

Inapproximability of Hypergraph Vertex Cover and Applications to Scheduling Problems

Nikhil Bansal

Subhash Khot

November 5, 2009

Abstract

Assuming the Unique Games Conjecture (UGC), we show optimal inapproximability results for two classic scheduling problems. We obtain a hardness of $2 - \varepsilon$ for the problem of minimizing the total weighted completion time in concurrent open shops. We also obtain a hardness of $2 - \varepsilon$ for minimizing the makespan in the assembly line problem.

These results follow from a new inapproximability result for the Vertex Cover problem on k -uniform hypergraphs that is stronger and simpler than previous results. We show that assuming the UGC, for every $k \geq 2$, the problem is inapproximable within $k - \varepsilon$ even when the hypergraph is *almost k -partite*.

1 Introduction

A k -uniform hypergraph $G = (V, E)$ consists of a collection of vertices $v \in V$ and a collection of edges $e \in E$ where an edge corresponds to a k element subset of V . A vertex cover of G is a subset S of vertices such that each edge contains at least one vertex from S (i.e. $e \cap S \neq \emptyset$ for each edge $e \in E$). The Ek -Vertex-Cover problem is to find the minimum vertex cover of G . This is equivalent to the Set Cover problem (by viewing edges as elements and vertices as sets) where each element lies in exactly k sets. An independent set in a k -uniform hypergraph is defined as the complement of a vertex cover, i.e. an independent set is a subset of vertices T such that no edge is completely contained in T .

The Ek -Vertex-Cover problem, and in particular the case $k = 2$ corresponding to the Vertex Cover problem on graphs, is extremely well-studied. There is an almost trivial k -approximation by finding a maximal matching and the best known algorithms achieve only a slight improvement of $(1 - o(1))k$ [9, 13]. On the inapproximability side, for the case $k = 2$, Dinur and Safra [5] obtained 1.36 NP-hardness improving on the $\frac{7}{6} - \varepsilon$ hardness by Håstad [10]. For larger k , a sequence of works obtained hardness factor $k^{1/19}$ (Trevisan [22]), $2 - \varepsilon$ for some constant k (Goldreich [8]), $2 - \varepsilon$ for $k = 4$, $\frac{3}{2} - \varepsilon$ for $k = 3$, $k^{1-o(1)}$ (Holmerin [12, 11]), $k - 3 - \varepsilon$ (Dinur et al [3]), and finally $k - 1 - \varepsilon$ (Dinur et al [4]) that superseded all earlier results for $k \geq 3$.

Assuming the Unique Games Conjecture [14] however, optimal $k - \varepsilon$ hardness is known due to Khot and Regev [15]. Recently, in [1], the authors of this paper strengthened the Khot-Regev result for the case $k = 2$: they showed that assuming the UGC, vertex cover in graphs is $2 - \varepsilon$ hard to approximate even when the graph is almost bipartite, i.e. when the graph has two disjoint independent sets $V_1, V_2 \subseteq V$ of size $(\frac{1}{2} - \varepsilon)|V|$ each. The techniques developed were also used to show a tight $2 - \varepsilon$ inapproximability result (assuming a variant of the UGC) for the classic problem of minimizing the weighted completion time with precedence constraints. They designed a PCP with one free bit, near-perfect completeness, and arbitrarily

low soundness, and used the equivalence between hardness of vertex cover problem and existence of PCPs with zero free bits [6, 2].

In this paper, we obtain a stronger UGC-based inapproximability result for Ek -Vertex-Cover which subsumes both [15, 1] and holds on hypergraphs that are almost k -partite (essential for the applications). Our result is conceptually different and much simpler than both these results (albeit using a recent invariance style theorem of Mossel [19]). Formal statement of our result is:

Theorem 1.1 *Assuming the Unique Games Conjecture, for any integer $k \geq 2$ and arbitrary constants $\varepsilon, \delta > 0$, given a k -uniform hypergraph $G = (V, E)$, distinguishing between the following cases is NP-hard:*

- (YES Case): *there exist disjoint subsets $V_1, \dots, V_k \subseteq V$, satisfying $|V_i| \geq \frac{1-\varepsilon}{k}|V|$ and such that no hyperedge contains two or more vertices from some V_i .*
- (NO Case): *every vertex cover has size at least $(1 - \delta)|V|$.*

Note that in the YES case, letting $V' = V \setminus (V_1 \cup \dots \cup V_k)$, the sets $V' \cup V_i$ for $i = 1, \dots, k$ are almost disjoint vertex covers of size at most $(\frac{1}{k} + \varepsilon)|V|$ each.

