Stochastic analysis of partitioning algorithms for matching problems

Ludger Rüschendorf and Gernot Sachs
University of Freiburg

Abstract

Partitioning algorithms for the Euclidean matching and for the semi-matching problem in the plane are introduced and analysed. The algorithms are analogues of Karp's well-known dissection algorithm for the traveling salesman problem. The algorithms are proved to run in time $n \log n$ and to approximate the optimal matching in the probabilistic sense. The analysis is based on the techniques developed in Karp (1977) and on the limit theorem of Redmond and Yukich (1993) for quasiadditive functionals.

1 Introduction

Define for $(x_1, \dots, x_n) \in \mathbb{R}^d, d \geq 1$ the weighted Euclidean matching functional

$$L(x_1, \dots, x_n) \;=\; \min_{\sigma \in S_n} \sum_{i=1}^{m} \|x_{\sigma(2i-1)} - x_{\sigma(2i)}\| \tag{1.1}$$

where S_n is the set of permutations of $\{1, \dots, n\}, m = \left[\frac{n}{2}\right]$, and $\|\|$ is the Euclidean norm. L is a quasiadditive functional, i.e. L is subadditive, continuous and there exists an approximating superadditive functional $\widehat{L} \leq L + 1$ on $[0,1]^d$ defined by the corresponding boundary matching functional which allows matching to boundary points (cf. Redmond and Yukich (1993) for an indication and Sachs (1997) for a detailed proof. See also the detailed discussion in the recent book of Steele (1997). Therefore as consequence of the general version of the Beardwood, Halton and Hammersly (BHH) Theorem of Redmond and Yukich (1994) for any *iid* sequence (X_i) on $[0,1]^d$ with $P^{X_i} = \mu$ holds

$$\lim_{n \to \infty} \frac{L(X_1, \dots, X_n)}{n^{(d-1)/d}} \;=\; \beta(L) \int f(x)^{\frac{(d-1)}{d}} d\lambda^d(x) \tag{1.2}$$

in the sense of complete convergence, where f is the density of the Lebesgue continuous part of μ. Papadimitriou (1978) proved (1.2) for the matching functional in the uniform case while Rhee (1993) gave a proof of a more general limit theorem based on a continuity property of L.

Karp (1977) introduced a partitioning algorithm for the traveling salesman problem (TSP) using a random subdivision of the domain (in the case d = 2). This partitioning algorithm subdivides the set of cities into small groups, constructs an optimum tour through each group and then patches the subtours together to obtain a tour through all the cities. Based on the BHH-Theorem, Karp could prove that this algorithm is asymptotically optimal in a model with randomly distributed cities in the sense that for any $\varepsilon > 0$ the probability that the tour length $T(X_1, \ldots, X_n)$ exceeds the optimal tour length by a factor more than $1+\varepsilon$ converges to zero. Moreover the algorithm is proved to run within time $D_1\varepsilon^2 d_1^{1/\varepsilon^2} n + 0(n \log n)$ with some constants d_1, D_1. In particular this was the first instance of a polynomial approximation to a NP-problem. Papadimitriou (1978) introduced and analysed a 3-phase algorithm for Euclidean matching and obtained some bounds for the constant β. Dyer and Frieze (1984) introduced a partitioning algorithm with a different fixed partitioning scheme for the Euclidean matching problem and proved that it approximates the optimal matching in a probabilistic sense for uniform distributions.

In this paper we introduce analogues of Karps algorithm for the Euclidean matching and for the semi-matching problem. In the semi-matching problem it is allowed that any point is matched to any number of points with sum of the weights equal to one. We prove that in both cases the proposed partitioning algorithm approximates the optimal solution in a probabilistic model and operates in running time $n \log n$ for general distributions. For the Euclidean matching it is useful to construct the partitioning in such a way that nearly all subproblems have an even number of points. In contrast to the TSP which is an NP-problem, the Euclidean matching and semi-matching problems can be formulated as linear programming problems and therefore have a polynomial running time exact solution of order $O(n^3)$. But the improvement of the order of the running time to $O(n \log n)$ by the partitioning algorithm discussed in this paper is of practical interest also for these combinatorial optimization problems.

