

On Minimizing Generalized Makespan on Unrelated Machines

Nikhil Ayyadevara ✉

University of Michigan, Ann Arbor, MI, USA

Nikhil Bansal ✉

University of Michigan, Ann Arbor, MI, USA

Milind Prabhu ✉

University of Michigan, Ann Arbor, MI, USA

Abstract

We consider the *Generalized Makespan Problem* (GMP) on unrelated machines, where we are given n jobs and m machines and each job j has arbitrary processing time p_{ij} on machine i . Additionally, there is a general symmetric monotone norm ψ_i for each machine i , that determines the load on machine i as a function of the sizes of jobs assigned to it. The goal is to assign the jobs to minimize the maximum machine load.

Recently, Deng, Li, and Rabani [8] gave a 3 approximation for GMP when the ψ_i are top- k norms, and they ask the question whether an $O(1)$ approximation exists for general norms ψ ? We answer this negatively and show that, under natural complexity assumptions, there is some fixed constant $\delta > 0$, such that GMP is $\Omega(\log^\delta n)$ hard to approximate. We also give an $\Omega(\log^{1/2} n)$ integrality gap for the natural configuration LP.

2012 ACM Subject Classification Theory of computation → Scheduling algorithms

Keywords and phrases Hardness of Approximation, Generalized Makespan

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.21

Category APPROX

Funding *Nikhil Bansal*: Supported in part by the NWO VICI grant 639.023.812.

1 Introduction

We consider a question by Deng, Li and Rabani [8] about scheduling jobs to minimize makespan on unrelated machines in the setting of more general norms. Recall that in the unrelated machines setting, we are given n jobs and m machines where each job $j \in [n]$ has some arbitrary processing time p_{ij} on machine $i \in [m]$. Given an assignment of jobs to machines, the load on a machine is the sum of the processing times of all the jobs assigned to it. In a seminal result, Lenstra, Shmoys and Tardos [14] gave a 2-approximation for minimizing the makespan (the maximum machine load). These results were later extended to the problem of minimizing the ℓ_p -norm of the machine loads [1, 4, 13, 16].

In a breakthrough work [5], Chakrabarty and Swamy introduced a substantial generalization of the problem to general symmetric monotone norms. Recall that a function $\phi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ is a norm if it satisfies (i) $\phi(u) = 0$ iff $u = 0^n$, (ii) $\phi(\alpha u) = |\alpha| \phi(u)$ for all u and $\alpha \in \mathbb{R}$ and, (iii) $\phi(u + v) \leq \phi(u) + \phi(v)$. We say that ϕ is *symmetric* if $\phi(u) = \phi(u')$ where u' is some permutation of u , and it is *monotone* if $\phi(u) \leq \phi(v)$ for all u, v satisfying $0 \leq u(i) \leq v(i)$ for all $i \in [m]$.

Surprisingly, they showed that for any arbitrary symmetric monotone norm $\phi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$, that determines the overall objective as a function of the individual machine loads, there is an $O(1)$ approximation. The approximation ratio was subsequently improved to $4 + \epsilon$ [6]



© Nikhil Ayyadevara, Nikhil Bansal, and Milind Prabhu;
licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023).

Editors: Nicole Megow and Adam D. Smith; Article No. 21; pp. 21:1–21:13



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

and more recently to $2 + \epsilon$ [11], which remarkably almost matches the bound for makespan. These results introduce several new ideas to handle general norms by relating them to the special class of top- k norms¹, and algorithmic techniques to work with them.

General inner and outer norms. An even further generalization was considered by Deng, Li and Rabani [8], which we refer to as the generalized load-balancing (GLB) problem. Here, additionally, each machine i also has an arbitrary symmetric monotone norm $\psi_i : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ referred to as the *inner-norm*, that determines the machine’s load as a function of the sizes of the jobs assigned to it (note that all the results described previously have inner norm ℓ_1). Formally, given an assignment $\rho : [n] \rightarrow [m]$ of jobs to machines, the load on machine i is $\text{load}(i) = \psi_i(p_i^\rho)$ where p_i^ρ is the vector of sizes of jobs assigned to i , i.e., $p_i^\rho(j) = p_{ij}$ if $\rho(j) = i$ and 0 otherwise. The overall objective is $\phi(\text{load})$, where $\text{load} = (\text{load}(1), \dots, \text{load}(m))$ is the vector of machine loads (where $\phi : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$ is the norm as in [5]). We shall refer to ϕ as the *outer-norm*. Throughout this paper a general norm always means a symmetric monotone norm.

In its full generality (when the ψ_i and ϕ are general), GLB becomes $\Omega(\log n)$ hard to approximate, as it generalizes² the Set Cover problem when $\psi = \ell_\infty$ and $\phi = \ell_1$. Interestingly, [8] gave a matching $O(\log n)$ approximation for general ψ and ϕ , based on solving and rounding a novel configuration LP.

Generalized Makespan Problem (GMP). Given the $\Omega(\log n)$ hardness for the general case, [8] also consider the interesting and natural special case where the outer-norm ℓ_∞ , but the inner norm is general (i.e., the goal is to minimize makespan but where the machine loads are given by general inner-norms ψ_i). We refer to this problem as the *Generalized Makespan Problem (GMP)*. In a sense, this problem can be considered as a “dual” of the problem considered by Chakrabarty and Swamy (where the inner-norm is ℓ_1 , but the outer-norm is general).

For GMP, Deng, Li and Rabani gave a 3-approximation for the special case when each ψ_i is a top- k norm. The main open question they ask is whether an $O(1)$ approximation is achievable for general inner-norms. A priori this seems quite plausible as there is close connection between top- k norms and general norms (see e.g. [5, 6]). Moreover, the $O(1)$ approximation of Chakrabarty and Swamy [5] for the “dual” problem, also suggests that GMP may have an $O(1)$ approximation.

Our Results

Our main result is that GMP does not admit an $O(1)$ -approximation under standard complexity-theoretic assumptions, answering the main open problem in [8].

Our starting point is an integrality gap instance for the natural configuration LP for GMP.

