On the Locality Properties of Space-Filling Curves

H.K. Dai and H.C. Su*

Computer Science Department, Oklahoma State University
Stillwater, Oklahoma 74078, U. S. A.
{dai, suh}@cs.okstate.edu

Abstract. A discrete space-filling curve provides a linear traversal or indexing of a multi-dimensional grid space. We present an analytical study of the locality properties of the m-dimensional k-order discrete Hilbert and z-order curve families, $\{H_k^m \mid k = 1, 2, \ldots\}$ and $\{Z_k^m \mid k = 1, 2, \ldots\}$, respectively, based on the locality measure L_δ that cumulates all index-differences of point-pairs at a common 1-normed distance δ. We derive the exact formulas for $L_\delta(H_k^m)$ and $L_\delta(Z_k^m)$ for $m = 2$ and arbitrary δ that is an integral power of 2, and $m = 3$ and $\delta = 1$. The results yield a constant asymptotic ratio $\lim_{k\to\infty} \frac{L_\delta(H_k^m)}{L_\delta(Z_k^m)} > 1$, which suggests that the z-order curve family performs better than the Hilbert curve family over the considered parameter ranges.

1 Preliminaries

Space-filling curves have many applications in algorithms, databases, and parallel computation, in which linearization techniques of multi-dimensional arrays or grids are needed. Sample applications include heuristics for Hamiltonian traversals, multi-dimensional space-filling indexing methods, image compression, and dynamic unstructured mesh partitioning.

For positive integer n, denote $[n] = \{1, 2, \ldots, n\}$. An m-dimensional (discrete) space-filling curve of length n^m is a bijective mapping $C : [n^m] \to [n]^m$, thus providing a linear indexing/traversal or total ordering of all grid points in $[n]^m$. An m-dimensional grid is said to be of order k if it has side-length $n = 2^k$; a space-filling curve has order k if its codomain is a grid of order k. The generation of a sequence of multi-dimensional space-filling curves of successive orders usually follows a recursive framework (on the dimensionality and order), which results in a few classical families, such as Gray-coded curves, Hilbert curves, Peano curves, and z-order curves (see, for examples, [AN00] and [MJFS01]).

Denote by H_k^m and Z_k^m an m-dimensional Hilbert and z-order, respectively, space-filling curve of order k. Figure 1 illustrates the recursive constructions of H_k^2 and Z_k^2 for $m = 2$ and $k = 1, 2$, and $m = 3$ and $k = 1$.

* Current address: Department of Computer Science, Arkansas State University, State University, Arkansas 72467, U.S.A.

T. Ibaraki, N. Katoh, and H. Ono (Eds.): ISAAC 2003, LNCS 2906, pp. 385–394, 2003.

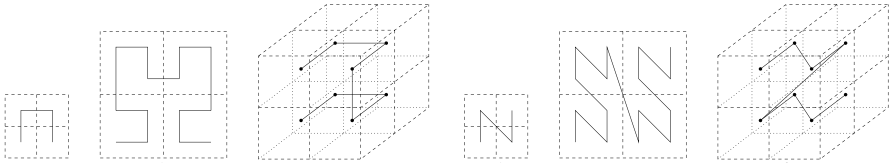


Fig. 1. Recursive constructions of Hilbert and z-order curves of higher order (H_k^m and Z_k^m, respectively) by interconnecting symmetric subcurves, via reflection and rotation, of lower order (H_{k-1}^m and Z_{k-1}^m, respectively): (a) H_1^2; (b) H_2^2; (c) H_1^3; (d) Z_1^2; (e) Z_2^2; (f) Z_1^3.

We measure the applicability of a family of space-filling curves based on their common structural characteristics, which are informally described as follows. Locality preservation reflects proximity between the grid points of $[n]^m$, that is, close-by points in $[n]^m$ are mapped to close-by indices/numbers in $[n^m]$, or vice versa. Clustering performance measures the distribution of continuous runs of grid points (clusters) over identically shaped subspaces of $[n]^m$, which can be characterized by the average number of clusters and the average inter-cluster distance (in $[n^m]$) within a subspace.

Empirical and analytical studies of clustering and inter-clustering performances of various low-dimensional space-filling curves have been reported in the literature (see [MJFS01] for details). Generally, the Hilbert and z-order curve families exhibit good performance in this respect.

Jagadish [Jag97] derives exact formulas for the mean numbers of clusters over all rectangular 2×2 and 3×3 subgrids of an H_k^2-structural grid space. Moon, Jagadish, Faloutsos, and Saltz [MJFS01] prove that in a sufficiently large m-dimensional H_k^m-structural grid space, the mean number of clusters over all rectilinear polyhedral queries with surface area $S_{m,k}$ approaches $\frac{1}{2}\frac{S_{m,k}}{m}$ as k approaches ∞. They also extend the work in [Jag97] to obtain the exact formula for the mean number of clusters over all rectangular $2^q \times 2^q$ subgrids of an H_k^2-structural grid space.

