# On the Approximability of the Maximum Interval Constrained Coloring Problem⋆

Stefan Canzar[1], Khaled Elbassioni[2], Amr Elmasry[2,3], and Rajiv Raman[4]

[1] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[2] Max-Planck-Institut für Informatik, Saarbrücken, Germany
[3] Datalogisk Institut, University of Copenhagen, Denmark
[4] DIMAP and Department of Computer Science, University of Warwick, UK

**Abstract.** In the Maximum Interval Constrained Coloring problem, we are given a set of intervals on a line and a $k$-dimensional requirement vector for each interval, specifying how many vertices of each of $k$ colors should appear in the interval. The objective is to color the vertices of the line with $k$ colors so as to maximize the total weight of intervals for which the requirement is satisfied. This $\mathcal{NP}$-hard combinatorial problem arises in the interpretation of data on protein structure emanating from experiments based on hydrogen/deuterium exchange and mass spectrometry. For constant $k$, we give a factor $O(\sqrt{|\text{Opt}|})$-approximation algorithm, where Opt is the *smallest-cardinality* maximum-weight solution. We show further that, even for $k = 2$, the problem remains APX-hard.

## 1 Introduction

The Interval Constrained Coloring (Icc) problem was introduced recently by Althaus et al. [1, 2] as the mathematical abstraction of a problem appearing in the interpretation of experimental data in biochemistry. Monitoring exchange rates via mass spectrometry is a method used to obtain information about the 3-dimensional structure of proteins. ICC captures the problem of increasing the resolution of the exchange data from peptic fragments to single residues. We refer the interested reader to Althaus et al. [2] for more on the biochemical background. Interval Constrained Coloring is a decision problem that asks for a coloring of an ordered sequence of $n$ vertices $V = [n]$ using $k$ colors such that a given set of requirements is satisfied. Each requirement is made up of a closed interval $I \subseteq [n]$, and a complete specification of how many elements in $I$ should be colored with each color.

More formally, let $\mathcal{I}$ be a set of $m$ intervals defined on $V = [n]$, let $[k]$ be a set of color classes, and $r : \mathcal{I} \times [k] \to \mathbb{Z}_+$ be a requirement function such that

$$\sum_{c\in[k]} r(I, c) = |I| \quad \text{for all } I \in \mathcal{I}. \tag{1}$$

⋆ Part of the work was done while the first, third and forth authors were members of Max-Planck Institute. A. Elmasry is on leave from Alexandria University of Egypt. R. Raman's research is supported by the Centre for Discrete Mathematics and its Applications (DIMAP) at University of Warwick, EPSRC award EP/D063191/1.

The given set of intervals $\mathcal{I}$ is said to be *feasible* (or *colorable*) if there exists a coloring $\chi : V \to [k]$ such that for every $I \in \mathcal{I}$ we have $N_\chi(I, c) = r(I, c)$, for all $c \in [k]$, where $N_\chi(I, c) = |\{i \in I : \chi(i) = c\}|$ denotes the number of vertices in $I$ assigned color $c$ by $\chi$. Clearly, not all sets of intervals are colorable. In the biochemical application from which our problem arises, the requirement function models exchange data collected in real experiments, which usually contain some noise. Hence, we might not obtain a colorable instance even if the real underlying instance is colorable. This motivates the study of the problem of extracting the largest subset $\mathcal{I}' \subseteq \mathcal{I}$ that is colorable. More precisely, we consider a more general version of the problem where each interval also has a weight, in addition to its color requirements, given by $w : \mathcal{I} \to \mathbb{R}_+$, and we wish to find a maximum-weight colorable subset of the intervals, as well as produce a feasible coloring. We define the problem formally below.

**Definition 1** (Max-Feasible-Coloring (Mfc)). *Given a set of intervals $\mathcal{I}$ with non-negative weights $w : \mathcal{I} \to \mathbb{R}_+$, a requirement function $r : \mathcal{I} \times [k] \to \mathbb{Z}^+$ satisfying (1), the* Max-Feasible-Coloring (Mfc) *problem asks for finding a maximum weight colorable subset $\mathcal{I}' \subseteq \mathcal{I}$.*

The Mfc problem can also be cast as a problem on linear systems, as observed by Byrka et al. [3]. A 0/1 matrix $A$ has the consecutive 1's property if in each row of $A$ the 1's appear consecutively. Consider the following system, where $A \in \{0, 1\}^{m \times n}$ is the consecutive 1's matrix derived in the natural manner from the Mfc problem, and $r^{(i)} \in \mathbb{Z}_+^m$ is a column vector of requirements for the $i$-th color from the $m$ intervals. Let $I$ be the $n \times n$ identity matrix.

$$\begin{bmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & A \\ I & \dots & & I \end{bmatrix} \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(k)} \end{bmatrix} = \begin{bmatrix} r^{(1)} \\ r^{(2)} \\ \vdots \\ r^{(k)} \\ 1 \end{bmatrix}$$

The Icc problem asks for a feasible solution to the above system with $x^{(i)} \in \{0, 1\}^n$, $i = 1, \dots, k$. The Mfc problem can then be stated as the problem of computing the maximum subset of rows of the matrix $A$ for which the above system has a feasible solution. In this light, the problem is similar to the Maximum Feasible Subsystem (Mfs) problem, where an infeasible linear system $Ax = b$ is given and we wish to find the largest subsystem that has a feasible solution. However, in the Mfs problem we may allow a rational vector $x$.

