Università degli Studi di Milano

Physics Department

MONOPARTITE EUCLIDEAN TRAVELLING SALESMAN PROBLEM IN ONE DIMENSION

Supervisor:
Prof. Sergio CARACCIOLO

Assistant Supervisors:
Andrea DI GIOACCHINO
Enrico M. MALATESTA

Candidate:
Carlo VANONI
Matr. n. 867304

A.Y. 2017-2018

Abstract

This thesis work deals with the *Travelling Salesman Problem*. Many versions of this problem have been defined through the years and many authors and researchers were involved in its study, essentially for two reasons: its intrinsic difficulty, since it is a $\mathcal{NP} - hard$ problem, and its infinitely large applications. Essentially, given N cities and the distance between each couple of them, one aims to find the shortest closed path that visits each city only once. Here we focus on the one-dimensional Euclidean version of it. In particular we obtain the exact optimal solutions for a class of cost functions. Moreover, for another class of cost functions, the C-functions, we show that the optimal solution always belongs to a subset, that we characterize, of all the possible Hamiltonian cycles of the given graph.

The 2-factor problem is also considered; it consists in finding the minimum-weight loop cover of an undirected graph. In particular, we consider this problem on the complete graph embedded in a one dimensional interval and we find the optimal 2-factor solution for a cost function belonging to the C-function class.

Finally, we study the average cost of the optimal Travelling Salesman Problem configuration in the limit of large number of points, showing the connection of this average properties with the ones of previously combinatorial optimization problems that have been studied.

To my family

Contents

Chapter 1

Introduction

If a salesman, starting from his home city, is to visit exactly once each city on a given list and then return home, it is plausible for him to select the order in which he visits the cities so that the total cost paid in travelling his tour is as small as possible. Let us assume he knows, for each pair of cities, the cost that must be paid in going from one to the other. Then he has all the data necessary to find the minimum, but it is by no means obvious how to use these data in order to get the answer. The problem of finding the optimal tour is known as the *Travelling Salesman Problem*.
The *Travelling Salesman Problem*, or *TSP*, is one of the most famous combinatorial optimization problems, because of the simplicity in its statement and the difficulty in its solution. In this work we study the *TSP* in its one dimensional version, i.e. supposing that all the cities lay on a line, introducing as the cost that must be paid in travelling from one city to the other a function of the Euclidean distance between the two cities.
The thesis work is organized as follows. In chapter 1 we give a general introduction to the problem, discussing its origins, briefly presenting some recent developments and giving some necessary notions. In chapter 2 we expose the problem we have considered and we give our solutions for the studied cases. In this chapter we also give all the proofs of the presented solutions. Moreover, we study the limit of large number of points and we briefly show the results of the numerical simulations performed. In chapter 3 we give some conclusions. In appendices A and B we present two results. The first regards a topological property of fundamental importance for the proofs of our solutions, while the second is the solution we found to the 2-factor problem in a specific case and that we used to solve the TSP in the same situation.

1.1 Background history

The origin of the travelling salesman problem are unclear. The general form of the TSP has been first studied by mathematicians during the 1930s in Vienna and at Harvard, in particular by Karl Menger. In 1950s and 1960s the problem became increasingly popular in scientific circles in Europe and in the USA, especially when the problem has been expressed as an integer linear program by G. Dantzig, D.R. Fulkerson and S.M. Johnson from the RAND Corporation. They published in 1954 a description of a method for solving the TSP and illustrated the power of this method by solving an instance with 49 cities, an impressive size at that time. They created this instance by picking one city from each of the 48 states in the U.S.A. (Alaska and Hawaii became states only in 1959) and adding Washington, D.C.; the costs of travel between these cities were defined by road distances. Over the years, many TSP solutions arose, with an always increasing number of instances. In 2006, William J. Cook and others computed an optimal tour through an 85,900-city instance given by a microchip layout problem, currently the largest solved TSPLIB instance.

An innovative approach to the study of the TSP and, in general, to combinatorial optimization problems, has been introduced by Marc Mezard and Giorgio Parisi in 1986 when they understood that many methods and concepts developed in the theory of spin-glasses could be applied in combinatorial optimization problems [13–17]. In the early 1980s has been realized and demonstrated that the determination of the ground state of an infinite range spin-glass is a $\mathcal{NP}$-*complete* problem; therefore the techniques developed in the study of this particular branch of statistical mechanics, e.g. the replica method and the cavity method, could be used to study other $\mathcal{NP}$-*complete* problems. In the theory of spin-glasses, the replica method is used to have informations on the properties of the ground state of the system even if they are not sufficient to determine the ground state (see Section1.5); in other words, the replica method allows one to solve the spin-glass model without having to compute the ground state. If the cost of the TSP is thought as the energy, the ground state, i.e. the minimum energy state, is the optimal solution of the problem. Mezard and Parisi in [14] used the replica method to derive analytical estimates for thermodynamic quantities, such as the length of the shortest path, for the TSP in which the distances between the cities are independent random variables. The study of the TSP with random weights can be seen as an infinite dimension version of it.

1.2 TSP and computational complexity

The computational complexity of a problem is a measure of the computational resources, typically time, required to solve the problem. Defined the *worst case time complexity T(n)* as:

$$T(n) = \max_{|x|=n} t(x) \tag{1.2.1}$$

where $t(x)$ is the running time of the algorithm for input data x, let us define:

- $T(n)$ is of order at most $g(n)$ and we write $T(n) = \mathcal{O}(g(n))$ if there exist positive constants c and n_0 such that $T(n) \leq cg(n)$ for all $n \geq n_0$.
- $T(n)$ is of order $g(n)$ and we write $T(n) = \Theta(g(n))$ if there exist positive constants c_1, c_2 and n_0 such that $c_1 g(n) \leq T(n) \leq c_2 g(n)$ for all $n \geq n_0$.

In computational complexity, all problems which can be solved by a polynomial algorithm, i.e. an algorithm with time complexity $\Theta(n^k)$ for some k, are lumped together and called *tractable*. Problems which can only be solved by algorithms with non-polynomial running time like $\Theta(2^n)$ or $\Theta(n!)$ are also lumped together and called *intractable*.
In order to give a precise definition of complexity classes it is needed to consider *decision problems*, i.e. problems whose solution is either "yes" or "no". Every optimization problem can be turned into a decision problem adding a bound B to the instance. In the case of the TSP this becomes: *given N cities, the set of distances between them and a number $B \geq 0$, is there a TSP solution such that the total cost needed to travel the tour is smaller or equal to B?*
Given this definition, we can now define the complexity classes of decision problems:

Definition 1.2.1. *A decision problem P is element of the class $\mathcal{P}$ if and only if it can be solved by a polynomial time algorithm.*

Definition 1.2.2. *A decision problem P is element of the class $\mathcal{NP}$ if and only if a solution can be verified to be correct in polynomial time.*

Note that "$\mathcal{NP}$" stands for *polynomial non-deterministic* algorithm, not for *non-polynomial*, even though every known algorithm for a $\mathcal{NP}$ problem solves the problem in exponential time.

An immediate consequence is that: $\mathcal{P} \subseteq \mathcal{NP}$. Let us give other two definitions that allow to give a hierarchy in the $\mathcal{NP}$:

Definition 1.2.3. *We say a problem P_1 is polynomially reducible to a problem P_2 and we write $P_1 \leq P_2$ if there exist a polynomial algorithm for P_1 provided that there is a polynomial algorithm for P_2.*

Definition 1.2.4. *A problem P is $\mathcal{NP}$- complete if $P \in \mathcal{NP}$ and $Q \leq P$ for all $Q \in \mathcal{NP}$.*

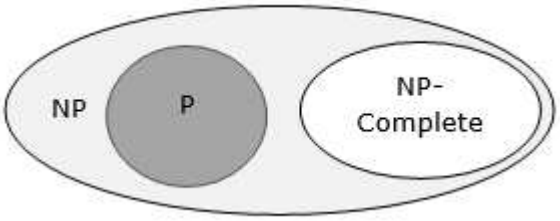


Figure 1.1: The map of computational complexity

The class of $\mathcal{NP}$-*complete* problems collects the hardest problems in $\mathcal{NP}$. The class $\mathcal{NP}$ not only contains a large number of problems with important applications, but represent a real challenge: all problems in $\mathcal{NP}$ still have a chance to be in $\mathcal{P}$. A proof of non-existence of a polynomial algorithm for a single problem from $\mathcal{NP}$ would establish that $\mathcal{P} \neq \mathcal{NP}$. As long as such a proof is missing,

$$\mathcal{P} \stackrel{?}{=} \mathcal{NP} \tag{1.2.2}$$

represent the most famous open conjecture in theoretical computer science and it is one of the seven Millennium Prize Problems stated by the Clay Mathematics Institute in 2000.
We say that an optimization problem is $\mathcal{NP}$-*hard* if its decision variant is $\mathcal{NP}$-*complete*. The decisional version of TSP is $\mathcal{NP}$-*complete* and it has been proved in [19] that the Euclidean TSP is a $\mathcal{NP}$-*complete* problem, even if its instances are restricted to be realizable by sets of points on the Euclidean plane. The problems in the $\mathcal{NP}$ class are intractable unless some mathematical insight provides us with a polynomial algorithm to avoid exhaustive search. Such an insight promotes a problem into the class $\mathcal{P}$ of polynomially soluble problems.

1.3 Graph Theory and Combinatorial Optimization Problems

In order to properly formulate the *TSP*, and, in general, a combinatorial optimization problem, some basic Graph Theory notions are needed. Thus, we present, in this paragraph, the fundamental definitions that are indispensable for our purpose and how some of the most important combinatorial optimization problems are defined in therms of graphs.

Definition 1.3.1. *A graph is an ordered pair $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is a finite set of points, called vertices, and $\mathcal{E}$ is a collection of 2-element subsets of $\mathcal{V}$, called edges.*

Usually, in a picture of a graph, vertices are represented by dots and edges are represented by lines. A graph is said to be *simple* if it does not contain loops and multiple edges, and it is said to be *undirected* if the couple of vertices in an edge is not ordered.
If v and w are vertices of a graph $\mathcal{G}$ and $e = \{v, w\}$ is an edge of $\mathcal{G}$, we say that e *joins* v and w, that v and w are *extrema* of e and that v and w are *adjacent*. If the edge e is oriented from v to w we write $e = (v, w)$, if it is not oriented we write $e = \{v, w\} = \{w, v\}$.

Definition 1.3.2. *A graph $\mathcal{G}' = \{\mathcal{V}', \mathcal{E}'\}$ is said to be a subgraph of $\mathcal{G}$ ($\mathcal{G}' \subseteq \mathcal{G}$) if $\mathcal{V}' \subseteq \mathcal{V}$ and $\mathcal{E}' \subseteq \mathcal{E}$. In case $\mathcal{V}' = \mathcal{V}$ the graph is said to be spanning.*

Definition 1.3.3. *A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is said to be bipartite if the set $\mathcal{V}$ of vertices can be bipartite in two disjoint subsets, $\mathcal{V}_1$ and $\mathcal{V}_2$ with $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$, and such that every edge in $\mathcal{E}$ has the form (v_1, v_2) or (v_2, v_1), where $v_1 \in \mathcal{V}_1$ and $v_2 \in \mathcal{V}_2$.*

Definition 1.3.4. *A graph is said to be complete if every two vertices are adjacent. The N-vertices complete graph is denoted $\mathcal{K}_N$. A bipartite graph is said to be complete if, for every couple of vertices $v_1 \in \mathcal{V}_1$ and $v_2 \in \mathcal{V}_2$ there exist an edge that joins them. The complete bipartite graph is denoted $\mathcal{K}_{N,M}$, where $N = |\mathcal{V}_1|$ and $M = |\mathcal{V}_2|$.*

A *walk* ω of length $|\omega| = k \geq 0$ connecting v_0 with v_k in $\mathcal{G}$ is a sequence $(v_0, e_1, v_1, \ldots, e_k, v_k)$ such that all $v_i \in \mathcal{V}$, all $e_i \in \mathcal{E}$ and $v_{i-1}, v_i \in e_i$ for $1 \leq i \leq k$. A *loop* is a walk connecting the starting point v_0 with itself. A *path* in $\mathcal{G}$ is a walk in which $v_0, v_1, \ldots, v_k$ are distinct vertices of $\mathcal{G}$, and $e_1, e_2, \ldots, e_k$ are distinct edges of $\mathcal{G}$. A *cycle* in $\mathcal{G}$ is such that:

- $v_0, v_1, \ldots, v_{k-1}$ are distinct vertices of $\mathcal{G}$, and $v_k = v_0$;
- $e_1, e_2, \ldots, e_k$ are distinct edges of $\mathcal{G}$;
- $k \geq 2$.