The space-filling index structure can support efficient query processing (such as range queries) provided that we minimize the average number of external fetch/seek operations, which is related to the clustering statistics. Asano, Ranjan, Roos, Welzl, and Widmayer [ARR+97] study the optimization of range queries over space-filling index structures, which aims at minimizing the number of seek operations (not the number of block accesses) — tradeoff between seek time to proper block (cluster) and latency/transfer time for unnecessary blocks (inter-cluster gap). Good bounds on inter-clustering statistics translate into good bounds on the average tolerance of unnecessary block transfers.

Dai and Su [DS03] obtain the exact formulas for the following three statistics for H_k^2 and Z_k^2: (1) the summation of all inter-cluster distances over all $2^q \times 2^q$ query subgrids, (2) the universe mean inter-cluster distance over all inter-cluster gaps from all $2^q \times 2^q$ subgrids, and (3) the mean total inter-cluster distance over all $2^q \times 2^q$ subgrids.

A few locality measures have been proposed and analyzed for space-filling curves in the literature. Generally, Hilbert and z-order curve families achieve good locality performances.

Denote by d and d_p the Euclidean and p-normed metric (Manhattan ($p = 1$), Euclidean ($p = 2$), and maximum metric ($p = \infty$)), respectively. Let $\mathcal{C}$ denote a family of m-dimensional curves of successive orders. For quantifying the proximity preservation of close-by points in the m-dimensional space $[n]^m$, Pérez, Kamata, and Kawaguchi [PKK92] employ an average locality measure:

$$L_{\mathrm{PKK}}(C) = \sum_{i,j \in [n^m] | i<j} \frac{|i-j|}{d(C(i), C(j))} \text{ for } C \in \mathcal{C},$$

and provide a hierarchical construction for a 2-dimensional $\mathcal{C}$ with good but suboptimal locality with respect to this measure.

Mitchison and Durbin [MD86] use a more restrictive locality measure parameterized by q:

$$L_{\mathrm{MD},q}(C) = \sum_{i,j \in [n^m] | i<j \text{ and } d(C(i),C(j))=1} |i-j|^q \text{ for } C \in \mathcal{C}$$

to study optimal 2-dimensional mappings for $q \in [0, 1]$. For the case $q = 1$, the optimal mapping with respect to $L_{\mathrm{MD},1}$ is very different from that in [PKK92]. For the case $q < 1$, they prove a lower bound for arbitrary 2-dimensional curve C:

$$L_{\mathrm{MD},q}(C) \geq \frac{1}{1+2q} n^{1+2q} + O(n^{2q}),$$

and provide an explicit construction for a 2-dimensional $\mathcal{C}$ with good but suboptimal locality. They conjecture that the space-filling curves with optimal locality (with respect to $L_{\mathrm{MD},q}$ with $q < 1$) must exhibit a "fractal" character.

For measuring the proximity preservation of close-by points in the indexing space $[n^m]$, Gotsman and Lindenbaum [GL96] approximate the locality measures:

$$L_{\mathrm{GL,min}}(C) = \min_{i,j \in [n^m] | i<j} \frac{d(C(i), C(j))^m}{|i-j|}, \text{ and}$$
$$L_{\mathrm{GL,max}}(C) = \max_{i,j \in [n^m] | i<j} \frac{d(C(i), C(j))^m}{|i-j|},$$

for arbitrary m-dimensional curve C, and obtain good bounds for Hilbert curve family. Alber and Niedermeier [AN00] generalize $L_{\mathrm{GL,max}}$ to $L_{\mathrm{AN},p}$ by employing the p-normed metric d_p, in place of the Euclidean metric d. They improve and extend the tight bounds for the 2-dimensional Hilbert curve family in [GL96] for $p = 1$, 2, and ∞.

The focus of this paper provides an analytical study of locality properties of the Hilbert and z-order curve families. The underlying locality measure is similar to $L_{\mathrm{MD},1}$ conditional on a 1-normed distance of δ between points in $[n]^m$:

$$L_\delta(C) = \sum_{i,j \in [n^m] | i<j \text{ and } d_1(C(i),C(j))=\delta} |i-j|.$$

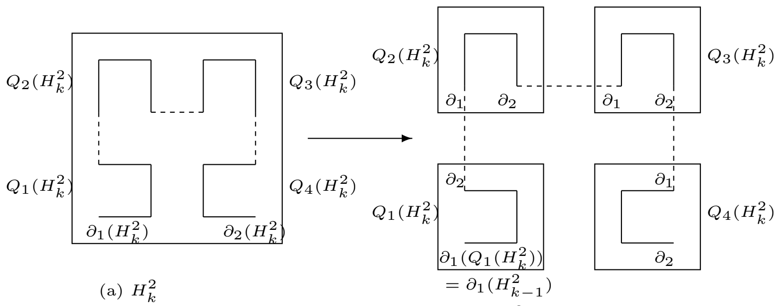
The locality statistics $L_\delta(C)$ cumulates all index-differences of point-pairs (distances traversed in the sequential index space) at a common 1-normed distance δ (local operation in the C-structural grid space). Note that for the three statistics: $L_\delta(C)$, $L_{[\delta]}(C) = \sum_{i=1}^{\delta} L_i(C)$, and the mean absolute index-difference over all point-pairs at a common 1-normed distance δ, the knowledge of one statistics for all δ yields the other two.

