FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Jan Soukup

Drawing geometric graphs on red-blue point sets

Department of Applied Mathematics

Supervisor of the master thesis: Mgr. Jan Kynčl, Ph.D.
Study programme: Computer Science
Study branch: Discrete Models and Algorithms

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

Title: Drawing geometric graphs on red-blue point sets

Author: Jan Soukup

Department: Department of Applied Mathematics

Supervisor: Mgr. Jan Kynčl, Ph.D., Department of Applied Mathematics

Abstract: Consider a set B of blue points and a set R of red points in the plane such that $R \cup B$ is in general position. A graph drawn in the plane whose edges are straight-line segments is called a geometric graph. We investigate the problem of drawing non-crossing properly colored geometric graphs on the point set $R \cup B$. We show that if $||B| - |R|| \leq 1$ and a subset of R forms the vertices of a convex polygon separating the points of B, lying inside the polygon, from the rest of the points of R, lying outside the polygon, then there exists a non-crossing properly colored geometric path on $R \cup B$ covering all points of $R \cup B$.

If $R \cup B$ lies on a circle, the size of the longest non-crossing geometric path is related to the size of the largest separated matching; a separated matching is a non-crossing properly colored geometric matching where all edges can be crossed by a line. A discrepancy of $R \cup B$ is the maximal difference between cardinalities of color classes of intervals on the circle. When the discrepancy of $R \cup B$ is at most 2, we show that there is a separated matching covering asymptotically $\frac{4}{5}$ of points of $R \cup B$. During this proof we use a connection between separated matchings and the longest common subsequences between two binary sequences where the symbols correspond to the colors of the points.

Keywords: geometric graph, bichromatic point set, non-crossing alternating path, longest common subsequence, graph drawing

Contents

Introduction

A graph $G = (V, E)$ is an ordered pair where V is a set, whose elements are called *vertices*, and E is a set of unordered pairs of vertices, whose elements are called *edges*. A graph drawn in the plane is called a *geometric graph* if all its edges are straight-line segments. A *non-crossing geometric graph* is a geometric graph whose no two edges except the ones having a common endpoint intersect.

In this thesis, we consider geometric graphs drawn on a set of points colored red and blue. Throughout this thesis, let B and R always denote the set of blue points and the set of red points, respectively. Moreover, we assume that B and R are always disjoint and that $R \cup B$ is always in general position (that is, no three points are collinear) unless specifically stated otherwise. A point set is in *convex position* if the points of this point set are vertices of a convex polygon. A graph with colored vertices is *properly colored* if no edge of this graph connects two vertices of the same color. We have only two colors, and thus we call properly colored graphs *alternating* graphs.

We study the existence of certain non-crossing geometric graphs with vertices in the given point set $R \cup B$. In particular, we are interested in paths, matchings, and trees. We often ask this question in an extremal way: What is the largest non-crossing geometric path (matching, tree) that exists on $R \cup B$? Furthermore, we sometimes impose further restrictions on the points set. For example, we can require the points to be in convex position. Additionally, we mostly study the cases when $|R| = |B|$. For a comprehensive survey about discrete geometry on colored points, that includes these and many different problems, we refer to Kano and Urrutia [12] and Kaneko and Kano [11].

One of the most studied configurations is the case of alternating geometric paths on point sets in convex position. Let $l(n)$ be the largest number such that for any set $R \cup B$ in convex position with $|R| = |B| = n$ there exists a non-crossing alternating geometric path covering at least $l(n)$ vertices. Without loss of generality, we can assume that the point set lies on a circle. The problem of determining $l(n)$ was first asked by Erdős and is listed as an open problem by Brass, Moser, and Pach [5, p. 409]. Erdős showed that $l(n) \leq \frac{3}{2}n + o(n)$ and conjectured that the bound is asymptotically tight. Several authors [13, 16, 2] disproved this conjecture by providing families of point sets and proving $l(n) \leq \frac{4}{3}n + o(n)$. The best upper bound $l(n) \leq (4 - 2\sqrt{2})n + o(n)$ is by Csóka et al. [8].

On the other hand, the trivial lower bound $l(n) \geq n$ was improved by sub-linear additive terms by Kynčl, Pach, and Tóth [13] and Mészáros and Hajnal [16]. The best currently known lower bound showing that there exists some $\varepsilon > 0$ such that $l(n) \geq (1 + \varepsilon)n$ is by Mulzer and Valtr [18].

Most of the authors have not proved theorems about non-crossing alternating geometric paths directly. Instead, they used a connection with so-called separated matchings. A non-crossing alternating geometric matching on a convex point set is called a *separated non-crossing alternating geometric matching* (or a *separated matching* for short) if there exists a line, which we will call an *axis*, intersecting the interiors of all the edges of the matching. Let $\mu(n)$ be the largest number such that for any set $R \cup B$ in convex position with $|R| = |B| = n$, there always exists a separated matching covering at least $\mu(n)$ vertices. Again, we may assume that the points lie on a circle. It is trivial to see that the edges of a separated matching can be joined by new edges to form a non-crossing alternating geometric path. Thus, $\mu(n) \leq l(n)$. Kynčl, Pach, and Tóth [13] showed that a variation of the opposite inequality also holds; they proved that if k is the number of color changes along the circle (or, equivalently, the number of maximal monochromatic intervals of points along the circle), then $l(n) - 4k - 1 \leq \mu(n)$. Furthermore, they showed that if there are k color changes along the circle, a non-crossing alternating geometric path of size $n + k - 1$ exists. Hence, if the number of color changes along the circle is in $\Omega(n)$, we can improve the lower bound on $l(n)$ by a constant multiplicative factor. Therefore, most authors investigated cases when the number of such color changes is in $o(n)$, and thus the asymptotics of $l(n)$ and $\mu(n)$ differ only by a function in $o(n)$. This means that all of the bounds on $l(n)$ above also hold for $\mu(n)$.

The problem of finding separated matchings is almost equivalent to the problem of finding anti-palindromic sequences in binary circular words. One can alternatively define the parameter $\mu(n)$ as the largest number such that in every circular binary word with n ones and n zeros, there always exists an anti-palindromic subsequence of such length. Müllner and Ryzhikov [17] independently proved that $\mu(n) \leq \frac{4}{3}n + o(n)$ together with other results about anti-palindromes and palindromes.

Our additional focus is on alternating geometric paths on red and blue point sets not necessarily in convex position. Let $l^g(n)$ be the largest number such that for any set $R \cup B$ with $|R| = |B| = n$ there exists a non-crossing alternating geometric path covering $l^g(n)$ vertices. Clearly, $l^g(n) \leq l(n)$ but other than that not much is known about $l^g(n)$. Abellanas et al. [1] showed that in the case when R can be separated by a line from B, and $||R| - |B|| \leq 1$, there exists a non-crossing alternating geometric path covering all points. This fact together with the existence of a line splitting $R \cup B$ in half implies that $l^g(n) \geq n$.

Cibulka et al. [7] look more closely on a configuration when R and B forms a double chain. A *convex* or a *concave chain* is a finite set of points in the plane lying on the graph of a strictly convex or a strictly concave function, respectively. A *double-chain* consists of a convex chain and a concave chain such that each point of the concave chain lies strictly below every line determined by the convex chain and, similarly, each point of the convex chain lies strictly above every line determined by the concave chain. Cibulka et al. [7] showed that if $||R| - |B|| \leq 1$ and each of the chains of the double-chain contains at least one fifth of all points, then there exists a non-crossing alternating geometric path on $R \cup B$ covering all points. Moreover, they showed that such a path does not exist if one chain contains approximately 28 times more points than the other.

Another specific configuration was investigated by Abellanas et al. [1]. They showed that if $||R| - |B|| \leq 1$, the points of R are vertices of a convex polygon, and all points of B are inside this polygon, then there exists a non-crossing alternating geometric path covering all points.

Our work

To better understand the behavior of $\mu(n)$ independently of $l(n)$, Mészáros [15] studied the cases when the number of color changes along the circle is linear in n. Particularly, they studied cases with small discrepancies. The *disbalance* of a set with colored elements is the difference between the cardinalities of its color classes. The *discrepancy* of a set of red and blue points on a circle is the largest integer d for which there exists an interval on the circle with disbalance d. There is also another motivation for studying cases with small discrepancies. Mészáros and Hajnal [16] implicitly showed that if a discrepancy of $R \cup B$ on a circle is d, then $\mu(n) \geq n + \frac{d}{2}$. It follows that in cases with discrepancy of the order $\Omega(n)$ the best lower bound on $\mu(n)$ can be improved by a constant multiplicative factor. Therefore, it seems that cases with discrepancy of the order $o(n)$ are the most interesting.

In the first chapter, we focus on cases with constant discrepancy. Let $\mu_d(n)$ be the largest number such that for any set $R \cup B$ on a circle with $|R| = |B| = n$ with discrepancy at most d there always exists a separated matching covering $\mu_d(n)$ points. Mészáros [15] showed that $\mu_2(n) \geq \frac{4}{3}n$ and that $\mu_3(n) \geq \frac{4}{3}n$.

There is one significant distinction between cases with small discrepancies and cases with unbounded discrepancies. When the discrepancy is unbounded, and an adversary fixes one intersection of the axis of a separated matching with the circle, we cannot beat the trivial lower bound on the size of such separated matching. To see this consider a cycle with n blue points on the left side and n red points on the right side and select the intersection point in the middle of the red points. Then every separated matching cover at most n points. On the other hand, the above-mentioned bounds $\mu_2(n) \geq \frac{4}{3}n$ and $\mu_3(n) \geq \frac{4}{3}n$ hold even if an adversary selects one intersection of the axis with the circle arbitrarily.

We investigate a closely related problem when both points of the intersection of the axis of a separated matching with the circle are given. In that case, the points on the circle are split by the axis into two intervals, and separated matchings match points from the opposite intervals.

These separated matchings can be described combinatorially as follows. Let $A = (a_1, a_2, \dots)$ and $C = (c_1, c_2, \dots)$ be two sequences of red and blue points. A matching M between A and C consists of $|M|$ mutually disjoint edges (pairs of vertices); each edge connecting one point from A with one point from C. Furthermore, we call a matching *homogeneous* if each edge connects equally colored points and *heterogeneous* if each edge connects differently colored points. A matching M is *non-crossing* if no two edges cross, that is, for every pair of distinct edges (a_i, b_j) and (a_k, b_l) it holds that $i < k \Leftrightarrow j < l$.

It is clear that if we put the sequences A and C onto two disjoint arcs of a circle separated by an axis s, one in a clockwise and one in a counter-clockwise direction, then every non-crossing heterogeneous matching between A and C corresponds to a separated matching on the circle with axis s and vice versa.

Furthermore, for every heterogeneous matching between two sequences A, C, there exists a homogeneous matching of the same size between A and the sequence formed from C by flipping the colors of all its points. Similarly, for every homogeneous matching, we can find a corresponding heterogeneous matching. Therefore, we can focus solely on the homogeneous case.

Homogeneous matchings have one clear advantage over heterogeneous ones;

they naturally correspond to common subsequences in the following way: If we replace every blue point with a symbol 1 and every red point with a symbol 0, we obtain a binary sequence. Let A and C be two sequences of red and blue points and let A' and C' be their corresponding binary sequences. Then for every non-crossing homogeneous matching M between A and C, we obtain a corresponding common subsequence of A' and C' by reading all matched points of A (or of C) from left to right. Similarly, for every common subsequence of A' and C', we can find a corresponding non-crossing homogeneous matching between A and C. Therefore, homogeneous matchings are easier to visualize or process programmatically. As far as we are aware, the problem of finding the longest common subsequence has not been studied in a way that would help understand separated matchings. One of the closest problems that received significant attention is determining the length of the longest common subsequence of two random binary strings [6, 9, 14].

The *discrepancy* of a sequence of points is defined in the same way as the discrepancy of points on a cycle. That is, it is the largest integer d for which there exists an interval of points from the sequence with disbalance d.

In chapter 1, our main result is the following.

Theorem 1. *Let A and C be sequences of red and blue points of length n with discrepancy at most 2. Then there exist prefixes A' of A and C' of C of combined length of at least n such that there exists a non-crossing homogeneous matching between A' and C' covering at least $\frac{4}{5}(|A'| + |C'|) - O(1)$ points.*

Furthermore, this bound is asymptotically tight.

Moreover, we show that this theorem implies the following.

Corollary 2. *We have, $\mu_2(n) \geq \frac{8}{5}n - o(n)$.*

Additionally, we investigate non-crossing homogeneous matchings between one sequence with discrepancy 1 and one sequence with bounded discrepancy. We believe that similar methods could be used for bounding $\mu_d(n)$ even for larger d.

In chapter 2 we extend the family of configurations of points for which there exists a non-crossing alternating geometric path covering all points. Specifically, we prove the following theorem.

Theorem 3. *Let R be a set of red points and B be a set of blue points such that $R \cup B$ is in general position. Let P be a polygon whose vertices are formed by a subset of R. Assume that the remaining points of R lie outside of P, points of B lie in the interior of P, and $||R| - |B|| \leq 1$. Then there exists a non-crossing alternating geometric path on $R \cup B$ covering all points of $R \cup B$.*

1. Separated matchings and matchings between sequences

1.1 Bounding cardinalities of separated matchings using matchings between two sequences

In order to find a connection between separated matchings and non-crossing homogeneous matchings, we can not simply look for the largest matchings between two sequences. Instead, we look for matchings large in comparison to the prefixes they are in. Thus, we need to introduce new terminology.

Let $A = (a_1, a_2, \dots, a_n)$ and $C = (c_1, c_2, \dots, c_m)$ be two sequence of red and blue points. The *matching efficiency* of A and C, denoted by $e(A, C)$, is the maximum of $\frac{2|M|}{|A|+|C|}$ over all non-crossing homogeneous matchings M between A and C. A pair of prefixes (A', C') of A and C is *semi-complete* if $|A'| + |C'| \geq \min(|A|, |C|)$. The *matching potential* of A and C, denoted by $p(A, C)$, is the maximum of matching efficiencies over all semi-complete pairs of prefixes of A, C. For the heterogeneous matchings we similarly define $e^{\text{het}}(A, C)$ and $p^{\text{het}}(A, C)$ but we will mostly work with just the homogeneous matchings.

We will also be using the following notation. If there exists a non-crossing homogeneous matching that covers the whole A, we say that A is *matchable* into C, and we denote it by $A \prec C$. Additionally, let $\oplus$ denote the operation of concatenating two sequences of points or the operation of concatenating a sequence of points with a single point.

Additionally, in figures and when presenting concrete examples, we represent every blue point by the symbol 1 and every red point by the symbol 0. For example, 101 represents an alternating sequence of blue, red, and blue points.

By the following trivial observation, we can do proofs on subsequences having nicer structures (usually by discarding the beginnings and endings of sequences).

Observation 4. *Let $A = (a_i)_{i=1}^n, C = (c_i)_{i=1}^m$ be two sequence of red and blue points. Let A^* be a subsequence of A and C^* a subsequence of C of sizes $n - O(1)$ and $m - O(1)$, respectively. Then $e(A, C)$ and $e(A^*, C^*)$ as well as $p(A, C)$ and $p(A^*, C^*)$ can differ only by the order of $O(\frac{1}{\min\{n,m\}})$.*

We are interested in the asymptotic behavior of matching potentials of sequences with bounded discrepancies. We define $p_{(k,l)}(n)$ as the minimum of $p(A, C)$ over all sequences A, C of length n that have discrepancy at most k and at most l, respectively. We define $p^{\text{het}}_{(k,l)}(n)$ for heterogeneous matchings analogously. Since the discrepancy of a sequence with flipped colors stays the same, we see that $p_{(k,l)}(n) = p^{\text{het}}_{(k,l)}(n)$. Thus, it is sufficient to work with only homogeneous matchings.

The next lemma shows that we can prove lower bounds on $\mu_d(n)$ using $p_{(d,d)}(n)$.

Lemma 5. *Let $p_{(d,d)}(n) \geq c - o(1)$ for some constant c. Then $\mu_d(n) \geq 2cn - o(n)$.*

Proof. Since $p_{(d,d)}(n) \geq c - o(1)$, we have $p^{\text{het}}_{(d,d)}(n) \geq c - o(1)$. Thus, in this proof we can use the results about homogeneous matchings for the heterogeneous ones.

Let R be a set of red points on a circle and B be a set of blue points on the circle such that $|R| = |B| = n$, the set $R \cup B$ is in general position, and $R \cup B$ has discrepancy at most d. We want to show that there exists a separated matching covering at least $2cn - o(n)$ points. Let C_1 be a sequence of all points along the circle in clockwise order. We inductively build sequences and matchings $C_i, A_i, A'_i, B_i, B'_i, M_i$ in the following way until $|C_t|$ is small enough (we will define the precise value later). Let A_i be the prefix of C_i containing $\left\lfloor \frac{|C_i|}{2} \right\rfloor$ points of C_i and let B_i be the reverse of the suffix of C_i containing $\left\lfloor \frac{|C_i|}{2} \right\rfloor$ points. By the definition of $p^{\text{het}}_{(d,d)}(n)$, there exists a semi-complete pair of prefixes (A'_i, B'_i) of A_i, B_i and a non-crossing heterogeneous matching M_i between A'_i, B'_i such that $e^{\text{het}}(A'_i, B'_i) = p^{\text{het}}_{(d,d)}\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right)$. Finally let C_{i+1} be the interval of C_i formed by removing A'_i and B'_i from C_i.

It is easy to see that all A'_i, B'_i are pairwise disjoint and that the points of all A'_i's form an interval on the circle, as do the points of all B'_i's. Moreover, these two intervals can be separated by a line. Hence, the union of all the matchings M_i is a non-crossing heterogeneous matching between these two intervals. Therefore, we get a non-crossing alternating geometric separated matching by representing the edges of this non-crossing heterogeneous matching by straight-line segments. It remains to bound its size. For precise computation let $g(n)$ be a function in $o(1)$ such that $p^{\text{het}}_{(d,d)}(n) \geq c - g(n)$.

We have

$$\begin{aligned}
\left|\bigcup_{i=1}^{t-1} M_i\right| &= \sum_{i=1}^{t-1} \left(e^{\text{het}}(A'_i, B'_i) \cdot (|A'_i| + |B'_i|)\right) \\
&\geq \sum_{i=1}^{t-1} \left(\left(c - g\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right)\right) \cdot (|A'_i| + |B'_i|)\right) \\
&= c \cdot \sum_{i=1}^{t-1} (|A'_i| + |B'_i|) - \sum_{i=1}^{t-1} \left(g\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right) \cdot (|A'_i| + |B'_i|)\right).
\end{aligned}$$

From the induction it is easy to see that $2n = |C_1| = \sum_{i=1}^{t-1} (|A'_i| + |B'_i|) + |C_t|$. If we build the sequences until $|C_t| << n$ (for example $|C_t| \leq 2 \log n$), then

$$c \cdot \sum_{i=1}^{t-1} (|A'_i| + |B'_i|) \geq 2cn - o(n).$$

Note that since (A'_i, B'_i) is a semi-complete pair of A_i, B_i, we see that that $|A'_i| + |B'_i| \geq \frac{|C_i|}{2} - 1$. Thus, it is possible to build the sequences until $|C_t| \leq 2 \log n$.

Let

$$f(n) = \sum_{i=1}^{t-1} \left(g\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right) \cdot (|A'_i| + |B'_i|)\right).$$

It remains to show that $f(n)$ is in $o(n)$.

Thus, we need to show that for every $\varepsilon > 0$ there exists n_0 such that for every $n \geq n_0$, we have $f(n) \leq \varepsilon n$. We know that $g(x)$ is in $o(1)$. Hence, there exists x_0 such that for every $x \geq x_0$, we have $g(x) \leq \frac{\varepsilon}{2}$. Let us build the sequences C_i so that C_t is the first for which $|C_t| \leq 2 \log n$. Hence, for every i smaller than t, we have $|C_i| \geq 2 \log n$. Let us select $n_0 = e^{x_0+1}$. Then for every i from $\{1, \ldots t-1\}$, we have

$$\left\lfloor \frac{|C_i|}{2} \right\rfloor \geq \lfloor \log n \rfloor \geq \lfloor \log n_0 \rfloor = \lfloor x_0 + 1 \rfloor \geq x_0.$$

Hence, $g\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right) \leq \frac{\varepsilon}{2}$ for every i from $\{1, \dots t-1\}$. Therefore,

$$f(n) = \sum_{i=1}^{t-1} \left(g\left(\left\lfloor \frac{|C_i|}{2} \right\rfloor\right) \cdot (|A'_i| + |B'_i|) \right) \leq \frac{\varepsilon}{2} \sum_{i=1}^{t-1} (|A'_i| + |B'_i|) \leq \varepsilon n$$

and the proof is finished.

□

Theorem 1 together with this lemma immediately implies Corollary 2.

1.2 Matching potential of sequences with discrepancy 1

We now look at cases when one of the sequences has discrepancy 1. Note that colors in sequences with discrepancy 1 alternate. We prove the following theorem.

Theorem 6. *Let k be a positive integer constant. Then $p_{(1,k)}(n) = \frac{2k}{3k-1} - O(\frac{1}{n})$.*

We show a more technical lemma that better represents the error term. The theorem then immediately follows. Alternatively, we could use Observation 4 and prove the Theorem without computing specific constants.

Lemma 7.

a) Let $n, m \geq 2$ be integers and let $k \geq 1$ be an integer. Assume that $A = (a_i)_{i=1}^n$ is a sequence of red and blue points with discrepancy at most k and $C = (c_i)_{i=1}^m$ is a sequence of red and blue points with discrepancy 1. Then there exists a semi-complete pair of prefixes (A', C') of A, C and a non-crossing homogeneous matching M' between A' and C' such that

$$2|M'| \geq \frac{2k}{3k-1} \cdot (|A'| + |C'| - 1).$$

b) Let $n, m \geq 2$ be integers and let $k \geq 1$ be an integer. Then there exist sequences of red and blue points $A = (a_i)_{i=1}^n$ and $C = (c_i)_{i=1}^m$ with discrepancies k and 1, respectively, such that for every semi-complete pair of prefixes (A', C') of A, C with a non-crossing homogeneous matching M' between A' and C', we have

$$2|M'| \leq \frac{2k}{3k-1} \cdot (|A'| + |C'|).$$

Proof. We start with the first part. We select A' as the longest prefix of A such that $A' \prec C$. Next, we select C' as the shortest prefix of C such that $A' \prec C'$. Finally, we select a maximal non-crossing homogeneous matching M' between A' and C'. Since $A' \prec C'$, then M' covers the whole A'. Thus, $|M'| = |A'|$.

