

# Quotient Geometric Crossovers and Redundant Encodings

Yourim Yoon<sup>†</sup>, Yong-Hyuk Kim<sup>‡</sup>, Alberto Moraglio<sup>\*</sup>, and Byung-Ro Moon<sup>†</sup>

<sup>†</sup> School of Computer Science & Engineering, Seoul National University

599 Gwanak-ro, Gwanak-gu, Seoul, 151-744, Korea

Email: {yryoon, moon}@soar.snu.ac.kr

<sup>‡</sup> Department of Computer Science & Engineering, Kwangwoon University

20 Kwangwoon-ro, Nowon-gu, Seoul, 139-701, Korea

Email: yhdfly@kw.ac.kr

<sup>\*</sup> School of Computing, University of Kent, UK

Email: A.Moraglio@kent.ac.uk

May 23, 2011

## Abstract

We extend a geometric framework for the interpretation of search operators to encompass the genotype-phenotype mapping derived from an equivalence relation defined by an isometry group. We show that this mapping can be naturally interpreted using the concept of quotient space, in which the original space corresponds to the genotype space and the quotient space corresponds to the phenotype space. Using this characterization, it is possible to define induced geometric crossovers on the phenotype space (called *quotient geometric crossovers*). These crossovers have very appealing properties for non-synonymously redundant encodings, such as reducing the size of the search space actually searched, removing the low locality from the encodings, and allowing a more informed search by utilizing distances better tailored to the specific solution interpretation. Interestingly, quotient geometric crossovers act on genotypes but have an effect equivalent to geometric crossovers acting directly on the phenotype space. This property allows us to actually implement them even when phenotypes cannot be represented directly. We give four example applications of quotient geometric crossovers for non-synonymously redundant encodings and demonstrate their superiority experimentally.

**Keywords:** Genotype-phenotype mapping, redundant representation, quotient metric space, geometric crossover, quotient geometric crossover.

## 1 Introduction

For many interesting problems in evolutionary computation, it is difficult to design effective search operators working directly on phenotypes, e.g., when phenotypes are rather complex structures. In such cases, encodings are normally used to assign phenotypes to standard genotypes, such as binary strings, upon which standard search operators can be used. For *redundant representation*, the same phenotype is represented by more than one genotype. There are quite a few search, optimization, and learning problems addressed with evolutionary algorithms for which it is difficult to represent one

phenotype by a single genotype using traditional representation, such as binary string or permutation, e.g., when phenotypes are graphs.

There have been some theoretical and empirical studies that analyze redundant encodings in terms of *synonymity* and *locality* [32, 33, 34]. These are known to be good (qualitative) indicators of problem hardness caused by the redundancy of the encoding for genetic algorithms. A representation is regarded as (non-)synonymously redundant if the genotypes that are assigned to the same phenotype are (not) similar to each other. Whereas synonymously redundant representations have a neutral effect on performances, non-synonymously redundant representations may lead to severe loss of performance in genetic algorithms, in particular, with respect to traditional crossovers<sup>1</sup> [7]. To alleviate the problems caused by non-synonymously redundant representations, some methods such as adaptive crossover have been proposed [9, 23, 30, 42]. Among them, a technique called *normalization*<sup>2</sup> is representative. Normalization transforms the genotype of a parent into another genotype that represents the same phenotype to be as similar as possible to the genotype of the other parent before performing traditional crossover. There have been many successful studies using normalization. Recently Choi and Moon [8] showed that normalization, which was designed heuristically according to the target problem, is useful for non-synonymously redundant representations, through experimental analyses on various problems. However, previous studies have explained its usefulness very informally and tackled the method in case-by-case design. In this paper, we formally design normalization as a tool for computing a more natural metric on search space, and show why normalization is important for non-synonymously redundant representations by re-interpreting it with formal geometric framework.

Geometric crossovers [28] are search operators defined geometrically using the distance associated with the search space. In this paper, we show how to derive formally geometric crossovers which are completely unaffected by the redundancy of the encoding. This is achieved by showing that a large class of genotype-phenotype mappings, each of which can be derived from an equivalence relation defined by an isometry group, can be naturally interpreted geometrically using the concept of *quotient metric space*. We then formally present the general relation between the notion of quotient metric space and genotype-phenotype mapping, and we derive the relation between genotype and phenotype spaces and their associated geometric crossovers. Finally, we show how to exploit the relation between geometric crossovers to derive an operational and representation-independent procedure that allows us to implement a geometric crossover on the phenotype space indirectly by appropriate manipulation of genotypes.

As a preliminary work [21, 27], we presented a well-designed metric suitable for the multiway graph partitioning problem, which presents a redundant encoding, and designed a geometric crossover defined on the corresponding metric space successfully. However, the idea of the previous work is applicable only to a specific problem (multiway graph partitioning) and does not apply to other problems with redundant encoding. In this paper, we propose a formal framework that generalizes the above mentioned work using the notion of quotient metric and makes it directly applicable to a large class of problems with redundant encodings. We will show that a number of problems can be easily tackled with our approach.

The remainder of the paper is organized as follows. In Section 2, we present some necessary math-

---

<sup>1</sup>Crossovers defined on binary and non-binary strings using crossover-masks.

<sup>2</sup>The term of *normalization* firstly appeared in [18]. However, it is based on the adaptive crossovers proposed in [23, 30].

emational preliminaries and the geometric framework. In Section 3, the notion of *quotient geometric crossover* is introduced in connection with the geometric interpretation of the genotype-phenotype mapping derived from an equivalence relation defined by an isometry group. In Section 4, we present several interesting examples of quotient geometric crossover for non-synonymously redundant encodings, together with experimental evidence showing their superiority.

## 2 Preliminaries

Given a set  $X$  and an equivalence relation  $\sim$  on  $X$ , the equivalence class of an element  $a$  in  $X$  is the subset of all elements in  $X$  that are equivalent to  $a$ . The set of all equivalence classes in  $X$  given an equivalence relation  $\sim$  is usually denoted by  $X/\sim$  and called the *quotient set* of  $X$  by  $\sim$ . Topologically, a point in the quotient space  $X/\sim$  identifies a set of all equivalent points in  $X$ .

A *group* is an algebraic structure consisting of a set together with a binary operator. To qualify as a group, the set and operator must satisfy group axioms: associativity, identity, and invertibility. In evolutionary computation, there have been some studies using group structure and *group action* to analyze and characterize crossover and mutation [35, 36]. In this paper, differently from the previous study, we use groups to construct equivalence relations which give rise to a well-behaved notion of quotient on metric spaces (see Section 3.1 for details).

We assume genotype and phenotype spaces to be metric spaces. Formally, the *metric*  $d$  - or *distance*  $d$  - denotes any real-valued function that conforms to the axioms of identity, symmetry, and triangular inequality. On a metric space  $(X, d)$ , the map  $f: X \rightarrow X$  satisfying the condition of  $d(f(x), f(y)) = d(x, y)$  for all  $x, y \in X$  is called an *isometry* of  $X$  [4]. The set of isometries on  $X$  is denoted by  $Iso(X)$ .  $Iso(X)$  forms a group under the function composition operator. In our study, an isometry subgroup  $\mathcal{G} \subseteq Iso(X)$  will be considered to generate an equivalence relation for the quotient metric space [4] (see Section 3.1).

We recall some known geometric definitions. In a metric space  $(X, d)$ , a *closed ball* is the set of the form  $B_d(x, r) = \{y \in X \mid d(x, y) \leq r\}$ , where  $x \in X$  is called the center of the ball and  $r \in \mathbb{R}^+$  is called the radius of the ball. A *line segment* (or closed interval) is the set of the form  $[x; y]_d = \{z \in X \mid d(x, z) + d(z, y) = d(x, y)\}$ , where  $x, y \in X$  are called extremes of the segment. Metric segment generalizes the familiar notion of segment in the Euclidean space to any metric space.

Geometric crossover [28] is an approach to generalize some standard search operators for the major representations used in evolutionary algorithms [26], such as binary string, real vector, permutation, permutation with repetitions, syntactic tree, sequence, and set. It is defined in geometric terms using the notions of line segment.<sup>3</sup> These notions and the corresponding genetic operators are well-defined once a notion of distance in the search space is defined. Defining search operators as functions of the search space is opposite to the standard way [17] in which the search space is seen as a function of the search operators employed. This viewpoint greatly simplifies the relationship between search operators and fitness landscape, and, as shown in this paper, also between search operators and genotype-phenotype mapping.