### 1.1 Our Results

We study the approximability of Mfc and present the first non-trivial approximation algorithm for the problem. In particular, we give a factor $O(\sqrt{|\textsc{Opt}|})$-approximation algorithm, where we denote throughout by $\textsc{Opt} \subseteq \mathcal{I}$ an optimal set of intervals, that is a maximum-weight colorable subset, with *smallest cardinality*. The main technique that we use is to decompose the problem into simpler instances using Dilworth's Theorem [4], and then show how we can solve these simpler instances in polynomial time. A similar technique was used earlier [5] to obtain an $O(\sqrt{|\text{OPT}|}\log n)$-approximation algorithm for the following MAXIMUM FEASIBLE SUBSYSTEM problem: Given an infeasible linear system $l \leq Ax \leq u$, $x \geq 0$, where $A$ is a consecutive 1's matrix, the problem is to find the largest subsystem for which there is a feasible solution satisfying the non-negativity constraints $x \geq 0$. However, the same decomposition does not work for the MFC problem. In particular, one needs to use a different poset definition when applying Dilworth's Theorem. As it turns out, the new decomposition can also be used to save the $\log n$-factor in the above approximation factor for MFS.

Note that the INTERVAL CONSTRAINED COLORING problem for $k = 2$ colors can be solved in polynomial time via linear programming. In contrast, we show that the maximization version of the problem, MFC, is APX-hard for $k = 2$. This is akin to the situation for 2SAT and MAX2SAT. We thus improve on the result in [3], which shows that MFC is APX-hard for $k = 3$.

### 1.2 Related work

The ICC problem has been introduced in [2] by Althaus et al. who formulated the problem of improving the resolution of exchange data as a linear minimization problem subject to integer linear constraints (ILP), and proposed a branch-and-bound approach to enumerate all optimal colorings (see also [6, 7]). In [1], Althaus et al. studied the problem from a theoretical point of view. They showed it to be NP-hard in general and developed algorithms that, given a *fractionally* feasible instance, find a coloring that satisfies all the requirements within $\pm 1$ of the prescribed value. Furthermore, they considered the maximization variant of the problem, MFC, and showed that if one is allowed to relax the coloring requirements by a small factor of $(1+\epsilon)$, then there is an algorithm that finds a coloring satisfying (with violations) the optimal number of intervals, and running in quasi-polynomial time if the number of colors is constant.

Komusiewicz et al. [8] showed that the problem is fixed-parameter tractable with respect to parameters such as the maximum interval length and the maximum number of intervals containing a given vertex. Very recently, the decision variant of the problem (ICC) was shown to be NP-hard even when $k = 3$ [3]; this settles the complexity of the problem since, for $k = 2$, the problem is polynomially solvable [2]. However, prior to the current paper, both inapproximability of MFC for $k = 2$, and non-trivial approximations, for constant $k$, were not known.

A related or specialized version of the INTERVAL CONSTRAINED COLORING problem also arises in discrete tomography. The goal in these problems is to reconstruct an image from the partial information that is available. In particular, we have an $m \times n$ matrix $A$ whose entries have to be colored with $k$ colors. The entries of the matrix are integers in the range $\{1, \ldots, k\}$. Further, we are given $k$ row vectors of dimension $m$, and $k$ column vectors of dimension $n$, which tell us the number of entries of each color in each row and column. The problem is then to reconstruct the coloring of the matrix $A$ with the information in the row and column vectors. This is a special case of a 2-dimensional version of our problem. See [9] for a recent survey, and [10] where the authors study the problem when the path in our case is replaced by a general graph. We believe that our techniques would be helpful in discrete tomography applications as well.

## 2 Preliminaries

In this section we recall some basic facts that we will use in our algorithm, and introduce some definitions. Let $(\mathcal{P}, \preceq)$ be a *partially ordered set (poset)*. A *chain* (respectively, anti-chain) is a set of pairwise comparable (respectively, incomparable) elements. The well-known Dilworth Theorem [4] states that, in any poset $\mathcal{P}$, the maximum size of anti-chain is equal to the minimum number of chain-covers (that is, a partition of the poset into chains). An immediate corollary is that $\mathcal{P}$ either contains a chain or an anti-chain of size at least $\sqrt{|\mathcal{P}|}$. Applying this recursively we obtain the following decomposition.

**Lemma 1.** *Let $(\mathcal{P}, \preceq)$ be any poset. Then, $\mathcal{P}$ can be decomposed into $k \leq 2\sqrt{|\mathcal{P}|}$ sets $\mathcal{P}_1, \ldots, \mathcal{P}_k$ such that, for each $i$, the induced subposet $(\mathcal{P}_i, \preceq)$ forms either a chain or an anti-chain.*

*Proof.* Let $\mathcal{P}_1$ be a chain or anti-chain of size $\sqrt{|\mathcal{P}|}$. Recurse on $(\mathcal{P} - \mathcal{P}_1, \preceq)$. The recurrence we get for the number of iterations is $f(p) \leq 1 + f(p - \sqrt{p})$, where $p = |\mathcal{P}|$. This is satisfied with $f(p) = 2\sqrt{p}$. □

**Corollary 1.** *Let $(\mathcal{P}, \preceq)$ be a poset and $w : \mathcal{P} \to \mathbb{R}_+$ be a weight function. Then there is either a chain or an anti-chain in $\mathcal{P}$ of size at least $\frac{w(\mathcal{P})}{2\sqrt{|\mathcal{P}|}}$.*

Our approximation algorithm is based on decomposing the problem into simpler instances, which can be solved in polynomial time. The instances are defined on special classes of intervals, described more precisely in the following definitions: A set of intervals $\mathcal{I}$ such that for every pair of intervals $I, I' \in \mathcal{I}$ either $I \subseteq I'$ or $I' \subseteq I$ will be called a *tower*. A set of intervals $\mathcal{I}$ such that for every pair of intervals $I, I' \in \mathcal{I}$ neither $I \subseteq I'$ nor $I' \subseteq I$ will be called an *anti-tower*. An anti-tower in which all the intervals intersect will be called a *staircase*.