Yukich (1995) introduced to any quasiadditive functional L with dual functional $\widehat{L}$ a functional

$$\widehat{L}^A(F) \;=\; \frac{1}{m} \sum_{i \le m^d} \widehat{L}(m[(F \cap Q_i) - q_i]), \tag{1.3}$$

$F \subset [0,1]^d$ a finite subset, (q_i) centering points and $(Q_i)_{i=1}^{m^d}$ the usual partition of $[0,1]^d$ into subcubes. He proved that

$$\begin{aligned} \widehat{L}^A(U_1, \ldots, U_n) &\le 1 + L(U_1, \ldots, U_n) \text{ and} \\ E\widehat{L}^A(U_1, \ldots, U_n) &\le 1 + EL(U_1, \ldots, U_n) \\ &\le 1 + E\widehat{L}^A(U_1 \ldots U_n) + C(\log n)^{-1/d} n^{(d-1)/d}. \end{aligned} \tag{1.4}$$

The expected execution time for calculating $\widehat{L}^A$ is $O(n^2 \log^{B-1} n)$, where B is a constant arising from the execution time of $\widehat{L}(F)$ in the form $A|F|^B\, 2^{|F|}$. So $\widehat{L}^A$ approximates the functional L in the sense of expectations and can be calculated in order $O(n^2 \log^{B-1} n)$. In contrast we construct an approximatively optimal matching. So our aim is different from that in Yukich's paper.

2 Weighted Euclidean matching in $\mathbb{R}^2$

In this section we introduce and analyse a partitioning algorithm for weighted Euclidean matching in $\mathbb{R}^2$. The construction and analysis of this algorithm is motivated by Karps (1977) treatment of the TSP.

2.1 Introduction of the partitioning algorithm

For n points $x_1, \ldots, x_n \in [a, b] \subset \mathbb{R}^2, a < b$ let $|T_n^*|$ denote the value of an optimal Euclidean matching

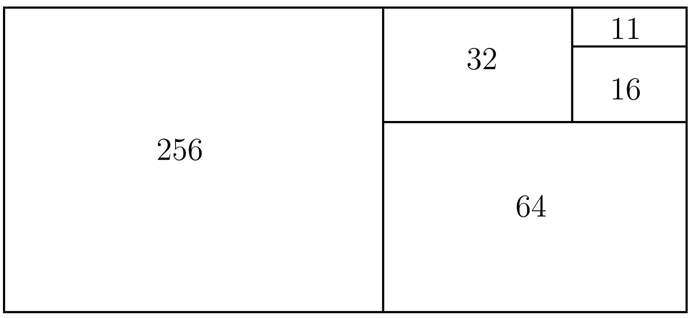
$$|T_n^*| \;=\; \min_{\sigma \in S_n} \sum_{i=1}^{m} \|x_{\sigma(2i-1)} - x_{\sigma(2i)}\| \qquad m = \left[\frac{n}{2}\right]. \tag{2.1}$$

The following partitioning algorithm proceeds by subdividing the rectangle $[a, b]$ into a number of subrectangles such that at most one of them has an odd number of points. Determining a matching in each of the subrectangles results in a matching of the $\{x_i\}$ which is approximatively optimal if the number of subdivisions is chosen suitably.

For $n = \sum_{i=0}^{m_n} \alpha_{n,i} 2^i, \alpha_{n,i} \in \{0, 1\}, \alpha_{n,m_n} = 1$ let $t = t_n = 2^{\ell_n}, 0 \leq \ell_n \leq m_n$ be an upper bound for the number of points in the subrectangles. We assume that all points $x_1, \ldots, x_n$ have different $x-$ and $y-$coordinates and let w.l.g. $b_1 - a_1 \geq b_2 - a_2$.

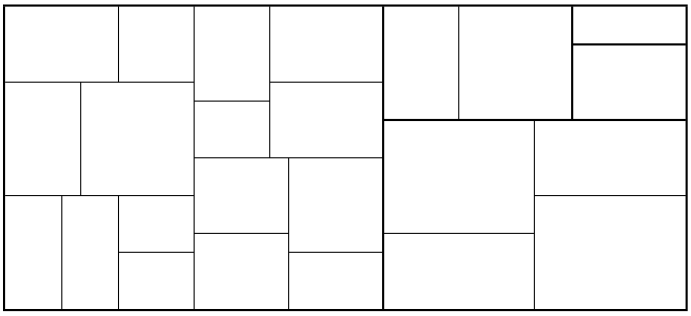
Specification of the algorithm: Cut the rectangle into two parts parallel to the $y-$ax is such that the left part contains 2^{m_n}, the right part $n - 2^{m_n}$ points. In the right rectangle cut parallel to the smaller side such that 2^k points, where $k = \max\{j < m_n;\ \alpha_{n,j} = 1\}$, are in one part of the new pair of rectangles. Continue cutting parallel to the smaller side inductively until the last constructed rectangle contains at most t_n points.