The *image*  $Im[XO]$  of a crossover operator  $XO$  is defined as the function mapping parents  $x$  and  $y$  to the set of all possible offspring produced by  $XO(x, y)$ . The following definition is *representation-independent*, in that it refers to the underlying representation only indirectly via a distance defined

---

<sup>3</sup>Geometric mutation, which is a generalization of mutation, is based on the notion of ball.

on a search space such as genotype space or phenotype space.

**Definition 1** (Geometric crossover). *A binary operator  $GX$  is a geometric crossover under the metric  $d$  if all offspring are in the segment between its parents  $x$  and  $y$ , i.e.,  $Im[GX](x, y) \subseteq [x; y]_d$ .*

For example, traditional mask-based crossovers, e.g., uniform and  $n$ -point crossovers, are geometric under the Hamming distance. A number of general properties for geometric crossover have been derived in [28].

### 3 Quotient Geometric Crossover

Let  $X$  and  $Y$  be the genotype space and the phenotype space, respectively. Consider a genotype-phenotype mapping  $\phi: X \rightarrow Y$  that is not injective (i.e., redundant representation). The mapping  $\phi$  induces a natural equivalence relation  $\sim$  on the set of genotypes: *genotypes that represent the same phenotype belong to the same class*. Hence  $Y$  comes to coincide (one-to-one mapping) to the quotient space  $X/\sim$ . Notice that under this interpretation a phenotype is *both* an element of  $Y$  and, at the same time, the set of all genotypes corresponding to that phenotype, which is an element of  $X/\sim$ .

We can formally define a geometric crossover on a space provided that a distance is defined on the space. Interestingly, if the equivalence relation  $\sim$  is derived from an isometry group, the quotient space  $X/\sim$  has a natural distance  $d_Y$  induced by the distance  $d_X$  of  $X$  (see Section 3.1). As we will show in Section 4, this type of equivalence relation appears to occur naturally in a number of non-synonymously redundant encodings. The geometric crossover under  $d_Y$  is a natural crossover on  $Y$  to consider since it reflects the structure of  $X$  by involving the distance  $d_X$  of  $X$ . By applying the formal definition of geometric crossover to the spaces  $(X, d_X)$  and  $(Y, d_Y)$ , we obtain the geometric crossovers  $GX$  and  $GY$ , respectively.  $GX$  searches the space of genotypes and  $GY$  searches that of phenotypes.

The crossover  $GY$  for non-synonymously redundant encodings turns out to be very interesting because, as we will show in Section 4, it presents many potential advantages: (i) it allows to search directly  $Y$  which is much smaller than  $X$ , hence quicker to search<sup>4</sup>; (ii) the induced phenotypic distance  $d_Y$  is better tailored to the underlying problem than the corresponding genotypic distance  $d_X$ , hence  $GY$  works better than  $GX$ ; (iii)  $Y$  has different geometric characteristics from  $X$ . These differences are reflected in different search biases of the corresponding geometric crossovers. In some cases,  $GY$  can be used instead of  $GX$  to remove unwanted biases.

Unfortunately, the crossover  $GY$  *cannot be directly implemented* because it recombines phenotypes that cannot be directly represented (or that are not convenient to represent directly). However, it turns out to be possible to exploit the affinity between the distances  $d_Y$  and  $d_X$  to search the space of phenotypes with the crossover  $GY$  *indirectly* by manipulating the genotypes  $X$  with the crossover  $GX$  after an appropriate formal transformation of  $GX$ .

#### 3.1 Definition and Properties

The genotype-phenotype mapping  $\phi$  naturally induces an equivalence relation  $\sim$  on the genotype space  $X$ . Under certain conditions, the quotient metric of genotype space  $(X, d_X)$  with respect to  $\sim$  returns

---

<sup>4</sup>It is true for non-synonymously redundant encodings. But, for synonymously redundant encodings, a reduction of the search space does not improve the performance of genetic algorithms [34].

a well-defined metric on the phenotype space  $(Y, d_Y)$ . However, not every equivalence relation gives a metric on its quotient space. The class of equivalence relations defined using isometry subgroups gives rise to a well-defined quotient metric space, as follows. Let  $(X, d_X)$  be a metric space and  $(\mathcal{G}, \cdot) \subseteq Iso(X)$  be an isometry subgroup, where ‘ $\cdot$ ’ denotes the function composition operator. We introduce a relation  $\sim$ :  $x$  and  $y$  in  $X$  are related if and only if  $x = g(y)$  for some  $g \in \mathcal{G}$ . Then  $\sim$  is an equivalence relation on  $X$  [4]. Once an equivalence relation is defined using an isometry subgroup, we can obtain a natural metric on its quotient space  $Y = X / \sim$ , which is induced by the original metric  $d_X$ . For  $x \in X$ ,  $\bar{x}$  denotes the equivalence class to which  $x$  belongs. That is,  $\bar{x} = \{g(x) \mid g \in \mathcal{G}\}$ . In [4], it is proven using the concept of *group action* that the metric  $d_Y$  on  $Y$  induced by the original metric  $d_X$  is as follows: the *quotient metric*  $d_Y(\bar{x}, \bar{y}) = \min\{d_X(x', y') \mid x' \in \bar{x}, y' \in \bar{y}\}$ .

The following proposition gives a characterization of the quotient metric  $d_Y$  that allows to compute it more efficiently. It says that to compute the quotient metric  $d_Y(\bar{x}, \bar{y})$  given  $x, y \in X$ , we can fix any representative element  $x$  of  $\bar{x}$  and have only to consider the distances to any element of  $\bar{y}$ , instead of considering the distances between all possible pairs of elements of  $\bar{x}$  and  $\bar{y}$  as in the original definition of quotient metric  $d_Y$ . This characterization is more practical than  $d_Y$  because the number of cases to consider is much smaller.

**Proposition 1.** *For any fixed  $x \in \bar{x}$ , let  $\tilde{d}_Y(\bar{x}, \bar{y}) := \min\{d_X(x, y') \mid y' \in \bar{y}\}$ . Then  $\tilde{d}_Y(\bar{x}, \bar{y}) = d_Y(\bar{x}, \bar{y})$ .*

*Proof.* Let  $\bar{x}, \bar{y} \in Y$ . It is clear that  $\tilde{d}_Y(\bar{x}, \bar{y}) \geq d_Y(\bar{x}, \bar{y})$  by definition. Now we will show that  $\tilde{d}_Y(\bar{x}, \bar{y}) \leq d_Y(\bar{x}, \bar{y})$ . Suppose that  $d_Y(\bar{x}, \bar{y}) = d_X(x_1, y_1)$ , where  $x_1 \in \bar{x}$  and  $y_1 \in \bar{y}$ . Since  $x$  and  $x_1$  belong to the same equivalence class, there exists  $g \in \mathcal{G}$  such that  $x = g(x_1)$ . Since  $g$  is an isometry,  $d_X(x_1, y_1) = d_X(g(x_1), g(y_1)) = d_X(x, g(y_1))$ .  $y \sim y_1$  and  $y_1 \sim g(y_1)$ . So  $y \sim g(y_1)$ . Hence  $\tilde{d}_Y(\bar{x}, \bar{y}) \leq d_X(x, g(y_1)) = d_X(x_1, y_1) = d_Y(\bar{x}, \bar{y})$ .  $\square$

Since a quotient metric space is a metric space, the line segment and the geometric crossover in this space are well-defined. However, the geometric crossover associated with this space is an *abstract operator*, in the sense that it cannot be directly implemented. This is a consequence of phenotypes, as a norm, being abstract entities themselves, which require to be mapped to genotypes to be represented in a computer. However, we can exploit the quotient relationship between the metric spaces of genotypes  $(X, d_X)$  and of phenotypes  $(Y, d_Y)$ , to determine a general representation-independent procedure to implement the geometric crossover on phenotypes *indirectly* acting on corresponding genotypes, which are entities that can be directly manipulated. In the following, such a general procedure is derived formally.

Given a genotype space  $(X, d_X)$  and its induced quotient phenotype space  $(Y, d_Y)$ , a *quotient closed ball* is the set of the form  $\overline{B}_{d_Y}(x, r) = \{y \in X \mid d_Y(\bar{x}, \bar{y}) \leq r\}$ , where  $r \in \mathbb{R}^+$ . A *quotient line segment* is the set of the form  $\overline{[x; y]}_{d_Y} = \{z \in X \mid d_Y(\bar{x}, \bar{z}) + d_Y(\bar{z}, \bar{y}) = d_Y(\bar{x}, \bar{y})\}$ . Notice that quotient closed balls and quotient line segments are sets on the genotype space defined as functions of closed ball and line segment in the phenotype space. Intuitively, they can be understood as the ‘‘shadow’’ of shapes in the phenotype space projected to the genotype space via the inverse of genotype-phenotype mapping  $(\phi^{-1})$ . The next proposition relates quotient closed balls and quotient line segments to closed balls and line segments in the genotype space, respectively.