We will use the following notation: For a set of intervals $\mathcal{I} \subseteq 2^V$, let $\mathbf{\Phi}(\mathcal{I})$, $\mathbf{\Psi}(\mathcal{I}, u, v)$ and $\mathbf{\Upsilon}(\mathcal{I})$ denote respectively the set of towers, the set of anti-towers starting at $u \in V$ and ending at $v \in V$, and the set of independent sets (that is, pairwise disjoint intervals) from $\mathcal{I}$.

An optimal tower $\mathcal{I}' \subseteq \mathcal{I}$ (respectively, anti-tower, or staircase) is one for which there exists a feasible coloring such that $w(\mathcal{I}') \stackrel{\text{def}}{=} \sum_{I \in \mathcal{I}'} w(I)$ is maximized among all such towers (respectively, anti-towers, or staircases). We will derive our main theorem in Section 3 form the following two lemmas, both of which assume that $k$ is a constant.

**Lemma 2.** *Given any instance of* Mfc*, we can find an optimal tower in polynomial time.*

**Lemma 3.** *Given any instance of* MFC *and two vertices $u$ and $v$, we can find an optimal staircase starting at $u$ and ending at $v$ in polynomial time.*

Two staircases are independent if the end point of one is less than the starting point of the other, i.e., all the intervals of one are independent (i.e., disjoint) of the intervals of the other. We will also need the following decomposition.

**Proposition 1 ([11]).** *Any anti-tower can be partitioned into two subsets of intervals, each of which is a set of independent staircases.*

## 3 The Approximation Algorithm

The algorithm proceeds as follows. We first find the largest-weight colorable tower (step 1). This can be done easily by dynamic programming, sketched in Section 3.1. We then find, for every pair of vertices $u, v \in V$, $u < v$, an optimal staircase starting at $u$ and ending at $v$, and define a new weight function $w'$ on every possible interval of $V$ (step 3). This can be done using the dynamic program presented in Section 3.2. Using this weight function, we compute a maximum wight independent set of intervals (step 5); we emphasize that we find this independent set *among all possible intervals* in $V$, not only form $\mathcal{I}$. This can also be done by dynamic programming (see e.g. [12]). Finally, the algorithm returns the larger among the two weights computed at steps 1 and 5. Note that it is straightforward to modify the algorithm to also return a coloring with the same weight that satisfies a subset of $\mathcal{I}$.

**Algorithm 1** Find-MFC($V, \mathcal{I}, r, w$)

1: $W_1 = \max_{\mathcal{I}' \in \mathbf{\Phi}(\mathcal{I})} w(\mathcal{I}')$ (cf. Section 3.1)
2: **for** every $u, v \in V$ s.t. $u < v$ **do**
3: $\quad w'_{u,v} = \max_{\mathcal{I}' \in \mathbf{\Psi}(\mathcal{I}, u, v)} w(\mathcal{I}')$ (cf. Section 3.2)
4: **end for**
5: Let $W_2 = \max_{\mathcal{I}' \in \mathbf{\Upsilon}(\{[u,v] : u,v \in V,\ u<v\})} w'(\mathcal{I}')$
6: **return** $\max\{W_1, W_2\}$

**Theorem 1.** *Algorithm Find-*MFC *outputs a feasible coloring for a subset of weight at least $\frac{w(\mathrm{OPT})}{4\sqrt{\mathrm{OPT}}}$.*

*Proof.* Consider the intervals in a minimum-cardinality optimal solution OPT. Define a poset $\mathcal{P} = (\mathrm{OPT}, \subseteq)$ by the containment relation on these intervals. A chain in such a poset is a tower, and an anti-chain is an anti-tower. By corollary 1, there is a tower or an anti-tower of weight at least $\frac{w(\mathrm{OPT})}{2\sqrt{|\mathrm{OPT}|}}$. If there is such an anti-tower, then, by Proposition 1, there is an independent set of staircases of weight at least $\frac{w(\mathrm{OPT})}{4\sqrt{|\mathrm{OPT}|}}$. Note that step 5 of the algorithm outputs the maximum weight of an independent set of staircases. Thus, in all cases, the total weight of intervals colored by the algorithm is as claimed. □

### 3.1 Finding an optimal tower; proof of Lemma 2

Let $\mathcal{I} = \{I_1, \dots, I_m\}$ be a given set of intervals, where we assume that $I_i = [a_i, b_i]$ for $i \in [m]$. Given the requirement function $r : \mathcal{I} \times [k] \to \mathbb{Z}_+$ satisfying (1), we construct a partially ordered set $(\mathcal{P}, \preceq)$, where $\mathcal{P} \subseteq \mathbb{Z}^{k+2}$ is defined as follows: there is one-to-one correspondence between $\mathcal{P}$ and $\mathcal{I}$; interval $I_i \in \mathcal{I}$ is mapped to the point $(-a_i, b_i, r(I_i, 1), \dots, r(I_i, k))$, and for two points $P, P' \in \mathcal{P}$, $P \preceq P'$ if and only if at each coordinate $P$ is at most the value of $P'$. The algorithm for finding an optimal tower is based on the following observation.

**Observation 1.** *$\mathcal{I}' \subseteq \mathcal{I}$ is a colorable tower if and only if it is a chain in $\mathcal{P}$.*

Thus, finding an optimal tower is equivalent to finding a maximum-weight chain, which can be done in polynomial time (see, e.g., [12]).

