Global optimality conditions for discrete and nonconvex optimization—With applications to Lagrangian heuristics and column generation

Torbjörn Larsson* and Michael Patriksson†

March 23, 2003

Abstract

The well-known and established global optimality conditions based on the Lagrangian formulation of an optimization problem are consistent if and only if the duality gap is zero. We develop a set of global optimality conditions which are structurally similar but which are consistent for any problem with continuous objective and constraint functions. This system characterizes a primal–dual optimal solution by means of primal and dual feasibility, primal Lagrangian ε-optimality, and, in the presence of inequality constraints, δ-complementarity, that is, a perturbed complementarity condition. The total size $\varepsilon+\delta$ of those two perturbations equals the size of the duality gap at an optimal solution. The system developed can therefore also be used to explain, to some degree, when and why Lagrangian heuristics for integer and combinatorial optimization are successful in reaching near-optimal solutions. Experiments on a set covering problem illustrate how the new optimality conditions can be utilized in the construction of Lagrangian heuristics. For more general integer programs, we outline possible uses of the optimality conditions in column generation algorithms and in the construction of core problems.

Key Words: Global optimality conditions; nonconvex optimization; integer programming; Lagrangian relaxation; Lagrangian heuristics; set covering problem; column generation; core problems.

1 Introduction

Classic optimality conditions for optimization problems are based on the fulfillment of primal–dual feasibility, primal Lagrangian stationarity, and complementarity conditions, associated with a particular Lagrangian function; global versions of them are fulfilled precisely at saddle-points of this function. These conditions are also the foundation of many algorithmic approaches for the search of (near-)optimal solutions. In one such class, the optimality conditions are approximated with simpler systems, such as in sequential quadratic (SQP) and linear (SLP) programming, and interior point methods. Other approaches are associated with satisfying a subset of the conditions while making adjustments in the primal–dual space in order to satisfy the rest. Among these we may count the simplex method for linear programming, while dual cutting plane algorithms perhaps more obviously belong to this class of methods. When there is a positive duality gap, however, which is typically the case with nonlinear, integer, and combinatorial optimization problems, the primal–dual system describing the set of saddle points of the Lagrangian function is inconsistent.

*Department of mathematics, Linköping University, SE-581 83 Linköping, Sweden. E-mail: tolar@mai.liu.se

†Department of mathematics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. E-mail: mipat@math.chalmers.se

We consider the problem of finding

$$f^* := \text{minimum} \;\; f(x), \tag{1a}$$
$$\text{subject to} \;\; g(x) \leq 0^m, \tag{1b}$$
$$x \in X, \tag{1c}$$

where $f : \Re^n \mapsto \Re$ and $g : \Re^n \mapsto \Re^m$ are continuous, and $X \subset \Re^n$ is compact. We assume that there exist feasible solutions to the problem, and hence also a compact set X^* of optimal solutions. An *ε-optimal* solution to any problem refers to any vector that is feasible and deviates in objective value at most ε from the optimal one; the set of such vectors in the above problem is denoted X^ε.

Let

$$\theta(u) := \underset{x \in X}{\text{minimum}} \left\{ f(x) + u^{\mathrm{T}} g(x) \right\}, \qquad u \in \Re^m, \tag{2}$$

be the Lagrangian dual function associated with the relaxation of the constraints in (1b), and

$$\theta^* := \text{maximum} \;\; \theta(u), \tag{3a}$$
$$\text{subject to} \;\; u \in \Re^m_+, \tag{3b}$$

be the Lagrangian dual problem. The duality gap for this primal–dual pair then is $\Gamma := f^* - \theta^*$.

Letting $(x, u) \in X \times \Re^m_+$, we define the *global optimality conditions* for the problem (1) as the combination of Lagrangian optimality, primal feasibility, and complementarity:

$$f(x) + u^{\mathrm{T}} g(x) \leq \theta(u), \tag{4a}$$
$$g(x) \leq 0^m, \tag{4b}$$
$$u^{\mathrm{T}} g(x) = 0. \tag{4c}$$

The following result establishes the consistency of the system (4). Similar results can be found in [Sha79a, Theorem 5.1] and [BSS93, Theorem 6.2.5].

THEOREM 1 (primal–dual optimality conditions). *Let $(x, u) \in X \times \Re^m_+$. Then, the following three statements are equivalent.*

(i) *The pair (x, u) satisfies the system* (4).

(ii) *The pair (x, u) is a saddle point of the Lagrangian function $(x, u) \mapsto L(x, u) := f(x) + u^{\mathrm{T}} g(x)$ over $X \times \Re^m_+$, that is,*

$$L(x, v) \leq L(x, u) \leq L(y, u), \qquad \forall (y, v) \in X \times \Re^m_+.$$

(iii) *x solves the primal problem* (1), *u solves the dual problem* (3), *and $f^* = \theta^*$.* □

COROLLARY 2 (a primal characterization of optimality). *Given any $u \in \Re^m_+$,*

$$\{ x \in X \mid (4) \text{ is satisfied} \} = \begin{cases} X^*, & \text{if } \theta(u) = f^*, \\ \emptyset, & \text{if } \theta(u) < f^* \end{cases}$$

holds. □

There are two possible cases in which $\theta(u) < f^*$, whence the system (4) is inconsistent: (a) the vector u is not optimal in the Lagrangian dual problem (3); (b) there is a positive duality gap, that is, $f^* > \theta^*$. Convex problems satisfying a Slater constraint qualification define a class of nonlinear programs for which the duality gap is zero (cf. [BSS93, Theorem 6.2.4]); linear programming is another important class of such problems. In the nonconvex case, Corollary 2 in general becomes a theorem about the empty set. Note further the following easy consequence of the above, which is a quite negative result: if there is a positive duality gap, then a primal optimal solution will *never* satisfy the global optimality conditions (4) for *any* value of $u \geq 0^m$. If, in particular, the relaxed constraints are equality constraints (see Section 2.2 for results corresponding to the above), then the solution to the Lagrangian relaxed problem can never yield a primal optimal solution, regardless of the choice of dual optimal solution.

Nevertheless, and somewhat surprising given that it is in fact based on the same system (4), *Lagrangian relaxation* is a popular and often successful approach to integer and combinatorial optimization problems. The interest in Lagrangian relaxation started with the publication of the papers by Held and Karp [HeK70, HeK71] on the application of subgradient optimization to the Lagrangian dual of a formulation of the traveling salesman problem, and grew with the subsequent publications in the 1970's and early 1980's (e.g., [Las70, HWC74, Geo74, MSW76, Erl78, Sha79a, Sha79b, Fis85]); see also the survey papers [Fis81, Bea93]. Until the 1980's, Lagrangian relaxation was often used as a bounding procedure in branch and bound algorithms (cf. [Fis81]). Since then, interest has shifted towards the combination of (a) Lagrangian relaxation, with a solution technique for (approximately) maximizing the dual function, and (b) a *Lagrangian heuristic* (cf. [Fis81] and [Bea93, Section 6.4]) that adjusts the (typically) infeasible Lagrangian subproblem solutions into feasible ones.

To be more precise, we shall define a Lagrangian heuristic as follows: *Initiated at a vector in the set defined by the non-relaxed constraints, it adjusts this vector by executing a finite number of steps that have the properties that (a) they utilize information from the Lagrangian dual problem, (b) the sequence of primal vectors generated remains within the set of non-relaxed constraints, and (c) the terminal vector is, if possible, primal feasible and hopefully also near-optimal in the problem* (1).

Included in this definition is the possibility that the heuristic does not terminate at the first primal feasible solution found, but continues with a primal local search. We remark that the most common means in which to comply with the property (a) are to initiate the heuristic at a (near-)optimal solution to a Lagrangian relaxed problem or to perform the adjustments guided by a merit function defined by the Lagrangian cost.

We especially distinguish between two important types of Lagrangian heuristics that we will analyze. The first, which will be referred to as *conservative*, has the properties that the initial vector is a (near-)optimal Lagrangian subproblem solution, and that the moves are local only, in the sense that the iterations retain near-optimality in the Lagrangian subproblem. The second, which will be referred to as *radical*, or *nonconservative*, has the property that it allows the resulting primal vector to be far from optimal in the Lagrangian relaxed problem. This type of heuristic includes those which are initiated at a point far from the Lagrangian subproblem solution, those that are defined by the solution of a restriction of the original problem, and large scale neighbourhood search.

Depending on the nature of the problem being attacked by Lagrangian relaxation, and the design of the heuristic, a conservative heuristic may provide very good primal feasible solutions, or no feasible solution at all. A new set of optimality conditions for general, possibly nonconvex, optimization (but which is structurally similar to those in the convex case) developed in the next section, provides an analysis of the success or failure of conservative Lagrangian heuristics. The decisive factor in this context is the size of the duality gap. In the event that conservative Lagrangian heuristics fail, this analysis motivates the use of radical Lagrangian heuristics, which may work in much larger neighbourhoods of the Lagrangian solution in the sense of the value of the Lagrangian function. A further consequence of the analysis which immediately follows from the appearance of the new optimality conditions, is that Lagrangian heuristics for problems with large duality gaps must take both subproblem optimality and complementarity fulfillment into account; the latter has to our knowledge never before been considered in Lagrangian heuristics.

In order to analyze Lagrangian heuristics, in particular for (mixed) integer programming problems, a primary goal of this paper is to reach a characterization of the set of primal–dual optimal solutions to (1), which has a structure similar to the system (4) but also covers the case of a nonzero duality gap. In contrast to (4), the system which will be developed contains perfect information about the primal–dual set of optimal solutions, and it lends itself very well to the construction of Lagrangian heuristics. Since the system (4) is not consistent for problems with a positive duality gap, it seems reasonable to investigate *relaxations* of it.

2 Global optimality conditions

2.1 Inequality constraints

We introduce nonnegative numbers ε and δ. Given the pair $(x, u) \in X \times \Re^m_+$, we define the *global optimality conditions* for the problem (1) as

$$f(x) + u^{\mathrm{T}} g(x) \leq \theta(u) + \varepsilon, \tag{5a}$$
$$g(x) \leq 0^m, \tag{5b}$$
$$u^{\mathrm{T}} g(x) \geq -\delta, \tag{5c}$$
$$\varepsilon + \delta \leq \Gamma, \tag{5d}$$
$$\varepsilon, \delta \geq 0. \tag{5e}$$

In this system, (5a) and (5c) define ε-optimality in the Lagrangian subproblem, and δ-complementarity, respectively. The systems (4) and (5) are equivalent precisely when the duality gap is zero.

The following theorem provides the analogous result to Theorem 1.

Theorem 3 (primal–dual optimality conditions). *Let $(x, u) \in X \times \Re^m_+$. Then, the following three statements are equivalent.*

(i) *Together with the pair (ε, δ), the pair (x, u) satisfies the system* (5).

(ii) *The pair $(\varepsilon, \delta) \geq (0, 0)$ satisfies $\varepsilon + \delta = \Gamma$, and the pair (x, u) satisfies the following saddle-point like condition for the Lagrangian function $(x, u) \mapsto L(x, u)$ over $X \times \Re^m_+$:*

$$L(x, v) - \delta \leq L(x, u) \leq L(y, u) + \varepsilon, \qquad \forall (y, v) \in X \times \Re^m_+. \tag{6}$$

(iii) *x solves the primal problem* (1) *and u solves the dual problem* (3).