Figure 1: Example of a subdivision with $n = 379$, $t_n = 16$



Finally each of the rectangles with 2^k points, where $\ell_n + 1 \leq k \leq m_n$ is divided in $k - \ell_n$ steps cutting parallel to the shorter sides into $2^{k-\ell_n}$ rectangles with each $t_n = 2^{\ell_n}$ points in it so one obtains a partition with many rectangles containing

2^{ℓ_n} points and possibly one rectangle with a number $\leq 2^{\ell_n}$ points. The resulting partition looks like
Figure 2: Final form of the subdivision.



The even number of points in the subrectangles allows a simplified analysis of this algorithm. In the subrectangles we determine an optimal matching which finally results in a matching of $\{x_i\}$ which we call "Partmatch". Let $|W_n|$ denote the value of this matching.

2.2 Analysis of the execution time

The partitioning algorithm ("Partmatch") in 2.1 subdivides the rectangle $[a, b]$ into subrectangles with at most t_n points and finds optimal matchings in the subrectangles.

We assume that we can solve an Euclidean matching problem with k points in time $\leq Dk^p$, with constants p, D. This is fulfilled for $p = 3$ using Papadimitriou and Steiglitz (1982), Theorem 11.3, Problem 14. Improvements of this order to $O(n^{2,5} \log n)$ for some geometric algorithms are given in Vajda (1989). Therefore, for our subdivided rectangles we need at most the time

$$\begin{aligned} Dt_n^p(2^{m_n-\ell_n} + 2^{m_n-\ell_n-1} + \cdots + 1 + 1) &= Dt_n^p 2^{m_n-\ell_n+1} \\ &\leq Dt_n^p \frac{2n}{t_n} = 2Dt_n^{p-1} \cdot n, \end{aligned} \tag{2.2}$$

for constructing the optimal matchings in the rectangles.

The partitioning algorithm divides successively a rectangle with k points into two subrectangles with $\left[\frac{k}{2}\right]$ and $\left[\frac{k}{2}\right] + 1$ points by cutting along the smaller side.

Putting the cities into two linked lists H according to increasing values of $x-$coordinates and V according to increasing values of $y-$coordinates a partitioning of the rectangle into two subrectangles cutting parallel to the smaller side is done by producing two sublists H_1, H_2 resp. V_1, V_2 giving the corresponding horizontal or vertical coordinates. This needs in each step time proportional to the number of points in the rectangle.

For the sorting step with Heapsort or Mergesort (cf. Ottmann and Widmayer (1990)) we need $O(n \log n)$ steps. For the subdivision of the rectangle into subrectangles with 2^k points $\ell_n + 1 \leq k \leq m_n$, we need, observing that the size of the succeeding subrectangle is smaller than $\left[\frac{n}{2}\right], \leq K_n(1 + \frac{1}{2} + \frac{1}{4} + \cdots + {}^1\!/_{2^{[n/2]}}) \leq 2K_n$ steps. Finally the rectangles with 2^k points are subdivided in $(k - \ell_n)$ subrectangles of 2^{ℓ_n} points each, which need for a rectangle with 2^k points at most

$$K(2^k + 2 \cdot 2^{k-1} + 4 \cdot 2^{k-2} + \cdots + 2^{k-\ell_n-1} \cdot 2^{\ell_n+1}) \;=\; O(2^k(k - \ell_n))$$

steps.
Together this results in

$$\sum_{i=\ell_n+1}^{m_n} O(2^i(i - \ell_n)) \;\leq\; C m_n \sum_{i=\ell_n+1}^{m_n} 2^i \;\leq\; C(\log n)n.$$

Theorem 2.1 *With suitable implementation the partitioning algorithm partmatch operates within the time bound*

$$2Dt_n^{p-1} n + O(n \log n) \tag{2.3}$$

2.3 Error analysis

To obtain an upper bound for the error $|W_n| - |T_n^*|$ of the partitioning algorithm to the optimal matching we use a result on cutting games from Karp (1977).