### 3.2 Finding an optimal staircase; proof of Lemma 3

Let $I_1, I_2, \dots, I_m$ be the sequence of intervals in $\mathcal{I}$, sorted in the order of their left endpoints, i.e., if $i < j$, then $a_i \leq a_j$. Assuming this ordering, let $I_s$ be an interval with $a_s = u$, where $u$ is the starting vertex of the optimal staircase we are looking for. Clearly, non of the intervals $I_\ell$ with $\ell < s$ can be contained in the optimal staircase starting at $u$. For simplicity, and w.l.o.g., let us therefore assume for the remainder of this section, that we want to find the optimal staircase starting at $a_1$. Similarly, if $I_t$ is an interval with $b_t = v$, non of the intervals $I_\ell$ with $\ell > t$ can be contained in the optimal staircase ending at $v$. W.l.o.g. we therefore assume in the following that we want to find the optimal staircase ending at $b_m$. Furthermore, we will assume that we have removed, in a preprocessing step, all intervals from the instance which do not form a staircase with $I_1$, i.e., intervals $I_\ell$ with $b_\ell \leq b_1$. Note that if such intervals $I_s$ or $I_t$ do not exist, the weight of an optimal staircase for this choice of $u$ and $v$ is defined to be $-\infty$ in the algorithm in Section 3. Similarly, if $I_s \cap I_t = \emptyset$ no staircase starting at $a_s$ and ending at $b_t$ exists and again its weight is defined to be $-\infty$. However, if $I_s$ or $I_t$ is not uniquely defined, we have to maximize over all possible choices of $I_s$ and $I_t$ with $I_s \cap I_t \neq \emptyset$.

We denote by $\mathcal{I}_t$ the set containing the subsequence $I_1, I_2, \dots, I_t$, for $t \leq m$, and by $\textsc{Opt}_t$ an optimal staircase to the instance induced by interval set $\mathcal{I}_t$ that contains intervals $I_1$ and $I_t$. If such a solution does not exist, we set $\textsc{Opt}_t \stackrel{\text{def}}{=} \emptyset$. Recall that for a given coloring $\chi$ and an arbitrary interval $I$ defined on $V$, $N_\chi(I, c)$ counts the number of vertices in $I$ colored $c$ by $\chi$. Accordingly, we define vector $\mathbf{N}_\chi(I) = (N_\chi(I, c))_{c \in [k]}$. Note that we denote vectors by boldface characters. For a vector $\bar{\mathbf{r}} \in \mathbb{R}^k$, $\textsc{Opt}_t(\bar{\mathbf{r}}, I)$ further constrains an optimal solution $\textsc{Opt}_t$ to be satisfiable by a coloring $\chi$ with $\mathbf{N}_\chi(I) = \bar{\mathbf{r}}$. For such a coloring to exist, the coloring requirement $\bar{\mathbf{r}}$ imposed on interval $I$ must be *valid* in the sense that $\|\bar{\mathbf{r}}\|_1 = |I|$. Since we will constrain optimal solutions $\textsc{Opt}_t$ only by valid requirements on a tail subinterval of $I_t$, we will denote $\textsc{Opt}_t(\bar{\mathbf{r}}, I)$ simply by $\textsc{Opt}_t(\bar{\mathbf{r}})$. Note that such a solution $\textsc{Opt}_t(\bar{\mathbf{r}})$ for a valid $\bar{\mathbf{r}}$ might not exist, in which case again $\textsc{Opt}_t(\bar{\mathbf{r}}) \stackrel{\text{def}}{=} \emptyset$.

Intuitively, the dynamic program exploits the following optimality property Consider a set $\mathcal{I}' \subseteq \mathcal{I}_{t'}$ with $I_1, I_{t'} \in \mathcal{I}'$ that can be satisfied by a coloring $\chi'$. Then the colors assigned by $\chi'$ to vertices within $[a_{t'}, b_1]$ can be shuffled arbitrarily without causing any interval in $\mathcal{I}'$ to be not satisfied. Therefore, for $\mathcal{I}' \cup \{I_t\}$ to be satisfiable it suffices to require that there exists a coloring $\chi$ satisfying both $I_{t'}$ and $I_t$ such that $\mathbf{N}_\chi([b_1, b_{t'}]) = \mathbf{N}_{\chi'}([b_1, b_{t'}])$. In particular, rearranging the colors of $\chi'$ within $[a_{t'}, b_1]$ such that it satisfies $I_t$ (under an appropriate coloring of the remaining vertices in $[b_{t'}+1, b_t]$) does not affect the intervals in $\mathcal{I}'$, since none of them starts or ends in this interval. Furthermore, the remainder of interval $I_{t'}$, namely $[b_{t'}+1, b_t]$, does not intersect any interval in $\mathcal{I}'$. In other words, if $I_{t'}$ is the predecessor of $I_t$ in $\text{OPT}_t$, realized by a coloring $\chi$, then the intervals in $\text{OPT}_t \setminus \{I_t\}$ must form a solution to $\mathcal{I}_{t'}$ of maximum weight among those solutions that can be realized by a coloring $\chi''$ that conforms with $\chi'$ in $[b_1, b_{t'}]$, i.e., for which $\mathbf{N}_{\chi''}([b_1, b_{t'}]) = \mathbf{N}_{\chi'}([b_1, b_{t'}])$ holds. An equivalent composition of optimal solutions can be established if we condition on the colorig within interval $[p, b_{t'}]$, $a_t \leq p \leq b_1$, i.e., when we extend the left boundary of $[b_1, b_{t'}]$ in the above observation to any $p \geq a_t$. More formally, an optimal solution exhibits the following optimal substructure.