Suppose that $A' \neq A$ and $|C'| \leq |C| - 2$. Then $A' \oplus a_{|A'|+1} \not\prec C$ (because A' is the longest prefix of A that is matchable into C). But at the same time we know that $A' \prec C'$. This implies that $A' \oplus a_{|A'|+1} \prec C' \oplus c_{|C'|+1} \oplus c_{|C'|+2}$ (because C contains points with alternating colors). That is a contradiction with $A' \oplus a_{|A'|+1} \not\prec C$. Therefore, either $|A'| = |A|$ or $|C'| = |C|$ or $|C'| = |C| - 1$. Since $|A'| \geq 1$, in all cases (A', C') is a semi-complete pair of prefixes of A and C. Thus, it is sufficient to show that $2|A'| \geq \frac{2k}{3k-1} \cdot (|A'| + |C'| - 1)$ or, equivalently, that $|C'| \leq \frac{2k-1}{k} \cdot |A'| + 1$.

For every a_i, let $M'(a_i)$ denote the point from C' matched to a_i. We count the number of unmatched points in C'. Suppose that there are two unmatched consecutive points c_x, c_{x+1} in C'. Then all the points a_y matched to some points c_z to the right side of c_{x+1} (i.e., $z > x + 1$) could be matched to the points c_{z-2} instead, and the matching would remain non-crossing and homogeneous. This is a contradiction because C' is the shortest prefix of C for which $A' \prec C'$. Thus, there is at most one unmatched point between every pair of matched points $(M'(a_i), M'(a_{i+1}))$ and at most one unmatched point before $M'(a_1)$. By the same argument, we see that there is no unmatched point of C' after the last matched point of C'.

Let a_i, a_{i+1} be two consecutive points of A' with distinct colors. Since M' is non-crossing, there are no matched points between $M'(a_i)$ and $M'(a_{i+1})$. Moreover, there is at most one unmatched point between $M'(a_i)$ and $M'(a_{i+1})$, but the colors of C' alternates, thus there can be none and $M'(a_i), M'(a_{i+1})$ are consecutive. Note that there are at least $\frac{1}{k} \cdot |A'| - 1$ such pairs (a_i, a_{i+1}) because A' has discrepancy at most k.

Overall there is at most one unmatched point of C' before $M'(a_1)$, and at most one additional one unmatched point of C' for every pair of consecutive points of A' with the same color. Thus, there are at most $1 + \frac{k-1}{k} \cdot |A'|$ unmatched points in C'.

Hence,

$$|C'| \leq |A'| + 1 + \frac{k-1}{k} \cdot |A'| = \frac{2k-1}{k} \cdot |A'| + 1$$

and the first part of the proof is complete.

It remains to prove part b). We will construct the desired sequences. Let A be formed by monochromatic intervals such that colors of consecutive intervals alternate and every interval, except possibly the last one, has size k. Furthermore, let C be a sequence of points of alternating color such that the first points of A and C have different colors. Clearly, A has discrepancy at most k and C has discrepancy 1.

Let (A', C') be a semi-complete pair of prefixes of A, C and let M' be a maximal matching between A' and C'.

Let A'_i denote the i-th monochromatic interval of A' and let t be the number of such intervals A'_is. Let m_i denote the number of matched points in A'_i. For every i except $i = 1$, let C'_i denote the minimal interval in C' that spans the points of C' that are matched with points of A'_i. Additionally, let C'_1 denote the minimal prefix of C' that spans the points of C' that are matched with points of A'_1. Note that no two intervals A'_i intersect, so no two C'_i intersect either. Thus, $\sum_i^t |A'_i| = |A'|$ and $\sum_i^t |C'_i| \leq |C'|$. Since every A'_i contains points of only one color and $C'i$ is alternating, then C'_i is of length at least $2m_i - 1$. Furthermore, every interval A'_i has length k except the last one that has length at least m_t.

For every i, $2 \le i \le t-1$ we have

$$\begin{aligned} 2m_i &= \frac{2m_i}{k+2m_i-1} \cdot (k+2m_i-1) \le \\ &\le \frac{2k}{3k-1} \cdot (k+2m_i-1) \le \frac{2k}{3k-1} \cdot (|A'_i| + |C'_i|). \end{aligned} \tag{1.1}$$

We used the inequality

$$\frac{2m_i}{k+2m_i-1} \le \frac{2k}{3k-1}.$$

that follows from the fact that $m_i \le k$.

Additionally, for m_t we have

$$\begin{aligned} 2m_t &= \frac{2m_t}{m_t+2m_t} \cdot (m_t+2m_t) \le \\ &\le \frac{2k}{3k-1} \cdot (m_t+2m_t) \le \frac{2k}{3k-1} \cdot (|A'_t| + |C'_t| + 1). \end{aligned}$$

And finally, for m_1 we known that either a_1 is not matched or c_1 is not matched (because they have different colors). If c_1 is matched than $m_1 = 0$ because the matching is non-crossing and c_1 have different color than all points of A'_1. If c_1 is not matched than $|C'_1| \ge 2m_1 - 1 + 1$. Thus, similarly as in (1.1), we have

$$2m_1 \le \frac{2k}{3k-1} \cdot (|A'_1| + |C'_1| - 1).$$

We finish the proof by putting these inequalities together .

$$\begin{aligned} 2|M'| = 2\sum_{i=1}^{t} m_i \le \frac{2k}{3k-1} \cdot (|A'_1| + |C'_1| - 1) + \sum_{i=2}^{t-1} \frac{2k}{3k-1} \cdot (|A'_i| + |C'_i|) + \\ + \frac{2k}{3k-1} \cdot (|A'_t| + |C'_t| + 1) \le \\ \le \frac{2k}{3k-1} \cdot \left(\sum_{i=1}^{t} |A'_i| + \sum_{i=1}^{t} |C'_i| \right) \le \frac{2k}{3k-1} \cdot (|A'| + |C'|)\,. \end{aligned}$$

□

Proof of Theorem 6. Let n be a natural number. Let $A = (a_i)_{i=1}^n$ be a sequence of red and blue points with discrepancy at most k and $C = (c_i)_{i=1}^n$ be a sequence of red and blue points with discrepancy 1. By Lemma 7 a) there exists a semi-complete pair of prefixes (A', C') of A, C and a non-crossing homogeneous matching M' between A' and C' such that $2|M'| \ge \frac{2k}{3k-1} \cdot (|A'| + |C'| - 1)$.

Thus,

$$p(A,C) \ge e(A',C') = \frac{|M'|}{|A'| + |C'|} \ge \frac{\frac{2k}{3k-1} \cdot (|A'| + |C'| - 1)}{|A'| + |C'|} \ge \frac{2k}{3k-1} - \frac{1}{n}.$$

Hence,

$$p_{(1,k)}(n) \ge \frac{2k}{3k-1} - O\left(\frac{1}{n}\right).$$

The upper bound is implied by Lemma 7 b) in a similar way. □

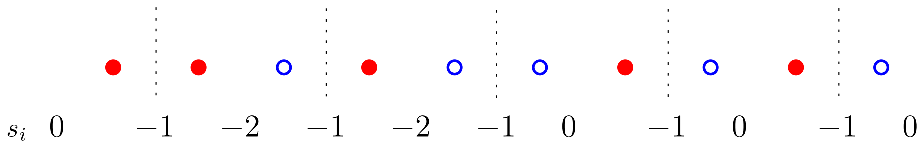


Figure 1.1: The structure of a sequence of points with discrepancy at most 2.

1.3 Matching potential of sequences with discrepancy 2

Our goal in this section is to prove Theorem 1. This theorem can be rephrased in our new notation as $p_{(2,2)}(n) = \frac{4}{5} - O\left(\frac{1}{n}\right)$.

Since every sequence with discrepancy 1 is also a sequence with discrepancy at most 2, Theorem 6 implies that $p_{(2.2)}(n) \leq \frac{4}{5} - O(\frac{1}{n})$. Thus, it remains to prove the lower bound.

First of all, we show that sequences with discrepancy at most 2 have nice structures.

Lemma 8. *Every sequence A of red and blue points with discrepancy at most 2 can be partitioned, possibly without its first and last element, into intervals of length 2 so that each interval contains one red and one blue point.*

Proof. For all i, $0 \leq i \leq n$, let s_i be the number of blue points minus the number of red points in the first i points of A (it is a "signed disbalance"). For every $i < j$, the values s_i and s_j differ by at most 2, otherwise the interval from the $(i+1)$-th to the j-th point would have disbalance greater than 2. Thus, there exists some integer x such that every s_i is either x, $x+1$, or $x+2$. Together with the fact that for every i, we have $|s_i - s_{i+1}| = 1$, it means that either exactly all the s_i with even index have value $x+1$, or exactly all the s_i with odd index have value $x+1$. If we split A in every position where "s_i has the value $x+1$", we obtain the desired partition. See Figure 1.1 for an illustration. □

We proceed with proving the main theorem.

Proof of Theorem 1. As we stated earlier it remains to prove the lower bound. Let $A = (a_i)_{i=1}^{n}$, and $C = (c_i)_{i=1}^{m}$ be sequences of red and blue points with discrepancy at most 2.

By Lemma 8 both A and C can be partitioned, possibly without its first and last element, into intervals of length 2; each containing one red and one blue point. Moreover, by Observation 4, we can ignore these possibly problematic first and last elements and assume that both A and C can be partitioned entirely.

We will show by double induction that there always exists a semi-complete pair of prefixes (A', C') of A, C and a non-crossing homogeneous matching M' between A' and C' such that $2|M'| \geq \frac{4}{5} \cdot (|A'| + |C'| - 22)$. Hence, $e(A', C') \geq \frac{4}{5} - \Theta(\frac{1}{\min{(n,m)}})$ and the theorem follows. Our approach actually shows that there always exists a semi-complete pair of prefixes (A', C') of A, C and a non-crossing homogeneous matching M' between A' and C' such that $2|M'| \geq \frac{4}{5} \cdot (|A'| + |C'| - 6)$, but the argumentation would be even more technical.

The first induction is on n and m (the lengths of A and C), ordered by product order; that is, $(n_1, m_1) \leq (n_2, m_2)$ if $n_1 \leq n_2$ and $m_1 \leq m_2$. If $n \leq 22$ or $m \leq 22$

then the statement clearly holds because we can select a semi-complete pair of prefixes of combined length 22 and an empty matching.

Next, assume there exist prefixes A', C' of even length of A and C, respectively, with efficiency at least $\frac{4}{5}$. Then there exists a non-crossing homogeneous matching M' between A' and C' such that $2|M'| \geq \frac{4}{5} \cdot (|A'| + |C'|)$. Furthermore, we can use the induction hypothesis on the remaining parts of A and C because they can be split into intervals of length 2 the same way as A and C can be split. Thus, there exists a semi-complete pair of prefixes (A^*, C^*) of the remaining parts of A, C and a non-crossing homogeneous matching M^* between A^* and C^* such that $2|M^*| \geq \frac{4}{5} \cdot (|A^*| + |C^*| - 22)$. Hence, $(A' \oplus A^*, C' \oplus C^*)$ is a semi-complete pair of prefixes of A, C. Moreover, for the matching $M' \cup M^*$ between $A' \oplus A^*, C' \oplus C^*$ we have $2|M' \cup M^*| \geq \frac{4}{5} \cdot (|A' \oplus A^*| + |C' \oplus C^*| - 22)$. Therefore, it would be sufficient to show that there always exists a pair of prefixes of A and C with efficiency at least $\frac{4}{5}$. Unfortunately, that is not always true.

Instead, we use a second induction. Let s be an even integer greater than or equal to 22, let P be a prefix of A of length s, and let O be a prefix of C of length s. We show by induction on s that there either exist prefixes of P and O of even length with efficiency $\frac{4}{5}$ or P and O contain one of the following semi-complete pairs of prefixes (depending on their residue modulo 6), possibly with all colors switched or with switched order of sequences inside the pairs, but these cases are symmetric:

1. if $s = 6t$:
   (a) $(011010(010101)^{t-1}, 100101(1001)^{t-1}1001)$
   (b) $(011010(010101)^{t-1}, 100101(1001)^{t-1}1010)$
2. if $s = 6t + 2$:
   (a) $(011010(010101)^{t-1}01, 100101(1001)^{t-1}1001)$
   (b) $(011010(010101)^{t-1}01, 100101(1001)^{t-1}1010)$
   (c) $(011010(010101)^{t-1}10, 100101(1001)^{t-1}1001)$
   (d) $(011010(010101)^{t-1}10, 100101(1001)^{t-1}1010)$
3. if $s = 6t + 4$:
   (a) $(011010(010101)^{t-1}1010, 100101(1001)^{t-1}100101)$
   (b) $(011010(010101)^{t-1}0101, 100101(1001)^{t-1}100110)$

For every pair of prefixes (A', C') of one of the form above, we can find a non-crossing homogeneous matching M' such that $2|M'| \geq \frac{4}{5} \cdot (|A'| + |C'| - 22)$ since the periodic parts can be matched together while having $\frac{4}{5}$ of all its points covered and it is easy to check that the best possible matching of the remaining prefixes and suffixes is big enough. Furthermore, when $s = \min(|A|, |C|)$ these pairs are semi-complete pairs of prefixes of A and C and the theorem follows. Thus, it remains to finish the induction on s.

The base case when $s = 22$ follows from the case analysis in Attachment A.1. For other s we know that O, P contains one pair from the group of pairs of prefixes we get from the induction hypothesis applied on the prefixes of O, P of

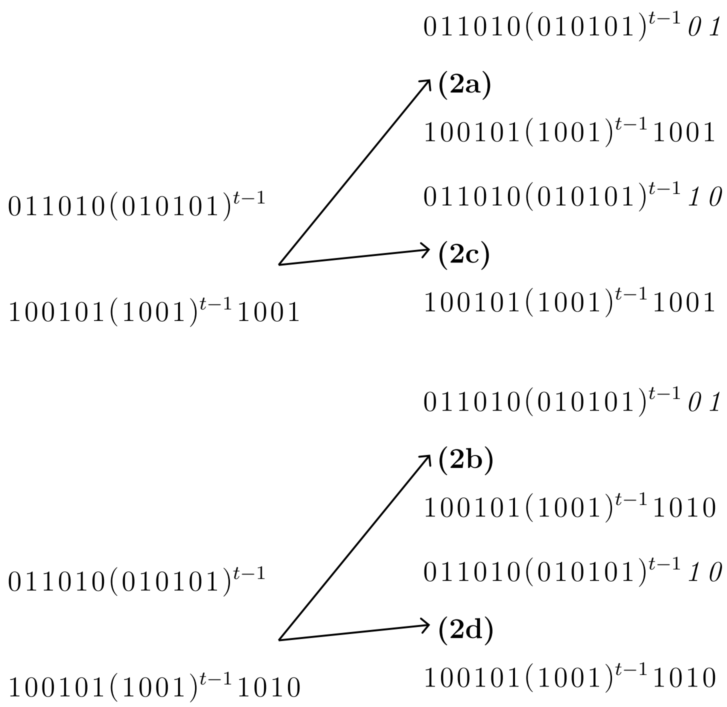


Figure 1.2: Possible extensions of pairs of prefixes in the case when $s = 6t + 2$.

size $s - 2$. We try to extend all of these possible pairs of prefixes. Moreover, we know that A, C can be split into intervals of size 2 containing one point of each color. Therefore, we can do the extension only by intervals represented by 01 or by 10. The analysis is split into cases according to the residue modulo 6. The points we used for extending the prefixes from the induction hypothesis are written in italics. The matchings in figures signify the existence of prefixes of even length with efficiencies of at least $\frac{4}{5}$.

First, we consider the case when $s \mod 6 = 2$. Let t be an integer such that $s = 6t+2$. We are extending pairs $(011010(010101)^{t-1}, 100101(1001)^{t-1}1001)$ and $(011010(010101)^{t-1}, 100101(1001)^{t-1}1010)$. By extending the first prefixes by 01 or 10, we get one of the semi-complete pairs (2a) – (2d) in all cases. See Figure 1.2 for graphical explanation.

Next, we consider the case when $s \mod 6 = 4$. Let t be an integer such that $s = 6t + 4$. We have four possible pairs of prefixes from the induction hypothesis. We try to extend all of them. We always either get the semi-complete pair of prefixes (3a) or (3b) or we find prefixes with efficiency at least $\frac{4}{5}$. See Figure 1.3 and Figure 1.4.

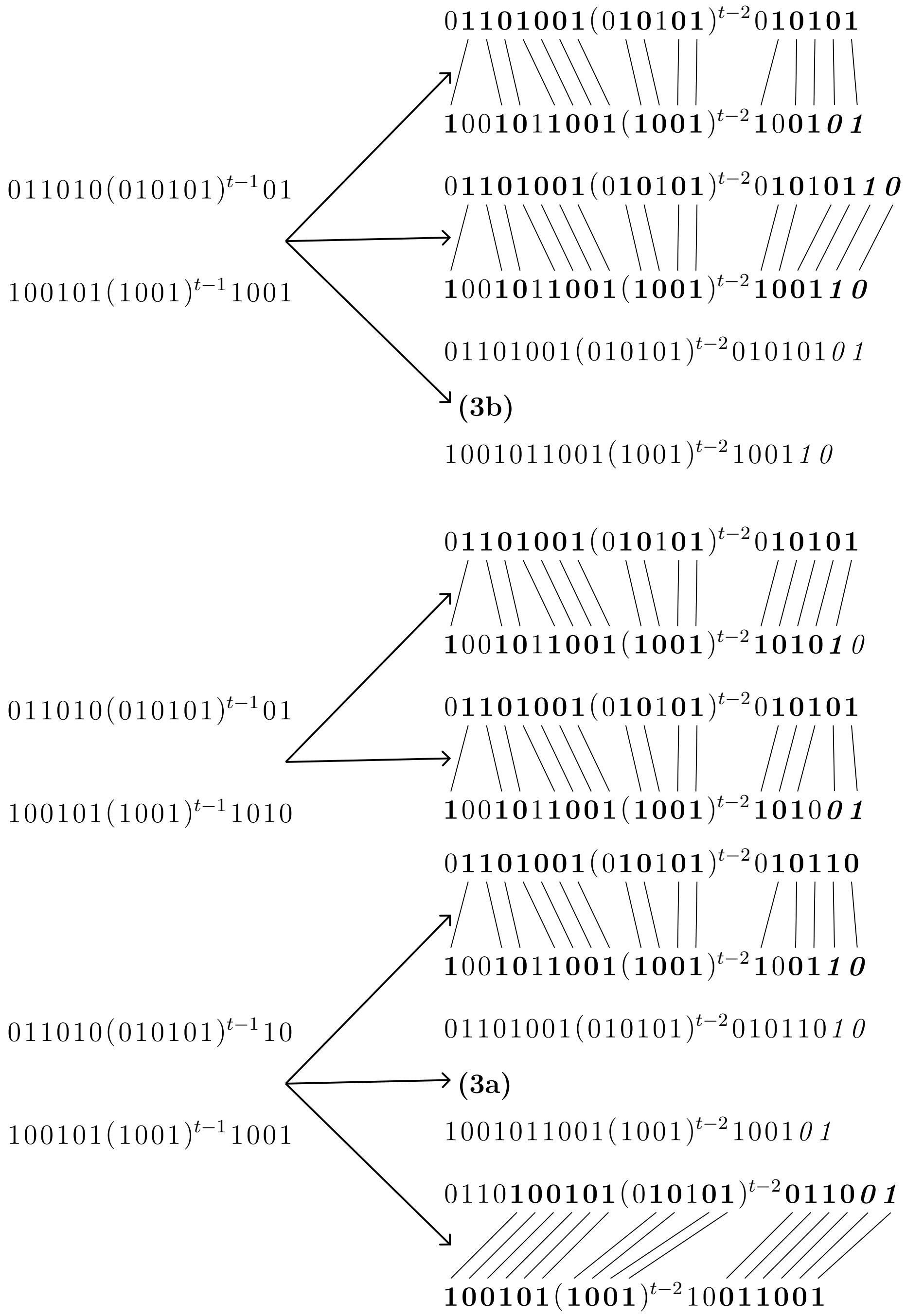


Figure 1.3: Possible extensions of pairs of prefixes in the case when $s = 6t + 4$.

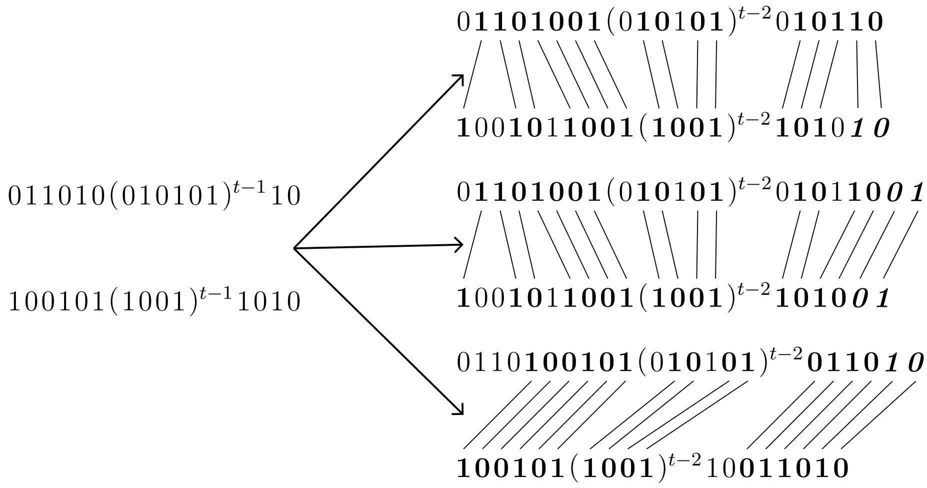


Figure 1.4: Possible extensions of pairs of prefixes in the case when $s = 6t + 4$.(cont.)

Finally, we consider the case when $l \mod 6 = 0$. Let t be an integer such that $s = 6t$. We proceed analogously to previous cases. See Figure 1.5.