**Proposition 2.** (i)  $\forall r \in \mathbb{R}^+, \forall x \in X, B_{d_X}(x, r) \subseteq \overline{B}_{d_Y}(x, r)$ .  
(ii)  $\forall x, y \in X$  and  $y^* \in \bar{y}$  such that  $d_Y(\bar{x}, \bar{y}) = d_X(x, y^*)$  then  $[x; y^*]_{d_X} \subseteq \overline{[x; y]}_{d_Y}$ .

*Proof.* (i) Let  $y \in B_{d_X}(x, r)$ . Then,  $d_X(x, y) \leq r$ , and so  $d_Y(\bar{x}, \bar{y}) \leq r$  by the definition of quotient metric. Hence  $y \in \overline{B}_{d_Y}(x, r)$ .

(ii) Let  $z \in [x; y^*]_{d_X}$ . Then,  $d_Y(\bar{x}, \bar{z}) + d_Y(\bar{z}, \bar{y}) \leq d_X(x, z) + d_X(z, y^*) = d_X(x, y^*) = d_Y(\bar{x}, \bar{y})$ . Since  $d_Y(\bar{x}, \bar{z}) + d_Y(\bar{z}, \bar{y}) \geq d_Y(\bar{x}, \bar{y})$  by the triangular inequality,  $d_Y(\bar{x}, \bar{z}) + d_Y(\bar{z}, \bar{y}) = d_Y(\bar{x}, \bar{y})$ . So,  $z \in \overline{[x; y]}_{d_Y}$ .  $\square$

**Definition 2** (Quotient geometric crossover). *A binary operator QGX on the genotype space  $X$  is a quotient geometric crossover under the metric  $d_X$  and the equivalence relation  $\sim$  if all offspring are in the quotient line segment between its parents  $x$  and  $y$ , i.e.,  $Im[QGX](x, y) \subseteq \overline{[x; y]}_{d_Y}$ .*

This is an important definition that requires some attention. Notice that the quotient geometric crossover QGX acts on genotypes and requires that the offspring (genotypes) lie in the “shadow” of the line segment in the phenotype space whose endpoints are the phenotypes of its parents. This, in effect, implements a geometric crossover GY in the phenotype space indirectly via acting on the genotype space because the corresponding phenotypes of the offspring lie necessarily in the line segment between its parents’ phenotypes (by projecting back the offspring in the “shadow” from the genotype space to the phenotype space via the mapping  $\phi$ ).

The quotient geometric crossover, although being in principle implementable because of acting on genotypes, has a non-operational definition that does not suggest how to actually construct the crossover. In the following, we define the *induced quotient crossover* (IQX) and we show that it is a quotient geometric crossover. Interestingly, this crossover can be derived by transforming the original geometric crossover GX on the genotype space under the distance  $d_X$ . The involved transformation is a generalization of the *normalization*, which we mentioned in the introduction. This transformation can be defined very generally, in a representation-independent way, given a metric space and an isometry subgroup on this space. The induced quotient crossover IQX turns out to be easily implemented because, as a norm, both the original geometric crossover GX on the genotype space and the normalization transformation are constructive.

**Definition 3** (Induced quotient crossover). *The operator IQX( $x, y$ ) with  $x, y \in X$  is as follows. First, normalize  $y$  with respect to  $x$  by finding the  $y^*$  equivalent to  $y$  (in  $\bar{y}$ ) which is at a minimum genotypic distance from  $x$ , i.e.,  $d_Y(\bar{x}, \bar{y}) = d_X(x, y^*)$ . Then, use the geometric crossover GX associated with  $X$  to recombine the first parent  $x$  and the normalized second parent  $y^*$ .*

**Corollary 1.** *Induced quotient crossover IQX is a quotient geometric crossover.*

*Proof.* Let  $y^* \in \bar{y}$  be a normalized second parent, i.e.,  $d_Y(\bar{x}, \bar{y}) = d_X(x, y^*)$ . Then,  $[x; y^*]_{d_X} \subseteq \overline{[x; y]}_{d_Y}$  by Proposition 2. This satisfies the definition of quotient geometric crossover.  $\square$

In summary, the induced quotient crossover IQX can be seen as a bridge between the original geometric crossover GX on the genotype space and the quotient geometric crossover QGX, since it is characterized in operational terms. In turn, QGX becomes a bridge between phenotype space and genotype space, as it characterizes on the genotype space a geometric crossover GY on the phenotype space.

## 4 Example Applications

In this section, we consider four example applications of genotype-phenotype encodings for different representations and we show how to use the quotient framework to derive crossover operators which are completely unaffected by the redundancy of the encoding.

In the first example about groupings in Section 4.1, a previously proposed crossover [27] is reinterpreted as a quotient geometric crossover. In the other examples, we derive new quotient geometric crossovers consistent with the genotype-phenotype mapping of the given problems.

The purpose of this section goes beyond illustrating the application of the quotient geometric crossover framework to specific examples. Importantly, it is to show that different types of representations, which are used for various problems, that have never been conceptually considered together are naturally unified and explained in a rigorous manner within this framework.

### 4.1 Groupings

*Grouping* problems [13] are commonly concerned with partitioning a given set of items into mutually disjoint subsets. Examples belonging to this class of problems are multiway graph partitioning, graph coloring, bin packing, and so on. This class of problems has been approached successfully with evolutionary algorithms [10, 12, 27].

Most normalization studies for grouping problems have focused on the  $k$ -way partitioning problem [23], which is roughly the problem of partitioning a given node set into  $k$  indistinguishable subsets (for the exact definition of the problem, see Section 4.5). In this problem, the  $k$ -ary representation has been generally used. In this representation, one phenotype (a  $k$ -way partition) is represented by  $k!$  different genotypes, which are all the possible ways of labeling the subsets. For this problem, a normalization method was used in [7, 18]. In the sense that normalization pursues the minimization of genotype inconsistency among parental chromosomes (position-wise mismatches), in [20], Kim and Moon proposed an optimal, efficient normalization method for grouping problems based on a novel distance measure, the *labeling-independent distance*, that eliminates dependency on the underlying labeling completely. In the following, we show that, in effect, the recombination that includes the normalization procedure followed by standard crossover on discrete vectors, is an instance of quotient geometric crossover.

Let  $a, b \in X = \{0, 1, \dots, k-1\}^n$  be  $k$ -ary genotypes (fixed-length vectors on a  $k$ -ary alphabet), where  $X$  is the genotype space. The phenotype space  $Y$  is the set of all possible ways of partitioning the set of items into  $k$  subsets, irrespective of the labeling of the subsets. Let  $\Sigma_k$  be the set of all permutations of length  $k$ . For each  $\sigma \in \Sigma_k$ , we can view  $\sigma$  as a function on  $X$  by defining  $\sigma(a)$  to be a permuted encoding of  $a$  by a permutation  $\sigma$ . For example, in the case that  $a = (0, 1, 2, 2, 1, 3, 0, 3)$  is a 4-ary genotype and  $\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 3 & 2 & 0 \end{pmatrix} \in \Sigma_4$ , then  $\sigma(a) = (1, 3, 2, 2, 3, 0, 1, 0)$ . The genotype  $\sigma(a)$  has the same phenotype as  $a$  for each permutation  $\sigma \in \Sigma_k$  because applying the permutation to the genotype  $a$  has only the effect of relabeling the groups and does not affect how the set of items is partitioned. Therefore, in this case, the equivalence relation  $\sim$  between genotypes sees genotypes as equivalent when they represent the same partitioning of the set of items.

The crossover on genotypes GX considered is the standard crossover on discrete vectors. This crossover is a geometric crossover on  $X$  under Hamming distance  $H$  [26], which is then the distance  $d_X$  on the genotype space to consider in this case.

