

Chapter 15

Mixed-Integer Linear Optimization: Primal–Dual Relations and Dual Subgradient and Cutting-Plane Methods



Ann-Brith Strömberg, Torbjörn Larsson, and Michael Patriksson

Abstract This chapter presents several solution methodologies for mixed-integer linear optimization, stated as mixed-binary optimization problems, by means of Lagrangian duals, subgradient optimization, cutting-planes, and recovery of primal solutions. It covers Lagrangian duality theory for mixed-binary linear optimization, a problem framework for which ultimate success—in most cases—is hard to accomplish, since strong duality cannot be inferred. First, a simple conditional subgradient optimization method for solving the dual problem is presented. Then, we show how ergodic sequences of Lagrangian subproblem solutions can be computed and used to recover mixed-binary primal solutions. We establish that the ergodic sequences accumulate at solutions to a convexified version of the original mixed-binary optimization problem. We also present a cutting-plane approach to the Lagrangian dual, which amounts to solving the convexified problem by Dantzig–Wolfe decomposition, as well as a two-phase method that benefits from the advantages of both subgradient optimization and Dantzig–Wolfe decomposition. Finally, we describe how the Lagrangian dual approach can be used to find near optimal solutions to mixed-binary optimization problems by utilizing the ergodic sequences in a Lagrangian heuristic, to construct a core problem, as well as to guide the branching in a branch-and-bound method. The chapter is concluded with a section comprising notes, references, historical downturns, and reading tips.

Keywords Mixed-binary linear optimization · Convexified problem · Lagrange dual · Non-smooth convex function · Subgradient method · Ergodic sequences · Cutting planes · Column generation · Dantzig–Wolfe decomposition · Core problem

A.-B. Strömberg (✉) · M. Patriksson
Chalmers University of Technology and the University of Gothenburg, Göteborg, Sweden
e-mail: anstr@chalmers.se; mipat@chalmers.se

T. Larsson
Linköping University, Linköping, Sweden
e-mail: torbjorn.larsson@liu.se

15.1 Introduction and Motivation

The aim of this chapter is to provide theory and methodology for Lagrangian dual approaches for solving mixed-integer linear optimization problems, when stated as mixed-binary problems.¹ It covers Lagrangian duality theory for mixed-binary linear optimization, generalizations of classical dual subgradient algorithms, cutting-plane methods, and (fractional) approximations of primal solutions from ergodic sequences, as well as the recovery of primal integer solutions. The chapter summarizes a research stream on subgradient methods for nondifferentiable optimization applications accomplished by the authors over two decades. While being based on a series of articles by the authors and co authors, some of the results presented here are, however, hitherto unpublished.

A strong motive for using Lagrangian dual methods in many applications of discrete and combinatorial optimization is that such problems may often be viewed as relatively easily solvable problems (being commonly highly structured) to which a number of complicating side constraints are added. An example is the (asymmetric) traveling salesperson problem, which can be viewed as that of finding a least cost trip assignment subject to subtour eliminating side constraints. In a Lagrangian dual method appropriate prices are assigned to the side constraints which then are included in the objective function. A solution to the resulting simpler problem yields a lower bound on the optimal value of the original problem, but does usually not fulfill the relaxed constraints. The prices are iteratively improved by means of some updating rule in an attempt to find prices such that the relaxed constraints become 'optimally fulfilled', that is, such that an optimal solution to the original problem is obtained. In discrete optimization, however, such prices usually do not exist.

Lagrangian dual methods are nevertheless increasingly popular tools in discrete optimization. Among their merits are their flexibility and applicability to many different problem structures and their ease of implementation. They also often produce higher lower bounds on the optimal value than does a continuous (linear programming) relaxation.² Lagrangian dual methods are most often used successfully in combination with other discrete optimization methodologies, such as branch-and-bound algorithms—within which they provide lower bounds—, local search methods, and primal heuristics. Among the latter, a popular combination is Lagrangian heuristics, which combine Lagrangian dual schemes with manipulations of primal infeasible solutions, aiming at producing near optimal and primal feasible solutions.

¹Note that any mixed-integer optimization problem can be transformed into a mixed-binary optimization problem with a finite number of binary variables, provided that the feasible region for the original integer variables is bounded.

²One must add, however, that it requires solving a convex and nondifferentiable optimization problem, which may be quite nontrivial.

We consider a Lagrangian dual approach for solving mixed-binary linear optimization problems. We outline the properties of the Lagrangian dual problem, its relations to the primal mixed-binary problem, and its optimality conditions. Further, primal–dual optimality conditions are presented, both for a convexified primal problem and for the mixed-binary problem.

It is described how the dual problem can be solved by means of a simple conditional subgradient optimization method. For this general method, we provide a convergence result for an ergodic (averaged) sequence of solutions to the Lagrangian subproblems. The ergodic sequence is constructed in the same manner as done in [44, 60] for the case of convex optimization, and which establishes that the ergodic sequences in the limit produce optimal solutions to the original problem. Here, however, we establish that the sequences accumulate at solutions to a convexified version of the original mixed-binary optimization problem.

We further present a cutting-plane approach to the Lagrangian dual problem; it amounts to solving the convexified problem utilizing Dantzig–Wolfe decomposition, that is, column generation with the columns being associated with the solutions to the Lagrangian subproblems. Then we describe an approach to generate high quality initial columns, obtained from subproblem solutions in a subgradient optimization method applied—in a prediction phase—to the Lagrangian dual.

In the following section, we present a Lagrangian dual approach to find feasible and near optimal solutions to mixed-binary optimization problems utilizing

- (1) a Lagrangian heuristic based on the ergodic sequences,
- (2) a core problem, which is constructed based on information from the ergodic sequences, and
- (3) the ergodic sequences to guide the branching in a branch-and-bound method.

The chapter is then concluded with an extensive section with notes, references, historical downturns and further reading tips.

15.2 Mixed-Binary Linear Optimization and Its Convexified Counterpart

We consider a general mixed-binary linear optimization problem. In our presentation and the derivation of methods to follow, the feasible set is described as the intersection of two sets. One set is characterized by general, explicit linear inequality constraints, which are to be *Lagrangian relaxed*. The other set is implicit and may be a Cartesian product set, resulting in one or several separable subproblems in the solution procedure(s); our description is general in that each subproblem may contain solely continuous, solely binary, or mixed variables.

Our mixed-binary linear optimization problem is defined as

$$z^* := \min_{\mathbf{x}_b, \mathbf{x}_c} \mathbf{c}_b^\top \mathbf{x}_b + \mathbf{c}_c^\top \mathbf{x}_c \quad (15.1a)$$

$$\text{subject to } A_b \mathbf{x}_b + A_c \mathbf{x}_c \geq \mathbf{b}, \quad (15.1b)$$

$$(\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top \in X, \quad (15.1c)$$

where z^* denotes the optimal value, $\mathbf{c}_b \in \mathbb{R}^{n_b}$, $\mathbf{c}_c \in \mathbb{R}^{n_c}$, $A_b \in \mathbb{R}^{m \times n_b}$, $A_c \in \mathbb{R}^{m \times n_c}$, $\mathbf{b} \in \mathbb{R}^m$, and $n_b, n_c, m \in \mathbb{Z}_+$. The set $X := \{(\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top \in \{0, 1\}^{n_b} \times \mathbb{R}_+^{n_c} \mid D_b \mathbf{x}_b + D_c \mathbf{x}_c \geq \mathbf{d}\}$, where $D_b \in \mathbb{R}^{k \times n_b}$, $D_c \in \mathbb{R}^{k \times n_c}$, $\mathbf{d} \in \mathbb{R}^k$, and $k \in \mathbb{Z}_+$, is assumed to be nonempty and bounded. By defining $\mathbf{x} := (\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top$, $\mathbf{c} := (\mathbf{c}_b^\top, \mathbf{c}_c^\top)^\top$, $A := (A_b, A_c)$, $D := (D_b, D_c)$, and $n := n_b + n_c$, the optimization problem (15.1) can be equivalently expressed as³

$$z^* := \min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x}, \quad (15.2a)$$

$$\text{subject to } A\mathbf{x} \geq \mathbf{b}, \quad (15.2b)$$

$$\mathbf{x} \in X, \quad (15.2c)$$

where $X = \{\mathbf{x} \in \{0, 1\}^{n_b} \times \mathbb{R}_+^{n_c} \mid D\mathbf{x} \geq \mathbf{d}\}$. We generally assume that the mixed-binary linear optimization problem is feasible, that is, that $\{\mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}\} \neq \emptyset$ holds, and denote by $X^* := \operatorname{argmin}_{\mathbf{x} \in X} \{\mathbf{c}^\top \mathbf{x} \mid A\mathbf{x} \geq \mathbf{b}\}$ its nonempty solution set.

By denoting an *extreme point* of the *convex hull* of the set X with \mathbf{x}^q and letting \mathcal{Q} be an index set for all such points, the convex hull can be expressed as

$$X_{\text{conv}} := \operatorname{conv} X = \operatorname{conv}_{q \in \mathcal{Q}} \{\mathbf{x}^q\}. \quad (15.3)$$

Any extreme point to X_{conv} can be expressed as $\mathbf{x}^q = ((\mathbf{x}_b^q)^\top, (\mathbf{x}_c^q)^\top)^\top$, where $\mathbf{x}_b^q \in \{0, 1\}^{n_b}$ and \mathbf{x}_c^q is an extreme point to the nonempty polyhedral set

$$X_c(\mathbf{x}_b^q) := \{\mathbf{x}_c \in \mathbb{R}_+^{n_c} \mid D_c \mathbf{x}_c \geq \mathbf{d} - D_b \mathbf{x}_b^q\}, \quad q \in \mathcal{Q}.$$

The linear programming (LP) relaxation of the set X is expressed as

$$X_{\text{LP}} := \left\{ (\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top \in [0, 1]^{n_b} \times \mathbb{R}_+^{n_c} \mid D_b \mathbf{x}_b + D_c \mathbf{x}_c \geq \mathbf{d} \right\} \quad (15.4a)$$

$$= \left\{ \mathbf{x} \in [0, 1]^{n_b} \times \mathbb{R}_+^{n_c} \mid D\mathbf{x} \geq \mathbf{d} \right\}. \quad (15.4b)$$

³The notation in (15.1) and (15.2) will be used interchangeably throughout this chapter.

It holds that $X \subseteq X_{\text{conv}} \subseteq X_{\text{LP}}$. Replacing the set X in (15.2c) with its convex hull X_{conv} results in the linear optimization problem defined as

$$z_{\text{conv}}^* := \min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}; \mathbf{x} \in X_{\text{conv}} \right\} \leq z^*. \quad (15.5)$$

By assumption, the set $\{\mathbf{x} \in X_{\text{conv}} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$ is nonempty and bounded. We let X_{conv}^* denote the solution set to the optimization problem (15.5). Replacing the set X in (15.2c) by X_{LP} yields

$$z_{\text{LP}}^* := \min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}; \mathbf{x} \in X_{\text{LP}} \right\} \leq z_{\text{conv}}^*, \quad (15.6)$$

and we let X_{LP}^* denote the solution set to the optimization problem (15.6).

For the case when the set X_{LP} possesses the *integrality property* with respect to the binary variables, that is, when all its extreme points have only integer valued variables \mathbf{x}_b , the equality $X_{\text{conv}} = X_{\text{LP}}$ holds, implying the equality $z_{\text{conv}}^* = z_{\text{LP}}^*$.

Remark 15.1 In many applications, the set X is a Cartesian product set, here denoted by $X := Y_1 \times Y_2 \times \dots \times Y_S = \prod_{s \in \mathcal{S}} Y_s$, where $\mathcal{S} = \{1, \dots, S\}$. It is then assumed that each set $Y_s \subset \{0, 1\}^{n_{b,s}} \times \mathbb{R}^{n_{c,s}}$ is defined over binary and/or continuous variables \mathbf{y}_s , such that $(\mathbf{y}_1^\top, \dots, \mathbf{y}_S^\top)^\top \equiv \mathbf{x}$, and such that the relations $n_{b,s}, n_{c,s} \in \mathbb{Z}_+$, $s \in \mathcal{S}$, $\sum_{s \in \mathcal{S}} n_{b,s} = n_b$, and $\sum_{s \in \mathcal{S}} n_{c,s} = n_c$ hold. The optimization problem (15.2) is then expressed as

$$z^* := \min_{\mathbf{y}_s, s \in \mathcal{S}} \sum_{s \in \mathcal{S}} \mathbf{c}_s^\top \mathbf{y}_s \quad (15.7a)$$

$$\text{subject to } \sum_{s \in \mathcal{S}} A_s \mathbf{y}_s \geq \mathbf{b}, \quad (15.7b)$$

$$\mathbf{y}_s \in Y_s, \quad s \in \mathcal{S}, \quad (15.7c)$$

where $\mathbf{c}_s \in \mathbb{R}^{n_s}$, $A_s \in \mathbb{R}^{m \times n_s}$, and $n_s = n_{b,s} + n_{c,s}$, $s \in \mathcal{S}$. The constraints (15.7b) are said to be *coupling*, since relaxing them will result in a Lagrangian subproblem that separates into one minimization problem for each $s \in \mathcal{S}$. The integrality property may be considered for each of the sets Y_s , as needed/being relevant.