**Theorem 2.** (Optimal Substructure)
*Assume that the optimal staircase* $\text{OPT}$ *starts at interval* $I_1$ *and ends at interval* $I_t$, *i.e.,* $I_1, I_t \in \text{OPT}$ *and* $I_\ell \notin \text{OPT}$ *for all* $\ell > t$. *Further assume that* $I_{t'}$ *is the predecessor of* $I_t$ *in* $\text{OPT}$, *i.e.,* $I_{t'} \in \text{OPT}$, *for some* $1 \leq t' < t$, *and for all* $I_\ell$ *with* $t' < \ell < t$, $I_\ell \notin \text{OPT}$. *Let* $\chi$ *be a coloring satisfying intervals in* $\text{OPT}$.*Then for any* $a_t \leq p \leq b_1$, $w(\text{OPT} \setminus \{I_t\}) = w(\text{OPT}_{t'}(\mathbf{N}_\chi([p, b_{t'}])))$.

*Proof.* Let $\chi$ be a coloring satisfying intervals in $\text{OPT}$ and let $a_t \leq p \leq b_1$ be fixed. We show that for any subset $\mathcal{I}' \subseteq \mathcal{I}_{t'}$ with $I_1, I_{t'} \in \mathcal{I}'$, that can be satisfied by a coloring $\chi'$ s.t. $\mathbf{N}_{\chi'}([p, b_{t'}]) = \mathbf{N}_\chi([p, b_{t'}])$, set $\mathcal{I}' \cup \{I_t\}$ is colorable. For that, we construct a coloring $\tilde{\chi}$ as follows.

$$\tilde{\chi}(u) = \begin{cases} \chi(u) & \text{if } u \in [b_{t'}+1, b_t] \text{ or } u \in [a_{t'}, p-1] \\ \chi'(u) & \text{if } u \in [p, b_{t'}] \text{ or } u \in [a_1, a_{t'}-1] \end{cases}$$

We show that coloring $\tilde{\chi}$ satisfies $\mathcal{I}' \cup \{I_t\}$. First, consider interval $I_t$. The only vertices spanned by $I_t$ whose coloring differs from $\chi$ lie in $[p, b_{t'}]$. These are colored according to $\chi'$, which satisfies $\mathbf{N}_{\chi'}([p, b_{t'}]) = \mathbf{N}_\chi([p, b_{t'}])$. Therefore $\mathbf{N}_{\tilde{\chi}}(I_t) = \mathbf{r}(I_t)$ and the interval is satisfied by $\tilde{\chi}$. Analogously, it can be shown that interval $I_{t'}$ is satisfied by $\tilde{\chi}$. Now consider an arbitrary interval $I_\ell \in \mathcal{I}'$ with $\ell < t'$. The only vertices spanned by $I_\ell$ whose coloring differs from $\chi'$ lie in $[a_{t'}, p-1]$. Since both $\chi$ and $\chi'$ satisfy $I_{t'}$, and $\mathbf{N}_\chi([p, b_{t'}]) = \mathbf{N}_{\chi'}([p, b_{t'}])$, we have $\mathbf{N}_\chi([a_{t'}, p-1]) = \mathbf{N}_{\chi'}([a_{t'}, p-1])$ and thus $\mathbf{N}_{\tilde{\chi}}(I_\ell) = \mathbf{r}(I_\ell)$. □

Let $[p, b_t]$ be a tail subinterval of $I_t$ with $a_t \leq p \leq b_1$, and let vector $\bar{\mathbf{r}}$ be a valid coloring requirement with $\bar{\mathbf{r}} \leq \mathbf{r}(I_t)$ on that subinterval. To obtain an optimal solution to $\mathcal{I}_t$ containing $I_1$ and $I_t$ that can be satisfied by a coloring $\chi$ with $\mathbf{N}_\chi([p, b_t]) = \bar{\mathbf{r}}$, the following recurrence "guesses" the predecessor $I_{t'}$ and the best allocation of $\bar{\mathbf{r}}$ to $I_\alpha = [p, b_{t'}]$ and the remaining interval $I_\beta = [b_{t'}+1, b_t]$, and recursively solves the resulting subproblem optimally.

$$w(\textsc{Opt}_t(\bar{\mathbf{r}})) = \max_{\substack{1 \le t' < t:\ a_{t'} < a_t \text{ and } b_{t'} < b_t \\ \bar{\bar{\mathbf{r}}} \in \mathbb{R}^k:\ ||\bar{\bar{\mathbf{r}}}||_1 = b_{t'} - p + 1 \text{ and } \bar{\bar{\mathbf{r}}} \le \bar{\mathbf{r}}, \\ \mathbf{r}(I_{t'}) - \bar{\bar{\mathbf{r}}} \ge \mathbf{r}(I_t) - \bar{\mathbf{r}}}} w(\textsc{Opt}_{t'}(\bar{\bar{\mathbf{r}}})). \tag{2}$$

If there is no coloring satisfying both $I_1$ and $I_t$ under the restriction $\mathbf{N}_\chi([p, b_t]) = \bar{\mathbf{r}}$, we set $\textsc{Opt}_t(\bar{\mathbf{r}}) \stackrel{\text{def}}{=} \emptyset$ and define $w(\emptyset) \stackrel{\text{def}}{=} -\infty$.

The recurrence considers only possible predecessors $I_{t'}$ that together with $I_t$ form a staircase. The tail subinterval $[p, b_t]$ is divided into interval $I_\alpha$ overlapping with $I_{t'}$ and the remaining interval $I_\beta$. A valid requirement $\bar{\bar{\mathbf{r}}}$ on $I_\alpha$ is chosen in such a way that there exists a coloring $\chi$ satisfying both $I_\alpha$ and $I_\beta$ under the constraints $\mathbf{N}_\chi([p, b_t]) = \bar{\mathbf{r}}$ and $\mathbf{N}_\chi([p, b_{t'}]) = \bar{\bar{\mathbf{r}}}$. Such a coloring $\chi$ exists if and only if requirement $\bar{\bar{\mathbf{r}}}$ on $[p, b_{t'}]$ is consistent with requirement $\bar{\mathbf{r}}$ on $[p, b_t]$ (i.e., $\bar{\bar{\mathbf{r}}} \le \bar{\mathbf{r}}$) and at the same time the resulting requirement on $[a_t, p-1]$ of a coloring satisfying $I_t$, namely $\mathbf{r}(I_t) - \bar{\mathbf{r}}$, is consistent with the resulting requirement on $[a_{t'}, p-1]$ of a coloring satisfying $I_{t'}$, namely $\mathbf{r}(I_{t'}) - \bar{\bar{\mathbf{r}}}$ (i.e., $\mathbf{r}(I_{t'}) - \bar{\bar{\mathbf{r}}} \ge \mathbf{r}(I_t) - \bar{\mathbf{r}}$).

