# On the weighted Euclidean matching problem in $\mathbb{R}^d$

Birgit Anthes and Ludger Rüschendorf[1]
University of Freiburg

March 3, 2000

**Abstract**

A partitioning algorithm for the Euclidean matching problem in $\mathbb{R}^d$ is introduced and analyzed in a probabilistic model. The algorithm uses elements from the fixed dissection algorithm of Karp and Steele (1985) and the Zig-Zag algorithm of Halton and Terada (1982) for the traveling salesman problem. The algorithm runs in expected time $n(\log n)^{p-1}$ and approximates the optimal matching in the probabilistic sense.

# 1 Introduction

Consider for $x_1, \ldots, x_n \in \mathbb{R}^d$, $d \geq 1$, the weighted Euclidean matching functional

$$L(x_1, \ldots, x_n) = \min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{m} |x_{\sigma(2i-1)} - x_{\sigma(2i)}| \tag{1.1}$$

where $\mathcal{S}_n$ is the set of all permutations of $\{1, \ldots, n\}$, $m = [\frac{n}{2}]$, and $\| \, \|$ is the Euclidean norm. $L$ is a quasi additive functional, i.e. $L$ is subadditive, continuous, and there exists an approximating superadditive functional $\widehat{L}$ with $\widehat{L} \leq L + 1$ on $[0,1]^d$. $\widehat{L}$ is defined by the corresponding boundary matching functional, which allows matching to boundary points (cf. Redmond and Yukich (1994)). Therefore, the general version of the Beardwood, Halton, and Hammersly (BHH) Theorem of Redmond and Yukich (1994) implies for any iid sequence $(X_i)$ on $[0,1]^d$ with $P^{X_i} = \mu$ that

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

in the sense of complete convergence, where $f$ is the density of the Lebesgue continuous part of $\mu$. In this paper we introduce a partitioning algorithm for the Euclidean matching problem and prove that it approximates the optimal solution in a probabilistic model and operates in expected running time $n(\log n)^{p-1}$. For the construction of the algorithm we combine elements of the fixed dissection algorithm

[1]Corresponding author
*1991 Mathematics Subject Classification*: 60F99, 68E99.
*Key words and phrases*: Euclidean matching problem, partitioning algorithm, probabilistic analysis of algorithms.

of Karp (1977), Karp and Steele (1985) and of the Zig-Zag algorithm of Halton and Terada (1982) as designed for the traveling salesman problem (TSP). In the two-dimensional case $p = 2$ an alternative partitioning algorithm analogue to Karp's (1977) bisection algorithm for the TSP where cells are constructed with a deterministic number of points has been introduced and analyzed in Rüschendorf and Sachs (1998) for general distributions. Dyer and Frieze (1984) considered a related fixed partitioning scheme for the uniform matching problem in $d = 2$. In his recent book on Euclidean optimization problems Yukich (1998) describes a general approach to an approximation result of subadditive Euclidean functionals (satisfying some conditions) by a partitioning heuristic as in Karp's TSP-heuristic (see Theorem 5.8). An essential implicit assumption in this approach (relation (5.15)) which describes the approximation of the heuristic by the independent sum of the values of the functional at the subcubes however is not discussed in this book. To establish a related estimate is the main work in this and also in the previous papers on this subject.

In contrast to the TSP which is an NP-problem the Euclidean matching can be formulated as linear programming problem and, therefore, has a polynomial running time exact solution of order $n^3$. (The fastest known exact solution for $d = 2$ is of the order $O(n^{2.5}(\log n)^4)$ (see Vaidya (1989)).) But the improvement of the order of the running time to $n(\log n)^{d-1}$ is considerable and of practical interest also for this combinatorial optimization problem. We analyze the partitioning algorithm in the random model where $\{X_i\}_{i \leq i \leq n}$ are independent random vectors uniformly distributed on the unit cube $Q = [0,1]^d$. In the final remark we extend this analysis to general distributions on $[0,1]^d$.

