European Journal of Operational Research 236 (2014) 78–84

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

In and out forests on combinatorial landscapes

A.R. Kammerdiner [a,*], E.L. Pasiliao [b]

[a] Department of Industrial Engineering, New Mexico State University, Las Cruces, NM 88003, United States
[b] Air Force Research Laboratory, Munitions Directorate, Eglin AFB, FL 32542, United States

ARTICLE INFO

Article history:
Received 8 May 2013
Accepted 20 November 2013
Available online 1 December 2013

Keywords:
Local search
Fitness landscape
Local optima
Directed graph
Out forest (spanning diverging forest)
In forest (spanning converging forest)

ABSTRACT

Fitness landscape theory is a mathematical framework for numerical analysis of search algorithms on combinatorial optimization problems. We study a representation of fitness landscape as a weighted directed graph. We consider out forest and in forest structures in this graph and establish important relationships among the forest structures of a directed graph, the spectral properties of the Laplacian matrices, and the numbers of local optima of the landscape. These relationships provide a new approach for computing the numbers of local optima for various problem instances and neighborhood structures.

1. Introduction

Combinatorial optimization is beneficial for minimizing transportation costs of different goods, finding protein folding structures having minimal energy, determining the best topology for a wireless network, or designing microprocessors to minimize the communication costs among individual transistors (Pardalos, Du, & Graham, 2013). Computational complexity is a common feature of many combinatorial optimization problems (e.g., the traveling salesman problem and the quadratic assignment problem are NP-hard). Another typical feature is a large size of the solution space for even problems of moderate size. For d-partite graph matching or the multidimensional assignment problem, for example, the number of feasible solutions grows exponentially in both dimensionality d and cardinality n of the problem.

Computational complexity and very large solution space make application of exact solution techniques restrictively slow or even infeasible for such problems. So, rather than obtaining globally optimal solution, the task is simplified to finding good quality approximate solutions. Search based methods, including iterative and stochastic local searches (Hirsch, Pardalos, & Resende, 2010; Hoos & Stützle, 2005; Stützle, 2006), simulated annealing (Kirkpatrick, Gelatt, & Vecchi, 1983; Pedamallu & Ozdamar, 2008), and genetic algorithm (Gonçalves, Mendes, & Resende, 2008; Holland, 1975), are widely utilized for finding approximate solutions of discrete and combinatorial optimization problems (Hvattum & Glover, 2009). A neighborhood of a given solution is a key concept in search algorithms. It is through the notion of neighborhood that the distance between two solutions, as well as the local optima (minima and maxima) are defined.

Fitness landscape theory is an important subfield of combinatorial optimization. It provides a framework for analysis of the search algorithm's behavior on different instances of a given problem (Krokhmal & Pardalos, 2009; Stadler, 2002). In addition to combinatorial optimization, fitness landscape analysis found application in the physics of disordered systems and evolutionary biology (Reidys & Stadler, 2002). As a matter of fact, fitness landscapes originated in theoretical biology as a way of visualizing evolutionary adaptation (Wright, 1932). Later, they appeared in analysis of spin glasses and other models of disordered physical systems (Binder & Young, 1986). Reidys and Stadler (2002) also acknowledge the similarity between fitness landscapes and the potential energy surfaces used to study the folding of biopolymers (e.g., nucleic acids and proteins) in theoretical chemistry (Mezey, 1987; Neumaier, 1997).

Combinatorial landscapes are not unlike their earth counterparts. If solutions are thought as places, and a solution's fitness (i.e., objective value) as an altitude, then local minima and maxima are like depths and peaks. Just like a walk is long if two places are far apart, the search algorithm's run time is long when the distance between initial solution and resulting local optimum is large. Similarly to difficulties encountered traversing a rugged terrain, search algorithms are challenged by combinatorial landscapes with many local optima.

As a result, a number of fitness landscape studies examine the trends in the numbers of local optima for various problem sizes, neighborhoods, and distributions of coefficients. Other studies look

* Corresponding author. Tel.: +1 115756462475.
E-mail address: alla@nmsu.edu (A.R. Kammerdiner).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.11.025

at the relationship between solution's fitness and its distance to the optima. It is widely acknowledged that careful analysis of combinatorial landscapes is a key to a deeper understanding of the behavior of the search based algorithms (Schiavinotto & Stützle, 2007).

Many important results and methodologies for fitness landscape analysis are reviewed in Reidys and Stadler (2002) and Stadler (2002). Combinatorial landscapes can be examined through a wide variety of perspectives, including probability theory and stochastic processes, dynamical systems, random graphs, and abstract algebra. In the majority of the literature on fitness landscapes, graph representations of landscapes and the respective discrete Laplacian operators (or graph Laplacians) arise in the context of undirected graphs. Undirected graphs assume the presence of edges in both directions, and so, they lead to symmetric graph Laplacians.

We argue that landscape representations via undirected graphs ignore or lose important information on which of the two solutions (edge endpoints) has a better fitness value. Yet, the theory of directed graphs (digraphs) is not nearly as well developed as for undirected graphs. Some key results in undirected graph theory do not hold for directed graphs and need to be extended for digraphs. Fortunately, a well-known algebraic result connecting the number of weak components with the multiplicities of undirected graph Laplacians was extended to digraphs in Agaev and Chebotarev (2005). It was shown that converging trees in a digraph form a directed forest known as in forest, and this structure serves as a digraph analog of an undirected graph's decomposition into weak components.

Here we extend this important result to alternative types of directed forests and their components (specifically, out forests and diverging trees). The main contribution of the paper is the fitness landscape representation via directed graph and the important connections between local optima on landscapes and certain nodes on directed trees. The connection between landscapes' optima and directed trees is visualized in Fig. 1, which serves as a graphical abstract.

