A Runtime Analysis Method unifying Evolutionary Algorithms on Discrete Search Spaces

Submitted by Tina Ambinintsoa Malalanirainy Rakotoson
to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Computer Science

December 2020

Signature:

Author declaration

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Tina A. MALALANIRAINY RAKOTOSON
December 2020

Acknowledgements

I would like to thank the University of Exeter for believing in me and for funding my PhD. I thank my supervisor Dr Alberto Moraglio for introducing me to Evolutionary Algorithms and most importantly to research. I thank Dr Dirk Sudholt, Dr Per Kristian Lehre, and Prof Benjamin Doerr for replying to my emails and helping me understand the intricacies of runtime analysis. I thank Dr Robin Chapman and Dr Samuele Giraudo for helping me revise some long forgotten properties of the permutation group. Finally, I thank Dr Khulood Alyahya and the Evolutionary Computation and Machine Learning group of the University of Exeter for their feedback on my work.

Abstract

This thesis aims to develop a unified runtime analysis of: EA[1] with no mutation and with a standard crossover, $(1+1)$ EA, and EA with both a mutation and a standard crossover. To this end, we determined for each algorithm a class of problems they efficiently solve. Polynomially quasi-concave problems on the Hamming (resp. Manhattan) space, that are already known to be easy for EA with no mutation and with a non-standard crossover, were shown to be easy for EA with no mutation and with a standard crossover. A class of problems that is determined by its balls of radius ρ, is defined for each the following algorithms: $(1+1)$ EA and EA with both a mutation and a standard crossover. Each of these classes is shown to only contain easy problems for an instantiation of a generalization of the algorithm they correspond to. Unlike the class of quasi-concave fitness landscapes, these classes are not affected by the choice of representation. We conclude that if the definition of a class of problems is built upon particular metrics, then the runtime result is affected by the choice of representation.

[1] In this thesis, 'evolutionary algorithm' and 'evolutionary algorithms' will both be written as 'EA'.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Evolutionary algorithms (EAs) [Hol75] are randomized algorithms that use a heuristic approach inspired by Darwinian evolution to solve a given problem. Namely, an initial set of solutions is evolved until an optimal solution is eventually found. Indeed, an EA is usually equipped with :

- A selection scheme to choose the solutions that will be evolved,
- One or two distinct genetic operators are used to create new solutions from the previously selected solutions,
- A replacement strategy to decide whether or not the new population contains solutions from the previous generation.

An illustration is given in Figure 1.1.

Selection

There exist various types of selection schemes in the literature.

- Ranking [Bak87] consists of selecting an individual in proportion with its fitness.

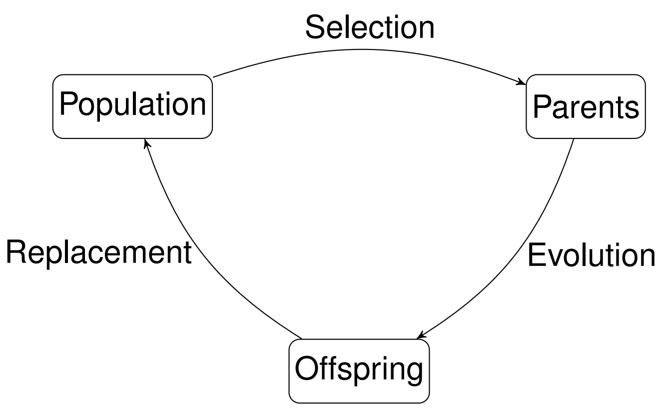


Figure 1.1: A typical EA

- Fitness proportion [Hol75] is a particular ranking where the selection probability of an individual is directly proportional to its fitness value.

- Tournament [Bac95, BT95] consists of randomly sampling a fixed number of solutions and selecting the best individual in the batch.

- Uniform selection [Sch95, BS02] consists of giving all individuals the same probability to be selected.

Genetic Operators

In order to evolve the parents, genetic operators are used. There are two different types of genetic operators: mutation and crossover (also called recombination). The mutation operator generates an offspring from a single parent, while the crossover operator generates an offspring from a population of parents. An EA uses at least one type of genetic operator and uses at most both types.

Replacement

There exist different sorts of replacement strategies that define a particular type of EA. Indeed, for a constant population size μ and a constant number λ of selected individuals from the previous population:

- A $(\lambda + \mu)$ EA corresponds to a population whose individuals are selected from the set of $\lambda+\mu$ individuals formed by the union of parents and offspring.

- A (λ, μ) EA corresponds to a population whose individuals are selected from the set of μ offspring that were previously generated.

- An elitist EA always keeps the current fittest individual in the replacement step.

1.1 Representation and Genetic Operators

Depending on the problem considered, solutions can be encoded as binary strings, strings on a finite alphabet, permutations, trees, etc ... Different representations give rise to different instances of EA and to different genetic operators. There are mainly three categories of representations: genetic algorithms, genetic programming, and evolution strategies. As this thesis is restricted to problems on discrete search spaces, we will only describe the first two representations.

	$\mathcal{A} = \{0, 1\}$	$\mathcal{A} = \mathbb{N}$
Mutation	Bit flip	Random resetting
	-	-
	-	Creep mutation [Dav91]
Crossover	1-point crossover [Hol75, DJ75]	1-point crossover
	N-point crossover	N-point crossover
	Uniform crossover [Sta02, Sys89]	Uniform crossover
	-	-

Table 1.1: Genetic operators for binary and integer representations

1.1.1 Genetic algorithm

A genetic algorithm (GA) [Hol75, Gol89] is an EA in which solutions are represented as finite strings of same length over an alphabet $\mathcal{A}$. We can distinguish three types of GA:

- binary representation for $\mathcal{A} = \{0, 1\}$ [Hol75],
- integer representation for $\mathcal{A} = \mathbb{N}$ [Dav91],
- permutation representation where solutions are strings of length n over the alphabet $\mathcal{A} = \{1, 2, \ldots, n\}$ with no repeating letters [GL$^+$85, Dav91, OSH87, BFM97].

Binary and Integer Representations

Each type of GA has its own mutation operator and its own crossover operator as summarized in Table 1.1. Operators with similar actions are placed on the same line.

Permutation Representation

The operators of a GA on permutations are nothing alike the operators of GA on common strings. Indeed, an operator must keep $\mathcal{S}$ closed (i.e., the value returned by a genetic operator must be an element of $\mathcal{S}$) [SM91]. In this regard, the mutation and crossover operators of a GA on permutations must always return a permutation. The main operators that are used for a GA on permutations are described below.

- Mutation
  - Swap and Insert mutations both work with two random letters of the parent permutation. While swap mutation swaps the two letters, insert mutation moves the right most chosen letter to the position right after the left most chosen letter.
  - Scramble and Inversion mutations both work on a random substring of the parent permutation. While scramble mutation randomly scrambles

the letters of the substring, inversion mutation reverses the order of the substring (the left most letter becomes the right most one).

- Crossover

  - Both the PMX (partially mapped crossover) [GL$^+$85] and the order crossover [Dav91] create offspring by changing a randomly chosen substring of one parent into the substring of the other parent at the same positions. The PMX uses the mappings defined by the randomly chosen substrings to determine the values that the remaining positions of the offspring should take. The order crossover fills in the blank from the right most position in an increasing order.
  - Cycle crossover [OSH87] works on the cycles defined by the two parents. The letters appearing in a randomly chosen cycle are kept and the remaining positions are filled with the other parent's letters at these positions.
  - Edge crossover [BFM97] consists of seeing the permutations as cycle graphs whose nodes are the letters (e.g., the permutation 312 is the cycle graph whose nodes are 1, 2 and 3 and whose edges are $\{3, 1\}$, $\{1, 2\}$, and $\{2, 3\}$). This gives a list of neighbours for each letter of a permutation (e.g., the neighbours of the letter 1 are 3 and 2 for the permutation 312). The lists of neighbours per letter of the two parents are merged, and one permutation is randomly created by assigning a neighbour to each letter.

1.1.2 Genetic programming

In a genetic programming (GP) [Koz92, Koz94, PLMK08], solutions are represented as syntax trees. The internal nodes of the tree are operations and the leaves of the tree are constants and variables. Trees do not necessarily have the same number of nodes. Mutation consists of substituting a randomly chosen subtree with a random one [Ang97]. Crossover consists of exchanging randomly chosen subtrees [Koz92] of their parents.

1.1.3 Representation and Fitness Landscapes

An EA is used to maximize/minimize a fitness function f, that assigns a fitness value to each element of the search space $\mathcal{S}$. A fitness landscape corresponding to the problem f is defined by endowing the search space $\mathcal{S}$ with a notion of connectedness [Sta02].

- A notion of connectedness that is induced by one of the operators of the EA of interest [J$^+$95] is obtained by defining the set of offspring that can be obtained from the application of the operator to any set of parents.

- A notion of connectedness that is solely dependent on the search space $\mathcal{S}$ [PA12] is obtained by defining a metric function D on the search space $\mathcal{S}$.

Regardless of the notion of connectedness considered, the representation used always impacts the resulting fitness landscape. The notion of connectedness that is solely dependent on the search space $\mathcal{S}$ will be used throughout this thesis.

1.2 Relevance and Importance

Representations not only affect the operators of the EA, but they also determine how the problem to be solved will be encoded. As a result, it seems impossible to avoid representations while analyzing EAs. However, in order to reach a more global understanding of EAs it is necessary to develop an analysis that *unifies* EAs across representations [DJ06].

A first unifying runtime analysis has been developed in [MS17], for EAs with no mutation and with a particular recombination operator. This was done in three steps. First, a formal EA with no mutation called Convex evolutionary Search (CS) has been defined. The generalization of the algorithm consisted of finding the geometrical object that is described by its operator across representations. Then, a class of fitness landscapes called *quasi-concave* has been defined. Finally, the runtime of the CS on a quasi-concave problem has been estimated. At this point, both the algorithm and the problem are considered from a representation free perspective. This yields a runtime result that can be viewed as representation free. This representation free runtime result mostly remains useless for practitioners until the representation used for the solutions (i.e., strings, permutations, etc ...) and the metric used to endow the corresponding search space with, are specified. In [MS17], the runtime result was only instantiated to strings of the metric spaces $M_{d,\mathrm{HD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$ and $M_{d,\mathrm{MD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$ where HD and MD respectively denote the Hamming and the Manhattan distances.

1.3 Questions and Objectives

The aim of this thesis is to extend the unifying runtime analysis of [MS17] to: EA with no mutation and with a standard crossover, $(1 + 1)$ EA, and EA with both a mutation and a standard crossover.

Definition 1. *A problem is easy for an algorithm if the expected runtime upper bound for the algorithm to find a global optimum is at most polynomial in the solution size.*

A unifying runtime analysis of a given algorithm can be used to identify a class of easy problems for this algorithm. Hence, by extending the unifying runtime analysis of [MS17], we aim to answer the following:

- Are polynomial quasi-concave problems easy for the CS, regardless of the representation used?
- Are polynomial quasi-concave problems easy for EA with no mutation and with a standard crossover, regardless of the representation used?
- What problems are easy for $(1+1)$ EA?
- What problems are easy for EA with both a mutation and a standard crossover?

The first question is answered for the CS through the instantiation of the runtime analysis of [MS17] to the usual metric spaces of permutations. However, new runtime analyses need be developed to answer the last three questions. Namely, a generalization across representations of each EA must be determined in the manner of [MS17]: the geometrical object described by the operator(s) of each algorithm across representations is determined. Then, a class of fitness landscapes that are efficiently solved by the instantiations of the generalized algorithm must be found.

1.4 Organisation of the Thesis

The summary of the runtime analysis of the CS on quasi-concave fitness landscapes on the metric spaces $M_{d,\mathrm{HD}}$ and $M_{d,\mathrm{MD}}$ done by [MS17], is given in Chapter 3. We show in Chapter 4 that quasi-concave fitness landscapes on the usual metric spaces of permutations, with at most polynomially many level sets, need not be efficiently solved by the CS.

The runtime analysis of the SES on quasi-concave landscapes is done in Chapter 5. We show in Chapter 6 that quasi-concave fitness landscapes on the metric spaces $M_{d,\mathrm{HD}}$ and $M_{d,\mathrm{MD}}$, with at most polynomially many level sets, are efficiently solved by the SES for a well chosen population size. We show in Chapter 7 that the analysis of the CS on quasi-concave fitness landscapes on the usual metric spaces of permutations does not extend to the SES. Hence, we bring the problem back to common strings by considering a bijection between permutations and common strings.

In Chapter 8, any polynomially ρ-improving fitness landscape with at most polynomially many level sets is shown to be efficiently solved by an instantiation of the generalized $(1+1)$ EA for a well chosen mutation parameter.

In Chapter 9, any ρ-improving fitness landscape with at most polynomially many level sets is shown to be efficiently solved by an instantiation of the generalized EA with a mutation and a standard crossover for a well chosen population size.

1.5 Publication

The contents of Chapters 5 and 6 form an extension of the early ideas published in the paper [MM19], for which I am a lead author. We showed in [MM19] that the SES can be seen as a particular CS, on metric spaces where the union of all segments that can be formed from any subset is always a convex set.

Chapter 2

Literature Review

In the context of EAs analysis, a theory that can be instantiated to different representations will be referred to as a unifying theory. A number of unifying theories have been developed to understand the working principles of EAs. These unifying theories use techniques from areas ranging from Biology, Physics, to Mathematics. The genetic operator(s) of EAs are defined from the mutation and the crossover operators found in Biology, regardless of the representation considered. The modelization of the evolving population can be studied through dynamical systems and statistical mechanics, for different representations. Representation free EAs can be formalized using Mathematical tools such as equivalence relations or geometry. We will look at different existing unifying theories, with a particular interest in the unifying theories on the runtime analysis of EAs.

2.1 Unifying Theories on the Analysis of EA

Genetic algorithms [Hol75], evolution strategies [Sch81], and genetic programming [Koz92] are different types of EA. While it is clear that these EA all use a mutation operator and/or a recombination operator, those operators often differ between representations. In [SR96], a formal EA that can be instantiated to any representation is defined using the findings of [Rad91]. The convex evolutionary search algorithm (CS) of [Mor11] is another formal EA with no mutation that unifies the algorithms across discrete search spaces. Unlike the formal EA of [SR96], the CS of [Mor11] is defined through geometry. The formalization of EAs across representations is one of many unifying theories on the analysis of EAs.

2.1.1 Schema Theory

Holland's schema theorem [Hol75] is the main result on the increase in the number of strictly improving solutions in one generation. It says that we can find some template (called schema) corresponding to a subset of fit solutions in each generation, that increases in size in the next generation. Holland's schema theorem can only be applied to problems where schemata can be defined and only holds for infinitely large populations. The schema theorem has been criticized in [Alt95] for only taking into account the case where a given schema is lost because of the disruptive effect of the genetic operators. Most importantly, the weakness of the schema theorem is due to the limitation of its scope to the one step variation in the number of individuals with a given schema [Vos99].

Radcliffe [Rad91] extended Holland's schema theorem to general non-string representations using equivalence relations. In [LP13], Holland's schema theorem has been extended to Genetic Programming (GP).

2.1.2 Convergence of the Population Sequence

The convergence of the population sequence is defined differently by different authors. In [Rud94a], the population sequence converges when a population whose best individual is a global optimum is obtained. In [Rud94b], populations of size one are studied. The population sequence converges when its corresponding sequence of fitness values converges. In [Mor11], the population sequence converges when a population whose individuals are all clones of each other is obtained. Notice that the (initial) population sequence may converge before a global optimum is found, in the last two cases. This is referred to as a premature convergence. Also, only the definition of convergence given in [Mor11] does not require the use of a fitness function. As a fitness function can only be defined when solutions are represented in some ways, the only definition of convergence that is representation free is the one given in [Mor11]. We will use this definition of convergence throughout the thesis.

2.1.3 Dynamics of the Adaptation Process

The main results that describe how an EA generates solutions that are fitter than their parents can be divided into three distinct categories: building block analysis, modelization of the evolving population, and fitness landscape theory.

Building Block Analysis

According to Godlberg [Gol89], there exist schemata whose elements always generate offspring that are fitter than their parents. In particular, Goldberg's "building blocks" are short, low order, and highly fit schemata. By using an exact evolution equation, [SW99] determined the "building blocks" of a GA and showed that they need not be as Goldberg's. Initial steps towards the theoretical analysis of building blocks for GP have been taken in [BAW$^+$15].

Modelization of the Evolving Population

The evolution of a population has been modelized as a dynamical system for GA [VL91, NV92] and GP [Koz92]. The initial population corresponds to the initial state. The total number of states is given by the number of generations. The population (or state) corresponding to the t-th generation determines the population (or state) corresponding to the $(t+1)$-th generation. A unifying framework that links the dynamics of the population of a GA to the dynamics of the population of a GP has been given in [SZ03].

In [J$^+$95], a directed graph (called *landscape graph*) has been used to model the effect of each operator of the EA of interest on the search space. A vertex of this direct graph corresponds to a multiset whose elements are taken from the search space. If the application of the operator to a multiset generates a multiset, then an edge must be drawn from the parent multiset to the offspring multiset. The effect of the EA on the search space is therefore modelized by a larger directed graph containing the landscape graphs of each operator of the EA. This model can be applied to any representation. Another graph modelization of the dynamics of EAs is given in [ZTV$^+$18] as a complex system.

In [PB97], statistical mechanics have been used to see all possible populations as points whose union makes up the *phase space*. Then, the evolution of a population can be seen as a *trajectory* in this *phase space*. This model is not limited to GA.

Fitness Landscape Theory

Fitness landscape theory can be used to determine how the choice of landscape affects the evolution of the population [RS02].

In genetics, epistasis is the interaction between genes [Cor02]. This notion has been adapted to the study of EA to measure the difficulty of a fitness landscape. Davidor [Dav90] was among the first to notice the potential of epistasis in measuring fitness landscape difficuly for GA. In [HW99], the use of epistasis is applied to fitness functions with a mathematical expression. It has been shown that the difficulty at which a GA solves a fitness landscape can be predicted in this case. Namely, a low order fitness function corresponds to an easy problem for GA.

A different approach [JF$^+$95] where a GA is identified to a heuristic search, showed that fitness landscapes with a high Fitness Distance Correlation (FDC) are often easily solved by GA.

The topography of a fitness landscape can also be analyzed to estimate the difficulty of searching it [MF00]. The ruggedness of a fitness landscape is defined differently by different authors. Ruggedness can correspond to the number of local optima [Pal93]. The trend of the fitness values of points along a random walk is another measure of the ruggedness of a fitness landscape: a good correlation implies a smoother landscape in the investigated area [Jon95, Wei90]. The average length of a path that is strictly going downhill (resp. upnhill) within a basin can also be used as a ruggedness measure [FSBB$^+$93]. We recall that a basin of a local minimum (resp. maximum) [Sta02], is the union of the solutions forming a path that is strictly going downhill (resp. upnhill) to the local minimum (resp. maximum). Roughly speaking, the difficulty of a fitness landscape for an EA is "proportional" to the ruggedness of the fitness landscape. This however need not be true. In [QRSS98], ridge functions have be shown to be easy for GA despite corresponding to a rugged fitness landscape (as classified as *highly misleading* by the FDC). In [MBK99], Job Shop Scheduling Problems have be shown to be difficult for GA despite corresponding to smooth fitness landscapes for harder instances.

Other studies focused on comparing the fitness landscapes induced by different operators. The fitness landscape induced by recombination and the fitness landscape induced by mutation have been shown to be homomorphic for GA (resp. GP) in [GW96]. A similar result was obtained in [SW97] through a different approach.

Artificial parametrized class of fitness landscapes can be defined for the sole purpose of being tested with EA. This is for instance the case of Nk-landscapes [DJFS97] and quasi-concave landscapes [Mor11].

Finally, dynamic fitness landscape analysis has been defined in [Ric13].

2.1.4 Runtime

In runtime analysis, we are interested in estimating the number of fitness evaluations needed by an EA to find the first optimal solution in a fitness landscape. In fact, runtime analysis can be seen as a mathematical proof for the performance of an EA on a fitness landscape. In this regard, research is focused on improving runtime analysis methods to obtain the most accurate possible result. However, these methods are often applied separately for different representations. To our knowledge, the first unifying runtime analysis method has been given in [MS17]. This was made possible by the development of a method for the runtime analysis of a formal EA (i.e., a generalization of EA across representation) on a class of fitness landscapes (i.e., a model of landscapes that can suit problems from different representations). Nonetheless, this unifying runtime analysis has only been applied to a particular formal EA with no mutation that does not perform standard crossover. We aim to extend this unifying runtime analysis to EA with both a mutation and a standard crossover.

2.2 Methods for Runtime Analysis

An overview on the different methods of runtime analysis can be found in [Weg03]. Some methods were specifically designed for a particular EA, a particular problem, and a particular representation. This is for instance the case of the Coupon Collector and the Gambler's ruin methods, which are well-suited to the study of $(1+1)$ EA on binary strings. Other methods are restricted to a single representation, such as the family tree technique [Wit06] that can only be used on pseudo-Boolean problems. There exist also broader techniques that are neither limited to specific EA nor limited to a particular representation.

2.2.1 Artificial fitness levels

The artificial fitness levels method [Weg03] is a general approach that uses a partition of the search space into fitness levels. It has been initially used to analyze the $(1+1)$ EA on various pseudo-Boolean problems in [Weg03]. Nonetheless, it is a powerful tool that can be used to analyze any EA on any representation. Indeed, the artificial fitness levels method has been used in [MS17] to analyze the runtime of a formal EA with no mutation on a class of fitness landscapes. The artificial fitness levels method has been improved into level-based analysis in [CDEL17].

The artificial fitness levels method is used for the runtime analysis of a $(1+1)$ EA on a fitness function $f : \mathcal{S} \longrightarrow \mathbb{R}$, that can take $q+1$ distinct values, in the following manner. Let $a_0 < a_1 < \cdots < a_q$ be the different values that the fitness function can take. A fitness level A_i is the subset of the search space $\mathcal{S}$ containing all individuals whose fitness value is equal to a_i. The search space $\mathcal{S}$ is therefore partitioned into fitness levels:

$$\begin{cases} \mathcal{S} = A_0 \cup A_1 \cup \cdots \cup A_q, \\ A_i \cap A_j = \varnothing \text{ if } i \neq j. \end{cases}$$

The probability that a randomly chosen point belongs to A_i is denoted $P(A_i)$. An element of A_i is said to *leave* A_i when its offspring belongs to $A_{i+1} \cup A_{i+2} \cup \cdots \cup A_q$. An illustration can be found in Figure 2.1. The probability that an element a of A_i *leaves* A_i is denoted $s(a)$. Let X_t be the random variable corresponding to the population at the t-th generation, and let $\tau = \min\{t | X_t \in A_q\}$ be the first hitting time of A_q. We have:

$$\begin{cases} \sum_{1 \le i \le m-1} \dfrac{P(A_i)}{\max\{s(a) \mid a \in A_i\}} \le E[\tau], \\ E[\tau] \le \sum_{1 \le i \le m-1} P(A_i) \left(\dfrac{1}{\min\{s(a) \mid a \in A_i\}} + \cdots + \dfrac{1}{\min\{s(a) \mid a \in A_{q-1}\}} \right). \end{cases}$$

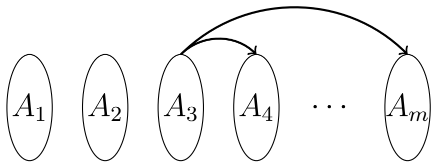


Figure 2.1: Leaving fitness level A_3: the arrows show fitness levels where the offspring might end up.

A Unifying Runtime Analysis of a particular EA with no mutation

To develop a unifying runtime analysis of a formal EA with no mutation, an artificial fitness levels method using the geometric properties of the recombination operator of the formal EA has been introduced in [MS17]. The recombination operator samples an offspring on the convex hull of the parents and the search performed by the formal EA with no mutation in the Euclidean space is illustrated in Figure 2.2. Using the same notations as in Section 2.2.1, the probability for *conquering* $A_{i+1} \cup A_{i+2} \cup \ldots \cup A_q$ from $A_i \cup A_{i+1} \cup \ldots \cup A_q$ is the probability that:

- the parents are contained in $A_i \cup A_{i+1} \cup \ldots \cup A_q$,
- the convex hull of the parents intersects $A_{i+1} \cup A_{i+2} \cup \ldots \cup A_q$.

In the worst case, the convex hull of the parents contains the union $A_i \cup A_{i+1} \cup \ldots \cup A_q$. The probability of *conquering* $A_{i+1} \cup A_{i+2} \cup \cdots \cup A_q$ from $A_i \cup A_{i+1} \cup \cdots \cup A_q$ replaces the probability of *leaving* A_i.

This unifying runtime analysis method need not hold for formal EA other than the one it has been designed for. Hence, similar methods need be developed for: formal EA with no mutation and with a standard crossover, formal EA with no crossover, and formal EA with a mutation and a standard crossover.

2.2.2 Potential functions

The potential function method [Weg03] is an extension of the artificial fitness levels method. It is used when computing the probability of leaving a fitness level is too costly. We first work with an easier function (which is the potential function), then we take into account the difficult fitness function. The potential function is used to measure the algorithm's progress, while the fitness function is used to decide whether an offspring is accepted or not.

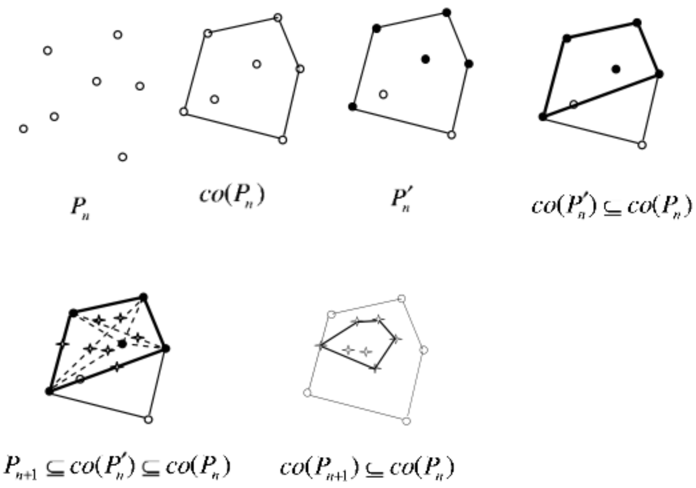


Figure 2.2: Convex search for an Euclidean distance [Mor11]

2.2.3 Drift Analysis

The drift analysis method [HY01] is a particular case of the potential function method. It uses a particular distance function d, to measure how far from the global optimum a population is. Let X_t denote the random variable corresponding to the population at the t-th generation, and let $d(X_t)$ be the random variable corresponding to the distance of the population at the t-th generation to the global optimum. The drift analysis method consists of bounding the one-step mean drift

$$E[d(X_t) - d(X_{t+1})|d(X_t) > 0] \tag{2.1}$$

from below, in order to get an upper bound of the expectation $E[\tau|X_0]$ of the running time $\tau = \min\{t|X_t \in \mathcal{E}_o\}$ of the EA, where $\mathcal{E}_o$ is the set of populations containing the global optimum.

Depending on the form of the positive lower bound of the one-step mean drift, different types of drift analysis are defined:

- For an additive drift [HY01], the lower bound does not depend on $d(X_t)$.
- For a variable drift [MRC09, Joh11], the lower bound is a function of $d(X_t)$.
- A multiplicative drift [DJW10, FOV08] is a variable drift where the lower bound is a multiple of $d(X_t)$.

A negative drift is also introduced in [Leh10].

Chapter 3

Background

The convex search algorithm is a generalization across representations of EAs with no mutation that has been introduced in [Mor11]. This generalization is based on the notion of geodesic convexity, that extends the traditional notion of convexity of Euclidean spaces to combinatorial spaces. A fitness landscape where offspring are at least as fit as their worst parent and whose canonical level sets are geodesically convex is quasi-concave [MS17]. The class of quasi-concave landscapes generalizes traditional quasi-concave functions on continuous domains to combinatorial spaces. We summarize the runtime analysis of the convex search algorithm on a quasi-concave landscape done in [MS17]. We first recall the notions of segments and geodesically convex sets in Section 3.1. Then, we recall the convex search algorithm in Section 3.3. After this, we recall the definition of a quasi-concave landscape in Section 3.4. We do this by explaining the notion of canonical level sets. Finally, we give a summary of the runtime analysis of [MS17] in Section 3.5.

3.1 Segments and Convex sets

Let S be a search space endowed with a metric D. We recall that a metric function D is a mapping from $\mathcal{S} \times \mathcal{S} \longrightarrow \mathbb{R}_+$ that satisfies for any $x,\ y$ and z in $\mathcal{S}$:

1. $D(x, y) = D(y, x)$,
2. $D(x, z) \leq D(x, y) + D(y, z)$,
3. $D(x, y) = 0$ if and only if $x = y$.

We start by recalling the notions of segments and convex sets in a discrete metric space $(\mathcal{S}, D)$. The discrete metric space $(\mathcal{S}, D)$ can be seen as a graph. The elements of $(\mathcal{S}, D)$ are the nodes of the graph and the distance between any two nodes is the length of the shortest paths between them. This length is the number of edges in the path.

Definition 2 (Segment). *Let $(\mathcal{S}, D)$ be a metric space, and let x and y be elements of $\mathcal{S}$. The segment between x and y is the union of the shortest paths between x and y. That is, $[x, y]_D = \{z \in \mathcal{S} \mid D(x, z) + D(z, y) = D(x, y)\}$. The points x and y are extremes of the segment $[x, y]_D$.*

Example 1. *In the two-dimensional Hamming space $(\{0, 1\}^2, \mathrm{HD})$, the segment $[00, 11]$ is the union of the shortest paths between 00 and 11. The shortest paths between 00 and 11 are: $\{00, 01, 11\}$ and $\{00, 10, 11\}$. Hence, $[00, 11] = \{00, 01, 10, 11\}$. Consequently, the same segment can have more than a pair of extremes, unlike the case of the Euclidean space. For instance, we have that $[00, 11] = [01, 10]$.*

We shall now recall the notion of convexity in a discrete metric space.

Definition 3 (Geodesic Convexity [vDV93]). *Let $(\mathcal{S}, D)$ be a metric space. A subset C of $\mathcal{S}$ is geodesically convex if all shortest paths between any two points of C are included in C. That is, $[x, y]_D \subseteq C$ for all $x,\ y$ in C.*

Example 2. *Let $n \geq 2$, the set $\{0, 1\}^n$ is geodesically convex for the Hamming (resp. Manhattan) distance. All singletons and segments of length one are geodesically convex for the Hamming (resp. Manhattan) distance.*

We will use the term *convex* set for *geodesically convex* set in the rest of the thesis. Let A be a subset of the metric space $(\mathcal{S}, D)$. We finally recall the notion of convex hull of a subset A, which is central to the analysis of the Convex Search algorithm.

Definition 4 (Convex hull [vDV93]). *Let $(\mathcal{S}, D)$ be a metric space. The convex hull of a subset A of $\mathcal{S}$ is the smallest convex set containing A. In particular, it is the intersection of all convex sets containing A. The convex hull of A is denoted $co(A)$.*

Example 3. *Let HD denote the Hamming distance. In the metric space $(\{0, 1\}^2, \mathrm{HD})$, the convex hull of the set $\{01\}$ is $co(\{01\}) = \{01\}$. The convex hull of the set $\{00, 10\}$ is itself and is equal to the segment $[00, 10]$. The convex hull of the set $\{01, 10\}$ is $co(\{01, 10\}) = \{0, 1\}^2$.*

3.2 Schemata for Strings

We start by recalling the notion of schemata [H+92] for strings on a finite alphabet.

Definition 5. *A schema in the set $\{0, 1, \ldots, d-1\}^n$ is a template with n positions where a position is either:*

- *Free to take any value in the set $\{0, 1, \ldots, d-1\}$,*

- *Restricted to take values in a non-empty strict subset of* $\{0, 1, \ldots, d-1\}$.

A free position is denoted $*$, *whereas a restricted position is denoted* $*_A$ *where* A *is the set of admissible values.*

Example 4. *All the elements of the set* $\{0,1,2\}^5$ *match the schema* $* * * * *$. *The smallest schema matching the elements* 00123 *and* 21103 *is* $*_{\{0,2\}} *_{\{0,1\}} 1 *_{\{0,2\}} 3 = *_{02} *_{01} 1 *_{02} 3$.

A schema can be seen as subset of the search space, whose elements are those matching it.

3.2.1 Hamming distance

Recall that the Hamming distance between x and y is the number of differing positions between them:

$$\mathrm{HD}(x,y) = \sum_{k=1}^{n} [1 - \delta_{x(k),y(k)}], \tag{3.1}$$

where $\delta_{i,j}$ is the Kronëcker delta. That is, $\delta_{i,j} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{otherwise.} \end{cases}$

Proposition 1. *In the metric space* $(\{0,1,\ldots,d-1\}^n, \mathrm{HD})$, *a segment* $[x,y]$ *is a schema such that the admissible values at position* i *are* $x(i)$ *and* $y(i)$.

Proof. Let S be the schema whose admissible values at position i are $x(i)$ and $y(i)$. We are going to show that $S = [x,y]$.

- '$[x,y] \subseteq S$': Let $z \in [x,y]$, we have:

  $$\mathrm{HD}(x,z) + \mathrm{HD}(z,y) = \mathrm{HD}(x,y). \tag{3.2}$$

  That is, z differs from x in exactly $\mathrm{HD}(x,z)$ positions and takes the values of y in these positions. Therefore, z belongs to the schema S.

- '$S \subseteq [x,y]$': Let $z \in S$. For each position i, the value of $z(i)$ is either $x(i)$ or $y(i)$. This means that there exists $0 \leq k \leq n$ such that z differs from x in exactly k positions and takes the values of y in these positions. As a result, $z \in [x,y]$.

$\square$

Proposition 2. *In the metric space* $(\{0,1,\ldots,d-1\}^n, \mathrm{HD})$, *the convex hull of a finite set of points* $\{x_1, x_2, \ldots, x_m\}$ *is a schema such that the set of admissible values at position* i *is* $\{x_1(i), x_2(i), \ldots, x_m(i)\}$.

Proof. Let $A = \{x_1, x_2, \ldots, x_m\}$ and let S denote the schema whose set of admissible values at position i is $\{x_1(i), x_2(i), \ldots, x_m(i)\}$. We are going to show that $S = co(A)$. To this end, we will show that S is a the smallest convex set containing A.

Any element x of A belongs to S because the value $x(i)$ of x at position i always belongs to the set of admissible values at position i of S. Let x and y be two elements of S. We show that $[x, y]$ is contained in S. Let $z \in [x, y]$, the value of $z(i)$ is either $x(i)$ or $y(i)$. As $x(i)$ and $y(i)$ belong to the set of admissible values at position i of S, so does $z(i)$. Therefore, $z \in S$. As a result, $[x, y]$ is contained in S. We conclude that S is a convex set containing A.

It remains to show that S is the smallest convex set containing A. To this end, we show that by removing a single element of S the set A is no longer contained in S.

Let $z \in S$, the removal of z yields a new schema S' whose admissible values at position i are $\{x_1(i), x_2(i), \ldots, x_m(i)\}\backslash\{z(i)\}$. For each position i, $z(i)$ is one of the admissible values at positions i of S. That is, there exists an index $m_i \in \{1, 2, \ldots, m\}$ such that:

$$z(i) = x_{m_i}(i). \tag{3.3}$$

The removal of $x_{m_i}(i)$ at position i breaks the belonging of the element x_{m_i} in the set A. Hence, the set A is not contained in the new schema S'. Therefore, S is the smallest convex set containing A. □

3.2.2 Manhattan distance

Recall that the Manhattan distance between x and y is

$$\mathrm{MD}(x, y) = \sum_{k=1}^{n} \mid x(k) - y(k) \mid . \tag{3.4}$$

Proposition 3. *In the metric space* $(\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$*, a segment* $[x, y]$ *is a schema such that the admissible values at position* i *are the* $k \in \{0, 1, \ldots, d-1\}$ *such that:*

$$\min[x(i), y(i)] \leq k \leq \max[x(i), y(i)]. \tag{3.5}$$

Proof. Let S be the schema whose admissible values at position i are as in Equation (3.5). We are going to show that $S = [x, y]$.

- '$[x, y] \subseteq S$': Let $z \in [x, y]$, we have:

$$\mathrm{MD}(x, z) + \mathrm{MD}(z, y) = \mathrm{MD}(x, y). \tag{3.6}$$

That is,

$$\sum_{i=1}^{n} | x(i) - z(i) | + \sum_{i=1}^{n} | z(i) - y(i) | = \sum_{i=1}^{n} | x(i) - y(i) |,$$
$$\sum_{i=1}^{n} \{| x(i) - z(i) | + | z(i) - y(i) | - | x(i) - y(i) |\} = 0.$$

Therefore, at each position i we have:

$$| x(i) - z(i) | + | z(i) - y(i) | - | x(i) - y(i) | = 0,$$
$$| x(i) - z(i) | + | z(i) - y(i) | = | x(i) - y(i) | .$$

As all the values are positive, this implies that:

$$\min[x(i), y(i)] \leq z(i) \leq \max[x(i), y(i)], \tag{3.7}$$

for each position i. Hence, $z \in S$.

- '$S \subseteq [x, y]$': Let $z \in S$. For each position i, the value of $z(i)$ is included between $x(i)$ and $y(i)$. This means that for each position i we have:

  $$| x(i) - z(i) | + | z(i) - y(i) | = | x(i) - y(i) | . \tag{3.8}$$

  As a result,

  $$\sum_{i=1}^{n} | x(i) - z(i) | + \sum_{i=1}^{n} | z(i) - y(i) | = \sum_{i=1}^{n} | x(i) - y(i) |,$$
  $$\mathrm{MD}(x, z) + \mathrm{MD}(z, y) = \mathrm{MD}(x, y).$$

  Thus, $z \in [x, y]$.

□

Proposition 4. *In the metric space* $(\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$*, the convex hull of a finite set of points* $\{x_1, x_2, \ldots, x_m\}$ *is a schema such that the set of admissible values at position* i *is the set of consecutive values between the smallest and the largest elements of* $\{x_1(i), x_2(i), \ldots, x_m(i)\}$.

Proof. Let $A = \{x_1, x_2, \ldots, x_m\}$ and let S denote the schema whose set of admissible values at position i is the set of consecutive values between the smallest and the largest elements of $\{x_1(i), x_2(i), \ldots, x_m(i)\}$. We are going to show that $S = co(A)$. To this end, we will show that S is a the smallest convex set containing A.

Any element x of A belongs to S because the value $x(i)$ of x at position i always belongs to the set of admissible values at position i of S. Let x and y

be two elements of S. We show that $[x, y]$ is contained in S. Let $z \in [x, y]$, the value of $z(i)$ is between $\min\{x(i), y(i)\}$ or $\max\{x(i), y(i)\}$ (extremes included). As $\min_{x,y \in A}\{x(i), y(i)\} \leq \min\{x(i), y(i)\}$ and $\max_{x,y \in A}\{x(i), y(i)\} \leq \max_{x,y \in A}\{x(i), y(i)\}$, $z(i)$ belongs to $[\min_{x,y \in A}\{x(i), y(i)\}, \max_{x,y \in A}\{x(i), y(i)\}]$. Therefore, $z \in S$. As a result, $[x, y]$ is contained in S. We conclude that S is a convex set containing A.

It remains to show that S is the smallest convex set containing A. To this end, we show that by removing a single element of S either breaks its convexity of the inclusion of A.

Let $z \in S$, the removal of z yields a new schema S' whose admissible values at position i are $[\min_{x,y \in A}\{x(i), y(i)\}, \max_{x,y \in A}\{x(i), y(i)\}] \setminus \{z(i)\}$. For each position i, $z(i)$ is one of the admissible values at positions i of S. That is, $z(i)$ is an element of $[\min_{x,y \in A}\{x(i), y(i)\}, \max_{x,y \in A}\{x(i), y(i)\}]$.

- If $z \in A$, then its removal breaks its belonging to A. Thus, S' does not contain A.

- If $z \notin A$, then $\min_{x,y \in A}\{x(i), y(i)\} < z(i) < \max_{x,y \in A}\{x(i), y(i)\}$. Its removal breaks the convexity of the set of admissible values at position i. Thus, S' is not a convex set.

We conclude that S is the smallest convex set containing A. $\square$

3.3 Convex Search Algorithm (CS)

EAs with no mutation and with an unusual population-based crossover are convex search algorithms. The convex search algorithm [MS17] is presented in this section. We start by defining the search operator used by the convex search algorithm.

Definition 6 (Convex hull recombination [MS17])**.** *The (uniform) convex hull recombination returns an offspring sampled uniformly at random from the convex hull formed by its parents.*

A pseudo-code corresponding to the Convex Search Algorithm [MS17] is shown in Algorithm 1.

Algorithm 1 Convex Search Algorithm

Input: μ, population size
Output: individual in the last population
Initialise population uniformly at random
while population has not converged to the same individual **do**
    Rank individuals on fitness
    if there are at least two fitness values in the current population **then**
        remove all individuals with the worst fitness
    end if
    Create new population:
    for counter in $\{1, 2, \ldots, \mu\}$ **do**
        Apply the CONVEX HULL RECOMBINATION to the remaining individuals in the current population to create an individual
    end for
end while
Return any individual in the last population

Let us denote P' the set of parents that are selected from a population P. The set of reachable solutions $R(P')$ from the set of parents P' is the set of solutions that can be reached by repeated application of a search operator to the set of parents P'. In particular, the set $R_{\mathrm{CS}}(P')$ of reachabale solutions for the convex hull recombination is:

$$R_{\mathrm{CS}}(P') = co(P'). \tag{3.9}$$

The offspring distribution is uniform on $R_{\mathrm{CS}}(P')$ for the Convex Search Algorithm. The probability for sampling an offspring in $R_{\mathrm{CS}}(P')$ is:

$$\frac{1}{\mid R_{\mathrm{CS}}(P') \mid} = \frac{1}{\mid co(P') \mid}. \tag{3.10}$$

3.4 Quasi-concave fitness landscapes

Quasi-concave landscapes [MS17] are a generalisation across representations of quasi-concave functions on continuous domain to combinatorial spaces.

Definition 7 (Canonical fitness level set [MS17])**.** *Let $\mathcal{S}$ denote the search space, and let f be a fitness function on $\mathcal{S}$. The codomain of the fitness function f is finite with values $a_0 < a_1 < \cdots < a_q$. The canonical level set $A_{\geq j}$ is defined for $0 \leq j \leq q$ as $\{x \in \mathcal{S} | f(x) \geq a_j\}$.*

This definition is different from Wegener's [Weg03], as Wegener's level set corresponds to $A_j = A_{\geq j} \backslash A_{\geq j+1} = \{x \in \mathcal{S} | f(x) = a_j\}$.

Example 5. *Let '$*$' denote the don't care symbol. In the search space* $\{0,1\}^4$*, the canonical level sets of* $\mathrm{LeadingOnes}$ *are:* $A_0 = \{0,1\}^4,\ A_1 = 1***,\ A_2 = 11**,\ A_3 = 111*,$ *and* $A_4 = \{1111\}$.

Definition 8 (Quasi-concave Landscape [MS17])**.** *A problem belongs to the class of quasi-concave problems iff:*

1. *All its canonical level sets are convex sets,*

2. *For a maximizing quasi-concave problem, we have for all sets* $C \subseteq \mathcal{S}$ *that* $f(z) \geq \min_{x\in C} f(x)$ *if* $z \in co(C)$.

