Chapter 20

RAMP: A NEW METAHEURISTIC FRAMEWORK FOR COMBINATORIAL OPTIMIZATION

César Rego
Hearin Center for Enterprise Science, School of Business Administration, University of Mississippi, University, MS 38677, USA, crego@bus.olemiss.edu;

Abstract: We propose a new metaheuristic framework embodied in two approaches, Relaxation Adaptive Memory Programming (RAMP) and its primal-dual extension (PD-RAMP). The RAMP method, at the first level, operates by combining fundamental principles of mathematical relaxation with those of adaptive memory programming, as expressed in tabu search. The extended PD-RAMP method, at the second level, integrates the RAMP approach with other more advanced strategies. We identify specific combinations of such strategies at both levels, based on Lagrangean and surrogate constraint relaxation on the dual side and on scatter search and path relinking on the primal side, in each instance joined with appropriate guidance from adaptive memory processes. The framework invites the use of alternative procedures for both its primal and dual components, including other forms of relaxations and evolutionary approaches such as genetic algorithms and other procedures based on metaphors of nature.

Keywords: RAMP, Scatter Search, Surrogate Constraints, Lagrangean Relaxation, Cross-Parametric Relaxation, Subgradient Optimization, Adaptive Memory, Metaheuristics, Combinatorial Optimization.

1. Introduction

Adaptive memory programming (AMP) has been the source of numerous important developments in metaheuristics in the last decade. The term refers to the appropriate integration of memory structures to effectively explore the solution space in optimization algorithms. Because AMP is the foundation of tabu search (TS), which appeared as the first method specially designed to exploit adaptive memory, the terms TS and AMP have often been used interchangeably. However, more recently the principles of AMP as introduced in tabu search have likewise been used to enhance other approaches as in the creation of hybrid algorithms that incorporate tabu search components. Many important examples exist of methods that integrate genetic algorithms and evolutionary computation methods with adaptive memory, but perhaps the most prominent example and successful use of adaptive memory outside of tabu search are the recent developments in scatter search and its generalization, the path-relinking approach.

On the other hand, relaxation techniques have been widely used in combinatorial optimization to provide bounds for tree search procedures as well as to produce heuristic algorithms. These techniques are based upon the solution of an auxiliary (or relaxed) problem derived from the original by fully dropping or diminishing the restrictiveness of some constraints. Since a feasible solution to the original problem is usually not available from the solution of the relaxed problem, constructive and local search procedures are typically used to search for complementary feasible and possibly improved solutions. The effectiveness of the improvement methods employed is critical for the performance of these relaxation-based approaches. Yet surprisingly, only relatively rudimentary forms of search strategies can be found in the literature on these methods, and are usually limited to descent algorithms involving greedy strategies and very simple neighborhoods. This is a critical gap in the current state-of-the-art of approximation algorithms since it is well established that in many cases in the history of optimization dual approaches outperform their primal counterpart. Therefore, these findings provide a significant motivation for the next generation of metaheuristics to take advantage of duality as well as the creation of metaheuristic approaches that can exploit the primal-dual relationships as a means for bridging the duality gap in combinatorial optimization.

This paper takes a first step toward integrating these two key developments by proposing a unified framework for the design of dual and primal-dual metaheuristics that take full advantage of adaptive memory programming. Such an initiative is influenced and reinforced by the following remarks.

An important conceptual difference exists between local search methods and relaxation techniques. Local search methods are characteristically *restrictive* approaches that confine the solution space to the solutions that are made accessible by the neighborhood structure employed. By contrast, relaxation techniques enlarge the solution space to include infeasible solutions that fall within the boundaries of the *relaxation* utilized. Ultimately, even in cases where local search methods allow for crossing feasibility boundaries as a form of strategic oscillation, neighborhood search and relaxation heuristics are conceptually complementary approaches for handling optimization models and generating search directions.

Local search methods are typically *primal* approaches in the sense that they explore the solution space by exploiting the context of the original problem to be solved. By contrast, relaxation heuristics work by primarily solving a distinct problem associated with the type of the relaxation. In particular, Lagrangean and surrogate constraint relaxation approaches are based on optimizing the Lagrangean or the surrogate *dual* problem, respectively, and then finding a complementary feasible primal solution. However, when it comes to integer programs the optimization of the dual problem relaxation does not necessarily result in a complementary feasible primal solution, and in many cases these methods do not converge in a reasonable amount of time, thus producing solutions with large duality gaps.

Our fundamental premise is that solving the Lagrangean or the surrogate problem affords relevant insights for the creation of adaptive memory structures by gathering information that cannot be obtained by primal based approaches. Consequently, a method that effectively gathers information from the primal and dual sides fulfills the concept of adaptive memory programming and suitably may provide a unified framework for solving difficult combinatorial optimization problems.

We propose two metaheuristic approaches based on these observations. The first approach couples surrogate and Lagrangean relaxations with tabu search and path-relinking as a means to create a Relaxation AMP (RAMP) method. In addition, we introduce a special relaxation technique giving rise to a cross-parametric relaxation method (CPRM) that combines the notion of *parametric subgradients* from surrogate constraint duality theory (Glover 1975) with the Lagrangean/Surrogate relaxation technique introduced in Narciso and Lorena (1999). More precisely, CPRM combines Lagrangean and surrogate relaxations by using a Lagrangean based subgradient search within a surrogate constraint framework to generate good surrogate constraints.