Given a rectangle $[a, b] \subset \mathbb{R}^2$ in the first round of the game the rectangle is divided into two subrectangles. Player 1 decides whether to cut parallel to the $x-$ or the $y-$axis, then player 2 decides where to put the cut.

In the ℓ-th round each of the $2^{\ell-1}$ subrectangles is divided according to the same principle. After k rounds the game ends and player 1 pays to player 2 an amount equal to the sum of the perimeters of the 2^k rectangles produced in round k. Call the short strategy for player 1 to cut parallel to the shorter side and the bisection strategy for player 2 to divide a rectangle into equal halves. Then with $F_k(a, b)$ the sum of the perimeters of a k round game, where both players play optimal (i.e. use a minimax strategy) the following result holds.

Theorem 2.2 *(cp. Karp (1977), Theorem 2, Corollary 1)*

a) The short strategy is optimal for player 1 the bisection strategy is optimal for player 2 (in the sense of minimax).

b) $F_k(a, b) = \min_{s+t=k} \; 2(2^t(b_1 - a_1) + 2^s(b_2 - a_2))$

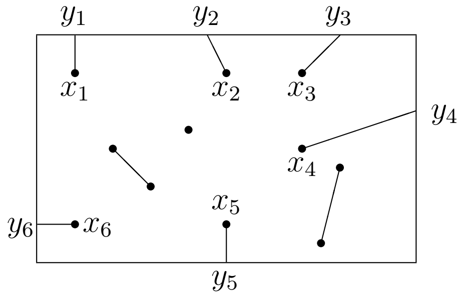
c) $\sup_k \frac{F_k(a,b)}{2^{k/2}} < \infty$

For $F \subset [a, b] \subset \mathbb{R}^2$ with $|F| = n$ and a rectangle Y let $\mathrm{per}(Y)$ denote the perimeter of Y, let $|T^*(Y)|$ denote the value of an optimal matching of $F \cap Y$, $|T_n^*| = |T^*([a, b])|$ and let $|T_n^* \cap Y|$ be the sum of segments of an optimal matching of F with both endpoints in Y.

Lemma 2.3 $|T^*(Y)| - |T_n^* \cap Y| \;\leq\; \mathrm{per}(Y)$.

Proof: Let $y_1, \ldots, y_k$ denote the cutting points of segments of the optimal matching of F with ∂Y, numbering w.l.g. the y_i clockwise. Let $x_1, \ldots, x_k$ denote the corresponding endpoints of the segments in Y.

Figure 3: Segments of an optimal matching.



<u>Case 1</u>

In Y is no unmatched point. Then w.l.g. $k \in 2\mathbb{N}$ and

$$\overline{y_1y_2} + \overline{y_3y_4} + \cdots + \overline{y_{k-1}y_k} \quad \leq \quad \overline{y_2y_3} + \overline{y_4y_5} + \cdots + \overline{y_ky_1}$$

where for two points w, z on ∂Y $\overline{wz}$ is the length of the shortest way on the boundary. So we obtain a matching of the points on the boundary of length $\leq \frac{1}{2}\mathrm{per}(Y)$. This results in a matching of the points in Y by keeping the matchings of points in Y, deleting the segments $[x_i, y_i]$ and joining $[x_{2i-1}, x_{2i}], 1 \leq i \leq \frac{k}{2}$. From the triangle inequality we obtain

$$|T^*(Y)| \;\leq\; |T_n^* \cap Y| + \frac{1}{2}\mathrm{per}(Y) \tag{2.4}$$

<u>Case 2</u>

There is an unmatched point in Y. For k even we argue as in case 1. For k odd we match $x_1, \ldots, x_{k-1}$ as in case 1 and match x_k with the unmatched point. This adds at most $\frac{1}{2}\mathrm{per}(Y)$ to the sum of all matchings. This implies the bound in the lemma. □

Lemma 2.4 *The sum of the perimeters of the rectangles of $[a, b]$ produced by the partmatch algorithm is bounded above by $F_{m_n - \ell_n + 1}(a, b)$.*