1.1 Applications

Our hypergraph vertex cover result was motivated by showing hardness results for two classic problems in scheduling. The first is the problem of minimizing weighted completion time of jobs in the so-called Concurrent Open Shop model (also referred to as the Order Scheduling model). The second is the problem of minimizing the makespan in the so-called assembly line problem.

Concurrent Open Shop: There is a collection of machines $M = \{1, \dots, m\}$ with each machine capable of producing its own unique type of component. There is a collection of jobs $J = \{1, \dots, n\}$ where each job requires a specific amount of components from each machine. Formally, each job $j \in J$ has weight w_j and requires a processing of p_{ij} units for each $i = 1, \dots, m$. In the concurrent open shop model, there is no restriction on how the p_{ij} units for a job must be scheduled. The components of a job are independent of each other, and for example, all of them can be scheduled in parallel. A job is completed when all its components are completed and the goal is to minimize the total weighted completion time of the jobs.

Given its generality, it is widely used to model various applications in manufacturing and supply chain (see for example [16] for a survey). Several 2 approximations, using various different techniques are known for this problem [17, 7, 18]. However, until recently only NP-Hardness was known. Garg et al showed that the problem is APX-Hard [7]. More recently, Mastrolilli et al [18] showed that the problem is NP-hard to approximate within $\frac{6}{5} - \varepsilon$, and that it is hard to approximate within $\frac{4}{3} - \varepsilon$ assuming the UGC. In this work we give a tight hardness result assuming the UGC and show that the problem is hard to approximate within $2 - \varepsilon$. Our result is based on combining the construction of [18] with Theorem 1.1.

The Assembly Line Problem: There are $m + 1$ machines $M = \{0, \dots, m\}$. Machine 0 is special and called the assembly machine. Each job j has an operation O_{ij} of size p_{ij} for each machine $i = 1, \dots, m$. These operations can be done in any order. After all these operations are completed, a final operation O_{0j} of size p_{0j} is performed on machine 0 (that assembles the operations O_{1j}, \dots, O_{mj} to produce the final product j). The job j is completed when O_{0j} is completed. The goal is to minimize the makespan, or equivalently, the latest completion time of a job.

The problem is a classic one [20]. It has a trivial 2-approximation: in fact any arbitrary schedule that does not waste time unnecessarily is a 2-approximation. This is because $\max_{i=0,\dots,m} \text{Load}(i)$ is a lower bound for the problem, where $\text{Load}(i)$ is the total size of operations on machine i . On the other hand, any schedule of the type described above achieves a makespan of $\text{Load}(0) + \max_{i=1,\dots,m} \text{Load}(i)$. A somewhat better $2 - \frac{1}{m}$ approximation is also known [20]. The problem was known to be strongly NP-hard even for $m = 2$ [20]. It featured in Schuurman and Woeginger's [21] well-known list of important open problems in scheduling. It was subsequently observed by several researchers (but unpublished) that an APX-Hardness follows from known hardness results for coloring k -colorable graphs. However, these results only yield a small hardness factor. In this paper, we show that the trivial 2-approximation algorithm described above is essentially the best one can hope for, assuming the UGC.

1.2 Techniques

Our hardness reduction for the hypergraph vertex cover problem is quite simple and described in Sections 2.2, 3, and 4. Here we give a quick overview. As is standard in UGC-based reductions, the crux of our construction is a *dictatorship test*. This can be viewed as a construction of a k -uniform hypergraph with vertex set Ω^n for a finite set Ω . The hypergraph has these two properties:

- (Completeness): If the hypergraph is partitioned into $|\Omega|$ sets $\{V_\omega\}_{\omega \in \Omega}$ based on j^{th} co-ordinate for any fixed j , then each hyperedge contains at most one vertex from V_ω for any $\omega \in \Omega$.
- (Soundness): Any large (i.e. linear sized) independent set in the hypergraph must have an *influential co-ordinate*. This follows directly from a theorem of Mossel [19] that gives gaussian bounds for noise correlation of functions.

Since in the completeness case, we need the hypergraph to be (almost) k -partite, it would be natural to take $|\Omega| = k$. To satisfy the completeness property, we are led to define $(x^1, \dots, x^k) \in (\Omega^n)^k$ to be an hyperedge iff for every co-ordinate j , (x_j^1, \dots, x_j^k) are distinct, i.e. this tuple is a permutation of Ω . However, this does not quite work as the set of all permutations of Ω is *disconnected* as a subset of Ω^k , whereas application of Mossel's theorem has a certain connectedness requirement. We get around this problem by a simple trick that seems quite useful for future applications: we introduce a dummy element called 0 and let $\Omega = \{0, 1, \dots, k\}$. The hyperedges are defined as before: $(x^1, \dots, x^k) \in (\Omega^n)^k$ is an hyperedge iff for every co-ordinate j , (x_j^1, \dots, x_j^k) are distinct. Now the set of all k -tuples of distinct elements of the $k + 1$ sized set Ω happens to be a connected subset of Ω^k and therefore Mossel's theorem is applicable. We let the element 0 to have negligible probability mass relative to the other elements so that the hypergraph is almost k -partite. Construction of this hypergraph (i.e. the dictatorship test) leads to a UGC-based hardness result in a straightforward manner.