Definition 1.3.5. *A graph $\mathcal{G}$ is said to be connected if every pair of vertices in $\mathcal{G}$ can be connected by a walk.*

Definition 1.3.6. *A forest is a graph that contains no cycles. A tree is a connected forest.*

Definition 1.3.7. *Given a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, a matching $\mathcal{M}$ in $\mathcal{G}$ is a set of pairwise non-adjacent edges, none of which are loops, i.e. no two edges share a common vertex. A maximal matching is a matching $\mathcal{M}$ of a graph $\mathcal{G}$ with the property that if any edge not in $\mathcal{M}$ is added to $\mathcal{M}$, it is no longer a matching. A perfect matching is a matching which matches all vertices of the graph.*

Using this definitions, we can now summarize the statements of the principal Euclidean combinatorial optimization problems:

- **Random Euclidean Matching Problem**: given $2N$ random points $x_1, \dots, x_{2N}$ on a Euclidean domain and defined a weight $w_{ij} = c(||x_i - x_j||)$ to the couple (x_i, x_j), the problem consists in finding the *perfect matching* μ that minimizes the cost $E(\mu) = \sum_{(x_i,x_j)\in\mu} w_{ij}$;
- **Random Euclidean Assignment Problem**: given 2 sets of N random points $x_1, \dots, x_N, y_1, \dots, y_N$ on a Euclidean domain, define a weight $w_{ij} = c(||x_i - y_j||)$ to the couple (x_i, y_j). In this case only points of different sets can be coupled, thus the problem consists in finding the permutation $\pi \in \mathcal{S}_N$ such that the cost $E(\pi) = \sum_{i=1}^{N} w_{i\pi(i)}$ is minimized;
- **Euclidean 2-factor (or 2-matching)**: given a generic undirected graph, a *factor* is a spanning subgraph, while a *k-factor* is a factor k-regular, that is in which each vertex belongs exactly to k edges. Let us suppose that a weight $w_e = c(||x_i - x_j||)$ is associated to the edge $e = (x_i, x_j)$ that belongs to the considered 2-factor ν. Then the problem consists in finding the 2-factor that minimizes the cost $E(\nu) = \sum_{e\in\nu} w_{ij}$;
- **Euclidean TSP**: see Section 2.1.

1.4 Recent results

Encouraged by the infinite-dimension results, professor Caracciolo and his PhD students started studying one dimensional versions of combinatorial optimization problems like the assignment, the matching, the 2-factor and the TSP, both in their complete and bipartite versions. The importance in knowing the solutions to these versions of the problems is also due to the fact that they help us understand what happens in higher dimensions, in particular in two dimensions. Here we report a list of the results obtained:

- Euclidean matching problem: equivalence relations between the Euclidean bipartite matching problem on the line and on the circle, computation of the correlation function and optimal cost of the original problem in the thermodynamic limit [4].

- Euclidean bipartite matching problem: a scaling hypothesis has been developed in collaboration with professor Parisi [7]: in particular it has been proposed the asymptotic behaviour of the average optimal cost in dimension $d = 1$, $d = 2$ and $d > 2$, verifying that in the limit $d \to \infty$ it coincides with the subleading scaling of the random assignment problem. Then it has been computed the two-point correlation function for spin configurations obtained by solving the Euclidean matching problem [2]. In particular, the optimal solution for both the assignment problem and the matching problem in one dimension and for a large class of cost functions has been found.

- Bipartite Euclidean *TSP* in one dimension: the optimal solution for a convex and increasing cost function has been found, and the results obtained have been compared with the assignment, showing that the cost for the optimal *TSP* in not smaller than twice the cost in the optimal assignment on the same set of points [6].

- Euclidean 2-matching in one dimension: it consists of finding the minimum-weight loop cover of an undirected graph. It has been solved for a convex cost function and both in its complete and bipartite version [5].

- Bipartite *TSP* and 2-factor problem in two dimensions: in this case the problems cited above are studied in two dimensions. In particular, in [3], it is shown that the average optimal cost of the bipartite *TSP* in two dimensions is simply related to the average cost of the assignment problem. As a byproduct the authors show that the 2-factor problem has the same average optimal cost.

1.5 The Replica Method and the *TSP* in Physics

Before discussing why the *TSP* has found a lot of interest in physics, let us briefly review the replica method, that will be useful for our purpose.
In statistical physics one is interested in average properties of a given system. In particular, in the canonical formalism the partition function is defined as:

$$Z(\beta) = \sum_{\mu \in \mathcal{M}} e^{-\beta E(\mu)} \tag{1.5.1}$$

where β is the inverse temperature of the system, μ is a particular configuration in the set $\mathcal{M}$ of all the possible configurations and $E(\mu)$ is the energy of the given configuration. If the partition function is given, for fixed $\mathcal{M}$, one is able to derive all the macroscopic quantities of interest, such as the energy, given by:

$$E = -\frac{\partial \ln Z(\beta)}{\partial \beta} \tag{1.5.2}$$

However, in disordered systems, the set $\mathcal{M}$ of all the possible configurations is an instance of the problem. Thus, if one is interested in averaging the macroscopic properties also over all the possible instances, the following quantity must be computed:

$$\overline{E} = -\frac{\partial \overline{\ln Z(\beta)}}{\partial \beta} \tag{1.5.3}$$

where $\overline{\bullet}$ denotes the average over all instances $\mathcal{M}$.

The average in (1.5) is hard to be computed analytically and the it has been approached using the *replica method*. It consists in expanding the logarithm as:

$$x^n = e^{n \ln x} \approx 1 + n \ln x \quad \text{as} \quad n \to 0 \tag{1.5.4}$$

Therefore:

$$\ln x = \lim_{n \to 0} \frac{x^n - 1}{n} \tag{1.5.5}$$

This way we can compute the average as:

$$\overline{\ln Z(\beta)} = \lim_{n \to 0} \frac{\overline{Z(\beta)^n - 1}}{n} = \lim_{n \to 0} \frac{\overline{Z(\beta)^n} - 1}{n} \tag{1.5.6}$$

This procedure is physically equivalent to considering n copies, or replicas, of the system and then average over it. In order to average over n copies of the system, n has to be an integer; in the limit of Eq. (1.5), instead, n is sent continuously to zero. This problem can be avoided considering an analytic continuation of $Z(\beta)^n$, where n is promoted to be a real number.

Parisi and Mezard understood [14] that this powerful procedure could be used also in studying combinatorial optimization problems, in particular TSP. In fact, if $\mathcal{M}$ is the set of all possible Hamiltonian cycles of a given instance of the problem and $E(\mu)$ is the cost of the cycle $\mu \in \mathcal{M}$, then the replica method makes it possible to find the average cost needed to perform the journey.

The particular attention that the TSP, and in general combinatorial optimization problems, gained in the theoretical physics community in the last decades has a double motivation.
Firstly, as we outline in Section1.1 and above in this paragraph, physicists understood that many concepts and methods developed in statistical mechanics, in particular in the theory of spin glasses, could be applied in the solution of this kind of problems, for example the replica method explained in this section. This way, physicists faced problems reserved to mathematicians and theoretical computer scientists in an innovative way that brought to important results, not only analytical, but also computational. In fact, the trick of mapping the optimization problem into a statistical system, were

the energy replaces the cost, and then find the minimum of the energy lowering the temperature is commonly used in computational simulations.
While statistical physics provides methods indispensable to derive average properties of combinatorial optimization problems' solutions, on the other side, the solutions, at fixed instance, to this kind of problems, as the one presented in this thesis work, can be used the opposite way compared to the replica method, i.e. to solve disordered systems problems. If a statistical mechanics problem can be mapped into a combinatorial optimization one whose solution in known, then also the physical problem can be solved.

Chapter 2

ETSP in One Dimension

2.1 The problem

Given a generic (undirected) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and a *cycle* of length k in $\mathcal{G}$, the cycle is *Hamiltonian* when the visited vertices are all different and the cardinality of the set of vertices $|\mathcal{V}|$ is exactly k for $k > 2$. In other terms, a Hamiltonian cycle is a closed path visiting all the vertices in $\mathcal{V}$ only once. The determination of the existence of an Hamiltonian cycle is an NP-complete problem (see Johnson and Papadimitriou in [12]). A graph that contains a Hamiltonian cycle is called a Hamiltonian graph. The complete graph with N vertices $\mathcal{K}_N$ is Hamiltonian for $N > 2$. The bipartite complete graph with $N + M$ vertices $\mathcal{K}_{N,M}$ is Hamiltonian for $M = N > 1$.

Let us denote by $\mathcal{H}$ the set of Hamiltonian cycles of the graph $\mathcal{G}$. Let us suppose now that a weight $w_e > 0$ is assigned to each edge $e \in \mathcal{E}$ of the graph $\mathcal{G}$. We can associate to each Hamiltonian cycle $h \in \mathcal{H}$ a total cost

$$E(h) := \sum_{e \in h} w_e \,. \tag{2.1.1}$$

In the (weighted) Hamiltonian cycle problem we search for the Hamiltonian cycle $h \in \mathcal{H}$ such that the total cost in (2.1.1) is minimized, i.e., the optimal Hamiltonian cycle $h^* \in \mathcal{H}$ is such that

$$E(h^*) = \min_{h \in \mathcal{H}} E(h) \,. \tag{2.1.2}$$

When the N vertices of $\mathcal{K}_N$ are seen as cities and the weight for each edge is the cost paid to cover the route distance between the cities, the search for h^* is the *travelling salesman problem*. For example, consider when the graph $\mathcal{K}_N$ is embedded in $\mathbb{R}^d$, that is for each $i \in [N] = \{1, 2, \ldots, N\}$ we associate a point $x_i \in \mathbb{R}^d$, and for $e = (i, j)$ with $i, j \in [N]$ we introduce a cost which is a function of their Euclidean distance $w_e = |x_i - x_j|^p$ with $p \in \mathbb{R}$. When $p = 1$, we obtain the usual *Euclidean TSP*.

The simplest way to introduce randomness in the problem is to consider the weights w_e independent and identically distributed random variables. In this case the problem is called *random TSP* and has been extensively studied by disordered system techniques such as replica and cavity methods [11, 14, 15, 18, 21, 22, 24] and by a rigorous approach [25]. In the random Euclidean TSP [1, 9, 10, 20, 23], instead, the positions of the points are generated at random and as a consequence the weights will be correlated. The typical properties of the optimal solution are of interest, and in particular the average optimal cost

$$\overline{E} := \overline{E(h^*)}\,, \tag{2.1.3}$$

where we have denoted by a bar the average over all possible realization of the disorder.
In (2.1.1) the weight w_e is, in general, a function of the Euclidean distance between two points and, in this work, we will use the function:

$$w_e = |x_i - x_j|^p\,; \qquad p \in \mathbb{R}. \tag{2.1.4}$$

However, as it is shown in [4], the cost function in (2.1.4), when $p < 0$, has the same matching properties of a class of functions usually called C-functions:

Definition 2.1.1. *We will say that a function $f : [0,1] \to \mathbb{R}$ is a C-function if, given $0 < z_1 < z_2 < 1$, for any $\eta \in (0, 1 - z_2), \mu \in (z_2, 1)$*

$$f(z_2) - f(z_1) \leq f(\eta + z_2) - f(\eta + z_1) \tag{2.1.5}$$

$$f(z_2) - f(z_1) \leq f(\mu - z_2) - f(\mu - z_1) \tag{2.1.6}$$

In case the cost is a C-function, the expression for the weigh in (2.1.1) becomes:

$$w_e = f(|x_i - x_j|) \tag{2.1.7}$$

where f is a C-function. This property will allow us to extend the results obtained for the $p < 0$ case to the situation in which the cost is a general C-function.

2.2 Hamiltonian cycles in the complete graph

In our problem, we shall consider the complete graph $\mathcal{K}_N$ with N vertices, that is with vertex set $V = [N] := \{1, \ldots, N\}$. A Hamiltonian cycle, as explained in Sect.2.1, is a unique closed walk on the graph which visits all the vertices. This graph has $\frac{(N-1)!}{2}$ Hamiltonian cycles. Indeed, each permutation π in the symmetric group of N elements, $\pi \in \mathcal{S}_N$, defines a Hamiltonian cycle on $\mathcal{K}_N$.

The sequence of points $(\pi(1), \pi(2), \ldots, \pi(N), \pi(1))$ defines a closed walk with starting point $\pi(1)$, but the same walk is achieved by choosing any other vertex as starting point and the walk in the opposite order, that is,

$$(\pi(1), \pi(N), \ldots, \pi(2), \pi(1)) \tag{2.2.1}$$

corresponds to the same Hamiltonian cycle. As the cardinality of $\mathcal{S}_N$ is $N!$ we get that the number of Hamiltonian cycles in $\mathcal{K}_N$ is $\frac{N!}{2 \cdot N} = \frac{(N-1)!}{2}$.

