Norm-Based Locality Measures of Two-Dimensional Hilbert Curves

H.K. Dai\textsuperscript{1(\textcopyright)} and H.C. Su\textsuperscript{2}

\textsuperscript{1} Computer Science Department, Oklahoma State University, Stillwater, OK 74078, USA
dai@cs.okstate.edu
\textsuperscript{2} Department of Computer Science, Arkansas State University, Jonesboro, AR 72401, USA
suh@astate.edu

Abstract. A discrete space-filling curve provides a 1-dimensional indexing or traversal of a multi-dimensional grid space. Applications of space-filling curves include multi-dimensional indexing methods, parallel computing, and image compression. Common goodness-measures for the applicability of space-filling curve families are locality and clustering. Locality reflects proximity preservation that close-by grid points are mapped to close-by indices or vice versa. We present an analytical study on the locality property of the 2-dimensional Hilbert curve family. The underlying locality measure, based on the \( p \)-normed metric \( d_p \), is the maximum ratio of \( d_p(u, v)^m \) to \( d_p(\tilde{u}, \tilde{v}) \) over all corresponding point-pairs \((u, v)\) and \((\tilde{u}, \tilde{v})\) in the \( m \)-dimensional grid space and 1-dimensional index space, respectively. Our analytical results identify all candidate representative grid-point pairs (realizing the locality-measure values) for all real norm-parameters in the unit interval \([1, 2]\) and grid-orders. Together with the known results for other norm-parameter values, we have almost complete knowledge of the locality measure of 2-dimensional Hilbert curves over the entire spectrum of possible norm-parameter values.

Keywords: Space-filling curves · Hilbert curves · z-order curves · Locality

1 Preliminaries

Discrete space-filling curves have many applications in databases, parallel computation, algorithms, 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 \( \mathbb{Z} = \{1, 2, \ldots, n\} \). An \( m \)-dimensional (discrete) space-filling curve of length \( n^m \) is a bijective mapping \( C : [n]^m \rightarrow [n]^m \), thus providing a linear indexing/traversal or total ordering of the 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.

One of the salient characteristics of space-filling curves is their “self-similarity”. 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 $H_k^m$ and $Z_k^m$ for $m = 2$, and $k = 1, 2$, and $m = 3$, and $k = 1$.

![Recursive constructions of Hilbert and z-order curves](image)

Fig. 1. Recursive constructions of Hilbert and z-order curves of higher order (respectively, $H_k^m$ and $Z_k^m$) by interconnecting symmetric subcurves, via reflection and/or rotation, of lower order (respectively, $H_{k-1}^m$ and $Z_{k-1}^m$) along an order-1 subcurve (respectively, $H_1^m$ and $Z_1^m$): (a) $H_2^1$; (b) $H_2^2$; (c) $H_3^1$; (d) $Z_2^2$; (e) $Z_2^2$; (f) $Z_3^3$.

We measure the applicability of a family of space-filling curves based on:

1. their common structural characteristics that reflect locality and clustering,
2. descriptional simplicity that facilitates their construction and combinatorial analysis in arbitrary dimensions, and
3. computational complexity in the grid space-index space transformation.

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 performances of various low-dimensional space-filling curves have been reported in the literature (see [4] and [6] for details). These studies show that the Hilbert and z-order curve families manifest good data clustering properties according to some quality clustering measures, robust mathematical formalism, and viable indexing techniques for querying multi-dimensional data, when compared with other curve families.

The locality preservation of a space-filling curve family is crucial for the efficiency of many indexing schemes, data structures, and algorithms in its applications, for examples, spatial correlation in multi-dimensional indexings, compression in image processing, and communication optimization in mesh-connected parallel computing. To analyze locality, we need to rigorously define its measures that are practical—good bounds (lower and upper) on the locality measure translate into good bounds on the declustering (locality loss) in one space in the presence of locality in the other space.
A few locality measures have been proposed and analyzed for space-filling curves in the literature. Denote by $d$ and $d_p$ the Euclidean metric and $p$-normed metric (rectilinear metric ($p = 1$) and maximum metric ($p = \infty$)), respectively. Let $\mathcal{C}$ denote a family of $m$-dimensional curves of successive orders.