We derive exact formulas for L_δ for the Hilbert curve family $\{H_k^m \mid k = 1, 2, \ldots\}$ and z-order curve family $\{Z_k^m \mid k = 1, 2, \ldots\}$ for $m = 2$ and arbitrary δ that is an integral power of 2, and $m = 3$ and $\delta = 1$. Note that we present the skeletons for proving the main results without the lengthly derivations. Complete proofs and verifying programs are available from the authors.

2 Locality Measures of Hilbert and z-Order Curves

One of the salient characteristics of Hilbert curves is their "self-similarity" — a Hilbert curve can be generated by interconnecting identical subcurves via reflection and rotation (see Figure 1). For 2-dimensional Hilbert curves, this self-similar structural property guides us to decompose H_k^2 into four identical H_{k-1}^2-subcurves (via reflection and rotation), which are amalgamated together by an H_1^2-curve. Following the linear order along this H_1^2-curve, we denote the four H_{k-1}^2-subcurves as $Q_1(H_k^2)$, $Q_2(H_k^2)$, $Q_3(H_k^2)$, and $Q_4(H_k^2)$.

For a 2-dimensional grid, the "orientation" of H_k^2 uniquely determines that of $Q_\alpha(H_k^2)$ for $\alpha = 1, 2, 3, 4$, and thus only one H_k^2 exists modulo symmetry (whereas there are 1536 structurally different 3-dimensional Hilbert curves [AN00]). For a 2-dimensional Hilbert curve H_k^2 indexing the grid $[2^k]^2$, with a canonical orientation shown in Figure 2(a), we denote by $\partial_1(H_k^2)$ and $\partial_2(H_k^2)$ the entry and exit, respectively, grid point in $[2^k]^2$ (with respect to the canonical orientation). Figure 2 depicts the decomposition of H_k^2 and the ∂_1- and ∂_2-labels of four H_{k-1}^2-subcurves.



(a) H_k^2

(b) H_1^2-interconnection

Fig. 2. Generation of H_k^2 from a H_1^2-interconnection of four H_{k-1}^2-subcurves.

We derive the exact formula for $L_\delta(H_k^2)$ as follows (similarly for $L_\delta(Z_k^2)$). The recursive decomposition (in k) of H_k^2 gives that

$$L_\delta(H_k^2) = 4L_\delta(H_{k-1}^2) + \sum_{\alpha,\beta\in\{1,2,3,4\}|\alpha<\beta} \Delta_\delta(Q_\alpha(H_k^2), Q_\beta(H_k^2)),$$

where $\Delta_\delta(Q_\alpha(H_k^2), Q_\beta(H_k^2))$ denotes the cumulative contribution of $|i-j|$ from the two subcurves $Q_\alpha(H_k^2)$ and $Q_\beta(H_k^2)$, that is, for all $i, j \in [2^{2k}]$ such that $i < j$, $d_1(H_k^2(i), H_k^2(j)) = \delta$, and i and j appear in (the index ranges of) $Q_\alpha(H_k^2)$ and $Q_\beta(H_k^2)$, respectively. In order to compute the closed-form solution for $L_\delta(H_k^2)$, we develop a suite of lemmas (Lemmas 4, 5, and 6) to compute $\Delta_\delta(Q_\alpha(H_k^2), Q_\beta(H_k^2))$ for all $\alpha, \beta \in \{1, 2, 3, 4\}$ with $\alpha < \beta$, in which we establish a recurrence system (in k — when $2^k > \delta$) of summations with a basis system of summations computed in Lemmas 1, 2, and 3. The following denotations illustrate the geometrical structures (rows, columns, and diagonals) involved in the computations.

With respect to the canonical orientation of H_k^2 shown in Figure 2(a), we cover the 2-dimensional k-order grid with:

1. 2^k rows $(R_{k,1}, R_{k,2}, \ldots, R_{k,2^k})$, indexed from the bottom,
2. 2^k columns $(C_{k,1}, C_{k,2}, \ldots, C_{k,2^k})$, indexed from the left,
3. $2^{k+1} - 1$ main diagonals $(D_{k,1}, D_{k,2}, \ldots, D_{k,2^k} = D'_{k,2^k}, D'_{k,2^k-1}, \ldots, D'_{k,1})$, indexed from the lower-right corner, and
4. $2^{k+1} - 1$ auxiliary diagonals $(A_{k,1}, A_{k,2}, \ldots, A_{k,2^k} = A'_{k,2^k}, A'_{k,2^k-1}, \ldots, A'_{k,1})$, indexed from the lower-left corner.

For $\alpha \in [2^k]$ and a grid point $p \in [2^k]^2$, we denote:

1. $\hbar_k(v, v') = |(H_k^2)^{-1}(v) - (H_k^2)^{-1}(v')|$, the index-difference between two grid points $v, v' \in [2^k]^2$.
2. $\Delta(X_{k,\alpha}, p) = \sum_{v\in X_{k,\alpha}} \hbar_k(v, p)$, where the symbol X denotes R, C, D, D', A, or A' (for example, $\Delta(R_{k,\alpha}, p) = \sum_{v\in R_{k,\alpha}} \hbar_k(v, p)$). That is, $\Delta(X_{k,\alpha}, p)$ cumulates all index-differences of all grid points in the structure $X_{k,\alpha}$ with respect to p; when $p = \partial_1(H_k^2)$, $\Delta(X_{k,\alpha}, p)$ is the index-cumulation of all grid points in $X_{k,\alpha}$.
3. $\mathcal{X}_{k,\alpha} = \sum_{\beta=1}^{\alpha} \sum_{v\in X_{k,\beta}} (\alpha + 1 - \beta)\hbar_k(v, \partial_1(H_k^2))$, where the symbol-pair $(\mathcal{X}, X)$ denotes $(\mathcal{R}, R)$, $(\mathcal{C}, C)$, $(\mathcal{D}, D)$, $(\mathcal{D}', D')$, $(\mathcal{A}, A)$, or $(\mathcal{A}', A')$ (for example, $\mathcal{R}_{k,\alpha} = \sum_{\beta=1}^{\alpha} \sum_{v\in R_{k,\beta}} (\alpha + 1 - \beta)\hbar_k(v, \partial_1(H_k^2)))$;
   $\mathcal{N}_{k,\alpha} = \sum_{\beta=1}^{\alpha} \sum_{v\in X_{k,\beta}} (\alpha + 1 - \beta)$, when X denotes D, D', A, or A' (independent of the choice); that is, $\mathcal{N}_{k,\alpha}$ cumulates the number of index (grid point v) references in the summation of $\mathcal{X}_{k,\alpha}$ (that is, $\mathcal{D}_{k,\alpha}$, $\mathcal{D}'_{k,\alpha}$, $\mathcal{A}_{k,\alpha}$, or $\mathcal{A}'_{k,\alpha}$).
   Note that when $\alpha = 0$, all cumulations degenerate to 0, that is, $\mathcal{X}_{k,0} = \mathcal{X}'_{k,0} = \mathcal{N}_{k,0} = 0$.
4. $\overline{\mathcal{X}}_{k,\alpha} = \sum_{\beta=1}^{\alpha} \sum_{v\in X_{k,\beta}} \hbar_k(v, \partial_1(H_k^2)) = \sum_{\beta=1}^{\alpha} \Delta(X_{k,\beta}, \partial_1(H_k^2))$, where the symbol-pair $(\overline{\mathcal{X}}, X)$ denotes $(\overline{\mathcal{R}}, R)$, $(\overline{\mathcal{C}}, C)$, $(\overline{\mathcal{D}}, D)$, $(\overline{\mathcal{D}'}, D')$, $(\overline{\mathcal{A}}, A)$, or $(\overline{\mathcal{A}'}, A')$;

$\overline{\mathcal{N}}_{k,\alpha} = \sum_{\beta=1}^{\alpha} \sum_{v \in X_{k,\beta}} 1$, when the symbol X denotes D, D', A, or A' (independent of the choice); that is, $\overline{\mathcal{N}}_{k,\alpha}$ cumulates the number of index (grid point v) references in the summation of $\overline{\mathcal{X}}_{k,\alpha}$ (that is, $\overline{\mathcal{D}}_{k,\alpha}$, $\overline{\mathcal{D}'}_{k,\alpha}$, $\overline{\mathcal{A}}_{k,\alpha}$, or $\overline{\mathcal{A}'}_{k,\alpha}$).
Note that $\mathcal{X}_{k,\alpha} = \mathcal{X}_{k,\alpha-1} + \overline{\mathcal{X}}_{k,\alpha}$ for $\alpha \in \{2, 3, \ldots, 2^k\}$, where the symbol $\mathcal{X}$ denotes $\mathcal{R}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{D}'$, $\mathcal{A}$, or $\mathcal{A}'$.

5. $\overline{\mathcal{W}}_k = \sum_{v \in H_k^2} \hbar_k(v, \partial_1(H_k^2))$, which cumulates the indices of all grid points of H_k^2 relative to $\partial_1(H_k^2)$. (For a grid point v, the membership "$v \in H_k^2$" abbreviates "$v \in H_k^2([2^{2k}])$".)

Figure 3 illustrates the organization of a 2-dimensional grid into the row, column, main-diagonal, auxiliary-diagonal structures and the coverages for $\mathcal{D}_{k,\alpha}$, $\mathcal{D}'_{k,\alpha}$, $\mathcal{A}_{k,\alpha}$, and $\mathcal{A}'_{k,\alpha}$.

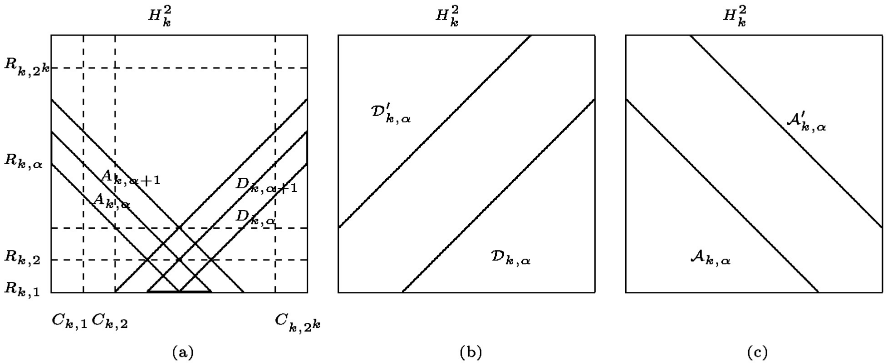


Fig. 3. (a) Organize a 2-dimensional grid $[2^k]^2$ into the row, column, main-diagonal, and auxiliary-diagonal structures; (b) coverages of $\mathcal{D}_{k,\alpha}$ and $\mathcal{D}'_{k,\alpha}$; (c) coverages of $\mathcal{A}_{k,\alpha}$ and $\mathcal{A}'_{k,\alpha}$.