Proof: The resulting partition of the partitioning algorithm may be regarded as a play of a $(m_n - \ell_n + 1)$-round cutting game. If at some rectangles no cut is made (in case $\alpha_{n,i} = 0$), the play can be regarded as one in which player 1 chooses the optimal short cut strategy while player 2 chooses his cutpoint at the boundary of the longer side of the rectangle. Thus by Theorem 2

$$\sum \mathrm{per}(Y_i) \;\leq\; F_{m_n - \ell_n + 1}(a, b).$$

□

As consequence we obtain the following error bound

Theorem 2.5 *An error bound of the partmatch algorithm is given by*

$$\begin{aligned} |W_n| - |T_n^*| &\leq F_{m_n-\ell_n+1}(a,b) \qquad (2.5) \\ &\leq F_{\left[\log_2 \frac{2n}{t_n}\right]}(a,b) = O\left(\sqrt{\frac{n}{t_n}}\right) \end{aligned}$$

Proof: Let $Y_1, \ldots, Y_{m_n-\ell_n+1}$ denote the subrectangles produced by partmatch with $2^{m_n}, \ldots, 2^{\ell_n}$ points and $Y_{m_n-\ell_n+2}$ the rectangle with less than 2^{ℓ_n} points (if $\alpha_{n,i} = 0$ then set $Y_i := \phi$). Furthermore, for $k = 1, \ldots, m_n - \ell_n + 1$ let $Y_k^1, \ldots, Y_k^{2^{m_n-\ell_n-k+1}}$ denote the subdivisions of Y_k into rectangles with 2^{ℓ_n} points (eventually defined to be empty). By Lemmas 2.3, 2.4 we obtain

$$\begin{aligned} |W_n| &= \sum_{k=1}^{m_n-\ell_n+2} \sum_{j=1}^{2^{m_n-\ell_n-k+1}} |T^*(Y_k^j)| \\ &\leq \sum_k \sum_j \left(|T_n^* \cap Y_k^j| + \mathrm{per}(Y_k^j) \right) \\ &= |T_n^*| + \sum_k \sum_j \mathrm{per}(Y_k^j) \\ &\leq |T_n^*| + F_{m_n+\ell_n+1}(a,b) \\ &\leq |T_n^*| + F_{\left[\log_2 \frac{2n}{t_n}\right]}(a,b) \end{aligned}$$

Since by Theorem 2.2, $F_k(a,b) = O(2^{k/2})$, we obtain

$$\begin{aligned} F_{\left[\log_2 \frac{2n}{t_n}\right]} &= O\left(2^{\frac{1}{2}\left[\log_2 \frac{2n}{t_n}\right]}\right) \\ &= O\left(\sqrt{\frac{2n}{t_n}}\right) = O\left(\sqrt{\frac{n}{t_n}}\right) \end{aligned}$$

$\square$

We next show that in a model with random independent points in $[0,1]^2$ we can choose the parameter t_n such that the corresponding partmatch algorithm approximates the optimal matching with high probability.

Theorem 2.6 *Let (X_i) be iid random points in $[a,b] \subset \mathbb{R}^2$ and let P^{X_1} be not singular w.r.t. λ^2. Then for some $D_1 > 0$ and $\varepsilon > 0$ there exists a $t_n = t_n(\varepsilon)$ such that the corresponding partmatch algorithm $W_n(t_n)$ satisfies:*

a) $P(|W_n(t_n)| \leq (1+\varepsilon)|T_n^*|) = 1 - o(1)$

b) With suitable implementation the running time is (for some $1 \leq p \leq 3$)
$\leq D_1(\frac{1}{\varepsilon^2})^{p-1} n + O(n \log n)$

Proof:

a) Let $\varepsilon > 0$ be given. By the Theorem of Redmond and Yukich (1993) (cp. (1.2))

$$\lim_{n\to\infty} \frac{|T_n^*|}{\sqrt{n}} = \beta \int f(x)^{\frac{1}{2}}\, \mathrm{d}\lambda^2(x) =: \widetilde{\beta} > 0$$

since P^{X_1} is not singular w.r.t. λ^2.
From Theorem 2.5 there exists a constant $C > 0$ such that $|W_n| - |T_n^*| < C\sqrt{\frac{n}{t_n}}$.
Choosing $t_n := 2^{\left[\log_2\left(4(C+1)^2/\widetilde{\beta}^2\varepsilon^2\right)\right]+1}$ we have on $\{|T_n^*| > \frac{\widetilde{\beta}}{2}\sqrt{n}\ \} =: A_n$

$$\begin{aligned} \frac{|W_n|}{|T_n^*|} &\leq \frac{|T_n^*| + C\sqrt{\frac{n}{t_n}}}{|T_n^*|} \\ &< 1 + 2C\frac{\sqrt{\frac{n}{t_n}}}{\widetilde{\beta}\sqrt{n}} \\ &< 1 + \frac{2C}{\widetilde{\beta}}\ \frac{\widetilde{\beta}\varepsilon}{2(C+1)} \\ &< 1 + \varepsilon. \end{aligned} \tag{2.6}$$