Although the RAMP approach constitutes a stand alone metaheuristic, we integrate this approach with scatter search and path-relinking to create a Primal-Dual metaheuristic approach, called PD-RAMP, which constitutes a major contribution of this paper. While RAMP is primarily a dual approach, PD-RAMP exploits the primal-dual relationships more thoroughly. For economy of terminology, we will refer to both forms of the RAMP method (the basic RAMP method and PD-RAMP) simply as RAMP, without added qualification, whenever no specific details are necessary regarding the level of the approach utilized.

The reminder of this paper is organized as follows. Section 2 reviews classical relaxation procedures that are deemed relevant in the context of this research and discusses the cross-parametric relaxation method. Section 3 describes the dual and primal-dual components of the RAMP method. Section 4 presents concluding remarks and perspectives for further developments.

2. Fundamentals of Cross-Parametric Relaxation

Surrogate constraints (SC) and Lagrangean relaxation (LR) form the building blocks of the cross-parametric relaxation method (CPRM).

Throughout this paper we define specific problems by reference to their value functions. Following this convention, consider the general 0-1 integer linear programming problem P defined by

$$v(P) = Min\ \{cx \mid Ax \leq b, Dx \leq e, x \in \{0,1\}\}$$

and assume that the constraints $Ax \leq b$ are the ones that make the problem difficult to solve (i.e., the form of the problem that excludes these constraints can be solved efficiently by known methods).

Lagrangean relaxation

The Lagrangean relaxation of P is obtained by *dualizing* the constraints $Ax \leq b$ to form the integer programming problem LP^{λ} defined by

$$v(LP^{\lambda}) = Min\ \{cx - \lambda(Ax - b) \mid Dx \leq e, x \in \{0,1\}\},$$

where λ represents a nonnegative vector of multipliers with one component for each row of A. The dual problem in this case consists of finding such a λ that maximizes the value $v(LP^{\lambda})$. A solution to LP^{λ} for any given vector $\lambda \geq 0$ provides a lower bound on the primal objective function $v(P)$ – a result known as *weak Lagrangean duality*. Whenever there is a *duality gap*[1], i.e., whenever the optimum values for the primal and dual problems are not the same, it is not possible to determine optimality for an integer linear programming problem by means of solving the dual. *Strong Lagrangean*

[1] The *duality gap* is sometimes called the *integrality gap* when referring to the dual of the linear programming relaxation of P.

duality, that gives rise to the *optimality conditions* for an Lagrangean primal-dual solution[2], includes *complementary slackness conditions*—i.e. for a given λ a primal solution x must satisfy $\lambda(Ax-b)=0$. If the optimal primal dual solution fails to satisfy the complementary slackness conditions, the solution is a *near-optimal solution* to P with a duality gap $v(P)-v(LP^{\lambda})=\lambda(Ax-b)$. Determining an optimal primal-dual solution underlies finding the optimal multipliers for LP^{λ} that result from solving the Lagrangean dual of the primal problem *P*, which may be more explicitly defined by

$$v(D^{\lambda}) = Max\ \{v(LP^{\lambda}) \mid \lambda \geq 0\}.$$

Although the problem is only restricted by non-negativity constraints that do not offer difficulties, it contains a nonlinear function with an implicit Maxmin objective that is rather costly to evaluate. In fact the problem is equivalent to

$$v(D^{\lambda}) = \underset{\lambda \geq 0}{Max}\ \underset{Dx \leq e}{Min}\ \{cx - \lambda(Ax-b) \mid x \in \{0,1\}\}.$$

Therefore, to determine $v(LP^{\lambda})$ for a given λ, one must solve the Lagrangran problem LP^{λ}. A useful property is that $v(LP^{\lambda})$ is a continuous piecewise linear concave function of λ (represented by a finite number of linear functions $cx-\lambda(Ax-b)$, one for each Lagrangean dual solution). The function is differentiable except at points λ where LP^{λ} has alternative optimal solutions. Subgradient optimization, which extends the gradient concept to nondifferentiable concave/convex functions, has proved effective to find optimal or near-optimal Lagrangean dual solutions. Likewise the method can be used to generate dual solutions for other types of relaxations as will be discussed later.

