Average Case Quality of Partitions Induced by the Lebesgue Indexing

Jens-Michael Wierum

Paderborn Center for Parallel Computing, PC^2
Fürstenallee 11, 33102 Paderborn, Germany
jmwie@upb.de
www.upb.de/pc2/

PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Technical Report PC^2
TR-002-01
November 2001

Abstract. This paper presents the quality of partitions induced by the Lebesgue curve in average case. The surface that surrounds an arbitrary index range is compared with the optimal partition in the grid, i. e. the square. The upper bound on the surface is asymptotically $\frac{5}{2 \cdot \sqrt{3}}$ times the optimal size.

1 Introduction

Data structures for maintaining sets of multidimensional points play an important role in many areas of computational geometry. While for example Voronoi diagrams have been established for efficient requests on neighbourhood relationships, data structures based on space-filling curves are often used for requests on axis aligned bodies of arbitrary size. The aim of the requests is to find all points located in such multidimensional intervals. Those types of requests are needed in many applications like N-body simulations [12], image compression and browsing [10, 4], databases [2], and contact search in finite element analysis [5]. An overview on this and other techniques for range searching in computational geometry is given in [1]. Space-filling curves have other locality properties which are e. g. useful in parallel finite element simulations [3, 7].

Space-filling curves are geometric representations of bijective mappings $M : \{1, \ldots, N^m\} \to \{1, \ldots, N\}^m$. The curve M traverses all N^m cells in the m-dimensional grid of size N. An (historic) overview on space-filling curves is given in [11]. Experimental work and theoretical analysis have shown, that algorithms based on space-filling curves behave well on most inputs, while they are unfitted for some special situations. Therefore, an analysis for the average case is often more important than for the worst case.

Due to the varying requirements on the *locality* properties, different metrics have been used to qualify, compare, and improve space-filling curves. A major metric for the analysis of the locality of space-filling curves is the ratio of index interval to maximum distance within this index range. Results are published for different indexing schemes like Hilbert, Lebesgue, and H-indexing for Manhattan metric, Euclidean metric, and maximum metric [6, 9].

Other examinations concentrate on the number of index intervals which have to be determined for a given request region. Sharp results are given in [2] for squares in two-dimensional space. The costs for arbitrary shaped regions is discussed in [8].

Here we examine the surface of a partition which is induced by an interval of a space-filling curve. Practical results of this relationship for uniform grids and unstructured meshes can be found in [13]. We define a *quality coefficient* which represents a normed value for the quality of the induced partition in an uniform grid of size $N \times N$. We use the shape of an optimal partition, the square, as a reference:

Definition 1 (quality coefficient). *Let curve be an indexing scheme, p an index range, $S(p)$ the surface of a partition and $V(p) = |p|$ the size (volume) of it. $C^{curve}(p)$ defines the* quality coefficient *of the partition given by index range p:*

$$C^{curve}(p) = \frac{S^{curve}(p)}{4 \cdot \sqrt{V(p)}} \tag{1}$$

This formulation can be extended to a quality coefficient *of an indexing scheme:*

$$C^{curve}_{\text{avg}} = \text{avg}_p\{C^{curve}(p)\} \tag{2}$$

Definition 1 implies that $C(p) \geq 1$ for all indexing schemes.

2 Lebesgue Curves

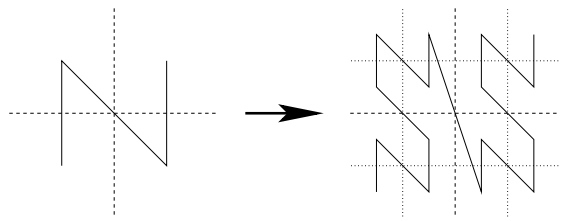

Fig. 1. Production rule for Lebesgue indexing.

Figure 1 illustrates the recursive definition of the Lebesgue indexing. The resulting curve is also known as bit interleaving or z-code. In the following the edges of cells are given a level, depending on the step they were introduced during the recursive construction. The lines of the final step are of level 0. In the shown example dashed lines are of level 1, dotted lines of level 0. It is obvious that an arbitrary edge is of level l with an asymptotic probability of $2^{-(l+1)}$.