Proof. We establish first that (i) and (ii) are equivalent. It is clear that the consistency of (5) implies that $\varepsilon + \delta = \Gamma$ holds [combine (5a), (5c), and the duality gap consequence that $f(x) - \theta(u) \geq \Gamma$ holds with (5d)]. That the second inequality in (6) is equivalent to (5a) is immediate. The first inequality in (6) is equivalent to

$$g(x)^{\mathrm{T}}(u - v) \geq -\delta, \qquad \forall v \in \Re^m_+. \tag{7}$$

With $v = 0^m$, we obtain (5c). To reach (5b), we note that if it is not satisfied, then there is some $i \in \{1, 2, \dots, m\}$ for which $g_i(x) > 0$; by letting $v_i \to +\infty$, we contradict (7). Conversely, for all $v \in \Re^m_+$,

$$\begin{aligned} f(x) + v^{\mathrm{T}} g(x) - \delta &= f(x) + u^{\mathrm{T}} g(x) + g(x)^{\mathrm{T}} (v - u) - \delta \\ &\leq f(x) + u^{\mathrm{T}} g(x), \end{aligned}$$

where the inequality follows from (5b)–(5c). This completes the first part of the proof.

Next, we establish that (i) and (iii) are equivalent. Suppose that (i) holds. Then, (5a), (5c), and (5d) imply that

$$f(x) \leq \theta(u) + \Gamma. \tag{8}$$

By definition, $\Gamma = f^* - \theta^*$. Therefore, (5) holds if and only if (x, u) is primal–dual optimal, whence (iii) follows.

Finally, suppose that (iii) holds. Then, (8) holds. Further, suppose that for the given pair (x, u), we choose ε and δ according to

$$\varepsilon := \varepsilon(x, u) = \Gamma + u^{\mathrm{T}} g(x) \qquad \text{and} \qquad \delta := \delta(x, u) = -u^{\mathrm{T}} g(x). \tag{9}$$

Adding $u^{\mathrm{T}} g(x)$ to both sides of the inequality (8) yields (5a). The inequality (5b) follows from the optimality of x in the primal problem (1), and (5c) is trivially satisfied, by the choice (9), and (i) follows. This completes the proof. □

Theorem 3 implies the following (cf. Corollary 2):

COROLLARY 4 (a primal characterization of optimality). *Given any* $u \in \Re^m_+$,

$$\{\, x \in X \mid (5) \text{ is satisfied} \,\} = \begin{cases} X^*, & \text{if } \theta(u) = f^* - \Gamma, \\ \emptyset, & \text{if } \theta(u) < f^* - \Gamma \end{cases}$$

holds. □

The two systems (4) and (5) both state that the (Cartesian product) set of primal–dual optimal solutions satisfy similar saddle-point conditions. The main difference between the convex and nonconvex case is that in the nonconvex case, primal–dual optimal solution are only near-saddle points. An interpretation is provided in Figure 1 in the context of the example of Section 2.3.

In the case of convex programming, the global optimality conditions (5) are related to, but much simpler than, those presented by Strodiot et al. [SNH83], which are based on an ε-subdifferential form of a global Karush–Kuhn–Tucker condition.

We next present a relaxation of the system (5), which is consistent also for near-optimal solutions in the primal and dual problems. We will use this system particularly when analyzing algorithms, such as Lagrangian heuristics, for integer programs later on in this paper. To this end, we introduce a nonnegative parameter κ, which defines the level of near-optimality allowed.

Given the pair $(x, u) \in X \times \Re^m_+$, we define the *relaxed global optimality conditions* for the problem (1) as

$$f(x) + u^{\mathrm{T}} g(x) \leq \theta(u) + \varepsilon, \tag{10a}$$
$$g(x) \leq 0^m, \tag{10b}$$
$$u^{\mathrm{T}} g(x) \geq -\delta, \tag{10c}$$
$$\varepsilon + \delta \leq \Gamma + \kappa, \tag{10d}$$
$$\varepsilon, \delta, \kappa \geq 0. \tag{10e}$$

We note immediately, with reference to the above result, that a consistent system (10) always has

$$\Gamma \leq \varepsilon + \delta \leq \Gamma + \kappa.$$

For this system, we state some immediate consequences in terms of the relations to near-optimal, and near-complementarity, solutions to the primal–dual problem.

PROPOSITION 5 (near-optimal solutions and the system (10)).

(a) (near-optimality in the primal problem (1)). *Let* $(x, u) \in X \times \Re^m_+$. *Suppose that, for some* $\varepsilon, \delta, \kappa \geq 0$, (10) *holds. Then,* x *is feasible in* (1), *and*

$$f(x) \leq \theta(u) + \Gamma + \kappa.$$

Suppose further that u *solves the dual problem* (3). *Then,*

$$f(x) \leq f^* + \kappa.$$

(b) (near-optimality in the Lagrangian subproblem (2)). *Suppose that* $(x, u) \in X \times \Re^m_+$ *is* β-*optimal and* α-*optimal, respectively, in the primal and dual problem* (1) *and* (3), *for some* β, $\alpha \geq 0$. *Then,*

$$\theta(u) \leq f(x) + u^{\mathrm{T}} g(x) \leq \theta(u) + \Gamma + \beta + \alpha.$$

Suppose further that $(x, u) \in X \times \Re^m_+$ *solves the primal and dual problems* (1) *and* (3), *respectively. Then,*

$$\theta^* \leq f(x) + u^{\mathrm{T}} g(x) \leq f^*. \tag{11}$$

(c) (near-complementarity). *Suppose that* $(x, u) \in X \times \Re^m_+$ *is* β-*optimal and* α-*optimal, respectively, in the primal and dual problem* (1) *and* (3), *for some* β, $\alpha \geq 0$. *Suppose further that* $\varepsilon \geq 0$ *is such that* (10a) *holds with equality. Then,* (10) *holds, with* $\delta := \Gamma - \varepsilon + \alpha + \beta \geq 0$ *and* $\kappa := \alpha + \beta$. *In fact,*

$$\Gamma + \beta + \alpha \leq \varepsilon + \delta \leq \Gamma + \kappa$$

always holds when (10) *is consistent.*

Suppose further that $(x, u) \in X \times \Re^m_+$ *solves the primal and dual problem* (1) *and* (3), *respectively. Then,* (10) *holds, with* $\delta := \Gamma - \varepsilon \geq 0$ *and* $\kappa = 0$. □

REMARK 6 (interpretations). The result in (a) states that vectors x that are near-optimal in the Lagrangian problem (2) and near-complementary also are near-optimal solutions to the primal problem (1), in particular so when the value of κ is small, that is, when the sum of the perturbations ε and δ are in the order of the size of the duality gap. It implies that the goal, when searching for a primal vector x which satisfies (5), should be to minimize κ, that is, essentially minimizing $\varepsilon + \delta$. A specialization of the result (a) to linear integer programming is found in [NeW88, Corollary II.3.6.9].

The result in (b) shows that a (near-)optimal solution to the primal problem (1) must also be near-optimal in the Lagrangian subproblem defined at a (near-)optimal dual solution. The example in Section 2.3 will show that either of (or neither of) the two inequalities in (11) may be tight for some optimal solutions. (In the case of equality constraints, the last inequality is always tight at optimal solutions.)

The result in (c) shows that a (near-)optimal solution to the primal problem (1) must also be near-complementary. It shows how closely related the two perturbations ε and δ are to the value of Γ, and it follows that the system (5) is always consistent at an optimal primal–dual solution. □

Theorem 3 also implies the following:

COROLLARY 7 (a primal characterization of near-optimality). *Let $u \in \Re^m_+$ be α-optimal in the dual problem* (3)*, for some $\alpha \geq 0$. Then,*

$$\{ x \in X \mid (10) \text{ is satisfied} \} = \begin{cases} X^{\kappa-\alpha}, & \text{if } \kappa \geq \alpha, \\ \emptyset, & \text{if } \kappa < \alpha \end{cases} \tag{12}$$

holds. □

REMARK 8 (interpretations). We characterize the optimal solution set X^* precisely when $\kappa = \alpha$. From this characterization, we see that primal optimal solutions can be obtained from non-optimal dual solutions, provided that the sum of the perturbations ε and δ matches precisely this non-optimality, in the sense that $\varepsilon + \delta = \Gamma + \kappa = \Gamma + \alpha$. (The sum is clearly unique, but not necessarily the values of ε and δ individually; cf. the example in Section 2.3.) Since this result is more general than Corollary 2, it follows that it is true also for convex problems.

Theorem 1 and Corollary 2 are special cases of Theorem 3 and Corollary 7, and follow when in addition $\Gamma = \varepsilon = \delta = \kappa = \alpha = 0$. Moreover, Corollary 4 is the special case of the above, for the case where $\kappa = \alpha = 0$.

The relation (12) combines the results of the above theorem with that of Proposition 5, in that we relate near-optimal solutions of the primal problem (1) and the dual problem (2) to each other. [One can of course also characterize the set of *dual* (near-)optimal solutions through a statement analogous to (12).] □

A special case of Corollary 7 to linear integer programming is found in [NeW88, Theorem II.3.6.7], however stated in terms of objective values. (That theorem traces back to results in Everett [Eve63], and is outlined in [Las70, Section 8.3.2].)

REMARK 9 (on the use of the system (5)). If the set X is discrete, then it is possible, in principle, to solve the problem (1) by enumerating the points in X according to an increasing value of the Lagrangian function $(x, u) \mapsto f(x) + u^{\mathrm{T}} g(x)$. Every time a feasible solution appears, we obtain an upper bound on f^*. It is easy to show that during the enumeration, the value of the Lagrangian function always underestimates the objective value of every feasible solution that still has not been found. In particular, if the enumeration continues until the value of the Lagrangian function becomes at least as large as the best upper bound found, then that solution is globally optimal. (As a special case, with the choice of $u = 0^m$, we recover the simple method, where the first feasible solution found is optimal, as then the enumeration is made in terms of the original cost.) If, on the other hand, the enumeration is terminated prior to this occurrence, then the terminal value of the Lagrangian function is a lower bound to f^*.

Handler and Zang [HaZ80] utilized a Lagrangian cost based ranking methodology to solve a knapsack constrained shortest path problem, from an optimal dual solution u. Recently, Caprara et al. [CFT02] have constructed a similar feasibility heuristic for a train timetabling problem. A similar methodology with the purpose of constructing core problems is developed in Section 5.1, for the case where X is a discrete Cartesian product set. □

2.2 Equality constraints

We next specialize the above main result to the case of equality constraints. So, suppose, locally in this section only, that (1b) is replaced by

$$h(x) = 0^{\ell}, \tag{1b}$$

where $h : \Re^n \mapsto \Re^\ell$ is continuous. The multiplier vector for these constraints is $v \in \Re^\ell$; the dual function $\theta : \Re^\ell \mapsto \Re$ is defined accordingly. The condition corresponding to (5) then is

$$
\begin{aligned}
f(x) + v^{\mathrm{T}} h(x) &\leq \theta(v) + \varepsilon, && (13a)\\
h(x) &= 0^\ell, && (13b)\\
0 \leq \varepsilon &\leq \Gamma. && (13c)
\end{aligned}
$$

THEOREM 10 (primal–dual optimality conditions). *Let $(x, v) \in X \times \Re^\ell$. Then, the following two statements are equivalent.*

(i) *Together with ε, the pair (x, v) satisfies the system* (13).

