The Approximability of the Binary Paintshop Problem

Anupam Gupta[1⋆], Satyen Kale[2], Viswanath Nagarajan[2], Rishi Saket[2], and Baruch Schieber[2]

[1] Dept. of Computer Science, Carnegie Mellon University, Pittsburgh PA 15213. anupamg@cs.cmu.edu

[2] IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598. {sckale, viswanath, rsaket, sbar}@us.ibm.com

Abstract. In the Binary Paintshop problem, there are m cars appearing in a sequence of length $2m$, with each car occurring twice. Each car needs to be colored with two colors. The goal is to choose for each car, which of its occurrences receives either color, so as to minimize the total number of color changes in the sequence. We show that the Binary Paintshop problem is equivalent (up to constant factors) to the Minimum Uncut problem, under randomized reductions. By derandomizing this reduction for hard instances of the Min Uncut problem arising from the Unique Games Conjecture, we show that the Binary Paintshop problem is $\omega(1)$-hard to approximate (assuming the UGC). This answers an open question from [4,9,3].

1 Introduction

The paintshop problem is defined as follows: we are given a $2m$ length sequence containing m cars, where each car appears twice. Each car need to be colored red in one occurrence, and blue in the other. We need to choose which occurrence for each car to color with which color — the goal is to minimize the number of times we need to change the current color. E.g., for $m = 3$, we may represent the 3 cars by x, y, z. If the sequence is $x_1x_2y_1z_1y_2z_2$, where the subscripts denote the first and second occurrence of each car, we could use the colors BRRRBB, to get two color changes, which is the minimum possible. This problem (along with generalizations) was introduced by Epping, Hottstättler and Oertel [6]; their motivation was a natural application in the automotive industry.

Let us formalize the definition: in the basic *Binary Paintshop* problem, the input is a sequence of length n (which is usually associated with the set $[n] := \{1, 2, \ldots, n\}$), along with a matching H on the points in $[n]$. A feasible coloring $f : [n] \to \{\mathsf{B}, \mathsf{R}\}$ of the vertices must ensure that the endpoints of each matching edge $(i, j) \in H$ are bi-colored—i.e., feasibility means $f(i) \neq f(j)$ for all $(i, j) \in$

⋆ This author thanks IBM T.J. Watson Research Center for the generous hospitality while this work was done.

H. The *cost* of a coloring, also called the number of *color changes* is the number of pairs $(i, i+1)$ for $i \in [n-1]$ that are bichromatic; i.e.,

$$\text{cost}(f) := \sum_{i=1}^{n-1} \mathbf{1}(f(i) \neq f(i+1)) \ .$$

For an instance of the Binary Paintshop problem $\mathit{\Gamma}$, we denote by $\mathsf{Opt}(\mathit{\Gamma})$ the cost of the minimum cost coloring. The goal is to find a feasible coloring f that (approximately) minimizes the cost. We refer to the edges of H as the *constraints* in the instance.

Epping et al. [6], along with defining the problem, gave an exponential-time dynamic programming algorithm to solve the problem exactly, and showed NP-hardness for it as well. Subsequently, Bonsma et al. [4] and Meunier and Sebö [9] showed the problem to be APX-hard. They posed the question of whether the problem admitted a constant-factor approximation algorithm. We resolve this question negatively assuming the Unique Games Conjecture (UGC) of Khot [7]:

Theorem 1. *Assuming the UGC, the (basic) Binary Paintshop problem is NP-hard to approximate within any constant factor.*

The above theorem follows via reduction from *Min Uncut* and is proved in Section 4. The Min Uncut problem is the complement of Maximum Cut, and is defined formally in Section 1.2. We present an approximation preserving (up to constant factors) reduction from Min Uncut to Binary Paintshop in Section 3.1. Assuming the Unique Games Conjecture, Khot et al. [8] showed that Min Uncut is NP-hard to approximate within any constant factor. Using their result and a specific instantiation of the reduction in Section 3.1 yields the proof of Theorem 1. A connection between the Binary Paintshop and Min Uncut problems was also noted by Meunier and Sebö [9], but they could only show an APX-hardness using this connection.

We also consider a generalization of this problem: in the *generalized Binary Paintshop* problem instead of being given a sequence of n points (which is naturally associated with a path graph on n nodes), we are given a general graph $G = (V, E)$. Moreover, instead of the constraints H forming a matching on V, we have a bipartite graph $H = (V, E_H)$ on V. The feasibility of a coloring $f : V \to \{\mathsf{B}, \mathsf{R}\}$ still requires that all constraint edges in H are bi-colored—i.e., $f(i) \neq f(j)$ for all $(i, j) \in E_H$. Note that the bipartiteness of H is essential to ensure feasibility, if H is not bipartite there is no feasible coloring. The interesting cases of this problem are when the constraint graph H has many components—e.g., when H is a matching and hence potentially has $n/2$ components.[3] The cost of f is defined naturally as

$$\text{cost}(f) := \sum_{(i,j) \in E(G)} \mathbf{1}(f(i) \neq f(j)) \ .$$

[3] This is because the problem can be solved in time exponential in the number of connected components of H (and polynomial in $n = |V(G)|$) by enumerating over the 2-colorings of each component of H.