15.2.1 The Lagrangian Dual

The *Lagrange function* $L : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ with respect to the relaxation of the constraints (15.1b) by means of the *price vector* $\mathbf{u} \in \mathbb{R}_+^m$, also called *dual variables*

or (*Lagrangian* or *dual*) *multipliers*, is defined as

$$L(\mathbf{x}, \mathbf{u}) := \mathbf{c}^\top \mathbf{x} + \mathbf{u}^\top (\mathbf{b} - A\mathbf{x}).$$

Its minimization over $\mathbf{x} \in X$ determines the *Lagrangian dual function* $h : \mathbb{R}^m \mapsto \mathbb{R}$, as defined by

$$\begin{aligned} h(\mathbf{u}) &:= \min_{\mathbf{x} \in X} L(\mathbf{x}, \mathbf{u}) = \mathbf{b}^\top \mathbf{u} + \min_{\mathbf{x} \in X} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x} \right\} \\ &= \mathbf{b}^\top \mathbf{u} + \min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x}^q \right\}. \end{aligned} \quad (15.8)$$

The function h is formed by the point wise minimum over $|\mathcal{Q}|$ affine functions, and it is therefore piecewise affine and concave, hence continuous but generally not everywhere differentiable. Letting $X(\mathbf{u})$ and $\mathcal{Q}(\mathbf{u})$ denote the optimal set to the respective inner minimization problem—the so-called *Lagrangian subproblem* or *Lagrangian relaxed problem*—in (15.8), the following relations hold:

$$X(\mathbf{u}) = \operatorname{argmin}_{\mathbf{x} \in X} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x} \right\} = \left\{ \mathbf{x}^q \right\}_{q \in \mathcal{Q}(\mathbf{u})}; \quad (15.9a)$$

$$X_{\text{conv}}(\mathbf{u}) := \operatorname{conv} X(\mathbf{u}) = \operatorname{argmin}_{\mathbf{x} \in X_{\text{conv}}} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x} \right\} = \operatorname{conv}_{q \in \mathcal{Q}(\mathbf{u})} \left\{ \mathbf{x}^q \right\}. \quad (15.9b)$$

The expression for $X(\mathbf{u})$ in (15.9a) can always be replaced by its *convexified* version $X_{\text{conv}}(\mathbf{u})$ in (15.9b), since for any linear objective there is an optimal extreme point to the set X_{conv} that is also optimal with respect to the set X .

By *weak duality*, the inequality $h(\mathbf{u}) \leq \mathbf{c}^\top \mathbf{x}$ holds whenever $\mathbf{u} \in \mathbb{R}_+^m$ and $\mathbf{x} = (\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top$ is feasible in (15.2) [and, consequently, in (15.1)]. In order to find the best possible underestimate of z^* , the prices \mathbf{u} should be chosen as to maximize the Lagrangian dual function, that is, to solve the *Lagrangian dual problem* defined as

$$h^* := \max_{\mathbf{u} \in \mathbb{R}_+^m} h(\mathbf{u}). \quad (15.10)$$

The problem (15.10) is a *convex* optimization problem having a concave and generally *nondifferentiable* objective function. By the assumption that the polyhedron $\{\mathbf{x} \in X_{\text{conv}} \mid A\mathbf{x} \geq \mathbf{b}\}$ is nonempty, also the optimal set of (15.10)—denoted U^* —is nonempty and polyhedral. Thus, by weak duality, the inequality $h^* \leq z^*$ holds. For most mixed-binary linear optimization problems, however, it holds that $h^* < z^*$, that is, the *duality gap* $z^* - h^*$ is nonzero.

Using the transformations in (15.8) along with a LP dualization, the Lagrangian dual (15.10) can be reformulated according to

$$h^* = \max_{\mathbf{u}} \left\{ \mathbf{b}^\top \mathbf{u} + \min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x}^q \right\} \mid \mathbf{u} \in \mathbb{R}^m \right\} \quad (15.11a)$$

$$= \max_{\mathbf{u}, v} \left\{ \mathbf{b}^\top \mathbf{u} + v \mid (A\mathbf{x}^q)^\top \mathbf{u} + v \leq \mathbf{c}^\top \mathbf{x}^q, q \in \mathcal{Q}; \mathbf{u} \in \mathbb{R}_+^m; v \in \mathbb{R} \right\} \quad (15.11b)$$

$$= \min_{\lambda} \left\{ \sum_{q \in \mathcal{Q}} (\mathbf{c}^\top \mathbf{x}^q) \lambda_q \mid \sum_{q \in \mathcal{Q}} (A\mathbf{x}^q) \lambda_q \geq \mathbf{b}; \sum_{q \in \mathcal{Q}} \lambda_q = 1; \lambda_q \geq 0, q \in \mathcal{Q} \right\} \quad (15.11c)$$

$$= \min_{\mathbf{x}, \lambda} \left\{ \mathbf{c}^\top \mathbf{x} \mid A\mathbf{x} \geq \mathbf{b}; \mathbf{x} = \sum_{q \in \mathcal{Q}} \lambda_q \mathbf{x}^q; \sum_{q \in \mathcal{Q}} \lambda_q = 1; \lambda_q \geq 0, q \in \mathcal{Q} \right\} \quad (15.11d)$$

$$= \min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} \mid A\mathbf{x} \geq \mathbf{b}; \mathbf{x} \in X_{\text{conv}} \right\} = z_{\text{conv}}^*. \quad (15.11e)$$

In summary, the above derivations of primal–dual connections imply the following *weak* and *strong duality* relations for the problems (15.1), (15.2), (15.5), (15.6), and (15.10):

$$z^* \geq z_{\text{conv}}^* = h^* \geq z_{\text{LP}}^*.$$

For typical—as well as the most interesting—applications of Lagrangian dualization the straightforward continuous relaxation of the set X does, however, not result in a set X_{LP} [defined in (15.4)] having integer extreme points. Hence, although the equality $z_{\text{conv}}^* = z_{\text{LP}}^*$ may hold, typically $X_{\text{conv}} \subset X_{\text{LP}}$ holds [X_{conv} being defined in (15.3)], which—in practice—most often implies the strict inequality $z_{\text{conv}}^* > z_{\text{LP}}^*$.

15.2.2 Optimality Conditions for the Convexified Problem

In order to derive optimality conditions for the convexified optimization problem (15.5) and, eventually, also for the original optimization problem (15.1) [or (15.2)], we first define the *subdifferential* of the concave function h at $\mathbf{u} \in \mathbb{R}^m$ as

$$\partial h(\mathbf{u}) := \left\{ \boldsymbol{\gamma} \in \mathbb{R}^m \mid h(\mathbf{v}) \leq h(\mathbf{u}) + \boldsymbol{\gamma}^\top (\mathbf{v} - \mathbf{u}), \mathbf{v} \in \mathbb{R}^m \right\},$$

the elements of which are called *subgradients*. The following characterization of the subdifferential holds for any Lagrangian dual objective function.

Proposition 15.1 (Subdifferential of the Dual Objective Function) *For each $\mathbf{u} \in \mathbb{R}^m$, it holds that $\partial h(\mathbf{u}) = \{\mathbf{b} - A\mathbf{x} \mid \mathbf{x} \in X_{\text{conv}}(\mathbf{u})\} = \text{conv}\{\mathbf{b} - A\mathbf{x}^q \mid q \in \mathcal{Q}(\mathbf{u})\}$. The function h is differentiable at \mathbf{u} if and only if $\partial h(\mathbf{u})$ is a singleton set, that is, if and only if $\mathbf{b} - A\mathbf{x}$ is constant on $X_{\text{conv}}(\mathbf{u})$, in which case $\nabla h(\mathbf{u}) = \mathbf{b} - A\mathbf{x}$ for any $\mathbf{x} \in X_{\text{conv}}(\mathbf{u})$.⁴*

The normal cone to the set \mathbb{R}_+^m at $\mathbf{u} \in \mathbb{R}^m$ is defined as

$$N_{\mathbb{R}_+^m}(\mathbf{u}) := \begin{cases} \{\mathbf{v} \in \mathbb{R}^m \mid \mathbf{v}^\top(\mathbf{v} - \mathbf{u}) \leq 0, \mathbf{v} \in \mathbb{R}_+^m\} \\ \quad = \{\mathbf{v} \in \mathbb{R}_-^m \mid \mathbf{v}^\top \mathbf{u} = 0\}, & \mathbf{u} \in \mathbb{R}_+^m, \\ \emptyset, & \mathbf{u} \notin \mathbb{R}_+^m. \end{cases}$$

Letting \mathbf{e}^i denote the i -th unit column, $N_{\mathbb{R}_+^m}(\mathbf{u}) = \text{cone}\{-\mathbf{e}^i \mid u_i = 0, i \in \{1, \dots, m\}\}$ holds for $\mathbf{u} \in \mathbb{R}_+^m$. The conditional subdifferential of h at $\mathbf{u} \in \mathbb{R}_+^m$, the elements of which will be referred to as *conditional subgradients*, is in our Lagrangian dual setting then defined as

$$\begin{aligned} \partial^{\mathbb{R}_+^m} h(\mathbf{u}) &:= \left\{ \boldsymbol{\gamma} \in \mathbb{R}^m \mid h(\mathbf{v}) \leq h(\mathbf{u}) + \boldsymbol{\gamma}^\top(\mathbf{v} - \mathbf{u}), \mathbf{v} \in \mathbb{R}_+^m \right\} \\ &= \partial h(\mathbf{u}) - N_{\mathbb{R}_+^m}(\mathbf{u}) \\ &= \text{conv}\{\mathbf{b} - A\mathbf{x}^q \mid q \in \mathcal{Q}(\mathbf{u})\} + \text{cone}\{\mathbf{e}^i \mid u_i = 0, i \in \{1, \dots, m\}\}. \end{aligned} \tag{15.12}$$

Clearly, $\partial^{\mathbb{R}_+^m} h(\mathbf{u}) \supseteq \partial h(\mathbf{u})$ holds for all $\mathbf{u} \in \mathbb{R}_+^m$. The next proposition is immediate.

Proposition 15.2 (Properties of the Conditional Subdifferential) *The conditional subdifferential $\partial^{\mathbb{R}_+^m} h(\mathbf{u})$ is nonempty, closed and convex for all $\mathbf{u} \in \mathbb{R}_+^m$. Further, $\partial^{\mathbb{R}_+^m} h(\mathbf{u})$ is unbounded whenever $\mathbf{u} \in \text{bd } \mathbb{R}_+^m$.*

Proposition 15.3 (Properties of the Lagrangian Dual) *The following statements are equivalent.*

- (i) *the Lagrangian dual problem (15.10) has a bounded optimal solution;*
- (ii) $\mathbf{0} \in \bigcup_{\mathbf{u} \in \mathbb{R}_+^m} \partial^{\mathbb{R}_+^m} h(\mathbf{u})$;
- (iii) $\{\mathbf{x} \in X_{\text{conv}} \mid A\mathbf{x} \geq \mathbf{b}\} \neq \emptyset$.

The optimality conditions for the Lagrangian dual (15.10) are expressed as follows.

Proposition 15.4 (Optimality Conditions for the Lagrangian Dual Problem) *A dual vector $\mathbf{u} \in \mathbb{R}_+^m$ is optimal, that is, $\mathbf{u} \in U^*$, if and only if $\partial h(\mathbf{u}) - N_{\mathbb{R}_+^m}(\mathbf{u}) \ni \mathbf{0}$, or equivalently, $\partial h(\mathbf{u}) \cap N_{\mathbb{R}_+^m}(\mathbf{u}) \neq \emptyset$, that is, if and only if there exists a $\boldsymbol{\gamma} \in \partial h(\mathbf{u})$ such that $\boldsymbol{\gamma} \leq \mathbf{0}$ and $\mathbf{u}^\top \boldsymbol{\gamma} = 0$ hold.*

⁴While Proposition 15.1 implies that the function h is differentiable at $\mathbf{u} \in \mathbb{R}^m$ if the set $X_{\text{conv}}(\mathbf{u})$ is a singleton, the opposite, however, does not hold in general.

Further, subgradients that verify the optimality of dual solutions correspond to optimal solutions to the convexified primal problem (15.11e), as expressed below.

Proposition 15.5 (Primal–Dual Optimality Conditions for the Convexified Problem) *Let $(\mathbf{x}, \mathbf{u}) \in X_{\text{conv}} \times \mathbb{R}_+^m$. Then $(\mathbf{x}, \mathbf{u}) \in X_{\text{conv}}^* \times U^*$ if and only if $\mathbf{x} \in X_{\text{conv}}(\mathbf{u})$ and $\mathbf{b} - A\mathbf{x} \in \partial h(\mathbf{u}) \cap N_{\mathbb{R}_+^m}(\mathbf{u})$ hold, that is, if and only if $\mathbf{x} \in X_{\text{conv}}(\mathbf{u})$, $A\mathbf{x} \geq \mathbf{b}$, and $\mathbf{u}^\top (\mathbf{b} - A\mathbf{x}) = 0$ hold.*

As is well known, a primal–dual pair $(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in X_{\text{conv}} \times \mathbb{R}_+^m$ satisfies the conditions of the proposition exactly when it is a *saddle point* of the Lagrange function L with respect to $X_{\text{conv}} \times \mathbb{R}_+^m$:

$$L(\bar{\mathbf{x}}, \mathbf{u}) \leq L(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \leq L(\mathbf{x}, \bar{\mathbf{u}}), \quad \mathbf{u} \in \mathbb{R}_+^m, \quad \mathbf{x} \in X_{\text{conv}}. \quad (15.13)$$

The mapping $\partial h \cap N_{\mathbb{R}_+^m}$ is constant on the dual solution set U^* . Hence, irrespective of the choice of dual solution $\mathbf{u}^* \in U^*$ the solution set to the primal problem (15.11e) may be expressed as

$$X_{\text{conv}}^* = \left\{ \mathbf{x} \in X_{\text{conv}}(\mathbf{u}^*) \mid A\mathbf{x} \geq \mathbf{b}; (\mathbf{u}^*)^\top (A\mathbf{x} - \mathbf{b}) = 0 \right\}. \quad (15.14)$$

In the typical situation, the subproblem solution set $X_{\text{conv}}(\mathbf{u}^*)$ is not a singleton, the dual objective function is nonsmooth on U^* , and finding a subgradient that verifies dual optimality is computationally expensive. Further, since not all points in $X_{\text{conv}}(\mathbf{u}^*)$ are optimal in the primal problem,⁵ finding a point $\mathbf{x} \in X_{\text{conv}}^*$ is nontrivial; this phenomenon is referred to as *non-coordinability*, and is relevant both when the original problem is an LP or a mixed-binary linear optimization problem.

