Multicolor traveling salesman problem: approximation and feasibility

Frédéric Meunier, Pauline Sarrabezolles

The Multicolor Traveling Salesman Problem: approximation and feasibility

Frédéric Meunier and Pauline Sarrabezolles

Université Paris Est, CERMICS,
77455 Marne-la-Vallée, France
{frederic.meunier,pauline.sarrabezolles}@enpc.fr

Abstract. The multicolor traveling salesman problem (MTSP) is defined on a complete graph whose vertex set is partitioned into k subsets, identified with colors. It aims to find a shortest Hamiltonian tour subject to restrictions: the number of vertices of the subtour between two consecutive vertices of the same color is bounded from above and from below.
In this work, we propose new approximation algorithms. Some special cases with two colors have already received attention: the bipartite traveling salesman problem and the black-and-white traveling salesman problem. Polynomial-time approximation algorithms are known for these problems. We cover new cases with two colors and a special case when all colors have same size. In addition, we find necessary conditions and sufficient conditions for the MTSP to have feasible solutions. Finally, we establish a connection between the balance properties of words and the existence of feasible solutions for the MTSP.

Keywords: approximation algorithms, balance properties of words, black-and-white traveling salesman problem, traveling salesman problem

1 Introduction

1.1 Problem

The Black-and-White Traveling Salesman Problem (BWTSP), introduced by Ghiani et al. [7], is defined on a complete graph whose vertex set is partitioned into black vertices and white vertices. The aim is to design a shortest Hamiltonian cycle on the graph subject to cardinality and length constraints: both the number of white vertices as well as the length of the tour between two consecutive black vertices are bounded from above. In 2010, Tresoldi et al. [12] proposed to generalize the problem by considering more than two colors. They defined the *Multicolor Traveling Salesman Problem* on a complete graph whose vertex set V is partitioned into an arbitrary number k of subsets of vertices, each of these subsets being identified with a color. The aim is again to find a shortest Hamiltonian cycle, subject to cardinality constraints: for $i = 1, \dots, k$, the number of vertices between two consecutive vertices of color i is bounded from above and from below. Note that there are no longer length constraints.

More precisely, the Multicolor Traveling Salesman Problem – or MTSP for short – requires in input the complete graph, the partition $V = V_1 \cup \cdots \cup V_k$ of its vertices, with $V_i \cap V_j = \emptyset$ for $i \neq j$, the distance function, and two integer vectors $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k) \in \mathbb{Z}_+^k$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k) \in \mathbb{Z}_+^k$. A Hamiltonian cycle is *feasible* if between two consecutive vertices of color i there are at most β_i and at least α_i vertices, for $i = 1, \dots, k$.

Multicolor Traveling Salesman Problem

Input. A complete graph $K_n = (V, E)$, an integer k, a partition $V = V_1 \cup \cdots \cup V_k$ of the vertex set, a distance function $d : E \to \mathbb{R}_+$ satisfying the triangle inequality, two integer vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}_+^k$.

Task. Decide whether a feasible Hamiltonian cycle exists, and if 'yes', find a shortest one.

Note that the BWTSP without length constraints is the MTSP with $k = 2$, $\alpha_1 = \alpha_2 = 0$ and $\beta_1 = +\infty$, where the color 1 is assumed to be white.

1.2 Complexity

The MTSP is clearly NP-hard since it contains the usual Traveling Salesman Problem as a special case. It also contains the Bipartite Traveling Salesman Problem, which is the special case with $k = 2$ and with $\beta_1 = \beta_2 = 1$ and $|V_1| = |V_2|$. The Bipartite Traveling Salesman Problem, which is also a special case of the BWTSP, is known to be NP-hard as well, see [5] for instance. However, we do not know whether the MTSP reduced to its decision version is NP-complete, even if k is fixed. In both subcases mentioned above – Traveling Salesman Problem and Bipartite Traveling Salesman Problem –, the problem is a priori known to be feasible. Note that in the decision version, the input can be reduced to the sizes $|V_i|$ of the subsets and to the vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}_+^k$, with the assumption that α_i and β_i are both smaller than $|V_i|$ for $i = 1, \dots, k$, since the distance function and the vertices are then useless. The input is then of size $O(k \log n)$, while it is of size $O(n^2 L)$ in its full optimization version with L being the logarithm of the largest distance in the graph, making the question of a fast algorithm for the decision problem much harder.

1.3 Some motivations

The BWTSP was originally motivated by optimization problems in aircraft routing arising when maintenance constraints have to be taken into account, not all airports being able to ensure maintenance operations. In addition to be a natural generalization of the BWTSP, Tresoldi et al. motivated the study of the MTSP by practical applications in the same spirit: routing where places with some features have to be visited not too often and no too seldom. For instance, a security agent has a series of buildings to visit during a night, each building must be visited a prefixed number of times and the number of buildings visited between two consecutive visits of the same building must lie in some interval.

1.4 Contribution and plan

The contributions of the present paper are threefold.

First, we propose new approximation algorithms, especially in the case when there are two colors. In this case, we use the indices W and B to denote the colors (white and black) instead of 1 and 2. A result of this type was given in 2007 by Bhattacharya et al. [4]. They proposed a polynomial-time $(4 - \frac{3}{2\beta_B})$-approximation algorithm for the BWTSP without length constraints (case when $\alpha_W = \alpha_B = 0$ and $\beta_W = +\infty$). The Bipartite Traveling Salesman Problem is 2-approximable in polynomial time [5] and corresponds to the case when $\beta_W = \beta_B = 1$ and $|V_W| = |V_B|$. In Section 2.2, among other results, we show Theorem 1 stating that there is a polynomial-time 2.5-approximation algorithm for the problem of finding a Hamiltonian cycle of minimal length in a two-colored complete graph with the constraint that each vertex has in the cycle at most one neighbor of its color. It is the two-color case of the MTSP with $\alpha_W = \alpha_B = 0$ and $\beta_W = \beta_B = 2$. Moreover, we propose a $(4.5 - 3/k)$-approximation algorithm for the problem of finding a Hamiltonian cycle of minimal length in a colored complete graph with the constraint that the cycle visits the colors in a periodic way. This case is precisely the MTSP with all V_i of same cardinality and all β_i equal to $k - 1$ (Proposition 2 in Section 2.3).

