldórsson, Hadas Shachnai, and Tami Tamir. On chromatic sums and distributed resource allocation. *Inf. Comput.*, 140(2):183–202, 1998.
5. Amotz Bar-Noy, Magnús M. Halldórsson, and Guy Kortsarz. A matched approximation bound for the sum of a greedy coloring. *Inf. Process. Lett.*, 71(3):135–140, 1999.
6. S. Burer and R. Monteiro. A projected gradient algorithm for solving the maxcut sdp relaxation. *Optimization Methods and Software*, 15:175–200, 2001.
7. Chandra Chekuri and Rajeev Motwani. Precedence constrained scheduling to minimize sum of weighted completion times on a single machine. *Discrete Applied Mathematics*, 98(1-2):29–38, 1999.
8. Uriel Feige, László Lovász, and Prasad Tetali. Approximating min sum set cover. *Algorithmica*, 40(4):219–234, 2004.
9. Martin Grotschel, László Laszlo Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization. Second edition. Algorithms and Combinatorics, 2.* Springer-Verlag, Berlin, 1993.
10. Leslie A. Hall, Andreas S. Schulz, David B. Shmoys, and Joel Wein. Scheduling to minimize average completion time: Off-line and on-line approximation algorithms. *Mathematics of Operations Research*, 22(3):513–544, 1997.
11. François Margot, Maurice Queyranne, and Yaoguang Wang. Decompositions, network flows, and a precedence constrained single-machine scheduling problem. *Operations Research*, 51(6):981–992, 2003.
12. Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.
13. Maurice Queyranne and Maxim Sviridenko. A $(2 + \varepsilon)$-approximation algorithm for the generalized preemptive open shop problem with minsum objective. *J. Algorithms*, 45(2):202–212, 2002.

14. Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*. Springer-Verlag, Berlin, 2003.
15. Andreas Schulz and Martin Skutella. Random-based scheduling: new approximations and lp lower bounds. In *RANDOM*, pages 119–133, 1997.
16. Martin Skutella and David P. Williamson. A note on the generalized min-sum set cover problem. *Operations Research Letters, to appear*, 2011.
17. Gerhard J. Woeginger. On the approximability of average completion time scheduling under precedence constraints. *Discrete Applied Mathematics*, 131(1):237–252, 2003.