Pareto multi-objective optimization for high locality-preserving space-filling curve identification

Patrick Franco, Rémy Mullot, Valentin Owczarek

Pareto multi-objective optimization for high locality-preserving space-filling curve identification

Patrick Franco*, Rémy Mullot and Valentin Owczarek

La Rochelle University - Laboratoire Informatique, Image, Interaction (L3i) UMR EA 2118 - , Avenue M. Crepeau,, La Rochelle, 17000, France

ARTICLE INFO

Keywords:
Space-filling curves
New patterns
Locality preserving
Multi-objective optimization
Pareto dominance
Evolutionary process

ABSTRACT

Space-filling curves are widely used in applications that take advantage of the locality-preserving property. The pattern (the first order curve) plays a central role in locality preserving property. This was highlighted in previous works well-received by the community. A formulation was established leading to defined new patterns, including competitive patterns, i.e., patterns carrying comparable (and sometimes better) locality-preserving level than the Hilbert curve, so far the reference. Nevertheless, the number of pattern solutions resulting from the given formulation exponentially grows up with the space dimension.
In this article, with the help of an evolutionary algorithm, an original approach dedicated to the identification of high locality-preserving multidimensional patterns is proposed. Our idea is to embed the problem in a multi-objective optimization framework guided by Pareto optimality. In a such framework, each locality score (through a standard criteria) obtained by a pattern at a specific radius value can be processed as an objective function (to be minimized). The overall multi-objective function then illustrates how well the objectives are met, i.e. how the locality is achieved as progressively the space is filled. So, several space radii of interest are taken into account in the locality estimation and not a single one. This track contributes to define an accurately process of pattern identification.
Comparative experimental results led on dimensions upper than three, seem to confirm that the proposed approach is *Reliable*, *Efficient* and *Flexible*.
Finally, being able to identify patterns that preserve locality at a competitive level compare to the referent Hilbert curve (RBG pattern-based) constitutes a real contribution and could greatly improve the effectiveness of applications.

1. Introduction

Space-filling curves like Peano [1], Lebesgue [2], or Hilbert [3], provide a way to connect all points of a multidimensional discrete space (without crossing) while preserving spatial point proximity (locality) as much as possible. This property is known as locality-preserving or distance-preserving. It refers to the ability of a curve to map close multidimensional points to close 1-D indexes, which is useful to manage multidimensional data.The Hilbert curve, which is known to have superior locality-preserving properties compared to other curves [1, 2, 3], was originally introduced in 1891 [3] to fill a 2-D space and was guided by an implicit rule. Recently, a new formulation was exposed in [4, 5], in which each resulting solution defines a typical way to fill a multidimensional space. Among the set of solutions, classical curves from literature can be found, such as the original 2-D Hilbert curve, and for multidimensional purposes, the RBG-based curves[1] like Butz[6], Lawder [7] or Jin[8]. However, other solutions have been discovered that realize a multidimensional extension of the popular Hilbert curve. Furthermore, it was shown in [5] that among the set of solutions now accessible, a subset can bring a comparable (and sometimes better) level of locality-preserving compared to the Hilbert curve.

In this article, an approach is proposed to extract good solutions, i.e. to identify the subset of D-Dimensional curves (derived from the formulation in [5]) which present an interest from a locality-preserving point of view, compared to the Hilbert curve, which is currently the reference.
The number of solutions resulting from [5] exponentially grows up with the space dimension that needs to be filled. On the one hand, this opens up the possibility of finding a lot of good solutions but clearly requires overcoming the

*Corresponding author
patrick.franco@univ-lr.fr (P. Franco)
ORCID(s):
[1] In binary coding, curves built from the Reflected Binary Gray code.

expensive computational cost of their identification. On the other hand, it could be relevant to classify the solutions according to how the locality is conserved as progressively the space is filled. So, several space radii of interest are considered here, this contributes to define an accurately process of identification.
Our idea is to embed the problem of the identification of highly locality-preserving curves in a Pareto Multi-objective optimization framework. In such framework, each locality-preserving score (through the standard Faloutsos criteria [9]) obtained by one solution at a specific space radius of interest can be processed as an objective function (to be minimized) . The overall multi-objective function then illustrates how well the locality-preserving property is achieved as deeply the space is filled. Thanks to the Pareto domination (dominated/non dominated solutions), it is possible to establish a classification between solutions according to how the objectives are met. This is a real contribution because it is possible to access :

- curves with a high-locality profile over successive radius values (Pareto optimal solution),
- curves with the maximum level of locality for a given space radius of interest (the best non-dominated solution for one query radius),
- new curves not yet previously identified in [5] (and to our knowledge in the literature).

Our proposition is organized as follows.
Section 2 presents a background on the Space-filling curves, including a detailed explanation of the locality-preserving property and a discussion on the role of the pattern in preserving locality. In Section 3 the promoted Pareto Multi-objective optimization framework is introduced. We provide details on the guiding idea, and we justify the well-founded framework by completing the problem formalization.
We note that only linearly independent solutions (i.e., curves non related by linear geometric transformations or isometry-free curves) can lead to different levels of locality. So, we need to detect if two solutions are linearly dependent in order to reject one of them. If we do not reject one of them, the locality will not be improved and consequently it will not future be Pareto optimal. In Section 3.3, we develop an isometry-free detector and estimate its computational cost.
Finally, we propose a powerful approach for identifying curves with a high level of locality-preserving. This approach involves filtering solutions using the isometry-free detector (Section 3.3), and then merging the selected solutions in a framework that is able to classify and arbitrate between them using the Pareto Multi-objective optimization (Section 3). Figure 5 summarizes this approach. Experimental results obtained on dimensions upper than three are presented in Section 4. These results are compared to classical multidimensional Hilbert curves (precisely, RBG-based curves like Butz [6], Lawder [7] or Jin [8]) and they tend to confirm our assumptions.

2. Multi-dimensional Space Filling curves

2.1. Background, Locality Preserving and Pattern

A Space-filling curve is a self-similar path which goes through every point of a given multidimensional discrete space. The curve connects all the D-dimensional points belonging to a regular grid space without crossing itself (each point is visited only once). For example, the Peano [1] and Hilbert [3] curves filling a $2-D$ space are illustrated in Figure 1. Many references are available on space-filling curves, to have an overview the reader can consult [10, 11]. Historical aspect are discussed starting with Kantor or Peano's first curve, passing through Hilbert's response, to more recent proposition (such as the Onion curve [12]).

An interesting aspect of this kind of curves is that a D-Dimensional point cannot necessary be represented (localized) with its space coordinates as usual, but by its position along the curve ($1-D$ index) when the curve is visiting the space (cf. Figure 1). Therefore, constructing a space-filling curve S_n provides a way to define a bijection (cf. Equation 1) between multi-dimensional points P that belong to the D-dimensional grid space and their $1-D$ indexes I on the curve :

$$\begin{aligned} S_n : [0, 2^n - 1]^D &\rightarrow [0, 2^{nD} - 1] \\ P &\longmapsto I = S_n(P) \end{aligned} \tag{1}$$

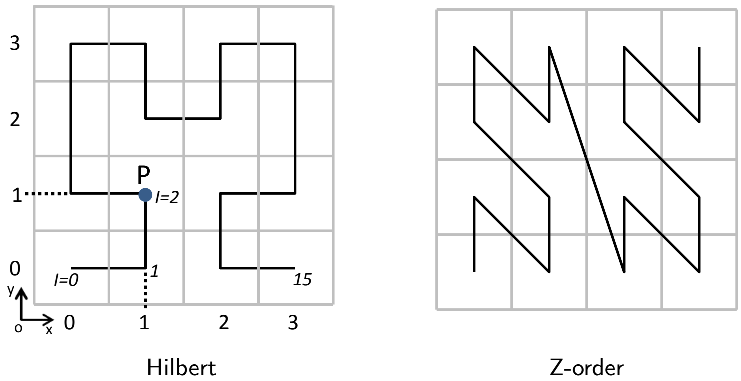

Figure 1: Illustration of space-filling curves including the famous Hilbert and Z-order curves filling a $2-D$ space. Configuration: the space dimension is $D = 2$, the space resolution is $n = 2$ (2^n points per dimension, here 4) so 2^{nD} points P to be connected (here 16). The space resolution n qualifies the order of the curve. Observations: (a) All the points are connected, and each curve define a typical path to fill the same $2-D$ space, and (b) The point P with coordinates $(x = 1, y = 1)$ can be now localised with its $1-D$ index $(I = 2)$ on the curve.

with P, a D-dimensional point and I its corresponding $1-D$ index through the n order curve S_n.