15.2.3 Conditions for Optimality and Near Optimality of Mixed-Binary Linear Optimization Problems

The optimality conditions in Proposition 15.5 and the characterization in (15.14) can be constructed because of the convexity of the problem (15.5), which yields that strong duality, that is, $h^* = z_{\text{conv}}^*$ holds. For the *nonconvex* mixed-binary linear optimization problem (15.2), for which $h^* \leq z^*$ (and typically $h^* < z^*$) holds, these conditions can be generalized through a well-defined relaxation.

⁵Usually, some points in $X_{\text{conv}}(\mathbf{u}^*)$ are infeasible in (15.2), while others are feasible but non-optimal.

Proposition 15.6 (Primal–Dual Optimality Conditions for Mixed-Binary Linear Optimization) *Let $(\mathbf{x}, \mathbf{u}) \in X \times \mathbb{R}_+^m$. Then $(\mathbf{x}, \mathbf{u}) \in X^* \times U^*$ if and only if the following system is consistent:*

$$A\mathbf{x} \geq \mathbf{b}, \quad (15.15a)$$

$$\mathbf{c}^\top \mathbf{x} + \mathbf{u}^\top (\mathbf{b} - A\mathbf{x}) \leq h(\mathbf{u}) + \varepsilon, \quad (15.15b)$$

$$-\mathbf{u}^\top (\mathbf{b} - A\mathbf{x}) \leq \delta, \quad (15.15c)$$

$$\varepsilon + \delta \leq z^* - h^*, \quad (15.15d)$$

$$\varepsilon, \delta \geq 0. \quad (15.15e)$$

The conditions stated in Proposition 15.6 characterize primal–dual optimal solutions through five properties:

- (i) primal feasibility [$\mathbf{x} \in X$ and (15.15a)];
- (ii) dual feasibility [$\mathbf{u} \in \mathbb{R}_+^m$];
- (iii) Lagrangian ε -optimality [(15.15b)];
- (iv) a relaxed (δ -)complementarity [(15.15c)];
- (v) a bounded sum of nonnegative perturbations [(15.15d)–(15.15e)].

The combination of the inequalities in (15.15) with the condition $(\mathbf{x}, \mathbf{u}) \in X \times \mathbb{R}_+^m$ leads to the equality $\varepsilon + \delta = z^* - h^*$ being fulfilled. Further, the system (15.15) is consistent for the primal–dual pair $(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \in X \times \mathbb{R}_+^m$ if and only if the inequalities

$$L(\bar{\mathbf{x}}, \mathbf{u}) - (z^* - h^*) + \varepsilon \leq L(\bar{\mathbf{x}}, \bar{\mathbf{u}}) \leq L(\mathbf{x}, \bar{\mathbf{u}}) + \varepsilon \quad (15.16)$$

hold for all $\mathbf{u} \in \mathbb{R}_+^m$ and all $\mathbf{x} \in X$, and for some $\varepsilon \in [0, z^* - h^*]$, meaning that $(\bar{\mathbf{x}}, \bar{\mathbf{u}})$ is a *near saddle point* to the Lagrange function L . For a zero duality gap, that is, when $z^* = h^*$, the conditions (15.15) reduce to those stated in Proposition 15.5 while the near saddle point property (15.16) reduces to that of an ordinary saddle point, that is, (15.13).

Using the conditions (15.15) and an optimal Lagrangian dual solution $\mathbf{u}^* \in U^*$, the optimal set of the mixed-binary linear optimization problem (15.2) is given by

$$X^* = \bigcup_{\substack{\delta = z^* - h^* - \varepsilon \\ \varepsilon \in [0, z^* - h^*]}} \left\{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}; \mathbf{c}^\top \mathbf{x} - h^* - \varepsilon \leq -(\mathbf{u}^*)^\top (\mathbf{b} - A\mathbf{x}) \leq \delta \right\}. \quad (15.17)$$

We define, for any $\mathbf{u} \in \mathbb{R}_+^m$, the sets of ε -optimal [cf. (15.9a)] and δ -complementary solutions to the Lagrangian subproblem by the following expressions:

$$X_\varepsilon^{\text{opt}}(\mathbf{u}) := \left\{ \mathbf{x} \in X \mid \mathbf{c}^\top \mathbf{x} + \mathbf{u}^\top (\mathbf{b} - A\mathbf{x}) \leq h(\mathbf{u}) + \varepsilon \right\}, \quad \varepsilon \geq 0,$$

$$X_{\delta}^{\text{comp}}(\mathbf{u}) := \left\{ \mathbf{x} \in X \mid -\mathbf{u}^{\top}(\mathbf{b} - A\mathbf{x}) \leq \delta \right\}, \quad \delta \geq 0.$$

The set X^* , as formulated in (15.17), can then for any $\mathbf{u}^* \in U^*$ be expressed as

$$X^* = \bigcup_{\varepsilon \in [0, z^* - h^*]} \left\{ \mathbf{x} \in X_{\varepsilon}^{\text{opt}}(\mathbf{u}^*) \mid A\mathbf{x} \geq \mathbf{b}; -(\mathbf{u}^*)^{\top}(\mathbf{b} - A\mathbf{x}) \leq z^* - h^* - \varepsilon \right\} \quad (15.18a)$$

$$= \bigcup_{\varepsilon \in [0, z^* - h^*]} \left\{ \mathbf{x} \in X_{\varepsilon}^{\text{opt}}(\mathbf{u}^*) \cap X_{z^* - h^* - \varepsilon}^{\text{comp}}(\mathbf{u}^*) \mid A\mathbf{x} \geq \mathbf{b} \right\}, \quad (15.18b)$$

where (15.18a) is derived as a generalization of (15.14).

Given an optimal dual solution, the conditions in the expression (15.14) can, in principle, be used to calculate an optimal solution to the convexified problem (15.5). In contrast, given an optimal dual solution, the conditions in the expressions (15.17) and (15.18) are in general not instrumental for finding optimal solutions to the problem (15.2), because the size of the duality gap $z^* - h^*$ is unknown.

The characterizations (15.17) and (15.18) can be generalized to allow for non-optimal choices of $\mathbf{u} \in \mathbb{R}_+^m$ and to describe near optimal solutions to the problem (15.2). For this purpose, we introduce the functions $\varepsilon, \delta : X \times \mathbb{R}_+^m \mapsto \mathbb{R}$, defined by

$$\varepsilon(\mathbf{x}, \mathbf{u}) := \mathbf{c}^{\top} \mathbf{x} + \mathbf{u}^{\top}(\mathbf{b} - A\mathbf{x}) - h(\mathbf{u}), \quad (15.19a)$$

and

$$\delta(\mathbf{x}, \mathbf{u}) := -\mathbf{u}^{\top}(\mathbf{b} - A\mathbf{x}). \quad (15.19b)$$

It holds that $\varepsilon(\mathbf{x}, \mathbf{u}) \geq 0$ when $\mathbf{x} \in X$, and that $\delta(\mathbf{x}, \mathbf{u}) \geq 0$ when $\mathbf{u} \in \mathbb{R}_+^m$ and $A\mathbf{x} \geq \mathbf{b}$. For any primal–dual pair $(\mathbf{x}, \mathbf{u}) \in X \times \mathbb{R}_+^m$, (15.19) can be used to characterize the ε -optimality and the δ -complementarity of a solution $\mathbf{x}^{\text{appr}}(\mathbf{u})$ that is near optimal in the Lagrangian subproblem (15.9a) and feasible in the primal problem (15.2), that is, values of $\varepsilon \geq 0$ and $\delta \geq 0$ such that $\mathbf{x}^{\text{appr}}(\mathbf{u})$ is included in the sets $X_{\varepsilon}^{\text{opt}}(\mathbf{u}) = \{ \mathbf{x} \in X \mid \varepsilon(\mathbf{x}, \mathbf{u}) \leq \varepsilon \}$ and $X_{\delta}^{\text{comp}}(\mathbf{u}) = \{ \mathbf{x} \in X \mid \delta(\mathbf{x}, \mathbf{u}) \leq \delta \}$, respectively. Then, for $\mathbf{u} \in \mathbb{R}_+^m$ and $\beta \geq 0$, the set of β -optimal solutions to the problem (15.2) can be expressed as

$$X_{\beta}^* := \left\{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}; \mathbf{c}^{\top} \mathbf{x} \leq z^* + \beta \right\} \quad (15.20a)$$

$$= \left\{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}; \varepsilon(\mathbf{x}, \mathbf{u}) + \delta(\mathbf{x}, \mathbf{u}) \leq z^* - h(\mathbf{u}) + \beta \right\} \quad (15.20b)$$

$$= \left\{ \mathbf{x} \in X_{\varepsilon}^{\text{opt}}(\mathbf{u}) \cap X_{\delta}^{\text{comp}}(\mathbf{u}) \mid A\mathbf{x} \geq \mathbf{b}; \varepsilon + \delta \leq z^* - h(\mathbf{u}) + \beta; \varepsilon, \delta \geq 0 \right\}. \quad (15.20c)$$

In particular, it holds that $X_0^* = X^*$. Further, for $\beta = 0$ and any $\mathbf{u} \in U^*$, the characterizations (15.17)–(15.18) are recovered.

Neither of the characterizations in (15.20) are practically useful for solving the problem (15.2), since the optimal value z^* is unknown, but it suggests a heuristic principle for searching for near optimal solutions. Given any value of $\mathbf{u} \in \mathbb{R}_+^m$ [preferably being (near) optimal in the Lagrangian dual (15.10)], a point $\mathbf{x} \in X$ such that $A\mathbf{x} \geq \mathbf{b}$ holds and such that the values of $\varepsilon(\mathbf{x}, \mathbf{u})$ and $\delta(\mathbf{x}, \mathbf{u})$ are both small, should be sought, that is, a primal feasible solution that is ε -optimal and δ -complementary, with small values of $\varepsilon \geq 0$ and $\delta \geq 0$. A natural starting point for such a heuristic search is a Lagrangian subproblem solution $\mathbf{x}(\mathbf{u}) \in X_0^{\text{opt}}(\mathbf{u}) = X(\mathbf{u})$ [for which $\varepsilon(\mathbf{x}(\mathbf{u}), \mathbf{u}) = 0$ holds], and the search should typically be restricted to the set X . The search should then strive for primal feasibility with respect to $A\mathbf{x} \geq \mathbf{b}$, while maintaining small values of $\varepsilon(\mathbf{x}, \mathbf{u})$ and $\delta(\mathbf{x}, \mathbf{u})$. This heuristic principle has shown to be effective for finding near optimal solutions for applications where the problem (15.2) possesses specific structures that can be exploited in a search which strive for primal feasibility.

Remark 15.2 In the case when the constraints (15.2b) are instead equalities, $\delta(\mathbf{x}, \mathbf{u}) = 0$ holds for any primal feasible solution. The characterization (15.20b) then reduces to

$$X_\beta^* = \{ \mathbf{x} \in X \mid A\mathbf{x} = \mathbf{b}; \varepsilon(\mathbf{x}, \mathbf{u}) \leq z^* - h(\mathbf{u}) + \beta \},$$

that is, primal feasibility and ε -optimality in the Lagrangian relaxed problem. In particular, then $X^* = \{ \mathbf{x} \in X \mid A\mathbf{x} = \mathbf{b}; \varepsilon(\mathbf{x}, \mathbf{u}) \leq z^* - h(\mathbf{u}) \}$. It follows that $A\mathbf{x} \neq \mathbf{b}$ must hold for any $\mathbf{x} \in X$ such that $\varepsilon(\mathbf{x}, \mathbf{u}) < z^* - h(\mathbf{u})$. Hence, a solution $\mathbf{x} \in X$ is optimal in (15.2) if and only if it is the most near optimal solution in the Lagrangian subproblem that is fulfills the relaxed constraints. This observation implies that (15.2) can be solved by enumerating elements of X with respect to increasing values of $\varepsilon(\mathbf{x}, \mathbf{u})$, whence the first one found that fulfills $A\mathbf{x} = \mathbf{b}$ is therefore also optimal.

For the case of inequality constraints in the problem (15.2), only

$$X^* \subseteq \{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}; \varepsilon(\mathbf{x}, \mathbf{u}) \leq z^* - h(\mathbf{u}) \}$$

holds, and there is no guarantee that a feasible solution with a minimal value of $\varepsilon(\mathbf{x}, \mathbf{u})$ is optimal, since the corresponding value of $\delta(\mathbf{x}, \mathbf{u})$ may be large [while an optimal solution may have a large value of $\varepsilon(\mathbf{x}, \mathbf{u})$ and a small value of $\delta(\mathbf{x}, \mathbf{u})$]. If, however, an upper bound $\bar{z} \geq z^*$ is at hand, by enumerating *all* elements in the set $\{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b}; \varepsilon(\mathbf{x}, \mathbf{u}) \leq \bar{z} - h(\mathbf{u}) \}$ an optimal solution to (15.2) will be found.