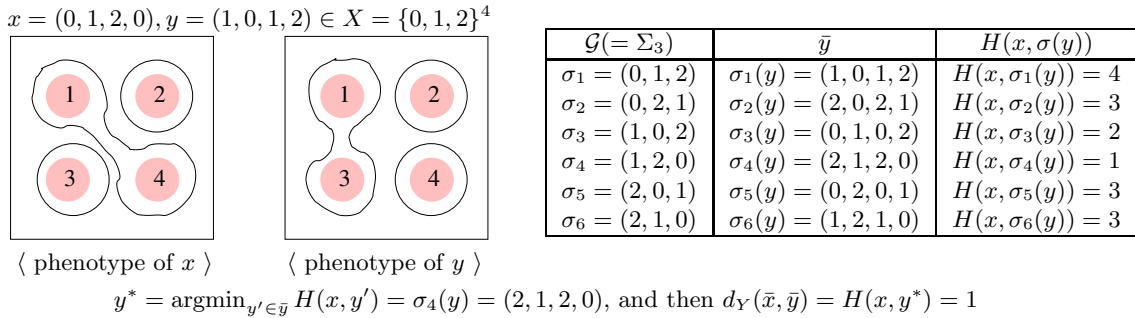


Figure 1: An example in 3-grouping of 4 items

It is well-known that permutations form a group [14, 35], which is known as *symmetric group*. Hence,  $\Sigma_k$  is a group. Moreover, it is easy to check that each  $\sigma \in \Sigma_k$  is an isometry on  $X$  with respect to the Hamming distance and that the equivalence relation  $\sim$  is associated with this isometry subgroup  $\Sigma_k$ . Since these pre-conditions hold, the phenotype space  $Y$  is associated with a well-defined quotient metric  $d_Y$ . By Proposition 1, the quotient metric so obtained coincides to the labeling-independent distance introduced in [20] and shown as follows:  $d_Y(\bar{a}, \bar{b}) = \min_{\sigma \in \Sigma_k} H(a, \sigma(b))$ . An example is shown in Figure 1. In this example, the phenotype space  $Y$  consists of all 3-groupings of 4 items. The figure shows that the phenotypic distance between two 3-groupings  $\bar{x} = \{\{1, 4\}, \{2\}, \{3\}\}$  and  $\bar{y} = \{\{2\}, \{1, 3\}, \{4\}\}$  is 1 by using the quotient metric  $d_Y$  on  $X/\sim$ .

The induced quotient crossover IQX can be obtained using the following: *normalize the second parent to the first under the Hamming distance, and then apply a geometric crossover using the first parent and the normalized second parent*. Corollary 1 then guarantees that this crossover can be understood as a geometric crossover GY associated with the phenotype space  $(Y, d_Y)$ . As anticipated, the presented IQX coincides to the labeling-independent crossover presented in [21].

In order to implement IQX, one needs a way to implement efficiently the normalization specifically for this class of genotypes and for the Hamming distance. A naïve method to implement the normalization is to enumerate all  $k!$  permutations and find the optimal one among them. However, for a large  $k$ , such a procedure is intractable. Fortunately, the normalization can be done in  $O(k^3)$  time using the *Hungarian method* proposed by Kuhn [22].<sup>5</sup>

## 4.2 Graphs

We consider any problem naturally defined over a graph in which the fitness of the solution does not depend on the labels on the nodes but only on the structural relationship between them (determined by how edges link nodes). Graphs are a very important representation with many applications as they are ubiquitous. Encodings for graphs, e.g., adjacency matrix, are inherently redundant, hence the application of the quotient framework is potentially very interesting as it allows us to design crossover operators which are not affected by the redundancy.

Formally, let  $A \in \mathfrak{M}_n$  be the adjacency matrix of a labeled graph with  $n$  nodes (genotype) and

<sup>5</sup>We can also consider some fast heuristic algorithms for normalizing solutions approximately [2]. They can be implemented in  $O(k^2)$  and experimental studies indicated that these methods are, in practice, very close to the optimum and much faster than the Hungarian method.



let  $\mathcal{P}_n$  be a set of all  $n \times n$  permutation matrices, where a permutation matrix is a  $(0, 1)$ -matrix with exactly one 1 in every row and column. Then, for each permutation matrix  $P \in \mathcal{P}_n$ , the matrix  $PAP^T$  corresponds to the labeled graph obtained by relabeling the labeled graph  $A$  according to the permutation represented by  $P$ . So, the graphs  $A$  and  $PAP^T$  are isomorphic (graph isomorphism is the equivalence relation on the genotype space in this case). The fitness  $w: \mathfrak{M}_n \rightarrow \mathbb{R}$  satisfies that for every  $A \in \mathfrak{M}_n$  and every permutation matrix  $P$ ,  $w(A) = w(PAP^T)$ . This requirement captures formally the idea that the fitness of a solution depends only on the underlying structure of the graph (phenotype) and not on the way its nodes are labeled.

Let  $(\mathfrak{M}_n, H)$  be a metric space on the labeled graphs under the Hamming distance  $H$  on their adjacency matrices<sup>6</sup>. Notice that this metric is labeling-dependent. In particular, the distance  $H(A, PAP^T)$  between  $A$  and  $PAP^T$  may not be zero although  $A$  and  $PAP^T$  represent the same structure. If  $B$  is equal to  $PAP^T$  for some permutation matrix  $P$ , we define  $A$  and  $B$  to be in relation  $\sim$ , i.e.,  $A \sim B$ . Since a set of permutation matrices  $\mathcal{P}_n$  forms an isometry subgroup, the relation  $\sim$  is an equivalence relation.

The equivalence class  $\bar{A}$  is represented as follows:  $\bar{A} = \{PAP^T \mid P \in \mathcal{P}_n\}$ . The equivalence class so defined naturally corresponds to an *unlabeled graph* and consequently the quotient space  $\mathfrak{M}_n / \sim$  can be understood as the *space of unlabeled graphs*. To make the link with the quotient geometric crossover explicit, we have that the space of adjacency matrices is the genotype space  $X$ , which we can represent and manipulate directly in the search, and the space of the unlabeled graphs is the phenotype space  $Y$ , which we actually would like to search but we cannot represent directly. The space  $\mathfrak{M}_n / \sim$  is a quotient metric space endowed with the metric:  $d_Y(\bar{A}, \bar{B}) = \min_{P \in \mathcal{P}_n} H(A, PBP^T)$ . An example for graph representation is shown in Figure 2. In this example,  $Y$  consists of all graphs with 3 nodes. The figure illustrates that for two 3-node graphs the phenotype distance  $d_Y$  between them is 2 through the normalization using permutation matrices.

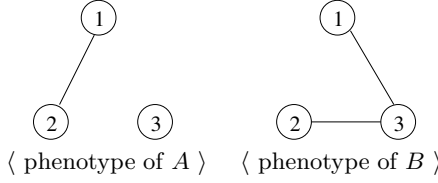
The distance  $d_Y(\bar{A}, \bar{B})$  between graphs coincides to the ins/del edge edit distance between (unlabeled) graphs with the same number of nodes, which is defined as the minimum number of edges to add to and delete from a graph to obtain the other graph. Similarly to the case of groupings, the geometric crossover under this phenotypic distance, on unlabeled graphs in this case, is an abstract operator that cannot be directly implemented because unlabeled graphs are abstract structures in the sense that they cannot be directly represented in a computer.

In the following, we present a recombination operator, the quotient geometric crossover, that implements the geometric crossover on the space of unlabeled graphs (phenotypes) indirectly by manipulating their adjacency matrices (genotypes). As for the case of groupings, this recombination operator can be broken down in two separate phases using the notion of induced quotient crossover (Definition 3) in a normalization phase followed by a crossover phase. The crossover phase consists of a geometric crossover on adjacency matrices under the Hamming distance. Operationally, this crossover can be shown to be a straightforward extension to matrices of the mask-based crossover for binary strings, where the mask is a matrix. Applying the formal definition of normalization to the specific case, the normalization phase consists of finding the adjacency matrix by relabeling the nodes of the second parent graph such that the Hamming distance to the adjacency matrix of the first parent graph is minimized. This coincides to the maximal graph matching problem [41]. Since

---

<sup>6</sup>Since the space of labeled graphs and the space of their corresponding adjacency matrices are isomorphic, we can use them interchangeably.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in X = \mathfrak{M}_3$$



$\mathcal{G}(= \mathcal{P}_3)$	$B$	$H(A, PBP^T)$
$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$P_1BP_1^T = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$	$H(A, P_1BP_1^T) = 6$
$P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$P_2BP_2^T = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$H(A, P_2BP_2^T) = 2$
$P_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$P_3BP_3^T = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$	$H(A, P_3BP_3^T) = 6$
$P_4 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	$P_4BP_4^T = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$H(A, P_4BP_4^T) = 2$
$P_5 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$P_5BP_5^T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$H(A, P_5BP_5^T) = 2$
$P_6 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$P_6BP_6^T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$H(A, P_6BP_6^T) = 2$

$$B^* = \operatorname{argmin}_{B' \in \bar{B}} H(A, B') = P_2BP_2^T, P_4BP_4^T, P_5BP_5^T, \text{ or } P_6BP_6^T, \text{ and then } d_Y(\bar{A}, \bar{B}) = H(A, B^*) = 2$$

Figure 2: An example in 3-node graph

the subgraph isomorphism is NP-complete [25], one drawback for the use of graph matching is from its time complexity and this is an inherent difficulty of the graph matching problem. However, a research effort has been made to develop computationally tractable graph matching algorithms [31], so such efficient heuristics can be employed to implement the normalization phase. In summary, the induced quotient crossover for graphs is as follows: *Do the graph matching of the second parent  $B$  to the first  $A$  under the Hamming distance  $H$ , i.e.,  $B^* := \operatorname{argmin}_{B' \in \bar{B}} H(A, B')$ , and then do a geometric crossover using the first parent  $A$  and the graph-matched second parent  $B^*$ .* The induced quotient crossover implements indirectly, using adjacency matrices, a geometric crossover on unlabeled graphs  $\mathfrak{M}_n / \sim$ . Therefore the space actually searched by this geometric crossover is the phenotype space, not the genotype space.

