On the approximability of the maximum feasible subsystem problem with 0/1-coefficients

Khaled Elbassioni* Rajiv Raman* Saurabh Ray † René Sitters‡

October 3, 2008

Abstract

Given a system of constraints $\ell_i \leq a_i^T x \leq u_i$, where $a_i \in \{0,1\}^n$, and $\ell_i, u_i \in \mathbb{R}_+$, for $i = 1, \ldots, m$, we consider the problem Mrfs of finding the largest subsystem for which there exists a feasible solution $x \geq 0$. We present approximation algorithms and inapproximability results for this problem, and study some important special cases. Our main contributions are :

1. In the general case, where $a_i \in \{0,1\}^n$, a sharp separation in the approximability between the case when $L = \max\{\ell_1, \cdots, \ell_m\}$ is bounded above by a polynomial in n and m, and the case when it is not.

2. In the case where A is an interval matrix, a sharp separation in approximability between the case where we allow a violation of the upper bounds by at most a $(1+\epsilon)$ factor, for any fixed $\epsilon > 0$ and the case where no violations are allowed.

Along the way, we prove that the induced matching problem on bipartite graphs is inapproximable beyond a factor of $\Omega(n^{\frac{1}{3}-\epsilon})$, for any $\epsilon > 0$ unless NP=ZPP. Finally, we also show applications of Mrfs to the recently studied pricing problems.

1 Introduction

In large real-life linear programs the main difficulty often lies not in the optimization, but in the formulation of the program. If the model, which may consist of thousands of constraints turns out to be infeasible, one wishes to resolve infeasibility by deleting as few constraints as possible, or equivalently, to keep a maximum number of constraints such that the system is feasible. This motivates the study of the *maximum feasible subsystem* problem (Mfs), also known as *max satisfy* or *max satisfying linear subsystem*, defined as follows. Given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$ we wish to find a largest subset of constraints of the system $Ax \diamond b$, where $\diamond$ is an operator in $\{=, \leq, <\}$, that has a feasible solution. Weighted and unweighted versions of this problem have a number of applications in various fields such as operations research, machine learning, computational geometry, statistical analysis, and computational biology. For an overview of Mfs and its applications we refer to the recent paper by Amaldi et al. [2] and the references therein. For heuristics to solve Mfs , see [4, 17].

In this paper, we study a relaxed version of Mfs, which we call Mrfs, where each constraint has both upper and lower bounds. It is easy to see that this version subsumes Mfs$^{\diamond}$, where $\diamond \in \{=, \leq, <\}$. Two important special cases we consider are when the constraint matrix has only 0/1 entries, and where the constraint matrix is an interval matrix, i.e., it has the consecutive ones property in the rows. For these cases, we present several new approximation algorithms and inapproximability results.

1.1 Related work

Amaldi and Kann [3] first studied the approximability of Mfs and showed that the problem with inequality constraints (Mfs$^{\leq}$) is APX-hard but can be approximated within a factor of 2 using a simple greedy algorithm similar to the greedy algorithm for maximum satisfiability. They also showed that the constrained variant of the problem, where a mandatory set of constraints has to be taken in every feasible solution, is as hard to approximate as maximum independent sets in graphs, and hence cannot be approximated within m^{ϵ}, for some $\epsilon > 0$, unless NP=ZPP.

Feige and Reichman [15] showed that the problem with equality constraints (Mfs$^{=}$ with non-0/1-coefficients) cannot be approximated within a factor $n^{1-\epsilon}$ for any positive ϵ, unless NP $\subset$ BPP. The best approximation algorithm for the problem is due

*Max-Planck-Institut für Informatik, Saarbrücken, Germany; ({elbassio,rraman}@mpi-inf.mpg.de)

†Universität des Saarlandes, Saarbrücken, Germany; (saurabh@cs.uni-sb.de)

‡Department of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands; (r.sitters@tue.nl). The author is supported by a research grant from the Netherlands Organization for Scientific Research (NWO-veni grant).

to Halldórsson [19], and achieves an approximation ratio of $O(n/\log n)$. Very recently, Guruswami and Raghavendra [25] have shown that $\text{MFS}^=$ with non-0/1-coefficients cannot be approximated within a factor of $n^{1-\epsilon}$ even when we have at most 3 variables per equation.

If the matrix A is totally unimodular then the problem remains NP-hard, but it is polynomial-time solvable if A and the right-hand-side b together form a totally unimodular matrix. [24].

1.2 Our results In this paper, we consider only maximum feasible subsystems where the constraint matrix is a 0/1 matrix with non-negative solutions. Under this restriction, we prove that $\text{MFS}^=$ with all coefficients of A and b in $\{0,1\}$ cannot be approximated within a factor $O(\log^\mu n)$, for some positive constant μ. Moreover, if the right-hand side is allowed to be arbitrarily large, then the threshold increases to $\Omega(n^{1/3-\epsilon})$ for any positive ϵ, assuming NP $\neq$ ZPP. On the positive side, we consider (α,β)-approximation algorithms that find a subsystem of size at least $1/\alpha$-fraction of the optimum, and violate the upper bounds by at most a factor of β. In particular, if $L = poly(n,m)$, we present a $(\log(nL/\epsilon), 1+\epsilon)$-approximation algorithm for any $\epsilon > 0$ that runs in time that is polynomial in $(n, m, \log L, 1/\epsilon)$.

We then consider the MRFS problem on interval matrices, i.e., matrices with the consecutive ones property in the rows. Here, show that the MRFS problem is APX-hard if we do not allow any violations in the upper bounds. On the other hand, for any $\epsilon > 0$, we give a $(1, 1+\epsilon)$ algorithm that runs in quasi-polynomial time (i.e., in time $2^{polylog(n,1/\epsilon)}$) when $L = poly(n,m)$. When L is not bounded by a polynomial in n and m, we give polynomial time algorithms that guarantee a $(\sqrt{OPT}\log n)$-approximation without violations, and an $O(\log^2 n \log\log(nL/\epsilon), 1+\epsilon)$-approximation.

Table 1 summarizes our results. Most notable in these results is the strict separation in approximability (i) in the general case, when L is polynomial in n and m versus the case when L is exponential, and (ii) the interval case, when violations are allowed (where a QPTAS is possible) versus the case when no violation is allowed (APX-hardness). In the general case, the upper and lower bounds are almost tight (up to constant factors in the exponent).

In our study of the approximability of the maximum feasible subsystem problem, we obtain along the way, the result that the *maximum induced matching problem* on *bipartite graphs* cannot be approximated within a factor $n^{1/3-\epsilon}$ for any positive ϵ, unless NP = ZPP, improving the previous APX-harness result of Duckworth et al. [11].

The paper is organized as follows: We start with preliminary definitions in Section 2. In Section 3, we present our results for the MRFS problem when the constraint matrix is a general 0/1 matrix, and also present the inapproximability result for the induced matching problem on bipartite graphs. We then present approximation algorithms and inapproximability results for interval constraint matrices in Section 4. Finally, in Section 5, we discuss how our results for MRFS can be applied to the profit-maximizing pricing problem, and conclude with open questions and discussion in Section 6. Due to lack of space, we only present sketches for most of the proofs. The final version of the paper will contain complete proofs.

