Chapter 1

State of the Art

The combined nature of LRPs closely relates them to a variety of other problems from CO. In particular, familiarity with location analysis, vehicle routing and assignment problems becomes crucial in the understanding of LRPs and can provide insight for the design of tools for dealing with them. The same happens with their stochastic versions.

The structure of this chapter in four sections is motivated by the variety in the topics related to our work. The first two sections are devoted to the main areas where the problems under study are contained, namely, Combinatorial Optimization and Stochastic Programming. In Section 1.3 we present an overview of some subproblems of LRPs: assignment problems, discrete location and vehicle routing. In each case, we review the existing work both, for deterministic and stochastic versions. Finally, Section 1.4 focuses on the LRP and its variants, in the deterministic as well as in the stochastic case.

1.1 Combinatorial Optimization

Combinatorial Optimization has been an active area of research for many decades. Several general methodologies as well as specialized algorithms for specific problems have been designed. Most of the problems contained in this area belong the $\mathcal{NP}$ class, where no polynomial time exact algorithms are known. In this context, heuristic approaches are often more suitable than exact algorithms to achieve real world needs, the reason being their capability of providing high quality solutions in reasonable amounts of time. A classical general text on combinatorial optimization is Nemhauser and Wolsey (1988). More recent books are, for instance, Cook, Cunningham, W. R, and Schrijver (1998), Schrijver (1998) and Wolsey (1998).

Generally, (mixed) integer programming models are used to formally describe combinatorial optimization problems. When it is possible, linear models are used. Optimization methodologies can be classified in exact and approximate or heuristic. Among heuristic approaches we can distinguish constructive methods, improvement techniques, Lagrangian relaxations and column generation.

Constructive heuristics are often designed *ad-hoc* to exploit the characteristics of each problem. The most popular methodology is the greedy one. Most of the classical improvement techniques can be seen as *Local Search* (LS) procedures with different definitions of the neighborhoods. One of the weaknesses of classical heuristics is early termination due to local optimality. More recent metaheuristic methods overcome this limitation by applying different strategies. The increase in the computational time that they require undoubtedly pays given the improvement in the quality of the obtained solutions. The most popular metaheuristics are Tabu Search (Glover 1986; Glover and Laguna 1997), Simulated Annealing (Cerny 1985; Kirpatrick, Gelatt, and Vecchi 1993), Genetic and Evolutive Algorithms (Holland 1975; Michalewicz 1992), Variable Neighborhood Search (Hansen and Mladenović 1997, 2001), Ant Colony Systems (Dorigo 1992; Dorigo and Di Caro 1999) and Greedy Randomized Adaptive Search Procedures (Feo and Resende 1989, 1995).

We next describe the principles of Tabu Search, which has been used in this work.

Tabu Search, first introduced in Glover (1986), is an extension of classical LS. In this method, to avoid getting trapped in local optima, the process is not terminated when none of the neighbors of the current solution improves the objective value. That is, even if the best neighbor of the current solution does not improve the value of the objective function, it is taken for the next iteration in the hope that in successive iterations it will lead to a better solution. A list of forbiden moves (tabu list) is mantained to prevent this approach from cycling.

In practice, instead of storing the complete solutions in the tabu list, only some of their attributes are kept. This makes storage as well as checking much more efficient. However, the fact that more solutions than the ones already visited are now forbidden, makes it necessary to remove tabu attributes from the list after a number of iterations. By doing this, a certain risk of cycling is again incurred. This risk is commonly reduced by using a tabu list of random length.

An other general strategy that has been proposed to improve the behavior of Tabu Search is to allow moves in the infeasible region. Infeasibilities are monitored by means of penalty terms included in the objective function which are associated with the violated constraints (Gendreau, Hertz, and Laporte 1994; Díaz and Fernández 2001). The weight of such penalties is commonly dynamic and depends on the evolution of the search process.

Lagrangian Relaxation has been widely used in the context of integer programming to obtain tight bounds, occasionally combined with simple heuristics. Broadly speaking, difficult constraints are eliminated from the problem and incorporated to the objective function, weighted by appropriate multipliers. Comprehensive surveys on Lagrangian relaxation for discrete programming are, for instance, Geoffrion (1974), Shapiro (1979) or Fisher (1981). Different techniques have been used to solve the resulting dual problem. One of the most popular is subgradient optimization (Held, Wolfe, and Crowder 1974; Shor 1985).