There is another way to associate permutations to Hamiltonian cycles. Let $\pi^k := \pi \circ \pi^{k-1}$ for integer k and π^0 be the identity function. Of course $\pi^N = \pi^0$, $\pi \in \mathcal{S}_N$. A permutation $\pi \in \mathcal{S}_N$ is said to be k-cycle if it formed by a unique cycle of length k and $N-k$ fixed points. There are $\frac{1}{k}\frac{(N)!}{(N-k)!}$ k-cycle in $\mathcal{S}_N$. Let us consider now the orbit of the point j under the action of π, that is the sequence of points $(\pi^0, \pi, \pi^2, \ldots, \pi^N)(j)$, with $j \in [N]$. This sequence defines a Hamiltonian cycle if and only if the permutation π is a N-cycle. If π is a N-cycle also π^{-1} is a N-cycle. It provides the same closed walk in the opposite direction. As the cardinality of the N-cycles in $\mathcal{S}_N$ is $(N-1)!$ we get, once more, that the number of Hamiltonian cycles in $\mathcal{K}_N$ is $\frac{(N-1)!}{2}$.

2.3 Optimal cycle for the bipartite complete graph, $p > 1$

We will resume here, first, the results for the optimal Hamiltonian cycle that have already been shown for the case of the bipartite graph $\mathcal{K}_{N,N}$ embedded on the interval $\Omega = [0, 1] \subset \mathbb{R}$.

Let us call $\mathcal{R} := \{r_i\}_{i=1,\ldots,N} \subset \Omega$ and $\mathcal{B} := \{b_j\}_{j=1,\ldots,N} \subset \Omega$ two sets of points, both of cardinality N, in the interval Ω. All points are supposed to be identically and independently distributed over Ω according to the flat distribution.

Remember that the bipartite complete graph $\mathcal{K}_{N,N}$ has the vertex set with the same cardinality but $\frac{N!\,(N-1)!}{2}$ Hamiltonian cycles. Indeed, given two permutations $\sigma, \pi \in \mathcal{S}_n$ the sequence of points

$$h[(\sigma, \pi)] := (r_{\sigma(1)}, b_{\pi(1)}, r_{\sigma(2)}, b_{\pi(2)}, \ldots, r_{\sigma(N)}, b_{\pi(N), r_{\sigma(1)}}) \tag{2.3.1}$$

defines a closed walk starting from $\sigma(1)$ and therefore if we fix $\sigma(1) = 1$ this sequence and the sequence obtained by reversing the order define a Hamiltonian cycles.

In [6] Caracciolo et al. proved that, if both red and blue points are ordered, i.e. $r_1 \leq \cdots \leq r_N$ and $b_1 \leq \cdots \leq b_N$ the optimal cycle is defined by $h^* = \tilde{h} := h[(\tilde{\sigma}, \tilde{\pi})]$ with

$$\tilde{\sigma}(i) := \begin{cases} 2i - 1 & i \leq (N+1)/2 \\ 2N - 2i + 2 & i > (N+1)/2 \end{cases} \tag{2.3.2}$$

and

$$\tilde{\pi}(i) = \tilde{\sigma}(N + 1 - i) \tag{2.3.3}$$

is exactly the reverse of $\tilde{\sigma}$ and has the same optimal cost.

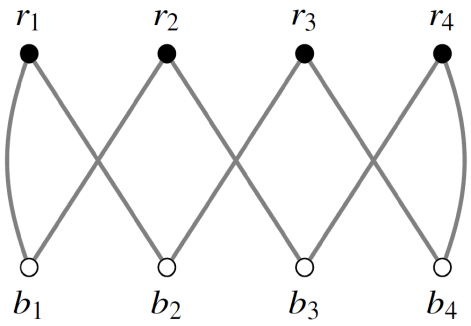


Figure 2.1: The optimal Hamiltonian cycle $\tilde{h}$ for $N = 4$. Blue and red points are chosen in the unit interval and sorted in increasing order.

Therefore at fixed ordered set of red and blue points the optimal cost is exactly

$$E_N(\tilde{h}) = |r_1 - b_1|^p + |r_N - b_N|^p + \sum_{i=1}^{N-1} \left[|b_{i+1} - r_i|^p + |r_{i+1} - b_i|^p \right]. \tag{2.3.4}$$

2.4 The $p > 1$ case: optimal cycle

Notation: we will say that an edge e has n crossings if there are n edges that intersects e. We will say that a graph has n matchings if the total number of intersections between the edges is n.
We can now prove that, given an ordered sequence $\mathcal{R} = \{r_i\}_{i=1,\dots,N}$ of N points in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, if

$$h^* = h[\tilde{\sigma}] = (r_{\tilde{\sigma}(1)}, r_{\tilde{\sigma}(2)}, \dots, r_{\tilde{\sigma}(N)}, r_{\tilde{\sigma}(1)}) \tag{2.4.1}$$

with $\tilde{\sigma}$ defined as in (2.3.2) (and the same Hamiltonian cycle is obtained using instead $\tilde{\pi}$ given in (2.3.3)), then

Proposition 2.4.1. *The Hamiltonian cycle which provides the optimal cost is h^*.*

Proof. For each $\sigma \in \mathcal{S}_N$, with $\sigma(1) = 1$, the cost the Hamiltonian cycle associated to it is given by

$$E_N(h[\sigma]) = |r_{\sigma(2)} - r_1|^p + |r_{\sigma(N)} - r_1|^p + \sum_{i=2}^{N-1} |r_{\sigma(i+1)} - r_{\sigma(i)}|^p \tag{2.4.2}$$

Let us introduce now a new set of ordered points $\mathcal{B} := \{b_j\}_{j=1,\dots,N} \subset [0,1]$ such that

$$b_i = \begin{cases} r_1 & \text{for} \quad i = 1 \\ r_{i-1} & \text{otherwise} \end{cases} \tag{2.4.3}$$

and consider the Hamiltonian cycle on the complete bipartite graph with vertex sets $\mathcal{R}$ and $\mathcal{B}$

$$h[(\sigma, \pi_\sigma)] := (r_1, b_{\pi_\sigma(1)}, r_{\sigma(2)}, b_{\pi_\sigma(2)}, \dots, r_{\sigma(N)}, b_{\pi_\sigma(N)}, r_{\sigma(1)}) \tag{2.4.4}$$

so that

$$\pi_\sigma(i) = \begin{cases} 2 & \text{for} \quad i = 1 \\ \sigma(i) + 1 & \text{for} \quad i < k \\ \sigma(i+1) + 1 & \text{for} \quad i \geq k \\ 1 & \text{for} \quad i = N \end{cases} \tag{2.4.5}$$

where k is such that $\sigma(k) = N$. We have therefore

$$\begin{aligned} &(b_{\pi_\sigma(1)}, b_{\pi_\sigma(2)}, \dots, b_{\pi_\sigma(k-1)}, b_{\pi_\sigma(k)}, \dots, b_{\pi_\sigma(N-1)}, b_{\pi_\sigma(N)}) \\ &\qquad = (r_1, r_{\sigma(2)}, \dots, r_{\sigma(k-1)}, r_{\sigma(k+1)}, \dots, r_{\sigma(N)}, r_1) \end{aligned} \tag{2.4.6}$$

The cost of this bipartite matching is:

$$\begin{aligned} E[h[(\sigma, \pi_\sigma)]] &= |r_1 - b_{\pi_\sigma(1)}|^p + |r_{\sigma(2)} - b_{\pi_\sigma(1)}|^p + |r_N - b_{\pi_\sigma(N)}|^p \\ &+ |r_{\sigma(1)} - b_{\pi_\sigma(N)}|^p + \sum_{i=2,}^{N-1} \left[|r_{\sigma(i)} - b_{\pi_\sigma(i)}|^p + |r_{\sigma(i+1)} - b_{\pi_\sigma(i)}|^p \right] . \end{aligned} \tag{2.4.7}$$

By construction (using (2.4.6)):

$$E_N(h[\sigma]) = E_N(h[(\sigma, \pi_\sigma)]) \geq E_N(h[(\tilde{\sigma}, \tilde{\pi})]) = E_N(h[(\tilde{\sigma}, \pi_{\tilde{\sigma}})]) = E_N(h[\tilde{\sigma}]), \tag{2.4.8}$$

where the fact that $\tilde{\pi} = \pi_{\tilde{\sigma}}$ can be checked using (2.3.2) and (2.3.3) and (2.4.5).

□

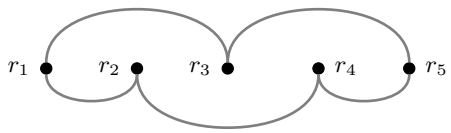


Figure 2.2: The optimal Hamiltonian cycle for $p > 1$ for $N = 5$

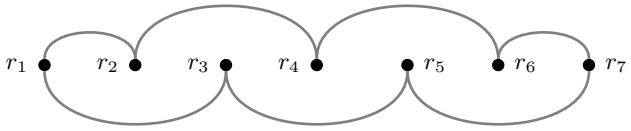


Figure 2.3: The optimal Hamiltonian cycle for $p > 1$ for $N = 7$

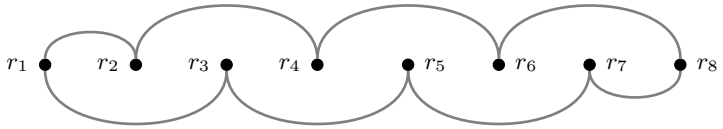


Figure 2.4: The optimal Hamiltonian cycle for $p > 1$ for $N = 8$

2.5 The $0 < p < 1$ case: optimal cycle

We now prove that, given an ordered sequence $\mathcal{R} = \{r_i\}_{i=1,\dots,N}$ of N points in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, if $0 < p < 1$ and if

$$h^* = h[\mathbb{1}] = (r_{\mathbb{1}(1)}, r_{\mathbb{1}(2)}, \dots, r_{\mathbb{1}(N)}, r_{\mathbb{1}(1)}) \tag{2.5.1}$$

where $\mathbb{1}$ is the identity permutation, i.e.:

$$\mathbb{1}(j) = j \tag{2.5.2}$$

then

Proposition 2.5.1. *The Hamiltonian cycle which provides the optimal cost is h^*.*

Of course we have:

$$h^* = h[\mathbb{1}] = (r_1, r_{\mathbb{1}(2)}, \dots, r_{\mathbb{1}(N)}, r_{(1)}) = (r_1, r_2, \dots, r_N, r_1) \tag{2.5.3}$$

Before proving this result, we will enunciate and demonstrate a lemma we will use in the proof of Proposition 5.1.

Lemma 1. *Given a Hamiltonian cycle with the connection between the points drown in the upper half-plain, let's suppose that two of the connections are crossing in a node. Then, there exist only one way of replacing this crossing matching with a non-crossing one without splitting the original cycle into two 2-factors; moreover, this new configuration is less expensive than the original one.*

Figure 2.5: Crossing matching

Proof. Let's consider a generic oriented Hamiltonian cycle and let's suppose it contains a matching as in Fig.2.5:

There are two possible orientations for the matching that correspond to this two oriented Hamiltonian cycles:
1) $(C_1 r_1 r_3 C_2 r_2 r_4 C_3)$,
2) $(C_1 r_1 r_3 C_2 r_4 r_2 C_3)$
(the other possibilities are the dual of this two, and thus they are equivalent).

In both cases, a priori, there are two choices to replace this crossing matching (r_1, r_3), (r_2, r_4) with a non-crossing one: (r_1, r_2), (r_3, r_4) or (r_1, r_4), (r_2, r_3). We now show, for the two possible prototypes of Hamiltonian cycles which is the right choice for the non-crossing matching, giving a general rule. Let's consider case 1): here, if we replace the crossing matching with (r_1, r_4), (r_2, r_3), the cycle will split; in fact we would have two cycles: $(C_1 r_1 r_4 C_3)$ and $(r_3 C_2 r_2)$. Instead, if we use the other non-crossing matching, we would have: $(C_1 r_1 r_2 [C_2]^{-1} r_3 r_4 C_3)$. This way we have removed the node without splitting the cycle. Let's consider now case 2): in this situation, using (r_1, r_4), (r_2, r_3) as the new matching, we would have: $(C_1 r_1 r_4 [C_2]^{-1} r_3 r_2 C_3)$; the other matching, on the contrary, gives: $(C_1 r_1 r_2 C_3)$ and $(r_3 C_2 r_4)$.
The general rule is the following: given the oriented matching, consider the four oriented lines going inward and outward the node. Then, the right choice for the non-crossing matching is obtained joining the two couples of lines with opposite orientation.

Since the difference between the cost of the original graph and the new one simply consists in the difference between a crossing matching and a non-crossing on, this is positive when $0 < p < 1$, as shown in [2].