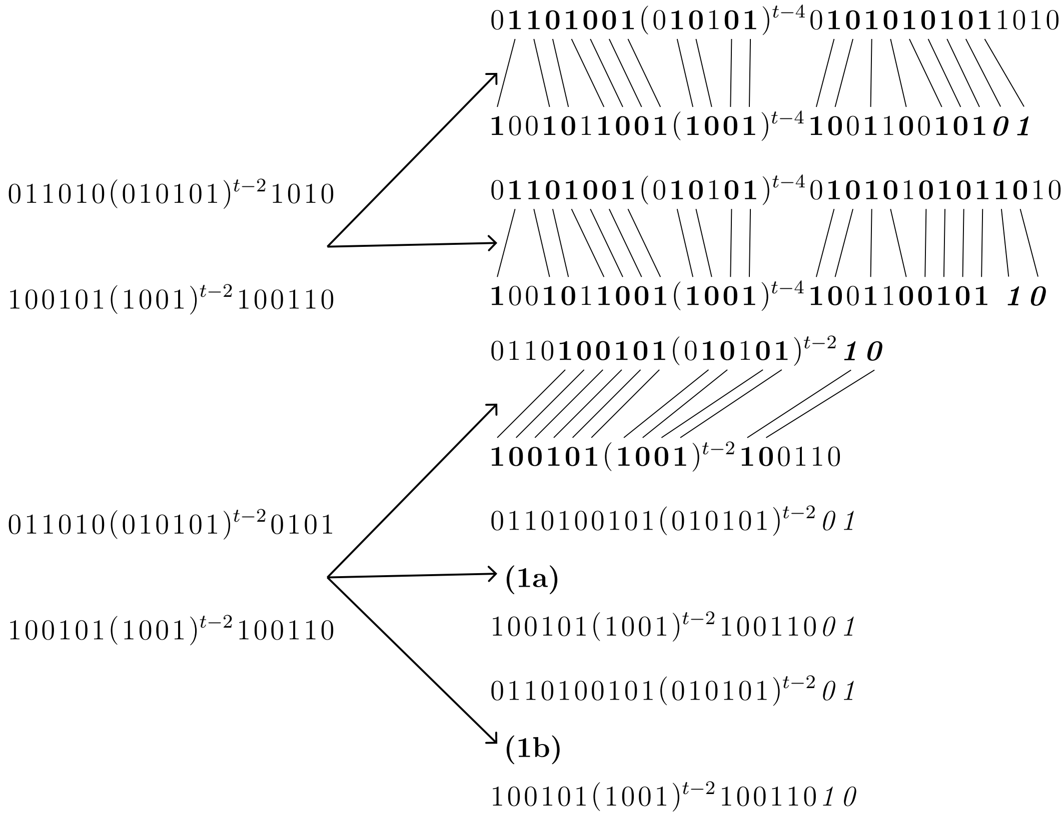


Figure 1.5: Possible extensions of pairs of prefixes in the case when $s = 6t$.

□

2. Alternating paths on points in non-convex configurations

A *polygon* in this chapter is a closed, possibly unbounded, region in the plane whose boundary consists of non-crossing strait-line segments or half-lines. A bounded polygon can also be defined by an ordered set of its vertices; in that case, we assume that the vertices lie on the boundary of the polygon in the clockwise direction. The *convex hull* of a set of points X, denoted by $\operatorname{conv}(X)$, is the smallest convex set that contains X. Recall that B, R always denote the set of blue points and the set of red points, respectively. Moreover, B and R are always disjoint, and $R \cup B$ is always in general position.

Our primary goal in this chapter is to prove the Theorem 3.

This theorem is a generalization of the following theorem proved by Abellanas et al. [1].

Theorem 9 ([1]). *Let R form the vertices of the polygon $\operatorname{conv}(R \cup B)$, the points of B lie in the interior of $\operatorname{conv}(R \cup B)$, and $||R| - |B|| \leq 1$. Then there exists a non-crossing alternating path on $R \cup B$ covering all points of $R \cup B$.*

Our improvement is that the polygon P can be formed by only a subset of R, whereas the remaining points of R remain outside of P. The approach in the proof of Theorem 9 in a case when $|R| = |B|$ is to partition the polygon into convex parts, each containing exactly one edge of the polygon and one blue point from inside the polygon, and then connect by straight-line segments each of the blue points to the end vertices of the edge that is inside the same part. In this way, alternating geometric paths of length two are formed inside each part of the partition. Moreover, they share their end vertices, and so they form an alternating cycle together. This cycle is non-crossing since each path lies in its own part of the partition.

We proceed similarly with only two significant distinctions. Firstly, we need to partition the whole plane into convex parts such that every edge of the polygon is a diagonal of one part of the partition, and each part of the partition contains a certain amount of red and blue points. Secondly, we need to apply a slightly stronger theorem to find a non-crossing alternating geometric path inside every part. These paths together will form an alternating cycle the same way as before.

Before we begin, we introduce some geometric notation needed in our arguments. For a directed line l, the close half-plane to the left of l is denoted by $\operatorname{left}(l)$, and the close half-plane to the right of l is denoted by $\operatorname{right}(l)$. For an edge e of a convex polygon, the closed half-plane to the side of e that is disjoint with the polygon's interior is denoted by $\operatorname{out}(e)$.

For a region T of the plane, $\|T\|_R$ and $\|T\|_B$ denotes the number of red points inside T and the number of blue points inside T, respectively. Generally, we count even the points on boundaries of regions, but sometimes, when specifically noted, we do not count them, usually when a point on boundaries of more regions is assigned to some other region.

2.1 Partitioning of the plane into convex polygons

The partitioning theorem we need to prove is the following.

Theorem 10. *Let $P = (p_1, \dots, p_s)$ be a convex polygon, B be a set of blue points in the interior of P, and R be a set of red points outside of P such that $s = |B| - |R|$ and $R \cup B \cup \{p_1, \dots p_s\}$ is in general position. Then there exists a partition of the plane into convex polygons $Q_1, \dots, Q_s$ such that $p_i p_{i+1}$ is a diagonal of Q_i and for every i, we have $\|Q_i\|_B - \|Q_i\|_R = 1$ (index arithmetic is modulo s). Moreover, every point of $R \cup B$ is counted in exactly one Q_i. That is, if a point of $R \cup B$ lies on the common boundary of more Q_i's it is assigned to only one of them.*

We believe that the following stronger version of this theorem holds.

Conjecture 1. *Let Q be a convex polygon, $P = (p_1, \dots, p_s)$ be a convex polygon inside Q, B be a set of blue points in the interior of P, and R be a set of red points outside P but inside Q such that $R \cup B \cup \{p_1, \dots p_s\}$ is in general position. Assume that $|B| - |R| = n(p_1p_2) + \cdots + n(p_sp_{s+1})$ where the index arithmetic is modulo s and all $n(p_ip_{i+1})$'s are integers such that for every nonempty cyclic interval of indices I, we have*

$$\sum_{i \in I} n(p_ip_{i+1}) \geq - \left\| Q \cap \bigcup_{i \in I} \mathrm{out}(p_ip_{i+1}) \right\|_R . \tag{2.1}$$

Then there exists a partition of Q into convex polygons $Q_1, \dots, Q_s$ such that for every i, the segment p_ip_{i+1} is a diagonal or an edge of Q_i and $\|Q_i\|_B - \|Q_i\|_R = n(p_ip_{i+1})$. Moreover, every point of $R \cup B$ is counted in exactly one Q_i.

For an example partition, see Figure 2.1.

Note that in a case when Q is the plane and all n_i's are equal to 1, Conditions (2.1) always hold, and the conjecture is equivalent with Theorem 10. Moreover, Conditions (2.1) are necessary. Otherwise, the parts of the partition corresponding to the index set for which a condition does not hold would contain too many blue points compared to red ones.

The case when there are no red points outside of P and all $n(p_ip_{i+1})$ are positive integers was already proved by García and Tejel [10] and later by Aurenhammer [3]. The case with points outside of P seems to be more difficult (for example, Conditions 2.1 hold implicitly if all $n(p_ip_{i+1})$ are positive integers).

We managed to prove Conjecture 1 only in the case $s = 3$, but that proved crucial in proving Theorem 10.

Lemma 11. *For $s = 3$, Conjecture 1 holds.*

In the proof, we need to use continuity arguments. For that purpose we will use Knaster–Kuratowski–Mazurkiewicz lemma. It is one of the well-known fixed-point theorems.

Lemma 12 (Knaster–Kuratowski–Mazurkiewicz lemma).
Let $S = \mathrm{conv}(e_1, e_2, e_3) \subset \mathbb{R}^2$ and $\{F_1, F_2, F_3\}$ be a family of closed subsets of S such that for $A \subseteq \{1, 2, 3\}$, we have

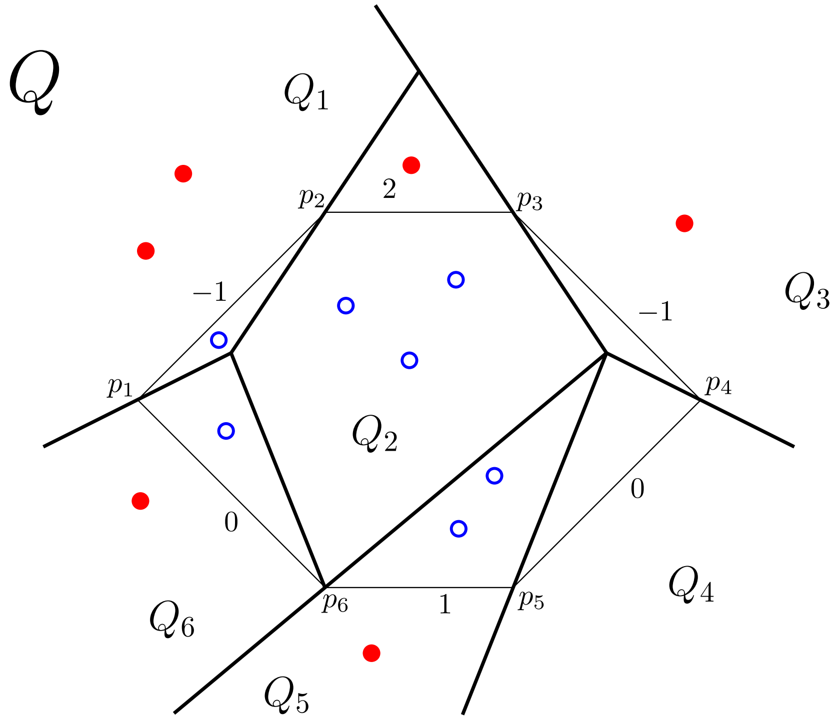


Figure 2.1: A partition of Q into convex polygons $Q_1, \ldots Q_6$ as in Conjecture 1. Blue point are drawn as circles and red points as discs. Numbers $n(p_i p_{i+1})$ are written next to their corresponding edges of the polygon.

$$\operatorname{conv}(e_i : i \in A) \subseteq \bigcup_{i \in A} F_i.$$

Then $\bigcap_{i=1}^{3} F_i$ *is compact and non-empty.*

For a simple proof of this lemma, see for example [4, Theorem 5.1]. This lemma also holds in an analogous form in higher dimensions, but we need just the planar version.

Proof of Lemma 11. For a point $x \in P \setminus \{p_1, p_2\}$, define Q_1^x as the convex polygon enclosed by the boundary of Q and by the lines xp_1, xp_2 so that Q_1^x contains edge p_1p_2 (or, in a degenerate case when x lies on p_1, p_2, define it as $\operatorname{out}(p_1p_2) \cap Q$). Define Q_2^x and Q_3^x analogously.

For every i, let $n_i = n(p_i p_{i+1})$. Let $\{b_1, \ldots, b_{|B|}\} = B$ and $\{r_1, \ldots, r_{|R|}\} = R$. In order to properly resolve points on the boundaries of polygons we need to substitute points with discs. Substitute every b_i by a disk b_i' with b_i in its center and substitute every r_i by a disk r_i' with r_i in its center. Furthermore, since the set $R \cup B \cup \{p_1, p_2, p_3\}$ is in general position we can do the substitution so that every disk has a same positive diameter ε and no line intersects more than two discs and vertices of P simultaneously. Let $\mu(T)$ denote the multiple of the area (standard Lebesgue measure) of a region T in the plane such that for each of our disks d, we have $\mu(d) = 1$.

For all i, $1 \leq i \leq 3$, let

$$F_i = \left\{ x \in P \setminus \{p_i, p_{i+1}\} : \sum_{j=1}^{|B|} \mu(Q_i^x \cap b_j') - \sum_{j=1}^{|R|} \mu(Q_i^x \cap r_j') \geq n_i \right\}.$$

Take x from $P \setminus \{p_1, p_2, p_3\}$. Since $|B| - |R| = n_1 + n_2 + n_3$ and for every such x, (Q_1^x, Q_2^x, Q_3^x) is a partition of Q, we have

$$\sum_{i=1}^{3} \sum_{j=1}^{|B|} \mu(Q_i^x \cap b'_j) - \sum_{i=1}^{3} \sum_{j=1}^{|R|} \mu(Q_i^x \cap r'_j) = |B| - |R| = n_1 + n_2 + n_3. \tag{2.2}$$

Therefore, $F_1 \cup F_2 \cup F_3$ covers the interior of P.

Take $x \in p_1p_2 \setminus \{p_1, p_2\}$. Then $\mathrm{out}(p_1p_2) \cap Q = Q_1^x$ and $\sum_{j=1}^{|B|} \mu(Q_1^x \cap b'_j) = 0$ since all the blue discs are in the interior of P. Moreover, by Inequalities (2.1) and the fact that the line p_1p_2 does not cross any discs, we have $\sum_{j=1}^{|R|} \mu(Q_1^x \cap r'_j) \geq -n_1$. Summing this with Equation (2.2) we have

$$\sum_{j=1}^{|B|} \mu(Q_2^x \cap b'_j) + \sum_{j=1}^{|B|} \mu(Q_3^x \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_2^x \cap r'_j) - \sum_{j=1}^{|R|} \mu(Q_3^x \cap r'_j) \geq n_2 + n_3.$$

Hence, $x \in F_2$ or $x \in F_3$. Therefore, $F_2 \cup F_3$ covers p_1p_2. Analogously, we see that $F_1 \cup F_2$ covers p_3p_1 and $F_3 \cup F_1$ covers p_2p_3.

Take $x = p_1$. Then Q can be partition into two polygons, Q_2^x and $Q \cap (\mathrm{out}(p_1, p_2) \cup \mathrm{out}(p_3, p_1))$. Moreover, Q_2^x contains all blue discs. Thus,

$$\sum_{j=1}^{|B|} \mu(Q_2^x \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_2^x \cap r'_j) - \|Q \cap (\mathrm{out}(p_1, p_2) \cup \mathrm{out}(p_3, p_1))\|_R = |B| - |R| =$$
$$= n_1 + n_2 + n_3.$$

By Conditions (2.1) we have

$$\|Q \cap (\mathrm{out}(p_1, p_2) \cup \mathrm{out}(p_3, p_1))\|_R \geq -n_1 - n_3.$$

Hence, we see that

$$\sum_{j=1}^{|B|} \mu(Q_2^x \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_2^x \cap r'_j) \geq n_2.$$

Therefore, F_2 covers p_1. Analogously, F_3 covers p_2 and F_1 covers p_3.

Furthermore, all sets F_i are closed, except possibly F_i in points p_i, p_{i+1} (index arithmetic is modulo 3). But since F_i covers p_{i-1} it has to cover even a small disc around p_{i-1} because points are in general position. Thus, we can remove this open disc from F_{i-1} and F_{i+1}. If we do this for all F_i, then all of them will be closed and we can apply Lemma 12.

Therefore, there exists a point $y \in P$ such that $y \in F_1 \cap F_2 \cap F_3$. Clearly, it is not any vertex of P. We claim that the polygons Q_1^y, Q_2^y, and Q_3^y form the desired partition of Q. Clearly, every p_ip_{i+1} is a diagonal of Q_i^y. Since the area of the intersection of any two Q_i^y, Q_j^y is zero, then by Equality (2.2) and the definitions of F_i, for every i, we have,

$$\sum_{j=1}^{|B|} \mu(Q_i^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_i^y \cap r'_j) = n_i. \tag{2.3}$$

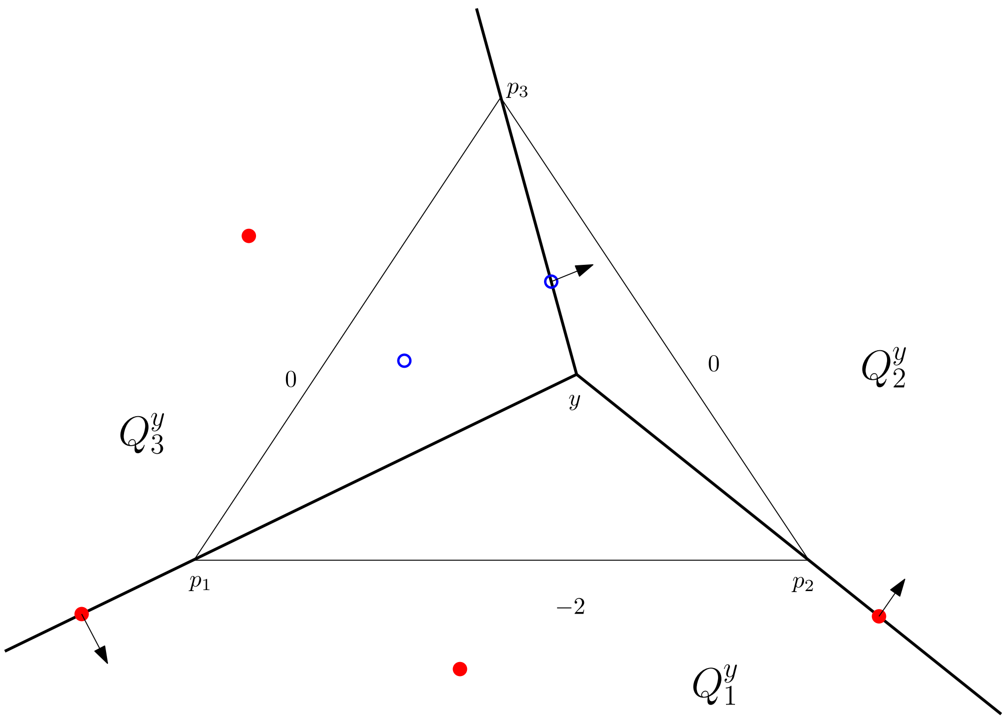


Figure 2.2: Assignment of points when discs are intersected by boundaries of polygons.

It remains to show that corresponding equalities holds also for points and not only for discs. That is, we need to show that for every i, we have

$$\|Q_i^y\|_B - \|Q_i^y\|_R = n_i. \tag{2.4}$$

Since the points are in general position, every line yp_i crosses at most one disc. Therefore, there are only several possibilities how the intersection of these lines with discs can look like.

- No line yp_i crosses any discs:

  In this case every disc is entirely inside some Q_i^y. Therefore,

  $$\sum_{j=1}^{|B|} \mu(Q_i^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_i^y \cap r'_j) = \|Q_i^y\|_B - \|Q_i^y\|_R$$

  and the theorem follows.

- Exactly one line, without loss of generality yp_1, crosses some disc:

  In this case $\sum_{j=1}^{|B|} \mu(Q_1^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_1^y \cap r'_j)$ is not an integer. A contradiction with (2.3).

- Two lines, without loss of generality yp_1, yp_2, cross some discs:

  In this case $\sum_{j=1}^{|B|} \mu(Q_2^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_2^y \cap r'_j)$ is not an integer. A contradiction with (2.3).

- All three lines yp_1, yp_2, yp_3, cross some discs:

If two of them, say yp_1, yp_2, cross the same disc, then $\sum_{j=1}^{|B|} \mu(Q_1^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_1^y \cap r'_j)$ is not an integer. A contradiction with (2.3).

Thus, each of them crosses a different disc, say yp_i crosses d'_i with the colored point d_i in its center. Hence, Q_1^y intersects d'_1 and d'_2, Q_2^y intersects d'_2 and d'_3, and Q_3^y intersects d'_3 and d'_1. Firstly, assume that some d_i, does not lie on its corresponding line yp_i. Then by Equalities (2.3) even the centers of the other two d'_i cannot lie on their corresponding lines. Thus, all points of $R \cup B$ are in the interiors of some Q_i^y's. Therefore, if we round to the nearest integer the contribution of every disc d'_k to the value $\sum_{j=1}^{|B|} \mu(Q_1^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_1^y \cap r'_j)$, we obtain the value $\|Q_1^y\|_B - \|Q_1^y\|_R$. Furthermore, only disks d'_1 and d'_2 are only partly in Q_1^y. Thus, we could not change the value by 1 or more by the rounding. Therefore,

$$\sum_{j=1}^{|B|} \mu(Q_1^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_1^y \cap r'_j) = \|Q_1^y\|_B - \|Q_1^y\|_R$$

since both sides are integers. Analogously, the same holds for Q_2^y and Q_3^y, and so Equations (2.4) are satisfied.

It remains to solve the case when each d_i lies on its corresponding line yp_i. In this case, each d_i lies on the boundaries of two Q_i^y's, and we have to assign each of them to exactly one Q_i^y such that Equations (2.4) would hold. If all d_i have the same color, we assign one to every Q_i^y. If two of them, say d_1 and d_2 have one color, and the remaining one, d_3 in our case, has a different color we assign d_2 and d_3 to Q_2^y and we assign d_1 to Q_1^y (see Figure 2.2 for a case when two of them are red). It is clear that by this assignment we have

$$\sum_{j=1}^{|B|} \mu(Q_i^y \cap b'_j) - \sum_{j=1}^{|R|} \mu(Q_i^y \cap r'_j) = \|Q_i^y\|_B - \|Q_i^y\|_R$$

for every i. Hence, Equations (2.4) are satisfied.

□

In order to prove Theorem 10, we want to use an induction and apply Lemma 11 to some base cases. The problem are Conditions (2.1) in Lemma 11. Therefore, we will prove by induction a slightly different statement where this condition is satisfied by default. After that, we will do some initial steps and reduce Theorem 10 to this new problem.

Lemma 13. *Let Q be a convex polygon and a be a point on the boundary of Q. Additionally, let $s \geq 1$ be an integer and $p_1, p_2, \ldots, p_{s+1}$ be points inside Q such that p_1 and p_{s+1} lie on the boundary of Q and $(a, p_1, \ldots, p_{s+1})$ is a convex polygon inside Q (we also allow cases when either $p_1 = a$, or $p_{s+1} = a$). Let P be the convex polygon enclosed by $p_1p_2, \ldots, p_sp_{s+1}$ and by the part of the boundary of Q from p_{s+1} through a to p_1 in clockwise direction. Moreover, let B be a set of blue points inside P, and R be a set of red points outside P but inside Q such that $s = |B| - |R|$. Assume that $R \cup B \cup \{a, p_1, p_2, \ldots p_{s+1}\}$ is in general position.*

Then there exists a partition of Q into convex polygons $Q_1, \ldots, Q_s$ such that s_1s_{i+1} is a diagonal of Q_i and for every i, we have $\|Q_i\|_B - \|Q_i\|_R = 1$. Moreover, every point of $R \cup B$ is counted in only one Q_i.