(ii) *The perturbation $\varepsilon = \Gamma$, and the pair (x, v) satisfies the following saddle-point like condition for the Lagrangian function $(x, v) \mapsto L(x, v) := f(x) + v^{\mathrm{T}} h(x)$ over $X \times \Re^\ell$:*

$$
L(x, w) \leq L(x, v) \leq L(y, v) + \varepsilon, \qquad \forall (y, w) \in X \times \Re^\ell.
$$

(iii) *x solves the primal problem* (1) *and v solves the dual problem* (3). □

COROLLARY 11 (a primal characterization of optimality). *Given any $v \in \Re^\ell$,*

$$
\{\, x \in X \mid (13) \text{ is satisfied} \,\} = \begin{cases} X^*, & \text{if } \theta(v) = f^* - \Gamma, \\ \emptyset, & \text{if } \theta(v) < f^* - \Gamma \end{cases}
$$

holds. □

The relaxed optimality conditions here are:

$$
\begin{aligned}
f(x) + v^{\mathrm{T}} h(x) &\leq \theta(v) + \varepsilon, && (14a)\\
h(x) &= 0^\ell, && (14b)\\
\varepsilon &\leq \Gamma + \kappa, && (14c)\\
\varepsilon, \kappa &\geq 0. && (14d)
\end{aligned}
$$

COROLLARY 12 (a primal characterization of near-optimality). *Let $v \in \Re^\ell$ be α-optimal in the dual problem* (3), *for some $\alpha \geq 0$. Then,*

$$
\{\, x \in X \mid (14) \text{ is satisfied} \,\} = \begin{cases} X^{\kappa - \alpha}, & \text{if } \kappa \geq \alpha, \\ \emptyset & \text{if } \kappa < \alpha \end{cases}
$$

holds. □

2.3 A numerical example

Consider the following linear integer programming problem:

$$
\begin{aligned}
f^* := \text{minimum} \quad & f(x) := -x_2, && (15a)\\
\text{subject to} \quad & g(x) := x_1 + 4x_2 - 6 \leq 0, && (15b)\\
& x \in X := \{\, x \in \mathcal{Z}^2 \mid 0 \leq x_1 \leq 4;\ 0 \leq x_2 \leq 2 \,\}. && (15c)
\end{aligned}
$$

The Lagrangian function associated with the dualization of the constraint (15b) with a multiplier $u \geq 0$ is $(x, u) \mapsto L(x, u) := ux_1 + (4u-1)x_2 - 6u$, and the dual problem has an objective function with the following form:

$$\theta(u) := \begin{cases} 2u - 2, & 0 \leq u \leq 1/4, \\ -6u, & 1/4 \leq u, \end{cases}$$

whose maximum over $\Re_+$ is attained at $u = 1/4$, with $\theta^* = \theta(u) = -3/2$.

The *linear relaxation* [$\Re^2$ replaces $\mathcal{Z}^2$ in (15c)] of the above problem has the same Lagrangian dual problem, and its primal solution is characterized by the system (4) as follows. At $u = 1/4$, the minimum of the Lagrangian function over the set $[0,4] \times [0,2]$ is the set $X(u) := \{ x \in \Re^2 \mid x_1 = 0;\ x_2 \in [0,2] \}$. [This is the set defined by (4a).] Together with primal feasibility [that is, (4b), or, in this case, (15b)], we obtain that x_2 is further restricted to be less than or equal to 3/2, while complementarity [that is, (4c)], forces x_2 to take on the value 3/2. So, from (4) we obtain that the primal–dual optimal solution set is the singleton set $\{(0, 3/2)^{\mathrm{T}}\} \times \{1/4\}$, with optimal (or, saddle) value $-3/2$.

Returning to the integer program, there are three optimal solutions, $x^1 = (0,1)^{\mathrm{T}}$, $x^2 = (1,1)^{\mathrm{T}}$, and $x^3 = (2,1)^{\mathrm{T}}$, with objective value $f^* = -1$. The duality gap is $\Gamma := f^* - \theta^* = 1/2$. In order to show how these optimal solutions will arise from an application of the system (5), we begin by noting that ε and δ must sum to $\Gamma = 1/2$, according to Theorem 3.

We first investigate the case where $\varepsilon = 0$, that is, the Lagrangian subproblem is solved exactly. Then, from (5a), we obtain that $X(u) = \{(0,0)^{\mathrm{T}}, (0,1)^{\mathrm{T}}, (0,2)^{\mathrm{T}}\}$. Primal feasibility [that is, (5b)] then dictates that x is either $(0,0)^{\mathrm{T}}$ or $(0,1)^{\mathrm{T}}$. Finally, we know that $u^{\mathrm{T}} g(x) \geq -1/2$ [that is, (5c)] since $\delta = 1/2$. The only primal vector of the two satisfying (5) with $\varepsilon = 0$ is $x^1 = (0,1)^{\mathrm{T}}$. (So, this solution violates the complementarity conditions.)

That the system (5) is consistent when $\varepsilon = 0$ is not necessarily the case; ε may need to take on positive values in order to reach an optimal solution. The optimal solution $x^2 = (1,1)^{\mathrm{T}}$ corresponds to letting $\varepsilon = \delta = 1/4$. In Figure 1, this optimal solution is contrasted with a non-optimal solution, and the values of ε and δ are given a further interpretation.

The optimal solution $x^3 = (2,1)^{\mathrm{T}}$ corresponds to letting $\varepsilon = 1/2$, while $\delta = 0$. This solution violates primal subproblem optimality even more than in the previous solution.

3 A dissection of Lagrangian heuristics

While Lagrangian heuristics are designed primarily to identify primal *feasible* solutions, a main goal is to also reach *near-optimal* solutions. We may interpret a Lagrangian heuristic as a procedure for attempting to satisfy the system (10); according to Theorem 3 and its corollary, the heuristic should also be designed to recover a primal feasible solution x such that the corresponding value of κ, hence of ε and δ, is small, relative to the size of the duality gap which their sum must not underestimate. If, and only if, this is possible, then near-optimal solutions to (1) are identified. Conservative Lagrangian heuristics (as defined in Section 1) for *equality* constrained problems will typically result in near-optimal solutions to (1), since making local moves will guarantee that subproblem near-optimality is retained; cf. Theorem 10.

We stress that for *equality* constrained problems, the value of ε is *fixed* at Γ for every optimal solution [statement (ii) in Theorem 10]. In the case of *inequality* constrained problems, however, only the sum $\varepsilon + \delta$ is fixed, at Γ [statement (ii) in Theorem 3]; the respective sizes of ε and δ may vary significantly, not only among problem instances, but even among optimal solutions for the same problem. (See the example in Section 2.3.) Therefore, whether a Lagrangian heuristic will be successful or not depends on several additional factors that are not so easy to determine the nature of in advance. For example, minimizing the value of ε in this context may result in an inconsistent system, thus making the heuristic fail to produce a feasible solution. This

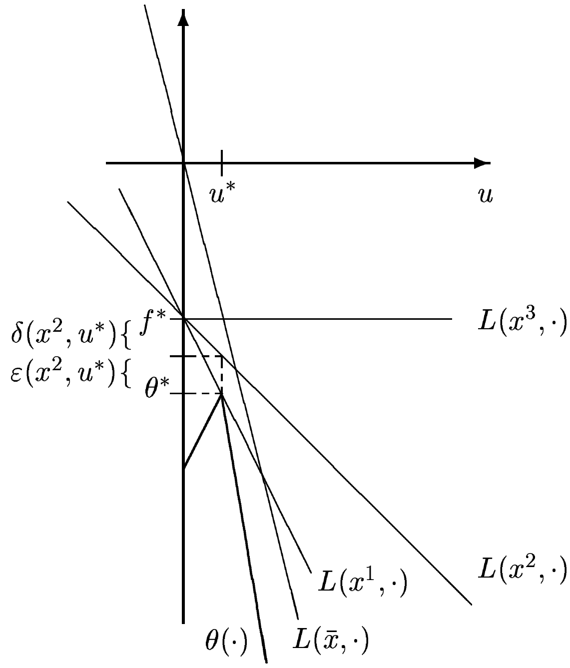


Figure 1: Illustration of the role of ε and δ in the characterization of optimality. For the optimal solution x^2, the value of $\varepsilon(x^2, u^*)$ equals the vertical distance between the two functions θ and $L(x^2, \cdot)$ at u^*. The remaining vertical distance to f^* equals minus the slope of the function $L(x^2, \cdot)$ at u^* [which is $g(x^2) = -1$] times u^*, that is, $\delta(x^2, u^*) = 1/4$. In the case of the candidate vector $\bar{x} := (2, 0)^{\mathrm{T}}$, the value of ε is $1/2$, and $\delta = 1$ [the slope of $L(\bar{x}, \cdot)$ at u^* is -4]; in this case, then, $\theta^* + \varepsilon + \delta = f(\bar{x}) = 0 > f^*$, so $\bar{x}$ cannot be optimal.

section collects some basic consequences of Theorem 3 in terms of the workings of a successful Lagrangian heuristic, depending on the type of problem being attacked through Lagrangian relaxation.

3.1 Small duality gap

We begin by describing the connection between the system (10) and conservative Lagrangian heuristics, according to the definition in Section 1. We have at hand some dual vector $u \in \Re^m_+$, which is α-optimal for some (unknown) $\alpha \geq 0$ in the problem (3). We attack the primal subproblem (2), obtaining a primal solution $\bar{x}(u) \in X$ which is ε_0-optimal in (2) for some (possibly unknown) $\varepsilon_0 \geq 0$. [Thus, we satisfy (10a), with $\varepsilon = \varepsilon_0$.] For future reference, we also introduce $\delta_0 := -u^{\mathrm{T}} g(\bar{x}(u)) \in (-\infty, \infty)$ to denote the level of complementarity fulfillment at $\bar{x}(u)$. If $\bar{x}(u)$ does not satisfy the relaxed constraints, then an attempt is made to attain primal feasibility through a manipulation of this primal solution, while, typically, remaining within the set X; sometimes, this manipulation is not terminated when a primal feasible solution has been found, but is instead followed by a primal local search heuristic. If successful, the result of this heuristic projection of the infeasible solution $\bar{x}(u)$ onto the feasible set is a feasible vector, $\bar{x}$. This vector is associated with the values $\varepsilon \geq 0$ and $\delta := -u^{\mathrm{T}} g(\bar{x}) \geq 0$, satisfying (10a) and (10c). The vector $\bar{x}$ is further β-optimal in (3), where $\beta \geq 0$ satisfies the relation in Corollary 7. [Obviously, $\beta \leq f(\bar{x}) - \theta(u)$.]

Suppose that Γ and α both are close to zero. Since $\varepsilon + \delta \leq \Gamma + \kappa$ holds in any solution to the system (10), according to Proposition 5(c), both ε and δ are then known to be close to zero at near-optimal primal solutions. Then, the near-optimal primal solutions all lie in a small neighbourhood [in the sense of the value of $L(\cdot, u)$] of the subproblem solution $x(u)$; consequently, when $\varepsilon_0 = 0$ holds, it is then sufficient to consider a feasibility heuristic which is conservative in the sense of the adjustment in the value of $L(\cdot, u)$. Since δ is small, the heuristic must also be able to ensure that the complementarity violation is kept down. As it is possible that complementarity is violated to a large degree at $x(u)$, the feasibility heuristic must in fact be designed so that it is able to reduce the value of δ substantially, if necessary.