□

Now we can prove Proposition 5.1:

Proof. Consider a generic Hamiltonian cycle and draw the connections between the points in the upper half-plain. Suppose to have a Hamiltonian cycle where there are, let's say, n intersections between arcs; thanks to *Lemma 1* and the result obtained in Appendix A, we can remove a crossing matching with a non-crossing one without splitting the Hamiltonian cycle and lowering the number of crossings between the arcs. Moreover, the cost

of this cycle is smaller than the cost of the starting one. Iterating this procedure, it follows that one can find a cycle with no crossings. Evidently, if the non-crossing cycle is unique, its cost is the lower possible.
Now we prove that there are no other cycles out of h^* and its dual with no crossings. This can be easily seen, since h^* (and its dual) is the only cycle that visits all the points, starting from the first, in order. This means that all the other cycles do not visit the points in order and, thus, they have a node, due to the fact that the point that is not visited in a first time, must be visited next, creating a node. □

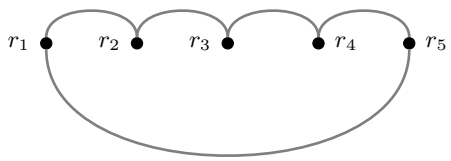


Figure 2.6: The optimal Hamiltonian cycle for $0 < p <> 1$ for $N = 5$

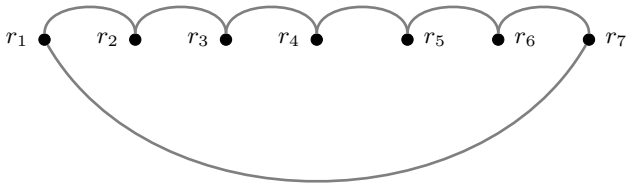


Figure 2.7: The optimal Hamiltonian cycle for $0 < p < 1$ for $N = 7$

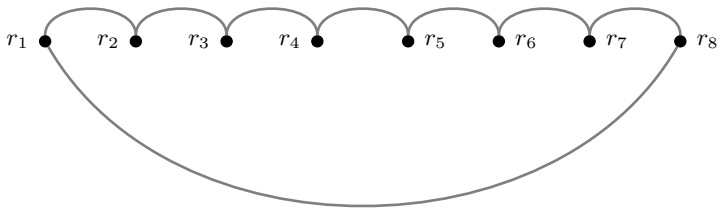


Figure 2.8: The optimal Hamiltonian cycle for $0 < p < 1$ for $N = 8$

2.6 The $p < 0$ case

In order to study the optimal cycle for $p < 0$, we first analyse the optimal matching, following the procedure used in [2].

Optimal matching This analysis goes through a case study, that reduces, thanks to the symmetries of the problem, to only three cases. We will denote them using this pictorial representation: [$\circ\circ\bullet\bullet$], [$\circ\bullet\circ\bullet$] and [$\circ\bullet\bullet\circ$], where the edges are supposed to link points of different colours, i.e., in [$\circ\circ\bullet\bullet$] the first point can be connected with the third one and the second with the fourth, but also the first with the fourth and the second with the third. For simplicity, let us say that $\bullet$ is a red point and $\circ$ is a blue point. In all the cases we will set the first point in the origin, the distance between the first and second point equal to 1, the distance between the second and third point equal to x_1 and the distance between the second and fourth point equal to x_2.

[$\circ\circ\bullet\bullet$] case Let T_1 be the cost of the matching in which the leftmost blue point goes with the leftmost red one, i.e. the ordered matching, and T_2 the cost of the other possible matching. Since $p < 0$, let us define $q := |p| = -p$, so that $q > 0$. Thus, the cost for the two possible matchings in this configuration is given by:

$$T_1 = \frac{1}{(1+x_1)^q} + \frac{1}{x_2^q} \tag{2.6.1}$$

$$T_2 = \frac{1}{(1+x_2)^q} + \frac{1}{x_1^q} \tag{2.6.2}$$

Comparing the costs, is easy to see that $T_2 \geq T_1$; in fact one has:

$$\frac{1}{(1+x_1)^q} + \frac{1}{x_2^q} \leq \frac{1}{(1+x_2)^q} + \frac{1}{x_1^q} \implies \frac{1}{x_2^q} - \frac{1}{x_1^q} \leq \frac{1}{(1+x_2)^q} - \frac{1}{(1+x_1)^q} \tag{2.6.3}$$

which is always true since $x_2 \geq x_1$ and $1/x^q$ is a monotone function whose second derivative is positive for every $q > 0$.

[$\circ\bullet\circ\bullet$] case In this situation, the cost for the two possible matchings is given by:

$$T_1 = 1 + \frac{1}{(x_2 - x_1)^q} \tag{2.6.4}$$

$$T_2 = \frac{1}{x_1^q} + \frac{1}{(1+x_2)^q} \tag{2.6.5}$$

This situation, in which both matchings are non-crossing, one finds that there is not a unique solution, but it depends on the distribution of points; explicitly, one should compare:

$$1 - \frac{1}{(1+x_2)^q} \tag{2.6.6}$$

and

$$\frac{1}{x_1^q} - \frac{1}{(x_2-x_1)^q} \tag{2.6.7}$$

and it is immediate to see that any relation of order between the two expressions is true only in a range of values of x_2.

[$\circ\bullet\bullet\circ$] case In this configuration, the cost of the two different matchings is:

$$T_1 = 1 + \frac{1}{(x_2-x_1)^q} \tag{2.6.8}$$

$$T_2 = \frac{1}{(1+x_1)^q} + \frac{1}{x_2^q} \tag{2.6.9}$$

In this case $T_1 \geq T_2$, in fact:

$$\begin{aligned} 1 + \frac{1}{(x_2-x_1)^q} \geq \frac{1}{(1+x_1)^q} + \frac{1}{x_2^q} \implies \\ \implies 1 - \frac{1}{(1+x_1)^q} \geq \frac{1}{x_2^q} - \frac{1}{(x_2-x_1)^q} \end{aligned} \tag{2.6.10}$$

which is always true since the left-hand side is always positive, while the right-hand side is negative.

Conclusion For $p < 0$ the optimal matching is *crossing*

Suppose to have a generic oriented Hamiltonian cycle and draw the connections between the vertices in the upper half plain. Suppose it is possible to identify a matching that is non-crossing, then the possible situations are the following:

For the first case there are two possible independent orientations:
1) $(r_1 r_4 C_2 r_2 r_3 C_3)$
2) $(r_1 r_4 C_2 r_3 r_2 C_3)$
It is obvious to see that the matching in the first cycle, if replaced by a crossing matching, would be: $(r_1 r_3 C_3)(r_2 [C_2]^{-1} r_4)$, and this is no more a Hamiltonian cycle. On the other hand, in the second cycle, the non-crossing matching can be replaced by a crossing one without breaking the cycle: $(r_1 r_3 [C_2]^{-1} r_4 r_2 C_3)$.

For the second case the possible orientations are:
1) $(r_1 r_2 C_2 r_4 r_3 C_3)$
2) $(r_1 r_2 C_2 r_3 r_4 C_3)$
By means of the same procedure used in the first case, one finds that the non-crossing matching in the second cycle can be replaced by a crossing one without splitting the cycle, while in the first case the cycle is divided by this operation.

We have proven that is not always possible to replace a non-crossing matching by a crossing one keeping unaltered the property of Hamiltonian cycle, but, in any case, this move preserves the graph property of being a 2-factor. This move is such that the cost of the new graph is lower than the cost of the old one: in fact, the cost difference between the two configurations coincides with the cost difference between the two matchings, and we have proved above that the crossing matching has a lower cost.

2.6.1 N odd case

Given an ordered sequence $\mathcal{R} = \{r_i\}_{i=1,\dots,N}$ of N points, with N odd, in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, if $p < 0$, consider the permutation σ defined as:

$$\sigma(i) = \begin{cases} 1 & \text{for} \quad i = 1 \\ \frac{2N-i+3}{2} & \text{for} \quad \text{odd i>1} \\ \frac{N-i+3}{2} & \text{for} \quad \text{even i>1} \end{cases} \tag{2.6.11}$$

This permutation defines the following Hamiltonian cycle:

$$h^* := h[\sigma] = (r_{\sigma(1)}, r_{\sigma(2)}, \dots, r_{\sigma(N)}). \tag{2.6.12}$$

Theorem 1. *The Hamiltonian cycle which provides the optimal cost is h^*. Moreover, h^* provides the optimal 2-factor solution.*

Proof. Let us consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, such that $|\mathcal{V}| = N$, with N odd, and a Hamiltonian cycle on it. Consider, in particular, the edge connecting two vertices $r_{\sigma(i)}$ and $r_{\sigma(i+1)}$: obviously both the edges $(r_{\sigma(i-1)}, r_{\sigma(i)})$ and $(r_{\sigma(i+1)}, r_{\sigma(i+2)})$ share a common vertex with $(r_{\sigma(i)}, r_{\sigma(i+1)})$, therefore they cannot, by definition, be part of a matching with $(r_{\sigma(i)}, r_{\sigma(i+1)})$. So, if we have N vertices, each edge has $N-3$ other edges that can be part of the same matching with it (obviously we don't have to consider the edge itself and the other two that share a vertex with him). Let us denote with $\mathcal{N}(\sigma(i))$ the number of edges that cross the edge $(r_{\sigma(i)}, r_{\sigma(i+1)})$ and let us define the sets:

$$A_j := \begin{cases} \{r_k\}_{k=\sigma(i)+1 \pmod N,\dots,\sigma(i+1)-1 \pmod N} & \text{for} \quad j=1 \\ \{r_k\}_{k=\sigma(i+1)+1 \pmod N,\dots,\sigma(i)-1 \pmod N} & \text{for} \quad j=2 \end{cases} \tag{2.6.13}$$

This two sets contain the points between $r_{\sigma(i)}$ and $r_{\sigma(i+1)}$.
In particular, the maximum number of crossings an edge can have is given by:

$$\max(\mathcal{N}(\sigma(i))) = \begin{cases} 2\min\{|A_j|, j=1,2\} & \text{for} \quad |A_1| \neq |A_2| \\ 2\min\{|A_j|, j=1,2\} - 1 & \text{for} \quad |A_1| = |A_2| \end{cases} \tag{2.6.14}$$

This is easily seen, since the maximum number of crossings an edge can have is obtained when all the points belonging to the smaller between A_1 and A_2 contributes with two crossings. This cannot happen when the cardinality of A_1 and A_2 is the same. Consider the particular case such that $\sigma(i) = a$ and $\sigma(i+1) = a + \frac{N-1}{2} \pmod N$ or $\sigma(i+1) = a + \frac{N+1}{2} \pmod N$. Then (2.6.14) in this cases is exactly equal to $N-3$, which means that the edges $(r_a, r_{a+\frac{N-1}{2} \pmod N})$ and $(r_a, r_{a+\frac{N+1}{2} \pmod N})$ can have the maximum number of crossings if the right configuration is chosen.
Clearly, if there is a 2-factor such that every edge has $N-3$ crossings, such a 2-factor is unique, because the only way of obtaining it is connecting the vertex r_a with $r_{a+\frac{N-1}{2} \pmod N}$ and $r_{a+\frac{N+1}{2} \pmod N}, \forall a$.

The Hamiltonian cycle h^* has exactly $N(N-3)/2$ crossings: in fact, the vertex a is connected with the vertices $a + \frac{N-1}{2} \mod N$ and $a + \frac{N+1}{2} \mod N$. The edge $(a, a + \frac{N-1}{2} \mod N)$ has $2\frac{N-3}{2} = N-3$ crossings due to the $\frac{N-3}{2}$ vertices $a+1 \mod N, a+2 \mod N, \dots, a + \frac{N-1}{2} - 1 \mod N$ that contribute with 2 edges each. This holds also for the edge $(a, a + \frac{N+1}{2} \mod N)$ and for each $a \in [1, N]$. Counting the total number of crossings and noting that each one of them is counted twice, we obtain the result.

Therefore, h^* is the only 2-factor, and, in particular, the only Hamiltonian cycle, such that the total number of crossings is saturated. This means that any other 2-factor has, at least, one crossing less, and so it cannot

be optimal, because the same 2-factor with the non-crossing edges replaced with two crossing ones has a lower cost. This procedure cannot be applied only to h^*, and so it is the optimal solution.