Proof. We will use induction on s.

First assume that $s = 1$. In this case, set $Q_1 = Q$ and assign all points of $B \cup R$ on the boundary of Q to Q_1. Since $s = |B| - |R|$, we have $\|Q_1\|_B - \|Q_1\|_R = 1$.

Now assume that $s \geq 2$. For every p_i, $2 \leq i \leq s$, we say that $s_i a$ is left-partitionable if

$$\|Q \cap \text{right}(p_i a)\|_B - \|Q \cap \text{right}(p_i a)\|_R \leq i - 1$$

and we say that $p_i a$ is right-partitionable if

$$\|Q \cap \text{left}(p_i a)\|_B - \|Q \cap \text{left}(p_i a)\|_R \leq s - i + 1.$$

Since $s = |B| - |R|$ and the points p_i together with a are in general position, $p_i a$ is left-partitionable if

$$\|Q \cap \text{left}(p_i a)\|_B - \|Q \cap \text{left}(p_i a)\|_R \geq s - i + 1$$

and $p_i a$ is right-partitionable if

$$\|Q \cap \text{right}(p_i a)\|_B - \|Q \cap \text{right}(p_i a)\|_R \geq i - 1.$$

Moreover, every such $p_i a$ is left or right-partitionable.

Assume that $p_2 a$ is right-partitionable. Note that this does not happen if $a = p_1$ because there are no blue point to the right of $p_2 p_1$. Let l be a line containing $p_2 a$. By rotating l in the clockwise direction around p_2 we can find a point x on the boundary of Q between a and p_1 different from p_1 such that

$$\|Q \cap \text{right}(p_2 x)\|_B - \|Q \cap \text{right}(p_2 x)\|_R = 1$$

since

$$\|Q \cap \text{right}(p_2 p_1)\|_B - \|Q \cap \text{right}(p_2 p_1)\|_R \leq 0$$

(again note that there are no blue or red points on the line $p_2 p_1$ because $R \cup B \cup \{p_2, p_1\}$ is in general position). Thus, we can set $Q_1 = Q \cap \text{right}(p_2 x)$ and apply induction on the polygon $Q \cap \text{left}(p_2 x)$ and points $p_2, \ldots p_{s+1}$. Then the partition of $Q \cap \text{left}(p_2 x)$ together with Q_1 forms the desired partition of Q. See Figure 2.3.

Therefore, we can assume that $p_2 a$ is left-partitionable. By similar analysis for $p_s a$ we solve the case when $p_s a$ is left-partitionable. Hence, we can assume that $p_s a$ is right-partitionable.

Thus, we can find index j, $2 \leq j \leq s - 1$ such that $p_j a$ is left-partitionable and $p_{j+1} a$ is right-partitionable. Let T be the triangle $a p_j p_{j+1}$. Let $B' = B \cap T$, $n(a p_j) = j - 1 - \|\text{right}(p_j a)\|_B$, $n(p_{j+1} a) = s - j - \|\text{left}(p_{j+1} a)\|_B$, and $n(p_j p_{j+1}) = 1$. We want to use Lemma 11 on Q, triangle T inside Q, set of blue points B', set of red points R, and the numbers $n(a p_j), n(p_{j+1} a), n(p_j p_{j+1})$.

Clearly, $B' \cup R \cup \{a, p_j, p_{j+1}\}$ are in general position, B' is in the interior of T, R is outside T, and $|B'| - |R| = n(a p_j) + n(p_{j+1} a) + n(p_j p_{j+1})$. See Figure 2.4 for an illustration. It remains to check that Conditions (2.1) hold.

Let us first check it for the edge $p_j a$. Since $p_j a$ is left-partitionable,

$$\|Q \cap \text{right}(p_j a)\|_B - \|Q \cap \text{right}(p_j a)\|_R \leq j - 1.$$

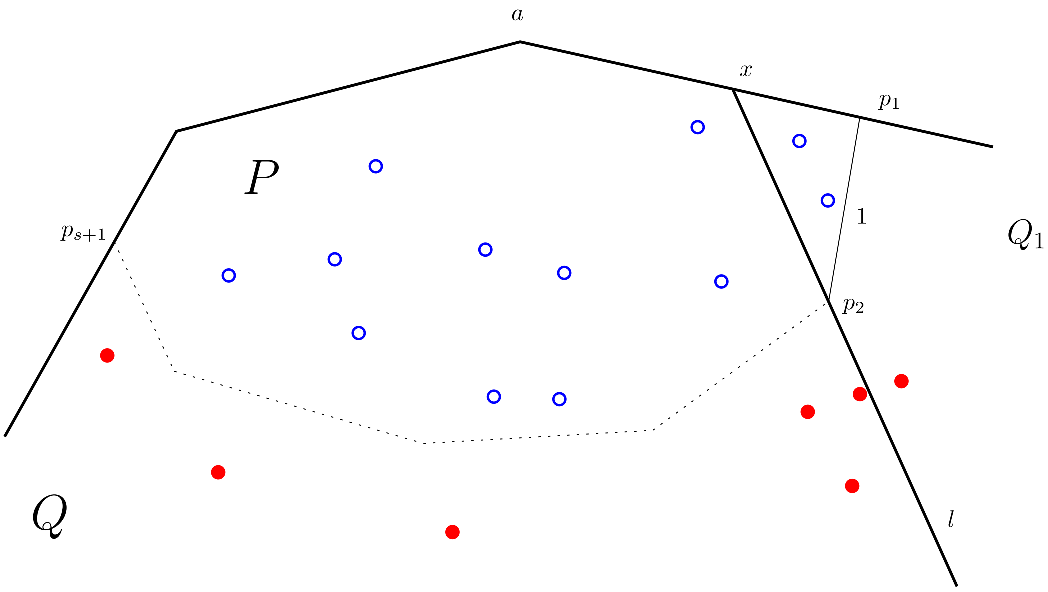


Figure 2.3: Induction step in Lemma 13 when p_2a is right-partitionable. Polygon Q_1 is the first part of the partition of Q. The remaining parts of the partition are obtained by induction hypothesis applied on the remaining part of Q.

Thus,

$$-\|Q \cap \text{right}(p_j a)\|_R \leq j - 1 - \|Q \cap \text{right}(p_j a)\|_B = n(ap_j).$$

Similarly, since $p_{j+1}a$ is right-partitionable, we have $-\|Q \cap \text{left}(p_{j+1}a)\|_R \leq n(ap_{j+1})$. Additionally, since a in on the boundary of Q, the interior of $\text{left}(p_{j+1}a) \cap \text{right}(p_j a)$ is entirely outside Q. Thus,

$$\begin{aligned} n(ap_{j+1}) + n(ap_j) &\geq -\|Q \cap \text{right}(p_j a)\|_R - \|Q \cap \text{left}(p_{j+1}a)\|_R = \\ &= -\|Q \cap (\text{right}(p_j a) \cup \text{left}(p_{j+1}a))\|_R . \end{aligned}$$

The remaining conditions follow immediately from these ones since $n(p_jp_{j+1})$ is a positive number. Thus, by Lemma 11, Q can be partition into convex polygons O_1, O_2, O_3 such that $ap_j, p_jp_{j+1}, ap_{j+1}$ are diagonals or edges of O_1, O_2, O_3, respectively, and $\|O_1\|_{B'} - \|O_1\|_R , \|O_2\|_{B'} - \|O_2\|_R , \|O_3\|_{B'} - \|O_3\|_R$ are equal to $n(ap_j), n(p_jp_{j+1}), n(ap_{j+1})$, respectively. See Figure 2.4 for an illustration. Hence,

$$\begin{aligned} \|O_1\|_B - \|O_1\|_R &= n(ap_j) + \|Q \cap \text{left}(ap_j)\|_B = j - 1 \\ \|O_2\|_B - \|O_2\|_R &= n(p_jp_{j+1}) + \|Q \cap \text{left}(p_jp_{j+1})\|_B = 1 \\ \|O_3\|_B - \|O_3\|_R &= n(ap_{j+1}) + \|Q \cap \text{left}(p_{j+1}a)\|_B = s - j \end{aligned}$$

Therefore, we can apply the induction hypothesis to the polygon O_1, point a on the boundary of O_1, points $p_1, \ldots, p_j$, the set of blue points $B \cap O_1$, and the set of red points $R \cap O_1$ (if some red or blue point are on the boundary of O_1, we include only the ones assigned to O_1 by Lemma 11) to obtain partition $Q_1, \ldots, Q_{j-1}$, of O_1. Similarly, we obtain a partition $Q_{j+1}, \ldots, Q_{s+1}$ of O_3 by induction hypothesis applied to O_3 and corresponding points.

Furthermore, since $n(p_jp_{j+1}) = 1$, p_j, p_{j+1} is a diagonal of O_2 and we can set $Q_j = O_2$. Finally, $Q_1, \ldots, Q_s$ is the desired partition of Q. □

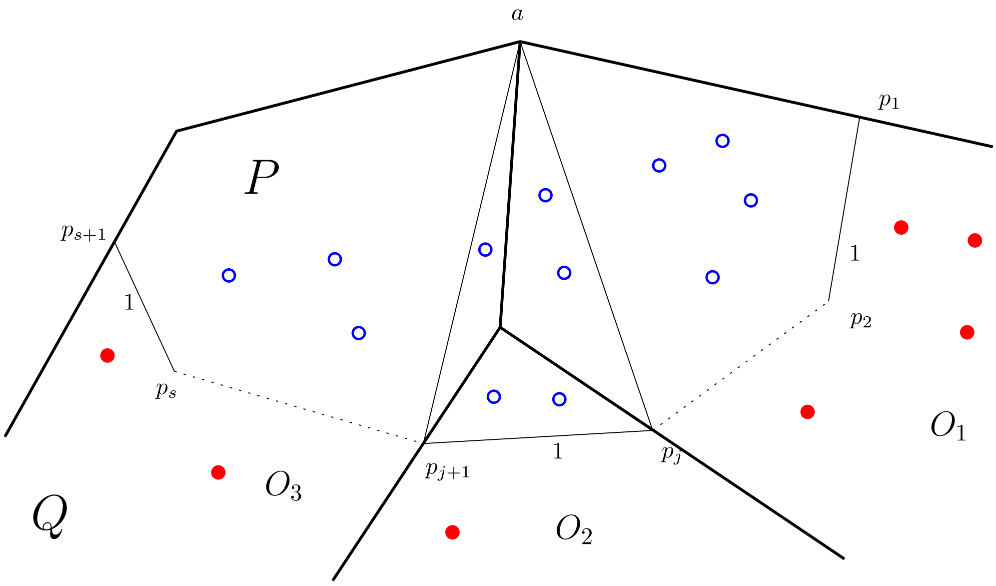


Figure 2.4: Induction step in Lemma 13. Polygon Q is split into three parts O_1, O_2, O_3. Induction hypothesis can then be applied on polygons O_1 and O_3 to obtain a complete partition of Q.

It remains to prove Theorem 10. We will proceed similarly as in the proof of the previous lemma. The only real difference is that previously we had one point of P on the boundary of Q. Now the first step will have to be a bit more complicated but still very similar.

Proof of Theorem 10. For every p_i, $2 \leq i \leq s$, we say that $p_i p_1$ is left-partitionable if

$$\|\text{left}(p_i p_1)\|_B - \|\text{left}(p_i p_1)\|_R \geq s - i + 1$$

or, equivalently, if

$$\|\text{right}(p_i p_1)\|_B - \|\text{right}(p_i p_1)\|_R \leq i - 1.$$

This equivalence holds since $s = |B| - |R|$.

Similarly, we say that $p_i p_1$ is right-partitionable if

$$\|\text{right}(p_i p_1)\|_B - \|\text{right}(p_i p_1)\|_R \geq i - 1$$

or, equivalently, if

$$\|\text{left}(p_i p_1)\|_B - \|\text{left}(p_i p_1)\|_R \leq s - i + 1.$$

Furthermore, every such $p_i p_1$ is left or right-partitionable.

Note that $p_2 p_1$ is left-partitionable because it is an edge of P, and so there are no blue points to the right of $p_2 p_1$. Similarly, $p_s p_1$ is right-partitionable. Therefore, we can find j, $2 \leq j \leq s - 1$ such that $p_j p_1$ is left-partitionable and $p_{j+1} p_1$ is right-partitionable.

Let T be the triangle $p_1 p_j p_{j+1}$. Let $B' = B \cap T$, $n(p_1 p_j) = j - 1 - \|\text{right}(p_j p_1)\|_B$, $n(p_{j+1} p_1) = s - j - \|\text{left}(p_{j+1} p_1)\|_B$, and $n(p_j p_{j+1}) = 1$. The situation is almost identical as in the previous proof. Thus, we would like to use Lemma 11 on the whole plane, triangle T, the set of blue points B', the set of red points R, and the numbers $n(p_1p_j), n(p_1p_{j+1}), n(p_jp_{j+1})$. All the assumptions needed in Lemma 11 are satisfied in the similar way as in the previous proof with the exception of the following condition:

$$n(p_1p_j) + n(p_1p_{j+1}) \geq -\left\|(\text{right}(p_jp_1) \cup \text{left}(p_{j+1}p_1))\right\|_R.$$

Hence, if this inequality holds, we can partition the plane into three convex polygons, two of which we can further partition by Lemma 13 as in the previous proof. In this way, we obtain the desired partition of the plane.

Now assume that

$$n(p_1p_j) + n(p_1p_{j+1}) < -\left\|\text{right}(p_jp_1) \cup \text{left}(p_{j+1}p_1)\right\|_R.$$

Note that since $p_{j+1}p_1$ is right-partitionable and p_jp_1 is left-partitionable, this can happen only if $3 \leq j \leq s-2$.

We can further split the region $\text{right}(p_jp_1) \cup \text{left}(p_{j+1}p_1)$ and write

$$\begin{aligned} n(p_1p_j) + n(p_1p_{j+1}) < &-\left\|\text{right}(p_jp_1) \cap \text{right}(p_{j+1}p_1)\right\|_R - \\ &- \left\|\text{left}(p_jp_1) \cap \text{left}(p_{j+1}p_1)\right\|_R - \left\|\text{right}(p_jp_1) \cap \text{left}(p_{j+1}p_1)\right\|_R. \end{aligned} \tag{2.5}$$

Since p_jp_1 is left-partitionable, $\left\|\text{right}(p_jp_1)\right\|_B - \left\|\text{right}(p_jp_1)\right\|_R \leq j-1$. By substituting $n(p_1p_j)$ for $j - 1 - \left\|\text{right}(p_jp_1)\right\|_B$, we have

$$n(p_1p_j) \geq -\left\|\text{right}(p_jp_1)\right\|_R. \tag{2.6}$$

Similarly, since $p_{j+1}p_1$ is right-partitionable, $n(p_1p_{j+1}) \geq -\left\|\text{left}(p_{j+1}, p_1)\right\|_R$. By combining this with Equation (2.5) we get

$$n(p_1p_j) < -\left\|\text{right}(p_jp_1) \cap \text{right}(p_{j+1}p_1)\right\|_R. \tag{2.7}$$

Equations (2.6) and (2.7) imply that there exists a directed half-line l starting at p_1 that splits the region $\text{right}(p_jp_1) \cap \text{left}(p_{j+1}p_1)$ in a way that l does not cross any point of $R \cup B$, and

$$n(p_1p_j) = -\left\|\text{right}(p_jp_1) \cap \text{right}(l)\right\|_R. \tag{2.8}$$

This together with Equation (2.5) implies that

$$n(p_1p_{j+1}) < -\left\|\text{left}(l) \cap \text{left}(p_{j+1}p_1)\right\|_R. \tag{2.9}$$

Therefore, the plane is split into two polygons; convex one $Q^* = \text{right}(p_jp_1) \cap \text{right}(l)$ and non-convex one $Q' = \text{left}(l) \cup \text{left}(p_jp_1)$. See Figure 2.5. If we substitute for the value of $n(p_1p_j)$ into Equation (2.8) we see that

$$j - 1 - \left\|\text{right}(p_jp_1)\right\|_B = -\left\|Q^*\right\|_R.$$

Furthermore, note that the blue points are only inside P. Thus,

$$j - 1 = \left\|Q^*\right\|_B - \left\|Q^*\right\|_R \tag{2.10}$$

Hence, we can apply Lemma 13 to the polygon Q^* with point p_1 on its boundary, points $p_1, \ldots, p_j$, set of blue points $B \cap Q^*$ and set of red points $R \cap Q^*$. This way we obtain a partition of Q^* into convex polygons $Q_1, \ldots, Q_{j-1}$.

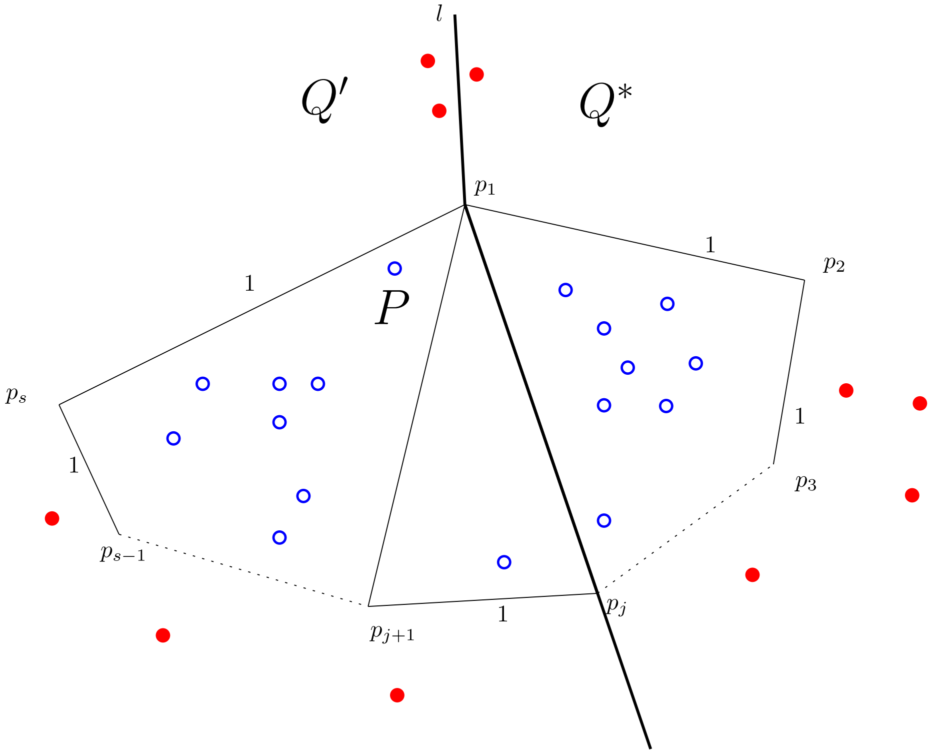


Figure 2.5: First step of partitioning the convex polygon P.

Thus, it suffice to partition Q'. Equation (2.10) together with the fact that $s = |B| - |R|$ implies that $s - j + 1 = \|Q'\|_B - \|Q'\|_R$. We would like to apply Lemma 13 to the polygon Q' with point p_1 on its boundary, points $p_j, \ldots, p_s, p_1$, set of blue points $B \cap Q'$ and set of red points $R \cap Q'$. The only problem is that Q' is not convex. Luckily, it turns out that it does not matter in this case because even if we applied the same approach, the parts of the partition would still be convex. Since the proof is basically the same as in the proof of Lemma 13 we present a more concise version.

Let $R' = R \cap Q'$ and $B' = B \cap Q'$. The important part is that Equation (2.9) implies that the diagonal $p_{j+1}p_1$ is left-partitionable in Q'. Furthermore, $p_s \neq p_{j+1}$ since $s \geq j + 2$. Therefore, we once again find some index k, $j + 1 \leq k \leq s - 1$, such that $p_k p_1$ is left-partition-able and $p_{k+1}p_1$ is right-partitionable. Let T' be the triangle $p_1 p_k p_{k+1}$. Let $n(p_1 p_k) = k - j - \|\text{right}(p_k p_1)\|_{B'}$, $n(p_{k+1}p_1) = s - k - \|\text{left}(p_{k+1}p_1)\|_{B'}$, and $n(p_k p_{k+1}) = 1$. We apply Lemma 11 on the whole plane, triangle T, set of blue points B', set of red points R', and the numbers $n(p_1 p_k), n(p_1 p_{k+1}), n(p_k p_{k+1})$. The Conditions (2.1) are satisfied since $p_k p_1$ is left-partition-able, $p_{k+1}p_1$ is right-partitionable, and $\text{right}(p_k p_1) \cap \text{left}(p_{k+1}p_1)$ does not contain any point of B' or R' (because its interior is disjoint with Q'). Thus, we can partition the plane into convex polygons O_1, O_2, O_3 such that $p_1 p_k, p_k p_{k+1}, p_1 p_{k+1}$ are diagonals or edges of O_1, O_2, O_3, respectively, and other conditions about number of blue and red points inside these polygon holds. It is immediate that all intersection $O_i \cap Q'$ are also convex polygons.

Therefore, we can apply Lemma 13 to the convex polygon $Q' \cap O_1$ with the point p_1 on its boundary, points $p_j, \ldots, p_k$, the set of blue points $B' \cap O_1$, and the set of red points $R' \cap O_1$ (if some red and blue point are on the boundary of O_1, we include only the ones assigned to O_1 by Lemma 11) to obtain a partition

$Q_j, \ldots, Q_{k-1}$, of $Q' \cap O_1$. Similarly, we obtain a partition $Q_{k+1}, \ldots, Q_{s+1}$ of O_3 by Lemma 13 applied to $Q' \cap O_3$ and corresponding points.

Finally, we can set $Q_k = Q' \cap O_2$, and the partition $Q_1, \ldots, Q_s$ is the desired partition of the plane.

□

2.2 Alternating paths covering all red and blue points

Before we proceed to prove Theorem 3 we need a result already proved by Abellanas et al. [1]. We provide our proof to keep the thesis self contained.

Theorem 14 ([1]). *Let R be a set of red points and B be a set of blue points such that $R \cup B$ is in general position. Assume that $||R| - |B|| \leq 1$ and that R can be separated from B by a line. Then there exists a non-crossing alternating geometric path covering the entire point set $R \cup B$ such that this path begins and ends in end vertices of the alternating edges of the convex hull of $R \cup B$.*

Proof. Denote the separating line by s. By rotating the whole plane, we can assume that s is vertical, R is on the left side of s, and B is on the right side of s. Let $T = R \cup B$. For every subset X of T containing at least one point of each color, there exists one or two edges of $\text{conv}(X)$ crossing s. These edges are alternating, and we call the top one *top alternating edge* of $\text{conv}(X)$ and the bottom one *bottom alternating edge* of $\text{conv}(X)$.