► **Theorem 1.** *There is an instance of GMP with a symmetric monotone norm ψ , for which the natural configuration LP has an integrality gap of $\Omega((\log n)^{1/2})$.*

¹ The top- k norm of a non-negative vector v is the sum of its largest k entries.

² Given subsets S_1, \dots, S_m of $[n]$, consider the scheduling instance on m machines (one per set) and n jobs (one per element), with $p_{ij} = 1$ iff element $j \in S_i$ and $p_{ij} = \infty$ otherwise. That is, only jobs in S_i can be assigned to i . The point is that as $\psi = \ell_\infty$, we have $\text{load}(i) = 0$ if no job is assigned to i , and exactly 1 otherwise (even if all jobs in S_i are assigned to i). As $\phi = \ell_1$, the objective is exactly the number of machines (sets) needed to cover all the jobs.

The gap instance is based on a probabilistic construction and a key idea is to work with a suitably chosen norm ψ , defined as the sum of top- k norms at various different scales, that interacts nicely with properties of random set cover instances. This construction is described in Section 3, and it forms a key gadget in our following hardness result.

► **Theorem 2.** *There is a universal constant $\delta > 0$, such that any polynomial-time approximation algorithm for GMP has approximation ratio $\Omega(\log^\delta n)$ provided that $\text{NP} \not\subseteq \text{ZTIME}(n^{O(\log \log n)})$.*

Our construction for this hardness result builds on the ideas of Lund and Yannakakis [15], who showed the $\Omega(\log n)$ hardness of Set Cover by a gap reduction from the Label Cover problem. However, we require some additional ideas as reducing the Label Cover problem to a scheduling instance and exploiting the properties of the norm ψ requires much more care. Specifically, even though our gadget in Theorem 1 has the $\Omega((\log n)^{1/2})$ gap, embedding it into Label Cover imposes extra constraints on the number of labels in the Label Cover instance, and only leads to an $\Omega(\log^\delta n)$ hardness, for some constant $\delta > 0$. Interestingly, this constant δ depends on the soundness parameter of label cover as a function of the number of labels, based on Raz’s parallel repetition theorem [18] and its subsequent improvements [10, 17].

► **Remark.** The constant δ can be computed explicitly, but we do not attempt to do this here. This construction is described in Section 4. We remark that there are extensive work on improving the soundness in PCP constructions as a function of the number of labels. The best known result in this direction is due to Chan [7] that achieves soundness roughly $L^{-1/2}$ for L labels. However, Chan’s result does not have perfect completeness and hence cannot be used in our constructions. Roughly speaking, this is because each job must be assigned to some machine (this is similar to the reason that one requires perfect completeness in reducing label cover to set cover, as each element must be covered).

► **Remark.** In personal communication, Amey Bhangale has pointed out that in an unpublished manuscript, they can show that assuming the 2-to-2 conjecture with perfect completeness, there is a label cover instance with L labels that has perfect completeness and soundness $L^{-1/2}$. Using a such a label cover, our construction in Section 4 would imply a hardness of $\Omega(\log^{1/8} n)$ in Theorem 2. We note that currently, proving the 2-to-2 conjecture with perfect completeness remains open, and in particular the breakthrough results of Khot, Minzer and Safra [12] on the 2-to-2 conjecture assume imperfect completeness.

2 Preliminaries

2.1 (n, m, ℓ, β) Set-System

The constructions of the integrality gap instance in Section 3 and the reduction from label cover to GMP presented in Section 4 both use the following set system as a building block.

► **Definition 3** ((n, m, ℓ, β) Set-system). *Let n, m, ℓ be positive integers, $\beta \in (0, 1)$, U be a set with $|U| = n$, and A_1, \dots, A_m be subsets of U . The sets $(U; A_1, \dots, A_m)$ form an (n, m, ℓ, β) set-system if for every set I of at most ℓ indices from $[m]$, $|\cup_{i \in I} B_i| \leq (1 - \beta)|U|$, where B_i is either A_i or $\overline{A_i}$.*

Intuitively, an (n, m, ℓ, β) set-system has the property that any set cover which uses at most ℓ subsets necessarily uses a complementary pair of subsets A_i and $\overline{A_i}$. Moreover, any collection of at most ℓ subsets that do not contain any complementary pair can cover at most

21:4 On Minimizing Generalized Makespan on Unrelated Machines

a $(1 - \beta)$ fraction of the elements in U . The following lemma shows that for a particular choice of parameters n, m, ℓ , and β , there is a simple and efficient randomized construction of an (n, m, ℓ, β) set-system.

► **Lemma 4.** *For a sufficiently large positive integer n and a positive integer $m \in [\sqrt{\log n}, 2\sqrt{\log n}]$ there exists an (n, m, ℓ, β) set-system $(U; A_1, \dots, A_m)$ with $\ell = m/10$ and $\beta = \exp(-m) = \exp(-O(\sqrt{\log n}))$. There is a polynomial-time algorithm that constructs such a set system with high probability.*

Proof. Let U be a set of n elements, and initialize m empty sets A_1, \dots, A_m . For each element $e \in U$, sample a random index set $J \subset [m]$ of size exactly $m/2$ and add e to sets A_j for $j \in J$.

We show that this construction gives an (n, m, ℓ, β) set-system with high probability. Consider an index set $I \subset [m]$ with $|I| = \ell$ and a collection of sets B_i for $i \in I$ such that each B_i is either A_i or \bar{A}_i . For a fixed $e \in U$, let p denote the probability that e is not contained in $\cup_{i \in I} B_i$, i.e., $p = \Pr[e \notin \cup_{i \in I} B_i]$. We have,

$$p \geq \binom{m-\ell}{m/2} / \binom{m}{m/2} \geq \left(\frac{m-\ell}{m/2}\right)^{m/2} / \left(\frac{em}{m/2}\right)^{m/2} \geq \left(\frac{m-\ell}{em}\right)^{m/2} \geq \exp(-0.6m),$$

where the second inequality uses that $\binom{n}{k}^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.