2 Notation and Preliminaries

The general problem we consider in this paper, the MRFS problem is defined as follows. Let $\mathcal{S} = \{S_1,\ldots,S_m\} \subseteq 2^{[n]}$ be a given (multi)set of subsets of $[n]$. For each $S \in \mathcal{S}$, let $\ell_S, u_S \in \mathbb{R}_+$ be given non-negative numbers such that $\ell_S \le u_S$, and $w_S \in \mathbb{R}_+$ be given non-negative weights. Let $a_S \in \{0,1\}^n$ be the characteristic vector of set $S \in \mathcal{S}$. The problem is to find the largest weight subset $\mathcal{T} \subseteq \mathcal{S}$, such that the system $l_S \le a_S^T x \le u_S$, $S \in \mathcal{T}$ has a feasible solution $x \ge 0$. In matrix notation, let $A \in \{0,1\}^{m\times n}$ be the matrix whose rows are $a_{S_1}^T,\ldots,a_{S_m}^T$ and let $\ell = (\ell_{S_1},\ldots,\ell_{S_m})$, $u = (u_{S_1},\ldots,u_{S_m})$, and $w = (w_{S_1},\ldots,w_{S_m})$. For a subset $\mathcal{T} \subseteq \mathcal{S}$, denote by $A[\mathcal{T}]$, the sub-matrix of A with rows a_S^T, for $S \in \mathcal{T}$, and similarly for a vector $v \in \mathbb{R}^m$, denote by $v[\mathcal{T}] = (v_S \;:\; S \in \mathcal{T})$ the restriction of v to S. Then the MRFS problem is :

(2.1)
$$\max_{\mathcal{T}\subseteq\mathcal{S}} \left\{ w(\mathcal{T}) \;:\; \{x \in \mathbb{R}^n_+ : \ell[\mathcal{T}] \le A[\mathcal{T}]x \le u[\mathcal{T}]\} \neq \emptyset \right\},$$

where $w(\mathcal{T}) = \sum_{S\in\mathcal{T}} w_S$.

In the rest of the paper, we talk mostly about the *unweighted* version of MRFS where $w(S) = 1$, $\forall S \in \mathcal{S}$. However, all our results extend also to the weighted version.

For $\alpha, \beta \ge 1$, an (α,β)-approximation is given by a pair $(\mathcal{T}, x)$ of a subset $\mathcal{T} \subseteq \mathcal{S}$, and a vector $x \ge 0$, such that $w(\mathcal{T}) \ge w(\mathrm{OPT})/\alpha$, $\ell[\mathcal{T}] \le A[\mathcal{T}]x \le \beta u[\mathcal{T}]$.

For the given family of subsets $\mathcal{S}$, let $L(\mathcal{S}) \stackrel{\text{def}}{=} \max\{\ell_S :\; S \in \mathcal{S}\}$, $\ell(\mathcal{S}) \stackrel{\text{def}}{=} \min\{\ell_S :\; S \in \mathcal{S}\}$, $U(\mathcal{S}) \stackrel{\text{def}}{=} \max\{u_S :\; S \in \mathcal{S}\}$, and $u(\mathcal{S}) \stackrel{\text{def}}{=} \min\{u_S :\; S \in \mathcal{S}\}$. We may assume without loss of generality (by scaling the bounds) that $\min\{\ell_S :\; S \in \mathcal{S},\; \ell_S \neq 0\} = 1$. We may also assume, as the following proposition shows, that $U(\mathcal{S}) \le nL(\mathcal{S})$.

PROPOSITION 2.1. *Consider an instance $(\mathcal{S}, w)$ of* MRFS. *There is an optimal solution $(\mathcal{T}, x) \subseteq 2^{\mathcal{S}} \times \mathbb{R}^n_+$,*

	Approximation		Inapproximability
	(α, β)	Running time	(α, β)
General 0/1-matrices	$(\log(nL/\epsilon), 1+\epsilon)$	$\text{poly}(n, m, \log L, \frac{1}{\epsilon})$	$(O(\log^{\mu} n), O(1))$
			$(O((\frac{\log L}{\log\log L})^{\frac{1}{3}-\epsilon}), O(1))$
Interval matrices	$(2\log n, 2)$	$\text{poly}(n, m)$	$(O(1), 1)$
	$(O(\log^2 n \log\log(nL/\epsilon)), 1+\epsilon)$	$\text{poly}(n, m, \log L, \frac{1}{\epsilon})$	
	$(1, 1+\epsilon)$	$(mL)^{O(\frac{\log L}{\epsilon}\log^2 m)}$	
	$(\sqrt{m}\log n, 1)$	$\text{poly}(n, m)$	

Table 1: Summary of positive and negative approximability results for Mrfs with 0/1-matrices: $\mu \in (0,1)$ is assumed to be some fixed constant, while ϵ is any arbitrary constant in $(0,1)$. The inapproximability result (f, g) should be interpreted as follows: under strongly believed complexity assumptions, any algorithm yielding a solution with violation $\beta = g$ cannot give a better approximation factor than f.

in which $x(S) \leq nL(\mathcal{T})$ for all $S \in \mathcal{T}$,

Proof. Let $(\mathcal{T}, y) \subseteq 2^{\mathcal{S}} \times \mathbb{R}^n_+$ be any optimal solution. We define another optimal solution $(\mathcal{T}, x)$ as follows: $x_i = L = L(\mathcal{T})$ if $y_i > L$, and $x_i = y_i$ otherwise. Then it easy to see that $\ell_S \leq x(S) \leq y(S) \leq u_S$ and $x(S) \leq |S|L$, for all $S \in \mathcal{T}$.

An interval matrix, or a matrix with consecutive ones property is a matrix with 0/1 entries such that each row contains at most one run of 1's. For ease of exposition, in later sections we view the rows of the matrix as a set of consecutive edges of a path. More precisely, let $\Pi = (V, E)$ be a path with $V = \{0, \cdots, n\}$, and edges $E = \{e_1, \cdots, e_n\}$, where $e_i = \{i-1, i\}$ (here n is the number of columns of the constraint matrix). There is a natural order on the edges of the path, viz. $e = \{u-1, u\} < e' = \{v-1, v\}$ if $u \leq v-1$. We denote by $[e, f]$ the set of edges e' such that $e \leq e' \leq f$. The set of rows of the interval matrix correspond to intervals $\mathcal{I} = \{I_1, \cdots, I_m\}$, with $I_j = [s_j, t_j] \stackrel{\text{def}}{=} \{\{s_j, s_j+1\}, \ldots, \{t_j-1, t_j\}\} \subseteq E$.
To any interval matrix, we can associate an interval graph such that the rows of the matrix correspond to the vertices of the graph with two intervals adjacent if they share an edge. The interval matrix, then represents the clique-vertex incidence matrix of this graph.

3 Mrfs with general 0/1-matrices

Amaldi and Kann [3] gave a reduction from EXACT COVER BY 3-SETS showing that $\text{Mfs}^=$ is NP-hard, even for the restricted version with $a_i \in \{0,1\}^n$ (in fact, their reduction can be used to show APX-hardness). Here we give a stronger inapproximability result by a reduction from the UNIQUE-COVERAGE problem [9].

Theorem 3.1. *Assuming $NP \not\subseteq \text{BPTIME}(2^{n^{\epsilon}})$ for an arbitrary small $\epsilon > 0$, there is a constant $\sigma(\epsilon)$ such that there is no (α, β)-approximation algorithm for* Mrfs*, with $\alpha = O(\log^{\mu} n)$, $\beta = O(\log^{\lambda} n)$ and $\sigma(\epsilon) = \lambda + \mu$, even if $\ell = u = \underline{1}$, where $\underline{1}$ is the vector of all ones.*

Proof. (*Sketch*) We can model the UNIQUE-COVERAGE problem as an instance of Mrfs. We then use a probabilistic argument similar to in Lemma A.1 in [9], to show that an (α, β)-approximation for Mrfs, implies a $2e\alpha\beta$-approximation (where e is the base of the natural logarithm) for UNIQUE-COVERAGE.