15.3 Conditional Subgradient Optimization

This section presents a generalization of subgradient optimization for solving a Lagrangian dual problem to utilize *conditional subgradients*. We discuss a particular choice of conditional subgradients, obtained through Euclidean projections, which leads to an easily implementable modification of traditional subgradient optimization schemes. Computational experiments have shown that the resulting subgradient projection method performs better than traditional subgradient optimization; in some cases the difference is considerable. This generalization of subgradient optimization is especially advantageous in the context of Lagrangian duals possessing many nonnegativity constraints, onto which Euclidean projections are simple. The section further presents a computationally cheap scheme for generating ergodic sequences of subproblem solutions and which is shown to converge to the optimal set of the convexified problem (15.5). This scheme is then enhanced, in terms of faster convergence to the optimal set. The section is then concluded by a heuristic scheme for the finite generation of feasible solutions that are ε -optimal, for an arbitrary $\varepsilon > 0$.

15.3.1 Basic Convergence in the Lagrangian Dual Problem

We consider solving the Lagrangian dual problem (15.10) by the *conditional subgradient optimization* method, which is given by the following. Choose a starting solution $\mathbf{u}^0 \in \mathbb{R}_+^m$ and compute iterates \mathbf{u}^t , $t = 0, 1, \dots$, according to

$$\mathbf{u}^{t+\frac{1}{2}} = \mathbf{u}^t + \alpha_t(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)), \quad \mathbf{u}^{t+1} = [\mathbf{u}^{t+\frac{1}{2}}]_+, \quad (15.21)$$

where $\mathbf{x}(\mathbf{u}^t) \in X(\mathbf{u}^t)$ solves the Lagrangian subproblem in (15.9) at $\mathbf{u}^t \in \mathbb{R}_+^m$, so that $\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) \in \partial h(\mathbf{u}^t)$ is a subgradient to h at \mathbf{u}^t , $\mathbf{v}(\mathbf{u}^t) \in N_{\mathbb{R}_+^m}(\mathbf{u}^t)$ is an element of the normal cone of \mathbb{R}_+^m at \mathbf{u}^t , $\alpha_t > 0$ is the step length chosen at iteration t , and $[\cdot]_+$ denotes the Euclidean projection onto the nonnegative orthant \mathbb{R}_+^m . Note that $\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t) \in \partial_{\mathbb{R}_+^m} h(\mathbf{u}^t)$, that is, the step direction belongs to the conditional subdifferential, defined in (15.12).

If $\{\mathbf{v}(\mathbf{u}^t)\} := \{\mathbf{0}\}$, then the method (15.21) reduces to the traditional subgradient optimization method.

The choices $\mathbf{v}(\mathbf{u}^t) := \operatorname{argmin}_{\mathbf{v} \in N_{\mathbb{R}_+^m}(\mathbf{u}^t)} \|\mathbf{v} - (\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t))\|$, where $\|\cdot\|$ denotes the Euclidean norm, define a special case of the method (15.21) called the *subgradient projection method*, which uses a *feasible direction* from every $\mathbf{u}^t \in \mathbb{R}_+^m$, as

$$b_i - \mathbf{A}_i \mathbf{x}(\mathbf{u}^t) - v_i(\mathbf{u}^t) = \begin{cases} [b_i - \mathbf{A}_i \mathbf{x}(\mathbf{u}^t)]_+, & \text{if } u_i^t = 0, \\ b_i - \mathbf{A}_i \mathbf{x}(\mathbf{u}^t), & \text{if } u_i^t > 0, \end{cases} \quad i = 1, \dots, m, \quad (15.22)$$

here $b_i/A_i/v_i(\cdot)$ denotes the i -th element/row of the vector/matrix $\mathbf{b}/A/v(\cdot)$.

Due to the nondifferentiability of the Lagrangian dual objective function, subgradient based optimization methods cannot rely on a one dimensional maximization for determining the step length in each iteration. Instead, to ensure convergence of the sequence of dual iterates to optimality, the step lengths must be computed according to a (typically predefined) rule. We next present convergence results for the method (15.21) under different step length rules. The first assures that the sequence of dual iterates tends to the set of solutions to the Lagrangian dual.

Theorem 15.1 (Convergence to the Solution Set by Divergent Series Step Lengths) *Let the method (15.21) be applied to the problem (15.10), with the sequence $\{\alpha_t\}$ of step lengths fulfilling the divergent series conditions*

$$\alpha_t > 0, \quad t = 0, 1, 2, \dots, \quad (15.23a)$$

$$\alpha_t \rightarrow 0, \quad \text{as } t \rightarrow \infty, \quad (15.23b)$$

and

$$\left\{ \sum_{s=0}^{t-1} \alpha_s \right\} \rightarrow \infty, \quad \text{as } t \rightarrow \infty. \quad (15.23c)$$

If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded then it holds that $\{h(\mathbf{u}^t)\} \rightarrow h^*$ and $\{\min_{\mathbf{u} \in U^*} \|\mathbf{u} - \mathbf{u}^t\|\} \rightarrow 0$ as $t \rightarrow \infty$.

Proof We show that the iterates will eventually belong to an arbitrarily small neighbourhood of the set of solutions to (15.10).

Let $\delta > 0$ be arbitrary and define $B^\delta := \{\mathbf{u} \in \mathbb{R}^m \mid \|\mathbf{u}\| \leq \delta\}$. Since the function h is piecewise affine and concave, the set \mathbb{R}_+^m is nonempty, closed and convex, and the set U^* is nonempty and polyhedral, there exists an $\varepsilon > 0$ such that the level set $U^\varepsilon := \{\mathbf{u} \in \mathbb{R}_+^m \mid h(\mathbf{u}) \geq h^* - \varepsilon\}$ fulfills $U^\varepsilon \subseteq U^* + B^{\delta/2}$. Further, the sequence $\{\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\}$ is bounded and $\alpha_t \rightarrow 0$. Hence, there exists an N_δ such that $\alpha_t \|\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\|^2 \leq \varepsilon$ and $\alpha_t \|\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\| \leq \delta/2$ for all $t \geq N_\delta$.

The sequel of the proof is based on induction. First we show that there exists a $t_\delta \geq N_\delta$ such that $\mathbf{u}^{t_\delta} \in U^* + B^\delta$. Then, we establish that if the inclusion $\mathbf{u}^t \in U^* + B^\delta$ holds for some value $t \geq N_\delta$, then also $\mathbf{u}^{t+1} \in U^* + B^\delta$.

For an arbitrary $\mathbf{u}^* \in U^*$, in each iteration t of the method (15.21) the relations

$$\begin{aligned} \|\mathbf{u}^* - \mathbf{u}^{t+1}\|^2 &= \left\| \mathbf{u}^* - [\mathbf{u}^t + \alpha_t(\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))] \right\|_+^2 \\ &\leq \left\| \mathbf{u}^* - \mathbf{u}^t - \alpha_t(\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)) \right\|^2 \\ &= \left\| \mathbf{u}^* - \mathbf{u}^t \right\|^2 - 2\alpha_t(\mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))^\top (\mathbf{u}^* - \mathbf{u}^t) \\ &\quad + \alpha_t^2 \left\| \mathbf{b} - A\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t) \right\|^2 \end{aligned} \quad (15.24)$$

hold, where the inequality follows from the projection property.⁶ Now suppose that, for all $s \geq N_\delta$, the inequality

$$2(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s))^\top (\mathbf{u}^* - \mathbf{u}^s) - \alpha_s \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s)\|^2 > \varepsilon \quad (15.25)$$

holds. Then by telescoping (15.24), we obtain that for every $t \geq N_\delta$, the inequality

$$\|\mathbf{u}^* - \mathbf{u}^{t+1}\|^2 < \|\mathbf{u}^* - \mathbf{u}^{N_\delta}\|^2 - \varepsilon \sum_{s=N_\delta}^t \alpha_s,$$

holds. Then, from (15.23c) it follows that the right-hand-side of this inequality tends to $-\infty$ as $t \rightarrow \infty$, which is clearly impossible. Therefore, the inequality

$$2(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))^\top (\mathbf{u}^* - \mathbf{u}^t) - \alpha_t \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\|^2 \leq \varepsilon \quad (15.26)$$

must hold for at least one iteration $t = t_\delta \geq N_\delta$. By definition of N_δ the inequality $\alpha_{t_\delta} \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^{t_\delta}) - \mathbf{v}(\mathbf{u}^{t_\delta})\|^2 \leq \varepsilon$ holds, which together with (15.26) implies the inequality $(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^{t_\delta}) - \mathbf{v}(\mathbf{u}^{t_\delta}))^\top (\mathbf{u}^* - \mathbf{u}^{t_\delta}) \leq \varepsilon$. Since \mathbf{u}^* , $\mathbf{u}^{t_\delta} \in \mathbb{R}_+^m$, the definition (15.12) implies the inequality $h(\mathbf{u}^*) - h(\mathbf{u}^{t_\delta}) \leq (\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^{t_\delta}) - \mathbf{v}(\mathbf{u}^{t_\delta}))^\top (\mathbf{u}^* - \mathbf{u}^{t_\delta})$. Hence, it holds that $h(\mathbf{u}^{t_\delta}) \geq h^* - \varepsilon$, that is, $\mathbf{u}^{t_\delta} \in U^\varepsilon \subseteq U^* + B^{\delta/2} \subset U^* + B^\delta$.

Now, suppose that $\mathbf{u}^t \in U^* + B^\delta$ for some $t \geq N_\delta$. If (15.25) holds, then (15.24) implies the inequality $\|\mathbf{u}^* - \mathbf{u}^{t+1}\| < \|\mathbf{u}^* - \mathbf{u}^t\|$ for any $\mathbf{u}^* \in U^*$. Defining the Euclidean projection of \mathbf{u}^t onto U^* as $\bar{\mathbf{u}}_{\text{proj}}^t := \operatorname{argmin}_{\mathbf{u} \in U^*} \|\mathbf{u} - \mathbf{u}^t\|$ then yields the inequalities

$$\|\bar{\mathbf{u}}_{\text{proj}}^{t+1} - \mathbf{u}^{t+1}\| \leq \|\bar{\mathbf{u}}_{\text{proj}}^t - \mathbf{u}^{t+1}\| < \|\bar{\mathbf{u}}_{\text{proj}}^t - \mathbf{u}^t\| \leq \delta, \quad t = 0, 1, \dots,$$

which imply the inclusion $\mathbf{u}^{t+1} \in U^* + B^\delta$. Otherwise, (15.26) must hold and, using the same arguments as above, we obtain the inequality $h(\mathbf{u}^t) \geq h^* - \varepsilon$, that is, $\mathbf{u}^t \in U^\varepsilon \subseteq U^* + B^{\delta/2}$. Since the relations

$$\begin{aligned} \|\mathbf{u}^{t+1} - \mathbf{u}^t\| &= \left\| [\mathbf{u}^t + \alpha_t(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))]_+ - \mathbf{u}^t \right\| \\ &\leq \left\| \mathbf{u}^t + \alpha_t(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)) - \mathbf{u}^t \right\| \\ &= \alpha_t \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\| \leq \frac{\delta}{2} \end{aligned}$$

hold whenever $t \geq N_\delta$, it follows that $\mathbf{u}^{t+1} \in U^* + B^{\delta/2} + B^{\delta/2} = U^* + B^\delta$. We conclude that, in either case, whenever $t \geq N_\delta$, $\mathbf{u}^t \in U^* + B^\delta$ implies that $\mathbf{u}^{t+1} \in U^* + B^\delta$.

⁶The projection property states that for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m$ and any convex set $U \subseteq \mathbb{R}^m$, the inequality $\|\operatorname{argmin}_{\mathbf{u} \in U} \|\mathbf{u} - \mathbf{v}\| - \operatorname{argmin}_{\mathbf{u} \in U} \|\mathbf{u} - \mathbf{w}\|\| \leq \|\mathbf{v} - \mathbf{w}\|$ holds.

By induction with respect to $t \geq t_\delta$, it follows that $\mathbf{u}^t \in U^* + B^\delta$ for all $t \geq t_\delta$. Since this holds for arbitrarily small values of $\delta > 0$ and since the function h is continuous, the theorem follows. \square

By requiring also that the sum of squared step lengths is convergent it is ensured that the sequence of dual iterates accumulates at a point in the dual solution set.

Theorem 15.2 (Convergence to a Solution by Quadratically Convergent Divergent Series Step Lengths) *Let the method (15.21) be applied to the problem (15.10), with the step lengths α_t fulfilling the conditions (15.23) as well as*

$$\left\{ \sum_{s=0}^{t-1} \alpha_s^2 \right\} \rightarrow p, \quad \text{as } t \rightarrow \infty, \quad (15.27)$$

where $p < \infty$. If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then $\{\mathbf{u}^t\} \rightarrow \mathbf{u}^\infty \in U^*$.