The probability that $\cup_{i \in I} B_i$ contains a fixed subset of cardinality greater than or equal to $(1 - \beta)n$ is at most $(1 - p)^{(1-\beta)n}$. By a union bound over at most $\binom{n}{\beta n} n$ possible subsets with cardinality at least $(1 - \beta)n$,

$$\begin{aligned} \Pr[|\cup_{i \in I} B_i| \geq (1 - \beta)n] &\leq \binom{n}{\beta n} n \cdot (1 - p)^{(1-\beta)n} \leq (e/\beta)^{\beta n} n \cdot e^{-(1-\beta)np} \\ &\leq \exp(3n\beta \log(1/\beta) - np/2) \leq \exp(-np/4) \leq \exp(-n^{0.9}), \end{aligned}$$

where in the second to last inequality we use that $3\beta \log(1/\beta) < p/4$.

A union bound over the at most $2^\ell \cdot \binom{m}{\ell} \leq \exp(\sqrt{\log n})$ possible choices to pick the ℓ sets B_i , gives that with high probability, the union of any ℓ sets B_i has cardinality less than $(1 - \beta)n$. ◀

2.2 Label Cover

In Section 4, we prove the hardness of approximation of GMP via a reduction from the standard label cover problem as defined below.

► **Definition 5.** *A label cover instance \mathcal{L} is defined by a tuple $((U, V, E), L, \Pi)$. Here (U, V, E) is a bipartite graph with vertices $U \cup V$ and edges $E \subseteq U \times V$; L is a positive integer and Π is a set of functions one for each edge $e \in E$ i.e., $\Pi = \{\pi_e : [L] \rightarrow [L] \mid e \in E\}$. A labeling of the vertices $\sigma : U \cup V \rightarrow [L]$ is said to satisfy an edge $e = (u, v)$ if $\pi_e(\sigma(u)) = \sigma(v)$. Given \mathcal{L} , the goal of the label cover problem is to find a labeling σ^* that satisfies the maximum number of edges in E . We use $\text{OPT}(\mathcal{L})$ to denote the fraction of the edges in E satisfied by σ^* .*

As we will need the explicit dependence between the number of labels and the soundness, for completeness we sketch below the precise gap version of the label cover problem that we will use.

► **Lemma 6 (Hardness of Gap Label Cover).** *Given a label cover instance $\mathcal{L} = ((U, V, E), L, \Pi)$ satisfying:*

- (i) $|U| = |V| = N$
- (ii) The degree of every vertex in $U \cup V$ is $d = O((\log N)^{c_1})$ for some constant c_1 .
- (iii) $L = \sqrt{\log N}$

There is some constant $c > 0$, for which there is no polynomial-time algorithm to decide if $OPT(\mathcal{L}) = 1$ or $OPT(\mathcal{L}) \leq (\log N)^{-c}$ provided that $\text{NP} \not\subseteq \text{DTIME}(N^{O(\log \log N)})$.

Proof. Using a standard argument (see for ex., [9, 2]) one can obtain a reduction from a 3SAT-5 instance ϕ with t variables to a Label Cover instance $\mathcal{L}_1 = ((U_1, V_1, E_1), 8, \Pi)$, where $|U_1| = |V_1| = O(t)$ and the graph (U_1, V_1, E) is 15-regular. The instance \mathcal{L}_1 has the following property: if ϕ has a satisfying assignment then $OPT(\mathcal{L}_1) = 1$; else if any assignment satisfies at most $(1 - \epsilon)$ fraction of clauses in ϕ , then $OPT(\mathcal{L}_1) \leq (1 - \Theta(\epsilon))$. By the PCP-theorem [3] it follows that, for some constant $\epsilon_0 > 0$, deciding if $OPT(\mathcal{L}_1) = 1$ or $OPT(\mathcal{L}_1) \leq 1 - \epsilon_0$ is NP-hard.

The following well-known construction [2] gives stronger inapproximability results for label cover. We define the k th power of the label cover instance $\mathcal{L}_k = ((U_k, V_k, E_k), 8^k, \Pi^k)$, where U_k, V_k are k -tuples of vertices in U_1, V_1 respectively, E_k is the set of all k -tuples of edges in E_1 . The resulting graph has $N = t^{O(k)}$ vertices and is $(15)^k$ -regular. The new set of labels³ consist of k -tuples of $\{1, \dots, 8\}$. For an edge $e = (e_1, \dots, e_k) \in E_k$, we define the function $\pi_e^k(a_1, \dots, a_k) = (\pi_{e_1}(a_1), \dots, \pi_{e_k}(a_k))$. Raz's Parallel Repetition Theorem [18], shows that for the label cover instance constructed above, there exists a constant α such that $OPT(\mathcal{L}_k) \leq (OPT(\mathcal{L}_1))^{\alpha k}$.

We now pick k so that $L = \sqrt{\log N}$. Since $L = 8^k$ and $N = t^{O(k)}$, this gives $k = \Theta(\log \log t)$. This choice of k ensures that $d = (15)^k = (\log N)^{c_1}$ for some constant c_1 . Moreover, if $OPT(\mathcal{L}_1) \leq (1 - \epsilon_0)$, then $OPT(\mathcal{L}_k) \leq (1 - \epsilon_0)^{\alpha k} \leq (\log t)^{-c'} \leq (\log N)^{-c}$ for some positive constants c, c' . ◀

3 Integrality Gap for Configuration LP

We begin by describing the configuration LP for GMP. We then explain the high-level ideas behind the gap construction and the properties we need from the norm ψ . We then describe the norm ψ and the integrality gap instance formally and then prove Theorem 1. As mentioned earlier, this gap instance and the norm ψ form the key gadget in our hardness construction in Section 4.1, and understanding it is crucial to the results in Section 4.

Configuration LP. The most natural LP relaxation of GMP is to consider assignment variables $x_{ij} \in [0, 1]$ that determine the fraction of job j assigned to machine i , and impose natural constraints. However, simple examples show that such an LP is too weak to handle general norms ψ . A stronger relaxation is the configuration LP, where we have exponentially many variables $x_{i,C}$, one for each machine i , and a possible subset of jobs C that can be feasibly assigned to machine i .