We [5] consider a locality measure conditional on a 1-normed distance of $\delta$ between points in $[n]^m$:

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

They derive exact formulas for $L_\delta$ 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 $\delta$ that is an integral power of 2, and $m = 3$ and $\delta = 1$ (lower-order terms collected in asymptotic form for brevity):

$$ L_\delta(H_k^2) = \begin{cases} \frac{17}{27} \cdot 2^{3k} + \mathcal{O}(2^{2k}) & \text{if } \delta = 1 \\ 2^{3k+2} \log \delta + \mathcal{O}(2^{2k+3} \log \delta) & \text{otherwise,} \end{cases} $$

$$ L_\delta(Z_k^2) = \begin{cases} 2^{3k} + \mathcal{O}(2^k) & \text{if } \delta = 1 \\ 2^{3k+2} \log \delta + \mathcal{O}(2^{2k+3} \log \delta) & \text{otherwise;} \end{cases} $$

$$ L_1(H_k^3) = \frac{67}{2 \cdot 31} \cdot 2^{5k} + \mathcal{O}(2^{3k}) \text{ and } L_1(Z_k^3) = 2^{5k} + \mathcal{O}(2^{2k}). $$

With respect to the locality measure $L_\delta$ 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 $\delta$-spectrum of integral powers of 2. When $\delta = 2^k$, the domination reverses. The superiority of the z-order curve family persists but declines for $m = 3$ with unit 1-normed distance for $L_\delta$.

For measuring the proximity preservation of close-by points in the indexing space $[n]^m$, Gotsman and Lindenbaum [7] consider the following measures: for $C \in \mathcal{C}$,

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

Alber and Niedermeier [1] generalize $L_{\max}$ to $L_p$ by employing the $p$-normed metric $d_p$ for real norm-parameter $p \geq 1$ in place of the Euclidean metric $d$, which is the locality measure studied in our work (and [5]). We summarize below: (1) the representative lower- and upper-bound results and exact formulas for the locality measure $L_p$ of the 2-dimensional Hilbert curve family $H_k^2$ for various norm-parameter $p$-values and grid-order $k$-values, and (2) the contribution of our studies:

1. For $p = 1$: Niedermeier, Reinhardt, and Sanders [8] give a lower bound for $L_1(H_k^2)$: for all $k \geq 1$,

$$ L_1(H_k^2) \geq \frac{(3 \cdot 2^{k-1} - 2)^2}{4^{k-1}}, $$

and Chochia et al. [3] provide a matching upper bound for $L_1(H_k^2)$ for all $k \geq 2$. We [5] also provide the exact formula for $L_1(H_k^2)$ for all $k \geq 2$. 


2. For \( p = 2 \): Gotsman and Lindenbaum [7] derive a lower and upper bounds for \( L_2(H^2_k) \): for all \( k \geq 6 \),

\[
\frac{(2^{k-1} - 1)^2}{2} \cdot 4^{k-2} + \frac{1}{3} \leq L_2(H^2_k) \leq 6 \cdot 2^{k-3} + 1,
\]

and Alber and Niedermeier [1] improves the upper bound for \( L_2(H^2_k) \): for all \( k \geq 1 \),

\[
L_2(H^2_k) \leq \frac{6}{2}.
\]

We [5] prove that the lower bound above [7] is the exact formula for \( L_2(H^2_k) \): for all \( k \geq 5 \),

\[
L_2(H^2_k) = 6 \cdot 2^{2k-3} - 2^{k-1} + 2^{-1} \frac{2^{2k-3} + 1}{2}.
\]

Bauman [2] obtains a matching lower and upper bounds for \( L_2(H^2_k) \) for \( k = \infty \):

\[
L_2(H^2_\infty) = 6.
\]

3. For \( 2 < p \leq \infty \): Due to the monotonicity of the underlying \( p \)-normed metric: for every grid-point pair \((v, u)\), the \( p \)-normed metric \( d_p(v, u) \) is strictly decreasing in \( p \in [1, \infty) \), we [5] prove the same exact formula for \( L_p(H^2_k) \) as for the case when \( p = 2 \):

\[
L_p(H^2_k) = 6 \cdot \frac{2^{2k-3} - 2^{k-1} + 2^{-1}}{2^{2k-3} + 1} \text{ for all reals } p \geq 2.
\]

When \( p = \infty \), Alber and Niedermeier [1] establish a lower and upper bounds for \( L_\infty(H^2_k) \), respectively:

\[
6(1 - O(2^{-k})) \leq L_\infty(H^2_k) \leq 6 \cdot 2^{k-3} + 1.
\]

Our proofs of the exact formulas of \( L_p(H^2_k) \) for \( p \in \{1, 2\} \) in [5] follow a uniform approach: identifying all the representative grid-point pairs, which realize the \( L_p(H^2_k) \)-value, for each \( p \in \{1, 2\} \). The analytical results close the gap between the current best lower and upper bounds with exact formulas for \( p \in \{1, 2\} \), and extend to all reals \( p \geq 2 \). The identifications of candidate representative grid-point pairs rely on sequences of reduction. A reduction of a grid-point pair to another pair is based on the dominance of the underlying locality-measure values of the corresponding grid-point pairs. The geometric characteristics of the underlying \( p \)-norms (rectilinear and Euclidean metrics of \( p = 1 \) and \( p = 2 \), respectively) help distinguish candidate representative grid-point pairs and verify tedious reductions.

