Local Optima in Landscapes of Combinatorial Optimization Problems

DIPLOMARBEIT

eingereicht von

Bärbel Krakhofer

zur Erlangung des akademischen Grades

Magister rerum naturalium

an der Formal- und Naturwissenschaftlichen Fakultät

der Universität Wien

Jänner 1995

Diese Arbeit entstand in der Zeit von Oktober 1993 bis Dezember 1994 am Institut für Theoretische Chemie der Universität Wien.
Ich möchte all jenen danken, die auf die eine oder andere Weise dazu beigetragen haben.

Peter Schuster danke ich für die freundliche Aufnahme in seine Gruppe und dafür, mein Interesse an Theoretischer Chemie geweckt zu haben.
Peter Stadler führte mich mit viel Geduld in wissenschaftliches Arbeiten ein und hatte stets kompetente Antworten auf alle aufgeworfenen Fragen.
Ivo Hofacker wußte Rat bei Computerproblemen und fand immer Zeit, um mir beim Ordnen meiner Gedanken zu helfen.
Erich Bornberg-Bauer stand bei meinen ersten Programmierversuchen mit Rat und Tat zur Seite.
Walter Fontana, Thomas Griesmacher, Robert Happel, Robert Hecht, Herbert Kratky, Kai Neumann und Alex Renner danke ich für das angenehme Arbeitsklima.
Vor allem danke ich meinen Eltern für ihre Unterstützung und dafür, daß sie mich dazu ermuntert haben, meine Studienzeit zu genießen.

Abstract

The landscapes of the symmetric and asymmetric Traveling Salesman Problem, the Graph Bipartitioning Problem, the Graph Matching Problem and the Sherrington-Kirkpatrick spin glass are investigated by performing adaptive and gradient walks. The number of local optima in these landscapes is obtained by random sampling. The number of configurations in a patch with radius correlation length in configuration space is computed and compared with the number of local optima.

The lengths of adaptive and gradient walks scale linearly with system size. Optima reached at the end of walks are better than randomly found optima. In most investigated landscapes there are O(1) local optima in a ball of radius correlation length.

Deutsche Zusammenfassung

Die Landschaften des symmetrischen und des asymmetrischen Traveling Salesman Problems, des Graph Bipartitioning Problems, des Graph Matching Problems und des Sherrington-Kirkpatrick Spinglases werden durch adaptive und gradient walks erforscht. Die Anzahl lokaler Optima in diesen Landschaften wird durch zufälliges Auswählen von Punkten bestimmt. Die Zahl der Konfigurationen in einem Ball mit Radius Korrelationslänge im Konfigurationsraum wird berechnet und mit der Anzahl der lokalen Optima verglichen.
Die Längen der adaptive und gradient walks steigen linear mit der Systemgröße an. Optima, die als Endpunkte von walks erreicht werden, haben bessere Werte der Kostenfunktion als zufällig gefundene Optima. In den meisten der untersuchten Landschaften sind O(1) lokale Optima in einem Ball mit Radius Korrelationslänge.

Contents

List of Figures

List of Tables

1 Introduction

Charles Darwin described the principles of biological evolution in his famous work "On the Origin of Species": errors occuring during the reproduction of organisms create variation; natural selection leads to the "survival of the fittest". The fittest organism is the one with the largest number of offspring reaching the age of fertility. In future generations those types of individuals which have more descendants will increase in number, the others will die out. Fitness of organisms can therefore be measured in terms of their reproductive success. It is very hard to describe the features which determine fitness instead of defining the fittest type as the one who survives in the end.

In the 1930s Sewall Wright introduced the notion of a fitness landscape. A fitness value is assigned to every genome - the genome can be seen as the set of parameter values of some fitness function. Variation occurs through small changes in the genome. The new possibility is accepted if it is better - which means that it has a higher fitness value - than its predecessor. The process of variation and selection thus involves a local search procedure in the space of possible genome sequences.

Evolutionary adaptation can be envisioned as a hill climbing process on such a complex fitness landscape. This strategy seems simple, but depending on the structure of the landscape the problem of finding the highest peak can become extremely hard. The biggest problem is to assign the fitness a priori instead of waiting for the outcome of the selection process in order to avoid the tautology of the survival of the survivor.

This intuitive concept of viewing optimization as a hill climbing process on a landscape can be applied to various combinatorial optimization problems. Landscapes arise in physics, biology and mathematics. Examples for biologically motivated landscapes are the free-energy-landscapes of RNA and proteins, affinity landscapes of antibodies, and fitness landscapes of RNA sequences based on estimated replication and degradation rate constants.

One has to be aware of the fact that these landscapes are high-dimensional, complex and rugged. Optimization on rugged landscapes, that is, landscapes with a large number of local optima, is very difficult. In most cases it will not be possible to reach the global optimum, so one has to settle for other goals, like finding a very good local optimum.

Local optima are traps for optimization algorithms. What method is the best to find a good minimum depends on the underlying structure of the landscape, so one has to find features which characterize a landscape. The landscapes of simple model systems can be explored and their statistical properties can be described in detail. Such model landscapes arise from combinatorial optimization problems like the Traveling Salesman Problem, the Graph Bipartitioning Problem and many others.

A thorough understanding of these model landscapes could lead to a better understanding of biologically motivated landscapes and the evolutionary processes which take place on them. To understand the process of evolution on rugged landscapes it is necessary to know the constraints of a local search procedure on such landscapes.

The overall of the landscape can be described by statistical measures like the number and fitness values of local optima, the lengths of gradient and adaptive walks, and the degree of ruggedness of the landscape.

One has to find a measure to compare landscapes which arise from different problems and have different properties; the final goal is to be able to predict the success of optimization on a novel landscape by the average behavior of the class of landscapes it belongs to. One such classification measure is the correlation length, the behavior of the autocorrelation function near the origin.

In this work the value landscapes of the Traveling Salesman Problem, the Graph Bipartitioning Problem, the Graph Matching Problem and the Sherrington-Kirkpatrick spin glass are investigated by performing adaptive and gradient walks, and the number and properties of local optima are obtained by random sampling. These data are compared with the features of the minimum free energy landscape of RNA molecules, which was subject of intensive studies carried out by Fontana et al [7].

2 Landscapes

In 1932 Sewall Wright introduced the notion of a landscape to describe evolution as an uphill walk on a fitness landscape [31]. This very intuitive concept of viewing an optimization process as walking on a landscape can be applied to various combinatorial optimization problems, which occur in fields as different as physics of spin glasses, theoretical biology and computer science. One has to be aware of the fact that these landscapes are high-dimensional and discrete.

A combinatorial optimization problem consists in finding a best configuration out of a very large but finite number of possible configurations. Usually one searches for the global minimum (or maximum) of a cost function which depends on many variables.

A configuration is an allowed combination (set) of these variables.

An instance of a combinatorial optimization problem consists of a set C of possible configurations and a cost function f that assigns a value to each configuration[20].

$$f : C \to R, \qquad c \to f(c)$$

The problem is to find an $\hat{x} \in C$ for which

$$f(\hat{x}) \leq f(y) \qquad \text{for all} \qquad y \in C$$

Such a point $\hat{x}$ is called a globally minimal solution to the given instance or a global minimum.

An optimization problem is a set I of instances of an optimization problem.

Maxima are defined analogous to minima; in the optimization problems that are subject of this work the cost function has to be minimized, so the term optimum in this context always means minimum.

To impose a structure on the space of possible configurations one can find a rule that determines whether two configurations are nearest neighbours. An other point of view is to define a move set, that is, operations which modify a configuration in a certain way. Nearest neighbours are reached by applying the move rule once. Examples for move rules are single point mutations or interchanges.

Configurations can be viewed as vertices in an abstract configuration space. If nearest neighbours are connected by edges, the set C of allowed configurations becomes a graph.

In case of binary sequences of length n with the move rule point mutation this configuration space is a hypercube of dimension n. The Hamming distance $d_H(x, y)$ is the number of positions in which the two sequences x and y differ. In the hypercube neighbouring sequences, which have Hamming distance $d_H = 1$, are connected by straight lines. Figure 1 shows the configuration space of binary sequences of length 4.

An instance of a landscape is defined by an instance of a combinatorial optimization problem and a rule that determines whether two configurations are nearest neighbours.

A natural metric distance $d(x, y)$ between any two configurations x and y on such a landscape is the minimum number of moves necessary to convert x to y.

Different movesets result in different topologies of the landscape. The choice of a different moveset generates a different set of neighbours, but does not change the set of allowed configurations.

Definition: A point (configuration) $x \in C$ is called a local minimum (local optimum) if

$$f(x) \leq f(y) \qquad \text{for all} \qquad y \in N(x)$$

where $N(x)$ denotes the neighbourhood of x, i.e. all $y \in N(x)$ are nearest neighbours of x,

$$N(x) = \{y|d(x, y) = 1\}.$$

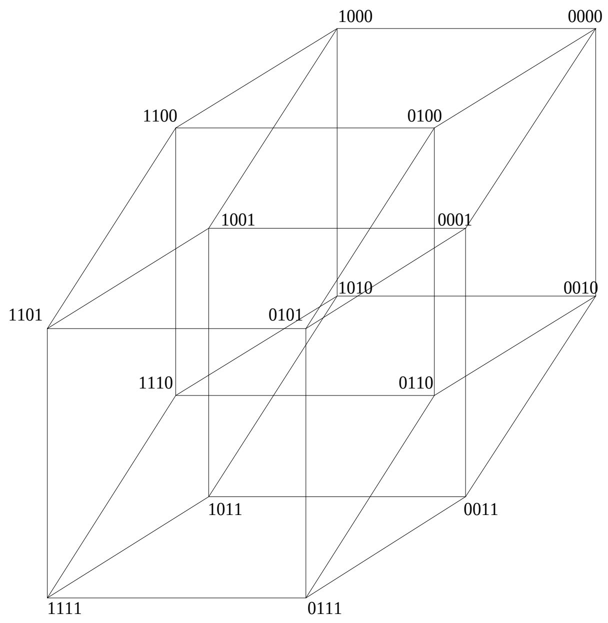

Figure 1: The configuration space of binary sequences of length 4 is a 4-dimensional hypercube

Since most of the landscapes of interest are rugged, which means that neighbouring configurations can have very different cost function values, optimization on these landscapes is a hard task. Even on very fast computers in most cases it is not possible to find the global optimum in reasonable timespans. So one has to be content with finding at least a very good suboptimal solution - a local optimum. To decide which optimization method will work best on the landscape considered, one has to gather information on the structure of the landscape.

3 Examples for landscapes

3.1 Biologically important landscapes

Evolutionary processes take place on fitness landscapes. Since the fitness of an organism in Darwin's sense - that means that fitness is measured by means of reproductive success - is hard to evaluate a priori, so far only simple model systems like RNA sequences have been investigated.

Energy landscapes are generated by plotting the free energy of the folded molecule vs. its sequence or its spatial structure. In the RNA case the secondary structure of a sequence can be computed by a recursive algorithm.

Properties like the ability of proteins to catalyze a specific reaction or the antigen binding capacity of antibodies give rise to other biologically motivated landscapes.

Landscapes arising from biological applications usually are rugged, that means that nearest neighbours can have very different values of the cost function.

3.1.1 Energy landscapes of RNA and proteins

The primary structure - or genotype - of an RNA molecule of length n can be represented by a sequence of n letters, with each letter chosen from a k-letter alphabet, i.e. the natural four-letter alphabet AUGC or the binary alphabet GC.

$$\mathbf{I} = \{\ \mathbf{AUGCGCGUACGUCGGACU} \ldots \mathbf{AGUCA}\}$$

RNA sequences are objects of combinatorial complexity: there are k^n possible sequences of length n. This enormous number of sequences can be ordered with the help of the concept of a sequence space. A natural moveset is point

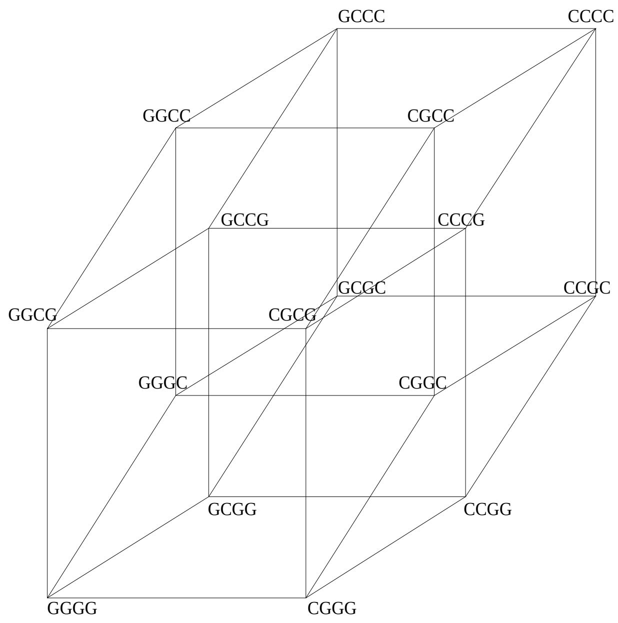

Figure 2: Sequence space of GC-sequences of length $n = 4$.

mutation, that is the exchange of a single letter in a single site. Since only sequences of constant length n are considered, insertions and deletions are not allowed. A metric on this landscape is the Hamming distance d_H, that is the number of positions in which two sequences differ. Neighbouring sequences have Hamming distance one. The sequence space of binary GC-sequences is a hypercube of dimension n (Figure 2), sequence spaces of four-letter sequences are more complicated objects. Every sequence can be reached from every other one by successive point mutations. The diameter of the sequence space (the maximum Hamming distance between two configurations) is n.

The analogue of the phenotype of an RNA molecule is its thermodynamically most stable spatial structure. The folding of an RNA molecule can be split into two steps:

- The folding of the string into a two-dimensional quasiplanar secondary structure by formation of complementary base pairings
- The formation of a three-dimensional tertiary structure from the secondary structure.

While no reliable model for predicting the tertiary structure exists so far, the secondary structure can be computed by a recursive algorithm which predicts the minimum free energy structure of the molecule [33],[32]. Moreover it is very hard to encode the tertiary structure in a compact form, while secondary structures are easily encoded and stored. Since the forces which stabilize the secondary structure, base pairing and stacking, are much stronger than the ones which influence the three-dimensional structure, the computation of RNA secondary structures is fairly reliable. For these reasons the investigation of phenotype properties was based on the predicted secondary structure.

A landscape is obtained by assigning to every sequence in sequence space the minimum free energy of its secondary structure. The minimum free energy landscape of RNA molecules was subject of extensive studies carried out by Fontana et al [4],[7].

In the case of the energy landscape of a single protein a configuration is the set of all spatial coordinates of the protein's atoms; the cost function assigns the potential energy of the molecule to each configuration. A protein usually has many different conformations which are nearly isoenergetic, so protein folding can be envisioned as an optimization process on a landscape with many local optima.