□

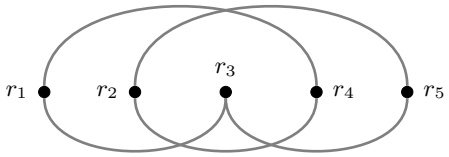


Figure 2.9: The optimal Hamiltonian cycle for $p < 0$ for $N = 5$

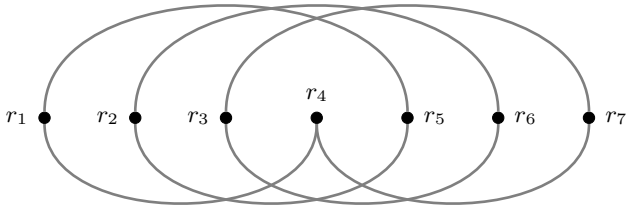


Figure 2.10: The optimal Hamiltonian cycle for $p < 0$ for $N = 7$

2.6.2 N even case

In this situation, differently from the above case, the solution in not unique. We will show how these solutions can be found and how they are related. In this section we will use the results obtained in Appendix B regarding the Monopartite Euclidean 2-factor for $p < 0$
Given the usual sequence of points $\mathcal{R} = \{r_i\}_{i=1,\dots,N}$ of N points, with N even, in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, if $p < 0$, consider the permutation σ such that:

$$\sigma(i) = \begin{cases} 1 & \text{for } i = 1 \\ \frac{N}{2} - i + 3 & \text{for even } i \leq \frac{N}{2} + 1 \\ N - i + 3 & \text{for odd } i \leq \frac{N}{2} + 1 \\ i - \frac{N}{2} & \text{for even } i > \frac{N}{2} + 1 \\ i & \text{for odd } i > \frac{N}{2} + 1 \end{cases} \tag{2.6.15}$$

This permutation generates the following connections between the vertices:

$$\begin{cases} 1 \longleftrightarrow \frac{N}{2}+1, \frac{N}{2}+2 \\ \frac{N}{2} \longleftrightarrow N, N-1 \\ i \longleftrightarrow \frac{N}{2}+i-1, \frac{N}{2}+i+1 \quad \text{for} \quad i \neq 1, \frac{N}{2}, \frac{N}{2}+1, N \\ \sigma(1)=1 \end{cases} \tag{2.6.16}$$

Given $\tau \in \mathcal{S}_N$ defined by $\tau(i) = i+1$ for $i \in [1, N-1]$ and $\tau(N) = 1$, we call Σ the set of permutations $\sigma_k, k = 1, ..., N$ defined as:

$$\sigma_k(i) = \tau^k(\sigma(i)) \tag{2.6.17}$$

where $\tau^k = \tau \circ \tau^{k-1}$. Thus we have the following result:

Proposition 2.6.1. *The set of Hamiltonian cycles that provides the optimal cost is*

$$h_k^* := h[\sigma_k] = (r_{\sigma_k(1)}, r_{\sigma_k(2)}, \ldots, r_{\sigma_k(N)}). \tag{2.6.18}$$

In Appendix B the optimal solution for the Euclidean 2-factor in obtained. In particular, we show how the solution is composed of a loop-covering of the graph. The idea for the proof of the TSP is to show how to join the loops in the optimal way in order to obtain the optimal TSP.

Proof. Let us begin, then, from the permutations that define the optimal solutions for the 2-factor:

- $4|N$:

$$\begin{aligned} \sigma_1 =& (r_1, r_{\frac{N}{2}+1}, r_2, r_{\frac{N}{2}+2}) \ldots \\ & \ldots (r_a, r_{a+\frac{N}{2}}, r_{a+1}, r_{a+\frac{N}{2}+1}) \ldots (r_{\frac{N}{2}-1}, r_{N-1}, r_{\frac{N}{2}}, r_N) \end{aligned} \tag{2.6.19}$$

$$\begin{aligned} \sigma_2 =& (r_1, r_{\frac{N}{2}+1}, r_N, r_{\frac{N}{2}}) \ldots \\ & \ldots (r_a, r_{a+\frac{N}{2}}, r_{a-1}, r_{a+\frac{N}{2}-1}) \ldots (r_{\frac{N}{2}-1}, r_{N-1}, r_{\frac{N}{2}-2}, r_{N-2}) \end{aligned} \tag{2.6.20}$$

- $4 \not| N$:

$$\pi_k(i) := \pi(i) + k \pmod N, k \in [0, N-1] \tag{2.6.21}$$

where

$$\begin{aligned} \pi =& (r_1, r_{\frac{N}{2}}, r_N, r_{\frac{N}{2}+1}, r_2, r_{\frac{N}{2}+2})(r_3, r_{\frac{N}{2}+3}, r_4, r_{\frac{N}{2}+4}) \ldots \\ & \ldots (r_{\frac{N}{2}-2}, r_{N-1}, r_{\frac{N}{2}-1}, r_{N-2}) \end{aligned} \tag{2.6.22}$$

In both cases, the optimal solution is formed only by edges of length $\frac{N}{2}-1$ and of length $\frac{N}{2}-2$. Since the optimal 2-factor is not a TSP, in order to obtain a Hamiltonian cycle from the 2-factor solution, couples of crossing edges need to became non-crossing, where one of the two edges belongs to one loop of the covering and the other to another loop. Now we show that the optimal way of joining the loops is replacing two edges of length $\frac{N}{2}-1$ with other two of length $\frac{N}{2}-2$.
Let us consider two adjacent 4-vertices loops, i.e. two loops of the form:

$$(r_a, r_{a+\frac{N}{2}}, r_{a+1}, r_{a+\frac{N}{2}+1}), (r_{a+2}, r_{a+2+\frac{N}{2}}, r_{a+3}, r_{a+\frac{N}{2}+3}) \tag{2.6.23}$$

and let us analyse the possible cases:

1. to remove two edges of length $\frac{N}{2}-2$, that can be replaced in two way:
   - either with an edge of length $\frac{N}{2}-2$ and one of length $\frac{N}{2}-4$; in this case the maximum number of crossings decreases by 4;
   - or with two edges of length $\frac{N}{2}-3$; also in this situation the maximum number of crossings decreases by 4.
2. to remove one edge of length $\frac{N}{2}-2$ and one of length $\frac{N}{2}-1$, and also this operation can be done in two ways:
   - either with an edge of length $\frac{N}{2}-2$ and one of length $\frac{N}{2}-3$; in this case the maximum number of crossings decreases by 3;
   - or with an edge of length $\frac{N}{2}-3$ and one of length $\frac{N}{2}-4$; in this situation the maximum number of crossings decreases by 7.
3. the last chance is to remove two edges of length $\frac{N}{2}-1$, and also this can be done in two ways:
   - either with two edges of length $\frac{N}{2}-3$; here the maximum number of crossings decreases by 6;
   - or with two edges of length $\frac{N}{2}-2$; in this situation the maximum number of crossings decreases by 2. This happens when we substitute two adjacent edges of length $\frac{N}{2}-1$, that is, edges of the form $(r_a, r_{\frac{N}{2}+a})$ and $(r_{a+1}, r_{\frac{N}{2}+a+1})$, with the non-crossing edges $(r_a, r_{\frac{N}{2}+a+1})$ and $(r_{a+1}, r_{\frac{N}{2}+a})$

It is then obvious that the last possibility is the optimal one, since our purpose is to find the maximum-crossing TSP, in order to conclude it has the lower cost. It is also obvious that the cases discussed above holds also for the 6-vertices loop and an adjacent 4-vertices loop of the $4 \not| N$ case. Moreover, we have considered here adjacent loops because, if the were not adjacent, clearly, the difference in maximum crossings would have been even

bigger.
Now we have a constructive pattern for building the optimal TSP. Let us call $\mathcal{O}$ the operation described in the second point of (3). Then, starting from the optimal 2-factor solution, if it is formed by n points, $\mathcal{O}$ has to be applied $\frac{n}{4} - 1$ times if $4|n$ and $\frac{n-6}{4}$ times if $4 \not| n$. In both cases it is easily seen that $\mathcal{O}$ always leaves two adjacent edges of length $\frac{N}{2} - 1$ invariant, while all the others have length $\frac{N}{2} - 2$. The multiplicity of solutions is given by the $\frac{N}{2}$ ways one can choose the two adjacent edges of length $\frac{N}{2} - 1$. In particular, the Hamiltonian cycles h_k^* saturates the maximum number of crossings that can be done, i.e., every time that $\mathcal{O}$ is applied, exactly 2 crossings are lost. We have proved, then, that h_k^* are the Hamiltonian cycles with the maximum number of crossings and, for fixed N, any other Hamiltonian cycle has a lower number of crossings. This means that any other Hamiltonian cycle must have

- either every edge of length $\frac{N}{2} - 2$;
- or at least one edge of length less than or equal to $\frac{N}{2} - 3$.

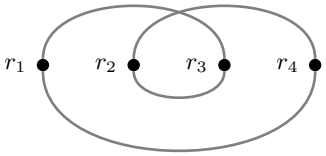
This is easily seen, since it is not possible to build a Hamiltonian cycle with more than two edges or only one edge of length $\frac{N}{2} - 1$ and all the others of length $\frac{N}{2} - 2$. It is also impossible to build a Hamiltonian cycle with two non-adjacent edges of length $\frac{N}{2} - 1$ and all the others of length $\frac{N}{2} - 2$: the proof is immediate.
Consider then the two cases presented above: in the first case the cycle (let us call it H) is clearly not optimal, since it differs from $h_k^*, \forall k$ by a matching that is crossing in h_k^* and non-crossing in H. Let us consider, then, the second case and suppose the shortest edge, let us call it b, has length $\frac{N}{2} - 3$: the following reasoning equally holds if the considered edge is shorter. The shortest edge creates two subsets of vertices: in fact, called x and y the vertices of the edge considered and supposing $x < y$, there are the subsets defined by:

$$A = \{r \in \mathcal{V} : x < r < y\} \tag{2.6.24}$$

$$B = \{r \in \mathcal{V} : r < x \vee r > y\} \tag{2.6.25}$$

Suppose, for simplicity, that $|A| < |B|$: then, necessarily $|A| = \frac{N}{2} - 3$ and $|B| = \frac{N}{2} + 1$. As an immediate consequence, there is a vertex in B whose edges have both vertices in $|B|$. As a consequence, fixed an orientation on the cycle, one of this two edges and b are obviously non-crossing and, moreover, have the right relative orientation so that they can be replaced by two crossing edges without splitting the Hamiltonian cycle. Therefore also in this case the Hamiltonian cycle considered is not optimal. □

Since both in the N odd case and in the N even case the results obtained only depend on the properties of the matching for $p < 0$, and since the



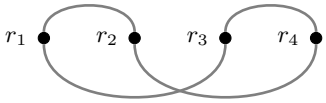
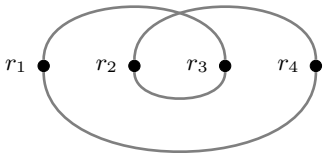
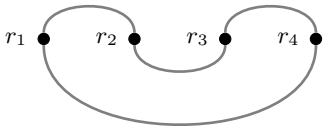


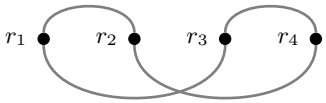
Figure 2.11: The optimal Hamiltonian cycles for $p < 0$ and $N = 4$



(a) $p < 0$



(b) $0 < p < 1$



(c) $p > 0$ and $p < 0$

Figure 2.12: Optimal Hamiltonian cycles for $N = 4$ and different p cases. The cycles represented are all the 3 possible $N = 4$ Hamiltonian cycles for $N = 4$.

weight w_e in (2.1.4) is a $C - function$ for $p < 0$, our results for this case are valid in general for a $C - function$ cost. It holds, then:

Proposition 2.6.2. *Given a cost function of the form:*

$$E(h) = \sum_{e \in h} w_e \tag{2.6.26}$$

where

$$w_e = f(|x_i - x_j|) \tag{2.6.27}$$

with f a $C - function$, the optimal cycle is given by (2.6.12) when N is odd and by (2.6.18) when N is even.

2.7 Evaluation of the average costs

If we have N random points chosen with flat distribution in the interval $[0, 1]$ and we order them in increasing position, the probability for finding the l-th

point in the interval $[x, x+dx]$ is given by the conditional probability:

$$p_l(x) = \binom{N}{l} x^l [1-x]^{N-l} \binom{l}{1} \frac{dx}{x} = \frac{\Gamma(N+1)}{\Gamma(l)\,\Gamma(N-l+1)}\, x^{l-1}(1-x)^{N-l}\, dx \tag{2.7.1}$$

where the therm

$$p_{l,1}(x \in dx) = \binom{N}{l} x^l [1-x]^{N-l} \tag{2.7.2}$$

represents the probability of having l points between 0 and x and the remaining ones between x and 1, while

$$p_{l,2}(x \in dx) = \binom{l}{1} \frac{dx}{x} \tag{2.7.3}$$

is the conditional probability that one of the l points between 0 and x is in the infinitesimal interval $[x, x+dx]$. Notice that the property of the Γ-function $\Gamma(n) = (n-1)!$ has been used.

The probability for finding the l-th point in the interval $[x, x+dx]$ and the s-th point in the interval $[y, y+dy]$ is given, for $s > l$ by

$$p_{l,s}(x,y) = \frac{\Gamma(N+1)}{\Gamma(l)\,\Gamma(s-l)\,\Gamma(N-s+1)}\, x^{l-1}(y-x)^{s-l-1}(1-y)^{N-s}\, \theta(y-x)\, dx\, dy \tag{2.7.4}$$

see for example [8, App. A]. The procedure for finding Eq.2.7.4 is analogous to the one used for Eq.2.7.1.