Let J denote the set of jobs, and $M = [m]$ be the set of machines. For a machine $i \in M$, and subset $C \subseteq J$, let $p_i[C] = (p_{ij} \cdot \mathbb{1}[j \in C])_{j \in J}$ denote the vector of sizes of jobs in C . Let T be a guess on the optimum makespan (we can do a binary search on T). Call a configuration C valid for machine i if the load $\psi_i(p_i[C])$ of C on i is at most T . We will slightly abuse notation and denote $\psi_i(p_i[C])$ by $\psi_i(C)$. Consider the following feasibility LP.

³ The labels are essentially numbers from 1 to 8^k .

21:6 On Minimizing Generalized Makespan on Unrelated Machines

$$\sum_{C \subseteq J} x_{i,C} \leq 1 \quad \forall i \in M \quad (1a)$$

$$\sum_{i \in M, C: j \in C} x_{i,C} = 1 \quad \forall j \in J \quad (1b)$$

$$x_{i,C} = 0 \quad \text{if } \psi_i(C) > T \quad (1c)$$

The first constraint says that each machine has at most one configuration. The second constraint says that each job is assigned to a machine, and the last constraint ensures that only valid configurations are considered and hence the generalized makespan on any machine is at most T . This LP can be solved efficiently for any desired accuracy $\epsilon > 0$ (so the configurations satisfy $\psi_i(C) \leq (1 + \epsilon)T$), see e.g., [8].

The main idea. We start with a simple instructive example, that motivates our choice of the norm and the choice of parameters that lead to the $\Omega(\sqrt{\log n})$ integrality gap. Consider a random set system on a universe U of n elements and m sets A_1, \dots, A_m where each element independently lies in $m/2$ randomly chosen sets. We will set $m \ll \log n$. Create an unrelated machine (in fact restricted assignment) scheduling instance I with m machines, where only the jobs in A_i can be assigned to machine i (all other jobs have infinite size on i). As each element lies in $m/2$ sets, the LP solution that picks the configuration A_i on each machine i with $x_{i,A_i} = 2/m$ is feasible, and uses only 2 machines in total.

However, in any integral solution, we claim that several machines must pick a non-trivial fraction of jobs from their sets A_i . Roughly, this is because the union of any $\ell \ll m$ sets A_i still leaves about $2^{-\ell}|U|$ uncovered (as each element lies in a set with probability $1/2$). Hence in any feasible integral solution, for any $\ell > 0$, at least ℓ machines must be assigned at least $2^{-\ell}|U|/m$ jobs.

To exploit this, suppose we define ψ as the top- k norm with $k \approx \Omega(2^{-\ell}n/m)$, so that any machine with $\geq k$ jobs incurs the same load, say T . Then in any integral solution, at least ℓ machines have load T , while fractionally at most 2 machines (in total) have load T . By creating $m/2$ disjoint copies $I^{(1)}, \dots, I^{(m/2)}$ of these set-cover instances, one would then expect that fractionally each machine has load T , while integrally the average load becomes $\Omega(\ell)$.

Unfortunately, this does not quite work as stated above, because once there are several instances $I^{(1)}, \dots, I^{(m/2)}$, as these jobs share the same machines, an algorithm can find a low makespan even if it cannot figure out the good underlying set cover solution. In fact, this is provably so, as [8] gave a 3-approximation when ψ is a Top_k norm.

However, our key observation is that this idea can still be made to work by choosing the instances $I^{(1)}, \dots, I^{(m/2)}$ at different scales (of the number of jobs and processing times) and defining the norm ψ as a suitable mixture of the top- k norms at these different scales. Roughly, this norm ψ still behaves as a top- k norm at each individual scale, but when jobs from different scales are combined, it takes on a large value, which be used to create a large gap in the reduction above.

Implementing this idea requires separating every two scales by $\Omega(2^\ell)$. So the ℓ scales leads to instances with size about 2^{ℓ^2} leading to the choice of $\ell = \Theta(\sqrt{\log n})$ to produce the $\Omega(\ell)$ gap. We now give the details.

The Instance. The instance will have n jobs and $m = \sqrt{\log n}$ machines. We first create $h = m/8$ disjoint set-cover instances $I^{(1)}, \dots, I^{(h)}$, where $I^{(s)} = (U(s); A_1(s), A_2(s), \dots, A_m(s))$ forms an (n_s, m, ℓ, β) set-system with $\ell = m/10, \beta = \exp(-m)$ and $n_s = (\sqrt{n}/2) \cdot \exp(4ms) = (\sqrt{n}/2)\beta^{-4s}$. Note that n_s increases as β^{-4s} with s , and $n_1 \geq \sqrt{n}$ and $n_h = n/2$, and it is easily checked that parameters for each $I^{(s)}$ satisfy the condition in Lemma 4.

For each $s \in [h]$, the elements in $U(s)$ correspond to jobs with size $p^{(s)} = \beta^{s-1}$ (or infinite if the job cannot be assigned to a machine). Each job j in $U(s)$ has size $p_{ij} = p^{(s)}$ on machine i iff $j \in A_i(s)$ and $p_{ij} = \infty$ otherwise. We abuse the notation slightly to refer to the resulting scheduling instance also as $I^{(s)}$. As the instances $I^{(1)}, \dots, I^{(h)}$ are at different scales in terms of the number of jobs and processing times, we refer to jobs in $I^{(s)}$ as being in the s -th *size class*⁴.

We define the inner norm $\psi = \sum_{s \in [h]} \psi^{(s)}$, where each $\psi^{(s)}$ is a scaled top- k norm given by

$$\psi^{(s)}(\mathbf{v}) = \frac{\text{Top}_{\beta^2 n_s}(\mathbf{v})}{(\beta^2 n_s p^{(s)})}. \quad (2)$$

Each machine i will have the same inner norm $\psi_i = \psi$. Informally, ψ has the following key property: for any subset C of $A_i(s)$ with at least $\beta^2 n_s$ jobs, $\psi_i^{(s)}(C) = 1$ and $\psi_i^{(s')}(C) \approx 0$ for $s' \neq s$. This in particular implies that $\psi_i(A_i(s)) \approx 1$ for any class s . Moreover, if C is an arbitrary subset of jobs with at least a β^2 fraction of jobs from r distinct sets among $A_i(1), A_i(2), \dots, A_i(h)$, then $\psi_i(C) \approx r$ (roughly, each such size-class s affects a different term $\psi^{(s)}$ of the norm). This property motivates the following definition.