Example 6. *Using the same example as above, we can see that each canonical level set of* $\mathrm{LeadingOnes}$ *satisfies:*

$$\begin{aligned} A_i &= \underbrace{1\cdots 1}_{i \text{ times}}\ \underbrace{*\cdots *}_{4-i \text{ times}}, \\ &\simeq \{0,1\}^{4-i}. \end{aligned}$$

That is, A_i *is isomorphic to the set* $\{0,1\}^{4-i}$*. As the set* $\{0,1\}^{4-i}$ *is convex for the Hamming (resp. Manhattan) distance,* $\mathrm{LeadingOnes}$ *belongs to the class of quasi-concave problems.*

The notion of convexity requires a metric D on the search space $\mathcal{S}$. Therefore, the resulting triplet $(\mathcal{S}, f, D)$ forms a fitness landscape [RS02]. A quasi-concave fitness landscape is determined by:

- The total number $q+1$ of distinct canonical level sets,
- The minimum ratio r between the size of two consecutive canonical level sets:
$$r = \min_{0\leq j\leq q} \frac{\mid A_{\geq j+1} \mid}{\mid A_{\geq j} \mid}. \tag{3.11}$$

Example 7. *Let* $n \geq 2$*, in the metric space* $(\{0,1\}^n, \mathrm{HD})$ *(resp.* $(\{0,1\}^n, \mathrm{MD})$*) the quasi-concave problem* $\mathrm{LeadingOnes}$ *is determined by:*

- $q+1 = n+1$,
- $r = 0.5$.

3.5 Runtime Analysis Method

In a quasi-concave landscape, two fit parents always produce a fit offspring. Namely, the offspring is at least as fit as the worst parent [MS17]. An upper bound on the runtime of the convex search algorithm on a quasi-concave landscape has been given in [MS17]. As both the convex search algorithm and the quasi-concave landscape are representation free, the resulting runtime analysis is also representation free. Consequently, the general runtime result needs to be instantiated to a specific metric space for a particular representation.

3.5.1 Summary of the Analysis

The runtime analysis used in [MS17] is based on the fitness levels method [Weg01]. The search space $\mathcal{S}$ is partitioned with respect to the fitness values of its elements. However, a fitness level is defined to contain all individuals with a fitness larger than or equal to some existing fitness value. This gives rise to a chain of level sets ordered by inclusion. The smallest level set in this chain corresponds to the set of global optima. Given m individuals sampled uniformly at random from a level set $A_{\geq j}$, the probability $P^{\text{Cov}}_{A_{\geq j}}(m)$ that the convex hull of these m elements covers $A_{\geq j}$ is also estimated. Using the same notations as in [MS17], we have:

$$P^{\text{Cov}}_{A_{\geq j}}(m) = \Pr[co(P') = A_{\geq j} \mid P' = \text{Unif}_m(A_{\geq j})], \tag{3.12}$$

$$\geq \Pr[co(P') = \mathcal{S} \mid P' = \text{Unif}_m(\mathcal{S})], \tag{3.13}$$

$$= P^{\text{Cov}}_{\mathcal{S}}(m). \tag{3.14}$$

A lower bound on the expected number m of strictly improving offspring in each level set is computed, by finding a lower bound on the probability to hit a higher level set. [MS17] showed that the probability to hit a higher level set is at least $\frac{1}{r}$. They also showed that for a population size of μ, the expected number of strictly improving offspring in each level set is at least μr. As the probability that the expected number of strictly improving offspring in each level set is less than $\frac{\mu r}{4}$ is exponentially small, [MS17] defined the worst-case typical behaviour to have exactly $\frac{\mu r}{4}$ strictly improving offspring in each level set. The worst case probability for finding a global optimum is therefore given by:

Theorem 1. *[MS17] The Convex Search Algorithm with population size μ finds a global optimum within q generations and μq fitness evaluations with probability at least*

$$\left[P^{\text{Cov}}_{\mathcal{S}}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right). \tag{3.15}$$

3.6 CS on Strings on a Finite Alphabet

The general runtime analysis of the convex search algorithm on a quasi-concave landscape has been thoroughly isntantiated to strings on a finite alphabets for the Hamming and the Manhattan distances in [MS17]. We recall their results in this section. We consider d-ary strings of length n on the alphabet $\{0, 1, 2, \ldots, d-1\}$ and two different metrics on the set $\{0, 1, 2, \ldots, d-1\}^n$:

- The Hamming distance HD,
- The Manhattan distance MD.

3.6.1 Hamming distance

The probability $P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}(m)$ is the probability that the schema matching all the m elements of P' with respect to the Hamming distance is $\underbrace{***\cdots*}_{n \text{ times}}$.

Lemma 1. *[MS17] We assume that $d \geq 2$, for any convex set C of the metric space $M_{d,\mathrm{HD}}$ we have $P^{\mathrm{Cov}}_C(m) \geq P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}(m)$ where,*

$$P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}(m) \geq 1 - dn\left(1 - \frac{1}{d}\right)^m. \tag{3.16}$$

A lower bound on the population size for which the success probability is at least 0.5 has been estimated in [MS17] using Theorem 1. The formula shown below is adapted from the formula of Theorem 11 and the formula of Corollary 12 of [MS17], where $q + 2$ should read $2q + 1$.

Theorem 2. *[MS17] Let $d \geq 2$, if the population size μ is at least:*

$$\frac{4d}{r}\ln[2dn(2q+1)], \tag{3.17}$$

then the convex search algorithm finds a global optimum on a quasi-concave landscape on the metric space $M_{d,\mathrm{HD}}$ with probability at least 0.5 within μq fitness evaluations.

3.6.2 Manhattan distance

The probability $P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}(m)$ is the probability that the schema matching all the m elements of P' with respect to the Manhattan distance is $\underbrace{***\cdots*}_{n \text{ times}}$.

Lemma 2. *[MS17] We assume that $d \geq 2$, for any convex set C of the metric space $M_{d,\mathrm{MD}}$ we have $P^{\mathrm{Cov}}_C(m) \geq P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}(m)$ where,*

$$P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}(m) \geq 1 - 2n\left(1 - \frac{1}{d}\right)^m. \tag{3.18}$$

A lower bound on the population size for which the success probability is at least 0.5 has been estimated in [MS17] using Theorem 1. The formula shown below is adapted from the formula of Theorem 14 and the formula of Corollary 15 of [MS17], where $q + 2$ should read $2q + 1$.

Theorem 3. *[MS17] Let $d \geq 2$, if the population size μ is at least:*

$$\frac{4d}{r} \ln[4n(2q + 1)], \tag{3.19}$$

then the convex search algorithm finds a global optimum on a quasi-concave landscape on the metric space $M_{d,\mathrm{MD}}$ with probability at least 0.5 within μq fitness evaluations.

3.7 Summary

The convex search algorithm is a generalization across representations of EAs with no mutation. The crossover operator of the convex search algorithm is population-based. A level set is defined to contain all individuals with a fitness greater than or equal to some existing fitness value. This gives rise to a chain of level sets ordered by inclusion. The smallest level set in this chain is the set of global optima, whereas the largest level set is the entire search space. We consider fitness landscapes where offspring are at least as fit as their worst parent and whose level sets are all convex. They make up the class of quasi-concave landscapes. A general analysis of the runtime of the convex search algorithm on a quasi-concave landscape has been developed in [MS17]. A lower bound on the population size for which the success probability is at least 0.5 has been computed. The corresponding runtime upper bound has also been estimated. The representation free analysis has been specified to strings on a finite alphabet for two distinct metrics: the Hamming and the Manhattan distances. The explicit computation of the specific parameters for these spaces was possible by noticing that the search space $\{0, 1, 2, \ldots, d-1\}^n$ is equal to the schema $\underbrace{* * * \cdots *}_{n \text{ times}}$ for both metrics.

Chapter 4

CS on Permutations

We specify the runtime analysis of the CS on a quasi-concave landscape of [MS17] to permutations. We are aiming to determine whether the runtime result obtained in [MS17] for strings on a finite alphabets (wrt the Hamming and the Manhattan distances) holds for permutations (wrt the Kendall's τ, Cayley, the Ulam metrics, and the reversal distance).

We start by recalling algebraic properties of permutations in Section 4.1. Then we define the metrics that will be used in our analysis in Section 4.2. We show that the Kendall's τ and the Cayley metrics are respectively linked to the left weak order and the strong right order in Section 4.3. This will be useful in Section 4.4, where we determine the convex sets corresponding to each metric. The runtime analysis is done in Section 4.5. We show that a permutation is uniquely determined by a common string in Section 4.6. We study permutations through their string form in Section 4.7.

4.1 Permutation Group

The set of the permutations of the elements of $[n] = \{1, 2, \ldots, n\}$ is denoted S_n. We recall that a permutation of S_n is a rearrangement of the elements of $[n]$. Let $\circ$ denote the composition law. We recall some known properties of the group $(S_n, \circ)$ that can be found in [Sag13]. We will start with the definition of a group.

Definition 9 (Group)**.** *A group $(G, .)$ is a set G equipped with a binary operation $. : G \times G \longrightarrow G$ that satisfies the following properties:*

- *There exists an* identity element $e \in G$ *such that:*

$$a.e = e.a = a, \tag{4.1}$$

  for all $a \in G$.

- *Each element $a \in G$ has an* inverse *denoted a^{-1} such that:*

$$a.a^{-1} = a^{-1}.a = e. \tag{4.2}$$

- *The binary operation '.' is* associative*:*

$$a.(b.c) = (a.b).c, \tag{4.3}$$

  for any elements $a,\ b$ and c of G.

- *The set G is* closed *under the binary operation '.' :*

$$a.b \in G, \tag{4.4}$$

  for any a and b in G.

Example 8. *The set of integers $\mathbb{Z}$ equipped with '$+$' is a group. The identity element is 0 and the inverse of an integer n is the integer $(-n)$. The sum of two integers remains an integer and '$+$' is associative.*

The set S_n of the permutations of the elements of $[n] = \{1, 2, \dots, n\}$ becomes a group when equipped with a composition law $\circ : S_n \times S_n \longrightarrow S_n$. Indeed, two permutations σ and τ of S_n can be composed into a permutation $\sigma \circ \tau$ of S_n.

Example 9. *Let $\sigma = 53241$ and let $\tau = 35241$ be two permutations of S_5. The permutation $\sigma \circ \tau$ is obtained by finding out where $1,\ 2,\ 3,\ 4,$ and 5 are respectively sent by first going through τ and then through σ. We have:*

$$\begin{aligned}
1 &\xrightarrow{\tau} 3 \xrightarrow{\sigma} 2, \\
2 &\xrightarrow{\tau} 5 \xrightarrow{\sigma} 1, \\
3 &\xrightarrow{\tau} 2 \xrightarrow{\sigma} 3, \\
4 &\xrightarrow{\tau} 4 \xrightarrow{\sigma} 4, \\
5 &\xrightarrow{\tau} 1 \xrightarrow{\sigma} 5.
\end{aligned}$$

Thus, $\sigma \circ \tau = 21345$.

In the group $(S_n, \circ)$, the identity element is $123 \dots n$. The computation of the inverse of a permutation requires the notion of *transpositions*. A *transposition* of S_n is a permutation that only rearrange two elements of $[n] = \{1, 2, \dots, n\}$.

Example 10. *The permutation 15342 of S_5 is a transposition.*

In fact, any permutation of S_n can be obtained through the composition of transpositions of S_n for $n \geq 2$. This composition of transpositions need not be unique.

Example 11. *The permutation* $\sigma = 53421$ *can be written as* $(15) \circ (23) \circ (34)$*. Indeed, we have:*

$$1 \xrightarrow{\sigma} 5 \xrightarrow{\sigma} 1,$$
$$2 \xrightarrow{\sigma} 3 \xrightarrow{\sigma} 4 \xrightarrow{\sigma} 2.$$

That is, $\sigma = (15) \circ (234)$*. Thus, by splitting* (234) *into a product of transpositions we obtain that* $\sigma = (15) \circ (23) \circ (34)$*.*

We say that the group $(S_n, \circ)$ is *generated* by its transpositions. This result enables us to:

- compute the inverse of any given permutation,
- compute the composition of any two permutations more easily.

By abuse of language we use the term ‘product’ instead of ‘composition’ while referring to transpositions. Moreover, the notation ‘$\circ$’ is omitted while composing transpositions.

Transpositions Product Rules

We recall the product rules of transpositions. Let $1 \leq i < j \leq k < l \leq n$ be integers, the product of two tranpositions follows the following rules:

- For any transposition (ij),
$$(ij)^{-1} = (ij). \tag{4.5}$$

- For any transpositions (ij) and (kl),
$$[(ij)(kl)]^{-1} = (kl)^{-1}(ij)^{-1}, \tag{4.6}$$
$$= (kl)(ij). \tag{4.7}$$

- If the sets $\{i, j\}$ and $\{k, l\}$ are disjoint then:
$$(ij)(kl) = (kl)(ij). \tag{4.8}$$

  We say that (ij) and (kl) *commute*.

Example 12. *The inverse of the permutation* $\sigma = 53421$ *in Example 11 is given* *by:*

$$
\begin{aligned}
\sigma^{-1} &= 53421^{-1},\\
&= [(15)(23)(34)]^{-1},\\
&= [(23)(34)]^{-1}(15)^{-1},\\
&= (34)^{-1}(23)^{-1}(15),\\
&= (34)(23)(15),\\
&= 54231.
\end{aligned}
$$

Example 13. *Let $\sigma = 53241$ and let $\tau = 35241$ be the two permutations of Example 9. We recompute $\sigma \circ \tau$ using transpositions:*

$$
\begin{aligned}
\sigma \circ \tau &= 53241 \circ 35241,\\
&= [(15)(23)](1325),\\
&= (15)(23)(1325),\\
&= (15)(23)(13)(32)(25),\\
&= 21345.
\end{aligned}
$$

We ended up to the same result as Example 9 without drawing any diagram.

Right and Left Transpositions

There are two different ways to compose (ij) with a permutation σ of S_n. The right transposition consists of composing (ij) with σ from the right (i.e., doing $\sigma \circ (ij)$). Whereas, the left transposition consists of composing (ij) with σ from the left (i.e, doing $(ij) \circ \sigma$). The right transposition swaps the letters at positions i and j in σ. Whereas, the left transposition swaps the letters i and j in σ.

Example 14. *In the group $(S_5, \circ)$, we have $(13) \circ 452\mathbf{13} = 452\mathbf{31}$ and $\mathbf{4}5\mathbf{2}13 \circ (13) = \mathbf{2}54\mathbf{1}3$.*

We are now ready to define the Kendall's τ, the Cayley, the Ulam metrics, and the reversal distance on the set S_n.

4.2 Metrics for Permutations

We shall consider four distinct metrics on S_n: the Kendall's τ metric K [Ken48, Cri12], the Cayley metric T [DH98, Cri12], the Ulam metric UL [AD99, Ula72], and the reversal distance R [BP96]. Each of these metrics uses the group property of $(S_n, \circ)$.

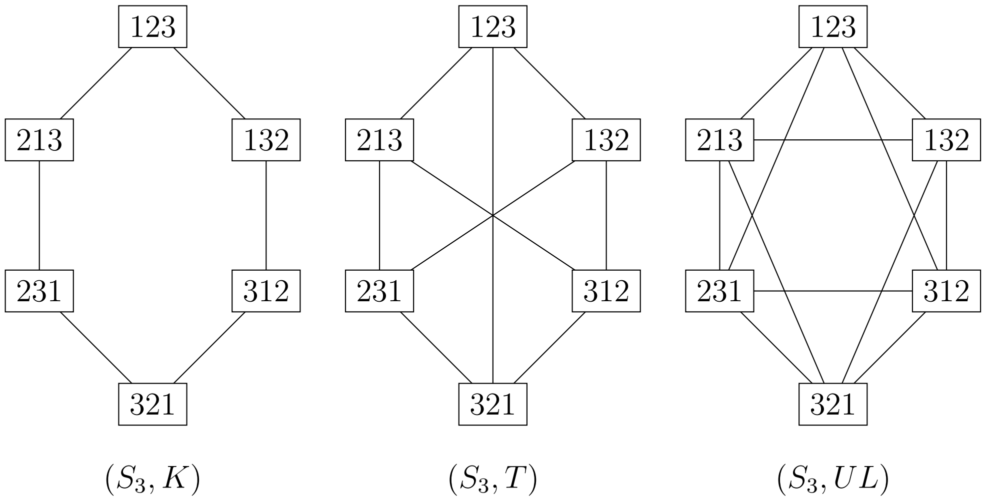


Figure 4.1: Illustration of the metric space (S_3, D) for different metrics D. An edge is drawn between two permutations whose distance from each other is one.

4.2.1 Kendall's τ metric

A tranposition that rearranges two consecutive values is called *adjacent*.

Example 15. *The transpositions* $(23) = 13245$ *and* $(45) = 12354$ *of* S_5 *are adjacent. Whereas, the transpositions* $(25) = 15342$ *and* $(13) = 32145$ *are not adjacent.*

The definition of the Kendall's τ metric uses *right adjacent transpositions*.

Definition 10 (Kendall's τ metric [Ken48, Cri12])**.** *Let* σ_1 *and* σ_2 *be permutations in* S_n*. The Kendall's* τ *distance* $K(\sigma_1, \sigma_2)$ *between* σ_1 *and* σ_2 *is the minimum number of* right adjacent transpositions *needed to obtain* σ_2 *from* σ_1*.*

This means that for the Kendall's τ metric a 1-neighbour is obtained by swapping any two letters whose positions are adjacent.

Example 16. *The permutation* 54123 *is a 1-neighbour of the permutation* 51423 *(i.e., their Kendall's* τ *distance is one) because one permutation is obtained from the other by swapping two letters whose positions are adjacent. More precisely:*

$$54123 = 51423 \circ (23),$$
$$51423 = 54123 \circ (23).$$

The Kendall's τ distance between two permutations σ_1 and σ_2 is therefore given by the minimum number of *adjacent transpositions* whose composition yields the permutation $\sigma_1^{-1} \circ \sigma_2$.

Example 17. *Let* $\sigma_1 = 32145$ *and let* $\sigma_2 = 45312$*, we have:*

$$\begin{aligned}
\sigma_1^{-1} \circ \sigma_2 &= 32145^{-1} \circ 45312, \\
&= (13)^{-1}(14)(25), \\
&= (13)(14)(25), \\
&= [(12)(23)(12)][(12)(23)(34)(23)(12)][(23)(34)(45)(34)(23)], \\
&= (12)(34)(23)(12)(23)(34)(45)(34)(23), \\
&= (34)\underbrace{(12)(23)(12)}(23)(34)(45)(34)(23), \\
&= (34)(23)(12)(23)(23)(34)(45)(34)(23), \\
&= (34)(23)(12)(34)(45)(34)(23).
\end{aligned}$$

The minimum number of adjacent transpositions whose composition yields the permutation $\sigma_1^{-1} \circ \sigma_2$ *is* 7*. Hence, the Kendall's* τ *distance between these two permutations is* 7*. Notice that this need not be the unique way to write* $\sigma_1^{-1} \circ \sigma_2$*.*

A more practical way to compute the Kendall's τ distance between two permutations σ_1 and σ_2 of S_n is given in [Ken48] as follows:

$$\begin{aligned}
K(\sigma_1, \sigma_2) =& \{(i, j) \in [n] \times [n] \mid [\sigma_1(i) < \sigma_1(j) \text{ AND } \sigma_2(i) > \sigma_2(j)] \\
& \text{OR } [\sigma_1(i) > \sigma_1(j) \text{ AND } \sigma_2(i) < \sigma_2(j)]\}.
\end{aligned}$$

In particular, $\max_{\sigma,\tau \in S_n} K(\sigma, \tau) = \frac{(n-1)n}{2}$.

4.2.2 Cayley metric

The definition of the Cayley metric uses left tranpositions.

Definition 11 (Cayley metric [DH98, Cri12])**.** *Let* σ_1 *and* σ_2 *be permutations in* S_n*. The Cayley distance* $T(\sigma_1, \sigma_2)$ *between* σ_1 *and* σ_2 *is the minimum number of* left transpositions *needed to obtain* σ_2 *from* σ_1*.*

This means that for the Cayley metric a 1-neighbour is obtained by swapping any two values.

Example 18. *The permutation* 52314 *is a 1-neighbour of the permutation* 54312 *(i.e., their Cayley distance is one) because one permutation is obtained from the other by swapping the values* 2 *and* 4*. More precisely:*

$$\begin{aligned}
52314 &= (24) \circ 54312, \\
54312 &= (24) \circ 52314.
\end{aligned}$$

The Cayley distance between two permutations σ_1 and σ_2 is therefore given by the minimum number of *transpositions* whose composition yields the permutation $\sigma_2 \circ \sigma_1^{-1}$.

Example 19. *Let $\sigma_1 = 32145$ and let $\sigma_2 = 45312$, we have:*

$$\begin{aligned}\sigma_2 \circ \sigma_1^{-1} &= 45312 \circ 32145^{-1},\\ &= (14)(25)(13)^{-1},\\ &= (14)(25)(13).\end{aligned}$$

The minimum number of transpositions whose composition yields the permutation $\sigma_2 \circ \sigma_1^{-1}$ is 3. Hence, the Cayley distance between these two permutations is 3. Notice that this need not be the unique way to write $\sigma_2 \circ \sigma_1^{-1}$.

A more practical way to compute the Cayley distance between two permutations σ_1 and σ_2 consists of determining the number of (disjoint) cycles in $\sigma_2 \circ \sigma_1^{-1}$ [DH98]. That is:

$$T(\sigma_1, \sigma_2) = n - \mid \{\text{Cycles in } \sigma_2 \circ \sigma_1^{-1}\} \mid . \tag{4.9}$$

Example 20. *Using the same permutations as in the previous example, we have that:*

$$\begin{aligned}\sigma_2 \circ \sigma_1^{-1} &= 45312 \circ 32145^{-1},\\ &= (14)(25)(13),\\ &= (14)(13)(25),\\ &= (134)(25).\end{aligned}$$

Hence, $\sigma_2 \circ \sigma_1^{-1}$ has two disjoint cycles. It follows that the Cayley distance between σ_1 and σ_2 is $5 - 2 = 3$

In particular, $\max_{\sigma,\tau \in S_n} T(\sigma, \tau) = n - 1$.

4.2.3 Ulam metric

A letter displacement consists of deleting a letter then inserting it in a different position. The definition of the Ulam metric uses letter displacements. A letter can be displaced through a chain of right transpositions.

A chain of right transpositions corresponding to a letter displacement keeps track of the consecutive positions taken by the moving letter until it reaches its final position. Hence, a letter of a permutation σ that moves from position i to position $j > i$ is the right composition of $(i\ i+1\ \cdots\ j) = (i\ i+1)(i+1\ i+2)\dots(j-1\ j)$ (position counting starts at 1) with σ. A permutation of the form $(i\ i+1\ \cdots\ j)$ is called an adjacent cycle [BD12]. The definition of the Ulam metric makes use of right adjacent cycles.

Definition 12 (Ulam metric [AD99, Ula72])**.** *Let σ_1 and σ_2 be permutations in S_n. The Ulam distance $UL(\sigma_1, \sigma_2)$ between σ_1 and σ_2 is the minimum number of right adjacent cycles (or letter displacements) needed to obtain σ_2 from σ_1.*

This means that a 1-neighbour is obtained through the displacement of one letter to a different position.

Example 21. *The permutations 25314 and 23154 are 1-neighbour (i.e., their Ulam distance is one) because one permutation is obtained from the other through the displacement of the letter 5. In particular,*

$$\begin{aligned} 23154 &= [25314 \circ (23)] \circ (34), \\ &= 25314 \circ [(23)(34)], \\ &= 25314 \circ (234). \end{aligned}$$

This shows that the letter at position 2 is first displaced to position 3, then displaced to position 4. All the transpositions appearing in the chain $(23)(34)$ are adjacents.

The Ulam distance between two permutations σ_1 and σ_2 is therefore given by the minimum number of *adjacent cycles* whose composition yields the permutation $\sigma_1^{-1} \circ \sigma_2$.

Example 22. *Let $\sigma_1 = 32145$ and let $\sigma_2 = 45312$, we have:*

$$\begin{aligned} \sigma_1^{-1} \circ \sigma_2 &= 32145^{-1} \circ 45312, \\ &= (13)^{-1}(14)(25), \\ &= (13)(14)(25), \\ &= [(12)(23)(12)][(12)(23)(34)(23)(12)][(23)(34)(45)(34)(23)], \\ &= (34)(23)(12)(34)(45)(34)(23), \\ &= (34)(23)(12)(345)(34)(23), \\ &= (34)(23)(345)(12)(34)(23), \\ &= (34)(2345)(34)(12)(23), \\ &= (34)(2345)(34)(123). \end{aligned}$$

The minimum number of adjacent cycles whose composition yields the permutation $\sigma_1^{-1} \circ \sigma_2$ is 4. Hence, the Ulam distance between these two permutations is 4. Notice that this composition need not be unique.

A more practical way to compute the Ulam distance between two permutations is not known yet. However, it has been shown that the Ulam metric is well approximated by edit distances on non repetitive strings of the same length (See [AN10]).

4.2.4 Reversal distance

A reversal consists of reversing the order of the values between two distinct positions of a permutation [BP96]. The definition of the reversal distance uses reversals. A reversal can be performed through a chain of disjoint left transpositions.

A chain of disjoint left transpositions corresponding to a reversal keeps track of the different pair of values that are consecutively swapped. Hence, reversing the values of a permutation σ between position i and j (where $\sigma(j) > \sigma(i)$) is the left composition of

$$(\sigma(i)\,\sigma(j))(\sigma(i)+1\,\sigma(j)-1)\ldots\left(\sigma(i)+\left[\tfrac{\sigma(j)-\sigma(i)}{2}\right]\,\sigma(j)-\left[\tfrac{\sigma(j)-\sigma(i)}{2}\right]\right) \tag{4.10}$$

with σ, where $\left[\frac{\sigma(j)-\sigma(i)}{2}\right]$ denotes the integer part of $\frac{\sigma(j)-\sigma(i)}{2}$. A permutation of this form will be referred to as a reversal.

Definition 13 (Reversal distance [BP96])**.** *Let σ_1 and σ_2 be permutations in S_n. The reversal distance $R(\sigma_1, \sigma_2)$ between σ_1 and σ_2 is the minimal number of reversals needed to obtain σ_1 from σ_2.*

This means that a 1-neighbour is obtained through one reversal.

Example 23. *The permutations 13425 and $\sigma = 12435$ are 1-neighbour (i.e., their reversal distance is one) because one permutation is obtained from the other through the reversal of the values between positions 2 and 4 (position counting starts at 1). In particular,*

$$\begin{aligned} 13425 &= \left(\sigma(2)+\left[\tfrac{\sigma(4)-\sigma(2)}{2}\right]\,\sigma(4)-\left[\tfrac{\sigma(4)-\sigma(2)}{2}\right]\right)\circ\sigma, \\ &= (23)\circ 12435. \end{aligned}$$

The reversal distance between two permutations σ_1 and σ_2 is therefore given by the minimum number of *reversals* whose composition yields the permutation $\sigma_2 \circ \sigma_1^{-1}$.

Example 24. *Let $\sigma_1 = 32145$ and let $\sigma_2 = 45312$, we have:*

$$\begin{aligned}
\sigma_2 \circ \sigma_1^{-1} &= 45312 \circ 32145^{-1}, \\
&= (14)(25)(13)^{-1}, \\
&= (14)(13)(25), \\
&= (134)(25), \\
&= (13)(34)(25), \\
&= [(13)(22)][(25)(34)].
\end{aligned}$$

The minimum number of reversals whose composition yields the permutation $\sigma_2 \circ \sigma_1^{-1}$ is 2. Hence, the reversal distance between these two permutations is 2. Notice that this composition need not be unique.

4.3 Order on Permutations

The permutations of the elements of $[n] = \{1, 2, \ldots, n\}$ can be partially ordered. There are two well known partial orders on the set S_n: the weak left order and the strong right order. We show that these partial orders are respectively linked to the Kendall's τ and the Cayley metrics. In particular, these partial orders define *convex subsets of the partially ordered set*. We shall see that for the weak left order, convex subsets of the partially ordered set are convex subsets of the metric space (S_n, K). This is not true for the strong right order and the metric space (S_n, T). However, the converse is always true: a convex subset of the metric space (S_n, T) (resp. (S_n, K)) is a convex subset of the partial ordered set wrt the strong right (resp. weak left) order.

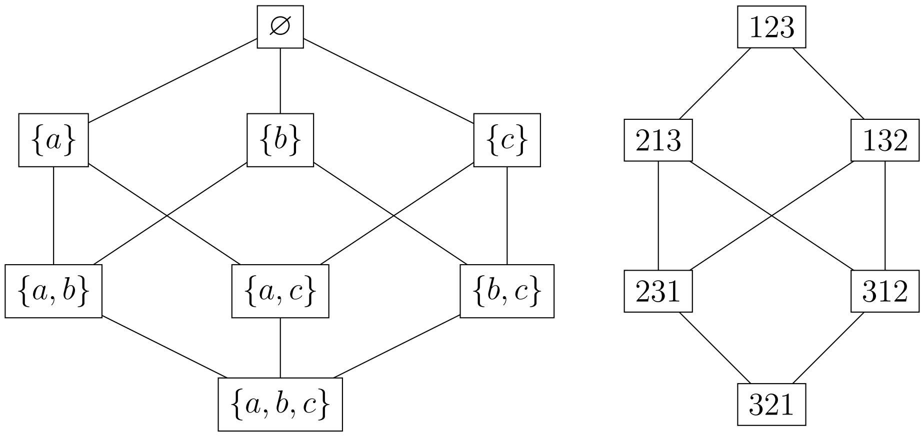
The notion of order convexity [Pel13] is not new. Moreover, it is not difficult to notice the relationship between the Kendall's τ (resp. Cayley) metric and the weak left (resp. strong right) order on S_n. However, the relationship between the Kendall's τ (resp. Cayley) metric convex sets and the left weak (resp. strong right) order convex sets in S_n is a new contribution. The new results presented in this section are not intermediate results to another main conclusion. Instead, they are useful tools for finding convex sets in the metric spaces (S_n, K) and (S_n, T).

We start by recalling the notions of partial order and lattice.

Definition 14 (Partial order)**.** *A partial order on a set $\mathcal{S}$ is a binary relation $\prec$ that is:*

1. *Reflexive*

$$x \prec x \text{ for any } x \in \mathcal{S}, \tag{4.11}$$



Hasse diagram of $(\{a, b, c\}, \subseteq)$ Hasse diagram of $(S_3, \prec)$

Figure 4.2: Examples of partially ordered sets.

2. *Antisymmetric*

$$\text{If } x \prec y \text{ and } y \prec x, \text{ then } x = y \text{ for any } x, y \in \mathcal{S}, \tag{4.12}$$

3. *Transitive*

$$\text{If } x \prec y \text{ and } y \prec z, \text{ then } x \prec z \text{ for any } x, y, z \in \mathcal{S}. \tag{4.13}$$

Example 25. *The inclusion $\subseteq$ is a partial order on the elements of the set $\{a, b, c\}$. The ordering of the elements of $\{a, b, c\}$ with respect to $\subseteq$ can be represented in a Hasse diagram. See Figure 4.2.*

Definition 15 (Lattice)**.** *A partially ordered set $(\mathcal{S}, \prec)$ is a lattice if each pair of elements has a unique least upper bound and a unique greatest lower bound.*

Example 26. *The pair $(\{a, b, c\}, \subseteq)$ forms a lattice.*

In particular, any non empty subset of a lattice has a least upper bound and a greatest lower bound.

4.3.1 Weak Left Order

Let $\prec_w$ denote the weak left order (also called weak order in [Dra05]). The weak order is a partial order on S_n that is defined as follows:

Definition 16. *[YO69] Let σ and τ be two elements of S_n. We say that τ* covers *σ and write $\sigma \xrightarrow{w} \tau$ iff there exists a positive integer $i < n$ such that:*

$$\sigma(i) = \tau(i+1) < \sigma(i+1) = \tau(i) \text{ and } \sigma(k) = \tau(k) \text{ for } k \neq i, i+1. \tag{4.14}$$

An element υ is said to be not smaller than σ with respect to $\prec_w$ (i.e., $\sigma \prec_w \upsilon$) iff there exist $\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_{m-1}$ in S_n such that:

$$\sigma = \sigma_0 \xrightarrow{w} \sigma_1 \xrightarrow{w} \sigma_2 \xrightarrow{w} \sigma_3 \xrightarrow{w} \cdots \xrightarrow{w} \sigma_{m-1} \xrightarrow{w} \sigma_m = \upsilon. \tag{4.15}$$

The weak order can be defined using the Kendall's τ distance. To this end, let us first recall from [Dra05] that an inversion in a permutation σ is a a pair of values $(\sigma(i), \sigma(j))$ such that: $\sigma(i) > \sigma(j)$ and $i < j$.

We can see that a permutation τ covers a permutation σ with respect to the weak order iff the Kendall's τ distance between the two permutations is one and τ has exactly one more inversion than σ. Therefore, the weak order can be used to order the elements of S_n by taking into account their Kendall's τ distance from each other. In this case, if $\sigma \prec_w \tau$ then the Kendall's τ segment between σ and τ is simply:

$$[\sigma, \tau]_K = \{\nu \mid \sigma \prec_w \nu \prec_w \tau\}, \tag{4.16}$$
$$= [\sigma, \tau]_{\prec_w}. \tag{4.17}$$

The weak order $\prec_w$ can be used to define convex subsets of the partially ordered set $(S_n, \prec_w)$. Namely, a subset C of S_n is convex iff for any $\sigma \prec_w \tau$ in C we have:

$$\sigma \prec_w \nu \prec_w \tau \Rightarrow \nu \in C. \tag{4.18}$$

Or equivalently,

$$[\sigma, \tau]_{\prec_w} \subseteq C. \tag{4.19}$$

Theorem 4. *A convex set of S_n with respect to the Kendall's τ metric is a convex subset of the partially ordered set $(S_n, \prec_w)$. Conversely, a convex subset of the partially ordered set $(S_n, \prec_w)$ is a convex set of S_n with respect to the Kendall's τ metric.*

Proof. Let C be a convex set of S_n with respect to the Kendall's τ metric. For any σ and τ in C, the segment $[\sigma, \tau]_K$ is contained in C because C is convex. Now, let σ' and τ' be two elements of C such that $\sigma' \prec_w \tau'$. We have:

$$\begin{aligned}[\sigma', \tau']_K &= \{\nu \in S_n \mid \sigma' \prec_w \nu \prec_w \tau'\}, \\ &= [\sigma', \tau']_{\prec_w}, \\ &\subseteq C.\end{aligned}$$

Thus, C is a convex subset in the partially ordered set $(S_n, \prec_w)$.

Conversely, let C be a convex subset in the partially ordered set $(S_n, \prec_w)$. For any $\sigma \prec_w \tau$ in C, the set

$$[\sigma, \tau]_{\prec_w} = \{\nu \in S_n \mid \sigma \prec_w \nu \prec_w \tau\}, \tag{4.20}$$

is always contained in C. Let us now consider two elements σ' and τ' of C.

If the elements σ' and τ' are comparable with respect to the weak order, then the Kendall's τ segment they form is contained in C. Indeed, without loss of generality let $\sigma' \prec_w \tau'$. We have that $[\sigma', \tau']_K = [\sigma', \tau']_{\prec_w}$ and is contained in C.

If the elements σ' and τ' are not comparable with respect to $\prec_w$, then we use the lattice property of $(S_n, \prec_w)$ [YO69] to show that the Kendall's τ segment $[\sigma', \tau']_K$ they form is either of the form:

- $[LUB(\sigma', \tau'), GLB(\sigma', \tau')]_{\prec_w}$,
- $[LUB(\sigma', \tau'), \sigma']_{\prec_w} \cup [LUB(\sigma', \tau'), \sigma']_{\prec_w}$,
- $[\sigma', GLB(\sigma', \tau')]_{\prec_w} \cup [\tau', GLB(\sigma', \tau')]_{\prec_w}$,

where GLB (resp. LUB) denotes the Greatest Lower Bound (resp. Lowest Upper Bound). Since σ' and τ' are already contained in C, it remains to show that the greatest lower bound (GLB) and the lowest upper bound (LUB) of the pair (σ', τ') belong to C.

C is a convex subset in the partially ordered set $(S_n, \prec_w)$ that contains σ' and τ'. If $GLB(\sigma', \tau')$ is not contained in C, then both the segments $[\sigma', GLB(\sigma', \tau')]_{\prec_w}$ and $[\tau', GLB(\sigma', \tau')]_{\prec_w}$ are not contained in C. Hence, the convex set C is either the set $\{\sigma'\}$ or the set $\{\tau'\}$. Either way, the convex set C can not contain both σ' and τ' unless $\sigma' = \tau'$. This contradicts the initial assumption. Consequently, $GLB(\sigma', \tau')$ must be contained in C. We show in a similar way that $LUB(\sigma', \tau')$ is contained in C.

As a result, the segment $[\sigma', \tau']_K$ is always contained in C whenever σ' and τ' are elements of C. Thus, C is a convex set of S_n with respect to the Kendall's τ metric. $\square$

Corollary 1. *Let $\sigma \prec_w \tau$ be two permutations of S_n, the set $[\sigma, \tau]_{\prec_w}$ is convex (with respect to the Kendall's τ metric) and is equal to $co(\{\sigma, \tau\})$.*

Proof. Since $\sigma \prec_w \tau$, we have:

$$[\sigma, \tau]_{\prec_w} = [\sigma, \tau]_K.$$

As $[\sigma, \tau]_{\prec_w}$ is a convex set of the metric space (S_n, K), the segment $[\sigma, \tau]_K$ is also a convex set of the metric space (S_n, K). Since σ and τ both belong to the set $co(\{\sigma, \tau\})$, then the segment $[\sigma, \tau]_K$ must be contained in the set $co(\{\sigma, \tau\})$

because of its convexity. The smallest convex set containing both σ and τ being $co(\{\sigma, \tau\})$ by definition, we must have that:

$$[\sigma, \tau]_K = co(\{\sigma, \tau\}).$$

The result follows. □

4.3.2 Strong Right Order

Let $\prec$ denote the strong right order (also called Bruhat order in [Inc04]). The strong order is a partial order on S_n that is defined as follows:

Definition 17. *[Inc04] Let σ and τ be two elements of S_n. We say that τ covers σ and write $\sigma \xrightarrow{1} \tau$ iff there exist two distinct positions $k_1 < k_2$ such that:*

$$\sigma(k_1) = \tau(k_2) < \sigma(k_2) = \tau(k_1), \text{ and } \sigma(i) = \tau(i) \text{ for all } i \neq k_1,\ k_2, \tag{4.21}$$

and for all $k_1 \leq k \leq k_2$ the value $\sigma(k)$ does not satify

$$\sigma(k_1) \leq \sigma(k) \leq \sigma(k_2). \tag{4.22}$$

An element υ is said to be not smaller than σ with respect to $\prec$ (i.e., $\sigma \prec \upsilon$) iff there exist $\sigma_1,\ \sigma_2,\ \ldots,\ \sigma_{m-1}$ in S_n such that:

$$\sigma = \sigma_0 \xrightarrow{1} \sigma_1 \xrightarrow{1} \sigma_2 \xrightarrow{1} \sigma_3 \xrightarrow{1} \cdots \xrightarrow{1} \sigma_{m-1} \xrightarrow{1} \sigma_m = \upsilon. \tag{4.23}$$

We can see that a permutation τ covers a permutation σ with respect to the strong order iff the Cayley distance between the two permutations is one and τ has exactly one more inversion than σ. Therefore, the strong order can be used to order the elements of S_n by taking into account their Cayley distance from each other. In this case, if $\sigma \prec \tau$ then the Cayley segment between σ and τ is simply:

$$\begin{aligned} [\sigma, \tau]_T &= \{\nu \mid \sigma \prec \nu \prec \tau\}, && (4.24)\\ &= [\sigma, \tau]_\prec. && (4.25) \end{aligned}$$

The strong order $\prec$ can be used to define convex subsets of the partially ordered set $(S_n, \prec)$. Namely, a subset C of S_n is convex iff for any $\sigma \prec \tau$ in C we have:

$$\sigma \prec \nu \prec \tau \Rightarrow \nu \in C. \tag{4.26}$$

Or equivalently,

$$[\sigma, \tau]_\prec \subseteq C. \tag{4.27}$$

Theorem 5. *A convex set of S_n with respect to the Cayley metric is a convex subset of the partially ordered set $(S_n, \prec)$, but the converse is not true in general.*

Proof. The proof to show that a convex set of S_n with respect to the Cayley metric is a convex subset of the partially ordered set $(S_n, \prec)$ is similar to the first part of the proof of Theorem 4.

The converse is not always true since the partially ordered set $(S_n, \prec)$ is not a lattice for $n \geq 3$ [Rea02]. $\square$

Example 27. *The segment $[123, 312]_\prec$ is a convex subset of the partially ordered set $(S_3, \prec)$ but it is not a convex set of S_3 with respect to the Cayley metric. Indeed, 213 and 132 belong to $[123, 312]_\prec$ but the element 231 of the segment $[213, 132]_T$ is not in $[123, 312]_\prec$. See Figure 4.2.*

4.4 Segments and Convex Hulls of Permutations

The convex hulls of the metric spaces of permutations considered in this thesis have not yet been determined in any previous study. Even if a full knowledge of the definitions of these convex hulls is not necessary to the runtime analysis, these sets are determined for completeness and for novelty. The following metrics on S_n are considered : Kendall's τ, Cayley, Ulam metrics, and the reversal distance. These metrics are referred to as length metrics [Dia88]. As such, they are bound to the group $(S_n, \circ)$ where '$\circ$' is the composition law. The results presented in this section are not the first to contribute to both convex analysis and abstract algebra (see for instance [AA18]). Nevertheless, the results presented in this section are the first to focus on the convex hulls of the metric spaces: $(S_n, K),\ (S_n, T),\ (S_n, UL)$, and (S_n, R).

4.4.1 Kendall's τ metric

The Kendall's τ metric uses right adjacent transpositions. We shall use subgroups of $(S_n, \circ)$ that are generated by adjacent transpositions to determine convex sets of the metric space (S_n, K).

Lemma 3. *Let σ and τ be two permutations of S_n, and let $\mathcal{P}_{at}(\sigma^{-1} \circ \tau)$ be the set of all possible ways to write $\sigma^{-1} \circ \tau$ as a minimal product of adjacent transpositions. An element ν of the segment $[\sigma, \tau]_K$ is of the form:*

$$\nu = \sigma \circ \prod_{i=1}^{k} \tau_i, \tag{4.28}$$

where $k \leq K(\sigma, \tau)$, $\prod_{i=1}^{0} \tau_i = \mathrm{id}$, and $\prod_{i=1}^{K(\sigma,\tau)} \tau_i \in \mathcal{P}_{at}(\sigma^{-1} \circ \tau)$.

Proof. Let σ and τ be two permutations of S_n. The minimum number of right adjacent transpositions needed to transform σ into τ (and conversely) is therefore $K(\sigma,\tau)$. Therefore, we can find $K(\sigma,\tau)$ transpositions τ_i such that:

$$\prod_{i=1}^{K(\sigma,\tau)} \tau_i = \sigma^{-1} \circ \tau. \tag{4.29}$$

Let $\mathcal{Z}_{\sigma^{-1}\circ\tau}$ be the centralizer of $\sigma^{-1} \circ \tau$, that is:

$$\mathcal{Z}_{\sigma^{-1}\circ\tau} = \{\nu \in S_n \mid \nu \circ (\sigma^{-1} \circ \tau) \circ \nu^{-1} = \sigma^{-1} \circ \tau\}. \tag{4.30}$$

There are $\mid \mathcal{Z}_{\sigma^{-1}\circ\tau} \mid$ different ways to write $\sigma^{-1} \circ \tau$ as a product of disjoint cycles. Therefore, there may be more than one product of $K(\sigma,\tau)$ adjacent transpositions that yield the permutation $\sigma^{-1} \circ \tau$. We denote $\mathcal{P}_{at}(\sigma^{-1} \circ \tau)$ the set of all distinct products of $K(\sigma,\tau)$ adjacent transpositions yielding the permutation $\sigma^{-1} \circ \tau$.