It follows that the mean distance between point r_l and point r_{l+k} is given by:

$$\overline{|r_{l+k} - r_l|^\alpha} = \int_0^1 dx\, dy\, (y-x)^\alpha\, p_{l,l+k}(x,y) = \frac{\Gamma(N+1)\,\Gamma(k+\alpha)}{\Gamma(N+\alpha+1)\,\Gamma(k)} \tag{2.7.5}$$

independently from l. Therefore, in the case $p > 1$ (see Sect.2.3), the optimal solution for N points has $N-2$ edges of the type (r_k, r_{k+2}) and two of the type (r_k, r_{k+1}). Consequently, using Eq.2.7.5, the expression for the average cost is:

$$\begin{aligned} \overline{E_N[h^*]} &= (N-2)\frac{\Gamma(N+1)\,\Gamma(p+2)}{\Gamma(N+p+1)\,\Gamma(2)} + 2\frac{\Gamma(N+1)\,\Gamma(p+1)}{\Gamma(N+p+1)\,\Gamma(1)} = \\ &= [(N-2)(p+1)+2]\, \frac{\Gamma(N+1)\,\Gamma(p+1)}{\Gamma(N+p+1)} \end{aligned} \tag{2.7.6}$$

and in particular for $p = 2$

$$\overline{E_N[h^*]} = \frac{2\,(3N-4)}{(N+1)(N+2)}\,, \tag{2.7.7}$$

and for $p = 1$ we get

$$\overline{E_N[h^*]} = \frac{2\,(N-1)}{N+1}\,, \tag{2.7.8}$$

In the case $0 < p < 1$, see Sect.2.5, the optimal solution has $N-1$ edges of the type (r_k, r_{k+1}) and the remaining one is (r_1, r_N); therefore the average cost is given by:

$$\begin{aligned}\overline{E_N[h^*]} &= (N-1)\frac{\Gamma(p+1)\Gamma(N+1)}{\Gamma(N+p+1)\Gamma(1)} + \frac{\Gamma(p+N-1)\Gamma(N+1)}{\Gamma(N+p+1)\Gamma(N-1)} = \\ &= \left[(N-1)\,\Gamma(p+1) + \frac{\Gamma(N+p-1)}{\Gamma(N-1)}\right]\frac{\Gamma(N+1)}{\Gamma(N+p+1)}\end{aligned} \tag{2.7.9}$$

which coincides at $p = 1$ with (2.7.8) and, at $p = 0$, provides $\overline{E_N[h^*]} = N$. For large N, we get

$$\lim_{N\to\infty} N^{p-1}\overline{E_N[h^*]} = \begin{cases}\Gamma(p+2) & \text{for } p \geq 1 \\ \Gamma(p+1) & \text{for } 0 < p < 1\,.\end{cases} \tag{2.7.10}$$

The asymptotic cost for large N and $p > 1$ is $2(p+1)$ times the average optimal cost of the complete matching problem [8]. For $p < 0$ and N odd the average optimal cost is

$$\overline{E_N[h^*]} = \left[(N-1)\frac{\Gamma\left(\frac{N+1}{2}+p\right)}{\Gamma\left(\frac{N+1}{2}\right)} + (N+1)\frac{\Gamma\left(\frac{N-1}{2}+p\right)}{\Gamma\left(\frac{N-1}{2}\right)}\right]\frac{\Gamma(N+1)}{2\Gamma(N+p+1)} \tag{2.7.11}$$

and for large N it behaves as

$$\lim_{N\to\infty}\frac{\overline{E_N[h^*]}}{N} = \frac{1}{2^p}\,, \tag{2.7.12}$$

which coincides with the scaling derived before for $p = 0$. Note that for large N the cost of the TSP problem is the same of the matching problem for $p < 0$.

For N even, instead, there are $N/2$ possible solutions. One can see $N/2-1$ of these share the same average energy, since they have the same number of links with the same k of equation (2.7.5). These solutions, in particular have 2 links with $k = N/2$, $N/2$ links with $k = N/2+1$ and $N/2-2$ links with $k = N/2+1$. We will call this configuration with h_1 and its average cost will be

$$\begin{aligned}\overline{E_N[h_1]} =& \frac{\Gamma(N+1)}{\Gamma(N+p+1)}\left[\frac{N}{2}\frac{\Gamma\left(\frac{N}{2}+p-1\right)}{\Gamma\left(\frac{N}{2}-1\right)} + \right. \\ &\left. + \left(\frac{N}{2}-2\right)\frac{\Gamma\left(\frac{N}{2}+p+1\right)}{\Gamma\left(\frac{N}{2}+1\right)} + 2\frac{\Gamma\left(\frac{N}{2}+p\right)}{\Gamma\left(\frac{N}{2}\right)}\right]\end{aligned} \tag{2.7.13}$$

The last possible solution, that we will call with h_2 has 2 links with $k = N/2-1$, $N/2$ links with $k = N/2+1$ and $N/2-1$ links with $k = N/2+1$

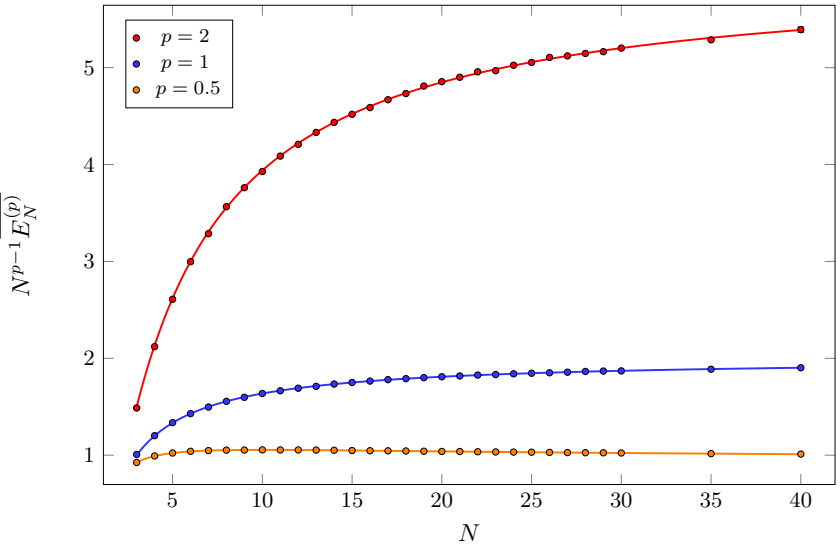


Figure 2.13: Rescaled average optimal cost for various values of $p > 0$. The points are the result of a numerical simulation whereas lines are theoretical predictions.

and its average cost will be

$$\begin{aligned}\overline{E_N[h_2]} =& \frac{\Gamma(N+1)}{\Gamma(N+p+1)} \left[\left(\frac{N}{2} - 1 \right) \frac{\Gamma\left(\frac{N}{2} + p - 1\right)}{\Gamma\left(\frac{N}{2} - 1\right)} + \right. \\ &\left. + \left(\frac{N}{2} - 1 \right) \frac{\Gamma\left(\frac{N}{2} + p + 1\right)}{\Gamma\left(\frac{N}{2} + 1\right)} + 2 \frac{\Gamma\left(\frac{N}{2} + p\right)}{\Gamma\left(\frac{N}{2}\right)} \right]\end{aligned} \tag{2.7.14}$$

2.7.1 General distribution of points

In this section we shall consider a more general distribution of points.

Let choose the points in the interval $[0, 1]$ according to the distribution ρ, which has no zero in the interval, and let

$$\Phi_\rho(x) = \int_0^x dt\, \rho(t) \tag{2.7.15}$$

its *cumulative*, which is an increasing function with $\Phi_\rho(0) = 0$ and $\Phi_\rho(1) = 1$.

In this case, the probability for finding the l-th point in the interval $[x, x + dx]$ and the s-th point in the interval $[y, y + dy]$ is given, for $s > l$ by

$$\begin{aligned}p_{l,s}(x, y)\, d\Phi_\rho(x)\, d\Phi_\rho(y) = & \frac{\Gamma(N+1)}{\Gamma(l)\,\Gamma(s-l)\,\Gamma(N-s+1)} \\ & \Phi_\rho^{l-1}(x)\, [\Phi_\rho(y) - \Phi_\rho(x)]^{s-l-1}\, [1 - \Phi_\rho(y)]^{N-s}\, \theta(y - x)\, d\Phi_\rho(x)\, d\Phi_\rho(y)\end{aligned} \tag{2.7.16}$$

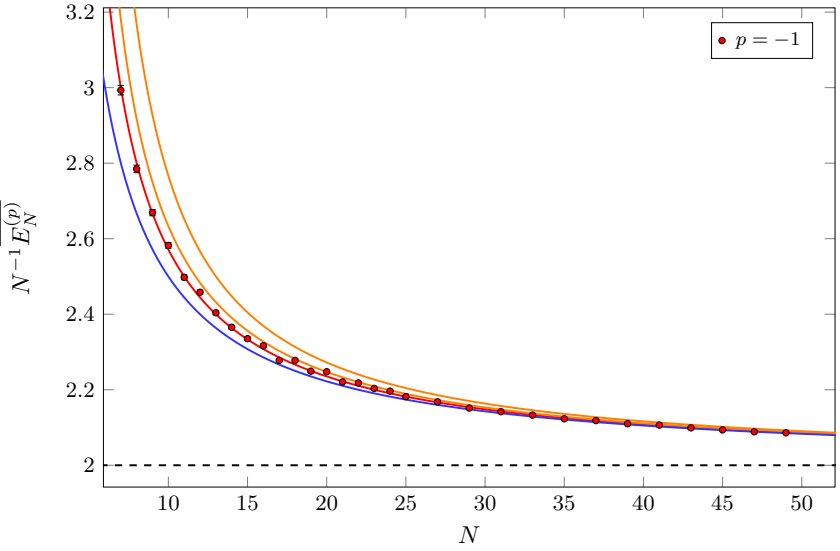


Figure 2.14: Rescaled average optimal cost in the $p = -1$ case. The red points and line are respectively the result of a numerical simulation and the theoretical prediction in the odd N case. The blue line is the 2 times the theoretic value of the optimal matching. The orange lines from the top the average costs $\overline{E_N[h_1]}$ and $\overline{E_N[h_2]}$ defined in equation (2.7.13) and (2.7.14) respectively. The dashed black line is the large N limit of all the curves.

We have that, in the case $p > 1$

$$\overline{E_N[h^*]} = \\ = \int (y-x)^p \left[p_{1,2}(x,y) + p_{N-1,N}(x,y) + \sum_{l=1}^{N-2} p_{l,l+2}(x,y) \right] d\Phi_\rho(x)\, d\Phi_\rho(y) \tag{2.7.17}$$

and

$$\sum_{l=1}^{N-2} p_{l,l+2}(x,y)\, d\Phi_\rho(x)\, d\Phi_\rho(y) = \frac{\Gamma(N+1)}{\Gamma(N-2)} \left[1 - \Phi_\rho(y) + \Phi_\rho(x)\right]^{N-3} \\ \left[\Phi_\rho(y) - \Phi_\rho(x)\right] \theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \tag{2.7.18}$$

while

$$\left[p_{1,2}(x,y) + p_{N-1,N}(x,y)\right] d\Phi_\rho(x)\, d\Phi_\rho(y) = \\ = \frac{\Gamma(N+1)}{\Gamma(N-1)} \left\{ \left[1 - \Phi_\rho(y)\right]^{N-2} + \Phi_\rho^{N-2}(x) \right\} \theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \tag{2.7.19}$$

For large N we can make the approximation

$$\overline{E_N[h^*]} \approx N^3 \int (y-x)^p \left[1 - \Phi_\rho(y) + \Phi_\rho(x)\right]^N \\ \left[\Phi_\rho(y) - \Phi_\rho(x)\right] \theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \tag{2.7.20}$$

and we remark that the maximum of the contribution to the integral comes from the region where $\Phi_\rho(y) \approx \Phi_\rho(x)$ and we make the change of variables

$$\Phi_\rho(y) = \Phi_\rho(x) + \frac{\epsilon}{N} \tag{2.7.21}$$

so that

$$y = \Phi_\rho^{-1}\left[\Phi_\rho(x) + \frac{\epsilon}{N}\right] \approx x + \frac{\epsilon}{N\rho(x)} \tag{2.7.22}$$

and we get

$$\overline{E_N[h^*]} \approx N^3 \int d\Phi_\rho(x) \int_0^\infty \frac{d\epsilon}{N} \left[\frac{\epsilon}{N\rho(x)}\right]^p \frac{\epsilon}{N}\, e^{-\epsilon} = \tag{2.7.23}$$

$$= \frac{\Gamma(p+2)}{N^{p-1}} \int dx\, \rho^{1-p}(x)\,. \tag{2.7.24}$$

When $0 < p < 1$

$$\overline{E_N[h^*]} = \int (y-x)^p \left[p_{1,N}(x,y) + \sum_{l=1}^{N-1} p_{l,l+1}(x,y) \right] d\Phi_\rho(x)\, d\Phi_\rho(y) \tag{2.7.25}$$

and

$$\sum_{l=1}^{N-1} p_{l,l+1}(x,y)\, d\Phi_\rho(x)\, d\Phi_\rho(y) = \\ = N\,(N-1)\,\left[1-\Phi_\rho(y)+\Phi_\rho(x)\right]^{N-2}\,\theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \quad (2.7.26)$$

while

$$p_{1,N}(x,y)\, d\Phi_\rho(x)\, d\Phi_\rho(y) = \\ = N\,(N-1)\,\left[\Phi_\rho(y)-\Phi_\rho(x)\right]^{N-2}\,\theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y)\,. \quad (2.7.27)$$

For large N we can make the approximation

$$\overline{E_N[h^*]} \approx N^2 \int (y-x)^p \left[1-\Phi_\rho(y)+\Phi_\rho(x)\right]^N\,\theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \approx \quad (2.7.28)$$

$$\approx N^2 \int d\Phi_\rho(x) \int_0^\infty \frac{d\epsilon}{N}\left[\frac{\epsilon}{N\rho(x)}\right]^p e^{-\epsilon} = \quad (2.7.29)$$

$$= \frac{\Gamma(p+1)}{N^{p-1}} \int dx\, \rho^{1-p}(x)\,. \quad (2.7.30)$$

Indeed the other term, for large N, gives a contribution

$$N^2 \int (y-x)^p \left[\Phi_\rho(y)-\Phi_\rho(x)\right]^N\,\theta(y-x)\, d\Phi_\rho(x)\, d\Phi_\rho(y) \quad (2.7.31)$$

so that, we will set

$$\Phi_\rho(y) = 1-\frac{\epsilon}{N}\,, \qquad \Phi_\rho(x) = \frac{\delta}{N}\,, \qquad y-x \approx 1 \quad (2.7.32)$$

and therefore we get a contribution

$$\int_0^\infty d\epsilon\, e^{-\epsilon} \int_0^\infty d\delta\, e^{-\delta} = 1 \quad (2.7.33)$$

which is of the same order of the other term only at $p=1$.