Our study of 2-dimensional curve family \( H^2_k \) is focused on the exact analysis of \( L_p(H^2_k) \) for all reals \( p \in [1, 2] \). The intrinsic mathematical appeal in completing the computation of \( L_p(H^2_k) \) for all possible norm-parameters \( p \) is our primary motivation. While the three most obviously important \( p \)-values: \( \{1, 2, \infty\} \) are
intimately related to intuitive concepts, in some cases the structure of applications of the Hilbert curves may suggest a different choice of \( p \)-value as the most natural setting for the underlying locality measure.

We present analytical and empirical studies on the locality measure \( L_p \) for the 2-dimensional Hilbert curve family for all reals \( p \in [1, 2] \). The underlying locality measure \( L_p \), based on the \( p \)-normed metric \( d_p \), is the maximum ratio of \( d_p(u, v)^m \) to \( d_p(\tilde{u}, \tilde{v}) \) over all corresponding point-pairs \((u, v)\) and \((\tilde{u}, \tilde{v})\) in the \( m \)-dimensional grid space and (1-dimensional) index space, respectively:

1. We identify all the candidate representative grid-point pairs for all norm-parameter \( p \)-values in \([1, 2]\) and grid-order \( k \)-values. Together with the known results for other norm-parameter values, we have almost complete knowledge of \( L_p(H_k^2) \) over the entire spectrum of possible norm-parameter values.

2. Our empirical study, which complements the analytical ones, shows that:
   (1) The analytical results are consistent with program verification over various norm-parameter \( p \)-values and sufficiently large grid-order \( k \)-values, and
   (2) As \( p \) increases over the real unit interval \([1, 2]\), the locations of candidate representative grid-point pairs agree with the intuitive interpolation effect over the two delimiting \( p \)-values.

3. A practical implication of our results on \( L_p(H_k^2) \) is that the exact formulas provide good bounds on measuring the loss in data locality in the index space, while spatial correlation exists in the 2-dimensional grid space.

We present a high-level approach to the main results without any derivations and proofs, supplemented with an empirical study that verifies the analytical results for various \( p \)-values and sufficiently large \( k \)-values. Complete results: illustrated figures, derivations, and proofs, and verifying computer programs are available from the authors.

2 Analytical Studies of \( L_p(H_k^2) \) with \( p \in [1, 2] \)

For 2-dimensional Hilbert curves, the 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 (quadrants) as \( Q_1(H_k^2) \), \( Q_2(H_k^2) \), \( Q_3(H_k^2) \), and \( Q_4(H_k^2) \). We extend the notion to identify all \( H_l^m \)-subcurves of a structured \( H_k^m \) for all \( l \in [k] \) inductively on the order in an obvious manner.

For a space-filling curve \( C \) indexing an \( m \)-dimensional grid space, the notation \("v \in C"\) refers to "grid point \( v \) indexed by \( C \)\", and \( C^{-1}(v) \) gives the index of \( v \) in the 1-dimensional index space. The locality measure in our study is for all reals \( p \geq 1 \),

\[
L_p(C) = \max_{\text{indices } i, j \in [m^n]} \frac{d_p(C(i), C(j))^m}{d_p(i, j)} = \max_{v, u \in C} \frac{d_p(v, u)^m}{|C^{-1}(v) - C^{-1}(u)|};
\]

When \( m = 2 \), we write \( \mathcal{L}_{C,p}(u, v) = \frac{d_p(u, v)^2}{|C^{-1}(v) - C^{-1}(u)|} \).
For subcurves \( C_1, C_2, C'_1, \) and \( C'_2 \) of \( C \), a grid-point pair \((v_1, v_2) \in C_1 \times C_2\) is reducible to a grid-point pair \((v'_1, v'_2) \in C'_1 \times C'_2\) if \( L_{C,p}(v_1, v_2) \leq L_{C,p}(v'_1, v'_2)\) — denoted by \((v_1, v_2) \preceq (v'_1, v'_2)\), and subcurve pair \( C_1 \times C_2 \) is reducible to subcurve pair \( C'_1 \times C'_2 \) if for every \((v_1, v_2) \in C_1 \times C_2\), there exists \((v'_1, v'_2) \in C'_1 \times C'_2\) such that \((v_1, v_2)\) is reducible to \((v'_1, v'_2)\) — denoted by \( C_1 \times C_2 \preceq C'_1 \times C'_2\). We define the strict reducibility, denoted by \( \prec \), for grid-point pairs and subcurve pairs via the strict inequality of \( L_{C,p} \)-values in an obvious manner.