3.1.2 RNA fitness landscapes

The original idea of Sewall Wright was to construct a landscape by assigning a fitness value to every genotype. Evolution can be considered as an optimization process taking place on this fitness landscape. The problem consists in finding the best out of a large number of individuals. The fitness criterion in Darwin's sense is the number of fertile descendants. Evolution is the interplay of mutation and selection; mutation acts on the genotype, whereas selection takes place on the phenotype level. In a stationary population variants with more or more fertile offspring will increase in number and in percentage of the population, the others will die out. In the end one single fittest type will survive - this is the meaning of the popular catchphrase "survival of the fittest". A measure for an individual's fitness has to be defined that does not depend on the outcome of selection in order to avoid the tautology of the survival of the survivor.

The hardest problems that arise when developing models for evolution is to find a function which maps the genotype to the phenotype and a function that assigns a fitness value to the phenotype. So far the prediction of phenotype properties from the genetic information is not possible for even the simplest organisms. Due to the fact that an a priori computation of the fitness of organisms (instead of waiting for the outcome of selection) is out of reach, a simpler model system had to be found.

A simple system fulfilling the minimum requirements for Darwinian evolution - reproduction, variation and selection that leads to the survival of the fittest - has been found in the reaction kinetics of RNA molecules. Spiegelman et al. [23] showed that RNA molecules can proliferate in a cell free medium containing nucleoside triphosphates and the RNA replicase of the bacteriophage Qβ.

The first theoretical model of molecular evolution was proposed by Eigen in his pioneering work [3]. This model for polynucleotide replication was based on ordinary differential equations derived from chemical kinetics:

$$I_j \xrightarrow{Q_{kj}A_j} I_j + I_k$$

An RNA template sequence I_j is replicated with the rate constant A_j, the probability for the outcome of this process being a sequence I_k is Q_{kj}. The probability for error-free replication of sequence I_j is Q_{jj}. Degradation of the molecules is taken into account by the degradation rate constant D_j. An unspecific dilution flux $\Phi(t)$ removes templates from the system. The results of this theory are: there exists a sharply defined minimum accuracy of replication, the error threshold. Below this threshold genetic information is unstable - random replication.
Stationary states of a population are characterized by distributions of sequences - the so-called quasispecies. Not a single fittest type, but the quasispecies is selected by the evolutionary process.

RNA molecules can be multiplied in vitro by a replication assay containing the Qβ-RNA replicase. In this case the phenotype is the 3D-structure of the folded RNA molecule which is recognized by the Qβ-replicase and determines the replication velocity and thus the fitness of the molecule. By means of this system the replication kinetics of RNA was explored in great detail [2]. Since only single stranded molecules are accepted as templates by the enzyme, the secondary structure has to melt to make replication possible. On the other hand single stranded regions of an RNA molecule can be easily attacked by hydrolytic agents or nucleases.

Since the secondary structure - or phenotype - of RNA molecules can be predicted fairly well from the sequence or genotype, an RNA fitness landscape based on estimated replication and degradation velocities of RNA molecules and a computer model of evolution were investigated by Fontana et al.[5],[6], [4].

3.1.3 Peptide space, affinity landscapes of antibodies

Maynard-Smith suggested that the evolution of proteins proceeds via single mutations; mutants which have higher fitness than their predecessors will be expressed at higher rates and ultimately replace them [16]. The configurations in a peptide space are all peptides of a specified length. Nearest neighbours differ by a single amino acid in a single site (move rule: amino acid substitution). The fitness of a peptide can be defined as its ability to bind a particular substrate or to catalyze a specific reaction.

If the considered peptide is the variable region of an antibody a natural measure of fitness is the antibody affinity to the immunizing antigen. The affinity of each antibody could be derived experimentally, but given the large number of sequences this is not feasible. To predict an antibody's affinity for a certain antigen from sequence, one would have to know the three-dimensional structure of both antibody and antigen, which is also not possible at the moment. A model for antibody evolution during an immune response was proposed by Macken et al.[15]. Every sequence is assigned a random fitness value, and evolution is modeled by a strictly uphill walk via fitter one-mutant neighbours.

3.2 Model landscapes

3.2.1 The Traveling Salesman Problem (TSP)

The TSP is one of the most extensively studied combinatorial optimization problems [12]. Starting from his home town a salesman has to visit $n-1$ cities exactly once and then has to return to his home town. Knowing all intercity distances he will try to find the shortest tour which fulfills these requirements.

A tour τ can be described by listing the visited cities in chronological order ($\tau \in S_n$, the group of all permutations of the n cities). The configuration space is the set of all possible tours.

The cost function $f(\tau)$ that has to be minimized assigns the total length of the tour to every possible ordering of the cities :

$$f(\tau) = \sum_{i=1}^{n-1} \eta_{\tau(i)\tau(i+1)} + \eta_{\tau(n)\tau(1)}$$

where τ is the permutation encoding the order of the n cities, $\tau(i)$ is the i-th city visited in the tour, η_{ab} is the distance between the cities a and b.

The distance matrix can be symmetric ($\eta_{ab} = \eta_{ba}$) or asymmetric. An other special case is to compute the distances of cities scattered uniformly in a unit square.

For an n-city-TSP the allowed configurations are given by the $n!$ permutations of the n cities. By symmetry one may choose a fixed starting city, so the problem is restricted to the permutations of the remaining $n-1$ cities. The number of possible tours grows so quickly with n, that the problem of finding the shortest tour is a very hard task.

The TSP belongs to the large class of NP-complete problems [8]; even the most efficient algorithms which can solve such a problem require a computational effort that grows faster than polynomially with the size of the

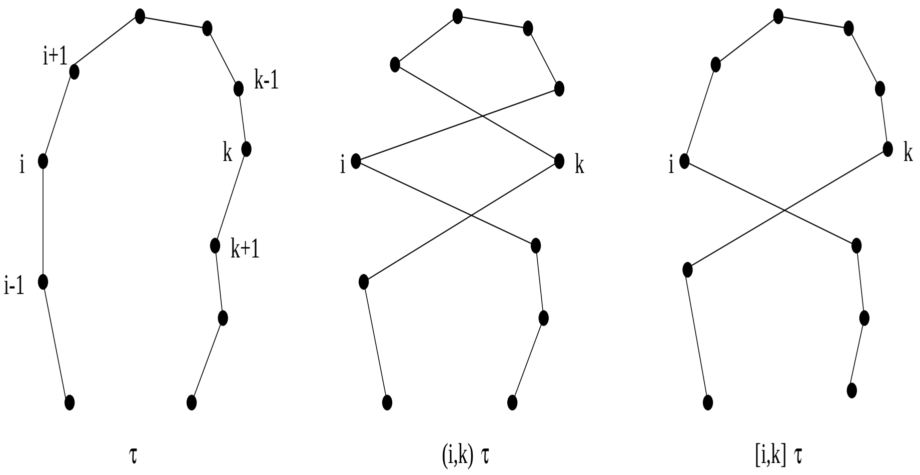

Figure 3: Move sets transposition and inversion for the TSP

problem. Computer time requirements for an exhaustive search - check all possible tours - for the best solution of a mid-size-TSP are prohibitive.

For the sTSP the cost function value $f(\hat{\tau})$ of the global optimum $\hat{\tau}$ for large n has been estimated [18]:

$$f(\hat{\tau}) \rightarrow 2.08$$

for large problems. A possible move rule which defines whether two tours are nearest neighbours of each other in configuration space is transposition $(i,k)\tau$, that is the exchange of two cities i and k in the tour. With this move rule applied to the symmetric problem every tour has $n(n-1)/2$ nearest neighbours (if initial city and direction are not specified).

If the choice of the cities to be exchanged is restricted to cities which follow each other in the tour, the move rule is called canonical transposition. An other possibility is the 2-opt move or inversion [13]: the inversion $[i,k]$ exchanges the cities i and k and reverts the path from i to k. Figure 3 shows the transposition and inversion moves.

For the move set transposition the maximal distance of two tours in configuration space is $\max d(x,y) = n-1$ [22]. For inversions the maximum distance is also $d(x, y) = n - 1$ [1].

A different way to encode a configuration is based on the intercity connections. A tour can be represented by a binary string, every position corresponding to the connection between two cities. In case of symmetric distance matrices the string has length $n(n-1)/2$. If two cities are connected in a tour, the corresponding position in the string is labelled by 1, otherwise by 0. A sequence encoding an allowed tour therefore has exactly n nonzero entries. A 2-opt move means that two intercity connections are removed and two other ones are inserted. Therefore the hamming distance d_H between two neighbouring tours is $d_H = 4$. The triangle inequality implies then

$$8d_{tr}(x, y) \geq d_H(x, y) \qquad \text{and} \qquad 4d_{inv}(x, y) \geq d_H(x, y).$$

3.2.2 The Graph Bipartitioning Problem (GBP)

Given is a graph V with an even number of vertices n and an $(n \times n)$-matrix H with h_{ij} being the distance between the vertices i and j (the weight of the edge connecting the vertices i and j) [26]. The graph has to be divided into two equal-sized subsets A and B such that the sum of the weights of all edges connecting the vertices contained in A with those contained in B is minimized:

$$f([A; B]) = \sum_{i \in A} \sum_{j \in B} h_{ij}$$

Special cases of the problem consider a symmetric distance matrix $h_{ij} = h_{ji}$ with h_{ij} being mutually independent random variables, or the vertices are uniformly scattered in a unit square.

The configuration space C of the GBP consists of the set of all partitions of the vertex set V into two equally sized subsets A and B. The partition can be encoded as a string of length n, the first $n/2$ vertices belonging to subset A. An other possibility is to encode the partition as a binary string, labelling a vertex by 1 if it is contained in subset A and by 0 if it is contained

in B. In the latter case it is possible to compute the hamming distance d_H between two configurations by counting the number of positions in which the two configurations differ.
A natural move set is to exchange two vertices: choose one vertex from the first half of the string (from A) and one from the second half (from B) and exchange them. Two partitions $[A, B]$ and $[C, D]$ of V are neighbours of each other if the symmetric differences

$$A \ominus C = B \ominus D = \{v_1, v_2\}$$

both equal the pair of exchanged vertices. The number of nearest neighbours for this case is $(n/2)^2$.

The definition of a neighbourhood imposes a structure on the configuration space C. If neighbouring partitions (configurations) are connected by edges, the set C of configurations becomes a graph. A metric for the distance $d([A, B], [C, D])$ between two configurations $[A, B]$ and $[C, D]$ on this configuration graph C is the minimum number of edges separating these two configurations.

This distance between two configurations is related to the hamming distance between the binary strings of the same configurations by

$$d([A, B], [C, D]) = \frac{1}{2} d_H([A, B], [C, D])$$

The distance sequence DS, that is the number of configurations with distance d from an arbitrary reference point is given by

$$DS(C, d) = \binom{n/2}{d}^2.$$

The number of possible configurations $\#C$, and the diameter of the configuration space, $diamC$, are

$$\#C = \binom{n}{n/2}$$
$$diamC = n/2$$

3.2.3 Graph Matching (GM)

An instance of this optimization problem is given by an even $(n \times n)$-matrix H with h_{ij} being the weight of the edge between vertices i and j. The problem consists in finding a bipartition of the vertices such that every vertex in the right part is connected to exactly one vertex in the left part and that the sum of the weights of the edges that connect the pairs of vertices becomes minimal [25]. A configuration can be encoded as an ordered list c of the vertices with vertices in positions 1 to $n/2$ belonging to the left part and in positions $n/2+1$ to n belonging to the right part, respectively. Vertex $c(i)$ in position i is connected to the vertex $c(i+n/2)$. In this case the cost function which has to be minimized is

$$f(c) = \sum_{i=1}^{n/2} h_{c(i),c(i+n/2)}$$

where $h_{c(a),c(b)}$ is the weight of the edge connecting the vertex in position a of the configuration (list) c with the vertex in position b.

A possible move set is transposition, that is the exchange of two vertices in the list. Different variants of the problem arise from choosing different construction rules for H. In the case of symmetric and asymmetric distance matrices the entries are drawn from a uniform distribution, in the symmetric case $h_{ij} = h_{ji}$. An other possibility is to scatter the n vertices randomly on the unit square and compute the distances between them.

If the matrix H is symmetric, exchanging vertices $c(i)$ and $c(i + n/2)$ is a neutral move. For the symmetric case the cost function value of the global optimum $\hat{c}$ has been estimated [17]:

$$f(\hat{c}) \to \frac{\pi^2}{12} \simeq 0.822$$

for large systems.

3.2.4 Spin glasses

Spin glasses are alloys of noble metals (Au, Ag, Cu, or Pt) with 0.1% to 10% transition metal ions (Fe or Mn), the latter randomly distributed in a matrix formed by the noble metal. The spins of neighbouring transition metal ions are coupled by magnetic interactions. The interaction between spins is sometimes ferromagnetic, which means that it is energetically favourable to align the spins, and sometimes antiferromagnetic - align spins with opposite signs - depending on their distance.

A configuration σ which describes the state of the system is given by the orientation of each of the n spins

$$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$$

with σ_i being the state of the i-th spin. The energy of a configuration is given in general by the Hamiltonian

$$H(\sigma) = \sum_{(i,j)} J_{ij}\sigma_i\sigma_j$$

where σ_i can have the values +1 or -1, the sum running over all pairs of interacting spins, and the constants J_{ij} - drawn randomly from a distribution - describing the interactions (coupling strength) between spins i and j. Special cases are the Ising models - interacting spins are neighbours in a two- or three-dimensional lattice - and the Sherrington-Kirkpatrick spin glass - a long range model with interaction possible between any two spins.

For the move set single spin flip the configuration space is an n-dimensional hypercube. Each configuration has n nearest neighbours. The distance between two configurations x, y in configuration space, the number of spin flips necessary to convert x to y, is the number of positions in which x and y differ, the hamming distance $d_H(x, y)$. The maximum distance in the space of configurations is n, the average distance is $n/2$.

Since the coupling constants J_{ij} are randomly distributed and vary in sign, every orientation of a spin favourable with respect to some spins means unfavourable coupling to others, a phenomenon known as frustration. The resulting landscapes are very rugged.

If p spins interact, the Hamiltonian has the form

$$H(\sigma) = \sum_{i_1 i_2 \ldots i_p} J_{i_1 i_2 \ldots i_p} \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_p}$$

The Sherrington-Kirkpatrick spin glass is the special case for $p = 2$.

3.2.5 Kauffman's n-k model

A frequently used statistical model of a landscape was proposed by Stuart Kauffman [9],[10],[11]. The so-called n-k model gives rise to a tunably rugged energy landscape. Consider a binary string c of n bits; the energy $F(c)$ of the string is the average of the contributions from each bit. The energy contribution f_i from bit i is a function of the state of this bit and the state of $k < n$ other bits. There are 2^{k+1} possible values for f_i; for every possible state of the $k+1$ bits on which f_i depends an independent random variable is chosen from a probability distribution. These assignments - the energy table - are generated independently for each of the n bits. The cost function is thus given by

$$F(c) = \frac{1}{n} \sum_{i=1}^{n} f_i.$$

The k sites that influence a given position can be chosen in different ways [7]:

- Adjacent Neighbourhood (AN) - the $k/2$ neighbouring bits on each side, assuming that the bits are arranged in a circle (periodic boundary conditions)

- Random Neighbourhood (RN) - k randomly picked sites

- Purely Random (PR) - f_i does not depend on i; all $k+1$ sites are chosen randomly.