If a permutation ν belongs to the segment $[\sigma,\tau]_K$, then the minimum number of right adjacent transpositions needed to transform σ into ν is:

$$K(\sigma,\nu) = K(\sigma,\tau) - K(\nu,\tau).$$

Thus, $K(\sigma,\nu) \leq K(\sigma,\tau)$. That is, there exists a product of $K(\sigma,\nu)$ adjacent transpositions such that:

$$\nu = \sigma \circ \prod_{i=1}^{K(\sigma,\nu)} \tau_i'. \tag{4.31}$$

Similarly, there exists a product of $K(\nu,\tau)$ adjacent transpositions that transform ν into τ:

$$\begin{aligned} \tau &= \nu \circ \prod_{i=1}^{K(\nu,\tau)} \tau_i'', \\ &= \sigma \circ \left[\prod_{i=1}^{K(\sigma,\nu)} \tau_i' \circ \prod_{i=1}^{K(\nu,\tau)} \tau_i'' \right]. \end{aligned}$$

Necessarily, there exists a product p in $\mathcal{P}_{at}(\sigma^{-1} \circ \tau)$ such that:

$$p = \prod_{i=1}^{K(\sigma,\nu)} \tau_i' \circ \prod_{i=1}^{K(\nu,\tau)} \tau_i''. \tag{4.32}$$

As a result, $\prod_{i=1}^{K(\sigma,\nu)} \tau_i'$ is obtained from p by deleting the right most $K(\nu,\tau)$ transpositions. This means that any element ν of the segment $[\sigma,\tau]_K$ whose distance from τ is k is of the form $\sigma \circ p'$, where p' is obtained from one of the minimal adjacent transpositions product expression of $\sigma^{-1} \circ \tau$ by deleting its k right most transpositions. $\square$

Let us denote $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$ the set:

$$\{\mathrm{id}\} \cup \bigcup_{k=1}^{K(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \middle| \prod_{i=1}^{K(\sigma,\tau)} \tau_i \in \mathcal{P}_{at}(\sigma^{-1} \circ \tau) \right\}. \tag{4.33}$$

Proposition 5. *In the metric space, (S_n, K) the segment formed by σ and τ is given by:*

$$[\sigma, \tau]_K = \{\sigma \circ a \mid a \in \mathcal{A}_{at}(\sigma^{-1} \circ \tau)\}. \tag{4.34}$$

Proof. By Lemma 3, we have:

$$\begin{aligned} [\sigma, \tau]_K &= \bigcup_{k=0}^{K(\sigma,\tau)} \left\{ \sigma \circ \prod_{i=1}^{k} \tau_i \middle| k \leq K(\sigma, \tau) \text{ and } \prod_{i=1}^{K(\sigma,\tau)} \tau_i \in \mathcal{P}_{at}(\sigma^{-1} \circ \tau) \right\}, \\ &= \{\sigma \circ a \mid a \in \mathcal{A}_{at}(\sigma^{-1} \circ \tau)\}. \end{aligned}$$

$\square$

Theorem 6. *In the metric space (S_n, K), a segment $[\sigma, \tau]_K$ is convex iff $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$ is a subgroup of $(S_n, \circ)$.*

Proof. The segment $[\sigma, \tau]_K$ is convex iff for any ν and ν' in $[\sigma, \tau]_K$, the segment $[\nu, \nu']_K$ is always contained in the segment $[\sigma, \tau]_K$. Let ν and ν' belong to the segment $[\sigma, \tau]_K$. There exist a and a' in $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$ such that:

$$\begin{aligned} \nu &= \sigma \circ a, \\ \nu' &= \sigma \circ a'. \end{aligned}$$

Let ν'' be an element of the segment $[\nu, \nu']_K$. There exists $a'' \in \mathcal{A}_{at}(\nu^{-1} \circ \nu')$ such that:

$$\begin{aligned} \nu'' &= \nu \circ a'', \\ &= (\sigma \circ a) \circ a'', \\ &= \sigma \circ aa''. \end{aligned}$$

Since

$$\begin{aligned} \nu^{-1} \circ \nu' &= (\sigma \circ a)^{-1} \circ (\sigma \circ a'), \\ &= a^{-1} \circ a', \end{aligned}$$

then $\mathcal{A}_{at}(\nu^{-1} \circ \nu') = \mathcal{A}_{at}(a^{-1} \circ a')$. Hence, a'' in $\mathcal{A}_{at}(a^{-1} \circ a')$. Thus,

$$aa'' \in \mathcal{A}_{at}(\sigma^{-1} \circ \tau)\mathcal{A}_{at}(a^{-1} \circ a'). \tag{4.35}$$

Notice that

$$a^{-1} \circ a' \in [\mathcal{A}_{at}(\sigma^{-1} \circ \tau)]^{-1} \mathcal{A}_{at}(\sigma^{-1} \circ \tau). \tag{4.36}$$

As a result,

$$\mathcal{A}_{at}(a^{-1} \circ a') \subseteq [\mathcal{A}_{at}(\sigma^{-1} \circ \tau)]^{-1} \cup [\mathcal{A}_{at}(\sigma^{-1} \circ \tau)]^{-1} \mathcal{A}_{at}(\sigma^{-1} \circ \tau). \tag{4.37}$$

Therefore,

$$[\mathcal{A}_{at}(\sigma^{-1} \circ \tau)]^{-1} \cup [\mathcal{A}_{at}(\sigma^{-1} \circ \tau)]^{-1} \mathcal{A}_{at}(\sigma^{-1} \circ \tau) \subseteq \mathcal{A}_{at}(\sigma^{-1} \circ \tau), \tag{4.38}$$

iff $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$ is a group for the composition law. In this case, $aa'' \in \mathcal{A}_{at}(\sigma^{-1} \circ \tau)$ and $\nu'' = \sigma \circ aa''$ is in the segment $[\sigma, \tau]_K$. That is, for any $\nu,\ \nu' \ \in \ [\sigma, \tau]_K$ the segment $[\nu, \nu']_K$ is always contained in $[\sigma, \tau]_K$. Thus, the segment $[\sigma, \tau]_K$ is a convex set for the Kendall's τ metric. □

Let $\langle \mathcal{A}_{at}(\sigma^{-1} \circ \tau) \rangle$ denote the smallest subgroup of $(S_n, \circ)$ containing the set $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$. We can see that this subgroup is generated by the adjacent transpositions appearing in a factorization of $\sigma^{-1} \circ \tau$ into a product (with respect to $\circ$) of $K(\sigma, \tau)$ adjacent transpositions. We have the following result:

Corollary 2. *In the metric space (S_n, K), the convex hull of two elements σ and τ is the left coset $\sigma\ \langle \mathcal{A}_{at}(\sigma^{-1} \circ \tau) \rangle$.*

Proof. We start with the result of Proposition 5, namely:

$$[\sigma, \tau]_K = \left\{ \sigma \circ a \middle| a \in \mathcal{A}_{at}(\sigma^{-1} \circ \tau) \right\}.$$

Then, we use Theorem 6 to deduce that the smallest convex set containing the segment $[\sigma, \tau]_K$ is obtained from the smallest group containing the set $\mathcal{A}_{at}(\sigma^{-1} \circ \tau)$. That is:

$$co([\sigma, \tau]_K) = \left\{ \sigma \circ g \middle| g \in \left\langle \mathcal{A}_{at}(\sigma^{-1} \circ \tau) \right\rangle \right\}. \tag{4.39}$$

The result follows as $co(\{\sigma, \tau\}) = co([\sigma, \tau]_K)$. □

Corollary 3. *In the metric space (S_n, K), the convex hull of m elements is the convex hull formed by the greatest lower bound and the least upper bound of these m elements with respect to the weak order '$\prec_w$':*

$$co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\}) = co(\{\sigma, \tau\}), \tag{4.40}$$

where $\sigma = \mathrm{GLB}_{1 \leq i \leq m}\{\sigma_i\}$ and $\tau = \mathrm{LUB}_{1 \leq i \leq m}\{\sigma_i\}$.

Proof. As $(S_n, \prec_w)$ is a lattice, the set $\{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ has a greatest lower bound $\sigma = \mathrm{GLB}_{1 \leq i \leq m}\{\sigma_i\}$ and a lowest upper bound $\tau = \mathrm{LUB}_{1 \leq i \leq m}\{\sigma_i\}$ such that:

$$[\sigma, \tau]_K = \{\nu \in S_n \mid \sigma \prec_w \nu \prec_w \tau\}. \tag{4.41}$$

It follows that $co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$ is contained in $co([\sigma, \tau]_K) = co(\{\sigma, \tau\})$.

Conversely, we show that σ and τ belong to the set $co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$. Let us consider $\sigma = \mathrm{GLB}_{1 \leq i \leq m}\{\sigma_i\}$. If $\sigma = \sigma_{i_0}$ for some $1 \leq i_0 \leq m$, then $\sigma \in co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$ and we are done. If $\sigma \neq \sigma_i$ for all $1 \leq i \leq m$, then there exists two distinct indices i_0 and i_1 such that σ_{i_0} and σ_{i_1} are not comparable and $\sigma = \mathrm{GLB}(\sigma_{i_0}, \sigma_{i_1})$. Therefore, $\sigma \in [\sigma_{i_0}, \sigma_{i_1}]$ and thus is an element of $co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$. We show in a similar way that τ is an element $co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$. As a result, $co(\{\sigma, \tau\})$ is contained in $co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\})$. □

Corollary 4. *Any convex set of the metric space (S_n, K) is of the form $[\sigma, \tau]_{\prec_w}$, where $\sigma \prec_w \tau$.*

Proof. By Corollary 3, any convex set is of the form $co(\{\sigma, \tau\})$ where $\sigma \prec_w \tau$. By Corollary 1, $co(\{\sigma, \tau\})$ is equal to $[\sigma, \tau]_{\prec_w}$. □

Covering S_n using the Kendall's τ metric.

The lowest element of $(S_n, \prec_w)$ is the identity permutation id because it does not contain any inversion. The greatest element of $(S_n, \prec_w)$ is $n\, n-1 \ldots 321$ as it contains all the possible inversions. Thus,

$$S_n = [\mathrm{id}, n\, n-1 \ldots 321]_{\prec_w}, \tag{4.42}$$

$$= co(\{\mathrm{id}, n\, n-1 \ldots 321\}). \tag{4.43}$$

Example 28. *The lowest and greatest elements of $(S_3, \prec_w)$ are respectively id and 321. Hence,*

$$S_3 = [\mathrm{id}, 321]_{\prec_w},$$

$$= co(\{\mathrm{id}, 321\}).$$

Any element of S_3 can be obtained from id by gradually adding inversions through right adjacent transpositions. Moreover, S_3 can be recovered through (the convex hull of) the elements id and 321 only.

4.4.2 Cayley metric

The Cayley metric uses left transpositions. We shall use subgroups of $(S_n, \circ)$ that are generated by transpositions to determine convex sets of the metric space (S_n, T).

Lemma 4. *Let σ and τ be two permutations of S_n, and let $\mathcal{P}_t(\tau \circ \sigma^{-1})$ be the set of all possible ways to write $\tau \circ \sigma^{-1}$ as a minimal product of transpositions. An* *element ν of the segment $[\sigma,\tau]_T$ is of the form:*

$$\nu = \prod_{i=1}^{k} \tau_i \circ \sigma, \tag{4.44}$$

where $k \leq T(\sigma,\tau)$, $\prod_{i=1}^{0} \tau_i = \mathrm{id}$, and $\prod_{i=1}^{T(\sigma,\tau)} \tau_i \in \mathcal{P}_t(\tau \circ \sigma^{-1})$.

Proof. Similar to that of Lemma 3 where minimal products of transpositions yielding $\tau \circ \sigma^{-1}$ are considered instead of minimial products of adjacent transpositions yielding $\sigma^{-1} \circ \tau$. $\square$

Let us denote $\mathcal{A}_t(\tau \circ \sigma^{-1})$ the set:

$$\{\mathrm{id}\} \cup \bigcup_{k=1}^{T(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \,\middle|\, \prod_{i=1}^{T(\sigma,\tau)} \tau_i \in \mathcal{P}_t(\tau \circ \sigma^{-1}) \right\}. \tag{4.45}$$

Proposition 6. *In the metric space, (S_n, T) the segment formed by σ and τ is given by:*

$$[\sigma,\tau]_T = \{a \circ \sigma \mid a \in \mathcal{A}_t(\tau \circ \sigma^{-1})\}. \tag{4.46}$$

Proof. By Lemma 4, we have:

$$\begin{aligned}
[\sigma,\tau]_T &= \bigcup_{k=0}^{T(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \circ \sigma \,\middle|\, k \leq T(\sigma,\tau) \text{ and } \prod_{i=1}^{T(\sigma,\tau)} \tau_i \in \mathcal{P}_t(\tau \circ \sigma^{-1}) \right\}, \\
&= \{a \circ \sigma \mid a \in \mathcal{A}_t(\tau \circ \sigma^{-1})\}.
\end{aligned}$$

$\square$

Theorem 7. *In the metric space (S_n, T), a segment $[\sigma,\tau]_T$ is convex iff $\mathcal{A}_t(\tau \circ \sigma^{-1})$ is a subgroup of $(S_n, \circ)$.*

Proof. Similar to that of Theorem 6. $\square$

Let $\langle \mathcal{A}_t(\tau \circ \sigma^{-1}) \rangle$ denote the smallest subgroup of $(S_n, \circ)$ containing the set $\mathcal{A}_t(\tau \circ \sigma^{-1})$. We can see that this subgroup is generated by the transpositions appearing in a factorization of $\tau \circ \sigma^{-1}$ into a product (with respect to $\circ$) of $T(\sigma,\tau)$ transpositions. We have the following result:

Corollary 5. *In the metric space (S_n, T), the convex hull of two elements σ and τ is the right coset $\langle \mathcal{A}_t(\tau \circ \sigma^{-1}) \rangle\ \sigma$.*

Proof. We start with the result of Proposition 6, namely:

$$[\sigma,\tau]_T = \left\{ a \circ \sigma \middle| a \in \mathcal{A}_t(\tau \circ \sigma^{-1}) \right\}.$$

Then, we use Theorem 7 to deduce that the smallest convex set containing the segment $[\sigma, \tau]_T$ is obtained from the smallest group containing the set $\mathcal{A}_t(\tau \circ \sigma^{-1})$. That is:

$$co([\sigma, \tau]_T) = \left\{ g \circ \sigma \middle| g \in \left\langle \mathcal{A}_t(\tau \circ \sigma^{-1}) \right\rangle \right\}. \tag{4.47}$$

The result follows as $co(\{\sigma, \tau\}) = co([\sigma, \tau]_T)$. □

Corollary 6. *In the metric space (S_n, T), the convex hull of $m \geq 2$ elements $\sigma_1,\ \sigma_2,\ \dots,$ and σ_m is:*

$$co\left(\bigcup \left\{ \left\langle \mathcal{A}_t(\tau \circ \sigma^{-1}) \right\rangle \sigma \middle| \sigma \in S_{\substack{\min,\prec \\ 1 \leq i \leq m}}(\sigma_i) \text{ and } \tau \in S_{\substack{\max,\prec \\ 1 \leq i \leq m}}(\sigma_i) \right\} \right), \tag{4.48}$$

where $S_{\substack{\max,\prec \\ 1 \leq i \leq m}}(\sigma_i)$ (resp. $S_{\substack{\min,\prec \\ 1 \leq i \leq m}}(\sigma_i)$) is the set of largest (resp. smallest) elements of the set $\{\sigma_1,\ \sigma_2,\ \dots,\ \sigma_m\}$ wrt to the strong order $\prec$.

Proof. We use the partial order $\prec$ on S_n to order the elements of $\{\sigma_1,\ \sigma_2,\ \dots,\ \sigma_m\}$. Once the largest and smallest elements are found, any element of the previous set belongs to a segment $[\sigma, \tau]_\prec$ where σ is a smallest element and τ is a largest element. Thus, $\{\sigma_1,\ \sigma_2,\ \dots,\ \sigma_m\}$ is always contained in the union of the segments $[\sigma, \tau]_\prec$. The first inclusion is obtained by considering their respective convex hulls. Conversely, $co([\sigma, \tau]_\prec)$ is always contained in $co(\{\sigma_1,\ \sigma_2,\ \dots,\ \sigma_m\})$ since $co([\sigma, \tau]_\prec) = co(\{\sigma, \tau\})$ and σ and τ are elements of $\{\sigma_1,\ \sigma_2,\ \dots,\ \sigma_m\}$. Thus,

$$\begin{aligned}
co(\{\sigma_1, \sigma_2, \dots, \sigma_m\}) &= co\left(\bigcup \left\{ co([\sigma, \tau]_\prec) \middle| \sigma \in S_{\substack{\min,\prec \\ 1 \leq i \leq m}}(\sigma_i) \text{ and } \tau \in S_{\substack{\max,\prec \\ 1 \leq i \leq m}}(\sigma_i) \right\} \right), \\
&= co\left(\bigcup \left\{ co(\{\sigma, \tau\}) \middle| \sigma \in S_{\substack{\min,\prec \\ 1 \leq i \leq m}}(\sigma_i) \text{ and } \tau \in S_{\substack{\max,\prec \\ 1 \leq i \leq m}}(\sigma_i) \right\} \right), \\
&= co\left(\bigcup \left\{ \left\langle \mathcal{A}_t(\tau \circ \sigma^{-1}) \right\rangle \sigma \middle| \sigma \in S_{\substack{\min,\prec \\ 1 \leq i \leq m}}(\sigma_i), \tau \in S_{\substack{\max,\prec \\ 1 \leq i \leq m}}(\sigma_i) \right\} \right).
\end{aligned}$$

□

Covering S_n using the Cayley metric.

The lowest element of the partially ordered set $(S_n, \prec)$ is id as it does not contain any inversion. The greatest element of $(S_n, \prec)$ is $n\,n-1\dots321$ as it contains all the possible inversions. We have:

$$\begin{aligned}
S_n &= [\mathrm{id}, n\,n-1\dots321]_\prec, &(4.49)\\
&= co\left(\{\mathrm{id}\} \cup \{c\}\right), &(4.50)
\end{aligned}$$

where $c = (c_1 c_2 \dots c_n)$ is a n-cycle of S_n. Indeed,

$$\begin{aligned} co\left(\{\mathrm{id}\} \cup \{c\}\right) &= co\left([\mathrm{id}, c]_T\right), \\ &= \left\{ g \circ \mathrm{id} \middle| g \in \left\langle \mathcal{A}_t(c \circ \mathrm{id}^{-1}) \right\rangle \right\}, \\ &= \left\langle (c_1 c_2), (c_2 c_3), \dots, (c_{n-1} c_n) \right\rangle, \\ &= S_n. \end{aligned}$$

Example 29. *The lowest and greatest elements of* $(S_3, \prec)$ *are respectively* id *and* 321*. We have:*

$$\begin{aligned} S_3 &= [\mathrm{id}, 321]_{\prec}, \\ &= co\left(\{\mathrm{id}, (123)\}\right), \\ &= co\left(\{\mathrm{id}, (132)\}\right). \end{aligned}$$

Any element of S_3 *can be obtained from* id *by gradually adding inversions through left transpositions. Moreover,* S_3 *can be recovered through (the convex hull) of* id *and a* 3*-cycle of* S_3*. Notice also that:*

$$\begin{aligned} co\left(\{\mathrm{id}, 321\}\right) &= \left\{ g \circ \mathrm{id} \middle| g \in \left\langle \mathcal{A}_t(321 \circ \mathrm{id}^{-1}) \right\rangle \right\}, \\ &= \left\langle (13) \right\rangle, \\ &\neq S_3. \end{aligned}$$

4.4.3 Ulam metric

The Ulam metric uses right adjacent cycles. We shall use subgroups of $(S_n, \circ)$ that are generated by adjacent cycles to determine convex sets of the metric space (S_n, UL).

Lemma 5. *Let* σ *and* τ *be two permutations of* S_n*, and let* $\mathcal{P}_{ac}(\sigma^{-1} \circ \tau)$ *be the set of all possible ways to write* $\sigma^{-1} \circ \tau$ *as a minimal product of adjacent cycles. An element* ν *of the segment* $[\sigma, \tau]_{UL}$ *is of the form:*

$$\nu = \sigma \circ \prod_{i=1}^{k} \tau_i, \tag{4.51}$$

where $k \leq UL(\sigma, \tau)$*,* $\prod_{i=1}^{0} \tau_i = \mathrm{id}$*, and* $\prod_{i=1}^{UL(\sigma,\tau)} \tau_i \in \mathcal{P}_{ac}(\sigma^{-1} \circ \tau)$*.*

Proof. Similar to that of Lemma 3 where minimal products of adjacent cycles yielding $\sigma^{-1} \circ \tau$ are considered instead of minimial products of adjacent transpositions. $\square$

Let us denote $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$ the set:

$$\{\text{id}\} \cup \bigcup_{k=1}^{UL(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \middle| \prod_{i=1}^{UL(\sigma,\tau)} \tau_i \in \mathcal{P}_{ac}(\sigma^{-1} \circ \tau) \right\}. \tag{4.52}$$

Proposition 7. *In the metric space (S_n, UL), the segment formed by σ and τ is given by:*

$$[\sigma, \tau]_{UL} = \{\sigma \circ a \mid a \in \mathcal{A}_{ac}(\sigma^{-1} \circ \tau)\}. \tag{4.53}$$

Proof. By Lemma 5, we have:

$$\begin{aligned} [\sigma, \tau]_{UL} &= \bigcup_{k=0}^{UL(\sigma,\tau)} \left\{ \sigma \circ \prod_{i=1}^{k} \tau_i \middle| k \leq UL(\sigma, \tau) \text{ and } \prod_{i=1}^{UL(\sigma,\tau)} \tau_i \in \mathcal{P}_{ac}(\sigma^{-1} \circ \tau) \right\}, \\ &= \{\sigma \circ a \mid a \in \mathcal{A}_{ac}(\sigma^{-1} \circ \tau)\}. \end{aligned}$$

□

Theorem 8. *In the metric space (S_n, UL), a segment $[\sigma, \tau]_{UL}$ is convex iff $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$ is a subgroup of $(S_n, \circ)$.*

Proof. Similar to the proof of Theorem 6. □

Let $\langle \mathcal{A}_{ac}(\sigma^{-1} \circ \tau) \rangle$ denote the smallest subgroup of $(S_n, \circ)$ containing the set $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$. We can see that this subgroup is generated by the adjacent cycles appearing in a factorization of $\sigma^{-1} \circ \tau$ into a product (with respect to $\circ$) of $UL(\sigma, \tau)$ adjacent cycles. We have the following result:

Corollary 7. *In the metric space (S_n, UL), the convex hull of two elements σ and τ is the left coset $\sigma\ \langle \mathcal{A}_{ac}(\sigma^{-1} \circ \tau) \rangle$.*

Proof. We start with the result of Proposition 7, namely:

$$[\sigma, \tau]_{UL} = \left\{ \sigma \circ a \middle| a \in \mathcal{A}_{ac}(\sigma^{-1} \circ \tau) \right\}.$$

Then, we use Theorem 8 to deduce that the smallest convex set containing the segment $[\sigma, \tau]_{UL}$ is obtained from the smallest group containing the set $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$. That is:

$$co([\sigma, \tau]_{UL}) = \left\{ \sigma \circ g \middle| g \in \left\langle \mathcal{A}_{ac}(\sigma^{-1} \circ \tau) \right\rangle \right\}. \tag{4.54}$$

The result follows as $co(\{\sigma, \tau\}) = co([\sigma, \tau]_{UL})$. □

Corollary 8. *In the metric space (S_n, UL), the convex hull of $m \geq 2$ elements $\sigma_1,\ \sigma_2,\ \ldots,$ and σ_m is the smallest convex set containing the union of all possible left coset $\sigma_i\ \left\langle \mathcal{A}_{ac}(\sigma_i^{-1} \circ \sigma_j \right\rangle$:*

$$co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\}) = co\left(\bigcup_{1 \leq i < j \leq m} \sigma_i\ \left\langle \mathcal{A}_{ac}(\sigma_i^{-1} \circ \sigma_j \right\rangle \right). \tag{4.55}$$

Proof. We have that:

$$\begin{aligned} co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\}) &= co\left(\bigcup_{1\le i<j\le m} co(\{\sigma_i, \sigma_j\})\right), \\ &= co\left(\bigcup_{1\le i<j\le m} \sigma_i \left\langle \mathcal{A}_{ac}(\sigma_i^{-1}\circ\sigma_j)\right\rangle\right). \end{aligned}$$

□

Covering S_n using the Ulam metric.

The Ulam metric does not induce a partial order on S_n in the same way the Kendall's τ and the Cayley metrics do. Nonetheless, we have that:

$$S_n = co(\{\sigma, (12\ldots n)\}), \tag{4.56}$$

where σ is an adjacent transposition. Indeed,

$$\begin{aligned} co(\{\sigma, (12\ldots n)\}) &= co([\sigma, (12\ldots n)]_{UL}), \\ &= \left\{\sigma \circ g \middle| g \in \left\langle \mathcal{A}_{ac}(\sigma^{-1}\circ(12\ldots n))\right\rangle\right\}, \\ &= \langle \sigma, (12\ldots n)\rangle, \\ &= S_n. \end{aligned}$$

4.4.4 Reversal distance

The reversal distance uses (left) reversals. We shall use subgroups of $(S_n, \circ)$ that are generated by reversals to determine convex sets of the metric space (S_n, R).

Lemma 6. *Let σ and τ be two permutations of S_n, and let $\mathcal{P}_r(\tau\circ\sigma^{-1})$ be the set of all possible ways to write $\tau\circ\sigma^{-1}$ as a minimal product of reversals. An element ν of the segment $[\sigma,\tau]_R$ is of the form:*

$$\nu = \prod_{i=1}^{k} \tau_i \circ \sigma, \tag{4.57}$$

where $k \le R(\sigma,\tau)$, $\prod_{i=1}^{0}\tau_i = \mathrm{id}$ and $\prod_{i=1}^{R(\sigma,\tau)}\tau_i \in \mathcal{P}_r(\tau\circ\sigma^{-1})$.

Proof. Similar to that of Lemma 3 where minimal products of reversals yielding $\tau\circ\sigma^{-1}$ are considered instead of minimial products of adjacent transpositions yielding $\sigma^{-1}\circ\tau$. □

Let us denote $\mathcal{A}_r(\tau \circ \sigma^{-1})$ the set:

$$\{\text{id}\} \cup \bigcup_{k=1}^{R(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \middle| \prod_{i=1}^{R(\sigma,\tau)} \tau_i \in \mathcal{P}_r(\tau \circ \sigma^{-1}) \right\}. \tag{4.58}$$

Proposition 8. *In the metric space,* (S_n, R) *the segment formed by* σ *and* τ *is given by:*

$$[\sigma, \tau]_R = \{a \circ \sigma \mid a \in \mathcal{A}_r(\tau \circ \sigma^{-1})\}. \tag{4.59}$$

Proof. By Lemma 6, we have:

$$\begin{aligned}
[\sigma, \tau]_R = & \bigcup_{k=0}^{R(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \circ \sigma \middle| k \leq R(\sigma, \tau) \text{ and } \prod_{i=1}^{R(\sigma,\tau)} \tau_i \in \mathcal{P}_r(\tau \circ \sigma^{-1}) \right\}, \\
= & \{a \circ \sigma \mid a \in \mathcal{A}_r(\tau \circ \sigma^{-1})\}.
\end{aligned}$$

□

Theorem 9. *In the metric space* (S_n, R)*, a segment* $[\sigma, \tau]_R$ *is convex iff* $\mathcal{A}_r(\tau \circ \sigma^{-1})$ *is a subgroup of* $(S_n, \circ)$*.*

Proof. Similar to that of Theorem 6. □

Let $\langle \mathcal{A}_r(\tau \circ \sigma^{-1}) \rangle$ denote the smallest subgroup of $(S_n, \circ)$ containing the set $\mathcal{A}_r(\tau \circ \sigma^{-1})$. We can see that this subgroup is generated by the reversals appearing in a factorization of $\tau \circ \sigma^{-1}$ into a product (with respect to $\circ$) of $R(\sigma, \tau)$ reversals. We have the following result:

Corollary 9. *In the metric space* (S_n, T)*, the convex hull of two elements* σ *and* τ *is the right coset* $\langle \mathcal{A}_r(\tau \circ \sigma^{-1}) \rangle\ \sigma$*.*

Proof. We start with the result of Proposition 8, namely:

$$[\sigma, \tau]_R = \left\{ a \circ \sigma \middle| a \in \mathcal{A}_r(\tau \circ \sigma^{-1}) \right\}.$$

Then, we use Theorem 9 to deduce that the smallest convex set containing the segment $[\sigma, \tau]_R$ is obtained from the smallest group containing the set $\mathcal{A}_r(\tau \circ \sigma^{-1})$. That is:

$$co([\sigma, \tau]_R) = \left\{ g \circ \sigma \middle| g \in \left\langle \mathcal{A}_r(\tau \circ \sigma^{-1}) \right\rangle \right\}. \tag{4.60}$$

The result follows as $co(\{\sigma, \tau\}) = co([\sigma, \tau]_R)$. □

Corollary 10. *In the metric space* (S_n, R)*, the convex hull of* $m \geq 2$ *elements* $\sigma_1,\ \sigma_2,\ \ldots,$ *and* σ_m *is the smallest convex set containing the union of all possible right coset* $\left\langle \mathcal{A}_r(\sigma_j \circ \sigma_i^{-1}) \right\rangle\ \sigma_i$*:*

$$co(\{\sigma_1, \sigma_2, \ldots, \sigma_m\}) = co\left(\bigcup_{1 \leq i < j \leq m} \left\langle \mathcal{A}_r(\sigma_j \circ \sigma_i^{-1}) \right\rangle\ \sigma_i \right). \tag{4.61}$$

Proof. We have that:

$$\begin{aligned} co(\{\sigma_1, \sigma_2, \dots, \sigma_m\}) &= co\left(\bigcup_{1\le i<j\le m} co(\{\sigma_i, \sigma_j\})\right), \\ &= co\left(\bigcup_{1\le i<j\le m} \left\langle \mathcal{A}_r(\sigma_j \circ \sigma_i^{-1})\right\rangle\ \sigma_i\right). \end{aligned}$$

□

Covering S_n using the reversal distance.

The reversal distance does not induce a partial order on S_n in the same way the Kendall's τ and the Cayley metrics do. Let

$$r_{i,j} = (i\, j)(i+1\, j-1)\dots\left(i+\left[\tfrac{j-i}{2}\right]\ j-\left[\tfrac{j-i}{2}\right]\right). \tag{4.62}$$

We have that:

$$S_n = co\left(\{\sigma, r_{1,n}\dots r_{1,3}r_{1,2}\,\sigma\}\right).$$

Indeed,

$$\begin{aligned} co\left(\{\sigma, r_{1,n}\dots r_{1,3}r_{1,2}\,\sigma\}\right) &= co\left([\sigma, r_{1,n}\dots r_{1,3}r_{1,2}\,\sigma]_R\right), \\ &= \left\langle \mathcal{A}_r(r_{1,n}\dots r_{1,3}r_{1,2}\,\sigma\circ\sigma^{-1})\right\rangle\ \sigma, \\ &= \langle \mathcal{A}_r(r_{1,n},\dots,r_{1,3},r_{1,2})\rangle\ \sigma, \\ &= \langle r_{1,n},\dots,r_{1,3},r_{1,2}\rangle\ \sigma, \\ &= S_n. \end{aligned}$$

4.5 Specification of the Analysis to Permutations

The runtime analysis of the CS on a quasi-concave landscape introduced in [MS17] and summarized in Chapter 3, is applied to permutations for the Kendall's τ, the Cayley, the Ulam metrics, and the reversal distance. We consider the set S_n of the permutations of the elements of $[n] = \{1, 2, \dots, n\}$.

We recall that $P^{\mathrm{Cov}}_{(S_n,D)}(m)$ denotes the probability that the convex hull (with respect to the metric D) of m elements sampled uniformly at random from S_n covers S_n:

$$P^{\mathrm{Cov}}_{(S_n,D)}(m) = \Pr[co(P') = S_n \mid P' = \mathrm{Unif}_m(S_n)]. \tag{4.63}$$

4.5.1 Kendall's τ metric

We saw in Section 4.4.1 that S_n can be recovered through the convex hull of id and $n\,n-1\ldots 321$ for the Kendall's τ metric. We shall use this result to estimate $P^{\mathrm{Cov}}_{(S_n,K)}(m)$.

Lemma 7. *For any convex set C of the metric space (S_n,K), we have $P^{\mathrm{Cov}}_{C}(m) \geq P^{\mathrm{Cov}}_{(S_n,K)}(m)$ where:*

$$P^{\mathrm{Cov}}_{(S_n,K)}(m) \geq 1 - 2\left(1-\frac{1}{n!}\right)^m. \tag{4.64}$$

Proof. The probability $P^{\mathrm{Cov}}_{(S_n,K)}(m)$ is bounded below by the probability for sampling id and $n\,n-1\ldots 321$ at least once from m trials.

Both id and $n\,n-1\ldots 321$ have probability $\frac{1}{n!}$ to be sampled. The probability that id or $n\,n-1\ldots 321$ is never sampled over the m trials is therefore $2\left(1-\frac{1}{n!}\right)^m$. Hence, the probability that they both appear at least once is:

$$1 - 2\left(1-\frac{1}{n!}\right)^m. \tag{4.65}$$

□

We apply Theorem 1 to a quasi-concave landscape on the metric space (S_n,K).

Theorem 10. *Let us consider a quasi-concave landscape on (S_n,K), whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. The CS with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1 - 3(q+1)\exp\left(-\frac{\mu r}{4n!}\right), \tag{4.66}$$

where $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$.

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{(S_n,K)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right)$.

$$\begin{aligned}
&\left[P^{\mathrm{Cov}}_{(S_n,K)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq \left[1 - 2\left(1-\frac{1}{n!}\right)^{\frac{\mu r}{4}}\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1 - 2(q+1)\left(1-\frac{1}{n!}\right)^{\frac{\mu r}{4}} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1 - 2(q+1)\exp\left(-\frac{\mu r}{4n!}\right) - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1 - [2(q+1)+q]\exp\left(-\frac{\mu r}{4n!}\right),\\
&\geq 1 - 3(q+1)\exp\left(-\frac{\mu r}{4n!}\right).
\end{aligned}$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x<0$. $\square$

Corollary 11. *Let us consider a quasi-concave landscape on* (S_n, K)*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The CS with population size:*

$$\mu \geq \frac{4n!}{r}\ln[6(q+1)], \tag{4.67}$$

finds a global optimum within at most q *generations and* μq *fitness evaluations with probability at least 0.5.*

Proof. The result follows from solving in μ the inequality:

$$1 - 3(q+1)\exp\left(-\frac{\mu r}{4n!}\right) \geq \frac{1}{2}.$$

$\square$

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 11, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5}=2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.5.2 Cayley metric

We saw in Section 4.4.2 that S_n can be recovered through the convex hull of id and a n-cycle for the Cayley metric. We shall use this result to estimate $P^{\mathrm{Cov}}_{(S_n,T)}(m)$.

Lemma 8. *For any convex set* C *of the metric space* (S_n,T)*, we have* $P^{\mathrm{Cov}}_C(m) \geq P^{\mathrm{Cov}}_{(S_n,T)}(m)$ *where:*

$$P^{\mathrm{Cov}}_{(S_n,T)}(m) \geq 1 - 2\left(1-\frac{1}{n!}\right)^m. \tag{4.68}$$

Proof. The probability $P^{\mathrm{Cov}}_{(S_n,T)}(m)$ is bounded below by the probability for sampling id and a n-cycle at least once from m trials.

The probability to sample id is $\frac{1}{n!}$. The probability to sample a n-cycle is $\frac{(n-1)!}{n!}$ because S_n has exactly $(n-1)!$ n-cycles.

The probability that id is never sampled over the m trials is $\left(1-\frac{1}{n!}\right)^m$. The probability that a n-cycle is never sampled over the m trials is $\left(1-\frac{1}{n}\right)^m$. Hence, the probability that id and all n-cycles are never sampled over the m trials is therefore:

$$\left(1-\frac{1}{n!}\right)^m + \left(1-\frac{1}{n}\right)^m. \tag{4.69}$$

Thus, the probability to sample id and a n-cycle at least once each is:

$$1 - \left(1-\frac{1}{n!}\right)^m - \left(1-\frac{1}{n}\right)^m \geq 1 - 2\left(1-\frac{1}{n!}\right)^m. \tag{4.70}$$

$\square$

We apply Theorem 1 to a quasi-concave landscape on the metric space (S_n, T).

Theorem 11. *Let us consider a quasi-concave landscape on* (S_n, T)*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. The CS with population size* μ *finds a global optimum within at most* q *generations and* μq *fitness evaluations with probability at least:*

$$1 - 3(q+1)\exp\left(-\frac{\mu r}{4n!}\right), \tag{4.71}$$

where $r = \min_{0 \leq j \leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*.*

Proof. Similar to that of Theorem 10. $\square$

Corollary 12. *Let us consider a quasi-concave landscape on* (S_n, T)*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0 \leq j \leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The CS with population size:*

$$\mu \geq \frac{4n!}{r} \ln[6(q+1)], \tag{4.72}$$

finds a global optimum within at most q *generations and* μq *fitness evaluations with probability at least 0.5.*

Proof. Similar to that of Corollary 11. $\square$

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 12, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.5.3 Ulam metric

We saw in Section 4.4.3 that S_n can be recovered through the convex hull of $(12\ldots n)$ and an adjacent transposition for the Ulam metric. We shall use this result to estimate $P^{\mathrm{Cov}}_{(S_n,UL)}(m)$.

Lemma 9. *For any convex set* C *of the metric space* (S_n, UL)*, we have* $P^{\mathrm{Cov}}_C(m) \geq P^{\mathrm{Cov}}_{(S_n,UL)}(m)$ *where:*

$$P^{\mathrm{Cov}}_{(S_n,UL)}(m) \geq 1 - 2\left(1 - \frac{1}{n!}\right)^m. \tag{4.73}$$

Proof. The probability $P^{\mathrm{Cov}}_{(S_n,UL)}(m)$ is bounded below by the probability for sampling $(12\ldots n)$ and an adjacent transposition at least once from m trials.

The probability to sample $(12\ldots n)$ is $\frac{1}{n!}$. The probability to sample an adjacent transposition is $\frac{n-1}{n!}$ because S_n has exactly $n-1$ adjacent transpositions.

The probability that $(12\ldots n)$ is never sampled over the m trials is $\left(1-\frac{1}{n!}\right)^m$. The probability that an adjacent transposition is never sampled over the m trials is $\left(1-\frac{n-1}{n!}\right)^m$. Hence, the probability that $(12\ldots n)$ and adjacent transpositions are never sampled over the m trials is therefore:

$$\left(1-\frac{1}{n!}\right)^m + \left(1-\frac{n-1}{n!}\right)^m. \tag{4.74}$$

Thus, the probability to sample $(12\ldots n)$ and an adjacent transposition at least once each is:

$$1-\left(1-\frac{1}{n!}\right)^m - \left(1-\frac{n-1}{n!}\right)^m \geq 1-2\left(1-\frac{1}{n!}\right)^m. \tag{4.75}$$

□

We apply Theorem 1 to a quasi-concave landscape on the metric space (S_n, UL).

Theorem 12. *Let us consider a quasi-concave landscape on (S_n, UL), whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. The CS with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1-3(q+1)\exp\left(-\frac{\mu r}{4n!}\right), \tag{4.76}$$

where $r=\min_{0\leq j\leq q}\frac{|A_{\geq j+1}|}{|A_{\geq j}|}$.

Proof. Similar to the proof of Theorem 10. □

Corollary 13. *Let us consider a quasi-concave landscape on (S_n, UL), whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r=\min_{0\leq j\leq q}\frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The CS with population size:*

$$\mu \geq \frac{4n!}{r}\ln[6(q+1)], \tag{4.77}$$

finds a global optimum within at most q generations and μq fitness evaluations with probability at least 0.5.

Proof. Similar to that of Corollary 11. □

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 13, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5}=2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.5.4 Reversal distance

We saw in Section 4.4.4 that S_n can be recovered through the convex hull of a permutation σ and the permutation $r_{1,n}\ldots r_{1,3}r_{1,2}\,\sigma$ for the reversal distance. We shall use this result to estimate $P^{\mathrm{Cov}}_{(S_n,R)}(m)$.

Lemma 10. *For any convex set C of the metric space (S_n, R), we have $P^{\mathrm{Cov}}_C(m) \geq P^{\mathrm{Cov}}_{(S_n,R)}(m)$ where:*

$$P^{\mathrm{Cov}}_{(S_n,R)}(m) \geq 1 - 2\left(1 - \frac{1}{n!}\right)^m. \tag{4.78}$$

Proof. The probability $P^{\mathrm{Cov}}_{(S_n,R)}(m)$ is bounded below by the probability for sampling a permutation σ and the permutation $r_{1,n}\ldots r_{1,3}r_{1,2}\,\sigma$ at least once from m trials.

Both permutations have probability $\frac{1}{n!}$ to be sampled. The rest of the proof is similar to that of Lemma 7. □

We apply Theorem 1 to a quasi-concave landscape on the metric space (S_n, R).

Theorem 13. *Let us consider a quasi-concave landscape on (S_n, R), whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. The CS with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1 - 3(q+1)\exp\left(-\frac{\mu r}{4n!}\right), \tag{4.79}$$

where $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$.

Proof. Similar to that of Theorem 10. □

Corollary 14. *Let us consider a quasi-concave landscape on (S_n, R), whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The CS with population size:*

$$\mu \geq \frac{4n!}{r}\ln[6(q+1)], \tag{4.80}$$

finds a global optimum within at most q generations and μq fitness evaluations with probability at least 0.5.

Proof. Similar to that of Corollary 11. □

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 14, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.6 Schemata for Permutations

The notion of schemata [H+92] is used to define the convex canonical level sets of a quasi-concave landscape in [MS17]. While the schemata used in [MS17] are focused on strings on a finite alphabet, the schemata defined below focus on permutations.

We recall that a permutation of S_n is a reordering of the elements $1,\ 2,\ \ldots,\ n$. Hence, a permutation σ can be seen as the list of pairs of distinct positions (i,j) where the order of the values $\sigma(i)$ and $\sigma(j)$ differs from the order of i and j. We say that an *inversion* occurred between the positions i and j.

Let $\mathbb{1}_{\mathrm{Inv}}(i,j)$ denote the indicator function of an inversion occuring between the positions i and j. That is:

$$\mathbb{1}_{\mathrm{Inv}}(i,j) = \begin{cases} 0 \text{ if no inversion occurs,} \\ 1 \text{ otherwise.} \end{cases} \tag{4.81}$$

Definition 18. *The matrix $\mathcal{M}_{\mathrm{Inv}}$ of the inversions of a permutation of S_n is a $(n-1)\times(n-1)$ upper triangular matrix such that:*

$$\mathcal{M}_{\mathrm{Inv}}(i,j-1) = \mathbb{1}_{\mathrm{Inv}}(i,j), \tag{4.82}$$

for $0 \le i \le n-2$ and $1 \le j \le n-1$.

Example 30. *The permutation 24315 of S_5 can be written as:*

$$\begin{matrix} \mathbb{1}_{\mathrm{Inv}}(0,1) & \mathbb{1}_{\mathrm{Inv}}(0,2) & \mathbb{1}_{\mathrm{Inv}}(0,3) & \mathbb{1}_{\mathrm{Inv}}(0,4) \\ & \mathbb{1}_{\mathrm{Inv}}(1,2) & \mathbb{1}_{\mathrm{Inv}}(1,3) & \mathbb{1}_{\mathrm{Inv}}(1,4) \\ & & \mathbb{1}_{\mathrm{Inv}}(2,3) & \mathbb{1}_{\mathrm{Inv}}(2,4) \\ & & & \mathbb{1}_{\mathrm{Inv}}(3,4) \end{matrix} = \begin{matrix} 0 & 0 & 1 & 0 \\ & 1 & 1 & 0 \\ & & 1 & 0 \\ & & & 0 \end{matrix} \tag{4.83}$$

Definition 19. *A mapping ψ from S_n to $\{0,1,\ldots,n-1\}\times\{0,1,\ldots,n-2\}\times\ldots\times\{0,1\}$ is defined as follows:*

$$\begin{aligned} \psi : S_n &\longrightarrow \{0,1,\ldots,n-1\}\times\{0,1,\ldots,n-2\}\times\ldots\times\{0,1\} \\ \sigma &\longmapsto \left(\sum_{j=0}^{n-2}\mathcal{M}_{\mathrm{Inv}}(0,j), \sum_{j=1}^{n-2}\mathcal{M}_{\mathrm{Inv}}(1,j), \ldots, \mathcal{M}_{\mathrm{Inv}}(n-2,n-2)\right) \end{aligned}$$

Example 31. *We have:*

$$\psi(24315) = (1,2,1,0).$$

Lemma 11. *The mapping ψ is a bijection.*

Proof. As the size of the domain and the codomain of ψ are both equal to $n!$, it remains to show that ψ is injective. That is $\psi(\sigma)=\psi(\sigma')$ implies that $\sigma=\sigma'$. We use a proof by induction on $n \ge 2$.