2 Preliminaries

2.1 The Unique Games Conjecture

Definition 2.1 An instance $\mathcal{L}(G(V, W, E), [n], \{\sigma(v, w)\}_{(v,w) \in E})$ of Unique Games consists of a regular bipartite graph $G(V, W, E)$ and a set $[n]$ of labels. For each edge $(v, w) \in E$ there is a constraint specified by a permutation $\sigma(v, w) : [n] \mapsto [n]$. The goal is to find a labelling $\ell : V \cup W \rightarrow [n]$ of the vertices such that as many edges as possible are satisfied, where an edge $e = (v, w)$ is said to be satisfied if $\ell(v) = \sigma(v, w)(\ell(w))$.

Definition 2.2 Given a Unique Games instance $\mathcal{L}(G(V, W, E), [n], \{\sigma_{v,w}\}_{(v,w) \in E})$ let $\text{OPT}(\mathcal{L})$ denote the maximum fraction of simultaneously satisfied edges of \mathcal{L} by any labeling, i.e.

$$\text{OPT}(\mathcal{L}) := \frac{1}{|E|} \max_{\ell: V \cup W \rightarrow [n]} |\{e \in E : \ell \text{ satisfies } e\}|.$$

A formal statement of the Unique Games Conjecture appears below; the conjecture has been widely employed to prove strong inapproximability results for several problems.

Conjecture 2.3 ([14]) For any constants $\zeta, \gamma > 0$, there is a sufficiently large constant $n = n(\zeta, \gamma)$ such that, for Unique Games instances \mathcal{L} with label set $[n]$ it is NP-hard to distinguish between

- $\text{OPT}(\mathcal{L}) \geq 1 - \zeta$,
- $\text{OPT}(\mathcal{L}) \leq \gamma$.

2.2 Gaussian bounds on Correlated Spaces

Let (Ω^k, \mathbb{P}) be a (so-called correlated) probability space where Ω is a finite set and \mathbb{P} is a probability distribution over Ω^k , and let the relation $R := \{x \in \Omega^k \mid \mathbb{P}(x) > 0\} \subseteq \Omega^k$ be the support of \mathbb{P} . Assume moreover that the marginal of \mathbb{P} on any co-ordinate is the same, denote it by \mathbb{P} as well. We say that R is a *connected relation* if for any $x, y \in R$, there exists a path $x = y(0), y(1), \dots, y(r) = y$ in R such that $y(i)$ and $y(i+1)$ differ in one coordinate only.

Consider a k -uniform hypergraph with the vertex set Ω^n . A tuple $(x^1, \dots, x^k) \in (\Omega^n)^k$ is defined to be an hyperedge if for every co-ordinate $1 \leq i \leq n$, we have $(x_i^1, \dots, x_i^k) \in R$. In fact we define a probability distribution on the set of hyperedges where a random hyperedge (x^1, \dots, x^k) is selected by selecting independently for $1 \leq i \leq n$, the tuple (x_i^1, \dots, x_i^k) from space (Ω^k, \mathbb{P}) . We will use a theorem of Mossel [19] stating that every *large* independent set in this hypergraph must have an *influential co-ordinate*. Specifically, let $A \subseteq \Omega^n$ be a set and $f_A : \Omega^n \mapsto \{0, 1\}$ be its indicator function. The fraction of hyperedges that lie entirely inside A is clearly,

$$\mathbb{E}_{(x^1, \dots, x^k)} \left[\prod_{j=1}^k f_A(x^j) \right].$$

Mossel's theorem states that if $\mathbb{E}[f_A]$ is *large* and all influences of f_A are sufficiently small, then this expectation is lower bounded by a constant, and therefore A cannot be an independent set. In fact, as is necessary for our application, the theorem works even for (non-boolean) functions $f : \Omega^n \mapsto [0, 1]$ and one only requires that *degree d influences* are sufficiently small. We state the theorem below and also define the notion of influence of variable.