Our second result, proved in Section 3.2, is that the generalized Binary Paintshop problem is no harder than the Min Uncut problem:

Theorem 2. *A ρ-approximation algorithm for Min Uncut can be used to give a ρ-approximation for the (generalized) Binary Paintshop problem.*

Using the algorithm of Agarwal et al. [1], we now get an $O(\sqrt{\log n})$-approximation for Binary Paintshop. Note that the hardness result is shown for the most restrictive (basic) Binary Paintshop problem, whereas the algorithm is for the generalized Binary Paintshop problem.

1.1 Related Work

Other than considering the complexity of the problem as mentioned above, previous work analyzed the performance of greedy algorithms for the paintshop problem, since this type of algorithms are actually used in real life instances of the problem. Meunier and Sebö [9] showed a class of instances for which the greedy algorithm is optimal. Amini et al. [2] showed that the greedy algorithm is optimal for even a larger class of instances and also proved that the expected number of color changes given by the greedy coloring on a random sequence is at most $2n/3$. Andres and Hottstättler [3] improved this upper bound to $n/2$. They also considered a hybrid greedy algorithm whose expected number of color changes is $2n/5$.

1.2 Notation and Definitions

In the *Min Uncut* problem, the input is an undirected, unweighted graph $G = (V, E)$ on $n := |V|$ vertices, with every vertex having degree at most $\text{poly}(n)$ (thus we allow parallel edges). The goal is to find a cut $(U, \overline{U})$ for $U \subseteq V$ to minimize the number of uncut edges—i.e., the edges in $E[U] \cup E[\overline{U}]$, where $E[S]$ is the set of edges both of whose edges lie within the set S. With some abuse of notation, for an instance of the Min Uncut problem G, we denote by $\mathsf{Opt}(G)$ the optimal value. The current best algorithm for Min Uncut is an $O(\sqrt{\log n})$-approximation due to Agarwal et al. [1].

2 Dispersive Permutations

For our hardness proofs, we will need that random permutations satisfy a certain property, which we call dispersion. In this section, we record this proof.

Definition 1. *Given a coloring $f : [m] \to \{\mathsf{B}, \mathsf{R}\}$, the* complement *coloring $\overline{f}$ is obtained by switching the assignments of f from* B *to* R *and vice versa. The coloring f is called M-*non-monochromatic *for $M = \min\left\{|f^{-1}(\mathsf{B})|, |f^{-1}(\mathsf{R})|\right\}$.*

Definition 2. *Given a coloring $f : [m] \to \{\mathsf{B}, \mathsf{R}\}$ that is M-non-monochromatic, a permutation $\sigma : [m] \to [m]$ is called* dispersive *for f if either f or $(\overline{f} \circ \sigma)$ has at least $M/16$ color changes. If a permutation is dispersive for all colorings, then it is simply called* dispersive.

Lemma 1. *For any $m \geq 10^8$, a uniformly random permutation $\sigma : [m] \to [m]$ is dispersive with probability at least $1 - \frac{1}{512m}$.*

Proof. Fix any $M \in [m/2]$ and let f be an M-non-monochromatic coloring. Observe that when $M < 16$, the coloring f trivially has at least $M/16$ color changes; so we assume $M \geq 16$ below.

We first count the number of colorings with at most $M/16$ color changes. For any number $r \leq M/16$, to estimate the number of colorings with exactly r color changes, we note that any such coloring f gives rise to a subset of r indices $\{i_1, i_2, \ldots, i_r\} \subseteq [m-1]$ where the coloring changes color, i.e. for each such i_k, we have $f(i_k) \neq f(i_k + 1)$. The number of such subsets is $\binom{m-1}{r}$, and each such subset can give rise to two colorings with r-color changes. Hence, the number of colorings with at most $M/16$ color changes is bounded by

$$N \;:=\; \sum_{r=0}^{M/16} 2\binom{m-1}{r} \;\leq\; \frac{M}{8}\binom{m-1}{M/16}.$$

Since for a uniformly random permutation σ, $(\overline{f} \circ \sigma)$ is a uniformly random M-non-monochromatic coloring with the opposite minority color of f, of which there are $\binom{m}{M}$, the probability that $(\overline{f} \circ \sigma)$ has at most $M/16$ color changes is bounded by

$$p \quad:=\quad \sum_{r=0}^{M/16} \left[\frac{2\binom{m-1}{r}}{\binom{m}{M}}\right] \quad\leq\quad \frac{M\binom{m-1}{M/16}}{8\binom{m}{M}}. \tag{1}$$