Let $N = n \times n$ be the size of the grid and $L = \log n$ the number of levels. Consider a partition of size V and a vertical level l. For the estimation of the number of inner edges we examine a region of size $2^{2(l+1)}$.

Example 1. Figure 2 shows the corresponding region for $l = 1$ and the order within it. Table 3 describes the dependency of partition size, position of first cell and number of inner edges on level 1 lines (red line in left figure). The left column states the position of the first cell relative to the level 1 line. Depending on this position is the minimal volume printed which is needed to get one, two, three, or four inner edges. Is 4 the first cell of the partition we need 21 cells to get 3 inner edges. Let the content of the table be given by $\delta(i, j)$, with i being the column (=*first cell*) and j the row (=*number of inner edges at this level*).

For cells close to the considered level (marked with green circles in the example) holds $\delta(i, 1) = 7$ $(= 4 \cdot \frac{4^l - 1}{3} + 2 + 1$, 1 cell more than the distance to the neighbouring cell on the right hand at level $l)$. Furthermore, for those cells $\delta(i, j) = \delta((i+1) \bmod 16, j-1)+1$ is true. The minimal partition size for the j-th inner edge is 1 larger than the minimal partition size for the $j-1$-th inner edge of the following cell. For all other cells holds $\delta(i, j) = \delta((i + 1) \bmod 16, j) + 1$. The partition size is directly dependent on the value of the following cell.

The formulation of δ only includes limited partition sizes. It can be extended to arbitrary partitions with the following formulation:

$$\gamma(V, l, i, j) = \left\lfloor \frac{V + 4^{(l+1)} - \delta(i, j)}{4^{(l+1)}} \right\rfloor \tag{3}$$

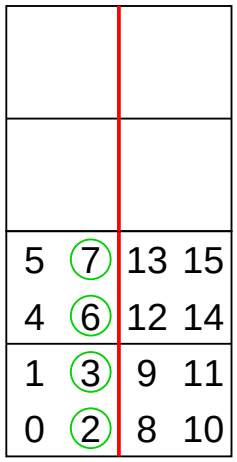

Fig. 2. Basis for the determination of inner edges for vertical lines of level 1.

	1.	2.	3.	4.
0	9	10	13	14
1	8	9	12	13
2	7	8	11	12
3	7	10	11	22
4	9	10	21	22
5	8	9	20	21
6	7	8	19	20
7	7	18	19	22
8	17	18	21	22
9	16	17	20	21
10	15	16	19	20
11	14	15	18	19
12	13	14	17	18
13	12	13	16	17
14	11	12	15	16
15	10	11	14	15

Fig. 3. Inner edges depending on partition size and position of first cell.

γ denotes the number of blocks, in which we expect j inner edges, depending on the relative position of the first cell.

The frequency of occurrence of a relative position to be first cell is given by:

$$\varphi(L,V,l,i) = \left\lfloor \frac{4^L - V + 4^{(l+1)} - i}{4^{(l+1)}} \right\rfloor \tag{4}$$

It depends on the volume of the partition and the size of the underlying grid given by L (number of recursive steps needed for generation).

From this follows for the number of inner edges in a partition of size V in average case:

$$E_l = \frac{\sum_{i=0}^{2^{2(l+1)}} \left(\varphi(L,V,l,i) \cdot \sum_{j=1}^{2^{l+1}} \gamma(V,l,i,j) \right)}{\sum_{i=0}^{2^{2(l+1)}} \varphi(L,V,l,i)} \tag{5}$$

Still missing is the description of special cells near the level l line in the definition of δ. Let us denote them as positions i_1, i_2, ... At first we analyze the distance between those positions ($\Delta_k = i_{k+1} - i_k$). Due to the regular structure of the Lebesgue curve it can be generated easily as:

$$\begin{aligned} \Delta_k &= (1,3,1,11,1,3,1,43,1,3,1,11,\ldots) \\ &= 2\frac{4^i - 1}{3} + 1 \ ; \ \text{for } k = 2^i + j \cdot 2^{i+1} \ , \ i,j \in I\!N_0 \end{aligned} \tag{6}$$

That means, the value $2\frac{4^i-1}{3}+1$ is given at position 2^i first, recurring at each 2^{i+1}-th position.

