Diskrete Mathematik und Optimierung

Stephan Dominique Andres, Winfried Hochstättler:

Some heuristics for the binary paint shop problem and their expected number of colour changes

Technical Report feu-dmo022.10
Contact: dominique.andres@fernuni-hagen.de,
winfried.hochstaettler@fernuni-hagen.de

FernUniversität in Hagen
Fakultät für Mathematik und Informatik
Lehrgebiet für Diskrete Mathematik und Optimierung
D – 58084 Hagen

2000 Mathematics Subject Classification: 90C59, 90B35, 68R15
Keywords: greedy algorithm, paint shop problem, average performance

Some heuristics for the binary paint shop problem and their expected number of colour changes

Stephan Dominique Andres[1]
Winfried Hochstättler[2]
[1,2]Fakultät für Mathematik und Informatik
FernUniversität in Hagen, D-58084 Hagen, Germany
[1,2]Email: {dominique.andres,winfried.hochstaettler}@fernuni-hagen.de

August 6, 2010

Abstract

In the binary paint shop problem we are given a word on n characters of length $2n$ where every character occurs exactly twice. The objective is to colour the letters of the word in two colours, such that each character receives both colours and the number of colour changes of consecutive letters is minimized. Amini et. al proved that the expected number of colour changes of the heuristic greedy colouring is at most $2n/3$. They also conjectured that the true value is $n/2$. We verify their conjecture and, furthermore, compute an expected number of $2n/3$ colour changes for a heuristic, named *red first*, which behaves well on some worst case examples for the greedy algorithm. From our proof method, finally, we derive a new recursive greedy heuristic which achieves an average number of $2n/5$ colour changes.

Key words: greedy algorithm, paint shop problem, average performance
MSC: 90C59, 90B35, 68R15

1 Introduction

Motivated by the problem of minimizing colour changes in a paint shop in an automobile plant Epping, Hochstättler and Oertel [3] introduced the following problem:

PPW(2,1)[Binary Paint Shop Problem]: Let Σ be a finite alphabet of cardinality n. We call the elements of Σ *characters* and their occurrence in a word a *letter*. An instance of $PPW(2,1)$ is a word $w = (w_1, \ldots, w_{2n}) \in \Sigma^{2n}$ in which each character occurs exactly twice. A *feasible colouring* is a colouring $(f_1, \ldots, f_{2n}) \in \{0,1\}^{2n}$ with the property that $i \neq j$ and $w_i = w_j$ implies

$\{f_i, f_j\} = \{0, 1\}$. The objective is to find a feasible colouring such that the number $\sum_{i=1}^{2n-1} |f_i - f_{i+1}|$ of colour changes is minimized.

Bonsma, Epping and Hochstättler [2] and, by a different reduction, Meunier and Sebő [4] showed that $PPW(2,1)$ is $\mathcal{APX}$-hard. It is not known whether $PPW(2,1)$ is also $\mathcal{APX}$-easy, i.e. whether it admits an approximation with a guaranteed constant ratio.

The *greedy algorithm* for $PPW(2,1)$ is a feasible colouring that colours w_1 with 0 and then, processing the word from left to right, the first occurrence of each other character with the colour of its predecessor, the second occurrence with the remaining colour.

Amini, Meunier, Michel and Mojaheri [1] showed that, assuming a uniform distribution on the words of length $2n$, the expected number $\mathbb{E}_n(g)$ of colour changes computed by the greedy algorithm is

$$\mathbb{E}_n(g) \leq \frac{2n}{3}.$$

They conjectured that $\mathbb{E}_n(g) \sim \frac{n}{2}$. We verify this conjecture and show:

Theorem 1. *If the instances of the paintshop problem of size $2n$ are uniformly distributed, then the expected number of colour changes computed by the greedy algorithm is given by*

$$\mathbb{E}_n(g) = \sum_{k=0}^{n-1} \frac{2k^2 - 1}{4k^2 - 1}.$$

Unfortunately, the greedy algorithm does not approximate the optimum value within a constant factor. We will present a class of examples where it behaves poorly. This class is coloured optimally by *red first*, a heuristic which colours the first occurence of each character by 0 and the other with 1. As is to be expected, this heuristic on average is not competitive with the greedy algorithm:

Theorem 2. *If the instances of the paintshop problem of size $2n$ are uniformly distributed, then the expected number of colour changes computed by red first is given by*

$$\mathbb{E}_n(rf) = \frac{2n+1}{3}.$$

The key idea in the analysis of the above heuristics inspired a third algorithm which is a recursive application of the greedy idea. This heuristic behaves substantially better than the greedy algorithm:

Theorem 3. *If the instances of the paintshop problem of size $2n$ are uniformly distributed, then the expected number of colour changes computed by the recursive greedy algorithm is bounded as follows.*

$$\frac{2}{5}n + \frac{8}{15} \leq \mathbb{E}_n(rg) \leq \frac{2}{5}n + \frac{7}{10}.$$

The paper is organized as follows. In Section 2 first we count the patterns of consecutive colours in a feasible colouring and use this to prove Theorem 1. Furthermore, we present a class of worst case examples with unbounded performance ratio for this heuristic. In Section 3 we proceed analogously with red first. In Section 4 we present the recursive greedy heuristic and prove Theorem 3. We conclude with some open questions.

2 The greedy algorithm

If $(w_1, \ldots, w_{2n})$ is an instance of $PPW(2,1)$ and $(f_1, \ldots, f_{2n})$ a feasible colouring, then for $i = 1, \ldots, 2n-1$ we call $w_i w_{i+1}$ a *gap of type* $f_i f_{i+1}$

Lemma 4. *Let w be an instance of the paint shop problem with length $2n$, which is feasibly coloured and where $f_1 = 0$. Let k be the number of colour changes in w. If k is even, then the number of gaps of the four possible types is given by the following table:*

type	01	10	11	00
#gaps	$\frac{k}{2}$	$\frac{k}{2}$	$n - \frac{k}{2}$	$n - 1 - \frac{k}{2}$

If k is odd, we have

type	01	10	11	00
#gaps	$\frac{k+1}{2}$	$\frac{k-1}{2}$	$n - \frac{k+1}{2}$	$n - \frac{k+1}{2}$

Proof. Assume that k is even. Clearly, half of the colour changes produce a gap of type 01 and the others of type 10. The number of gaps of type 00 or 11 in total must be $2n - 1 - k$ and hence be odd. Since we have the same number of 0s and 1s but the first and the last letter are coloured with 0, we have one more gap of type 11 than of 00.

If k is odd, we have one more gap of type 01 than of 10 and the same number of type 11 as of type 00. □

We now turn to a first simple observation about the behaviour of the greedy colouring.

Lemma 5. *Assume the instances of the paint shop problem of size $2n$ are uniformly distributed, w is such an instance and Z the character at position $2n$. If greedy colouring is applied to w, then the probability that the first occurrence of Z is coloured with 0 is $\frac{n}{2n-1}$, and with probability $\frac{n-1}{2n-1}$ it is coloured with 1.*

Proof. Let w' be the word obtained from w by deleting both occurrences of Z. There are $2n-1$ words w leading to the same word w' since the first occurrence of Z can be placed after each character of w' and at the beginning of w'. The first occurrence of Z is colored with 0 if it is placed in a gap of type 00 or 01 or at the beginning, or if k is even and it is placed at the end. Otherwise it is coloured with 1. Thus, the claim follows from Lemma 4. □

As a consequence of the recursive construction in the proof of Lemma 5 we conclude:

Proposition 6. *The number of different instances of the paint shop problem of length $2n$ is*

$$\prod_{i=1}^{n}(2i-1).$$

Using Lemma 5 we can prove Theorem 1.