Surrogate constraint relaxation

A surrogate relaxation of P consists of replacing the constraints $Ax \leq b$ by a nonnegative linear combination of these constraints weighted by a vector of multipliers w. This replaces the constraints $Ax \leq b$ by a single *surrogate constraint*, $w(Ax-b) \leq 0$, thus producing the *surrogate problem* SP^{w} defined by

[2] A Lagrangean primal-dual solution is an optimal solution for LP^{λ} that also satisfies the primal $Ax \leq b$ constraints.

$$v(SP^w) = Min\,\{cx \mid w(Ax-b) \le 0, Dx \le e, x \in \{0,1\}\}.$$

Since SP^w is a relaxation of P (for w nonnegative), $v(SP^w)$ cannot exceed the optimal objective function value for P and it approaches this value more closely as $w(Ax-b) \le 0$ becomes a more accurate representation of the polyhedron defined by the constraints $Ax \le b$. The associated surrogate dual is the one that yields

$$v(D^w) = Max\,\{v(SP^w) \mid w \ge 0\}.$$

Although surrogate constraint relaxation has not been employed as widely as Lagrangean relaxation, it is theoretically more powerful—the surrogate duality gap is always at least as good, and often better than, the Lagrangean duality gap. In fact, the Lagrangean dual value can be equal to the surrogate dual value only if the complementary slackness conditions hold for every Lagrangean multiplier (Greenberg and Pierskalla 1970), which is very unlikely for most integer programming problems. Another important result is that, as surrogate constraints incorporate the corresponding original primal constraints as a special case, any optimal solution to the surrogate problem that is feasible for the primal is automatically optimal for the primal. This result contrasts advantageously with Lagrangean relaxation because no complementary slackness conditions are required to determine optimality.

Subgradient Optimization

The subgradient method is a generalization of the gradient method to nondifferentiable concave/convex functions. Subgradient optimization is a well established technique to solve the Lagrangean dual and likewise has been recently proved effective in finding good weights for surrogate constraint relaxations (see for example, Lorena and Lopes 1994).

Let λ^* denote the optimal solution for the Lagrangean dual problem D^λ. Beginning at some point λ^0 (e.g. $\lambda^0 = 0$) the method iteratively generates a sequence of points

$$\lambda^{k+1} = \lambda^k + \theta^k (b - Ax^k),$$

where θ^k is a positive scalar called the step size. If θ^k is small enough, the point λ^{k+1} will be closer (in a Euclidian distance sense) to λ^* than λ^k. In fact, although each step of the subgradient method does not guarantee an increase in $v(LP^\lambda)$, it can be shown (see Held, Wolfe and Crowder 1974) that under the assumption that $\lim_{k\to\infty}\theta^k = 0$ and $\sum_{k=0}^{\infty}\theta^k = \infty$, the sequence of $v(LP^{\lambda^k})$ values converges to $v(D^{\lambda^*})$. Therefore, the question is how to select θ^k in order to guarantee convergence. It is common to compute θ^k as

$$\theta^k = \frac{\beta^k[v(D^\lambda) - v(LP^\lambda)]}{\| b - Ax^k \|^2},$$

where β^k is a parameter between 0 and 2 and $v(D^\lambda)$ is an upper bound on the optimal value $v(D^{\lambda^*})$, with x^k being an optimal solution for LP^λ. In practice, by weak duality $v(D^\lambda)$ is replaced by the value of a feasible solution to P and parameter β^k is initialized by setting $\beta^0 = 2$ and halving its current value whenever $v(D^\lambda)$ has not improved for a fixed number of iterations. Likewise, the method stops when no substantial improvement has been verified for a predefined number T of iterations, e.g. $| v(LP^{\lambda^k}) - v(LP^{\lambda^{k+t}}) | < \varepsilon$ for a given $\varepsilon > 0$ and $1 \leq t \leq T$.

Cross-Parametric Relaxation

Cross-parametric relaxation combines surrogate and Lagrangean relaxations coupled with Lagrangean-based subgradient search to generate good surrogate constraints. In terms of graph theory, the method can be defined as using a classical subgradient search with a *Lagrangean substitution* as a way to produce *parametric subgradients*, as defined in Glover (1975).

The parametric subgradient feature of the method can be sketched as follows. The subgradient method is used to generate the vector of surrogate multipliers for the relaxed constraints of the primal to create the corresponding surrogate problem. Then, the surrogate vector is used as a *parameter* vector in the subgradient search carried out on the Lagrangean relaxation of the surrogate problem aimed at determining a surrogate dual solution. It should be noticed here that if only the surrogate constraint is relaxed the Lagrangean multiplier is a scale factor rather than a vector. However, there are cases where relaxing more than one component under this Lagrangean substitution framework may be advantageous as will be discussed later. In any event, the new surrogate dual solution and the primal solution (obtained by projecting the surrogate dual solution on the primal-feasible region) yield the new lower and upper bounds, respectively. These bounds in turn are used in computing the next subgradient-based multipliers to generate a new surrogate problem, thus completing the loop for one iteration of a parametric subgradient search.

In fact, the method is parametric in multiple senses: first in the sense that the subgradient search itself is a search for good parameter (weight) values; second in the sense that the method is substituting the classical Lagrangean based subgradient search inside the surrogate constraint framework to provide what are called parametric subgradients. Although the resulting surrogate relaxation is not guaranteed to produce a smallest duality gap, this method is usefully designed to yield an effective and efficient dual approach for solving difficult integer programming problems.

In other words, the cross-linking of Lagrangean and surrogate relaxations by using (1) Lagrangean-based subgradient directions for the solution of the surrogate dual and (2) subgradient-based surrogate weights for the surrogate relaxation, gives rise to what we call *cross-parametric relaxation.* The method can be viewed as a generalization of the Lagrangean/surrogate relaxation considered in Narciso and Lorena (1999) and the parametric subgradient method as just described.