## 2 Introduction of the partitioning algorithm.

For $n$ points $x_1, \ldots, x_n \in [0,1]^d$ let $T_n$ denote an optimal Euclidean matching with value $|T_n| = L(x_1, \ldots, x_n)$. The construction of the partitioning algorithm $I_n$ uses the following steps:

a) Subdivision of the cube

$Q = [0,1]^d$ is divided into $M := m^d$ congruent cubes of side length $\frac{1}{m}$. Here

$$m = 2\left(\left[\frac{1}{2}\left(\frac{2n}{\log \sigma(n)}\right)^{\frac{1}{d}}\right] + 1\right) \tag{2.1}$$

with an increasing function $\sigma$ satisfying $\sigma(n) \to \infty$ as $n \to \infty$. Then

$$M = m^d \sim \frac{2n}{\log \sigma(n)}. \tag{2.2}$$

Any side of $Q$ is partitioned in $m^{d-1}$ congruent $(d-1)$-dimensional cubes and generates a cubic lattice with $M$ $d$-dimensional cubes $C_j$, $1 \leq j \leq M$.

b) Ordering of the $C_j$

The $M$ cubes $C_j$ are ordered by the Zig-Zag algorithm of Halton and Terada (1982) in a time of order $M$ resulting in a sequence $C_1 C_2 \cdots C_M$ (after relabeling of the cells). This ordering is established as follows: For each lattice vector $\mathbf{a} = (a_1, \dots, a_d)$ with $a_i \in L := \{0, 1, \dots, m-1\}$ define the cell

$$C(\mathbf{a}) := \left\{ x \in [0,1]^d; \quad \frac{a_i}{m} \le x_i < \frac{a_i + 1}{m}, \; 1 \le i \le d \right\}. \tag{2.3}$$

This defines a bijection of the $M$ lattice vectors $\mathbf{a}$ to the cells $C_i$, $1 \le i \le M$. So an ordering of the lattice vectors $\mathbf{a}$ induces an ordering of the cells $C_i$. Define

$$r_i = r_i(\mathbf{a}) = (-1)^{1 + a_1 + \ldots + a_{i-1}}, \quad 2 \le i \le d.$$

For each lattice vector $\mathbf{a}$ there exists at most one $t$ such that

$$\begin{array}{lll} a_i + r_i \notin L & \text{for} & i = t+1, \dots, d \\ a_t + r_t \in L & \text{and} & t \ge 3. \end{array} \tag{2.4}$$

If a $t$ with (2.4) exists, then define the successor of $\mathbf{a}$ as $\mathbf{a}' = \mathbf{a} + r_t \mathbf{e}_t$ where $a_i' = a_i$, $i \neq t$ and $a_t' = a_t + r_t$ and $\mathbf{e}_t$ is the unit vector in the $t$-th coordinate direction.

If no $t$ with (2.4) exists then define the successor $\mathbf{a}'$ of $\mathbf{a}$ by:

$$\begin{array}{lll} \mathbf{a}' = \mathbf{a} - \mathbf{e}_1 & \text{if} & (a_1 = 1, a_2 = 0) \text{ or } (a_1 > 1, a_2 \text{ even}) \\ \mathbf{a}' = \mathbf{a} + \mathbf{e}_1 & \text{if} & (a_1 = 0, a_2 = m-1) \text{ or } (0 < a_1 < m-1, a_2 \text{ odd}) \\ \mathbf{a}' = \mathbf{a} - \mathbf{e}_2 & \text{if} & (a_1 = 1, a_2 \neq 0 \text{ even}) \text{ or } (a_1 = m-1, a_2 \text{ odd}) \\ \mathbf{a}' = \mathbf{a} + \mathbf{e}_2 & \text{if} & (a_1 = 0, a_2 < m-1). \end{array} \tag{2.5}$$

Halton and Terada (1982) prove that this definition generates a tour through the set of lattice vectors $\mathbf{a}$ in $L^d$.