Conclusion When the dual solution at hand is near-optimal and the duality gap is small, then in order to be able to find near-optimal primal feasible solutions, it is always *sufficient* to consider Lagrangian heuristics that are conservative in the value of the Lagrangian function, provided that they are also able to reduce any complementarity violations, if necessary. (In the case of equality constraints, the last remark may be stricken.)

If the value of ε_0 is large, then it is necessary that feasibility and Lagrangian optimality are improved simultaneously in the heuristic. This case might arise, for example, if the Lagrangian relaxed problem is a difficult discrete optimization problem.

Some examples of specialized conservative heuristics which have yielded very good results for certain combinatorial problems are given in [BGB81, JLV90, CaM91, Fis94].

A special case of the above is *convex programming*, where $\Gamma = \varepsilon = \delta = 0$ (if $\alpha = 0$). Although it is not easy to state exactly what a well-designed Lagrangian heuristic should be, it is quite simple to find examples of what it should *not* be. A feasibility heuristic which is not based on an optimization in the original objective or on a Euclidean-like projection operation may clearly result in large values of ε, and therefore in solutions of low quality when applied at near-optimal dual solutions.

An example problem where this has been observed is in dual solution procedures for strictly convex minimum cost network flow problems. The feasible set is then of the form $\{\, \ell \leq x \leq c \mid Ax = b \,\}$, where x_j is the flow on a directed link j which is subject to bounds and flow conservation constraints, $A \in \{-1, 0, 1\}^{m \times n}$ and $b \in \Re^m$ being the node–link incidence matrix and the demand vector, respectively. The objective function is often modelled as separable, that is, of the form $f(x) = \sum_{j=1}^{n} f_j(x_j)$, where each function $f_j : \Re \mapsto \Re$ is strictly convex and coercive. A Lagrangian relaxation of the flow conservation constraints leads to a strictly convex Lagrangian dual problem, with a multiplier for each node in the network. The classic algorithm for this problem is the Gauss–Seidel (coordinate ascent) algorithm, which amounts to balancing the flow through one node at a time by a line search in the associated multiplier, cf., for example, [BeT89, Chapter 5]. We add that due to strict convexity, at any dual vector u^* the (unique) Lagrangian subproblem solution is a (unique) solution to the original problem. Few articles are devoted to the generation of primal feasible flows in this application. One of them is Ventura [Ven91]. From a dual solution, u^t, his Lagrangian heuristic works as follows. Unless u^t is optimal, the resulting subproblem solution, $x(u^t)$, does not satisfy all the flow conservation constraints. A *linear* network flow problem is then constructed, where the demand vector is the residual $b' := b - Ax(u^t)$, and where the linear cost vector is $\nabla f(x(u^t))$. Letting x' be a solution to this problem, the vector $\bar{x} := x(u^t) + x'$ is feasible in the original problem. We note, however, that the quality of this solution may be poor, since the use of a linear cost makes the solution extremal, whereas the original problem will typically have an optimal solution in the relative interior of the feasible set. In the notation of this paper, we conclude that the heuristic is not conservative (ε may be large), since the size of x' can still be substantial even when we are close to a dual optimum (and, further, it does not tend to zero as $\{u^t\} \to u^*$). This observation lead Marklund [Mar93], in a master's project supervised by the authors, to devise heuristic projections based on conservative node imbalance-reducing graph search techniques, which in comparison yield feasible flows of a much better quality. (See further [Pat94, Chapter 4.3] for the so far only published account of the procedures in [Mar93].)

3.2 Large duality gap

We take as a starting point the inequality constrained case. Suppose that Γ, and possibly also α, is large. Then, in general, since we do not know beforehand whether the identity (assuming that $\beta = 0$) $\varepsilon + \delta = \Gamma + \alpha$ requires ε or δ, or both, to be large, we cannot guarantee that a conservative heuristic (in terms of the value of ε) will be able to produce near-optimal solutions to (1). (In some cases it *may* be successful, since violating complementarity could compensate for a small value of ε. In the equality constrained case, however, conservative heuristics cannot yield feasible solutions.) This implies that it is necessary to choose a heuristic which is radical in that it allows ε to become, or remain, large.

Further, since the value of δ is unknown, the heuristic must also be radical in the sense of allowing, if necessary, δ to take on both small and large values in order to reach good solutions.

Conclusion When the dual solution at hand is far from being optimal (that is, the value of α is large), or the duality gap is large, then in order to be able to obtain near-optimal primal feasible solutions, it is *necessary* to consider Lagrangian heuristics that are radical with respect to the value of the Lagrangian function and allow for both small and large complementarity violations. (In the case of equality constraints, the last remark may be stricken.)

Because of the radical, and therefore global, nature of the above type of heuristic, it may be more appropriate to think of them as being global heuristic optimization procedures, as opposed to the local nature of conservative heuristics. One example of a radical type of heuristic is very large-scale neighbourhood search (see [AEOP02]). Another example of a heuristic which may be designed to be radical in our sense is the class of *greedy algorithms* for discrete optimization. Another type of technique which may serve as a radical Lagrangian heuristic is available in integer optimization problems with two groups of variables, such as mixed-integer linear programs: the *Benders subproblem* (e.g., [Las70]). Suppose that $x = (x_1, x_2)$ where the subvector x_1 is required to be integer valued whereas the subvector x_2 may take on fractional (continuous) values. Such models are frequent in design type problems, where the integer variables are associated with design decisions, while the continuous variables are associated, for example, with network flows. The Lagrangian heuristic is to solve the (linear) Benders subproblem for the original problem, which here means that the original objective function is optimized over the continuous variables x_2 while the integer variables x_1 are fixed to their values $\bar{x}_1(u)$ from the Lagrangian subproblem. Typically, the Lagrangian relaxation is such that the resulting solution $\bar{x}_1$ will be feasible; otherwise, the integer solution may have to be heuristically updated and the LP resolved. This heuristic clearly qualifies as a radical heuristic, as it allows for large moves in x_2, and therefore also in the value of ε. (Depending on the problem instance, even a large move in some variables can amount to a small adjustment in the value of ε, however.) A successful case is given in [MuC79] for a location problem.

4 Experiments on the set covering problem

4.1 The set covering problem and its dual

The set covering problem is to find

$$f^* := \text{minimum} \sum_{j=1}^{n} c_j x_j, \tag{16a}$$

$$\text{subject to} \sum_{j=1}^{n} a_j x_j \geq 1^m, \tag{16b}$$

$$x \in \{0,1\}^n, \tag{16c}$$

where $c_j \in \Re$ and $a_j \in \{0,1\}^m$, $j = 1, \ldots, n$. Its Lagrangian with respect to the relaxation of the linear constraints (16b) has the form $L(x,u) := (1^m)^{\mathrm{T}} u + \bar{c}^{\mathrm{T}} x$, $u \in \Re^m$, where we have defined the reduced cost vector $\bar{c} := c - A^{\mathrm{T}} u$. Here, $c = (c_j)_{j=1}^n$ and $A = (a_1\, a_2\, \cdots\, a_n)$.

We define the Lagrangian dual problem to find

$$\begin{aligned} \theta^* :=\ & \text{maximum} \quad \theta(u), \\ & \text{subject to} \quad u \geq 0^m, \end{aligned}$$

where

$$\theta(u) := (1^m)^{\mathrm{T}} u + \sum_{j=1}^{n} \underset{x_j \in \{0,1\}}{\text{minimum}}\ \bar{c}_j x_j, \qquad u \geq 0^m$$

is the dual function; the Lagrangian subproblem is of course solved such that

$$x_j(u) \begin{cases} = 1, & \text{if } \bar{c}_j < 0, \\ \in \{0,1\}, & \text{if } \bar{c}_j = 0, \\ = 0, & \text{if } \bar{c}_j > 0. \end{cases} \tag{17}$$

In the following, we will report upon two experiments with Lagrangian heuristics for set covering problems. Our main goal with these experiments is to illustrate the potential of utilizing our theoretical findings, in particular by considering also complementarity fulfillment in Lagrangian heuristics for such problems. Throughout, we have worked with the set covering problem `rail507`, for which we have the following data: $n = 63,009$, $m = 507$, and the best bounds reported in the literature are $[172.1456, 174]$ (see [AMRT01] and [CFT99], respectively; the lower bound of 172.4 reported in [CNS98] is probably incorrect.) We note that the experiments that we have performed on the similar set covering problems `rail516` and `rail582` corroborate the conclusions that are made below in Sections 4.4 and 4.5.

4.2 A generic primal greedy heuristic

A primal greedy heuristic is often a main component of a set covering algorithm. In our experiments, we will use several such algorithms, some of which are classic, and all of which can be written (at least essentially) as instances of the following generic primal greedy heuristic:

(**Input**) A primal vector $\bar{x} \in \{0,1\}^n$ and a cost vector $p \in \Re^n$.

(**Output**) A vector $\hat{x} \in \{0,1\}^n$, feasible in (16).

(**Starting phase**) Given $\bar{x}$, delete all rows i in (16b) that are covered. Delete all variables x_j with $\bar{x}_j = 1$.

(**Greedy insertion**) Identify an undeleted variable x_τ which has the minimal value of p_j relative to the number (k_j) of uncovered, undeleted rows which it covers. Set $x_\tau := 1$. Delete all rows in (16b) which have been covered. Delete x_τ. If any uncovered rows in (16b) remain, then repeat this step; otherwise, let $\tilde{x} \in \{0,1\}^n$ denote the feasible solution found.

(**Greedy deletion in over-covered rows**) Identify a variable x_τ with $\tilde{x}_\tau = 1$ which is present only on rows which are over-covered at $\tilde{x}$, and which has the maximal value of p_j relative to k_j. Set $\tilde{x}_\tau := 0$. If any such variable remains, then repeat this step; otherwise, let $\hat{x} \in \{0,1\}^n$ denote the feasible solution found, and terminate.

We can identify several known instances of the above algorithm:

(I) Let $\bar{x} := 0^n$ and $p := c$. This procedure is described by Chvátal [Chv79].

(II) Let $\bar{x} := 0^n$ and $p := \bar{c}$, defined at some dual vector u (see above). This is essentially the heuristic PRIMAL5 of Balas and Ho [BaH80].

(III) Let $\bar{x} := x(u)$, whose component $x_j(u)$ is given by (17), and let $p := c$. This heuristic is described by Beasley [Bea87, Bea93] and Wolsey [Wol98, Section 10.4].

(IV) Let $\bar{x} := x(u)$ and $p := \bar{c}$. This is essentially the heuristic ERCGH of Balas and Carrera [BaC96].

The greedy selection criterion utilized in our description above is often of the form p_j/k_j, but can combine the entities p_j and k_j in different ways; such versions of the procedure (I) can be found in Balas and Ho [BaH80], and Vasko and Wilson [VaW84], where c_j/k_j is replaced by, among other choices, $c_j/\log_2 k_j$. (See [BaC96, Section 4] for an account of numerical experience with such heuristics.)