The adjacent neighbourhood model corresponds to a one dimensional, short range spin glass, the random neighbourhood corresponds to a dilute, long range spin glass. Depending on the value of k the landscapes range from highly correlated single peaked to fully random uncorrelated landscapes. For $k = 0$ the energy contribution of each bit is completely independent of the state of the other bits, so each bit has an optimal state. The single global optimum is a string consisting of the optimal states of all its bits. For the maximum value of k ($k = n - 1$) the landscape is fully uncorrelated, which means that one-mutant neighbours in sequence space have completely random energy values with respect to each other. The n-k model is probably the best studied model landscape [29]; this model has been applied to investigate evolution on rugged landscapes. Fontana et al. compared the n-k model with the RNA free energy landscape [7].

4 Statistical analysis of landscapes

Since most of the landscapes of interest are rugged, optimization on these landscapes is a hard task. Even on very fast computers in most cases it is not possible to find the global optimum in reasonable timespans. So one has to be content with finding at least a very good suboptimal solution - a local optimum. To decide which optimization method will work best on the landscape considered, one has to gather information on the structure of the landscape.

Interesting questions are:

- Number of local optima
- Cost function values of random optima and optima reached after a walk
- Correlation between depth and width of optima
- Length of different walks
- Correlation - or ruggedness - of the landscape
- What optimization algorithm leads to the best results on a given landscape

4.1 Characterization of probability distributions

A distribution is specified by its infinitely many moments. Usually only the first two moments are of interest, since the cost function values of the model landscapes of interest are Gaussian distributed. The first two moments are the expectation value $\langle X_i \rangle$ and the variance $var(X_i)$

Expectation value:

$$\langle X_i \rangle = \frac{1}{N} \sum_{i=1}^{N} X_i$$

Variance:

$$\begin{aligned} var(X_i) &= \langle (X_i - \langle X_i \rangle)^2 \rangle \\ &= \langle X_i^2 \rangle - \langle X_i \rangle^2 \end{aligned}$$

The expectation value and the variance have the dimension of X_i, X_i^2 respectively.

An instance of a combinatorial optimization problem is usually generated by assigning lots of parameters at random, drawn from some distribution. The entries in the distance matrices used in this work are drawn from a uniform distribution (random numbers uniformly distributed in the interval $[0, 1]$; coupling constants for the SK spin glass are drawn from the interval $[-1, +1]$).

For this reason there are two different types of averages for combinatorial optimization problems:

- The average over configurations for one single instance with fixed parameters
- The ensemble average over different instances of the problem

Some properties like the autocorrelation function are self-averaging for larger system sizes, which means that both averages give the same result.

4.2 The autocorrelation function

The moments of a landscape do not give any information about the change in the value of the cost function when moving from one point in configuration space to another. To characterize the relationship between the distance of two points (number of moves necessary to convert one into the other), and the difference between the two cost function values, one has to define an appropriate measure, for example the autocorrelation function [4]:

$$\begin{aligned} r(k) &= \frac{cov(X_i, X_{i+k})}{\sqrt{var(X_i)var(X_{i+k})}} \\ &= \frac{\langle (X_i - \langle X_i \rangle)(X_{i+k} - \langle X_{i+k} \rangle) \rangle}{\sqrt{\langle (X_i - \langle X_i \rangle)^2 \rangle \langle (X_{i+k} - \langle X_{i+k} \rangle)^2 \rangle}} \end{aligned}$$

X_i and X_{i+k} belong to the same stochastic process. X_{i+k} is the value of the random variable exactly k steps after i. $cov(X_i, X_{i+k})$ is the covariance of X_i and X_{i+k} defined as

$$cov(XY) = \langle XY \rangle - \langle X \rangle \langle Y \rangle$$

If stationarity of the stochastic process is assumed, the expectation values, variances and higher moments of X_i and X_{i+k} become identical: $\langle X_i \rangle = \langle X_{i+k} \rangle$ and $var(X_i) = var(X_{i+k})$. The autocorrelation function can thus be rewritten as

$$r(k) = \frac{\langle (X_i - \langle X_i \rangle)(X_{i+k} - \langle X_{i+k} \rangle) \rangle}{\langle (X_i - \langle X_i \rangle)^2 \rangle}$$

or as

$$r(k) = 1 - \frac{\langle X_i^2 \rangle - \langle X_i X_{i+k} \rangle}{\langle X_i^2 \rangle - \langle X_i \rangle^2}$$

The limits of the autocorrelation function can easily be seen from this equation: $\lim_{k \to 0} r(k) = 1$, and $\lim_{k \to \infty} \langle X_i X_{i+k} \rangle = \langle X_i \rangle \langle X_{i+k} \rangle = \langle X_i \rangle^2$, since the two random variables X_i and X_{i+k} become independent for large k. This gives us $\lim_{k \to \infty} r(k) = 0$.

If X_i is the value of the cost function of some configuration i and X_j the cost function value of a configuration j which has distance d from the first one, then the autocorrelation function $\rho(d)$

$$\rho(d) = 1 - \frac{\langle X_i^2 \rangle - \langle X_i X_j \rangle_{d(i,j)=d}}{\langle X_i^2 \rangle - \langle X_i \rangle^2}$$

gives us the relationship between the distance in configuration space and the difference in the values of the cost function und is thus a measure for the ruggedness of the landscape.

The correlation structure of a landscape can be investigated by simple random walks: An initial configuration c_0 is chosen randomly, and by iterated application of the move rule the series $\{c_k\}$ of configurations is obtained; the series $\{X_k\}$ of cost function values of these configurations is called the random walk on the landscape.

The autocorrelation function $r(s)$ of a random walk is given by

$$r(s) = \frac{\langle (X_t - \langle X_t \rangle)(\langle X_{t+s} - \langle X_t \rangle) \rangle}{\langle (X_t - \langle X_t \rangle)^2 \rangle}$$

where X_{t+s} is the value of the cost function exactly s steps after t. Note that

$$d(c_t, c_{t+s}) \leq s$$

since applying the move rule once can increase the distance to the initial configuration by one at most, but may also decrease the distance by one. The autocorrelation functions $r(s)$ and $\rho(d)$ are related by the system of linear equations

$$r(s) = \sum_{d=0}^{n} \phi_{sd} \rho(d)$$

with ϕ_{sd} being the probability that a random walk is at distance d from the initial configuration after s steps [25].

For some cases the autocorrelation function has the shape of a decaying exponential, $r(s) = e^{-\lambda s}$. In this case the reciprocal value of the decay constant

λ, the correlation length $l = \lambda^{-1}$ contains all information about the autocorrelation function. The correlation length represents the value of k at which the autocorrelation function has dropped to e^{-1}.

Since $\phi_{11} = 1$ (after one step the random walk is at distance one from the initial configuration)

$$\rho(1) = r(1).$$

In this work the following definition of the correlation length is used:

$$l = -\frac{1}{\ln r(1)}.$$

Smoother landscapes have larger correlation lengths.

Problem	Metric	max $d(x,y)$	l	$\rho(d)$	$r(s)$	Reference
symmetric TSP	Transposition	$n-1$	$n/4$	?	$\sim e^{-4s/n}$	[27]
	Inversion	$n-1$	$n/2$	?	$\sim e^{-2s/n}$	[27],[1]
	Canonical Tr.	$\frac{n(n-1)}{2}$	$n/2$	?	$\sim e^{-2s/n}$	[30]
asymmetric TSP	Transposition	$n-1$	$n/4$	?	$\sim e^{-4s/n}$	[27]
	Inversion	$n-1$	$*$	?	$\sim (1/2)e^{-2s/n}$	[27]
	Canonical Tr.	$\frac{n(n-1)}{2}$	$n/3$	?	$\sim e^{-3s/n}$	[30]
Graph Bipartitioning	Exchange	$n/2$	$(n-3)/8$	$1-\frac{n-1}{n-2}[8\frac{d}{n}+16\left(\frac{d}{n}\right)^2]$	$\left(1-\frac{8}{n}+\frac{8}{n^2}\right)^s$	[26]
Graph Matching	Transposition	$n-1$	$n/4$	?	$\sim e^{-4s/n}$	[25]
n-k model AN	Hamming	n	$n/(k+1)$	$1-\frac{d(k+1)}{n}+\frac{d}{n\left(\binom{n-1}{d-1}\right)}\sum_{l=1}^{k}(k+1-l)\left(\binom{n-l-1}{d-2}\right)$		[7],[30]
n-k model RN	Hamming	n	$n/(k+1)$	$\left(1-\frac{d}{n}\right)\left(1-\frac{k}{n-1}\right)^d$		[7]
n-k model PR	Hamming	n	$n/(k+1)$	$\left(1-\frac{k+1}{n}\right)^d$		[7]
SK-spin glass	Hamming	n	$n/4$	$1-\frac{n}{n-1}[4\frac{d}{n}-4\left(\frac{d}{n}\right)^2]$	$\left(1-\frac{4}{n}\right)^s$	[30]

Table 1: Combinatorial optimization problems and their autocorrelation functions (* not defined). See also [24].

4.3 Exploration of landscapes

4.3.1 Adaptive and gradient walks

Information about the structure of a landscape can be attained by performing uphill (or downhill) walks on the landscape. These walks are restricted to increasing (or decreasing) values of the cost function and always end in local optima, so they provide information on the distribution of local optima.

A series of configurations is created by successively applying the move rule:

$$c_0 \to c_1 \to c_2 \to \ldots \to c_i \to c_{i+1} \to \ldots \to c_{i+k} \to \ldots$$

with the restriction of decreasing cost function values

$$f(c_0) > f(c_1) > f(c_2) > \ldots > f(c_i) > f(c_{i+1}) > \ldots > f(c_{i+k}) \ldots$$

Subsequent configurations c_i, c_{i+1} in this series have distance one

$$d(c_i, c_{i+1}) = 1$$

so the walk occurs along the edges of configuration space.

Two kinds of walks are of interest:

- **Adaptive walks** start at a randomly chosen configuration, choose one neighbouring configuration at random and accept it, if it has a better value of the cost function (first improvement). As long as a better solution exists, it is adopted and the neighbourhood search is repeated from the new solution. The walk stops if no neighbour has to offer a better value of the cost function, which means that a local optimum is reached.

- **Gradient walks** are deterministic walks of step size one. They also start at a randomly chosen configuration, but all neighbours are checked and the one with the best value of the cost function is accepted (steepest descent).

4.3.2 Uniform sampling

To get information about the correlation in the landscape one has to know the average differences in the values of the cost function for pairs of configurations with different distances from each other. Since random walks provide good samples only for small distances, a better sampling method is needed: A start configuration is chosen randomly, and a sample of mutants is generated for distances $d = 1$ to the maximum distance $d = d_{max}$. The value of the cost function is computed for every mutant, and the averaged differences are used to compute the autocorrelation function $\rho(d)$.

4.3.3 About neighbourhoods

The size of the neighbourhood of a configuration - defined by the move set - strongly influences optimization procedures which are based on local searching. A larger neighbourhood might promise better optima but takes longer to search, so the number of optima found in a certain computer-timespan will be relatively low. A neighbourhood has a certain strength, that means that optima found by local search have a certain average quality [20]. This strength seems to be related to the correlation of the cost function values of the start configurations with those of the local optima. The quality of local optima found with a strong neighbourhood should be independent of the quality of the starting configuration, while in weaker neighbourhoods the correlation between the two cost function values is supposed to be high. The ruggedness of a landscape strongly depends on the definition of neighbourhood. The term ruggedness is not exactly defined - a landscape is rugged if it has many local optima; Palmer [19] calls a landscape rugged if the number of local optima increases at least exponentially with system size. Adaptive and gradient walks on rugged landscapes are short compared to the diameter of the landscape, and the nearest neighbour correlation is low [28],[10]. The sTSP landscape is much more rugged for moveset canonical transposition than for transpositions, as one can see from the numerical data in the next chapter.

5 Numerical Results

The following five model landscapes were investigated:

- symmetric Traveling Salesman Problem (sTSP)
- asymmetric Traveling Salesman Problem (aTSP)
- symmetric Graph Bipartitioning Problem (GBP)
- symmetric Graph Matching Problem (GM)
- Sherrington-Kirkpatrick spin glass (SK)

The probability of finding a local optimum by random sampling was obtained and compared with the number of configurations in a ball of radius correlation length; adaptive and gradient walks were carried out. The autocorrelation function for the symmetric TSP with transposition distance metric was computed.

The probability of finding a local optimum by random sampling decreases with increasing number of neighbours and with increasing correlation between neighbours.

Stadler conjectures that there is $O(1)$ local optimum in configuration space in a patch of radius correlation length [27].

5.1 The Traveling Salesman Problem

5.1.1 The autocorrelation function for the symmetric TSP

Permutations

S_n, the group of all permutations of the numbers 1 to n, is the natural support of an n-city TSP [27].

$$| S_n | = n!$$

A permutation x can be denoted by displaying its values:

$$\mathbf{x} = \begin{pmatrix} 1 & 2 & 3 & \dots \\ x(1) & x(2) & x(3) & \dots \end{pmatrix}$$

Definition: Let $i_1, i_2, \dots, i_r$ be distinct integers between 1 und n. If $x \in S_n$ fixes the other integers $(x(j) = j)$ and

$$x(i_1) = i_2, \qquad x(i_2) = i_3, \dots, \qquad x(i_{r-1}) = i_r, \qquad x(i_r) = i_1,$$

then x is an $r - cycle$ (a cycle of length r)[21].

A way to denote an r-cycle is

$$\mathbf{x} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 4 & 1 \end{pmatrix} = \dots$$

There are r such notations for an $r - cycle$.
All $1 - cycles$ equal the identity permutation, which is denoted by (1).

Definition: Two permutations x, y are disjoint, if every i moved by one is fixed by the other.

Theorem [21]: Every permutation $x \in S_n$ is the product of disjoint cycles; this factorization is unique except for the order in which the cycles are written.

Definition: A $2 - cycle$ is also called transposition.

Example:

$$\mathbf{x} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 1 & 5 & 2 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 4 & 5 \end{pmatrix}\begin{pmatrix} 6 \end{pmatrix}$$

Consider two possible tours (permutations) x,y of length n (number of cities); The distance d between these two permutations is the minimum number of transpositions necessary to convert x to y (or y to x, which is the same for symmetry reasons). There is exactly one permutation which turns x to y; the minimum number of transpositions necessary to get this permutation from identity is equal to the distance between x and y and is given by

$$d = n - k(xy^{-1})$$

where $k(xy^{-1})$ is the number of cycles of the permutation xy^{-1} (cycles of length 1 are also counted!) [22].

