Chapter 1

Introduction

The shortest path problem is one of the fundamental problems in computer science. It has been well studied and efficient polynomial time algorithms are known. Shortest path problems frequently arise in practice since in a variety of application settings we wish to send some material (*e.g.*, a data packet, a telephone call, or a vehicle) between two specified points in a network as quickly, as cheaply, or as reliably as possible.

However, in a practical setting we are often not only interested in a cheapest path or a quickest path but rather in a combination of different criterias, *e.g.* we want to have a path that is both cheap and quick. This is known as the bi- or multicriteria shortest path problem. Since optimizing over all criteria at once is not possible we choose one criteria as the cost function that we want to minimize, the others as resource functions and impose resource (or budget) limits on the maximal resource consumption of a path. The *constrained shortest path problem* is to find a minimum cost path between two nodes whose resource consumptions satisfy the resource limits.

Applications

The constrained shortest path problem has a large number of practical applications:

The first application that comes to mind is *route planning* in traffic networks. We want to go from A to B and for example want to minimize the possibility of traffic congestion while imposing a length constraint on the path (see Figure 1.1). Alternatively, we want to go from A to B as fast as possible but have budget constraints on fuel consumption and road fees.

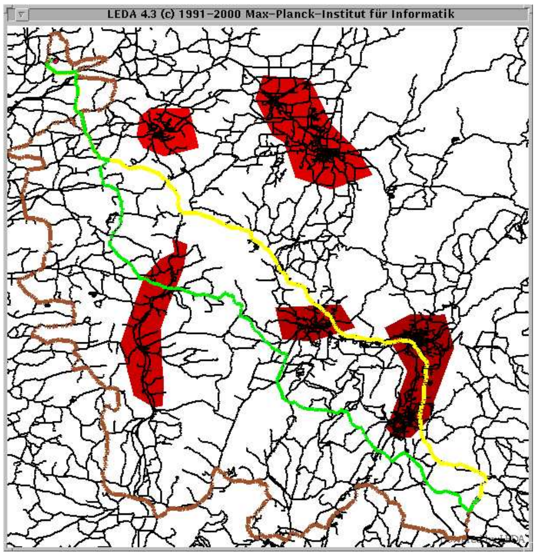

Figure 1.1: Minimum congestion path satisfying length constraint (areas of congestion are shaded). The minimum cost path is brown, the minimum resource path yellow and the constrained shortest path green.

In communication networks, we have the *quality of service (QoS)* routing problem. We are searching for a path of minimum costs that obeys given delay or reliability bounds. This is again a constrained shortest path problem. Further information can be found in the papers of Orda (1998) and Xue (2000).

There are also other applications that can be modeled as constrained shortest path problems:

Elimam and Kohler (1997) show how to model two engineering applications as (multiple resource) constrained shortest path problems: optimal sequences for the treatment of wastewater processes and minimum cost energy-efficient composite wall and roof structures.

Dahl and Realfsen (2000) and Nygaard (2000) studied the *linear curve approximation* problem and showed how to model this as a constrained shortest path problem (see Section 3.5). They find a path that corresponds to a minimal error approximation using a limited number of breakpoints (see Figure 1.2). Applications are data compression in areas like cartography, computer graphics, and image processing. In the area of traffic and communication networks this is known as routing problem with *hop constraints*, *i.e.*, a limit on the number of path links.