Sorting multi-dimensional points according to their $1-D$ position on the curve is an attractive idea. Firstly, handling $1-D$ data offers a real advantages in term of algorithmic complexity, which naturally interests numerous applications. Secondly, analyzing the distribution of $1-D$ indexes can reveal information about the D-dimensional data points, or can give access to a view of the data landscape, making it useful for data analysis applications.

However, we emphasize that all these tracks are reliable under the assumption that distances between indexes and points are conserved. In other words, we must ensure that through a given bijection S_n, close $1-D$ indexes correspond to close D-dimensional points (and vice versa), which is the property of conservation of locality (locality-preserving).

For example, the Faloutsos criterion [9] is a standard measure to estimate the level of locality conservation of a curve.

$$L_r(S_n) = \frac{1}{2^n D} \sum_{k,l \in [2^n D],\ k<l,\ \mathfrak{d}(k,l) \leq \mathbf{r}} \max\{\mathfrak{d}(P(k), P(l))\} \tag{2}$$

$L_r(S_n)$, is level of locality preservation achieved by a n order curve S_n, that fills a D-dimensional space. $L_r(S_n)$, is estimated under a given metric $\mathfrak{d}$, and for a specified radius value r. $L_r(S_n)$ is a relevant measure because it checks whether two close indexes k and l correspond to points $P(k)$ and $P(l)$ effectively located not too far away in space ($max()$). In other words, $L_r(S_n)$ verifies if indexes within a radius of interest r correspond to D-dimensional points that are really inside the same radius in space. As $L_r(S_n)$ tends towards 1, better is the level of locality preservation.

Several comparative studies have been conducted on different space-filling curves (such as Peano, Hibert, Gray, Scan, Sweep etc.), and they have concluded that the Hilbert curve is the curve that maximizes the level of locality preservation [13, 9, 14, 15, 16].

Nevertheless, a curve going through every point of a given multidimensional discrete space, without crossing, without favouring any direction, and that would completely preserve locality does not exist. Topology breaks between D-Dimensional and $1-D$ spaces remain and induce a loss of locality, which is taken into account by the Faloutsos criterion. Illustrations of topology breaks on space-filling curves are presented in Figure 2, including the recently proposed Onion curve [12].

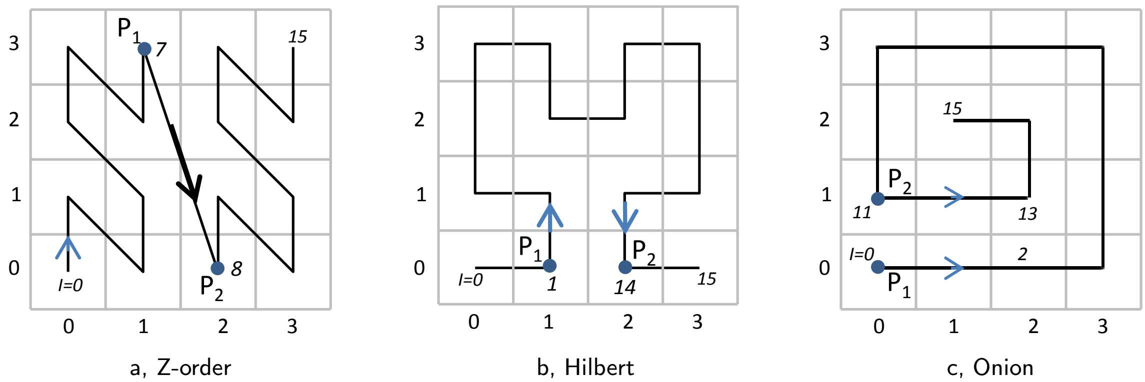

Figure 2: Illustrations of topology breaks on Space-Filling Curves that result in a loss of locality. In (a), two close indexes $I_1 = 7$ and $I_2 = 8$ correspond to points P_1 and P_2 far away in space this is expressed by a jump in the grid space. In (b), two close points P_1 and P_2 corresponding to split indexes $I_1 = 1$ and $I_2 = 14$, which is the worst case for the Hilbert curve. In (c), similar cases can be observed on the Onion curve recently proposed [12].

2.2. Role of the pattern on locality-preserving

In previous works [4, 5], our focus was on the role of the pattern (i.e., the first order curve $n = 1$) in designing a curve with an overall competitive level of locality preserving. We have shown that defining a good pattern - based on locality preserving criteria - is a key to design a curve bringing, across order ($n = 2, 3, ...$), an interesting level of locality. This is due to self-similarity property of space-filling curves. This is illustrated in Figure 3 for both Hilbert and Z-order curves.

Precisely, starting from a D-Dimensional basic pattern at the first order $n = 1$, the pattern is copied $2^{(n-1)D}$ times (and sometimes transformed) to built a $n = 2, 3$ order curve. For example, for $n = 2$, the 2-Dimensional Hilbert curve is set up from $2^D = 4$ copies of the the basic pattern and 16 copies at order $n = 3$. Because the basic pattern is (i) the initialisation of the process of construction, so it contributes to determine 2^D points ordering and, (ii) is widely copied through orders, so the pattern fixes the ordering of $2^{(n-1)D}$ sub-spaces. Therefore, the choice of the pattern plays a central role in locality preserving.

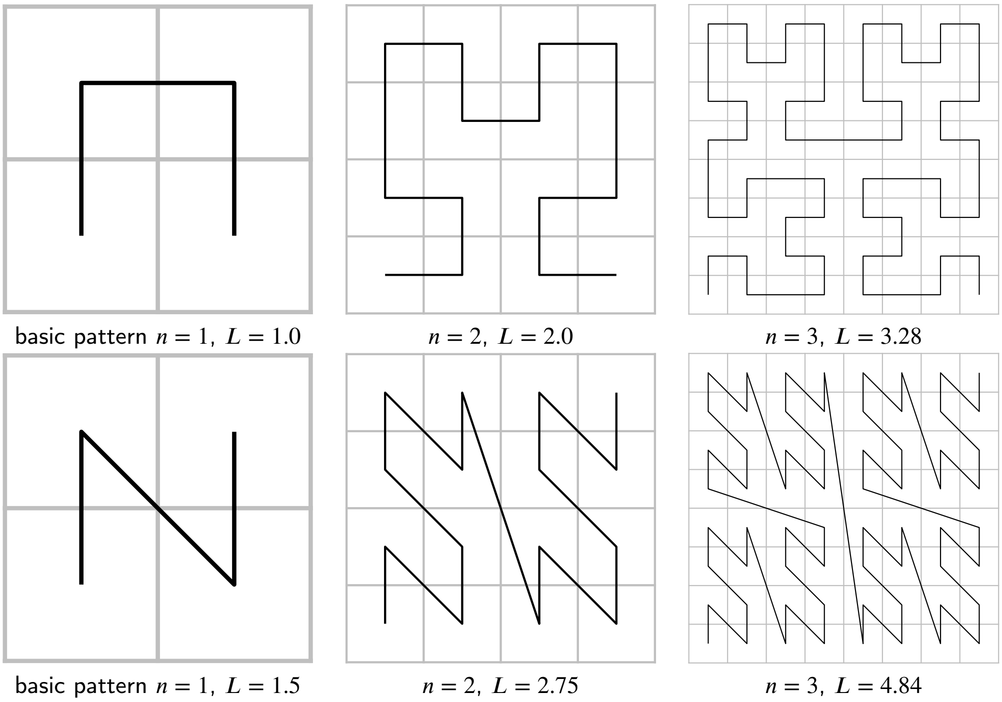

Figure 3: Illustration of the role of the pattern (the first order curve $n = 1$) in locality preservation. Selecting a good locality-preserving pattern leads to design, through orders, an overall better locality preserving curve. Illustration on the $2 - D$ Hilbert and Z-order space-filling curves for $n = 1, 2,$ and 3. The level of locality-preservation of the n order curve, estimated using the standard Faloutsos criteria described in Section 2.1 (at unit radius $r = 1$, under Manhattan metric), is denoted by L. An L value close to 1 indicates better locality preservation. Compared to Z-order curve, the Hilbert curve exhibits superior locality preservation at the initial order (50%) that is maintained across higher orders (37.5%, 47.5%).

3. Contribution : Multi-objective Pareto optimization for high locality-preserving pattern identification

3.1. Motivations, Idea, and justifications

As justified in Section 2.2, the choice of high locality-preserving pattern contributes to designing, trough orders, a curve with an overall interesting level of locality. In [4, 5], a study was oriented toward the Hilbert curve, i.e. toward the curve known to bring superior locality score compared to other curves [13, 9, 14, 15, 16]. Particularly, the rule witch drives the construction of the original $2-D$ Hilbert[2] curve was revisited, and a multidimensional extension was proposed. Precisely, a formulation to define adjacent patterns was exposed, leading to the identification of alternative patterns bringing comparable and sometimes better locality preserving scores than the Hilbert curve[3]. For example, one alternative pattern that reached the same locality-preserving score as the classical RBG is presented in Table 1 ($4-D$ case).