The relative position of the first cell (i_1), which depends on the investigated level l, is given by:

$$\Delta_0 = 2\frac{4^l-1}{3} \tag{7}$$

The inductive definition of δ shows, that the resulting values depend on the distance to the j-th special cell. This distance is also given by the sequence Δ_k. For this purpose it is necessary to split the summation on relative positions in intervals, which are induced by those special cells. Let $i_I = \sum_{k=0}^{I-1} \Delta_k$ be the relative positions of special cells. We get three summations for $i \leq i_1$, $i_1 < i \leq i_{2^{l+1}}$, and $i_{2^{l+1}} < i$. For simplicity of terms we set $\Phi(L,V,l) = \sum_{i=0}^{2^{2(l+1)}} \varphi(L,V,l,i)$. It holds:

$$\begin{aligned}
E_l &= \sum_{i=0}^{i_1} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma(V,l,i,j) \right) \\
&\quad + \sum_{I=1}^{2^{(l+1)}-1} \sum_{i=i_I+1}^{i_{(I+1)}} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma(V,l,i,j) \right) \\
&\quad + \sum_{i=i_{2^{(l+1)}}+1}^{2^{2(l+1)}} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma(V,l,i,j) \right) \\
&= \sum_{i=0}^{i_1} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma_1(V,l,I,i,j) \right) \\
&\quad + \sum_{I=1}^{2^{(l+1)}-1} \sum_{i=i_I+1}^{i_{(I+1)}} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma_2(V,l,I,i,j) \right) \\
&\quad + \sum_{i=i_{2^{(l+1)}}+1}^{2^{2(l+1)}} \left(\frac{\varphi(L,V,l,i)}{\Phi(L,V,l)} \cdot \sum_{j=1}^{2^{l+1}} \gamma_3(V,l,I,i,j) \right)
\end{aligned} \tag{8}$$

with

$$\begin{aligned}
\gamma_1(V,l,I,i,j) &= \left\lfloor \frac{V + 4^{(l+1)} - (R_l + i_j - i)}{4^{(l+1)}} \right\rfloor \\
\gamma_2(V,l,I,i,j) &= \left\lfloor \frac{V + 4^{(l+1)} - \left(R_l + (i_{(j+I) \bmod 2^{l+1}} - i) \bmod 4^{(l+1)}\right)}{4^{(l+1)}} \right\rfloor \\
&= \left\lfloor \frac{V + 4^{(l+1)} - \left(R_l + i_{(j+I)} - i\right)}{4^{(l+1)}} \right\rfloor
\end{aligned}$$

$$\gamma_3(V,l,I,i,j) = \left\lfloor \frac{V + 4^{(l+1)} - \left(R_l + i_j + 4^{(l+1)} - i\right)}{4^{(l+1)}} \right\rfloor \quad .$$

The exact formulation for E_l does not allow any simplifications. We can only reduce the complexity by an asymptotic examination: For very large grids $\varphi(L,V,l,i)$ converges to $\varphi(L,V,l,0)$. I. e. it goes independent from the relative position of the partition's first cell.

$$\lim_{L\to\infty} \left(\frac{\varphi(L,V,l,i)}{\sum_{i=0}^{2^{2(l+1)}} \varphi(L,V,l,i)} \right) = \frac{1}{2^{2(l+1)}} \tag{9}$$