There are several applications in operations research that involve constrained shortest paths as a subproblem in column generation methods:

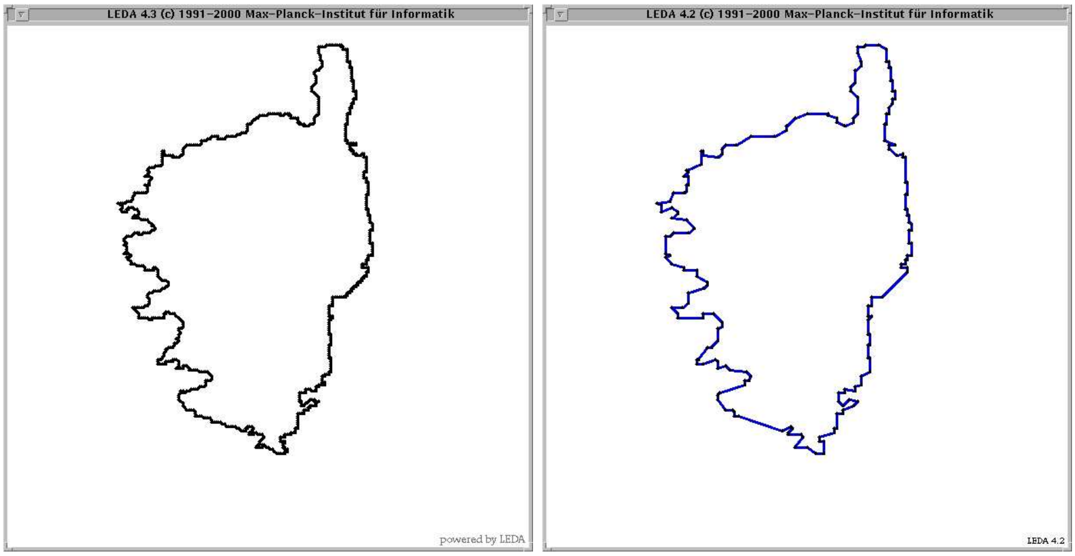

Figure 1.2: Coastline of Corsica (800 points) and minimum error approximation using only 200 points

Borndörfer and Löbel (2001) consider the *duty scheduling* problem and propose an adaptive column generation algorithm that needs to solve multiple resource constrained shortest paths as a subproblem.
Jahn, Möhring, and Schulz (1999) study a *route guidance* problem. They consider the optimal routing of traffic flows with length restrictions in road networks with congestion, and also propose a column generation method with constrained shortest paths as subproblem.
A similar problem in the QoS area was investigated by Holmberg and Yuan (1997) who also need constrained shortest paths as a subproblem in a column generation approach.
Lübbecke and Zimmermann (2000) proposed a column generation method for the *scheduling of switching engines* that also needs to solve constrained shortest paths as subproblem.

We see that the constrained shortest path problem is of immense practical interest in different areas of operations research.

Unfortunately, the introduction of even a single resource constraint turns the problem into a *hard* problem where we do not know a polynomial time algorithm to solve it. However, regarding its huge practical importance we would still like to solve the problem (or at least get an approximation) as efficiently as possible.

In this thesis we study the constrained shortest path problem both theoretically and experimentally. We also consider related problems like constrained minimum span-