Proof Let $\mathbf{u}^* \in U^*$ be arbitrary and let $t \geq 1$. Telescoping (15.24) yields the inequality

$$\begin{aligned} \|\mathbf{u}^* - \mathbf{u}^t\|^2 &\leq \|\mathbf{u}^* - \mathbf{u}^0\|^2 - 2 \sum_{s=0}^{t-1} \alpha_s (\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s))^\top (\mathbf{u}^* - \mathbf{u}^s) \\ &\quad + \sum_{s=0}^{t-1} \alpha_s^2 \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s)\|^2. \end{aligned} \quad (15.28)$$

Since $\mathbf{u}^* \in U^*$, $\mathbf{u}^s \in \mathbb{R}_+^m$, and $\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s) \in \partial \mathbb{R}_+^m h(\mathbf{u}^s)$ for all $s \geq 0$ we obtain the inequalities

$$h(\mathbf{u}^s) \leq h(\mathbf{u}^*) \leq h(\mathbf{u}^s) + (\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s))^\top (\mathbf{u}^* - \mathbf{u}^s), \quad (15.29)$$

and hence that the inequality $(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s))^\top (\mathbf{u}^* - \mathbf{u}^s) \geq 0$ holds for all $s \geq 0$. We define $c := \sup_t \{\|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\|\}$, so that the inequality $\|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s)\| \leq c$ holds for all $s \geq 0$. From (15.27) and (15.23a) follow that $\sum_{s=0}^{t-1} \alpha_s^2 < p$ for all $t \geq 1$. By inserting this in (15.28), we then conclude that $\|\mathbf{u}^* - \mathbf{u}^t\|^2 < \|\mathbf{u}^* - \mathbf{u}^0\|^2 + pc^2$ for any $t \geq 1$; it follows that the sequence $\{\mathbf{u}^t\}$ is bounded.

Assume now that there is no subsequence \mathcal{T} such that $\{(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))^\top (\mathbf{u}^* - \mathbf{u}^t)\}_{t \in \mathcal{T}} \rightarrow 0$. Then, there exists an $\varepsilon > 0$ and a $t_\varepsilon > 1$ such that the inequality $(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s))^\top (\mathbf{u}^* - \mathbf{u}^s) \geq \varepsilon$ for all $s \geq t_\varepsilon$. By (15.28) and (15.23c) then follow that $\{\|\mathbf{u}^* - \mathbf{u}^t\|\} \rightarrow -\infty$, which is clearly impossible. Therefore, there is a subsequence \mathcal{T} such that $\{(\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t))^\top (\mathbf{u}^* - \mathbf{u}^t)\}_{t \in \mathcal{T}} \rightarrow 0$. From (15.29) then follows that $\{h(\mathbf{u}^t)\}_{t \in \mathcal{T}} \rightarrow h^*$.

The boundedness of the sequence $\{\mathbf{u}^t\}$ implies the existence of an accumulation point of the subsequence $\{\mathbf{u}^t\}_{t \in \mathcal{T}}$, say \mathbf{u}^∞ , and by the continuity of h then follows that $\mathbf{u}^\infty \in U^*$.

To show that \mathbf{u}^∞ is the only accumulation point of the sequence $\{\mathbf{u}^t\}$, let $\delta > 0$ and choose an $N_\delta \geq 0$ such that the inequalities $\|\mathbf{u}^\infty - \mathbf{u}^{N_\delta}\|^2 \leq \delta/2$ and $\sum_{s=N_\delta}^\infty \alpha_s^2 \leq \delta/(2c^2)$ hold. Consider any $t > N_\delta$. Analogously to the derivation of (15.28), and using (15.29), we then obtain that

$$\begin{aligned} \|\mathbf{u}^\infty - \mathbf{u}^t\|^2 &\leq \|\mathbf{u}^\infty - \mathbf{u}^{N_\delta}\|^2 + \sum_{s=N_\delta}^{t-1} \alpha_s^2 \|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^s) - \mathbf{v}(\mathbf{u}^s)\|^2 \\ &< \frac{\delta}{2} + \frac{\delta}{2c^2}c^2 = \delta. \end{aligned}$$

Since this holds for arbitrarily small values of $\delta > 0$, the theorem follows. \square

In subgradient optimization, the important *Polyak step length* rule has documented practical usefulness. Convergence to an optimal solution with this step length rule relies on the optimal objective value h^* . This result extends to the case of the conditional subgradient optimization method (15.21), for which the Polyak step length formula is defined by

$$\alpha_t := \frac{\theta_t (h^* - h(\mathbf{u}^t))}{\|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\|^2}, \quad 0 < \epsilon_1 \leq \theta_t \leq 2 - \epsilon_2 < 2, \quad t = 0, 1, 2, \dots \quad (15.30)$$

Proposition 15.7 (Convergence to a Solution by Polyak Step Lengths) *Let the method (15.21) be applied to the problem (15.10), with the step lengths α_t fulfilling the conditions (15.30). If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then $\{h(\mathbf{u}^t)\} \rightarrow h^*$ and $\{\mathbf{u}^t\} \rightarrow \mathbf{u}^\infty \in U^*$.*

Remark 15.3 With step lengths defined by (15.30), the case of subgradient projection according to (15.22) yields actual steps in (15.21) (that is, $\mathbf{u}^{t+1} - \mathbf{u}^t$) that are longer, as compared with the case of plain subgradients, that is, when $\{\mathbf{v}^t\} \equiv \{\mathbf{0}\}$ holds.

For an $\varepsilon > \bar{h} - h^* \geq 0$, finite convergence to ε -optimality can be achieved by replacing h^* in (15.30) by an upper bound $\bar{h} \geq h^*$ and letting $\theta_t \equiv 1$.

Proposition 15.8 (Finite ε -Optimality by Polyak Step Lengths) *Let the method (15.21) be applied to the problem (15.10), with the step lengths $\{\alpha_t\}$ defined by*

$$\alpha_t := \frac{\theta_t (\bar{h} - h(\mathbf{u}^t))}{\|\mathbf{b} - \mathbf{A}\mathbf{x}(\mathbf{u}^t) - \mathbf{v}(\mathbf{u}^t)\|^2}, \quad t = 0, 1, 2, \dots, \quad (15.31)$$

where $\theta_t \equiv 1$ and $\bar{h} > h^*$. If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then for any $\varepsilon > 0$, there is a $t_\varepsilon > 0$ such that $h(\mathbf{u}^{t_\varepsilon}) \geq 2h^* - \bar{h} - \varepsilon$.

Our results to follow in Sect. 15.3.2—on ergodic convergence to a primal solution—rely on the divergent series conditions (15.23), (15.27) on the step lengths in the method (15.21). By ensuring that the sequence $\{\alpha_t\}$ fulfills the conditions

$$\frac{\underline{a}}{b+t} \leq \alpha_t \leq \frac{\bar{a}}{b+t}, \quad 0 < \underline{a} \leq \bar{a}, \quad b > 0, \quad t = 0, 1, \dots, \quad (15.32)$$

convergence in terms of Theorem 15.2 can, however, be shown.⁷

Corollary 15.1 (Convergence to a Solution by Generalized Harmonic Series Step Lengths) *Let the method (15.21), where the sequence $\{\alpha_t\}$ fulfills (15.32), be applied to the problem (15.10). If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then $\{\mathbf{u}^t\} \rightarrow \mathbf{u}^\infty \in U^*$ hold.*

15.3.2 Ergodic Convergence in the Primal Problem

The conditional subgradient optimization method (15.21) constructs a sequence $\{\mathbf{x}(\mathbf{u}^t)\}$ of solutions to the Lagrangian subproblem (15.9). Due to the non-coordinability of the Lagrangian subproblem (see Sect. 15.2.2) this sequence is, however, not convergent. We propose a scheme for generating an ergodic sequence of subproblem solutions, which is shown to converge to the solution set X_{conv}^* . The generation of the ergodic sequence is computationally cheap and its storage requires a relatively small amount of memory. The sequence is defined by convexity weights that are proportional to the step lengths α_t ; the latter requirement is then generalized and improved.

From Propositions 15.2 and 15.4 follow that the set $\partial h(\mathbf{u}^\infty) \cap N_{\mathbb{R}_+^m}(\mathbf{u}^\infty)$ is nonempty. The next proposition establishes that the sequence $\{\mathbf{b} - A\mathbf{x}(\mathbf{u}^t)\}$ of subgradients to the dual objective function converges in an ergodic sense to an element that verifies optimality of the Lagrangian dual, in terms of Proposition 15.4.

Proposition 15.9 (Ergodic Subgradients Converge to the Optimality-Verifying Set) *Apply the method (15.21), (15.23) to the problem (15.10) and define the sequence $\{\bar{\mathbf{g}}^t\}$ as*

$$\bar{\mathbf{g}}^t := \frac{1}{\sum_{s=0}^{t-1} \alpha_s} \sum_{s=0}^{t-1} \alpha_s (\mathbf{b} - A\mathbf{x}(\mathbf{u}^s)), \quad t = 1, 2, \dots$$

If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then $\{\min_{\boldsymbol{\gamma} \in \partial h(\mathbf{u}^\infty) \cap N_{\mathbb{R}_+^m}(\mathbf{u}^\infty)} \|\boldsymbol{\gamma} - \bar{\mathbf{g}}^t\|\} \rightarrow 0$.

⁷For any sequence $\{\alpha_t\}$ of step lengths as defined in (15.31) [or according to (15.23), (15.27)], there exists constants \underline{a} , \bar{a} , and b such that the conditions (15.32) are fulfilled.

The ergodic sequence $\{\bar{\mathbf{x}}^t\}$ is then defined as weighted averages (that is, convex combinations) of the subproblem solutions found up to iteration t of (15.21), as

$$\bar{\mathbf{x}}^t := \frac{1}{\sum_{s=0}^{t-1} \alpha_s} \sum_{s=0}^{t-1} \alpha_s \mathbf{x}(\mathbf{u}^s), \quad t = 1, 2, \dots \quad (15.33)$$

The convergence of the sequence $\{\bar{\mathbf{x}}^t\}$ to the set X_{conv}^* , as expressed in (15.14), is then established in terms of fulfilment of the optimality conditions in Proposition 15.5.

Theorem 15.3 ($\{\bar{\mathbf{x}}^t\}$ Converges to the Optimal Set of the Convexified Problem)

Let the method (15.21), (15.23) be applied to the problem (15.10), the set X_{conv}^ and the sequence $\{\bar{\mathbf{x}}^t\}$ be given by (15.14) and (15.33), respectively, and suppose that the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded. Then, $\{\min_{\mathbf{x} \in X_{\text{conv}}^*} \|\mathbf{x} - \bar{\mathbf{x}}^t\|\} \rightarrow 0$.*

Efficient updates of the ergodic iterates $\bar{\mathbf{x}}^t$ requires only the previous ergodic iterate $\bar{\mathbf{x}}^{t-1}$ and subproblem solution $\mathbf{x}(\mathbf{u}^{t-1})$, according to the convex combination

$$\bar{\mathbf{x}}^t := \frac{\sum_{s=0}^{t-1} \alpha_s - \alpha_{t-1}}{\sum_{s=0}^{t-1} \alpha_s} \bar{\mathbf{x}}^{t-1} + \frac{\alpha_{t-1}}{\sum_{s=0}^{t-1} \alpha_s} \mathbf{x}(\mathbf{u}^{t-1}), \quad t = 2, 3, \dots \quad (15.34)$$

with $\bar{\mathbf{x}}^1 := \mathbf{x}(\mathbf{u}^0)$.

15.3.3 Enhanced Primal Ergodic Convergence

The convergence of the ergodic sequence of subproblem solutions according to Theorem 15.3 is, however, typically very slow. Efforts have therefore been put into enhancing the convergence speed, by exploiting more information from later subproblem solutions than from earlier ones. We next present a more general pattern for constructing the ergodic sequences; the ergodic sequence $\{\tilde{\mathbf{x}}^t\}$ is defined by

$$\tilde{\mathbf{x}}^t := \sum_{s=0}^{t-1} \mu_s^t \mathbf{x}(\mathbf{u}^s); \quad \sum_{s=0}^{t-1} \mu_s^t = 1; \quad \mu_s^t \geq 0, \quad s = 0, \dots, t-1, \quad (15.35)$$

where the convexity weights μ_s^t are defined as

$$\mu_s^t := \gamma_s^t \alpha_s, \quad s = 0, \dots, t-1, \quad t = 1, 2, \dots, \quad (15.36a)$$

and the parameters γ_s^t fulfil the requirements ($N > 0$ being a constant)

$$\gamma_s^t \geq \gamma_{s-1}^t, \quad s = 1, \dots, t - 1, \quad t = 2, 3, \dots, \tag{15.36b}$$

$$\left\{ \max_{s \in \{1, \dots, t-1\}} \{ \gamma_s^t - \gamma_{s-1}^t \} \right\} \rightarrow 0, \quad \text{as } t \rightarrow \infty, \tag{15.36c}$$

$$\gamma_0^t \rightarrow 0, \quad \text{as } t \rightarrow \infty, \tag{15.36d}$$

$$\gamma_{t-1}^t \leq N, \quad t = 1, 2, \dots \tag{15.36e}$$

The requirement (15.36b), together with the definition (15.36a), implies the inequality $\mu_s^t (\mu_{s-1}^t)^{-1} \geq \alpha_s (\alpha_{s-1})^{-1}$, which means that the ratio of any two consecutive convexity weights should be no less than that of the corresponding step lengths. The requirement (15.36c) implies that the difference between consecutive pairs of subsequent convexity weights tends to zero as t increases, meaning that no primal iterate may be completely neglected. The requirement (15.36d) implies, however, that the weight (that is, μ_0^t) of the first iterate [that is, $\mathbf{x}(\mathbf{u}^0)$] tends to zero when t increases. The requirements (15.36e), (15.36d), and (15.36b) assure that, for decreasing step lengths α_s , the convexity weights μ_s^t decrease at a rate not slower than that of the step lengths. Efficient updates of the ergodic iterates $\tilde{\mathbf{x}}^t$ can be made, analogously to (15.34).