A pair of grid points \( v \) and \( u \) indexed by \( C \) is representative for \( C \) with respect to \( L_p \), if \( L_{C,p}(v, u) = L_p(C) \), or, equivalently, for all \( v', u' \in C \), \((v', u') \preceq (v, u)\). The identifications of candidate representative grid-point pairs for \( C \) often involve sequences of reductions — successive considerations of two grid-point pairs and the comparisons of their \( L_{C,p} \)-values. Our studies of \( L_p(H^2_k) \) cover all norm-parameters \( p \geq 1 \). However, for all reals \( p \in (1, 2) \), the lack of geometric clarity for interpreting \( L_p \)-values can adversely increase the complexity: (1) of identifying candidate representative grid-point pairs, and (2) in comparing \( L_{H^2_{k,p}} \)-values for reductions due to the complex interplay of the norm-parameter \( p \)-value and grid-order \( k \)-value.

### 2.1 Reductions of Grid-Point Pairs and Subcurve Pairs

For two grid-point pairs \((v_1, v_2)\) and \((v'_1, v'_2)\) (two subcurve pairs \( C_1 \times C_2 \) an \( C'_1 \times C'_2 \) of \( H^2_k \)), the reduction \((v_1, v_2) \preceq (v'_1, v'_2)\) \((C_1 \times C_2 \preceq C'_1 \times C'_2\), respectively) eliminates \((v_1, v_2)\) \((C_1 \times C_2\), respectively) from the candidacy for representative grid-point pairs. We develop various sufficient conditions for reduction with an example below.

For the grid space \([2^k]^2\) of a 2-dimensional Hilbert curve \( H^2_k \) with a referenced \((x, y)\)-coordinate system (with origin \((1, 1)\)) in a canonical orientation (see Fig.1(a) and (b)), we denote the \( x \)- and \( y \)-coordinates of a grid point \( v \) by \( x(v) \) and \( y(v) \), respectively.

**Lemma 1.** For all norm-parameters \( p \in [1, 2] \) and three arbitrary grid points \( u, v, v' \in H^2_k \) such that: (1) the sequence of three grid points: \((u, v, v')\) is in indexing order (that is, \((H^2_k)^{-1}(u) \leq (H^2_k)^{-1}(v) \leq (H^2_k)^{-1}(v')\) or \((H^2_k)^{-1}(u) \geq (H^2_k)^{-1}(v) \geq (H^2_k)^{-1}(v')\)), and (2) the two sequences of their \( x \)- and \( y \)-coordinates: \((x(u), x(v), x(v'))\) and \((y(u), y(v), y(v'))\) have the same monotone property (both increasing or both decreasing), if \(|(H^2_k)^{-1}(u) - (H^2_k)^{-1}(v)| + |(H^2_k)^{-1}(v) - (H^2_k)^{-1}(v')| \geq 0 \) \((>|v(u) - v(v)||v(v) - v(v')| + |y(u) - y(v)||y(v) - y(v')| \geq 0 \) \((/>0)\), then \((u, v) \leq (u, v')\) \((u, v) \prec (u, v')\) via \( L_{H^2_{k,p}}(u, v) \leq L_{H^2_{k,p}}(u, v')\) \((L_{H^2_{k,p}}(u, v) < L_{H^2_{k,p}}(u, v'),\) respectively).

Note that the sufficient condition for the reduction is independent of the \( p \)-value for \( L_{H^2_{k,p}} \).

For reductions of grid-point pairs, we mostly use various \( p \)-independence sufficient conditions as the one in Lemma 1. For reductions of subcurve pairs, simple
ones are realized by symmetry arguments with regard to relative subcurve-orientations or succinct geometric interpretations of the $\mathcal{L}_{H_k^2,p}$-computation if possible.

For subcurves in the form of nested subquadrants of $H_k^2$, we may prove the reduction between subcurve pairs $C_1 \times C_2 \preceq C'_1 \times C'_2$ with a divide-and-conquer approach by considering all possible reductions between quadrant-subcurve pairs $Q_{i_1}(C_1) \times Q_{i_2}(C_2)$ (for all $i_1, i_2 \in [4]$) to $Q_{j_1}(C'_1) \times Q_{j_2}(C'_2)$ (for some $j_1, j_2 \in [4]$). Some reductions of quadrant-subcurve pairs may be resolved by simple symmetry/geometric arguments, while others may entail further reductions of subquadrant-subcurve pairs. These nested reductions generally arrive at some forms of recursive patterns, and mathematical induction is applied to resolve the reductions.