However, when the space dimension increases, there are numerous patterns that can satisfy the formulation given in [5]. Precisely, for $D > 2$ several paths exist to connect 2^D points, without crossing itself, without prioritizing any space direction, under the constraint that two successive points are separated by an unit distance, which could potentially impact the level of locality.

Consequently, some remarks can be pointed out :

- *Numerous returnable solutions :*
  Being able to access several patterns opens the possibility to discover alternative way to fill a space, and competitive patterns can emerge (cf. Table 1). However, the number of patterns resulting from (derived from) the formulation in [5] exponentially grows with the space dimension. The number of possible solutions is upper-bounded[4] by $2^{D-1} \times (D-1)^{2^D-2}$. Then, an exhaustive search to identify good locality-preserving patterns is clearly not reasonable. To bypass this limitation, we propose here to investigate the development of an optimization approach founded on evolutionary algorithm (EA).
  Starting with tuning procedures in early 90s, evolutionary algorithms dedicated to multi-objective optimisation problems (MOEA) are now an established field of research. From the use of genetic algorithms in [17], to the role of elitism strategy on optimality in [18], through the attempt to define an unified model in [19] (including archiving and re-insertion strategies), the reader can find a state of the art covering the last 30 years in [20, 21]. Related to our problem, EA have yet demonstrated their capabilities to find optimal (or near-optimal) solutions, especially in difficult contexts. For example, when the function to be optimized is not continuous and/or not derivable. That is the case here.
  Firstly, an analytic model linking a solution (pattern) and its locality score (Faloutsos criteria, cf. Equation 2) is not easily accessible, and discontinuities on locality values can occur. Consequently, that avoid the use of traditional optimization approaches (such as gradient descent or simulated annealing). Secondly, an algorithm for generating patterns for a given space dimension is available in [5]. Therefore, a set of initial solutions can be generated (initial population) that constitute the right initial condition to start the evolutionary algorithm (EA), and evaluation, selection, crossover, and mutation operators can then be applied. This approach leads to discover new high locality-preserving patterns that have not yet been identified in [5].

- *Locality-preserving through several space radius of interest :*
  The locality-preserving property is here not only estimated for a single radius of interest (cf. Section 2, Equation 2) but for successive radius values. This expresses the ability of a pattern to conserve locality as progressively the space is filled. *This is suitable for applications where good locality score is required for a specific radius of interest, usually a short one in order to detect if data are located close to a D-dimensional query point. However, applications may also require good scores for successive radius values, for example, in range query problems (range query matches).* Therefore, our idea is to embed the identification of high locality-preserving pattern in a multi-objective Pareto optimization framework, where each objective is the locality preservation achieved at a specific radius value Lr. Each Lr is then processed as an objective function to be minimized. Patterns solutions previously obtained via EA (as mentioned above) can then be classify according to the Pareto

[2] Hilbert curve [3] was introduced in 1891 to originally fill $2-D$ space (a plane) and it was guided by an implicit rule.
[3] Rigorously compare to the RBG pattern, i.e the multidimensional Hilbert-based RBG pattern as we can find in [6], [7] or [8].
[4] When all the solutions start to the same first point 0^D.

Pattern	Locality score: L_r L_1	L_2	L_3	L_4	Points list
RBG	1.0	2.0	2.75	2.75	0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0
Alternative Pattern $\mathcal{P}$	1.0	2.0	2.75	2.75	0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0

Table 1
Example of two $4-D$ patterns, which are solutions of the formulation presented in previous works [5]. Each pattern define a typical way to fill a space, here a $4-D$ space (see Points List). The RBG pattern is a classical one in literature used to build a multidimensional extension of the Hilbert curve. $\mathcal{P}$ is an alternative pattern. L_r represents the level of locality score (Faloutsos criteria, cf. Equation 2, under Manhattan metric) reached by a pattern for several radii of interest, $r = 1, 2, 3, 4$. Both patterns lead to an identical level of locality preservation across radii. So, competitive patterns can emerged.

domination (dominated and non-dominated solutions). This track leads to the identification of patterns according to their capabilities to satisfy objectives, i.e. to their abilities to preserve locality as the space is explored more deeply.

- *All patterns do not improve locality :*
  As mentioned in [5], not all patterns improve the locality-preserving property. Patterns linked by a linear transformation reach the same locality score, so locality is not improved. In Figure 6, an example of $3-D$ patterns related by a linear transformation (Isometry as reflection, or rotation) is illustrated.
  Therefore, in the evolutionary process of construction of new patterns proposed, we need to detect if two solutions are linearly dependent to reject one of them because the locality will not be improved, and consequently it will not be Pareto optimal. An isometry-free detector is developed in Section 3.3, and its computational cost is estimated. This approach seems to ensure a good convergence of the process towards interesting solutions from locality-preserving objectives.

The proposed approach (Synthesized in Figure 5) leads to discover new high locality-preserving patterns, that have not yet been identified with the help of the algorithm in [5]. This is confirmed by the experimental results provided in Section 4.
In the next Section, details about the promoted framework, including formalization, Pareto optimality, and archive update rule are given.

3.2. Pareto Multi-objective Optimization for High Locality-preserving Pattern Identification: formalization

The standard formalization of a multi-objective optimization problem has been adapted and applied to our challenge, i.e., to the task of identifying D-dimensional high locality-preserving pattern $\mathcal{P}$. In particular, the identification process is not limited to a fixed radius of interest r, but encompasses several radius in order to evaluate locality preservation as the space is increasingly explored (filled).

Figure 5 provides an overview of the proposed approach. The identification of high locality-preserving pattern involves the search of a solution $\mathcal{P}$ that *minimizes* :

$$\mathbf{L}(\mathcal{P}) = [L_1(\mathcal{P}), L_2(\mathcal{P}), ..., L_r(\mathcal{P}), ..., L_N(\mathcal{P})] \tag{3}$$

L_r represents the objective functions that need to be minimized, which are the Locality Faloutso's criteria (cf. Equation 2) estimated at each radius of interest $r \in [1, 2, ..., N]$.
N is the number of radius r considered, which is arbitrarily set to $N = 2^D/4$, thereby fixing the number of handled objective functions.
$\mathcal{P}$ represents a pattern solution, which is a list of 2^D sorted D-Dimensional points respecting the multi-dimensional Hilbert extension in [5]. $\mathcal{P}$ is expressed in binary coding as a sequence of 2^D binary words (cf. Table 1), which is provided iteratively through an evolutionary process and classified based on Pareto optimality (as described in Equation 4).
To summarize, $\mathbf{L}(\mathcal{P})$ is a vector composed of the N Faloutsos 's score functions L_r obtained by one solution $\mathcal{P}$ for successive radius values $r \in [1, 2, ..., N]$.
A priori, it is possible that a solution $\mathcal{P}$ which minimizes all objective functions L_r may not exit, and discontinuities in the L_r score may also be observed. Pareto optimality provides a framework for defining classes of solution based on the concept of dominant and dominated solutions. Let $\mathcal{P}^*$ and $\mathcal{P}$ be two solutions of (3), $\mathcal{P}^*$ is said to dominate $\mathcal{P}$ if the following conditions are satisfied :

$$\begin{array}{c} (1) \quad L_r(\mathcal{P}^*) \leq L_r(\mathcal{P}) \textit{ for all radius } r = 1, 2, ..., N \\ (2) \quad L_r(\mathcal{P}^*) < L_r(\mathcal{P}) \textit{ for at least one radius } r,\ r \in [1, 2, ..., N] \end{array} \tag{4}$$

A solution $\mathcal{P}^*$ is Pareto optimal if it is not dominated by any other solutions.
All of the Pareto optimal solutions are grouped into a set called Pareto front. Depending on the dimension of the solution space (i.e., the number of objective functions being handled), the Pareto front is an hypersurface on which no solution lies below. Otherwise, it would be considered Pareto optimal according to Equation 4.

Regarding our proposal, all the Pareto optimal solutions (patterns) observed at each iteration will be stored in an archive called the Pareto archive (see Figure 5). The Pareto archive is then updated across iterations, and its content evolves from a population of solutions to population until convergence of the algorithm. Practically, the archive is updated with the following rule : the non dominated candidates observed at the current iteration are added to the archive, and all dominated solutions previously stored are removed. In the field of evolutionary algorithms, this is a fairly common population-based elitism search. For example, a Pareto Archive obtained from the application of the proposed approach is illustrated in Figure 4.