Theorem 15.4 ($\{\tilde{\mathbf{x}}^t\}$ Converges to the Optimal Set of the Convexified Problem)

Apply the method (15.21), (15.23), (15.27) to the problem (15.10) and define the set X_{conv}^* and the sequence $\{\tilde{\mathbf{x}}^t\}$ by (15.14) and (15.35), (15.36), respectively. If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded, then $\{\min_{\mathbf{x} \in X_{\text{conv}}^*} \|\mathbf{x} - \tilde{\mathbf{x}}^t\|\} \rightarrow 0$.

Remark 15.4 For any fixed value of $s \in \{0, \dots, t - 1\}$, the requirements (15.36b)–(15.36e) imply that $\gamma_s^t \leq \gamma_0^t + s \cdot \max_{r \in \{1, \dots, t-1\}} \{\gamma_r^t - \gamma_{r-1}^t\} \rightarrow 0$ as $t \rightarrow \infty$. This yields that $\mu_s^t = \gamma_s^t \alpha_s \rightarrow 0$ as $t \rightarrow \infty$, since $\alpha_s < \infty, s = 0, 1, \dots$

Remark 15.5 The ergodic sequence $\{\bar{\mathbf{x}}^t\}$, defined in (15.33), is equivalent to the special case of (15.35), (15.36) defined by $\gamma_s^t := (\sum_{r=0}^{t-1} \alpha_r)^{-1}$ (being independent of s). For this choice of $\{\gamma_s^t\}$, the requirements (15.36b)–(15.36e) are fulfilled with $N := (\alpha_0)^{-1}$.

Remark 15.6 For the special case of (15.35), (15.36) given by (15.32) with $a := \underline{a} = \bar{a}$, that is, *modified harmonic series* step lengths

$$\alpha_s := \frac{a}{b + s}, \quad a > 0, \quad b > 0, \quad s = 0, 1, \dots, t - 1, \tag{15.37}$$

and choosing $\gamma_s^t := (t\alpha_s)^{-1}$, the convexity weights become $\mu_s^t = t^{-1}$ (then, $\tilde{\mathbf{x}}^t$ equals a simple average of the subproblem solutions found). For these choices of

$\{\gamma_s^t\}$ and $\{\alpha_s\}$, the requirements (15.36b)–(15.36e) are fulfilled, with $N := a^{-1} \cdot \max\{b, 1\}$.

By choosing the step lengths α_t according to (15.37) and letting the sequence $\{\mu_s^t\}$ of convexity weights fulfil the requirements

$$\mu_s^t \geq \mu_{s-1}^t, \quad s = 1, \dots, t-1, \quad t = 2, 3, \dots, \quad (15.38a)$$

$$\left\{ t \cdot \max_{s \in \{1, \dots, t-1\}} \{\mu_s^t - \mu_{s-1}^t\} \right\} \rightarrow 0, \quad \text{as } t \rightarrow \infty, \quad (15.38b)$$

$$t \mu_{t-1}^t \leq M, \quad t = 1, 2, \dots, \quad (15.38c)$$

where $M > 0$ is a constant, it can be shown that also the requirements (15.36b)–(15.36e) are fulfilled, with $N = a^{-1}M \cdot \max\{b, 1\}$.

The so-called s^k -rule, for which the corresponding convexity weights, μ_s^t , fulfill the requirements (15.38), is defined by

$$\mu_s^t := \frac{(s+1)^k}{\sum_{r=0}^{t-1} (r+1)^k}, \quad s = 0, \dots, t-1, \quad t = 1, 2, \dots, \quad k \geq 0. \quad (15.39)$$

For $k > 0$, the s^k -rule results in an ergodic sequence (15.35) in which the later iterates are assigned higher weights than the earlier ones. For larger values of k , the weights are shifted towards later iterates. Given that step lengths α_t according to (15.37) are utilized in the method (15.21) applied to the Lagrangian dual (15.10), it can be shown that weights according to (15.39) fulfill the requirements (15.38), so that convergence for the resulting primal ergodic sequence $\{\tilde{\mathbf{x}}^t\}$ to the optimal set X_{conv}^* can be established.

Remark 15.7 Note that the s^0 -rule yields $\mu_s^t = t^{-1}$ [cf. Remark 15.6, where $\gamma_s^t = (t\alpha_s)^{-1}$]. For $k > 0$, the s^k -rule results in an ergodic sequence in which later iterates are assigned higher weights than earlier ones. For larger values of k , the weights are shifted towards increasingly later iterates.

Remark 15.8 Since the conditional subgradient optimization method (15.21) is memory-less, without loss of generality the computation of the ergodic sequences $\{\bar{\mathbf{x}}^t\}$, $\{\bar{\mathbf{g}}^t\}$, and $\{\tilde{\mathbf{x}}^t\}$ may be postponed a finite number of iterations. If the postponement is “long enough”, each $\bar{\mathbf{x}}^t$ (or $\tilde{\mathbf{x}}^t$) will be a solution to the Lagrangian subproblem at the optimal dual point \mathbf{u}^∞ , as defined in Theorem 15.2.

15.3.4 Finite Primal Feasibility and Finite ε -Optimality

Theorem 15.3 establishes optimality in the limit for the sequence $\{\bar{\mathbf{x}}^t\}$. While dual feasibility holds for the sequence $\{\mathbf{u}^t\}$, in general neither primal feasibility (that is, $A\bar{\mathbf{x}}^t \geq \mathbf{b}$) nor complementarity [that is, $(\mathbf{u}^t)^\top(\mathbf{b} - A\bar{\mathbf{x}}^t) = 0$] will be finitely satisfied. Eventually, however, $\bar{\mathbf{x}}^t$ will be both near feasible and near complementary. Whenever finite primal feasibility is required, a procedure can be applied that converts any $\bar{\mathbf{x}}^t$ into a feasible solution to (15.5), for example, by computing the Euclidean projection of $\bar{\mathbf{x}}^t$ onto the feasible set, as

$$\bar{\mathbf{x}}_{\text{proj}}^t := \operatorname{argmin}_{\mathbf{x} \in X_{\text{conv}}} \left\{ \|\mathbf{x} - \bar{\mathbf{x}}^t\| \mid A\mathbf{x} \geq \mathbf{b} \right\}. \quad (15.40)$$

Solving (15.40) regularly may, however, be computationally too expensive. Instead, we consider a heuristic procedure, preferably exploiting the structure of the set $\{\mathbf{x} \in X_{\text{conv}} \mid A\mathbf{x} \geq \mathbf{b}\}$ in the search for a feasible and near optimal solution to (15.40). Let the function $\delta : \mathbb{R}_+ \mapsto \mathbb{R}_+$ be continuous and such that $\delta(\beta) > 0$ whenever $\beta > 0$ and $\lim_{\beta \rightarrow 0^+} \delta(\beta) = 0$. Define a *heuristic projection* $\bar{\mathbf{x}}_{\text{heur}}^t$ of $\bar{\mathbf{x}}^t \in X_{\text{conv}}$ by the inclusion

$$\bar{\mathbf{x}}_{\text{heur}}^t \in \{\mathbf{x} \in X_{\text{conv}} \mid A\mathbf{x} \geq \mathbf{b}\}, \quad (15.41a)$$

and such that

$$\left\| \bar{\mathbf{x}}_{\text{heur}}^t - \bar{\mathbf{x}}_{\text{proj}}^t \right\| \leq \delta(\beta) \quad \text{whenever} \quad \left\| \bar{\mathbf{x}}^t - \bar{\mathbf{x}}_{\text{proj}}^t \right\| \leq \beta. \quad (15.41b)$$

Theorem 15.5 (Convergence to Primal Optimality by Heuristic Projections)

Let the method (15.21), (15.23), (15.27) be applied to the problem (15.10). Let the set X_{conv}^* and the sequences $\{\bar{\mathbf{x}}^t\}$, $\{\bar{\mathbf{x}}_{\text{proj}}^t\}$, and $\{\bar{\mathbf{x}}_{\text{heur}}^t\}$ be defined by (15.14), (15.33), (15.40), and (15.41a), respectively. If the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded and the conditions (15.41b) hold, then $\left\{ \min_{\mathbf{x} \in X_{\text{conv}}^*} \|\mathbf{x} - \bar{\mathbf{x}}_{\text{heur}}^t\| \right\} \rightarrow 0$.

We can now construct an algorithm employing heuristic projections and yielding convergence to the optimal value in both the primal and dual procedures.

Corollary 15.2 (Finite Termination at ε -Optimality) *Given the assumptions of Theorem 15.5, for every $\varepsilon > 0$ there is a $t_\varepsilon > 0$ such that $\mathbf{c}^\top \bar{\mathbf{x}}_{\text{heur}}^t - h(\mathbf{u}^t) \leq \varepsilon$ holds for all $t \geq t_\varepsilon$.*

This projection principle is thus a way to recover primal feasibility, and eventually also optimality, in an otherwise purely dual method.

15.4 Dual Cutting-Planes: Dantzig–Wolfe Decomposition

Assuming that a nonempty subset $\overline{Q} \subseteq Q$ is given, the Lagrangian dual function is outer approximated by the function $\overline{h} : \mathbb{R}^m \mapsto \mathbb{R}$ as defined by

$$\overline{h}(\mathbf{u}) := \mathbf{b}^\top \mathbf{u} + \min_{q \in \overline{Q}} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x}^q \right\}. \quad (15.42)$$

This function is formed by the point-wise minimum of $|\overline{Q}|$ affine functions. Hence, it is piecewise affine and concave, as is the Lagrangian dual function h . Clearly, $\overline{h}(\mathbf{u}) \geq h(\mathbf{u})$ holds for any $\mathbf{u} \in \mathbb{R}^m$. Hence, the problem

$$\overline{h}^* := \max_{\mathbf{u} \in \mathbb{R}_+^m} \overline{h}(\mathbf{u}), \quad (15.43)$$

is a *relaxation* of the Lagrangian dual problem (15.10), so that $\overline{h}^* \geq h^*$ holds. The subdifferential of the function \overline{h} at \mathbf{u} is given by

$$\partial \overline{h}(\mathbf{u}) := \text{conv}_{q \in \overline{Q}(\mathbf{u})} \{ \mathbf{b} - A\mathbf{x}^q \}, \quad \mathbf{u} \in \mathbb{R}_+^m,$$

where $\overline{Q}(\mathbf{u})$ denotes the optimal index set of the inner minimization in (15.42), and its conditional subdifferential is given by

$$\partial^{\mathbb{R}_+^m} \overline{h}(\mathbf{u}) = \partial \overline{h}(\mathbf{u}) - N_{\mathbb{R}_+^m}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}_+^m.$$

The relaxed dual problem (15.43) has a bounded optimal solution if and only if

$$\mathbf{0} \in \bigcup_{\mathbf{u} \in \mathbb{R}_+^m} \partial^{\mathbb{R}_+^m} \overline{h}(\mathbf{u}). \quad (15.44)$$

Remark 15.9 A sufficient condition for the relaxed Lagrangian dual problem (15.43) to have a bounded optimal solution is that some point $\mathbf{x}^{\overline{q}}, \overline{q} \in \overline{Q}$, is feasible in the original mixed-binary linear optimization problem (15.2), since this implies that

$$\begin{aligned} \overline{h}^* &= \max_{\mathbf{u} \in \mathbb{R}_+^m} \left\{ \min_{q \in \overline{Q}} \left\{ \mathbf{c}^\top \mathbf{x}^q + \mathbf{u}^\top (\mathbf{b} - A\mathbf{x}^q) \right\} \right\} \\ &\leq \max_{\mathbf{u} \in \mathbb{R}_+^m} \left\{ \mathbf{c}^\top \mathbf{x}^{\overline{q}} + \mathbf{u}^\top (\mathbf{b} - A\mathbf{x}^{\overline{q}}) \right\} \\ &= \mathbf{c}^\top \mathbf{x}^{\overline{q}} \end{aligned}$$

holds, where the second equality holds because $A\mathbf{x}^{\overline{q}} \geq \mathbf{b}$.

Assuming that the condition (15.44) holds, we find a $\bar{\mathbf{u}} \in \mathbb{R}_+^m$ such that the inclusion $\mathbf{0} \in \partial^{\mathbb{R}_+^m} \bar{h}(\bar{\mathbf{u}})$ holds, implying that $\bar{\mathbf{u}}$ is optimal in (15.42). Hence, $\bar{h}^* = \bar{h}(\bar{\mathbf{u}})$ holds. The value of the Lagrangian dual function at this point is

$$h(\bar{\mathbf{u}}) = \mathbf{b}^\top \bar{\mathbf{u}} + \min_{\mathbf{x} \in X} \left\{ (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x} \right\} = \mathbf{b}^\top \bar{\mathbf{u}} + \min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^q \right\}.$$

From the Lagrangian dual problem (15.10) we conclude that $h(\bar{\mathbf{u}}) \leq h^*$. To summarize, the relations $h(\bar{\mathbf{u}}) \leq h^* \leq \bar{h}^*$ hold, that is, the values \bar{h}^* and $h(\bar{\mathbf{u}})$ provide upper and lower bounds, respectively, on the optimal value h^* .