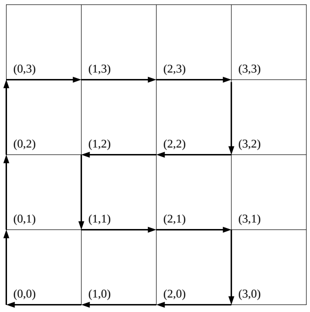


Figure 1: Zig-Zag tour in $d = 2$, $m = 4$.

c) Optimal matching within $C_j$

In the first step an optimal matching is constructed in $C_1$. If the number of points in $C_1$ is odd then the nonmatched point is matched with any point in $C_2$. Then the remaining points in $C_2$ are optimally matched. The next cell is $C_3$ and the ordering of the cells by the Zig-Zag algorithm is followed also when no points in a cell $C_i$ are present until finally an optimal matching is generated in $C_M$.

Let $I_n$ denote the matching generated by this partitioning algorithm and let $\ell(I_n)$ denote the value of this matching which consists of the sum of the lengths of the optimal matchings within $C_i$ and additionally the sum of the lengths of connections of possibly unmatched points between neighbouring cells.

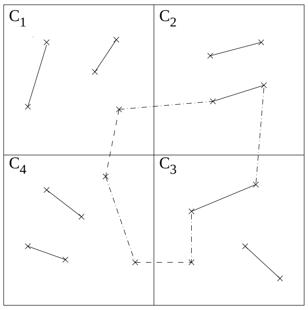


Figure 2: ×—× and ×- - -× describe $T_n$; ×—× and ×-·-·-× describes $I_n$

Note that two neighbouring blocks have $d-1$ edges of length $\frac{1}{m}$ and one edge of length $\frac{2}{m}$. The diameter is

$$\frac{1}{m}\left((d-1)+2^2\right)^{\frac{1}{2}} = \frac{1}{m}\sqrt{d+3}. \tag{2.6}$$

Therefore, these connections contribute at most $M \cdot \frac{1}{m}\sqrt{d+3}$ to the value of the matching $I_n$. If a cell is empty then the diagonal length is contained in the bound above.

# 3 Running time of the partitioning algorithm.

The running time of the partitioning algorithm is determined by two components. Firstly the running time of the Zig-Zag algorithm to produce the sequential order of the cubes $C_j$. By Halton and Terada (1982) this is of the order

$$O(M) = O\left(\frac{n}{\log \sigma(n)}\right). \tag{3.1}$$

In each cell $C_i$ an optimal matching is produced. We assume that we can solve an Euclidean matching problem with $k$ points in time $f(k) := Dk^p$ with constants

$p$, $D$. This is fulfilled for $p = 3$ by Papadimitriou and Steiglitz (1982, Theorem 11.3, Problem 14). Improvements of this order to $0(n^{2.5}(\log n)^4)$ are given in Vaidya (1989) to some more general class of geometric algorithms in dimension two. By the assumption that $X_1, \dots, X_n$ are independent, uniformly distributed on $[0,1]^d$ the number $n_i$ of points in $C_i$ are binomial distributed random variables $P(n_i = k) = \binom{n}{k}\left(\frac{1}{m^d}\right)^k\left(1-\frac{1}{m^d}\right)^{n-k}$. With the bound $f(k) = Dk^p$ for the optimal matching of $k$ points we obtain

$$Ef(k_i) \sim Dn^p\left(\frac{1}{m^d}\right)^p. \tag{3.2}$$

The optimal matching of the $M = m^d$ cubes therefore needs the expected running time

$$\sum_{i=1}^{m^d} Ef(k_i) \sim Dn\left(\frac{n}{m^d}\right)^{p-1}.$$

Choosing $\sigma(n) = \exp(2\log n)$ we obtain

$$\sum Ef(k_i) = O\left(n(\log n)^{p-1}\right). \tag{3.3}$$

The variance of $f(k_i)$ is of the order

$$\operatorname{Var} f(k_i) \sim D^2 n^{2p}\left(\frac{1}{m^d}\right)^{2p} \tag{3.4}$$

which implies by McDiarmid's inequality

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{m^d} f(k_i)\right) &\leq \sum \operatorname{Var} f(k_i) \\
&\sim D^2 n^{2p}\left(\frac{1}{m^d}\right)^{2p-1} \\
&= \left(O(\log n)^{2p-1}\right)
\end{aligned}$$

for the choice of $\sigma(n)$ as above. Since the Zig-Zag algorithm has a deterministic upper bound $O\left(\frac{n}{\log n}\right)$ we obtain.