Thus, for any fixed M-non-monochromatic coloring f, at most p fraction of permutations σ yield a "bad" coloring $(\overline{f} \circ \sigma)$, i.e. $(\overline{f} \circ \sigma)$ has at most $M/16$ color changes. However, we are only concerned with colorings f which are "bad" to begin with, i.e. f has at most $M/16$ color changes. As above, the number of such colorings is bounded by N. Thus, by the union bound, the probability that a random permutation σ *is not* dispersive for some coloring f that is M-non-monochromatic is at most

$$Np \;\leq\; \frac{M^2\binom{m-1}{M/16}^2}{64\binom{m}{M}} \;\leq\; \frac{(m/2)^2}{64m^4} \;\leq\; \frac{1}{256m^2}, \tag{2}$$

where the second inequality uses Claim 3 below. Summing this probability bound for all $m/2$ integer values for M, we complete the proof of the lemma.

Claim 3 *Suppose $m \geq 10^8$. For any integers $16 \leq M \leq \frac{m}{2}$, we have $\binom{m-1}{M/16}^2 \leq \frac{1}{m^4} \cdot \binom{m}{M}$.*

Proof. The proof is by the following two calculations.

Suppose $M \geq 32 \ln m$. We have $\binom{m-1}{M/16}^2 \leq \left(\frac{16em}{M}\right)^{M/8}$ and $\binom{m}{M} \geq \left(\frac{m}{M}\right)^M$. So the ratio of these is at most:

$$\left(\frac{16em}{M}\right)^{M/8} \cdot \left(\frac{M}{m}\right)^M \;=\; \left(16e(M/m)^7\right)^{M/8} \;\leq\; \left(16e(1/2)^7\right)^{M/8} \;\leq\; e^{-M/8} \;\leq\; \frac{1}{m^4},$$

where the first inequality follows because $M \leq m/2$, and the last because $M \geq 32 \ln m$.

Now suppose $M \leq 32 \ln m$. We have $\binom{m-1}{M/16}^2 \leq m^{M/8}$ and $\binom{m}{M} \geq \left(\frac{m}{M}\right)^M$. Thus the ratio is at most

$$m^{M/8} \cdot \frac{M^M}{m^M} \leq \frac{m^{M/2}}{m^M} \leq \frac{1}{m^4},$$

where the first inequality follows because for $m \geq 10^8$, we have $M \leq 32 \ln(m) \leq m^{3/8}$, and the second because $M \geq 8$.

3 Relationship to Min Uncut

In this section we show formal connections between the Min Uncut problem and the Binary Paintshop problem. Recall the Min Uncut problem from Section 1.2, in which we want to find a cut that minimizes the number of uncut edges. We show that the two problems have the same asymptotic approximability under randomized reductions.

3.1 Reducing Min Uncut to Binary Paintshop

Theorem 4. *A ρ-approximation algorithm for Binary Paintshop implies a randomized algorithm for Min Uncut that returns an $O(\rho)$-approximation with probability at least* 0.99*.*

The success probability can be boosted to arbitrarily close to 1 by repeating the algorithm with different random seeds and returning the best solution found.

The proof works as follows. Given an instance of the Min Uncut problem, we give a gadget transformation to an instance of the Binary Paintshop problem. This gadget has a block of nodes for each node in the Min Uncut instance, and matchings between different blocks represent edges in the Min Uncut instance. Since a solution to the Binary Paintshop instance is forced to color the endpoints of each matching edge differently, within every block we can interpret the different colors assigned as "votes" for the side of the cut that the node in the Min Uncut instance should lie on. One can then round this solution to the Binary Paintshop instance to a solution for the Min Uncut instance by taking the majority vote within each block. To ensure that the cost of the obtained solution to the Min Uncut can be bounded in terms of the cost of the solution to the Binary Paintshop instance, we need to relate the minority vote in each block to the number of color changes. To do this, we add an additional block for each node together with a matching provided by a dispersive permutation into the original block which serves to "mix up" the coloring of the original block. This ensures that the total number of the color changes within the two blocks for each node is at least a constant fraction of the minority vote. The details follow.

The Binary Paintshop Instance $\varGamma_G$. We are given a graph $G(V, E)$ as input to the Min Uncut problem. Let $n = |V|$ where vertices are indexed $\{1, 2, \ldots, n\}$. Let $d(i)$ denote the degree of vertex $i \in [n]$. We choose an integer parameter $T \leq poly(n)$ to be specified later, and consider the multigraph G' obtained by making T copies of each edge in G, so each vertex i has degree $Td(i)$ in G', and we order the corresponding edges arbitrarily. For each $i \in [n]$, we choose a random permutation σ_i on $Td(i)$ elements. Our instance $\varGamma_G$ of Binary Paintshop contains for each vertex $i \in [n]$, two *sequences* of points R_i and S_i. (See Figure 1.)