We will inductively build a sequence $(p_1, p_2, \ldots, p_{|T|})$ of all points of T such that connecting these points in the given order by straight-line segments forms a non-crossing alternating geometric path. We set p_1 to be one of the two vertices of the top alternating edge of $\text{conv}(T)$ whose color is the more numerous one (if $|R| = |B|$ it does not matter which).

Assume that we have already built the sequence up to p_i. Let S_i be the set containing the remaining points. We set p_{i+1} to be the vertex of the top alternating edge of $\text{conv}(S_i)$ colored by different color than p_i. If $|T \setminus P_i| = 2$ we set $s_{|T|}$ to be the last unselected point and the path is complete. Clearly, the formed path is alternating. Moreover, in every step the edge $p_i p_{i+1}$ lies in $\text{conv}(S_i \cup \{p_i\}) \setminus \text{conv}(S_i)$ since p_i is a vertex of the top alternating edge of $S_i \cup \{p_i\}$ and p_{i+1} is a vertex of the top alternating edge of S_i, see Figure 2.6 for illustration. Therefore, the formed path is also non-crossing.

Furthermore, it is clear from the construction that the first vertex is a vertex of the top alternating edge of $\text{conv}(T)$ and the last one is a vertex of the bottom alternating edge of $\text{conv}(T)$. □

We have everything ready to finish the proof of Theorem 3.

Proof of Theorem 3. Let s be the number of vertices of P and $(p_1, \ldots, p_s)$ be the vertices of P. Let $R' = R \setminus P$. That is, R' contains exactly the points of R that are not vertices of P. We will assume that $|R| = |B|$ and we will show that there exists a non-crossing alternating geometric cycle that covers all vertices. If $|R|$ and $|B|$ differ by one, we can add one temporary point of the appropriate color,

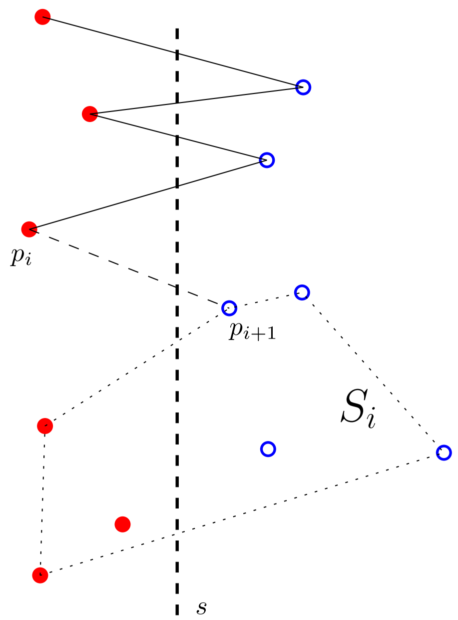


Figure 2.6: Inductive forming of a non-crossing alternating geometric path when the color classes are separated by a line.

and after we remove it in the end, we will still have a non-crossing alternating geometric path. Therefore, $s = |B| - |R'|$.

By Theorem 10 applied on the polygon P, set of blue points B and set of red points R', there exists a partition of the plane into convex polygons $Q_1, \ldots, Q_s$ such that for every i, the edge $p_i p_{i+1}$ is a diagonal of Q_i, and for every i, we have $\|Q_i\|_B - \|Q_i\|_{R'} = 1$ (index arithmetic is modulo s). Moreover, every point of $R' \cup B$ is counted in exactly one Q_i. That is, if a point of $R \cup B$ is on boundaries of more Q_i's it is assigned to exactly one of them.

Therefore, we want to apply Theorem 14 to each Q_i separately. Each Q_i contains one more blue point than it contains points of R'. Moreover, these red and blue points are separated by line $p_i p_{i+1}$. Additionally, there are two more red points p_i, p_{i+1}; the end vertices of edge $p_i p_{i+1}$. Thus, by Theorem 14 there exists a non-crossing alternating geometric path with ends in p_i and p_{i+1} covering all these red and blue points inside Q_i. Note that this path is inside the convex polygon Q_i since it consists of straight-line segments.

Clearly, these paths are connected in the end vertices p_i. Therefore, together they form an alternating geometric cycle covering all points of $R \cup B$. Furthermore, this cycle is non-crossing since each path is in its own convex polygon Q_i, points are in general position, and every point of $R' \cup B$ is assigned to exactly one Q_i. See Figure 2.7 for an illustration. □

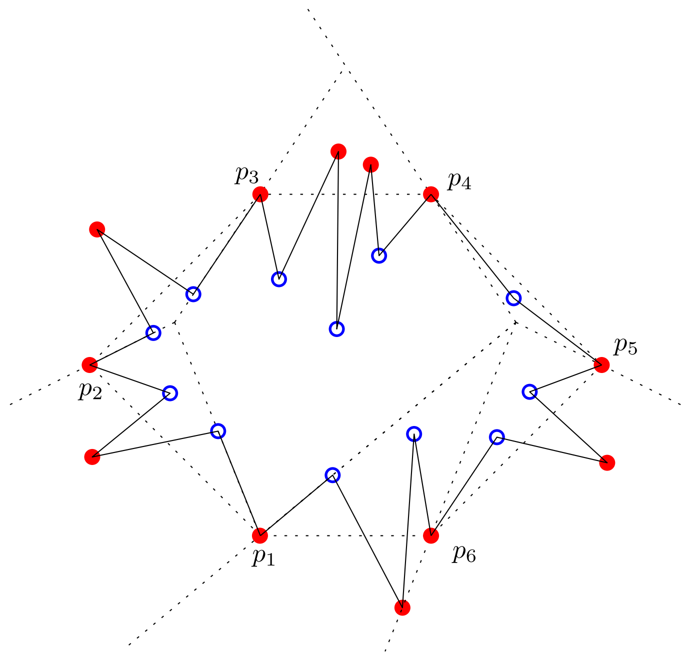


Figure 2.7: A non-crossing alternating geometric cycle in a case when 6 red points form a polygon separating the remaining 6 red points from 12 blue points lying inside the polygon.

Conclusion

In this thesis, we managed to find a connection between separated matchings and common subsequences of two binary words. With this connection, we improved lower bounds on the sizes of separated matchings on convex point sets with discrepancy at most 2. We believe that similar results could be achieved even for larger discrepancies by more thorough analysis. Specifically, we believe the following to be true:

Conjecture 2. *Let k be a positive integer constant. Then $p_{(k,k)}(n) = \frac{2k}{3k-1} - o(1)$.*

Theorem 6 shows the corresponding upper bound. Thus, it suffices to show that every pair of sequences with discrepancy at most k of length n have matching potential at least $\frac{2k}{3k-1} - o(1)$. By Lemmma 5, it would imply that $\mu_k(n) \geq \frac{4k}{3k-1}n - o(n)$. Thus, if true, it would mean that the long believed to be true conjecture $\mu(n) \geq \frac{4}{3}n + o(n)$ holds true in cases with discrepancy at most k where k is a constant. The only example that disproves this conjecture (in [8]) has discrepancy approximately $n^{\frac{4}{5}}$.

We also believe that the connection between the matching potential and the parameter $\mu(n)$ is much tighter, and $\mu_d(n)$ can be bounded from the above by $p_{(d,d)}(n)$. Specifically, we believe the following to be true.

Conjecture 3. *Let d and c be positive constants. Let $p_{(d,d)}(n) \leq c + o(1)$. Then $\mu_d(n) \leq 2cn + o(n)$.*

In a case with unbounded discrepancies, the situation seems to be more difficult. For example, if we allow discrepancy n, then the matching potential of two sequences of lengths $2n$ can be 0. Nevertheless, Theorem 5 works even for unbounded discrepancies, and it is quite likely that a variation of Conjecture 3 with some additional dependence on d would also hold true. Since all the worst cases seem to arise when the discrepancy is sublinear in n, it could prove vital in determining even the general parameter $\mu(n)$.

Our additional focus was on cases with point sets in general position. We managed to found a new configuration of points when a non-crossing alternating geometric path covering all points exists. During our attempts to prove Theorem 3, we encountered an interesting partitioning problem and formulated Conjecture 1. Even though we were ultimately unable to prove this conjecture in a general case, we still believe it to be true. Unfortunately, we were also unable to use our results to better understand the behavior of the parameter $l^g(n)$. With additional time we believe that at least sublinear improvements on the lower bound on $l^g(n)$ is possible and would not be that difficult to prove.

During our research, we also tried to look into problems concerning geometric trees on bicolored point sets. Most of our initial findings turned out to be already known and published before. Thus, we decided we need more time to research the related problems further.

Bibliography

[1] M. Abellanas, J. García, G. Hernández, M. Noy, and P. Ramos. "Bipartite embeddings of trees in the plane". *Discrete Appl. Math.* 93.2-3 (1999), pp. 141–148. DOI: 10.1016/S0166-218X(99)00042-6.

[2] M. Abellanas, A. Olaverri, F. Hurtado, and J. Tejel. "Caminos alternantes". Spanish. In: *X Encuentros de Geometría Computacional.* Sevilla, 2003, pp. 7–12.

[3] F. Aurenhammer. "Weighted skeletons and fixed-share decomposition". *Comput. Geom.* 40.2 (2008), pp. 93–101. ISSN: 0925-7721. DOI: 10.1016/j.comgeo.2007.08.002.

[4] K. C. Border. *Fixed point theorems with applications to economics and game theory.* Cambridge University Press, Cambridge, 1989, pp. viii+129. ISBN: 0-521-38808-2.

[5] P. Brass, W. Moser, and J. Pach. *Research problems in discrete geometry.* Springer, New York, 2005, pp. xii+499. ISBN: 978-0387-23815-8; 0-387-23815-8.

[6] V. Chvátal and D. Sankoff. "Longest common subsequences of two random sequences". *J. Appl. Probability* 12 (1975), pp. 306–315. ISSN: 0021-9002. DOI: 10.2307/3212444.

[7] J. Cibulka, J. Kynčl, V. Mészáros, R. Stolař, and P. Valtr. "Universal sets for straight-line embeddings of bicolored graphs". In: *Thirty essays on geometric graph theory.* Springer, New York, 2013, pp. 101–119. DOI: 10.1007/978-1-4614-0110-0_8.

[8] E. Csóka, Z. L. Blázsik, Z. Király, and D. Lenger. *Upper bounds for the necklace folding problems.* 2020. arXiv: 2005.12603v1 [math.CO].

[9] J. D. Dixon. *Longest common subsequences in binary sequences.* 2013. arXiv: 1307.2796 [math.GR].

[10] A. García and J. Tejel. "Dividiendo una nube de puntos en regiones convexas". Spanish. In: *Actas VI Encuentros de Geometría Computacional.* Barcelona, 1995, pp. 169–174.

[11] A. Kaneko and M. Kano. "Discrete geometry on red and blue points in the plane—a survey". In: *Discrete and computational geometry.* Vol. 25. Algorithms Combin. Springer, Berlin, 2003, pp. 551–570. DOI: 10.1007/978-3-642-55566-4_25.

[12] M. Kano and J. Urrutia. "Discrete geometry on colored point sets in the plane—a survey". *Graphs Combin.* 37.1 (2021), pp. 1–53. DOI: 10.1007/s00373-020-02210-8.

[13] J. Kynčl, J. Pach, and G. Tóth. "Long alternating paths in bicolored point sets". *Discrete Math.* 308.19 (2008), pp. 4315–4321. DOI: 10.1016/j.disc.2007.08.013.

[14] G. S. Lueker. "Improved bounds on the average length of longest common subsequences". *J. ACM* 56.3 (2009), Art. 17, 38. ISSN: 0004-5411. DOI: 10.1145/1516512.1516519.

[15] V. Mészáros. "Separated matchings and small discrepancy colorings". In: *Computational geometry*. Vol. 7579. Lecture Notes in Comput. Sci. Springer, Cham, 2011, pp. 236–248. DOI: 10.1007/978-3-642-34191-5_23.

[16] V. Mészáros and P. Hajnal. "Note on noncrossing path in colored convex sets". 2010. URL: https://infoscience.epfl.ch/record/175677/files/?ln=en.

[17] C. Müllner and A. Ryzhikov. "Palindromic subsequences in finite words". In: *Language and automata theory and applications*. Vol. 11417. Lecture Notes in Comput. Sci. Springer, Cham, 2019, pp. 460–468. DOI: 10.1007/978-3-030-13435-8_3.

[18] W. Mulzer and P. Valtr. *Long Alternating Paths Exist*. 2020. arXiv: 2003.13291v1 [cs.CG].

List of Figures

A. Attachments

A.1 First Attachment

Assume that P and O are sequences of red and blue points of length 22 such that they can be separated into intervals of length 2; each interval containing one red and one blue point. We need to show that either they contain prefixes of even length with efficiency at least $\frac{4}{5}$ or they contain a pair of prefixes of one of the following forms, possibly with all colors switched or with switched order of sequences inside the pairs, but these cases are symmetric:

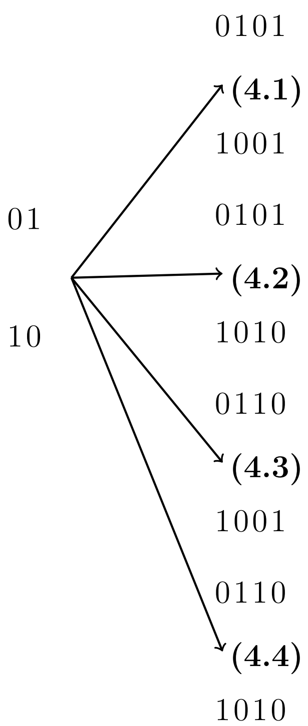
1. $(0110100101010101011010, 10010110011001100101)$

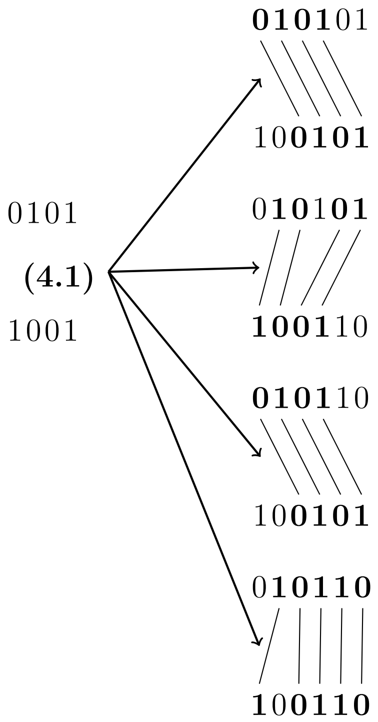
2. $(0110100101010101010101, 10010110011001100110)$

We can assume that P starts with 01 and O starts with 10. If they started with the same colored pair of points, they would contain prefixes of even length with efficiency 1. The remaining case is symmetric with this one.

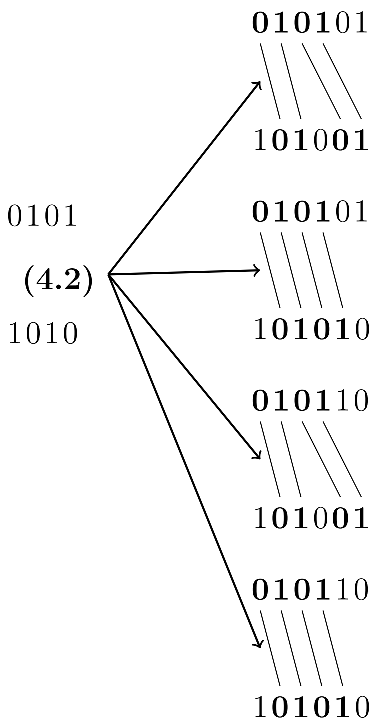
We iteratively try to extend these possible starts with intervals 01 or 10. Whenever there is a pair of prefixes of even length with an efficiency of at least $\frac{4}{5}$, we graphically show the corresponding matching and stop expanding this particular pair. In this way, we construct all possible pairs of prefixes of P and O of the length 4, then of the length 6, and so on until we construct all possible pairs of prefixes of P and O one of length 22 and the other of length 20. We see that indeed the only pairs of prefixes of P and O one of length 22 and the other of length 20 that does not contain any pair of prefixes of even length with efficiency at least $\frac{4}{5}$ are pairs (1) and (2).

For better readability, we label each pair of sequences that we still need to extend. That way it is easier to find how the pairs are extended next.

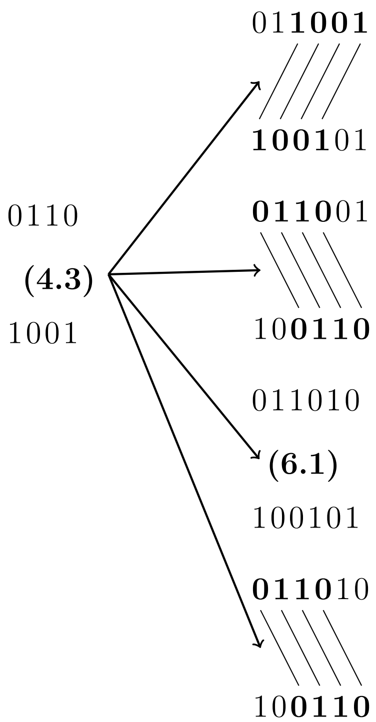


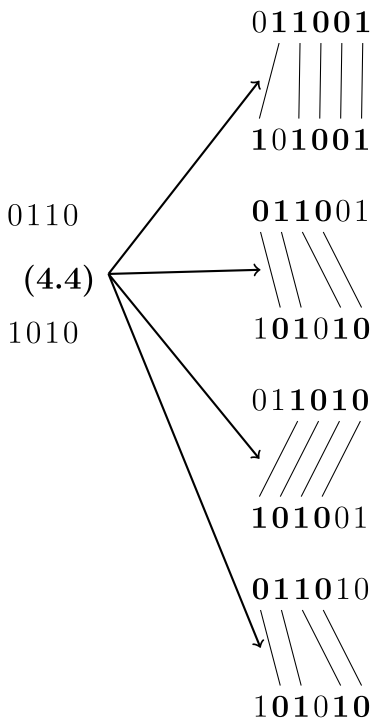


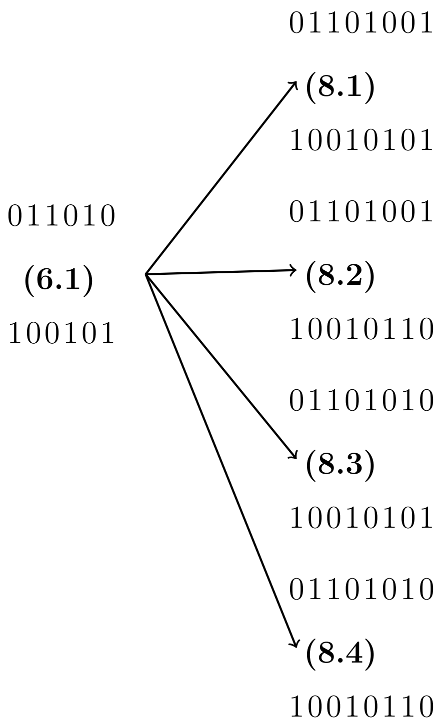
010101
100101
0101
(4.1)
1001
010101
100110
010110
100101
010110
100110

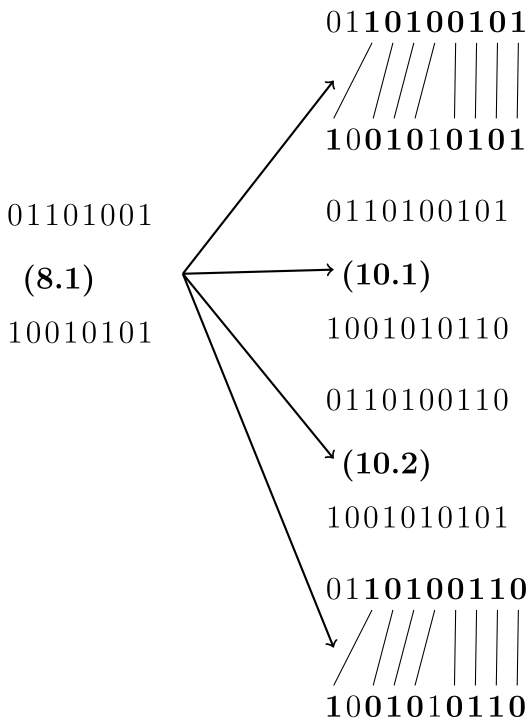


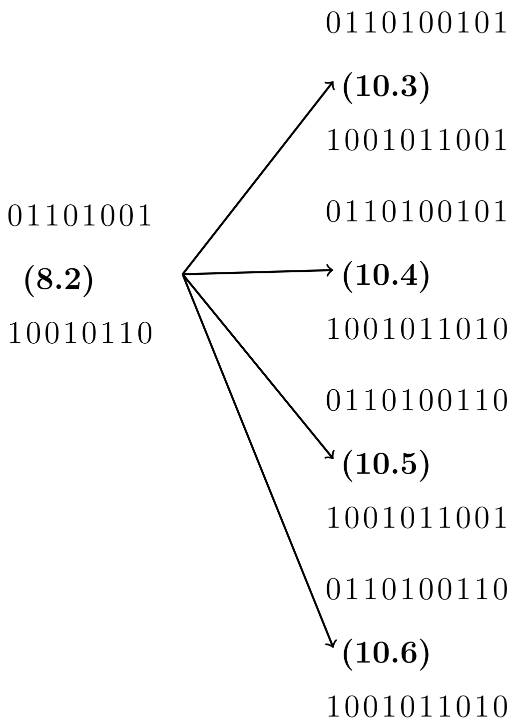
010101
101001
0101
(4.2)
1010
010101
101010
010110
101001
010110
101010