The permutation xy^{-1} is computed by first applying y^{-1}, then x. The number of permutations in S_n with exactly k cycles is given by the signless Stirling number $ss(n,k)$ [14]:

$$\begin{aligned} ss(n,k) &= ss(n-1,k-1) + (n-1) * ss(n-1,k) \\ \sum_{k=1}^{n} ss(n,k) &= n! \end{aligned}$$

One possibility p of cycle composition for $n = 8, k = 4$ is (1)(1)(1)(5), that means that the first cycle has length 1, the fourth one length 5. The number of permutations in S_n which have a certain cycle composition is

$$a(n,k,p) = \frac{n!}{\prod\limits_{i=1}^{k} c_l(i) \prod\limits_{c_l} m(c_l)!}$$

where $c_l(i)$ denotes the length of the $i-th$ cycle of p and $m(c_l)$ is the number of cycles with equal length c_l (product runs over all c_l with $m(c_l) \geq 1$). The sum over all p gives the signless Stirling number [21].

$$\sum_p a(n,k,p) = ss(n,k)$$

Computation of the autocorrelation function

The autocorrelation function for landscapes of combinatorial optimization problems has the form [26]

$$\rho(d) = 1 - \frac{\langle (f(x) - f(y))^2 \rangle_{d(x,y)=d}}{\langle (f(x) - f(y))^2 \rangle_{random}}$$

For every possible d (k) 10000 permutations xy^{-1} are randomly generated, distributed over all cycle composition possibilities according to their weights $a(n,k,p)/ss(n,k)$

For every xy^{-1} a permutation y is generated ($n!$ possibilities), the permutation x is then exactly defined ($n!/2$ possibilities to choose pairs of permutations x, y which give exactly the considered xy^{-1}).

$$\sum_k \sum_p \frac{n! a(n,k,p)}{2} = \sum_k \frac{n! ss(n,k)}{2} = \frac{n! n!}{2}$$

(Number of possibilities to choose x and y from S_n). The differences in the values of the cost function for x and y are computed and used to obtain the value of the autocorrelation function. The autocorrelation function $\rho(d)$ for the symmetric TSP with transposition metric, system size $n = 100$ was computed (up to $d = 42$). Figure 4.

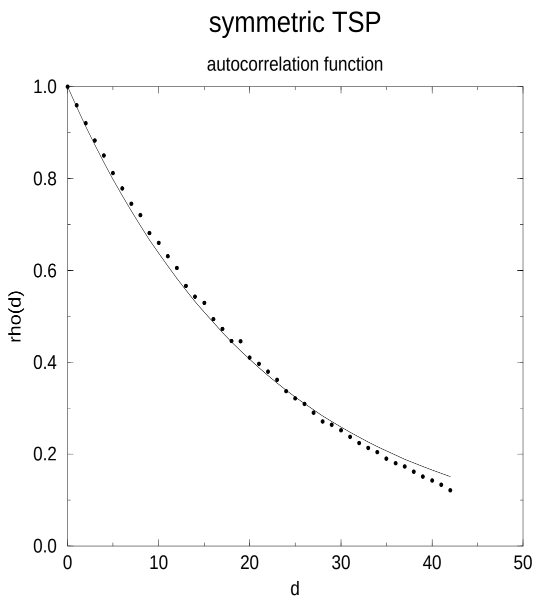

Figure 4: Autocorrelation function for the symmetric TSP, transposition metric. The solid line is the exponential fit :$\rho(d) = exp(-0.04503 * d)$.

5.1.2 Number of local optima

Symmetric TSP:

The highest number of local optima is found for canonical transpositions - a configuration has n nearest neighbours, the hamming distance between neighbouring configurations is always 4, which means that 2 bonds between cities in a tour are removed and 2 other bonds are inserted. In the landscapes arising from the move sets transposition and inversion a configuration has $n(n-1)/2$ nearest neighbours, but the nearest neighbour correlation is higher for inversion (hamming distance between nearest neighbours is always 4) than for transposition ($d_H = 8$ if there are at least 2 cities between the exchanged ones in the tour - true for $n(n-5)/2$ neighbours - otherwise $d_H = 4$, true for $2n$ neighbours). For this reason the inversion landscape has fewer optima than the transposition landscape, as one can see in Figure 5.

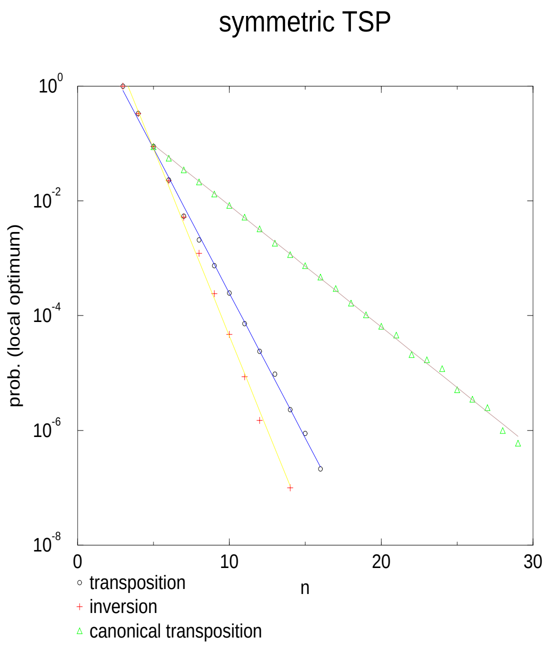

Figure 5: Probability of finding a local minimum by random sampling vs. system size (number of cities) for the symmetric TSP for move sets transposition, inversion and canonical transposition. The lines are least square exponential fits.

Asymmetric TSP:
The canonical transposition landscape again has the highest number of local optima (n neighbours), and more than for the symmetric problem, since $d_H = 6$ for neighbouring configurations. The number of optima for transposition metric is higher than in the symmetric case, since $d_H = 8$ if there is at least 1 city between the exchanged ones in the tour - true for $n(n-3)/2$ neighbours, else $d_H = 6$; this means that in the asymmetric case nearest neighbours on average differ in more intercity connections, hence the correlation in the landscape will be lower and the number of local optima will be higher than in the symmetric case. For a 15-city-TSP the probability of randomly finding a local optimum is 8.8e-07 in the symmetric case and 2.8e-06 in the asymmetric case. For the move set inversion there are much more optima than in the symmetric case, even more than for asymmetric transposition, since the hamming distance between neighbours is $2(r+1)$ with r being the number of cities lying between the exchanged ones in the tour. Figure 6.

5.1.3 Number of configurations in a patch of radius correlation length for the sTSP

It is known that the correlation length along the steps of a random walk for the symmetric TSP with move set transposition scales linearly with problem size [27]

$$l = 0.246n - 0.583.$$

The exact number of configurations C_{l-ex} with distances $d = 0$ to $d = n/4$ from an arbitrary reference point is

$$C_{l-ex} = \sum_{d=0}^{n/4} ss(n, n-d)$$

with $ss(n,i)$ being the signless Stirling numbers. The number of configurations $V(d)$ that have distance d from an arbitrary configuration (number of

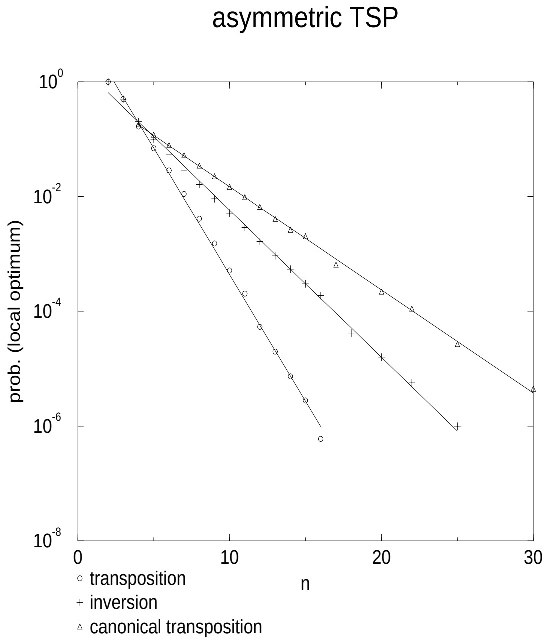

Figure 6: Probability of finding a local minimum by random sampling vs. system size (number of cities) for the asymmetric TSP for move sets transposition, inversion and canonical transposition.

permutations of n numbers with exactly $k = n - d$ cycles) has been approximated by Sorkin [22]

$$V(d) \simeq \left(n^2/2\right)^d/d! \qquad \text{for} \qquad d \ll n$$

giving an estimate for C_l

$$C_{l-app1} = \sum_{d=0}^{n/4} \left(n^2/2\right)^d/d!.$$

This sum can be approximated by its largest term

$$C_{l-app2} = \left(n^2/2\right)^l/l!.$$

The correlation length l is

$$l = n/4 = xn.$$

By applying Stirling's formula the largest term of the sum can be rewritten as

$$\begin{aligned}
\frac{\left(\frac{n^2}{2}\right)^l}{l!} &= \frac{\left(\frac{n^2}{2}\right)^{xn}}{(xn)!} = \frac{n^{2xn}}{2^{xn}\,(xn)!} \simeq \\
&\simeq \frac{n^{2xn}}{2^{xn}\left(\frac{xn}{e}\right)^{xn}\sqrt{2\pi xn}} = \frac{n^{2xn}e^{xn}}{2^{xn}x^{xn}n^{xn}}\frac{1}{\sqrt{2\pi xn}} = \\
&= \left(\frac{ne}{2x}\right)^{xn}\frac{1}{\sqrt{2\pi xn}} \simeq \left(\frac{ne}{2x}\right)^{xn} = \\
&= \left(\frac{nx}{e}\right)^{xn}\left(\frac{e^2}{2x^2}\right)^{xn} \simeq (xn)!\left(\frac{e^2}{2x^2}\right)^{xn}
\end{aligned}$$

$$C_{l-app3} = (xn)!\left(\frac{e^2}{2x^2}\right)^{xn} \simeq \Gamma(xn)\left(\frac{e^2}{2x^2}\right)^{xn}$$

The quality of the estimates can be seen in figure 7.

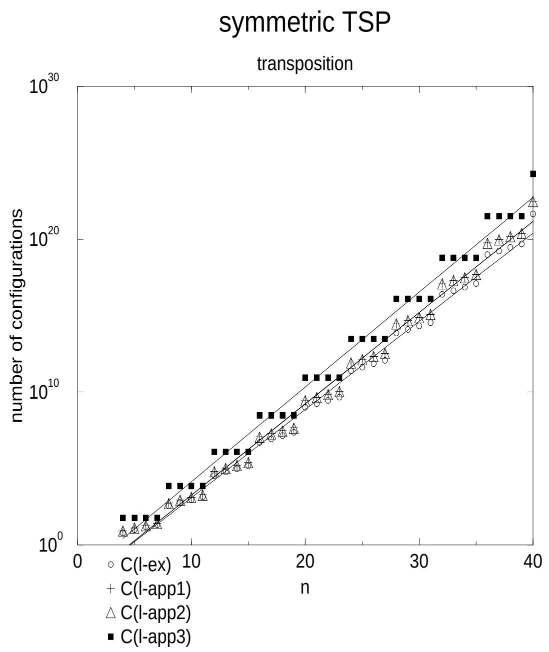

Figure 7: Number of configurations in a patch of radius l (and approximations) vs. system size for the symmetric TSP, transposition metric.

The Stirling approximation C_{l-app3} was computed for $n = 4$ to $n = 30$ and compared with the experimentally obtained number of optima. There is one local optimum in a ball of radius correlation length as conjectured by Stadler [27]. Figure 8.

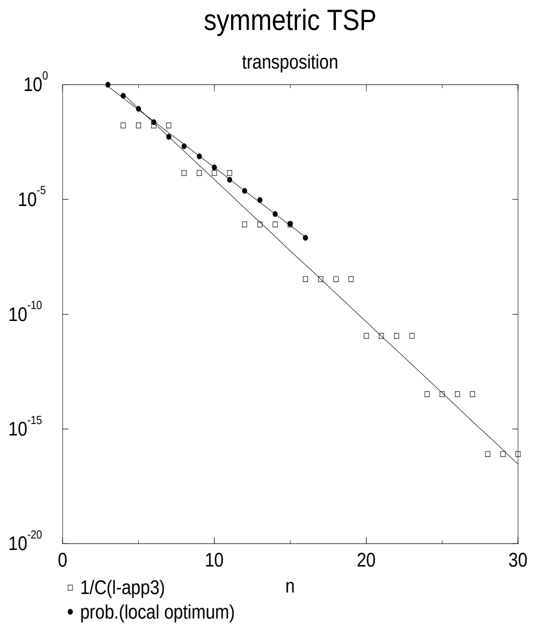

Figure 8: Density of local optima and $1/C_{l-app3}$ for the symmetric TSP, transposition metric. There is one local optimum in a ball of radius correlation length.

5.1.4 Adaptive and gradient walks

Results for the symmetric TSP:

All average walk lengths grow about linear with system size. The shortest walks are observed with the moveset canonical transposition, which has the lowest number of neighbours, therefore the walks are quickly trapped in a local optimum, followed by transposition and inversion. Inversion walks are twice as long as transposition walks.

For the symmetric TSP, movesets inversion and transposition, adaptive walks are about three times as long as gradient walks. For the move set canonical transposition adaptive walks are only one third longer than gradient walks. Figure 9.

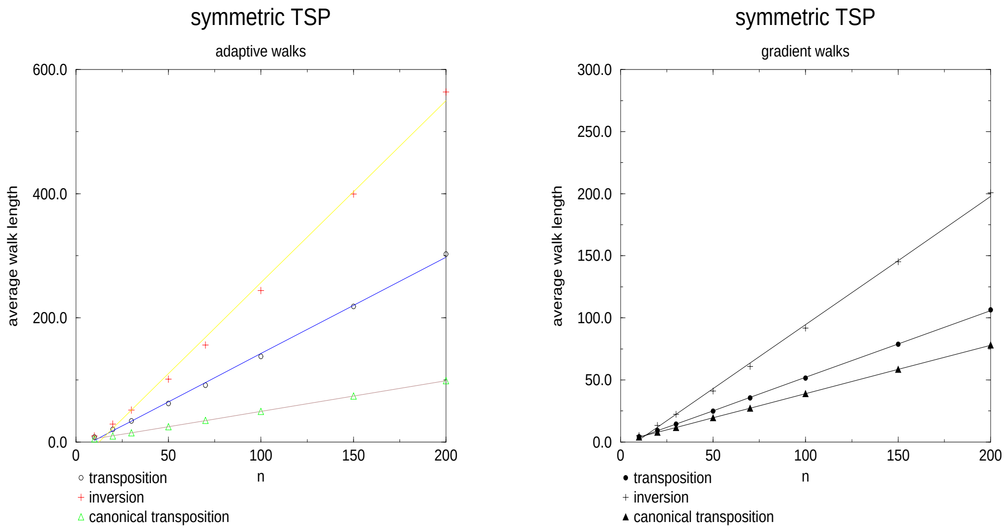

Figure 9: Average lengths of adaptive and gradient walks vs. system size (number of cities) for the symmetric TSP for move sets transposition, inversion and canonical transposition.

The average tour lengths at the end of walks (quality of optima) are the same for adaptive and gradient walks. The best optima are found with inversions, followed by transpositions; walks with move set canonical transposition find the worst optima. Figure 10.