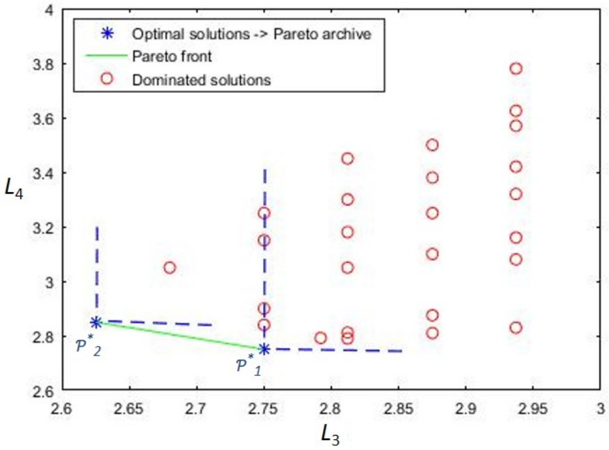

Figure 4: Application of the proposed approach, synthesized in Figure 5, to identify hight-locality Pattern $\mathcal{P}$. Example of Pareto archive obtained after the algorithm converges in the $4-D$ case. The approach involves minimizing $N = 4$ objectives functions, $\mathbf{L}(\mathcal{P}) = [L_1(\mathcal{P}), L_2(\mathcal{P}), L_3(\mathcal{P}), L_4(\mathcal{P})]$. The solutions are illustrated in the plane $[L_3, L_4]$, as the Pareto optimal solutions found reach the same score for $L_1(\mathcal{P}) = 1.0$ and $L_2(\mathcal{P}) = 2.0$. After convergence, two solutions $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ dominate all other solutions previously stored in the archive, with respective score of $\mathbf{L}(\mathcal{P}_1^*) = [1.0, 2.0, 2.75, 2.75]$ and $\mathbf{L}(\mathcal{P}_2^*) = [1.0, 2.0, 2.62, 2.87]$. $\mathcal{P}_1^*$ and $\mathcal{P}_2^*$ are two Pareto optimal patterns. Among the two identified patterns, it will be shown later that $\mathcal{P}_1^*$ is the RBG pattern already known, and $\mathcal{P}_2^*$ is a new one, not previously identified in [5].

In the evolutionary process of construction of new patterns solutions, we need to detect if two solutions are linked by a linear geometric transformation (e.g., isometry as reflection, and/or rotation) in order to reject one of them. This is because the locality will not be improved, and consequently, it will not be Pareto optimal. An isometry-free detector is developed in Section 3.3. Contextually, in the multi-objective Pareto optimization framework (as promoted in Figure 5), an isometry-free test between pattern candidates (the current solutions) and non-dominated solutions (previously stored in the Pareto archive) is performed at each iteration (population).

First, we emphasize that only linearly independent candidates (isometry-free patterns) are taking into account, meaning only candidates that could potentially become Pareto optimal. Secondly, regarding the algorithmic cost, we can note that the isometry-free test is computed before the evaluation stage ($\mathbf{L}(\mathcal{P})$, cf. Equation 3) and before the Pareto classification (cf. Equation 4). This track seems to ensure a rather rapid convergence of the archive towards interesting solutions from locality-preserving point of view. Practically, the algorithm is stopped when no new solutions appear in the Pareto Archive (stabilized Pareto front).

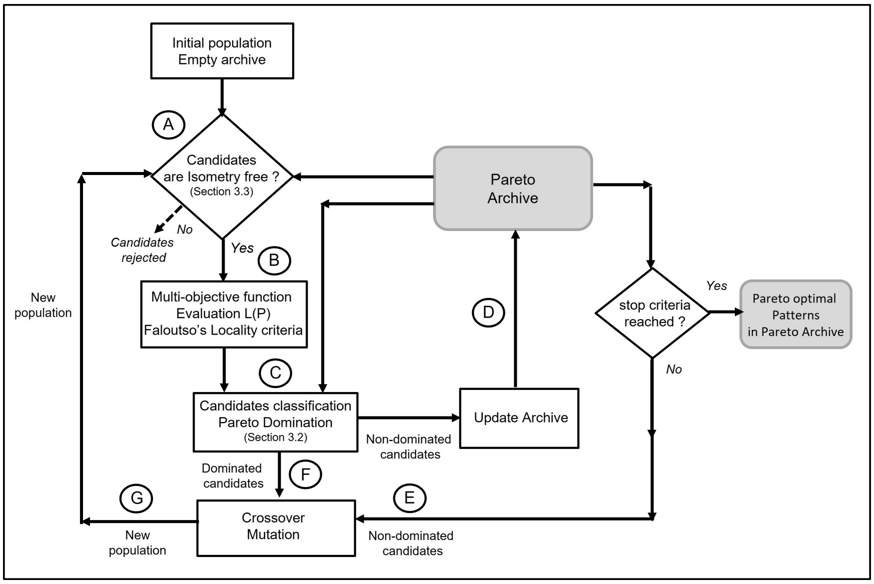

Figure 5: Pareto Multi-objective Optimization for High Locality-preserving Pattern Identification: Overview of the Proposed Approach.

3.3. Prediction of non-Pareto optimal solutions : Isometry-free detection

The detection of a geometrical link, such as linear transformation or isometry, between patterns is a complex task but is needed if we want to find out Pareto optimal solutions. The problem can be defined as follow, let $\mathcal{P}$ and $\mathcal{P}'$ be two D-dimensional patterns. In the proposed approach (cf. Figure 5), $\mathcal{P}$ is a candidate that arises from the current population, while $\mathcal{P}'$ is a non-dominated solution previously stored in the Pareto archive. The goal is to determine whether an isometry, denoted $\mathfrak{Iso}$, exists between $\mathcal{P}$ and $\mathcal{P}'$, such that $\mathfrak{Iso}(\mathcal{P}) = \mathcal{P}'$. An isometry is a linear transformation that consists of two basic operations: a reflection (denoted $\mathfrak{Ref}$) and a rotation (denoted $\mathfrak{Rot}$). When applied to a pattern $\mathcal{P}$, these operations change the coordinate values (in binary coding, the digits) of the D-dimensional points that make up the pattern $\mathcal{P}$. Details on reflection and rotation operations can be found in [4, 22]. If an isometry is detected, $\mathcal{P}$ will be rejected, because it is a linearly dependent solution. Consequently it cannot improve $\mathbf{L}(\mathcal{P}')$ and it cannot then become Pareto optimal.
To determine whether an isometry $\mathfrak{Iso}$ links $\mathcal{P}$ and $\mathcal{P}'$ or not, one naive process is to compute all possible $\mathfrak{Iso}$. This operation is composed of two sub-tasks, depending on the number of all possible reflections $\mathfrak{Ref}$ and rotations $\mathfrak{Rot}$. The number of all possible reflections is equal to the sum of the combinations composed of a set of D elements :

$$\begin{aligned} \mathbf{card}(\mathfrak{Ref}) &= 1 + \binom{D}{1} + \binom{D}{2} + ... + \binom{D}{D} \\ &= 1 + \sum_{i=1}^{D} \binom{D}{i} \\ &= 1 + 2^D - 1 \\ &= 2^D \end{aligned} \tag{5}$$

The number of all possible rotations is equal to the number of possible permutations in a set of D elements :

$$\mathbf{card}(\mathfrak{Rot}) = D \times (D-1) \times ... \times 1 = D! \tag{6}$$

Then, the total number of possible isometries is equal to :

$$\mathbf{card}(\mathfrak{Iso}) = D! \times 2^D \tag{7}$$

Therefore, such a naive algorithm would work with a prohibitive complexity $O(D! \times 2^D)$.
In order to reduce algorithmic complexity, an isometry-free test is set-up from an original signature, noted $\mathbf{Q}$. Ideally, $\mathbf{Q}$ should satisfy desirable invariance properties, such as :

$$\begin{aligned} &\mathbf{Q}(\mathcal{P}) \neq \mathbf{Q}(\mathcal{P}^{'}) \text{ if } \mathcal{P}, \mathcal{P}^{'} \textit{ are isometric-free patterns} \\ &\mathbf{Q}(\mathcal{P}) = \mathbf{Q}(\mathcal{P}^{'}) \text{ if } \mathcal{P}, \mathcal{P}^{'} \textit{ are non isometric-free patterns} \end{aligned} \tag{8}$$