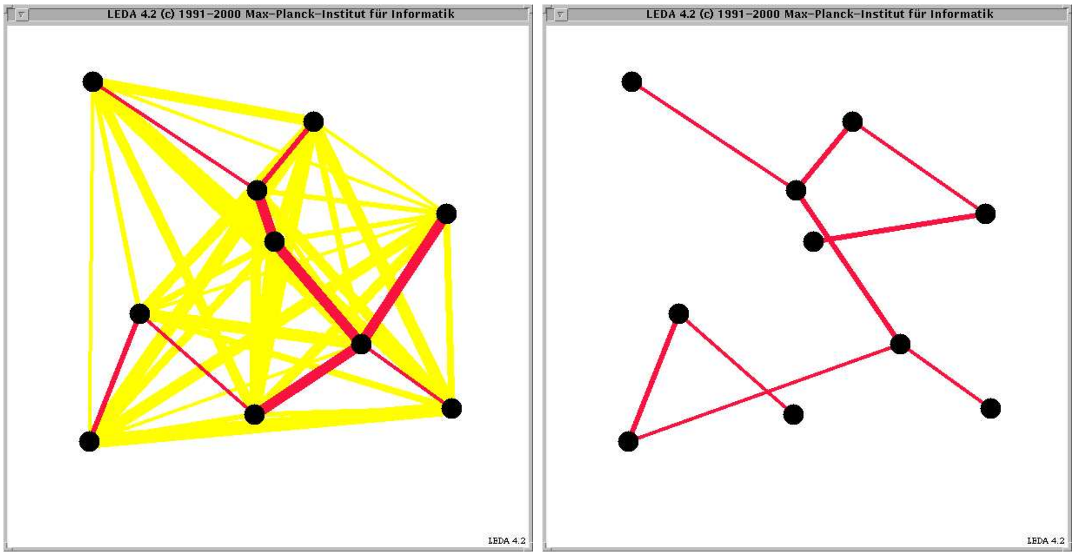

Figure 1.3: Minimum cost spanning tree and Minimum cost reliability constrained spanning tree. Width of edges corresponds to fault probability.

ning trees where we want to find a spanning tree of minimal cost while its resource consumptions satisfy the resource limits (see Figure 1.3).

Our Contribution

There is a variety of work on constrained shortest paths coming from different communities like operations research, algorithms, communication networks, and even signal processing. Almost all papers come up with essentially the same algorithm solving a relaxation of the problem for the single resource case. They only differ in the presentation of the method. Some derive it from geometric intuition, others adopt the Lagrangean relaxation viewpoint. Starting from a new ILP formulation of the problem, we will combine geometric intuition and linear programming theory to obtain a unified understanding of the method. Using this combined view, we are the first to prove a tight polynomial runtime bound for this method. We will show that the relaxation can be solved with $O(\log(nRC))$ parametric shortest path computations, where n is the number of nodes in the network and C and R denote the maximal cost and resource consumption of an edge, respectively. Our reformulation also allows us to extend the method to the multiple resource case, which has been an open problem up to now.

Solving the relaxation gives us upper and lower bounds for our problem. Previous papers suggested different gap closing steps to obtain a 2-step method for constrained shortest paths. We again give a geometric intuition of the gap closing step and propose a special labeling approach to close the gap.

Then we experimentally compare all state of the art methods for constrained shortest paths on different benchmarks. This is the first detailed experimental runtime comparison for constrained shortest paths.

We then show that the 2-step method can be generalized to a broader class of constrained network optimization problems. All we need is a function returning the unconstrained optimum and a function ranking solutions. The single resource runtime bound for the relaxation extends. We illustrate the generic method using three examples: constrained minimum spanning trees, table layout, and constrained geodesic shortest paths.

We have developed a software package CNOP that implements the generic 2-step approach. A user only has to specify a function solving the corresponding unconstrained problem and a function ranking problem solutions. Additionally, CNOP offers all state of the art methods for constrained shortest paths and can be used as a testbed to see which approach is most suited for a special application. While several implementations of different methods exist, this is the first package that makes all state of the art methods publically available. It also offers the first publicly available implementation for the constrained minimum spanning tree problem. The flexibility of the CNOP package allows the user to experiment with other bi- or multicriteria network optimization problems.

We presented our results at the 8th Annual European Symposium on Algorithms 2000 in Saarbrücken and the 3rd Workshop on Algorithm Engineering and Experiments 2001 in Washington, DC.

Outline

We first review some basic notation, central algorithms and techniques in Chapter 2. Chapter 3 deals with the constrained shortest path problem. We propose a new relaxation formulation that allows us to derive simple combinatorial algorithms to solve the relaxation in the single and the multiple resource case. We discuss gap closing methods to obtain a 2-step approach and finally evaluate the different methods experimentally. In Chapter 4 we extend our 2-step approach to the more general class of constrained network optimization problems and discuss three examples: constrained minimum spanning trees, the table layout problem, and constrained geodesic shortest paths. In Chapter 5 we present our software package CNOP that implements the generic 2-step approach. Finally, we close with a discussion of our results and open problems in Chapter 6.