Second, we find necessary conditions and sufficient conditions for the MTSP to have a feasible solution. There is an easy necessary condition (Proposition 6 in Section 3.4), namely that

$$\alpha_i \leq \frac{\sum_{j \neq i} |V_j|}{|V_i|} \leq \beta_i \quad \text{for all } i = 1, \ldots, k. \tag{1}$$

In the case of the BWTSP, and more generally in the case of the MTSP with two colors, the condition (1) is also a sufficient one (Proposition 3 in Section 3.2). For three or more colors, things become less clear. The example with

$$k = 3, \quad |V_1| = 3, \quad |V_2| = 2, \quad |V_3| = 1$$
$$\alpha_1 = \alpha_2 = \alpha_3 = 0, \quad \beta_1 = 1, \beta_2 = 2, \beta_3 = 5$$

satisfies the necessary condition (1) but has no feasible solution, since it does not satisfy another necessary condition stated in Section 3.4 (Proposition 7).

Third, we reformulate the question of the existence of a feasible solution for the MTSP as a word problem (Section 4). Words are already used in Section 3 as a useful tool to establish necessary or sufficient conditions. In Section 4, we go further with words and show the links between the MTSP and notions like the balance property or the partial derivative for words.

All the proofs are given in Appendix.

1.5 Tools

Let $v_1, \ldots, v_n$ be n distinct vertices of a graph $G = (V, E)$ with $v_i v_{i+1} \in E$ for $i = 1, \ldots, n$. We use the convention $v_{n+1} = v_1$. Then we denote by $(v_1, \ldots, v_n)$ the Hamiltonian cycle with edge set $\{v_i v_{i+1} : i = 1, \ldots, n\}$.

2 Approximation algorithms

2.1 Preliminaries

Christofidès' heuristics [6] is used for each of the approximation algorithms proposed in our paper. It is a well-known polynomial-time approximation algorithm for the Traveling Salesman Problem. The approximation factor is 1.5 in its best version. We use this latter without further mention.

2.2 Case with two colors

As mentioned in the Section 1.4, the two colors are assumed to be black and white, denoted respectively B and W. As already noted, the case with two colors, $\alpha_W = \alpha_B = 0$, and $\beta_W = +\infty$ has been proved to be $(4 - \frac{3}{2\beta_B})$-approximable in polynomial-time by Bhattacharya et al. [4]. The case $\beta_W = \beta_B = 1$ and $|V_W| = |V_B|$ is the Bipartite Traveling Salesman Problem and is 2-approximable in polynomial time. It is a result by

Chalasani and Motwani [5]. Before, a 2.5-approximation algorithm was proposed by Anily and Hassin [2]. We prove new approximation results.

The first approximation algorithm deals with the case $\alpha_B = \alpha_W = 0$ and $\beta_B = \beta_W = 2$ (Theorem 1). It corresponds to the problem of finding a minimal length Hamiltonian cycle in a two-colored complete graph such that each vertex has in the cycle at most one neighbor of its color.

The algorithm uses the construction of a map $f : E \to \{0, 1, 2\}$ such that $\sum_{e \in \delta(v)} f(e) = 2$ for all $v \in V$ and satisfying an additional property we explain now. Note that the edges e with $f(e) \neq 0$ form a collection $\mathcal{C}_f$ of cycles and edges, all being pairwise vertex-disjoint. We require that each cycle in $\mathcal{C}_f$ satisfies the same constraint as the Hamiltonian cycle we look for: each cycle has no subpath of three or more vertices of the same color. We also require that each edge in $\mathcal{C}_f$ has one of its endpoint in black and the other in white. The set of all maps satisfying those constraints is denoted $\mathcal{F}$.

Lemma 1. *The problem* $\min_{f \in \mathcal{F}} \sum_{e \in E} d(e)f(e)$ *can be solved in polynomial time.*

Using Lemma 1, we are able to prove the following theorem, see Appendix.

Theorem 1. *In the two-color case of the MTSP, if* $\alpha_B = \alpha_W = 0$*, and* $\beta_B = \beta_W = 2$*, there is a polynomial-time* 2.5*-approximation algorithm.*

Another approximation result is the following one.

Proposition 1. *In the two-color case of the MTSP, if* $\beta_W = 1$ *or* $\alpha_B > 0$*, there is a polynomial-time* $\left(2\left\lceil \frac{\beta_B}{2} \right\rceil + 1.5\right)$*-approximation algorithm.*

Theorem 1, Proposition 1, and the results of the literature cover all cases with $\alpha_B, \alpha_W, \beta_B, \beta_W$ simultaneously in $\{0, 1, 2\}$. Indeed, suppose without loss of generality that $|V_B| \leq |V_W|$. There is no feasible solution if $\alpha_W = 2$ or $\beta_B = 0$. The case with $\alpha_W = 1$ or $\beta_B = 1$ is the Bipartite TSP. The case with $\beta_W = 0$ is the usual TSP. The remaining cases are contained in Theorem 1 or Proposition 1.

2.3 Case with three colors and more

When there are strictly more than two colors, i.e. $k > 2$, things seem to be more difficult. The following proposition deals with a special case of the MTSP that generalizes the Bipartite Traveling Salesman Problem.

Proposition 2. *There is a polynomial-time* $(4.5 - \frac{3}{k})$*-approximation algorithm for the MTSP when* $|V_1| = \cdots = |V_k|$ *and* $\beta_i = k - 1$ *for* $i = 1, \ldots, k$.

Note that the feasible solutions in this case are exactly the Hamiltonian cycles visiting the V_i in a periodic way. The case $k = 2$ is the Bipartite Traveling Salesman Problem.

When $k = 3$, it is possible to improve slightly the approximation ratio to $(3 - 1/6)$. When $k = 2$, a similar argument gives a ratio equal to 2.5, which is Anily and Hassin's result [2].

3 Feasibility and circular words

In this section, we provide necessary conditions and sufficient conditions for an instance of the MTSP to have a feasible solution. Except when there are two colors, i.e. $k = 2$, finding compact conditions for the existence of solutions is still an open question. We believe this question to be difficult, since it resembles to difficult questions in word theory, see Section 4.

3.1 Reformulation with words

As noted earlier, the existence of a feasible Hamiltonian cycle only depends on the number of colors k, the number of vertices of each color $|V_i|$, and the two integer vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$. This leads us to consider a new formulation of the decision problem as a problem of words. This point of view turns out to be useful in the proofs of existence results.

We remind basic notions from word theory, see [9] for more background.

An *alphabet* Σ is a nonempty finite set, the elements of which are called *letters*. *Words* are finite sequences of letters. The *length* $|w|$ of a word w is the number of its letters, and $|w|_a$ denotes the number of occurrences of a letter a in w. A word x is a *factor* of w if there exist two words u and v such that $w = uxv$. A *a-factor* is a factor of w not containing the letter a, it is said to be *maximal* if it is maximal for the order relation "being a factor of". Any subsequence of w is called a *subword* of w.