**Theorem 3.1** *In the random uniform model the partitioning algorithm with the choice $\sigma(n) = \exp(2\log n)$ operates in expected running time $O\left(n(\log n)^{p-1}\right)$. The variance of the running time is $O\left(n(\log n)^{2p-1}\right)$.*

In comparison to the bisection algorithm as considered in Rüschendorf and Sachs (1998) for $d = 2$ we have for the fixed dissection algorithm in this paper only a random running time behaviour with expectation of order $O\left(n(\log n)^{p-1}\right)$. Under a stronger condition on $\sigma(n)$ Halton and Terada (1982) derive an a.s. running time behaviour of the corresponding algorithm in the TSP which could be also given in this context.

# 4 Approximation of the optimal matching.

In the first part of this section we compare the value of the partitioning algorithm with the value of the optimal matching and establish an upper estimate for $\ell(I_n)-|T_n|$, where $T_n$ denotes the optimal tour through $\{X_1,\dots,X_n\}$. For each face $F_{ij}$, $1\le j\le 2d$, of the cube $C_i$ we consider the set of marks $M_{ij}$ where an edge of $T_n$ which connects a point of $C_i\cap\{X_1,\dots,X_n\}$ to some point in a different cube intersects $F_{ij}$.

According to the construction of our partitioning algorithm an optimal matching is constructed in cell $C_i$ and a possibly unmatched point in $C_i$ is connected to an arbitrary point in the next nonempty cell according to the Zig-Zag tour through the cells. It is not obvious how to bound $\ell(I_n)$ from above by the length of the optimal matching $T_n$ plus some correction term. To that purpose we introduce a matching $\widehat{T}_n(i)$ of the points in cell $C_i$ which is induced by the matchings $T_{ij}$ of the marks $M_{ij}$ in the faces of $C_i$ by the following steps:

Firstly match those points in $C_i$ where the corresponding marks generated by the optimal matching $T_n$ in the faces are connected in the optimal matching $T_{ij}$ of $M_{ij}$ respectively are matched in the second step for the remaining marks in the faces.

Secondly use the part of the matching of points inside $C_i$ by the optimal matching $T_n$.

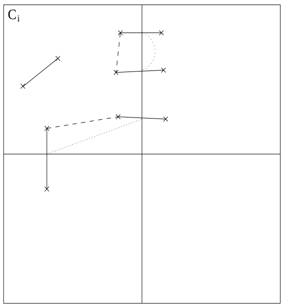


Figure 3: ×—× describes $T_n$, ×- - -× describes new parts of $\widehat{T}_n(i)$, $\cdots$ describes $T_{ij}$

Obviously the value of this matching $\widehat{T}_n(i)$ is an upper bound for the value of the optimal matching in cell $C_i$ used in our partitioning algorithm $I_n$. We will next prove that the length of the induced total matching $\widehat{T}_n=\sum \widehat{T}_n(i)$ can be estimated from above by the value of the optimal matching $T_n$ plus some terms of lower order.

By construction we obtain from the triangle inequality (see also Figure 3)

$$\begin{aligned}\ell(I_n) &\le |\widehat{T}_n| + M\frac{\sqrt{d+3}}{m} &(4.1)\\ &\le |T_n| + 2\sum_i\sum_j |T_{ij}| + M\frac{\sqrt{d+3}}{m} + \sum_{i=1}^{m^d} C_2(2d)^{\frac{d-1}{d}}\frac{1}{m}\end{aligned}$$

where the term $M\frac{\sqrt{d+3}}{m}$ is from the matching of the remaining points in the cells $C_i$ connected to the neighbour cells and where the last term comes from the matching of the points which remain unmatched in the faces in the first step with a universal constant $C_2$ and with the scaling $\frac{1}{m}$ for the side length (see Figure 3).