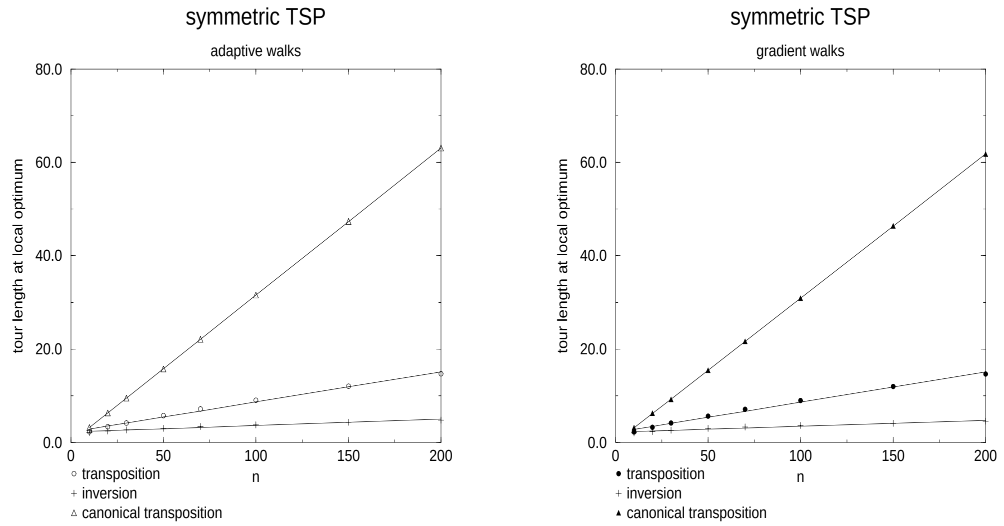

Figure 10: Average tour lengths at local minima attained after adaptive and gradient walks vs. system size (number of cities) for the symmetric TSP for move sets transposition, inversion and canonical transposition.

For the move set transposition the optima obtained after a walk are better than randomly found optima. Figure 11.

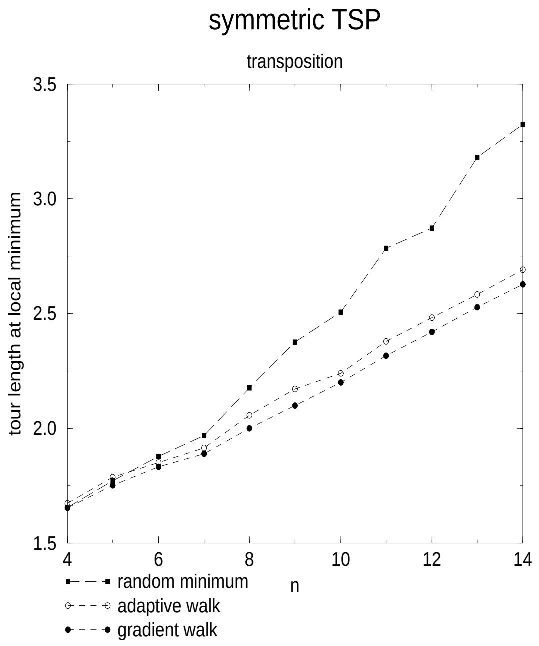

Figure 11: Average tour lengths at randomly found local minima and at minima attained after walks vs. system size (number of cities) for the symmetric TSP for move set transposition.

For constant system size $n = 150$, move set transposition, the following observations can be made: The longer the walk, the better the optimum found by this walk, and the larger the difference in energy to the start point of the walk, but also the worse the starting point. Figure 12.

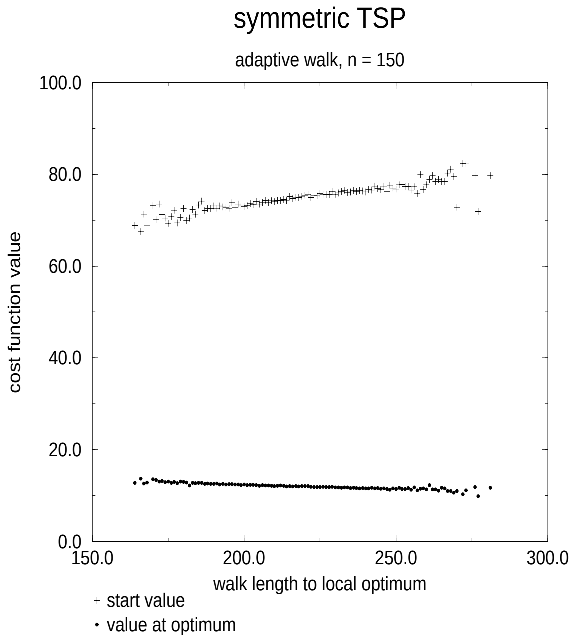

Figure 12: Average values of the cost function of starting configurations and of local optima attained after adaptive walks vs. walk length to the local optimum for the symmetric TSP, move set transposition, n = 150. 10 000 walks were carried out, 10 walks per instance. Deviations at the sides arise because walk lengths are Gaussian distributed, and hence the statistics is pure for extremal walk lengths.

For constant system size $n = 150$, adaptive walk: With move sets transposition and inversion the average start energy is about the same for all optima - a walk ending in a good optimum has the same average start energy as a walk ending in a bad optimum (strong neighbourhood).

This is not true for walks using the move set canonical transposition: better optima have already started at configurations with better cost function values (weak neighbourhood). The better the optimum, the longer the average walk that reached it.

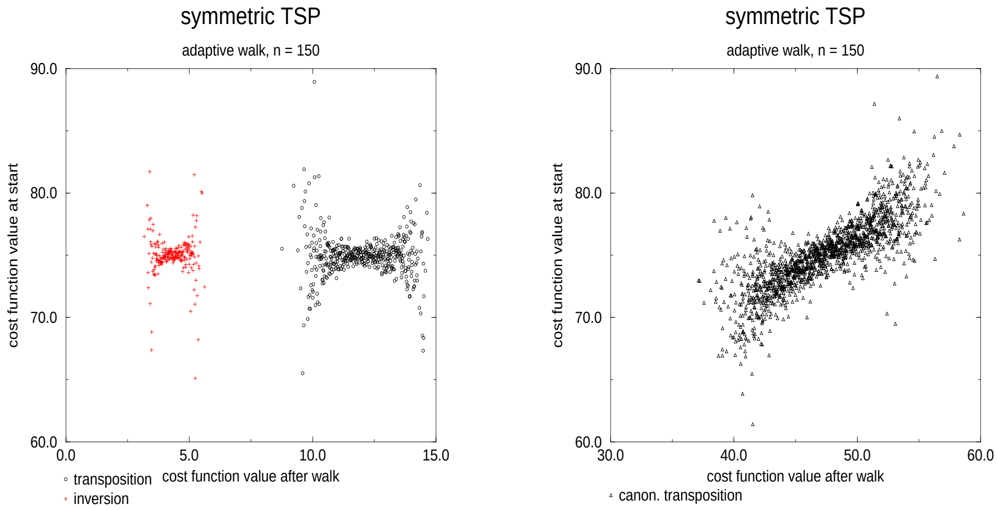

Figure 13: Average values of the cost function at the starting point vs. values at the end of an adaptive walk on the symmetric TSP, n=150, for move sets transposition, inversion and canonical transposition. Deviations at the sides arise from pure statistics because of the Gaussian distribution of cost function values.

Results for the asymmetric TSP:

Adaptive walks with move sets transposition and canonical transposition are roughly as long as for the symmetric problem, while inversion walks are very much shorter. This seems reasonable because in the asymmetric case the inversion of half of the tour is equivalent to randomisation of half of the tour, while in the symmetric case only two bonds are exchanged.

Gradient walks with move set transposition are about as long as in the symmetric case, canonical transposition walks are roughly 17 percent shorter than in the symmetric case, while the length of inversion walks reaches only one third of that of the symmetric case. Figure 14.

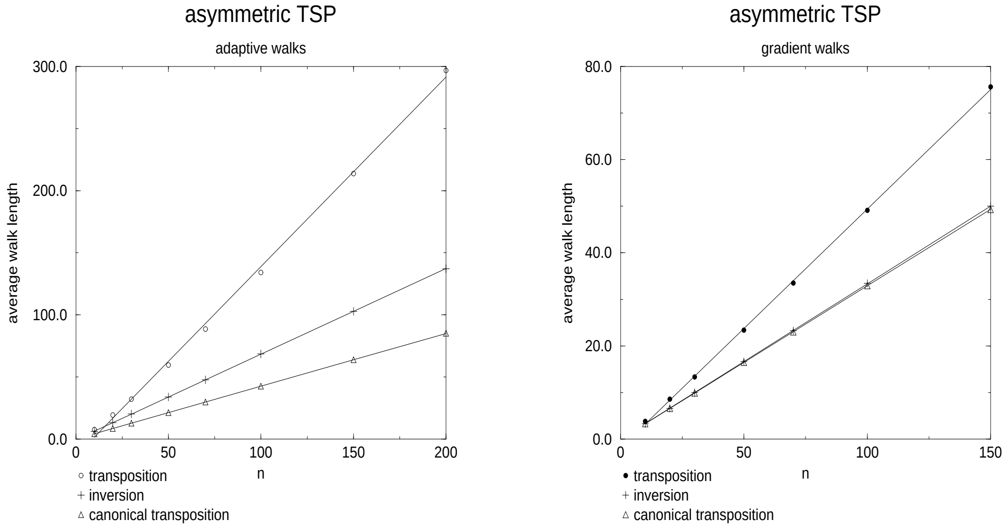

Figure 14: Average lengths of adaptive and gradient walks vs. system size (number of cities) for the asymmetric TSP for move sets transposition, inversion and canonical transposition.

The average tour lengths at the end of walks are the same for adaptive and gradient walks. Transposition and canonical transposition walks find optima of the same quality as in the symmetric case, while the tour length at the end of inversion walks is longer than in the symmetric case. Figure 15.

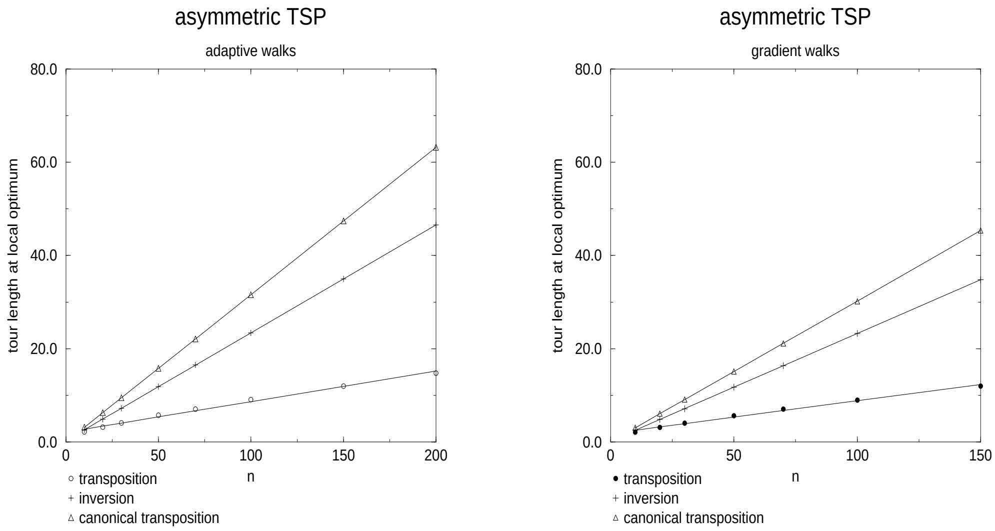

Figure 15: Average tour lengths at local minima attained after adaptive and gradient walks vs. system size (number of cities) for the asymmetric TSP for move sets transposition, inversion and canonical transposition.

5.2 The Graph Bipartitioning Problem

5.2.1 The autocorrelation function for the GBP

The autocorrelation function $\rho(d)$ for the GBP can be obtained by uniform sampling.

$$\rho(d) = 1 - \frac{\langle (f(x) - f(y))^2 \rangle_{d(x,y)=d}}{\langle (f(x) - f(y))^2 \rangle_{random}}$$

$$\langle (f(x) - f(y))^2 \rangle_{random} = \sum_d p(d) \langle (f(x) - f(y))^2 \rangle_d$$

where

$$p(d) = DS(d)/\#C$$

$$\sum_{d=0}^{n/2} DS(d)\rho(d) = 0$$

The analytical solution for the autocorrelation function for the GBP is known [26]:

$$\rho(d) = 1 - 8\frac{d}{n}\left(1 + \frac{1}{n-2}\right) + 16\left(\frac{d}{n}\right)^2\left(1 + \frac{1}{n-2}\right)$$

The autocorrelation function along the steps of a random walk is given by [26]:

$$r(s) = \left(1 - \frac{8}{n} + \frac{8}{n^2}\right)^s$$

The correlation length along the steps of a random walk is [26]

$$l_{walk} = -\frac{1}{\ln(1 - \frac{8}{n} + \frac{8}{n^2})} = \frac{n-3}{8} + O\left(\frac{1}{n}\right),$$

The correlation length l_{ex} for the exchange metric can be computed from the (empirical or theoretical) autocorrelation function $\rho(d)$; l_{ex} is the value of the distance d at which the autocorrelation function $\rho(d)$ has dropped to e^{-1},

$$\rho(l_{ex}) = e^{-1}$$

and is smaller than l_{walk}, as expected.

$$l_{ex} = 0.098n - 0.132 \approx \frac{n-1}{10}$$

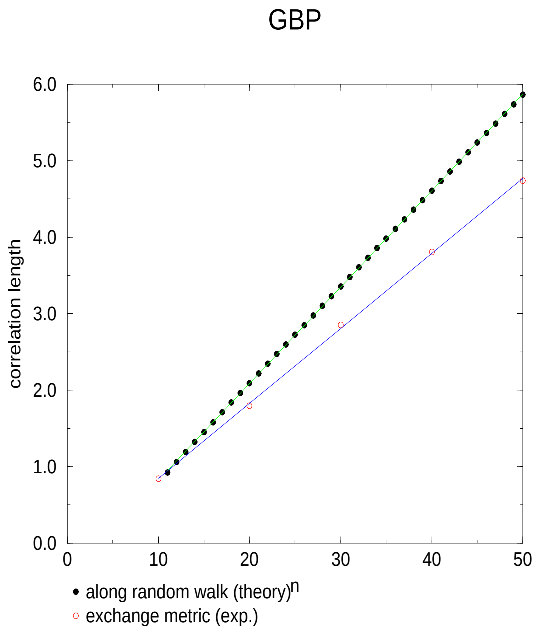

Figure 16: Correlation lengths vs. system size for the GBP. Correlation lengths along a random walk (theoretical prediction) are taken from [26], correlation lengths for the exchange metric are computed from empirical autocorrelation functions.

5.2.2 Number of local optima

Every configuration has $n^2/4$ nearest neighbours with hamming distance $d_H = 2$. A move replaces $(n-2)$ edges in the graph by $(n-2)$ other ones, which means that $(n-2)$ out of $n^2/4$ bonds are exchanged. Figure 18 shows the probability of hitting a local optimum at random.