- For $n=2$, we have $\psi(12)=0$ and $\psi(21)=1$. Hence, ψ is injective.
- Similarly, ψ is injective for $n=3$ as:
  $\psi(123)=(0,0),\ \psi(213)=(1,0),\ \psi(132)=(0,1),\ \psi(231)=(1,1),$
  $\psi(312)=(2,0),\ \psi(321)=(2,1).$

- We assume that ψ is injective for n.

- Let us show that ψ remains injective for $n+1$. If $\psi(\sigma) = \psi(\sigma')$, then:

$$\begin{aligned}
&\left(\sum_{j=0}^{n-1} \mathcal{M}_{\text{Inv}}(0,j), \sum_{j=1}^{n-1} \mathcal{M}_{\text{Inv}}(1,j), \ldots, \mathcal{M}_{\text{Inv}}(n-1,n-1)\right) \\
&= \left(\sum_{j=0}^{n-1} \mathcal{M'}_{\text{Inv}}(0,j), \sum_{j=1}^{n-1} \mathcal{M'}_{\text{Inv}}(1,j), \ldots, \mathcal{M'}_{\text{Inv}}(n-1,n-1)\right).
\end{aligned}$$

That is:

$$\begin{aligned}
&\left(\sum_{\substack{j=0 \\ j\neq n-2}}^{n-1} \mathcal{M}_{\text{Inv}}(0,j), \sum_{\substack{j=1 \\ j\neq n-2}}^{n-2} \mathcal{M}_{\text{Inv}}(1,j), \ldots, 0, \mathcal{M}_{\text{Inv}}(n-1,n-1)\right) \\
&+ \left(\mathcal{M}_{\text{Inv}}(0,n-2), \mathcal{M}_{\text{Inv}}(1,n-2), \ldots, \sum_{j=n-2}^{n-1} \mathcal{M}_{\text{Inv}}(n-2,j), 0\right) \\
&= \left(\sum_{\substack{j=0 \\ j\neq n-2}}^{n-1} \mathcal{M'}_{\text{Inv}}(0,j), \sum_{\substack{j=1 \\ j\neq n-2}}^{n-2} \mathcal{M'}_{\text{Inv}}(1,j), \ldots, 0, \mathcal{M'}_{\text{Inv}}(n-1,n-1)\right) \\
&+ \left(\mathcal{M'}_{\text{Inv}}(0,n-2), \mathcal{M'}_{\text{Inv}}(1,n-2), \ldots, \sum_{j=n-2}^{n-1} \mathcal{M'}_{\text{Inv}}(n-2,j), 0\right).
\end{aligned}$$

Let ν_{n-1} (resp. ν'_{n-1}) denote the subsequence of length n that is obtained from σ (resp. σ') by deleting the position $n-1$. Let also τ_j (resp. τ'_j) denote the subsequence of length 2 that is obtained from σ (resp. σ') by only keeping the positions j and n, where $0 \leq j \leq n-1$. We have:

$$0 = \psi(\tau'_{n-1}) - \psi(\tau_{n-1}). \tag{4.84}$$

By assumption, ψ is injective for permutations of length less than or equal to n. Therefore, $\tau_{n-1} = \tau'_{n-1}$.

Let now ν_{n-2} (resp. ν'_{n-2}) denote the subsequence of length n that is obtained from σ (resp. σ') by deleting the position $n-2$. By a similar reasoning as above we find that $\tau_{n-2} = \tau'_{n-2}$.

More generally, by considering the subsequence ν_j (resp. ν'_j) of length n that is obtained from σ (resp. σ') by deleting the position j and using the same reasoning as above we find that $\tau_j = \tau'_j$ for $0 \leq j \leq n-1$.

As the permutations σ and σ' share the same subsequence of length 2 when only positions j and $n-1$ are kept for $0 \leq j \leq n-1$, then they are necessarily equal. Consequently, the mapping ψ is injective for permutations of length

$n+1$.

Thus, ψ is always injective.

As a result, ψ is a bijection. $\square$

Consequently, a permutation of S_n is uniquely determined by the $(n-1)$-uplet corresponding the sum of the lines of its inversion matrix. This result gives a template for the permutations of $\psi(S_n)$.

Theorem 14. *A schema corresponding to $\psi(S_n)$ is a template with $n-1$ positions, where the admissible values at position $n-i$ are: $0,\ 1,\ \ldots,\ i-1$. That is,*

$$\psi(S_n) = *_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,1]}, \tag{4.85}$$

where $[0,i]$ denotes the set $\{0,\ 1,\ \ldots,\ i\}$.

Proof. By Lemma 11, a permutation of S_n can be identified to the $(n-1)$-uplet of the sums of the rows of its matrix of inversions $\mathcal{M}_{\mathrm{Inv}}$ through ψ. The rows of $\mathcal{M}_{\mathrm{Inv}}$ have respectively: $1,\ 2,\ \cdots,\ n-1$ entries. Therefore, the sum of the row with i entrie(s) is at least 0 and at most i. $\square$

4.7 Permutations as strings of the Hamming and Manhattan spaces

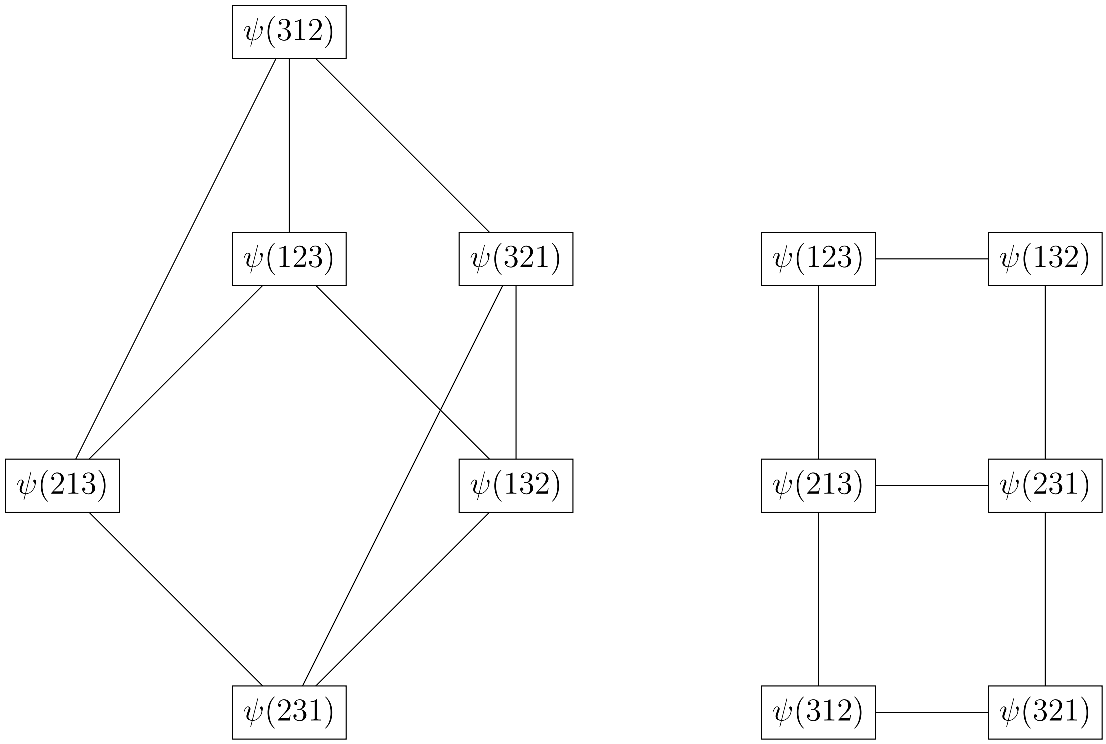
A permutation of S_n is uniquely determined by a $(n-1)$-uplet of $\{0,1,\ldots,n-1\} \times \{0,1,\ldots,n-2\} \times \ldots \times \{0,1\}$, which is a strict subset of $\{0,1,\ldots,n-1\}^{n-1}$. Hence, one can work on the metric spaces:

- $(\{0,1,\ldots,n-1\} \times \{0,1,\ldots,n-2\} \times \ldots \times \{0,1\}, \mathrm{HD})$,
- $(\{0,1,\ldots,n-1\} \times \{0,1,\ldots,n-2\} \times \ldots \times \{0,1\}, \mathrm{MD})$.

Example 32. *Let us consider $S_3 = \{123, 213, 132, 231, 312, 321\}$. We have that: $\psi(123) = (0,0)$, $\psi(213) = (1,0)$, $\psi(132) = (0,1)$, $\psi(231) = (1,1)$, $\psi(312) = (2,0)$, $\psi(321) = (2,1)$. To ease the notation, each 2-uplet (a,b) will simply be written as ab. We have:*

$$\begin{aligned}\{00,\ 10,\ 01,\ 11,\ 20,\ 21\} &= \{0,1,2\} \times \{0,1\},\\ &= \{ab \mid ab \in \{0,1,2\} \times \{0,1\}\},\\ &= \{\psi(\sigma) \mid \sigma \in S_3\},\\ &= \psi(S_3).\end{aligned}$$

The set S_3 can therefore be seen as the set $\{0,1,2\} \times \{0,1\}$. The latter can be endowed with the Hamming (resp. Manhattan) distance. See Figure 4.3.



$(\psi(S_3), \mathrm{HD}) = (\{0,1,2\} \times \{0,1\}, \mathrm{HD})$ $\qquad$ $(\psi(S_3), \mathrm{MD}) = (\{0,1,2\} \times \{0,1\}, \mathrm{MD})$

Figure 4.3: Illustration of the metric space $(\psi(S_3), D)$ for different metrics D. An edge is drawn between two strings whose distance from each other is one.

By studying $(\psi(S_n), \mathrm{HD})$ (resp. $(\psi(S_n), \mathrm{MD})$), we bring the problem back to strings of the Hamming (resp. Manhattan) space. Indeed, the case of $M_{d,\mathrm{HD}}$ (resp. $M_{d,\mathrm{MD}}$) has already been dealt with in Section 3.6 of Chapter 3.

We shall now analyse the runtime of the CS on a quasi-concave landscape of permutations, by seeing each permutation of S_n as the string of $\{0, 1, \ldots, n-1\} \times \{0, 1, \ldots, n-2\} \times \ldots \times \{0, 1\}$ that uniquely determines it. Two metrics will be considered: the Hamming and the Manhattan distance.

We recall that $P^{\mathrm{Cov}}_{(\psi(S_n),D)}(m)$ denotes the probability that the convex hull (with respect to the metric D) of m elements sampled uniformly at random from $\psi(S_n)$ covers $\psi(S_n)$:

$$P^{\mathrm{Cov}}_{(\psi(S_n),D)}(m) = \Pr[co(P') = \psi(S_n) \mid P' = \mathrm{Unif}_m(\psi(S_n))]. \tag{4.86}$$

The schema corresponding to $\psi(S_n)$ is $* *_{[0,n-2]} \cdots *_{[0,1]}$. Therefore, $P^{\mathrm{Cov}}_{(\psi(S_n),D)}(m)$ is the probability for obtaining the schema $* *_{[0,n-2]} \cdots *_{[0,1]}$ from the convex hull of m elements (sampled uniformly at random) from $\psi(S_n)$. To ease the notation, we will simply write $P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ instead of $P^{\mathrm{Cov}}_{(\psi(S_n),D)}(m)$. The metric D will be specified by the context.

4.7.1 Hamming distance

We start by noticing that all schemata correspond to a convex set in the metric space $(\psi(S_n), \mathrm{HD})$.

Corollary 15. *Any schema in the metric space* $(\psi(S_n), \mathrm{HD})$ *is a convex set.*

Proof. Similar to that of Corollary 23 by noticing that the metric space $(\psi(S_n), \mathrm{HD})$ is contained in the metric space $(\{0, 1, \ldots, n-1\}^{n-1}, \mathrm{HD})$. $\square$

We estimate $P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ which is a lower bound on the probability for covering a convex set C of $(\psi(S_n), \mathrm{HD})$ with m samples from C.

Lemma 12. *For any convex set* C *of the metric space* $(\psi(S_n), \mathrm{HD})$*, we have* $P^{\mathrm{Cov}}_{C}(m) \geq P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ *where:*

$$P^{\mathrm{Cov}}_{\psi(S_n)}(m) \geq 1 - n^2 \left(1 - \frac{1}{n}\right)^m . \tag{4.87}$$

Proof. We saw in Corollary 15 that any schema corresponds to a convex set in the metric space $(\psi(S_n), \mathrm{HD})$. In particular, the schema corresponding to the entire search space is the only schema with the largest number of positions that are free to take more than one value. Moreover, each of these free positions take the maximum number of possible values. Therefore, the schema corresponding to any other convex set has at most $n-i$ free positions as position i (counting starts at 1).

Let us now compute the probability for covering the entire search space from sampling m points from it. The schema corresponding to the entire search space is $* *_{[0,n-2]} \cdots *_{[0,1]}$.

The symbol $*_{[0,i]}$ is obtained at position $n-i$, when each of the values $0,\ 1,\ \ldots,\ i$ appears at least once at this position. The probability that a value appears at this position is $\frac{1}{n}$. The probability that this value never appears at this position is therefore $1 - \frac{1}{n}$. The probability that this value never appears at this position in m trials is therefore:

$$\left(1 - \frac{1}{n}\right)^m . \tag{4.88}$$

The probability that the value 0 never appears at this position OR the value 1 OR ... OR the value i is:

$$(i+1)\left(1 - \frac{1}{n}\right)^m . \tag{4.89}$$

Hence, the probability that each value appears at least once at that position is:

$$1 - (i+1)\left(1 - \frac{1}{n}\right)^m . \tag{4.90}$$

Thus, the probability for obtaining the schema $* *_{[0,n-2]} \cdots *_{[0,1]}$ is:

$$\prod_{i=1}^{n-1}\left[1-(i+1)\left(1-\frac{1}{n}\right)^m\right] \geq \left[1-n\left(1-\frac{1}{n}\right)^m\right]^{n-1}, \tag{4.91}$$

$$\geq 1-(n-1)n\left(1-\frac{1}{n}\right)^m, \tag{4.92}$$

$$\geq 1-n^2\left(1-\frac{1}{n}\right)^m, \tag{4.93}$$

using Bernoulli's inequality in (4.92). The result follows. □

We apply Theorem 1 to a quasi-concave landscape on the metric space $(\psi(S_n), \mathrm{HD})$, whose parameters r and q are defined as in Section 3.4.

Theorem 15. *Let us consider a quasi-concave landscape on* $(\psi(S_n), \mathrm{HD})$*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The CS with population size* μ *finds a global optimum within at most* q *generations and* μq *fitness evaluations with probability at least:*

$$1-2(q+1)n^2\exp\left(-\frac{\mu r}{4n}\right). \tag{4.94}$$

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{\psi(S_n)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right)$.

$$\begin{aligned}
&\left[P^{\mathrm{Cov}}_{\psi(S_n)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq \left[1-n^2\left(1-\frac{1}{n}\right)^{\frac{\mu r}{4}}\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1-(q+1)n^2\left(1-\frac{1}{n}\right)^{\frac{\mu r}{4}} - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1-(q+1)n^2\exp\left(-\frac{\mu r}{4n}\right) - q\exp\left(-\frac{9\mu r}{32}\right),\\
&\geq 1-[(q+1)n^2+q]\exp\left(-\frac{\mu r}{4n}\right),\\
&\geq 1-(q+1)(n^2+1)\exp\left(-\frac{\mu r}{4n}\right),\\
&\geq 1-2(q+1)n^2\exp\left(-\frac{\mu r}{4n}\right).
\end{aligned}$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x < 0$. □

Corollary 16. *Let us consider a quasi-concave landscape on* $(\psi(S_n), \mathrm{HD})$*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The CS with population size:*

$$\mu \geq \frac{8n}{r}\ln\left(2n\sqrt{q+1}\right), \tag{4.95}$$

finds a global optimum within at most q generations and μq fitness evaluations with probability at least 0.5.

Proof. The result follows from solving in μ the inequality:

$$1 - 2(q+1)n^2 \exp\left(-\frac{\mu r}{4n}\right) \geq \frac{1}{2}.$$

$\square$

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 16, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.7.2 Manhattan distance

We start by finding all schemata corresponding to a convex set in the metric space $(\psi(S_n), \mathrm{MD})$.

Corollary 17. *Let $[k,l]$ denote the set $\{k, k+1, \ldots, l-1, l\}$. The only convex schemata of the metric space $(\psi(S_n), \mathrm{MD})$ are those that use symbols $*_{[k,l]}$ such that $[k,l] \subseteq [0,i]$ at position $n-i$, for $1 \leq i \leq n-1$.*

Proof. Similar to that of Corollary 24 by noticing that the metric space $(\psi(S_n), \mathrm{MD})$ is contained in the metric space $(\{0, 1, \cdots, n-1\}^{n-1}, \mathrm{MD})$. $\square$

We estimate $P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ which is a lower bound on the probability for covering a convex set C of $(\psi(S_n), \mathrm{MD})$ with m samples from C.

Lemma 13. *For any convex set C of the metric space $(\psi(S_n), \mathrm{MD})$, we have $P^{\mathrm{Cov}}_{C}(m) \geq P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ where:*

$$P^{\mathrm{Cov}}_{\psi(S_n)}(m) \geq 1 - 2(n-1)\left(1 - \frac{1}{n}\right)^m. \tag{4.96}$$

Proof. We saw in Corollary 17 that schemata corresponding to convex sets only use a symbol $*_{[k,l]}$ such that $[k,l] \subseteq [0, n-i]$ at position i (counting starts at 1). In particular, the schema corresponding to the entire search space is the only schema with the largest number of free positions. Moreover, each of these free positions takes the maximum number of possible values.

Let us now compute the probability for covering the entire search space from sampling m points from it. The schema corresponding to the entire search space is $*_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,1]}$.

The symbol $*_{[0,n-i]}$ is obtained at position i, when each of the values 0 and $n-i$ appears at least once at this position. The probability that a value appears at this position is $\frac{1}{n}$. The probability that this value never appears at this position is therefore $1-\frac{1}{n}$. The probability that this value never appears at this position in m trials is therefore:

$$\left(1-\frac{1}{n}\right)^m. \tag{4.97}$$

The probability that the value 0 never appears at this position OR the value $n-i$ never appears at this position is:

$$2\left(1-\frac{1}{n}\right)^m. \tag{4.98}$$

Hence, the probability that each value appears at least once at that position is:

$$1-2\left(1-\frac{1}{n}\right)^m. \tag{4.99}$$

Thus, the probability for obtaining the schema $* *_{[0,n-2]} \cdots *_{[0,1]}$ is:

$$\prod_{i=1}^{n-1}\left[1-2\left(1-\frac{1}{n}\right)^m\right] \geq \left[1-2\left(1-\frac{1}{n}\right)^m\right]^{n-1}, \tag{4.100}$$

$$\geq 1-2(n-1)\left(1-\frac{1}{n}\right)^m, \tag{4.101}$$

using Bernoulli's inequality in the last line. The result follows. □

We apply Theorem 1 to a quasi-concave landscape on the metric space $(\psi(S_n), \mathrm{MD})$, whose parameters r and q are defined as in Section 3.4.

Theorem 16. *Let us consider a quasi-concave landscape on $(\psi(S_n), \mathrm{MD})$, whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The CS with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1-2n(q+1)\exp\left(-\frac{\mu r}{4n}\right). \tag{4.102}$$

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{\psi(S_n)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right)$.

$$
\begin{aligned}
&\left[P^{\mathrm{Cov}}_{\psi(S_n)}\left(\frac{\mu r}{4}\right)\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right), \\
&\geq \left[1 - 2(n-1)\left(1-\frac{1}{n}\right)^{\frac{\mu r}{4}}\right]^{q+1} - q\exp\left(-\frac{9\mu r}{32}\right), \\
&\geq 1 - 2(n-1)(q+1)\left(1-\frac{1}{n}\right)^{\frac{\mu r}{4}} - q\exp\left(-\frac{9\mu r}{32}\right), \\
&\geq 1 - 2(n-1)(q+1)\exp\left(-\frac{\mu r}{4n}\right) - q\exp\left(-\frac{9\mu r}{32}\right), \\
&\geq 1 - [2(n-1)(q+1)+q]\exp\left(-\frac{\mu r}{4n}\right), \\
&\geq 1 - [2(n-1)+1](q+1)\exp\left(-\frac{\mu r}{4n}\right), \\
&\geq 1 - 2n(q+1)\exp\left(-\frac{\mu r}{4n}\right).
\end{aligned}
$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x < 0$. □

Corollary 18. *Let us consider a quasi-concave landscape on* $(\psi(S_n), \mathrm{MD})$*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The CS with population size:*

$$
\mu \geq \frac{2n}{r}\ln[8n(q+1)], \tag{4.103}
$$

finds a global optimum within at most q *generations and* μq *fitness evaluations with probability at least 0.5.*

Proof. The result follows from solving in μ the inequality:

$$
1 - 2n(q+1)\exp\left(-\frac{\mu r}{4n}\right) \geq \frac{1}{2}.
$$

□

Let one run of the CS be performed in q generations. If the population size satisfies the condition of Corollary 18, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

4.8 Application Example

Let us consider the metric space (S_n, K) and a fixed minimal product $\prod_{i=1}^{l}\tau_i$ of $l = \frac{n(n-1)}{2}$ adjacent transpositions yielding the permutation $n\ n-1\ \cdots\ 321$.

Example 33. *For $n = 4$, we have:*

$$4321 = (12)(23)(34)(23)(12)(23). \tag{4.104}$$

This is one of the many ways to write 4321 as the minimal product of $6 = \frac{4(4-1)}{2}$ adjacent transpositions.

Let us call each adjacent transposition τ_i a *letter* and let the product $\prod_{i=1}^{l} \tau_i$ be a *word*. A *suffix* is obtained from the word $\prod_{i=1}^{l} \tau_i$ by deleting its first $k \leq l$ letters.

Example 34. $(23)(12)(23)$ *is a suffix of* $(12)(23)(34)(23)(12)(23)$*, that is obtained by deleting its first three letters* $(12)(23)(34)$*.*

We shall now define a fitness function on the set of permutations S_n. First, we fix a minimal product $\mathcal{P} = \prod_{i=1}^{l} \tau_i$ of $l = \frac{n(n-1)}{2}$ adjacent transpositions yielding the permutation $n\ n-1\ \cdots\ 321$. Then, the fitness $SX_{\mathcal{P}}(\nu)$ of a permutation ν is given by the length of the longest suffix of $\mathcal{P}$ that appears as a suffix in a writing of ν as a minimal product of adjacent transpositions.

Example 35. *In S_4, let us consider the minimal product $\mathcal{P} = (12)(23)(34)(23)(12)(23)$ and let $\nu = 3142$. The permutation ν has exactly two distinct writings as a minimal product of adjacent transpositions:*

$$\begin{aligned} \nu &= (34)(12)(23), \\ &= (12)(34)(23). \end{aligned}$$

The longest suffix of $\mathcal{P} = (12)(23)(34)(23)(12)(23)$ appearing as a suffix in a writing of ν as a minimal product of adjacent transpositions is therefore $(12)(23)$. It appears in the first writing. As a result, the fitness of ν is $SX_{\mathcal{P}}(\nu) = 2$.

Proposition 9. *Let $\mathcal{P}$ be a fixed minimal product of $l = \frac{n(n-1)}{2}$ adjacent transpositions yielding the permutation $n\ n-1\ \cdots\ 21$. The fitness landscape $(S_n, SX_{\mathcal{P}}, K)$ is quasi-concave with parameters $r \geq \frac{1}{n}$ and* $q = \begin{cases} \frac{l}{2} - 1 & \text{if } l \text{ is even,} \\ \frac{l-1}{2} - 1 & \text{if } l \text{ is odd.} \end{cases}$

Proof. It is enough to define the canonical level sets of the problem and to show that they are convex sets in the metric space (S_n, K). Let $\prod_{i=1}^{l} \tau_i$ denote the fixed minimal product $\mathcal{P}$. The possible lengths of a suffix of the word $\prod_{i=1}^{l} \tau_i$ are: $l,\ l-1,\ \ldots,\ 2,\ 1,$ and 0 when the suffix is the identity $12 \cdots n = \mathrm{id}$. A permutation of S_n can be written as a minimial product of either: $0,\ 1,\ \ldots$, or l adjacent transpositions. The identity permutation can not be written as a product of adjacent transpositions.

Let $A_{\geq j}$ be the canonical level set containing all permutations whose fitness value is at least j. This means that an element of $A_{\geq j}$ that can be written as a minimal product of k_0 adjacent transpositions can be written as:

$$\tau'_{l-k_0+1}\cdots\tau'_{l-j}\tau_{l-j+1}\cdots\tau_{l-1}\tau_l, \tag{4.105}$$

where $j \leq k_0 \leq l$. Let σ denote the permutation $n\ n-1\ \cdots 21$. We have the following results:

$$\begin{aligned}
A_{\geq 0} &= [\mathrm{id}, \sigma]_K \\
A_{\geq 1} &= [\tau_l, \tau_1 \circ \sigma]_K \\
A_{\geq 2} &= [\tau_{l-1}\tau_l, \tau_2\tau_1 \circ \sigma]_K \\
&\cdots \\
A_{\geq k} &= [\tau_{l-k+1}\cdots\tau_{l-1}\tau_l, \tau_k\cdots\tau_2\tau_1 \circ \sigma]_K \\
&\cdots
\end{aligned}$$

If l is even then $k \leq \frac{l-2}{2}$. Otherwise, $k \leq \frac{l-3}{2}$ if l is odd. Each canonical level set is a segment of the metric space (S_n, K), built in such a way that one extreme is always smaller than the other with respect to the weak left order $\prec_w$. By Corollary 1, this implies that the canonical level sets are convex sets of the metric space (S_n, K). Moreover, $A_{\geq j+1}$ is always contained in $A_{\geq j}$ by construction:

$$\tau_{l-j+1}\cdots\tau_{l-1}\tau_l \prec_w \tau_{l-(j+1)+1}\cdots\tau_{l-1}\tau_l, \tag{4.106}$$

$$\tau_{j+1}\cdots\tau_2\tau_1 \circ \sigma \prec_w \tau_j\cdots\tau_2\tau_1 \circ \sigma. \tag{4.107}$$

The number $q+1$ of distinct level sets is therefore $\frac{l}{2}$ if l is even and $\frac{l-1}{2}$ if l is odd. The number of distinct level sets is always polynomial in n. The smallest ratio r between the sizes of two consecutive canonical level sets is still an open problem. Indeed, a closed form for the cardinality of a segment of length k of (S_n, K) is not known yet. We shall therefore estimate a lower bound on r instead. Let us consider the level sets $A_{\geq k}$ and $A_{\geq k-1}$. The length of the segment $[\tau_{l-k+1}\cdots\tau_{l-1}\tau_l, \tau_k\cdots\tau_2\tau_1 \circ \sigma]_K = A_{\geq k}$ is denoted l_k. By construction, the length of the segment $[\tau_{l-k+2}\cdots\tau_{l-1}\tau_l, \tau_{k-1}\cdots\tau_2\tau_1 \circ \sigma]_K = A_{\geq k-1}$ is l_k+2. The smallest number of points that need to be added to $A_{\geq k}$ to obtain $A_{\geq k-1}$ is two. The largest number of points that need to be added to $A_{\geq k}$ to obtain $A_{\geq k-1}$ is bounded above

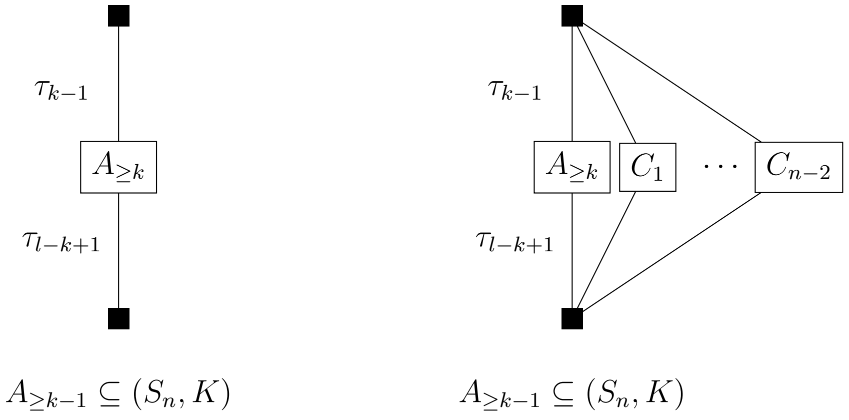


Figure 4.4: $A_{\geq k-1}$ is obtained from $A_{\geq k}$ by adding two extra nodes. The extremes of the possible structures of $A_{\geq k-1}$ are illustrated above, where C_i is a convex segment with the same length as the segment yielding $A_{\geq k}$. There are at most $n-1$ distinct adjacent transpositions that can be used in S_n. One of them is already used to connect one of the extra nodes to $A_{\geq k}$. Hence, there are at most $n-2$ other possible connections from that extra node left.

by $(n-2) \mid A_{\geq k} \mid +2$. An illustration is given in Figure 4.4. Hence, we have:

$$r = \min_{0 \leq k \leq q} \frac{\mid A_{\geq k} \mid}{\mid A_{\geq k-1} \mid}, \tag{4.108}$$

$$\geq \min_{0 \leq k \leq q} \frac{\mid A_{\geq k} \mid}{(n-2) \mid A_{\geq k} \mid +2}, \tag{4.109}$$

$$\geq \min_{X > 1} \frac{X}{(n-2)X + 2}, \tag{4.110}$$

$$\geq \frac{1}{n}. \tag{4.111}$$

$\square$

Remark 1. *Segments need not be convex sets in the metric spaces:* (S_n, T), (S_n, UL) *and* (S_n, R). *Therefore, segments can not be used as canonical level sets in these metric spaces.*

Theorem 17. *In the metric space* (S_n, K) *where* $\frac{n(n-1)}{2}$ *is even, if the population size is at least:*

$$\mu \geq 4n!n \ln \left[\frac{3n(n-1)}{2} \right], \tag{4.112}$$

then the CS finds the fixed minimal product of $\frac{n(n-1)}{2}$ *adjacent transpositions (yielding the permutation* $n\ n-1\ \cdots 21$*) with probability at least* 0.5*, while searching for longest common suffixes with the fixed minimal product of* $\frac{n(n-1)}{2}$ *adjacent transpositions.*

Proof. By Proposition 9 and Corollary 11. $\square$

Let one run of the CS be performed in $q = \frac{n(n-1)}{4} - 1$ generations. If the population size satisfies the condition of Theorem 17, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q = \frac{n(n-1)}{2} - 2$ and $2\mu q = O\left(n! n^3 \ln\left[2n\right]\right)$.

4.8.1 From Permutations to Strings

Using the bijection $\psi : S_n \longrightarrow \{0, 1, \ldots, n-1\} \times \{0, 1, \ldots, n-2\} \times \ldots \times \{0, 1\}$ of Definition 19, we bring the problem of permutations to strings. We shall now consider the metric space $(\psi(S_n), \mathrm{HD})$ and a fixed string $a_1 a_2 \cdots a_{n-1}$ of $\psi(S_n)$.

Example 36. *In $\psi(S_4) = \{0, 1, 2, 3\} \times \{0, 1, 2\} \times \{0, 1\}$, 201 is a fixed string.*

The fitness $SX_a(b)$ of a string b is given by the length of the longest suffix of a that is also a suffix of b.

Example 37. *In $\psi(S_5)$, let $a = 2311$ and let $b = 4321$. The longest suffix of a that is also a suffix of b is 1. Hence, the fitness $SX_a(b)$ of b is 1.*

Proposition 10. *Let a be a fixed string of $\psi(S_n)$. The fitness landscape $(\psi(S_n), SX_a, \mathrm{HD})$ (resp. $(\psi(S_n), SX_a, \mathrm{MD})$) is quasi-concave with parameters $q = n-1$ and $r = \frac{1}{n+1}$.*

Proof. It is enough to define the canonical level sets of the problem and to show that they are convex sets in the metric space $(\psi(S_n), \mathrm{HD})$ (resp. $(\psi(S_n), \mathrm{MD})$). Let $a = a_1 a_2 \cdots a_{n-1}$, the possible lengths of a suffix of a are: $n-1,\ n-2,\ \ldots,\ 2,\ 1,$ and 0.

Let $A_{\geq j}$ be the canonical level set containing all strings whose fitness value is at least j. This means that an element of $A_{\geq j}$ is of the form:

$$a'_1 \cdots a'_{n-j-1} a_{n-j} \cdots a_{n-2} a_{n-1}, \tag{4.113}$$

where $1 \leq j \leq n-1$. We have the following results, using the same notations as in Theorem 14:

$$\begin{aligned}
A_{\geq 0} &= *_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,1]} \\
A_{\geq 1} &= *_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,2]}\ a_{n-1} \\
A_{\geq 2} &= *_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,3]}\ a_{n-2} a_{n-1} \\
&\cdots \\
A_{\geq k} &= *_{[0,n-1]} *_{[0,n-2]} \cdots *_{[0,k+1]}\ a_{n-k} \cdots a_{n-1} \\
&\cdots \\
A_{\geq n-1} &= \{a_1 a_2 \cdots a_{n-1}\}
\end{aligned}$$

By a similar proof to that of Proposition 2 (resp. Proposition 4), we find that each canonical level set is a convex set in the metric space $(\psi(S_n), \mathrm{HD})$ (resp.

$(\psi(S_n), \mathrm{MD})$). Moreover, $A_{\geq j+1}$ is always contained in $A_{\geq j}$ by construction. The number $q+1$ of distinct level sets is therefore n. The smallest ratio r between the sizes of two consecutive canonical level sets is $\frac{1}{n+1}$. □

Remark 2. *Let a be a fixed string of $\psi(S_n)$. For any string b of $\psi(S_n)$, let $SP_a(b)$ be the length of the longest prefix of a that is also a prefix of b. Using the same reasoning as above, we also find that the fitness landscape $(\psi(S_n), SP_a, \mathrm{HD})$ (resp. $(\psi(S_n), SP_a, \mathrm{MD})$) is quasi-concave.*

Theorem 18. *In the metric space $(\psi(S_n), \mathrm{HD})$, if the population size is at least:*

$$\mu \geq 8n(n+1)\ln(2n\sqrt{n}), \tag{4.114}$$

then the CS finds the longest suffix of a fixed string of $\psi(S_n)$ appearing as a suffix in a writing of a permutation σ as a string $\psi(\sigma)$, with probability at least 0.5.

Proof. By Proposition 10 and Corollary 16. □

Let one run of the CS be performed in $q = n-1$ generations. If the population size satisfies the condition of Theorem 18, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q = 2(n-1)$ and $2\mu q = 16(n-1)n(n+1)\ln(2n\sqrt{n})$.

4.9 Summary

We showed that the runtime analysis of the CS on a quasi-concave landscape that has been introduced in [MS17] can be specified to permutations.

We showed that for the metrics considered (Kendall's τ, Cayley, Ulam metrics, and reversal distance), the convex hulls can only be formed from the permutations involved. This is different from the convex hulls of strings on a finite alphabet (wrt the Hamming and the Manhattan distance), which can be formed from 'pieces' of strings that do not necessarily equal the strings involved.

Consequently, the runtime result for strings on a finite alphabets (wrt the Hamming and the Manhattan distance) does not extend to permutations (wrt the Kendall's τ, the Cayley, the Ulam metrics, and the reversal distance).

We also found that a permutation of S_n is uniquely determined by a $(n-1)$-uplet of $[0, n-1] \times [0, n-2] \times \cdots \times [0, 1]$, where $[0, i] = \{0, 1, 2, \ldots, i-1, i\}$. This enabled us to bring back the study of permutations to that of strings of the Hamming (resp. Manhattan) space.

We conclude that the runtime analysis of the CS on a quasi-concave landscape of [MS17] can be specified to any representation whose metric space is discrete. However, its runtime result may differ for the representations whose convex hulls can not be formed from ‘pieces’ of solutions.

Chapter 5

Standard Evolutionary Search Algorithm [1] (SES)

EA with crossover and no mutation have been generalized across representations as Convex Search Algorithms (CS) [Mor11]. The search operator used by the CS samples the convex hull of the selected population. Whereas, [Mor08] showed that most crossover operators sample the metric segment of two parents. In particular, the geometric crossover [MP04] has been defined as a generalization across representations of crossover operators sampling the metric segment of two parents. Hence, a more accurate generalization of EA with crossover and no mutation across representations should make use of geometric crossover. We aim to introduce a unifying runtime analysis of EA with no mutation and with a standard crossover, by:

- defining a generalization of EAs with crossover and no mutation across representations that makes use of geometric crossover,
- analyzing its runtime using a similar approach to that used for the CS in [MS17].

The general algorithm is given in Section 5.2. The runtime analysis is done in Section 5.3. We will start by studying the properties of the set where metric segments are sampled from in Section 5.1.

5.1 Union $Seg(A)$ of the segments of a set A

Let A be a subset of the metric space $(\mathcal{S}, D)$. A segment whose extremes are elements of A is referred to as a *segment of* A. We start by defining the set of points belonging to a segment of A.

[1] This algorithm is different from the algorithms with the same name that can be seen in other papers such as [MS02]. Indeed, here the word 'standard' refers to the standard crossover which is the only genetic operator used by the EA.

Definition 20. *Let $(\mathcal{S}, D)$ be a metric space and let $A \subseteq (\mathcal{S}, D)$. The set $Seg(A)$ is the union of all the segments of A. That is, the union of all the segments that can be made out of the elements of the set A.*

Example 38. *In the two-dimensional Hamming space $(\{0,1\}^2, \mathrm{HD})$, let us consider the subset $A = \{00, 01, 11\}$. The set $Seg(A)$ is the union of the segments $[00,00], [01,01], [11,11], [00,01], [00,11]$ and $[01,11]$. Hence, $Seg(A) = \{0,1\}^2$.*

5.1.1 Relationship between $Seg(A)$ and $co(A)$

Proposition 4.1.2. of [vDV93] gives the following result for any subset A of a metric space $(\mathcal{S}, D)$:

$$co(A) = A \cup Seg(A) \cup Seg(Seg(A)) \cup \cdots \cup Seg(\cdots(Seg(A))\cdots) \cup \cdots$$

This implies the following proposition:

Proposition 11. *Let $(\mathcal{S}, D)$ be a metric space and let A be a subset of $\mathcal{S}$. The set $Seg(A)$ is always included in the set $co(A)$.*

Besides the proof of Proposition 4.1.2. given in [vDV93], an alternative proof for Proposition 11 is given below for the sake of clarity.

Proof. By definition, $co(A)$ is the smallest convex set containing A. Hence $A \subseteq co(A)$. It follows that $Seg(A) \subseteq Seg(co(A))$. Since the set $co(A)$ is convex, all segments whose extremes are points of $co(A)$ are included in $co(A)$. Therefore, $Seg(co(A)) \subseteq co(A)$. Thus, $Seg(A) \subseteq co(A)$. □

Lemma 14. *Let $(\mathcal{S}, D)$ be a metric space and let A be a subset of $\mathcal{S}$, we have $co(Seg(A)) = co(A)$.*

Proof. On the one hand, we have $Seg(A) \subseteq co(A)$ by Proposition 11. This implies that $co(Seg(A)) \subseteq co(co(A)) = co(A)$. On the other hand, $A \subseteq Seg(A) \subseteq co(Seg(A))$. As a result, $co(Seg(A))$ is a smaller convex set that contains A. Necessarily, $co(Seg(A)) = co(A)$. □

Theorem 19. *Let $(\mathcal{S}, D)$ be a metric space and let A be a subset of $\mathcal{S}$. The following statements are equivalent:*

1. *$Seg(A)$ is a convex set,*
2. *$Seg(A) = co(A)$.*

Proof. On the one hand, if $Seg(A) = co(A)$ then the set $Seg(A)$ is convex as $co(A)$ is. On the other hand, if $Seg(A)$ is a convex set then $Seg(A) = co(Seg(A))$. The result follows from Lemma 14. □

This result will be useful in comparing the SES with the CS. In particular, we will be interested in metric spaces where the set $Seg(A)$ is always convex regardless of the choice of the subset A.

5.1.2 Ratio of segments of A strictly contained in $co(A)$

Let A be a finite set in a discrete metric space. We aim to compute a lower bound on the probability for sampling a pair of elements of A forming a segment that is strictly included in $co(A)$ when pairs are uniformly distributed. To this end, we first estimate the ratio of segments of A covering its convex hull $co(A)$.

Lemma 15. *The ratio of segments of A equating $co(A)$ is bounded above by $1/3$ whenever A contains at least two distinct elements.*

Proof. Let $[x_1, y_1]$ and $[x_2, y_2]$ be two segments of A equating $co(A)$.

- We show that if two segments equating $co(A)$ share an endpoint then they must share the other endpoint.

  If $x_1 = x_2$ and $y_1 \neq y_2$, then the segment $[y_1, y_2]$ is included in $co(A)$. This is because both y_i belong to $co(A)$ and $co(A)$ is a convex set. We have,

  $$[x_1, y_1] = [x_1, y_2] \text{ and } [y_1, y_2] \subseteq [x_1, y_1]. \tag{5.1}$$

  Therefore, $y_2 \in [x_1, y_1]$. As $[x_1, y_1] = [x_1, y_2]$, then y_2 must be equal to y_1. This contradicts the initial assumption. Therefore, whenever $x_1 = x_2$ then $y_1 = y_2$ when $[x_1, y_1]$ and $[x_2, y_2]$ are both equal to $co(A)$.

- Let $x_1 \neq x_2, y_2$ and $y_1 \neq x_2, y_2$. The segments $[x_1, y_1]$ and $[x_2, y_2]$ do not share any endpoint (though they may be equal in some specific metric spaces). We show that if $[x_1, y_1]$ and $[x_2, y_2]$ are both equal to $co(A)$, then the segments $[x_1, x_2]$, $[y_1, y_2]$, $[x_1, y_2]$, and $[x_2, y_1]$ are not equal to $co(A)$.

  Without loss of generality, let us show that $[x_1, x_2]$ is not equal to $co(A)$. Let us assume that $[x_1, x_2] = co(A)$. We have,

  $$[x_1, x_2] = [x_1, y_1], \tag{5.2}$$

  because $[x_1, y_1] = co(A)$ by assumption. This implies that $x_1 = y_2$. Since, $[x_1, y_1] = [x_2, y_2]$ we have:

  $$[y_2, y_1] = [x_2, y_2], \tag{5.3}$$

  by replacing x_1 by y_2 in the left hand side. Consequently, $x_2 = y_1$. As a result we have:

$$\begin{cases} x_1 &= y_2, \\ x_2 &= y_1, \end{cases} \tag{5.4}$$

which contradicts the initial assumption. Therefore, $[x_1, x_2]$ is not equal to $co(A)$.

We conclude that whenever two segments of A that do not share endpoints are both equal to $co(A)$, there exist at least four segments of A that are not equal to $co(A)$. Thus, the ratio of segments of A equating $co(A)$ is at most $\frac{2}{2+4} = \frac{1}{3}$.

If $A = \{x_1, x_2, x_3\}$ where the three elements are distinct, and the segment $[x_1, x_2]$ is equal to $co(A)$, then the segments $[x_2, x_3]$ and $[x_3, x_1]$ can not be equal to $co(A)$. Indeed, if they were they would be equal to $[x_1, x_2]$ and A would only contain two distinct elements instead of three. Hence, the ratio of segments of A equating $co(A)$ is at most $\frac{1}{3}$.