Sequence R_i: This contains $2Td(i)$ points and is given by

$$\langle x_{i,1}, y_{i,1}, x_{i,2}, y_{i,2}, \cdots, x_{i,Td(i)}, y_{i,Td(i)} \rangle.$$

There are $Td(i)$ x-points corresponding to edges incident to vertex i in G', and the remaining are y-points which will be used to enforce a feasible coloring.

Sequence S_i: This contains $Td(i)$ points

$$\langle z_{i,1}, z_{i,2}, \cdots, z_{i,Td(i)} \rangle.$$

Final sequence W: The final sequence W of the instance $\varGamma_G$ is just a concatenation of the sequences constructed above.

$$W := R_1 \circ R_2 \circ \cdots \circ R_n \circ S_1 \circ S_2 \circ \cdots \circ S_n. \tag{3}$$

The instance $\varGamma_G$ of Binary Paintshop now consists of the path whose vertices correspond to the points in W, in that order. Note the length of this path is $3T \sum_{i=1}^{n} d(i) = 6T|E|$, which is polynomial in the size of G. Now for the constraints in $\varGamma_G$—recall these must form a matching H on the points in W. There are two kinds of matching pairs: *edge pairs* and *permutation pairs*, which are defined below.

Edge pairs: For each edge e in G', if e is the rth edge incident to vertex i and the sth edge incident to vertex j, then we define $\{x_{i,r}, x_{j,s}\}$ to be an edge pair.

Permutation pairs: For each vertex $i \in [n]$ and each $\ell \in \{1, 2, \cdots, Td(i)\}$, we have a permutation pair $\left\{y_{i,\ell}, z_{i,\sigma_i(\ell)}\right\}$, where σ_is are the random permutations chosen above.

Relating $\mathsf{Opt}(\varGamma_G)$ and $\mathsf{Opt}(G)$. We now relate the optimal solutions on any Min Uncut instance G, and the Binary Paintshop instance $\varGamma_G$ created by the above process.

Lemma 2. *Given any feasible solution to the Min Uncut problem on G with M uncut edges, we can construct a feasible solution to the Binary Paintshop instance $\varGamma_G$ of cost at most $2MT + 2n$. Thus $\mathsf{Opt}(\varGamma_G) \leq 2T \cdot \mathsf{Opt}(G) + 2n$.*

Proof. Let $(U, \overline{U})$ be a cut in G with M uncut edges. We now construct a feasible solution to the Paintshop instance $\varGamma_G$ of cost at most $2MT + 2n$. We first

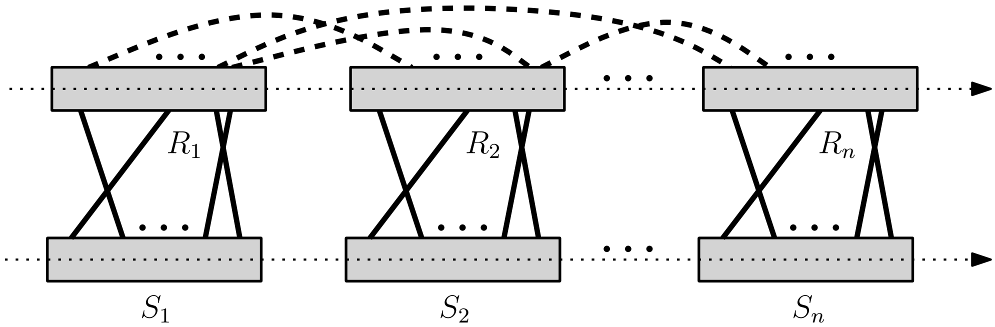

Fig. 1. High level view of the construction. The rectangles represent the sequences R_i and S_i, and the dotted line represents their concatenation in the final sequence W. The edge pairs are represented by the dashed lines, whereas the permutation pairs are solid lines.

construct an initial coloring $F_{initial}$ of the points in Γ_G as follows. (This initial coloring will be infeasible, but we shall see that incurring a small extra cost will make it feasible.) For each $1 \leq i \leq n$, the points in the sequence R_i are colored R if $i \in U$ and B otherwise. Similarly, all points in the sequence S_i are colored B if $i \in U$ and R otherwise, for $1 \leq i \leq n$. (In Figure 1, all points within each rectangle have the same color, and any pair of vertically aligned rectangles has opposite colors.) Note that every permutation pair is bichromatic and hence satisfied. Before we fix the monochromatic edge pairs, let us first count the number of color changes in $F_{initial}$. Clearly, since each of the sequences S_i and R_i are monochromatic, the only color changes are between adjacent sequences of the form R_i and R_{i+1}, S_i and S_{i+1}, or R_n and S_1. So the total number of color changes in $F_{initial}$ is at most $2n$.