Since $\sum_n P\left(A_n^c\right) < \infty$ we conclude that $\sum_n P\left(\frac{|W_n|}{T_n^*} \geq 1 + \varepsilon\right) < \infty$.
Therefore,

$$\frac{|W_n|}{|T_n^*|} \longrightarrow 1 \quad a.s. \tag{2.7}$$

and in (2.7) even complete convergence holds.

b) By Theorem 2.1 the running time is bounded by

$$\begin{aligned} &2Dt_n^{p-1} \cdot n + O(n\log n) \\ &\quad\leq\ 2D\left(\frac{8(C+1)^2}{\widetilde{\beta}^2\varepsilon^2}\right)^{p-1} \cdot n + O(n\log n) \\ &\quad\leq\ D_1\left(\frac{1}{\varepsilon^2}\right)^{p-1} \cdot n + O(n\log n) \end{aligned}$$

□

From the proof of Theorem 2.6 we obtain a somewhat stronger form of convergence using the (possible) choice $p = 3$ and $\varepsilon = \varepsilon_n \sim 1/\sqrt[4]{\log n},\ t = t_n \sim \sqrt{\log n}$

Corollary 2.7 *Under the assumptions of Theorem 2.6 choosing $t_n \sim \sqrt{\log n}$ and $\varepsilon_n \sim 1/\sqrt[4]{\log n}$ we obtain* *a)* $\sum_{n=1}^{\infty} P\left(|W_n(t_n)| \geq (1+\varepsilon_n)|T_n^*|\right) < \infty$ *;*

in particular $\frac{|W_n(t_n)|}{|T_n^*|} \longrightarrow 1$ *almost surely.*

b) The running time of the partmatch algorithm with $t = t_n$ *is bounded by* $O(n \log n)$.

Remarks:

a) For the practical application of the algorithm, the choice C using the bound in Theorem 2.5 should be much too pessimistic and also an optimal matching in the subrectangles could be replaced by a good branch and bound approximation.

b) The Euclidean metric is not used in an essential way and we could take any ℓ_p-metric on $\mathbb{R}^{\mathrm{d}}, p \geq 1$, $\mathrm{d} = 2$ for a similar result, the constant $\beta(L)$ of course will be dependent on the metric. For the ℓ_p-metric obviously the matching functional is quasiadditive.

c) Karp's (1977) partitioning algorithm and its analysis has been generalized by Halton and Terada (1982) to the d-dimensional case (for uniform random variables, cf. also Karp and Steele (1985)). It seems possible to analyse also similar extensions of the Euclidean matching to the d-dimensional case.

□

3 Weighted semi-matching in $\mathbb{R}^2$

For $x_1, \ldots, x_n \in [a,b] \subset \mathbb{R}^2, a < b$ let $L_s(x_1, \ldots, x_n)$ denote the semi-matching functional i.e. the optimal solution of the following linear program

$$(SM) : \min \sum_{\substack{i,j=1 \\ i<j}} x_{ij}\, e_{ij} \tag{3.1}$$

$$\sum_{j=1}^{i-1} x_{ji} + \sum_{j=i+1}^{n} x_{ij} = 1, \quad \forall \quad i = 1, \ldots, n$$

$$x_{ij} \geq 0, \quad \forall \quad i,j = 1, \ldots, n;\ i < j$$

where $e_{ij} = \|x_i - x_j\|$.

For an optimal solution of (SM) it is known that $x_{ij} \in \{0, \frac{1}{2}, 1\}, \forall\, i,j = 1, ..., n$; $i < j$ (cf. Lovasz and Plummer (1986), pg. 291).

If there is a circle of even number with all weights equal to $\frac{1}{2}$, say $u_1, \ldots, u_{2m}$, then for $\|u_1 - u_2\| + \ldots + \|u_{2m-1} - u_{2m}\| \leq \|u_2 - u_3\| + \ldots + \|u_{2m} - u_1\|$ replace the matching of $u_1, \ldots, u_{2m}$ by connecting u_{2i-1} and $u_{2i}, 1 \leq i \leq m$. This reduces the value of the matching. Therefore, we can assume for an optimal solution w.l.g., that it consists of pairs of points with weight one and of a set of circles with an uneven number of points with weight $\frac{1}{2}$.