4.3 A simple dual algorithm

To find a good lower bound on θ^*, we apply *conditional subgradient optimization* ([LPS96]), using Polyak [Pol69] step lengths, that is, starting from a $u^0 \in \Re^m_+$,

$$u^{t+1} := \left[u^t + \ell_t \gamma^t_+(u^t)\right]_+, \qquad t = 0, 1, \dots,$$

where $[\cdot]_+$ denotes the Euclidean projection onto $\Re^m_+$,

$$(\gamma^t_+(u^t))_i := \begin{cases} 0, & \text{if } u^t_i = 0 \text{ and } a^i x(u^t) > 1, \\ 1 - a^i x(u^t), & \text{otherwise,} \end{cases} \qquad i = 1, 2, \dots, m,$$

is a *projected subgradient* of θ with respect to $\Re^m_+$ at u^t, a^i being row i of the matrix A, and where the step length is

$$\ell_t := \nu_t \frac{UBD - \theta(u^t)}{\|\gamma^t_+(u^t)\|^2}, \qquad t = 0, 1, \dots,$$

where the parameter $\nu_t := 1.5 \cdot 0.99^t$, $t = 0, 1, \dots$.

The upper bound *UBD* used in the step length formula was calculated a priori by applying the greedy heuristic (I) above. For the instance `rail507`, the primal heuristic (I) produces $UBD = 209$. (The feasible solution found after the greedy insertion phase has the objective value 216.)

This subgradient algorithm is, in our experiments, terminated after a fixed number of iterations, a number which is altered among the experiments. The dual vector u chosen at termination was the final iteration point.

4.4 Experiment I

We have little information a priori about the violation of the complementarity conditions and subproblem optimality at an optimal solution. To ensure that all possibilities are considered we will use an objective in the heuristic search which combines the Lagrangian function $[L(\cdot, u)]$ and complementarity fulfillment $[-u^{\mathrm{T}}g(\cdot)]$, that is,

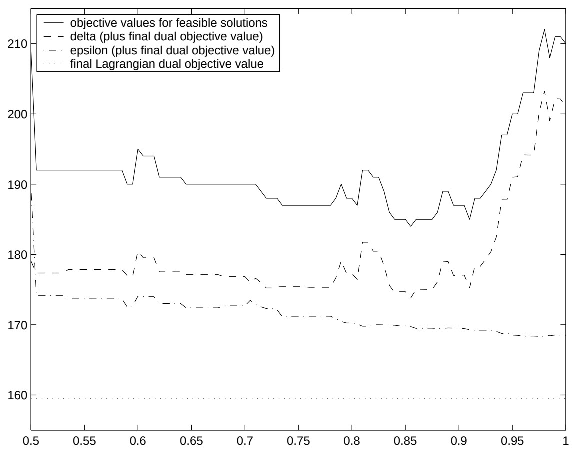
$$h(x) := \lambda[f(x) + u^{\mathrm{T}}g(x)] + (1-\lambda)[-u^{\mathrm{T}}g(x)], \qquad 1/2 \le \lambda \le 1. \tag{18}$$

In the set covering application, $h(x) = [\lambda\bar{c} + (1-\lambda)A^{\mathrm{T}}u]^{\mathrm{T}}x$ holds. We obtain the original cost c by the choice $\lambda := 1/2$, while the Lagrangian reduced cost $\bar{c}$ follows from the choice $\lambda := 1$. As remarked above, we will also consider values in between.

In order to motivate the lower bound of 1/2 on λ in (18), consider the problem to minimize $f(x)$ over $\{ x \in X \mid (10) \text{ is satisfied} \}$, for a given $u \in \Re^m_+$. If κ is chosen large enough, then this problem has X^* as its solution set. The objective in (18), which can be rewritten as $\lambda f(x) + (2\lambda - 1)u^{\mathrm{T}}g(x)$, is equivalent to the Lagrangian function corresponding to the relaxation of (10a) with a multiplier μ and of (10c) with a zero multiplier; the condition $\lambda \ge 1/2$ then stems from the requirement that $\mu \ge 0$.

Based on the generic primal heuristic of Section 4.2, we then define a set of heuristics which use a cost vector of the form $p := \lambda\bar{c} + (1-\lambda)A^{\mathrm{T}}u$. Note that all four instances (I)–(IV) described above use cost coefficients that are defined at the end-points of the interval for λ: (I) and (III) correspond to $\lambda := 1/2$, while (II) and (IV) correspond to $\lambda := 1$. In this first experiment, we define $\bar{x} := 0^n$, so the heuristics in this first test are all radical, according to our definition; the experiments in the next subsection also look at more conservative heuristics, where $\bar{x} := x(u)$.

We ran two tests, the first with $t = 200$ as the final iteration, and the second with $t = 500$. For each of these two final values, we ran the above heuristic with values of λ in $[1/2, 1]$, with an increment of 0.005. (We also ran the same problem with values of λ less than 1/2, but the solutions obtained were inferior.) The results are shown in Figure 2.



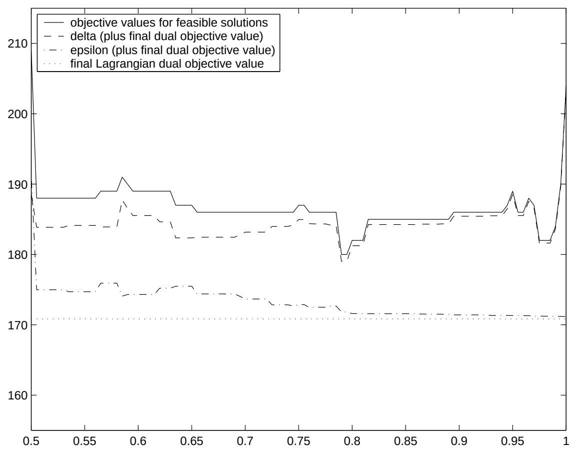


Figure 2: The Lagrangian heuristic for the set covering problem `rail507`. (a) $t = 200$; (b) $t = 500$.

Looking at Figure 2(a), the horizontal, dotted line is the value of the last lower bound found, which in this case was $\theta(u^{200}) = 159.53$. The three other lines, taken from the highest to the lowest, show the objective values of the feasible solutions obtained by the proposed Lagrangian heuristic, the values of δ, and the values of ε, as the value of λ ranges from 1/2 to 1. The scale on the y-axis only applies to the original cost; for the latter two lines, the value zero corresponds to the dotted line. This line is, for Figure 2(b), at the level $\theta(u^{500}) = 170.84$. Both figures also illustrate that $f(x) = \theta(u) + \delta + \varepsilon$.

After 200 iterations, the quality of the lower bound is quite poor, and running the primal heuristic from better dual solutions, we have found that better primal solutions are then always provided.

According to the Figures 2(a) and (b), the value of ε clearly decreases with an increase in the value of λ, which is expected, since for $\lambda = 1$ the merit function used is the Lagrangian. The variation in δ (which always dominates in value over ε here) is less regular, except for larger values of λ when it increases rapidly, again as expected.

The value of ε is very small for solutions that are of high quality, meaning that near-optimal solutions to this set covering problem violate complementarity to a large extent. From runs with the heuristics used in our experiments on other, similar, instances of the set covering problem, we have experienced a similar behaviour. It has indeed been observed that, for this class of problems, often several rows are over-covered in an optimal solution ([Tak01]).

The result of applying the heuristic (I) is a feasible solution with the cost 209; this corresponds to the height of the uppermost line at $\lambda = 1/2$ in both figures. The result of using the heuristic (II) can be seen as the height of the uppermost line at $\lambda = 1$ in both figures (with the objective values 210 and 204, respectively). Clearly, both of these choices are inferior to using values of λ in the open interval $(1/2, 1)$. The best solutions are found for relatively large values of λ, as long as they are not very close to 1. At both endpoints of the interval for λ, we can observe from the appearance of (18) that the violation of complementarity is ignored. The above observations lead us to advocate the use of heuristics based neither on the original cost, nor on the Lagrangian cost, but on a combination of the two of them, because this combination does take complementarity violation into consideration.

The following experiment takes these observations as its starting point.

4.5 Experiment II

Based on the previous experiment, we chose to set $\lambda := 0.9$, and performed a second experiment. We ran three primal greedy heuristics at every iteration of the dual algorithm, starting from $t = 200$ and terminating at $t = 500$. For each of these, we recorded the objective value of the feasible solution $\hat{x}$ obtained, and created histograms, as can be seen in Figure 3. The top one was obtained by using the heuristic (III), that is, starting at the solution $\bar{x} := x(u^t)$ and using the original cost coefficients, $p := c$. This is a conservative heuristic. The middle one was obtained by the instance of the generic heuristic where $\bar{x} := x(u^t)$ but where $p := \lambda\bar{c}^t + (1-\lambda)A^{\mathrm{T}}u^t$ with $\lambda = 0.9$. This is also a conservative heuristic, which however is based on a better cost function. The bottom one, finally, was created by the use of the primal heuristic which also uses this cost vector p, but which is radical because it takes $\bar{x} := 0^n$ as the starting point.

We observe from the figure a quite remarkable difference especially between the two conservative heuristics and the radical one which consistently provides feasible solutions of rather high quality. For each of the three respective histograms, we have the following minimum, mean, and maximum objective values:

$$(192, 203.99, 221); \qquad (182, 194.45, 212); \qquad (182, 186.55, 195).$$

The radical heuristic produces solutions the worst of which is nearly as good as the best outcome of the heuristic (III).

We have also, in Figure 4, for each iteration plotted the moving average of the objective values over 30 iterations. (The value plotted for iteration 230 corresponds to the average of the iterations 201–230.) Figure 4 reveals that the radical heuristic provides relatively good primal solutions already at an early stage of the dual algorithm, and it clearly improves upon the other

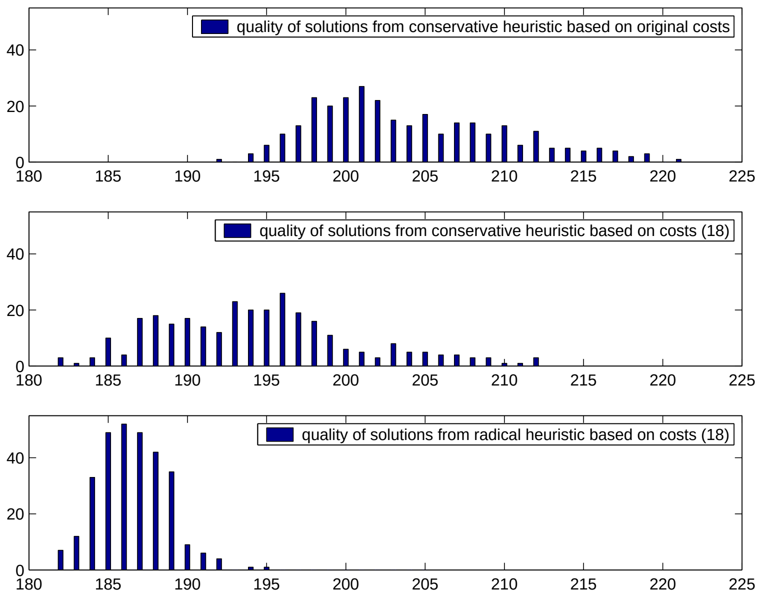


Figure 3: Quality of primal solutions obtained by three greedy heuristics. (a) $\bar{x} := x(u^t)$ and $p := c$ [heuristic (III)]; (b) $\bar{x} := x(u^t)$ and $p := 0.9\bar{c}^t + 0.1A^{\mathrm{T}}u^t$; (c) $\bar{x} := 0^n$ and $p := 0.9\bar{c}^t + 0.1A^{\mathrm{T}}u^t$.

two. Further, the second of the conservative heuristics (which is also new) is in turn much better than the first, while being a very simple modification thereof.