However, as mentioned above, some of the edge pairs may be monochromatic. If an edge $e = \{i, j\}$ is not separated by the cut $(U, \overline{U})$, then all T pairs, corresponding to the copies of this edge in G', are monochromatic. To fix this, we flip the color of one of the endpoints in every such pair. Call this new coloring F^*, which is feasible by construction. Observe that this process incurs an additional cost of at most $2T$ for each uncut edge in G, since flipping the color of a point may make both its neighboring edges in W two-colored. Since the total number of uncut edges in the Min Uncut instance is M, and handling each one incurs an extra cost of at most $2T$, the number of color changes in F^* is at most $2MT$ greater than that in $F_{initial}$, and hence at most $2MT + 2n$.

Lemma 3. *Suppose all the permutations σ_i chosen in the reduction are dispersive. Then given a feasible coloring for Γ_G with C color changes, we can construct a feasible solution to minimum uncut on G with cost at most $33\,(C/T)$. Thus we have* $\mathsf{Opt}(G) \leq \frac{33}{T} \cdot \mathsf{Opt}(\Gamma_G)$.

Proof. Consider some feasible coloring F for Γ_G, with cost (a.k.a. number of color changes) C. We will construct a feasible solution to minimum uncut on G where the number of uncut edges is at most $O(C/T)$.

For each $i \in [n]$, define the *majority color* of i (under the feasible coloring F for Γ_G) to be the one that is represented at least $|R_i|/2$ times among the points in R_i; the *minority color* of i is defined to be the opposite color. Also, for $i \in [n]$, let $q_i \leq |R_i|/2$ denote the number of points in R_i colored with the minority color. Let $U \subseteq [n]$ denote the vertices i having majority color B, and hence $\overline{U}$ is the set of vertices having majority color R. In the rest of the proof, we show that the cost of the solution $(U, \overline{U})$ to Min Uncut is $D \leq 33 \cdot \frac{C}{T}$.

Consider any uncut edge (u, v) in this solution $(U, \overline{U})$. There are T edge-pairs between points of R_u and R_v. By the feasibility of F, these must be colored with opposite colors. Since u and v are either both in U or both in $\overline{U}$, they have the same majority color: so the $2T$ points in $R_u \cup R_v$ corresponding to edge (u, v) contribute at least T minority colors at u and v. Thus the total number of points colored with the minority color is $\sum_{i=1}^{n} q_i \geq T \cdot D$. Rearranging,

$$D \quad \leq \quad \frac{1}{T} \sum_{i=1}^{n} q_i. \tag{4}$$

Fix any $i \in [n]$. Since q_i equals the number of minority colors in R_i, by Definition 1, the coloring $F[R_i]$ is q_i-non-monochromatic. Now we can use the dispersive property of the permutation σ_i to prove the following claim.

Claim 5 *For any $i \in [n]$, let the total number of color changes in R_i and S_i be C_i. Then $q_i \leq 33\, C_i$.*

Proof. Let $R_i := X_i \cup Y_i$ where X_i and Y_i consist of the x-points and y-points, respectively. Let $F[Y_i]$ be the coloring F restricted to the subsequence of the y-points. Say the minority color in $F[Y_i]$ is B, and let B_y denote the number of B-colored points in Y_i. This implies that $F[Y_i]$ is B_y-non-monochromatic.

We first claim that $B_y \leq 16C_i$. Suppose (for a contradiction) that $B_y > 16C_i$. Then since σ_i is dispersive, either $F[Y_i]$ or $\overline{F}[Y_i \circ \sigma_i]$ has at least $B_y/16 > C_i$ color changes. Note that $\overline{F}[Y_i \circ \sigma_i]$ is precisely the coloring $F[S_i]$ since the points in Y_i and S_i are paired (under permutation σ_i) and so have opposite colors by the feasibility of F. Note also that the number of color changes in $F[R_i]$ is at least that number in $F[Y_i]$ since Y_i is a subsequence of R_i. It follows that the number of color changes in $F[R_i]$ and $F[S_i]$ is greater than C_i, giving us a contradiction. Thus we must have $B_y \leq 16C_i$.

Finally, the sequence R_i is obtained by alternating between X_i and Y_i. Since the number of color changes in $F[R_i]$ is at most C_i, the number of B-colored points in $F[X_i]$ and $F[Y_i]$ must differ by at most C_i. The latter quantity is at most $16C_i$ by the previous argument, so the number of B-colored points in $F[X_i]$ is at most $17C_i$. It follows that the total number of B-colored points in R_i is at most $33C_i$, and hence $q_i \leq 33C_i$.