Finally, an approximate method that has proven to be quite effective for problems with a large number of columns (variables) is Column Generation (Ford and Fulkerson 1958; Dantzig and Wolfe 1960; Gilmore and Gomory 1961). Its principle is to consider only a subset of the actual variables, and successively generate new columns when they can improve current solutions.

The most extended exact methodologies are based on implicit enumeration schemes, where blocks of solutions are successively eliminated by using bounding techniques (Johnson, Nemhauser, and Savelsbergh 2000). Such methods often differ from each other in the relaxation used (linear, Lagrangian, etc.) to compute the bounds. Additionally, the initial model can be successively enhanced by the inclusion of valid inequalities, preferably facets of the convex hull of feasible solutions. This is the basic principle of branch and cut (Padberg and Rinaldi 1987, 1991). An other interesting implicit enumeration scheme is obtained when combining branch and bound with column generation, as proposed in Barnhart, Johnson, Nemhauser, Savelsbergh, and Vance (1988). Also in this case the addition of extra valid constraints has been considered.

1.2 Stochastic Programming

Stochastic programming models can be seen as an extension of linear and nonlinear programming models where the stochastic nature of part of the data is explicitely taken into account. The introduction of such models in optimization literature is generally attributed to a work by G. B. Dantzig in 1955, although some earlier sporadic contributions can be found, that fit into the stochastic programming scope in a broad sense.

Authors resort to different schemes to introduce in the models probabilistic representations of those parameters that are not known at the moment when decisions have to be taken. Disregarding those models based on queuing theory, two main strategies can be found in the literature: the use of probabilistic constraints, and two stage or recourse models.

Since decisions must be taken before the actual values of some of the parameters defining the problem are known, it can turn out that, once those values become available, the proposed solution is infeasible. *Probabilistic constrained* or *chance constrained* models allow these infeasibilities to occur only with a pre-specified probability. *Two stage models*, on the other hand, consider the possibility of taking some adaptive actions to correct these infeasibilities when they occur. In those models, a cost is associated with each possible adaptive action, and the expected cost of such actions is included in the objective function. Models where different subsets of the random parameters are known at different moments in time are refered to as multi stage models. Some examples exist where both, recourse actions and probabilistic constraints, have been included in the same model.

In what follows, we will concentrate in the existing literature on recourse models, since this is the modeling strategy we have used in the stochastic problems considered in this work. Stochastic problems with linear recourse can be modeled as:

$$\min\{cx + \mathcal{Q}(x) | Ax = b, x \in \mathbb{R}^n_+\}$$

where

$\mathcal{Q}(x) = \mathbb{E}_\xi v(x, \xi)$,

$v(x, \xi) = \min\{q(\xi)y | W(\xi)y = p(\xi) - T(\xi)x, y \in \mathbb{R}^{n'}_+\}$.

The function $v(x, \xi)$, refered to as the second stage problem, determines the recourse action associated with the solution x and the outcome ξ of the random parameters, and gives its associated cost. Its expectation, $\mathcal{Q}(x)$, is called the recourse function. The solution x, corresponding to the first decision, is often refered to as the *a priori* solution, since it is determined before the values of the random parameters become available.

A general overview of stochastic programming and of recourse models in particular can be found in Kall and Wallace (1994), Prékopa (1995), Birge and Louveaux (1997) or Klein Haneveld and van der Vlerk (2001). For more insight in the recent tendencies we refer the reader to the extensive bibliography compiled in van der Vlerk (2002).

The study of theoretical properties of the recourse function has attracted many researchers. Although some general results have been proven, assumptions on the second stage problem are most often needed. In particular, the so called technology matrix, $W(\xi)$, is often requested to be deterministic, *i.e.*, $W(\xi) = W \, \forall \xi$. In those cases, the recourse problem is called *fixed recourse* problem. The most fruitful result concerning problems with fixed recourse is the fact that the recourse function is convex. Moreover, if the vector of random parameters is discrete, $\mathcal{Q}(x)$ is piecewise linear. Convexity of the recourse function is the basis of the most widely spread general algorithm for stochastic linear programs: the L-shaped algorithm (Van Slyke and Wets 1969). This algorithm is an adaptation of the Benders decomposition principle to the particular case of stochastic programs (Benders 1962).