Proof of Theorem 1. We prove $\mathbb{E}_n(g) = \sum_{k=0}^{n-1} \frac{2k^2-1}{4k^2-1}$ by induction on n. For $n = 1$ it is trivial. Now let $n > 1$ and the assertion be proved for $n-1$. Let w be a word of length $2n$ and Z be the last letter of w. Let w' be the word obtained from w by deleting both occurrences of Z. Let Z' be the last letter of w'. Then w has the same number of colour changes as w' if the first occurrence of Z and the last occurrence of Z' are coloured differently, and w has one additional colour change if the first occurrence of Z and the last occurrence of Z' are coloured the same, since in the latter case there is an additional colour change before the second occurrence of Z.

By Lemma 5, the first occurrence of Z is 0 with probability $\frac{n}{2n-1}$, and the last occurrence of Z' is 0 with probability $\frac{n-2}{2n-3}$. This gives a probability of

$$p_0 = \frac{n}{2n-1} \cdot \frac{n-2}{2n-3} = \frac{n^2-2n}{4n^2-8n+3}$$

for the case that both are 0. Analogously, by Lemma 5, the first occurrence of Z is 1 with probability $\frac{n-1}{2n-1}$, and the last occurrence of Z' is 1 with probability $\frac{n-1}{2n-3}$. This gives a probability of

$$p_1 = \frac{n-1}{2n-1} \cdot \frac{n-1}{2n-3} = \frac{n^2-2n+1}{4n^2-8n+3}$$

for the case that both are 1. In total we have one additional colour change with probability

$$p_0 + p_1 = \frac{2n^2-4n+1}{4n^2-8n+3} = \frac{2(n-1)^2-1}{4(n-1)^2-1}.$$

This implies, using the induction hypothesis, that

$$\mathbb{E}_n(g) = \mathbb{E}_{n-1}(g) + \frac{2(n-1)^2-1}{4(n-1)^2-1} = \sum_{k=0}^{n-1} \frac{2k^2-1}{4k^2-1}.$$

Thus the theorem is established. □

The following corollary was conjectured by Amini et al. [1].

Corollary 7.

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}_n(g) = \frac{1}{2}.$$

Proof. This follows from

$$\sum_{k=0}^{n-1} \frac{2k^2-1}{4k^2-1} = \frac{n}{2} + O(1).$$

□

While the average performance of the greedy algorithm seems to be quite reasonable, it is not a constant factor approximation algorithm. This is seen considering the following class of instances.

Let $A_1, \ldots, A_{2k}$ be different characters and

$$a = A_1, \ldots, A_k, A_k, A_{k+1}, \ldots, A_{2k}, A_{2k}, A_1, A_{k+1}, A_2, A_{k+2}, \ldots, A_{k-1}, A_{2k-1}.$$

Greedy colouring requires $2k = n$ colour changes, the optimal solution only 3 (between $A_k A_k$, $A_k A_{k+1}$, and $A_{2k} A_{2k}$).

3 Red first colouring

The worst case examples for the greedy colouring are coloured optimally, if we color the first occurrence of each character with 0. We call this strategy *red first colouring* since in the previous work on the paint shop problem the colours where usually named red and blue.

Again we consider the uniform distribution on the instances of the paint shop problem on words of length $2n$. We denote by p_{2k+1}^n the probability that red first colouring produces exactly $2k+1$ colour changes. In particular,

$$p_{-1}^n = 0 \quad \text{and} \quad p_{2n-1}^{n-1} = 0. \tag{1}$$

Note that the number of colour changes of red first colouring is always odd since the first letter is coloured 0, the last 1.