2.2 Identification of Candidate Representative Grid-Point Pairs

The upper-bound argument [5] in establishing the exact formulas for $L_p(H_k^2)$ for $p \in \{1, 2\}$ does not translate into a viable application for $p \in (1,2)$. For identifying all possible candidate representative grid-point pairs in $H_k^2$, we consider all grid-point pairs in $Q_i(H_k^2) \times Q_j(H_k^2)$ with $1 \leq i < j \leq 4$ and their possible systematic reductions. Due to a simple reduction ($Q_1(H_k^2) \times Q_4(H_k^2) \preceq Q_2(H_k^2) \times Q_3(H_k^2)$) and geometric symmetry ($Q_2(H_k^2) \times Q_4(H_k^2)$ to $Q_1(H_k^2) \times Q_3(H_k^2)$ and $Q_3(H_k^2) \times Q_4(H_k^2)$ to $Q_2(H_k^2) \times Q_1(H_k^2)$), three cases remain: $Q_1(H_k^2) \times Q_2(H_k^2)$, $Q_1(H_k^2) \times Q_3(H_k^2)$, and $Q_2(H_k^2) \times Q_3(H_k^2)$. An involved analysis of $Q_1(H_k^2) \times Q_3(H_k^2)$ reveals that the quadrant-subcurve pair is void of any candidate representative grid-point pairs.

We summarize the findings below in Theorem 1, in which the sources of (candidate) representative grid-point pairs (named $A$, $B$, and $C$) are illustrated in Fig. 2 and elaborated with (local) $(x, y)$-coordinates and $\mathcal{L}_{H_k^2,p}$-values in Table 1. For brevity we omit the symmetry ones.

**Theorem 1.** Consider the following cases determined by the interplay of the grid-order $k \geq 1$ and norm-parameter $p \in [1,2]$ of $H_k^2$:

1. **Case when $k = 1$:**
   
   $\forall p \in \{1,2\}$: One representative grid-point pair with coordinates $((1,1), (2^k, 2^k))$ and its symmetry.

   $\forall p = 2$: Three representative grid-point pairs with coordinates $((1,1), (1, 2^k), (1, 2^k, 2^k)), ((1,1), (2^k, 2^k))$, and $((1,2^k), (2^k, 2^k))$, and their symmetries.

2. **Case when $k \in \{2,3\}$:**

   $\forall p \in [1,2]$: One representative grid-point pair $B$ and its symmetry.

3. **Case when $k = 4$:** The $p$-interval $[1,2]$ is decomposed into two $p$-subintervals: $[1, \rho)$ and $(\rho, 2]$, where $\rho \approx 1.825$.

   $\forall p \in [1, \rho)$: One representative grid-point pair $B$ and its symmetry.

   $\forall p \in (\rho, 2]$: One representative grid-point pair $A$ and its symmetry.

For $p = \rho$: Two representative grid-point pairs $B$ and $A$, and their symmetries.
4. Case when \( k \geq 5 \): For all \( p \in [1, 2] \): \( 1+(k-2)+(k-4) = 2k - 5 \) candidate representative grid-point pairs \( B, C_1, D_1, C_2, \ldots, C_{k-5}, D_{k-5}, C_{k-4}, D_{k-4}, C_{k-3}, C_{k-2} \), and their symmetries.

Refined analysis with further reductions eliminates \( D_1, \ldots, D_{k-5} \) and \( D_{k-4} \), and their symmetries from the candidacy for representative grid-point pairs.

**Theorem 2.** For all grid-orders \( k \geq 5 \) and norm-parameters \( p \in [1, 2] \) of \( H_k^2 \), the candidate representative grid-point pairs are \( B, C_1, C_2, \ldots, C_{k-5}, C_{k-4}, D_{k-4}, C_{k-3}, C_{k-2} \), and their symmetries.

For all norm-parameters \( p \in [1, 2] \), there exists a sufficiently large grid-order \( k_0 \geq 5 \) such that for all grid-orders \( k \geq k_0 \) of \( H_k^2 \), the candidate representative grid-point pairs are \( B, C_1, C_2, \ldots, C_{k-5}, C_{k-4}, C_{k-3}, C_{k-2} \), and their symmetries.

Our future work will be focused on establishing analytically the association of representative grid-point pairs in \( \{B, C_1, C_2, \ldots, C_{k-5}, C_{k-4}, C_{k-3}, C_{k-2}\} \) with their \( L_{H_k^2,p} \)-dominance \( p \)-subintervals and relevant grid-orders.