As it is the case for deterministic mathematical programming, real applications behind problems require, in many occasions, the use of integer variables. The inclusion of integer variables in the first stage problem does not introduce any complications essentially different from those faced in deterministic mixed integer programming. When integer variables are introduced in the second stage problem, on the contrary, most properties of the recourse function are lost. In particular convexity cannot be stated for this case.

Stochastic Integer Programming has given raise to an extensive literature in the last decades, and many approaches have been proposed (Wollmer 1980; Laporte and Louveaux 1992; Louveaux and van der Vlerk 1993; Klein Haneveld, Stougie, and van der Vlerk 1996; Carøe and Schultz 1999; Schultz, Stougie, and van der Vlerk 1998; Norkin, Ermoliev, and Ruszczyński 1998; Ahmed, Tawarmalani, and Sahinidis 2000). The works Schultz, Stougie, and van der Vlerk (1996), Stougie and van der Vlerk (1997), and Klein Haneveld and van der Vlerk (1999) give a good overview of the research on this topic.

1.3 Some Subproblems of Location Routing Problems

In this section we give an overview of the state of the art on some of the subproblems of the LRP; namely, the GAP, Discrete Location Problems and VRPs. In all cases, the amount of work on the stochastic problems is really small, as compared to what has been done for their deterministic counterparts.

The Generalized Assignment Problem

Assignment Problems consist of assigning jobs or tasks to agents or machines at a minimal cost. The GAP is an extension of such problems where the processing of each job by a given agent requires a certain amount of some resource, and assignments have to be done taking into account the resource availability of each agent. This kind of decisions must be taken in many real situations, such as capital investment (De Maio and Roveda 1971), design of shipyards (Gross and Pinkus 1972), assignment of jobs to computer networks (Balachandran 1976), terminals to concentrators (Mirzaian 1985; Chardaire 1999), user nodes to processing sites (Pirkul 1986), etc. It also appears as a subproblem of other problems such as VRPs (Fisher and Jaikumar 1981) or location problems (Ross and Soland 1977; Díaz and Fernández 2001; Díaz 2001). Comprehensive bibliography reviews on the GAP are Cattrysse and Van Wassenhove (1992), Martello and Toth (1990) or Romero Morales (2000).

The GAP was proven to be $\mathcal{NP}$-hard in Fisher, Jaikumar, and Van Wassenhove (1986). Moreover, its associated feasibility problem is $\mathcal{NP}$-complete (Martello and Toth 1990). However, authors have proposed several exact approaches for it, based on enumerative schemes. A lot of studies exist concerning (linear, Lagrangian or surrogate) relaxations of the GAP, and, in most cases, they have been embeded in Branch and Bound Algorithms (Ross and Soland 1975; Benders and van Nunen 1983; Fisher, Jaikumar, and Van Wassenhove 1986; Guignard and Rosenwein 1989; Karabakal, Bean, and Lohmann 1992). There also exist some works where Branch and Cut (Cattrysse, Degraeve, and Tistaert 1998) and Branch and Price (Savelsbergh 1997) schemes are proposed for the GAP.

A wide range of heuristics have been developed for the GAP (Osman 1995). An important proportion of them consist of a constructive phase, generally based on solutions of a relaxation of the problem, followed by some kind of local search (Klastorin 1979; Cattrysse 1990; Trick 1992; Lorena and Narciso 1996; Narciso and Lorena 1999). Also, a variety of metaheuristic methods have been applied to the GAP; either alone or within hybrid schemes. Racer and Amini (1994) and Yagiura, Tamaguchi, and Ibaraki (1998) propose different Variable Depth Search algorithms, while an application of Ant Colony Systems as well as a GRASP procedure are presented in Ramailhinho and Serra (2002). Also, a TS heuristic is proposed in Díaz and Fernández (2001), an ejection chain approach is presented in Yagiura, Ibaraki, and Glover (1999) and Genetic Algorithms are proposed in Chu and Beasley (1997), Wilson (1997), and Lorena, Narciso, and Beasley (2002). Examples of hybrid heuristics are the combination of greedy with local search (Martello and Toth 1990; Romeijn and Romero Morales 2000), Simulated Annealing with Tabu Search in (Osman 1995) or VDS with the heuristic proposed by Martello and Toth (Amini and Racer 1995).