We compute the recurrence relation by a dynamic programming approach as follows. An entry in the dynamic program matrix $\mathbf{D}$ is indexed by an interval index $t$, the left boundary $p$ of a tail subinterval of $I_t$, and a valid coloring requirement vector $\bar{\mathbf{r}}$ on this subinterval $[p, b_t]$. Then table entry $\mathbf{D}[t, p, \bar{\mathbf{r}}]$ stores the weight of an optimal solution to $\mathcal{I}_t$ that contains $I_1$ and $I_t$ and that can be satisfied by a coloring $\chi$ with $\mathbf{N}_\chi([p, b_t]) = \bar{\mathbf{r}}$, more precisely

$$\mathbf{D}[t, p, \bar{\mathbf{r}}] = w(\textsc{Opt}_t(\bar{\mathbf{r}})).$$

Concerning the base case $t = 1$, we simply have to decide, for each $p = a_1, a_2, \ldots, a_m$ and each valid coloring requirement $\bar{\mathbf{r}}$ on tail subinterval $[p, b_1]$ of $I_1$, whether interval $I_1$ is satisfiable under $\bar{\mathbf{r}}$:

$$\mathbf{D}[1, p, \bar{\mathbf{r}}] = \begin{cases} w(\{I_1\}) & \text{if } \bar{\mathbf{r}} \le \mathbf{r}(I_1), \\ -\infty & \text{otherwise.} \end{cases}$$

For all $2 \le t \le m$, $p = a_t, a_3, \ldots, a_m$, and valid coloring requirements $\bar{\mathbf{r}}$ on $[p, b_t]$, we simply "guess" the predecessor of $I_t$ in $\textsc{Opt}_t(\bar{\mathbf{r}})$ according to (2):

$$\mathbf{D}[t, p, \bar{\mathbf{r}}] = \max_{\substack{1 \le t' < t:\ a_{t'} < a_t \text{ and } b_{t'} < b_t \\ \bar{\bar{\mathbf{r}}} \in \mathbb{R}^k:\ ||\bar{\bar{\mathbf{r}}}||_1 = b_{t'} - p + 1 \text{ and } \bar{\bar{\mathbf{r}}} \le \bar{\mathbf{r}}, \\ \mathbf{r}(I_{t'}) - \bar{\bar{\mathbf{r}}} \ge \mathbf{r}(I_t) - \bar{\mathbf{r}}}} \mathbf{D}[t', p, \bar{\mathbf{r}}],$$

where $I_\alpha = [p, b_{t'}]$ and $I_\beta = [b_{t'} + 1, b_t]$, as above. The desired value of the optimal staircase starting at $a_1$ and ending at $b_m$ is $\mathbf{D}[m, a_m, \mathbf{r}(I_m)]$.

## 4 APX-hardness

We show that MFC is APX-hard even with $k = 2$ colors. The reduction is from MAX2ESAT which was shown to be APX-hard by Håstad [13]. The input to the MAX2ESAT problem is a boolean formula $\phi$ in conjunctive normal form with variables $x_1, \ldots, x_n$, and clauses $C_1, \ldots, C_m$, where each clause $C_i$ is a disjunction of exactly 2 literals. The problem asks for an assignment of boolean values to the variables that satisfies the maximum number of clauses(where a clause is satisfied if it evaluates to 1). For a variable $i$, we let $m_i$ denote the number of clauses that the variable appears in. The reduction is inspired by the APX-hardness proof for the maximum feasible subsystem on interval matrices [5]. We construct gadgets for each variable and each clause. Let $P$ be a path on $8n+2$ vertices, where the vertices are labeled $\{v_{-4n-1}, v_{-4n}, v_{-4n+1}, \ldots, v_{-1}, v_1, v_2, \ldots, v_{4n}, v_{4n+1}\}$. In the figures, we represent the path having vertices on integer points on the number line. However, the origin is not a vertex of $P$. Let the two colors we use be BLACK and WHITE. We start with a description of the variable gadgets.

*Variable Gadget:* The gadget for variable $x_i$ consists of $2m_i$ copies of the following *basic gadget* that we describe, i.e., each interval of the basic gadget for $x_i$ is replicated $2m_i$ times to obtain the variable gadget for $x_i$.

A basic gadget for variable $x_i$, $BG_i$ consists of 4 sets of intervals $L_i, R_i, C_i$, and $B_i$, nemonic for *left, right, center* and *boundary* respectively. The sets $L_i, R_i$ and $C_i$ consist of 2 intervals each, labeled $L_i^0$ and $L_i^1$, and so on, and the set $B_i$ consists of two intervals labeled $B_i'$ and $B_i''$. Figure 1 gives a graphic representation of the gadget along with the requirements and coordinates of the intervals. The intervals in set $L_i$, for $i = 1, \ldots, n$ have their right end-point at $-1$, and the intervals of set $R_i$ have their left end-point at 1. The set $C_i$ is symmetric around 0. A variable gadget consists of $2m_i$ copies of the basic gadget $BG_i$ for variable $x_i$. Assume we fix the colors of the vertices $v_{-1}$ and $v_1$ so that they have distinct colors. Then, in any solution that satisfies the maximal number of intervals of $BG_i$, the two sets of vertices $S_A = \{v_{-4i-1}, v_{-4i}, v_{4i}, v_{4i+1}\}$, and $S_B = \{v_{-4i+1}, v_{-4i+2}, v_{4i-2}, v_{4i-1}\}$ receive two consecutive BLACK and two consecutive WHITE colors in the same order, i.e., there are two possible ways in which this happens encoding TRUE and FALSE.