Using the bound from Claim 5 for each $i \in [n]$, along with (4),

$$D \quad \leq \quad \frac{1}{T} \sum_{i=1}^{n} q_i \quad \leq \quad \frac{33}{T} \sum_{i=1}^{n} C_i \quad \leq \quad 33 \frac{C}{T}.$$

This completes the proof of Lemma 3.

Completing the proof of Theorem 4. Having related the optima for the Min Uncut instance G and the Binary Paintshop instance Γ_G, we can now prove Theorem 4.

Proof. We may assume that $\mathsf{Opt}(G) \geq 1$: the case $\mathsf{Opt}(G) = 0$ corresponds exactly to checking that G is bipartite and this can be easily done in polynomial time. By the assumption of the theorem, there is a ρ-approximation algorithm for Binary Paintshop, where $\rho \geq 1$. Choosing $T = \max\{10^8, n\}$ means that the probability of some fixed permutation chosen in the gadget not being dispersive is at most $\frac{1}{512n}$; by a union bound, the probability that all permutations chosen are dispersive is at least $1 - 1/512 > 0.99$. If we run the claimed ρ-approximation algorithm for the Binary Paintshop instance Γ_G, we get a feasible coloring for Γ_G with $C \leq \rho\,\mathsf{Opt}(\Gamma_G)$ color changes. By Lemma 2 we know that $\mathsf{Opt}(\Gamma_G) \leq 2T\,\mathsf{Opt}(G) + 2n$. Using the "decoding" algorithm from Lemma 3, we can now construct a feasible solution to the Min Uncut instance G with cost at most

$$33\,\frac{C}{T} \;\leq\; 33\rho\,\frac{\mathsf{Opt}(\Gamma_G)}{T} \;\leq\; 33\rho\left(\frac{2T\,\mathsf{Opt}(G) + 2n}{T}\right) \;\leq\; 132\,\rho\,\mathsf{Opt}(G),$$

since $\mathsf{Opt}(G) \geq 1$ and $T \geq n$. This completes the proof of Theorem 4.

3.2 Reducing Binary Paintshop to Minimum Uncut

We now give a reduction in the opposite direction, showing that the Binary Paintshop problem is essentially a special case of Min Uncut. In fact, we show that even the *generalized* Binary Paintshop problem can be solved using an algorithm for Min Uncut.

Theorem 6. *Given a ρ-approximation algorithm for the Min Uncut problem, we get a ρ-approximation algorithm for generalized Binary Paintshop.*

Proof. Consider an instance of generalized Binary Paintshop: recall that this consists of a graph $G = (V, E_G)$, and the constraint graph $H = (V, E_H)$, and we want to find a coloring that bi-colors each of the edges in E_H while cutting the fewest edges in E_G. Let C denote the cost of the optimal solution to this instance; let $|V_G| = n$ and hence $C \leq |E_G| \leq \binom{n}{2}$.

Let $H_1, H_2, \ldots, H_k$ be the k connected components of $H = (V, E_H)$. Since H is bipartite, it is easy to see that each H_i is bipartite with a unique bipartition of vertices we denote by (V_i^0, V_i^1), $i = 1, \ldots, k$. The instance $I = (U, E_I)$ of Min Uncut is constructed as follows. The vertex set $U := \cup_{i=1}^k \{u_i^0, u_i^1\}$. For each $i = 1, \ldots, k$, add $t := \rho n^2$ edges between u_i^0 and u_i^1 in graph I.

Consider an edge $e \in E_G$ whose end points lie in V_i^a and V_j^b for some $i \leq j$ and $a, b \in \{0, 1\}$. Add a corresponding edge in graph I between $u_i^{a'}$ and u_j^b, where $a' = 1 - a$. Note that this may lead to self-loops in I. The number of edges in I is $|E_G| + tk$.

Consider a feasible solution to the Binary Paintshop instance. This solution colors all the nodes in V_i^0 the same color, and the opposite color for all the nodes in V_i^1, for $i = 1, \ldots, k$. To construct a solution for the Min Uncut instance I, add each u_i^a to one side of the cut according to the color of V_i^a for $i = 1, \ldots, k$ and $a \in \{0, 1\}$. This separates all the edges between u_i^0 and u_i^1. The only edges that may remain uncut are those corresponding to edges in E_G. An edge $e \in E_G$ with end points in V_i^a and V_j^b for some $i \leq j$ and $a, b \in \{0, 1\}$ corresponds to an edge $e' \in E_I$ between $u_i^{a'}$ and u_j^b, where $a' = 1 - a$. Thus, e' is cut in the uncut solution if and only if e was monochromatic in the Binary Paintshop solution. That is, the number of edges in E_I that are not separated (the uncut objective value) is exactly equal to the number of non-monochromatic edges in E_G (the paintshop objective value).