More formally, the cross-parametric relaxation can be written as

$$v(L_\lambda SP^w) = Min\ \{cx - \lambda w(Ax - b) \mid Dx \leq e, x \in \{0,1\}\},$$

where w is a vector of surrogate multipliers and λ is a scale factor representing the Lagrangean multiplier associated with the surrogate constraint.

The corresponding cross-parametric dual problem is the optimization problem in λ and w

$$v(D^{\lambda w}) = Max\ \{v(L_\lambda SP^w) \mid \lambda, w \geq 0\}.$$

It is immediate that by setting $\varphi = \lambda w$, problem $D^{\lambda w}$ is the Lagrangean dual D^φ, thus identifying the same optimal dual solution.

The purpose of the cross-parametric relaxation is to approach a solution for D^φ through a *decomposition method* consisting of solving a sequence of locally optimal dual problems

$$v(D_\lambda^w) = Max\ \{v(L_\lambda SP^w) \mid \lambda \geq 0\},$$

in which w is a *parameter* associated with a vector of surrogate multipliers determined by the subgradient method (or some pre-defined vector used to initialize the method). In other words, a solution to D_λ^w for a fixed w corresponds to a surrogate dual solution. The method progresses by maintaining an appropriate interaction between the multipliers λ and w in such a way that changes in λ leading to surrogate dual solutions implicitly induce changes in w. This interaction materializes by cross-linking the Lagrangean and surrogate approaches, using Lagrangean based subgradient search within a surrogate constraint relaxation and using subgradient directions (inferred by surrogate dual values and the corresponding projected primal solution values) to determine new surrogate multipliers.

It should be stressed that even though the cross-parametric and traditional Lagrangean relaxations are theoretically equivalent, the optimization processes underlying the two methods are significantly different, yielding different patterns and rates of convergence. This derives from the fact that cross-parametric relaxation is based on optimizing a composite dual variable $\varphi = \lambda w$ that reflects the interaction between two interrelated dual variables. This process rests on the solution of surrogate dual problems that do not arise in traditional Lagrangean relaxation. Also, since no assumption is made about the value of w when choosing λ, it is very unlikely that a Lagrangean substitution search over φ for a fixed w will lead to the same dual solutions produced by subgradient directions over λ without the parameter w. Therefore, one can expect that only for optimal combinations of w and λ, problems D_λ^w and D^φ produce the same bounds. In consequence, these composite relaxation strategies embedded in the cross-parametric relaxation approach can find potentially better local bounds than traditional Lagrangean relaxation alone.

A general framework for the cross-parametric relaxation approach is depicted in the flowchart of Figure 1. In the diagram $X(.)$ denotes the set of feasible solutions for a problem ".". Also, x_λ and x_w represent solutions to the problems $L_\lambda SP^w$ and SP^w, respectively. Consequently, it is assumed that if x_λ is feasible for SP^w, both x_λ and x_w stand for the same solution.

As the figure shows, the information contained in each surrogate solution may be used in a constructive process for building a feasible solution to the primal problem. In addition, an improvement method is used in an attempt to find an enhanced solution, which thereby may be used to replace the current upper bound for the next subgradient iteration. By this means, a sequence of Lagrangean multipliers generates lower bounds, while a sequence of feasible and enhanced solutions defines upper bounds. Without loss of generality, we will refer to the block "constructive plus improvement" simply as the *projection method*, because of its primary function of projecting a dual solution into primal feasible space. Some advances on the design of effective projection methods will be given in the next section.

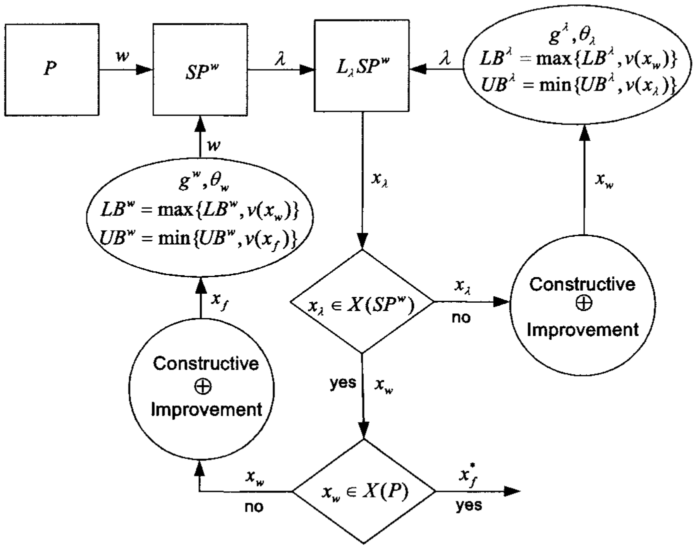

Figure 1. Cross-Parametric Relaxation Procedure

To conclude this section it is useful to say a few words about the rationale of the cross-parametric relaxation. Besides the conceptual relaxation design discussed earlier concerning some potentially misleading similarities with traditional Lagrangean subgradient search, cross-parametric relaxation rests on theoretical foundations stemming from properties of surrogate constraints. Consequently, they are exploited using concepts tied to surrogate duality theory, in particular those associated with the notion of parametric subgradients (Glover 1975). More precisely, cross-parametric relaxation is founded on the following observations.