Theorem 2.4 (Mossel [19]) Let (Ω^k, \mathbb{P}) be a correlated space such that the support $R \subseteq \Omega^k$ of \mathbb{P} is connected and the minimum probability of any atom in R is at least $\alpha \leq \frac{1}{2}$. Then there exist continuous functions $\underline{\Gamma} : (0, 1) \mapsto (0, 1)$ and $\overline{\Gamma} : (0, 1) \mapsto (0, 1)$ such that the following holds: for every $\varepsilon > 0$, there exists a $\tau > 0$ and integer d such that for any function $f : \Omega^n \mapsto [0, 1]$, satisfying

$$\forall 1 \leq i \leq n, \quad \text{Infl}_i^{\leq d}(f) \leq \tau,$$

we have:

$$\underline{\Gamma}(\mathbb{E}[f]) - \varepsilon \leq \mathbb{E}_{(x^1, \dots, x^k)} \left[\prod_{j=1}^k f(x^j) \right] \leq \overline{\Gamma}(\mathbb{E}[f]) + \varepsilon.$$

Remark 2.5 We have stated Theorem 2.4 in the form that we need (we only need the lower bound). This is essentially the same statement as Theorem 6.4 in [19]. Therein, the theorem is stated for a multi-function version whereas we only need the single-function version. The bound $\rho(\Omega_i^{(1)}, \dots, \Omega_i^{(k)}) \leq \rho < 1$ required there follows from the fact that the relation R is connected, see Lemma 2.9 therein. The author also gives quantitative bounds for $\underline{\Gamma}(), \bar{\Gamma}(), \tau$, but we don't necessarily need specific bounds.

For the sake of quick reference, we note the definition of influence and low degree influence (see [19, Section 3]). We assume that (Ω, P) is a finite probability space and distribution on Ω^n is the product distribution P^n . Let $f = \sum_{\phi} \hat{f}(\phi) X_{\phi}$ be the multi-linear representation of f (analogue of the standard Fourier representation).

Definition 2.6 For a function $f : \Omega^n \mapsto \mathbb{R}$, and a co-ordinate $1 \leq i \leq n$,

$$\text{Infl}_i(f) := \mathbb{E}[\text{Var}(f) | x_1, x_{i-1}, x_{i+1}, \dots, x_n] = \sum_{\phi: \phi_i \neq 0} \hat{f}(\phi)^2.$$

Definition 2.7 For a function $f : \Omega^n \mapsto \mathbb{R}$, a co-ordinate $1 \leq i \leq n$, and integer d , the degree d influence is defined as:

$$\text{Infl}_i^{\leq d}(f) := \sum_{\phi: \phi_i \neq 0, |\phi| \leq d} \hat{f}(\phi)^2.$$

Lemma 2.8 If $f : \Omega^n \mapsto [0, 1]$, then the number of co-ordinates i such that $\text{Infl}_i^{\leq d}(f) \geq \tau$ is at most d/τ .

Proof: This follows from:

$$\sum_{i=1}^n \text{Infl}_i^{\leq d}(f) = \sum_{i=1}^n \sum_{\phi: \phi_i \neq 0, |\phi| \leq d} \hat{f}(\phi)^2 \leq d \cdot \sum_{\phi} \hat{f}(\phi)^2 = d \cdot \|f\|_2^2 \leq d.$$

■

3 Dictatorship Test

In this section we define the appropriate dictatorship test that we will later use as a gadget to show the UGC based Hardness. Let Ω be a finite set consisting of $k + 1$ elements, $\Omega := \{0, 1, \dots, k\}$ (the element 0 is special as we see next). Let \tilde{P} be a distribution on Ω where the elements $1, \dots, k$ have probability mass $(1 - \delta)/k$ and element 0 has mass δ . Later we will set $\delta \ll 1/k$.

Let R denote the following relation on Ω^k : A k -tuple $(a^1, \dots, a^k) \in R$ if and only if elements $\{a^1, \dots, a^k\}$ are all distinct. It is easy to see that one can define a probability distribution P on R such that on each co-ordinate its marginal is \tilde{P} and moreover each atom in R has probability mass at least $\alpha > 0$.

Claim 3.1 R is a connected relation.

Proof: Suppose $(a^1, \dots, a^k), (b^1, \dots, b^k) \in R$. Let us first consider the case when $|\{a^1, \dots, a^k, b^1, \dots, b^k\}| = k + 1$. Then there is some index i such that $b^i \notin \{a^1, \dots, a^k\}$. We set $a^i = b^i$ and proceed inductively on the remaining $k - 1$ coordinates (without ever considering coordinate i again). Now suppose that $|\{a^1, \dots, a^k, b^1, \dots, b^k\}| = k$. In this case we arbitrarily set some coordinate a^i to $[k + 1] \setminus \{b^1, \dots, b^k\}$, and reduce to the previous case. ■