When $L = \mathit{poly}(n, m)$, we complement the hardness result above, with a logarithmic approximation algorithm.

Theorem 3.2. *Given any instance of* Mrfs*, there exists a $(\frac{\log(nL)}{\epsilon}, 1+\epsilon)$-approximation, whose running time is bounded by $\text{poly}(n, m, \log L, \frac{1}{\epsilon})$, for any $\epsilon > 0$.*

Proof. Let $R_{min} = \min\{\ell_S/|S| : S \in \mathcal{S}, \ell_S \neq 0\}$, $R_{max} = \max\{\ell_S/|S| : S \in \mathcal{S}\}$ and let $h = \lceil \log_{1+\epsilon}(R_{max}/R_{min}) \rceil$. Partition $\mathcal{S}$ into $h+1$ groups : $G_0 = \{S \in \mathcal{S} : \ell_S = 0\}$, and $G_i = \{S \in \mathcal{S} : (1+\epsilon)^{i-1} R_{min} \leq \ell_S/|S| < (1+\epsilon)^i R_{min}\}$, for $i = 1, \ldots, h$. Clearly, we can satisfy all the inequalities in G_0 by setting $x = 0$. Likewise, setting $x = (1+\epsilon)^{k+1} R_{min}\underline{1}$. Clearly, we can satisfy all the inequalities in one of the groups G_i, $i = 1, \cdots, h$, possibly violating the upper bounds by at most an ϵ factor. Since one of the groups G_i has size at least $OPT/(h+1)$, and $R_{max} \leq L$ and $R_{min} \geq 1/n$, the theorem follows.

It may seem that the dependence of the running time on L is an artifact of the approximation algorithm. However, that is not the case as we now show that when L is not bounded by a polynomial in n or m, Mrfs is inapproximable beyond $n^{\frac{1}{3}-\epsilon}$ for any $\epsilon > 0$ unless $NP = ZPP$. We start by proving an inapproximability result for the *maximum induced matching* problem on bipartite graphs. We then define the *maximum semi-induced* *matching* problem on bipartite graphs, and observe that the same reduction implies inapproximability for the semi-induced matching problem. We then use this reduction to show hardness of approximation for MRFS.

DEFINITION 3.1. *Maximum Induced Matching(*MIMP*) Given a graph $G = (V, E)$, and induced matching is a matching M, such that the graph induced by the vertices in M is a matching. i.e., if $\{u, v\}, \{u', v'\} \in M$, then none of $\{u, v'\}, \{u', v\}, \{u, u'\}, \{v, v'\} \in E$. The maximum induced matching problem is to find an induced matching of maximum cardinality.*

Induced matchings are well-studied in discrete mathematics especially as a subtask of finding a strong edge-colorings (see e.g. [14] and [23]). Duckworth et al. [11] studied the hardness of MIMP. On general graphs, they showed MIMP to be as hard to approximate as the maximum independent set problem, while on bipartite graphs they showed that the problem is APX-hard. See [11] for a recent overview on induced matchings. Here we prove the following stronger hardness result.

THEOREM 3.3. *The maximum induced matching problem on bipartite graphs with n vertices cannot be approximated within a factor $O(n^{\frac{1}{3}-\epsilon})$, for any $\epsilon > 0$, unless $NP = ZPP$.*

Proof. We reduce from the maximum independent set problem in general graphs (MIS). Given an instance $G = (V, E)$ of the maximum independent set problem, we define a bipartite graph $H = (W \cup W', F_V \cup F_E)$ on $2n := 2K|V|$ vertices, where K is a large number to be specified later. For each vertex $v_i \in V$ we define vertices $v_{i1}, v_{i2}, \ldots, v_{iK}$ in W and vertices $v'_{i1}, v'_{i2}, \ldots, v'_{iK}$ in W'. For every vertex $v_i \in V$ there is an edge between the vertices v_{ik} and v'_{ik} for every $k \in \{1, 2, \ldots, K\}$. Denote this set of edges by F_V. For every edge $\{v_i, v_j\} \in E$, we add an edge between v_{ik} and v'_{jl}, and between v_{jk} and v'_{il} for every pair of indices $k, l \in \{1, 2, \ldots, K\}$. Hence, for every edge in E we define $2K^2$ edges in F. Denote this set of edges by F_E. This completes the reduction.

Let $S \subseteq V$ be an independent set in G. Then, there is an induced matching of size $K|S|$ in the graph H. This matching consists of the edges $\{\{v_{ik}, v'_{ik}\} \mid v_i \in S, k \in \{1, \ldots, K\}\}$, i.e., the edges in F_V that correspond to the vertices in S. This gives a lower bound on the size, $\mathrm{O}PT_{MIMP}$, viz.

$$\mathrm{O}PT_{MIMP} \geq K \cdot \mathrm{O}PT_{MIS}. \quad (3.2)$$

Now let $M \subseteq F_V \cup F_E$ be an induced matching in H. First, note that only a limited number of the edges in M can be in F_E. i.e., $|M \cap F_E| \leq 2|E|$, since for every $\{v_i, v_j\} \in E$ the matching can contain only one edge from the K^2 edges $\{\{v_{ik}, v'_{jl}\} \mid k, l \in \{1, 2, \ldots, K\}\}$ and one from the set $\{\{v_{jk}, v'_{il}\} \mid k, l \in \{1, 2, \ldots, K\}\}$. Hence, $|M \cap F_V| = |M| - |M \cap F_E| \geq |M| - 2|E|$. Second, note that two edges in $M \cap F_V$ that correspond to different vertices in V must correspond to independent vertices in G, since otherwise these edges are connected by an edge in H. Hence, there must be an independent set in G of size at least $|M \cap F_V|/K \geq |M|/K - 2|E|/K$. If we choose $K \geq 2|E|$, we get $\mathrm{O}PT_{MIS} \geq \mathrm{O}PT_{MIMP}/K - 1$. Combined with (3.2) we get

$$\mathrm{O}PT_{MIMP}/K \geq \mathrm{O}PT_{MIS} \geq \mathrm{O}PT_{MIMP}/K - 1.$$

Since the maximum independent set is hard to approximate within a factor $|V|^{1-\epsilon}$, we conclude that $\mathrm{O}PT_{MIMP}/K$, and consequently $\mathrm{O}PT_{MIMP}$, is hard to approximate within this factor. The number of vertices in H is $2n = 2K|V| = O(|V|^3)$. Hence, $|V|^{1-\epsilon} = \Omega((2n)^{(1-\epsilon)/3})$, and the theorem follows.

DEFINITION 3.2. *Maximum Semi-induced matching (*SIM*) Let $G = (U, V, E)$ be a bipartite graph, with a total order on the elements of U, A matching $M \subseteq E$ is a semi-induced matching if for any $u_i, u_j \in U$ that are in the matching M, with $i < j$, there is no edge $\{u_j, v\} \in E$, where v is a neighbor of u_i in M. The maximum semi-induced matching problem is to find a semi-induced matching of maximum cardinality.*

We can define a weighted version of SIM in the natural way. There is a weight function $w : E \to \mathbb{R}_+$ on the edges of G, and the problem is to find a semi-induced matching of maximum weight. It is not hard to see that the reduction in Theorem 3.3 also shows inapproximability for the semi-induced matching problem on bipartite graphs.