It has been proved in Steele (1982) and Yukich (1995) that L_s is quasiadditive. Therefore, the convergence theorem of Redmond and Yukich (1993) applies to L_s.

3.1 The algorithm

As for Euclidean matching we divide the rectangle $[a, b]$ in smaller subrectangles with at most t_n points each. Then we solve the problem in the subrectangles exactly and get a semi-matching. We assume that no points have identical $x-$ or $y-$ coordinates and determine an upper bound t_n, $3 \leq t_n \leq n$ for the number of points in the subrectangles. Let $k_n := \left[\log_2 \frac{2(n-2)}{t_n-2}\right]$.

We divide $[a, b]$ into two rectangles with $\left[\frac{n}{2}\right]$ resp. $n - \left[\frac{n}{2}\right]$ points, cutting along the shorter side. Then we repeat this step k_n times with each of the rectangles. In this way we obtain in the first step two rectangles with at most $\left[\frac{n}{2}\right] + 1 \leq \frac{n+2}{2}$ points, in the second step 4 rectangles with at most $\left[\frac{n}{2}\right] + 1/2 + 1 \leq \frac{n+6}{4}$ points. Generally in the ℓ- th step we obtain 2^ℓ rectangles with at most $(n - 2 + 2^{\ell+1})/2^\ell$ points. After k_n steps each rectangle contains at most $\frac{n-2}{2^{k_n}} + 2 \leq (t_n - 2) + 2 = t_n$ points. In comparison to Euclidean matching an uneven number of points in the subrectangles does not cause a problem in semi-matching. Let $|U_n|$ denote the value of the matching consisting of optimal semi-matchings in all of the subrectangles.

Since the algorithm constructs $2^{k_n} \leq 2\frac{n-2}{t_n-2}$ rectangles with at most t_n points, and assuming that each subrectangle can be solved in time $\leq Dt_n^p$, we obtain the execution time

$$\leq 2\frac{n-2}{t_n-2} D t_n^p + O(n \log n). \tag{3.2}$$

3.2 Error analysis

For $F \in [a, b] \subset \mathbb{R}^2$ with $|F| = n$ and a rectangle Y in $[a, b]$ let $|S^*(Y)|$ denote the value of an optimal semi-matching of $F \cap Y$, $|S_n^*|$ the value of an optimal semi-matching of F and $|S_n^* \cap Y|$ the sum of the segments of an optimal semi-matching of F in Y.

Lemma 3.1 $|S^*(Y)| - |S_n^* \cap Y| \leq \frac{3}{4}\mathrm{per}(Y)$.

Proof: Consider at first those segments of an optimal semi-matching in Y which cut the boundary in $y_1, \ldots, y_{2m}$ and which lie on circles with weight $\frac{1}{2}$ and are ordered clockwise. W.l.g. let

$$\|y_1 - y_2\| + \|y_3 - y_4\| + \ldots + \|y_{2m-1} - y_{2m}\| \leq \|y_2 - y_3\| + \ldots + \|y_{2m} - y_1\|$$

We now construct a semi-matching joining all these points as follows. Start with y_1 and take the polygonsegment in Y to the next boundary point, say y_j. If j is even then we connect it to y_{j-1}, otherwise to y_{j+1}. If we reach in this way an already visited point, we match the points in Y along the constructed route and continue with a new unmatched point on the boundary. In the other case we start from y_{j-1} resp. y_{j+1} until we reach an already visited point. This leads to a semi-matching whose value is bounded above by $1/2$ times the length of the circle segments in Y plus $1/2$ times the length of the way on the boundary, i.e. $\frac{1}{2}\|y_1 - y_2\| + \ldots + \|y_{2m-1} - y_{2m}\| \leq \frac{1}{4}\mathrm{per}(Y)$.

If there are points in Y which are matched with weight 1 with points outside Y, then match them with weight 1 if there are two points of this kind. Otherwise match them in a circle of weight $\frac{1}{2}$. In both cases the value of this semi-matching is at most $\frac{1}{2}\mathrm{per}(Y)$.