0110
(4.3)
1001
011001
100101
011001
100110
011010
(6.1)
100101
011010
100110

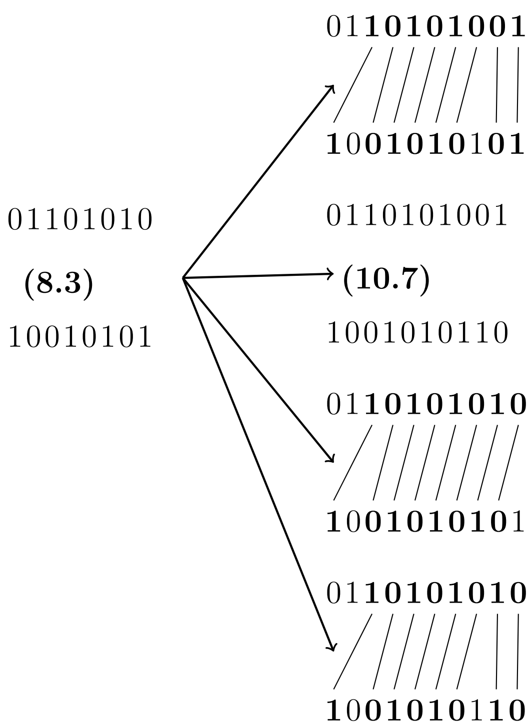

0110
(4.4)
1010
011001
101001
011001
101010
011010
101001
011010
101010

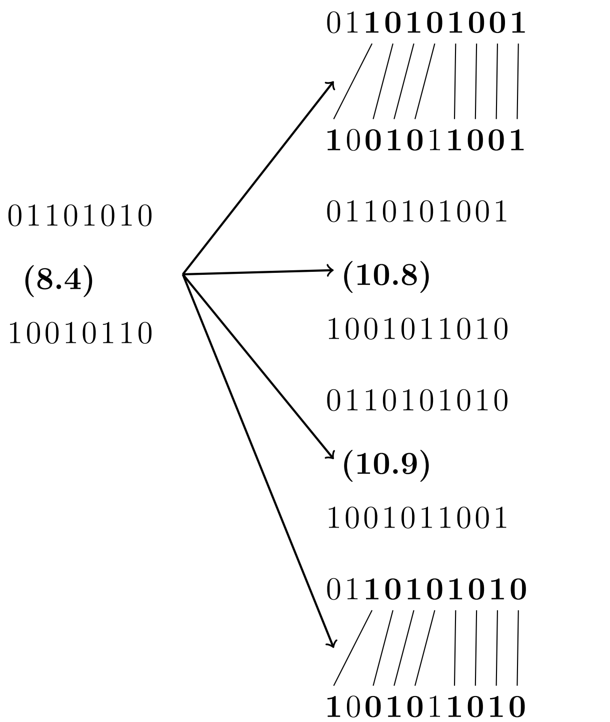

01101001
(8.1)
10010101
011010
(6.1)
100101
01101001
(8.2)
10010110
01101010
(8.3)
10010101
01101010
(8.4)
10010110

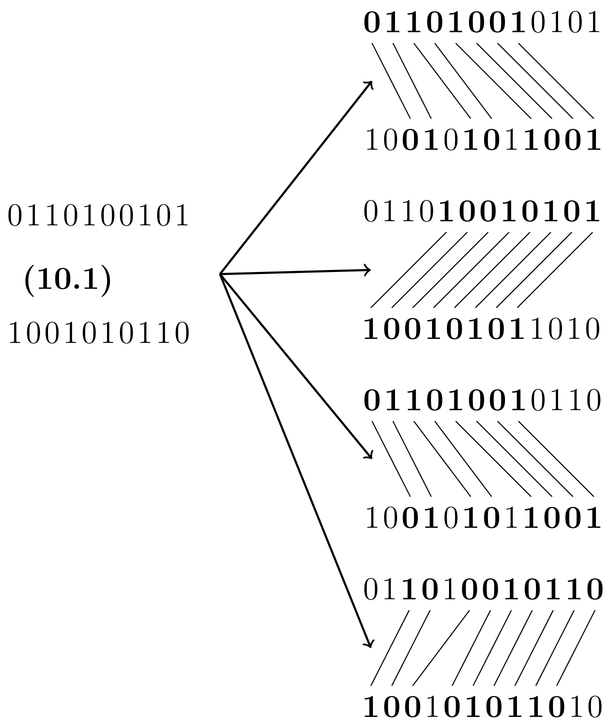

0110100101
1001010101
01101001
(8.1)
10010101
0110100101
(10.1)
1001010110
0110100110
(10.2)
1001010101
0110100110
1001010110

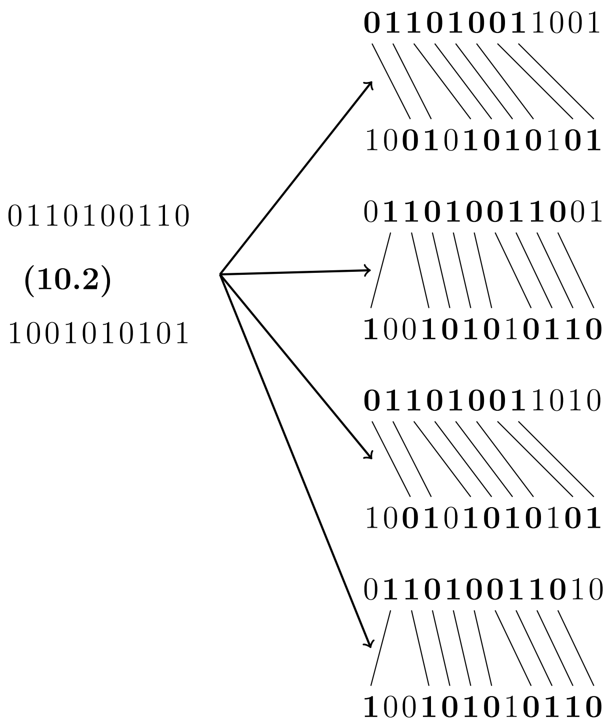


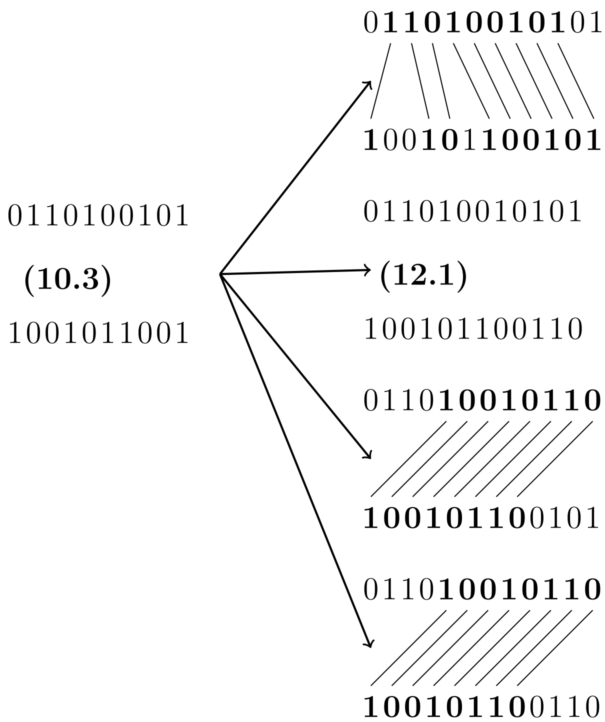
When we switch all colors and switch the order of sequence in the pair $(0110100101, 1001011001)$ we get the pair $(0110100110, 1001011010)$. Therefore, these cases are symmetric and we can continue expanding only the first of these pairs.

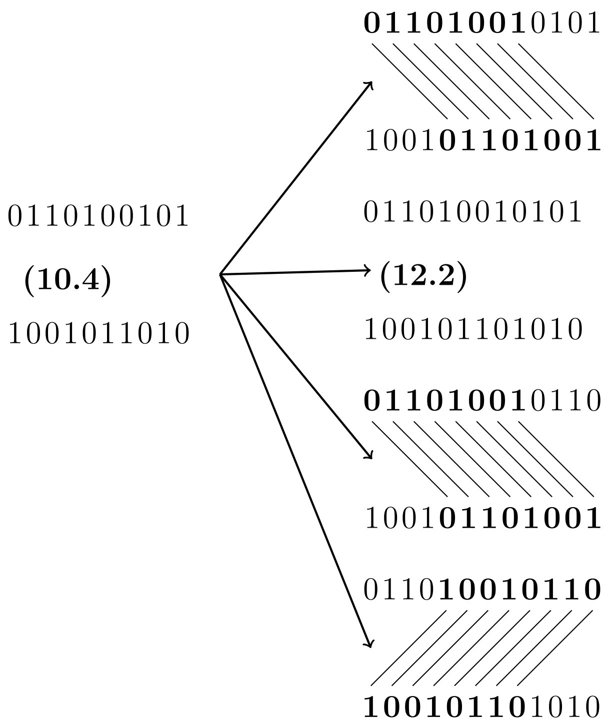


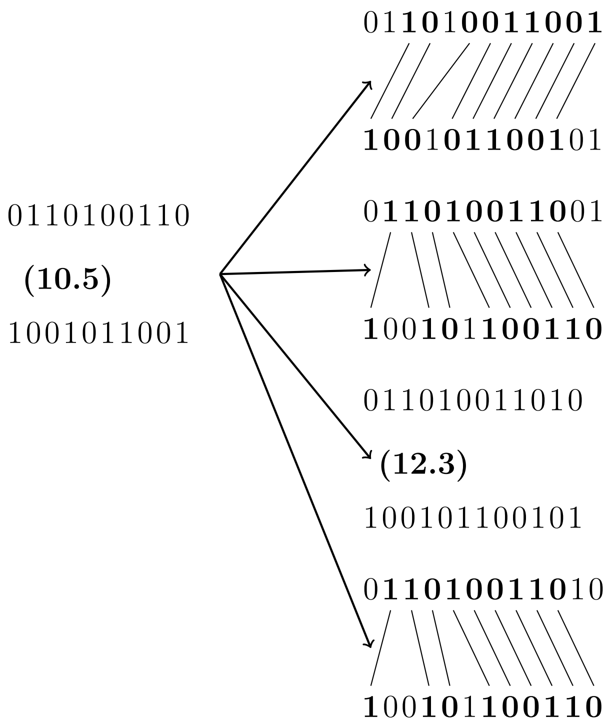

0110101001
1001011001
01101010
(8.4)
10010110
0110101001
(10.8)
1001011010
0110101010
(10.9)
1001011001
0110101010
1001011010

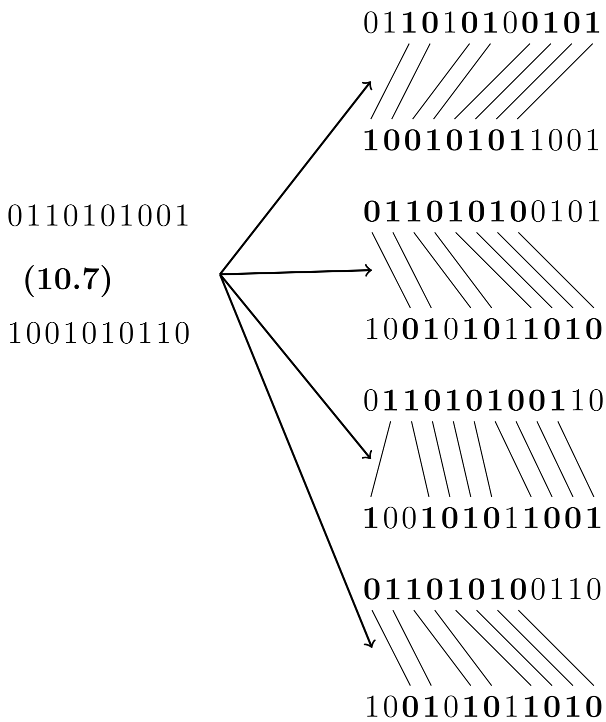

0110100101
(10.1)
1001010110
011010010101
100101011001
011010010101
100101011010
011010010110
100101011001
011010010110
100101011010


0110100110
(10.2)
1001010101
011010011001
100101010101
011010011001
100101010110
011010011010
100101010101
011010011010
100101010110

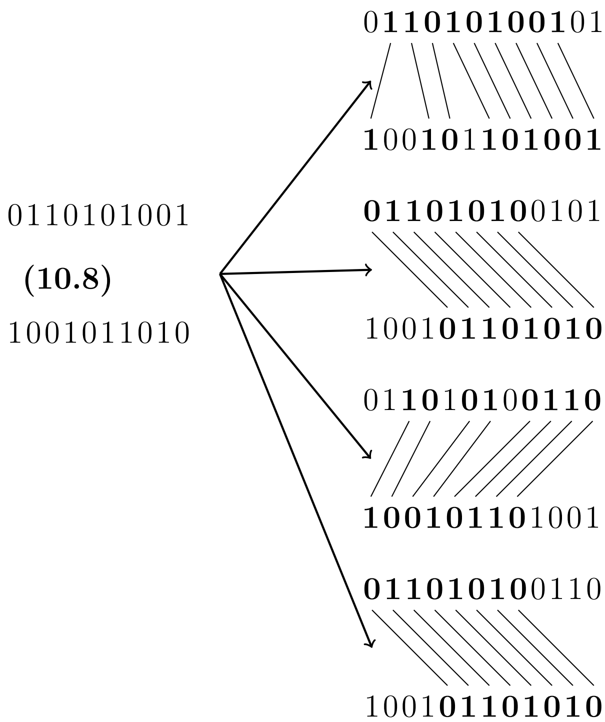

0110100101
(10.3)
1001011001
011010010101
100101100101
011010010101
(12.1)
100101100110
011010010110
100101100101
011010010110
100101100110


011010010101
100101101001
0110100101
(10.4)
1001011010
011010010101
(12.2)
100101101010
011010010110
100101101001
011010010110
100101101010

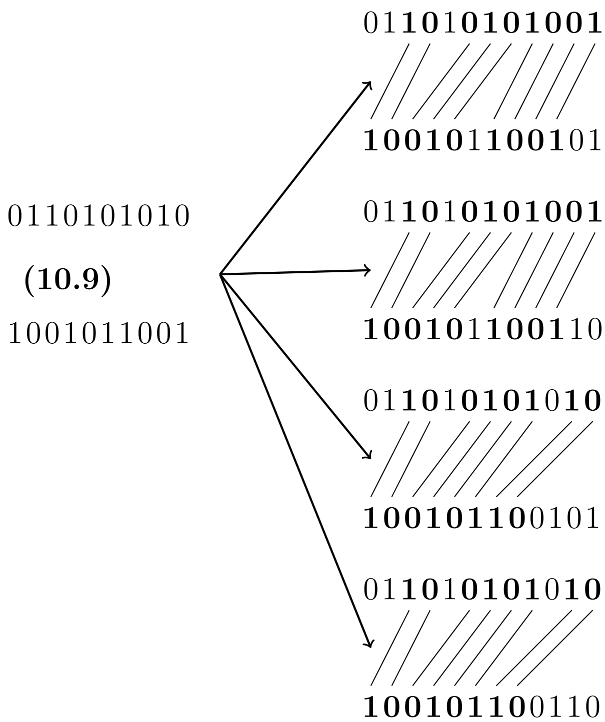

011010011001
100101100101
0110100110
(10.5)
1001011001
011010011001
100101100110
011010011010
(12.3)
100101100101
011010011010
100101100110

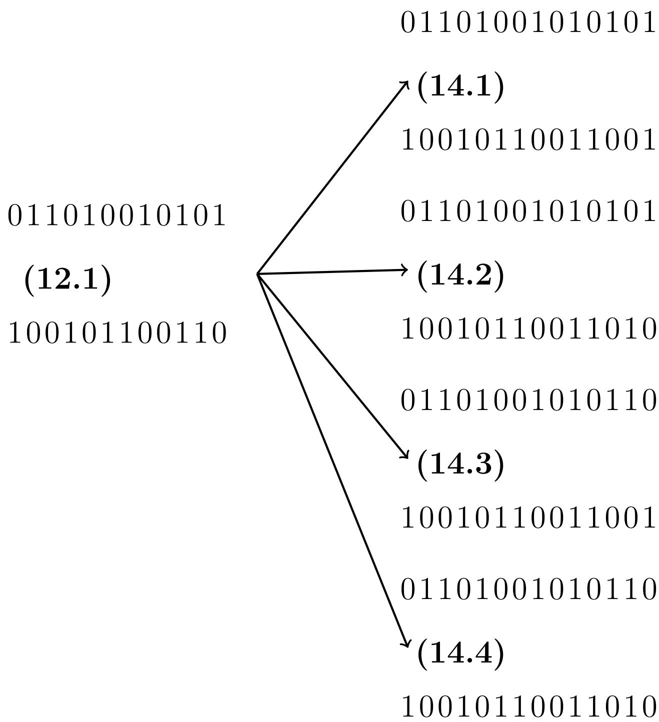

0110101001
1001010110
01101010100101
10010101101001
01101010100101
10010101011010
01101010100110
10010101011001
01101010100110
10010101011010

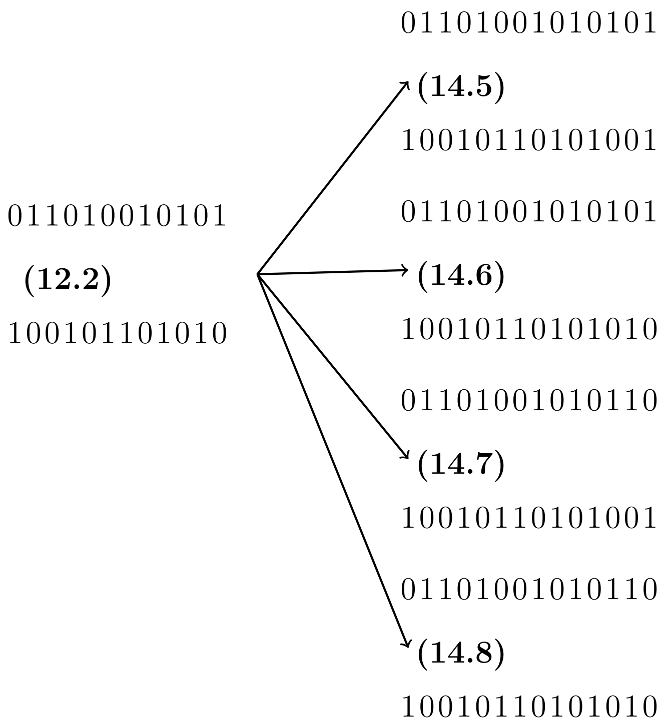
(10.7)

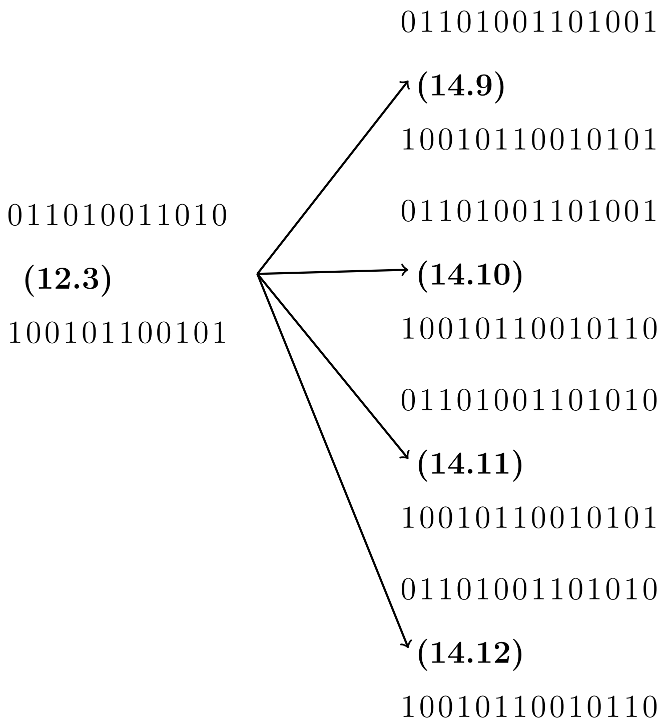

0110101001
1001011010
01101010010101
10010110101001
01101010100101
10010110101010
01101010100110
10010110101001
01101010100110
10010110101010

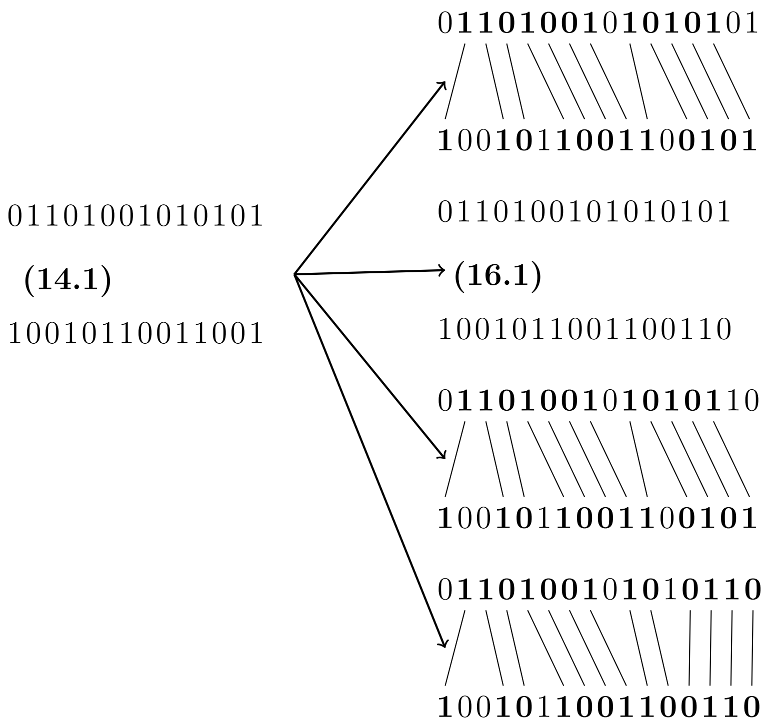
(10.8)

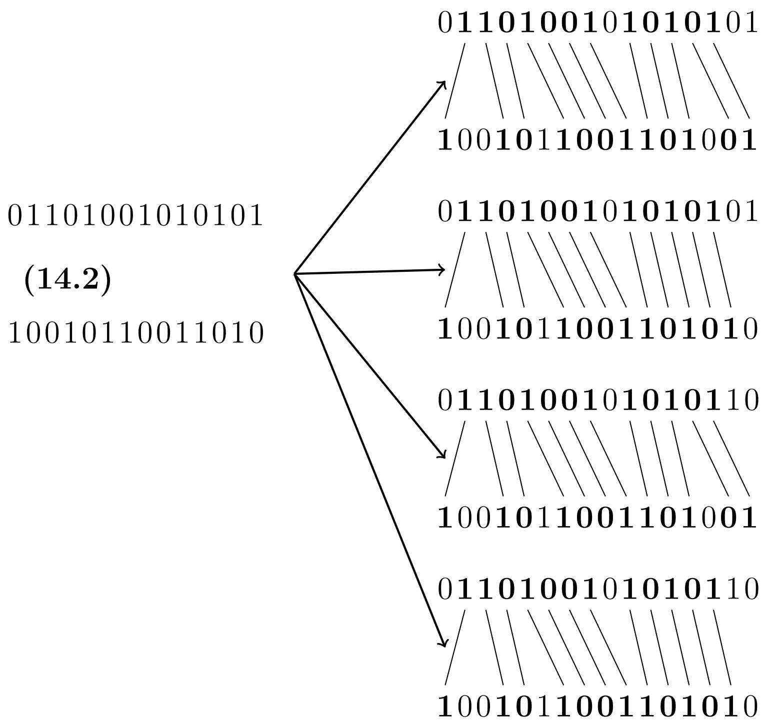

0110101010
(10.9)
1001011001
011010101001
100101100101
011010101001
100101100110
011010101010
100101100101
011010101010
100101100110