Theoretically, a solution for a surrogate problem can be difficult to obtain. In fact, if the primal problem P is NP-complete, any surrogate problem SP^w derived from it is also NP-Complete. Yet in practice a problem relaxation having fewer constraints than the original problem is usually easier to solve. Therefore one must be concerned with finding an appropriate surrogate relaxation for which effective methods exist to solve the corresponding surrogate problem. In an extreme situation, if all constraints are relaxed to become replaced by the surrogate constraint, the surrogate problem becomes a knapsack problem. Although the knapsack problem is NP-Complete, in many cases it can be solved efficiently by specialized algorithms, particularly by drawing on the work of Pisinger (1999a, 1999b, 2004), and Martello, Pisinger and Toth (2000). Nevertheless, from the theory of NP-Completeness one might find instances (not necessarily very large) for which finding an optimal solution may be computationally very expensive. Cross-parametric relaxation becomes especially relevant in such a case. When the surrogate problem is difficult to solve, the step of relaxing the surrogate constraint in Lagrangean fashion, especially using parametric subgradients, may produce an effective procedure to find optimal or near-optimal surrogate dual solutions. This assumption is in part supported by the interesting results obtained by Lorena and Lopes (1994), and Narciso and Lorena (1999) in their applications to the set covering and generalized assignment problems, employing a special case of the cross-parametric relaxation proposed here and relatively simple greedy heuristic projection methods.

The motivation for the use of parametric subgradients is also reinforced by the fact that in many practical situations, problems are so complex that the corresponding formulations may involve a relatively large set of difficult constraints of the type exemplified by $Ax \leq b$ in our general formulation. In these circumstances, a surrogate constraint model can be generated by grouping sets of constraints of the same type, such as those modeling similar requirements as a way to create multiple surrogate constraints. Under this framework a cross-parametric relaxation may involve the Lagrangean relaxation of all the surrogate constraints, and in this case λ is no longer a scale factor but a vector of multipliers; hence subgradient optimization becomes particularly relevant to perform a multiple-dimensional search of improved dual solutions. The potential value of generating multiple surrogate constraints has been suggested in Glover (1965, 1968) and cross-parametric relaxation affords a means to exploit such a possibility and develop a more powerful relaxation-based algorithm.

3. The RAMP Method

The RAMP method is presented in two stages, representing different levels of sophistication. The first level focuses on exploiting a dual framework, which is confined to creating an adaptive memory relaxation based on the type of relaxations described in Section 2, making use of tabu search and path-relinking strategies. The second level of sophistication embodies the primal-dual framework, which establishes the relationship between the dual approach (defined in the first level of sophistication) and a primal approach represented by scatter search and path-relinking. The role of path-relinking at this second level is different from the one represented in the first level, as will be clear later in this section.

3.1 Relaxation Adaptive Memory Programming

We begin by noting some similarities and contrasts between a subgradient search and a path-relinking approach, which gives the foundation for a path-relinking projection method as a fundamental component for a more advanced Relaxation AMP approach.

Subgradients of the Lagrangean dual function $v(D^{\lambda})$ at each investigated multiplier λ^{k+1} define convex linear combinations of vectors $(Ax^k - b)$ for solutions x^k solving LP^{λ}. From this perspective, subgradient optimization with Lagrangean substitution is a search method that links partially feasible primal solutions (that only satisfy $Dx \leq e$) to the subset of solutions that are primal feasible (i.e. that also satisfy $Ax \leq b$). Thus, from a more abstract perspective we may consider that the inequality system $Ax \leq b$ implicitly defines a reference set of solutions that the method is moving towards. This contrasts with a path-relinking approach where the reference set is explicitly defined by solutions that have already been visited. Also, we may say that solutions are linked to one another by means of subgradient directions.

In a path-relinking approach, moves from one solution to the next are derived from weighted transformations that drive the search toward a subset of solutions that meet certain requirements (defined by the level of importance, with regard to optimality, of each of their own attributes). By analogy, subgradient directions are derived from weighted combinations generated with the goal of converging to the best solutions that possess the requirements (or attributes) specified by $Ax \leq b$. In this sense, the method can be viewed as a special case of a path-relinking approach where a move from one solution to another in the path is given by a subgradient direction. During a path-relinking trajectory each move typically avoids reconsidering attributes that have been previously deleted (or rejected) throughout the path—this is implemented by defining appropriate *tabu restrictions* associated with the type of neighborhood utilized. Similarly, the subgradient approach creates dual solutions (representing Lagrangean multipliers) that are meant to penalize violated primal constraints, thus keeping the method from generating the same sets of solution attributes that would reiterate over repeated $(Ax^k - b)$ vectors—this is accomplished by choosing adequate *step sizes* to ensure appropriate convergence.

The foregoing observations underlie the foundation of a path-relinking projection method, as amplified in the following discussion.

Simple Relaxation AMP.