Referring to Equation 3, $\mathcal{P}$ and $\mathcal{P}^{'}$ are two sequences of 2^D binary words, each composed of D digits (cf. Table 1 in the $4-D$ case).
The proposed signature $\mathbf{Q}$, computed on a pattern $\mathcal{P}$ denoted $\mathbf{Q}(\mathcal{P})$, is the sorted set of cardinality D, where each element of the set q_n is the count list of consecutive equal digits c_1^n observed in each n dimension. Specifically, $\mathbf{Q}(\mathcal{P}) = \{q_1, q_2, ..., q_n, ..., q_D\}$, where q_n is equal to $[c_1^n, c_2^n, ..., c_m^n]$.
To compute $\mathbf{Q}(\mathcal{P})$, q_n is first computed for $1 < n \leq D$, and then the q_n are sorted according to lexical order (or dictionary order). That is, let q_i and q_j be equal to $[c_1^i, c_2^i, ..., c_m^i]$ and $[c_1^j, c_2^j, ..., c_l^j]$ with $1 < i, j \leq D$,

$$\begin{aligned} &q_i \leq q_j \text{ if } \forall\ k \in [0, \min(m, l)]\ c_k^i \leq c_k^j \\ &\qquad\qquad \text{else } q_i > q_j \text{ in other cases.} \end{aligned} \tag{9}$$

The signature $\mathbf{Q}(\mathcal{P})$ is defined to be invariant to reflection and rotation, as confirmed by the respect of the Equation 8.
Supplementary explanations on invariance properties are given below.
A didactic example of the calculation of the proposed $\mathbf{Q}$ signature can be found in Appendix 8.

Reflection-invariant: The reflection operation $\mathfrak{Ref}$, applied according to a given axis on a pattern $\mathcal{P}$ (denoted $\mathfrak{Ref}(\mathcal{P})$), changes every digits of $\mathcal{P}$ located at the given axis (1 into 0 and 0 into 1). As a result, q_n will not change, leading to $\mathbf{Q}(\mathfrak{Ref}(\mathcal{P})) = \mathbf{Q}(\mathcal{P})$.

Rotation-invariant: The rotation operation $\mathfrak{Rot}$, applied on a pattern $\mathcal{P}$ (denoted$\mathfrak{Rot}(\mathcal{P})$), creates a permutation of the coordinates of $\mathcal{P}$. Consequently, only the order of q_n is changed, leading to $\mathbf{Q}(\mathfrak{Rot}(\mathcal{P})) = \mathbf{Q}(\mathcal{P})$ after lexical ordering.

The proposed isometry-free test detection based on $\mathbf{Q}$ signature is illustrated in Figure 6 for three patterns $\mathcal{P}$, $\mathcal{P}^{'}$ and $\mathcal{P}^{''}$ linked by a geometric linear transformation (a reflection and a rotation, respectively), where the RBG pattern $\mathcal{P}$ is taken as reference.
Finally, with the help of the Isometry-free detection, it is possible to filter out linearly dependent candidates, focusing instead on pattern solutions that can potentially improve locality and become Pareto optimal. Moreover, we can note that the complexity of the proposed signature is equal to the time to browse all the points of the candidate and the time to sort the different q_n, resulting in a complexity of $O(D + D\log(D))$. This achieves a significant algorithmic complexity gain compare to an exhaustive search of isometry (as shown in Equation 7).

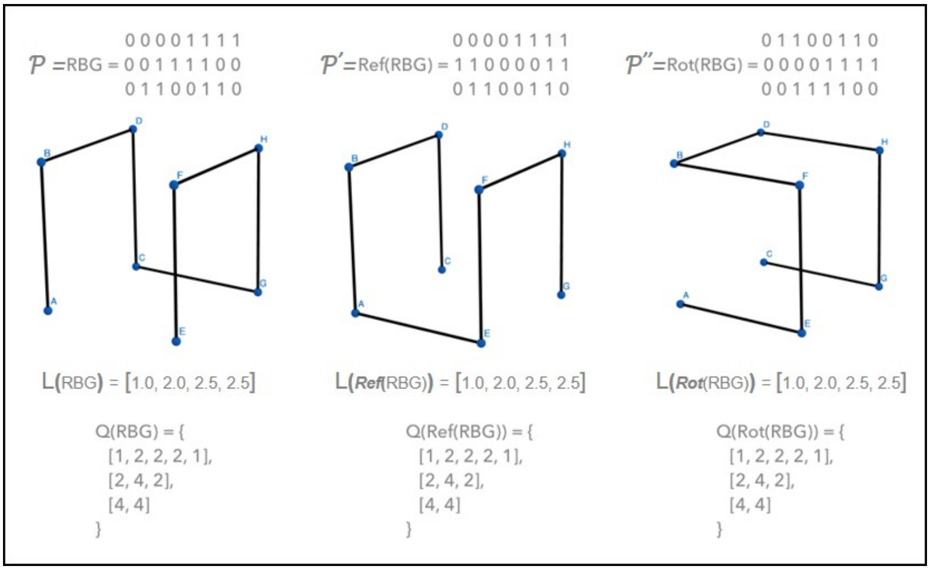

Figure 6: Illustration of the proposed Isometry-free detection. Example of three patterns $\mathcal{P}$, $\mathcal{P}'$ and $\mathcal{P}''$ linked with a linear transformation (respectively $\mathfrak{Ref}$, and $\mathfrak{Rot}$) in a $3-D$ case. As predicted, they reach the same level of locality $\mathbf{L}(RBG) = \mathbf{L}(\mathfrak{Ref}(RBG)) = \mathbf{L}(\mathfrak{Rot}(RBG)) = [1, 2, 2.5, 2.5]$, and objective functions L_r are not improved. The linear dependence between $\mathcal{P}$, $\mathcal{P}'$ and $\mathcal{P}''$ is well detected by the $\mathbf{Q}$ signature : $\mathbf{Q}(RBG) = \mathbf{Q}(\mathfrak{Ref}(RBG)) = \mathbf{Q}(\mathfrak{Rot}(RBG))$.

4. Pareto Multi-objective Optimization for High Locality-preserving Pattern Identification: Results

4.1. Details of the algorithm and Experimental conditions

In this Section, the capabilities of the proposed approach (cf. Figure 5) to identify high locality multidimensional patterns is assessed. However, for clarification purposes, firstly details of the algorithm and experimental conditions are given below.

The algorithm starts from an initial population and an empty archive. The initial population is a set of initial Pattern solutions provided thanks to the algorithm in [11]. The cardinality, noted I, was adapted according to the targeted space dimension ($I = 20$ and $I = 40$ respectively for $4-D$ and $5-D$ cases).

The function to be optimized is $\mathbf{L}(\mathcal{P})$ (Evaluation Function, cf. Equation 3). As a reminder, $\mathbf{L}(\mathcal{P})$ is a vector composed of the N objectives function to be minimized $L_r(\mathcal{P})$. Each $L_r(\mathcal{P})$ evaluate the level of locality - according to the Faloutsos criteria (cf. Equation 2, under Manhattan metric) - preserved by a pattern solution $\mathcal{P}$, at a specific radius r. As $L_r(\mathcal{P})$ tends towards 1, better is the level of locality preservation (cf. Section 2.1). By scanning a range of radii of interest $r \in [1, 2, ..., N]$ the overall function $\mathbf{L}(\mathcal{P})$ evaluates then how much the locality is conserved as deeply the space is filled. The number of radius N is arbitrarily set to $N = 2^D/4$. N, fixes the number of handled objective functions according the space dimension D ($N = 4$ and $N = 8$ respectively for the $4-D$ and $5-D$ cases).

In the first iteration (Pareto archive empty), the initial population is directly evaluated through the overall objective function $\mathbf{L}(\mathcal{P})$ (cf. Figure 5, Stage B). Precisely, the test of linear dependence (Stage A, Figure 5) is by-passed in order to keep as much diversity of solutions from the start.

Then, to pointing out which are the dominant and dominated solutions (belonging to the population), the equation of Pareto Optimality 4 is computed (Stage C). The Pareto archive is then updated with the dominant solutions (Stage D).
Next, the stop criterion not reached, all the dominant[5] solutions (yet in archive, Stage E) and a sub-set of dominated solutions Stage E (uniformly randomly selected) are selected to crossover operation. The crossover operator used is the Distance Preserving Crossover (DPX) proposed in [23, 24]. The solutions resulting from the crossover between dominant and dominated solutions then define a new population (Stage G). A mutation is then applied, as usual, it involves switching a binary value on a randomly selected digit (low probability) from the whole new population. Practically, we try to maintain the size of the population (I) - as much as possible - throughout the evolutionary process.