THEOREM 3.4. *The maximum semi-induced matching problem on bipartite graphs cannot be approximated within a factor of $O(n^{1/3-\epsilon})$ for any $\epsilon > 0$ unless NP=ZPP.*

We are now ready to prove hardness of approximation for MRFS.

THEOREM 3.5. *Unless $NP = ZPP$, there is no (α, β)-approximation algorithm for* MRFS*, with $\alpha = O(n^{\frac{1}{3}-\epsilon})$ and $\beta = O(1)$, for any $\epsilon > 0$.*

Proof. (*Sketch*) Given an instance $G = (V, E)$ of the maximum independent set problem, define the same instance H of maximum induced matching as in Theorem 3.3. Next, we define from H an instance of MRFS.

Given the graph H, denote the vertices of W by w_i ($i = 1 \dots n$), where w_i is the i^{th} vertex in the sequence $v_{11}, v_{12}, \dots, v_{1K}, v_{21}, \dots, v_{|V|K}$. Similarly, denote the vertices of W' by w'_i ($i = 1 \dots n$). Let $a_{ij} = 1$ if there is an edge between w_i and w'_j in H, and let $a_{ij} = 0$ otherwise. Now consider the following system $\mathcal{S}$ of equalities:

$$\sum_{j=1}^{n} a_{ij} x_j = (nB)^i, \qquad \text{for } i \in \{1, 2, \dots, n\},$$
$$x_j \geq 0, \qquad \text{for } j \in \{1, 2, \dots, n\},$$

where $B > \beta$. This completes the reduction.

We show that if there is an independent set $S \subseteq V$, then we can obtain a feasible system of size at least $K|S|$ by setting the variables corresponding to the vertices in S to $(nB)^i$, and the others to 0. To show the reverse direction, we show that an optimal solution to the Mrfs instance corresponds to a semi-induced matching in H of the same size. Using this, and a proof similar to Theorem 3.3 the result follows.

4 Maximum feasible subsystems with interval matrices

We now turn our attention to the Mrfs problem on interval matrices. Recall from Section 2 that we view the rows of the constraint matrix as consecutive sets of edges of a path. We start by showing that for L polynomially bounded, Mrfs for interval matrices admits a QPTAS if we allow any $\epsilon > 0$ factor violation in the upper bounds. Thus, it is unlikely that we can obtain an APX-hardness result for (α, β)-approximations like in the general case.

Theorem 4.1. *Consider an instance of* Mrfs *with an interval matrix A, and $\ell, u \in \mathbb{R}^n_+$ such that $L \leq$ quasi-poly(m). Then we can find a $(1, 1+\epsilon)$-approximation in quasi-polynomial time, for any $\epsilon > 0$.*

Proof. (*Sketch*) The algorithm is similar to the one in [12] for the highway problem. We use a divide and conquer strategy which starts by picking an edge in the middle and guesses the points at which the feasible solution x at optimality increases by factors of $(1+\epsilon)$ relative to that middle edge. Having guessed such "increment points", the algorithm picks a *superset* of the optimal set of intervals containing the middle edge, then recurses independently on the two subproblems to the left and right of the middle edge, making sure that all subsequent guesses are consistent with the initial guess.

On the other hand, if we do not allow any violations in the upper bounds, the problem becomes APX-hard. Thus Theorem 4.1 is the best possible (modulo improving the running time to polynomial). The proof of Theorem 4.2 is presented at the end of this section.

Theorem 4.2. *There exists a constant $\alpha > 1$ such that, unless if $P = NP$, there is no $(\alpha, 1)$-approximation algorithm for* Mrfs*, even if $L =$ poly(n, m), and the constraint matrix A is an interval matrix representing a clique.*

We next present polynomial time approximation algorithms for the problem, both with and without violations allowed. If no violations are allowed, the best guarantee is a $(\sqrt{OPT} \log n)$-approximation algorithm, while allowing violations of $(1+\epsilon)$, for any $\epsilon > 0$, we can guarantee a poly-log approximation factor. Note that these algorithms do not depend on whether L is polynomially bounded in (n, m).

4.1 Approximation algorithms

We start with a proposition that will be used in the main theorem. This was proved by Broersma, et al. [8].

Proposition 4.1. *([8]) Given an interval graph $G = (V, E)$ on n vertices, it can be partitoned into at most $\lfloor \log n \rfloor + 1$ sets, each of which is a disjoint union of cliques.*

Theorem 4.3. *Consider an instance of* Mrfs *with an interval matrix A, and any $\ell, u \in \mathbb{R}^n_+$. Then we can find:*

(i) a $(\sqrt{OPT} \log n, 1)$-approximation in poly(n, m) *time.*

(ii) a $(2 \log n, 2)$-approximation in poly(n, m) *time,*

(iii) a $(2 \log^2 n \log\log_{1+\epsilon}(nL)/\log(1+\epsilon)), 1+\epsilon)$-approximation in poly$(n, m, \log L, \frac{1}{\epsilon})$ *time, for any $\epsilon > 0$.*

Proof. (*Sketch*) (i) Assume the instance is a clique with $l_I = u_I$ for all $I \in \mathcal{I}$. We obtain a $\sqrt{OPT}$-approximation by the observation that a clique with l_I monotonic non-increasing in order of leftmost edges, or monotonic non-decreasing in rightmost edges can be realized, and finding the largest such set can be done using dynamic programming. The approximation claimed, then follows from the Erdös Szekeres theorem [13], and noting that OPT contains no *bad-containment pairs*, i.e., a pair of intervals with $I \subseteq J$, and $l_I > l_J$. The case where $l_I \leq u_I$ can be reduced to the case with equality, by replacing each interval with a set of $O(m)$ intervals that form pairwise bad-containments. Using the partitioning from Proposition 4.1, solving the clique problem for each of the disjoint cliques within a set, and selecting the largest, the result follows.

(ii) The result follows from using Proposition 4.1, and solving the problem for each clique. A $(2, 2)$-approximation for a clique is obtained by noting that

we can split OPT into two sets - those intervals that have at least half their length on the left, and those with at least half their lenght on the right.
(iii) We start with some definitions required for the rest of the proof.

DEFINITION 4.1. *Variable-Clique. A* variable-clique *is a subset of intervals, all of which have at least an edge in common. i.e.,* $\mathcal{I}' \subseteq \mathcal{I}$ *is a* variable-clique *if* $[\delta_v(\mathcal{I}'), \delta'_v(\mathcal{I}')] \stackrel{\text{def}}{=} \bigcap_{I\in\mathcal{I}'} I \neq \emptyset$.

DEFINITION 4.2. *Bound-Clique. A* bound-clique *is a subset of intervals whose length ranges have a common intersection. i.e.,* $\mathcal{I}' \subseteq \mathcal{I}$ *is a* bound-clique *if* $[\delta_b(\mathcal{I}'), \delta'_b(\mathcal{I}')] \stackrel{\text{def}}{=} \bigcap_{I\in\mathcal{I}'} [\ell_I, u_I] \neq \emptyset$.

Proof of 4.3-(III): Let us call a collection of bound-cliques $\mathcal{I}_1, \ldots, \mathcal{I}_r \subseteq \mathcal{I}$ *well-separated* if

(W) for $i, j \in [r]$, $i < j$, and $I \in \mathcal{I}_i, J \in \mathcal{I}_j$, we have $n \cdot u_I < \ell_J$.