### 3 Empirical Study on \( L_p(H_k^2) \) with \( p \in [1, 2] \)

To complement the analytical results for \( L_p(H_k^2) \) for all reals \( p \in [1, 2] \), we conduct an empirical study on \( L_p(H_k^2) \) for all \( k \in \{2, 3, \ldots, 12\} \) and some reals \( p \in [1, 2] \). We cover the grid space \( [2^k]^2 \) of a 2-dimensional Hilbert curve \( H_k^2 \) in a canonical orientation with Cartesian coordinates: \( 2^k \) columns (respectively, rows) indexed by \( x \)-coordinates (respectively, \( y \)-coordinates) \( 1, 2, \ldots, 2^k \). For every grid-order \( k \in \{2, 3, \ldots, 12\} \) and real \( p \in [1, 2] \) with granularity of 0.01 (for \( 2 \leq k \leq 12 \)), we locate with computer programs all representative grid-point pairs for \( H_k^2 \) with respect to \( L_p \). Figure 2(a) illustrates the three sources \( \{A, B, C\} \) of candidate representative grid-point pairs for \( k \geq 2 \).

Source \( A \) identifies the grid-point pair \( (u_A, v_A) = ((1, \frac{1}{4} \cdot 2k + 1), (1, 2^k)) \) and its symmetry. The pair \( (u_A, v_A) \) serves as the representative grid-point pair “briefly” – for \( k = 4 \) and \( 1.83 \leq p \leq 2.00 \).

Source \( B \) identifies the grid-point pair \( (u_B, v_B) = ((2k - 1, 1), (1, 2^k)) \) and its symmetry. The pair \( (u_B, v_B) \) serves as the representative grid-point pair for every \( k \in \{2, 3, \ldots, 12\} \) and all reals \( p \) of a (shrinking) prefix-interval \( [1, \rho_k) \subseteq [1, 2] \) – with \( \rho_k \) decreasing as \( k \) increases.

Source \( C \) identifies a sequence \( \{C_1, C_2, \ldots, C_{k-2}\} \) of grid-point pairs:

\[
C_t = (u_{C_t}, v_{C_t}) = ((\frac{1}{4} \cdot 2^k + 1, 2^{k-1} + 1), (\frac{3}{4} \cdot 2^k, 2^{k-1} + 2^{k-2} - 1))
\]

for \( t = 1, 2, \ldots, k - 2 \), and their symmetries, with:

\[
x(v_{C_{t+1}}) = x(v_{C_t}) \quad \text{and} \quad y(v_{C_{t+1}}) = 2^{k-1} - \frac{y(v_{C_t}) - 2^{k-1}}{2},
\]
and eventually $v_{C_t}$ converges to $v_{C_{k-2}}$. Note that, for $t = 0$, the grid-point pair $C_0 = (u_{C_0}, v_{C_0}) = ((\frac{1}{4} \cdot 2^k + 1, 2^{k-1} + 1), (\frac{3}{4} \cdot 2^k, 2^{k-1} + 2^{k-2}))$ is not included in $C$ since $C_0$ can not be a candidate representative grid-point pair (for any $k$ and real $p \in [1, 2]$):

$$\mathcal{L}_{H^2_k, p}(u_B, v_B) = \frac{(2^{k-1} - 1)^p + (2^k - 1)^p}{2^{2k-2}}$$

$$> \mathcal{L}_{H^2_k, p}(u_{C_0}, v_{C_0}) = \frac{(2^{k-1} - 1)^p + (2^k - 2 - 1)^p}{\frac{1}{3} \cdot 2^{2k-3} + \frac{1}{3} \cdot 2^{2k-4}}.$$

Empirically, for all $k \in \{5, 6, \ldots, 12\}$ and all reals $p$ of the (growing) suffix-interval $(\rho_k, 2] \subseteq [1, 2]$, all the representative grid-point pairs form a subsequence $C'$ of $C$ composed of: (1) a prefix of $C$ and (2) $(u_{C_{k-2}}, v_{C_{k-2}})$. The suffix-interval $(\rho_k, 2]$ is partitioned into disjoint successive $p$-subintervals, each of which supports a grid-point pair in the subsequence $C'$ as the representative grid-point pair for $H^2_k$ (for all reals $p$ of the subinterval). The length of $C'$ (number of all representative grid-point pairs from the source $C$) should depend on $k$ in general, and on the $p$-granularity in our empirical setting. Figure 2(b) depicts the sequence of candidate representative grid-point pairs from the source $C$.