5.2.3 Number of configurations in a patch of radius correlation length

The number DS of configurations with distance d from any configuration is given by

$$DS(d) = \binom{n/2}{d}^2.$$

The number of configurations C_l in a patch of radius correlation length $l = (n-3)/8$ is thus given by:

$$C_l = \sum_{d=0}^{(n-3)/8} \binom{n/2}{d}^2$$

This sum can be approximated very well by its largest term

$$C_l = \sum_{d=0}^{l} \binom{n/2}{d}^2 \simeq \binom{n/2}{l}^2$$

Set $l = xn$, then the largest term of the sum can be rewritten as

$$\binom{n/2}{xn}^2 = \left(\frac{\left(\frac{n}{2}\right)!}{(xn)!\left(\frac{n}{2}-xn\right)!}\right)^2 = \left(\frac{\left(\frac{n}{2}\right)!}{(xn)![\left(\frac{1}{2}-x\right)n]!}\right)^2$$

By using Stirling's formula this can be rewritten as

$$\left(\frac{\left(\frac{n}{2}\right)!}{(xn)![\left(\frac{1}{2}-x\right)n]!}\right)^2 \simeq \left(\frac{\left(\frac{n}{2e}\right)^{n/2}\sqrt{\pi n}}{\left(\frac{xn}{e}\right)^{xn}\sqrt{2\pi xn}\left(\frac{(1/2-x)n}{e}\right)^{(1/2-x)n}\sqrt{2\pi\left(\frac{1}{2}-x\right)n}}\right)^2 =$$

$$
\begin{aligned}
&= \frac{\left(\frac{n}{2e}\right)^{n} \pi n}{\left(\frac{xn}{e}\right)^{2xn} 2\pi xn \left(\frac{(1/2-x)n}{e}\right)^{(1-2x)n} 2\pi\left(\frac{1}{2}-x\right) n} = \\
&= \left[\frac{\left(\frac{n}{2e}\right)}{\left(\frac{xn}{e}\right)^{2x}\left(\frac{(1/2-x)n}{e}\right)^{(1-2x)}}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} = \\
&= \left[\frac{n}{2\,(xn)^{2x}\left[\left(\frac{1}{2}-x\right) n\right]^{(1-2x)}}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} = \\
&= \left[\frac{n}{2x^{2x} n^{2x}\left(\frac{1}{2}-x\right)^{(1-2x)} n^{(1-2x)}}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} = \\
&= \left[\frac{n}{2\left(\frac{x}{\frac{1}{2}-x}\right)^{2x}\left(\frac{1}{2}-x\right) n^{2x} n^{(1-2x)}}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} = \\
&= \left[\frac{n}{\left(\frac{x}{\frac{1}{2}-x}\right)^{2x}(1-2x)\,n}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} = \\
&= \left[\frac{\left(\frac{1-2x}{2x}\right)^{2x}}{(1-2x)}\right]^{n} \frac{1}{2\pi xn\left(\frac{1}{2}-x\right)} \simeq \\
&\simeq \left[\frac{\left(\frac{1-2x}{2x}\right)^{2x}}{(1-2x)}\right]^{n}
\end{aligned}
$$

Figure 17 shows the quality of the approximations.

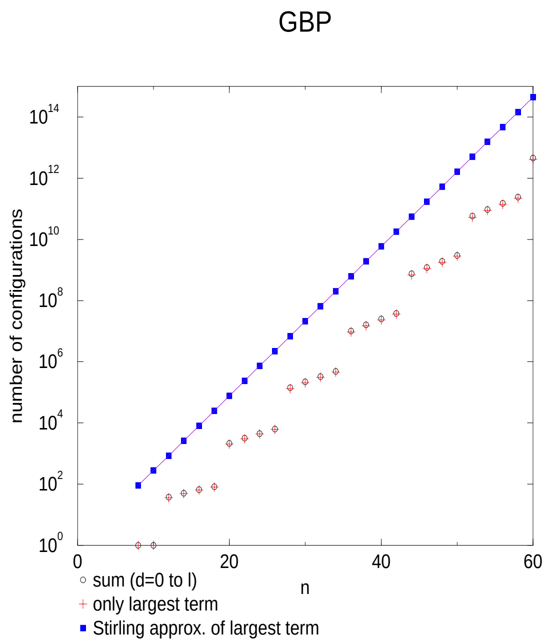

Figure 17: Number of configurations in a patch of radius l (and approximations) vs. system size for the GBP.

Figure 18 compares the probability of finding a local optimum at random with the Stirling approximation for the number of configurations in a ball of radius correlation length.

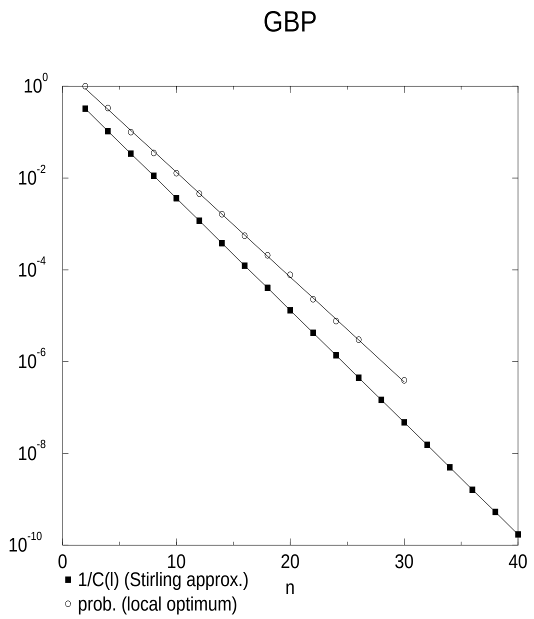

Figure 18: Density of local optima and $1/C_l$ (Stirling approximation of largest term) for the GBP, exchange metric. There is $O(1)$ local optimum in a ball of radius correlation length.

5.2.4 Adaptive and gradient walks

The lengths of adaptive and gradient walks scale linearly with system size. Adaptive walks are about three times as long as gradient walks. Figure 19.

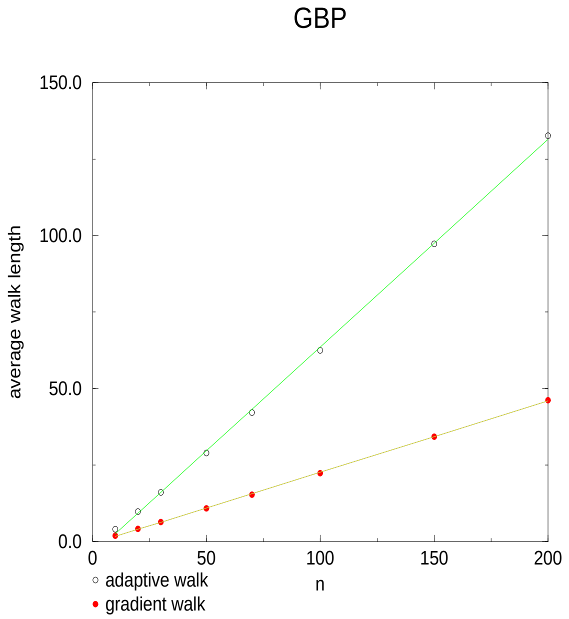

Figure 19: Average lengths of adaptive and gradient walks vs. system size (number of vertices) for the symmetric GBP, move set exchange.

The values of the cost function at local minima reached after adaptive and gradient walks are the same and scale linearly with the squared system size n^2. Minima reached at the end of a walk are only slightly better than minima found by random sampling (Figure 20).

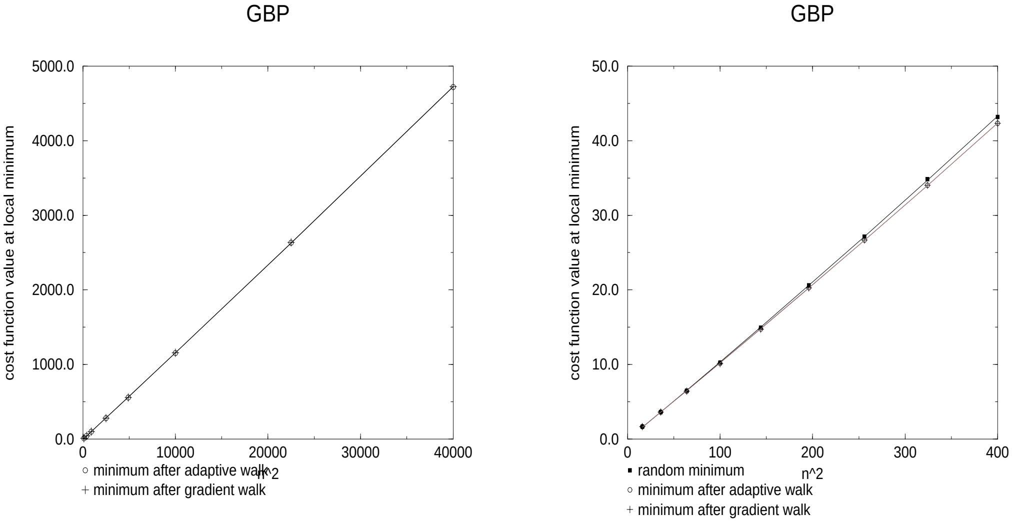

Figure 20: Average cost function values at minima attained after adaptive and gradient walks vs. system size (squared number of vertices) and at randomly found local minima for the GBP, move set exchange.

For constant system size $n = 150$, move set exchange, the following observations can be made: The longer the walk, the better the optimum found by this walk, and the larger the difference in energy to the start point of the walk, but also the worse the starting point. Figure 21.

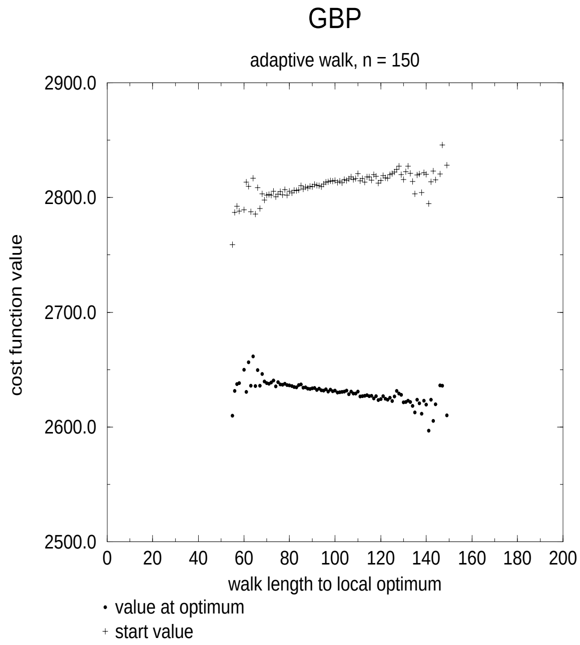

Figure 21: Average values of the cost function of starting configurations and of local optima attained after adaptive walks vs. walk length to the local optimum for the symmetric GBP, move set exchange, n = 150. 10 000 walks were carried out, 10 walks per instance.

The cost function value at the end of an adaptive walk depends on the quality of the configuration from which that walk started.

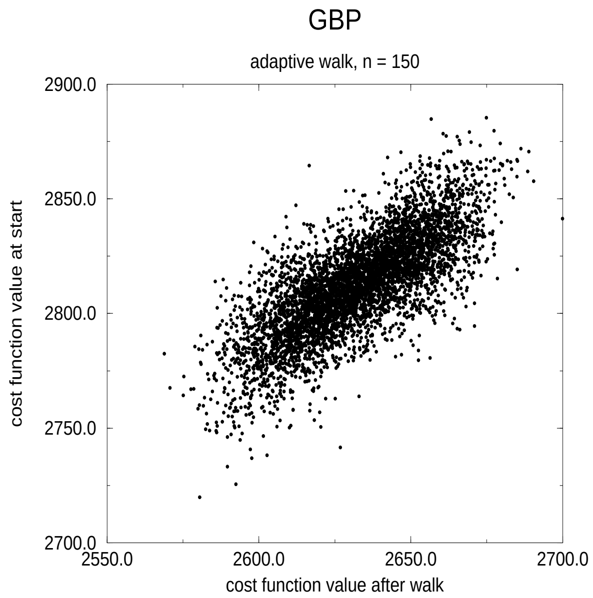

Figure 22: Average values of the cost function at the starting point vs. values at the end of an adaptive walk on the GBP landscape, n=150, for move set exchange.

5.3 The Graph Matching Problem

5.3.1 The autocorrelation function for the Graph Matching Problem

The autocorrelation function along a random walk has been computed by Stadler [25].

5.3.2 Number of local optima

A configuration has $n(n-2)/4$ nearest neighbours, $d_H = 4$. There are $n(n-1)/2$ neighbouring permutations, but $n/2$ of them are neutral, and of the remaining $n(n-2)/2$ neighbours each permutation has a corresponding permutation coding for the same configuration, thus the number of nearest neighbours has to be divided by two. A transposition exchanging the positions of vertices i and j in the list of numbers that encodes the configuration affects two bonds : the bond between i and its partner $i + n/2$ (if $i < n/2$, otherwise $i - n/2$) and the bond between j and its partner. This move is equal to the transposition exchanging the positions of the partners of i and j. Figure 23 shows the probability of randomly finding a local optimum.

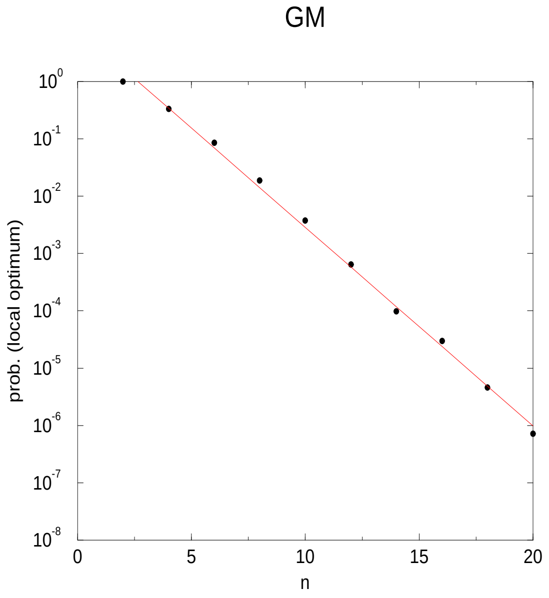

Figure 23: Probability of finding a local minimum by random sampling vs. system size (number of vertices) for the Graph matching problem, move set transposition.

5.3.3 Number of configurations in a patch of radius correlation length

The correlation lengths for the GM problem, move set transposition, have been computed from the nearest neighbour correlation (non-neutral move) by Stadler [25]

$$l(n) = \frac{n}{4} - 0.6.$$

One can try to estimate the number $DS(n, d)$ of configurations with distance d from any configuration as for the TSP, moveset transposition. Thus an upper bound for $DS(n,d)$ is given by:

$$DS(n,d) = ss(n, n-d).$$

The number of configurations in a ball of radius correlation length is

$$C_l \leq \sum_{d=0}^{n/4} ss(n, n-d).$$

As one can see in the left part of figure 24 the probability for finding a local optimum at random is much higher than the estimate. This fact does not contradict the conjecture by Stadler, which can be explained as follows:

First of all the number of different non-neutral neighbours is much lower (less than half) than in the TSP case. The number of permutations P_{id} of n numbers representing one configuration is

$$P_{id} = \left(\frac{n}{2}\right)! 2^{n/2}.$$

When a non-neutral nearest neighbour of a configuration is reached (by performing a transposition), the next transposition can lead to a permutation that has transposition distance $d = 2$ from the starting point, but is coding for the same configuration as this starting point.