If $A = \{x_1, x_2\}$ where the two elements are distinct, and the segment $[x_1, x_2]$ is equal to $co(A)$, then the segments $[x_1, x_1] = \{x_1\}$ and $[x_2, x_2] = \{x_2\}$ can not be equal to $co(A)$. Hence, the ratio of segments of A equating $co(A)$ is $\frac{1}{3}$. □

Theorem 20. *We assume that the pairs of elements of A are uniformly distributed on $co(A)$. If A contains at least two distinct elements, then the probability for sampling a pair of elements of A forming a segment equating $co(A)$ is bounded above by $1/3$.*

Proof. The probability for sampling a pair of elements of A forming a segment that is equal to $co(A)$ is the ratio of segments of A equating $co(A)$. By Lemma 15, this ratio is bounded above by $1/3$. The result follows. □

Corollary 19. *We assume that the pairs of elements of A are uniformly distributed on $co(A)$. If A contains at least two distinct elements, then the probability for sampling a segment of A that is strictly included in $co(A)$ is bounded below by $2/3$.*

Proof. This is the complementary of the event of sampling a segment of A equating $co(A)$. As the probability of its complementary is at most $1/3$, its probability is at least $1 - 1/3 = 2/3$. □

5.2 Standard Evolutionary Search Algorithm

In [MS17], the convex evolutionary search algorithm (CS) is defined as the EA with no mutation with a convex hull recombination. We define the standard evolutionary search algorithm (SES) as the EA with no mutation with a standard crossover. Standard crossovers are instantiations of the geometric crossover in a specific representation [MP04].

Definition 21 (Geometric crossover [MP04]). *The (uniform) geometric crossover returns an offspring sampled uniformly at random from the segment formed by its two parents.*

Example 39. *Let us consider the elements* $x = 010$ *and* $y = 110$ *of the metric space* $(\{0,1\}^3, \mathrm{HD})$. *The segment* $[x, y]$ *is equal to the schema* $*10$. *The geometric crossover of the elements* x *and* y *consists of sampling an element of* $*10 = \{010, 110\}$ *uniformly at random.*

A pseudo-code corresponding to the SES [MM19] is shown in Algorithm 2.

Algorithm 2 Standard Evolutionary Search Algorithm

```
1:  Input: μ, population size
2:  Output: individual in the last population
3:  Initialise population uniformly at random
4:  while population has not converged to the same individual do
5:      Rank individuals on fitness
6:      if there are at least two fitness values in the current population then
7:          remove all individuals with the worst fitness
8:      end if
9:      Create new population:
10:     for counter in {1, 2, ..., μ} do
11:         Randomly and uniformly pick two individuals from the remaining individ-
            uals in the current population
12:         Recombine them through GEOMETRIC CROSSOVER to create a new indi-
            vidual
13:     end for
14: end while
15: Return any individual in the last population
```

Lines 11 and 12 tell us that a pair of individuals is sampled uniformly at random out of the set of all possible pairs of selected individuals. This means that the distribution of pairs of selected individuals is uniform on the set of selected individuals.

Offspring are sampled from a segment. As the notion of segment can be defined for any representation, the SES is representation independent.

5.2.1 Offspring distribution

Let us denote P' the set of parents that are selected from a population P. The set of reachable solutions $R(P')$ from the set of parents P' is the set of solutions that can be reached by repeated application of a search operator to the set of parents P'. In particular, the set $R_{\mathrm{SES}}(P')$ of reachable solutions for the geometric crossover is the union of all the segments that can be formed out of the elements of P':

$$R_{\mathrm{SES}}(P') = Seg(P'). \tag{5.5}$$

The offspring distribution need not be uniform on $R_{\mathrm{SES}}(P')$. Indeed, if x and y are elements of P' then the probability for sampling an offspring in the segment $[x, y]$ is $\frac{1}{|[x,y]|}$. Let $\alpha_{s,P'}$ be the number of pairs of elements of P' yielding the segment s. The total number of pairs that can be formed out of the elements of P' is $\mid P' \mid^2$. Hence, the probability for sampling the segment s is not uniform and is given by:

$$\frac{\alpha_{s,P'}}{\mid P' \mid^2}. \tag{5.6}$$

Example 40. *In $(\{0,1\}^2, \mathrm{HD})$, let $P' = \{00, 01, 10, 11\}$.*

- *Let s be the segment $[00, 11]$. We have $\alpha_{s,P'} = 4$. Indeed,*

$$s = [00, 11] = [11, 00] = [01, 10] = [10, 01]. \tag{5.7}$$

  The probability for sampling the segment s is therefore $\frac{4}{4^2} = \frac{1}{4}$.

- *Let s_{00} be the segment $[00, 00]$, we have $\alpha_{s_{00},P'} = 1$. The probability for sampling s_{00} is therefore $\frac{1}{4^2}$. This is also the probability for sampling each of the segments $s_{01},\ s_{10}$ and s_{11}.*

- *Let $s_{\{00,10\}}$ be the segment $[00, 10]$, we have $\alpha_{s_{\{00,10\}},P'} = 2$. The probability for sampling the segment $s_{\{00,10\}}$ is $\frac{2}{4^2}$. This is also the probability for sampling each of the segments $s_{\{00,01\}},\ s_{\{10,11\}}$, and $s_{\{01,11\}}$.*

There are 9 distinct segments that can be formed out of the elements of P'. Those segments are: $s,\ s_{00},\ s_{01},\ s_{10},\ s_{11},\ s_{\{00,10\}},\ s_{\{00,01\}},\ s_{\{10,11\}}$, and $s_{\{01,11\}}$. We obtain one by adding up the probabilities for sampling each one of them. In particular, we have:

$$Seg(P') = s \cup s_{00} \cup s_{01} \cup s_{10} \cup s_{11} \cup s_{\{00,10\}} \cup s_{\{00,01\}} \cup s_{\{10,11\}} \cup s_{\{01,11\}}. \tag{5.8}$$

More generally, the set $Seg(P')$ can be rewritten as the union of the distinct segments that can be formed out of the elements of P'. That is, there exists $p \leq \mid P' \mid^2$ such that:

$$Seg(P') = \bigcup_{i=1}^{p} s_i. \tag{5.9}$$

As $s_1,\ s_2,\ \ldots,\ s_p$ are the only segments that can be formed out of the elements of P', we have:

$$\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{\mid P' \mid^2} = 1. \tag{5.10}$$

Theorem 21. *Let z be a reachable solution and let $\mathbb{1}_s$ be the indicator function on the segment s. The probability for sampling z is given by:*

$$Pr(z \in Seg(P')) = \sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{\mid P' \mid^2} \cdot \frac{\mathbb{1}_{s_i}(z)}{\mid s_i \mid}. \tag{5.11}$$

We also have:

$$\sum_{z \in Seg(P')} Pr(z \in Seg(P')) = \sum_{z \in Seg(P')} \sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{\mid P' \mid^2} \cdot \frac{\mathbb{1}_{s_i}(z)}{\mid s_i \mid}, \tag{5.12}$$

$$= 1. \tag{5.13}$$

5.2.2 Recycling the analysis of the CS for the SES

In metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for all A, the analysis of the CS on quasi-concave landscapes can be reused to analyse the SES on quasi-concave landscapes.

Recall that the search operator used by the CS returns an offspring sampled uniformly at random from the convex hull formed by the selected individuals (see Definition 6). Moreover, the convex hull formed by the selected individuals forms the set of reachable solutions for the CS (see Equation (3.9)).

For the SES, the set of reachable solutions is only convex in metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for all A. In particular, for the same set of parents P' the sets of reachable solutions for the CS and SES both coincide with $co(P')$. However, the distribution of the reachable solutions on $co(P')$ differs for both algorithms. Offspring are uniformly distributed on $co(P')$ for the CS [Mor11]. Whereas, offspring are not uniformly distributed on $co(P')$ for the SES (see Equation (5.11)).

Consequently, we must restrict our study to metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for all A in order to reuse the analysis of the CS on quasi-concave landscapes for the SES on quasi-concave landscapes.

5.3 Runtime Analysis

We compute an upper bound on the runtime of the SES on a quasi-concave landscape in a metric space where $Seg(A) = co(A)$ for any subset A, where the analysis used in [MS17] for the CS can be used as a guideline.

The SES finds a global optimum if the convex hull formed by the selected individuals always covers a higher level set than the one containing them. As level sets form a decreasing chain of sets with respect to the 'contains' order (see Definition 7), the condition above is satisfied whenever the convex hull formed by the selected individuals is always equal to the level set containing them. In combinatorial spaces, the latter happens with probability at least 0.5 for a well chosen population size. Indeed, the distribution of the offspring is not uniform on the level set. However, the distribution of pairs of parents is uniform on the level set and each offspring is created from a pair of parents. In this case, the SES finds a global optimum within $2q$ generations where $q+1$ is the total number of distinct level sets.

We start by estimating a lower bound on the probability of sampling a strictly improving offspring in a quasi-concave landscape. We recall that:

$$r = \min_{0\leq j\leq q-1}\left(\frac{\mid A_{\geq j+1}\mid}{\mid A_{\geq j}\mid}\right). \tag{5.14}$$

Theorem 22. *The probability for sampling a strictly improving offspring from any selected population with at least two distinct individuals is bounded below by* $2r^2/3$.

Proof. Let P' denote the set of selected individuals. We assume that P' is contained in the canonical level set $A_{\geq j}$. In the worst case, $co(P')$ is equal to the level set $A_{\geq j}$ containing it. In this case, the probability for sampling an offspring belonging to $A_{\geq j+1}$ (which is a strict subset of $A_{\geq j}$) is given by:

$$\sum_{z\in A_{\geq j+1}}\sum_{i=1}^{p}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}\cdot\frac{\mathbb{1}_{s_i}(z)}{\mid s_i\mid} \tag{5.15}$$

$$=\sum_{i=1}^{p}\sum_{z\in A_{\geq j+1}}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}\cdot\frac{\mathbb{1}_{s_i}(z)}{\mid s_i\mid}, \tag{5.16}$$

$$=\sum_{i=1}^{p}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}\cdot\frac{\mid A_{\geq j+1}\cap s_i\mid}{\mid s_i\mid}, \tag{5.17}$$

$$=\sum_{s_i\subseteq A_{\geq j+1}}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}\cdot\frac{\mid A_{\geq j+1}\cap s_i\mid}{\mid s_i\mid} \tag{5.18}$$

$$+\sum_{\substack{s_i\nsubseteq A_{\geq j+1}\\ s_i\cap A_{\geq j+1}\neq\varnothing}}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}\cdot\frac{\mid A_{\geq j+1}\cap s_i\mid}{\mid s_i\mid}, \tag{5.19}$$

$$\geq\sum_{s_i\subseteq A_{\geq j+1}}\frac{\alpha_{s_i,P'}}{\mid P'\mid^2}. \tag{5.20}$$

The bound in Inequality (5.20) is the probability to sample a segment of P' that is strictly included in $A_{\geq j}=co(P')$. By Corollary 19, the probability for sampling a segment of P' that is strictly included in $co(P')$ is bounded below by $2/3$ given that $Seg(P')=co(P')$ and P' contains at least two distinct elements. Pairs are sampled uniformly at random from the set of all possible pairs that can be made out of the elements of $P'\subseteq A_{\geq j}$. Thus, the probability for sampling a pair that is included in $A_{\geq j+1}$ is given by:

$$\left(\frac{\mid A_{\geq j+1} \mid}{\mid A_{\geq j} \mid}\right)^2 \geq \left[\min_{0 \leq j \leq q-1} \left(\frac{\mid A_{\geq j+1} \mid}{\mid A_{\geq j} \mid}\right)\right]^2, \tag{5.21}$$

$$= r^2. \tag{5.22}$$

Consequently, the probability to sample a segment of P' that is strictly included in $A_{\geq j} = co(P')$ is bounded below by $\frac{2r^2}{3}$ given that $Seg(P') = co(P')$ and P' contains at least two distinct elements. □

Corollary 20. *The expected number of strictly improving offspring for a population size of μ is at least:*

$$\frac{2r^2\mu}{3}, \tag{5.23}$$

if at least two distinct individuals are selected at each generation.

Proof. In the worst case, all strictly improving offspring of the selected population have the same least probability of Theorem 22 to be sampled. The total number of offspring that is created is given by the population size μ. Consequently, the expected number of strictly improving offspring among the μ offspring is at least $\frac{2r^2\mu}{3}$. □

The offspring are not uniformly distributed on $Seg(P')$ as seen in Equation (5.11). In metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for any subset A, any offspring has a non-zero probability to be sampled. This is because the sets $Seg(P')$ and $co(P')$ coincide, where $co(P')$ is the level set containing the offspring.

Let m be a positive integer and let C be a non empty convex set whose elements are distributed as in Equation (5.11). The set of m points drawn from C is denoted $\mathrm{NonUnif}_m(C)$.

Definition 22. *Let C be a convex set in a metric space $(\mathcal{S}, D)$ whose elements are distributed as in Equation* (5.11)*. The probability that the union of all the segments that can be made out of m points drawn from C equals C is:*

$$P^{\mathrm{Cov}}_{C,Seg}(m) = \Pr[Seg(P) = C \mid P = \mathrm{NonUnif}_m(C)]. \tag{5.24}$$

In metric spaces where $Seg(A) = co(A)$ for any subset A, the probability $P^{\mathrm{Cov}}_{C,Seg}(m)$ is equal to the probability that the convex hull of m points drawn from C equals C. That is:

$$P^{\mathrm{Cov}}_{C,Seg}(m) = \Pr[Seg(P) = C \mid P = \mathrm{NonUnif}_m(C)], \tag{5.25}$$

$$= \Pr[co(P) = C \mid P = \mathrm{NonUnif}_m(C)], \tag{5.26}$$

$$\geq \min_{C \in C_{\mathcal{S}}} \Pr[co(P) = C \mid P = \mathrm{NonUnif}_m(C)], \tag{5.27}$$

where $C_{\mathcal{S}}$ denotes the set of convex sets on the entire search space $\mathcal{S}$. Let us denote $P_{\mathcal{S}}^{\text{Cov}}(m)$ the probability $\min_{C \in C_{\mathcal{S}}} \Pr[co(P) = C \mid P = \text{NonUnif}_m(C)]$. As in [MS17], the probability $P_{\mathcal{S}}^{\text{Cov}}(m)$ is monotone increasing in m because additional samples can only increase the convex hull.

We assume a quasi-concave fitness function on the metric space $(\mathcal{S}, D)$ with fitness levels $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_{\geq q}$. Let P'_t denote the parents of generation t. The following lemma gives a lower bound on the probability that $co(P'_{t+1})$ is equal to some $A_{\geq j}$ given that $co(P'_t)$ is equal to $A_{\geq i}$ and $i < j$.

Lemma 16. *The probability that the next generation of parents covers a higher level set than the level set covered by the current generation of parents is at least:*

$$P_{\mathcal{S}}^{\text{Cov}}\left(\frac{2r^2\mu}{3}\right) - \exp\left(-\frac{r^2\mu}{18}\right). \tag{5.28}$$

Proof. The probability $P_{\mathcal{S}}^{\text{Cov}}(m)$ is monotone increasing in m. For a population size of μ, m is at least $\frac{2r^2\mu}{3}$ by Corollary 20. Hence, $P_{\mathcal{S}}^{\text{Cov}}\left(\frac{2r^2\mu}{3}\right)$ is a lower bound on $P_{\mathcal{S}}^{\text{Cov}}(m)$. Using Chernoff bound [MR95], the probability that the number of strictly improving offspring is smaller than $\frac{2r^2\mu}{3}$, is at most:

$$Pr\left(\mid P' \mid \leq \frac{2r^2\mu}{3}\right) \leq \exp\left[-\frac{r^2\mu}{2} \cdot \left(\frac{1}{3}\right)^2\right]. \tag{5.29}$$

We define the worst-case typical behaviour to have exactly $\frac{2r^2\mu}{3}$ strictly improving offspring in each level set as in Corollary 20. □

Theorem 23. *The SES with population size μ finds a global optimum within q generations and μq fitness evaluations with probability at least*

$$\left[P_{\mathcal{S}}^{\text{Cov}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right). \tag{5.30}$$

Proof. The reasoning is the same as in [MS17]. We assume that the probability for covering different level sets are independent. Each level set is visited taking into account $A_{\geq 0}$. Then, the probability that less than $\frac{2r^2\mu}{3}$ strictly improving offspring are generated is removed at each step. □

The next step is to explicitly compute

$$P_C^{\text{Cov}}(m) = \Pr[co(P) = C \mid P = \text{NonUnif}_m(C)] \tag{5.31}$$

for specific representations.

When the selection of the m elements of C is uniform, then they equally contribute to the creation of their convex hull $co(P)$. This is for example the case for the CS [MS17].

When the m elements of C are not selected uniformly at random, they need not equally contribute to the creation of their convex hull $co(P)$. Each element must contribute at least once in the making of their convex hull $co(P)$. It remains to determine the maximum number of contributions. To this end, we introduce the notion of weight to measure the number of contributions of each of the m elements of C.

Definition 23. *Let $e_1,\ e_2,\ \dots,\ e_m$ be m samples from a non-empty convex set C. For each $e \in C$, we denote $p(e)$ the probability to select e. The weight of the element e_i of C is defined as:*

$$w_i = p(e_i) \cdot \mathrm{lcd}_{e \in C} p(e), \tag{5.32}$$

where lcd stands for least common denominator.

Example 41. *For the CS, samples are selected uniformly at random from a non-empty level set $A_{\geq j}$ [MS17]. Each of them has the same probability $p(e) = \frac{1}{|A_{\geq j+1}|}$ to be selected. Indeed, an element $e \in A_{\geq j}$ is selected if it belongs to $A_{\geq j+1}$. Hence, $p(e)$ is the probability to sample an element e of $A_{\geq j}$ given that this element belongs to $A_{\geq j+1}$. The result follows as offspring are uniformly distributed on $A_{\geq j}$ with probability $\frac{1}{|A_{\geq j}|}$. We have:*

$$\begin{aligned} w &= p(e) \cdot \mathrm{lcd}_{e \in C} p(e), & (5.33) \\ &= 1. & (5.34) \end{aligned}$$

Those m samples correspond to the selected individuals that will make up the set of parents of the next generation. This means that they are strictly improving offspring with respect to the current set of offspring. Hence, the selection probability (of the m selected individuals) is the probability to sample a strictly improving offspring in the convex hull of the current selected population.

Let P'_t be the current selected population. The elements of $Seg(P'_t) = co(P'_t)$ are distributed as in Equation (5.11). Moreover, the set $co(P'_t)$ is equal to a level set $A_{\geq j}$ in our case study. That is, $P = P'_{t+1}$ and $C = A_{\geq j+1}$ in Equation (5.31). The selection probability is therefore the probability for sampling an element of $A_{\geq j}$ that belongs to $A_{\geq j+1}$.

Proposition 12. *The selection probability of elements of $A_{\geq j}$ is at least:*

$$\frac{1}{\mid A_{\geq j+1} \mid +1}. \tag{5.35}$$

Proof. We assume that the offspring are distributed on $co(P') = A_{\geq j}$ as in (5.11). Let t be the current generation, we have $P' = P'_t$. The parents P'_{t+1} of the next generation are selected from $A_{\geq j+1}$. The probability to select an offspring z_0 from $co(P'_t) = A_{\geq j}$ that belongs to $A_{\geq j+1}$ is given by:

$$\frac{\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{\mathbb{1}_{s_i}(z_0)}{|s_i|}}{\sum_{z \in A_{\geq j+1}} \sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{\mathbb{1}_{s_i}(z)}{|s_i|}} = \frac{\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}}{\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|}}. \tag{5.36}$$

We determine a lower bound on (5.36).

$$\begin{aligned}
&\frac{\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}}{\sum_{i=1}^{p} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|}}, \\
=&\frac{\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|} + \sum_{\substack{s_i \nsubseteq A_{\geq j+1} \\ s_i \cap A_{\geq j+1} \neq \varnothing}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}}{\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|} + \sum_{\substack{s_i \nsubseteq A_{\geq j+1} \\ s_i \cap A_{\geq j+1} \neq \varnothing}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|}}.
\end{aligned}$$

By Corollary 19, the probability for sampling a segment s_i that is strictly included in $A_{\geq j}$ is bounded below by $2/3$ when P' contains at least two distinct individuals. Hence, a segment s_i is either contained in $A_{\geq j+1}$ or equal to $A_{\geq j+1}$ in the typical case. Therefore,

$$\sum_{\substack{s_i \nsubseteq A_{\geq j+1} \\ s_i \cap A_{\geq j+1} \neq \varnothing}} \frac{\alpha_{s_i,P'}}{| P' |^2} \cdot \frac{| A_{\geq j+1} \cap s_i |}{| s_i |} \leq \sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{| P' |^2} \cdot \frac{| A_{\geq j+1} \cap s_i |}{| s_i |}. \tag{5.37}$$

Consequently, a lower bound on (5.36) is given by:

$$\begin{aligned}
&\frac{\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}}{\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|} + \sum_{\substack{s_i \nsubseteq A_{\geq j+1} \\ s_i \cap A_{\geq j+1} \neq \varnothing}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|}}, \\
\geq&\frac{\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}}{\left[\sum_{s_i \subseteq A_{\geq j+1}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|\{z_0\} \cap s_i|}{|s_i|}\right] \max_{\substack{s_i \subseteq A_{\geq j+1} \\ s_i \cap \{z_0\} \neq \varnothing}} \frac{|A_{\geq j+1} \cap s_i|}{|\{z_0\} \cap s_i|} + \sum_{\substack{s_i \nsubseteq A_{\geq j+1} \\ s_i \cap A_{\geq j+1} \neq \varnothing}} \frac{\alpha_{s_i,P'}}{|P'|^2} \cdot \frac{|A_{\geq j+1} \cap s_i|}{|s_i|}}, \\
\geq&\frac{1}{\max_{\substack{s_i \subseteq A_{\geq j+1} \\ s_i \cap \{z_0\} \neq \varnothing}} \frac{|A_{\geq j+1} \cap s_i|}{|\{z_0\} \cap s_i|} + 1}, \\
\geq&\frac{1}{| A_{\geq j+1} | + 1}.
\end{aligned}$$

□

Proposition 13. *The selection probability of elements of $A_{\geq j}$ is at most:*

$$\frac{2}{| A_{\geq j+1} | + 1}. \tag{5.38}$$

Proof. The probabilities for selecting an offspring belonging to $A_{\geq j+1}$ add up to one for each element of $A_{\geq j+1}$. An upper bound on the probability for selecting an offspring belonging to $A_{\geq j+1}$ is obtained when all the remaining $\mid A_{\geq j+1} \mid -1$ offspring have the least probability of Proposition 12 to be selected. Hence, the largest probability for selecting an offspring belonging to $A_{\geq j+1}$ (i.e., a parent for the next generation) is:

$$1 - \frac{\mid A_{\geq j+1} \mid -1}{\mid A_{\geq j+1} \mid +1} = \frac{2}{\mid A_{\geq j+1} \mid +1}.$$

□

By Proposition 12 and Proposition 13 we have:

Corollary 21. *The weight of a sample is at most two for the SES.*

Proof. The lcd of all the probabilities for selecting a parent (i.e., an offspring belonging to $A_{\geq j+1}$) is given by the denominator of the least possible probability given in Proposition 12. Any other probability is bounded above by the largest possible probability given in Proposition 13. Hence, all weights are bounded above by:

$$\frac{2}{\mid A_{\geq j+1} \mid +1} \cdot (\mid A_{\geq j+1} \mid +1) = 2.$$

□

As a result, each of the m samples of Equation (5.31) contributes between once and twice in the making of their convex hull in our analysis.

5.4 Summary

We defined a generalization of EAs with no mutation and with a standard two-parents crossover, called SES. By restricting our analysis to metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for all A, we computed an upper bound on the runtime of the SES on a quasi-concave landscape by adjusting the analysis of [MS17] on the CS.

The SES finds a global optimum if each set of offspring generates the level set containing them at each generation. We determined a lower bound on the probability for sampling a strictly improving offspring in a selected population. The latter was used to determine a lower bound on the expected number of strictly improving offspring for a given population size. Then, we used that result to estimate a lower bound on the success probability of the SES. Finally, we estimated the selection probability in order to specify the runtime results to specific representations.

It remains to find a new analysis across representations of the SES on quasi-concave landscapes, that can deal with metric spaces where the sets $Seg(A)$ and $co(A)$ do not necessarily coincide for all A.

Chapter 6

SES on Strings on a Finite Alphabet

We specify the analysis of the SES on quasi-concave landscapes to strings on finite alphabets in metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for any subset A. We will consider the same metrics used in [MS17] for the analysis of the CS:

- The Hamming distance HD,
- The Manhattan distance MD.

We aim to compare the runtime result of the SES to that of the CS. The metric space $(\{0,1,2,\cdots,d-1\}^n, D)$ is denoted $M_{d,D}$.

We first determine whether the sets $Seg(A)$ and $co(A)$ coincide for any subsets for each of the metric spaces $M_{d,\mathrm{HD}}$ and $M_{d,\mathrm{MD}}$ in Section 6.1. Then, we determine the schemata corresponding to convex sets for each of the metrics considered in Section 6.2. The runtime analysis is done in Section 6.3. Finally, the SES and the CS are compared in Section 6.4.

6.1 Convexity of the Set of Reachable Solutions

We study the convexity of the set of reachable solutions in the metric spaces $M_{d,\mathrm{HD}}$ and $M_{d,\mathrm{MD}}$. That is, we determine whether the sets $Seg(A)$ and $co(A)$ always coincide for any subset A.

6.1.1 Hamming distance

We show that $Seg(A) = co(A)$ for any subset A of the metric space $M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$. We also show that $Seg(A)$ need not be equal to $co(A)$ for any subset A of the metric space $M_{d,\mathrm{HD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{HD})$ where $d > 2$.

Proposition 14. *Any segment of the metric space $M_{d,\mathrm{HD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{HD})$ is a convex set.*

Proof. Any segments of the n-dimensional Hamming space $(\{0,1,\ldots,d-1\}^n, \mathrm{HD})$ is the Cartesian product of n segments of the one-dimensional space $(\{0,1,\ldots,d-1\}, \mathrm{HD})$ [vDV93]. A segment of $(\{0,1,\ldots,d-1\}, \mathrm{HD})$ is either a single element or the union of two distinct elements. Hence, a segment of $(\{0,1,\ldots,d-1\}, \mathrm{HD})$ is always a convex set. Since a Cartesian product of convex sets remains convex [vDV93], any segment of the n-dimensional Hamming space $(\{0,1,\ldots,d-1\}^n, \mathrm{HD})$ is also a convex set. □

Let A be a set in the metric space $M_{d,\mathrm{HD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{HD})$ and let $s_1, s_2, \ldots, s_p$ be the p distinct segments that can be formed out of the elements of A:

$$Seg(A) = \bigcup_{1 \leq j \leq p} s_j. \tag{6.1}$$

A segment $s_j = [x_j, y_j]$ corresponds to the schema $*_{A_{s_j}(1)} *_{A_{s_j}(2)} \ldots *_{A_{s_j}(n)}$, where $A_{s_j}(i) = \{x_j(i), y_j(i)\}$ is the set of admissible values at position i. We also recall that in the schema corresponding to $co(A)$, the admissible values at position i are the elements of $\bigcup_{1 \leq j \leq p} A_{s_j}(i)$. We have the following result:

Lemma 17. *The union $Seg(A)$ of all the segments that can be formed out of the elements of A is a convex set if there exists $1 \leq j \leq p$ such that:*

$$\bigcup_{1 \leq j \leq p} A_{s_j}(i) = A_{s_j}(i), \tag{6.2}$$

at each position i.

Proof. If Equation (6.2) is satisfied then there exists a segment s_j such that $co(A) = s_j$. Consequently, $co(A)$ is contained in $Seg(A) = \bigcup_{1 \leq j \leq p} s_j$. Therefore, the sets $co(A)$ and $Seg(A)$ are necessarily equal. □

Corollary 22. *In the metric space $M_{d,\mathrm{HD}}$, the set $Seg(A)$:*

- *is always convex for any subset A when $d = 2$,*
- *need not be convex for any subset A when $d > 2$*

Proof. In $M_{2,\mathrm{HD}}$, the set $\bigcup_{1 \leq j \leq p} A_{s_j}(i)$ contains either one or two elements. In both cases, we have:

$$\bigcup_{1 \leq j \leq p} A_{s_j}(i) = \{x_j(i)\} \cup \{y_j(i)\}, \tag{6.3}$$

$$= A_{s_j}(i). \tag{6.4}$$

Consequently, the set $Seg(A)$ is always convex in $M_{2,\mathrm{HD}}$.

When $d > 2$, the set $\bigcup_{1 \leq j \leq p} A_{s_j}(i)$ may contain more than two elements. In this case, it can not correspond to a set $A_{s_j}(i)$. As a result, the set $Seg(A)$ need not be convex in $M_{d,\mathrm{HD}}$ when $d > 2$. □

Example 42. *In the metric space* $(\{0,1,2\}^4, \mathrm{HD})$*, let* $A = \{0012, 2110, 2011\}$*. We have that:*

$$Seg(A) = [0012, 2110] \cup [0012, 2011] \cup [2110, 2011], \tag{6.5}$$

$$= *_{02} *_{01} 1 *_{02} \ \cup \ *_{02} 01 *_{12} \ \cup 2 *_{01} 1 *_{01}, \tag{6.6}$$

and $co(A) = *_{02} *_{01} 1*$*. We can see that* $0111 \in co(A)$ *but* $0111 \notin Seg(A)$*. Hence,* $Seg(A) \subsetneq co(A)$*.*

6.1.2 Manhattan distance

We show that $Seg(A) = co(A)$ for any subset A in the metric space $M_{d,\mathrm{MD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{MD})$.

Proposition 15. *Any segment in the metric space* $M_{d,\mathrm{MD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{MD})$ *is a convex set.*

Proof. Any segments of the n-dimensional Manhattan space $(\{0,1,\ldots,d-1\}^n, \mathrm{MD})$ is the Cartesian product of n segments of the one-dimensional space $(\{0,1,\ldots,d-1\}, \mathrm{MD})$ [vDV93]. A segment of $(\{0,1,\ldots,d-1\}, \mathrm{MD})$ is either a single element, two consecutive elements, three consecutive elements, ..., or d consecutive elements. Hence, a segment of $(\{0,1,\ldots,d-1\}, \mathrm{MD})$ is always a convex set. Since a Cartesian product of convex sets remains convex [vDV93], any segment of the n-dimensional Manhattan space $(\{0,1,\ldots,d-1\}^n, \mathrm{MD})$ is also a convex set. □

Let A be a set in the metric space $M_{d,\mathrm{MD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{MD})$ and let $s_1, s_2, \ldots, s_p$ be the p distinct segments that can be formed out of the elements of A:

$$Seg(A) = \bigcup_{1 \leq j \leq p} s_j. \tag{6.7}$$

A segment $s_j = [x_j, y_j]$ corresponds to the schema $*_{A_{s_j}(1)} *_{A_{s_j}(2)} \ldots *_{A_{s_j}(n)}$, where $A_{s_j}(i) = [\min\{x_j(i), y_j(i)\}, \max\{x_j(i), y_j(i)\}]$ is the set of admissible values at position i. We also recall that in the schema corresponding to $co(A)$, the admissible values at position i are the elements of $[\min_{x,y \in A}\{x(i), y(i)\}, \max_{x,y \in A}\{x(i), y(i)\}]$. We have the following result:

Lemma 18. *In the metric space* $M_{d,\mathrm{MD}} = (\{0,1,\ldots,d-1\}^n, \mathrm{MD})$*, the set* $Seg(A)$ *is convex for any subset* A*.*

Proof. In the schema corresponding to $co(A)$, the admissible values at position i are the elements of $[\min_{x,y \in A}\{x(i), y(i)\}, \max_{x,y \in A}\{x(i), y(i)\}]$. This means that there exists a segment s_j such that $co(A) = s_j$. As a result, $Seg(A) = co(A)$ and is therefore a convex set. □

Theorem 24. *Let A be a set, the union $Seg(A)$ of all the segments that can be formed out of the elements of A is equal to the convex hull $co(A)$ of A in the metric space $M_{d,\mathrm{MD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$.*

Proof. Since the set $Seg(A)$ is convex, it is equal to the set $co(A)$ by Theorem 19. □

Example 43. *In the metric space $(\{0, 1, 2\}^4, \mathrm{MD})$, let $A = \{0012, 2110, 2011\}$. We have that:*

$$Seg(A) = [0012, 2110] \cup [0012, 2011] \cup [2110, 2011], \tag{6.8}$$

$$= * *_{01} 1 * \ \cup \ * 01 *_{12} \ \cup 2 *_{01} 1*_{01}, \tag{6.9}$$

and $co(A) = * *_{01} 1*$. We can see that $co(A) = [0012, 2110]$ and $Seg(A) = co(A)$.

6.2 Schemata corresponding to Convex Sets

The notion of schemata [H+92] is used to define the convex canonical level sets of a quasi-concave landscape [MS17]. Schemata corresponding to convex sets of the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$ and the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$ are determined. We shall:

- Prove that any schema corresponds to a convex set for the Hamming distance,
- Determine the schemata corresponding to a convex set for the Manhattan distance.

6.2.1 Hamming distance

We first show that all schemata are convex sets in the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$.

Corollary 23. *Any schema in the metric space $M_{d,\mathrm{HD}}$ is a convex set.*

Proof. Let S be a schema in the metric space $M_{d,\mathrm{HD}}$, whose admissible values at position i are the elements of a subset $A_S(i)$ of $\{0, 1, \ldots, d-1\}$ for $0 \leq i \leq n-1$. Let x and y be two elements of S. We show that the segment $[x, y]$ is contained in S.

Let $z \in [x, y]$, the value of $z(i)$ is either $x(i)$ or $y(i)$. As both x and y belong to S, then $x(i)$ and $y(i)$ belong to the set $A_S(i)$ of admissible values at position i of S. Hence, $z(i)$ also belongs to $A_S(i)$. Thus, $[x, y]$ is contained in S. Therefore, the schema S is a convex set. □

6.2.2 Manhattan distance

We now determine the schemata that are convex sets in the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$.

Corollary 24. *Let $0 \leq k \leq l \leq d-1$ and let $[k, l]$ denote the set $\{k, k+1, \ldots, l-1, l\}$. The only convex schemata of the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{MD})$ are those that only use symbols $*_{[k,l]}$ and/or $*$ and/or fixed values.*

Proof. Let S be a schema in the metric space $M_{d,\mathrm{MD}}$, whose admissible values at position i are the elements of a subset $A_S(i)$ of $\{0, 1, \ldots, d-1\}$ for $0 \leq i \leq n-1$.

Let x and y be two elements of S. We determine the conditions under which the segment $[x, y]$ is contained in S.

Let $z \in [x, y]$, the value of $z(i)$ belongs to $[\min\{x(i), y(i)\}, \max\{x(i), y(i)\}]$. As both x and y belong to S, then $x(i)$ and $y(i)$ belong to the set $A_S(i)$ of admissible values at position i of S. Hence, $z(i)$ belongs to $A_S(i)$ if:

$$[\min\{x(i), y(i)\}, \max\{x(i), y(i)\}] \subseteq A_S(i). \tag{6.10}$$

In order to ensure that $[x, y]$ is contained in S for any $x, y \in S$, we must ensure that the inclusion above holds for any $x(i), y(i) \in A_S(i)$. Necessarily, $A_S(i)$ must be a set of consecutive values such that:

$$A_S(i) = \left[\min_{x(i),y(i)\in A_S(i)} \{x(i), y(i)\}, \max_{x(i),y(i)\in A_S(i)} \{x(i), y(i)\}\right]. \tag{6.11}$$

□

Consequently, the number of distinct level sets $A_{\geq j}$ is at most $q = n$ when the only symbol used is $*$ along with fixed values. In this case, the smallest ratio between the size of two consecutive level sets is:

$$\begin{aligned} r &= \min_{0\leq j\leq n-2} \{\mid A_{\geq j+1} \mid / \mid A_{\geq j} \mid\}, \\ &= \mid \{0, 1, \ldots, d-1\}^{n-1} \mid / \mid \{0, 1, \ldots, d-1\}^n \mid, \\ &= 1/d. \end{aligned}$$

An illustration is given in Figure 6.1 for $(\{0, 1, 2\}^5, \mathrm{MD})$.

6.3 Specification of the Runtime Analysis to Strings on a Finite Alphabet

We specify the runtime result of the SES on a quasi-concave landscape of Chapter 5 to strings on a finite alphabet, by considering the metric spaces $M_{2,\mathrm{HD}}$ and

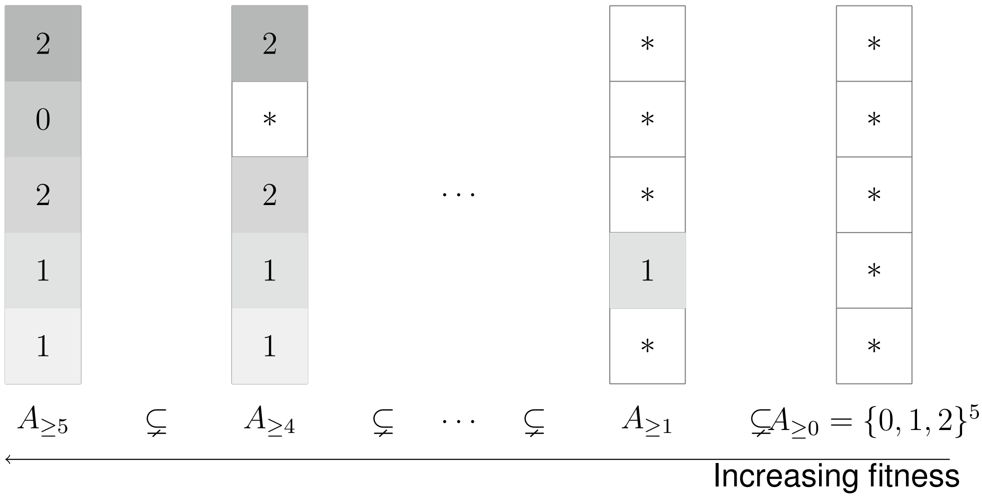


Figure 6.1: Convex canonical level sets in the metric space $(\{0,1,2\}^5, \mathrm{MD})$. A position that can take any value belonging to $\{0,1,2\}$ is marked with $*$. Two different positions may be fixed to the same value or to two distinct values.

$M_{d,\mathrm{MD}}$ for $d \geq 2$.

6.3.1 Hamming distance

The sets $Seg(A)$ and $co(A)$ coincide for any subset A of the metric space $M_{2,\mathrm{HD}}$. This need not be the case in the metric spaces $M_{d,\mathrm{HD}}$, where $d \geq 3$. Thus, we restrict our analysis to the metric space $M_{2,\mathrm{HD}}$.

We first estimate $P^{\mathrm{Cov}}_{M_{2,\mathrm{HD}}}(m)$ which is a lower bound on the probability for covering a convex set C of $M_{2,\mathrm{HD}}$ with m samples from C.

Lemma 19. *For any convex set C of the metric space $M_{2,\mathrm{HD}}$ we have $P^{\mathrm{Cov}}_{C}(m) \geq P^{\mathrm{Cov}}_{M_{2,\mathrm{HD}}}(m)$, where:*

$$P^{\mathrm{Cov}}_{M_{2,\mathrm{HD}}}(m) \geq 1 - 2n\left(1 - \frac{1}{4}\right)^{m}.$$

Proof. We will estimate:

$$P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}(m) = \Pr[co(P) = M_{d,\mathrm{HD}} \mid P = \mathrm{NonUnif}_m(M_{d,\mathrm{HD}})], \qquad (6.12)$$

for $d = 2$.

We saw in Corollary 23 that any schema corresponds to a convex set in the metric space $M_{d,\mathrm{HD}}$. In particular, the schema corresponding to the entire search space is the only schema with the largest number of positions that are free to take more than one value. Moreover, each of these free positions take the maximum number of possible values. Therefore, the schema corresponding to any other convex set has at most n free positions. Each of these positions is free to take at most d values.

Let us now compute the probability $\Pr[co(P) = M_{d,\text{HD}} \mid P = \text{NonUnif}_m(M_{d,\text{HD}})]$ for covering the entire search space from sampling m points from it. The schema corresponding to the entire search space is $\underbrace{* * * \cdots *}_{n \text{ times}}$.

The don't care symbol is obtained at some position when each of the values $0,\ 1,\ \ldots,\ d-1$ appears at least once at this position. The probability that a value appears at this position in e_i is $\frac{1}{d^{w_i}}$. The probability that this value never appears at this position in e_i is therefore $1 - \frac{1}{d^{w_i}}$. The probability that this value never appears at this position in e_1, e_2, ..., and e_m is therefore:

$$\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right). \tag{6.13}$$

The probability that the value 0 never appears at this position OR the value 1 OR ... OR the value $d-1$ is:

$$d\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right) \leq d\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.14}$$

Hence, the probability that each value appears at least once at that position is:

$$1 - d\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right) \geq 1 - d\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.15}$$

Thus, the probability for obtaining the don't care symbol at n positions is:

$$\left[1 - d\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right)\right]^n \geq \left[1 - d\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m\right]^n. \tag{6.16}$$

Hence, the probability for obtaining the schema $\underbrace{* * * \cdots *}_{n \text{ times}}$ is at least:

$$\left[1 - d\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m\right]^n \geq 1 - dn\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m, \tag{6.17}$$

using Bernoulli's inequality. The probability for obtaining a schema with n' free positions where each free position can take at most d' values is at least:

$$1 - d'n'\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.18}$$

As $d' \leq d$ and $n' \leq n$, we have:

$$1 - d'n'\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m \geq 1 - dn\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.19}$$

By Corollary 21, all weights are bounded above by two. The result follows. $\square$

Theorem 25. *Let us consider a quasi-concave landscape on $M_{2,\mathrm{HD}}$, whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The SES with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1 - 2n(2q+1)\exp\left(-\frac{1}{18}\cdot r^2\mu\right). \tag{6.20}$$

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right)$ for $d = 2$.

$$\begin{aligned}
&\left[P^{\mathrm{Cov}}_{M_{d,\mathrm{HD}}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1 - dn\left(1-\frac{1}{d^2}\right)^{\frac{2r^2\mu}{3}}\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1 - dn(q+1)\left(1-\frac{1}{d^2}\right)^{\frac{2r^2\mu}{3}}\right] - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1 - dn(q+1)\exp\left(-\frac{2r^2\mu}{3d^2}\right) - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1 - dn(2q+1)\exp\left[-\min\left(\frac{2}{3d^2}, \frac{1}{18}\right)\cdot r^2\mu\right].
\end{aligned}$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x < 0$. $\square$

Corollary 25. *Let us consider a quasi-concave landscape on $M_{2,\mathrm{HD}}$, whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The SES with population size:*

$$\mu \geq \frac{18\ln\left[4n(2q+1)\right]}{r^2}. \tag{6.21}$$

finds a global optimum within $2q$ expected generations and $2\mu q$ expected fitness evaluations with probability at least 0.5.

Proof. The result follows from solving in μ the inequality:

$$1 - 2n(2q+1)\exp\left(-\frac{1}{18}\cdot r^2\mu\right) \geq \frac{1}{2}. \tag{6.22}$$

$\square$

Let one run of the SES be performed in q generations. If the population size satisfies the condition of Corollary 25, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

6.3.2 Manhattan distance

We first estimate the probability $P^{\text{Cov}}_{M_{d,\text{MD}}}(m)$ which is a lower bound on the probability for covering a convex set C of $M_{d,\text{MD}}$ with m samples from C.

Lemma 20. *We assume that $d \geq 2$, for any convex set C of the metric space $M_{d,\text{MD}}$ we have $P^{\text{Cov}}_{C}(m) \geq P^{\text{Cov}}_{M_{d,\text{MD}}}(m)$, where:*

$$P^{\text{Cov}}_{M_{d,\text{MD}}}(m) \geq 1 - 2n\left(1 - \frac{1}{d^2}\right)^m.$$

Proof. We saw in Corollary 24 that schemata using only the symbol $*$ and/or $*_{[k,l]}$ and/or fixed values correspond to a convex set in the metric space $M_{d,\text{MD}}$. In particular, the schema corresponding to the entire search space is the only schema with the largest number of positions that are free to take more than one value. Moreover, each of these free positions take the maximum number of possible values. Therefore, the schema corresponding to any other convex set has at most n symbols $*$.