PROPOSITION 4.2. *Let* $\mathcal{I}_1, \ldots, \mathcal{I}_r \subseteq \mathcal{I}$ *be a set of well-separated bound-cliques, and* $x \in \mathbb{R}^n_+$ *be a vector satisfying* $\ell_I \le x(I) \le u_I$*, for all* $I \in \cup_{j=1}^r \mathcal{I}_j$*. For* $i < j$*,* $I \in \mathcal{I}_i$ *and* $J \in \mathcal{I}_j$*, if* $e \in \text{argmax}_{f\in J}\{x_f\}$*, then* $e \notin I$.

Proof. By the definition of e we have $x_e \ge \ell_J/|J| \ge \ell_J/n$. By (W), we have $x(I) \le u_I < \ell_J/n \le x_e$, and hence $e \notin I$.

We start by solving a special case of MRFS where the given instance $(\mathcal{I}, w)$, $\mathcal{I} = \cup_{j=1}^r \mathcal{I}_j$ is a variable-clique, which is also a disjoint union of well-separated bound-cliques $\mathcal{I}_1, \ldots, \mathcal{I}_r$. From the proof of 4.3 (II), we obtain a $(2, 2)$-approximation by solving two instances where all intervals start at the same edge. If we further assume that the bound cliques are *well-seperated*, then we can achieve the same factor with a violation of at most $(1 + \frac{1}{n-1})$ in the upper bounds.
For the MRFS instance $(\mathcal{I}, w)$, we define two SIMPLE instances $(\mathcal{I}', w')$ and $(\mathcal{I}'', w'')$ as follows: Let $\delta_v(\mathcal{I}) = \{u, u+1\}$, and for each $I = [s, t] \in \mathcal{I}$ define $I' = [s, u]$ and $I'' = [u, t]$ to be left and right sub-intervals of I, respectively. Then we set $\mathcal{I}'_j = \{I' : I \in \mathcal{I}_j\}$, $\mathcal{I}''_j = \{I'' : I \in \mathcal{I}_j\}$, and $w'(I') = w''(I'') = w(I)$ for all $I \in \mathcal{I}$.

LEMMA 4.1. *Consider a well-separated instance of* MRFS *described by a set of bound-cliques* $(\mathcal{I} = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r, w)$*, such that* $\mathcal{I}$ *is a variable-clique, and let* $(\mathcal{I}', w')$ *and* $(\mathcal{I}'', w'')$ *be the corresponding* SIMPLE *instances described above. Then any polynomial-time algorithm that returns the maximum of the optima of these instances gives a* $(2, 1 + \frac{1}{n-1})$*-approximation for the* MRFS *instance.*

Proof. Consider an optimum solution $(\text{O}PT, x)$ of the given instance of MRFS. For every $I \in \text{O}PT \cap \mathcal{I}_j$, there exists $e_I \in I$ such that $x_{e_I} \ge \ell(\mathcal{I}_j)/n$. Let $e'_j = \max\{e_I : I \in \mathcal{I}_j \cap \text{O}PT,\ e_I < \delta_v(\mathcal{I})\}$ (i.e. the right-most edge to the left of the intersection point of all intervals in $\mathcal{I}$) and $e''_j = \min\{e_I : I \in \mathcal{I}_j \cap \text{O}PT,\ e_I \ge \delta_v(\mathcal{I})\}$, for $j = 1, \ldots, r$. By definition, any interval $I \in \mathcal{I}_j \cap \text{O}PT$ must either contain e'_j or e''_j. Define $\text{O}PT' = \{I \in \text{O}PT :\ e'_j \in I,$ for some $j \in [r]\}$ and $\text{O}PT'' = \text{O}PT \setminus \text{O}PT'$, and suppose without loss of generality that $w(\text{O}PT') \ge w(\text{O}PT)/2$.

Define $x' \in \mathbb{R}^n_+$ as follows: $x'_{e'_j} = L(\mathcal{I}_j)$, and $x'_e = 0$ if $e \neq e'_j$ (If $e'_j = \emptyset$, all $x_e = 0$ for $I \in \mathcal{I}_j$). Define also, for each j such that $e'_j \neq \emptyset$, $\mathcal{I}'_j = \{I \in \mathcal{I}_j \cap \text{O}PT :\ e'_j \in I$ and $e'_i \notin I$ for all $i > j\}$, and let $\mathcal{I}' = \cup_{j:\mathcal{I}'_j \neq \emptyset} \mathcal{I}'_j$. We claim that $(\mathcal{I}', x')$ is a $(2, 1 + \frac{1}{n-1})$-approximation of the given instance of MRFS. To see this, fix $j \in [r]$ and consider any $J \in \mathcal{I}'_j$. Trivially, $x'(J) \ge \ell_J$. Moreover, by Proposition 4.2, for any $i > j$, we must have $e'_i \notin J$, and thus $e'_j > e'_i$. In particular,

$$
\begin{aligned}
x'(J) &\le \sum_{i\le j} L(\mathcal{I}_i) \\
&\le L(\mathcal{I}_j)(1 + \frac{1}{n} + \frac{1}{n^2} + \cdots) \qquad (4.3)\\
&< (1 + \frac{1}{n-1})L(\mathcal{I}_j) \le (1 + \frac{1}{n-1})u_J,
\end{aligned}
$$

where the second inequality follows from (W), and the last one follows from the fact that $\mathcal{I}_j$ is a bound-clique. Thus the set $\{e'_1, \ldots, e'_r\}$ is a feasible solution for the constructed instance $(\mathcal{I}', w')$ of SIMPLE with weight at least $w(\text{O}PT)/2$.

Conversely, given any feasible solution $\{e'_1, \ldots, e'_r\}$ to the SIMPLE instance $(\mathcal{I}', w')$, with weight $w'(\text{O}PT')$, we can construct a solution with exactly the same weight to the MRFS instance $(\mathcal{I}, w)$ as follows. For each $j \in [r]$ if $e'_j \neq \emptyset$, then we set $x_{e'_j} = L(\mathcal{I}_j)$ and $\mathcal{I}'_j = \{J \in \mathcal{I}_j : e'_j \in I$ and $e'_i \notin I$ for all $i > j\}$. Finally we define $\mathcal{I}' = \cup_{j=1}^r \mathcal{I}'_j$. Then $w(\mathcal{I}') = w'(\text{O}PT')$, and we can argue as in inequality (4.3) that for $I \in \mathcal{I}'$, x violates any right bound u_I by at most $u_I \cdot 1/(n-1)$. Thus $(\mathcal{I}', x)$ is a $(2, 1 + 1/(n-1))$-approximation for the MRFS instance $(\mathcal{I}, w)$.

An optimal solution to the SIMPLE instance can be computed by dynamic programming, similar to the dynamic program used for part (ii).
Now we are ready to finish the proof of Theorem 4.3-(iii). We begin by rounding down (respectively up) all the $\ell'_I s$ (respectively, the u_I's) to the nearest power of $(1 + \epsilon)$. In doing so we only lose a factor of $(1 + \epsilon)^2$ in the upper bounds in the final solution.