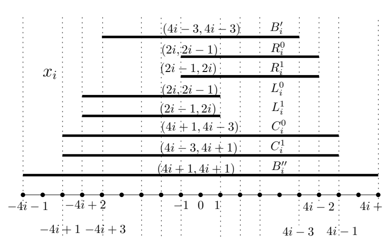


**Fig. 1.** The basic gadget for variable $x_i$

**Lemma 4.** *The maximum number of intervals of a basic gadget $BG_i$ that can be satisfied simultaneously is* 5. *If we assume that the vertices $v_{-1}$ and $v_1$ recieve distinct colors, then there are exactly two colorings of the sets $S_A$ and $S_B$ (modulo the colors on $v_{-1}$ and $v_1$).*

*Proof.* Consider a basic gadget $BG_i$. Each set $L_i, R_i$ and $C_i$ consist of a pair of mutually exclusive intervals, and we can satisfy at most one interval from each set. This leaves us with the two intervals in the set $B_i$. Hence, we can satisfy at most 5 intervals simultaneously, viz. one each from $L_i, R_i, C_i$, and both intervals from $B_i$. Note that such a set can always be satisfied as follows. Color the vertices in the range of $B'_i$ arbitrarily with equal numbers of each color. Now, suppose we color the vertices in the sets $S_A$ the same color (either BLACK or WHITE), and the set $S_B$ the same color. This gives a feasible coloring satifying 5 intervals. Note that any optimal solution must select both intervals in the set $B_i$ and one interval from each of the remaining sets for an optimal solution.

Now consider a coloring that assigns distinct colors to the vertices $-1$ and 1. We claim that there is no optimal solution that simultaneously satisfies both $R_i^0$ and $L_i^1$, and similarly there is no optimal solution simultaneously satisfying both the intervals $R_i^1$ and $L_i^0$. Since we have assumed that $v_{-1}$ and $v_1$ recieve distinct colors, the set of vertices $S = \{v_{-4i+3}, v_{-4i+2}, \ldots, v_{-3}, v_{-2}\} \cup \{v_2, v_3, \ldots, v_{4i-3}\}$ must satisfy the color requirement $(4i-4, 4i-4)$ in order to satisfy $B'_i$. Further, assume that we can satisfy $R_i^0$. Then, the range $\{v_2, \ldots, v_{4i-2}\}$ satisfy the requirement $(2i-1, 2i-2)$. Now, if we satisfy $L_i^1$, the range $\{v_{-4i+2}, \ldots, v_{-2}\}$ have color requirements $(2i-2, 2i-1)$. Summing them up, the range $S' = S \cup \{v_{-4i+2}, v_{4i-2}\}$ has colors $(4i-3, 4i-3)$. But, $S' \cup \{v_{-1}, v_1\}$ has colors $(4i-2, 4i-2)$, and we can not satisfy any interval in the set $C_i$. Hence, we can not have both $R_i^0$ and $L_i^1$ in an optimal solution. A symmetric argument holds for the case where we have $R_i^1$ and $L_i^0$ together in an optimal solution. □

For a variable gadget then, we can satisfy $10m_i$ intervals simultaneously. We abuse notation and use $L_i, L_i^0$, etc. to refer either to the intervals of the basic gadget, or to the $2m_i$ copies of the variable gadget. An optimal solution for a variable gadget consisting of $\{B_i, L_i^0, R_i^0, C_i^0\}$ is said to be in a FALSE configuration, and an optimal solution for a variable gadget consisting of $\{B_i, L_i^1, R_i^1, C_i^1\}$ is said to be in a TRUE configuration.

We can show easily by induction on $n$ that the variable gadgets can each be in TRUE or FALSE configuration independently, i.e., for any choice of an optimal configuration for each variable gadget corresponding to variables $x_1, \ldots, x_n$, there exists a coloring that satisfies all of them simultaneously. This is encoded in Lemma 5.

**Lemma 5.** *The variable gadgets can be independently set to a* TRUE *or* FALSE *configuration, with $v_{-1}$ and $v_1$ receiving distinct colors.*

*Clause Gadgets:* We have 4 different clause gadgets corresponding to the 4 different types of clauses. If a clause $C_p$ is of the form $(x_i \vee x_j)$ or $(\overline{x_i} \vee \overline{x_j})$, then the gadget for this clause consists of two intervals $I_\alpha^p, I_\beta^p$. If $C_p$ is of the form

$(\overline{x_i} \vee x_j)$ or $(x_i \vee \overline{x_j})$ the gadget consists of 3 intervals $I^p_\alpha, I^p_\beta$ and $I^p_\gamma$. Table 1 shows the gadgets for the 4 different types of clauses. The intervals corresponding to a clause gadget are mutually exclusive, and satisfy the property that exactly one interval will be satisfied if and only if the corresponding interval is satisfied. We show this in the next lemma.