Lemma 8. *For all $n \geq 2$,*

$$p_{2k+1}^n = \frac{n-k}{2n-1} p_{2k-1}^{n-1} + \frac{n+k}{2n-1} p_{2k+1}^{n-1}.$$

Proof. Let w be a word of length $2n$ for which red first colouring uses exactly $2k+1$ colour changes. Let Z be the last letter of w, and w' be the word obtained from w by deleting both occurrences of Z. There are two additional colour changes if and only if the first occurrence of Z is in a gap of type 11 or at the end of w'. By Lemma 4 the probability that Z is inserted into such a gap in a word of length $2n-2$ with $2k-1$ color changes is

$$\frac{1}{2n-1}\left(n-1-\frac{2k-1+1}{2}+1\right) = \frac{n-k}{2n-1}.$$

Analogously, the probability that the first occurrence of Z is inserted into a gap of type 01, 10, 00 or at the beginning in a word of length $2n-2$ with $2k+1$ color changes – causing no additional colour change – is

$$\frac{1}{2n-1}\left(\frac{2k+1+1}{2}+\frac{2k+1-1}{2}+n-1-\frac{2k+1+1}{2}+1\right)=\frac{n+k}{2n-1}.$$

The claim follows, since the probability that w' is coloured with $2k-1$ resp. $2k+1$ colours is p^{n-1}_{2k-1} resp. p^{n-1}_{2k+1}. □

Applying this recursion we now prove Theorem 2.

Proof of Theorem 2. We prove $\mathbb{E}_n(rf)=\frac{2n+1}{3}$ by induction on n. For $n=1$ the assertion is trivial. Now let $n>1$ and the assertion be proved for $n-1$. Using Lemma 8 and the induction hypothesis we obtain

$$\begin{aligned}
\mathbb{E}_n(rf) &= \sum_{k=0}^{n-1}(2k+1)p^n_{2k+1} \\
&\overset{\text{Lemma 8}}{=} \sum_{k=0}^{n-1}(2k+1)\left(\frac{n-k}{2n-1}p^{n-1}_{2k-1}+\frac{n+k}{2n-1}p^{n-1}_{2k+1}\right) \\
&= \sum_{k=-1}^{n-2}(2k+3)\frac{n-k-1}{2n-1}p^{n-1}_{2k+1}+\sum_{k=0}^{n-1}(2k+1)\frac{n+k}{2n-1}p^{n-1}_{2k+1} \\
&\overset{(1)}{=} \sum_{k=0}^{n-2}\left((2k+1)\frac{2n-1}{2n-1}-\frac{2k+1}{2n-1}+\frac{2n-1}{2n-1}\right)p^{n-1}_{2k+1} \\
&= \frac{2n-2}{2n-1}\sum_{k=0}^{n-2}(2k+1)p^{n-1}_{2k+1}+\sum_{k=0}^{n-2}p^{n-1}_{2k+1} \\
&= \frac{2n-2}{2n-1}\mathbb{E}_{n-1}(rf)+1 \\
&\overset{\text{i.h.}}{=} \frac{2n-2}{2n-1}\cdot\frac{2n-1}{3}+1 \\
&= \frac{2n+1}{3}.
\end{aligned}$$

This proves the theorem. □

Corollary 9.

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_n(rf)=\frac{2}{3}.$$

Red first colouring is not a constant factor approximation, too. Consider the word

$$b=B_1,B_2,\ldots,B_k,B_1,B_{k+1},B_2,B_{k+2},\ldots,B_k,B_{2k},B_{k+1},B_{k+2},\ldots,B_{2k}.$$

Red first colouring yields $2k+1 = n+1$ colour changes, while the optimal solution requires only two (between $B_k B_1$ and B_{2k}, B_{k+1}). Note that greedy colouring obtains the optimal solution in this example.

If we concatenate the word a from Section 2 and the word b (assuming that $B_i \neq A_j$ for all i, j), then we obtain a class of examples which show that the heuristic "best of" greedy colouring and red first colouring is not a constant factor approximation to the optimal solution.