Two words w and w' are *conjugate* if there exists two words x and y such that $w = xy$ and $w' = yx$. Conjugation is an equivalence relation. We identify conjugacy classes and *circular words*, which are finite circular sequences of letters. Given a word w, its conjugacy class is denoted (w). The length of a circular word (w) is the length of w. A *factor* (resp.

a-factor) of a circular word (w) is a factor (resp. a-factor) of some conjugate of w. Equivalently, a factor (resp. a-factor) of (w) is a factor (resp. a-factor) of ww of length not greater than $|w|$.

Interpreting the set of colors as an alphabet $\{1,\ldots,k\}$, the circular sequence of colors visited by a Hamiltonian tour $T=(v_1,\ldots,v_n)$ is a circular word (w). Formally, the circular word (w) is such that $w=i_1\cdots i_n$ and $v_j\in V_{i_j}$ for $j=1,\ldots,n$.

Note that in general there are two possible circular words induced by T since $(i_1\cdots i_k)$ is in general distinct from $(i_k\cdots i_1)$. We arbitrarily choose one of them. If T is feasible, the number of vertices between two consecutive vertices of color i is at most β_i and at least α_i, for $i=1,\ldots,k$. It translates on (w) in the following property: any maximal i-factor of (w) contains at least α_i and at most β_i letters, for $i=1,\ldots,k$.

3.2 Case with two colors

When there are only two colors, there is a necessary and sufficient condition. As already mentioned, the two colors are assumed to be black and white, denoted respectively B and W.

Proposition 3. *In the two-color case, the MTSP has a feasible solution if and only if*

$$\alpha_W \le \frac{|V_B|}{|V_W|} \le \beta_W \quad \textit{and} \quad \alpha_B \le \frac{|V_W|}{|V_B|} \le \beta_B.$$

3.3 Case with three colors

Contrary to what happens with two colors, the necessary condition (1), given in Section 1.4, is no longer sufficient when there are three colors, as already noted in the introduction. However, with an additional condition, condition (1) turns out to be sufficient.

Proposition 4. *Assume that $k=3$ and that two of the V_i's have same cardinality. Then the MTSP has a feasible solution if and only if*

$$\alpha_i \le \frac{\sum_{j\neq i}|V_j|}{|V_i|} \le \beta_i \qquad \textit{for all } i\in\{1,2,3\}.$$

The fact that it is a sufficient condition in this case is a consequence of a result by Altman et al. [1] for balanced words, see Section 4.

We have another sufficient condition, dealing with the case when there is a color more represented than the two others together.

Proposition 5. *Assume $k = 3$, $|V_1| \geq |V_2| + |V_3|$, $|V_2| \geq |V_3|$, and $\alpha_1 = \alpha_2 = \alpha_3 = 0$. If*

$$\beta_1 \geq 1, \quad \beta_2 \geq 2\left(\left\lceil \frac{|V_1|}{|V_2| + |V_3|} \right\rceil\right) + 1,$$

$$\text{and} \quad \beta_3 \geq \left(\left\lceil \frac{|V_2|}{|V_3|} \right\rceil + 1\right)\left(\left\lceil \frac{|V_1|}{|V_2| + |V_3|} \right\rceil + 1\right) - 1,$$

then the MTSP has a feasible solution.

3.4 General case

We start with two necessary conditions. The first one is condition (1) of Section 1.4. The second one allows to prove that the example of Section 1.4 is not feasible. Note that the cardinalities of the V_i are not involved in this latter.

Proposition 6. *If the MTSP has a feasible solution, then*

$$\alpha_i \leq \frac{\sum_{j \neq i} |V_j|}{|V_i|} \leq \beta_i \quad \text{for all } i = 1, \ldots, k.$$

Proposition 7. *If the MTSP has a feasible solution, then*

$$\sum_{\ell=1}^{j} \beta_{i_\ell} \geq j^2 \qquad \text{for all } j = 1, \ldots, k-1 \text{ and all } 1 \leq i_1 < \cdots < i_j < k.$$

The two following propositions give sufficient conditions for the existence of solutions. The proof of Proposition 8 uses a circular version of the so-called "billiard words" [3] and is actually a straightforward generalization of the proof of Proposition 3.

Proposition 8. *If we have*

$$\alpha_i \leq \sum_j \left\lfloor \frac{|V_j|}{|V_i|} \right\rfloor \quad \text{and} \quad \beta_i \geq \sum_j \left\lceil \frac{|V_j|}{|V_i|} \right\rceil \quad \text{for all } i = 1, \ldots, k,$$

then the MTSP has a feasible solution.

Proposition 9. *Suppose that $|V_1| \geq \cdots \geq |V_k|$ and that $|V_i|$ divides $\sum_{j=1}^{i-1} |V_j|$ for $i = 2, \ldots, k$. If we have*

$$\alpha_i \leq \frac{\sum_{j=1}^{i-1} |V_j|}{|V_i|} \quad \text{and} \quad \beta_i \geq \frac{\sum_{j=1}^{i-1} |V_j|}{|V_i|} + k - i \quad \text{for all } i = 1, \ldots, k,$$

then the MTSP has a feasible solution.

4 Words, partial derivative, and balance properties

In this section the connection between the question of existence of solutions for the MTSP and the word theory is detailed. There is a notion in word theory that fits perfectly for the MTSP: the partial derivatives. This notion provides a straightforward formulation of the decision problem as a problem of 'integration on words'.

4.1 Reformulation as an integration problem

The *partial derivative* $\partial_a(w)$ of a circular word (w) with respect to a letter a is a circular word over the alphabet $\mathbb{Z}_+$ of length $|w|_a$. For a circular word $(w) = (af_1a \cdots af_r)$, where the f_i's are maximal a-factors, we have $\partial_a(w) = (|f_1| \cdots |f_r|)$. We use the notation Σ^* to refer to the set of all words over an alphabet Σ.

The problem of deciding whether there is a feasible solution for the MTSP is reformulated as a problem of 'integration on words' as follows. Given an integer k, a set of integers $(n_1, \ldots, n_k)$, and two integer vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}_+^k$, decide whether there exists a circular word (w) over the alphabet $\{1, \ldots, k\}$ satisfying

$$|\partial_i(w)| = n_i \quad \text{and} \quad \partial_i(w) \in \{\alpha_i, \ldots, \beta_i\}^* \text{ for } i = 1, \ldots, k.$$

4.2 Links with the balance properties

Similarities between the decision problem of the MTSP and the *balance property problem* are now emphasized.