Consider a hypergraph $G = (V, E)$ defined on Ω^n as follows: Each element $a \in \Omega^n$ corresponds to a vertex in G . A k -tuple $(a^1, \dots, a^k) \in (\Omega^n)^k$ is an hyperedge if and only if the relation R is satisfied for each coordinate j , i.e. for each $1 \leq j \leq n$, $(a_j^1, a_j^2, \dots, a_j^k) \in R$. The instance satisfies the following completeness and soundness properties:

Completeness: Fix any coordinate j , and consider the partition of the vertices based on the coordinate. That is, for $i = 0, \dots, k$, let V_i denote the set of vertices $V_i = \{v = (a_1, \dots, a_n) : a_j = i\}$. Clearly, $|V_0| = \delta|V|$ and $V_i = (1 - \delta)|V|/k$ for $i = 1, \dots, k$.

Claim 3.2 Any hyperedge contains at most one vertex from any set V_i .

Proof: Consider any hyperedge $e = (a^1, \dots, a^k)$ and consider the k -tuple (a_j^1, \dots, a_j^k) of the j -th coordinates of the vertices in e . By definition, this tuple satisfies R , and hence no two of them can be i , and hence lie in V_i . ■

Soundness:

Claim 3.3 For every $\beta > 0$, there exists $\tau > 0$ and integer d , such that if $A \subseteq \Omega^n$ is an independent set of size $\beta|V|$, then A (or rather its indicator function) has a coordinate with degree d influence at least τ .

Proof: This follows by a direct application of Theorem 2.4, by choosing $\varepsilon = \underline{\Gamma}(\beta)/2$. ■

4 UGC based Hardness

We assume the following variant of Unique Games Conjecture.

Hypothesis 4.1 For arbitrarily small constants $\zeta, \gamma > 0$, there exists an integer $n = n(\zeta, \gamma)$ such that for a Unique Games instance $\mathcal{L}(G(V, W, E), [n], \{\sigma_{v,w}\}_{(v,w) \in E})$, it is NP-hard to distinguish between:

- (YES Case:) There is a set $W' \subseteq W$ such that $|W'| \geq (1 - \zeta)|W|$ and a labeling $\ell : V \cup W \mapsto [n]$ that satisfies every edge (v, w) for $v \in V$ and $w \in W'$.
- (NO Case:) No labeling satisfies even a γ fraction of edges.

Moreover, we can assume that the graph is both left and right regular. This variant is equivalent to the original Unique Games Conjecture as shown by Khot and Regev [15, Lemma 3.6].

Given the UGC instance \mathcal{L} , we construct the following hypergraph instance $H = (X, Y)$. With each vertex $v \in V$ and $w \in W$, we associate the hypercube Ω^n , where Ω is chosen as in the gadget in section 3. Let $\Omega^n(v)$ denote the hypercube for vertex v . The set of vertices X for the hypergraph is defined as $X := \cup_{w \in W} \Omega^n(w)$.

Before we describe the edges, we give some notation. Let $\sigma : [n] \rightarrow [n]$ be a permutation. For an element $x = (x_1, \dots, x_n) \in \Omega^n$, let $x \circ \sigma$ denote the string $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$. For each vertex $v \in V$ and every subset of its k neighbors $w^1, \dots, w^k \in W$, we define a block $B(v, w^1, \dots, w^k)$. There will be a collection of hyperedges for each block defined as follows: For each $x^1, x^2, \dots, x^k \in \Omega^n(v)$ such that $(x^1, \dots, x^k) \in R^n$, we have a hyperedge $y = (x^1 \circ \sigma(v, w^1), \dots, x^k \circ \sigma(v, w^k))$. Here R is the relation defined in the gadget in section 3.

In other words, we identify the coordinates of $\Omega^n(w^1), \dots, \Omega^n(w^k)$ via the permutations $\sigma(v, w^j)$, and add an hyperedge among k vertices if they satisfy the relation R on each coordinate.

Completeness: Consider the labeling ℓ in the YES case, and let W' be the set of vertices satisfying the condition in the UGC Hypothesis 4.1.

For $i = 0, 1, \dots, k$, we define

$$X_i = \bigcup_{w \in W'} \{x : x \in \Omega^n(w), x_{\ell(w)} = i\}.$$

Note that for each hypercube corresponding to $w \in W'$, we are partitioning its vertices based on their $\ell(w)$ coordinate.

Claim 4.2 *The sets X_0, X_1, \dots, X_k are pairwise disjoint. $|X_i| \geq (1-\delta)(1-\zeta)|X|/k$ for each $i = 1, \dots, k$ and no hyperedge contains more than one vertex from X_i .*

Proof: The disjointness of the X_i follows by construction. Since the elements $1, \dots, k$ have probability mass $(1-\delta)/k$ in Ω , it follows that $|X_i \cap \Omega^n(w)| = (1-\delta)|\Omega^n(w)|/k$ for each $w \in W'$ and $i = 1, \dots, k$. As $|W'| \geq (1-\zeta)|W|$ this implies the claimed bound on the size of X_i .