Let us now compute the probability $\Pr[co(P) = M_{d,\text{MD}} \mid P = \text{NonUnif}_m(M_{d,\text{MD}})]$ for covering the entire search space from sampling m points from it. The schema corresponding to the entire search space is $\underbrace{* * * \cdots *}_{n \text{ times}}$.

The don't care symbol is obtained at some position when each of the values 0 and $d-1$ appears at least once at this position. The probability that a value appears at this position in e_i is $\frac{1}{d^{w_i}}$. The probability that this value never appears at this position in e_i is therefore $1 - \frac{1}{d^{w_i}}$. The probability that this value never appears at this position in e_1, e_2, ..., and e_m is therefore:

$$\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right). \tag{6.23}$$

The probability that the value 0 never appears at this position OR the value $d-1$ never appears at this position is:

$$2\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right) \leq 2\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.24}$$

Hence, the probability that each value appears at least once at that position is:

$$1 - 2\prod_{i=1}^{m}\left(1 - \frac{1}{d^{w_i}}\right) \geq 1 - 2\left(1 - \frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.25}$$

Thus, the probability for obtaining the don't care symbol at n positions is:

$$\left[1-2\prod_{i=1}^{m}\left(1-\frac{1}{d^{w_i}}\right)\right]^n \geq \left[1-2\left(1-\frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m\right]^n, \tag{6.26}$$

$$\geq 1-2n\left(1-\frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m, \tag{6.27}$$

using Bernoulli's inequality in the second line. The probability for obtaining a schema with n' free positions is at least:

$$1-2n'\left(1-\frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.28}$$

As $n' \leq n$, we have:

$$1-2n'\left(1-\frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m \geq 1-2n\left(1-\frac{1}{d^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{6.29}$$

By Corollary 21, the largest w_i is bounded above by two. The result follows. □

Theorem 26. *Let us consider a quasi-concave landscape on $M_{d,\mathrm{MD}}$, whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The SES with population size μ finds a global optimum within at most q generations and μq fitness evaluations with probability at least:*

$$1-2n(2q+1)\exp\left[-\min\left(\frac{2}{3d^2},\frac{1}{18}\right)\cdot r^2\mu\right]. \tag{6.30}$$

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right)$.

$$\begin{aligned}
&\left[P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1-2n\left(1-\frac{1}{d^2}\right)^{\frac{2r^2\mu}{3}}\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1-2n(q+1)\left(1-\frac{1}{d^2}\right)^{\frac{2r^2\mu}{3}}\right] - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1-2n(q+1)\exp\left(-\frac{2r^2\mu}{3d^2}\right) - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1-2n(2q+1)\exp\left[-\min\left(\frac{2}{3d^2},\frac{1}{18}\right)\cdot r^2\mu\right].
\end{aligned}$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x<0$. □

Corollary 26. *Let us consider a quasi-concave landscape on $M_{d,\mathrm{MD}}$, whose canonical level sets are: $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$. Let also $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$. The SES with population size:*

$$\mu \geq \frac{\ln\left[4n(2q+1)\right]}{\min\left(\frac{2}{3d^2}\ ,\frac{1}{18}\right)\cdot r^2}. \tag{6.31}$$

finds a global optimum within $2q$ expected generations and $2\mu q$ expected fitness evaluations with probability at least 0.5.

Proof. The result follows from solving in μ the inequality:

$$1 - 2n(2q+1)\exp\left[-\min\left(\frac{2}{3d^2}\ ,\frac{1}{18}\right)\cdot r^2\mu\right] \geq \frac{1}{2}. \tag{6.32}$$

□

Let one run of the SES be performed in q generations. If the population size satisfies the condition of Corollary 26, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

6.3.3 Application to Leading Ones

Leading Ones (that we denote LO) is a pseudo-boolean function that returns the number of leading ones in a binary string. In the metric space $M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$, the canonical level sets of Leading Ones are of the form:

$$A_{\geq j} = \{x \in \{0,1\}^n \mid \mathrm{LO}(x) \geq j\}, \tag{6.33}$$

$$= \underbrace{11\cdots 1}_{j \text{ times}}\underbrace{**\cdots*}_{n-j \text{ times}}. \tag{6.34}$$

Leading Ones is therefore a quasi-concave problem where $q = n$ and $r = 0.5$ [MS17].

Theorem 27. *In the metric space $M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$, Leading Ones is solved in $O(n \ln n)$ fitness evaluations by the SES when the population size is at least $72\ln[4n(2n+1)]$.*

Proof. We apply the result of Corollary 25 to Leading Ones by replacing q and r with their respective values for Leading Ones and by replacing d with 2. □

6.3.4 Other Applications

The fitness $SX_a(b)$ of a string b is given by the length of the longest suffix of a that is also a suffix of b.

Example 44. *In* $\{0,1,2\}^4$*, let* $a = 2011$ *and let* $b = 1021$*. The longest suffix of* a *that is also a suffix of* b *is* 1*. Hence, the fitness* $SX_a(b)$ *of* b *is* 1*.*

Proposition 16. *Let* a *be a fixed string of* $\{0, 1, \cdots, d-1\}^n$*. The fitness landscape* $(\{0, 1, \cdots, d-1\}^n, SX_a, \mathrm{MD})$ *(resp.* $(\{0, 1, \cdots, d-1\}^n, SX_a, \mathrm{HD})$*) is quasi-concave with parameters* $q = n$ *and* $r = \frac{1}{d}$*.*

Proof. It is enough to define the canonical level sets of the problem and to show that they are convex sets in the metric space $(\{0, 1, \cdots, d-1\}^n, \mathrm{MD})$. Let $a = a_1 a_2 \cdots a_n$, the possible lengths of a suffix of a are: $n,\ n-1,\ \ldots,\ 2,\ 1$, and 0.

Let $A_{\geq j}$ be the canonical level set containing all strings whose fitness value is at least j. This means that an element of $A_{\geq j}$ is of the form:

$$a'_1 \cdots a'_{n-j} a_{n-j+1} \cdots a_{n-1} a_n, \tag{6.35}$$

where $1 \leq j \leq n$. We have the following results, using the same notations as in Theorem 14:

$$\begin{aligned}
A_{\geq 0} &= * * \cdots * \\
A_{\geq 1} &= * * \cdots * \, a_n \\
A_{\geq 2} &= * * \cdots * \, a_{n-1} a_n \\
&\cdots \\
A_{\geq k} &= * * \cdots * \, a_{n-k+1} \cdots a_n \\
&\cdots \\
A_{\geq n} &= \{a_1 a_2 \cdots a_n\}
\end{aligned}$$

By a similar proof to that of Proposition 4), we find that each canonical level set is a convex set in the metric space $(\{0, 1, \cdots, d-1\}^n, \mathrm{MD})$. Moreover, $A_{\geq j+1}$ is always contained in $A_{\geq j}$ by construction. The number $q+1$ of distinct level sets is therefore $n+1$. The smallest ratio r between the sizes of two consecutive canonical level sets is $\frac{1}{d}$. The same reasoning is used along Proposition 2 for the metric space $(\{0, 1, \cdots, d-1\}^n, \mathrm{HD})$). □

Remark 3. *Let* a *be a fixed string of* $\{0, 1, \cdots, d-1\}^n$*. For any string* b *of* $\{0, 1, \cdots, d-1\}^n$*, let* $SP_a(b)$ *be the length of the longest prefix of* a *that is also a prefix of* b*. Using the same reasoning as above, we also find that the fitness landscape* $(\{0, 1, \cdots, d-1\}^n, SP_a, \mathrm{MD})$ *(resp.* $(\{0, 1, \cdots, d-1\}^n, SP_a, \mathrm{HD})$*) is quasi-concave. In particular, for* $d = 2$ *and* $a = 11 \cdots 1$ *the fitness function* SP_a *is Leading Ones.*

Theorem 28. *In the metric space* $(\{0, 1, \cdots, d-1\}^n, \mathrm{MD})$*, if the population size is at least:*

$$\mu \geq \frac{d^2}{\min\left(\frac{2}{3d^2}, \frac{1}{8}\right)} \ln[4n(2n+1)], \tag{6.36}$$

then the SES finds the longest suffix of a fixed string of $\{0, 1, \cdots, d-1\}^n$ *appearing as a suffix, with probability at least* 0.5*.*

Proof. By Proposition 16 and Corollary 26. □

Let one run of the SES be performed in $q = n$ generations. If the population size satisfies the condition of Theorem 28, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q = 2n$ and $2\mu q = 3d^4 n \ln[4n(2n+1)]$ for $d \geq 3$.

6.4 SES versus CS

We compare the SES to the CS by looking at the probability for covering the search space entirely, and the smallest population size for which a global optimum is found.

6.4.1 Hamming distance

$M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$ is the only metric space on strings on finite alphabets where the sets $Seg(A)$ and $co(A)$ coincide for any subset A. As a result, the analysis of the SES can only be done on $M_{2,\mathrm{HD}}$ while the analysis of the CS can be done on $M_{d,\mathrm{HD}}$ for $d \geq 2$.

Probability for covering the search space entirely

The probability $P^{\mathrm{Cov}}_{M_{2,\mathrm{HD}}}(m)$ that the convex hull of m elements sampled uniformly at random from $\{0,1\}^n$ is equal to $\{0,1\}^n$ is bounded below by

$$1 - 2n\left(1 - \frac{1}{4}\right)^m$$

for the SES, while it is bounded below by

$$1 - 2n\left(1 - \frac{1}{2}\right)^m$$

for the CS. The probability for covering $\{0,1\}^n$ entirely is higher for CS.

Population Size Threshold for finding a Global Optimum

We saw that when the population size of the CS (resp. SES) exceeds a certain threshold μ_0, then the probability for finding a global optimum also exceeds 0.5.

In particular, if one run of the CS (resp. SES) is performed in q generations, then the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu_0 q$. We are interested in comparing the population size threshold of the CS to that of the SES.The population size threshold of the SES is

$$\mu_0 = \frac{18 \ln [4n(2q+1)]}{r^2},$$

while the population size threshold of the CS is

$$\mu_0 = \frac{8 \ln[4n(2q+1)]}{r}.$$

The smallest population size threshold is obtained for the CS.

6.4.2 Manhattan distance

The sets $Seg(A)$ and $co(A)$ coincide for any subset A of the metric space $M_{d,\mathrm{MD}}$ for $d \geq 2$. Hence, the analysis of the SES does not require any dimension restriction.

Probability for covering the search space entirely

The probability $P^{\mathrm{Cov}}_{M_{d,\mathrm{MD}}}(m)$ that the convex hull of m elements sampled uniformly at random from $\{0, 1, \ldots, d-1\}^n$ is equal to $\{0, 1, \ldots, d-1\}^n$ is bounded below by

$$1 - 2n\left(1 - \frac{1}{d^2}\right)^m$$

for the SES, while it is bounded below by

$$1 - 2n\left(1 - \frac{1}{d}\right)^m$$

for the CS. The probability for covering $\{0, 1\}^n$ entirely is higher for CS.

Population Size Threshold for finding a Global Optimum

We saw that when the population size of the CS (resp. SES) exceeds a certain threshold μ_0, then the probability for finding a global optimum also exceeds 0.5. In particular, if one run of the CS (resp. SES) is performed in q generations, then the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu_0 q$. We are interested in comparing the population size threshold of the CS to that of the

SES.The population size threshold of the SES is

$$\mu_0 = \frac{\ln\left[4n(2q+1)\right]}{\min\left(\frac{2}{3d^2}\ ,\frac{1}{18}\right)\cdot r^2},$$

while the population size threshold of the CS is

$$\mu_0 = \frac{4d\ln[4n(2q+1)]}{r}.$$

The smallest population size threshold is obtained for the CS when $d \geq 5$.

6.4.3 Summary

Due to the restrictions imposed by the choice of analysis used, the SES has only been studied on $M_{2,\mathrm{HD}}$ for the Hamming distance. For the Manhattan distance however, the SES could be studied on any metric space $M_{d,\mathrm{MD}}$ for $d \geq 2$.

The probability for covering the search space entirely is larger for the CS for both metric spaces. The smallest population size threshold for which a global optimum is found is smaller for the CS in $M_{2,\mathrm{HD}}$. In the metric space $M_{d,\mathrm{MD}}$, the smallest population size threshold for which a global optimum is found:

- depends on the values of d and r when $2 \leq d \leq 4$,
- is obtained for the CS for $d \geq 5$.

In the metric space $(\{0,1\}^n, \mathrm{HD})$, both the CS and the SES solve Leading Ones within $O(n \ln n)$ fitness evaluations.

We conclude that in general, the CS and the SES need not be equivalent on a quasi-concave landscape of strings for the Hamming and the Manhattan distance.

Chapter 7

SES on permutations

We aim to specify the analysis of the SES on quasi-concave landscapes to permutations, in metric spaces where the sets $Seg(A)$ and $co(A)$ coincide for any subset A. However, none of the usual metrics on S_n satisfy this property. Hence, we bring the problem back to strings by considering a bijection ψ between S_n and a subset of $\{0, 1, \ldots, n-1\}^{n-1}$. We will consider the same metrics used in [MS17] for the analysis of the SES:

- The Hamming distance HD,
- The Manhattan distance MD.

We then compare the runtime result of the SES to that of the CS.

We show that the sets $Seg(A)$ and $co(A)$ do not coincide for all subsets of the metric spaces (S_n, T), (S_n, UL) and (S_n, R) in Section 7.1.1. We show that the metric space $(\psi(S_n), \mathrm{MD})$ is the only metric space satisfying the desired property for $n \geq 2$ in Section 7.1.2. The runtime analysis is then done in Section 7.2. Finally, the SES and the CS are compared in Section 7.2.1.

7.1 Convexity of the Set of Reachable solutions

For the needs of our analysis, we must restrict our study to metric spaces where the set of reachable solution is always a convex set. That is, the set $Seg(A)$ always coincides with the set $co(A)$ for any subset A. Hence, we will test various metrics on permutations for this property. We will also test the Hamming and the Manhattan distance, which are metrics on the string-form of permutations (see Section 4.7 of Chapter 4).

7.1.1 Usual metrics on Permutations

We study the convexity of the set of reachable solutions in the metric spaces (S_n, K), (S_n, T), (S_n, UL), and (S_n, R). That is, we determine whether the sets

$Seg(A)$ and $co(A)$ coincide for any subset A.

Kendall's τ metric

We show a condition on a subset A of the metric space (S_n, K) under which the sets $Seg(A)$ and $co(A)$ coincide. Recall that $\prec_w$ denotes the weak order on S_n.

Proposition 17. *Let A be a subset in the metric space (S_n, K). The union $Seg(A)$ of all the segments that can be formed out of the elements of A is equal to the convex hull $co(A)$ of A if both the lowest upper bound (LUB) and the greatest lower bound (GLB) of A (with respect to the weak order $\prec_w$) belong to A.*

Proof. As $\mathrm{LUB}(A) \prec_w \mathrm{GLB}(A)$, we have:

$$[\mathrm{LUB}(A), \mathrm{GLB}(A)]_K = [\mathrm{LUB}(A), \mathrm{GLB}(A)]_{\prec_w}. \tag{7.1}$$

We also have that:

$$A \subseteq [\mathrm{LUB}(A), \mathrm{GLB}(A)]_{\prec_w}. \tag{7.2}$$

By Theorem 4, $[\mathrm{LUB}(A), \mathrm{GLB}(A)]_{\prec_w}$ is a convex set with respect to the Kendall's τ metric. Hence,

$$co(A) \subseteq [\mathrm{LUB}(A), \mathrm{GLB}(A)]_{\prec_w}. \tag{7.3}$$

If both $\mathrm{LUB}(A)$ and $\mathrm{GLB}(A)$ belong to A, then $[\mathrm{LUB}(A), \mathrm{GLB}(A)]_{\prec_w}$ is contained in $Seg(A)$. In this case, $Seg(A)$ and $co(A)$ must coincide as the set $Seg(A)$ is always contained in the set $co(A)$ for any subset A. $\square$

Cayley metric

We show that there exist subsets A of the metric space (S_n, T) such that the sets $Seg(A)$ and $co(A)$ do not coincide. Recall that $\prec$ denotes the strong order on S_n.

Proposition 18. *Let σ and τ be two elements of the metric space (S_n, T) such that $\sigma \prec \tau$. The set $Seg(\{\sigma, \tau\}) = [\sigma, \tau]_T$ need not be equal to the set $co(\{\sigma, \tau\})$.*

Proof. Let σ and τ be two elements of S_n such that $\alpha \prec \tau$, we have:

$$[\sigma, \tau]_T = [\sigma, \tau]_{\prec}. \tag{7.4}$$

By Theorem 5, $[\sigma, \tau]_{\prec}$ need not be a convex set with respect to the Cayley metric. As a result, $Seg(\{\sigma, \tau\})$ need not be equal to $co(\{\sigma, \tau\})$. $\square$

Ulam metric

We show that there exist subsets A of the metric space (S_n, UL) such that the sets $Seg(A)$ and $co(A)$ do not coincide. Recall that $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$ is the set:

$$\{\mathrm{id}\} \cup \bigcup_{k=1}^{UL(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \,\middle|\, \prod_{i=1}^{UL(\sigma,\tau)} \tau_i \in \mathcal{P}_{ac}(\sigma^{-1} \circ \tau) \right\},$$

where $\mathcal{P}_{ac}(\sigma^{-1} \circ \tau)$ is the set of all possible ways to write $\sigma^{-1} \circ \tau$ as a minimal product of adjacent cycles.

Proposition 19. *Let σ and τ be two elements of the metric space (S_n, UL) such that $\tau = \sigma c_1 c_2$, where the c_i are two disjoint adjacent cycles. The set $Seg(\{\sigma, \tau\}) = [\sigma, \tau]_{UL}$ is not equal to $co(\{\sigma, \tau\})$.*

Proof. By Theorem 8, $[\sigma, \tau]_{UL}$ is a convex set iff $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$ is a subgroup of $(S_n, \circ)$. As the cycles c_i are disjoint, the only other way to write the product $c_1 c_2$ is $c_2 c_1$. Thus,

$$\mathcal{A}_{ac}(\sigma^{-1} \circ \tau) = \{\mathrm{id}, c_1, c_2, c_1 c_2\}. \tag{7.5}$$

The set $\mathcal{A}_{ac}(\sigma^{-1} \circ \tau)$ can never be a subgroup of $(S_n, \circ)$. Indeed, the only possible inverse for c_1 is $c_1 c_2$. In this case, the inverse of c_2 has to be c_1^2 and c_1^2 is not in the set. $\square$

Reversal metric

We show that there exist subsets A of the metric space (S_n, R) such that the sets $Seg(A)$ and $co(A)$ do not coincide. Recall that $\mathcal{A}_r(\tau \circ \sigma^{-1})$ is the set:

$$\{\mathrm{id}\} \cup \bigcup_{k=1}^{R(\sigma,\tau)} \left\{ \prod_{i=1}^{k} \tau_i \,\middle|\, \prod_{i=1}^{R(\sigma,\tau)} \tau_i \in \mathcal{P}_r(\tau \circ \sigma^{-1}) \right\},$$

where $\mathcal{P}_r(\tau \circ \sigma^{-1})$ is the set of all possible ways to write $\tau \circ \sigma^{-1}$ as a minimal product of reversals.

Proposition 20. *Let σ and τ be two elements of the metric space (S_n, R) such that $\tau = r_1 r_2 \sigma$, where the r_i are two disjoint reversals. The set $Seg(\{\sigma, \tau\}) = [\sigma, \tau]_R$ is not equal to the set $co(\{\sigma, \tau\})$.*

Proof. By Theorem 9, $[\sigma, \tau]_R$ is a convex set iff $\mathcal{A}_r(\tau \circ \sigma^{-1})$ is a subgroup of $(S_n, \circ)$. Each reversal r_i is a product of disjoint transpositions. As the reversal are themselves disjoint, the only other way to write the product $r_1 r_2$ is $r_2 r_1$. Thus,

$$\mathcal{A}_r(\tau \circ \sigma^{-1}) = \{\mathrm{id}, r_1, r_2, r_1 r_2\}. \tag{7.6}$$

The set $\mathcal{A}_r(\tau \circ \sigma^{-1})$ can never be a subgroup of $(S_n, \circ)$. Indeed, the only possible inverse for r_1 is $r_1 r_2$. In this case, the inverse of r_2 has to be r_1^2 and r_1^2 is not in the set. $\square$

7.1.2 Hamming and Manhattan distances

Recall that a permutation of S_n is uniquely determined by a $(n-1)$-uplet of $\{0,1,\ldots,n-1\} \times \{0,1,\ldots,n-2\} \times \ldots \times \{0,1\}$, which is a strict subset of $\{0,1,\ldots,n-1\}^{n-1}$. The bijection ψ between S_n and $\{0,1,\ldots,n-1\} \times \{0,1,\ldots,n-2\} \times \ldots \times \{0,1\}$ has been defined in Definition 19. By studying $(\psi(S_n), \mathrm{HD})$ (resp. $(\psi(S_n), \mathrm{MD})$), we bring the problem back to strings of the Hamming (resp. Manhattan) space.

We study the convexity of the set of reachable solutions in the metric spaces $(\psi(S_n), \mathrm{HD})$ and $(\psi(S_n), \mathrm{MD})$. That is, we determine whether the sets $Seg(A)$ and $co(A)$ coincide for any subset A.

Hamming distance

We show that the equality of the sets $Seg(A)$ and $co(A)$ depends on n:

Proposition 21. *Let A be a subset in the metric space $(\psi(S_n), \mathrm{HD})$. The set $Seg(A)$ need not be equal to $co(A)$ for $n > 3$.*

Proof. The metric space $(\psi(S_n), \mathrm{HD})$ is contained in the metric space $(\{0,1,\ldots,n-1\}^{n-1}, \mathrm{HD})$. Let A be a subset in the metric space $(\{0,1,\ldots,n-1\}^{n-1}, \mathrm{HD})$. By Corollary 22, the set $Seg(A)$ need not be convex when $n-1 > 2$. As a result, the sets $Seg(A)$ and $co(A)$ need not coincide. $\square$

Manhattan distance

We have the following result:

Proposition 22. *Let A be a subset in the metric space $(\psi(S_n), \mathrm{MD})$. The set $Seg(A)$ is always equal to the set $co(A)$.*

Proof. The metric space $(\psi(S_n), \mathrm{MD})$ is contained in the metric space $(\{0,1,\ldots,n-1\}^{n-1}, \mathrm{MD})$. By Theorem 24, the set $Seg(A)$ is always convex for any subset A of $(\{0,1,\ldots,n-1\}^{n-1}, \mathrm{MD})$. As a result, the sets $Seg(A)$ and $co(A)$ always coincide. $\square$

7.2 Runtime Analysis

Among all the metric spaces considered, we showed that the set of reachable solutions is always convex for the metric spaces: $(\psi(S_2), \mathrm{HD}), (\psi(S_3), \mathrm{HD})$, and

$(\psi(S_n), \mathrm{MD})$. Hence, we will restrict the analysis of the SES to the metric space $(\psi(S_n), \mathrm{MD})$.

We first estimate the probability $P^{\mathrm{Cov}}_{\psi(S_n)}(m)$ which is a lower bound on the probability for covering a convex set C of $(\psi(S_n), \mathrm{MD})$ with m samples from C.

Lemma 21. *We assume that $n \geq 2$, for any convex set C of the metric space $(\psi(S_n), \mathrm{MD})$ we have $P^{\mathrm{Cov}}_{C}(m) \geq P^{\mathrm{Cov}}_{\psi(S_n)}(m)$, where:*

$$P^{\mathrm{Cov}}_{\psi(S_n)}(m) \geq 1 - 2(n-1)\left(1 - \frac{1}{n^2}\right)^m.$$

Proof. We saw in Corollary 17 that schemata corresponding to a convex set of the metric space $(\psi(S_n), \mathrm{MD})$, only use a symbol of the form $*_{[k,l]}$ where $[k,l] \subseteq [0, n-i]$ at position $1 \leq i \leq n-1$. The schema corresponding to the entire search space is $* *_{[0,n-2]} *_{[0,n-3]} \cdots *_{[0,1]}$. It is the only schema with the largest number of positions that are free to take more than one value in the metric space $(\psi(S_n), \mathrm{MD})$. Moreover, each of its free positions takes the maximum number of possible values. Therefore, the schema corresponding to any other convex set has at most $n-i$ possible values at position i (counting starts at 1).

Let us now compute the probability $\Pr[co(P) = \psi(S_n) \mid P = \mathrm{NonUnif}_m(S_n)]$ for covering the entire search space from sampling m points from it.

The symbol $*_{[0,n-i]}$ is obtained at position i when each of the values 0 and $n-i$ appears at least once at this position. The probability that a value appears at this position in a sample e_i is $\frac{1}{n^{w_i}}$, where w_i denotes the weight of the sample e_i. The probability that this value never appears at this position in e_i is therefore $1 - \frac{1}{n^{w_i}}$. The probability that this value never appears at this position in m samples e_1, e_2, ..., and e_m is therefore:

$$\prod_{i=1}^{m}\left(1 - \frac{1}{n^{w_i}}\right). \tag{7.7}$$

The probability that the value 0 never appears at this position OR the value $n-i$ never appears at this position is:

$$2\prod_{i=1}^{m}\left(1 - \frac{1}{n^{w_i}}\right) \leq 2\left(1 - \frac{1}{n^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{7.8}$$

Hence, the probability that each value appears at least once at that position is:

$$1 - 2\prod_{i=1}^{m}\left(1 - \frac{1}{n^{w_i}}\right) \geq 1 - 2\left(1 - \frac{1}{n^{\max_{1\leq i\leq m} w_i}}\right)^m. \tag{7.9}$$

Thus, the probability for obtaining the symbol $*_{[0,n-i]}$ at each position $1 \leq i \leq n-1$

is:

$$\left[1-2\prod_{i=1}^{m}\left(1-\frac{1}{n^{w_i}}\right)\right]^{n-1} \geq \left[1-2\left(1-\frac{1}{n^{\max_{1\leq i\leq m} w_i}}\right)^{m}\right]^{n-1}, \tag{7.10}$$

$$\geq 1-2(n-1)\left(1-\frac{1}{n^{\max_{1\leq i\leq m} w_i}}\right)^{m}, \tag{7.11}$$

using Bernoulli's inequality in the second line. By Corollary 21, the largest w_i is bounded above by two. The result follows. □

Theorem 29. *Let us consider a quasi-concave landscape on* $(\psi(S_n), \mathrm{MD})$*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The SES with population size* μ *finds a global optimum within at most* q *generations and* μq *fitness evaluations with probability at least:*

$$1-2(n-1)(2q+1)\exp\left[-\min\left(\frac{2}{3n^2}\,,\frac{1}{18}\right)\cdot r^2\mu\right]. \tag{7.12}$$

Proof. We estimate a lower bound on $\left[P^{\mathrm{Cov}}_{\psi(S_n)}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right)$.

$$\begin{aligned}
&\left[P^{\mathrm{Cov}}_{M_{\psi(S_n)}}\left(\frac{2r^2\mu}{3}\right)\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1-2(n-1)\left(1-\frac{1}{n^2}\right)^{\frac{2r^2\mu}{3}}\right]^{q+1} - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq \left[1-2(n-1)(q+1)\left(1-\frac{1}{n^2}\right)^{\frac{2r^2\mu}{3}}\right] - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1-2(n-1)(q+1)\exp\left(-\frac{2r^2\mu}{3n^2}\right) - q\exp\left(-\frac{r^2\mu}{18}\right),\\
&\geq 1-2(n-1)(2q+1)\exp\left[-\min\left(\frac{2}{3n^2}\,,\frac{1}{18}\right)\cdot r^2\mu\right].
\end{aligned}$$

The third line follows from Bernouilli's inequality. The fourth line is due to the fact that $\ln(1+x)$ is bounded above by x whenever $x<0$. □

Corollary 27. *Let us consider a quasi-concave landscape on* $(\psi(S_n), \mathrm{MD})$*, whose canonical level sets are:* $A_{\geq 0},\ A_{\geq 1},\ \ldots,\ A_q$*. Let also* $r = \min_{0\leq j\leq q} \frac{|A_{\geq j+1}|}{|A_{\geq j}|}$*. The SES with population size:*

$$\mu \geq \frac{\ln\left[4(n-1)(2q+1)\right]}{\min\left(\frac{2}{3n^2}\,,\frac{1}{18}\right)\cdot r^2}. \tag{7.13}$$

finds a global optimum within $2q$ *expected generations and* $2\mu q$ *expected fitness evaluations with probability at least 0.5.*

Proof. The result follows from solving in μ the inequality:

$$1 - 2(n-1)(2q+1)\exp\left[-\min\left(\frac{2}{3n^2}\,,\frac{1}{18}\right)\cdot r^2\mu\right] \geq \frac{1}{2}. \tag{7.14}$$

□

Let one run of the SES be performed in q generations. If the population size satisfies the condition of Corollary 27, then the expected number of runs before finding a global optimum (i.e., the expected hitting time) is at most $\frac{1}{0.5} = 2$. Hence, the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu q$.

7.2.1 SES versus CS

We compare the SES to the CS by looking at the probability for covering the search space entirely, and the smallest population size for which a global optimum is found.

Probability for covering the search space entirely

The probability $P^{\text{Cov}}_{\psi(S_n)}(m)$ that the convex hull of m elements sampled uniformly at random from $\psi(S_n)$ is equal to $\psi(S_n)$ is bounded below by

$$1 - 2(n-1)\left(1 - \frac{1}{n^2}\right)^m$$

for the SES, while it is bounded below by

$$1 - 2(n-1)\left(1 - \frac{1}{n}\right)^m$$

for the CS. The probability for covering $\psi(S_n)$ entirely is higher for CS.

Population Size Threshold for finding a Global Optimum

We saw that when the population size of the CS (resp. SES) exceeds a certain threshold μ_0, then the probability for finding a global optimum also exceeds 0.5. In particular, if one run of the CS (resp. SES) is performed in q generations, then the expected number of generations and the expected number of fitness evaluations needed for finding a global optimum are respectively $2q$ and $2\mu_0 q$. We are interested in comparing the population size threshold of the CS to that of the SES.The population size threshold of the SES is

$$\mu_0 = \frac{\ln\left[4(n-1)(2q+1)\right]}{\min\left(\frac{2}{3n^2}\,,\frac{1}{18}\right)\cdot r^2},$$

while the population size threshold of the CS is

$$\mu_0 = \frac{2n}{r} \ln[8n(q+1)].$$

The smallest population size threshold is obtained for the CS.

7.3 Summary

Due to the restrictions imposed by the choice of analysis used, the SES has only been studied on $(\psi(S_n), \mathrm{MD})$. The probability for covering the search space entirely is larger for the CS. Moreover, the smallest population size threshold for which a global optimum is found is obtained for the CS. We conclude that in general, the CS and the SES need not be equivalent.

Chapter 8

Generalized $(1+1)$ EA

We aim to introduce a unifying runtime analysis of $(1+1)$ EA across representations, by:

- defining a generalization of $(1+1)$ EA across representations,
- defining a class of fitness landscapes that are solved in polynomial time by instantiations of the generalized $(1+1)$ EA.

To this end, we will use metric spheres and metric balls whose notions are recalled in Section 8.1. The class of improving fitness landscapes that are parametrized by a radius $\rho > 0$ is defined in Section 8.2. The generalization of $(1+1)$ EA across representations is given in Section 8.3. Finally, the runtime analysis is done in Section 8.4. We find that the expected runtime of the generalized $(1+1)$ EA on a polynomially ρ-improving fitness landscape with at most polynomially many level sets, is at most polynomial for a well chosen mutation parameter. The runtime result is specified to metric spaces of strings on a finite alphabet and to a metric space of permutations.

8.1 Spheres and Balls

We recall the notions of sphere and ball in a metric space $(\mathcal{S}, D)$, that are central to the study of the generalized $(1+1)$ EA.

Definition 24 (Sphere)**.** *Let $R \geq 0$ and let x be a point of the metric space $(\mathcal{S}, D)$. The sphere of radius R centred at x is the set of points of $\mathcal{S}$ whose distance to x equals R:*

$$S_x(R) = \{y \in \mathcal{S} \mid D(x, y) = R\}. \tag{8.1}$$

The notion of sphere can then be used to define the notion of ball.

Definition 25 (Ball)**.** *Let $R \geq 0$ and let x be a point of the metric space $(\mathcal{S}, D)$. The (closed) ball of radius R centred at x is the set of points of $\mathcal{S}$ whose distance* *to x is less than or equal to R:*

$$B_x(R) = \{y \in \mathcal{S} \mid D(x,y) \leq R\}, \tag{8.2}$$
$$= \bigcup_{0\leq l\leq R} \{y \in \mathcal{S} \mid D(x,y) = l\}, \tag{8.3}$$
$$= \bigcup_{0\leq l\leq R} S_x(l). \tag{8.4}$$

8.1.1 Strings on a finite alphabet

Let $d \geq 2$, we shall determine the spheres and the balls of the metric space $M_{d,D} = (\{0,1,\ldots,d-1\}^n, D)$ for two different metrics:

- the Hamming distance,
- the Manhattan distance.

Hamming distance

Let x be an element of the metric space $M_{d,\mathrm{HD}}$. The sphere $S_x(R)$ is the set of strings of length n that differ from x by exactly R positions.

Example 45. *Let $n = 3$, and $x = 002$ is an element of $M_{3,\mathrm{HD}}$, we have:*

$$S_x(2) = *_{12} *_{12} 2 \ \cup\ *_{12}0 *_{01} \ \cup 0 *_{12} *_{01}.$$

The ball $B_x(2)$ is therefore given by:

$$\begin{aligned} B_x(2) &= \bigcup_{0\leq l\leq 2} S_x(2), \\ &= S_x(0) \ \cup\ S_x(1) \ \cup\ S_x(2), \\ &= \{002\} \ \cup\ (*_{12}02 \ \cup\ 0 *_{12} 2 \ \cup\ 00*_{01}) \ \cup\ (*_{12} *_{12} 2 \ \cup\ *_{12}0 *_{01} \ \cup 0 *_{12} *_{01}). \end{aligned}$$

Manhattan distance

Let x be an element of the metric space $M_{d,\mathrm{MD}}$. The sphere $S_x(R)$ is the set of strings y of length n that satisfy:

$$\sum_{i=1}^{n} \mid x_i - y_i \mid = R. \tag{8.5}$$

Example 46. *Let $n = 3$, and $x = 002$ is an element of $M_{3,\mathrm{MD}}$, we have:*

$$S_x(2) = \{202, 022, 112, 011, 101\}.$$

The ball $B_x(2)$ is therefore given by:

$$
\begin{aligned}
B_x(2) &= \bigcup_{0 \le l \le 2} S_x(2), \\
&= S_x(0) \,\cup\, S_x(1) \,\cup\, S_x(2), \\
&= \{002\} \,\cup\, \{001, 102, 012\} \,\cup\, \{202, 022, 112, 011, 101\}.
\end{aligned}
$$

8.1.2 Permutations

We shall determine the spheres and the balls of the metric space (S_n, D), for the following metrics:

- The Kendall's τ metric,
- The Cayley metric,
- The Ulam metric,
- The reversal distance.

Kendall's τ metric

Let σ be an element of the metric space (S_n, K). The sphere $S_\sigma(R)$ is the set of permutations of S_n that are obtained from σ by swapping a minimum of R pairs of values whose positions are adjacent (see Definition 10).

Example 47. *Let $\sigma = 213$ be an element of (S_3, K), we have:*

$$S_\sigma(2) = \{132, 321\}.$$

The ball $B_\sigma(2)$ is therefore given by:

$$
\begin{aligned}
B_\sigma(2) &= \bigcup_{0 \le l \le 2} S_\sigma(2), \\
&= S_\sigma(0) \,\cup\, S_\sigma(1) \,\cup\, S_\sigma(2), \\
&= \{213\} \,\cup\, \{123, 231\} \,\cup\, \{132, 321\}.
\end{aligned}
$$

Cayley metric

Let σ be an element of the metric space (S_n, T). The sphere $S_\sigma(R)$ is the set of permutations of S_n that are obtained from σ by swapping a minimum of R pairs of values (see Definition 11).

Example 48. *Let $\sigma = 213$ be an element of (S_3, T), we have:*

$$S_\sigma(2) = \{132, 321\}.$$

The ball $B_\sigma(2)$ is therefore given by:

$$\begin{aligned} B_\sigma(2) &= \bigcup_{0\leq l\leq 2} S_\sigma(2), \\ &= S_\sigma(0) \ \cup \ S_\sigma(1) \ \cup \ S_\sigma(2), \\ &= \{213\} \ \cup \ \{123, 231, 312\} \ \cup \ \{132, 321\}, \\ &= S_3. \end{aligned}$$

Ulam metric

Let σ be an element of the metric space (S_n, UL). The sphere $S_\sigma(R)$ is the set of permutations of S_n that are obtained from σ by performing a minimum of R character displacements (see Definition 12).

Example 49. *Let $\sigma = 213$ be an element of (S_3, UL), we have:*

$$S_\sigma(2) = \{312\}.$$

The ball $B_\sigma(2)$ is therefore given by:

$$\begin{aligned} B_\sigma(2) &= \bigcup_{0\leq l\leq 2} S_\sigma(2), \\ &= S_\sigma(0) \ \cup \ S_\sigma(1) \ \cup \ S_\sigma(2), \\ &= \{213\} \ \cup \ \{123, 231, 132, 321\} \ \cup \ \{312\}, \\ &= S_3. \end{aligned}$$

Reversal distance

Let σ be an element of the metric space (S_n, R). The sphere $S_\sigma(R)$ is the set of permutations of S_n that are obtained from σ by performing a minimum of R reversals (see Definition 13).

Example 50. *Let $\sigma = 213$ be an element of (S_3, R), we have:*

$$S_\sigma(2) = \{132, 321\}.$$

The ball $B_\sigma(2)$ is therefore given by:

$$\begin{aligned} B_\sigma(2) &= \bigcup_{0\leq l\leq 2} S_\sigma(2), \\ &= S_\sigma(0) \ \cup \ S_\sigma(1) \ \cup \ S_\sigma(2), \\ &= \{213\} \ \cup \ \{123, 231, 312\} \ \cup \ \{132, 321\}, \\ &= S_3. \end{aligned}$$

Notice that the spheres and the balls of the metric space (S_n, R) coincide with the spheres and the balls of the metric space (S_n, T) for $n = 3$. This need not be true for n larger than or equal to 4.

8.2 ρ-improving fitness landscapes

Recall that a landscape is rugged if the number of local optima is high [GT12]. We define a parametrized class of fitness landscapes that that captures the number of local optima per metric ball of some fixed radius. We shall consider two distinct parameters:

- A radius ρ that delimits the size of the metric balls of interest,
- The minimal number of strictly fitter solutions in a metric ball of radius ρ. Fitnesses are compared to the fitness of the centre of the ball.

We will start with the following definition:

Definition 26. *Let $\rho > 0$, a solution x in a fitness landscape $(\mathcal{S}, f, D)$ is ρ-improving iff there exists a strictly fitter solution y (with respect to x) on the sphere $S_x(\rho)$.*

We can use the word *increasing* (resp. *decreasing*) instead of the word *improving* for a maximizing (resp. minimizing) problem.

Example 51. *In the fitness landscape $(\{0, 1\}^3, \mathrm{LeadingOnes}, \mathrm{HD})$, the solution 010 is 1-increasing. Indeed, 110 is a strictly fitter solution than 010 and $\mathrm{HD}(010, 110) = 1$.*

Definition 27. *A fitness landscape $(\mathcal{S}, f, D)$ is ρ-improving iff all its elements that can improve (i.e., that are not global optima) are l-improving for $0 < l \leq \rho$.*

Indeed, a global maximum (resp. minimum) can not get any more fitter for a maximizing (resp. minimizing) problem. Let $\mathrm{Imp}_x(l)$ denote the number of solutions y that are strictly fitter than x and such that $D(x, y) = l$.

Definition 28. *A polynomially ρ-improving landscape is a ρ-improving landscape such that the number of strictly fitter solutions on a sphere of radius l is at least a polynomial fraction of the size of the sphere $S_x(l)$ for $0 < l \leq \rho$.*

The number of local optima in a ball of radius ρ of a polynomially ρ-improving fitness landscape, is at least a polynomial fraction of the size of the largest ball of radius ρ.

We shall now give examples of polynomially ρ-improving landscapes from the litterature.

8.2.1 Strings on a finite alphabet

Let $d \geq 2$, we shall consider fitness landscapes on the metric space $M_{d,\mathrm{HD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$.

Leading Ones

Leading Ones returns the largest number of successive ones starting from the first position of a binary string.

Lemma 22. *Let x be an element of the fitness landscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$. The number of solutions y that are strictly fitter than x and satisfying $\mathrm{HD}(x, y) = 1$, is denoted $\mathrm{Imp}_x(1)$. We have:*

$$\mathrm{Imp}_x(1) \geq 1, \tag{8.6}$$

for any solution x that is not the global maximum.

Proof. We shall determine the minimal number of strictly fitter solutions, with respect to an element x of the fitness lanscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$, on the sphere $S_x(1)$. We will consider a maximizing problem.

In the metric space $M_{2,\mathrm{HD}}$, the sphere $S_x(1)$ has $\binom{n}{1}$ distinct elements. Indeed, $\mathrm{HD}(x, y) = 1$ iff x and y differ in 1 position. Moreover, there is $2 - 1$ value to choose from for each differing position of y (the value that is already taken by x must be removed). Among these $\binom{n}{1}$ elements of $S_x(1)$, a strictly fitter (or higher) solution y is obtained when $\mathrm{LeadingOnes}(y)$ is strictly larger than $\mathrm{LeadingOnes}(x)$. We shall determine the smallest number of strictly higher solutions that can be obtained overall x in the fitness landscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$.

- If $x_i = 1$ for all positions i, then x is the global maximum and $\mathrm{Imp}_x(1) = 0$.
- If there exists at least least one position i_0 such that $x_{i_0} = 0$, then $\mathrm{Imp}_x(1) \geq 1$.

Thus, $\mathrm{Imp}_x(1) \geq 1$ unless x is the global maximum. □

Theorem 30. *The fitness landscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$ is polynomially 1-increasing.*

Proof. We estimate the ratio $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ for any solution x that is not a global maximum, by using the results of Lemma 22. We have:

$$\begin{aligned}\frac{\mid S_x(1) \mid}{\mathrm{Imp}_x(1)} &= \frac{n}{\mathrm{Imp}_x(1)}, \\ &\leq n.\end{aligned}$$

Hence, $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ is at most polynomial in n for any solution x that is not a global maximum. The result follows. □

Linear functions

Let $w_0,\ w_1,\ \dots,\ w_n$ be non-zero real numbers. We consider the following linear function [DJS11]:

$$\begin{aligned} f:\{0,1,\dots,d-1\}^n &\longrightarrow && \mathbb{R}, \\ x &\longmapsto && w_0+\sum_{i=1}^{n} w_i x_i. \end{aligned}$$

Lemma 23. *Let x be an element of the fitness landscape $(\{0,1,\dots,d-1\}^n, f, \mathrm{HD})$. The number of solutions y that are strictly fitter than x and satisfying $\mathrm{HD}(x,y)=l$, is denoted $\mathrm{Imp}_x(l)$. We have:*

- *If $x_i=0$ for all $w_i<0$ and $x_i=d-1$ for all $w_i>0$, then x is a global maximum and $\mathrm{Imp}_x(l)=0$.*

- *If $x_i=d-1$ for all $w_i>0$ and there exists at least one $w_{i_0}<0$ such that $x_{i_0}\neq 0$, then:*
$$\mathrm{Imp}_x(l) \geq \binom{\#\{w_i<0\}}{l}. \tag{8.7}$$

- *If $x_i=0$ for all $w_i<0$ and there exists at least one $w_{i_0}\geq 0$ such that $x_{i_0}\neq d-1$, then:*
$$\mathrm{Imp}_x(l) \geq \binom{\#\{w_i>0\}}{l}. \tag{8.8}$$

- *If there exists at least one $w_{i_0}<0$ such that $x_{i_0}\neq 0$ and there exists at least one $w_{i_1}\geq 0$ such that $x_{i_1}\neq d-1$, then:*
$$\mathrm{Imp}_x(l) \geq \binom{\#\{w_i<0\}}{l} + \binom{\#\{w_i>0\}}{l}. \tag{8.9}$$

Proof. We shall determine the minimal number of strictly fitter solutions, with respect to an element x of the fitness lanscape $(\{0,1,\dots,d-1\}^n, f, \mathrm{HD})$, on each of the spheres $S_x(l)$ for $0<l\leq n$. We will consider a maximizing problem.