If the equality $h(\bar{\mathbf{u}}) = \bar{h}^*$ holds, it follows that $h(\bar{\mathbf{u}}) = h^*$, implying that $\bar{\mathbf{u}}$ is optimal in (15.10). In the case when $h(\bar{\mathbf{u}}) < \bar{h}^*$ holds, we consider any solution to the minimization in (15.8) for $\mathbf{u} = \bar{\mathbf{u}}$, which we denote by $\mathbf{x}^{\bar{q}}$, where $\bar{q} \in \mathcal{Q}(\bar{\mathbf{u}})$, such that $h(\bar{\mathbf{u}}) = \mathbf{b}^\top \bar{\mathbf{u}} + (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^{\bar{q}}$ holds. Then, the relations

$$\mathbf{b}^\top \bar{\mathbf{u}} + (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^{\bar{q}} < \bar{h}^* = \mathbf{b}^\top \bar{\mathbf{u}} + \min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^q \right\}$$

hold, which yields that $\bar{q} \notin \mathcal{Q}$. By then augmenting the set \mathcal{Q} with the index \bar{q} , an improved outer approximation of the Lagrangian dual function is obtained. By resolving the problem (15.43) and repeating, an iterative procedure for solving the Lagrangian dual problem (15.10) is obtained. Its convergence is finite, since the set \mathcal{Q} is finite and since a point $\mathbf{x}^{\bar{q}}$ can never be regenerated.

The above procedure is commonly described as a *cutting-plane* or *constraint generation* procedure for the LP formulation (15.11b) of the Lagrangian dual problem. This formulation can be restated as

$$h^* = \max_{(\mathbf{u}, v) \in \mathbb{R}_+^m \times \mathbb{R}} \left\{ \mathbf{b}^\top \mathbf{u} + v \mid \mathbf{c}^\top \mathbf{x}^q - (A\mathbf{x}^q)^\top \mathbf{u} - v \geq 0, q \in \mathcal{Q} \right\}. \quad (15.45)$$

Let $(\bar{\mathbf{u}}, \bar{v}) \in \mathbb{R}_+^m \times \mathbb{R}$ be optimal in the relaxed problem

$$\bar{h}^* = \max_{(\mathbf{u}, v) \in \mathbb{R}_+^m \times \mathbb{R}} \left\{ \mathbf{b}^\top \mathbf{u} + v \mid \mathbf{c}^\top \mathbf{x}^q - (A\mathbf{x}^q)^\top \mathbf{u} - v \geq 0, q \in \bar{\mathcal{Q}} \right\}. \quad (15.46)$$

Then $\bar{h}^* = \mathbf{b}^\top \bar{\mathbf{u}} + \bar{v}$. The question then is if all constraints in the problem (15.45) are satisfied or not at the point $(\mathbf{u}, v) = (\bar{\mathbf{u}}, \bar{v})$. This is determined by finding the *most violated* constraint, if any, and amounts to solving

$$\begin{aligned} \min_{q \in \mathcal{Q}} \left\{ \mathbf{c}^\top \mathbf{x}^q - (A\mathbf{x}^q)^\top \bar{\mathbf{u}} - \bar{v} \right\} &= \min_{q \in \mathcal{Q}} \left\{ \mathbf{b}^\top \bar{\mathbf{u}} + (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^q - \bar{h}^* \right\} \\ &= \mathbf{b}^\top \bar{\mathbf{u}} + (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^{\bar{q}} - \bar{h}^* \\ &= h(\bar{\mathbf{u}}) - \bar{h}^*, \end{aligned} \quad (15.47)$$

where $\mathbf{x}^q \in X(\bar{\mathbf{u}})$ solves the minimization in (15.8) at $\mathbf{u} = \bar{\mathbf{u}}$. If the minimum value $h(\bar{\mathbf{u}}) - \bar{h}^*$ is negative, then a most violated constraint has been identified. Otherwise, $h(\bar{\mathbf{u}}) = \bar{h}^*$ holds and $\bar{\mathbf{u}} \in \mathbb{R}_+^m$ solves the dual problem (15.10).

The dual cutting-plane procedure sketched above is dually equivalent to applying *Dantzig–Wolfe decomposition* to the convexified problem (15.5), which by the reformulation (15.11c) is equivalent to the *complete Dantzig–Wolfe master problem*

$$\min_{\lambda} \left\{ \sum_{q \in \mathcal{Q}} (\mathbf{c}^\top \mathbf{x}^q) \lambda_q \mid \sum_{q \in \mathcal{Q}} (A\mathbf{x}^q) \lambda_q \geq \mathbf{b}; \sum_{q \in \mathcal{Q}} \lambda_q = 1; \lambda_q \geq 0, q \in \mathcal{Q} \right\}, \tag{15.48}$$

while also being the LP dual of the problem (15.45). Since

$$X_{\text{conv}} = \left\{ \mathbf{x} = \sum_{q \in \mathcal{Q}} \lambda_q \mathbf{x}^q \mid \sum_{q \in \mathcal{Q}} \lambda_q = 1; \lambda_q \geq 0, q \in \mathcal{Q} \right\}$$

holds, the variables λ_q in (15.48) are commonly referred to as *convexity variables* and the equality $\sum_{q \in \mathcal{Q}} \lambda_q = 1$ as a *convexity constraint*. In Dantzig–Wolfe decomposition, this problem is solved using the linear programming technique called *column generation*, in which, for this application, a column corresponds to the data of a convexity variable, that is, the values of the scalar $\mathbf{c}^\top \mathbf{x}^q$, the vector $A\mathbf{x}^q$, and a 1.

Using the analogous reformulation as in (15.11), we obtain

$$\bar{h}^* = \max_{\mathbf{u} \in \mathbb{R}_+^m} \left\{ \mathbf{b}^\top \mathbf{u} + \min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x}^q \right\} \right\} \tag{15.49a}$$

$$= \min_{\lambda} \left\{ \sum_{q \in \overline{\mathcal{Q}}} (\mathbf{c}^\top \mathbf{x}^q) \lambda_q \mid \sum_{q \in \overline{\mathcal{Q}}} (A\mathbf{x}^q) \lambda_q \geq \mathbf{b}; \sum_{q \in \overline{\mathcal{Q}}} \lambda_q = 1; \lambda_q \geq 0, q \in \overline{\mathcal{Q}} \right\} \tag{15.49b}$$

$$= \min_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} \mid A\mathbf{x} \geq \mathbf{b}; \mathbf{x} \in \text{conv}\{ \mathbf{x}^q \}_{q \in \overline{\mathcal{Q}}} \right\}. \tag{15.49c}$$

The problem (15.49b) is the *restricted Dantzig–Wolfe master problem*, referring to the fact that this problem includes the variables λ_q , with the corresponding columns $(\mathbf{c}^\top \mathbf{x}^q, (A\mathbf{x}^q)^\top, 1)^\top$, only for $q \in \overline{\mathcal{Q}} \subseteq \mathcal{Q}$, which is clearly equivalent to imposing the restrictions $\lambda_q = 0$ for $q \in \mathcal{Q} \setminus \overline{\mathcal{Q}}$ in (15.48).

Let $(\bar{\mathbf{u}}, \bar{v}) \in \mathbb{R}_+^m \times \mathbb{R}$ be an optimal solution to the LP dual of the restricted master problem. Using strong duality for LP, it then holds that $\bar{h}^* = \mathbf{b}^\top \bar{\mathbf{u}} + \bar{v}$. The LP reduced cost for any variable $\lambda_q, q \in \mathcal{Q}$, in the complete master problem can

then be expressed as

$$\bar{c}^q := \mathbf{c}^\top \mathbf{x}^q - \bar{\mathbf{u}}^\top A \mathbf{x}^q - \bar{v} = \mathbf{b}^\top \bar{\mathbf{u}} + (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^q - \bar{h}^*.$$

The problem of finding the variable λ_q , $q \in \mathcal{Q}$, with the most negative reduced cost reduces to find

$$\min_{q \in \mathcal{Q}} \left\{ (\mathbf{c} - A^\top \bar{\mathbf{u}})^\top \mathbf{x}^q \right\} = \min_{\mathbf{x} \in X} \left\{ (\mathbf{c} - A^\top \mathbf{u})^\top \mathbf{x} \right\}, \quad (15.50)$$

which is in this context known as a *column generation problem*, or *Dantzig–Wolfe subproblem*. If the solution to (15.50) corresponds to a negative reduced cost $\bar{c}^q < 0$, $\bar{q} \in \mathcal{Q}$,⁸ then $\bar{\mathcal{Q}} := \mathcal{Q} \cup \{\bar{q}\}$ and the problem (15.49b) is resolved. When no more columns having negative reduced costs can be found, which clearly occurs finitely, the convexified problem (15.5) has been solved.

Note that the problem (15.50) is equivalent to that of finding the most violated cutting-plane in the dual space [cf. (15.47)], since each constraint in the cutting-plane formulation of the Lagrangian dual corresponds to a convexity variable in the master problem. Note also that relaxing the dual problem (15.45) into (15.46) is dually equivalent to restricting the complete Dantzig–Wolfe master problem (15.48) into (15.49b).

Remark 15.10 Many textbooks derive the Dantzig–Wolfe decomposition method for an LP problem with two sets of explicit affine constraints, in our notation corresponding to $A \mathbf{x} \geq \mathbf{b}$ and $\mathbf{x} \in X_{\text{conv}}$. The method can, however, be applied also when no explicit representation of the set X_{conv} in terms of affine constraints is available, as long as the subproblem (15.50) can be solved by some means.

Remark 15.11 If the mixed-binary linear optimization problem (15.2) has a Cartesian product structure, as in Remark 15.1, then a separate set of convexity variables and a convexity constraint can be defined for each of the sets in the product. Letting $\{\mathbf{y}_s^q\}_{q \in \mathcal{Q}_s}$ denote the extreme points of $\text{conv } Y_s$, $s \in \mathcal{S}$, the complete Dantzig–Wolfe master problem for the problem (15.7) becomes

$$z^* = \min_{\lambda_{sq}, q \in \mathcal{Q}_s, s \in \mathcal{S}} \sum_{s \in \mathcal{S}} \sum_{q \in \mathcal{Q}_s} (\mathbf{c}_s^\top \mathbf{y}_s^q) \lambda_{sq} \quad (15.51a)$$

$$\text{subject to } \sum_{s \in \mathcal{S}} \sum_{q \in \mathcal{Q}_s} (A_s \mathbf{y}_s^q) \lambda_{sq} \geq \mathbf{b}, \quad (15.51b)$$

⁸It then holds that $\bar{q} \in \mathcal{Q} \setminus \bar{\mathcal{Q}}$, since $\bar{c}^q \geq 0$ holds for all $q \in \bar{\mathcal{Q}}$.

$$\sum_{q \in \mathcal{Q}_s} \lambda_{sq} = 1, \quad s \in \mathcal{S}, \quad (15.51c)$$

$$\lambda_{sq} \geq 0, \quad q \in \mathcal{Q}_s, s \in \mathcal{S}. \quad (15.51d)$$

When applying column generation to the master problem (15.51), there will be one Dantzig–Wolfe subproblem for each $s \in \mathcal{S}$.

Since each vector y_s , $s \in \mathcal{S}$, is here described by a separate set of convexity variables, the master problem (15.51) is based on a *disaggregate representation*, while the master problem (15.48) is based on an *aggregate representation*. Whenever the feasible set of the original problem (15.2) is described by a Cartesian product set and coupling constraints, as in (15.7), a disaggregate representation is typically favourable. This is due to (1) that the disaggregate and aggregate complete master problems contain $\sum_{s \in \mathcal{S}} |\mathcal{Q}_s|$ and $\prod_{s \in \mathcal{S}} |\mathcal{Q}_s|$ convexity variables, respectively, and (2) that a disaggregate restricted master problem is a relaxation of the corresponding aggregate restricted master problem, and hence the former typically provides stronger bounds on z^* .

15.5 A Two-Phase Method: Subgradient Optimization and Dantzig–Wolfe Decomposition

Subgradient optimization and Dantzig–Wolfe decomposition as means for solving the Lagrangian dual problem (15.10) possess their distinct advantages and disadvantages, which are consequences of the nondifferentiability of the Lagrangian dual function and the inherent properties of these solution principles.

In *subgradient optimization*, the update of the dual iterate is inexpensive once the Lagrangian relaxed problem has been solved. The method lacks, however, a good termination criterion, and in practice it is therefore often run for a preset number of iterations. Further, primal feasible solutions are typically not easily obtained, neither to the mixed-binary linear optimization problem (15.2) nor to its convexified version (15.5). It is, however, quite common to use a Lagrangian heuristic—tailored to each specific application—to convert Lagrangian subproblem solutions into feasible solutions to the original mixed-binary problem.

Dantzig–Wolfe decomposition converges finitely and produces feasible solutions to the convexified problem (15.5). The latter property allows early termination when the upper and lower bounds on the optimal value of (15.5) are close enough. Each iteration of the method is, however, expensive, since the next dual iterate is found by reoptimizing the LP restricted master problem. Further, due to the inherent instability of cutting-plane approaches, the method typically shows a poor convergence behaviour, in the sense that successive dual iterates may be very far apart. This phenomenon is commonly prevented by the introduction of a stabilization mechanism, such as *trust regions* in the dual space.