4 Recursive greedy colouring

The analysis of the greedy colouring indicates that it is suboptimal only to consider the predecessor of a letter. If we insert the first occurrence of our "last letter" Z into a gap of type 01 or 10, an additional colour change can always be avoided. This idea yields a feasible colouring with only two colour changes when applied to the word $ACABBC$, while the greedy algorithm needs three colour changes. Formally, we arrive at the recursive algorithm displayed in Table 1.

```
recursive_ greedy_colouring(w)
  if  length(w) = 2:
    colour w with 01
  else:
    let w' arise from w by deleting both occurences of the last character Z
    recursive_greedy_colouring(w')
    if  the first occurrence of Z is at the beginning of w':
      colour the first Z with 0
    if  the first occurrence of Z is at the end of w':
      colour the first Z with the colour of its predecessor
    let k denote the number of colour changes of w'
    if  k is odd:
      if  the first occurrence of Z is in a gap of type 11:
        colour the first Z with 1
      if  the first occurrence of Z is in a gap of type 00, 01 or 10:
        colour the first Z with 0
    else:
      if  the first occurrence of Z is in a gap of type 00:
        colour the first Z with 0
      if  the first occurrence of Z is in a gap of type 11, 01 or 10:
        colour the first Z with 1
    colour the second Z accordingly
```

Table 1: The recursive greedy algorithm

Note that in one recursive call an additional colour change will occur only if either k is odd and the first Z is inserted into a gap of type 11 or at the end or k is even and the first Z is in a gap of type 00 or in the beginning or at the end.

Again, we consider the uniform distribution on the instances of the paint shop problem on words of length $2n$. We denote by q_k^n the probability that recursive greedy colouring produces exactly k colour changes. In particular,

$$q_{-1}^n = 0 \quad \text{and} \quad q_{2n-1}^{n-1} = 0. \tag{2}$$

Lemma 10. *For $k \geq 2$,*

$$q_{k+1}^n = \frac{n - \lceil \frac{k}{2} \rceil}{2n-1} q_k^{n-1} + \frac{n + \lfloor \frac{k}{2} \rfloor}{2n-1} q_{k+1}^{n-1}.$$

Proof. Let w be a word of length $2n$ and w' be the word where both occurrences of the last character Z are deleted. As in Lemma 5, there are exactly $2n-1$ words w leading to the same word w'. Let $k+1$ be the number of colour changes in w by recursive greedy colouring.

If k is even, then w has an additional colour change wrt. w' if and only if the first occurrence of Z is inserted into a gap of type 00 or at the beginning or at the end. By Lemma 4 this happens in

$$\left(n - 2 - \frac{k}{2}\right) + 1 + 1 = n - \frac{k}{2}$$

cases. w has the same number of colour changes as w' if the first occurrence of Z is inserted between 01, 10, 00 or at the beginning. Since $k+1$ is odd this gives, by Lemma 4,

$$\frac{k+2}{2} + \frac{k}{2} + \left(n - 1 - \frac{k+2}{2}\right) + 1 = n + \frac{k}{2}$$

cases. In total we obtain

$$q_{k+1}^n = \frac{n - \frac{k}{2}}{2n-1} q_k^{n-1} + \frac{n + \frac{k}{2}}{2n-1} q_{k+1}^{n-1}.$$

If k is odd, then w has an additional colour change wrt. w' if the first occurrence of Z is inserted between 11 or at the end. This gives, by Lemma 4,

$$\left(n - 1 - \frac{k+1}{2}\right) + 1 = n - \frac{k+1}{2}$$

cases. w has the same number of colour changes as w' if the first occurrence of Z is inserted between 11, 10, or 01. Since $k+1$ is even Lemma 4 yields,

$$\frac{k+1}{2} + \frac{k+1}{2} + \left(n - 1 - \frac{k+1}{2}\right) = n + \frac{k-1}{2}$$

cases. In total we obtain

$$q_{k+1}^n = \frac{n - \frac{k+1}{2}}{2n-1} q_k^{n-1} + \frac{n + \frac{k-1}{2}}{2n-1} q_{k+1}^{n-1}.$$