A reason to be confident in the stronger performance of the quotient geometric crossover on graphs over its corresponding crossover on adjacency matrices, is that the edit distance upon which the phenotypic geometric crossover is based is arguably a more natural choice of metric on graphs than the Hamming distance on the adjacency matrix, in the sense that it is a better measure of distance between what a graph is meant to represent. In this sense, the phenotypic distance embeds more problem-knowledge in the search and the corresponding geometric crossover is better tailored to graphs. Hence the locality of the problem gets higher and the quotient geometric crossover creates more meaningful offspring.

$x = (1, 4, 5), y = (3, 0, 6) \in X = \mathbb{R}^3$	$\mathcal{G}(= \Sigma_3)$	$\bar{y}$	$E(x, \sigma(y))$
$\{1, 4, 5\}$ ⟨ phenotype of $x$ ⟩	$\sigma_1 = (1, 2, 3)$	$\sigma_1(y) = (3, 0, 6)$	$E(x, \sigma_1(y)) = \sqrt{21}$
$\{3, 0, 6\}$ ⟨ phenotype of $y$ ⟩	$\sigma_2 = (1, 3, 2)$	$\sigma_2(y) = (3, 6, 0)$	$E(x, \sigma_2(y)) = \sqrt{33}$
	$\sigma_3 = (2, 1, 3)$	$\sigma_3(y) = (0, 3, 6)$	$E(x, \sigma_3(y)) = \sqrt{3}$
	$\sigma_4 = (2, 3, 1)$	$\sigma_4(y) = (0, 6, 3)$	$E(x, \sigma_4(y)) = 3$
	$\sigma_5 = (3, 1, 2)$	$\sigma_5(y) = (6, 3, 0)$	$E(x, \sigma_5(y)) = \sqrt{51}$
	$\sigma_6 = (3, 2, 1)$	$\sigma_6(y) = (6, 0, 3)$	$E(x, \sigma_6(y)) = 3\sqrt{5}$

$y^* = \operatorname{argmin}_{y' \in \bar{y}} E(x, y') = \sigma_3(y) = (0, 3, 6)$ , and then  $d_Y(\bar{x}, \bar{y}) = E(x, y^*) = \sqrt{3}$

Figure 3: An example in 3-variable symmetric function under the Euclidean distance

### 4.3 Symmetric Functions

A *symmetric function* on  $n$  variables  $x_1, x_2, \dots, x_n$  is a function with the property that its returned values remain unchanged under any permuted order of its input variables. That is, if  $f(x_1, x_2, \dots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$  for any permutation  $\sigma$ , the function  $f$  is called symmetric function. For example, the one-max problem [38], which is a simple problem consisting in maximizing the number of ones of a bitstring, is a symmetric function. In this subsection, we consider problems whose fitness function is symmetric. Some evolutionary studies have been made on such problems over sets [44] or real vector space [39]. More properties about specific symmetric functions are introduced in [6, 40].

The form of solutions of problems defined using symmetric functions are  $n$ -dimensional vectors, i.e., length- $n$  strings corresponding to the vector of input variables of the symmetric function. However, the vectorial representation encodes the order of the input variables that, in effect, is irrelevant information for the purpose of determining the fitness of a solution because any order of the input variables leaves the fitness of the solution unchanged, as the fitness is a symmetric function. In this case, the quotient geometric crossover can be applied to search the much smaller space of the “unordered” input vectors, in effect, treating the input vectors as a (multi-)sets of fixed size, instead of the space of input vectors. We illustrate this idea in the following.

Let  $X$  be the solution space (genotype space) corresponding to the domain of a given symmetric function. Let  $\Sigma_n$  be the set of all permutations of length  $n$ . Similarly to the example of groupings in Section 4.1, a permutation  $\sigma \in \Sigma_n$  can be understood as a function which reorders the elements of an input vector  $x$ . For example, in the case that  $x = (x_1, x_2, x_3, x_4)$  and  $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} \in \Sigma_4$ ,  $\sigma(x) = (x_2, x_4, x_3, x_1)$ .  $\Sigma_n$  is a well-known symmetric group [14, 35], and analogously as the example of groupings, it can be shown that  $\Sigma_n$  is a group and each  $\sigma$  is an isometry. In the following, we consider an example case of real input vectors.

In the case that the genotype space  $X$  is a real vector space, we can endow it with the Euclidean distance  $E$  (or other Minkowski distances<sup>7</sup> such as the Manhattan distance). The induced quotient metric on  $X/\sim$  specified for the Euclidean distance is as follows:  $d_Y(\bar{x}, \bar{y}) = \min_{\sigma \in \Sigma_n} E(x, \sigma(y))$ . Figure 3 is helpful to understand the quotient metric space for this case. The figure shows an example in symmetric functions over  $\mathbb{R}^3$ . The phenotypic distance  $d_Y$  between two real vectors  $x = (1, 4, 5)$  and  $y = (3, 0, 6)$  becomes  $\sqrt{3}$ .

The induced quotient crossover, that implements the geometric crossover that searches the quotient space of “unordered” real vectors indirectly via manipulating the corresponding (ordered) real vectors,

<sup>7</sup>The underlying distance is a designer’s choice. Different distances are suited to different problems.

can be formally derived by plugging the specific distance and the specific equivalence relation for this example in Definition 3. The derivation of an operational definition of this operator can be done analogously to the case of groupings. The induced quotient crossover for the specific case comprises a normalization phase that can be performed in  $O(n^3)$  time using the Hungarian method, followed by a geometric crossover on real vectors under the Euclidean distance, which operationally coincides to the traditional blend crossover on real vectors. However, in the special case of Minkowski distances, it is easy to see that there exists a more efficient way to do the normalization, as it can be done by just sorting the two parents. This method returns an optimal arrangement of elements of the parent genotypes which minimizes the distances as required and its time complexity is  $O(n \log n)$ . However this method can only be applied to the case of real or integer vectors and it is not applicable to general symmetric functions.

#### 4.4 Circular Permutations

We consider the case in which the solutions of a problem are naturally represented as circular permutations such as in the traveling salesman problem (TSP). By gluing the head and the tail of a permutation we obtain a circular permutation. Circular permutations cannot be represented directly. They are typically represented with simple permutations. Then each circular permutation is represented by more than one permutation. For example, permutations  $(1, 2, 3)$ ,  $(2, 3, 1)$ , and  $(3, 1, 2)$  represent the same phenotype, i.e., circular permutation. In such problems, the genotype space is the set of permutations and the phenotype space is the set of circular permutations. We can consider this problem in view of genotype-phenotype mapping using the concept of quotient space. More specifically, our aim is to use the quotient geometric crossover framework to derive formally a recombination operator (the induced quotient crossover) acting on simple permutations to simulate a phenotypic geometric crossover that searches the smaller space of circular permutations.

Let  $\Sigma_n$  be the set of all permutations of length  $n$ . Let us define a function  $s_k : \Sigma_n \rightarrow \Sigma_n$  that applies to the input permutation  $k$ -step circular shift operations to right. For example,  $s_2(1, 2, 3) = (2, 3, 1)$ . The set of all shift operations  $S_n = \{s_k \mid k = 0, 1, 2, \dots, n - 1\}$  is a group. When we use position-independent distances on  $X (= \Sigma_n)$  such as the Hamming distance or swap distance, it is easy to check that each  $s_k \in S_n$  is an isometry on  $\Sigma_n$ , i.e.,  $d_X(s_k(\sigma_1), s_k(\sigma_2)) = d_X(\sigma_1, \sigma_2)$  for each  $\sigma_1, \sigma_2 \in \Sigma_n$ . From the position-independent distance on  $\Sigma_n$ ,  $\Sigma_n / \sim$  has an induced quotient metric.