Subgradient directions pointing to solutions that satisfy $Dx \leq e$ are attractive to provide starting points for a method that projects these solutions on the domain of feasibility of the original primal problem P. We emphasize that the type of projection methods that are referred to in the context of this paper are heuristic rather than exact and in particular are structured to exploit the notion of adaptive memory programming. As a prelude to other more elaborate strategies, we start by proposing the use of frequency-based tabu search memory to construct a primal solution from the dual solution and also to improve this solution further after reaching primal feasibility. Lower and upper bounds are then updated based on the new values of the dual and primal solutions respectively, and these values are used to compute a new subgradient direction form the dual, if it applies.

Even though this method constitutes the simplest form of a Relaxation AMP proposed here, we have found it can be quite effective compared to currently popular state-of-the-art constructive or local search projection methods based on simple greedy approaches. Also, the Simple Relaxation AMP can be enhanced by the use of neighborhood structures incorporating adaptive and dynamic search such as filter-and-fan or ejection chain methods. (For a detailed description of these methods see Rego and Glover (2002). Recent applications of filter-and-fan and ejection chain methods are reported in Greistorfer, Rego and Alidaee (2003), Renato, Rego and Gamboa (2004), and Rego, Li and Glover (2004).)

Advanced Relaxation AMP.

We now describe a more advanced approach that establishes a stronger connection between the primal and dual. To do this, we retain a collection of *elite* primal solutions selected among those projected from the dual to define a reference set for a *path-relinking projection method*, generating paths between dual solutions and their primal counterpart elite solutions. Here, the term "elite" is employed as in tabu search to refer to the output of a reference set update method utilizing appropriate measures of solution quality and diversity. An abbreviated version of path relinking projection consists of keeping a single solution in the reference set represented by the best upper bound found so far.

The method is structured in two phases. A first phase creates an initial reference set based upon a tabu search projection method, with the difference that, for the purpose of speeding up this initialization phase, only a rudimentary long term memory is employed, or even none at all if short-term memory is sufficient to achieve a reference set with desired levels of diversity and solution quality. The second phase replaces the constructive phase of the tabu search projection method by a path-relinking projection method to transform dual solutions (generated by the subgradient search) into primal feasible solutions. The analogy between the subgradient search and the path-relinking strategy constitutes a primary motivation for the development of this path-relinking projection method. The method operates by starting from a trial solution obtained from the dual and moving toward pre-defined subsets of guiding solutions in the current reference set. The definition of the aforementioned subset is established by an appropriate subset generation method as considered in the classical path-relinking/scatter search template (Glover 1997). To take fuller advantage of adaptive memory programming, tabu search is used as an improvement method over the best primal solutions obtained after combinations. A useful variation of such an improvement method can be based on integrating path-relinking and ejection chain strategies, as initially proposed in Rego and Glover (2002) and recently demonstrated to be effective for the solution of generalized assignment problems by Yagiura, Ibaraki, and Glover (2004a).

We should note that in these Relaxation AMP methods, the primal algorithm is confined to an improvement method executed by a tabu search procedure and a selection process empowered by a reference set update method. Therefore, it is considered a dual-based metaheuristic approach. A more sophisticated method that takes full advantage of the primal-dual relationships and adaptive memory programming is the Primal-Dual RAMP approach presented next.

3.2 Primal-Dual Relaxation Adaptive Memory Programming

Primal-Dual Relaxation Adaptive Memory Programming (PD-RAMP) goes beyond the customary notions of primal-dual relationships used in linear programming (where the primal and dual have the same optimum values). The method is enhanced by adaptive memory strategies that affect both sides of the primal-dual connection, as a means for bridging the duality gap that exists in combinatorial optimization.

More precisely, PD-RAMP refers to the exploitation of primal-dual relationships, by an adaptive process that takes advantage of the primal side using scatter search and path relinking, and takes advantage of the dual side using surrogate constraint and Lagrangian relaxations. It extends the basic (first level) form of the RAMP method by interconnecting its dual component with an appropriate primal component that utilizes adaptive memory programming more thoroughly. This is accomplished by integrating memory and learning through the use of a reference set of solutions that is common to and updated by both the primal and the dual approaches.

The main function of the dual approach is to generate new solutions for the reference set as specified by the first level RAMP approach. For a better understanding of the conceptual design of the PD-RAMP approach, we underline a few important concepts concerning the foundations of the surrogate constraint relaxations and contrast them with those of scatter search. (Similar concepts hold for surrogate constraints and the cross-parametric relaxation, therefore for the sake of simplifying the explanation we restrict attention to surrogate constraints.)

As previously noted, surrogate constraint approaches explore the solution space by generating solutions that are derived from the creation of nonnegative linear combinations of constraints. The process seeks to capture relevant information contained in individual constraints and integrate it into new surrogate constraints. The result is to generate composite decision rules (characterized by each surrogate problem) leading to associated new trial solutions. By contrast, scatter search combines vectors of solutions rather than combining vectors of constraints, and likewise is organized to capture information not contained separately in the original vectors. These observations set the stage for the primal-dual metaheuristic approach identified by the PD-RAMP method. Although there are other possible and also interesting variations to create primal-dual strategies under the RAMP model, we start off considering scatter search and surrogate constraints as the basic methods for the primal and dual approaches, respectively.