An analysis of the recursive construction of δ shows, that each value in interval $\left[4\frac{4^l-1}{3}+3, 4\frac{4^{l+1}-1}{3}+2\right]$ is given 2^{l+1} times $(4\frac{4^{l+1}-1}{3}+2-(4\frac{4^l-1}{3}+2) = 4^{l+1})$. From (5) follows:

$$\begin{aligned}
E_l &= \frac{1}{2^{2(l+1)}} \cdot \sum_{i=0}^{2^{2(l+1)}} \sum_{j=1}^{2^{l+1}} \gamma(V,l,i,j) \\
&= \frac{1}{2^{2(l+1)}} \cdot 2^{l+1} \cdot \sum_{i=4\frac{4^l-1}{3}+3}^{4\frac{4^{l+1}-1}{3}+2} \left\lfloor \frac{V + 2^{2(l+1)} - i}{2^{2(l+1)}} \right\rfloor \\
&= \frac{2^{l+1}}{2^{2(l+1)}} \cdot \left(V - 4\frac{4^l-1}{3} - 2\right) \\
&= \frac{1}{2^{l+1}} \cdot \left(V - 4\frac{4^l-1}{3} - 2\right)
\end{aligned} \tag{10}$$

A comparable analysis for horizontal edges of level h leads to

$$E_h = \frac{1}{2^{h+1}} \cdot \left(V - 2\frac{4^h-1}{3} - 1\right) \tag{11}$$

Now we are able to subtract all inner edges twice from the total number of cell's edges $(V \cdot 4)$ to get the size of the surface S:

$$S = 4V - 2\sum_{i=0}^{\infty} \max\{E_l, 0\} - 2\sum_{i=0}^{\infty} \max\{E_h, 0\} \tag{12}$$

Lemma 1. *For the surface of a partition induced by the Lebesgue curve holds in average case:*

$$S \leq \frac{3}{2^k}V + \frac{8}{3}2^k - \frac{5}{3}\frac{1}{2^k} \quad \text{for } V \in \left[2\frac{4^k-1}{3}+1, 4\frac{4^k-1}{3}+2\right[\text{ and}$$

$$S \leq \frac{4}{2^k}V + 2\cdot 2^k - 2\cdot\frac{1}{2^k} \quad \text{for } V \in \left[4\frac{4^{k-1}-1}{3}+2, 2\frac{4^k-1}{3}+1\right[, \text{with } k \in \mathbb{N} \ .$$

Proof. To be assure to accumulate only non-negative values, we have to split the evaluated space into two classes of intervals: $I_1 = [2\frac{4^k-1}{3}+1, 4\frac{4^k-1}{3}+2[$ and $I_2 = [4\frac{4^{k-1}-1}{3}+2, 2\frac{4^k-1}{3}+1[$. The size of the surface for all partitions p in intervals of class I_1 is given by:

$$\begin{aligned} S &\leq 4V - \sum_{i=0}^{k}\frac{1}{2^i}\left(V - 2\frac{4^i-1}{3} - 1\right) - \sum_{i=0}^{k-1}\frac{1}{2^i}\left(V - 4\frac{4^i-1}{3} - 2\right) \\ &= 4V - \left(V - \frac{1}{3}\right)\left(2 - \frac{1}{2^k}\right) + \frac{2}{3}\left(2^{k+1}-1\right) \\ &\quad - \left(V - \frac{2}{3}\right)\left(2 - \frac{1}{2^{k-1}}\right) + \frac{4}{3}\left(2^{k-1}\right) \\ &= \frac{3}{2^k}V + \frac{8}{3}2^k - \frac{5}{3}\frac{1}{2^k} \end{aligned} \tag{13}$$

For interval class I_2 holds:

$$\begin{aligned} S &\leq 4V - \sum_{i=0}^{k-1}\frac{1}{2^i}\left(V - 2\frac{4^i-1}{3} - 1\right) - \sum_{i=0}^{k-1}\frac{1}{2^i}\left(V - 4\frac{4^i-1}{3} - 2\right) \\ &= 4V - \left(V - \frac{1}{3}\right)\left(2 - \frac{1}{2^{k-1}}\right) + \frac{2}{3}\left(2^{k}-1\right) \\ &\quad - \left(V - \frac{2}{3}\right)\left(2 - \frac{1}{2^{k-1}}\right) + \frac{4}{3}\left(2^{k-1}\right) \\ &= \frac{4}{2^k}V + 2\cdot 2^k - 2\cdot\frac{1}{2^k} \end{aligned} \tag{14}$$

□

Theorem 1. *The quality coefficient of the average case for the Lebesgue indexing scheme is less than or equal to* $\frac{5}{2\cdot\sqrt{3}}$.