For a non-negative integer m and a circular word (w) over an alphabet Σ, a letter $a \in \Sigma$ is *m-balanced* in (w) if any pair (f, f') of factors of w such that each of f and f' is preceded and followed immediately by a letter a in (w) and such that $|f|_a = |f'|_a$ satisfies $||f| - |f'|| \leq m$. A word is m-balanced on Σ, if for all $a \in \Sigma$, the letter a is m-balanced. This notion was introduced by Sano et al. [10] and used in [8] with the terminology *m-uniform distribution*. When $m = 1$, we use *balanced* instead of 1-balanced.

Given an alphabet Σ, a family of integers $(n_a)_{a \in \Sigma}$, and an integer m, the *balance property problem* aims to decide whether there exists an m-balanced word (w) with $|w|_a = n_a$ for $a \in \Sigma$.

Mantaci et al. [8] gave the following characterization of m-balanced words. The jth partial derivatives of a circular word (w) with respect to a letter a is defined by

$$\partial_a^j(w) = \Big((\sum_{\ell=1}^{j} |f_\ell|)(\sum_{\ell=2}^{j+1} |f_\ell|)\cdots(\sum_{\ell=r}^{j+r-1} |f_\ell|) \Big),$$

with the convention $f_{r+i} = f_i$. Note that the jth partial derivative with respect to a is a circular word of length $|w|_a$ and it is not the derivative of the $(j-1)$th partial derivative.

Proposition 10. *A letter a is m-balanced in the word (w) if and only if for all $j \in \mathbb{Z}_+$, there exists an integer d_a^j such that $\partial_a^j(w) \in \{d_a^j, \ldots, d_a^j + m\}^*$.*

An m-balanced word (w) over the alphabet $\{1, \ldots, k\}$ induces a feasible solution for the MTSP with k colors, with $|V_i| = |w|_i$, and with $\alpha_i \leq d_i^1$ and $\beta_i \geq d_i^1 + m$, where d_i^1 is the integer of Proposition 10. However, the existence of m-balanced words does not insure the existence of solutions for the MTSP with the $|V_i|$'s and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ given a priori. We can nevertheless deduce from Proposition 10 the following proposition, see Appendix.

Proposition 11. *Assume that no $|V_i|$ divides $\sum_{j \neq i} |V_j|$ for $i = 1, \ldots, k$. A balanced circular word (w) over $\{1, \ldots, k\}$ with $|w|_i = |V_i|$ for all $i = 1, \ldots, k$ induces a feasible solution for the MTSP for any $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ such that*

$$\alpha_i \leq \frac{\sum_{j \neq i} |V_j|}{|V_i|} \leq \beta_i \quad \text{for all } i = 1, \ldots, k.$$

In other words, if the condition of Proposition 11 is satisfied and if there exists a balanced circular word (w) over $\{1, \ldots, k\}$ with $|w|_i = |V_i|$ for all $i = 1, \ldots, k$, the condition (1) is necessary and sufficient.

Given the n_i's, deciding whether there is a balanced word is difficult in general. This problem has been solved for a size of alphabet equal to two or three, and partially for a size equal to four. For a two-letter alphabet there exists a balanced word for any choice of n_a's. The balance property problem over a three-letter alphabet has been studied by Altman et al. [1].

Theorem 2. *Given $n_1, n_2, n_3 \in \mathbb{Z}_+$, there exists a balanced circular word (w) over the alphabet $\{1, 2, 3\}$ with $|w|_i = n_i$ for $i = 1, 2, 3$ if and only if*

$$\left(\frac{n_1}{n_1 + n_2 + n_3}, \frac{n_2}{n_1 + n_2 + n_3}, \frac{n_3}{n_1 + n_2 + n_3} \right) = \left(\frac{4}{7}, \frac{2}{7}, \frac{1}{7} \right)$$

or two n_i's are equal.

Applying this theorem with Proposition 11, we get the following corollary.

Corollary 1. *Assume that no* $|V_i|$ *divides* $\sum_{j \neq i} |V_j|$ *for* $i = 1, 2, 3$. *If* $(\frac{|V_1|}{|V|}, \frac{|V_2|}{|V|}, \frac{|V_3|}{|V|}) = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$ *or two of the* $|V_i|$*'s are equal, then there exists a feasible solution for the MTSP for any* $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ *such that*

$$\alpha_i \leq \frac{\sum_{j \neq i} |V_j|}{|V_i|} \leq \beta_i \quad \textit{for all } i = 1, 2, 3.$$

References

1. Altman, E., Gaujal, B., Hordijk, A.: Balanced Sequences and Optimal Routing. J. ACM. 47, 752–775 (1999)
2. Anily, S., Hassin, R.: The swapping problem. Networks. 22, 419–433 (1992)
3. Arnoux, P., Mauduit, C., Shiokawa, I., Tamura, J.: Complexity of sequences defined by billiards in the cube. Bull. Soc. Math. France. 122, 1-12 (1994)
4. Bhattacharya, B., Hu, Y., Kononov, A.: Approximation Algorithms for the Black and White Traveling Salesman Problem. In: COCOON 2007 , pp. 559–567. Banff, Canada (2007)
5. Chalasani, P., Motwani, R.: Approximating capacited routing and delivery problem. SIAM Journal on Computing. 28, 2133–2149 (1999)
6. Christofides, N.: Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem. Management sciences research report. Defense Technical Information Center. (1976)
7. Ghiani, G., Laporte, G., Semet, F.: The Black and White Traveling Salesman Problem. Operations Research. 54, 366–378 (2006)
8. Mantaci, R., Mantaci, S., Restivo, A.: Balance Properties and Distribution of Squares in Circular Words. Developments in Language Theory. 504–515 (2008)
9. Perrin, D.: Combinatorics on Words, Cambridge Mathematical Library. Cambridge University Press (2007)
10. Sano, S., Miyoshi, N., Kataoka, R.: m-Balanced words: A generalization of balanced words. Theoretical Computer Science. 314, 12, 97–120 (2004)
11. Schrijver, A.: Combinatorial Optimization, Volume A. Springer, Verlag Berlin Heidelberg (2003)
12. Tresoldi, E., Wolfler Calvo, R., Borne, S.: Solving the Multicolor TSP. In: 24th European Conference on Operational Research (EURO XXIV). Lisbon, Portugal (2010)

Appendix

Proof (of Lemma 1). We build a new graph G, containing the original graph $K_n = (V, E)$ as an induced subgraph, as follows. We make a copy V'_W of V_W and a copy V'_B of V_B. We keep the original edges. We add all edges uv with $u \in V_W \cup V'_W$ and $v \in V'_B$ and all edges uv with $u \in V'_W$ and $v \in V_B \cup V'_B$. The edges of G are weighted: the weight $w(uv)$ of an edge uv of G is $d(\bar{u}\bar{v})$, where $\bar{u} \in V_W$ is u itself or the vertex u is the copy of, depending on whether $u \in V_W$ or $u \in V'_W$, and where $\bar{v} \in V_B$ is v itself or the vertex v is the copy of, depending on whether $v \in V_B$ or $v \in V'_B$.