The method starts by creating an initial reference set of solutions produced by a surrogate approach as laid down by the initial phase of the RAMP approach. At this stage the reference set contains relevant information that was made available by the dual approach. Although the surrogate constraint model is provided with the ability to integrate information extracted from individual constraints, each piece of information, generated at each iteration of the surrogate constraint method, is represented in a single solution that results from solving the associated surrogate problem (and the application of the complementary projection method). These solutions may be viewed as individual memory structures that can be integrated to create more complex compound memory structures. This result is accomplished by the scatter search method through generating weighted combinations of these solutions (and so their attributes) to create new composite solutions. The method is structured to alternate between the primal and dual approaches, both updating the reference set in an evolutionary fashion.

Figure 2 provides an illustration of the general RAMP model that is completed by the Primal-Dual RAMP approach.

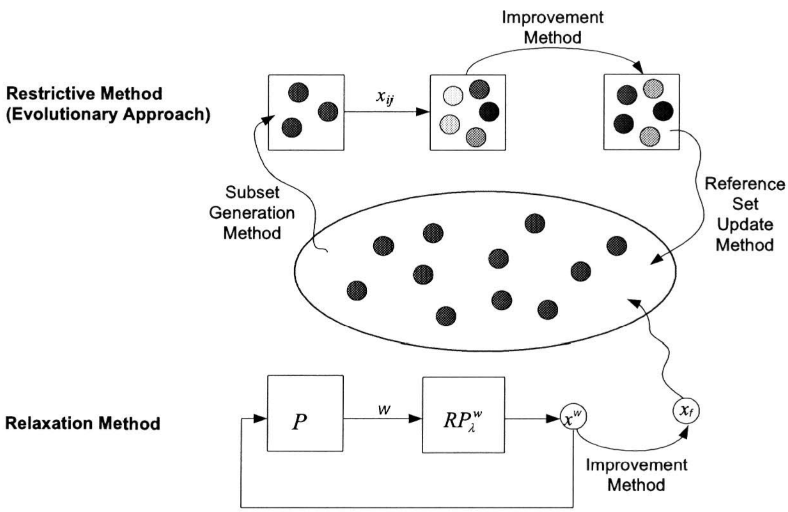

Figure 2. Relational AMP model

For the sake of providing a more comprehensive illustration, the complete PD-RAMP method uses cross-parametric relaxation in the primal side and scatter search in the dual side, though other variations of the method are possible. For example, scatter search can be naturally replaced by its path-relinking generalization. Ultimately a hybrid approach can be created by using other evolutionary approaches such as genetic algorithms (GAs) or evolution strategies (ESs), preferentially provided with adaptive memory components. Likewise, traditional surrogate constraint and Lagrangean relaxation can be used as stand alone dual procedures rather than combined in a cross-parametric approach. As far as the improvement method is concerned, there are advantages for using the adaptive memory orientation of tabu search. In addition, as described earlier path-relinking is also particularly relevant to create a projection method to bridge the gap between dual and primal solutions.

4. Conclusions and Perspectives

The complete RAMP method (embodying its primal-dual form as well as its more basic form) is organized to take advantage of a number of existing methods, using them as components or building blocks to form more advanced search strategies. The resulting procedure affords advantages that cannot be obtained by more customary metaheuristic approaches alone, or even by current hybrid methods. The component methods within RAMP are strategically articulated in a unified design rather than used as independent add-on components to be called whenever other components are unable to make progress in finding improved solutions. The cross-parametric relaxation method, the path-relinking projection method (in connection with subgradient optimization) and the integration of surrogate constraints and scatter search constitute primary factors of such an articulated methodology.

A key contribution of the RAMP method is its capability to exploit duality as well as primal-dual relationships that have been largely neglected in the field of metaheuristics. Moreover, the method includes a learning process that relies on adaptive memory rather than kicking-off or re-starting the search through randomized-based processes. Although memory is explicitly structured by reference to solution attributes, the RAMP method implicitly includes an automatic learning process embodied in an evolutionary framework that is enhanced by a cohesive integration of primal and dual approaches.

Finally, the variety of possibilities to create intriguing variations of RAMP algorithms opens new doors for research in metaheuristics. Opportunities for advances emerge from discoveries about primal-dual coordination strategies, including the coordination of parallel primal dual search. Advances can also be derived from combinations of scatter search and path-relinking strategies (or other evolutionary approaches) on the primal side and Lagrangean, surrogate or cross-parametric relaxations on the dual side. New neighborhood structures may likewise be exploited in RAMP strategies. In this connection, recent advances in ejection chain methods and combinations of these with path-relinking and filter-and-fan approaches (Rego and Glover, 2002) reveals an interesting path of research for creating advanced variants of RAMP.