Conversely, any solution to the Min Uncut instance I that separates each pair $\{u_i^0, u_i^1\}$ $(i = 1, \ldots, k)$ can be turned into a feasible solution for the Binary Paintshop instance of the same cost by coloring each V_i^a according to the side of the cut containing u_i^a for $a \in \{0, 1\}$. Now, the output of the ρ-approximation algorithm for Min Uncut on I must separate each pair $\{u_i^0, u_i^1\}$: else, its cost is at least $t = \rho n^2 > \rho C$, contradicting the fact that the output cut is a ρ-approximation.

Combining Theorems 4 and 6, we get that the approximability of these two problems is the same (up to constant factors) under randomized reductions.

Remark. It is easy to extend our results to weighted versions of the Binary Paintshop problem where each adjacent pair in the sequence comes with some cost.

4 UGC Hardness of Approximation

In this section we shall prove the desired Unique Games Conjecture based hardness of approximation for the Binary Paintshop problem via the above connection to the Min Uncut problem. We begin by stating the current best UGC based inapproximability result for Min Uncut. The following theorem is based on UGC for *regular* Unique Games – where the degree is a constant depending on the *completeness* and *soundness* of the instance[4].

Theorem 7 ([8]). *Assuming the Unique Games Conjecture [7] the following holds. For every constant $\varepsilon > 0$, there is a positive integer $d := d(\varepsilon)$ such that given a d-regular n-vertex graph G as an instance of Min Uncut, it is NP-hard to decide between the following two cases:*

1. *YES Case:* $\mathsf{Opt}(G) \leq \frac{1}{2}\varepsilon n d$.
2. *NO Case:* $\mathsf{Opt}(G) \geq \frac{1}{4}\sqrt{\varepsilon} n d$.

The formal statement of the reduction is given below and, along with Theorem 7, implies Theorem 1.

[4] Applying a pre-processing step of Dinur (Lemma 4.1 of [5]) using constant degree expanders followed by Parallel Repetition [10], a general instance of Unique Games can be deterministically converted to a regular instance and thus, one can assume UGC holds for regular instances of Unique Games.

Theorem 8. *There is a polynomial-time (deterministic) reduction from instances G of Min Uncut given by Theorem 7 with parameter $\varepsilon > 0$, to instances Γ of the Binary Paintshop problem on a sequence of length $N = poly(n)$ such that,*

1. *YES Case: If G is a YES instance of Min Uncut then $\mathsf{OPT}(\Gamma) \leq \varepsilon N$*
2. *NO Case: If G is a NO instance of Min Uncut then $\mathsf{OPT}(\Gamma) \geq \frac{1}{400}\sqrt{\varepsilon}N$.*

Proof. We instantiate the reduction of Section 3.1 with $T = \max\{10^8, 1/\varepsilon\}$. Since all nodes have the same degree d, we only require a single dispersive permutation $\sigma : [dT] \to [dT]$. Since dT is a constant (depending on ε), such a permutation can be found using a brute-force search. Note that the length of the Binary Paintshop instance generated is $N = 3ndT$.

Now if G is a YES instance of Min Uncut then Lemma 2 implies that

$$\mathsf{Opt}(\Gamma) \;\leq\; 2T\mathsf{Opt}(G) + 2n \;\leq\; \varepsilon ndT + 2n \;\leq\; \varepsilon N.$$

If G is a NO instance of Min Uncut then Lemma 3 implies that

$$\mathsf{Opt}(\Gamma) \;\geq\; \frac{T}{33}\mathsf{Opt}(G) \;\geq\; \frac{1}{132}\sqrt{\varepsilon}ndT \;\geq\; \frac{1}{400}\sqrt{\varepsilon}N.$$

5 A Ternary Paintshop Problem

A natural extension of the (generalized) Binary Paintshop problem is to the case where there are three or more colors. The goal is again to ensure that the constraint edges are not monochromatic, and to minimize the number of bichromatic edges in G. One natural hurdle in this case is that checking whether there exists a feasible solution becomes NP-hard, since we would have to check whether the given constraint graph H is k-colorable for $k \geq 3$. However, even when the constraints are trivially k-colorable, and the graph G is very simple, we show that the problem remains very hard to approximate.