If the crude estimate for $DS(d)$ is corrected by a factor 2^d for the identical neighbours, the value of $1/C_l$ (corrected) fulfills our hopes a little bit better, as one can see in the right part of figure 24.

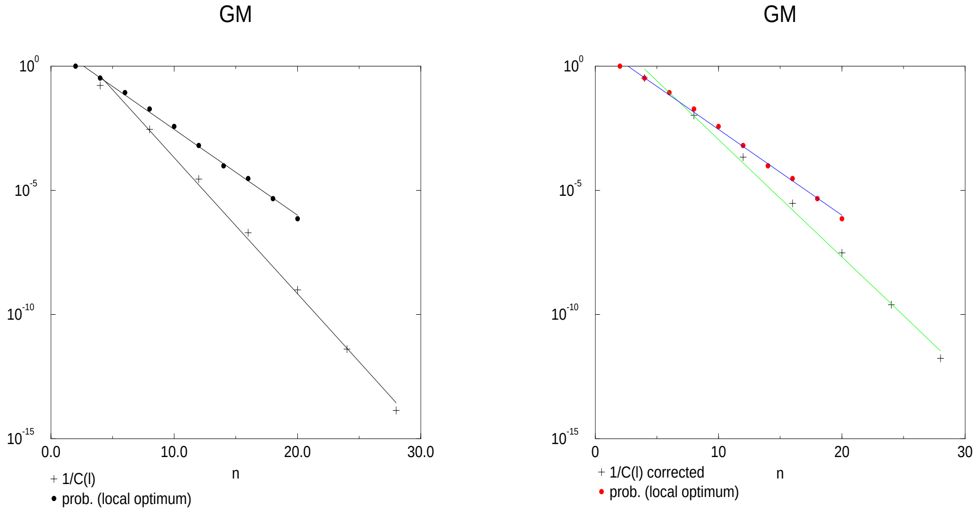

Figure 24: Density of local optima and $1/C_l$ for the GM problem, move set transposition.

5.3.4 Adaptive and gradient walks

Adaptive walks are 2.6 times longer than gradient walks (Figure 25).

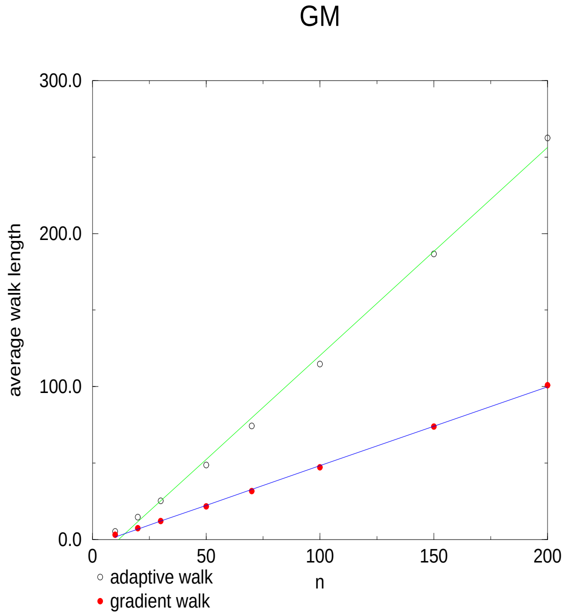

Figure 25: Average lengths of adaptive and gradient walks vs. system size (number of vertices) for the Graph matching problem, move set transposition. At least 10 000 walks were performed per system size, 1 walk per instance.

Gradient walks find slightly better optima (Figure 26).

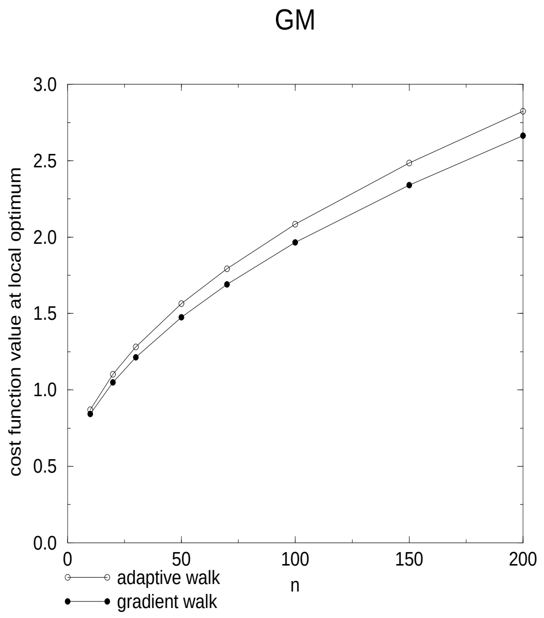

Figure 26: Average values of the cost function at local minima attained after adaptive and gradient walks vs. system size for the Graph matching problem, move set transposition.

Optima reached after a walk are much better than random optima. Figure 27.

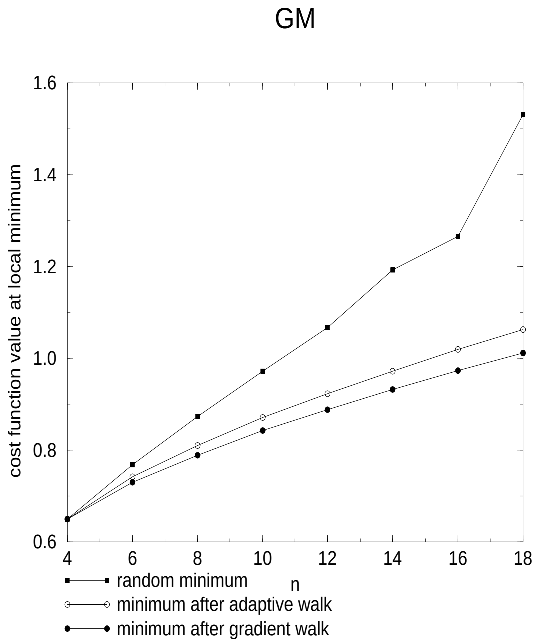

Figure 27: Average cost function values at randomly found local minima and at minima attained after walks vs. system size (number of cities) for the Graph matching problem, move set transposition.

The longer a walk, the better the optimum it finds, and the worse the starting solution. Figure 28.

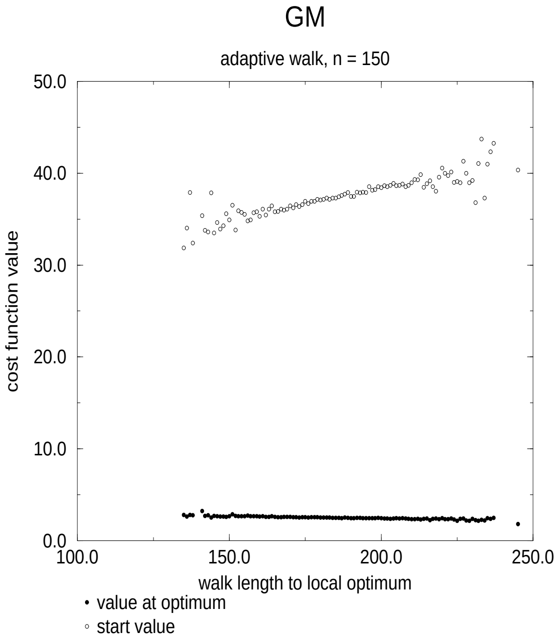

Figure 28: Average values of the cost function of starting configurations and of local optima attained after adaptive walks vs. walk length to the local optimum for the Graph matching problem, move set transposition, n = 150. 10 000 walks were carried out, 10 walks per instance.

The quality of the solution reached after a walk does not depend on the starting configuration. Figure 29.

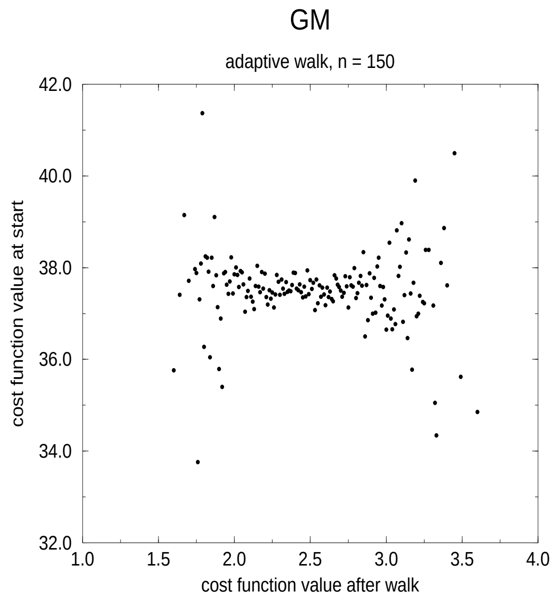

Figure 29: Average values of the cost function at the starting point vs. values at the end of an adaptive walk on the GM landscape, n=150, for move set transposition.

5.4 The Sherrington-Kirkpatrick spin glass

5.4.1 The autocorrelation function for the SK spin glass

The autocorrelation function for the SK spin glass is known to be [30]

$$\rho(d) = 1 - 4\frac{d}{n-1} + 4\frac{d^2}{n(n-1)}$$

with correlation length $l = n/4$.

5.4.2 Number of local optima

A configuration has n nearest neighbours with hamming distance $d_H = 1$. Figure 31 shows the probability of finding a local optimum at random.

5.4.3 Number of configurations in a patch of radius correlation length

The number DS of configurations with distance d from any configuration is given by

$$DS(d) = \binom{n}{d}.$$

The number of configurations C_l in a patch of radius correlation length $l = n/4$ is thus given by:

$$C_l = \sum_{d=0}^{n/4} \binom{n}{d}$$

This sum can be approximated by its largest term

$$C_l = \sum_{d=0}^{l} \binom{n}{d} \simeq \binom{n}{l}$$

(the sum is 1.45 times the largest term for $n = 100$).

Set $l = xn$, then the largest term of the sum can be rewritten as

$$\binom{n}{xn} = \frac{n!}{(xn)!\,(n-xn)!} = \frac{n!}{(xn)![(1-x)\,n]!}$$

By using Stirling's formula this can be rewritten as

$$\begin{aligned}
\frac{n!}{(xn)![(1-x)\,n]!} &\simeq \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{xn}{e}\right)^{xn} \sqrt{2\pi xn} \left(\frac{(1-x)n}{e}\right)^{(1-x)n} \sqrt{2\pi\,(1-x)\,n}} = \\
&= \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{x^{xn} \left(\frac{n}{e}\right)^{xn} (1-x)^{(1-x)n} \left(\frac{n}{e}\right)^{(1-x)n} 2\pi n\sqrt{x\,(1-x)}} = \\
&= \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{n}{e}\right)^n (1-x)^n \left(\frac{x}{1-x}\right)^{xn} 2\pi n\sqrt{x\,(1-x)}} = \\
&= \frac{1}{(1-x)^n \left(\frac{x}{1-x}\right)^{xn} \sqrt{2\pi n}\sqrt{x\,(1-x)}} = \\
&= \left[\left(\frac{1-x}{x}\right)^x \frac{1}{1-x}\right]^n \frac{1}{\sqrt{2\pi x\,(1-x)}}\frac{1}{\sqrt{n}} \simeq \\
&\simeq \left[\left(\frac{1-x}{x}\right)^x \frac{1}{1-x}\right]^n
\end{aligned}$$

Figure 30 shows the quality of the approximations.

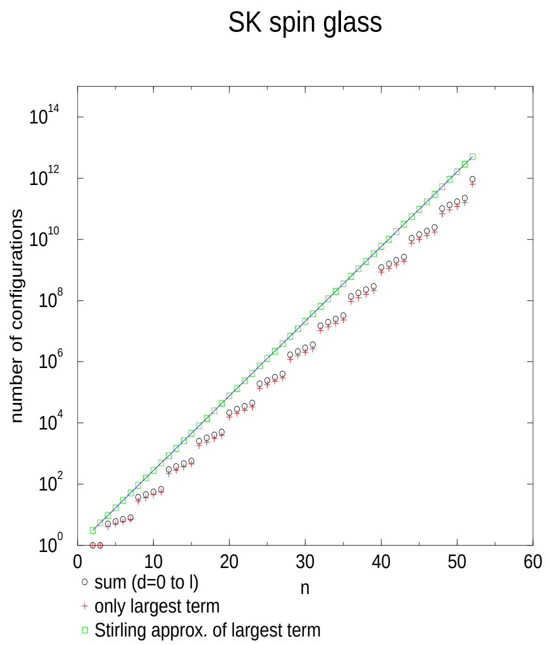

Figure 30: Number of configurations in a patch of radius l (and approximations) vs. system size for the SK spin glass.

Figure 31 compares the probability of finding a local optimum at random with the number of configurations in a ball of radius correlation length.

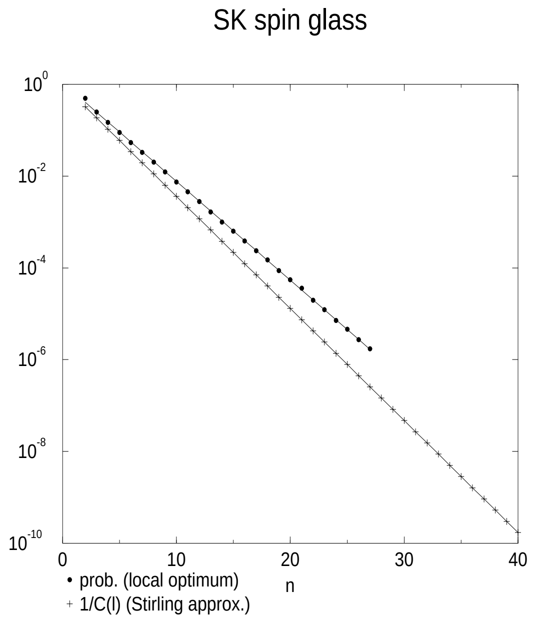

Figure 31: Density of local optima and $1/C_l$ (Stirling approximation of largest term) for the SK spin glass, move set single spin flip.

5.4.4 Adaptive and gradient walks

Adaptive walks are nearly twice as long as gradient walks. Figure 32.

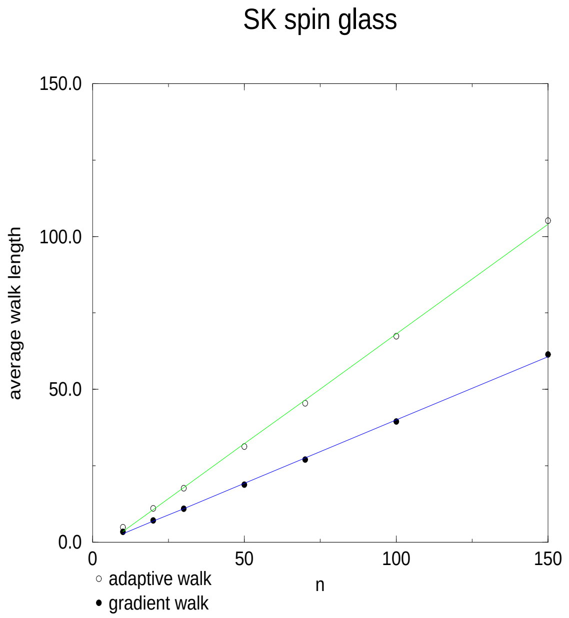

Figure 32: Average lengths of adaptive and gradient walks vs. system size for the SK spin glass, move set single spin flip. At least 10 000 walks were performed per system size, 1 walk per instance.