► **Definition 7** (Heavy Size Class). *Given an assignment of jobs $\rho : J \rightarrow M$, we say that size class $s \in [h]$ is heavy on machine $i \in M$, if at least $\beta^2 n_s$ jobs from $A_i(s)$ are assigned to machine i .*

We now formally state and prove the property of the norm described above.

► **Lemma 8.** *For any $s \in [h]$ and $i \in M$, $\psi_i(A_i(s)) = 1 + o(1)$. Furthermore, for an assignment $\rho : J \rightarrow M$, if C is the set of jobs assigned to machine i , then $\psi_i(C)$ is at least the number of heavy size classes $s \in [h]$ on machine i .*

Proof. We first show that $\psi_i(A_i(s)) = 1 + o(1)$ by computing the value of $\psi^{(s')}$ for different s' . By the definition of $\psi^{(s')}$,

$$\psi_i^{(s')}(A_i(s)) = \frac{\text{Top}_{\beta^2 n_{s'}}(p_i[A_i(s)])}{(\beta^2 n_{s'} p^{(s')})} = \frac{\min\{\beta^2 n_{s'}, |A_i(s)|\} \cdot p^{(s)}}{(\beta^2 n_{s'} p^{(s')})} = \left(\min \left\{ 1, \frac{|A_i(s)|}{\beta^2 n_{s'}} \right\} \right) \frac{p^{(s)}}{p^{(s')}}$$

For $s = s'$, this exactly equals 1 (as $\beta \ll 1$ and $|A_i(s)| \approx \frac{n_s}{2}$).

For $s' < s$, we have $\psi_i^{(s')}(A_i(s)) \leq \frac{p^{(s)}}{p^{(s')}} = \beta^{(s-s')} \leq \beta$.

Finally, for $s' > s$, we have

$$\psi_i^{(s')}(A_i(s)) \leq \frac{|A_i(s)|}{\beta^2 n_{s'}} \cdot \frac{p^{(s)}}{p^{(s')}} \leq \frac{n_s}{\beta^2 n_{s'}} \cdot \frac{p^{(s)}}{p^{(s')}} = \beta^{3(s'-s)-2} \leq \beta.$$

Let us now consider an arbitrary set of jobs C assigned to machine i . By the monotonicity of the norm,

$$\psi_i^{(s)}(C) \geq \psi_i^{(s)}(C \cap A_i(s)) = \min \left\{ 1, \frac{|C \cap A_i(s)|}{\beta^2 n_s} \right\} \geq \mathbb{1}[s \text{ is heavy on } i].$$

As $\psi_i(C) = \sum_{s \in [h]} \psi_i^{(s)}(C)$, it follows that $\psi_i(C)$ is at least the number of heavy size classes on i . ◀

⁴ The total number of jobs in the h instances is $\sum_{i=1}^h n_s \approx n/2$. To make the total number of jobs n , we add dummy jobs that have processing time 0 on all machines.

Fractional Solution. We claim that the following solution is feasible for the configuration LP Equation (1) with $T = 2$: for each machine $i \in M$, set $x_{i,A_i(s)} = 2/m$ for each $s \in [h]$.

Clearly, Equation (1a) is satisfied for each $i \in M$ as there are $h < m/2$ sets of jobs $A_i(1), A_i(2), \dots, A_i(h)$, each with value $2/m$. Equation (1b) is satisfied as each job $j \in U(s)$, for each $s \in [h]$, lies in $m/2$ sets in $A_1(s), A_2(s), \dots, A_m(s)$. Equation (1c) is also satisfied as $\psi_i(A_i(s)) = 1 + o(1) < 2 = T$ by Lemma 8.

Integral Solution. We now show that any integral solution has generalized makespan $\Omega(\sqrt{\log n})$.

► **Lemma 9.** *For any assignment $\rho : J \rightarrow M$, there is some machine with load $\Omega(\sqrt{\log n})$.*

Proof. We first show that each size class is heavy on at least $\ell = \sqrt{\log n}/10$ machines. Suppose this is not true for some size class $s \in [h]$. Let $H \subseteq M$ be the set of machines on which s is heavy, and so $|H| < \ell$. As $I^{(s)}$ forms an (n_s, m, ℓ, β) set-system, by Definition 3, $|\cup_{i \in H} A_i(s)| \leq (1 - \beta)n_s$, and hence at least βn_s jobs from $U(s)$ are assigned to machines $i \notin H$. However, as any machine $i \notin H$ can have at most $\beta^2 n_s$ jobs from $A_i(s)$, the machines in $i \notin H$ can have at most $m \cdot \beta^2 n_s < \beta n_s$, contradicting that each job in $U(s)$ was assigned to some machine.

By averaging over machines, there exists a machine i on which at least $(h\ell/m) = \Omega(\sqrt{\log n})$ size classes are heavy. By Lemma 8, this implies that $\psi_i = \Omega(\sqrt{\log n})$. ◀

This concludes the proof of Theorem 1.

4 Reduction from Label-Cover

We now prove Theorem 2 by reducing Label Cover to a GMP instance.

Overview. Our construction builds on the ideas used by Lund and Yannakakis [15] to show the $\Omega(\log n)$ hardness of set cover, using a gadget based on a natural $\Omega(\log n)$ integrality gap instance. We first give a rough sketch of their idea (see e.g. [2] for an excellent exposition), and then explain the additional steps needed in our setting and why we only get a $\log^\delta n$ hardness for some small $\delta > 0$, despite the $\Omega(\log^{1/2} n)$ integrality gap instance above.

Consider a label cover instance $\mathcal{L} = ((U, V, E), L, \Pi)$, as defined in Definition 5, with label set $[L]$ and vertex sets U and V . The main idea in [15] is the following. For each edge $e = (u, v) \in E$, one associates a (n, m, ℓ, β) -system $I^e = (U^e; A_1^e \dots, A_m^e)$ (with disjoint universes for each edge). The sets will be associated with labels for vertices (so that $L = m$) and we associate the set $A_{\pi_e(a)}^e$ with label a for u and \overline{A}_b^e with label b for vertex v . The point is that in the completeness case, where the labels a' and b' for u and v satisfy e (i.e., $\pi_e(a') = b'$), U^e can be covered by just the two corresponding sets $A_{\pi_e(a')}^e$ and $\overline{A}_{b'}^e$. Conversely, in the soundness case, if U^e is covered using less than ℓ sets, there must be a pair of sets that are complements of each other, which can be used to produce a good labeling for \mathcal{L} .