Now, we show that a f in $\mathcal{F}$ provides a perfect matching M in G such that $w(M) = \sum_{e \in E} d(e)f(e)$, and conversely that any perfect matching M in G provides a f in $\mathcal{F}$ such that $\sum_{e \in E} d(e)f(e) = w(M)$. Once this equivalence has been shown, the conclusion follows, since a perfect matching of minimal weight can be computed in polynomial time.

Let f be a map in $\mathcal{F}$. We deal first with the cycles of $\mathcal{C}_f$ and then its edges.

Let C be a cycle in $\mathcal{C}_f$. We choose an orientation for C and we make a round trip starting from an arbitrary vertex. Each time a black-black arc (b, b') is encountered, with $b, b' \in V_B$, we select the edge bb' with $b, b' \in V_B$ in G. The same holds for the white-white edges: each time a white-white arc (w, w') is encountered, with $w, w' \in V_W$, we select the edge ww' with $w, w' \in V_W$ in G.

Each time a black-white arc (b, w) is encountered, two possibilities. If the previous arc is a black-black one, we select in G the edge $b'w'$ where $b' \in V'_B$ is the copy of b and $w' \in V'_W$ is the copy of w. If the previous arc is a white-black one, we select in G the edge bw' where $w' \in V'_W$ is the copy of w.

Each time a white-black arc (w, b) is encountered, two possibilities. If the previous arc is a white-white one, we select in G the edge $w'b'$ where $w' \in V'_W$ is the copy of w and $b' \in V'_B$ is the copy of b. If the previous edge is a black-white one, we select in G the edge wb' where $b' \in V'_B$ is the copy of b.

Let bw be an edge in $\mathcal{C}_f$. We select in G the edge bw and the edge $b'w'$, with $b \in V_B$, $b' \in V'_B$, $w \in V_W$, and $w' \in V'_W$.

The set of all edges selected in G by this process is a perfect matching whose weight is the cost of f.

Conversely, any perfect matching M in G gives a map f in $\mathcal{F}$ such that $\sum_{e\in E} d(e)f(e) = w(M)$: for $uv \in E$, we define

$$f(uv) = \Big|\{uv, u'v, uv', u'v' : \text{with } u' \text{ being the copy of } u \text{ and } v' \text{ being the copy of } v\} \cap M\Big|.$$

□

Proof (of Theorem 1). The algorithm proceeds in three steps.

During the first step we compute a map $f^* \in \mathcal{F}$ minimizing $\sum_{e\in E} d(e)f(e)$ over all $f \in \mathcal{F}$. This is done with the polynomial algorithm of Lemma 1.

The second step consists in picking a black vertex in each element – cycle or edge – of $\mathcal{C}_{f^*}$. According to the property required for $\mathcal{C}_{f^*}$, we know that such vertices exist. We compute then a tour $T_B = (b_1, \ldots, b_r)$ on these black vertices using Christofidès' heuristics. The element of $\mathcal{C}_{f^*}$ the vertex b_i belongs to is denoted C_i. Note that each b_i has at least one neighbor w_i of color white on C_i. If C_i is not a simple edge, we denote u_i the neighbor of b_i on C_i that is distinct from w_i; otherwise, u_i is defined as being w_i.

In the third step, we build the Hamiltonian cycle T we look for: take the union of all elements C_i of $\mathcal{C}_{f^*}$; delete from this union all edges u_ib_i for the C_i being cycles, leading to a collection of r open paths; finally, add to this collection the edges u_ib_{i+1} to get the tour T, with the convention $b_{r+1} = b_1$. Note that the tour T_B is used to determine the cyclic order on the black vertices picked from each component of $\mathcal{C}_{f^*}$.

Figure 1 depicts the construction of the approximate solution.

It remains to show that the length of T is less than 2.5 times the best achievable length. We denote T^* an optimal solution. We have

$$\begin{aligned} d(T) &= \sum_{i=1}^{r} d(C_i) - \sum_{i:\ C_i \text{ is a cycle}} d(u_ib_i) + \sum_{i=1}^{r} d(u_ib_{i+1}) \\ &\leq \sum_{i=1}^{r} d(C_i) - \sum_{i:\ C_i \text{ is a cycle}} d(u_ib_i) + \sum_{i=1}^{r} \Big(d(u_ib_i) + d(b_ib_{i+1})\Big) \\ &= \sum_{e\in E} d(e)f^*(e) + d(T_B) \end{aligned}$$

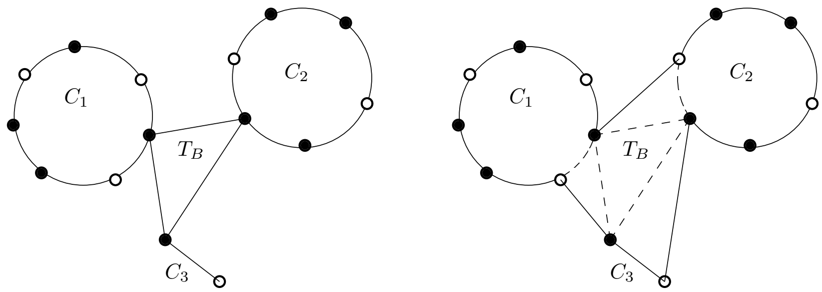

Fig. 1. Two-color case: construction of the approximate solution from a black tour and a collection of cycles and edges when $\alpha_B = \alpha_W = 0$ and $\beta_B = \beta_W = 2$

With the help of the triangle inequality and since Christofidès' heuristics achieves a 1.5-approximation, we get

$$d(T_B) \leq 1.5\ d(T^*).$$

Let $\tilde{f}$ be the map $E \to \{0, 1, 2\}$ such that $\tilde{f}(e) = 1$ if e in T^* and 0 otherwise. This map $\tilde{f}$ is in $\mathcal{F}$. Since f^* has been chosen in order to minimize $\sum_{e \in E} d(e)f(e)$, we have

$$\sum_{e \in E} d(e)f^*(e) \leq \sum_{e \in E} d(e)\tilde{f}(e) = d(T^*).$$

The conclusion follows. □

Proof (of Proposition 1). If $\beta_B = 0$, the problem admits a feasible solution only if $V_W = \emptyset$ and then coincides with the usual Traveling Salesman Problem for which Christofidès' heuristics achieves a 1.5-approximation. We suppose now that $\beta_B \geq 1$. The algorithm proceeds in two steps.