The following three lemmas study the cumulation of indices of grid points in the row, column, diagonal, and auxiliary-diagonal structures of H_k^2.

Lemma 1. *The index-cumulation of a row structure of H_k^2 is independent of its row-number: for all $\alpha \in [2^k]$,*

$$\Delta(R_{k,\alpha}, \partial_1(H_k^2)) = \Delta(R_{k,\alpha}, \partial_2(H_k^2)) = \frac{1}{2} \cdot 2^{3k} - \frac{1}{2} \cdot 2^k \qquad \textit{(independent of } \alpha\textit{)}.$$

Lemma 2. *The binary representation of the column-number of a column structure of H_k^2 helps compute its index-cumulation as follows:*

1. *For all $\alpha \in [2^{k-1}]$, a recurrence for $\Delta(C_{k,\alpha}, \partial_1(H_k^2))$ is:*

$$\Delta(C_{k,\alpha}, \partial_1(H_k^2)) = \Delta(R_{k-1,\alpha}, \partial_1(H_{k-1}^2)) + \Delta(C_{k-1,\alpha}, \partial_1(H_{k-1}^2)) + \frac{1}{2^3} \cdot 2^{3k}.$$

For $\alpha = 1$, a closed-form solution for $\Delta(C_{k,1}, \partial_1(H_k^2))$ from the recurrence above (in k) is:

$$\frac{3}{2 \cdot 7} \cdot 2^{3k} - \frac{1}{2} \cdot 2^k + \frac{2}{7}.$$

For those α satisfying $2^{q-1} < \alpha \leq 2^q$ for some integer $q \in [k-1]$ (that is, $q = \lceil \log \alpha \rceil$), the recurrence above (in k) yields a recurrence:

$$\Delta(C_{k,\alpha}, \partial_1(H_k^2)) = \Delta(C_{q,\alpha}, \partial_1(H_q^2)) + \sum_{\eta=q+1}^{k} (\frac{3}{2^4} \cdot 2^{3\eta} - \frac{1}{2^2} \cdot 2^{\eta}),$$

where the summation is:

$$\frac{3}{2 \cdot 7}(2^{3k} - 2^{3q}) - \frac{1}{2}(2^k - 2^q) = \frac{3}{2 \cdot 7} \cdot 2^{3k} - \frac{1}{2} \cdot 2^k - \frac{3}{2 \cdot 7} \cdot 2^{3\lceil \log \alpha \rceil} + \frac{1}{2} \cdot 2^{\lceil \log \alpha \rceil}.$$

2. *For all $\alpha, \beta \in [2^k]$ such that $\alpha < \beta$ and the binary representations of $\alpha - 1$ and $\beta - 1$ differ only at the i-th low-order bit, where $i \in \{0, 1, \ldots, k-1\}$ (that is, $(\alpha - 1) \oplus (\beta - 1) = 2^i$, where $\oplus$ denotes the binary exclusive-or operator),*

$$\Delta(C_{k,\beta}, \partial_1(H_k^2)) = \Delta(C_{k,\alpha}, \partial_1(H_k^2)) + 2^2 \cdot 2^{3i}.$$

Lemma 3. *For $k \geq 1$,*

1. $\Delta(A_{k,2^k}, \partial_1(H_k^2)) + \Delta(D_{k,2^k}, \partial_1(H_k^2)) = 2^{3k} - 2^k$, *and*
2. $\Delta(A_{k,2^k}, \partial_1(H_k^2)) = \frac{2}{3} \cdot 2^{3k} - \frac{2}{3} \cdot 2^k$.

Now we partition the summation $\sum_{\alpha,\beta \in \{1,2,3,4\} | \alpha < \beta} \Delta_\delta(Q_\alpha(H_k^2), Q_\beta(H_k^2))$ according to the two cases: for $\alpha, \beta \in \{1, 2, 3, 4\}$, contiguous subcurves ($\alpha + 1 \equiv \beta \pmod 4$) with four similar subcases, and diagonal subcurves ($\alpha + 2 \equiv \beta \pmod 4$) with two similar subcases. The summations resulting from restricting to these six subcases are investigated in the following lemma. A common thread to the computations in the six subcases is to express $\Delta_\delta(Q_\alpha(H_k^2), Q_\beta(H_k^2))$ as summations of index-cumulations of neighboring geometrical structures in $Q_\alpha(H_k^2)$ and $Q_\beta(H_k^2)$. For the case of contiguous subcurves, the boundary rows/columns and the boundary corners (of main- and auxiliary-diagonals) are involved; and for the case of diagonal subcurves, the boundary corners (of main- and auxiliary-diagonals) are involved.