**Table 1.** The Clause gadgets corresponding to the 4 different types of clauses. We assume that $i < j$.

| Clause | Intervals | Requirement |
|---|---|---|
| $C_p = (x_i \vee x_j)$ | $I^p_\alpha = [v_{-4j+2}, v_{4i-2}]$ | $(2i+2j-3, 2i+2j-1)$ |
| | $I^p_\beta = [v_{-4j+2}, v_{4i-2}]$ | $(2i+2j-2, 2i+2j-2)$ |
| $C_p = (\overline{x_i} \vee x_j)$ | $I^p_\alpha = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j, 2i+2j-3)$ |
| | $I^p_\beta = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j-2, 2i+2j-1)$ |
| | $I^p_\gamma = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j-3, 2i+2j)$ |
| $C_p = (x_i \vee \overline{x_j})$ | $I^p_\alpha = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j, 2i+2j-3)$ |
| | $I^p_\beta = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j-1, 2i+2j-2)$ |
| | $I^p_\gamma = [v_{-4j+1}, v_{4i-2}]$ | $(2i+2j-3, 2i+2j)$ |
| $C_p = (\overline{x_i} \vee \overline{x_j})$ | $I^p_\alpha = [v_{-4j+2}, v_{4i-2}]$ | $(2i+2j-1, 2i+2j-3)$ |
| | $I^p_\beta = [v_{-4j+2}, v_{4i-2}]$ | $(2i+2j-2, 2i+2j-2)$ |

**Lemma 6.** *Let $C$ be a clause, and assume that the gadgets for all the variables are all optimally satisfied with $v_{-1}$ and $v_1$ receiving distinct colors. Then, exactly one interval of $C$ is satisfied if and only if the corresponding clause is satisfied with the truth assignment implied by the satisfied variable gadgets.*

Assume that all the variable gadgets are optimally satisfied. Then, the lemma follows directly by counting the colors of the vertices contained in the intervals corresponding to the clause gadget; the details are omitted.

The final part of the reduction, we show how we can ensure the assumption in Lemmas 4, 5 and 6 that $v_{-1}$ and $v_1$ recieve distinct colors. We do this by the addition of a set of $24m$ $Z$ intervals spanning $[v_{-1}, v_1]$, with requrirement $(1, 1)$. This completes the reduction. We are now ready to prove the main theorem.

**Theorem 3.** MFC *is APX-hard with $k \geq 2$ colors.*

*Proof.* (Sketch) Given an MAX2ESAT instance with $n$ variables and $m$ clauses that has an assignment satisfying $k$ clauses, we construct a solution to the corresponding MFC as described in the previous paragraph, encoding TRUE and FALSE that satisfies at least $44m + k$ intervals. We can show that in the reverse direction that in any optimal solution to the MFC instance, all $Z$ intervals are satisfied, and all intervals corresponding to each variable gadget are satisfied optimally. Hence, correspoding to each satisfied clause, exactly one interval corresponding to the clause gadget is satsified. Since the instance has at most $O(m)$ intervals, the reduction is gap preserving. □

The reduction above heavily uses mutually exclusive intervals with the same end-points in order to achieve the gap preserving reduction. However, we believe that we can modify the gadgets, albeit making the reduction more complicated so that there are no mutually exlusive intervals with the same end-points, and yet the reduction is gap preserving.

## References


1. Althaus, E., Canzar, S., Elbassioni, K.M., Karrenbauer, A., Mestre, J.: Approximating the interval constrained coloring problem. In: Proceedings of the 11th Scandinavian Workshop on Algorithm Theory. (2008) 210–221
2. Althaus, E., Canzar, S., Emmett, M., Karrenbauer, A., Marshall, A., Meyer-Bäse, A., Zhang, H.: Computing H/D-exchange speeds of single residues from data of peptic fragments. In Wainwright, R., Haddad, H., eds.: SAC, ACM (2008) 1273–1277
3. Byrka, J., Karrebauer, A., Sanità, L.: Hardness of interval constrained coloring. In: Proceedings of the 9th Latin American Theoretical Informatics Symposium. (2009)
4. Dilworth, R.: A decomposition theorem for partially ordered sets. Annals of Mathematics **51** (1950) 161–166
5. Elbassioni, K.M., Raman, R., Ray, S., Sitters, R.: On the approximability of the maximum feasible subsystem problem with 0/1-coefficients. In Mathieu, C., ed.: SODA, SIAM (2009) 1210–1219
6. Althaus, E., Canzar, S., Ehrler, C., Emmett, M.R., Karrenbauer, A., Marshall, A.G., Meyer-Bäse, A., Tipton, J., Zhang, H.: Discrete fitting of hydrogen-deuterium-exchange-data of overlapping fragments. In: Proceedings of the 4th International Conference on Bioinformatics & Computational Biology. (2009) 23–30
7. Canzar, S., Elbassioni, K., Mestre, J.: A polynomial delay algorithm for enumerating approximate solutions to the interval constraint coloring problem. In: ALENEX, SIAM (2010) 23–33
8. Komusiewicz, C., Niedermeier, R., Uhlmann, J.: Deconstructing intractability: A case study for interval constrained coloring. In Kucherov, G., Ukkonen, E., eds.: CPM. Volume 5577 of Lecture Notes in Computer Science., Springer (2009) 207–220
9. de Werra, D., Costa, M.C., Picouleau, C., Ries, B.: On the use of graphs in discrete tomography. 4OR **6** (2008) 101–123
10. Bentz, C., Costa, M.C., de Werra, D., Picouleau, C., Ries, B.: On a graph coloring problem arising from discrete tomography. Networks **51** (2008) 256–267
11. Albertson, M., Jamison, R., Hedetniemi, S., Locke, S.: The subchromatic number of a graph. Discrete Mathematics **74** (1989) 33–49
12. Schrijver, A.: Combinatorial Optimization: Polyhedra and Efficiency, Algorithms and Combinatorics. Volume 24. Springer, New York (2003)
13. Håstad, J.: Some optimal inapproximability results. J. ACM **48** (2001) 798–859