Let us fix a radius l. In the metric space $M_{d,\mathrm{HD}}$, the sphere $S_x(l)$ has $(d-1)^l\binom{n}{l}$ distinct elements. Indeed, $\mathrm{HD}(x,y)=l$ iff x and y differ in l positions. Moreover, there are $d-1$ values to choose from for each differing position of y (the value that is already taken by x must be removed). Among these $(d-1)^l\binom{n}{l}$ elements of $S_x(l)$, a strictly fitter (or higher) solution y is obtained when $f(y)$ is strictly larger than $f(x)$. We shall determine the smallest number of strictly higher solutions that can be obtained overall x in the fitness landscape $(\{0,1,\dots,d-1\}^n, f, \mathrm{HD})$. We have:

$$f(x) = w_0 + \sum_{i=1}^{n} w_i x_i, \tag{8.10}$$

$$= w_0 + \sum_{w_i>0} \mid w_i \mid x_i - \sum_{w_i<0} \mid w_i \mid x_i. \tag{8.11}$$

We aim to increase the value of $f(x)$ by changing the values at l positions of x. If the position changed corresponds to a positive coefficient w_i, then x_i must take a larger value. There are therefore $(d-1) - x_i$ possible new values for a position corresponding to a positive coefficient w_i. If the position changed corresponds to a negative coefficient w_i, then x_i must take a smaller value. There are therefore x_i possible new values for a position corresponding to a negative coefficient w_i. The l changing positions are taken from the n positions of x. A position either corresponds to a positive coefficient, or to a negative coefficient.

The number of strictly higher solutions y with respect to x such that $\mathrm{HD}(x, y) = l$ is therefore given by:

$$\begin{aligned}
&\mathrm{Imp}_x(l) \\
&= \sum_{k=0}^{l} \left\{ \left(\sum_{w_i<0} x_i \right)^{l-k} \left(\sum_{w_i>0} (d-1) - x_i \right)^{k} \binom{\#\{w_i < 0\}}{l-k} \binom{\#\{w_i > 0\}}{k} \right\}, \\
&\geq \left(\sum_{w_i<0} x_i \right)^{l} \binom{\#\{w_i < 0\}}{l} + \left(\sum_{w_i>0} (d-1) - x_i \right)^{l} \binom{\#\{w_i > 0\}}{l}.
\end{aligned}$$

The result follows. □

Theorem 31. *The fitness landscape* $(\{0, 1, \dots, d-1\}^n, f, \mathrm{HD})$ *is polynomially* ρ*-increasing for any constant* $\rho \geq 1$.

Proof. We estimate the ratio $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ for any radius $0 < l \leq n$ and any solution x that is not a global maximum, by using the results of Lemma 23. Then, we determine the values of l for which $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ is at most polynomial in n.

- If $x_i = d - 1$ for all $w_i > 0$ and there exists at least one $w_{i_0} < 0$ such that $x_{i_0} \neq 0$, then:

$$\begin{aligned}
\frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)} &\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i<0\}}{l}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i<0\}}{1}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\#\{w_i < 0\}}, \\
&\leq (d-1)^l \binom{n}{l}.
\end{aligned}$$

- If $x_i = 0$ for all $w_i < 0$ and there exists at least one $w_{i_0} \geq 0$ such that $x_{i_0} \neq d-1$, then:

$$\begin{aligned}
\frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)} &\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i>0\}}{l}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i>0\}}{1}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\#\{w_i > 0\}}, \\
&\leq (d-1)^l \binom{n}{l}.
\end{aligned}$$

- If there exists at least one $w_{i_0} < 0$ such that $x_{i_0} \neq 0$ and there exists at least one $w_{i_1} \geq 0$ such that $x_{i_1} \neq d-1$, then:

$$\begin{aligned}
\frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)} &\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i<0\}}{l} + \binom{\#\{w_i>0\}}{l}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\binom{\#\{w_i<0\}}{1} + \binom{\#\{w_i>0\}}{1}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{\#\{w_i < 0\} + \#\{w_i > 0\}}, \\
&\leq \frac{(d-1)^l \binom{n}{l}}{2}.
\end{aligned}$$

It follows that $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ is polynomial in n if l is a constant. □

Remark 4. *The fitness landscape* $(\{0,1,\ldots,d-1\}^n, f, \mathrm{HD})$ *is also polynomially* ρ*-decreasing for any constant* $\rho \geq 1$*. This can be seen by noticing that the number of strictly lower solutions* y *with respect to a solution* x *such that* $\mathrm{HD}(x,y) = l$ *is given by:*

$$\begin{aligned}
&\mathrm{Imp}_x(l) \\
=&\sum_{k=0}^{l} \left\{ \left(\sum_{w_i<0} x_i \right)^{l-k} \left(\sum_{w_i>0} (d-1) - x_i \right)^{k} \binom{\#\{w_i > 0\}}{l-k} \binom{\#\{w_i < 0\}}{k} \right\}.
\end{aligned}$$

Hence, the global minimum is obtained when $x_i = 0$ *for all* $w_i > 0$*, and* $x_i = d-1$ *for all* $w_i < 0$*.*

Case of OneMax

OneMax is a linear function where $d = 2$, $w_0 = 0$, and $w_i = 1$ for all $1 \leq i \leq n$.

Lemma 24. *Let x be an element of the fitness landscape $(\{0,1\}^n, \mathrm{OneMax}, \mathrm{HD})$. The number of solutions y that are strictly fitter than x and satisfying $\mathrm{HD}(x,y) = l$, is denoted $\mathrm{Imp}_x(l)$. We have:*

- *If $x_i = 0$ for all w_i, then x is a global maximum and $\mathrm{Imp}_x(l) = 0$.*
- *If there exists at least one w_{i_0} such that $x_{i_0} \neq 1$, then:*

$$\mathrm{Imp}_x(l) \geq \binom{\#\{w_i > 0\}}{l}. \tag{8.12}$$

Proof. Similar to the proof of Lemma 23. □

Theorem 32. *The fitness landscape $(\{0,1\}^n, \mathrm{OneMax}, \mathrm{HD})$ is polynomially ρ-increasing for any constant $\rho \geq 1$.*

Proof. We estimate the ratio $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ for any radius $0 < l \leq n$ and any solution x that is not a global maximum, by using the results of Lemma 24. Then, we determine the values of l for which $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ is at most polynomial in n.

If there exists at least one w_{i_0} such that $x_{i_0} \neq d-1$, then:

$$\begin{aligned}
\frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)} &\leq \frac{\binom{n}{l}}{\binom{\#\{w_i>0\}}{l}}, \\
&\leq \frac{\binom{n}{l}}{\binom{\#\{w_i>0\}}{1}}, \\
&\leq \frac{\binom{n}{l}}{\#\{w_i > 0\}}, \\
&= \frac{\binom{n}{l}}{n}.
\end{aligned}$$

It follows that $\frac{\mathrm{Imp}_x(l)}{|S_x(l)|}$ is at least a polynomial fraction in n for any constant $l \geq 1$. □

Pseudo-Boolean monotone increasing polynomial

Let the w_A's be positive real numbers. We consider the following pseudo Boolean monotone increasing polynomial [WW05]:

$$\begin{aligned}
f : \{0,1\}^n &\longrightarrow \mathbb{R}, \\
x &\longmapsto \sum_{A \subseteq \{1,2,\ldots,n\}} w_A \prod_{i \in A} x_i.
\end{aligned}$$

The degree of a monomial $\prod_{i \in A} x_i$ is simply $\mid A \mid$. Whereas, the degree of f is given by:

$$\deg(f) = \max_{w_A > 0} \mid A \mid. \tag{8.13}$$

We recall the notions of *activation* and *deactivation* used in [WW05]. A monomial $\prod_{i \in A} x_i$ is activated when all its x_i's are 1's. A monomial is deactivated when at least one of its x_i's is 0.

Lemma 25. *Let x be an element of the fitness landscape $(\{0,1\}^n, f, \mathrm{HD})$. The number of solutions y that are strictly fitter than x and satisfying $\mathrm{HD}(x,y) = l$, is denoted $\mathrm{Imp}_x(l)$. We have:*

- *If x deactivates all the monomials, then x is a global minimum and*

$$\mathrm{Imp}_x(l) = 0.$$

- *If x activates at least one monomial, then*

$$\mathrm{Imp}_x(l) \geq \binom{\min_{w_A > 0} \mid A \mid}{l}. \tag{8.14}$$

Proof. We shall determine the minimal number of strictly fitter solutions, with respect to an element x of the fitness lanscape $(\{0,1\}^n, f, \mathrm{HD})$, on each of the spheres $S_x(l)$ for $0 < l \leq n$. We will consider a minimizing problem.

Let us fix a radius l. In the metric space $M_{2,\mathrm{HD}}$, the sphere $S_x(l)$ has $\binom{n}{l}$ distinct elements. Indeed, $\mathrm{HD}(x,y) = l$ iff x and y differ in l positions. Moreover, there is only one value to choose from for each differing position of y (the value that is already taken by x must be removed). Among these $\binom{n}{l}$ elements of $S_x(l)$, a strictly fitter (or lower) solution y is obtained when $f(y)$ is strictly smaller than $f(x)$. We shall determine the smallest number of strictly lower solutions that can be obtained overall x in the fitness landscape $(\{0,1\}^n, f, \mathrm{HD})$. We have:

$$f(x) = \sum_{A \subseteq \{1,2,\ldots,n\}} w_A \prod_{i \in A} x_i. \tag{8.15}$$

We aim to decrease the value of $f(x)$ by changing the values at l positions of x. If the position changed corresponds to an activated monomial (i.e., x_i appears in at least one of the activated monomials), then x_i is a 1. Changing x_i into a 0 deactivates all activated monomials in which it appears. Moreover, changing x_i into a 0 can not activate any of the deactivated monomials in which it appears. Therefore, changing a value at a position corresponding to an activated monomial always yields a solution with a smaller fitness. If the position changed does not correspond to an activated monomial (i.e., x_i does not appear in any of the activated monomials), then changing x_i can not deactivate any of the activated monomials. However, changing x_i may activate the deactivated monomials in which it appears. Consequently, changing a value at a position that does not correspond to any activated monomial never yields a solution with a smaller fitness. A position either corresponds to an activated monomial (A.M.) or not. The number of strictly lower solutions y with respect to x such that $\mathrm{HD}(x, y) = l$ is therefore given by:

$$\mathrm{Imp}_x(l) = \binom{\#\{x_i \mid x_i \text{ appears in an A.M.}\}}{l}.$$

We obtain the following results:

- If x deactivates all the monomials, then x is a global minimum and the set of A.M. is empty. Thus, $\mathrm{Imp}_x(l) = 0$.

- If x activates at least one monomial, then $\#\{x_i \mid x_i \text{ appears in an A.M.}\}$ is at least the degree of this activated monomial. Thus,

$$\mathrm{Imp}_x(l) \geq \binom{\min_{w_A > 0} \mid A \mid}{l}. \tag{8.16}$$

□

Theorem 33. *The fitness landscape* $(\{0,1\}^n, f, \mathrm{HD})$ *is polynomially* ρ*-decreasing for any constant* $\rho \geq 1$*.*

Proof. We estimate the ratio $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ for any radius $0 < l \leq n$ and any solution x that is not a global minimum, by using the results of Lemma 25. Then, we determine the values of l for which $\frac{|S_x(l)|}{\mathrm{Imp}_x(l)}$ is at most polynomial in n.

If x activates at least one monomial, then:

$$\begin{aligned}
\frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)} &= \frac{\binom{n}{l}}{\binom{\min_{w_A>0}|A|}{l}}, \\
&= \frac{n(n-1)\ldots(\min_{w_A>0} \mid A \mid +1)}{(n-l)(n-l-1)\ldots(\min_{w_A>0} \mid A \mid -l+1)}.
\end{aligned}$$

$\frac{\mathrm{Imp}_x(l)}{|S_x(l)|}$ is at least a polynomial fraction in n if l is a positive constant. □

Remark 5. *The fitness landscape* $(\{0,1\}^n, f, \mathrm{HD})$ *is not* ρ*-increasing for any* $\rho < \min_{w_A>0} \mid A \mid$*. Indeed, increasing the fitness consists of activating at least one monomial. However, all the factors of a monomial must be set to* 1 *in order to activate it. This means that the shortest distance between a solution* x *and a strictly higher solution* y *is* $\min_{w_A>0} \mid A \mid$*.*

8.2.2 Permutations

We shall consider a fitness landscape on the metric space (S_n, R), where R is the reversal distance (see Definition 13).

Euclidean TSP (cities in convex position)

We consider a set of n cities, that are in convex position in the Euclidean space. In other words, each city is a vertex of the convex hull formed by the n cities in the Euclidean space $(\mathbb{R}^2, d)$.

The Euclidean TSP consists of minimizing the fitness function:

$$\begin{aligned} f : S_n &\longrightarrow \mathbb{R}, \\ \sigma &\longmapsto d(x_{\sigma(n)}, x_{\sigma(1)}) + \sum_{i=1}^{n-1} d(x_{\sigma(i)}, x_{\sigma(i+1)}). \end{aligned}$$

[DHO$^+$06] showed that all permutations that respect the ordering of the cities around their convex hull are global minima.

Lemma 26. *Let σ be an element of the fitness landscape (S_n, f, R). The number of solutions τ that are strictly fitter than σ and satisfying $R(\sigma, \tau) = 1$, is denoted $\mathrm{Imp}_\sigma(1)$. We have:*

$$\mathrm{Imp}_\sigma(1) \geq 1, \tag{8.17}$$

for any solution σ that is not a global minimum.

Proof. We shall determine the minimal number of strictly fitter solutions, with respect to an element σ of the fitness lanscape (S_n, f, R), on the sphere $S_\sigma(1)$. We will consider a minimizing problem.

In the metric space (S_n, R), the sphere $S_\sigma(1)$ has $\binom{n}{2}$ distinct elements. Indeed, $R(\sigma, \tau) = 1$ iff σ is obtained from τ through one reversal. As a reversal reverses the order of the x_i between two distinct positions, the number of reversals that can be applied to a permutation of S_n is $\binom{n}{2}$. Among these $\binom{n}{2}$ elements of $S_\sigma(1)$, a strictly fitter (or lower) solution τ is obtained when $f(\tau)$ is strictly smaller than $f(\sigma)$. We shall determine the smallest number of strictly lower solutions that can be obtained overall σ in the fitness landscape (S_n, f, R).

- If $\sigma = x_1 x_2 \cdots x_n$ is a global minimum, then any reversal will increase the distance travelled. Hence, $\mathrm{Imp}_\sigma(1) = 0$.

- If $\sigma = x_1 x_2 \cdots x_n$ is not a global minimum, then it must contain at least one pair of distinct positions (i_1, i_2) such that reversing the order of the x_i between them yields a shorter distance travelled. Hence, $\mathrm{Imp}_\sigma(1) \geq 1$.

Thus, $\mathrm{Imp}_\sigma(1) \geq 1$ unless σ is a global minimum. $\square$

Theorem 34. *The fitness landscape (S_n, f, R) is polynomially 1-decreasing.*

Proof. We estimate the ratio $\frac{|S_\sigma(1)|}{\mathrm{Imp}_\sigma(1)}$ for any solution σ that is not a global minimum, by using the results of Lemma 26. We have:

$$\begin{aligned}\frac{\mid S_\sigma(1)\mid}{\mathrm{Imp}_\sigma(1)} &= \frac{\binom{n}{2}}{\mathrm{Imp}_\sigma(1)},\\ &\leq n^2.\end{aligned}$$

Hence, $\frac{|S_\sigma(1)|}{\mathrm{Imp}_\sigma(1)}$ is at most polynomial in n for any solution σ that is not a global minimum. The result follows. □

8.2.3 Representation-Independent Solutions

We shall consider a fitness landscape on a metric space $(\mathcal{S}, D)$, where $\mathcal{S}$ is the search space and D is a metric on $\mathcal{S}$. The solutions considered in this section encompass all possible representations: strings on a finite alphabet, permutations, etc ...

Cone Fitness Landscape

Cone fitness landscapes [Mor08] are simple examples of representation-independent classes of landscapes, that were introduced to capture the notion of simple unimodal landscape across representations.

Let x_* denoted a fixed element of $\mathcal{S}$. The fitness function of a cone fitness landscape on the metric space $(\mathcal{S}, D)$ is:

$$\begin{aligned} f : \mathcal{S} &\longrightarrow && \mathbb{R}_+,\\ x &\longmapsto && D(x_*, x).\end{aligned}$$

Lemma 27. *Let x be an element of the fitness landscape $(\mathcal{S}, f, D)$. The number of solutions y that are strictly fitter than x and satisfying $D(x,y) = l$, is denoted $\mathrm{Imp}_x(l)$. We have:*

- *If $x = x_*$ or $l = 0$, then x is a global minimum and $\mathrm{Imp}_x(l) = 0$.*
- *If $D(x, x_*) \geq \min\{D(x,y) \mid D(x,y) > 0\}$, then $\mathrm{Imp}_x(l) \geq 1$ for any $l \geq \min\{D(x,y) \mid D(x,y) > 0\}$.*

Proof. We shall determine the minimal number of strictly fitter solutions, with respect to an element x of the fitness lanscape $(\mathcal{S}, f, D)$, on each of the spheres $S_x(l)$ for $0 < l \leq \max_{x,y\in\mathcal{S}} D(x,y)$. We will consider a minimizing problem.

Let us fix a radius l. In the metric space $(\mathcal{S}, D)$, the sphere $S_x(l)$ has $\mid S_x(l) \mid$ distinct elements. Among these elements of $S_x(l)$, a strictly fitter (or lower) solution y is obtained when $f(y)$ is strictly smaller than $f(x)$. We shall determine the smallest number of strictly lower solutions that can be obtained overall x in the fitness landscape $(\mathcal{S}, f, D)$. We have:

$$f(x) = D(x, x_*). \tag{8.18}$$

We aim to decrease the value of $f(x)$ by replacing x with a solution y such that $D(x, y) = l$. This is satisfied when $y \in [x_*, x]_D \cap S_x(l)$. The number of strictly lower solutions y with respect to x such that $D(x, y) = l$ is denoted $\mathrm{Imp}_x(l)$. We obtain the following results:

- If $x = x_*$ or $l = 0$, then $\mathrm{Imp}_x(l) = 0$.

- Otherwise, the set $[x_*, x]_D \cap S_x(l)$ strictly contains $\{x\}$.
  In this case, $\mathrm{Imp}_x(l) =\mid [x_*, x]_D \cap S_x(l) \mid$ and is bounded below by one. The strict inclusion is guaranteed when:

$$D(x, x_*) \geq \min\{D(x, y) \mid D(x, y) > 0\}. \tag{8.19}$$

□

Theorem 35. *The fitness landscape $(\mathcal{S}, f, D)$ is ρ-decreasing for any $\rho \geq \min\{D(x, y) \mid D(x, y) > 0\}$.*

Proof. Follows from Lemma 27. □

8.3 Generalized $(1 + 1)$ EA

In a local search algorithm, individual offspring are always sampled within a ball centered at the parent. This is called a ball mutation [MP04].

Definition 29 (Ball mutation [MP04])**.** *Let $R > 0$, a ball mutation of radius R samples an offspring from a ball of radius R centred at the parent.*

A ball mutation is uniform [MP04] when each element of the ball has the same probability to be sampled. Otherwise, the ball mutation is not uniform.

In the literature, $(1+1)$ EAs are the simplest example of EAs with no crossover. We will show that the mutation of a $(1+1)$ EA can be seen as a ball mutation with non-uniform probability distribution. Let us first recall the pseudo code of a $(1+1)$ EA.

Algorithm 3 $(1+1)$ EA

Sample an individual x uniformly at random in the search space
parent $\leftarrow x$
while x is not a global optimum **do**
    $y \leftarrow \mathrm{Mutate}(x)$
    if y is at least as fit as x **then**
        parent $\leftarrow y$
    end if
end while

In order to determine the probability distribution of the ball mutation of a $(1+1)$ EA, we shall look at the mutation of $(1+1)$ EA for two different representations: binary strings and permutations.

8.3.1 Binary Strings

The traditional bit-wise mutation operator over binary strings can be seen as a non-uniform ball mutation, where the radius follows a binomial distribution.

Let us consider binary strings of the metric space $M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$ and let $0 < p < 1$ be the mutation probability. Given a binary string x, $\mathrm{Mutate}(x)$ flips each bit of x with probability p. Let Y be the random variable corresponding to the number of bits of x that are flipped. We have:

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{8.20}$$

Let X be the random variable for the distance between x and its offspring $\mathrm{Mutate}(x)$. As the Hamming distance between x and $\mathrm{Mutate}(x)$ is the number of bits of x that has been flipped to obtain $\mathrm{Mutate}(x)$, we have $X = Y$ and :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{8.21}$$

Hence, a $(1+1)$ EA with a traditional bit-wise mutation performs a non-uniform ball mutation in the metric space $M_{2,\mathrm{HD}} = (\{0,1\}^n, \mathrm{HD})$. Indeed, an offspring is always sampled from the ball of radius n centred at its parent. The probability for sampling an offspring depends on its distance $0 \leq X \leq n$ to its parent.

8.3.2 Permutations

The 2-opt-mutation [SN12] over permutations can be seen as a non-uniform ball mutation, where the radius minus one follows a Poisson distribution.

Let us consider permutations of the metric space (S_n, R) where R denotes the reversal distance. Given a permutation σ, a number s is drawn from a Poisson distribution of parameter $L = 1$. Then, $\mathrm{Mutate}(\sigma)$ performs $s + 1$ reversals consecutively on the permutation σ. Let Y be the random variable corresponding to the number of reversals applied to σ, we have:

$$P(Y = s + 1) = P(Y - 1 = s), \tag{8.22}$$
$$= \frac{e^{-1}}{s!}. \tag{8.23}$$

Let X be the random variable for the distance between σ and its offspring $\mathrm{Mutate}(\sigma)$. If the number $k \geq 1$ of reversals applied to σ is equal to the reversal distance between σ and $\mathrm{Mutate}(\sigma)$ then $X = k$ and we have:

$$P(X = k) = P(Y = k), \tag{8.24}$$
$$= \frac{e^{-1}}{(k-1)!}. \tag{8.25}$$

Indeed, the number of reversals applied to a permutation σ in order to obtain a permutation τ need not be equal to the reversal distance between σ and τ in general. This is because the reversal distance between σ and τ is the minimal number of reversals needed to transform σ into τ.

Example 52. *Let us consider the permutation* 123 *of* (S_3, R)*. The* 3 *following reversals can be performed consecutively on* 123*:* $123 \to 213 \to 312 \to 321$*. However, a single reversal suffices to obtain* 321 *from* 123*. Hence, the reversal distance between* 123 *and* 321 *is one. One is different from three, the number of reversals applied to* 123 *in order to obtain* 321*.*

Hence, a $(1{+}1)$ EA with a 2-opt-mutation performs a non-uniform ball mutation in the metric space (S_n, R). Indeed, an offspring $\mathrm{Mutate}(\sigma)$ is always sampled from the ball of radius $\max_{\sigma,\tau \in S_n} R(\sigma, \tau)$ centred at its parent σ. The probability for sampling an offspring depends on its distance to its parent.

8.3.3 Generalization of the $(1 + 1)$ EA

We generalize the $(1 + 1)$ EA of Algorithm 3 across representations. The generalization is done through the definition of a non-uniform ball mutation, that is a probability distribution on the radius of the ball mutation. In the representations considered above:

- the radius of the ball mutation follows a binomial distribution for binary strings,
- the radius minus one of the ball mutation follows a Poisson distribution for permutations.

Let $\mathcal{S}$ denote the search space, D is a metric on $\mathcal{S}$, and X is the random variable corresponding to the distance between a parent and its offspring in the metric space $(\mathcal{S}, D)$. In other words, X is the random variable corresponding to the radius of the ball mutation of the generalized $(1+1)$ EA. As the Poisson distribution is a limit of the binomial distribution, we set the probability distribution of X to a Poisson law of parameter $L > 0$:

$$P(X = k) = \frac{L^k e^{-L}}{k!}, \tag{8.26}$$

for $0 \leq k \leq \max_{x,y \in \mathcal{S}} D(x, y)$. This makes the generalized $(1+1)$ EA a non-uniform ball mutation, whose radius follows a Poisson distribution of parameter $L > 0$.

Algorithm 4 Generalized $(1 + 1)$ EA

Input a parameter $L > 0$
Sample an individual x uniformly at random in the search space
parent $\leftarrow x$
while x is not a global optimum **do**
  Draw a radius k from a Poisson distribution of parameter L
  if $k \leq \max_{x,y \in \mathcal{S}} D(x, y)$ **then**
    Sample an offspring y on the sphere $S_x(k)$ uniformly at random
    if y is at least as fit as x **then**
      parent $\leftarrow y$
    end if
  end if
end while

The Poisson distribution is defined for any radius $k \geq 0$, while the distance between a parent and its offspring lies between 0 and $\max_{x,y \in \mathcal{S}} D(x, y)$ in the metric space $(\mathcal{S}, D)$. Therefore, the generalized $(1 + 1)$ EA may sample a radius $k > \max_{x,y \in \mathcal{S}} D(x, y)$ corresponding to an empty sphere. In this case, we keep making new samplings until a radius $k \leq \max_{x,y \in \mathcal{S}} D(x, y)$ is obtained.

Lemma 28. *Let R denote a random variable following the Poisson law of parameter $L > 0$ and let ρ be a positive integer such that $\rho > L$. The probability to sample a value that is strictly smaller ρ is at least 0.5 if:*

$$L \leq (-1 + \sqrt{2}) \cdot \rho. \tag{8.27}$$

Proof. Let R denote a random variable following the Poisson law of parameter

$L > 0$. We have:

$$P(0 \leq R < \rho) = P\left(0 \leq R < L + \frac{\rho - L}{\sqrt{L}} \cdot \sqrt{L}\right), \tag{8.28}$$

$$= P\left(\mid R - L \mid < \frac{\rho - L}{\sqrt{L}} \cdot \sqrt{L}\right), \tag{8.29}$$

$$= 1 - P\left(\mid R - L \mid \geq \frac{\rho - L}{\sqrt{L}} \cdot \sqrt{L}\right), \tag{8.30}$$

$$\geq 1 - \frac{L}{(\rho - L)^2}, \tag{8.31}$$

for any $\rho > L$. Indeed, the equality in (8.29) follows as a random variable following a Poisson law can only take values that are greater than or equal to zero. And the inequality on the last line follows from Bienaymé-Chebychev's inequality. The result follows from solving in L the inequality:

$$1 - \frac{L}{(\rho - L)^2} \geq \frac{1}{2}. \tag{8.32}$$

□

Corollary 28. *The expected waiting time for sampling a radius strictly less than ρ, is at most two for a generalized $(1+1)$ EA of parameter*

$$0 < L \leq (-1 + \sqrt{2}) \cdot \rho. \tag{8.33}$$

Proof. Follows from Lemma 28. □

8.4 Runtime of the Generalized $(1+1)$ EA

Let $\rho > 0$, we compute an upper bound of the expected runtime of the generalized $(1+1)$ EA on a polynomially ρ-improving fitness landscape using the fitness levels method [Weg01]. A global optimum is found if a strictly fitter solution is sampled from a sphere of radius less than or equal to ρ at each step. We set the parameter L of the generalized $(1+1)$ EA as follows:

$$0 < L \leq (-1 + \sqrt{2}) \cdot (\rho + 1). \tag{8.34}$$

By Corollary 28, this ensures that the expected waiting time for sampling a radius less than or equal to ρ is at most two.

Waiting time for a single improvement

Let x be an element of the polynomially ρ-improving fitness landscape, that can still improve (i.e., x is not a global optimum). The expected waiting time for sampling a strictly fitter solution on the sphere $S_x(l)$ is at most:

$$\max_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)}. \tag{8.35}$$

Therefore, the expected waiting time for sampling a strictly fitter solution on a sphere $S_x(l)$ where x is not a global optimum and $0 < l \leq \rho$ is at most:

$$2 \max_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)}. \tag{8.36}$$

Runtime Upper Bound

We obtain the following result through the fitness levels method [Weg01].

Theorem 36. *A generalized $(1+1)$ EA of parameter $0 < L \leq (-1+\sqrt{2}) \cdot (\rho+1)$ finds a global optimum of a polynomially ρ-improving fitness landscape with $q+1$ distinct level sets within at most:*

$$2q \cdot \max_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mid S_x(l) \mid}{\mathrm{Imp}_x(l)}, \tag{8.37}$$

fitness evaluations.

Proof. An upper bound on the expected runtime is obtained by multiplying the expected waiting time by the largest number of distinct level sets that can be visited. □

Corollary 29. *A generalized $(1+1)$ EA of parameter $0 < L \leq (-1+\sqrt{2})(\rho+1)$ finds a global optimum of a polynomially ρ-improving fitness landscape with at most polynomially many distinct level sets in polynomial time.*

Proof. Follows from Theorem 36. □

We apply the runtime result to various polynomially ρ-improving fitness landscapes from the literature. Then, we compare the runtime results whenever possible.

8.4.1 Strings on a finite alphabet

Let $d \geq 2$, we shall consider fitness landscapes on the metric space $M_{d,\mathrm{HD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$.

Leading Ones

Leading Ones returns the largest number of successive ones starting from the first position of a binary string.

Corollary 30. *The generalized* $(1+1)$ *EA of parameter* $0 < L \leq 2(-1+\sqrt{2})$ *finds the global maximum of leading ones within* $2n^2$ *fitness evaluations.*

Proof. The fitness landscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$ is polynomially 1-increasing. For any solution x that is not a global maximum, $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ is bounded above by $\binom{n}{1}$. The result follows from Theorem 36 as LeadingOnes has $q+1 = n+1$ distinct level sets. $\square$

The upper bound for the expected runtime of the generalized $(1+1)$ EA is tight for Leading Ones. Indeed, the upper bound for the expected runtime of the $(1+1)$ EA for Leading Ones in [Weg03] is $e \cdot n^2$.

Linear functions

Let $w_0,\ w_1,\ \ldots,\ w_n$ be non-zero real numbers. We consider the following linear function [DJS11]:

$$\begin{aligned} f : \{0,1,\ldots,d-1\}^n &\longrightarrow \mathbb{R}, \\ x &\longmapsto w_0 + \sum_{i=1}^{n} w_i x_i. \end{aligned}$$

Corollary 31. *The generalized* $(1+1)$ *EA of parameter* $0 < L \leq 2(-1+\sqrt{2})$ *finds a global maximum of a linear function with* $q+1$ *distinct level sets within at most:*

$$2(d-1)^2 n^2, \tag{8.38}$$

fitness evaluations.

Proof. The fitness landscape $(\{0,1,\ldots,d-1\}^n, f, \mathrm{HD})$ is polynomially 1-increasing. For any solution x that is not a global maximum, $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ is bounded above by $(d-1)\binom{n}{1}$.

In order to estimate the number q of level sets to be visited, we shall extend the technique used in [Weg03] for pseudo-Boolean linear functions to linear functions on a finite alphabet $\{0,1,\cdots,d-1\}$ where $d \geq 2$.

Recall that:

$$f(x) = w_0 + \sum_{w_i>0} \mid w_i \mid x_i - \sum_{w_i<0} \mid w_i \mid x_i. \tag{8.39}$$

By replacing all the x_i such that $w_i < 0$ with $\bar{x_i} = x_i - (d-1)$, we obtain a new linear function g such that:

$$g(x) = w_0 + \sum_{w_i > 0} \mid w_i \mid x_i + \sum_{w_i < 0} \mid w_i \mid [(d-1) - x_i]. \tag{8.40}$$

The generalized $(1+1)$ EA has the same behaviour on both f and g. Thus, the number of level sets of f is equal to the number of level sets of g. Hence, we shall estimate the number of level sets of g.

For $1 \leq i \leq n$, each term of $g(x)$ that is not w_0 takes one of the values: $0, \mid w_i \mid, 2 \mid w_i \mid, \ldots, (d-1) \mid w_i \mid$. By ordering and numbering these values in a non-increasing order, we obtain a sequence of at most $(d-1)n$ positive numbers when removing 0 from the list. Thus, the number q of fitness levels to be visited from any level set is at most $(d-1)n$.

The result follows from Theorem 36. □

The upper bound for the expected runtime of the generalized $(1+1)$ EA is tight for pseudo-Boolean linear functions. Indeed, [Weg03] showed that the number q of of level sets to be visited in this case is at most n. We recall that the upper bound for the expected runtime of the $(1+1)$ EA for pseudo-Boolean linear functions in [Weg03] is $e \cdot n^2$.

The upper bound for the expected runtime of the generalized $(1+1)$ EA is loose for linear functions on the finite alphabet $\{0, 1, 2\}$. The upper bound for the expected runtime of the $(1+1)$ EA for linear functions on the finite alphabet $\{0, 1, 2\}$ in [DJS11] is $O(n \log n)$. It is worth noticing however that [DJS11] did not use the fitness level method to derive this upper bound. They used the multiplicative drift theorem.

Case of OneMax

OneMax is a linear function where $d = 2$, $w_0 = 0$, and $w_i = 1$ for all $1 \leq i \leq n$.

Corollary 32. *The generalized* $(1+1)$ *EA of parameter* $0 < L \leq 2(-1+\sqrt{2})$ *finds the global maximum of OneMax within at most* $2n$ *fitness evaluations.*

Proof. OneMax is a particular linear function with at most n distinct level sets to visit from any starting level set. Moreover, the corresponding fitness landscape is polynomially 1-increasing. We have:

$$\begin{aligned} \frac{\mid S_x(1) \mid}{\mathrm{Imp}_x(1)} &= \frac{\binom{n}{1}}{\#\{w_i > 0\}}, \\ &= \frac{\binom{n}{1}}{n}, \\ &= 1. \end{aligned}$$

The result follows. $\square$

The upper bound for the expected runtime of the generalized $(1+1)$ EA is tight for OneMax. Indeed, the upper bound for the expected runtime of the $(1+1)$ EA for OneMax in [Weg03] is $e \cdot n \cdot (\ln n + 1)$.

Pseudo-Boolean Monotone Increasing function

Let the w_A's be positive real numbers. We consider the following pseudo Boolean monotone increasing polynomial [WW05]:

$$\begin{aligned} f : \{0,1\}^n \longrightarrow & \quad \mathbb{R}, \\ x \longmapsto & \sum_{A \subseteq \{1,2,\ldots,n\}} w_A \prod_{i \in A} x_i. \end{aligned}$$

Corollary 33. *The generalized* $(1+1)$ *EA of parameter* $0 < L \leq 2(-1+\sqrt{2})$ *finds a global minimum of a pseudo-Boolean monotone increasing function with* N *non-vanishing weights within at most:*

$$2N \cdot n, \tag{8.41}$$

fitness evaluations.

Proof. The fitness landscape $(\{0,1\}^n, f, \mathrm{HD})$ is polynomially 1-decreasing. We have:

$$\begin{aligned} \frac{\mid S_x(1) \mid}{\mathrm{Imp}_x(1)} &= \frac{\binom{n}{1}}{\binom{\min_{w_A > 0} |A|}{1}}, \\ &= \frac{n}{\min_{w_A > 0} \mid A \mid}, \\ &\leq n. \end{aligned}$$

By ordering and numbering the non-vanishing weights in a non-decreasing order, [Weg03] showed that the number q of fitness levels to be visited is at most the number of non-vanishing weights N. The result follows from Theorem 36. $\square$

The upper bound for the expected runtime of the generalized $(1+1)$ EA is tight for pseudo-Boolean monotone increasing functions. Indeed, the upper bound for the expected runtime of the $(1+1)$ EA for a pseudo-Boolean monotone increasing function with N non-vanishing weights in [Weg03] is $e.n^{\deg(f)}.N$.

In [WW05], the expected runtime of the $(1+1)$ EA for a pseudo-Boolean monotone increasing function f with N non-vanishing weights is at most:

$$O\left(N \cdot (n/\deg(f)) \cdot 2^{\deg(f)}\right), \tag{8.42}$$

if $\deg(f) \leq 2\log n - 2\log\log n - \alpha$ for some constant α.

8.4.2 Permutations

We shall consider a fitness landscape on the metric space (S_n, R), where R is the reversal distance (see Definition 13).

Euclidean TSP (cities in convex position)

We consider a set of n cities, that are in convex position in the Euclidean space. In other words, each city is a vertex of the convex hull formed by the n cities in the Euclidean space $(\mathbb{R}^2, d)$. This class of Euclidean TSP is solvable in polynomial time.

The Euclidean TSP consists of minimizing the fitness function:

$$\begin{aligned} f : S_n &\longrightarrow \mathbb{R}, \\ \sigma &\longmapsto d(x_{\sigma(n)}, x_{\sigma(1)}) + \sum_{i=1}^{n-1} d(x_{\sigma(i)}, x_{\sigma(i+1)}). \end{aligned}$$

We shall use the same notations as [SN12], where d_{max} and d_{min} are respectively the maximal and minimal Euclidean distances between any two cities. For any three cities $u,\ v,$ and w, θ denotes the angle formed by the line from u to v and the line from v to w. The angle ϵ satisfies $0 < \epsilon < \theta < \pi - \epsilon$. Moreover,

$$\gamma(\epsilon) = \left(\frac{d_{max} - d_{min}}{d_{min}}\right)\left(\frac{\cos\epsilon}{1 - \cos\epsilon}\right). \tag{8.43}$$

Corollary 34. *The generalized* $(1+1)$ *EA of parameter* $0 < L \leq 2(-1+\sqrt{2})$ *finds a global minimum of an Euclidean TSP, where the cities are in convex position, within at most:*

$$n^3 \cdot \gamma(\epsilon), \tag{8.44}$$

fitness evaluations.

Proof. The fitness landscape (S_n, f, R) is polynomially 1-decreasing. For any solution σ that is not a global minimum, $\frac{|S_\sigma(1)|}{\mathrm{Imp}_\sigma(1)}$ is bounded above by $\binom{n}{2}$. The result follows from Theorem 36 as an upper bound on the number level sets to be visited is given in [SN12] as:

$$\begin{aligned} q &= n\left(\frac{d_{max} - d_{min}}{2d_{min}}\right)\left(\frac{\cos\epsilon}{1 - \cos\epsilon}\right), \\ &= \frac{n}{2} \cdot \gamma(\epsilon). \end{aligned}$$

□

The upper bound for the expected runtime of the generalized $(1+1)$ EA is tight for the Euclidean TSP with cities in convex position. Indeed, the upper bound for the expected runtime of the $(1+1)$ EA for the same problem in [SN12] is $O(n^3 \cdot \gamma(\epsilon) + n \cdot \gamma(\epsilon))$.

8.5 Summary

We generalized the $(1+1)$ EA across representations as a non-uniform ball mutation whose radius follows a Poisson law of parameter $L > 0$.

We defined a class of fitness landscapes parametrized by $\rho > 0$, called polynomially ρ-improving. This landscape captures the number of solutions per sphere of radius $0 < l \leq \rho$, that are strictly fitter than the centre of the sphere. For a polynomially ρ-improving fitness landscape, this number is at least a polynomial fraction of the size of the sphere.

We found that the generalized $(1+1)$ EA of parameter $0 < L \leq (-1+\sqrt{2})\cdot(\rho+1)$ finds a global optimum of any polynomially ρ-improving fitness landscape with at most polynomially many distinct level sets in polynomial time.

We specified the runtime results to polynomially ρ-improving fitness landscapes from the literature. The upper bound for the expected runtime of the generalized $(1+1)$ EA is always tight when compared to runtime upper bounds derived through the fitness level method. However, the upper bound for the expected runtime of the generalized $(1+1)$ EA is loose when compared to a runtime upper bound derived through drift analysis.

As we used the fitness level method throughout the analysis, we conclude that the generalized $(1+1)$ EA captures well the behaviour of a $(1+1)$ EA regardless of the representation used.

It would be interesting to re-do the analysis with a drift analysis instead of the fitness level methods. Indeed, this may yield a tighter upper bound.

Chapter 9

A Representation free EA with both a Mutation and a Crossover

We aim to introduce a unifying runtime analysis of EA with both a mutation and a standard crossover across representations, by:

- defining a representation free EA with both a mutation and a standard crossover operators,
- finding a class of fitness landscapes that are solved in polynomial time by instantiations of the representation free EA with both a mutation and a standard crossover operators.

To this end, we use the mutation operator of the generalized $(1+1)$ EA of Chapter 8 and the crossover operator of the SES (see Chapter 5) to define the representation free EA with both a mutation and a standard crossover in Section 9.1. The notion of ρ-improving fitness landscapes that has been introduced in Chapter 8, is recalled in Section 9.2. Then, the runtime analysis of the representation free EA on ρ-improving fitness landscapes is done in Section 9.3. We find that the representation free EA solves a ρ-improving fitness landscape with $q + 1$ distinct level sets in $2q$ fitness evaluations for a well chosen population size. The runtime result is then specified to strings on a finite alphabet in Section 9.3.1 and to permutations in Section 9.3.2, by considering polynomially ρ-improving fitness landscapes from the literature. We find that Leading Ones, linear functions (including OneMax), pseudo-Boolean functions, and Euclidean TSP with cities in convex positions are solved in polynomial time by the representation free EA for any population size larger than or equal to one.

9.1 A Representation free EA

We defined a generalized mutation operator in Section 8.3 of Chapter 8. The mutation offspring is sampled uniformly at random on a sphere centred at the parent.

The radius of the sphere is sampled from a Poisson law. A generalized standard two-parents crossover called geometric crossover has been defined by [MP04] (see Section 5.2 of Chapter 5). The crossover offspring is sampled uniformly at random from the metric segment formed by the two parents. Each parent is sampled uniformly at random from the selected population. In this section, we define a representation free EA from the generalized mutation operator of Chapter 8 and the geometric crossover of [MP04]. A pseudocode corresponding to the representation free EA is given in Algorithm 5.

Let $\mathcal{S}$ denote the search space and let D denote a metric on $\mathcal{S}$. We recall that $S_x(l)$ denotes the sphere centred at $x \in \mathcal{S}$ and of radius $l > 0$:

$$S_x(l) = \{z \in \mathcal{S} \mid D(x, z) = l\}. \tag{9.1}$$

Algorithm 5 Representation free EA

Input: population size μ, generalized mutation parameter L
Output: individual in the last population
Initialise population uniformly at random
while optimum is not in the population **do**
  Rank individuals on fitness
  if there are at least two distinct fitness values in the current population **then**
    remove all individuals with the worst fitness
  end if
  for counter in $\{1, 2, \ldots, \mu\}$ **do**
    Randomly and uniformly pick two individuals from the remaining individuals
    Recombine them through GEOMETRIC CROSSOVER to create a new individual x
    Draw a radius l from a Poisson distribution of parameter L
    Sample an offspring y uniformly at random on $S_x(l)$
  end for
end while
Return the best individual in the population

An offspring is sampled uniformly at random on a sphere, whose center is an element of a segment of the selected population P'. The center of the sphere is sampled uniformly at random on the segment. The extremes of the segment are also sampled uniformly at random from P'. However, the sphere is not sampled uniformly at random from the set of all spheres with the same centre. Indeed, the radii of the spheres sharing the same center follow a Poisson law of parameter L. By abuse of language, spheres sharing the same centre follow a Poisson law of parameter L by identifying a sphere with its radius.