In the following we consider several alternative distances on permutations and derive the corresponding induced quotient crossovers. Let us start by considering the most typical distance for discrete spaces, the Hamming distance  $H$ . It is known that the cycle crossover for permutations is geometric under this distance [29]. The quotient metric for the Hamming distance relative to the equivalence relation induced by the group  $S_n$  is:  $d_Y(\bar{x}, \bar{y}) = \min_{s \in S_n} H(x, s(y))$ . An example case is shown in Figure 4. The figure illustrates an example of circular permutation of length 6. We can see that the phenotype distance  $d_Y$  between two circular permutations  $x = (2, 4, 5, 1, 6, 3)$  and  $y = (4, 6, 1, 5, 3, 2)$  becomes 2.

Using the framework of quotient geometric crossover, we can formally derive the following induced quotient crossover for this specific case: *normalize the second parent to the first under the Hamming distance  $H$ , and then do the cycle crossover using the first parent and the normalized second parent.* The normalization procedure in the crossover defined above is obtained by specifying the formal definition of normalization for the equivalence relation induced by the isometry subgroup  $S_n$  and the

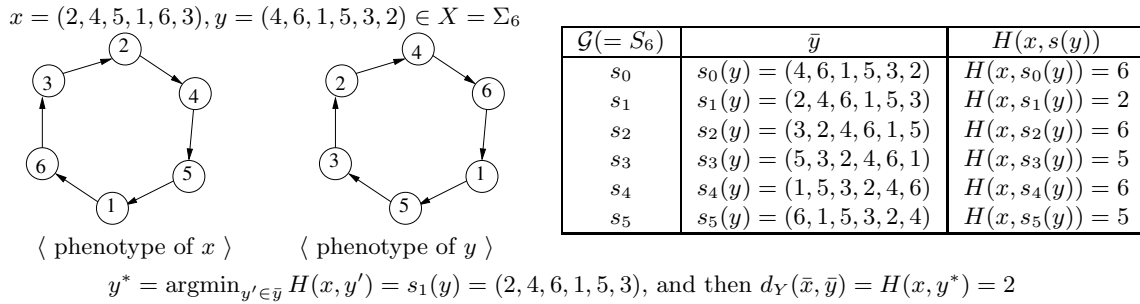


Figure 4: An example in circular permutation under the Hamming distance

Hamming distance. This normalization can be easily implemented. It takes  $O(n)$  time to normalize the second parent because the equivalence class of the second parent has exactly  $n$  elements by shift operations.

Cycle crossover is also geometric under the swap distance [29]. The induced quotient crossover can be defined in a similar way using the swap distance instead of the Hamming distance.

We can use another well-known distance on permutations, the *reversal distance*. Its neighborhood structure is based on the 2-opt move (reversal). The reversal move selects any two points along the permutation and then reverses the subsequence between these points. This move induces a graphic distance between permutations, the reversal distance, defined as the minimum number of reversals to transform one permutation into the other. There is a version of the reversal distance for circular permutations defined in an analogous way. The geometric crossovers associated with these distances belong to the family of sorting crossovers [29]: they pick offspring on the minimum sorting trajectory between parent (circular) permutations sorted by reversals. Applying the framework of quotient geometric crossover to the reversal distance on permutation with equivalence relation induced by the isometry subgroup  $S_n$ , we obtain the following induced quotient crossover: *normalize the second parent to the first under the reversal distance, and then do the crossover based on sorting by reversals for permutation using the first parent and the normalized second parent.*

The framework shows that this recombination implements the geometric crossover under reversal distance on *circular* permutations using the geometric crossover under reversal distance on *simple* permutations preceded by normalization. Interestingly, this example of quotient geometric crossover illustrates how to obtain a geometric crossover for a transformed representation (circular permutation) starting from a geometric crossover for the original representation (permutation). So in this case quotient geometric crossover is used as a tool to build a new crossover for a derivative representation from a known geometric crossover for the original representation.

There is a problem in implementing the geometric crossover under the reversal distance. Sorting linear or circular permutations by reversals is NP-hard [5]. So the geometric crossover under the reversal distance cannot be implemented efficiently. However, a good approximation of this crossover can be implemented in quadratic time [19]. From [29], we can know that the sorting-by-reversals crossover for permutations is an excellent crossover for TSP.

## 4.5 Experimental Support

We made experiments to compare the performance of geometric crossover and its induced quotient geometric crossover. Among the presented example applications, we chose three cases: groupings, symmetric functions, and circular permutations. As a representative problem of each class, we adopted a well-known NP-hard problem. Multiway graph partitioning (MGP) [37], the maximum  $k$ -coverage problem (MKCP) [15], and the traveling salesman problem (TSP) [24] are selected for groupings, symmetric functions, and circular permutations, respectively. In the following, we give a brief description of each problem.

– (Multiway graph partitioning). Let  $G = (V, E)$  be an unweighted undirected graph, where  $V$  is the set of vertices and  $E$  is the set of edges.  $k$ -way partition is a partitioning of the vertex set  $V$  into  $k$  disjoint subsets. A  $k$ -way partition is said to be *balanced* if the difference of cardinalities between the largest and the smallest subsets is at most one. The *cut size* of a partition is defined to be the number of edges with endpoints in different subsets of the partition. The  $k$ -way partitioning problem is the problem of finding a  $k$ -way balanced partition with the minimum cut size.

– (Maximum  $k$ -coverage problem). MKCP is a generalized version of covering problem. Let  $A = (a_{ij})$  be an  $m \times n$  zero-one matrix and weight  $w_i$  be given to the  $i$ -th row of  $A$  for each  $i$ . The objective of MKCP is to select  $k$  columns that *cover* the rows for their weight-sum to be maximized. Here *cover* means that for each row the value of at least one column among selected  $k$  columns is 1. The weight  $w_i$  can be different according to  $i$ , however, we will deal with the case that each  $w_i$  is set to be the same value in our experiments. Then MKCP becomes the problem of finding a solution that covers as many rows as possible using a fixed number of columns.

– (Traveling salesman problem). Given  $n$  cities and their pair-wise distances, TSP is the problem of finding the shortest tour visiting all the cities.

To classify different types of redundant representations, Rothlauf proposes to measure the degree of similarity of genotypes with the same phenotype [34]. Then he defines a representation to be *synonymously redundant* if the genotypes that are assigned to the same phenotype are similar to each other. Otherwise, the representation is said to be *non-synonymously redundant*. We measured the *synonymity* of each representation to have an indication of the hardness of the genetic search based on the traditional redundant encoding. Refining a previous measure of *synonymity* [33], we define the *synonymity* as follows:  $\frac{1}{\sum_{\bar{x} \in Y} |\bar{x}|(|\bar{x}|-1)} \sum_{\bar{x} \in Y} \sum_{x_1, x_2 \in \bar{x} (x_1 \neq x_2)} d_X(x_1, x_2)$ .

The above formula has the following probabilistic interpretation: it is the expected distance between two (distinct) genotypes  $x_1$  and  $x_2$  sampled uniformly at random in the genotype space *provided* that we know that they have the same phenotype. If this distance is significantly smaller than the average distance between genotypes in the entire genotype space then the representation is synonymously redundant (otherwise it is non-synonymously redundant). This interpretation is sound in that knowing that the genotypes correspond to the same phenotype implies that they are (expected to be) closer to each other (more similar) than when not having such information, hence the representation is synonymously redundant. Clearly, the smaller the measure of synonymity, the more synonymously redundant the representation.

Table 1 gives the synonymity of the representation in each test problem. For this test, we generate just one random instance for each problem. The synonymity depends only on the problem size because it is a function of only the metric on  $X$  and not of the specific fitness function. Hence, it is enough to measure the synonymity of just one instance for each test problem. In all test problems,

Table 1: Synonymity of Representations in Test Problems

Problem type	Problem	Instance size	Distance $d_X$ in $X$	Max distance	Ave distance <sup>†</sup>	Synonymity <sup>‡</sup>
Groupings	MGP	$ V  = 500$ $k = 50$	Hamming distance	500 (= $ V $ )	490.00 ( $\approx  V  \cdot (1 - \frac{1}{k})$ )	490.00 ( $\approx  V  \cdot (1 - \frac{1}{k})$ )
Symmetric functions	MKCP	$n = 1000$ $k = 20$	Hamming distance	20 (= $k$ )	19.98 ( $\approx k$ )	19.00 (= $k - 1$ )
Circular permutations	TSP	$n = 500$	Hamming distance	500 (= $n$ )	499.00 (= $n - 1$ )	<b>500.00</b> (= $n$ )
			swap distance	499 (= $n - 1$ )	493.20	493.20

<sup>†</sup> Average from  $10^7$  random genotype pairs  $(x_1, x_2) \in X \times X$  ( $x_1 \neq x_2$ ).