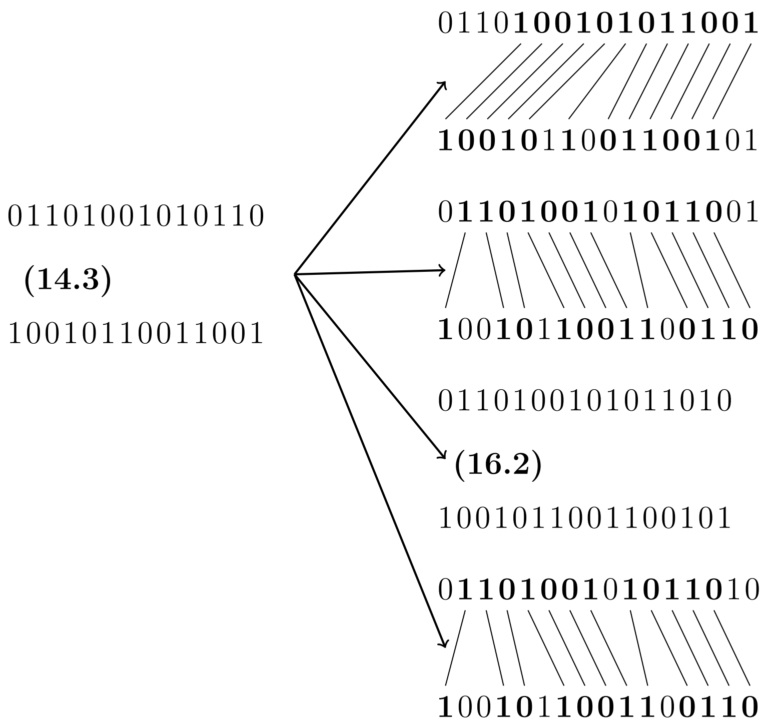

011010010101
(12.1)
100101100110
01101001010101
(14.1)
10010110011001
01101001010101
(14.2)
10010110011010
01101001010110
(14.3)
10010110011001
01101001010110
(14.4)
10010110011010

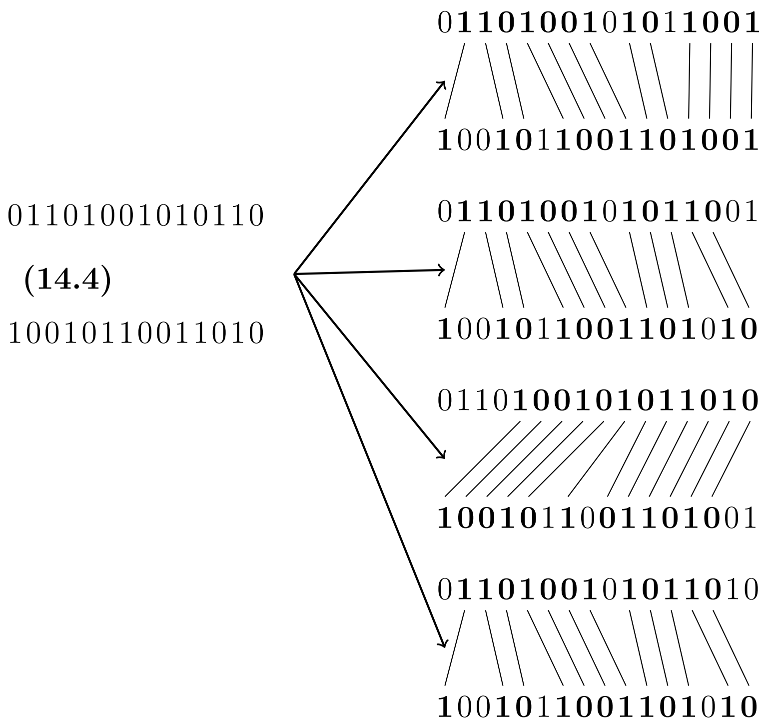

011010010101
(12.2)
100101101010
01101001010101
(14.5)
10010110101001
01101001010101
(14.6)
10010110101010
01101001010110
(14.7)
10010110101001
01101001010110
(14.8)
10010110101010

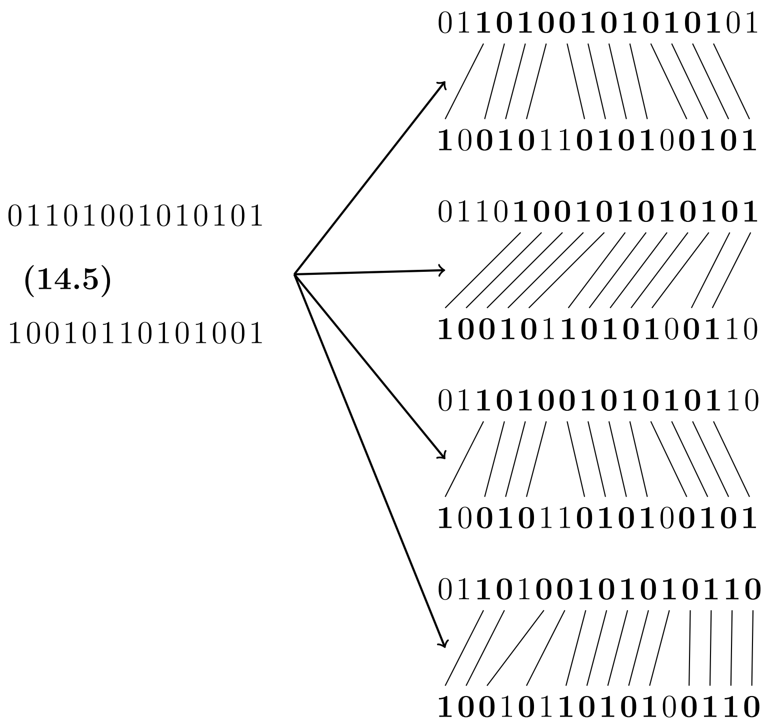

011010011010
(12.3)
100101100101
01101001101001
(14.9)
10010110010101
01101001101001
(14.10)
10010110010110
01101001101010
(14.11)
10010110010101
01101001101010
(14.12)
10010110010110

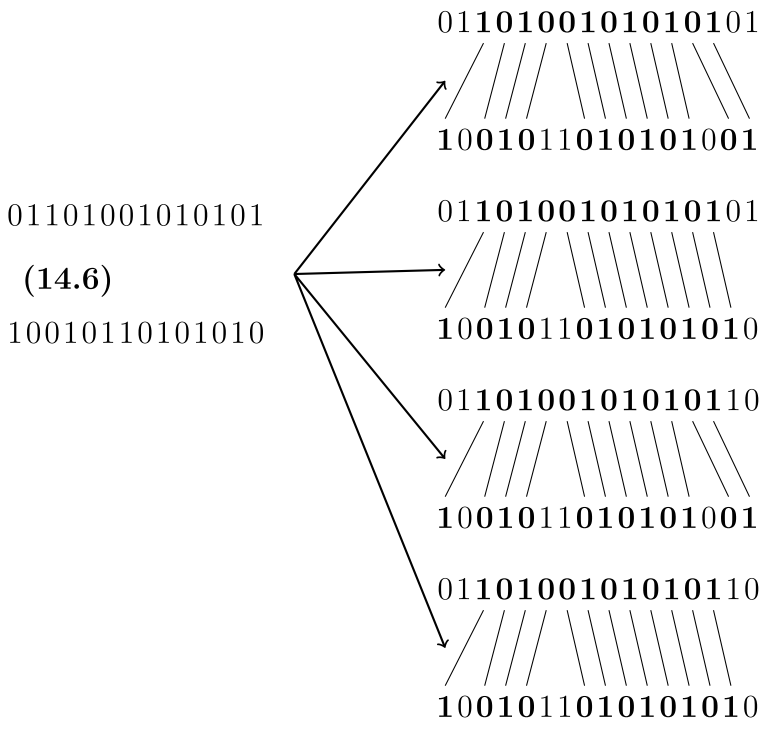

01101001010101
(14.1)
10010110011001
0110100101010101
1001011001100101
0110100101010101
(16.1)
1001011001100110
0110100101010110
1001011001100101
0110100101010110
1001011001100110

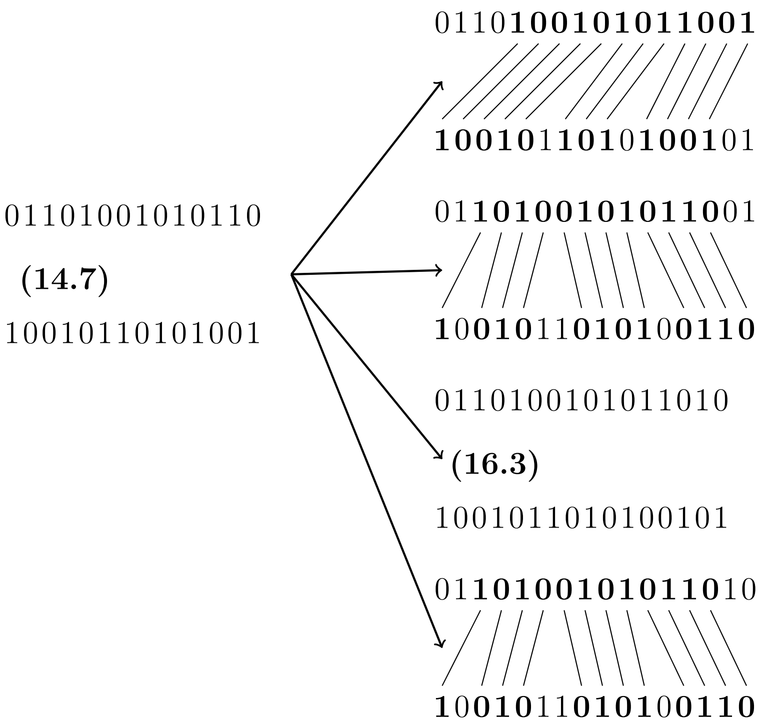

01101001010101
(14.2)
10010110011010
0110100101010101
1001011001101001
0110100101010101
1001011001101010
0110100101010110
1001011001101001
0110100101010110
1001011001101010

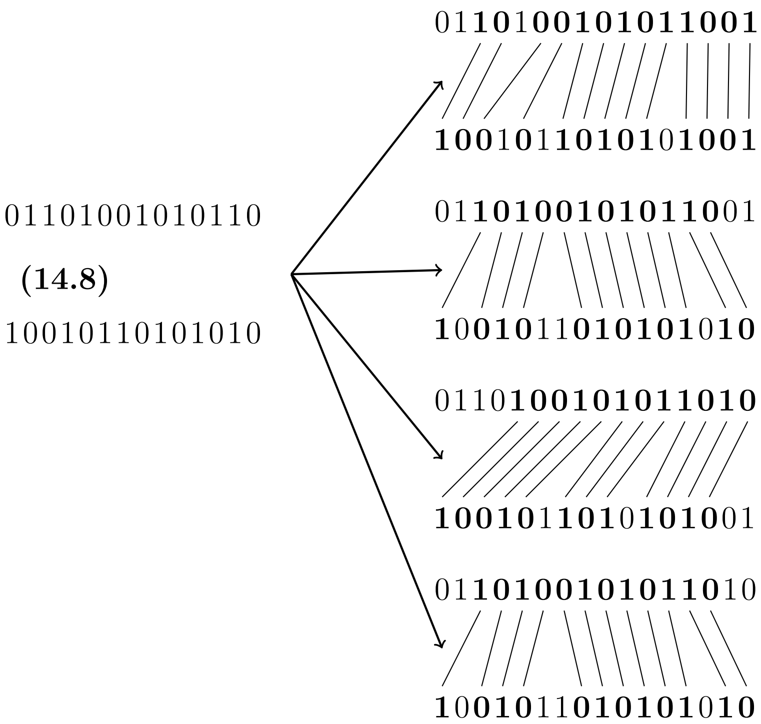

01101001010110
(14.3)
10010110011001
0110100101011001
1001011001100101
0110100101011001
1001011001100110
0110100101011010
(16.2)
1001011001100101
0110100101011010
1001011001100110

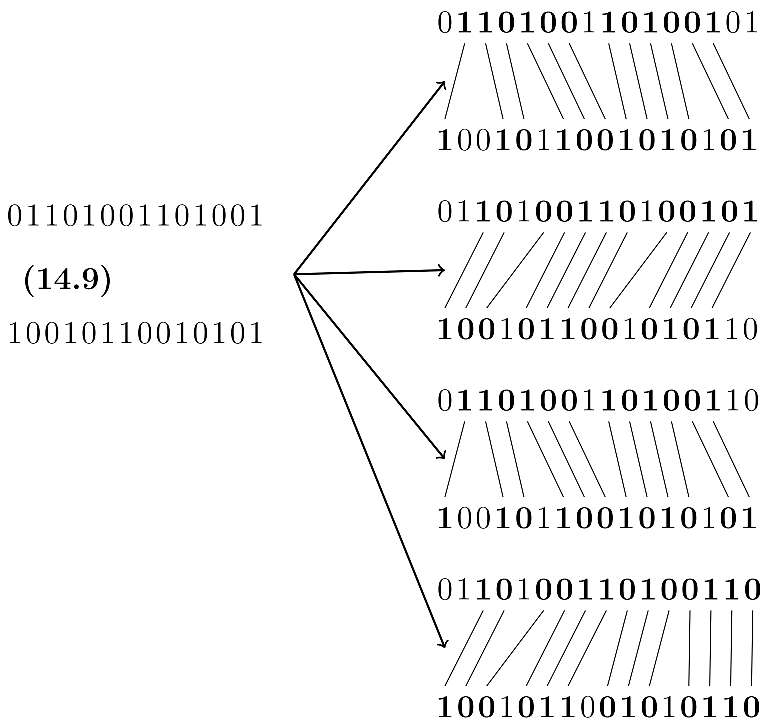

01101001010110
(14.4)
10010110011010
0110100101011001
1001011001101001
0110100101011001
1001011001101010
0110100101011010
1001011001101001
0110100101011010
1001011001101010

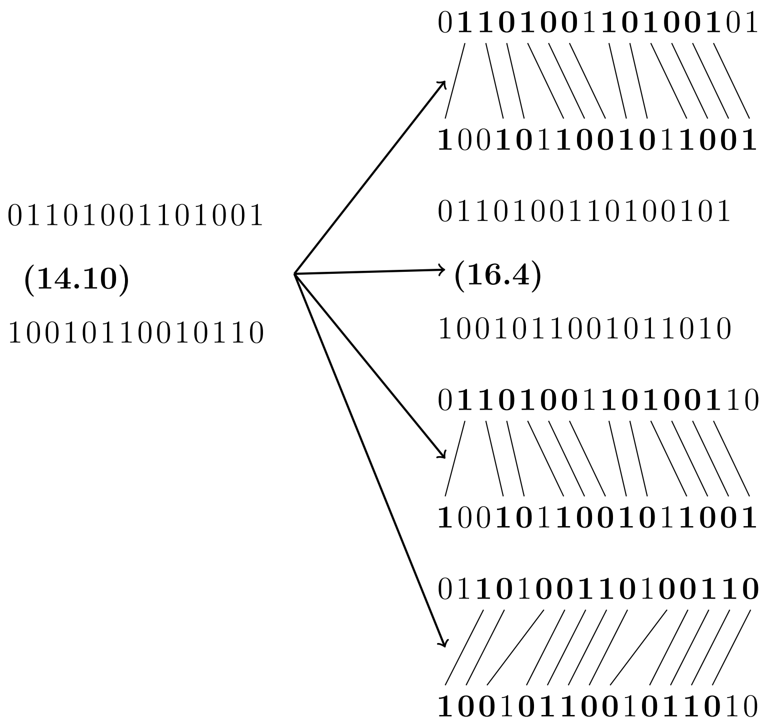

01101001010101
(14.5)
10010110101001
01101001010101
10010110101001
01101001010101
10010110100110
01101001010110
10010110101001
01101001010110
10010110100110

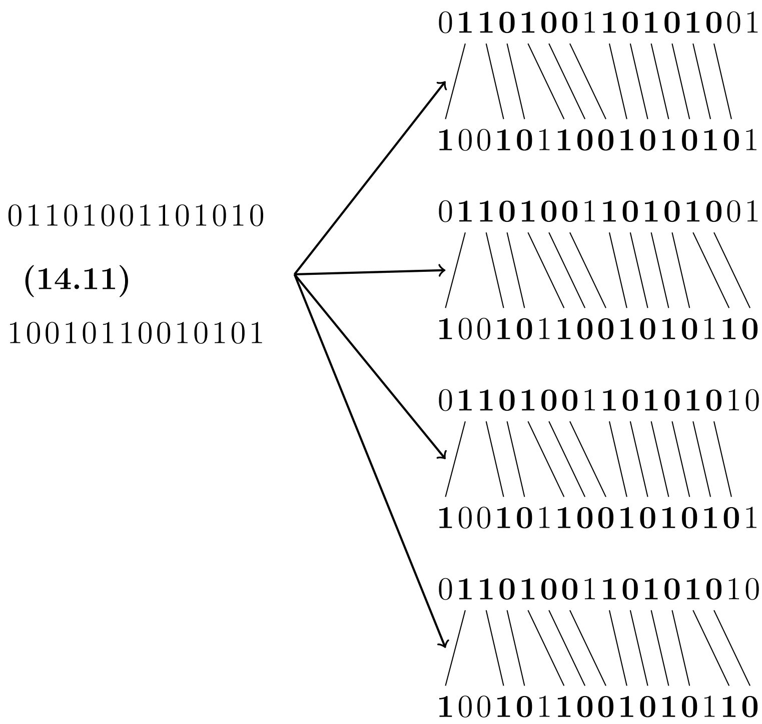

01101001010101
(14.6)
10010110101010
01101001010101
10010110101001
01101001010101
10010110101010
01101001010110
10010110101001
01101001010110
10010110101010

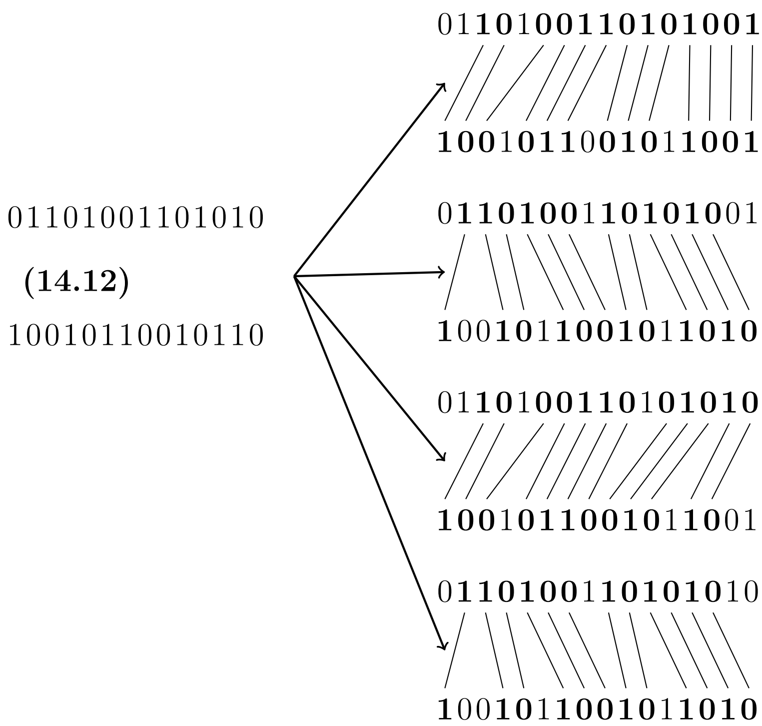

01101001010110
(14.7)
10010110101001
0110100101011001
1001011010100101
0110100101011001
1001011010100110
0110100101011010
(16.3)
1001011010100101
0110100101011010
1001011010100110

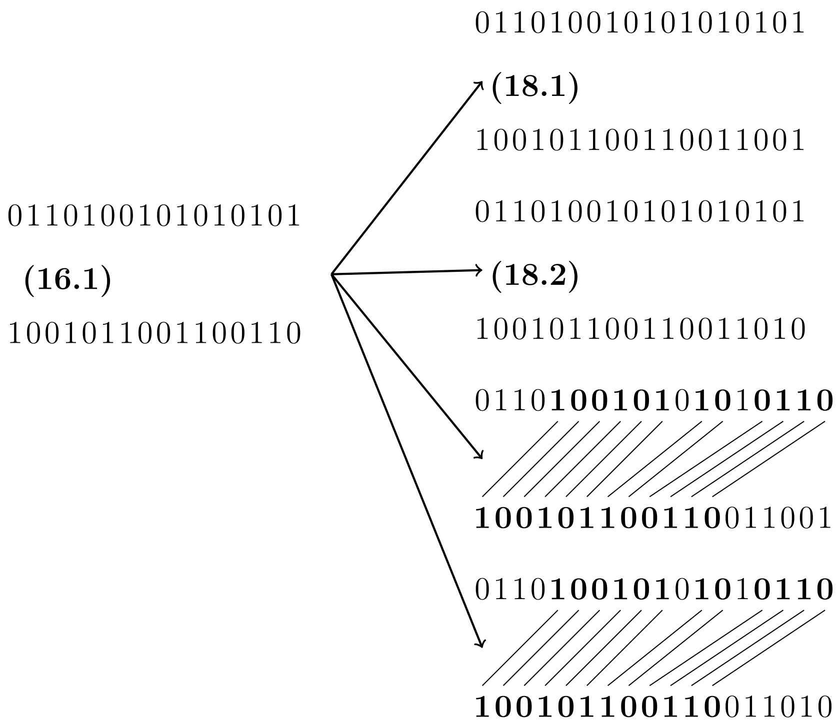

01101001010110
(14.8)
10010110101010
0110100101011001
1001011010101001
0110100101011001
1001011010101010
0110100101011010
1001011010101001
0110100101011010
1001011010101010