During the first step, we compute a subset $F \subseteq E$ such that each edge in F has one white endpoint and one black endpoint and such that

$$\deg_F(v) = 1 \text{ if } v \in V_W \text{ and } \max(1, \alpha_B) \leq \deg_F(v) \leq \beta_B \text{ if } v \in V_B \quad (2)$$

and with minimal $d(F)$. Such a subset exists necessarily if there is a feasible solution to the MTSP. The one with minimal $d(F)$ can be computed in polynomial-time. Indeed this is a special case of the problem studied p.353 of *Combinatorial Optimization* [11] which seeks an optimal edge subset in a bipartite graph that is simultaneously a "b-matching" and a "a-vertex cover" and reduces to a minimum cost circulation.

The idea behind the computation of F is roughly the following. An optimal Hamiltonian tour T^* of the MTSP with $\beta_W = 1$ or $\alpha_B > 0$ has no adjacent black vertices. There exists thus a subset $\tilde{F}$ of edges with an endpoint in each color, satisfying condition (2), and with a total length not too far from $d(T^*)$: indeed, define $\tilde{F}$ to be the subset of E such that $wb \in \tilde{F}$ if $w \in V_W$, $b \in V_B$, and b is the closest black vertex to w in T^*.

During the second step, we build the Hamiltonian cycle T we look for. We compute a tour $T_B = (b_1, \ldots, b_{|V_B|})$ on the vertices of V_B using Christofidès' heuristics. The neighbors of b_i in (V, F) are denoted $w^i_1, \ldots, w^i_{r_i}$ where $r_i = \deg_F(b_i)$. We define P_i to be the path whose vertex set is $\{b_i, w^i_1, \ldots, w^i_{r_i}\}$ and whose edge set is $\{b_i w^i_1, w^i_1 w^i_2, \ldots, w^i_{r_i-1} w^i_{r_i}\}$. We define then T to be

$$T = P_1 \cup \{w^1_{r_1} b_2\} \cup P_2 \cup \cdots \cup P_{|V_B|} \cup \{w^{|V_B|}_{r_{|V_B|}} b_1\}$$

identifying the tour T and its edge set. Figure 2 depicts this construction.

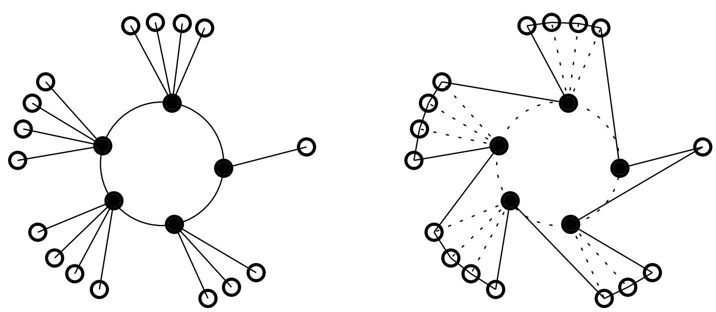

Fig. 2. Two-color case: construction of an approximate solution when $\beta_W = 1$ or $\alpha_B > 0$

It remains to show that the length of T is less than $(2\lceil \beta_B/2 \rceil + 1.5)$ times the best achievable length. We denote T^* an optimal solution of the MTSP.

$$
\begin{aligned}
d(T) &= \sum_{i=1}^{|V_B|} \big(d(P_i) + d(w^i_{r_i} b_{i+1})\big) \\
&\le \sum_{i=1}^{|V_B|} \left(d(b_i w^i_1) + \sum_{j=2}^{r_i - 1} (d(b_i w^i_j) + d(b_i w^i_{j+1})) \right) + \sum_{i=1}^{|V_B|} (d(w^i_{r_i} b_i) + d(b_i b_{i+1})) \\
&\le 2\ d(F) + d(T_B).
\end{aligned}
$$

With the help of the triangle inequality and since Christofidès' heuristics achieves a 1.5-approximation in its best version, we get

$$d(T_B) \le 1.5\ d(T^*).$$

It remains to bound $d(F)$. We use the subset $\tilde{F}$ already mentioned: $\tilde{F}$ is the subset of E such that $wb \in \tilde{F}$ if $w \in V_W$, $b \in V_B$, and b is the closest black vertex to w in T^* in terms of the number of vertices between them. There is a tie precisely when there is an odd number of white vertices between two black vertices. In this case, for the central white vertex w, any of the two black vertices closest to w could be selected. We break the tie by orienting the tour and by selecting the black vertex being forward.

By the triangle inequality we have

$$d(\tilde{F}) \le \left\lceil \frac{\beta_B}{2} \right\rceil\ d(T^*).$$

Since $\tilde{F}$ satisfies the same constraints as F, and since F has been chosen of minimal length, we have

$$d(F) \le \left\lceil \frac{\beta_B}{2} \right\rceil\ d(T^*).$$

The conclusion follows. □

Proof (of Proposition 2).

The algorithm proceeds in four steps.

In the first step, we consider the complete bipartite graph with vertex partition given by $V_i \cup V_j$ for each pair i, j with $i \neq j$. The weight $w(e)$ of an edge e is its length $d(e)$ in the input graph. For each of these $k(k-1)/2$ bipartite graphs, we compute a minimal-weight perfect matching M_{ij}.

The second step considers the complete graph K_k whose vertices are identified with the colors $1, 2, \ldots, k$. Each edge ij in K_k gets a length $l(ij) = w(M_{ij})$. We compute a Hamiltonian tour $C = (j_1, \ldots, j_k)$ in K_k with the help of Christofidès' heuristics.

In the third step, we build a Hamiltonian tour H_i on the vertices in V_{j_i}, again with the help of Christofidès' heuristics, for each $i = 1, \ldots, k$.