Lemma 4. *For δ that is an integral power of 2, and $1 \leq \delta < 2^k$,*

$$\begin{aligned}
\Delta_\delta(Q_1(H_k^2), Q_4(H_k^2)) &= 2(\overline{\mathcal{R}}_{k-1,\delta} + 2\mathcal{R}_{k-1,\delta-1} - \mathcal{D}'_{k-1,\delta-1} - \mathcal{A}'_{k-1,\delta-1}) \\
&\quad + (\frac{1}{2} \cdot 2^{2k} + 1)(2^{k-1}\delta^2 - 2\mathcal{N}_{k-1,\delta-1}), \\
\Delta_\delta(Q_1(H_k^2), Q_2(H_k^2)) &= \Delta_\delta(Q_3(H_k^2), Q_4(H_k^2)) \\
&= (\overline{\mathcal{C}}_{k-1,\delta} + 2\mathcal{C}_{k-1,\delta-1} - \mathcal{D}'_{k-1,\delta-1} - \mathcal{A}_{k-1,\delta-1}) + (\overline{\mathcal{R}}_{k-1,\delta} \\
&\quad + 2\mathcal{R}_{k-1,\delta-1} - \mathcal{A}_{k-1,\delta-1} - \mathcal{D}_{k-1,\delta-1}) + (2^{k-1}\delta^2 - 2\mathcal{N}_{k-1,\delta-1}), \\
\Delta_\delta(Q_2(H_k^2), Q_3(H_k^2)) &= 2(\overline{\mathcal{C}}_{k-1,\delta} + 2\mathcal{C}_{k-1,\delta-1} - \mathcal{D}'_{k-1,\delta-1} - \mathcal{A}_{k-1,\delta-1}) \\
&\quad + (2^{k-1}\delta^2 - 2\mathcal{N}_{k-1,\delta-1}), \ \textit{and} \\
\Delta_\delta(Q_1(H_k^2), Q_3(H_k^2)) &= \Delta_\delta(Q_2(H_k^2), Q_4(H_k^2)) \\
&= (\frac{1}{2^2} \cdot 2^{2k} + 1)\mathcal{N}_{k-1,\delta-1} + \mathcal{D}'_{k-1,\delta-1} + \mathcal{A}_{k-1,\delta-1}.
\end{aligned}$$

The following two lemmas allow us to simplify the overall summation of $L_\delta(H_k^2)$.

Lemma 5. *For all integers q with $0 \le q \le k$,*

1. $\overline{\mathcal{W}}_k = \frac{1}{2} \cdot 2^{4k} - \frac{1}{2} \cdot 2^{2k}$,
2. $\overline{\mathcal{N}}_{k,2^q} = \frac{1}{2} \cdot 2^{2q} + \frac{1}{2} \cdot 2^q$,
3. $\overline{\mathcal{R}}_{k,2^q} = 2^q \Delta(R_{k,1}, \partial_1(H_k^2))$,
4. $\overline{\mathcal{C}}_{k,2^q} = 2^q \Delta(C_{k,1}, \partial_1(H_k^2)) + \frac{2}{7}(2^{4q} - 2^q)$,
5. $\overline{\mathcal{A}}_{k,2^k} + \overline{\mathcal{A}'}_{k,2^k} = \overline{\mathcal{W}}_k + \Delta(A_{k,2^k}, \partial_1(H_k^2))$,
6. $\overline{\mathcal{A}}_{k,2^k} = 2\overline{\mathcal{W}}_{k-1} + \Delta(A_{k-1,2^{k-1}}, \partial_1(H_{k-1}^2)) + 2^{2k}\overline{\mathcal{N}}_{k-1,2^{k-1}}$,
7. *Each of the pairs: $(\overline{\mathcal{D}}_{k,2^q}, \overline{\mathcal{A}}_{k,2^q})$ and $(\overline{\mathcal{D}'}_{k,2^q}, \overline{\mathcal{A}'}_{k,2^q})$ are related via $\overline{\mathcal{N}}_{k,2^q}$ as follows:*
$$\overline{\mathcal{D}}_{k,2^q} + \overline{\mathcal{A}}_{k,2^q} = \overline{\mathcal{D}'}_{k,2^q} + \overline{\mathcal{A}'}_{k,2^q} = (2^{2k} - 1)\overline{\mathcal{N}}_{k,2^q},$$
and
8. $\overline{\mathcal{D}'}_{k,2^q} = \frac{1}{3}(2^{2k} - 2^{2q})\overline{\mathcal{N}}_{k,2^q} + \overline{\mathcal{D}'}_{q,2^q}$.

Computations of various $\mathcal{X}_{k,2^q}$ are similar to those for $\overline{\mathcal{X}}_{k,2^q}$.