We denote that rounded set of intervals by $\tilde{\mathcal{I}}$. This gives $h \leq \log P/\log(1+\epsilon)+1$ points $p_0 = 0, p_1 = 1, p_2 = (1+\epsilon), \ldots, p_{h+1} = (1+\epsilon)^h$ on the real line, at which the bounds of intervals from $\tilde{\mathcal{I}}$ can begin or end. We next partition the set of intervals in $\tilde{\mathcal{I}}$ into $k \leq \log h$ groups $G_1, G_2, \ldots$, where $G_1 = \{I \in \tilde{\mathcal{I}} : [\ell_I, u_I] \ni p_{\lfloor h/2 \rfloor}\}$, $G_2 = \{I \in \tilde{\mathcal{I}} \setminus G_1 : [\ell_I, u_I] \cap \{p_{\lfloor h/4 \rfloor}, \lfloor p_{3h/4} \rfloor\} \neq \emptyset\}$, $G_3 = \{I \in \tilde{\mathcal{I}} \setminus (G_1 \cup G_2) : [\ell_I, u_I] \cap \{p_{\lfloor h/8 \rfloor}, \lfloor p_{3h/8} \rfloor, \lfloor p_{5h/8} \rfloor, \lfloor p_{7h/8} \rfloor\} \neq \emptyset\}, \ldots$ We solve k independent problems one for each group of intervals G_i, $i = 1, \ldots, k$, and return from among these the solution of maximum weight.

Fix a group G_i, $i \in [k]$. Note that the intervals in G_i have the property that they can be decomposed into a number, say r, of disjoint bound-cliques $\mathcal{I}_1, \ldots, \mathcal{I}_r$: $[\ell_I, u_I] \cap [\ell_J, u_J] = \emptyset$ for $I \in \mathcal{I}_i, J \in \mathcal{I}_j$, $i \neq j$. We may assume thus that these cliques are numbered from left to right, i.e., $\delta_b(\mathcal{I}_1) < \delta_b(\mathcal{I}_2) < \cdots < \delta_b(\mathcal{I}_r)$. It is easy to see, by the way these cliques are constructed that

$$U(\mathcal{I}_i) \leq (1+\epsilon)^{j-i-1}\ell(\mathcal{I}_j), \text{ for } i \leq j-1.$$

In particular, we get $U(\mathcal{I}_i) \leq n\ell(\mathcal{I}_j)$ for $j - i - 1 \geq \frac{\log n}{\log(1+\epsilon)}$. This observation allows us to further partition G_i into $\frac{\log n}{\log(1+\epsilon)}$ subsets $G_i^1, G_i^2, \ldots$, such that the cliques in each subset are well-separated. Finally we further decompose each such subset G_i^j into $\log n$ variable-cliques as in Proposition 4.1. This gives us a set of instances which can be solved using Lemma 4.1, and the final solution will be the maximum of the obtained solutions. The bound on the approximation ratio follows.

4.2 Proof of Theorem 4.2 In this section we prove that $\textsc{Mfs}^=$ with an interval constraint matrix is APX-hard, and hence does not admit a PTAS. It is easier to visualize the $\textsc{Mfs}^=$ problem with an interval matrix as that of drawing intervals with specified lengths given an order on their end-points. Hence, we state the hardness result in terms of drawing intervals with a given linear order on their end-points. We call the problem MID. We prove this by presenting a *gap-preserving* reduction from MAX-2-SAT, which was shown to be APX-hard by Håstad [21]. The reduction consists of a gadget for each variable and a gadget for each clause, where each gadget is a collection of intervals with specified lengths, and a linear order on their end-points specified. Let the MAX-2-SAT instance consists of n variables $\{x_1, \cdots, x_n\}$ and m clauses $\{C_1, \cdots, C_m\}$, with the variables numbered $1, \cdots, n$. We now describe the construction of the variable and clause gadgets.

Variable Gadget : A variable gadget consists of two sets X and Y of intervals. Set X consists of the seven intervals $\{a, \cdots, g\}$, and Set Y, the five intervals $\{p, \cdots, t\}$. The interval z is *not* part of the variable gadget, but is common to all the gadgets. The intervals c, d, e, f are called *short intervals*, and the rest are the *long intervals*. In the proof, we have interval z set to have length $\delta = 0$, but the proof goes through for any $\delta < 1/2$.

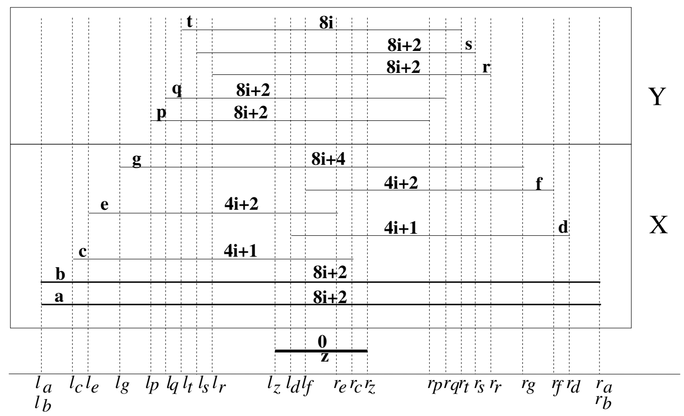

Figure 1: The variable gadget for variable x_i consists of 12 intervals. The lenghts of the intervals are a function of the index of the variable. The order of the end-points is shown by dotted lines. The interval z is *not* part of the gadget.

Figure 1 shows the gadget for variable x_i. The order of the end-points of the intervals corresponding to the variable gadget are shown by the dotted vertical lines. The order of the left end-points is : $l_a = l_b \leq l_c \leq l_e \leq l_g \leq l_p \leq l_q \leq l_t \leq l_s \leq l_r \leq l_z \leq l_d \leq l_f$, and the right end-points are ordered as : $r_e \leq r_c \leq r_z \leq r_p \leq r_q \leq r_t \leq r_s \leq r_r \leq r_g \leq r_f \leq r_d \leq r_a = r_b$. Note that some of the intervals are mutually exclusive. For example, the pair of intervals c and e cannot both be realized, since $L(e) > L(c)$, and $l(c) \leq l(e) \leq r(e) \leq r(c)$. We call such pairs *bad-containment* pairs. We now show that there are exactly two optimal solutions for MID for a variable gadget.

LEMMA 4.2. *If interval z is required to be realized, there are two optimal solutions to* MID *for a variable gadget, both consisting of* 8 *intervals (excluding z).*

Proof. (*Sketch*). From the Set X, notice that we can realize either $\{a, b, c, d\}$, or $\{e, f, g\}$ if the interval z is required to be realized. If we select $\{a, b, c, d\}$, then the interval t in Set Y can not be realized. We then realize the intervals $\{a, b, c, d, p, q, r, s\}$. If we instead realize $\{e, f, g\}$, then all intervals $\{p, q, r, s, t\}$ can be realized. In any other case, we can only realize fewer than 8 intervals.

We call the optimal solutions $\{a, b, c, d, p, q, r, s\}$ and $\{e, f, g, p, q, r, s, t\}$ TRUE and FALSE configurations respectively.

Remark 1: Note that in both the TRUE and FALSE configurations, the pair $\{p, q\}$ and the pair $\{r, s\}$ have their left and right end-points aligned. We will use this fact to bind the end-points of the clause gadgets.

Let m_i denote the number of clauses that variable i appears in. A variable gadget consists of $2m_i$ copies of each interval of the initial gadget shown in Figure 1. The interval z is replicated $2m$ times, where m is the number of clauses. Having described the gadget for a variable, we now describe how these are combined. For a variable x_i, let $l(x_i)$ denote the set of left end-points of all intervals of the gadget of x_i *except* d and e, and let $r(x_i)$ denote the set of right end-points of all the intervals corresponding to x_i, except c and e. The order of the end-points is then :

$$\begin{aligned} &l(x_n) \preceq l(x_{n-1}) \preceq \cdots \leq l(x_1) \preceq l(z) \\ &\preceq r(z) \preceq r(x_1) \cdots \preceq r(x_{n-1}) \preceq r(x_n) \end{aligned}$$

The left end-point of the interval d is the same for all x_i, and the same holds for f and the right end-points of c and e. Since the set of all left end-points precede the set of all right end-points, it is clear that the collection of intervals induce a clique.