To adapt this to our setting, suppose we create a job for each element in U^e , and m machines per vertex, one for each label in $[L]$. Also suppose that for labels a, b , we set the processing times of jobs in the set $A_{\pi_e(a)}^e$ to be finite on the a -th machine of vertex u and the processing times of jobs in set \overline{A}_b^e to be finite on the b -th machine of vertex v . In the completeness case, any perfect labeling gives an assignment where jobs are assigned to exactly one out of the m machines per vertex (similar to the value of the LP solution in

the gap example in Section 3). However, the soundness argument fails, as a low makespan assignment can spread jobs from U^e on multiple machines, and not give any information to recover a good labeling (this is similar to the reason we needed multiple size classes in the gap example in Section 3).

To get around this, for each e , we will use h different set systems $I^{e,1}, \dots, I^{e,h}$ (of geometrically increasing sizes) where the s -th set system is the following $I^{e,s} = (U^e(s); A_1^e(s), \dots, A_m^e(s))$. Each vertex will have m machines, one for each label. The intended solution is that if vertex u is assigned label a , then we pick the sets $A_{\pi_e(a)}^e(1), \dots, A_{\pi_e(a)}^e(h)$, and assign the corresponding jobs to the m machines for u with small makespan (this requires some care so that for any label a , jobs from different classes s can be assigned to different machines, but we ignore this issue here).

Using the properties of the norm ψ and the arguments in Section 3 for the integrality gap, one can show that if $h = \Omega(m)$, then given any schedule with low makespan, one can construct a good labeling thereby proving soundness. A key new idea here beyond [15] is to show that there is some fixed size class s^* , such that the assignment of jobs in class s^* gives a small set of good candidate labels for a large fraction of edges. However, as the hardness of Label-Cover is only $\Omega(L^c)$ for some small constant c as a function of the number of labels, our resulting hardness is only $\Omega(\log^{c'} n)$ for some small $c' > 0$, instead of $\Omega(\log^{1/2} n)$.

4.1 The Reduction

Suppose that we are given a label cover instance, $\mathcal{L} = ((U, V, E), L, \Pi)$ satisfying the properties of Lemma 6, i.e., the number of vertices $|U| = |V| = N$, the degree of every vertex is $d = O((\log N)^{c_1})$ and the number of labels $L = \sqrt{\log N}$. We now describe a polynomial-time (randomized) reduction from \mathcal{L} to a GMP instance \mathcal{I} with machines M , jobs J and assign processing times for each machine $i \in M$ and job $j \in J$.

Machines. For each vertex $w \in U \cup V$, we create $m = L = \sqrt{\log N}$ machines. We denote the set of machines corresponding to vertex w by $M_w = \{w_1, \dots, w_m\}$. We denote the set of all machines by $M = \bigcup_{w \in U \cup V} M_w$. In total, we have $2N\sqrt{\log N}$ machines.

Jobs. For each edge $e \in E$, we create $O(N)$ jobs and partition them into $h = m/8$ size-classes $U^e(1), U^e(2), \dots, U^e(h)$ of geometrically increasing size. More precisely, we pick the number of jobs in the s -th set to be $|U^e(s)| = \sqrt{N} \cdot \exp(4s\sqrt{\log N})$ which is always $O(N)$ since $s \leq \sqrt{\log N}/8$. Therefore, the total number n of jobs created is $\text{poly}(N)$. We also remark that these sets can be constructed efficiently by the randomized procedure described in Lemma 4 and that this is the only randomized step of the reduction.

Processing times. To assign processing times, for each edge $e \in E$ and $s \in [h]$, we construct a $(|U^e(s)|, m, \ell, \beta)$ set system $I^{e,s} = (U^e(s); A_1^e(s), \dots, A_m^e(s))$ with $\ell = m/10$ and $\beta = \exp(-m)$. For every vertex $u \in U$, label $a \in [L]$ and $s \in [h]$, define the set

$$S_{u,a,s} = \bigcup_{e \in \delta(u)} A_{\pi_e(a)}^e(s). \tag{3}$$

Similarly for every vertex $v \in V$, label $b \in [L]$ and $s \in [h]$, define the set

$$S_{v,b,s} = \bigcup_{e \in \delta(v)} \overline{A}_b^e(s). \tag{4}$$

21:10 On Minimizing Generalized Makespan on Unrelated Machines

For every vertex $w \in U \cup V$, label $a \in [L]$ and size class $s \in [h]$, we choose the processing times of all the jobs in $S_{w,a,s}$ to be $p^{(s)} = \beta^{s-1}$ on the machine w_i if the index i satisfies $i \equiv (a + s) \pmod{m}$. This way of assigning processing times ensures that for a fixed machine w_i and a label a , there is at most one set of jobs among $S_{w,a,1}, S_{w,a,2}, \dots, S_{w,a,h}$ that have finite processing time on w_i . This is a useful property to have when we prove the completeness of our reduction.

Norm. Define the norm $\psi = \sum_{s \in [h]} \psi^{(s)}$, where each $\psi^{(s)}$ is a scaled top- k norm given by

$$\psi^{(s)}(\mathbf{v}) = \frac{\text{Top}_{\beta^2 n_s}(\mathbf{v})}{(\beta^2 n_s p^{(s)})} \quad (5)$$

where $n_s = \left| \bigcup_{e \in \delta(w)} U^e(s) \right|$ is the number of jobs of size class s contained in edges incident to any vertex. Note that since the graph is d -regular this number is the same for each vertex.

The above reduction takes $\text{poly}(N)$ time because there are only polynomially many jobs, machines and the (n, m, ℓ, β) set systems required can be constructed efficiently by Lemma 4.

