



A genetic algorithm to solve a space-filling curve problem

Valentin Owczarek, Patrick Franco, Rémy Mullot

► To cite this version:

Valentin Owczarek, Patrick Franco, Rémy Mullot. A genetic algorithm to solve a space-filling curve problem. SLS2019: International Workshop on Stochastic Local Search Algorithms, Sep 2019, Lille, France. pp.5-6. hal-02297854

HAL Id: hal-02297854

<https://hal.science/hal-02297854>

Submitted on 26 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A genetic algorithm to solve a space-filling curve problem

V. Owczarek, P. Franco and R. Mullot

La Rochelle université, Laboratoire L3i, 17031 La Rochelle
valentin.owczarek1@univ-lr.fr

1 Space-filling curve: A pattern driven algorithm

A space-filling curve (SFC) is an ordering function which creates a bijection between integers I and points X lying in a D -dimensional regular grid. Let consider a point $\mathbf{X}_a \in \mathbf{X} = [0, 2^n - 1]^D$, the associate integer corresponds to $\mathbf{I}_a \in \mathbf{I} = [0, 2^{nD} - 1]$ where n is the order of the curve. This ordering has the property to conserve the distance, i.e. close points in the space have close indices. This property known as the locality preserving. At this time, the Hilbert curve defined by D. Hilbert [1] in 1891 achieve the best locality preserving level [2]. In various domain, many applications are drawn on the locality preservation properties of space-filling curves. For example, one could concern global optimization, data visualization, image compression and encryption.

Recently, G. Nguyen [3] proposed a new algorithm to compute SFCs where the curve at order $n = 1$ is used to recursively define the bijection between the D -dimensional grid and the indices. This new formulation is able to define any Hilbert like SFC, by using any pattern that respects the adjacency rule: if $\mathbf{I}_a$ and $\mathbf{I}_b$ are separated by one unit then their associate points $\mathbf{X}_a$ and $\mathbf{X}_b$ are separated by a unit distance. The well know Hilbert curve is defined by using the Reflected binary gray code (RBG).

Using this new definition of SFC, relevant studies have been performed by P. Franco [4] to investigate the influence of the pattern for the locality preserving level of the n order curve. The results proved that it is possible to create comparable or better SFC than the regular Hilbert curve which is considered, in the literature, as the best locality preserving curve.

In fact, finding patterns which have a high level of locality preservation level is equivalent to find all Hamiltonian paths in a hypercube graph, which is NP-complete according to R. Karp [5]. A hypercube graph in dimension D is defined by 2^D nodes along with $D \cdot 2^{D-1}$ edges, and each node has a degree of D . Obviously, this problem can not be solved by regular graph traversal techniques when $D > 5$. Moreover, our goal is not to find one unique pattern but a set of non-dominated solutions according to the Faloutsos criteria [2] as

$$F_r(\mathbf{X}) = \frac{1}{2^D} \sum_{k, l \in [2^D], k < l, d(k, l) \leq r} \max\{d(\mathbf{X}_k, \mathbf{X}_l)\}, \quad (1)$$

where $\mathbf{X}$ is the set of points in the D -dimensional regular grid ordered by a SFC and d is the Manhattan distance.

In the next section, different stochastic local search algorithms are briefly explained and results of our experiments are discussed.

2 Stochastic local search algorithms for hamiltonian chain

In our work, genetic algorithms (GAs)[6, 7] are used to find hamiltonian chains. GAs use successively mutation/crossover and selection on a population of solutions to find the optima. In every algorithm, the mutation operator is 2-opt defined by Georges A. Croes [8] and the crossover is single point. The algorithms described in Table 1, are different in terms of selection and fitness function. The distance of the chain is denoted by $|C|$ and the Faloutsos criteria at radius α by F_α .

In the elitist selection method “Top 100” the first hundred hamiltonian chain corresponding at each generation to the hundred lowest fitness score are conserved to be the next generation. Whereas “Non-dominated set” stands for the conservation over generations of the set of all non-dominated solutions.

Table 1. The different GAs used to find hamiltonian paths, according to $|C|$ the length of a path and F_α , the Faloutsos criteria.

Name	Type	Selection method	Fitness Function
<i>Classic</i>	Single-objective	Top 100	$ C $
<i>Sum</i>	Single-objective	Top 100	$ C ^2 + \sum_{i=3}^6 F_i$
<i>Multi</i>	Multi-objective	Non-dominated set	$\{F_3, C , \dots, F_6, C \}$

According to Table 1, Each of them differs in term of selection and fitness function. *Classic* uses only $|C|$ whereas *Sum* and *Multi* use the Faloutsos criteria F_α (Equation 1). *Multi* is a multi-objective GA version with a non-dominated set selection method.

For each algorithm the non-dominated set of the results of 1000 runs, for $D = 5$, is saved. Thus three created fronts are compared, as proposed in [9], using the hypervolume (to be maximized) and the additive epsilon (to be minimized) indicators.

For the hypervolume indicator, *Multi* is better with a score of 0.1254 compared to 0.1197 and 0.0772 for respectively *Sum* and *Classic*. The additive epsilon indicator point out the same performances: *Multi* reached the first place with a score of 0.0625 whereas *Sum* and *Classic* respectively obtained 0.0937 and 0.1875 scores.

The results are promising: *Multi* provided close sets to the exact known front, according to the hypervolume and the additive epsilon indicators. Similarly, *Sum* reached comparable results which would be consistent for further improvements. Indeed extended investigations will be ensure to confirm the results on grids of higher dimension. Moreover, the non-dominated set found during this study, on dimension $D = 5$, shows the non-optimality of the Hilbert curve pattern: the associated Faloutsos scores are dominated by several other patterns.

References

1. David Hilbert. Ueber die stetige Abbildung einer Line auf ein Flächenstück. *Mathematische Annalen*, 38(3):459–460, 1891.
2. C. Faloutsos and S. Roseman. Fractals for Secondary Key Retrieval. In *Proceedings of the Eighth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS ’89, pages 247–252. ACM, 1989.
3. Giap Nguyen. *Courbes remplissant l’espace et leur application en traitement d’images*. PhD thesis, Université de La Rochelle, 2013.
4. Patrick Franco, Giap Nguyen, Remy Mullot, and Jean-Marc Ogier. Alternative patterns of the multi-dimensional Hilbert curve. *Multimedia Tools and Applications*, 77(7):8419–8440, 2018.
5. Richard M. Karp. *Reducibility among Combinatorial Problems*, pages 85–103. Springer US, Boston, MA, 1972.
6. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.
7. Bilel Derbel, Arnaud Liefoghe, Qingfu Zhang, Sébastien Verel, Hernan Aguirre, and Kiyoshi Tanaka. A set-oriented MOEA/D. In *GECCO 2018 - Genetic and Evolutionary Computation Conference*, pages 617–624, Kyoto, Japan, July 2018. ACM Press.
8. G. A. Croes. A method for solving traveling-salesman problems. *Operations Research*, 6(6):791–812, 1958.
9. J. Knowles, L. Thiele, and E. Zitzler. A Tutorial on the Performance Assessment of Stochastic Multi-objective Optimizers. TIK Report 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, 2006.