Chapter 3

Conclusions

3.1 Obtained results and perspectives

In this thesis work we have studied the Travelling Salesman Problem, in particular its one-dimensional version and, as cost function, we have considered a function of the Euclidean distance between the points (see Eq. 2.1.4). The main results we have obtained are the following: firstly, we derived the exact optimal tour for $p \geq 0$, described is Sect. 2.4 ($p > 1$) and 2.5 ($0 < p < 1$).

Secondly, also for $p < 0$ and odd N (see 2.6.1), we have found the exact optimal solution, showing that it is also the 2-factor optimal solution for $p < 0$. On the other hand, for $p < 0$ and even N (see 2.6.2), we have been able to isolate a subset, that we have characterized, of all the possible Hamiltonian cycles whose cardinality grows linearly with the number of points and we have proved that the optimal solution is always contained in this subset. Since, for $p < 0$, only the properties of the matching have been used to demonstrate our results and since all $C - functions$ have the same matching properties, knowing that the cost function with $p < 0$ is a $C-function$, then our results automatically hold for a generic $C-function$.

Thirdly, in App. B, we have described the solution of the 2-factor problem for $p < 0$ and even N. Also in this case, as for the TSP in the same conditions, we have able to isolate the solution into a subset, that we have characterized, of cardinality that grows linearly with the number of points.

To conclude, we have calculated the average cost of the TSP solutions, supposing, firstly, that the points are included in the unit interval and distributed with the flat probability distribution, calculating the average optimal cost in the large N limit for $p > 0$ and for $p < 0$ and odd N: in particular we have shown that, in the large N limit and for $p > 1$, the average optimal cost is $2(p + 1)$ times the average optimal cost of the matching problem, while for $p < 0$ and odd N the cost obtained coincides with the matching one. Then we computed the average cost for a generic distribution of points, and, obviously, the results explicitly depend on the probability distribution.

The work we have presented represents one of the remaining steps in one-dimensional combinatorial optimization problems: since now, in fact, only the matching problem has been analysed in the one-dimensional monopartite version. The importance of the solutions found in this work lies in the fact that, not only we give the scaling for quantities of interest, e.g. the average optimal cost, but also the exact optimal solutions for the cost functions we have considered.

The results obtained may be used to study combinatorial optimization problems in higher dimensions, specially two, e.g. in [3].

3.2 Acknowledgements

This thesis work is the last step of a three-year journey: certainly it was not easy, but, at the same time, it has been very challenging. First of all, I have to thank my family for the opportunity they gave me and for their support; in particular I thank my parents for their indispensable and never commonplace advices and my brothers, with whom I'm spending beautiful years.
I want to give particular thanks to my supervisor, professor Caracciolo: I had the good fortune to follow his course in Quantum Mechanics and I am honoured to have had the possibility of working with him.
I also want to thank my assistant supervisors, Andrea Di Gioacchino and Enrico M. Malatesta, for the fruitful discussions we had about the problem and for their patience in listening to my ideas.
A thank goes, as well, to my friends, inside and outside university, for the good time spent together and the opinions shared about an infinite variety of arguments.

I would also like to thank professor Giorgio Parisi and the organizers of the conference "Disordered Serendipity: a glassy path to discovery" held in Rome from September 19 to 22 and that I attended: it has been a precious chance for meeting leading experts from all around the world in physics of complex systems. I'm grateful to my supervisor, professor Caracciolo, for having mentioned the results obtained in this thesis work in the talk he gave during the conference.

Appendix A

A result about crossing matchings

Given a Hamiltonian cycle, in general is not obvious that replacing a non-crossing matching with a crossing one, the total number of intersections increases: there could be the chance that one or more nodes are removed in the operation of substituting the matching we are interested in. However, we now show that holds the following

Proposition A.0.1. *Given a generic graph with a matching that is non-crossing, if it is replaced by a crossing one, the total number of intersections always increases. Vice versa, if a crossing matching is replaced by a non-crossing one, the total number of crossings always decreases.*

Proof. This is a topological property we will demonstrate for cases, using the representation on the circle in (A.1).

□

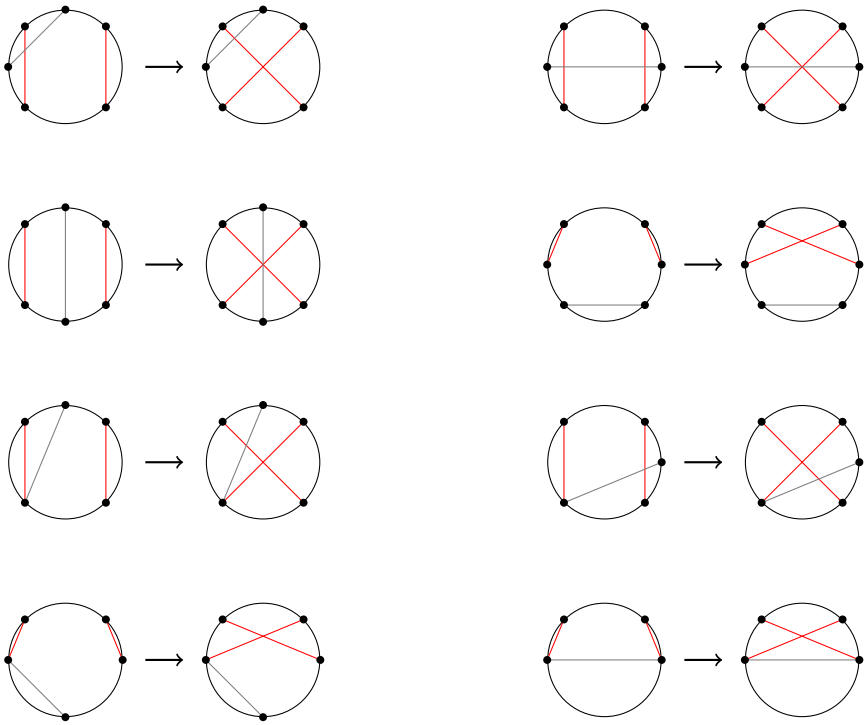


Figure A.1: The proof of (A): from left to right the non-crossing red edges are replaced with two crossing edges: the number of crossings always increases. The opposite result is obtained simply inverting the arrows.

Appendix B

The 2-factor solution for $p < 0$ and even N

We study here the 2-factor solution for $p < 0$. The odd-N case has already been solved with the TSP solution since, in that case, the optimal 2-factor is the optimal TSP. We focus here to the even-N case.

Notation: given $\sigma \in \mathcal{S}_N$, we say the edge $(r_{\sigma(i)}, r_{\sigma(i+1)})$ has length $L \in \mathbb{N}$ if:

$$L = \mathcal{L}(i) := \min\{|A_j(i)|, j = 1, 2\} \tag{B.0.1}$$

where

$$A_j(i) := \begin{cases} \{r_k\}_{k=\sigma(i)+1 \pmod N,\ldots,\sigma(i+1)-1 \pmod N} & \text{for} \quad j = 1 \\ \{r_k\}_{k=\sigma(i+1)+1 \pmod N,\ldots,\sigma(i)-1 \pmod N} & \text{for} \quad j = 2 \end{cases} \tag{B.0.2}$$

The optimal solution for the 2-factor and even N is divided into two possible cases, that are analysed in the following sections: in particular, we will study the case in which $\frac{N}{2}$ is even, thus $N = 4n$, $n \in \mathbb{N}$ and the case in which $\frac{N}{2}$ is odd, thus $N = 4n + 2$, $n \in \mathbb{N}$.

B.1 $N/2$ is even

Let us consider the sequence of points $\mathcal{R} = \{r_i\}_{i=1,\ldots,N}$ of N points, with N a multiple of 4, in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, consider the permutations σ_j, $j = 1, 2$ defined by the following cyclic decomposition:

$$\begin{aligned} \sigma_1 =& (r_1, r_{\frac{N}{2}+1}, r_2, r_{\frac{N}{2}+2}) \ldots (r_a, r_{a+\frac{N}{2}}, r_{a+1}, r_{a+\frac{N}{2}+1}) \ldots \\ & \ldots (r_{\frac{N}{2}-1}, r_{N-1}, r_{\frac{N}{2}}, r_N) \end{aligned} \tag{B.1.1}$$

$$\begin{aligned} \sigma_2 =& (r_1, r_{\frac{N}{2}+1}, r_N, r_{\frac{N}{2}}) \ldots (r_a, r_{a+\frac{N}{2}}, r_{a-1}, r_{a+\frac{N}{2}-1}) \ldots \\ & \ldots (r_{\frac{N}{2}-1}, r_{N-1}, r_{\frac{N}{2}-2}, r_{N-2}) \end{aligned} \tag{B.1.2}$$

for integer $a \in [1, \frac{N}{2} - 1]$.

Defined $h_1^* := h[\sigma_1]$ and $h_2^* := h[\sigma_2]$, holds the following:

Proposition B.1.1. *h_1^* and h_2^* are the 2-factors that contain the maximum number of crossings between the arcs.*

Proof. An edge can be involved, at most, in $N - 3$ crossing matchings. In the even N case, this number is achieved by the edges of the form $(r_a, r_{a+\frac{N}{2} \pmod N})$, i.e. by the edges of length $\frac{N}{2} - 1$. Obviously, there can be at most $\frac{N}{2}$ edges of this form in a 2-factor. Thus, in order to maximise the number of crossings, the other $\frac{N}{2}$ edges must be of the form $(r_a, r_{a+\frac{N}{2}+1 \pmod N})$ or $(r_a, r_{a+\frac{N}{2}-1 \pmod N})$, i.e. of length $\frac{N}{2} - 2$. It is immediate to verify that both h_1^* and h_2^* have this property; we have to prove they are the only ones with this property.
Consider, then, to have already fixed the $\frac{N}{2}$ edges $(r_a, r_{a+\frac{N}{2} \pmod N}) \forall a \in [1, N]$. Suppose to have fixed also the edge $(r_1, r_{\frac{N}{2}})$ (the other chance is to fix the edge $(r_1, r_{\frac{N}{2}+2})$: this brings to the other 2-factor). Consider now the point $r_{\frac{N}{2}+1}$: suppose it is not connected to the point r_N, but to the point r_2, i.e., it has a different edge from the cycle h_2^*. We now show that this implies it is not possible to construct all the remaining edges of length $\frac{N}{2} - 2$. Consider, indeed, of having fixed the edges $(r_1, r_{\frac{N}{2}})$ and $(r_2, r_{\frac{N}{2}+1})$ and focus on the vertex $r_{\frac{N}{2}+2}$: in order to have an edge of length $\frac{N}{2} - 2$, this vertex must be connected either with r_1 or with r_3, but r_1 already has two edges, thus, necessarily, there must be the edge $(r_3, r_{\frac{N}{2}+2})$. By the same reasoning, there must be the edges $(r_4, r_{\frac{N}{2}+3})$, $(r_5, r_{\frac{N}{2}+4}), \ldots, (r_{\frac{N}{2}-1}, r_{N-2})$. Proceeding this way, we have constructed $N-1$ edges; the remaining one is uniquely determined, and it is (r_{N-1}, r_N), which has null length.
Therefore the edge $(r_2, r_{\frac{N}{2}+1})$ cannot be present in the optimal 2-factor and so, necessarily, there is the edge $(r_{\frac{N}{2}+1}, r_N)$; this creates the cycle $(r_1, r_{\frac{N}{2}}, r_N, r_{\frac{N}{2}+1})$. Proceeding the same way on the set of the remaining vertices $\{r_2, r_3, \ldots, r_{\frac{N}{2}-1},$
$r_{\frac{N}{2}+2}, \ldots, r_{N-1}\}$, one finds that the only way of obtaining $\frac{N}{2}$ edges of length $\frac{N}{2} - 1$ and $\frac{N}{2}$ edges of length $\frac{N}{2} - 2$ is generating the loop coverings of the graph h_1^* or h_2^*. □