Proof. For the determination of the upper bound for the average case the limits of the intervals of classes I_1 and I_2 has to be examined. For the upper limit of I_1 we get:

$$V = 4\frac{4^{k-1}-1}{3} + 2 \Leftrightarrow \sqrt{3V-2} = 2^k \tag{15}$$

Using the result of Lemma 1 gives:

$$S \leq \frac{4}{2^k}V + 2\left(2^k - \frac{1}{2^k}\right) = \frac{4V-2}{\sqrt{3V-2}} + 2 \cdot \sqrt{3V-2} = \frac{10V-6}{\sqrt{3V-2}}$$
$$\leq \frac{10}{\sqrt{3}}\sqrt{V} \tag{16}$$

The corresponding lower limit of I_1 (eq. to upper limit of I_2) results in:

$$S \leq \frac{7\sqrt{2}}{\sqrt{3}} \cdot \sqrt{V} \tag{17}$$

The surface size is obviously larger in the first case. The quality coefficient for the average case is:

$$C_{\text{avg}}^{\text{Lebesgue}} \leq \frac{10}{\sqrt{3}} \cdot \frac{1}{4} = \frac{5}{2 \cdot \sqrt{3}} \approx 1.44 \tag{18}$$

□

References

1. Pankaj K. Agarwal and Jeff Erickson. Geometric range searching and its relatives. *Advances in Discrete and Computational Geometry*, 1998.
2. Tetsuo Asano, Desh Ranjan, Thomas Roos, Emo Welzl, and Peter Widmayer. Space-filling curves and their use in the design of geometric data structures. *Theoretical Computer Science*, 181:3–15, 1997.
3. Jörn Behrens and Jens Zimmermann. Parallelizing an unstructured grid generator with a space-filling curve approach. In A. Bode, T. Ludwig, W. Karl, and R. Wismüller, editors, *Euro-Par 2000*, LNCS 1900, pages 815–823. Springer, 2000.
4. Scott Craver, Boon-Lock Yeo, and Minerva Yeung. Multilinearization data structure for image browsing. In *SPIE – The International Society for Optical Engineering*, pages 155–, 1998.
5. Ralf Diekmann, Jan Hungershöfer, Michael Lux, Lars Taenzer, and Jens-Michael Wierum. Using space filling curves for efficient contact searching. In *Proc. IMACS*, 2000.
6. C. Gotsman and M. Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797, May 1996.
7. Michael Griebel and Gerhard Zumbusch. Parallel multigrid in an adaptive PDE solver based on hashing and space-filling curves. *Parallel Computing*, 25:827–843, 1999.
8. Bongki Moon, H. V. Jagadish, Christos Faloutsos, and Joel H. Saltz. Analysis of the clustering properties of the Hilbert space-filling curve. *IEEE Transaction on Knowledge and Data Engineering*, 13(1), Jan/Feb 2001.
9. Rolf Niedermeier, Klaus Reinhardt, and Peter Sanders. Towards optimal locality in mesh-indexings. *LNCS 1279*, 1997.
10. Renato Pajarola and Peter Widmayer. An image compression method for spatial search. *IEEE Transactions on Image Processing*, 9(3):357–365, 2000.
11. H. Sagan. *Space Filling Curves*. Springer, 1994.

12. Shang-Hua Teng. Provably good partitioning and load balancing algorithms for parallel adaptive N-body simulation. *SIAM Journal on Scientific Computing*, 19(2):635–656, 1998.
13. Gerhard Zumbusch. On the quality of space-filling curve induced partitions. *Zeitschrift für Angewandte Mathematik und Mechanik*, 81, SUPP/1:25–28, 2001.