To the best of our knowledge, the only work independent of this thesis dealing with a SGAP is Mine, Fukushima, Ishikawa, and Sawa (1983), where a heuristic is proposed for a GAP with stochastic side constraints.

Discrete Location Problems

Location problems consist of determining the optimal location of a fixed or variable number of facilities in order to achieve optimality with respect to some economical or social measure, while guaranteeing a prespecified service level. The interested reader is addressed to Daskin (1995), Drezner and Hamacher (2002) or Mirchandani and Francis (1990) for surveys of different types of location models and algorithms. In what follows, we will restrict to the case of discrete location.

Different discrete location models are distinguished, among others, by:

- The existence or not of fixed costs associated with each candidate location.

  When such costs do not exist, normally the number of facilities to locate is given. If no side constraints are added such problems are known as p-median problems (Hakimi 1965).

- The consideration of capacities for each facility. Uncapacitated Facility Location Problem (UFLP) or Simple Plant Location Problem *versus* Capacitated Plant Location Problem (CPLP).

  The crucial difference between both problems is that the allocation of customers to facilities goes from being trivial in the case of the UFLP to a much more involved problem in the capacitated case. Nevertheless, even the UFLP has been proved to be $\mathcal{NP}$-hard (Cornuéjols, Nemhauser, and Wosley 1989).

- The allowance of serving a customer from different facilities.

  The use of multiple sourcing for a given customer is only justified in the case of capacitated problems. In this case, the complexity of the problem increases notably when each customer is requested to be completely serviced by a single facility. Again, the difference relies on the allocation of customers to facilities; while for the CPLP is a Transportation Problem it becomes a Generalized Assignment Problem (GAP) in the Single Source Capacitated Plant Location Problem (SSCPLP).

  A review of solution methods for the CPLP can be found in Sridharan (1995). For insight on the SSCPLP we refer the reader to Díaz (2001).

As opposed to the deterministic case, the literature on stochastic discrete location is scarce. Louveaux (1993) presents a review of the literature devoted to stochastic location problems.

Stochasticity in location problems appears typically in the location of customers, the presence or absence of each customer, the travel costs or times, or the level of demand.

Vehicle Routing

There are a variety of problems included the VRP family. The classical problem is defined on a graph with a differentiated vertex (depot) that has travel costs or times associated with each arc. A set of m vehicle routes starting and ending at the depot, with minimal total cost, have to be designed in such a way that each of the remaining vertices is visited by exactly one vehicle. The value of m can be part of the data or of the decision variables. Other variants are obtained by the consideration of additional side constraints such as capacity and/or duration constraints or time windows and also depending on the kind of service the vehicles provide (pickups, deliveries, both, etc.). When there are more than one depot, the problem is refered to as the Multi-Depot VRP. Another family of problems closely related to VRPs are Arc Routing Problems (Dror 2000). They are also defined on a graph with a differentiated vertex and costs associated with the arcs, but this time are the arcs (all or a subset of them) that need to be visited by the vehicles, at least once.

The number of applications of VRPs has motivated a rich literature. Recent comprehensive reviews of the VRP and several of its variants, together with the study of many applications, can be found in Crainic and Laporte (1998) and Toth and Vigo (2002b). Both books are mainly devoted to deterministic models, with the exception of one of the applications in the latter, and the chapter by Laporte and Louveaux in the former, where the authors review applications of a general stochastic integer programming algorithm to VRPs.

Several exact algorithms have been proposed for the VRP and most of its deterministic variants. Most contributions are based on applying Integer Programming techniques to different modelizations of the problem. Existing formulations can be classified into vehicle flow, commodity flow and set partitioning. Basically, vehicle flow formulations have only variables related to the utilization of each arc by the vehicles, while commodity flow formulations have also flow variables indicating the load of the vehicles when crossing a given arc. Finally, variables in set partitioning formulations are related to feasible routes; due to the number of such variables, column generation schemes are often used in this context. For more insight in exact methods for the we refer the reader to Toth and Vigo (2002a), Naddef and Rinaldi (2002) and Bramel and Simchi-Levi (2002).