While the above experiments were certainly not performed in order to establish the superiority of these new Lagrangian heuristics for solving large-scale set covering problems, the result is encouraging for their use, especially when taking into account how simple it was to incorporate them into the well established greedy strategy for set covering problems.

5 Applications to column generation and core problems

5.1 Column generation

The principle of column generation is most frequently used for attacking discrete optimization problems. (See, for example, the recent survey in [Wil01], and [Wol98, Chapter 11].) Since, however, column generation is founded on linear programming duality, it is merely a continuous relaxation that can be solved by means of this principle. The results to be presented below introduce a certain control over the integer programming quality of a column generation scheme.

Consider a discrete optimization problem with a feasible region that is defined by a finite Cartesian product set and a number of linear and coupling side constraints, that is, a problem of the form

$$f^* := \text{minimum} \sum_{j=1}^{n} c_j^{\mathrm{T}} x_j, \tag{19a}$$

$$\text{subject to} \sum_{j=1}^{n} A_j x_j \geq b, \tag{19b}$$

$$x_j \in X_j, \qquad j = 1, \dots, n, \tag{19c}$$

where the sets $X_j \subset \Re^{n_j}$, $j = 1, \dots, n$, are finite, $c_j \in \Re^{n_j}$ and $A_j \in \Re^{m \times n_j}$, $j = 1, \dots, n$, and $b \in \Re^m$. The problem is assumed to have a feasible solution.

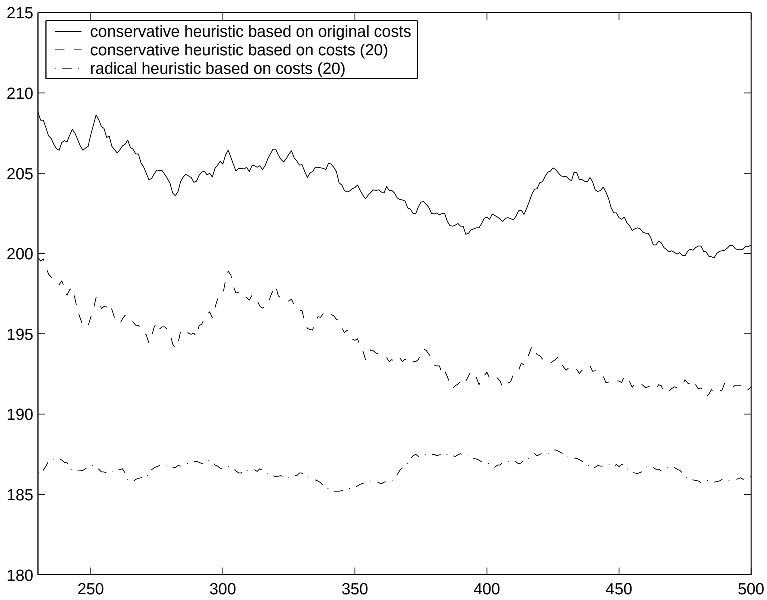


Figure 4: Moving averages of the solution quality for three greedy heuristics.

In many applications which give rise to a model of this form, the sets X_j, $j = 1, \ldots, n$, are described by linear constraints and integrality restrictions. The result to be presented below does not require this description to have the integrality property. Further, the objective function and the side constraints are stated as being linear in order to ease the presentation only. (In the case of nonlinearities, additive separability over the Cartesian product is however required.)

Denote, for $j = 1, \ldots, n$, by P_j the number of points in the set X_j, and denote these points by x_j^i, $i = 1, \ldots, P_j$. The problem (19) is then equivalent to the disaggregated master problem

$$f^* = \text{minimum} \sum_{j=1}^{n} \sum_{i=1}^{P_j} \left(c_j^{\mathrm{T}} x_j^i\right) \lambda_j^i, \tag{20a}$$

$$\text{subject to} \sum_{j=1}^{n} \sum_{i=1}^{P_j} \left(A_j x_j^i\right) \lambda_j^i \geq b, \tag{20b}$$

$$\sum_{i=1}^{P_j} \lambda_j^i = 1, \qquad j = 1, \ldots, n, \tag{20c}$$

$$\lambda_j^i \in \{0, 1\}, \qquad i = 1, \ldots, P_j, \qquad j = 1, \ldots, n. \tag{20d}$$

Let $u \in \Re_+^m$ be multipliers associated with the side constraints (19b), define the Lagrangian subproblem

$$\theta(u) := b^{\mathrm{T}} u + \sum_{j=1}^{n} \theta_j(u), \tag{21}$$

with

$$\theta_j(u) := \underset{x_j \in X_j}{\text{minimum}} \left(c_j^{\mathrm{T}} - u^{\mathrm{T}} A_j\right) x_j, \qquad j = 1, \ldots, n,$$

and suppose that $\bar{u} \in \Re_+^m$ is near-optimal in the Lagrangian dual problem to maximize the value of $\theta(u)$ over $u \in \Re_+^m$.

Suppose further that $p_j \geq 1$ points in the respective sets X_j, $j = 1, \dots, n$, are available explicitly, let $\delta \in \Re_+$, and consider the following restricted master problem.

$$f_r^* := \text{minimum} \sum_{j=1}^{n} \sum_{i=1}^{p_j} \left(c_j^{\mathrm{T}} x_j^i\right) \lambda_j^i, \tag{22a}$$

$$\text{subject to} \quad \sum_{j=1}^{n} \sum_{i=1}^{p_j} \left(A_j x_j^i\right) \lambda_j^i \geq b, \tag{22b}$$

$$\sum_{j=1}^{n} \sum_{i=1}^{p_j} \left(\bar{u}^{\mathrm{T}} A_j x_j^i\right) \lambda_j^i \leq \bar{u}^{\mathrm{T}} b + \delta, \tag{22c}$$

$$\sum_{i=1}^{p_j} \lambda_j^i = 1, \qquad j = 1, \dots, n, \tag{22d}$$

$$\lambda_j^i \in \{0, 1\}, \qquad i = 1, \dots, p_j, \qquad j = 1, \dots, n. \tag{22e}$$

The purpose of this nonstandard formulation of a restricted master problem is that any feasible solution to it will satisfy near-complementarity [cf. (5c)].

This problem can be built up by, for example, enumerating (ranking) points in the product sets according to increasing objective values, or by applying column generation (to optimality or truncated) to the linear programming relaxation of the master problem (20). Another alternative is to apply subgradient optimization to the above defined Lagrangian dual problem, and during the course of this scheme accumulate optimal solutions to the relaxed problems (all of them or some only). (The use of subgradient optimization for accumulating columns to a linear programming restricted master problem is justified by, for example, [LaL97, Proposition 7]; see also [PeP97, LPS98, LPS99, LPS03] for similar results for subgradient optimization applied to more general problems, and [Kiw95, FeK00] for primal recovery results using proximal dual subgradient methods.) Still another alternative for creating a restricted master problem is to enumerate the points in the product sets according to the reduced costs that are obtained within the column generation or Lagrangian relaxation approaches (cf. the discussion in Remark 9). One might of course also consider combinations of all or some of these strategies. [Note that the result to be given below is valid whichever principle has been used for constructing the restricted master problem (22).]

We mention as a special example of the above the restricted master problem defined by Sweeney and Murphy [SwM79] in their decomposition method for integer programs; they built their restricted master problem by enumerating the vectors in the respective sets X_j according to a ranking based on Lagrangian reduced costs. [Their master problem is however the standard restriction of the problem (20).]

In order to state the main result, we define the (Lagrangian) reduced costs

$$\bar{c}_j^i := (c_j^{\mathrm{T}} - \bar{u}^{\mathrm{T}} A_j) x_j^i - \theta_j(\bar{u}), \qquad i = 1, \dots, p_j, \qquad j = 1, \dots, n, \tag{23}$$

for the variables in the restricted master problem. (Note that $\bar{c}_j^i \geq 0$ always holds.)

THEOREM 13 (quality of restricted master problem). *Let $\bar{u} \in \Re_+^m$ and $\delta \in \Re_+$. Suppose that the restricted master problem* (22) *has a feasible solution. Then,*

$$f_r^* \leq \theta(\bar{u}) + \varepsilon + \delta,$$

where

$$\varepsilon := \sum_{j=1}^{n} \max_{i=1,\dots,p_j} \bar{c}_j^i.$$

PROOF. We utilize Proposition 5(a), as follows. Let $x_j = x_j^{i(j)}$, where $i(j) \in \{1, \dots, p_j\}$, for $j = 1, \dots, n$, be a feasible solution to the problem (19) that corresponds to an optimal solution to (22). Denote this solution by $\bar{x}$. By using (23) and (21) we then have that

$$\begin{aligned} L(\bar{x}, \bar{u}) &:= b^{\mathrm{T}} \bar{u} + \sum_{j=1}^{n} \left(c_j^{\mathrm{T}} - \bar{u}^{\mathrm{T}} A_j \right) x_j^{i(j)} = b^{\mathrm{T}} \bar{u} + \sum_{j=1}^{n} \left(\theta_j(\bar{u}) + \bar{c}_j^{i(j)} \right) \\ &\leq \theta(\bar{u}) + \sum_{j=1}^{n} \max_{i=1,\dots,p_j} \bar{c}_j^i = \theta(\bar{u}) + \varepsilon. \end{aligned}$$

Hence, (10a) holds. Further, (10b) holds by assumption. Finally, with $g(x) := b - \sum_{j=1}^{n} A_j x_j$, (22c) gives that $\bar{u}^{\mathrm{T}} g(\bar{x}) \geq -\delta$ holds, whence (10c) follows. Hence, the pair $(\bar{x}, \bar{u})$ satisfies (10), and the conclusion follows. □

In the case where the inequality constraints (19b) are replaced by equalities, the restricted master problem is modified accordingly, (22c) is not present, and $\delta = 0$.

The theorem immediately implies that

$$f_r^* - f^* \leq \left(\sum_{j=1}^{n} \max_{i=1,\dots,p_j} \bar{c}_j^i \right) + \delta$$

holds.

The below result provides a (limited) possibility to assess the quality of the restrification which lead to the restricted master problem. It can also be used as a guide to its adjustment; especially, it describes a property of the restricted master problem such that its feasible set contains an optimal solution to (19). Its proof is rather straightforward, and is therefore omitted.

PROPOSITION 14 (variable fixing and optimality). *Let $\bar{u} \in \Re_+^m$. Suppose that the restricted master problem* (22) *has a feasible solution, and let $\bar{x}$ be the feasible solution to* (19) *that corresponds to an optimal solution to* (22). *Suppose further that we know an upper bound $\bar{f} \geq f^*$ (for example, $\bar{f} = f_r^*$).*

(a) *If a $j \in \{1, \dots, n\}$ is such that $\bar{c}_j^i > \bar{f} - \theta(\bar{u})$ holds for $i = p_j + 1, \dots, P_j$, then $x_j^* = \bar{x}_j$ in every optimal solution x^* to* (19).