The fourth step aims to build a Hamiltonian tour T_i on the whole input graph, for $i = 1, \ldots, k$. We explain the construction for T_1, the other T_i's being built in a similar way. Let $H_1 = (v_1, \ldots, v_s)$. The edges of $M_{j_1 j_2} \cup M_{j_2 j_3} \cup \cdots \cup M_{j_{k-1} j_k}$ are partitioned into s vertex-disjoint paths, each of them having an endpoint in V_{j_1} and an endpoint in V_{j_k}. For each vertex $v_\ell \in V_{j_1}$, we denote by P_ℓ the path among them having v_ℓ as an endpoint, and by $w_\ell \in V_{j_k}$ its other endpoint. We define T_1 to be

$$T_1 = P_1 \cup \{w_1 v_2\} \cup P_2 \cup \{w_2 v_3\} \cup \cdots \cup P_s \cup \{w_s v_1\}.$$

The other T_i's are obtained by taking H_i in place of H_1 and by considering the $M_{j_t j_{t+1}}$ except for $t = i - 1$.

The T_i with minimal $d(T_i)$ is the tour T we look for.

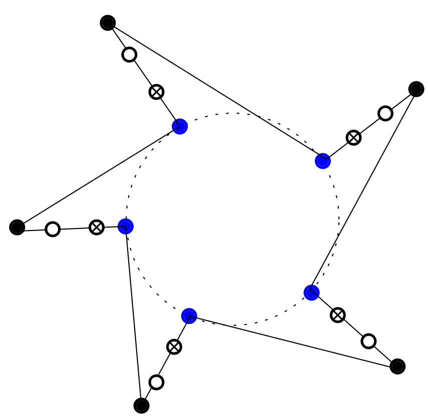

Fig. 3. The construction of T_1

Figure 3 depicts this construction.

It remains to show that the length of T is less than $\left(4.5 - \frac{3}{k}\right)$ times the best achievable length. We denote T^* an optimal solution of the MTSP. Using the triangle inequality we have

$$\begin{aligned} d(T) &\leq d(H_1) + 2\sum_{\ell=1}^{s} d(P_\ell) \\ &\leq 1.5\ d(T^*) + 2\sum_{t=1}^{k-1} w(M_{j_t j_{t+1}}) \end{aligned}$$

We have similar inequalities when we replace T_1 by T_i. Summing all these k inequalities, we obtain

$$\begin{aligned} k\ d(T) &\leq 1.5k\ d(T^*) + 2(k-1)\sum_{t=1}^{k} w(M_{j_t j_{t+1}}) \\ &\leq 1.5k\ d(T^*) + 2(k-1)l(C) \\ &\leq 1.5k\ d(T^*) + 3(k-1)l(C^*) \end{aligned}$$

where C^* is an optimal Hamiltonian tour on the complete graph K_k whose vertices are identified with the colors and with length function l.

Since any feasible solution visits the colors in a periodic way, the optimal tour T^* induces a Hamiltonian tour C' on the complete graph on the colors K_k. Moreover, T^* induces a perfect matching M'_{ij} on the complete bipartite graph with vertex partition $V_i \cup V_j$ for each edge ij in C'. We have thus

$$d(T^*) = \sum_{e \in E(C')}^{k} w(M'_e) \geq \sum_{e \in E(C')}^{k} w(M_e) = l(C') \geq l(C^*).$$

The conclusion follows. □

Proof (of Proposition 3). Consider a feasible Hamiltonian cycle and the associated circular word (w) on the alphabet $\{W, B\}$. Each occurrence of the letter W is in exactly one maximal B-factor of (w). Therefore we have

$$|V_W| = \sum_{f \in F_B} |f|, \tag{3}$$

where F_B is the set of all maximal B-factors. Note that $|F_B| = |V_B|$. Since $\alpha_B \leq |f| \leq \beta_B$ for all $f \in F_B$, we have $\alpha_B |V_B| \leq |V_W| \leq \beta_B |V_B|$. A similar reasoning exchanging the roles played by the two colors leads to

$\alpha_W|V_W| \leq |V_B| \leq \beta_W|V_W|$.

Conversely, suppose that the inequalities are satisfied. Consider the trigonometric circle on which we put a mark W every $\frac{2\pi}{|V_W|}$ and a mark B every $\frac{2\pi}{|V_B|}$, such that no two marks have same position. By a perturbation argument, such a construction is possible. Reading the marks on the circle in an arbitrary direction, we obtain a circular word over $\{W, B\}$. Between two consecutive W's, there is at most $\lceil |V_B|/|V_W| \rceil$ and at least $\lfloor |V_B|/|V_W| \rfloor$ occurrences of B's. Similarly, the number of occurrences of W's between two consecutive B's is between $\lceil |V_W|/|V_B| \rceil$ and $\lfloor |V_W|/|V_B| \rfloor$. Hence this word provides a feasible solution for the MTSP. □

Proof (of Proposition 4). It is the necessary condition (1). It is a sufficient condition as a consequence of Corollary 1 in Section 4. □

Proof (of Proposition 5). We consider a circular word (w') on the alphabet $\{1, X\}$ with $|V_1|$ occurrences of letter '1' and $|V_2| + |V_3|$ occurrences of letter X, such that there is at most one X between two '1' and at most $\left\lceil \frac{|V_1|}{|V_2|+|V_3|} \right\rceil$ occurrences of letter '1' between two consecutive occurrences of letter X. Such a circular word exists according to Proposition 3. Now, we replace the occurrences of X by $|V_2|$ occurrences of letter '2' and $|V_3|$ occurrences of letter '3' with the conditions that there is at most one occurrence of letter '3' between two occurrences of letter '2' and at most $\left\lceil \frac{|V_2|}{|V_3|} \right\rceil$ occurrences of letter '2' between two occurrences of letter '3'. It is possible, again because of Proposition 3. A straightforward calculation shows that the obtained circular word on alphabet $\{1, 2, 3\}$ corresponds to a feasible solution for the MTSP with the β_i satisfying the constraint of the statement. □

Proof (of Proposition 6). Consider a feasible Hamiltonian cycle, the associated circular word (w), and a letter $i \in \{1, \ldots, k\}$. Each letter distinct from i is in exactly one maximal i-factor. Therefore

$$\sum_{j \neq i} |V_j| = \sum_{f \in F_i} |f|, \tag{4}$$

where F_i is the set of all maximal i-factors of (w). Note that $|F_i| = |V_i|$. Since $\alpha_i \leq |f| \leq \beta_i$ for all $f \in F_i$, we have $\alpha_i|V_i| \leq \sum_{j \neq i} |V_j| \leq \beta_i|V_i|$. □

Proof (of Proposition 7). Let j be an integer in $\{1, \ldots, k-1\}$ and let $i_1, \ldots, i_j$ be integers such that $1 \leq i_1 < \cdots < i_j < k$. We start with a small claim.