In both cases this implies the assertion. □

Lemma 11. *Let rg be the number of colour changes for recursive greedy colouring. Then, for all $n \geq 2$,*

$$\mathbb{E}_n(rg) = \frac{4n-3}{4n-2}\mathbb{E}_{n-1}(rg) + \frac{n - \frac{1}{4} - \frac{1}{4}\sum_{k=0}^{n-2}(-1)^k q_{k+1}^{n-1}}{2n-1}.$$

Proof. For $n \geq 2$, by the previous lemma,

$$\begin{aligned}
\mathbb{E}_n(rg) &= \sum_{k=0}^{n-1}(k+1)q_{k+1}^n \\
&= \frac{1}{2n-1}\sum_{k=0}^{n-1}(k+1)\left(\left(n-\left\lceil\frac{k}{2}\right\rceil\right)q_k^{n-1} + \left(n+\left\lfloor\frac{k}{2}\right\rfloor\right)q_{k+1}^{n-1}\right) \\
&\overset{(2)}{=} \frac{1}{2n-1}\left(\sum_{k=1}^{n-1}(k+1)\left(n-\left\lceil\frac{k}{2}\right\rceil\right)q_k^{n-1}\right. \\
&\quad \left. + \sum_{k=0}^{n-2}(k+1)\left(n+\left\lfloor\frac{k}{2}\right\rfloor\right)q_{k+1}^{n-1}\right) \\
&= \frac{1}{2n-1}\left(\sum_{k=0}^{n-2}(k+2)\left(n-\left\lceil\frac{k+1}{2}\right\rceil\right)q_{k+1}^{n-1}\right. \\
&\quad \left. + \sum_{k=0}^{n-2}(k+1)\left(n+\left\lfloor\frac{k}{2}\right\rfloor\right)q_{k+1}^{n-1}\right) \\
&= \frac{1}{2n-1}\sum_{k=0}^{n-2}(k+1)(2n-1)q_{k+1}^{n-1} \\
&\quad + \frac{1}{2n-1}\sum_{k=0}^{n-2}\left(n-\left\lceil\frac{k+1}{2}\right\rceil\right)q_{k+1}^{n-1} \\
&= \mathbb{E}_{n-1}(rg) + \frac{n}{2n-1} - \frac{1}{2n-1}\sum_{k=0}^{n-2}\left(\frac{k+1}{2}+\frac{1}{4}+\frac{1}{4}(-1)^k\right)q_{k+1}^{n-1} \\
&= \frac{4n-3}{4n-2}\mathbb{E}_{n-1}(rg) + \frac{n-\frac{1}{4}-\frac{1}{4}\sum_{k=0}^{n-2}(-1)^k q_{k+1}^{n-1}}{2n-1}.
\end{aligned}$$

□

Lemma 12. *(a) For $n \geq 1$,*

$$\forall k \in \mathbb{N}_0 : \sum_{\ell=k}^{\infty}\left(q_{2\ell+1}^n - q_{2\ell+2}^n\right) \geq 0,$$

moreover

$$\forall 0 \leq k \leq \frac{n-1}{2} : \sum_{\ell=k}^{\infty}\left(q_{2\ell+1}^n - q_{2\ell+2}^n\right) > 0.$$

(b) For $n \geq 2$,

$$\sum_{\ell=0}^{\infty} \left(q^n_{2\ell+1} - q^n_{2\ell+2}\right) \leq \frac{1}{3}.$$

Proof. First note that all occurring sums have only a finite number of nonzero terms. We prove the two assertions of (a) and the assertion of (b) simultaneously by induction on n. For $n = 1$ we have

$$\sum_{\ell=k}^{\infty} \left(q^n_{2\ell+1} - q^n_{2\ell+2}\right) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \geq 1, \end{cases}$$

thus both assertions of (a) hold. For $n = 2$ we have

$$\sum_{\ell=0}^{\infty} \left(q^n_{2\ell+1} - q^n_{2\ell+2}\right) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3},$$

thus the assertion of (b) holds.