Preliminary computational results obtained by the RAMP and PD-RAMP methods are very encouraging. Applied to two challenging and well-studied problems, the Generalized Assignment Problem (GAP) and the Multi-Resource Generalized Assignment Problem (MRGAP), the RAMP method yields results rivaling the best in the literature, while the PD-RAMP method dominates the performance of the previously best methods for these problems. In particular, the PD-RAMP finds solutions that are always as good or better than the leading approaches of Yagiura, Ibaraki and Glover (2004a, 2004b) for the GAP problem and of Yagiura et al. (2004) for the MRGAP problem, while using 10% on average of the time required by these algorithms. (The referenced methods achieve their leading position by also using path-relinking and ejection chains, but without embedding them within the primal-dual RAMP framework proposed here.) Compared to CPLEX 8.1, the PD-RAMP approach finds optimal solutions to those problems that are simple enough to enable CPLEX to solve them to optimality within a reasonable amount of time, but generally requires only 5% of the solution time required by CPLEX to obtain these solutions. For larger and more difficult problems, PD-RAMP obtains solutions whose quality is superior to that of the best solutions CPLEX is able to find when it is allowed to run up to 70 times longer than PD-RAMP. In particular, PD-RAMP finds the optimal solution for about 85% of these larger problems tested while CPLEX succeeds in finding optimal solutions for less than 19% of these problems under its significantly larger allotted time span. More detailed results can be found in Rego et al. (2004), currently in process.

References

Duarte, R., C. Rego and D. Gamboa (2004) "A Filter and Fan Approach for the Job Shop Scheduling Problem: A Preliminary Study," in Proceedings: International Conference on Knowledge Engineering and Decision Support, 401-406.

Glover, F. (1965) "A Multiphase-Dual Algorithm for the Zero-One Integer Programming Problem," Operations Research, 13:879-919.

Glover, F. (1968) "Surrogate Constraints," *Operations Research*, 16(4):741-749.

Glover, F. (1975) "Surrogate Constraint Duality in Mathematical Programming," *Operations Research*, 23(3):434-451.

Glover, F. (1997) "A Template for Scatter Search and Path Relinking," in Lecture Notes in Computer Science, 1363, J.K. Hao, E. Lutton, E. Ronald, M. Schoenauer, D. Snyers (Eds.), 13-54.

Greenberg, H.J. and W.P. Pierskalla (1970) "Surrogate Mathematical Proramming," *Operations Research*, 18:1138-1162.

Greistorfer, P., C. Rego and B. Alidaee "Simple Filter-and-Fan Approach for the Facility Location Problem," Research Report, Hearin Center for Enterprise Science, School of Business Administration, University of Mississippi, USA.

Held, M., P. Wolfe and H.P. Crowder (1975) "Validation of Subgradient Optimization," *Mathematical Programming*, 6:62-88.

Lorena, L.A.N. and M.G. Narciso (1999) "Lagrangean/Surrogate Relaxations for the Generalized Assignment Problems," *European Journal of Operational Research*, 114(1): 165-177.

Lorena, L.A.N. and F.B. Lopes (1994) "A Surrogate Heuristic for Set Covering Problems," *European Journal of Operational Research*, 79:138-150.

Martello, S., D. Pisinger and P. Toth (2000), "New Trends in Exact Algorithms for the 0-1 Knapsack Pproblem," *European Journal of Operational Research*, 123:325-332.

Pisinger, D. (1999a) "Core problems in Knapsack Algorithms," *Operations Research*, 47:570-575.

Pisinger, D. (1999b) "An Exact Algorithm for Large Multiple Knapsack Problems", *European Journal of Operational Research*, 114:528-541.

Pisinger, D. (2004) "Where are the hard knapsack problems?," *Computers and Operations Research*, to appear.

Poljak, B.T. (1969) "Minimization of Unsmooth Functionals," USSR Computational Mathematics and Mathematical Physics, 9:14-29.

Rego, C. and F. Glover (2002) "Local Search and Metaheuristics for the Traveling Salesman Problem," in book The Traveling Salesman Problem and its Variations, G. Gutin and A. Punnen Editors, Kluwer Academic Publishers, 309-368.

Rego, C., H. Li, and F. Glover (2004) "Adaptive and Dynamic Search Neighborhoods for Protein Folding in a 2D-HP Lattice Model", Research Report, Hearin Center for Enterprise Science, School of Business Administration, University of Mississippi, USA.

Rego, C., L. Sagbansua, B. Alidaee, and F. Glover (2004) "RAMP and Primal-Dual RAMP Algorithms for the Multi-Resource Generalized Assignment Problem," Research Report, Hearin Center for Enterprise Science, School of Business Administration, University of Mississippi, USA.

Yagiura, M., T. Ibaraki, and F. Glover (2004a) "A Path Relinking Approach with Ejection Chains for the Generalized Assignment Problem," *European Journal of Operational Research*, to appear.

Yagiura, M., T. Ibaraki and F. Glover (2004b) "An Ejection Chain Approach for the Generalized Assignment Problem," *INFORMS Journal on Computing*, 16(2):133-151.

Yagiura, M., S. Iwasaki, T. Ibaraki and F. Glover (2004) "A Very Large-Scale Neighborhood Search Algorithm for the Multi-Resource Generalized Assignment Problem," *Discrete Optimization*, 1:87-98.