Specifically, consider the ternary case with $k = 3$ colors, where the underlying graph G is a collection of disjoint paths, and the constraint graph H is a matching. We show it is NP-hard to identify whether the optimal cost is zero or not, and hence NP-hard to approximate to any factor. Indeed, take a graph $G_c = (V_c, E_c)$ that is an instance of 3-coloring, and construct an instance (G, H) of ternary paintshop as follows: for each vertex $v \in V_c$, construct $\delta(v)$ vertices in $V(G)$, with one copy corresponding to each edge $e \in E_c$ incident to v, and connect all these $\delta(v)$ copies by a path P_v. These $\sum_{v \in V_c} \delta(v) = 2|E_c|$ vertices and $\sum_{v \in V_c}(\delta(v) - 1) = 2|E_c| - |V_c|$ edges form the graph G. Now for each edge $e = (u, v) \in E_c$, add a constraint edge in H between the corresponding copies of u and v in $V(G)$ — hence the constraint edges H form a matching. Now G_c has a valid 3-coloring if and only if the ternary paintshop instance (G, H) has a solution that monochromatically colors each of the $|V_c|$ paths and cuts zero edges.

In the above reduction we crucially used the fact that G was a forest, and had disconnected components. This can be remedied to show a slightly weaker hardness result. Define the *(basic) ternary paintshop* problem as follows: given a sequence of length n (again associated with the integers $[n] = \{1, 2, \ldots, n\}$), and a matching H on the points $[n]$, find a coloring f that ensures that each edge in H is bichromatic, and minimizes the number of bichromatic pairs $(i, i+1)$.

Theorem 9. *For any constant $\varepsilon > 0$, the basic ternary paintshop problem is NP-hard to approximate to within $n^{1-\varepsilon}$ in polynomial time.*

Proof. Consider the same reduction from a 3-coloring instance $G_c = (V_c, E_c)$ to the ternary paintshop instance (G, H) on disconnected paths, but now take T copies of the graphs (G, H). Obtain a sequence of length $N := T \cdot |V(G)|$ by considering the vertices of V_c in some order, and laying down all the paths P_v for the same vertex in each of these T copies consecutively; the constraints are inherited from the original instances (G, H). This gives the instance Γ for the basic ternary paintshop problem with sequence length N. Note that a 3-coloring for G_c naturally gives a ternary coloring of $[N]$ with at most $|V_c| - 1$ many color changes. On the other hand, if G_c is not 3-colorable, each of the T instances of (G, H) must incur at least one color change, and the number of color changes in Γ is at least T. If $n := |V_c|$, then setting $T = n^{2/\varepsilon}$ means it is NP-hard to distinguish the case when the optimum is at most $n \approx N^{\varepsilon}$ and when it is at least $T \approx N^{1-\varepsilon}$, giving us the claimed hardness.

A different version of the ternary paintshop problem (e.g., from [9]) is where the constraints are hyperedges of size 3, and also form a matching — i.e., none of the hyperedges share vertices from $[n]$. A constraint $\{i, j, k\}$ now means the three vertices i, j, k must be given distinct colors. The reduction from Theorem 9 easily extends to show hardness for this variant too: for each constraint $e = \{i, j\}$ in that reduction, add a new dummy vertex v_e and use the constraint $\{i, j, v_e\}$. Other extensions considered in previous papers, with constraints of the form *"the set $S \subseteq [n]$ must contain exactly t_{iS} nodes of color i for each color $i \in [k]$, where $\sum_{i \in [k]} t_{iS} = |S|$"*, are thus at least as hard.

References

1. Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{\log n})$ approximation algorithms for min UnCut, min 2CNF deletion, and directed cut problems. In 37*th Annual ACM Symposium on Theory of Computing*, pages 573–581, 2005.
2. Hadis Amini, Frédéric Meunier, Héloïse Michel, and Atefeh Mohajeri. Greedy colorings for the binary paintshop problem. *Journal of Discrete Algorithms*, 8(1):8–14, 2010.
3. Stephan Dominique Andres and Winfried Hochstättler. Some heuristics for the binary paint shop problem and their expected number of colour changes. *J. Discrete Algorithms*, 9(2):203–211, 2011.
4. Paul S. Bonsma, Thomas Epping, and Winfried Hochstättler. Complexity results on restricted instances of a paint shop problem for words. *Discrete Applied Mathematics*, 154(9):1335–1343, 2006.

5. Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3):12, 2007.
6. Thomas Epping, Winfried Hochstättler, and Peter Oertel. Some results on a Paint Shop problem for words. *Electronic Notes in Discrete Mathematics*, 8:31–33, 2001.
7. Subhash Khot. On the power of unique 2-prover 1-round games. In *34th Annual ACM Symposium on the Theory of Computing*, pages 767–775, July 2002.
8. Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.
9. Frédéric Meunier and András Sebö. Paintshop, odd cycles and necklace splitting. *Discrete Applied Mathematics*, 157(4):780–793, 2009.
10. Ran Raz. A parallel repetition theorem. *SIAM Journal of Computing*, 27(3):763–803, 1998.