The set of reachable solutions from a set P' of selected individuals, is therefore the union of all the spheres of any possible radii whose centres belong to a segment of P'. Let $\delta(\mathcal{S})$ denote $0 \leq l \leq \max_{x,y \in \mathcal{S}} D(x, y)$, we have:

$$R(P') = \bigcup_{\substack{x \in [p_1, p_2] \\ p_1, p_2 \in P'}} \left[\bigcup_{0 \leq l \leq \delta(\mathcal{S})} S_x(l) \right], \tag{9.2}$$

$$= \bigcup_{x \in Seg(P')} \left[\bigcup_{0 \leq l \leq \delta(\mathcal{S})} S_x(l) \right], \tag{9.3}$$

$$= \bigcup_{x \in Seg(P')} B_x(\delta(\mathcal{S})). \tag{9.4}$$

We saw in Section 5.2.1 of Chapter 5 that the probability to sample a crossover offspring x in $R_{\text{cross.}}(P') = Seg(P')$ is:

$$Pr(x \in R_{\text{cross.}}(P')) = \sum_{i=1}^{p} \frac{\alpha_{s_i, P'}}{\mid P' \mid^2} \cdot \frac{\mathbb{1}_{s_i}(x)}{\mid s_i \mid}, \tag{9.5}$$

where $\alpha_{s_i, P'}$ is the number of pairs of elements of P' yielding the segment s_i and $Seg(P') = s_1 \cup s_2 \cup \cdots \cup s_p$. Each crossover offspring x then becomes a mutation parent.

Let X denote the random variable corresponding to the radius of the ball mutation of the generalized $(1+1)$ EA of Chapter 8. The random variable X follows a Poisson law of parameter L:

$$P(X = l) = \frac{L^l e^{-L}}{l!}. \tag{9.6}$$

The probability to sample a mutant offspring z in $R_{\text{mut.}}(P') = \bigcup_{x \in P'} B_x(\delta(\mathcal{S}))$ is:

$$Pr(z \in R_{\text{mut.}}(P')) = \sum_{x \in P'} \sum_{0 \leq l \leq \delta(\mathcal{S})} P(X = l) \frac{\mathbb{1}_{S_x(l)}(z)}{\mid S_x(l) \mid}, \tag{9.7}$$

$$= \sum_{x \in P'} \frac{P(X = d(x, z))}{\mid S_x(d(x, z)) \mid}. \tag{9.8}$$

Consequently, the probability to sample a mutant offspring z whose parent is a crossover offspring x is given by:

$$Pr(z \in R(P')) = \sum_{x \in Seg(P')} \frac{P(X = d(x, z))}{\mid S_x(d(x, z)) \mid}, \tag{9.9}$$

$$= \sum_{x \in R_{\text{cross.}}(P')} \frac{P(X = d(x, z))}{\mid S_x(d(x, z)) \mid}, \tag{9.10}$$

$$= Pr(z \in R_{\text{mut.}}(R_{\text{cross.}}(P'))). \tag{9.11}$$

The offspring distribution on $R(P')$ is not uniform and is reminiscent of the offspring distribution of $R_{\mathrm{mut.}}(P')$ in Equation (9.8) where P' is replaced with $R_{\mathrm{cross.}}(P')$.

9.2 ρ-improving fitness landscapes

We recall from Section 8.2 of Chapter 8 that a ρ-improving fitness landscape $(\mathcal{S}, f, D)$ with $q+1$ distinct fitness values, is a fitness landscape where any solution x that is not a global optimum is l-improving (i.e., there exists a strictly fitter solution y with respect to x on the sphere $S_x(l)$) for any radius $0 < l \leq \rho$. The number of l-improving solutions on the sphere $S_x(l)$ is denoted by:

$$\mathrm{Imp}_x(l). \tag{9.12}$$

Hence, the ratio of l-improving solutions on the sphere $S_x(l)$ is:

$$\frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}. \tag{9.13}$$

When this ratio is at least a polynomial fraction for $0 < l \leq \rho$, then the fitness landscape is said to be polynomially ρ-improving.

Let $A_0 = \mathcal{S},\ A_1,\ \ldots,\ A_q$ denote the $q+1$ distinct level sets of the fitness landscape $(\mathcal{S}, f, D)$. A canonical level set $A_{\geq j}$ [MS17] is the union of the level sets $A_j,\ A_{j+1},\ \ldots,\ A_q$. Moreover, the canonical level sets form a chain :

$$\mathcal{S} = A_{\geq 0} \supsetneq A_{\geq 1} \supsetneq \cdots \supsetneq A_{\geq q-1} \supsetneq A_q,$$

where A_q is the set of global optima. We no longer require the canonical level sets to be convex sets with respect to the metric D as in Chapter 8. The following examples of ρ-improving landscapes have been given in Section 8.2 of Chapter 8.

Example 53. *The fitness landscape* $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$ *is polynomially* 1-*increasing.*

Example 54. *Let* f *be a linear function, the fitness landscape* $(\{0,1,\ldots,d-1\}^n, f, \mathrm{HD})$ *is polynomially* ρ*-increasing for any constant* $\rho \geq 1$.

Example 55. *The fitness landscape* $(\{0,1\}^n, \mathrm{OneMax}, \mathrm{HD})$ *is also polynomially* ρ*-increasing for any constant* $\rho \geq 1$.

Example 56. *Let* f *be a pseudo-Boolean function, the fitness landscape* $(\{0,1\}^n, f, \mathrm{HD})$ *is polynomially* ρ*-decreasing for any constant* $\rho \geq 1$.

Example 57. *Let* $f : S_n \to \mathbb{R}$ *denote the objective function of an Euclidean TSP, where the* n *cities are in convex positions. The set* S_n *is endowed with the reversal distance* R*. The fitness landscape* (S_n, f, R) *is polynomially* 1*-decreasing.*

9.3 Runtime Analysis

We use the fitness levels method [Weg01] to estimate an upper bound on the expected runtime of the representation free EA on a ρ-improving fitness landscape.

We first compute a lower bound on the probability for sampling a strictly improving solution. That is, if the population of selected individuals is contained in the canonical level set $A_{\geq j}$ then we compute the probability for sampling an offspring that belongs to $A_{\geq j+1}$.

Lemma 29. *The probability for sampling a strictly improving offspring is at least:*

$$Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}. \tag{9.14}$$

Proof. We assume that the selected population P' is included in $A_{\geq j}$. Assuming that the offspring are uniformly distributed on $R(P')$ with the least probability of Equation (9.9), the probability to sample a strictly improving offspring is at least:

$$\begin{aligned}
&\sum_{z\in A_{\geq j+1}} \sum_{x\in Seg(P')} \frac{P[X=d(x,z)]}{\mid S_x[d(x,z)] \mid},\\
\geq &\sum_{z\in A_{\geq j+1}} \sum_{x\in Seg(P')} \frac{P[X=d(x,z)]}{\mid S_x[d(x,z)] \mid} \cdot \mathbb{1}_{B_x(\rho)}(z),\\
= &\sum_{x\in Seg(P')} \sum_{\substack{0\leq l\leq\delta(\mathcal{S}) \\ S_x(l)\neq\varnothing}} \left[\sum_{z\in\{A_{\geq j+1} \mid d(x,z)=l\}} \frac{P(X=l)}{\mid S_x(l) \mid} \cdot \mathbb{1}_{B_x(\rho)}(z) \right],\\
= &\sum_{x\in Seg(P')} \sum_{\substack{0\leq l\leq\rho \\ S_x(l)\neq\varnothing}} \left[\frac{P(X=l)}{\mid S_x(l) \mid} \cdot \mid S_x(l) \cap A_{\geq j+1} \mid \right],\\
\geq &\sum_{x\in Seg(P')\cap A_{\geq j+1}} \sum_{\substack{0\leq l\leq\rho \\ S_x(l)\neq\varnothing}} \left[\frac{P(X=l)}{\mid S_x(l) \mid} \cdot \mid S_x(l) \cap A_{\geq j+1} \mid \right],\\
= &\sum_{x\in Seg(P')\cap A_{\geq j+1}} \sum_{\substack{0\leq l\leq\rho \\ S_x(l)\neq\varnothing}} \left[P(X=l) \cdot \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right],\\
= &\sum_{\substack{0\leq l\leq\rho \\ S_x(l)\neq\varnothing}} \left[P(X=l) \sum_{x\in Seg(P')\cap A_{\geq j+1}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right],\\
\geq &\sum_{\substack{0<l\leq\rho \\ S_x(l)\neq\varnothing}} \left[P(X=l) \cdot \min_{x \text{ impr.}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right],
\end{aligned}$$

$$
\begin{aligned}
&\geq \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \sum_{\substack{0\leq l\leq\rho \\ S_x(l)\neq\varnothing}} P(X=l), \\
&\geq Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}.
\end{aligned}
$$

Indeed, in any discrete metric space $(\mathcal{S}, D)$ the set $\{l \mid S_x(l) \neq \varnothing\}$ of all possible radii of a sphere is given by $\{0, 1, 2, \cdots, \delta(\mathcal{S})\}$ where $\delta(\mathcal{S})$ is the diameter of $\mathcal{S}$. □

Lemma 30. *The probability for sampling at least one strictly improving offspring is at least:*

$$
1 - \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}\right]^{\mu}, \tag{9.15}
$$

for a population size μ.

Proof. The probability for never sampling a strictly improving offspring among the μ individuals of the resulting population is at most:

$$
\left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}\right]^{\mu}. \tag{9.16}
$$

The result follows. □

Lemma 31. *The expected waiting time for sampling at least one strictly improving offspring is at most two, if the population size is at least:*

$$
\mu_{\min}(L, \rho) = \frac{1}{2\left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{|S_x(l)|}\right]}. \tag{9.17}
$$

Proof. It is enough to solve in μ the inequality:

$$
1 - \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid}\right]^{\mu} \geq \frac{1}{2}. \tag{9.18}
$$

The result follows as the left hand side is a lower bound on the probability for sampling at least one strictly improving offspring. We have:

$$1 - \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right]^\mu \geq \frac{1}{2},$$

$$\frac{1}{2} \geq \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right]^\mu \geq 1 - \mu \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right],$$

$$\mu \left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{\mid S_x(l) \mid} \right] \geq \frac{1}{2},$$

$$\mu \geq \frac{1}{2\left[1 - Le^{-L} \min_{\substack{x \text{ impr.} \\ 0<l\leq\rho \\ S_x(l)\neq\varnothing}} \frac{\mathrm{Imp}_x(l)}{|S_x(l)|} \right]}.$$

□

An upper bound on the expected runtime is estimated through the fitness levels method [Weg01].

Theorem 37. *A representation free EA of mutation parameter L and with a population size $\mu \geq \mu_{\min}(L, \rho)$, finds a global optimum of a ρ-improving fitness landscape with $q + 1$ distinct level sets within at most $2q$ fitness evaluations.*

Proof. In the worst case, each level set is visited before finding a global optimum. □

Corollary 35. *A representation free EA of mutation parameter L and with a population size $\mu \geq \mu_{\min}(L, \rho)$, finds a global optimum of a ρ-improving fitness landscape with at most polynomially many distinct level sets in polynomial time.*

Proof. Follows from Theorem 37. □

Remark 6. *In Theorem 37 and Corollary 35, the value of the parameter ρ is at least one. This is because the smallest strictly positive radius in a discrete metric space is one.*

We apply the runtime result to various polynomially ρ-improving fitness landscapes from the literature, that are also ρ-improving.

9.3.1 Strings on a finite alphabet

Let $d \geq 2$, we shall consider fitness landscapes on the metric space $M_{d,\mathrm{HD}} = (\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$.

Leading Ones

Leading Ones returns the largest number of successive ones starting from the first position of a binary string.

Corollary 36. *The representation free EA of parameter $L = 1$ finds the global maximum of leading ones within $2n$ fitness evaluations if the population size is at least:*

$$\mu \geq \frac{1}{2\left(1 - \frac{1}{en}\right)}. \tag{9.19}$$

Proof. The fitness landscape $(\{0,1\}^n, \mathrm{LeadingOnes}, \mathrm{HD})$ is polynomially 1-increasing. For any solution x that is not a global maximum, $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ is bounded above by $\binom{n}{1}$. The result follows from Theorem 37 as LeadingOnes has $q + 1 = n + 1$ distinct level sets. □

The upper bound for the expected runtime of the representation free EA is smaller than the upper bound for the expected runtime of the generalized $(1+1)$ EA for Leading Ones. We recall that the expected runtime of the generalized $(1+1)$ EA for Leading Ones is $2n^2$.

Linear functions

Let $w_0,\ w_1,\ \ldots,\ w_n$ be non-zero real numbers. We consider the following linear function [DJS11]:

$$\begin{aligned} f : \{0, 1, \ldots, d-1\}^n \longrightarrow & \qquad \mathbb{R}, \\ x \longmapsto & \qquad w_0 + \sum_{i=1}^{n} w_i x_i. \end{aligned}$$

Corollary 37. *The representation free EA of parameter L finds a global maximum of the linear function f within at most $2(d-1)n$ fitness evaluations if the population size is at least:*

$$\mu \geq \frac{1}{2\left[1 - \frac{L}{e^L(d-1)n}\right]}. \tag{9.20}$$

Proof. The fitness landscape $(\{0, 1, \ldots, d-1\}^n, f, \mathrm{HD})$ is polynomially 1-increasing. For any solution x that is not a global maximum, $\frac{|S_x(1)|}{\mathrm{Imp}_x(1)}$ is bounded above by $(d-1)\binom{n}{1}$.

In order to estimate the number q of level sets to be visited, we shall extend the technique used in [Weg03] for pseudo-Boolean linear functions to linear functions on a finite alphabet $\{0, 1, \cdots, d-1\}$ where $d \geq 2$.

Recall that:

$$f(x) = w_0 + \sum_{w_i > 0} \mid w_i \mid x_i - \sum_{w_i < 0} \mid w_i \mid x_i. \tag{9.21}$$

By replacing all the x_i such that $w_i < 0$ with $\bar{x_i} = x_i - (d-1)$, we obtain a new linear function g such that:

$$g(x) = w_0 + \sum_{w_i > 0} \mid w_i \mid x_i + \sum_{w_i < 0} \mid w_i \mid [(d-1) - x_i]. \tag{9.22}$$

The representation free EA has the same behaviour on both f and g. Thus, the number of level sets of f is equal to the number of level sets of g. Hence, we shall estimate the number of level sets of g.

For $1 \leq i \leq n$, each term of $g(x)$ that is not w_0 takes one of the values: $0, \mid w_i \mid, 2 \mid w_i \mid, \ldots, (d-1) \mid w_i \mid$. By ordering and numbering these values in a non-increasing order, we obtain a sequence of at most $(d-1)n$ positive numbers when removing 0 from the list. Thus, the number q of fitness levels to be visited from any level set is at most $(d-1)n$.

The result follows from Theorem 37. □

The upper bound for the expected runtime of the representation free EA is smaller than the upper bound for the expected runtime of the generalized $(1+1)$ EA for the linear function f. We recall that the expected runtime of the generalized $(1+1)$ EA for the linear function f is $2(d-1)^2n^2$.

Case of OneMax

OneMax is a linear function where $d = 2$, $w_0 = 0$, and $w_i = 1$ for all $1 \leq i \leq n$.

Corollary 38. *The representation free EA of parameter $L = 1$ finds the global maximum of OneMax within at most $2n$ fitness evaluations if the population size is at least:*

$$\mu \geq \frac{1}{2\left(1 - \frac{1}{e}\right)}. \tag{9.23}$$

Proof. OneMax is a particular linear function with at most n distinct level sets to visit from any starting level set. Moreover, the corresponding fitness landscape is polynomially 1-increasing. We have:

$$\begin{aligned} \frac{\mid S_x(1) \mid}{\mathrm{Imp}_x(1)} &= \frac{\binom{n}{1}}{\#\{w_i > 0\}}, \\ &= \frac{\binom{n}{1}}{n}, \\ &= 1. \end{aligned}$$

The result follows. □

The representation free EA and the generalized $(1+1)$ EA have the same expected runtime upper bound for OneMax. We recall that the expected runtime of the generalized $(1+1)$ EA for OneMax is $2n$.

Pseudo-Boolean Monotone Increasing function

Let the w_A's be positive real numbers. We consider the following pseudo Boolean monotone increasing polynomial [WW05]:

$$\begin{aligned} f : \{0,1\}^n \longrightarrow & \quad \mathbb{R}, \\ x \longmapsto & \sum_{A \subseteq \{1,2,\ldots,n\}} w_A \prod_{i \in A} x_i. \end{aligned}$$

Corollary 39. *The representation free EA of parameter $L = 1$ finds a global minimum of a pseudo-Boolean monotone increasing function with N non-vanishing weights within at most $2N$ fitness evaluations if the population size is at least:*

$$\mu \geq \frac{1}{2\left(1 - \frac{1}{en}\right)}. \tag{9.24}$$

Proof. The fitness landscape $(\{0,1\}^n, f, \mathrm{HD})$ is polynomially 1-decreasing. We have:

$$\begin{aligned} \frac{\mid S_x(1) \mid}{\mathrm{Imp}_x(1)} &= \frac{\binom{n}{1}}{\binom{\min_{w_A > 0} |A|}{1}}, \\ &= \frac{n}{\min_{w_A > 0} \mid A \mid}, \\ &\leq n. \end{aligned}$$

By ordering and numbering the non-vanishing weights in a non-decreasing order, [Weg03] showed that the number q of fitness levels to be visited is at most the number of non-vanishing weights N. The result follows from Theorem 37. □

The upper bound for the expected runtime of the representation free EA is smaller than the upper bound for the expected runtime of the generalized $(1+1)$ EA for a pseudo-Boolean function with N non-vanishing weights. We recall that the expected runtime of the generalized $(1+1)$ EA for a pseudo-Boolean function with N non-vanishing weights is $2N \cdot n$.

9.3.2 Permutations

We shall consider a fitness landscape on the metric space (S_n, R), where R is the reversal distance (see Definition 13).

Euclidean TSP (cities in convex position)

We consider a set of n cities, that are in convex position in the Euclidean space. In other words, each city is a vertex of the convex hull formed by the n cities in the

Euclidean space $(\mathbb{R}^2, d)$. This class of Euclidean TSP is solvable in polynomial time.

The Euclidean TSP consists of minimizing the fitness function:

$$\begin{aligned} f : S_n \longrightarrow & \quad \mathbb{R}, \\ \sigma \longmapsto & \quad d(x_{\sigma(n)}, x_{\sigma(1)}) + \sum_{i=1}^{n-1} d(x_{\sigma(i)}, x_{\sigma(i+1)}). \end{aligned}$$

We shall use the same notations as [SN12], where d_{max} and d_{min} are respectively the maximal and minimal Euclidean distances between any two cities. For any three cities $u,\ v,$ and w, θ denotes the angle formed by the line from u to v and the line from v to w. The angle ϵ satisfies $0 < \epsilon < \theta < \pi - \epsilon$. Moreover,

$$\gamma(\epsilon) = \left(\frac{d_{max} - d_{min}}{d_{min}}\right)\left(\frac{\cos\epsilon}{1 - \cos\epsilon}\right). \tag{9.25}$$

Corollary 40. *The representation free EA of parameter $L = 1$ finds a global minimum of an Euclidean TSP, where the cities are in convex position, within at most $n \cdot \gamma(\epsilon)$ fitness evaluations if the population size is at least:*

$$\mu \geq \frac{1}{2\left[1 - \frac{1}{e\binom{n}{2}}\right]}. \tag{9.26}$$

Proof. The fitness landscape (S_n, f, R) is polynomially 1-decreasing. For any solution σ that is not a global minimum, $\frac{|S_\sigma(1)|}{\mathrm{Imp}_\sigma(1)}$ is bounded above by $\binom{n}{2}$. The result follows from Theorem 36 as an upper bound on the number level sets to be visited is given in [SN12] as:

$$\begin{aligned} q &= n\left(\frac{d_{max} - d_{min}}{2d_{min}}\right)\left(\frac{\cos\epsilon}{1 - \cos\epsilon}\right), \\ &= \frac{n}{2} \cdot \gamma(\epsilon). \end{aligned}$$

□

The upper bound for the expected runtime of the representation free EA is smaller than the upper bound for the expected runtime of the generalized $(1 + 1)$ EA for the Euclidean TSP with cities in convex position. We recall that the expected runtime of the generalized $(1 + 1)$ EA for the Euclidean TSP with cities in convex position is $n^3 \cdot \gamma(\epsilon)$.

9.4 Summary

We defined a representation free EA with:

- a truncation selection [SVM93] that removes all individuals with the worst fitness value,
- the crossover operator of the SES,
- the mutation operator of the generalized $(1+1)$ EA that is parametrized by $L > 0$.

We defined a class of fitness landscapes parametrized by $\rho > 0$, called ρ-improving. This landscape captures the number of solutions per sphere of radius $0 < l \leq \rho$, that are strictly fitter than the centre of the sphere. For a ρ-improving fitness landscape of a discrete metric space, this number is at least one.

We found that there exists a lower bound $\mu_{\min}(L, \rho)$ on the population size, such that ρ-improving fitness landscapes with at most polynomially many level sets are solved in polynomial time by the representation free EA with both a mutation and a standard crossover when the population size is at least $\mu_{\min}(L, \rho)$.

We specified the runtime result to polynomially ρ-improving fitness landscapes from the literature, that are also ρ-improving: Leading Ones, linear functions (including OneMax), pseudo-Boolean functions, and Euclidean TSP with cities in convex position. We found that for any population size $\mu \geq 1$, each of these problems is solved in $2q$ expected fitness evaluations, where $q+1$ denotes the number of level sets of the problem. Moreover, the upper bound of the expected runtime of the representation free EA is smaller than or equal to that of the generalized $(1+1)$ EA for each of these problems. As bounds can not be used to compare algorithms, we can not yet conclude that the representation free EA outperforms the generalized $(1+1)$ EA on polynomially ρ-improving fitness landscapes. However, our result shows that the representation free EA solves any ρ-improving fitness landscape with at most polynomially fitness level sets in polynomial time. Whereas, the generalized $(1+1)$ EA only solves polynomially ρ-improving fitness landscape with at most polynomially fitness level sets in polynomial time. In metric spaces where polynomially ρ-improving fitness landscapes are ρ-improving fitness landscapes, the representation free EA solves at least as many problems as the $(1+1)$ EA. It remains to investigate how the representation free EA compares to EAs (with both a mutation and a standard two-parents crossover) from the literature, in particular (λ, μ) EAs.

Chapter 10

Conclusion

We defined:

- a representation free $(1+1)$ EA that generalizes $(1+1)$ EA on strings and $(1+1)$ EA on permutations,
- a representation free EA with no mutation and with a standard two-parents crossover (called SES) that generalizes EA with no mutation and with a standard two-parents crossover on strings and EA with no mutation and with a standard two-parents crossover on permutations,
- a representation free EA with the mutation operator of the generalized $(1+1)$ EA and the crossover operator of the SES.

Easy problems (of a given algorithm) were defined to be solved (by the algorithm) within at most polynomial time in the solution size. Then, a collection (or a class) of easy problems has been determined for each representation free algorithm. We obtained the following results:

- Polynomially ρ-improving problems (where $\rho \geq 1$) with at most polynomially many level sets, have been shown to be easy for $(1+1)$ EA on strings and $(1+1)$ EA on permutations.
- Quasi-concave problems with at most polynomially many level sets have been shown to be easy for EA with no mutation and with a standard crossover on strings. However, quasi-concave problems with at most polynomially many level sets need not be easy for EA with no mutation and with a standard crossover on permutations.
- ρ-improving problems (where $\rho \geq 1$) with at most polynomially many level sets, have been shown to be easy for the instantiations of the representation free EA with both a mutation and a standard crossover to strings and to permutations.

This divergence in the results is due to the difference in which the class of problems have been determined.

The class of quasi-concave problems has been obtained through the observation of specific problems on strings on a finite alphabet. Indeed, the class of quasi-concave problems was defined to contain these specific problems. Hence, the definition of the class of quasi-concave problems implicitly depends on representations where solutions are strings.

However, the class of ρ-improving problems has been defined using balls. A ball can be defined for any representation. Hence, the definition of the class of ρ-improving problems, does not depend on a particular representation.

10.1 Contributions of the Thesis

We defined a class of fitness landscapes that is parametrized by a radius $\rho > 0$ and determined by the balls of radius ρ that are not centred at a global optimum. We are particularly interested in the spheres of these balls (i.e., the spheres contained in a ball and having the same centre as the ball) and in the elements of the spheres which are strictly fitter than their centre.

- A polynomially ρ-improving fitness landscapes is obtained if the number of strictly fitter solutions on each sphere (of a ball of radius ρ that is not centred at a global optimum) is at least a polynomial fraction of the size of the sphere.

- A ρ-improving fitness landscapes is obtained if the number of strictly fitter solutions on each sphere (of a ball of radius ρ that is not centred at a global optimum) is at least one.

We showed that in the metric space $(\{0,1\}^n, \mathrm{HD})$ the fitness landscapes of Leading Ones, OneMax, and a pseudo-Boolean function are respectively polynomially 1-increasing, polynomially ρ-increasing for any $\rho \geq 1$, and polynomially ρ-decreasing for any $\rho \geq 1$.

In the metric space $(\{0, 1, \ldots, d-1\}^n, \mathrm{HD})$, the fitness landscape of a linear function is polynomially ρ-increasing for any $\rho \geq 1$.

In the metric space (S_n, R) where R denotes the reversal distance on permutations, the fitness landscape of an Euclidean TSP is polynomially 1-decreasing.

Each of these fitness landscapes are also ρ-improving for $\rho \geq 1$.

10.1.1 EA with no mutation

We generalized EA with no mutation and with a standard crossover across representations, as to generate offspring on the metric segment of its two parents.

Moreover, the probability distribution of the possible offspring on the metric segment is set to be uniform. This new generalized EA with no mutation is referred to as a Standard Evolutionary Search (SES).

We showed that the unifying runtime analysis of the CS can only be extended to the SES for problems on metric spaces where convex sets are union of segments. As a result, polynomial quasi-concave problems on the metric spaces $M_{d,\mathrm{HD}}$ and $M_{d,\mathrm{MD}}$ are efficiently solved by the SES. Moreover, the unifying runtime analysis of the CS could not be extended to the SES for problems on the usual metric spaces of permutations. This is because convex sets need not be union of segments in these metric spaces.

By specifying the unifying runtime analysis of the CS to permutations, we found that polynomial quasi-concave problems on the usual metric spaces of permutations need not be efficiently solved by the CS.

10.1.2 $(1+1)$ EA

We generalized $(1+1)$ EA across representations by considering a mutation that describes a metric ball centred at the parent: the centre of the ball is the parent and the radius of the ball is the largest possible. The probability distribution of the possible offspring on the ball is a function of their metric distance to the centre of the ball. We approximated this probability distribution with a Poisson law of parameter L for the generalized $(1+1)$ EA, after determining the probability distributions obtained for a $(1+1)$ EA on binary strings and a $(1+1)$ EA on permutations.

We showed that any polynomially ρ-improving fitness landscape with at most polynomially many level sets is efficiently solved by an instantiation of the generalized $(1+1)$ EA for a well chosen mutation parameter. Moreover, the runtime upper bounds obtained are tight when compared to the runtime upper bounds obtained through the fitness levels method in the literature.

10.1.3 EA with both a mutation and a standard crossover

We defined a representation free EA with both a mutation and a crossover as to have:

- the same mutation operator as the generalized $(1+1)$ EA,
- the same crossover operator as the SES.

We showed that any ρ-improving fitness landscape with at most polynomially many level sets is efficiently solved by an instantiation of the generalized EA for a well chosen population size.

10.2 Limitations

- The analysis of the SES presented in this work only holds on metric spaces where the union of the segments that can be formed from any subset is always a convex set. Hence, the current analysis of SES could not cover the case of problems on the usual metric spaces of permutations.

- The runtime upper bounds obtained for the instantiations of the generalized $(1+1)$ EA were tight with respect to the runtime upper bounds obtained for the $(1+1)$ EA from the literature, when restricting the runtime results to those obtained through the fitness levels method. However, the runtime upper bound of an instantiation of the generalized $(1+1)$ EA was looser than the runtime upper bound of its literature counterpart that has been deduced from a multiplicative drift method.

- We defined a representation free EA with both a mutation and a crossover. Then, we determined a class of problems whose elements are easy for an instantiation of the algorithm. However, we can not yet tell how the instantiations of the representation free algorithm compare with (λ, μ) EA from the literature.

10.3 Recommendations for Future Work

- Determine a class of easy problems for the SES using a similar approach to that used for $(1+1)$ EA and EA with both a mutation and a standard crossover.

- Determine how the chosen runtime analysis method affects the runtime result, by using drift analysis instead of the fitness levels method for the unifying runtime analysis.

- Compare the instantiations of the representation free EA with a mutation and a crossover to (λ, μ) EAs from the literature by:
  - finding papers on the runtime analysis of (λ, μ) EAs using the fitness levels method,
  - determining whether the problems that are efficiently solved by the (λ, μ) EAs are ρ-improving for some $\rho \geq 1$ and with at most polynomially many level sets,
  - comparing the runtime upper bounds of the papers to the runtime upper bounds of the instantiations of the representation free EA with a mutation and a crossover.

Bibliography

[AA18] Nada Mohammed Abbas and Ruma Kareem K Ajeena. Mathematical analysis on the cosets of subgroup in the group of e-convex sets. In *Journal of Physics: Conference Series*, volume 1003, page 012045. IOP Publishing, 2018.

[AD99] David Aldous and Persi Diaconis. Longest increasing subsequences: from patience sorting to the baik-deift-johansson theorem. *Bulletin of the American Mathematical Society*, 36(4):413–432, 1999.

[Alt95] Lee Altenberg. The schema theorem and prices theorem. *Foundations of genetic algorithms*, 3:23–49, 1995.

[AN10] Alexandr Andoni and Huy L Nguyen. Near-optimal sublinear time algorithms for ulam distance. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 76–86. SIAM, 2010.

[Ang97] Peter J Angeline. Subtree crossover: Building block engine or macromutation. *Genetic Programming*, 97:9–17, 1997.

[Bac95] Thomas Back. Generalized convergence models for tournament-and selection. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 2–8, 1995.

[Bak87] James E Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the second international conference on genetic algorithms*, pages 14–21, 1987.

[BAW+15] Bogdan Burlacu, Michael Affenzeller, Stephan Winkler, Michael Kommenda, and Gabriel Kronberger. Methods for genealogy and building block analysis in genetic programming. In *Computational Intelligence and Efficiency in Engineering Systems*, pages 61–74. Springer, 2015.

[BD12] Richard A Brualdi and Emeric Deutsch. Adjacent q-cycles in permutations. *Annals of Combinatorics*, 16(2):203–213, 2012.

[BFM97] T Back, DB Fogel, and Z Michalewicz. Handbook of evolutionary computation.–ny, 1997.

[BP96] Vineet Bafna and Pavel A Pevzner. Genome rearrangements and sorting by reversals. *SIAM Journal on Computing*, 25(2):272–289, 1996.

[BS02] Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1(1):3–52, 2002.

[BT95] Tobias Blickle and Lothar Thiele. A comparison of selection schemes used in genetic algorithms, 1995.

[CDEL17] Dogan Corus, Duc-Cuong Dang, Anton V Eremeev, and Per Kristian Lehre. Level-based analysis of genetic algorithms and other search processes. *IEEE Transactions on Evolutionary Computation*, 22(5):707–719, 2017.

[Cor02] Heather J Cordell. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human molecular genetics*, 11(20):2463–2468, 2002.

[Cri12] Douglas E Critchlow. *Metric methods for analyzing partially ranked data*, volume 34. Springer Science & Business Media, 2012.

[Dav90] Yuval Davidor. Epistasis variance: Suitability of a representation to genetic algorithms. *Complex Systems*, 4(4):369–383, 1990.

[Dav91] Lawrence Davis. Handbook of genetic algorithms. 1991.

[DH98] Michael Deza and Tayuan Huang. Metrics on permutations, a survey. In *Journal of Combinatorics, Information and System Sciences*. Citeseer, 1998.

[DHO+06] Vladimir G Deı, Michael Hoffmann, Yoshio Okamoto, Gerhard J Woeginger, et al. The traveling salesman problem with few inner points. *Operations Research Letters*, 34(1):106–110, 2006.

[Dia88] Persi Diaconis. Group representations in probability and statistics. *Lecture notes-monograph series*, 11:i–192, 1988.

[DJ75] Kenneth Alan De Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, Ann Arbor, MI, USA, 1975. AAI7609381.

[DJ06] Kenneth A De Jong. *Evolutionary computation: a unified approach*. MIT press, 2006.

[DJFS97] Kenneth De Jong, D Fogel, and Hans-Paul Schwefel. Handbook of evolutionary computation. *Department of Industrial Engineering, University of Pittsburgh, USA: Oxford University Press und IOP Publishing. Kap. A history of evolutionary computation, Section A*, 2, 1997.

[DJS11] Benjamin Doerr, Daniel Johannsen, and Martin Schmidt. Runtime analysis of the (1+ 1) evolutionary algorithm on strings over finite alphabets. In *Proceedings of the 11th workshop proceedings on Foundations of genetic algorithms*, pages 119–126. ACM, 2011.

[DJW10] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Drift analysis and linear functions revisited. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–8. IEEE, 2010.

[Dra05] Brian Drake. The weak order on pattern-avoiding permutations. *FPSAC PROCEEDINGS 2005 ACTES SFCA 2005 Taormina ITALIA*, page 611, 2005.

[FOV08] Simon Fischer, Lars Olbrich, and Berthold Vöcking. Approximating wardrop equilibria with finitely many agents. *Distributed Computing*, 21(2):129–139, 2008.

[FSBB+93] Walter Fontana, Peter F Stadler, Erich G Bornberg-Bauer, Thomas Griesmacher, Ivo L Hofacker, Manfred Tacker, Pedro Tarazona, Edward D Weinberger, and Peter Schuster. Rna folding and combinatory landscapes. *Physical review E*, 47(3):2083, 1993.

[GL+85] David E Goldberg, Robert Lingle, et al. Alleles, loci, and the traveling salesman problem. In *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, volume 154, pages 154–159. Lawrence Erlbaum, Hillsdale, NJ, 1985.

[Gol89] David E Goldberg. Genetic algorithms in search, optimization and machine learning addison-wesley, 1989. *Reading, MA*, 1989.

[GT12] Ashish Ghosh and Shigeyoshi Tsutsui. *Advances in evolutionary computing: theory and applications*. Springer Science & Business Media, 2012.

[GW96] Paul Gitchoff and Günter P Wagner. Recombination induced hypergraphs: A new approach to mutation-recombination isomorphism. *Complexity*, 2(1):37–43, 1996.

[H+92] John Henry Holland et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[Hol75] John Holland. Adaptation in artificial and natural systems. *Ann Arbor: The University of Michigan Press*, 1975.

[HW99] Robert B Heckendorn and Darrell Whitley. Predicting epistasis from mathematical models. *Evolutionary computation*, 7(1):69–101, 1999.

[HY01] Jun He and Xin Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial intelligence*, 127(1):57–85, 2001.

[Inc04] Federico Incitti. The bruhat order on the involutions of the symmetric group. *Journal of Algebraic Combinatorics*, 20(3):243–261, 2004.

[J$^+$95] Terry Jones et al. One operator, one landscape. *Santa Fe Institute Technical Report*, pages 95–02, 1995.

[JF$^+$95] Terry Jones, Stephanie Forrest, et al. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *ICGA*, volume 95, pages 184–192, 1995.

[Joh11] Daniel Johannsen. Random combinatorial structures and randomized search heuristics. 2011.

[Jon95] Terry Jones. *Evolutionary algorithms, fitness landscapes and search*. PhD thesis, Citeseer, 1995.

[Ken48] Maurice George Kendall. Rank correlation methods. 1948.

[Koz92] John R Koza. *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press, 1992.

[Koz94] John R Koza. Genetic programming ii: Automatic discovery of reusable subprograms. *Cambridge, MA, USA*, 1994.

[Leh10] Per Kristian Lehre. Negative drift in populations. In *International Conference on Parallel Problem Solving from Nature*, pages 244–253. Springer, 2010.

[LP13] William B Langdon and Riccardo Poli. *Foundations of genetic programming*. Springer Science & Business Media, 2013.

[MBK99] Dirk C Mattfeld, Christian Bierwirth, and Herbert Kopfer. A search space analysis of the job shop scheduling problem. *Annals of Operations Research*, 86:441–453, 1999.

[MF00] Peter Merz and Bernd Freisleben. Fitness landscape analysis and memetic algorithms for the quadratic assignment problem. *IEEE transactions on evolutionary computation*, 4(4):337–352, 2000.

[MM19] Tina Malalanirainy and Alberto Moraglio. Runtime analysis of abstract evolutionary search with standard crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 293–294. ACM, 2019.

[Mor08] Alberto Moraglio. *Towards a geometric unification of evolutionary algorithms*. PhD thesis, University of Essex, 2008.

[Mor11] Alberto Moraglio. Abstract convex evolutionary search. In *Proceedings of the 11th workshop proceedings on Foundations of genetic algorithms*, pages 151–162. ACM, 2011.

[MP04] Alberto Moraglio and Riccardo Poli. Topological interpretation of crossover. In *Genetic and Evolutionary Computation Conference*, pages 1377–1388. Springer, 2004.

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge university press, 1995.

[MRC09] Boris Mitavskiy, Jonathan Rowe, and Chris Cannings. Theoretical analysis of local search strategies to optimize network communication subject to preserving the total number of links. *International Journal of Intelligent Computing and Cybernetics*, 2(2):243–284, 2009.

[MS02] Jesús Marín and Ricard V Solé. Controlling chaos in unidimensional maps using macroevolutionary algorithms. *Physical Review E*, 65(2):026207, 2002.

[MS17] Alberto Moraglio and Dirk Sudholt. Principled design and runtime analysis of abstract convex evolutionary search. *Evolutionary computation*, 25(2):205–236, 2017.

[NV92] Allen E Nix and Michael D Vose. Modeling genetic algorithms with markov chains. *Annals of mathematics and artificial intelligence*, 5(1):79–88, 1992.

[OSH87] IM Oliver, DJd Smith, and John RC Holland. Study of permutation crossover operators on the traveling salesman problem. In *Genetic algorithms and their applications: proceedings of the second International Conference on Genetic Algorithms: July 28-31, 1987 at the Massachusetts Institute of Technology, Cambridge, MA*. Hillsdale, NJ: L. Erlhaum Associates, 1987., 1987.

[PA12] Erik Pitzer and Michael Affenzeller. A comprehensive survey on fitness landscape analysis. In *Recent Advances in Intelligent Engineering Systems*, pages 161–191. Springer, 2012.

[Pal93] Richard Palmer. Optimization on rugged landscapes. In *SANTA FE INSTITUTE STUDIES IN THE SCIENCES OF COMPLEXITY-PROCEEDINGS VOLUME-*, volume 14, pages 3–3. ADDISON-WESLEY PUBLISHING CO, 1993.

[PB97] Adam Prügel-Bennett. Modelling evolving populations. *Journal of Theoretical Biology*, 185(1):81–95, 1997.

[Pel13] Ignacio M Pelayo. *Geodesic convexity in graphs*. Springer, 2013.

[PLMK08] Riccardo Poli, William B Langdon, Nicholas F McPhee, and John R Koza. *A field guide to genetic programming*. Lulu. com, 2008.

[QRSS98] RJ Quick, Victor J Rayward-Smith, and GD Smith. Fitness distance correlation and ridge functions. In *International Conference on Parallel Problem Solving from Nature*, pages 77–86. Springer, 1998.

[Rad91] Nicholas J Radcliffe. Equivalence class analysis of genetic algorithms. *Complex systems*, 5(2):183–205, 1991.

[Rea02] Nathan Reading. Order dimension, strong bruhat order and lattice properties for posets. *Order*, 19(1):73–100, 2002.

[Ric13] Hendrik Richter. Dynamic fitness landscape analysis. In *Evolutionary computation for dynamic optimization problems*, pages 269–297. Springer, 2013.

[RS02] Christian M Reidys and Peter F Stadler. Combinatorial landscapes. *SIAM review*, 44(1):3–54, 2002.

[Rud94a] Günter Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE transactions on neural networks*, 5(1):96–101, 1994.

[Rud94b] Günter Rudolph. Convergence of non-elitist strategies. In *Proceedings of the First IEEE International Conference on Evolutionary Computation*, pages 63–66. IEEE Press, 1994.

[Sag13] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*, volume 203. Springer Science & Business Media, 2013.

[Sch81] Hans-Paul Schwefel. *Numerical optimization of computer models*. John Wiley & Sons, Inc., 1981.

[Sch95] Hans-Paul Schwefel. Evolution and optimum seeking. sixth-generation computer technology series, 1995.

[SM91] H.P. Schwefel and R. Männer. *Parallel problem solving from nature: 1st workshop, PPSN I, Dortmund, FRG, October 1-3, 1990 : proceedings*. Lecture notes in computer science. Springer-Verlag, 1991.

[SN12] Andrew M Sutton and Frank Neumann. A parameterized runtime analysis of evolutionary algorithms for the euclidean traveling salesperson problem. In *AAAI*, 2012.

[SR96] Patrick Surry and Nicholas Radcliffe. Formal algorithms+ formal representations= search strategies. *Parallel Problem Solving from NaturePPSN IV*, pages 366–375, 1996.

[Sta02] Peter F Stadler. Fitness landscapes. In *Biological evolution and statistical physics*, pages 183–204. Springer, 2002.

[SVM93] Dirk Schlierkamp-Voosen and H Mühlenbein. Predictive models for the breeder genetic algorithm. *Evolutionary Computation*, 1(1):25–49, 1993.

[SW97] Peter F Stadler and Günter P Wagner. Algebraic theory of recombination spaces. *Evolutionary computation*, 5(3):241–275, 1997.

[SW99] Chris Stephens and Henri Waelbroeck. Schemata evolution and building blocks. *Evolutionary computation*, 7(2):109–124, 1999.

[Sys89] Gilbert Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 2–9, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[SZ03] Christopher R Stephens and Adolfo Zamora. EC theory: A unified viewpoint. In *Genetic and Evolutionary Computation Conference*, pages 1394–1405. Springer, 2003.

[Ula72] SM Ulam. Some ideas and prospects in biomathematics. *Annual review of biophysics and bioengineering*, 1(1):277–292, 1972.

[vDV93] Marcel LJ van De Vel. *Theory of convex structures*, volume 50. Elsevier, 1993.

[VL91] Michael D Vose and Gunar E Liepins. Punctuated equilibria in genetic search. *Complex systems*, 5(1):31–44, 1991.

[Vos99] Michael D Vose. *The simple genetic algorithm: foundations and theory*, volume 12. MIT press, 1999.

[Weg01] Ingo Wegener. Theoretical aspects of evolutionary algorithms. In *International Colloquium on Automata, Languages, and Programming*, pages 64–78. Springer, 2001.

[Weg03] Ingo Wegener. Methods for the analysis of evolutionary algorithms on pseudo-boolean functions. In *Evolutionary optimization*, pages 349–369. Springer, 2003.

[Wei90] Edward Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological cybernetics*, 63(5):325–336, 1990.

[Wit06] Carsten Witt. Runtime analysis of the (μ+ 1) ea on simple pseudo-boolean functions. *Evolutionary Computation*, 14(1):65–86, 2006.

[WW05] Ingo Wegener and Carsten Witt. On the optimization of monotone polynomials by simple randomized search heuristics. *Combinatorics, Probability and Computing*, 14(1-2):225–247, 2005.

[YO69] Takemi Yanagimoto and Masashi Okamoto. Partial orderings of permutations and monotonicity of a rank correlation statistic. *Annals of the Institute of Statistical Mathematics*, 21(1):489–506, 1969.

[ZTV+18] Ivan Zelinka, Lukas Tomaszek, Pandian Vasant, Tran Trong Dao, and Duy Vo Hoang. A novel approach on evolutionary dynamics analysis–a progress report. *Journal of Computational Science*, 25:437–445, 2018.