Claim. Let u be a word on alphabet $\{i_1,\ldots,i_j,k\}$ with one occurrence of k and two occurrences of i_ℓ for $\ell=1,\ldots,j$. If the letters i_ℓ are such that $i_\ell k i_\ell$ is a subword of u for $\ell=1,\ldots,j$, then

$$\sum_{\ell=1}^{j}\lambda_{i_\ell}=j^2,$$

where λ_{i_ℓ} is the length of the i_ℓ-factor between the two occurrences of letter i_ℓ.

Proof (of the claim). Let $u=a_1a_2\cdots a_{2j+1}$. Note that $a_{j+1}=k$. Define $\chi(s,\ell)$ to be 1 if a_s is strictly between the two occurrences of the letter i_ℓ, and 0 otherwise. We have then

$$\sum_{\ell=1}^{j}\lambda_{i_\ell}=\sum_{\ell=1}^{j}\sum_{s=1}^{2j+1}\chi(s,\ell)=\sum_{s=1}^{2j+1}\sum_{\ell=1}^{j}\chi(s,\ell).$$

If $s\le j$, we have $\sum_{\ell=1}^{j}\chi(s,\ell)=s-1$. If $s\ge j+2$, we have $\sum_{\ell=1}^{j}\chi(s,\ell)=2j+1-s$. Moreover, $\sum_{\ell=1}^{j}\chi(j,\ell)=j$. Therefore,

$$\sum_{\ell=1}^{j}\lambda_{i_\ell}=\sum_{s=1}^{j}(s-1)+j+\sum_{s=j+2}^{2j+1}(2j+1-s)=2\sum_{s=1}^{j}(s-1)+j=j^2.$$

The claim is proved.

Now, take a feasible solution of the MTSP, and (w) an associated circular word. Choose an occurrence of the letter k in (w) and for each i_ℓ with $\ell=1,\ldots,j$, consider the closest occurrences of the letter i_ℓ on the left and on the right of this occurrence of k (the word being circular it can be the same occurrence). It provides a finite word u exactly as in the statement of the claim. The length λ_{i_ℓ} on this word is necessarily bounded by β_{i_ℓ} for $\ell=1,\ldots,j$. Finally, we obtain

$$\sum_{\ell=1}^{j}\beta_{i_\ell}\ge\sum_{\ell=1}^{j}\lambda_{i_\ell}=j^2,$$

as required. □

Proof (of Proposition 8). Consider the trigonometric circle on which we put a mark i every $\frac{2\pi}{|V_i|}$ for $i=1,\ldots,k$ such that no two marks have same position. By a perturbation argument, such a construction is possible.

Reading the marks in an arbitrary direction gives a circular word in which any maximal i-factor f satisfies $\sum_{j\neq i}\left\lfloor\frac{|V_j|}{|V_i|}\right\rfloor \leq |f| \leq \sum_{j\neq i}\left\lceil\frac{|V_j|}{|V_k|}\right\rceil$ for any i. The length of the arc of the circle between two consecutive marks i being $\frac{2\pi}{|V_i|}$, there are indeed at least $\left\lfloor\frac{|V_j|}{|V_i|}\right\rfloor$ and at most $\left\lceil\frac{|V_j|}{|V_i|}\right\rceil$ marks j between two consecutive i's. □

Proof (of Proposition 9). The proof works by induction on k.

If $k = 2$ and $|V_2|$ divides $|V_1|$, then the circular word with one occurrence of letter 2 every $|V_1|/|V_2|$ occurrences of letter 1 corresponds to a feasible solution for the MTSP.

Assume now that $k > 2$ and that $|V_i|$ divides $\sum_{j=1}^{i-1}|V_j|$ for all $i = 2,\ldots,k$. We build a word as follows. We consider a solution for the MTSP with $k-1$ colors as in the statement, which exists by induction, and take the associated circular word. We now insert the letter k exactly $|V_k|$ times in a periodic way along the circular word, that is every $\frac{\sum_{j=1}^{k-1}|V_j|}{|V_k|}$ letters.

Any maximal k-factor f satisfies $|f| = \frac{\sum_{j=1}^{k-1}|V_j|}{|V_k|}$. Consider now $i \neq k$ and a maximal i-factor f. We note $\bar{f}$ the subword of f we get by removing all occurrences of k. By induction, we have

$$\frac{\sum_{j=1}^{i-1}|V_j|}{|V_i|} \leq |\bar{f}| \leq \frac{\sum_{j=1}^{i-1}|V_j|}{|V_i|} + k - 1 - i. \tag{5}$$

Suppose for a contradiction that $|f|_k \geq 2$. By construction, the factor $\bar{f}$ is then such that $|\bar{f}| \geq \frac{\sum_{j=1}^{k-1}|V_j|}{|V_k|}$. As $\frac{\sum_{j=1}^{k-1}|V_j|}{|V_k|} \geq \frac{\sum_{j=1}^{i-1}|V_j|}{|V_k|} + \frac{\sum_{j=i}^{k-1}|V_j|}{|V_k|} \geq \frac{\sum_{j=1}^{i-1}|V_j|}{|V_i|} + k - i$, it would contradict Equation (5). Therefore $|f|_k \leq 1$, and the i-factor f satisfies $|\bar{f}| \leq |f| \leq |\bar{f}| + 1$, which, combined with Equation (5), leads to the desired conclusion. □

Proof (of Proposition 11). Consider a balanced circular word (w) over $\{1,\ldots,k\}$ with $|w|_i = |V_i|$ for all $i = 1,\ldots,k$. According to Proposition 10, there is an integer d_i^1 such that the maximal i-factors of (w) are of size d_i^1 or $d_i^1 + 1$. By a straightforward argument similar to the one used in the proof of Proposition 6, we have $d_i^1 < \frac{\sum_{j\neq i}|V_j|}{|V_i|} < d_i^1 + 1$ for all $i = 1,\ldots,k$. If

$$\alpha_i \leq \frac{\sum_{j\neq i}|V_j|}{|V_i|} \leq \beta_i \quad \text{for all } i = 1,\ldots,k,$$

then we have $d_i^1 \geq \alpha_i$ and $d_i^1 + 1 \leq \beta_i$ for $i = 1, \ldots, k$ and (w) induces a feasible solution for the MTSP.