Many authors have developed heuristic methods for different VRPs (Laporte and Semet 2002) and also most metaheuristics have been applied to them (Geandreau, Laporte, and Potvin 2002).

Again, the literature concerning stochastic VRP's is more limited than in the deterministic case. For an overview of the existing works on stochastic VRPs, we refer the reader to Stewart and Golden (1983), Dror, Laporte, and Trudeau (1989),Laporte and Louveaux (1998) and Hadjiconstantinou and Roberts (2002).

1.4 Location Routing Problems

In the location problems considered in Section 1.3 the service to customers is either performed in the facility site, requiring the own customers to travel there (hospitals, schools, post offices, etc.) or it's performed at the customers' sites, but requires a complete return trip form the plant to each customer (ambulances, fire stations, etc.) In other situations, however, services are given at the clients' sites and several services can be performed without any need to return to the facility site after each one (mail deliver, repairmen, etc.). Now, the operational level requires making additional decisions concerning the management of the fleets, giving rise to LRPs. In addition to the previously mentioned characteristics, LRPs take into account the distances between clients as well as other characteristics of the systems such as the number of available vehicles, their capacities, etc.

In many practical applications, the location of facilities not only affects locational costs, but also has a major impact on routing costs (Webb 1968; Salhi and Rand 1989). In these situations the overall location-routing decision is, in fact, a multilevel problem that should be addressed jointly, so a combined LRP is appropriate. Nevertheless, in practice, due to the difficulty of both location and routing problems, the two problems have been very often solved independently. In the last years, however, several works have addressed combined LRPs of various characteristics. Yet, as the model becomes more general, the solution methodology that is applied often resorts to iterative heuristic approaches where at each iteration a location problem plus a routing problem are solved independently.

There are many different aspects that characterize the various types of LRPs. Laporte (1988) reviews examples, applications and algorithms for deterministic LRPs and gives a complete classification. A more recent survey where stochastic models are also considered is Min, Jayaraman, and Srivastava (1998). In this work we focus on two static (one single period) problems, a deterministic and a stochastic one. Broadly speaking, static LRPs differ from each other in:

- the type of facilities to be located: *primary* if they are the origins and destinations of vehicle journeys, *secondary* when they can only be intermediate depots;
- the number of facilities to be located and whether they have capacities or not;
- the number of available vehicles and whether they have capacities or not.
- the nature of the data, that can be deterministic or stochastic

Different exact algorithms for LRPs with a single primary facility and uncapacitated vehicles have been proposed (Ghosh 1981; Stowers and Palekar 1993; Averbakh and Berman 1994). A heuristic algorithm for the case of capacitated vehicles is presented in Chien (1993).

The case of multiple facilities with uncapacitated vehicles, has also been considered in the literature. Primary locations have been addressed heuristically (Srivastava 1993) and with exact methods (Zografos and Samara 1989; List and Mirchandani 1991; ReVelle, Cohon, and Shobrys 1991). Exact algorithms for secondary facilities have been proposed by Laporte, Nobert, and Pelletier (1983) and by Laporte and Dejax (1989).

LRPs with multiple facilities and capacitated vehicles are more complicated. We are aware of only one work (Nambiar, Gelders, and Van Wassenhove 1981) where primary locations have been considered. Secondary facilities have been addressed heuristically both in the uncapacitated case (Gillett and Johnson 1976; Jacobsen and Madsen 1980; Srivastava and Benton 1990) and when facilities have capacities (Perl and Daskin 1984; Perl and Daskin 1985; Bookbinder and Reece 1988). To the best of our knowledge problems with capacitated secondary facilities have only been solved exactly in Laporte, Nobert, and Arpin (1986) and in Laporte, Nobert, and Taillefer (1988).

Both LRPs that we study in this work consider capacitated primary facilities and one single uncapacitated vehicle associated with each open facility. They differ in the nature of the demands. In one case, they are deterministic, and in the other they are stochastic, and they follow a Bernoulli distribution.