Now let $n \geq 2$ (for the proof of (a)) resp. $n \geq 3$ (for the proof of (b)) and all assertions be proved for $n-1$. Then, for any $k \in \mathbb{N}_0$, by Lemma 10

$$\begin{aligned}
& \sum_{\ell=k}^{\infty} \left(q^n_{2\ell+1} - q^n_{2\ell+2}\right) \\
\overset{\text{Lem. 10}}{=} & \sum_{\ell=k}^{\infty} \left(\frac{n-\ell}{2n-1} q^{n-1}_{2\ell} + \frac{n+\ell}{2n-1} q^{n-1}_{2\ell+1} - \frac{n-\ell-1}{2n-1} q^{n-1}_{2\ell+1} - \frac{n+\ell}{2n-1} q^{n-1}_{2\ell+2}\right) \\
= & \underbrace{\frac{n-k}{2n-1} q^{n-1}_{2k}}_{=:X(k)} + \sum_{\ell=k}^{\infty} \left(\frac{n-\ell-1}{2n-1} q^{n-1}_{2\ell+2} - \frac{n+\ell}{2n-1} q^{n-1}_{2\ell+2} + \frac{2\ell+1}{2n-1} q^{n-1}_{2\ell+1}\right) \\
= & X(k) + \sum_{\ell=k}^{\infty} \frac{2\ell+1}{2n-1} \left(q^{n-1}_{2\ell+1} - q^{n-1}_{2\ell+2}\right) \\
=: & X(k) + Y(k).
\end{aligned}$$

In order to prove (b) we note that $X(0) = 0$. Thus we are left to estimate the sum $Y(0)$ which can be done using the induction hypothesis in the following way:

$$Y(0) \leq \frac{2n-1}{2n-1} \sum_{\ell=0}^{\infty} \left(q^{n-1}_{2\ell+1} - q^{n-1}_{2\ell+2}\right) \overset{\text{i.h.}}{\leq} \frac{1}{3}.$$

Thus (b) is established.

For (a), we continue our estimations.

$$\begin{aligned}
& X(k) + Y(k) \\
\geq & Y(k)
\end{aligned}$$

$$\begin{aligned} &= \frac{2k+1}{2n-1}\sum_{\ell=k}^{\infty}\left(q_{2\ell+1}^{n-1}-q_{2\ell+2}^{n-1}\right)+\sum_{s=k+1}^{\infty}\frac{2}{2n-1}\sum_{\ell=s}^{\infty}\left(q_{2\ell+1}^{n-1}-q_{2\ell+2}^{n-1}\right) \\ &\overset{\text{i.h.}}{\geq} 0. \end{aligned}$$

This proves the first assertion of (a). Note that in the proof in the first step $q_{2k}^{n-1}=0$ for $k>\frac{n-1}{2}$ and $q_{2k}^{n-1}>0$ for $1\leq k\leq\frac{n-1}{2}$, which leads to a strict inequality in the first step. Furthermore if $k=0$, then

$$\sum_{\ell=k}^{\infty}\left(q_{2\ell+1}^{n-1}-q_{2\ell+2}^{n-1}\right)>0$$

by the induction hypothesis for the second assertion, which leads to a >0 in the last step. The last two observations together imply the second assertion of (a). □

Corollary 13. *Assume that the instances of length $2n$ are coloured by recursive greedy. Then the probability that the last letter is coloured with 1 is greater than the probability that the last letter is coloured with 0.*

Proof. This follows from Lemma 12 (a) by setting $k=0$. □

Lemma 14. *For all $n\geq 3$,*

$$\frac{4n-3}{4n-2}\mathbb{E}_{n-1}(rg)+\frac{3n-1}{6n-3}\leq\mathbb{E}_n(rg)\leq\frac{4n-3}{4n-2}\mathbb{E}_{n-1}(rg)+\frac{4n-1}{8n-4}$$

Proof. This follows by the formula of Lemma 11, since by Lemma 12 (a) resp. (b)

$$\frac{1}{4}\overset{(a)}{\leq}\frac{1}{4}+\frac{1}{4}\sum_{k=0}^{n-2}(-1)^k q_{k+1}^{n-1}\overset{(b)}{\leq}\frac{1}{4}+\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{3}.$$

□

Now we are able to prove our main theorem.