4.2 Analysis

The set of jobs $S_{w,a,s}$ for any vertex w , label a and size class s , only contribute to the s -th top- k term of ψ on machine w_i . Similar to Definition 7, we define heavy and light size classes and state a lemma whose proof is analogous to that of Lemma 8.

► **Definition 10 (Heavy Size Class).** For a vertex w , label a and size class s , consider the set $S_{w,a,s}$ of jobs and the machine w_i satisfying $i \equiv (a + s) \pmod{m}$. For an assignment $\rho : J \rightarrow M$, we say that size class s is heavy on machine w_i , if ρ assigns at least $\beta^2 n_s$ jobs from $S_{w,a,s}$ to machine w_i ; otherwise, we say the size class s is light on machine w_i .

► **Lemma 11.** For any vertex w , label a , size class s , and a machine w_i satisfying $i \equiv a + s \pmod{m}$, the norm $\psi_{w_i}(S_{w,a,s}) = 1 + o(1)$. Furthermore, for an assignment $\rho : J \rightarrow M$, let S be the set of jobs assigned to machine w_i . Then $\psi_{w_i}(S)$ is at least the number of heavy size classes heavy on w_i .

4.2.1 Completeness

Given a labeling σ for the label cover instance \mathcal{L} which satisfies all the edges, we use it to construct an assignment of jobs ρ with a low makespan.

► **Lemma 12.** If the label cover instance \mathcal{L} satisfies $\text{OPT}(\mathcal{L}) = 1$, the instance \mathcal{I} has an assignment $\rho : J \rightarrow M$ with makespan less than 2.

Proof. Let σ be the labeling of vertices that satisfies all edges in the label cover instance \mathcal{L} . Consider the assignment ρ of jobs to machines constructed using σ in the following way: for every vertex $w \in U \cup V$, and size class $s \in [h]$, assign the jobs in $S_{w,\sigma(w),s}$ to machine w_i where the index i satisfies $i \equiv \sigma(w) + s \pmod{m}$.

For an edge $e = (u, v) \in E$ we first show that each job in the sets $U^e(1), U^e(2), \dots, U^e(h)$ is assigned to some machine. For a size class $s \in [h]$, the jobs in $S_{u,\sigma(u),s}$ are assigned to a machine in M_u , and similarly, the jobs in $S_{v,\sigma(v),s}$ are assigned to a machine in M_v . Since $\pi_e(\sigma(u)) = \sigma(v)$, we infer that $A_{\sigma(v)}^e(s) = A_{\pi_e(\sigma(u))}^e(s) \subseteq S_{u,\sigma(u),s}$ and $\overline{A_{\sigma(v)}^e}(s) \subseteq S_{v,\sigma(v),s}$. It follows that each job in $U^e(s)$ is assigned to some machine.

We now bound the makespan of the assignment by showing that each machine is assigned jobs from at most one size class. For a vertex w and $i \in [m]$ consider the machine w_i . Since $h < m$, there is at most one $s \in [h]$ for which $\sigma(w) + s \equiv i \pmod{m}$. Therefore, w_i is assigned jobs from at most one of the sets $S_{w,\sigma(w),1}, S_{w,\sigma(w),2}, \dots, S_{w,\sigma(w),h}$. Therefore, by Lemma 11 its norm is at most $1 + o(1) < 2$. ◀

4.2.2 Soundness

Next, we show that if the instance \mathcal{I} has a makespan much less than ℓ , then $OPT(\mathcal{L})$ is large. Towards this end, we first prove some useful lemmas. Consider an assignment ρ with makespan $T \leq \ell/100$. Call a size class s to be *good* for a vertex w if it is heavy on at most $32T$ of the machines w_1, \dots, w_m ; if not define it to be bad. We first show that a large fraction of size classes are good for any vertex.

► **Lemma 13.** *There are at least $3h/4$ good size classes for each vertex.*

Proof. Let b be the number of bad size classes for some vertex w . By averaging over the m machines on vertex w , there is a machine w_i for which at least $(32Tb)/m$ size classes are heavy. By Lemma 11, w_i has norm at least $(32Tb)/m$. As any machine has norm at most T , we get $b \leq m/32 = h/4$ and hence the claim follows. ◀

Call a size class *good* for an edge (u, v) if it is good for both u and v . By Lemma 13 each edge e has at least $h/2$ good size classes. By averaging over the edges $e \in E$, there must exist a size class $s^* \in [h]$ which is good for at least $|E|/2$ edges.

We will fix the class s^* henceforth, and use it to construct a good label cover solution by assigning a suitable label to each vertex. These labels will only depend on the class s^* .

Constructing a good labeling. For each vertex $w \in U \cup V$, we define $L(w)$ to be the set of all labels a such that size-class s^* is heavy on w_i where $i \equiv (a + s) \pmod{m}$. If no such label exists, add an arbitrary label to $L(w)$.

► **Lemma 14.** *Let $e = (u, v)$ be an edge for which s^* is good. There exists $a \in L(u)$ and $b \in L(v)$ such that $\pi_e(a) = b$.*

Proof. Assume there are no such labels $a \in L(u)$ and $b \in L(v)$ for which $\pi_e(a) = b$. Since $|L(u)| + |L(v)| \leq 64T < \ell$, then the union of all the sets $A_{\pi_e(a)}^e(s^*)$ and $\overline{A}_b^e(s^*)$ such that $a \in L(u)$ and $b \in L(v)$ covers at most $(1 - \beta)|U^e(s^*)|$ from Definition 3.

From the definition of a light size class, for all labels $a \notin L(u)$ (resp. $b \notin L(v)$), at most $\beta^2 n_{s^*} = \beta^2(|U^e(s^*)| \cdot d)$ jobs from the sets $A_{\pi_e(a)}^e(s^*)$ (resp. $\overline{A}_b^e(s^*)$) are assigned to some machines on vertex u (resp. v). Notice that the degree of the graph (U, V, E) is $d = O((\log N)^{c_1})$. So, the union of all the jobs assigned from the sets $A_{\pi_e(a)}^e(s^*)$ (resp. $\overline{A}_b^e(s^*)$) such that $a \notin L(u)$ (resp. $b \notin L(v)$) has at most $(2m \cdot \beta^2 \cdot |U^e(s^*)| \cdot d) < \beta|U^e(s^*)|$ jobs which is a contradiction. ◀

For each vertex w , set $\sigma(w)$ to be a label selected uniformly at random from $L(w)$. We show that σ satisfies a large fraction of edges of \mathcal{L} completing the proof of soundness.