Remark 2: Note that the end-point orders and the lengths of the intervals are assigned in such a way that the subset of intervals selected from x_i to be realized does not affect the selection of a realizable subset from x_j, $j \neq i$.

Clause Gadget : Each clause gadget is a pair of intervals that form a bad containment. There are 4 types of clause gadgets corresponding to the 4 types of clauses, $(x_i \vee x_j)$, $(\overline{x_i} \vee x_j)$, $(x_i \vee \overline{x_j})$, $(\overline{x_i} \vee \overline{x_j})$. The intervals of a clause with variables x_i and x_j have their left and right end-points lie between the left and right end-points of the p, q, r and s intervals of the two variable gadgets. The gadget for each clause is shown in Table 2.

The intuitive idea behind the reduction is that in any optimal solution to Mid all the z intervals are realized, and this forces all variable gadgets to be in either a TRUE or FALSE configuration. This determines a specific length between the ends of some of the intervals so that exactly one of the clause intervals is realized if and only if the corresponding clause is satisfied. Since all clause gadgets contain the interval z, the set of all left end-points still precede the set of all right end-points, and the resulting intersection graph is still a clique.

Lemma 4.3. *For a clause gadget, of a clause with variables x_i, and x_j, if the variable gadgets are in a TRUE or FALSE configuration, and the z intervals are all realized, then exactly one interval P or Q of the clause gadget can be realized if and only if this assignment to the variables satisfies the corresponding clause.*

Now we can state the main theorem.

Theorem 4.4. *Given an instance of* Mid *with m intervals, there are constants $\epsilon > \delta > 0$ such that it is NP-hard to distinguish between the case where the optimal solution has size at least $(1-\epsilon)m$, and the case where the optimal solution has size at most $(1-\epsilon-\delta)m$.*

Proof. (*Sketch*) If the MAX-2-SAT instance has k satisfied clauses, choose a TRUE configuration for the gadgets of variables set to TRUE, FALSE for the rest of the variables. It follows from the previous discussion and Lemma 4.3 that all the variable gadgets, along with the z intervals and one interval for each satisfied clause can be realized. This yields a solution of size at least $34m + k(8\sum_{i=1}^{n} 2m_i + 2m + k)$.

For the reverse direction, first note that for any interval we can realize either all, or none of the copies of the interval. From Lemma 4.2, and the observation that each clause is a bad-containment pair, we can realize at most $33m$ intervals from the clause and variable gadgets combined. Hence, realizing the z intervals and the maximum number of intervals from the variable gadgets is necessary to get the count up to $34m$, and the rest come from the satisfied clauses. Noting that the total number of intervals in the reduction is $O(m)$, the claim follows.

5 Application to pricing problems

The pricing problem is a natural problem arising in several applications. The problem has recently attracted a lot of attention, and several authors have studied the complexity of this problem, and several special cases [1, 5, 6, 7, 22, 10, 12, 16, 18, 20]. The problem essentially is of setting prices for goods on sale, so as to maximize the profit obtained from selling the goods to customers. In this section, we show how (α, β)-approximation algorithms for Mrfs are related to the pricing problem with *single minded* customers. i.e., each customer is interested in buying exactly one bundle (subset) of the goods and will definitely buy her bundle if the total price of the bundle is within her budget. The problem is formally defined as follows: Let E be a finite set of n items and $\mathcal{S} = \{S_1, \ldots, S_m\} \subseteq 2^E$ be a (multi)set of subsets of E. Set S_j represents the bundle customer $j \in [m] \stackrel{\text{def}}{=} \{1, \ldots, m\}$ is interested in buying. With each set $S_j \in \mathcal{S}$, we are given a non-negative number B_{S_j} representing the budget of customer j, i.e., the

Clause	End-point order	Length
$(x_i \vee x_j)$	$l(s_i) \le l(P) \le l(Q) \le l(t_i) \le r(p_j) \le r(Q) \le r(P) \le r(q_j)$	$L(P) = 4i + 4j + 1$ $L(Q) = 4i + 4j + 2$
$(\overline{x_i} \vee x_j)$	$l(p_i) \le l(P) \le l(Q) \le l(q_i) \le r(p_j) \le r(Q) \le r(P) \le r(q_j)$	$L(P) = 4i + 4j + 2$ $L(Q) = 4i + 4j + 3$
$(x_i \vee \overline{x_j})$	$l(s_i) \le l(P) \le l(Q) \le l(t_i) \le r(s_j) \le r(Q) \le r(P) \le r(r_j)$	$L(P) = 4i + 4j + 2$ $L(Q) = 4i + 4j + 3$
$(\overline{x_i} \vee \overline{x_j})$	$l(p_i) \le l(P) \le l(Q) \le l(q_i) \le r(s_j) \le r(Q) \le r(P) \le r(r_j)$	$L(P) = 4i + 4j + 3$ $L(Q) = 4i + 4j + 4$

Table 2: This table shows the lengths and end-point orders, and the lengths of the pair of intervals making up a clause gadget for the four different kinds of clauses.

maximum amount of money she is willing to pay for her bundle. Given a *price* vector $p \in \mathbb{R}^n_+$ of the items, a customer j will *definitely buy* her bundle, if it is priced within her budget, and will pay $p(S_j) \stackrel{\text{def}}{=} \sum_{i \in S_j} p_i$. The objective of this problem, denoted Pp, is to assign a non-negative number (price) $p_i \in \mathbb{R}_+$ to each item $i \in E$, and to find a subset $\mathcal{S}' \subset \mathcal{S}$, so as to maximize $\sum_{S \in \mathcal{S}'} p(S)$, subject to the budget constraints $p(S) \le B_S, \quad \text{for all } S \in \mathcal{S}'$.

We show that at the loss of a factor of 4 in the approximation ratio, one can solve Pp as a special instance of Mrfs.

Proposition 5.1. *If there is an (α, β)-approximation algorithm for* Mrfs*, then there exists a $(4\alpha\beta)$-approximation algorithm for* Pp.

Proof. Consider an instance $(E, \mathcal{S}, B)$ of the pricing problem. Let $\text{O}PT$, p denote respectively, the optimal solution and corresponding pricing for Pp. Then, there exists a pricing p' and a subset of bundles $\mathcal{S}' \subseteq \mathcal{S}$ such that $B_S/2 \le p'(S) \le B_S$ for all $S \in \mathcal{S}'$, and $p'(\mathcal{S}') \ge p(\text{O}PT)/2$. Indeed, such a pricing can be found by iterating the following two steps: 1. let $\mathcal{S}_1 = \{S \in \mathcal{S} \ : \ p(S) \le B_S/2\}$ and $\mathcal{S}_2 = \{S \in \mathcal{S} \ : \ p(S) > B_S/2\}$; 2. if $p(\mathcal{S}_1) \ge p(\text{O}PT)/2$, then set $p_i \leftarrow 2p_i$ for $i \in E$, and $\text{O}PT \leftarrow \mathcal{S}_1$; until $p(\mathcal{S}_2) \ge p(\text{O}PT)/2$ is obtained. Clearly, at each iteration $\mathcal{S}_2 \ne \emptyset$ (since otherwise the current pricing is not optimal), and hence the procedure must terminate with a pricing p' and a set of bundles $\mathcal{S}'$ satisfying the claim.