The paper is organized as follows. The key concepts related to fitness landscapes and forest structures on directed graphs are introduced in Section 2. These definitions are used to obtain general results on digraphs and landscapes in Section 3. Finally, Section 4 presents the conclusions.

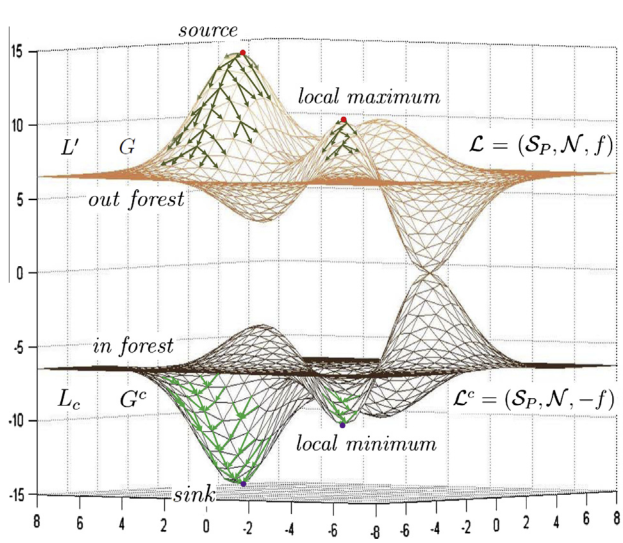

Fig. 1. Graphical abstract.

2. Notation

2.1. Combinatorial landscapes

Fitness landscape analysis is an important tool in combinatorial optimization. A search algorithm imposes a structure on the space of all feasible solutions. Landscapes theory studies how the imposed structure impacts the behavior and performance of the algorithm. To define a fitness landscape, we first introduce several related concepts, starting with a notion of a neighborhood of a given solution.

In this paper, we follow the terminology of (Schiavinotto & Stützle, 2007; Stadler, 2002). Suppose that $\mathcal{P}$ is a combinatorial optimization problem (e.g., the quadratic assignment problem or the traveling salesman problem). Let P be a specific instance of $\mathcal{P}$, in other words $P \in \mathcal{P}$. Denote by $\mathcal{S}_P$ the set of all feasible solutions of instance P. Then $\mathbf{2}^{\mathcal{S}_P}$ represents the space of all subsets of set $\mathcal{S}_P$.

A *neighborhood* is a mapping $\mathcal{N} : \mathcal{S}_P \to \mathbf{2}^{\mathcal{S}_P}$ that assigns to each solution $s \in \mathcal{S}_P$ its subset $\mathcal{N}(s) \in \mathbf{2}^{\mathcal{S}_P}$ of solutions, called *neighbors*.

Usually, it is assumed that $s \notin \mathcal{N}(s)$. For a neighborhood to be useful in search landscape analysis, the way a mapping $\mathcal{N}$ is defined must correspond to the "look around" step the search algorithm makes as it moves through the solution set $\mathcal{S}_P$. Hence, a neighborhood should be defined through an operator that describes all allowable moves for a single step of the algorithm.

An *operator* Δ is a set of the operator functions $\delta : \mathcal{S}_P \to \mathcal{S}_P$ that satisfy

$$s_0 \in \mathcal{N}(s) \text{ if and only if } \exists \delta \in \Delta \text{ such that } s_0 = \delta(s).$$

To ensure that $s \notin \mathcal{N}(s)$, it is usually assumed that $\delta(s) \neq s$ for all $s \in \mathcal{S}_P$. The application of some operator function $\delta \in \Delta$ is called a *move*.

A distance between two solutions is another key concept closely related to the ideas of a neighborhood, an operator, and a move. When the algorithm moves from one solution to another in search for optimal solutions, it crosses multiple neighborhoods via successive application of a sequence of moves. In other words, given any two solutions $s, t \in \mathcal{S}_P$, the search starting in s will arrive in t if and only if $\exists \{\delta_1, \ldots, \delta_k\} \subset \Delta$ such that $t = \delta_k(\ldots(\delta_1(s))\ldots)$. Notice that the number of neighborhoods crossed by applying the above sequence of moves is k. Clearly, alternative ways to move from one solution to another might exist. In such case, the number of neighborhoods the search crosses on its way from s to t might vary. The smallest possible number defines a distance between s and t.

Typically, the neighborhood $\mathcal{N}$ is defined in such a way that $s_1 \in \mathcal{N}(s_2) \iff s_2 \in \mathcal{N}(s_1)$ (i.e., it is commutative). As a result, we can get from t to s by simply reversing the sequence of moves, and so, the distance from t to s is the same as the distance from s to t. Thus, to guarantee that the distance is symmetric, the considered neighborhood must be commutative.

Suppose that $\mathcal{N}$ is, in fact, commutative. Then a distance between solutions of P can be formally defined as follows.

Given an operator Δ, a *distance* between two solutions $s, t \in \mathcal{S}_P$ is the minimum number of moves $\delta \in \Delta$ needed to reach s from t or t from s, i.e.,

$$d(s,t) = \min\{k : t = \delta_k(\ldots(\delta_1(s))\ldots) \text{ or } s = \delta_1(\ldots(\delta_k(t))\ldots), \delta_i \in \Delta \, \forall i\}$$

The neighborhood together with the corresponding operator and distance impose a certain structure on the solution space. A fitness landscape (sometimes also referred as combinatorial or search landscape) adorns the solution space structure of a problem instance P established by the choice of neighborhood $\mathcal{N}$ with the objective (or *fitness*) function $f : \mathcal{S}_P \to \mathbb{R}$ of P.