Here we use that by a simple induction argument (see Karp and Steele (1985)) one obtains that the length of an optimal matching of $n$ points in the $d$-dimensional unit cube is bounded above by

$$C_2 n^{\frac{d-1}{d}} \quad \text{for some } C_2 > 0. \tag{4.2}$$

Since the marks $M_{ij}$ are matched in the $(d-1)$-dimensional faces of side length $\frac{1}{m}$, we obtain

$$\ell(I_n) \le |T_n| + 2\sum_{i=1}^{m^d}\sum_{j=1}^{2d} C_2 |M_{ij}|^{\frac{d-2}{d-1}}\frac{1}{m} + M\frac{\sqrt{d+3}}{m} + \sum_{i=1}^{m^d} C_2 (2d)^{\frac{d-1}{d}}\frac{1}{m}. \tag{4.3}$$

Since $\sum_{i=1}^{m^d}\sum_{j=1}^{2d}|M_{ij}| \le 2n$ and since $x \to x^{\frac{d-2}{d-1}}$ is concave we obtain

$$\begin{aligned}
\ell(I_n) - |T_n| - M\frac{\sqrt{d+3}}{m} &\le 2\sum_{i=1}^{m^d}\sum_{j=1}^{2d} C_2\left(\frac{2n}{dm^d}\right)^{\frac{d-2}{d-1}}\frac{1}{m} + C_2 m^d (2d)^{\frac{d-1}{d}}\frac{1}{m} \\
&= K_1 (2n)^{\frac{d-2}{d-1}} m^{\frac{1}{d-1}} + K_2 m^{d-1} \\
&\sim K_1 (2n)^{\frac{d-2}{d-1}}\left(\frac{2n}{\log\sigma(n)}\right)^{\frac{1}{d-1}} + K_2\left(\frac{2n}{\log\sigma(n)}\right)^{\frac{d-1}{d}} \\
&= O\left(n^{\frac{d-1}{d}}(\log\sigma(n))^{-\frac{1}{d(d-1)}}\right).
\end{aligned}$$

With $M = m^d \sim \frac{2n}{\log\sigma(n)}$ we obtain

$$\ell(I_n) \le |T_n| + O\left(n^{\frac{d-1}{d}}(\log\sigma(n))^{-\frac{d-1}{d}}\right) + O\left(n^{\frac{d-1}{d}}(\log\sigma(n))^{-\frac{1}{d(d-1)}}\right).$$

Thus the value of the partitioning algorithm $I_n$ is bounded from above by the value of the optimal matching plus some deterministic bound of lower order.

**Theorem 4.1** *The length of the partitioning algorithm for Euclidean matching of $n$ points $X_1, \dots, X_n$ in $[0,1]^d$ is estimated from above by*

$$\ell(I_n) \le |T_n| + O\left(n^{\frac{d-1}{d}}(\log\sigma(n))^{-\frac{1}{d(d-1)}}\right). \tag{4.4}$$

To estimate the relative error compared to the optimal matching $T_n$ we use the following lemma.

**Lemma 4.2** *There exist constants $A$, $C > 0$, $0 < \varrho < 1$ such that in the random data model $X_1, \dots, X_n$*

$$P\left(|T_n| < An^{\frac{d-1}{d}}\right) \leq C\varrho^n, \quad \forall n \tag{4.5}$$

**Proof:** Since $|T_n|$ is a quasiadditive functional one obtains from Rhee (1993)

$$P(|T_n - ET_n| \geq t) \leq C \exp\left(-\frac{1}{Cn}\left(\frac{t}{C_1}\right)\right)^{\frac{2d}{d-1}}$$

for some constants $C, C_1 > 0$. Further for $\{X_i\}$ independent and uniformly on $[0,1]^d$ distributed one obtains

$$E \min\{|X_i - X_j|; \quad i,j \leq n, \ i \neq j\} \geq cn^{-\frac{1}{d}}$$

for some $c > 0$. This implies the lower bound $ET_n \geq \frac{n}{2}cn^{-\frac{1}{d}} = \frac{c}{2}n^{\frac{d-1}{d}}$. Therefore,

$$\begin{aligned}
P\left(T_n < An^{\frac{d-1}{d}}\right) &= P\left(T_n - ET_n < An^{\frac{d-1}{d}} - ET_n\right) \\
&\leq P\left(ET_n - T_n > \left(\frac{c}{2} - A\right) n^{\frac{d-1}{d}}\right) \\
&\leq P\left(|ET_n - T_n| > Bn^{\frac{d-1}{d}}\right) \\
&\leq C \exp\left(-\frac{1}{Cn}\left(\frac{Bn^{\frac{d-1}{d}}}{C_1}\right)^{\frac{2d}{d-1}}\right) \\
&= C\exp(-Dn) = C\varrho^n, \quad \text{with } \varrho = e^{-D} \\
&\phantom{=} \text{(all constants are } > 0).
\end{aligned}$$

□

As a result we obtain complete convergence of $\frac{\ell(I_n)}{T_n}$ to 1.