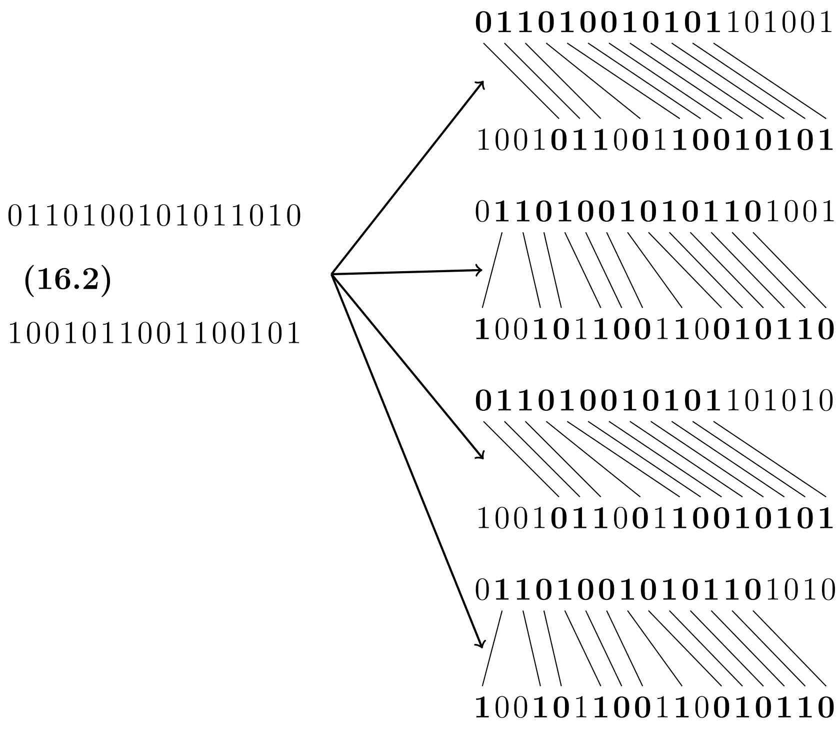

01101001101001
(14.9)
10010110010101
0110100110100101
1001011001010101
0110100110100101
1001011001010110
0110100110100110
1001011001010101
0110100110100110
1001011001010110

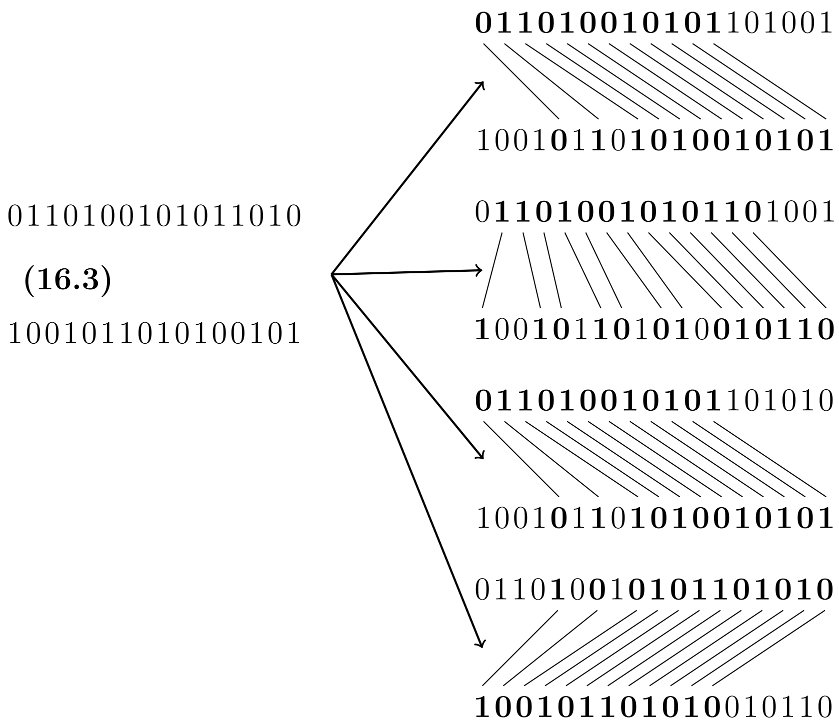

01101001101001
(14.10)
10010110010110
0110100110100101
1001011001011001
0110100110100101
(16.4)
1001011001011010
0110100110100110
1001011001011001
0110100110100110
1001011001011010


01101001101010
(14.11)
10010110010101
0110100110101001
1001011001010101
0110100110101001
1001011001010110
0110100110101010
1001011001010101
0110100110101010
1001011001010110

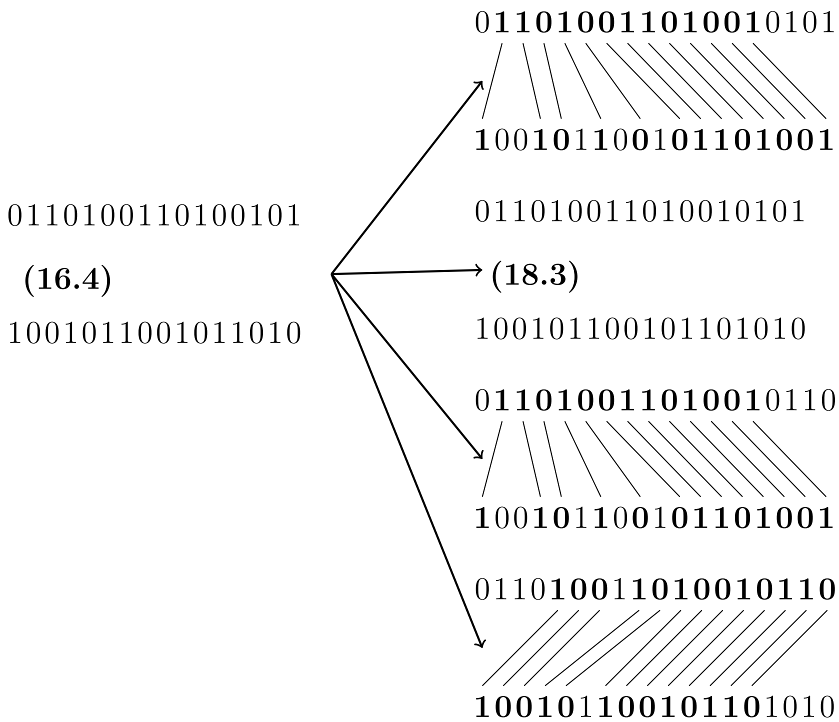

01101001101010
(14.12)
10010110010110
0110100110101001
1001011001011001
0110100110101001
1001011001011010
0110100110101010
1001011001011001
0110100110101010
1001011001011010


0110100101010101
(16.1)
1001011001100110
0110100101010101
(18.1)
1001011001100110 01
0110100101010101
(18.2)
1001011001100110 10
0110100101010110
1001011001100110011001
0110100101010110
1001011001100110011010


0110100101011010
(16.2)
1001011001100101
01101001010 1101001
1001011001100 10101
0110100101011 01001
1001011001100 10110
01101001010 1101010
1001011001100 10101
0110100101011 01010
1001011001100 10110



0110100101011010
(16.3)
1001011010100101
01101001010110 1001
1001011010010101
0110100101011010 01
100101101010010110
01101001010110 1010
1001011010010101
0110100101011010 10
1001011010 10010110



0110100110100101
(16.4)
1001011001011010
011010011010010101
100101100101101001
011010011010010101
(18.3)
100101100101101010
011010011010010110
100101100101101001
011010011010010110
100101100101101010

011010010101010101
(18.1)
100101100110011001

→ 01101001010101010101
(20.1)
100101100110011001

↘ 01101001010101010110
(20.2)
100101100110011001

011010010101010101
(18.2)
100101100110011010

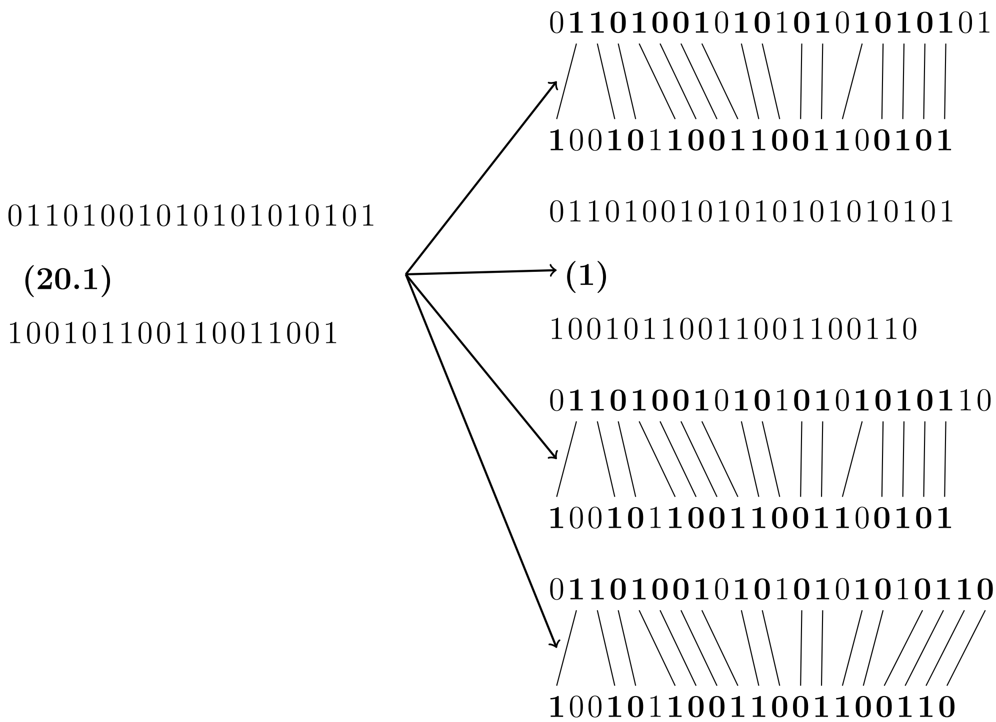
→ 01101001010101010101
(20.3)
100101100110011010

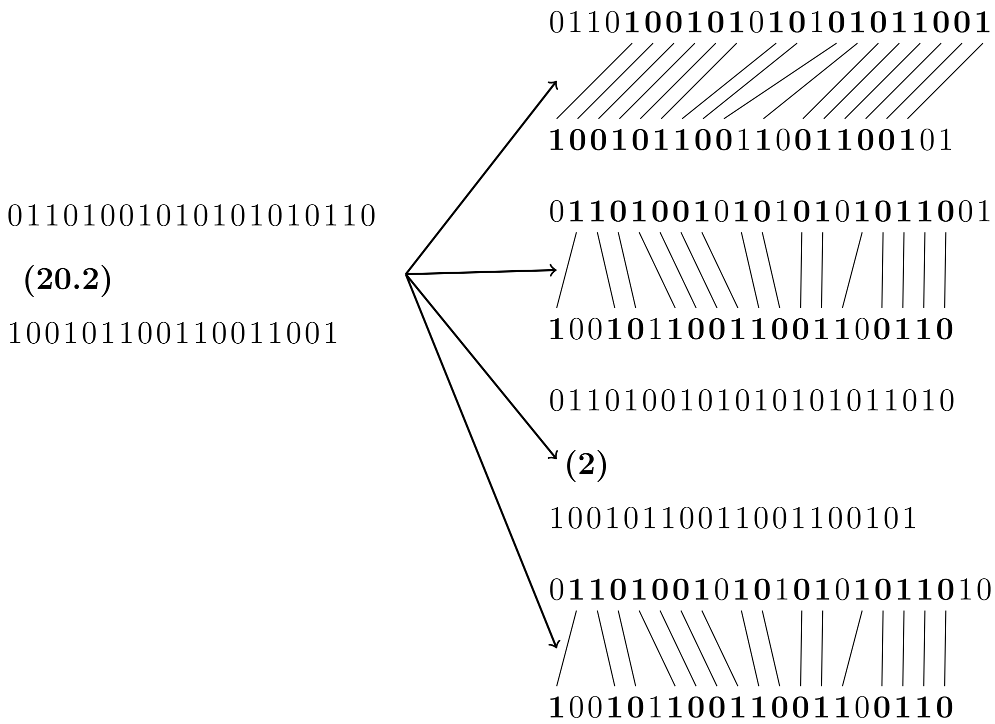
↘ 01101001010101010110
(20.4)
100101100110011010

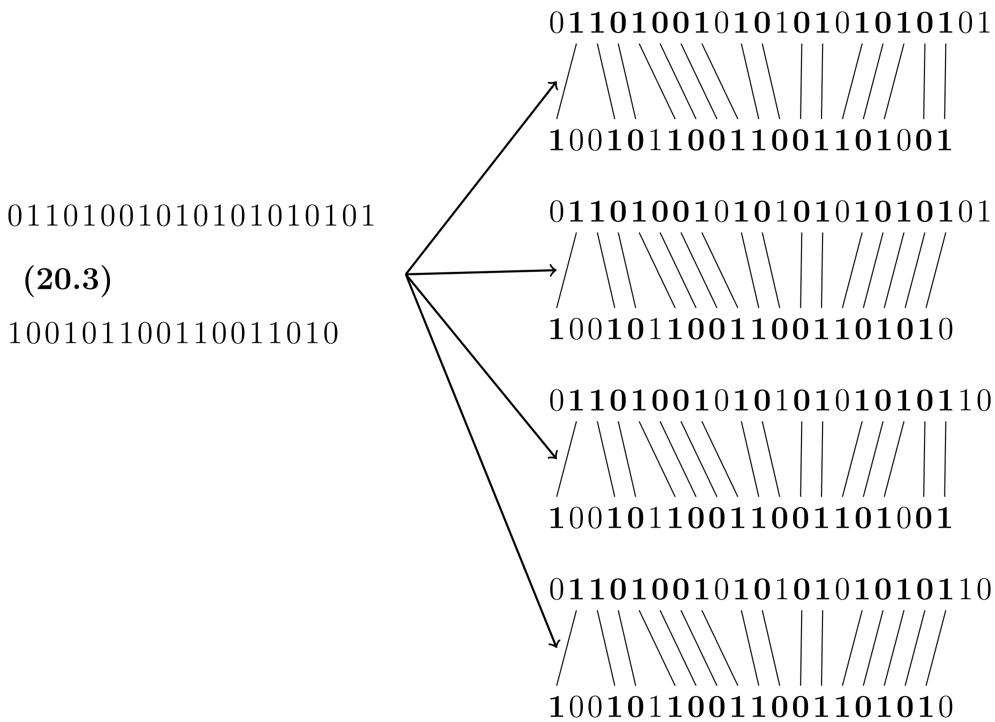
011010011010010101
(18.3)
100101100101101010

→ 01101001101001010101
(20.5)
100101100101101010

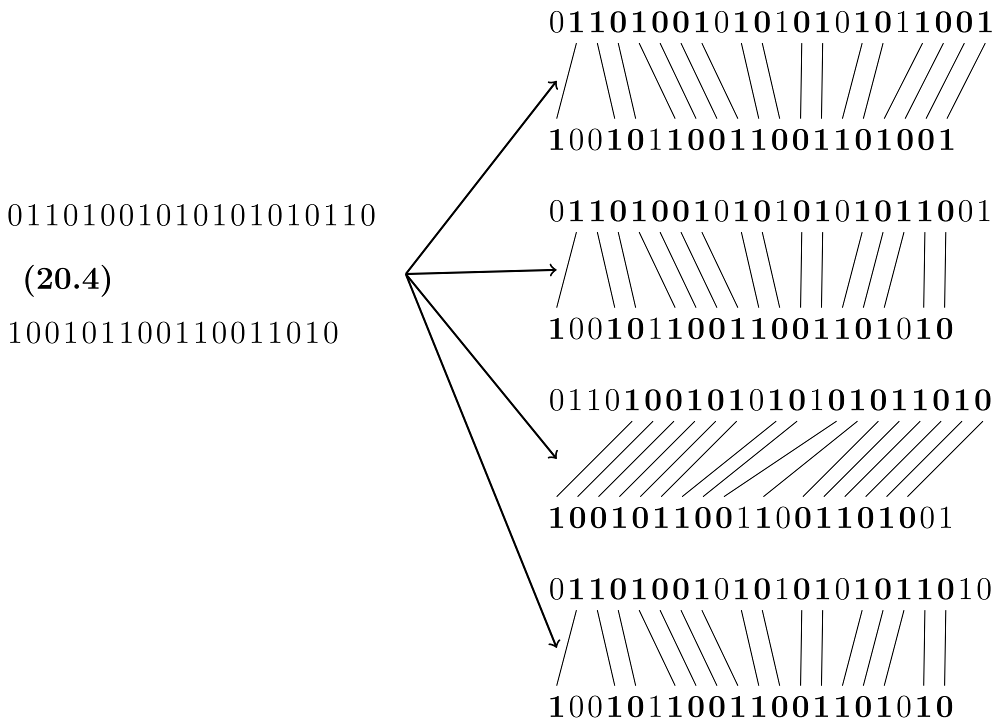
↘ 01101001101001010110
(20.6)
100101100101101010


01101001010101010101
(20.1)
100101100110011001
01101001010101010101
10010110011001100101
01101001010101010101
(1)
10010110011001100110
01101001010101010110
10010110011001100101
01101001010101010110
10010110011001100110

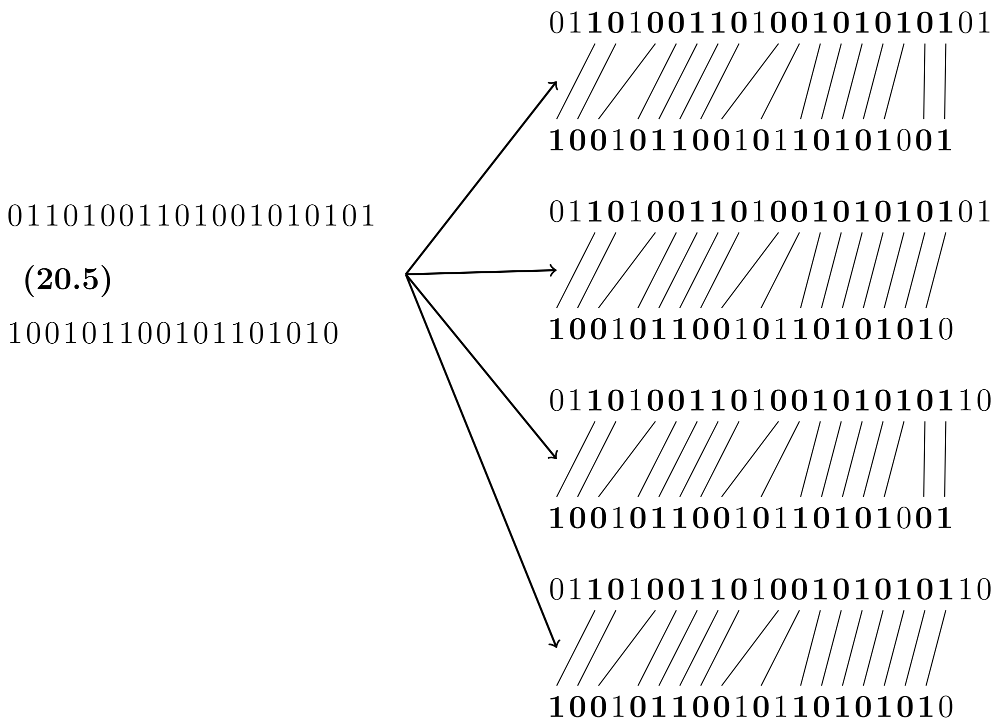

01101001010101010110
(20.2)
100101100110011001
01101001010101011001
10010110011001100101
01101001010101011001
10010110011001100110
01101001010101011010
(2)
10010110011001100101
01101001010101011010
10010110011001100110

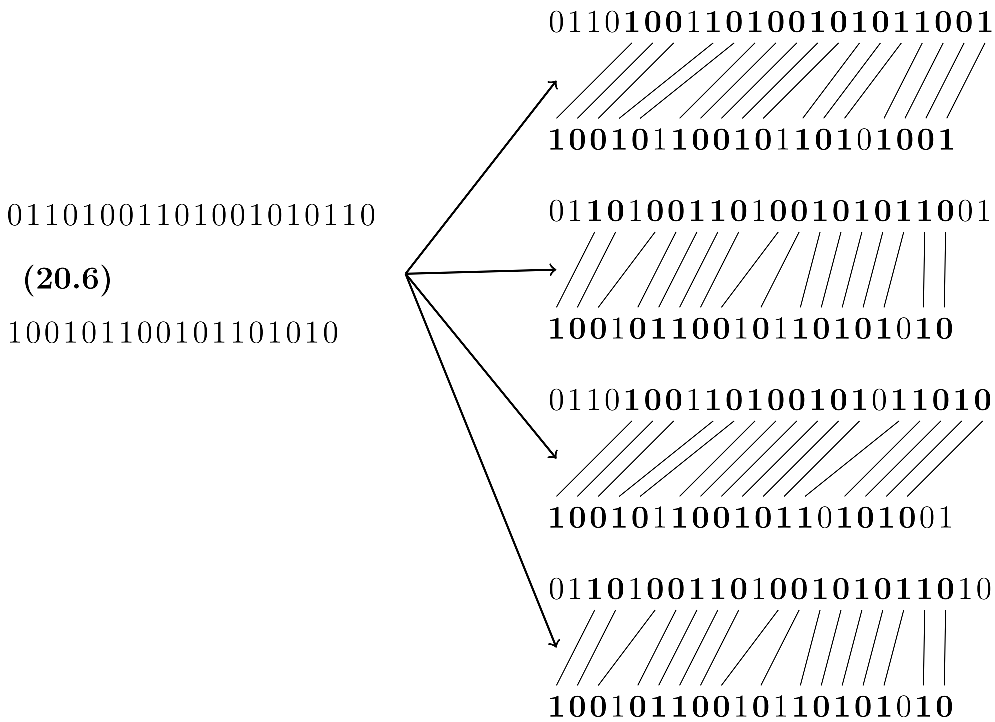

01101001010101010101
100101100110011010
0110100101010101010101
10010110011001101001
0110100101010101010101
10010110011001101010
0110100101010101010110
10010110011001101001
0110100101010101010110
10010110011001101010

(20.3)


01101001010101010110
100101100110011010
0110100101010101011001
10010110011001101001
0110100101010101011001
10010110011001101010
0110100101010101011010
10010110011001101001
0110100101010101011010
10010110011001101010

(20.4)


0110100110100101010101
10010110010110101001
0110100110100101010101
10010110010110101010
0110100110100101010110
10010110010110101001
0110100110100101010110
10010110010110101010
01101001101001010101
(20.5)
100101100101101010


0110100110100101011001
10010110010110101001
0110100110100101011001
10010110010110101010
0110100110100101011010
10010110010110101001
0110100110100101011010
10010110010110101010
01101001101001010110
(20.6)
100101100101101010