![Fig. 2. Candidate representative grid-point pairs for $H^2_k$ with respect to $L_p$ for $k \geq 2$: (a) three sources $\{A, B, C\}$ of candidate representative grid-point pairs; (b) detailed view of the source $C$.](image)

Table 1 tabulates: (1) for each $k \in \{2, 3, \ldots, 12\}$, the partitioning $p$-subintervals of $[1, 2]$, and the corresponding representative grid-point pair and its source; and (2) $\mathcal{L}_{H^2_k, p}(u, v)$ ($= L_p(H^2_k)$) for a representative grid-point pair $(u, v)$ in the three sources $A$, $B$, and $C$:
<table>
<thead>
<tr>
<th>$k$</th>
<th>$p$</th>
<th>$(x, y)$-coordinates</th>
<th>representative grid-point pair coordinates in terms of $k$</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>[1.00, 2.00]</td>
<td>(2, 1), (1, 4)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>[1.00, 2.00]</td>
<td>(4, 1), (1, 8)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>1.00, 1.82</td>
<td>(8, 1), (1, 16)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>1.83, 2.00</td>
<td>(15, 1), (1, 16)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>[1.00, 1.61]</td>
<td>(16, 1), (1, 32)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>1.62, 2.00</td>
<td>(19, 7), (24, 17)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_3$</td>
</tr>
<tr>
<td>7</td>
<td>1.60, 1.51</td>
<td>(31, 2), (1, 64)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_4$</td>
</tr>
<tr>
<td>8</td>
<td>1.60, 1.41</td>
<td>(64, 1), (1, 128)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>1.42, 1.57</td>
<td>(33, 65), (96, 80)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_1$</td>
</tr>
<tr>
<td></td>
<td>1.58, 1.66</td>
<td>(33, 65), (96, 72)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_2$</td>
</tr>
<tr>
<td></td>
<td>1.67, 1.67</td>
<td>(33, 65), (96, 68)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_3$</td>
</tr>
<tr>
<td></td>
<td>1.68, 2.00</td>
<td>(33, 65), (96, 65)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_5$</td>
</tr>
<tr>
<td>10</td>
<td>1.00, 1.36</td>
<td>(128, 1), (1, 256)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>1.34, 1.58</td>
<td>(129, 257), (384, 320)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_1$</td>
</tr>
<tr>
<td></td>
<td>1.59, 1.69</td>
<td>(129, 257), (384, 288)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_2$</td>
</tr>
<tr>
<td></td>
<td>1.70, 1.75</td>
<td>(129, 257), (384, 272)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_3$</td>
</tr>
<tr>
<td></td>
<td>1.76, 1.77</td>
<td>(129, 257), (384, 264)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_4$</td>
</tr>
<tr>
<td></td>
<td>1.78, 2.00</td>
<td>(129, 257), (384, 257)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_5$</td>
</tr>
<tr>
<td>11</td>
<td>1.00, 1.32</td>
<td>(512, 1), (1, 1024)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>1.33, 1.58</td>
<td>(257, 513), (768, 640)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_1$</td>
</tr>
<tr>
<td></td>
<td>1.59, 1.70</td>
<td>(257, 513), (768, 576)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_2$</td>
</tr>
<tr>
<td></td>
<td>1.71, 1.76</td>
<td>(257, 513), (768, 544)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_3$</td>
</tr>
<tr>
<td></td>
<td>1.77, 1.79</td>
<td>(257, 513), (768, 528)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_4$</td>
</tr>
<tr>
<td></td>
<td>1.80, 1.80</td>
<td>(257, 513), (768, 520)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_5$</td>
</tr>
<tr>
<td></td>
<td>1.81, 2.00</td>
<td>(257, 513), (768, 513)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_6$</td>
</tr>
<tr>
<td>12</td>
<td>1.00, 1.31</td>
<td>(1024, 1), (1, 2048)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>1.32, 1.58</td>
<td>(513, 1025), (1536, 1280)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_1$</td>
</tr>
<tr>
<td></td>
<td>1.59, 1.70</td>
<td>(513, 1025), (1536, 1152)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_2$</td>
</tr>
<tr>
<td></td>
<td>1.71, 1.76</td>
<td>(513, 1025), (1536, 1088)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_3$</td>
</tr>
<tr>
<td></td>
<td>1.77, 1.80</td>
<td>(513, 1025), (1536, 1056)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_4$</td>
</tr>
<tr>
<td></td>
<td>1.81, 1.82</td>
<td>(513, 1025), (1536, 1040)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_5$</td>
</tr>
<tr>
<td></td>
<td>1.83, 2.00</td>
<td>(513, 1025), (1536, 1025)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_6$</td>
</tr>
<tr>
<td></td>
<td>1.90, 1.81</td>
<td>(513, 1025), (1536, 1012)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_7$</td>
</tr>
<tr>
<td></td>
<td>1.91, 1.80</td>
<td>(513, 1025), (1536, 980)</td>
<td>$(2^{k-1}, 1), (1, 2^k)$</td>
<td>C$_8$</td>
</tr>
</tbody>
</table>