**Theorem 4.3** *In the random data model the relative length of the matching of the partitioning algorithm to the optimal length converges completely to one*

$$\frac{\ell(I_n)}{|T_n|} \xrightarrow{c} 1,$$

*i.e. for all $\varepsilon > 0$ holds $\sum_{n=1}^{\infty} P\left(\frac{\ell(I_n)}{|T_n|} \geq 1 + \varepsilon\right) < \infty$.*

**Proof:** Using the estimates in Theorem 4.1, Lemma 4.2 we obtain:

$$\begin{aligned}
P\left(\frac{\ell(I_n)}{|T_n|} \geq 1 + \varepsilon\right) &\leq P\left(|T_n| < An^{\frac{d-1}{d}}\right) + P\left(\ell(I_n) - |T_n| \geq \varepsilon An^{\frac{d-1}{d}}\right) \\
&\leq C\varrho^n + P(\ell(I_n) - |T_n| \geq \varepsilon An^{\frac{d-1}{d}}.
\end{aligned}$$

Therefore,

$$\sum_{n=1}^{\infty} P\left(\frac{\ell(I_n)}{|T_n|} \geq 1+\varepsilon\right) \leq C \sum_{n=1}^{\infty} \varrho^n + \sum_{n=1}^{\infty} P\left(\ell(I_n) - |T_n| \geq \varepsilon A n^{\frac{d-1}{d}}\right) < \infty$$

since the last sum has only finitely many nonzero summands. □

**Remark: (General distributions.)** The normalized minimum matching functional converges also for general distribution (see (1.2)). The estimate in Theorem 4.1 is valid for any set of deterministic points in $[0,1]^d$. Therefore, using (1.2) the normalized partitioning algorithm converges completely to the same constant $\beta(L)\int f(x)^{\frac{d-1}{d}}\lambda^d(dx)$ as well. This implies complete convergence of the quotient $\frac{\ell(I_n)}{|T_n|} \longrightarrow 1$ (see also Yukich (1998, proof of Theorem 5.8)). Therefore, we obtain an approximation of the matching functional by the partitioning algorithm in the probabilistic sense fore general distributions on $[0,1]^d$ with nontrivial Lebesgue continuous part. □

As a result of the analysis we obtain that the partitioning algorithm yields an approximative optimal matching and uses expected running time $O(n(\log n)^{p-1})$. This improves considerably on the running time of exact deterministic algorithms.

## References

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

Halton, J. H. and R. Terada (1982). A fast algorithm for the Euclidean traveling salesman problem optmal with probability one. *SIAM J. Comp. 11*, 28–46.

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

Karp, R. M. and J. Steele (1985). Probabilistic analysis of heuristics. In E. Lawler et al. (Eds.), *The Traveling Salesman Problem*, pp. 185–205.

Papadimitriou, C. H. and K. Steiglitz (1982). *Combinatorial Optimization, Algorithms and Complexity*. Prentice Hall N.J.

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

Rhee, W. S. (1993). A matching problem and subadditive Euclidean functionals. *Ann. Appl. Prob. 3*, 794–801.

Rüschendorf, L. and G. Sachs (1998). On a partitioning algorithm for weighted Euclidean matching. Preprint, to appear in: J. Appl. Prob.

Vaidya, P. M. (1989). Geometry helps in matching. *SIAM J. Comp. 18*, 1201–1225.

Yukich, J. E. (1998). *Probability Theory of Classical Euclidean Optimization Problems.* Lect. Notes Math. 1675. Springer.

Institut für Mathematische Stochastik
Universität Freiburg
Eckerstr. 1
79104 Freiburg
Germany
E-mail: ruschen@stochastik.uni-freiburg.de