Lemma 6. *For all integers q with $0 \le q \le k$,*

1. $\mathcal{N}_{k,2^q} = \frac{1}{2\cdot 3} \cdot 2^{3q} + \frac{1}{2} \cdot 2^{2q} + \frac{1}{3} \cdot 2^q$,
2. $\mathcal{R}_{k,2^q} = \Delta(R_{k,1}, \partial_1(H_k^2)) \sum_{\beta=1}^{2^q} \beta$,
3. $\mathcal{C}_{k,2^q} = \frac{3}{2^2\cdot 7} \cdot 2^{3k+2q} + \frac{3}{2^2\cdot 7} \cdot 2^{3k+q} - \frac{1}{2^2} \cdot 2^{k+2q} - \frac{1}{2^2} \cdot 2^{k+q} + \frac{2^3}{3\cdot 5\cdot 7} \cdot 2^{5q} + \frac{1}{7} \cdot 2^{4q} + \frac{1}{3\cdot 5} \cdot 2^q$,
4. *Each of the pairs: $(\mathcal{D}_{k,2^q}, \mathcal{A}_{k,2^q})$ and $(\mathcal{D}'_{k,2^q}, \mathcal{A}'_{k,2^q})$ are related via $\mathcal{N}_{k,2^q}$ as follows:*
$$\mathcal{D}_{k,2^q} + \mathcal{A}_{k,2^q} = \mathcal{D}'_{k,2^q} + \mathcal{A}'_{k,2^q} = (2^{2k} - 1)\mathcal{N}_{k,2^q},$$
5. $$\mathcal{A}_{k,2^k} = \begin{cases} 0 & \textit{if } k = 0 \\ \frac{7}{2^2\cdot 3^2\cdot 5} \cdot 2^{5k} + \frac{3}{2^4} \cdot 2^{4k} + \frac{5}{2^2\cdot 3^2} \cdot 2^{3k} - \frac{1}{2^2} \cdot 2^{2k} - \frac{2^3}{3^2\cdot 5} \cdot 2^k & \textit{otherwise} \end{cases}$$
6. $\mathcal{A}_{k,2^q} = \mathcal{A}_{q,2^q}$,
7. $$\mathcal{D}'_{k,2^k} = \begin{cases} 0 & \textit{if } k = 0 \\ \frac{11}{2^2\cdot 3^2\cdot 5} \cdot 2^{5k} + \frac{3}{2^4} \cdot 2^{4k} + \frac{1}{2^2\cdot 3^2} \cdot 2^{3k} - \frac{1}{2^2} \cdot 2^{2k} - \frac{2^2}{3^2\cdot 5} \cdot 2^k & \textit{otherwise} \end{cases}$$
8. $\mathcal{D}'_{k,2^q} = \frac{1}{3}(2^{2k} - 2^{2q})\mathcal{N}_{k,2^q} + \mathcal{D}'_{q,2^q}$.

Theorem 1. *For $\delta \in [2^k]$ that is an integral power of 2,*

$$L_\delta(H_k^2) = \begin{cases} \frac{17}{2\cdot 7} \cdot 2^{3k} - \frac{5}{2\cdot 3} \cdot 2^{2k} - \frac{2^3}{3\cdot 7} & \textit{if } \delta = 1 \\ \frac{17}{2\cdot 7} \cdot 2^{3k+2\log\delta} - \frac{2^3\cdot 3\cdot 5^2\cdot 7(k-\log\delta)+5\cdot 7\cdot 383}{2^4\cdot 3^3\cdot 5\cdot 7} \cdot 2^{2k+3\log\delta} & \\ \quad + \frac{2\cdot 3\cdot 5(k-\log\delta)-1}{2^2\cdot 3^3} \cdot 2^{2k+\log\delta} - \frac{2^2\cdot 41}{3^3\cdot 5\cdot 7} \cdot 2^{5\log\delta} - \frac{2}{3^3} \cdot 2^{3\log\delta} & \\ \quad - \frac{2}{3\cdot 5} \cdot 2^{\log\delta} & \textit{otherwise.} \end{cases}$$

For the z-order curve family, we follow a similar approach to develop a suite of lemmas to establish the following formula for its locality measure.

Theorem 2. *For $\delta \in [2^k]$ that is an integral power of* 2,

$$L_\delta(Z_k^2) = \begin{cases} 2^{3k} - 2^k & \text{if } \delta = 1 \\ 2^{3k+2\log\delta} - (\frac{2}{3^2}(k - \log\delta) + \frac{1949}{2^5 \cdot 3^3 \cdot 7})2^{2k+3\log\delta} & \\ \quad + (\frac{2}{3^2}(k-\log\delta) + \frac{7}{2^2 \cdot 3^3})2^{2k+\log\delta} + \frac{19}{2^2 \cdot 3 \cdot 7} \cdot 2^{2k} - \frac{2^2}{7} \cdot 2^{k+4\log\delta} & \\ \quad - \frac{3}{7} \cdot 2^{k+\log\delta} + \frac{2 \cdot 5}{3^3 \cdot 7} \cdot 2^{5\log\delta} - \frac{2^2}{3^3} \cdot 2^{3\log\delta} + \frac{2}{3 \cdot 7} \cdot 2^{2\log\delta} & \text{otherwise.} \end{cases}$$

For the 2-dimensional Hilbert and z-order curve families, the complexity of deriving $L_\delta(H_k^2)$ and $L_\delta(Z_k^2)$ lies in the parameterized 1-normed distance δ. For the case of dimensionality 3, we are able to derive exact formulas for $L_\delta(H_k^2)$ and $L_\delta(Z_k^2)$ for $\delta = 1$ (each for one of many structurally different curve families).