(b) *If $\delta \geq \bar{f} - \theta(\bar{u})$ and if $\bar{c}_j^i \geq \bar{f} - \theta(\bar{u})$ holds for $i = p_j + 1, \dots, P_j$ and for every $j \in \{1, \dots, n\}$, then $x^* = \bar{x}$ is an optimal solution to* (19). □

REMARK 15 (observations). The obvious way to tighten the upper bound in the theorem is to delete from the restricted master problem the columns with maximal reduced costs (23), and to decrease the value of δ; such further restrifications of (22) might however cause the problem to become infeasible. Note however that the continuous relaxation of (22) always is feasible.

The columns corresponding to the solutions to the relaxed problems

$$\theta_j(\bar{u}) := \underset{x_j \in X_j}{\text{minimum}} \left(c_j^{\mathrm{T}} - \bar{u}^{\mathrm{T}} A_j \right) x_j, \qquad j = 1, \dots, n,$$

do not need to be present in (22) [although it is of course quite natural that they are].

Whenever a feasible solution to (19) is known, it can be utilized to define vectors in X_j and a value of δ such that the restricted master problem (22) always has a feasible solution. □

We end with a corollary to the above result, which stems from the application of column generation to the LP relaxation of (20). Suppose we have the solved to (near-)optimality the linear program that is the LP relaxation of the restricted master problem (22), that we have a primal basic feasible solution (BFS) to this problem with objective value f_{RMP}, and that $(\bar{u}, \bar{v})$ is a complementary dual solution. Let the best column obtained in the column generation phase have the reduced cost $\bar{c}_j^{p_j+1}$ $(j = 1, \ldots, n)$. (If the restricted master problem was solved to optimality and the current solution is not optimal, then this corresponds to a column not previously generated.) Consider then the quality of the next restricted master problem:

COROLLARY 16 (solution quality in column generation). *In the current setting, we have the estimate*

$$f_{\mathrm{RMP}} + \sum_{j=1}^{n} \bar{c}_j^{p_j+1} \leq f^* \leq f_r^* \leq f_{\mathrm{RMP}} + \sum_{j=1}^{n} \left(\bar{c}_j^{p_j+1} + \max_{i=1,\ldots,p_j} \bar{c}_j^i \right) + \delta$$

of the optimal value. □

5.2 Core problems

The formulation and solution of core problems is a more sophisticated means to utilize Lagrangian duality to find primal feasible and near-optimal solutions, compared to simpler, manipulative heuristics such as those that were presented in Section 4. As such, core problems are special Lagrangian heuristics which may or may not be conservative, depending on the principle with which the core problem is defined, and the size of a resulting core problem. Core problems lie at the heart of the set covering heuristics in [CNS98, CFT99], and efficient "core algorithms" also exist for binary knapsack problems (e.g., [BaZ80, MaT90, Pis95]). These core problems, and their optimization, are devised primarily on the (linear programming-)reduced costs of the variables, and so they can be said to focus on Lagrangian near-optimality, as opposed to also incorporating complementarity near-fulfillment. Our below analysis of core problems in relation to the global optimality conditions, suggests that complementarity near-fulfillment can, and should, be introduced in core problems.

Consider the binary problem

$$f^* := \text{minimum} \sum_{j=1}^{n} c_j x_j, \tag{24a}$$

$$\text{subject to} \sum_{j=1}^{n} a_j x_j \geq b, \tag{24b}$$

$$\sum_{j=1}^{n} d_j x_j \geq e, \tag{24c}$$

$$x \in \{0,1\}^n, \tag{24d}$$

where b, a_j $(j = 1, \ldots, n) \in \Re^m$ and e, d_j $(j = 1, \ldots, n) \in \Re^r$. Suppose that the problem is feasible. We propose solving this problem by means of a Lagrangian relaxation of the (complicating) constraints (24b), the multipliers being $u \in \Re_+^m$. We assume that the resulting Lagrangian subproblem

$$\theta(u) := b^{\mathrm{T}} u + \underset{\substack{\sum_{j=1}^n d_j x_j \geq e \\ x_j \in \{0,1\}}}{\text{minimum}} \sum_{j=1}^{n} (c_j - u^{\mathrm{T}} a_j) x_j, \qquad u \geq 0^m, \tag{25}$$

has the integrality property, so that each constraint $x_j \in \{0,1\}$ can be replaced by $0 \leq x_j \leq 1$, for all $j = 1, \ldots, n$. We denote the corresponding linear programming dual multipliers for the constraints (24c) by $v \in \Re_+^r$.

At a near-optimal $\bar{u} \in \Re^m_+$ in the Lagrangian dual problem to maximize the value of $\theta(u)$ over $u \in \Re^m_+$, let $(x(\bar{u}), \bar{v})$ be the optimal primal–dual solution to the Lagrangian subproblem (25). The value of $x(\bar{u})$ is used to predict the optimal values of the variables x_j in the problem (24). We denote by $\mathcal{J}_0$ ($\mathcal{J}_1$) those indices $j \in \mathcal{J} := \{1, \ldots, n\}$ for which the prediction is that $x_j^* = 0$ (respectively, $x_j^* = 1$). A core problem is a restriction of the original problem (24) wherein the variables in $\mathcal{J}_0 \cup \mathcal{J}_1$ are fixed to their predicted values, and the remaining variables, $\mathcal{J}_f := \mathcal{J} \setminus (\mathcal{J}_0 \cup \mathcal{J}_1)$, are free. The integrality property is imposed upon the subproblem (25) in order to be able to utilize reduced costs in the ranking of the variables when deciding on these predictions.

Let $\Delta_1 \in \Re^m_+$, $\Delta_2 \in \Re^r_+$. The core problem is

$$f_c^* := \sum_{j \in \mathcal{J}_1} c_j + \text{minimum} \sum_{j \in \mathcal{J}_f} c_j x_j, \tag{26a}$$

$$\text{subject to } b - \sum_{j \in \mathcal{J}_1} a_j + \Delta_1 \geq \sum_{j \in \mathcal{J}_f} a_j x_j \geq b - \sum_{j \in \mathcal{J}_1} a_j, \tag{26b}$$

$$e - \sum_{j \in \mathcal{J}_1} d_j + \Delta_2 \geq \sum_{j \in \mathcal{J}_f} d_j x_j \geq e - \sum_{j \in \mathcal{J}_1} d_j, \tag{26c}$$

$$x_f \in \{0, 1\}, \qquad j \in \mathcal{J}_f. \tag{26d}$$

As was the case with the restricted master problem (22), the purpose with this construction of a core problem is that feasible solutions are near-complementary. It is to be noted that the means by which we enforce near-complementarity is slightly different from the one in (22); a constraint like (22c) is here avoided because it would destroy any favourable structure inherent in the constraints (26b)–(26c).

Although in principle the subsets $\mathcal{J}_0$ and $\mathcal{J}_1$ can be defined quite arbitrarily, it is natural to choose them such that

$$\mathcal{J}_0 \subseteq \{\, j \in \mathcal{J} \mid \bar{c}_j > 0 \,\}, \tag{27a}$$

$$\mathcal{J}_1 \subseteq \{\, j \in \mathcal{J} \mid \bar{c}_j < 0 \,\}, \tag{27b}$$

where

$$\bar{c}_j := c_j - \bar{u}^{\mathrm{T}} a_j - \bar{v}^{\mathrm{T}} d_j, \qquad j \in \mathcal{J} \tag{28}$$

define the (linear programming-)reduced cost vector in the Lagrangian subproblem. These relations will be used below.

THEOREM 17 (quality of core problem). *Let $\bar{u} \in \Re^m_+$, and let $(x(\bar{u}), \bar{v})$ solve the subproblem (25). Suppose that the prediction satisfies (27) and that the core problem (26) has a feasible solution. Then,*

$$f_c^* \leq \theta(\bar{u}) + \varepsilon + \delta,$$

where

$$\varepsilon := \sum_{j \in \mathcal{J}_f} |\bar{c}_j| + \bar{v}^{\mathrm{T}} \Delta_2,$$

$$\delta := \bar{u}^{\mathrm{T}} \Delta_1.$$

PROOF. The proof utilizes Proposition 5(a), as follows.

Let $\bar{x} := \left(\bar{x}_{\mathcal{J}_0}, \bar{x}_{\mathcal{J}_1}, \bar{x}_{\mathcal{J}_f}\right)$ be the primal vector corresponding to the predictions and the optimal solution to the core problem (26).

To establish that (10a) holds, we use below that

$$\bar{v}^{\mathrm{T}}\left(\sum_{j\in\mathcal{J}} d_j x_j(\bar{u}) - e\right) = 0, \qquad \bar{v}^{\mathrm{T}}\left(e - \sum_{j\in\mathcal{J}} d_j \bar{x}_j + \Delta_2\right) \geq 0$$

holds, the equality by complementarity in the linear programming subproblem equivalent to (25), and the second by the feasibility of $\bar{x}$ in (26); in particular, then,

$$\bar{v}^{\mathrm{T}} \sum_{j\in\mathcal{J}} d_j[x_j(\bar{u}) - \bar{x}_j] + \bar{v}^{\mathrm{T}}\Delta_2 \geq 0. \tag{29}$$

We therefore have that

$$\begin{aligned} L(\bar{x},\bar{u}) &:= b^{\mathrm{T}}\bar{u} + \sum_{j\in\mathcal{J}}(c_j - \bar{u}^{\mathrm{T}}a_j)\bar{x}_j = \theta(\bar{u}) + \sum_{j\in\mathcal{J}}(c_j - \bar{u}^{\mathrm{T}}a_j)[\bar{x}_j - x_j(\bar{u})] \\ &\leq \theta(\bar{u}) + \sum_{j\in\mathcal{J}} \bar{c}_j[\bar{x}_j - x_j(\bar{u})] + \bar{v}^{\mathrm{T}}\Delta_2 \quad [(28) \text{ and } (29)] \\ &\leq \theta(\bar{u}) + \varepsilon, \end{aligned}$$

where the last inequality follows from the assumption that (27) holds.

That (10b) holds follows by assumption.

Finally, from (26b), we obtain that, with $g(x) := b - \sum_{j\in\mathcal{J}} a_j x_j$, the relation $g(\bar{x}) \geq -\Delta_1$ holds, and so we obtain, because $\bar{u} \in \Re^m_+$, that $\bar{u}^{\mathrm{T}}g(\bar{x}) \geq -\bar{u}^{\mathrm{T}}\Delta_1 = -\delta$, whence (10c) follows. Hence, the pair $(\bar{x},\bar{u})$ satisfies (10).

Proposition 5(a) then yields the desired result. □

The theorem immediately implies that

$$f_c^* - f^* \leq \sum_{j\in\mathcal{J}_f} |\bar{c}_j| + \bar{u}^{\mathrm{T}}\Delta_1 + \bar{v}^{\mathrm{T}}\Delta_2$$

holds.

Note that if the core problem is constructed from an optimal dual solution (which can be found by solving the continuous relaxation of the original problem, due to the integrality property) and the values of Δ_1 and Δ_2 are taken to be at least as large as the LP optimal slacks, then the continuous relaxation of the core problem is feasible.