In the following iterations, the detection of linearly dependence is of course carried out (Stage A). Precisely, the test of isometry link between the solutions belonging to the new population and the dominant solutions (previously stored in Pareto archive) is estimated via the $\mathbf{Q}(\mathcal{P})$ signature (cf. Equation 9). Practically, the process is stopped when no new solutions appear in the Pareto archive (stabilized Pareto front).

These investigations firstly lead to the identification of new interesting patterns and confirmed patterns previously found in [5]. Secondly, the results obtained in $5-D$ (cf. Table 5) showed that the RBG pattern is not Pareto optimal. So, within the experimental conditions, the classical RBG pattern does not appear in Pareto archive (after several runs), on the contrary alternative patterns emerged.

[5]It is an evolutionary elitist strategy because all the dominant solutions are involved in crossover operation.

	Locality level : objectives score $\mathbf{L}(\mathcal{P}^*)$				
Pareto optimal solution $\mathcal{P}^*$	L_1	L_2	L_3	L_4	Points list
$\mathcal{P}_1^* = \text{RBG}$	1.0	2.0	2.75	2.75	0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0
$\mathcal{P}_{1-1}^*$	1.0	2.0	2.75	2.75	0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 0 1 1 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0
$\mathcal{P}_{1-2}^*$	1.0	2.0	2.75	2.75	0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0
$\mathcal{P}_{1-3}^*$	1.0	2.0	2.75	2.75	0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 1 0 0
$\mathcal{P}_2^*$	1.0	2.0	2.625	2.875	0 0 0 0 0 0 1 1 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 1 1 0 0 1 0 1 1 0 0 1 1 1 1 0 0 0 0 0 1 1

Table 2
Application of the proposed approach synthesized in Figure 5 to the identification of high-locality patterns in $4 - D$ case. $N = 4$ objectives function need to be minimized $L_r(\mathcal{P})$ with $r \in [1, 2, ..., N]$.

4.2. High-locality $4 - D$ pattern identification: results

Table 2 shows an example of Pareto optimal solutions $\mathcal{P}^*$ in the $4 - D$ case. The Pareto optimal solutions $\mathcal{P}^*$ (the first column) are extracted from the Pareto archive after the algorithm converges.
Results from several runs are reported along with the score of objectives (to be minimized, $\mathbf{L}(\mathcal{P}^*) = [L_1(\mathcal{P}^*), L_2(\mathcal{P}^*), L_3(\mathcal{P}^*), L_4$
Here, $\mathbf{L}(\mathcal{P}^*)$ illustrates how well the Pareto optimal pattern $\mathcal{P}^*$ preserves the locality as deeply the $4 - D$ space is explored.
These investigations lead to the identification :

- of patterns previously found in [5], such as $\mathcal{P}_1^* = RBG$ and $\mathcal{P}_{1-1}^*$ (cf. the two first rows in Table 2).
- but also new interesting patterns, such as $\mathcal{P}_{1-2}^*$, $\mathcal{P}_{1-3}^*$, and $\mathcal{P}_2^*$.

We can note that $\mathcal{P}_{1-2}^*$ and $\mathcal{P}_{1-3}^*$ are competitive one refer to the RBG pattern classically used to build multidimensional extension of the Hilbert curve. Precisely, $\mathcal{P}_{1-2}^*$ and $\mathcal{P}_{1-3}^*$ achieve an identical locality score than RBG with $\mathbf{L}(\mathcal{P}_{1-2}^*) = \mathbf{L}(\mathcal{P}_{1-3}^*) = \mathbf{L}(\mathcal{P}_1^* = RBG) = [1.0, 2.0, 2.75, 2.75]$ while exploring the space differently (cf. Points List).

	Pareto Archive browsing	
Radius query r	Best objective score at r radius : $L_r(\mathcal{P}^*)$	Pareto optimal solution $\mathcal{P}^*$: Points list
3	2.625	0 1 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 0 0 1 1 1 1 0 0 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1
4	2.875	0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
5, 6, 7	2.937	0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8	3.50	0 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 0 0 1 1 1 1 1 0 0 0 1 1 1 0 0 0 0 1 1 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 0 0 1 1 0 0 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0

Table 3
Application of the proposed approach (synthesized in Figure 5) to the identification of high-locality $5 - D$ pattern for one query space radius of interest.

4.3. High-locality $5 - D$ pattern identification: results

Our proposal can be seen as a evolutionary population based approach (MOEA, cf. [20]) given access to a set of solutions (here stored in the Pareto archive) in contrast of more standard optimization schemas. The Pareto optimal solutions create a trade-off between all handling objectives, considering all objectives as equally. On the one hand, this broadly helps to reveal important knowledge about a problem, here, how the locality is met as deeper the space is explored by a D-Dimensional pattern. On the other hand, it opens up the possibility for practitioners to select a solution (here, a pattern) by browsing the Pareto archive. For example, the user can now select a high-locality pattern for a specific space radius of interest (cf. Table 3), or they may be interested in finding a pattern with a high locality profile through successive radius values (cf. Table 6). The results obtained in Dimension 5 are presented to highlight this ability.

4.3.1. The best pattern for one query radius

An example of Pareto optimal solutions $\mathcal{P}^*$ corresponding to a given space radius of interest r is provided in Table 3. The user can identify high-locality patterns for one specific query space radius r by browsing off-line the Pareto archive (after the algorithm converges).
$\mathcal{P}^*$ (third column) is the pattern witch reach the best locality score (second column) for a radius r (first column). In the promoted Pareto Multi-objective optimization framework (cf. Section 3.2), $\mathcal{P}^*$ is a Pareto Optimal solution that minimize the objective $L_r(\mathcal{P})$.
For example, in the first row, the pattern witch reach the best locality level at query radius $r = 3$ is $\mathcal{P}^*$ (Points List provided in the third column) with the score $L_{r=3}(\mathcal{P}^*) = 2.625$.

	Objectives score reached by the RBG Pattern $\mathbf{L}(\mathcal{P})$						
Pattern $\mathcal{P}$	L_3	L_4	L_5	L_6	L_7	L_8	Points list
RBG	2.875	2.875	3	3.75	3.75	3.75	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0

Table 4
Objectives score reached by the RBG Pattern in $5 - D$ case. $N = 8$ objective functions need to be minimized $L_r(\mathcal{P})$ for $r \in [1, 2, ..., N]$. For ease of reading, the scores for L_1 and L_2 are not shown in Table, as all the solutions found achieve the same locality score for $r = 1$ and $r = 2$ such that $L_1 = 1$ and $L_2 = 2$.

	Locality level : Objectives score reached $\mathbf{L}(\mathcal{P})$					
Pattern $\mathcal{P}$	L_3	L_4	L_5	L_6	L_7	L_8
RBG	2.875	2.875	3	3.75	3.75	3.75
A Pareto optimal solution : $\mathcal{P}^*$	2.875	2.875	3	**3.5625**	**3.5625**	3.75
Pareto optimal solution $\mathcal{P}^*$						
0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 0 1 1 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0						

Table 5
An example of $5 - D$ Pareto optimal solution $\mathcal{P}^*$ that achieve better locality preservation than classical RBG pattern. In $5 - D$ case the RBG pattern is not Pareto optimal.

Compared to the referent RBG pattern (cf. Table 4) there are Pareto optimal solutions $\mathcal{P}^*$ that exhibit better performance. For example, with a query radius $r = 3$, the locality score reached is $L_{r=3}(\mathcal{P}^*) = 2.625$ (cf. Table 3) compared to $L_3 = 2.875$ for the RBG pattern. As reminder, as $L_r(\mathcal{P})$ tends towards 1, better is locality preservation (cf. Section 2.1).

Furthermore, the results shown that the RBG pattern is not Pareto optimal in $5-D$ (cf. Table 5). Precisely, the proposed approach allows to identify a Pareto optimal pattern $\mathcal{P}^*$ that outperforms the locality preservation. According to Pareto optimality (cf. Equation 4), $\mathcal{P}^*$ totally dominates the pattern RBG. There is at least one objective (here, two objectives L_6 and L_7, in bold in Table 5) that achieves a better score compared with RBG (respectively $L_6(\mathcal{P}^*) = 3.5625$ and $L_7(\mathcal{P}^*) = 3.5625$ vs $L_6(\mathcal{RBG}) = 3.75$ and $L_7(\mathcal{RBG}) = 3.75$) without the other objectives being downgraded (here, the scores of the other objectives remain equal for both patterns, cf. Table 5).