Theorem 3. *For a 3-dimensional Hilbert curve family,*

$$L_1(H_k^3) = \frac{67}{2 \cdot 31} \cdot 2^{5k} - \frac{11}{2 \cdot 7} \cdot 2^{3k} - \frac{2^6}{7 \cdot 31}.$$

Theorem 4. *For a 3-dimensional z-order curve family,*

$$L_1(Z_k^3) = 2^{5k} - 2^{2k}.$$

3 Comparison, Verification, and Conclusion

For sufficiently large k and $\delta \ll 2^k$,

$$\frac{L_\delta(H_k^m)}{L_\delta(Z_k^m)} \approx \begin{cases} \frac{17}{2 \cdot 7} \approx 1.2143 \text{ for } m = 2 \text{ and } \delta \text{ an integral power of } 2 \\ \frac{67}{2 \cdot 31} \approx 1.0806 \text{ for } m = 3. \end{cases}$$

With respect to the locality measure L_δ and for sufficiently large k and $\delta \ll 2^k$, the z-order curve family performs better than the Hilbert curve family for $m = 2$ and over the δ-spectrum of integral powers of 2.

When $\delta = 2^k$, the domination reverses as:

$$L_\delta(H_k^2) = \frac{37}{240} \cdot 2^{5k} - \frac{1}{12} \cdot 2^{3k} - \frac{2}{15} \cdot 2^k, \text{ and}$$
$$L_\delta(Z_k^2) = \frac{107}{672} \cdot 2^{5k} - \frac{1}{12} \cdot 2^{3k} - \frac{3}{28} \cdot 2^{2k}.$$

These give that

$$\frac{L_\delta(H_k^m)}{L_\delta(Z_k^m)} \approx \frac{2 \cdot 7 \cdot 37}{5 \cdot 107} \approx 0.9682.$$

The superiority of the z-order curve family persists but declines for $m = 3$ with unit 1-normed distance for L_δ.

For the extreme case $m = 2$ and $\delta = 1$, the locality measure L_δ in our study degenerates to $L_{\mathrm{MD},1}$ in [MD86]. Their analysis shows that for a 2-dimensional curve C for the grid $[n]^2$, $L_{\mathrm{MD},1}$ attains its minimum $\frac{4-\sqrt{2}}{3}n^3 + O(n^2)$ ($\approx 0.8619n^2 + O(n^2)$) when C and its equivalent variants assume the following characteristics: (1) Within the four $(1-\frac{1}{\sqrt{2}})n \times (1-\frac{1}{\sqrt{2}})n$ corner-subgrids, the sequence of 1-normed distances between adjacent points in $[n]^2$, $d_1(C(i), C(i+1))$, incrementally increases and/or decreases in the range $[1, (1-\frac{1}{\sqrt{2}})n]$ (while interleaving with segments of 1s), and (2) Within the central region interconnecting the corner-subgrids, the 1-normed distances are in $\{(1-\frac{1}{\sqrt{2}})n, n\}$ (while interleaving with segments of 1s). As the z-order curve family shares some of these characteristics, the asymptotic ratios (constants greater than 1) obtained above are not surprising.

We have verified all the exact formulas (intermediate and final) involved in the derivations with computer programs over various grid-orders and 1-normed distances: ($m = 2$, $k \in \{1, 2, 3, 4, 5, 6, 7\}$, and $\delta \in \{1, 2^1, 2^2, 2^3, \ldots, 2^k\}$), and ($m = 3$, $k \in \{1, 2, 3, 4, 5, 6\}$, and $\delta = 1$).

A similar study with dimensions greater than 3 appears to be much more difficult due to the loss of geometric intuition. Alber and Niedermeier [AN00] provide a simple mathematical mechanism to describe and analyze the combinatorial properties of continuous curves such as Hilbert curves and non-continuous ones such as z-order curves in arbitrary dimensions. This structure-theoretic viewpoint may shed some light on our current study with arbitrary dimensions and unit distance.

References

[AN00] J. Alber and R. Niedermeier. On multi-dimensional curves with Hilbert property. *Theory of Computing Systems*, 33(4):295–312, 2000.

[ARR+97] T. Asano, D. Ranjan, T. Roos, E. Welzl, and P. Widmayer. Space-filling curves and their use in the design of geometric data structures. *Theoretical Computer Science*, 181(1):3–15, 1997.

[DS03] H. K. Dai and H. C. Su. Approximation and analytical studies of inter-clustering performances. To appear in *Proceedings of the 2003 Discrete Random Walks Conference*, September 2003.

[GL96] C. Gotsman and M. Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797, 1996.

[Jag97] H. V. Jagadish. Analysis of the Hilbert curve for representing two-dimensional space. *Information Processing Letters*, 62(1):17–22, 1997.

[MD86] G. Mitchison and R. Durbin. Optimal numberings of an $N \times N$ array. *SIAM Journal on Algebraic and Discrete Methods*, 7(4):571–582, 1986.

[MJFS01] B. Moon, H. V. Jagadish, C. Faloutsos, and J. H. Saltz. Analysis of the clustering properties of the Hilbert space-filling curve. *IEEE Transactions on Knowledge and Data Engineering*, 13(1):124–141, 2001.

[PKK92] A. Pérez, S. Kamata, and E. Kawaguchi. Peano scanning of arbitrary size images. In *Proceedings of the International Conference on Pattern Recognition*, pages 565–568. IEEE Computer Society, 1992.