<sup>‡</sup> Average from  $10^7$  random genotype pairs  $(x_1, x_2) \in \bar{x} \times \bar{x}$  ( $x_1 \neq x_2, \bar{x} \in Y$ ).

the synonymity value is the same as or very close to the average distance value, and so each representation is non-synonymously redundant. When using non-synonymously redundant representation, genetic operators like crossover or mutation can produce offspring that are phenotypically completely different from its parents (by measuring phenotypic difference using the quotient metric). Hence, non-synonymously redundant representations do not allow genetic operators to work properly and they reduce the efficiency of evolutionary search by making problems more difficult[33]. Quotient geometric crossover solves this problem as it removes redundancy in such encodings and it makes problems easier to solve using genetic algorithms. Extensive experiments in the following support this fact.

First, we performed experiments on the multiway graph partitioning problem (as an example of grouping problem). We used the steady-state genetic algorithms (GAs) of [27] on well-known benchmark data set from [16]. To see the effect of crossover, we did not use any local optimization. Following the genetic framework of [27], we used swap mutation with rate 0.005. For fair comparison, all GAs were stopped after 1,000 generations. We conducted tests on 32-way and 128-way partitioning. We compared a geometric crossover (GX) based on the Hamming distance restricted to permutations with repetitions and its corresponding induced quotient crossover (IQX) based on the labeling-independent distance. The specific geometric crossover on the genotype space we use is the generalized cycle crossover [27]. The other genetic parameters were the same as in [27]. Table 2 shows the average results obtained from 100 runs. We also carried out one-tailed  $t$ -tests of the null hypothesis that IQX is worse than or equal to GX. Under the same number of generations, we got clear improvements for all the tested instances in terms of the quality of solutions achieved. Moreover, we obtained more improvement on problem instances with larger number of groups ( $k = 128$ ).

Next, we tested on the maximum  $k$ -coverage problem (as an example of problem with symmetric fitness function). The basic evolutionary model we used is the genetic algorithm of [44], which is quite similar to the CHC [11] that uses a restart technique without mutation. Our experiments were conducted on 6 set covering test instances with various sizes and densities from OR-library [3]. All GAs terminated after 1,000 generations and output the best solution found so far. We conducted tests on  $k = 10$  and  $k = 20$ . We compared a geometric crossover (GX) based on the Hamming distance and its corresponding induced quotient crossover (IQX). As a geometric crossover on genotype space, we use the uniform crossover [28]. Table 3 shows the average results obtained from 100 runs. It also shows that the results of one-tailed  $t$ -tests of the null hypothesis that IQX is worse than or equal to

Table 2: Results of Multiway Graph Partitioning

Instance	Instance size ( $ V ,  E $ )	32-way partitioning ( $k = 32$ )				128-way partitioning ( $k = 128$ )			
		GX Ave(SD)	IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (ms) <sup>†</sup>	GX Ave(SD)	IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (ms) <sup>†</sup>
G500.2.5	$ V  = 500$ $ E  = 625$	578.19 (2.85)	576.88 (2.70)	3.34 (5.95E-04)	3.50	605.85 (1.60)	598.58 (2.07)	27.79 (3.95E-49)	81.82
G500.20	$ V  = 500$ $ E  = 5000$	4901.10 (5.39)	4898.48 (4.41)	3.76 (1.42E-04)	3.55	5056.70 (2.70)	5046.56 (2.85)	25.83 (2.43E-46)	80.73
U500.05	$ V  = 500$ $ E  = 1250$	1202.93 (5.08)	1200.49 (4.14)	3.72 (1.63E-04)	3.44	1254.14 (2.27)	1244.85 (2.58)	27.03 (4.49E-48)	81.28
U500.40	$ V  = 500$ $ E  = 10000$	8434.28 (10.43)	8429.54 (9.73)	3.32 (6.22E-04)	3.62	8697.89 (3.71)	8684.25 (4.41)	23.67 (4.48E-43)	80.80

GX is the generalized cycle crossover [27]. The smaller, the better.

<sup>†</sup> Average CPU time (ms) needed for normalization per each generation (on Intel Xeon CPU 2.40 GHz).

We used the Hungarian method for normalization.

Table 3: Results of Maximum  $k$ -Coverage Problem

Instance	Instance size	$k = 10$				$k = 20$			
		GX Ave(SD)	IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (s) <sup>†</sup>	GX Ave(SD)	IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (s) <sup>†</sup>
scp51	$m = 200$ $n = 2000$	84.14 (1.09)	87.31 (1.26)	19.10 (1.93E-35)	0.14	137.22 (2.18)	142.38 (2.26)	16.46 (1.58E-30)	0.95
scp61	$m = 200$ $n = 1000$	137.83 (1.72)	139.75 (1.34)	8.81 (1.98E-14)	0.17	187.97 (2.10)	190.90 (1.86)	10.44 (5.46E-18)	1.14
scpc1	$m = 400$ $n = 4000$	154.54 (3.08)	157.76 (2.90)	7.62 (7.33E-12)	0.14	248.55 (3.26)	255.90 (3.39)	15.64 (6.34E-29)	0.91
scpd1	$m = 400$ $n = 4000$	252.24 (3.91)	256.63 (3.65)	8.20 (4.24E-13)	0.19	352.36 (3.64)	358.42 (3.25)	12.42 (2.78E-22)	1.36
scpnrg1	$m = 1000$ $n = 10000$	312.61 (4.92)	319.49 (4.42)	10.40 (6.45E-18)	0.15	520.48 (7.07)	532.94 (6.12)	13.33 (3.32E-24)	1.00
scpnrh1	$m = 1000$ $n = 10000$	542.09 (5.55)	549.20 (5.55)	9.06 (5.60E-15)	0.21	800.86 (5.87)	812.71 (5.48)	14.75 (3.86E-27)	1.61

GX is the uniform crossover [28]. The larger, the better.

<sup>†</sup> Average CPU seconds needed for normalization per each generation (on Intel Xeon CPU 2.40 GHz).

We used the Hungarian method for normalization.

GX. IQX significantly improved GX for all the tested instances.

Finally, we did experiments for the traveling salesman problem, which is based on the circular permutation representation. We used the genetic algorithm of GALib [43]. Our experiments were conducted on 6 test instances with various sizes from TSPLIB [1]. To see the effect of crossover, we did not use any local optimization. Following the genetic framework of [43], we used swap mutation with rate 0.05. For fair comparison, all GAs were stopped after 1,000 generations. We also compared a geometric crossover (GX) under the Hamming distance on permutations (cycle crossover) and its corresponding induced quotient crossover (IQX). Table 4 shows the average results obtained from 100 runs together with the results of one-tailed  $t$ -tests. We got improvements for all of the tested instances



Table 4: Results of Traveling Salesman Problem

Instance	# of cities ( $n$ )	GX Ave(SD)	Hamming distance			Swap distance		
			IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (s) <sup>†</sup>	IQX Ave(SD)	$t$ -value ( $p$ -value)	Time (s) <sup>†</sup>
eil101	101	1017.16 (39.77)	1013.12 (39.48)	0.72 (2.36E-01)	0.03	1013.92 (39.36)	0.58 (2.82E-01)	0.04
lin318	318	196089.48 (4813.64)	194298.30 (4900.46)	2.61 (5.26E-03)	0.25	195891.46 (6044.17)	0.26 (3.99E-01)	0.41
pcb442	442	292489.40 (6573.36)	288970.58 (7359.65)	3.57 (2.79E-04)	0.49	290699.54 (6833.56)	1.89 (3.10E-02)	0.79
att532	532	580522.66 (13897.82)	576099.86 (13079.59)	2.32 (1.13E-02)	0.70	575308.16 (13388.88)	2.70 (4.04E-03)	1.15
rat575	575	43323.40 (872.22)	42670.79 (973.35)	4.99 (1.26E-06)	0.82	42701.23 (824.29)	5.18 (5.66E-07)	1.36
u724	724	366244.38 (7146.64)	361603.06 (6506.50)	4.80 (2.76E-06)	1.29	362136.44 (6234.07)	4.33 (1.76E-05)	2.15

GX is the cycle crossover [29]. The smaller, the better.

† Average CPU seconds needed for normalization per each generation (on Intel Xeon CPU 2.40 GHz).

and both distances, and in 9 of the 12 cases the improvements were statistically significant at the 0.05 level.