Finally, an isolated point in Y can be included in a circle at the cost of at most $\mathrm{diam}(Y) \leq \frac{1}{2}\mathrm{per}(Y)$. Altogether

$$\begin{aligned} |S^*(Y)| - |S^* \cap Y| &\leq \frac{1}{4}\mathrm{per}(Y) + \frac{1}{2}\mathrm{per}(Y) \\ &= \frac{3}{4}\mathrm{per}(Y) \end{aligned}$$

□

Corollary 3.2 $|U_n| - |S_n^*| = O(\sqrt{\frac{n}{t_n}})$

Proof: This follows as in section 2, noting that

$$\begin{aligned} |U_n| &\leq |S_n^*| + \frac{3}{4}F_{k_n}(a,b) \text{ and} \\ F_{k_n}(a,b) &= O(2^{k_n/2}) = O\left(\sqrt{\frac{n-2}{t_n-2}}\right) = O\left(\sqrt{\frac{n}{t_n}}\right) \end{aligned}$$

□

As consequence we obtain

Theorem 3.3 *Let (X_i) be iid random points in $[a,b] \subset \mathbb{R}^2$ and let P^{X_1} be not singular w.r.t. λ^2. Then for some $D_1 > 0$ and any $\varepsilon > 0$ there exists a $t_n = t_n(\varepsilon)$ such that*

$$P(|U_n(t_n)| \leq (1+\varepsilon)|S_n^*|) = 1 + o(1).$$

With suitable implementation the running time is (for some $1 \leq p \leq 3$)

$$\leq D_1 \left(\frac{1}{\varepsilon^2}\right)^{p-1} n + O(n \log n).$$

Proof: The proof is analog to that of Theorem 2.6. Observe that by small shifts of the points we can avoid identical $x-$ or $y-$coordinates. □

The analogue of Corollary 2.7 concerning a.s. convergence therefore is also true for semi-matchings.

Acknowledgement: The authors thank M.Dyer and A. Frieze for pointing out their related paper of 1984.

References

[1] Dyer, M.E. and Frieze, A.M. A partitioning algorithm for minimum Euclidean matching. *Inf. Process. Lett.* **18**, 59-62 (1984)

[2] Halton, J.H. and Terada, R.: A fast algorithm for the Euclidean traveling salesman problem optimal with probability one. *SIAM J. Comp.* **11**, 28-46 (1982)

[3] Karp, R.M.: The probabilistic analysis of partitioning algorithms for the traveling salesman problem in the plane. *Math. Oper. Res.* **2**, 209-224 (1977)

[4] Karp, R.M. and Steele, J.M. : Probabilistic analysis of heuristics, In: The Traveling Salesman Problem. Eds. E.L. Lawler et al. 185-205 (1985)

[5] Ottmann, T. and Widmayer, P.: Algorithmen und Datenstrukturen, BI-Wiss. Verlag Mannheim (1990)

[6] Papadimitriou, C.H.: The probabilistic analysis of matching heuristics. In: Proc. 15th Ann. conf. comm. control computing, University of Illinois, Urbana, 363-378 (1978)

[7] Papadimitriou, C.H. and Steiglitz, K.: Combinatorial Optimization, Algorithms and Complexity. Prentice Hall N.J. (1982)

[8] Redmond, C. and Yukich, J.E.: Limit theorems and rates of convergence for Euclidean functionals. *Ann. Appl. Prob.* **4**, 1057-1073 (1994)

[9] Rhee, W.S.: A matching problem and subadditive Euclidean functionals. *Ann. Appl. Prob.* **3**, 794-801 (1993)

[10] Sachs, G.: Matching Probleme und Approximationsalgorithmen. Diplomarbeit Freiburg (1997)

[11] Steele, J.M.: Probability Theory and Combinatorial Optimization. *CBMS-NSF Regional Conference Series in Appl. Mathematics. SIAM.*

[12] Vajda, P.M.: Geometry helps in matching. *SIAM J. Comput.* **18**, 1201-1225 (1995)

[13] Yukich, J.E.: Quasi-additive Euclidean functionals. In: Discrete Probability and Algorithms. Eds. D. Aldous et al., 149-158 (1995)

Address: Institut für Mathematische Stochastik, Universität Freiburg, Eckerstr. 1, D-79104 Freiburg, Germany

e-mail: ruschen@stochastik.uni-freiburg.de