The cost function values at optima reached after adaptive and gradient walks are nearly the same (Figure 33). The lines in the plot show the least square quadratic fits:

$$f_{adap} = 16.98 - 2.108n - 0.018n^2$$
$$f_{grad} = 11.22 - 1.768n - 0.021n^2$$

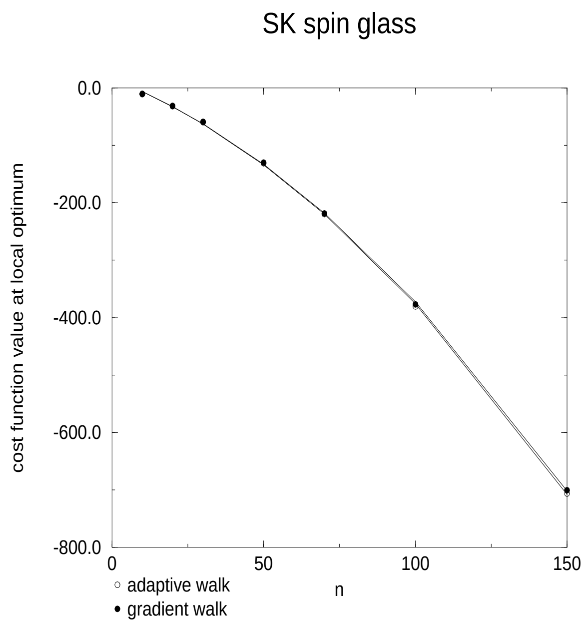

Figure 33: Average values of the cost function at local minima attained after adaptive and gradient walks vs. system size for the SK spin glass, move set single spin flip.

Optima reached at the end of a walk are better than randomly found optima. Figure 34.

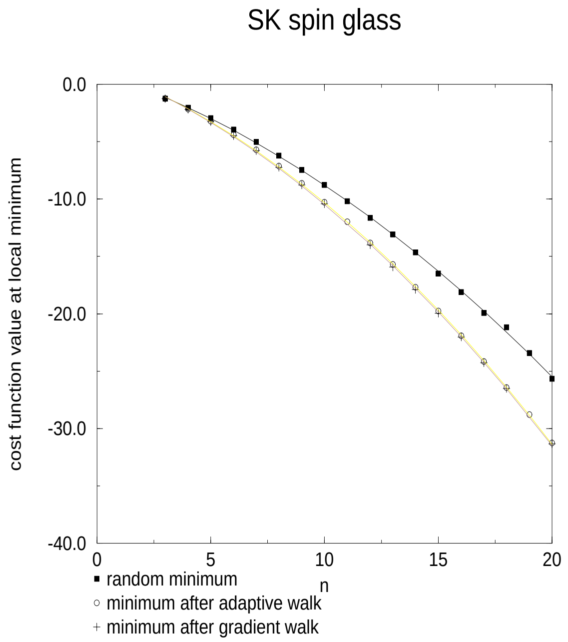

Figure 34: Average cost function values at randomly found local minima and at minima attained after walks vs. system size for the SK spin glass, move set single spin flip.

Longer walks find deeper optima and start at worse solutions. Figure 35.

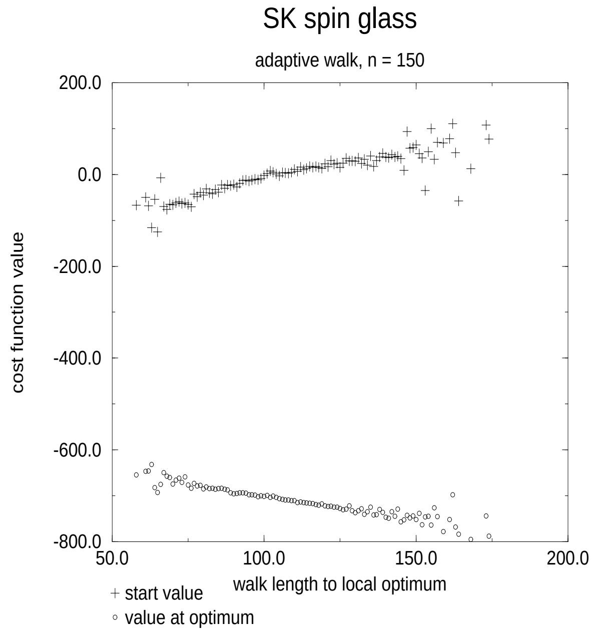

Figure 35: Average values of the cost function of starting configurations and of local optima attained after adaptive walks vs. walk length to the local optimum for the SK spin glass, move set single spin flip, n = 150. 10 000 walks were carried out, 10 walks per instance.

5.5 Comparison

Table 2 shows the least square exponential fits for the probability of hitting a local optimum at random vs. system size n:

$$\text{prob. (local optimum)} = b_{opt} e^{-c_{opt} n} = b_{opt} A_{opt}^{-n}$$

and the corresponding data c_{conf} and A_{conf} for the fits of the plots $1/C_l$ vs. n. For the sTSP, the GBP and the SK spin glass the Stirling approximation of the largest term was taken as value of C_l, for the GM problem the corrected value of C_l was used.

The lengths of adaptive and gradient walks scale linearly with system size. The least square fits of the plots are compared in table 3.

The landscapes of the sTSP, the aTSP and the GM problem with move set transposition are based on the same underlying configuration space. The probability of hitting a local optimum at random is highest in the GM landscape, since the number of non-neutral different neighbours is $n(n-2)/4$; a transposition exchanges 2 bonds out of $n/2$. For the sTSP and aTSP the number of nearest neighbours is $n(n-1)/2$, hence the probability that a configuration is a local optimum is lower. The difference in the number of local optima between these two landscapes arises from the fact that in the asymmetric case more intercity connections are exchanged by a single move (see chapter 5.1.2).
The lengths of adaptive walks are roughly the same for sTSP and aTSP; adaptive walks on the GM landscape are about 12 percent shorter. The lengths of gradient walks are roughly the same on all three landscapes.

The investigated landscapes share some features with the RNA minimum free energy landscape; the lengths of adaptive and gradient walks scale linearly with system size, as does the correlation length.

In all landscapes for which the number of configurations in a ball of radius correlation length was computed, O(1) local optima were found in this patch.

Problem	move set	b_{opt}	c_{opt}	A_{opt}	c_{conf}	A_{conf}
symmetric TSP	Transposition	27.667	1.163	3.199	1.428	4.169
	Inversion	151.130	1.505	4.504		
	can. Transposition	1.090	0.487	1.628		
asymmetric TSP	Transposition	11.290	1.016	2.762		
	Inversion	2.137	0.591	1.806		
	can. Transposition	1.076	0.414	1.513		
GBP	Exchange	2.562	0.526	1.692	0.562	1.755
Graph matching	Transposition	8.292	0.797	2.219	1.089	2.972
SK spin glass	spin flip	1.099	0.496	1.643	0.562	1.755
RNA mfe GCAU	point mutation		> 0.3	> 1.35	0.814	2.257

Table 2: Regression data for probabilities of hitting local optima at random.

In the RNA mfe landscape the length of a gradient walk is equal to the correlation length; this is not true for the other landscapes except for the aTSP with move set canonical transposition.

Problem	move set	l/n	L_{grad}/n	L_{adapt}/n	A_{opt}
symmetric TSP	Transposition	0.25	0.538	1.553	3.199
	Inversion	0.5	1.032	2.931	4.504
	canonical Transposition	0.5	0.390	0.493	1.628
asymmetric TSP	Transposition	0.25	0.514	1.527	2.762
	Inversion		0.334	0.689	1.806
	canonical Transposition	0.33	0.329	0.425	1.513
GBP	Exchange	0.125	0.233	0.679	1.691
Graph matching	Transposition	0.25	0.517	1.359	2.219
SK spin glass	spin flip	0.25	0.414	0.718	1.643
n-k model for GCAU from [7]	spin flip		0.32	0.60	1.82
RNA mfe GCAU	point mutation	0.26	0.26	0.44	> 1.35

Table 3: Correlation lengths, walk lengths and A for combinatorial optimization problems. The RNA mfe data are taken from [7].

6 Conclusion and outlook

6.1 Conclusion

The landscapes arising from five different combinatorial optimization problems (sTSP, aTSP, GBP, GM, SK spin glass) were explored by performing adaptive and gradient walks; the number of local optima on these landscapes was obtained by random sampling.

The results can be summarized as follows:

- The probability of hitting a local optimum at random scales down with system size as a decaying exponential on all landscapes investigated.
- There are roughly O(1) local optima in a ball of radius correlation length in configuration space.
- The lengths of adaptive and gradient walks scale linearly with system size.
- Local optima reached at the end of adaptive and gradient walks have very similar cost function values and are better than randomly found optima. Adaptive and gradient walks do not reach the global optimum.
- Longer walks find better optima.
- The correlation between cost function values at the start and end of walks depends on the quality of the neighbourhood. If a good move set is chosen, the cost function value at a local optimum reached after a walk should not depend on the quality of the starting configuration.

These results confirm the following conjectures by Stadler: Optimization by algorithms is easier when the correlation increases, since the number of local optima, which are traps for algorithms, decreases with increasing correlation length. A move set should be chosen which maximizes correlation length while keeping constant the number of nearest neighbours.

6.2 Outlook

By performing simple local search procedures such as adaptive and gradient walks one can derive important statistical features of landscapes arising from combinatorial optimization problems.

As far as estimates for the cost function value of the global optimum are known, adaptive and gradient walks clearly do not reach the latter.
Other, more sophisticated algorithms should not only find better solutions, but their performance on different landscapes might also provide deeper insight into the underlying structures of these landscapes.

7 Data Tables

		k									
		1	2	3	4	5	6	7	8	9	10
n	1	1									
	2	1	1								
	3	2	3	1							
	4	6	11	6	1						
	5	24	50	35	10	1					
	6	120	274	225	85	15	1				
	7	720	1764	1624	735	175	21	1			
	8	5040	13068	13132	6769	1960	322	28	1		
	9	40320	109584	118124	67284	22449	4536	546	36	1	
	10	362880	1026576	1172700	723680	269325	63273	9450	870	45	1

Table 4: Signless Stirling numbers $ss(n,k)$

References

[1] Bafna and Pevzner. Unpublished preprint.

[2] C. K. Biebricher, M. Eigen, and R. Luce. Kinetic analysis of template instructed and *de novo*RNA synthesis by $q\beta$-replicase. *Journal of Molecular Biology*, 148:391, 1981.

[3] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 10:465–523, 1971.

[4] W. Fontana, T. Griesmacher, W. Schnabl, P. Stadler, and P. Schuster. Statistics of landscapes based on free energies, replication and degredation rate constants of RNA secondary structures. *Monatshefte der Chemie*, 122:795–819, 1991.

[5] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaption. *Physical Review A*, 40(6):3301–3321, Sep. 1989.

[6] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophysical Chemistry*, 26:123–147, 1987.

[7] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. RNA folding and combinatory landscapes. *Phys. Rev. E*, 47(3):2083 – 2099, March 1993.

[8] M. Garey and D. Johnson. *Computers and Intractability. A Guide to the Theory of $\mathcal{NP}$ Completeness.* Freeman, San Francisco, 1979.

[9] A. S. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J.Theor.Biol.*, 128:11–45, 1987.

[10] S. A. Kauffman. *The Origin of Order.* Oxford University Press, New York, Oxford, 1993.

[11] S. A. Kauffman and E. D. Weinberger. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.*, 141:211–245, 1989.

[12] E. L. Lawler, J. K. Lenstra, A. H. G. R. Kan, and D. B. Shmoys. *The Traveling Salesman Problem. A Guided Tour of Combinatorial Optimization.* John Wiley & Sons, 1985.

[13] S. Lin and B. Kernighan. An effective heuristic algorithm for the traveling salesman problem. *Oper. Res.*, 21:498–516, 1973.

[14] H. Lueneburg. *Tools and Fundamental Constructions of Combinatorial Mathematics.* BI-Wissenschaftsverlag,Mannheim, Wien, Zuerich, 1989.

[15] C. A. Macken, P. S. Hagan, and A. Perelson. Evolutionary walks on rugged landscapes. *SIAM J.Appl.Math.*, 51:799–827, 1991.

[16] J. Maynard-Smith. Natural selection and the concept of a protein space. *Nature*, 225:563–564, 1970.

[17] M. Mezard and G. Parisi. Replicas and optimization. *J. Physique Lett.*, 46:L–771 – L–778, 1985.

[18] M. Mezard and G. Parisi. A replica analysis of the travelling salesman problem. *J. Physique*, 47:1285–1296, 1986.

[19] R. Palmer. Optimization on rugged landscapes. In A. Perelson and S.A.Kauffman, editors, *Molecular Evolution on Rugged Landscapes: Proteins, RNA, and the Immune System*, pages 3–25. Addison Wesley, Redwood City (CA), 1991.

[20] C. H. Papadimitrou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, 1982.

[21] J. J. Rotman. *The Theory of Groups:An Introduction.* Allyn and Bacon,Inc., 1965.

[22] G. B. Sorkin. Combinatorial optimization, simulated annealing, and fractals. *IBM Research Report RC13674 (no. 61253)*, 1988.

[23] S. Spiegelman. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 17:213, 1971.

[24] P. Stadler and R.Happel. Canonical approximations of landscapes. SFI Preprint, 1994.

[25] P. F. Stadler. Correlation in landscapes of combinatorial optimization problems. *Europhys. Lett.*, 20:479–482, 1992.

[26] P. F. Stadler and R. Happel. Correlation structure of the landscape of the graph-bipartitioning-problem. *J. Phys. A.: Math. Gen.*, 25:3103–3110, 1992.

[27] P. F. Stadler and W. Schnabl. The landscape of the traveling salesman problem. *Phys. Letters A*, 161:337–344, 1992.

[28] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol.Cybern.*, 63:325–336, 1990.

[29] E. D. Weinberger. Local properties of kauffman's n-k model: A tunably rugged energy landscape. *Physical Review A*, 44(10):6399–6413, 1991.

[30] E. D. Weinberger and P. F. Stadler. Why some fitness landscapes are fractal. *J.Theor.Biol.*, 163:255–275, 1993.

[31] S. Wright. The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In D. F. Jones, editor, *int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.

[32] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bulletin of Mathematical Biology*, 46(4):591–621, 1984.

[33] M. Zuker and P. Stiegler. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

Curriculum vitae

Bärbel Krakhofer

$*$ 1966-12-17, Wien

1973-1977	:	Volksschule, Bruck/L. and Eisenstadt
1977-1985	:	Bundesgymnasium Eisenstadt
1985	:	Matura mit Auszeichnung
1985-1995	:	Studies of Biochemistry at the Universität Wien
10/1993-12/1994	:	Diplomarbeit with Prof. Peter Schuster at the Institute of Theoretical Chemistry, Universität Wien