Table 1. Representative grid-point pairs for $H_k^2$ with respect to $L_p$ for $k \in \{2, 3, \ldots, 12\}$ and $p \in [1.00, 2.00]$ with granularity of 0.01
\[
L_{H^2_k, p}(u, v) = \begin{cases} \\
\frac{(3 \cdot 2^{k-2} - 1)^2}{2^{2k-4} - 1} & \text{if } (u, v) \text{ is in } A \\
\frac{(2^{k-1} - 1)^p + (2^k - 1)^p}{2^{2k-2}} & \text{if } (u, v) \text{ is in } B \\
\frac{((2^{k-1} - 1)^p + (2^{k-2} - t - 1)^p)^2}{2^{2k-3} + \frac{1}{3} \cdot 2^k - \frac{1}{3} \cdot 2^{2k-4} - 2t} & \text{if } (u, v) = (u_{C_t}, v_{C_t}) \text{ in } C, \text{ where } t = 1, 2, \ldots, k - 2.
\end{cases}
\]

Figure 3(a) and (b) show the graphs, using the mathematical software Maple, of the locality measure \( L_{H^2_k, p}(u, v) \) for \( k = 4 \) and 12, respectively, for all reals \( p \in [1, 2] \) and all \((u, v)\) in the three sources \( A, B, \text{ and } C \). Our future work will involve determining, for each \( k \), the dominant functions/measures over successive subintervals of \( [1, 2] \), whose piece-wise combination yields the (overall) locality measure \( L_p(H^2_k) \) for all reals \( p \in [1, 2] \).

![Fig. 3. Locality measures corresponding to the grid-point pairs in: (a) A, B, and C = \{C_2\} for k = 4 and p-granularity of 0.01; (b) B and C = \{C_t \mid 1 \leq t \leq k - 2\} for k = 12 and p-granularity of 0.01. (Color figure online)](image)

For the extreme case of \( k = 4 \) with \( p \)-granularity of 0.01, two representative grid-point pairs emerge from the sources \( B \) and \( A \) over the partitioning subintervals \([1.00, 1.82]\) and \([1.83, 2.00]\), respectively.

For a more general case of \( k = 12 \) with \( p \)-granularity of 0.01, the representative grid-point pairs are from the sources \( B \) and \( C \) over the partitioning subintervals \([1.00, 1.31]\) and \([1.32, 2.00]\), respectively. Observe that the subsequence \( C' \) of all representative grid-point pairs (from the source \( C = \{C_t \mid 1 \leq t \leq k - 2\} \)) is \( \{C_1, C_2, C_3, C_4, C_5, C_6, C_{10}\} \).

4 Conclusion

Our analytical study of the locality properties of the Hilbert curve family, \( \{H^2_k \mid k = 1, 2, \ldots\} \), is based on the locality measure \( L_p \), which is the maximum ratio
of $d_p(u, v)^m$ to $d_p(\tilde{u}, \tilde{v})$ over all corresponding point-pairs $(u, v)$ and $(\tilde{u}, \tilde{v})$ in the $m$-dimensional grid space and index space, respectively. Our analytical results identify all the candidate representative grid-point pairs of $H^2_k$ from the three sources $A$, $B$, and $C$ (which realize $L_p(H^2_k)$-values) for all norm-parameters $p \in [1, 2]$ and grid-orders $k$, which enable us to have almost complete knowledge of $L_p(H^2_k)$ for all $p \geq 1$ – except for the relation between the candidate grid-point pairs and their dominance $p$-subintervals. For all real norm-parameters $p \in [1, 2]$ with sufficiently small granularity and grid-orders $k \in \{2, 3, \ldots, 12\}$, our empirical study reveals the three major sources $(A, B, \text{and } C)$ of representative grid-point pairs $(v, u)$ that give $L_{H^2_k, p}(v, u) = L_p(H^2_k)$. The results also suggest that all the representative grid-point pairs of $B$ and $C$ are from $B$ and $C'$, which is a prefix-subsequence of $C$ together with $C_{k-2}$ for some sufficiently large grid-orders $k \in \{5, 6, \ldots, 12\}$. The study has shed some light on a continuing study of determining the interplay pattern between the norm-parameter $p$ and grid-order $k$ for emerging representative grid-point pairs.

References

Algorithmic Aspects in Information and Management
11th International Conference, AAIM 2016, Bergamo, Italy, July 18-20, 2016, Proceedings
Dondi, R.; Fertin, G.; Mauri, G. (Eds.)
2016, XIV, 223 p. 36 illus., Softcover
ISBN: 978-3-319-41167-5