It remains to show that no hyperedge contains two or more vertices from X_i . Consider some hyperedge y and suppose it lies in some block $B(v, w^1, \dots, w^k)$, then $y = (x^1 \circ \sigma(v, w^1), \dots, x^k \circ \sigma(v, w^k))$ for some $(x^1, \dots, x^k) \in R^n$. Let S denote the set of indices j such that $w^j \in W'$.

Since $X_i \cap \Omega^n(w) = \emptyset$ for $w \notin W'$, it follows that if $j \notin S$ then the vertex $x^j \circ \sigma(v, w^j)$ cannot lie in X_i . Thus we need to consider only the indices in S . Now for any $j \in S$, $x^j \circ \sigma(v, w^j)$ lies in X_i iff $(x^j \circ \sigma(v, w^j))_{\ell(w^j)} = i$. Since the labeling ℓ satisfies the constraints (v, w^j) for $j \in S$, we have $\sigma(v, w^j)_{\ell(w^j)} = \ell(v)$. Thus $(x^j \circ \sigma(v, w^j))_{\ell(w^j)} = (x^j)_{\ell(v)}$ for $j \in S$. So if there exist $j \neq j' \in S$ such that both $x^j \circ \sigma(v, w^j)$ and $x^{j'} \circ \sigma(v, w^{j'})$ lie in X_i , then this implies that $(x^j)_{\ell(v)} = (x^{j'})_{\ell(v)}$ contradicting the fact that $(x^1, \dots, x^k) \in R^n$. ■

Soundness: Suppose A is an independent set and that $|A| \geq \beta|X|$. Let $A_w = \Omega^n(w) \cap A$ be the vertices of A that lie in $\Omega^n(w)$ for $w \in W$. Define the boolean function $f_w : \Omega^n(w) \rightarrow \{0, 1\}$ where $f_w(x) = 1$ if $x \in A_w$ and is 0 otherwise. Let $N(v) \subseteq W$ denote the set of neighbors of $v \in V$. Let

$$f_v(x) = \mathbb{E}_{w \in N(v)} [f_w(x \circ \sigma(v, w))].$$

Since the unique games graph is regular, $\mathbb{E}_{v,x} [f_v(x)] \geq \beta$. Call a vertex $v \in V$ good if $\mathbb{E}_x [f_v(x)] \geq \beta/2$. By an averaging argument, at least $\beta/2$ fraction of vertices in V are good.

Since A is independent set, for any hyperedge $y = (x^1, \dots, x^k)$ in block $B(v, w_1, \dots, w_k)$, it must be that not all $f_{w^i}(x^i)$ are 1. Thus for any $v \in V$ and any k neighbors w^1, \dots, w^k of v , we have

$$\mathbb{E}_{(x^1, \dots, x^k)} \left[\prod_{i=1}^k f_{w^i}(x^i \circ \sigma(v, w^i)) \right] = 0. \quad (1)$$

Averaging over all k -tuples w^1, \dots, w^k of neighbors of v , we have

$$\begin{aligned} & \mathbb{E}_{(x^1, \dots, x^k)} \left[\prod_{i=1}^k f_v(x^i) \right] \\ &= \mathbb{E}_{(x^1, \dots, x^k)} \mathbb{E}_{w^1, \dots, w^k \in N(v)} \left[\prod_{i=1}^k f_{w^i}(x^i \circ \sigma(v, w^i)) \right] \\ &= 0. \end{aligned} \quad (2)$$

Applying Theorem 2.4 with $\varepsilon = \lfloor(\beta/2)/2$, it follows that f_v must have a co-ordinate with degree d influence at least τ . Let $S(v) \neq \emptyset$ be the set of all such influential co-ordinates.

For every good vertex $v \in V$ we give it an arbitrary label j from its (non-empty) set $S(v)$. For any $j \in S(v)$,

$$\begin{aligned} \tau &\leq \text{Infl}_j^{\leq d}(f_v) = \sum_{\phi_j \neq 0, |\phi| \leq d} \widehat{f}_v(\phi)^2 = \sum_{\phi_j \neq 0, |\phi| \leq d} \left(\mathbb{E}_w \left[\widehat{f}_w(\sigma(v, w)^{-1}(\phi)) \right] \right)^2 \\ &\leq \sum_{\phi_j \neq 0, |\phi| \leq d} \mathbb{E}_w \left[\widehat{f}_w(\sigma(v, w)^{-1}(\phi))^2 \right] = \mathbb{E}_w \left[\text{Infl}_{\sigma(v, w)^{-1}(j)}^{\leq d}(f_w) \right]. \end{aligned} \quad (3)$$