In summary, from our tests on three different classes of problems, the proposed quotient geometric crossovers were able to alleviate the difficulty of the genetic search arising from non-synonymously redundant representation.

## 5 Concluding Remarks

We have analyzed mathematically the mapping between genotype and phenotype spaces in a general setting and across representations. We showed that the phenotype space can be regarded as quotient space induced by the genotype-phenotype mapping. Geometric crossovers can be formally defined once a distance is defined. They have the advantage that they allow us to connect a solution space - seen as a metric space - and crossovers in a simple way. Moreover, geometric crossover based on an appropriate distance for the problem at hand can inform the search and improve performance. We showed that the phenotype space naturally inherits a metric, the quotient metric, from the genotype space through the genotype-phenotype mapping derived from an equivalence relation defined by an isometry group. Since the quotient metric is a part of the phenotype space structure, the geometric crossover on this metric reflects the properties of the phenotype space more effectively than the original geometric crossover on the genotype space. Due to the abstract nature of phenotypes, the geometric crossover on the phenotype space is well-defined but it is an abstract operator that cannot be directly implemented. We developed a formal framework that allows us to derive search operators defined on the genotype space that, in effect, implement geometric crossovers to search the phenotype space indirectly.

As shown in application examples, quotient geometric crossovers for non-synonymously redundant encodings are not only theoretically significant but have also a practical effect of making the search

more effective by reducing the search space and removing the low locality from the encodings. In the example of groupings, we newly reinterpreted geometric crossover [27] that was previously proposed by the authors, under more general framework of the quotient geometric crossover. In the examples of graphs, symmetric functions, and circular permutations, we induced quotient geometric crossovers better tailored to phenotype space by performing normalization. We tested the proposed phenotypic crossovers on well-known NP-hard optimization problems and were able to show their superiority to their corresponding crossovers on the genotype space. More examples and applications for each example case are left for future study.

## Acknowledgments

The authors would like to thank the referees for their valuable suggestions in improving this paper.

## References

- [1] TSPLIB. <http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/>.
- [2] D. Avis. A survey of heuristics for the weighted matching problem. *Networks*, 13:475–493, 1983.
- [3] J. E. Beasley. OR-library: Distributing test problems by electronic mail. *Journal of the Operational Research Society*, 41:1069–1072, 1990.
- [4] D. Burago, Y. Burago, S. Ivanov, and Iu. D. Burago. *A Course in Metric Geometry*. American Mathematical Society, 2001.
- [5] A. Caprara. Sorting by reversals is difficult. In *Proceedings of the First International Conference on Computational Molecular Biology*, pages 75–83, 1997.
- [6] S.-S. Choi, Y.-K. Kwon, and B.-R. Moon. Properties of symmetric fitness functions. *IEEE Transactions on Evolutionary Computation*, 11(6):743–757, 2007.
- [7] S.-S. Choi and B.-R. Moon. Normalization in genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 862–873, 2003.
- [8] S.-S. Choi and B.-R. Moon. Normalization for genetic algorithms with nonsynonymously redundant encodings. *IEEE Transactions on Evolutionary Computation*, 12(5):604–616, 2008.
- [9] R. Dorne and J. K. Hao. A new genetic local search algorithm for graph coloring. In *Proceedings of the Fifth Conference on Parallel Problem Solving from Nature*, pages 745–754, 1998.
- [10] W. Erben. A grouping genetic algorithm for graph colouring and exam timetabling. In *Practice and Theory of Automated Timetabling III*, pages 132–156, 2001.
- [11] L. J. Eshelman. The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. In *Foundations of Genetic Algorithms*, pages 265–283. Morgan Kaufmann, 1991.
- [12] E. Falkenauer. A hybrid grouping genetic algorithm for bin packin. *Journal of Heuristics*, 2(1):5–30, 1996.
- [13] E. Falkenauer. *Genetic Algorithms and Grouping Problems*. John Wiley & Sons, 1998.
- [14] J. B. Fraleigh. *A First Course in Abstract Algebra*. Addison-Wesley, 2003.
- [15] D. S. Hochbaum and A. Pathria. Analysis of the greedy approach in problems of maximum  $k$ -coverage. *Naval Research Logistics*, 45(6):615–627, 1998.

- [16] D. S. Johnson, C. Aragon, L. McGeoch, and C. Schevon. Optimization by simulated annealing: An experimental evaluation, Part 1, graph partitioning. *Operations Research*, 37:865–892, 1989.
- [17] T. Jones. *Evolutionary Algorithms, Fitness Landscapes and Search*. PhD thesis, University of New Mexico, 1995.
- [18] S.-J. Kang and B.-R. Moon. A hybrid genetic algorithm for multiway graph partitioning. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 159–166, 2000.
- [19] J. Kececioglu and D. Sankoff. Exact and approximate algorithms for sorting by reversals, with applications to genome rearrangement. *Algorithmica*, 13(1-2):180–210, 1995.
- [20] Y.-H. Kim and B.-R. Moon. New topologies for genetic search space. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1393–1399, 2005.
- [21] Y.-H. Kim, Y. Yoon, A. Moraglio, and B.-R. Moon. Geometric crossover for multiway graph partitioning. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1217–1224, 2006.
- [22] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Res. Logist. Quart.*, 2:83–97, 1955.
- [23] G. Laszewski. Intelligent structural operators for the  $k$ -way graph partitioning problem. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 45–52, 1991.
- [24] E. Lawler, J. Lenstra, A. Kan, and D. Shmoys. *The Traveling Salesman Problem*. Wiley-Interscience Publication, 1985.
- [25] K. Mehlhorn. *Graph Algorithms and NP-completeness*. Springer Verlag, New York, Inc., 1984.
- [26] A. Moraglio. *Towards a Geometric Unification of Evolutionary Algorithms*. PhD thesis, University of Essex, UK, 2007.
- [27] A. Moraglio, Y.-H. Kim, Y. Yoon, and B.-R. Moon. Geometric crossover for multiway graph partitioning. *Evolutionary Computation*, 15(4):445–474, 2007.
- [28] A. Moraglio and R. Poli. Topological interpretation of crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1377–1388, 2004.
- [29] A. Moraglio and R. Poli. Abstract geometric crossover for the permutation representation. *Technical Report CSM-429, Department of Computer Science, University of Essex*, 2005.
- [30] H. Mühlenbein. Parallel genetic algorithms in combinatorial optimization. In *Computer Science and Operations Research: New Developments in Their Interfaces*, pages 441–456, 1992.
- [31] R. Raveaux, J.-C. Buriea, and J.-M. Ogiera. A graph matching method and a graph matching distance based on subgraph assignments. *Pattern Recognition Letters*, 31(5):394–406, 2010.
- [32] F. Rothlauf. On the locality of representations. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1608–1609, 2003.
- [33] F. Rothlauf. *Representations for Genetic and Evolutionary Algorithms*. Springer Verlag, second edition, 2006.
- [34] F. Rothlauf and D. E. Goldberg. Redundant representations in evolutionary computation. *Evolutionary Computation*, 11(4):381–415, 2003.
- [35] J. E. Rowe, M. D. Vose, and A. H. Wright. Group properties of crossover and mutation. *Evolutionary Computation*, 10(2):151–184, 2002.

- [36] J. E. Rowe, M. D. Vose, and A. H. Wright. Representation invariant genetic operators. *Evolutionary Computation*, 18(4):635–660, 2010.
- [37] L. A. Sanchis. Multiple-way network partitioning. *IEEE Transactions on Computers*, 38(1):62–81, 1989.
- [38] J.D. Schaffer and L.J. Eshelman. On crossover as an evolutionary viable strategy. In *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 61–68, 1991.
- [39] J.-H. Seo, Y.-H. Kim, H.-B. Ryou, S.-H. Cha, and M. Jo. Optimal sensor deployment for wireless surveillance sensor networks by a hybrid steady-state genetic algorithm. *IEICE Transactions on Communications*, E91-B(11):3534–3543, 2008.
- [40] R. P. Stanley. *Enumerative Combinatorics*. Cambridge University Press, 1999.
- [41] J. R. Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM*, 23(1):31–42, 1976.
- [42] C. van Hoyweghen, B. Naudts, and D. E. Goldberg. Spin-flip symmetry and synchronization. *Evolutionary Computation*, 10:317–344, 2002.
- [43] M. Wall. GALib: A C++ library of genetic algorithm components. *Mechanical Engineering Department, Massachusetts Institute of Technology*, 1996.
- [44] Y. Yoon, Y.-H. Kim, and B.-R. Moon. Feasibility-preserving crossover for maximum  $k$ -coverage problem. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 593–598, 2008.