Now, construct an instance of Mrfs by setting the inequality $B_S/2 \le x(S) \le B_S$ of weight $B_S/2$ for each bundle $S \in \mathcal{S}'$. Let $\text{O}PT'$ be the optimum solution of the constructed Mrfs. Then setting $x(S) \leftarrow p'(S)$ for all $S \in \mathcal{S}'$ gives a feasible subsystem for Mrfs with total weight $B(\mathcal{S}')/2 \ge p'(\mathcal{S}')/2 \ge p(\text{O}PT)/4$, and thus $w(\text{O}PT') \ge p(\text{O}PT)/4$.

Suppose that $(\mathcal{S}'' \subseteq \mathcal{S}, x \in \mathbb{R}^n_+)$ is an (α, β)-approximation of the constructed instance of Mrfs. Then $B_S/2 \le x(S) \le \beta B_S$ for all $S \in \mathcal{S}''$, and $w(\mathcal{S}'') \ge w(\text{O}PT')/\alpha$. We construct a feasible pricing p'', for Pp where $p''_i = x_i/\beta$. Then $p''(S) \le B_S$ for all $S \in \mathcal{S}''$ and

$$p''(\mathcal{S}'') = \frac{x(\mathcal{S}'')}{\beta} \ge \frac{w(\mathcal{S}'')}{\beta} \ge \frac{w(\text{O}PT')}{\alpha\beta} \ge \frac{p(\text{O}PT)}{4\alpha\beta}.$$

The special variant of Pp when bundles are paths on the line and is called the *highway problem* in [18], and can be modelled by an Mrfs with an interval constraint matrix, and hence the (α, β)-approximation algorithms that we obtained for the Mrfs problem with interval matrices in Section 4 can be used in conjunction with the above Proposition to yield efficient approximation algorithms for the highway problem.

6 Conclusion

We have given upper and lower bounds on the approximability of the Mrfs problem with general 0/1-coefficients and for interval matrices. It seems somewhat surprising that even for a clique the problem is APX-hard if no violation is allowed, contrary to the intuition that such a clique instance should be easy. On the other hand, getting any poly-logarithmic approximation for this case or even for the general 0/1-case, without violation, remains an interesting open question. Of independent interest is our APX-hardness construction which might prove useful for proving other APX-hardness results on intervals.

References

[1] G. Aggarwal and J. D. Hartline, *Knapsack auctions*, SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (New York, NY, USA), ACM Press, 2006, pp. 1083–1092.

[2] E. Amaldi, M. Bruglieri, and G. Casale, *A two-phase relaxation-based heuristic for the maximum feasible subsystem problem*, Computers & Operations Research **35** (2008), no. 5, 1465–1482.

[3] E. Amaldi and V. Kann, *The complexity and approximability of finding maximum feasible subsystems of linear relations*, Theor. Comput. Sci. **147** (1995), no. 1-2, 181–210.

[4] E. Amaldi, M.E. Pfetsch, and Jr. L.E. Trotter, *Some structural and algorithmic properties of the maximum feasible subsystem problem*, Proceedings of the 7th International IPCO Conference on Integer Programming and Combinatorial Optimization (London, UK), Springer-Verlag, 1999, pp. 45–59.

[5] M.F. Balcan and A. Blum, *Approximation algorithms and online mechanisms for item pricing*, Theory of Computing **3** (2007), 179–195.

[6] P. Briest and P. Krysta, *Single-minded unlimited supply pricing on sparse instances*, SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (New York, NY, USA), ACM Press, 2006, pp. 1093–1102.

[7] ———, *Buying cheap is expensive: Hardness of non-parametric multi-product pricing*, Proc. 17th Annual ACM-SIAM Symposium on Discrete Algorithms, ACM-SIAM, 2007.

[8] Hajo Broersma, Fedor V. Fomin, Jaroslav Nesetril, and Gerhard J. Woeginger, *More about subcolorings.*, WG, 2002, pp. 68–79.

[9] E. D. Demaine, U. Feige, M. T. Hajiaghayi, and M. R. Salavatipour, *Combination can be hard: approximability of the unique coverage problem*, SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (New York, NY, USA), ACM Press, 2006, pp. 162–171.

[10] E. D. Demaine, M. T. Hajiaghayi, U. Feige, and M. R. Salavatipour, *Combination can be hard: approximability of the unique coverage problem*, SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm (New York, NY, USA), ACM Press, 2006, pp. 162–171.

[11] W. Duckworth, D.F. Manlove, and M. Zito, *On the approximability of the maximum induced matching problem*, Journal of Discrete Algorithms **3** (2005), 79–91.

[12] K.M. Elbassioni, R.A. Sitters, and Y. Zhang, *A quasi-PTAS for profit-maximizing pricing on line graphs*, ESA (L. Arge, M. Hoffmann, and E. Welzl, eds.), Lecture Notes in Computer Science, vol. 4698, Springer, 2007, pp. 451–462.

[13] P. Erdös and G. Szekeres, *A combinatorial problem in geometry*, Compositio Mathematica **2** (1935), 463–470.

[14] R.J. Faudree, A. Gyárfas, R.H. Schelp, and Z.Tuza, *Induced matchings in bipartite graphs*, Discrete Math. **78** (1989), 83–87.

[15] U. Feige and D. Reichman, *On the hardness of approximating max-satisfy*, Inf. Process. Lett. **97** (2006), no. 1, 31–35.

[16] P. W. Glynn, B. Van Roy, and P. Rusmevichientong, *A nonparametric approach to multi-product pricing*, Operations Research **54** (2006), to appear.

[17] O. Guieu and J.W. Chinneck, *Analyzing infeasible mixed-integer and integer linear programs*, INFORMS J. on Computing **11** (1999), no. 1, 63–77.

[18] V. Guruswami, J. D. Hartline, A. R. Karlin, D. Kempe, C. Kenyon, and F. McSherry, *On profit-maximizing envy-free pricing*, SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms (Philadelphia, PA, USA), Society for Industrial and Applied Mathematics, 2005, pp. 1164–1173.

[19] M.M. Halldórsson, *Approximations of weighted independent set and hereditary subset problems*, J. Graph Algorithms Appl. **4** (2000), no. 1, 1–16.

[20] J. D. Hartline and V. Koltun, *Near-optimal pricing in near-linear time*, Algorithms and Data Structures - WADS 2005 (F. K. H. A. Dehne, A. López-Ortiz, and J.-R. Sack, eds.), Lecture Notes in Computer Sciences, vol. 3608, Springer, 2005, pp. 422–431.

[21] J. Håstad, *Some optimal inapproximability results*, J. ACM **48** (2001), no. 4, 798–859.

[22] M.Cheung and C.Swamy, *Approximation algorithms for single-minded envy-free profit-maximization problems with limited supply*, FOCS, 2008, pp. 595–604, to appear.

[23] M.R. Salavatipour, *A polynomial time algorithm for strong edge coloring of partial k-trees*, Discrete Appl. Math. **143** (2004), no. 1-3, 285–291.

[24] J.K. Sankaran, *A note on resolving infeasibility in linear programs by constraint relaxation*, Operations Research Letters **13** (1993), no. 1, 19–20.

[25] V.Guruswami and P.Raghavendra, *A 3-query pcp over integers*, STOC, 2007, pp. 198–206.