For every $w \in W$ define the set of candidate labels for w to be

$$\text{Cand}[w] = \{i \in [n] : \text{Infl}_i^{\leq d} \geq \tau/2\}.$$

Since the sum of degree d influences is at most d , the number of candidates $|\text{Cand}[w]|$ is at most $2d/\tau$. Since at least $\beta/2$ fraction of vertices $v \in V$ are good, and by (3) for each good vertex at least $\tau/2$ fraction of its neighbors in W have $\text{Infl}_{\sigma(v, w)^{-1}(j)}^{\leq d}(f_w) \geq \tau/2$. Now if we label each vertex $w \in W$ by choosing a random element of $\text{Cand}[w]$ (or any label if this set is empty), it follows that among the set of edges adjacent to good vertices v , at least a $(\tau/2)(\tau/2d)$ -fraction are satisfied in expectation. Thus it follows that there is a labeling for all vertices which satisfies a $(\beta/2)(\tau/2)(\tau/2d)$ fraction of all edges. This completes the proof of soundness.

5 Applications

5.1 Total weighted completion time concurrent open shop

Recall the problem definition from Section 1.1. We give a reduction from Ek -Vertex-Cover. Given a k -uniform hypergraph instance $G = (V, E)$ satisfying the properties of Theorem 1.1. We construct an instance I of the concurrent open shop problem as follows: A vertex $v \in V$ corresponds to a job J_v in I . Each hyperedge $e \in E$ corresponds to a machine m_e in I . Let E_v denote the set of hyperedges in G that contain the vertex v . For each edge in $e \in E_v$ we introduce an operation of unit size for job J_v that needs to be executed on machine m_e . All other operations of J_v have size 0. Each job has weight 1.

We claim that in the YES case, the optimum average weighted completion time is at most $(1 + \varepsilon)(k + 1)/2$. Let V_1, \dots, V_k be the disjoint subsets of V satisfying the properties of G stated in theorem 1.1. Consider the following schedule. For each job $v \in V_i$ schedule all its unit size operations at time i (the operations of size 0 can all be completed instantaneously at beginning schedule, and do not play any role in the proof). This gives a valid schedule for the jobs corresponding to vertices in $V_1 \cup \dots \cup V_k$, because any hyperedge $e \in E$ contains at most one vertex from V_i , and hence no two operations are assigned to a machine at the same time. For the remaining jobs in $V \setminus (V_1 \cup \dots \cup V_k)$, we place their operations arbitrarily in the gaps in the schedule. Since it is a k -uniform hypergraph, each machine has load at most k , and hence each job completes by time k . Thus the average completion time is at most

$$\frac{1 - \varepsilon}{k}(1 + 2 + \dots + k) + \varepsilon k = (1 - \varepsilon)(k + 1)/2 + \varepsilon k \leq (1 + \varepsilon)(k + 1)/2.$$

In the NO case, in any schedule consider the collection V' of jobs (vertices) that complete on or before time $k - 1$. Now V' forms an independent set, since for any machine (hyperedge) e at most $k - 1$ unit size

operations could have been completed by time $k - 1$, and hence at most $k - 1$ vertices from e could lie in V' . Thus, if the average completion time is less than $(1 - \delta)k$, this would imply that at least δ fraction of the jobs finish by time $k - 1$ or sooner, which implies the existence of an independent set of size $\delta|V|$ in G .

Choosing k large enough and ε, δ arbitrarily small, by Theorem 1.1 it follows that the problem is hard to approximate within a factor arbitrarily close to 2 assuming the UGC.

5.2 Makespan minimization in the Assembly Line Problem

Recall the problem definition from Section 1.1. We give a reduction from Ek -Vertex-Cover. Consider a k -uniform hypergraph instance $G = (V, E)$ satisfying the properties of Theorem 1.1. We construct an instance I of the assembly line problem as follows: A vertex $v \in V$ corresponds to a job J_v in I . A hyperedge $e \in E$ corresponds to a machine m_e in I . Let E_v denote the set of hyperedges in G that contain the vertex v . For each edge in $e \in E_v$ we introduce an operation of unit size for job J_v that must be processed on machine m_e . On machine 0, the job J_v has an operation of size $k/|V|$. All other operations are of size 0.