Proposition B.1.2. *The 2-factor that provides the optimal cost is either h_1^* or h_2^*.*

Proof. We have already proven that $h_{1,2}^*$ are the 2-factors that have the maximum number of crossing matchings and there are no others out of them; then, automatically, the optimal solution is one of them, because any other 2-factor must have at least a non-crossing matching that can be replaced by a crossing one, lowering the cost. □

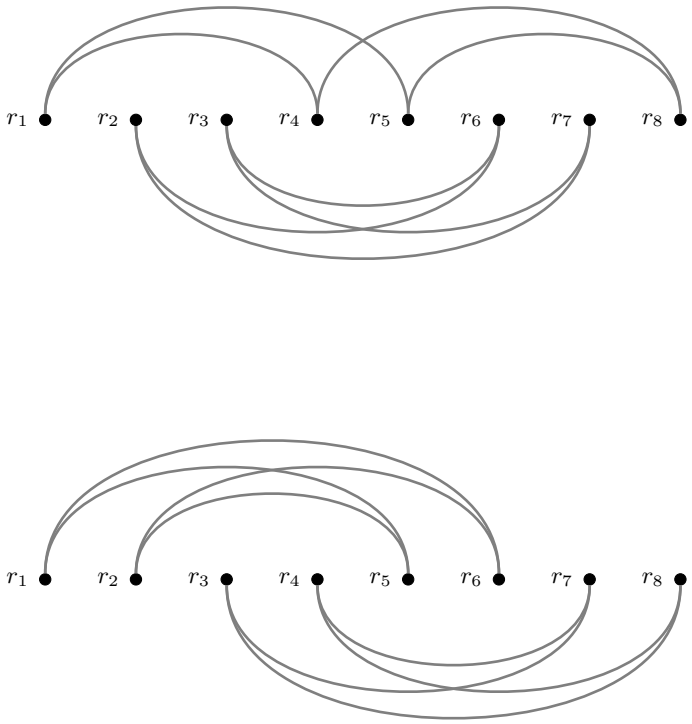


Figure B.1: The optimal 2-factor solutions for $N = 8$ and $p < 0$

B.2 $N/2$ is odd

Let us consider the usual sequence $\mathcal{R} = \{r_i\}_{i=1,\dots,N}$ of N points, with even N but not a multiple of 4, in the interval $[0, 1]$, with $r_1 \leq \cdots \leq r_N$, consider the permutation π defined by the following cyclic decomposition:

$$\begin{aligned}\pi =&(r_1, r_{\frac{N}{2}}, r_N, r_{\frac{N}{2}+1}, r_2, r_{\frac{N}{2}+2})(r_3, r_{\frac{N}{2}+3}, r_4, r_{\frac{N}{2}+4}) \dots \\ &\dots (r_{\frac{N}{2}-2}, r_{N-1}, r_{\frac{N}{2}-1}, r_{N-2})\end{aligned} \tag{B.2.1}$$

Defined

$$\pi_k(i) := \pi(i) + k \pmod N, k \in [0, N-1] \tag{B.2.2}$$

and

$$h_k^* := h[\pi_k] \tag{B.2.3}$$

the following proposition holds:

Proposition B.2.1. *h_k^* are the 2-factors that contain the maximum number of crossings between the arcs.*

Proof. Also in this case the observations done in the proof of proposition B.1 holds. Thus, in order to maximize the number of crossing matchings, one considers, as in the previous case, the $\frac{N}{2}$ edges of length $\frac{N}{2}-1$, i.e. of the form $(r_a, r_{a+\frac{N}{2} \pmod N})$, and then tries to construct the remaining $\frac{N}{2}$ edges of length $\frac{N}{2}-2$, likewise the $4|N$ case. Again, if one fixes the edge $(r_1, r_{\frac{N}{2}})$, the edge $(r_2, r_{\frac{N}{2}+1})$ cannot be present, by the same reasoning done in the proof of Proposition B.1. The fact that, in this case, N is not a multiple of 4 makes it impossible to have a 2-factor formed by 4-vertices loops, as in the previous case. The first consequence is that, given $\frac{N}{2}$ edges of length $\frac{N}{2}-1$, it is not possible to have $\frac{N}{2}$ edges of length $\frac{N}{2}-2$. In order to find the maximum-crossing solution, one has the following options:

- to take a 2-factor with $\frac{N}{2}$ edges of length $\frac{N}{2}-1$, $\frac{N}{2}-1$ edges of length $\frac{N}{2}-2$ and one edge of length $\frac{N}{2}-2$: in this case the theoretical maximum number of crossing matchings is $\frac{N(N-3)}{2}+(\frac{N}{2}-1)(N-4)+N-6 = N^2-\frac{7N}{2}-2$;
- to take a 2-factor with $\frac{N}{2}-1$ edges of length $\frac{N}{2}-1$, $\frac{N}{2}+1$ edges of length $\frac{N}{2}-2$: in this case the theoretical maximum number of crossing matchings is $(\frac{N}{2}-1)(N-3)+(\frac{N}{2}+1)(N-4) = N^2-\frac{7N}{2}-1$.

Clearly the second option is better, at least in principle, than the first one. The cycles h_k^* belong to the second case and saturate the number of crossing matchings. Suppose, then, to be in this case. Let us fix the $\frac{N}{2}-1$ edges of length $\frac{N}{2}-1$; this operation leaves two vertices without any edge, and this vertices are of the form r_a, $r_{a+\frac{N}{2} \pmod N}, a \in [1, N]$ (this is the motivation for the degeneracy of solutions). By the reasoning done above, the edges that link this vertices must be of length $\frac{N}{2}-2$, and so they are uniquely determined. They form the 6-points loop:

$$\begin{aligned}&(r_a, r_{a-1+\frac{N}{2} \pmod N}, r_{N-1+a \pmod N}, r_{a+\frac{N}{2} \pmod N},\\ &r_{a+1 \pmod N}, r_{a+1+\frac{N}{2} \pmod N})\end{aligned} \tag{B.2.4}$$

The remaining $N-6$ points, since $4|(N-6)$, by the same reasoning done in the proof of Proposition B.1., necessarily form the $\frac{N-6}{4}$ 4-points loops given by the permutations (B.2.2) □

Proposition B.2.2. *The 2-factor h^* that provides the optimal cost is such that $h^* \in \{h_k^*\}_{k=1}^N$.*

Proof. We have already proven that $h_k^*, k \in [1, N]$ are the 2-factors that have the maximum number of crossing matchings and there are no others out of them; then, automatically, the optimal solution is one of them, because any other 2-factor must have at least a non-crossing matching that can be replaced by a crossing one, lowering the cost. □

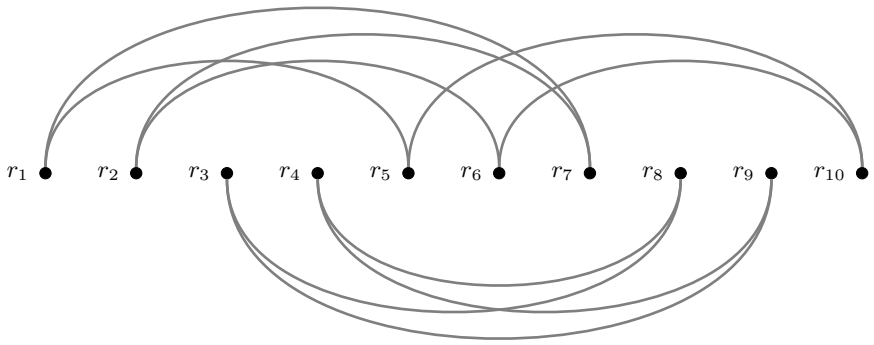


Figure B.2: One of the optimal 2-factor solutions for $N = 10$ and $p < 0$; the others are obtainable cyclically permuting this configuration

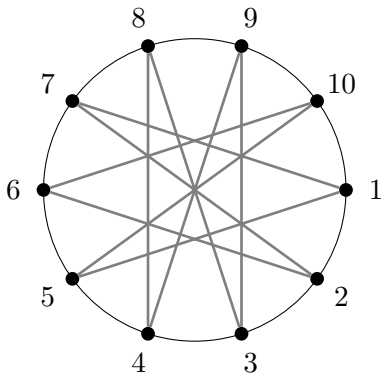


Figure B.3: The same optimal 2-factor solution of B.2, but represented on a circle, where the symmetries of the solutions are more easily seen

Bibliography

[1] J. Beardwood, J. H. Halton, and J. M. Hammersley. The shortest path through many points. *Proc. Cambridge Philos. Soc.*, 55(299):299–327, 1959.

[2] E. Boniolo, S. Caracciolo, and A. Sportiello. Correlation function for the grid-poisson euclidean matching on a line and on a circle. *J. Stat. Mech.*, 11:P11023, 2014.

[3] R. Capelli, S. Caracciolo, A. D. Gioacchino, and E. M. Malatesta. Exact value for the average optimal cost of bipartite traveling-salesman and 2-factor problems in two dimensions. 2018.

[4] S. Caracciolo, M. P. D'Achille, and G. Sicuro. Random euclidean matching problems in one dimension. *Phys. Rev. E*, 96:042102, 2017.

[5] S. Caracciolo, A. Di Gioacchino, and E. M. Malatesta. Plastic number and possible optimal solutions for an euclidean 2-matching in one dimension. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(8):083402, 2018.

[6] S. Caracciolo, A. D. Gioacchino, M. Gherardi, and E. M. Malatesta. Solution for a bipartite euclidean traveling-salesman problem in one dimension. *Phys. Rev. E*, 97:052109, 2018.

[7] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Phys. Rev. E*, 90:012118, 2014.

[8] S. Caracciolo and G. Sicuro. On the one dimensional euclidean matching problem: exact solutions, correlation functions and universality. *Phys. Rev. E*, 90(042112), 2014.

[9] N. J. Cerf, J. H. Boutet de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus. The random link approximation for the Euclidean traveling salesman problem. *Journal de Physique I*, 7(January):117–136, 1997.

[10] R. M. Karp and M. Steele. *The Travelling Salesman Problem*. John Wiley and Sons, New York, 1985.

[11] W. Krauth and M. Mézard. The Cavity Method and the Travelling-Salesman Problem. *Europhysics Letters*, 8(3):213–218, 1989.

[12] E. Lawler, D. Shmoys, A. Kan, and J. Lenstra. *The Traveling Salesman Problem*. John Wiley & Sons, Incorporated, 1985.

[13] M. Mézard and G. Parisi. Replicas and optimization. *Journal de Physique Lettres*, 46(17):771–778, 1985.

[14] M. Mézard and G. Parisi. A replica analysis of the travelling salesman problem. *Journal de Physique*, 47(1986):1285–1296, 1986.

[15] M. Mézard and G. Parisi. Mean-field equations for the matching and the travelling salesman problems. *Europhysics Letters*, 2(12):913–918, 1986.

[16] M. Mézard and G. Parisi. On the solution of the random link matching problems. *Journal de Physique*, 48(9):1451–1459, 1987.

[17] M. Mézard and G. Parisi. The Euclidean matching problem. *Journal de Physique*, 49:2019–2025, 1988.

[18] H. Orland. Mean-field theory for optimization problems. *Le Journal de Physique - Lettres*, 46(17), 1985.

[19] C. H. Papadimitriou. The euclidean travelling salesman problem is np-complete. *Theoretical computer science*, 4(3):237–244, 1977.

[20] A. G. Percus and O. C. Martin. Finite size and dimensional dependence in the Euclidean traveling salesman problem. *Physical Review Letters*, 76(8):1188–1191, feb 1996.

[21] S. Ravanbakhsh, R. Rabbany, and R. Greiner. Augmentative Message Passing for Traveling Salesman Problem and Graph Partitioning. *Advances in Neural Information Processing Systems*, 1(January):289–297, jun 2014.

[22] N. Sourlas. Statistical Mechanics and the Travelling Salesman Problem. *Europhysics Letters*, 2(12):919–923, 1986.

[23] M. Steele. *Ann. Probability*, 9:365–376, 1981.

[24] J. Vannimenus and M. Mézard. On the statistical mechanics of optimization problems of the travelling salesman type. *Journal de Physique Lettres*, 45(24):L1145–L1153, 1984.

[25] J. Wastlund. The mean field traveling salesman and related problems. *Acta Mathematica*, 204(1):91–150, 2010.