► **Lemma 15.** *If the instance \mathcal{I} has an assignment $\rho : J \rightarrow M$ with makespan $T \leq \ell/100$, then $OPT(\mathcal{L}) \geq 1/(2048T^2)$.*

Proof. Consider an edge $e = (u, v)$ for which size class s^* is good. In this case, we have $|L(u)| \leq 32T$ and $|L(v)| \leq 32T$. Also, by Lemma 14 there exists $a \in L(u)$ and $b \in L(v)$ for which $\pi_e(a) = b$. Therefore, $\pi_e(\sigma(u)) = \sigma(v)$ with probability at least $1/(32T)^2$. We have by the analysis above that the number of edges for which s^* is good is at least $|E|/2$. Therefore, the expected number of edges satisfied by σ is at least $|E|/(2(32)^2T^2) = |E|/(2048T^2)$ and the lemma follows. ◀

► **Theorem 2.** *There is a universal constant $\delta > 0$, such that any polynomial-time approximation algorithm for GMP has approximation ratio $\Omega(\log^\delta n)$ provided that $\text{NP} \not\subseteq \text{ZTIME}(n^{O(\log \log n)})$.*

Proof. Suppose that we have an algorithm that in polynomial time can decide if the GMP instance constructed has a makespan of at least T or at most 2. By the reduction above, from Lemmas 12 and 15 this algorithm can also distinguish between label cover instances that have value 1 and those that have value at most $1/2048T^2$. Due to the hardness of label cover (Lemma 6), this is not possible if $1/2048T^2 \geq 1/(\log N)^c$, i.e., if $T \leq O((\log N)^{c/2}) = O((\log n)^{c/2})$ since the number of jobs $n = \text{poly}(N)$. This, in particular, implies that any approximation algorithm for GMP has an approximation ratio of at least $\Omega((\log n)^{c/2})$ provided that $\text{NP} \not\subseteq \text{ZTIME}(n^{O(\log \log n)})$. ◀

5 Concluding Remarks

We conjecture that GMP admits an $O(\sqrt{\log n})$ approximation, based on suitably rounding the configuration LP. However, we are unable to prove any $o(\log n)$ approximation even in the restricted assignment case (note that the integrality gap and hardness instances in this paper only use restricted assignment). Finding a $o(\log n)$ approximation algorithm for any of these variants would be extremely interesting.

References

- 1 Noga Alon, Yossi Azar, Gerhard J Woeginger, and Tal Yadid. Approximation schemes for scheduling on parallel machines. *Journal of Scheduling*, 1(1):55–66, 1998.
- 2 Sanjeev Arora and Carsten Lund. Hardness of approximations. In *Approximation algorithms for NP-hard problems*, pages 399–446. PWS Publishing Company, Boston, 1997.
- 3 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, May 1998. doi:10.1145/278298.278306.
- 4 Yossi Azar and Amir Epstein. Convex programming for scheduling unrelated parallel machines. In *Symposium on Theory of Computing*, pages 331–337, 2005. doi:10.1145/1060590.1060639.
- 5 Deeparnab Chakrabarty and Chaitanya Swamy. Approximation algorithms for minimum norm and ordered optimization problems. In *Symposium on Theory of Computing*, pages 126–137, 2019. doi:10.1145/3313276.3316322.
- 6 Deeparnab Chakrabarty and Chaitanya Swamy. Simpler and better algorithms for minimum-norm load balancing. In *27th Annual European Symposium on Algorithms*, volume 144 of *LIPICs*, pages 27:1–27:12, 2019. doi:10.4230/LIPICs.ESA.2019.27.
- 7 Siu On Chan. Approximation resistance from pairwise-independent subgroups. *J. ACM*, 63(3):27:1–27:32, 2016. doi:10.1145/2873054.
- 8 Shichuan Deng, Jian Li, and Yuval Rabani. Generalized unrelated machine scheduling problem. In *Symposium on Discrete Algorithms (SODA)*, pages 2898–2916. SIAM, 2023.
- 9 Uriel Feige. A threshold of $\ln n$ for approximating set cover (preliminary version). In *Symposium on the Theory of Computing*, pages 314–318, 1996. doi:10.1145/237814.237977.

- 10 Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Symposium on Theory of Computing*, pages 411–419, 2007.
- 11 Sharat Ibrahimpur and Chaitanya Swamy. Minimum-norm load balancing is (almost) as easy as minimizing makespan. In *48th International Colloquium on Automata, Languages, and Programming*, volume 198 of *LIPICs*, pages 81:1–81:20, 2021. doi:10.4230/LIPICs.ICALP.2021.81.
- 12 Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *Symposium on Foundations of Computer Science, FOCS*, pages 592–601, 2018. doi:10.1109/FOCS.2018.00062.
- 13 V. S. Anil Kumar, Madhav V. Marathe, Srinivasan Parthasarathy, and Aravind Srinivasan. A unified approach to scheduling on unrelated parallel machines. *J. ACM*, 56(5):28:1–28:31, 2009. doi:10.1145/1552285.1552289.
- 14 Jan Karel Lenstra, David B. Shmoys, and Éva Tardos. Approximation algorithms for scheduling unrelated parallel machines. *Math. Program.*, 46:259–271, 1990. doi:10.1007/BF01585745.
- 15 Carsten Lund and Mihalis Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM (JACM)*, 41(5):960–981, 1994.
- 16 Konstantin Makarychev and Maxim Sviridenko. Solving optimization problems with diseconomies of scale via decoupling. *J. ACM*, 65(6):42:1–42:27, 2018. doi:10.1145/3266140.
- 17 Anup Rao. Parallel repetition in projection games and a concentration bound. In *Symposium on Theory of Computing*, pages 1–10, 2008.
- 18 Ran Raz. A parallel repetition theorem. In *Symposium on Theory of computing*, pages 447–456, 1995.