The following result corresponds to Proposition 14.

PROPOSITION 18 (variable fixing and optimality). *Let $\bar{u} \in \Re^m_+$. Suppose that the prediction satisfies (27), that the core problem (26) has a feasible solution, and that $\bar{x}$ is an optimal solution to it. Suppose further that we know an upper bound $\bar{f} \geq f^*$ (for example, $\bar{f} = f_c^*$).*

(a) *If a $j \in \mathcal{J}_0 \cup \mathcal{J}_1$ is such that $|\bar{c}_j| > \bar{f} - \theta(\bar{u})$, then $x_j^* = \bar{x}_j$ in every optimal solution x^* to (24).*

(b) *If*

$$(\Delta_1)_i \begin{cases} \geq (\bar{f} - \theta(\bar{u}))/\bar{u}_i, & \text{if } \bar{u}_i > 0, \\ = \infty, & \text{otherwise,} \end{cases} \qquad i = 1,\ldots,m,$$

$$(\Delta_2)_i \begin{cases} \geq (\bar{f} - \theta(\bar{u}))/\bar{v}_i, & \text{if } \bar{v}_i > 0, \\ = \infty, & \text{otherwise,} \end{cases} \qquad i = 1,\ldots,r,$$

and if $|\bar{c}_j| \geq \bar{f} - \theta(\bar{u})$ holds for every $j \in \mathcal{J}_0 \cup \mathcal{J}_1$, then $x^* = \bar{x}$ is an optimal solution to (24). □

An implication of this result is the well-known fact that core problems should be built up with variables having small reduced costs.

As an ending note, we provide an important comment on the construction of the master and core problems in these last two subsections. [The comments are made for the case of core problems, but the same arguments apply for the side constrained master problem (22).] The value of f_c^* is reduced by an increase in the value of each of the elements of Δ_1 and Δ_2, as the feasible set of the problem (26) would then increase, and therefore one might ask what the purpose of the additional restrictions in (26b) and (26c) is? The answer is that there is a trade-off between obtaining good objective values with a core problem, and the complexity of solving it; the additional restriction makes the problem easier to solve by restricting the feasible set. Further, the additional restriction introduced in the problem (26) which serves to control the violation of complementarity, incorporates explicitly a measure that we have shown previously to be of utmost importance in forcing Lagrangian heuristics to strive for an optimal solution, and a term which hitherto has not been present in core problems, as only the Lagrangian optimality-based term defined by the vector $\bar{c}$ of reduced costs is normally used.

Acknowledgements

Dr. Laura Wynter provided valuable comments on an early draft. Lic. Phil. Tuomo Takkula pointed out the relations between our Proposition 5(b) and Corollary 7 to [NeW88, Corollary II.3.6.9 and Theorem II.3.6.7]. Discussions with Lic. Phil. Myrna Palmgren inspired the study of relations to the work in [HaZ80] which lead to Remark 9. Lic. Phil. Andreas Westerlund drew our attention to the paper [SwM79].

The first author was partially financed by grant # 621-2001-2716 from the Swedish Science Foundation.

References

[AEOP02] R. K. Ahuja, O. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large scale neighborhood search techniques. *Discrete Applied Mathematics*, 123 (2002), pp. 75–102.

[AMRT01] E. D. Andersen, J. E. Mitchell, C. Roos, and T. Terlaky. A homogenized cutting plane method to solve the convex feasibility problem. In: *Optimization Methods and Applications*, X.-Q. Yang, K.L. Teo, and L. Caccetta (eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, pp. 167–190.

[BaC96] E. Balas and M. C. Carrera. A dynamic subgradient-based branch-and-bound procedure for set covering. *Operations Research*, 44 (1996), pp. 875–890.

[BaH80] E. Balas and A. Ho. An algorithm for set covering problems. *Mathematical Programming Study*, 12 (1980), pp. 37–60.

[BaZ80] E. Balas and E. Zemel. An algorithm for large zero-one knapsack problems. *Operations Research*, 28 (1980), pp. 1130–1154.

[BSS93] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms.* John Wiley & Sons, New York, NY, second edition, 1993.

[Bea87] J. E. Beasley. An algorithm for set covering problems. *European Journal of Operational Research*, 31 (1987), pp. 85–93.

[Bea93] J. E. Beasley. Lagrangian relaxation. Chapter 6 in *Modern Heuristic Techniques for Combinatorial Problems*, C.R. Reeves (ed.), Blackwell Scientific Publications, Oxford, U.K., 1993.

[BeT89] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Prentice-Hall, London, 1989.

[BGB81] G. G. Brown, A. Geoffrion, and G. Bradley. Seasonal production and sales planning with limited shared tooling at the key operation. *Management Science*, 27 (1981), pp. 247–259.

[CaM91] G. M. Campbell and V. A. Mabert. Cyclical schedules for capacitated lot sizing with dynamic demands. *Management Science*, 17 (1991), pp. 409–427.

[CFT99] A. Caprara, M. Fischetti, and P. Toth. A heuristic method for the set covering problem. *Operations Research*, 47 (1999), pp. 730–743.

[CFT02] A. Caprara, M. Fischetti, and P. Toth. Modeling and solving the train timetabling problem. *Operations Research*, 50 (2002), pp. 851–861.

[CNS98] S. Ceria, P. Nobili, and A. Sassano. A Lagrangian-based heuristic for large-scale set covering problems. *Mathematical Programming*, 81 (1998), pp. 215–228.

[Chv79] V. Chvátal. A greedy heuristic for the set-covering problem, *Mathematics of Operations Research*, 4 (1979), pp. 233–235.

[Erl78] D. Erlenkotter. A dual-based procedure for uncapacitated facility location. *Operations Research*, 26 (1978), pp. 992–1009.

[Eve63] H. Everett, III. Generalized Lagrange multiplier method for solving problem of optimal allocation of resources. *Operations Research*, 11 (1963), pp. 399–518.

[FeK00] S. Feltenmark and K. C. Kiwiel. Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization*, 14 (2000), pp. 697–721.

[Fis81] M. L. Fisher. The Lagrangian relaxation method for solving integer programming problems. *Management Science*, 27 (1981), pp. 1–18.

[Fis85] M. L. Fisher. An applications oriented guide to Lagrangian relaxation. *Interfaces*, 15 (1985), pp. 10–21.

[Fis94] M. L. Fisher. Optimal solution of vehicle-routing problems using minimum K-trees. *Operations Research*, 42 (1994), pp. 626–642.

[Geo74] A. M. Geoffrion. Lagrangian relaxation for integer programming. *Mathematical Programming Study*, 2 (1974), pp. 82–114.

[HaZ80] G. Y. Handler and I. Zang. A dual algorithm for the constrained shortest path problem. *Networks*, 10 (1980), pp. 293–310.

[HeK70] M. Held and R. M. Karp. The traveling salesman problem and minimum spanning trees: Part I. *Operations Research*, 18 (1970), pp. 1138–62.

[HeK71] M. Held and R. M. Karp. The traveling-salesman problem and minimum spanning trees: Part II. *Mathematical Programming*, 1 (1971), pp. 6–25.

[HWC74] M. Held, P. Wolfe and H. P. Crowder. Validation of subgradient optimization. *Mathematical Programming*, 16 (1974), pp. 62–88.

[JLV90] Ö. Jonsson, T. Larsson, and P. Värbrand. A Lagrangean relaxation scheme for a scheduling problem. *Asia-Pacific Journal of Operations Research*, 7 (1990), pp. 155–162.

[Kiw95] K. C. Kiwiel. Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities. *Mathematical Programming*, 69 (1995), pp. 89–109.

[LaL97] T. Larsson and Z.-W. Liu. A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows. *Optimization*, 40 (1997), pp. 247–284.

[LPS96] T. Larsson, M. Patriksson, and A.-B. Strömberg. Conditional subgradient optimization—theory and applications. *European Journal of Operational Research*, 88 (1996), pp. 382–403.

[LPS98] T. Larsson, M. Patriksson, and A.-B. Strömberg. Ergodic convergence in subgradient optimization. *Optimization Methods & Software*, 9 (1998), pp. 93–120.

[LPS99] T. Larsson, M. Patriksson, and A.-B. Strömberg. Ergodic, primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86 (1999), pp. 283–312.

[LPS03] T. Larsson, M. Patriksson, and A.-B. Strömberg. On the convergence of conditional ε-subgradient methods for convex programs and convex–concave saddle-point problems. *European Journal of Operational Research* (to appear).

[Las70] L. S. Lasdon. *Optimization Theory for Large Systems*. Macmillan, New York, NY, 1970.

[MSW76] T. L. Magnanti, J. F. Shapiro, and M. H. Wagner. Generalized linear programming solves the dual. *Management Science*, 21 (1976), pp. 1195–1203.

[Mar93] J. Marklund. A study of Lagrangian heuristics for convex network flow problems. Master's thesis, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden, Linköping, Sweden, 1993.

[MaT90] S. Martello and P. Toth. *Knapsack Problems: Algorithms and Computer Implementations.* Wiley, Chichester, UK, 1990.

[MuC79] J. Mulvey and H. Crowder. Cluster analysis: An application of Lagrangian relaxation. *Management Science*, 25 (1979), pp. 329–340.

[NeW88] G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization.* John Wiley & Sons, New York, NY, 1988.

[Pat94] M. Patriksson. *The Traffic Assignment Problem—Models and Methods.* Topics in Transportation. VSP BV, Utrecht, The Netherlands, 1994.

[PeP97] J. Petersson and M. Patriksson. Topology optimization of sheets in contact by a subgradient method. *International Journal of Numerical Methods in Engineering*, 40 (1997), pp. 1295–1321.

[Pis95] D. Pisinger. An expanding-core algorithm for the exact 0-1 knapsack problem. *European Journal of Operational Research*, 87 (1995), pp. 175–187.

[Pol69] B. T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9 (1969), pp. 14–29.

[Sha79a] J. F. Shapiro. *Mathematical Programming: Structures and Algorithms.* John Wiley & Sons, New York, NY, 1979.

[Sha79b] J. F. Shapiro. A survey of Lagrangian techniques for discrete optimization. *Annals of Discrete Mathematics*, 5 (1979), pp. 113–138.

[SNH83] J.-J. Strodiot, V. H. Nguyen, and N. Heukemes. ε-optimal solutions in nondifferentiable convex programming and some related questions. *Mathematical Programming*, 25 (1983), pp. 307–328.

[SwM79] D. J. Sweeney and R. A. Murphy. A method of decomposition for integer programs. *Operations Research*, 27 (1979). pp. 1128–1141.

[Tak01] T. Takkula. Private communication, 2001.

[VaW84] F. J. Vasko and G. R. Wilson. An efficient heuristic for large set covering problems. *Naval Research Logistics*, 31 (1984), pp 163–171.

[Ven91] J. A. Ventura. Computational development of a Lagrangian dual approach for quadratic networks. *Networks*, 21 (1991), pp. 469–485.

[Wil01] W. E. Wilhelm. A technical review of column generation in integer programming. *Optimization and Engineering*, 2 (2001), pp. 159–200.

[Wol98] L. A. Wolsey. *Integer Programming.* John Wiley & Sons, New York, 1998.