We claim that in the YES case, the optimum makespan is at most $k + 1 + \varepsilon k$. Let V_1, \dots, V_k be the disjoint collection of vertices satisfying the properties of Theorem 1.1. Consider the following schedule. For each job $v \in V_i$ schedule all its unit size operations (except those on machine 0) during the time slot $[i - 1, i]$. For each of these jobs, we schedule their final operation on machine 0 arbitrarily during the interval $[i, i + 1]$. This gives a valid schedule for jobs corresponding to vertices in $V_1 \cup \dots \cup V_k$ because no hyperedge contains more than one vertex in any V_i and hence no two unit size operations are assigned to a machine corresponding to an hyperedge at the same time. Since $V_i \leq (1 - \varepsilon)|V|/k$, all the final operations of jobs in V_i can also be completed during $[i, i + 1]$. For the remaining jobs in $V \setminus (V_1 \cup \dots \cup V_k)$, we place their operations on machines $1, \dots, m$ arbitrarily, and schedule their final operations on machine 0 arbitrarily after time $k + 1$. Thus the makespan is at most

$$k + 1 + \varepsilon|V| \cdot k/|V| = k + 1 + \varepsilon k.$$

In the NO case, in any schedule let V' denote the subset of vertices corresponding to the jobs for which the final operation is scheduled before time $k - 1$. We claim that V' corresponds to an independent set, because for any machine (hyperedge) at most $k - 1$ operations of unit size could have been completed by time $k - 1$, and hence at most $k - 1$ vertices from the hyperedge could lie in V' . Thus, if the makespan is less than $2k - 1 - \delta k = k - 1 + (1 - \delta)|V| \cdot k/|V|$, this implies that $|V'| \geq \delta|V|$ and hence there is an independent set of size $\delta|V|$.

Choosing k large enough and ε, δ arbitrarily small, by Theorem 1.1 it follows that the problem is hard to approximate within a factor arbitrarily close to 2 assuming the UGC.

References

- [1] N. Bansal and S. Khot. An optimal long code test with one free bit. In *FOCS*, 2009.
- [2] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, June 1998.
- [3] I. Dinur, V. Guruswami, and S. Khot. Vertex cover on k -uniform hypergraphs is hard to approximate within factor $(k - 3 - \varepsilon)$. In *Electronic Colloquium on Computational Complexity, Technical Report TR02-027*, 2002.

- [4] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM Journal on Computing*, 34(5):1129–1146, 2005.
- [5] Irit Dinur and Shmuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1), 2005. Preliminary version in STOC 2002.
- [6] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.
- [7] Naveen Garg, Amit Kumar, and Vinayaka Pandit. Order scheduling models: Hardness and algorithms. In *FSTTCS*, pages 96–107, 2007.
- [8] O. Goldreich. Using the FGLSS-reduction to prove inapproximability results for minimum vertex cover in hypergraphs. ECCC Technical Report TR01-102, 2001.
- [9] E. Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM Journal on Computing*, 31(5):1608–1623, 2002.
- [10] J. Håstad. Some optimal inapproximability results. *Journal of ACM*, 48(4):798–859, 2001.
- [11] J. Holmerin. Improved inapproximability results for vertex cover on k -regular hyper-graphs. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 1005–1016, 2002.
- [12] J. Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within $2 - \epsilon$. In *Proc. 34th ACM Symp. on Theory of Computing (STOC)*, pages 544–552, 2002.
- [13] G. Karakostas. A better approximation ratio for the vertex cover problem. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1043–1050, Lisboa, Portugal, July 2005.
- [14] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, 2002.
- [15] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.
- [16] J. Leung, H. Li, and M. Pinedo. *Multidisciplinary Scheduling: Theory and Applications. Chapter Order Scheduling Models: an overview.* 2005.
- [17] J.Y.T. Leung, H. Li, and M. Pinedo. Scheduling orders for multiple product types to minimize total weighted completion time. *Discrete Appl. Math.*, 155:945–970, 2007.
- [18] Monaldo Mastrolilli, Maurice Queyranne, Andreas S. Schulz, Ola Svensson, and Nelson A. Uhan. Minimizing the sum of weighted completion times in a concurrent open shop. In *Preprint*.
- [19] E. Mossel. Gaussian bounds for noise correlation of functions. To Appear in GAFA. Current version on arxiv/math/0703683, 2009.
- [20] C. N. Potts, S. V. Sevastianov, V. A. Strusevish, L. N. Van Wassenhove, and C. M. Zwaneveld. The two-stage assembly scheduling problem: Complexity and approximation. *Operations Research*, 43:346–355, 1995.

- [21] P. Schuurman and G. J. Woeginger. Polynomial time approximation algorithms for machine scheduling: ten open problems. *Journal of Scheduling*, 2:203–213, 1999.
- [22] L. Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Proc. 33rd ACM Symp. on Theory of Computing (STOC)*, pages 453–461, 2001.