<table>
<tr><td></td><td colspan="6">Pareto Archive browsing</td></tr>
<tr><td rowspan="2">Range radius query [r]</td><td colspan="6">Objectives score [$L_r(\mathcal{P}^*)$]</td></tr>
<tr><td>L_3</td><td>L_4</td><td>L_5</td><td>L_6</td><td>L_7</td><td>L_8</td></tr>
<tr><td>[3; 5]</td><td>2.875</td><td>2.875</td><td>2.937</td><td>3.312</td><td>3.375</td><td>3.937</td></tr>
<tr><td></td><td colspan="6">Pareto optimal solution $\mathcal{P}^*$</td></tr>
<tr><td></td><td colspan="6">0 1 1 0 0 1 1 1 0 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 0 1 1 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0
0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 1 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0 0 1 1 0 0 0 1 1 1 1</td></tr>
<tr><td rowspan="2">Range radius query [r]</td><td colspan="6">Objectives score [$L_r(\mathcal{P}^*)$]</td></tr>
<tr><td>L_3</td><td>L_4</td><td>L_5</td><td>L_6</td><td>L_7</td><td>L_8</td></tr>
<tr><td>[5; 7]</td><td>2.937</td><td>2.937</td><td>2.937</td><td>2.937</td><td>3</td><td>4</td></tr>
<tr><td></td><td colspan="6">Pareto optimal solution $\mathcal{P}^*$</td></tr>
<tr><td></td><td colspan="6">0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0
0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0
0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0</td></tr>
</table>

Table 6
Application of the proposed approach to the identification of high-locality $5 - D$ patterns over successive radius values of interest. Example of two Pareto optimal solutions $\mathcal{P}^*$ each corresponding to a given range of radii [r].

4.3.2. High-locality pattern over successive radius values

The archive-based evolutionary process proposed can opens up several opportunities for the user (use cases). After the ability to select a good pattern for a specific query space radius (cf. Section 4.3.1), the user can off-line browse the Pareto archive to pick up a pattern with a high-locality profile over a range of radii (if such a pattern exists).
The results in Table 6 show an example of Pareto optimal solutions $\mathcal{P}^*$ corresponding to a given range of radii [r]. Precisely, $\mathcal{P}^*$ is a Pareto Optimal solution that minimizes the successive objective functions $L_r(\mathcal{P})$ with $r \in [r]$. For example, in the first row, the pattern that achieves the best locality score over the range of radii [3; 5] is $\mathcal{P}^*$ (Points List provided at the bottom) with the scores $L_{r=3}(\mathcal{P}^*) = 2.875$, $L_{r=4}(\mathcal{P}^*) = 2.875$ and $L_{r=5}(\mathcal{P}^*) = 2.937$.

5. Discussion

Space-filling curves, particularly the Hilbert curve, are widely used in applications that require locality preservation. For example, in the field of bio-informatics [25], the Hilbert curve is used for genome analysis. Conventional techniques tend to form a very long $1 - D$ vector, in which each components characterizes the chromosome under study. Theses 1-D components are arranged onto a 2-D Hilbert curve, where similar components values are mapped closest along the Hilbert curve taking thus advantage of the preservation of locality (i.e. spatial proximity). This gives rise to a useful tool (called HCV) to explore genomic data. In the domain of robotic, trajectory planning in presence of obstacles is a challenging task. In [26], the self-similarity and locality-preserving properties of Hilbert curve are used to plan evasive strategies. By changing the way point of robot, the resulting strategies seem to be able to avoid obstacles with the constraint to visit the entire region. In image representation and compression, an original image encoder-based on the Hilbert curve that can achieve a great compression rate is presented in [27]. The authors propose to store in a 24-bit standard color image format both the geometrical information (captured via a phase map) and the texture of a given 3-D object. The mapping of the 2-D phase map into the Hilbert curve seems to achieve a good compression rate while maintaining a fair quality of image recovery (decoding step). Once again, the capability to preserve locality leads to better results (comparative results of encoded phase map and recovered geometry over several space-filling curves are shown).

More recently, applications dealing with massive point cloud management and attribute compression can be found in [28, 29]. Other applications could be targeted. For example, in machine learning and computer vision, Convolutional Neural Networks (CNNs) are useful for standard tasks (classification, segmentation, object detection etc.). A recent paper (2023, [30]) deals with the use of Hilbert curve as preprocessing step before implement a CNN. For design and training purpose, high-dimensional input must be converted into one dimensional string. To preserve local correlation within high-dimensional data, the mapping into Hilbert curve is assessed and compared to others transformations. The authors conclude that *Our findings quantify the importance of locality preservation when transforming data before training a one-dimensional CNN and show that the Hilbert space-filling curve is a preferential transformation to achieve this goal.* Back to our proposal, the existence of a set of alternative curves and use cases could translate into competitive conservation of the local correlations observed in the CNN entries. This could contribute to an efficient training phase (and maybe to improved finally CNN performance (testing phase)).

In [5], the role of the pattern (the first-order curve) on locality preservation is justified and highlighted (cf. reminder in Section 2.2). Selecting a better locality-preserving pattern (or here, identify new ones) leads to design, through orders, an overall better locality preserving curve.

To our knowledge, few works in the literature are paradoxically inscribed in that direction. In [31], new ways to fill a 2-D space, which, together with Hilbert and Moore curves, define a set called “Complete set of the Hilbert curve” are proposed. Four alternatives (in reference to Hilbert) sets of geometric transformations are suggested to arrange the 2-D RBG pattern in order to build an upper order curve. According to the set of transformation selected (L configuration) new patterns are resulting. This is a relevant study, but it still focuses on 2-D space and, as usual, it starts from RBG pattern (called Hilbert pattern in the article); in the same direction we can find [32]. In [33], a definition of a novel compact Hilbert index dedicated to spaces with unequal size dimensions is provided. That is helpful in applications where the cardinality of dimensions can be very different, giving an alternative to padding all dimensions to the cardinality of the largest one, which often conducts to manipulating "inflated" values. The line segments usually used on 2-D Hilbert curve to symbolize points ordering are replaced by arcs. We can note that this geometric change does not modify the order in which points (cells in the article) are visited. The cell ordering is still guided by the standard RBG code, meaning that it follows the standard RBG pattern.

Therefore, our work positioning is original, and the development of an approach to identify new multidimensional patterns that bring an interesting level of locality in a multi-objective optimization way constitutes a real contribution. In Section 4, the results have shown that it is possible to identify patterns according to their capabilities to satisfy objectives, specifically, their abilities to preserve locality as deeply as the space is filled. This leads to a better accuracy in pattern selection, which could finally benefit various applications.

6. Conclusion

As highlighted in previous well-received works [4, 5], Patterns, which are the first-order space-filling curves, play a central role on locality preserving property (cf. Section 2.2). In this article, we propose an original approach for identifying high locality-preserving multidimensional patterns. Rather than considering only one query radius, we consider several space radii of interest. That conducts to take into account how the locality is conserved as deeply the space is filled, giving birth to a more accurate identification process.
Technically, the problem is embedded in the Pareto multi-objective optimization framework (cf. Section 3.2). Each locality score reached by a pattern at a given radius value is processed as an objective function to be minimized. The problem is solved help with an elitist evolutionary algorithm using an archive strategy (Pareto archive).
Details on the guiding ideas and the justifications for the well-founded framework and the soundness of the proposed approach were provided (cf. Section 3). The Pareto optimality allows to establish a hierarchy between pattern-solutions (dominated / non dominated solutions) according to how the objectives are meet (i.e. how the locality is preserved), which is useful when conflicting objectives can be observed. Furthermore, adaptation on the evolutionary process have been made to design an approach better suited to our problem. For example, an isometry free test was developed in order to reject linearly dependent solutions (cf. Section 3.3), as only linearly independent solutions can improve locality. Experiments conducted on 4-D an 5-D (cf. Section 4) confirm, within the experimental conditions, that the proposed approach is :

- reliable : among the solutions, it includes patterns that were already identified in previous paper [5], as well as new patterns to our knowledge not previously mentioned in the literature (cf. Table 2). We can also note that the classical RBG pattern is not Pareto optimal in the $5-D$ case (cf. Table 5), on the contrary alternative patterns emerged.
- flexible : after the algorithm converges, the user can off-line browse the Pareto archive to extract solutions according to his own criteria. For example, the user can select the best pattern for a specific space radius value of interest (cf. Table 3), or it can find a pattern with high locality profile over a range of query radius values (cf. Table 6).
- efficient : an isometry-free test was developed in order to reject linearly dependent solutions, which seems to ensure good convergence of the approach toward Pareto optimal solutions (cf. Section 3.3).