Theorem 3. *For all $n\geq 1$,*

$$\frac{2}{5}n+\frac{8}{15}\leq\mathbb{E}_n(rg)\leq\frac{2}{5}n+\frac{7}{10}.$$

Proof. We proceed by induction on n. For $n=1$, $\mathbb{E}_n(rg)=1$, thus the assertion holds. For $n=2$, $\mathbb{E}_n(rg)=\frac{4}{3}=\frac{2}{5}\cdot 2+\frac{8}{15}$, thus the assertion holds. Now assume $n>2$ and the assertion be proved for $n-1$. Then by using the upper bound of the previous lemma and the induction hypothesis

$$\begin{aligned} \mathbb{E}_n(rg) &\leq \frac{4n-3}{4n-2}\left(\frac{2}{5}(n-1)+\frac{7}{10}\right)+\frac{4n-1}{8n-4} \\ &= \frac{(16n^2-28n+12)+(28n-21)+(20n-5)}{40n-20} \end{aligned}$$

$$\begin{aligned} &= \frac{16n^2 - 8n}{40n - 20} + \frac{28n - 14}{40n - 20} \\ &= \frac{2}{5}n + \frac{7}{10}. \end{aligned}$$

By the lower bound of the previous lemma and the induction hypothesis

$$\begin{aligned} \mathbb{E}_n(rg) &\geq \frac{4n-3}{4n-2}\left(\frac{2}{5}(n-1) + \frac{8}{15}\right) + \frac{3n-1}{6n-3} \\ &= \frac{(24n^2 - 42n + 18) + (32n - 24) + (30n - 10)}{60n - 30} \\ &= \frac{24n^2 - 12n}{60n - 30} + \frac{32n - 16}{60n - 30} \\ &= \frac{2}{5}n + \frac{8}{15}. \end{aligned}$$

This proves the theorem. □

Corollary 15.

$$\lim_{n \to \infty} \frac{1}{n}\mathbb{E}_n(rg) = \frac{2}{5}.$$

5 Open problems

We remark that recursive greedy colouring solves the second worst case example b to optimality, while it also fails on example a producing the same colouring as the non-recursive greedy colouring.

In particular the following question remains open.

Problem 16. *Is there a constant factor approximation for $PPW(2,1)$?*

Furthermore, Theorem 3 leaves a gap of $\frac{1}{6}$. This is optimal in the following sense. The lower bound $\frac{2}{5}n + \frac{8}{15}$ is attained for $n = 2$ and $n = 3$. Numerical simulations suggest that the difference $E_n(rg) - \frac{2}{5}n$ tends (very slowly) to $\frac{7}{10}$ when $n \longrightarrow \infty$. Therefore the upper bound seems to be best possible.

Another open question, which may be hard, is the following.

Problem 17. *What is the expected value of optimal colouring?*

References

[1] H. Amini, F. Meunier, H. Michel, A. Mohajeri, Greedy colorings for the binary paintshop problem, J. Discrete Algorithms 8 (2010) 8–14

[2] P. Bonsma, Th. Epping, W. Hochstättler, Complexity results on restricted instances of a paint shop problem for words, Discrete Appl. Math. 154 (2006) 1335–1343

[3] Th. Epping, W. Hochstättler, P. Oertel, Complexity result on a paint shop problem, Discrete Appl. Math. 136 (2004) 217–226

[4] F. Meunier, A. Sebő, Paintshop, odd cycles and necklace splitting, Discrete Appl. Math. 157 (2009) 780–793