Finally, being able to identify patterns that preserve locality at a competitive level compare to the referent Hilbert curve (RBG pattern-based) constitutes a real contribution. This can contribute to reduce topological breaks between spaces (cf. Figure 2) and could greatly improve the effectiveness of applications (cf. Section 5).
For future work, a new based-decomposition optimization schema could be investigated. A multi-objective optimization problem can be decomposed into aggregated subproblems, which are optimized simultaneously in a collaborative way (using information from their neighboring subproblems) [34, 35]. Number of Multi-Objective Evolutionary Algorithms (MOEA) have been successfully proposed to decomposition (MOEA/D) and frameworks are available.

7. Declarations

- **Competing Interests :**
  the authors declare that they have no competing interests.
- **Funding :**
  the authors did not receive support from any organization for the submitted work.
- **Data availability :**
  the authors declare that the data supporting this proposition are available within the paper.

References

[1] G. Peano, Sur une courbe, qui remplit toute une aire plane, Mathematische Annalen 36 (1) (1890) 157–160.
[2] H. Lebesgue, Leçons sur l'intégration, Paris, Gauthier-Villars, 1904.
[3] D. Hilbert, Ueber die stetige abbildung einer line auf ein flächenstück, Mathematische Annalen 38 (3) (1891) 459–460.

[4] G. Nguyen, Space-filling curves and their application in image processing, Ph.D. thesis, University of La Rochelle (France), Laboratoire Informatique Image et Interactions (L3I). HAL Id: tel-01174960. (2013).
[5] P. Franco, G. Nguyen, R. Mullot, J.-M. Ogier, Alternative patterns of the multidimensional hilbert curve, Multimedia Tools and Applications (MTAP) 77 (7) (2018) 8419–8440. doi:10.1007/s11042-017-4744-4.
[6] A. Butz, Alternative algorithm for hilbert's space-filling curve, IEEE Transactions on Computers 20 (4) (1971) 424–426.
[7] J. Lawder, P. King, Using space-filling curves for multi-dimensional indexing, in: Advances in Databases, Vol. 1832, 2000, pp. 20–35.
[8] G. Jin, J. Mellor-Crummey, Sfcgen: A framework for efficient generation of multi-dimensional space-filling curves by recursion, ACM Transactions on Mathematical Software (TOMS) 31 (2005) 120–148.
[9] C. Faloutsos, S. Roseman, Fractals for secondary key retrieval, in: Proceedings of the eighth ACM SIGACT-SIGMOD-SIGART symposium on Principles of Database Systems (PODS), 1989, pp. 247–252.
[10] H. Sagan, Space-Filling Curves, Springer Science and Business Media, 2012. doi:10.1007/978-1-4612-0871-6.
[11] M. Bader, Space-Filling Curves: An Introduction With Applications in Scientific Computing, Springer Berlin, Heidelberg, 2013. doi:10.1007/978-3-642-31046-1.
[12] P. Xu, C. Nguyen, S. Tirthapura, Onion curve: A space filling curve with near-optimal clustering, in: 2018 IEEE 34th International Conference on Data Engineering (ICDE), 2018, pp. 1236–1239. doi:10.1109/ICDE.2018.00119.
[13] G. Mitchison, R. Durbin, Optimal numberings of an n*n array, SIAM Journal on Algebraic Discrete Methods 7 (4) (1986) 571–582.
[14] C. Gotsman, M. Lindenbaum, On the metric properties of discrete space-filling curves, IEEE Transactions on Image Processing 5 (5) (1996) 794–797. doi:10.1109/83.499920.
[15] M. F. Mokbel, W. G. Aref, Analysis of multi-dimensional space-filling curves, GeoInformatiqua 7 (3) (2003) 179–209.
[16] B. Moon, H. Jagadish, C. Faloutsos, J. Saltz, Analysis of the clustering properties of the hilbert space-filling curve, IEEE Transactions on Knowledge and Data Engineering 13 (1) (2001) 124–141.
[17] C. Fonseca, P. J. Fleming, et al., Genetic algorithms for multiobjective optimization : formulation, discussion and generalization, in: Fifth International Conference on Genetic Algorithms (ICGA), San Francisco, USA, Vol. 93, 1993, pp. 416–423.
[18] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: Nsga-ii, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197.
[19] M. Laumanns, E. Zitzler, L. Thiele, A unified model for multi-objective evolutionary algorithms with elitism, in: Proceedings of the IEEE Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512), Vol. 1, 2000, pp. 46–53. doi:10.1109/CEC.2000.870274.
[20] K. Deb, Multi-objective Evolutionary Optimisation for Product Design and Manufacturing, Springer London, 2011, Ch. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction, pp. 3–34. doi:10.1007/978-0-85729-652-8_1.
[21] A. Zhou, B.-Y. Qu, H. Li, all, Multiobjective evolutionary algorithms: A survey of the state of the art, Swarm and Evolutionary Computation 1 (1) (2011) 32–49. doi:10.1016/j.swevo.2011.03.001.
[22] F. Combes, Algèbre et géométrie : agrégation, CAPES, licence-maîtrise, Bréal, 1998.
[23] B. Freisleben, P. Merz, New genetic local search for the travelling sale man problem, in: Springer-Verlag (Ed.), Lecture Notes in Computer Science, Vol. 1141, 1996, pp. 890–899.
[24] R. Mais Haj, C. Bloch, W. Ramdane Cherif Khettaf, P. Chatonnay, Différents operateurs évolutionnaires de permutation: selections, croisements et mutations, Tech. rep., Laboratoire Informatique - Université Franche-Comté (France). RR 210-07. (2010).
[25] S. Anders, Visualization of genomic data with the hilbert curve, Bioinformatics 25 (10) (2009) 1231–1235. doi:10.1093/bioinformatics/btp152.
[26] S. H. Nair, A. Sinha, L. Vachhani, Hilbert's space-filling curve for regions with holes, in: 56th IEEE Conference on Decision and Control (CDC), 2017, pp. 313–319. doi:10.48550/ARXIV.1709.02938.
[27] X. Chen, S. Zhang, Three dimensional range geometry and texture data compression with space-filling curves, International Journal Optics Express 25 (21) (2017) 26103–26117. doi:10.1364/OE.25.026103.
[28] X. Guan, P. Van Oosterom, B. Cheng, A parallel n-dimensional space-filling curve library and its application in massive point cloud management, ISPRS International Journal of Geo-Information 7 (8) (2018). doi:10.3390/ijgi7080327.
[29] J. Chen, L. Yu, W. Wang, Hilbert space filling curve based scan-order for point cloud attribute compression, IEEE Transactions on Image Processing 31 (2022) 4609–4621. doi:10.1109/TIP.2022.3186532.
[30] B. Verbruggen, V. Ginis, Improving image classification of one-dimensional convolutional neural networks using hilbert space-filling curves, Applied Intelligence 53 (2023) 26655–26671. doi:10.1007/s10489-023-04945-2.
[31] X. Liu, Four alternative patterns of the hilbert curve, Applied mathematics and computation 147 (3) (2004) 741–752.
[32] C. Pérez-Demydenko, I. Brito-Reyes, B. Aragón Fernández, E. Estevez-Rams, The complete set of homogeneous hilbert curves in two dimensions, Applied Mathematics and Computation 234 (2014) 531–542. doi:10.1016/j.amc.2013.12.193.
[33] C. H. Hamilton, A. Rau-Chaplin, Compact hilbert indices: Space-filling curves for domains with unequal side lengths, Information Processing Letters 105 (5) (2008) 155–163.
[34] Q. Zhang, H. Li, Moea/d: A multiobjective evolutionary algorithm based on decomposition, IEEE Transactions on Evolutionary Computation 11 (6) (2007) 712–731. doi:10.1109/TEVC.2007.892759.
[35] A. Liefooghe, M. Basseur, L. Jourdan, E. Talbi, Paradiseo-moeo: A framework for evolutionary multi-objective optimization, in: H. Springer, Berlin (Ed.), Lecture Notes in Computer Science, Vol. 4403, 2007, p. 386–400. doi:10.1007/978-3-540-70928-2_31.

8. Appendix

Didactic example of the calculation of the proposed signature Q (cf. Section 3.3) on a $3 - D$ pattern $\mathcal{P}$. The case $\mathcal{P} = RBG$ is considered.

$$RBG = \begin{array}{cccccccc} 0&0&0&0&1&1&1&1 \\ 0&0&1&1&1&1&0&0 \\ 0&1&1&0&0&1&1&0 \end{array} \tag{10}$$

The q_n are equal to:

$$\begin{aligned} q_1 &= [4, 4] \\ q_2 &= [2, 4, 2] \\ q_3 &= [1, 2, 2, 2, 1] \end{aligned} \tag{11}$$

Then, after sorting the q_n according to lexical ordering :

$$Q(RBG) = \{[1, 2, 2, 2, 1], [2, 4, 2], [4, 4]\} \tag{12}$$