Another way to improve a Dantzig–Wolfe decomposition scheme is to heuristically generate an initial set of columns of high quality. We here describe a specific means to generate such columns, leading to a *two-phase method* that benefits from the advantages of both subgradient optimization and Dantzig–Wolfe decomposition. A first *prediction phase* employs subgradient optimization, in which Lagrangian subproblem solutions are stored. At termination, a lower bound on the optimal value of the convexified problem as well as a number of Lagrangian subproblem solutions are at hand. A second *solution phase* uses these solutions to construct an initial restricted master problem, whereafter the Dantzig–Wolfe method is employed.

The prediction phase aims at setting up a high quality initial restricted master problem, such that the solution phase can attain and verify optimality in fewer column generations. If the prediction phase works perfectly, then it is enough to solve one restricted master problem and make one column generation in order to reach and verify optimality. Hence, the solution phase can alternatively be viewed as an evaluation of the outcome of the prediction phase, and if needed compensate for its shortcoming. The rationale for the prediction phase is that the subgradient optimization method asymptotically finds columns that are optimal in the complete master problem, if the step lengths are appropriately chosen, as demonstrated below.

We consider solving the problem (15.10) by using the conditional subgradient method (15.21), with a sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ that is bounded and with step lengths α_t fulfilling the conditions (15.23) and (15.27), such that the assumptions of Theorem 15.2 are fulfilled. This method will then in the limit find a dual solution \mathbf{u}^∞ , such that $(\mathbf{u}^\infty, v^\infty)$ is optimal in the LP dual (15.45) of the complete master problem (15.48).

Define the index sets

$$\mathcal{T}_q := \left\{ t \in \mathbb{Z}_+ \mid \mathbf{x}(\mathbf{u}^t) = \mathbf{x}^q \right\}, \quad q \in \mathcal{Q},$$

and

$$\widehat{\mathcal{Q}} := \left\{ q \in \mathcal{Q} \mid \mathcal{T}_q \text{ is infinite} \right\} \subseteq \mathcal{Q},$$

that is, $\widehat{\mathcal{Q}}$ contains the indices $q \in \mathcal{Q}$ such that \mathbf{x}^q solves the Lagrangian subproblem an infinite number of times. Consider then the corresponding restricted master problem

$$\widehat{h}^* := \min_{\lambda} \left\{ \sum_{q \in \widehat{\mathcal{Q}}} (\mathbf{c}^\top \mathbf{x}^q) \lambda_q \mid \sum_{q \in \widehat{\mathcal{Q}}} (A\mathbf{x}^q) \lambda_q \geq \mathbf{b}; \sum_{q \in \widehat{\mathcal{Q}}} \lambda_q = 1; \lambda_q \geq 0, q \in \widehat{\mathcal{Q}} \right\}. \tag{15.52}$$

Theorem 15.6 (Asymptotic Generation of Optimal Columns) *Let the method (15.21) be applied to the problem (15.10), with the step lengths $\{\alpha_t\}$ fulfilling the conditions (15.23) and (15.27), and assume that the sequence $\{\mathbf{v}(\mathbf{u}^t)\}$ is bounded. Then $\widehat{h}^* = h^*$ holds.*

Further, let $\mathbf{u}^\infty \in U^*$ be the limit point for the sequence $\{\mathbf{u}^t\}$ (cf. Theorem 15.2) and let $v = v^\infty$ be optimal in (15.45) given that $\mathbf{u} = \mathbf{u}^\infty$, so that $(\mathbf{u}^\infty, v^\infty)$ is an optimal dual solution to the complete master problem (15.48). For each Lagrangian subproblem solution $\mathbf{x}(\mathbf{u}^t)$, $t \in \mathbb{Z}_+$, define the reduced cost $\bar{c}_t := \mathbf{c}^\top \mathbf{x}(\mathbf{u}^t) - (\mathbf{u}^\infty)^\top \mathbf{A} \mathbf{x}(\mathbf{u}^t) - v^\infty$. Then $\bar{c}_t = 0$ holds for every t that is sufficiently large.

Proof Consider the ergodic sequence $\{\bar{\mathbf{x}}^t\}$ of subproblem solutions given by (15.33), that is, $\bar{\mathbf{x}}^t = \left(\sum_{s=0}^{t-1} \alpha_s\right)^{-1} \sum_{s=0}^{t-1} \alpha_s \mathbf{x}(\mathbf{u}^s)$, $t = 1, 2, \dots$. Define the convexity weights

$$\lambda_q^t := \frac{1}{\sum_{s=0}^{t-1} \alpha_s} \sum_{s \in \mathcal{T}_q^t} \alpha_s, \quad q \in \mathcal{Q}, \quad t = 1, 2, \dots, \tag{15.53}$$

where $\mathcal{T}_q^t := \{s \in \{0, 1, \dots, t-1\} \mid \mathbf{x}(\mathbf{u}^s) = \mathbf{x}^q\}$, $q \in \mathcal{Q}$, $t = 1, 2, \dots$. Clearly, $\lambda_q^t \geq 0$, $q \in \mathcal{Q}$, and $\sum_{q \in \mathcal{Q}} \lambda_q^t = 1$ hold for $t = 1, 2, \dots$. Then, the ergodic solution $\bar{\mathbf{x}}^t$ can alternatively be expressed as the convex combination

$$\bar{\mathbf{x}}^t = \sum_{q \in \mathcal{Q}} \lambda_q^t \mathbf{x}^q, \quad t = 1, 2, \dots \tag{15.54}$$

The sequence $\{\lambda^t\}_{t=1}^\infty$, where $\lambda^t := (\lambda_q^t)_{q \in \mathcal{Q}}$, is contained in the unit simplex in $\mathbb{R}^{|\mathcal{Q}|}$, in which it thus has some accumulation point, say $\bar{\lambda}$. The sequence $\{\bar{\mathbf{x}}^t\}$ then has an accumulation point at $\bar{\mathbf{x}} := \sum_{q \in \mathcal{Q}} \bar{\lambda}_q \mathbf{x}^q$. From Theorem 15.3 follows that $\bar{\mathbf{x}} \in X_{\text{conv}}^*$ must hold, which implies that $\bar{\lambda}$ is optimal in the complete master problem (15.48).

For any $q \notin \widehat{\mathcal{Q}}$, the set \mathcal{T}_q is finite and the formula (15.53) together with the divergent series conditions (15.23) yield that $\bar{\lambda}_q = 0$ holds. Hence, $\bar{\lambda}$ is feasible in the restricted master problem (15.52). It follows that $\widehat{h}^* = h^*$ holds.

From the optimality of $(\mathbf{u}^\infty, v^\infty)$ in the LP dual (15.45) of the complete master problem (15.48) follows that

$$h^* = \mathbf{b}^\top \mathbf{u}^\infty + v^\infty. \tag{15.55}$$

Since the Lagrangian dual function h is polyhedral, for large enough values of t , $\mathbf{x}(\mathbf{u}^t)$ solves the Lagrangian subproblem at \mathbf{u}^t as well as at \mathbf{u}^∞ . Hence, it holds that $\mathbf{b} - \mathbf{A} \mathbf{x}(\mathbf{u}^t) \in \partial h(\mathbf{u}^t) \cap \partial h(\mathbf{u}^\infty)$, and it follows that the inequalities

$$h(\mathbf{u}^\infty) \leq h(\mathbf{u}^t) + (\mathbf{b} - \mathbf{A} \mathbf{x}(\mathbf{u}^t))^\top (\mathbf{u}^\infty - \mathbf{u}^t)$$

and

$$h(\mathbf{u}^t) \leq h(\mathbf{u}^\infty) + (\mathbf{b} - \mathbf{A} \mathbf{x}(\mathbf{u}^t))^\top (\mathbf{u}^t - \mathbf{u}^\infty)$$

hold. By combining these inequalities and simplifying we obtain the equalities

$$\begin{aligned} h(\mathbf{u}^\infty) &= h(\mathbf{u}^t) + (\mathbf{b} - A\mathbf{x}(\mathbf{u}^t))^\top (\mathbf{u}^\infty - \mathbf{u}^t) \\ &= \mathbf{c}^\top \mathbf{x}(\mathbf{u}^t) + (\mathbf{b} - A\mathbf{x}(\mathbf{u}^t))^\top \mathbf{u}^t + (\mathbf{b} - A\mathbf{x}(\mathbf{u}^t))^\top (\mathbf{u}^\infty - \mathbf{u}^t) \\ &= \mathbf{c}^\top \mathbf{x}(\mathbf{u}^t) - (A\mathbf{x}(\mathbf{u}^t))^\top \mathbf{u}^\infty + \mathbf{b}^\top \mathbf{u}^\infty, \end{aligned}$$

which together with the equality (15.55) and $h^* = h(\mathbf{u}^\infty)$ yield that

$$\mathbf{c}^\top \mathbf{x}(\mathbf{u}^t) - (A\mathbf{x}(\mathbf{u}^t))^\top \mathbf{u}^\infty - v^\infty = 0$$

holds when t is large enough. \square

According to Theorem 15.6, all columns that are needed to solve the complete master problem (15.48) are eventually found in the subgradient optimization. In case of nondegeneracy at a dual optimum, late iterations will provide exactly the optimal basic columns of the complete master problem.⁹ Further, columns found in the early iterations are never indispensable. These properties justify the use of subgradient optimization for predicting optimal basic columns before applying column generation.

In practice, the two-phase method works as follows. In the prediction phase, \bar{t} subgradient optimization iterations are performed, with step lengths fulfilling the conditions (15.23) and (15.27). Lagrangian subproblem solutions are collected and columns in the restricted master problem are constructed from iteration $\underline{t} \leq \bar{t}$. Upon termination of this phase, an initial restricted master problem is available. Thereafter, Dantzig–Wolfe decomposition is used as usual. In order to benefit from the two-phase method, the Lagrangian relaxed problem must be computationally cheap in comparison with the restricted master problems, since the prediction phase involves a relatively large number of subgradient optimization iterations.

Theorem 15.3 suggests an ergodic sequence computed within the subgradient optimization method (15.21) that asymptotically solves the problem (15.5), and hence also (15.48). Finitely generated ergodic solutions are, however, typically infeasible. In contrast, a Dantzig–Wolfe restricted master problem—finitely generated within the prediction phase—can be used to find feasible near optimal solutions to (15.5). The key difference is that in the former approach, the primal solution is defined by (15.54) with the convexity weights given a priori by (15.53), while in the latter approach the values of the convexity weights are optimized by the restricted master problem.

⁹In a degenerate dual optimum, nonbasic columns with zero reduced costs can also be obtained.

15.6 Recovery of Primal Integer Solutions by Means of Lagrangian Dual Methods

As an application of Theorem 15.3 we give a partial motivation for the success of Lagrangian heuristics in cases where the Lagrangian lower bound is strong, that is, when the difference $z^* - h^* \geq 0$ is small. We establish that the extent to which one may expect such heuristics to generate feasible solutions of high quality is governed by the same factors as those determining the quality of lower bounds. Hence, a solution strategy that yields a high quality Lagrangian lower bound $h^* = z_{\text{conv}}^* \leq z^*$ is also likely to yield a high quality upper bound $\bar{z} \geq z^*$.

15.6.1 Primal Integer Solutions From Lagrangian Heuristics

The basic idea behind Lagrangian heuristics is to use the information obtained from the Lagrangian dual problem (15.10) to construct feasible solutions to the original problem (15.2). A Lagrangian heuristic commonly works as follows. It is initiated at a solution that is feasible with respect to the non-relaxed constraints (15.2c). This solution is gradually adjusted through a finite number of steps that (1) use information from the Lagrangian dual problem, (2) retain feasibility in the non-relaxed constraints (15.2c), and (3) strive for feasibility in the relaxed constraints (15.2b). If the heuristic is successful, the final solution is feasible in (15.2). Lagrangian heuristics are, however, often not guaranteed to find feasible solutions.

To comply with the required initial feasibility in the non-relaxed constraints (15.2c), Lagrangian heuristics are commonly initiated with (near) optimal solutions to the subproblem in (15.8). Appropriate adjustments of solutions are necessarily problem specific and range from simple roundings, via elaborate manipulations of solutions, to solving a mixed-integer linear optimization problem. The information used from the Lagrangian dual problem is typically a near optimal dual solution, obtained by, for example, subgradient optimization. The adjustments made when striving for feasibility in the relaxed constraints (15.2b) are often guided by a merit function defined by original costs or Lagrangian cost, such that the final solution, if feasible, is likely also near optimal. The heuristic may also continue with a local search or meta-heuristic search in the original problem, after feasibility is reached.

We distinguish between two types of Lagrangian heuristics: *conservative* and *radical*. The latter type allows the solution finally found to be far from optimal in the subproblem. A radical heuristic can, for example, solve a *restriction* of the original problem (e.g., a Benders subproblem), which yields a feasible solution.

In conservative heuristics, which are the more common, the initial solution is (near) optimal in the subproblem, the adjustments made are local and such that near optimality in the subproblem is retained, and the number of adjustments is required to be rather small. Due to these limitations, such a heuristic may produce very good

feasible solutions, or frequently fail to find feasible solutions, depending on the characteristics of the original problem and the details of the heuristic.

Initialize a conservative Lagrangian heuristic by a solution $\tilde{\mathbf{x}}$ to the subproblem (15.9). Then, the effect of the adjustments made can be characterized by the inclusion

$$\mathbf{x}_{\text{heur}}(\tilde{\mathbf{x}}) \in \{ \mathbf{x} \in X \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}; \text{ “the distance } \|\mathbf{x} - \tilde{\mathbf{x}}\| \text{ is small”} \}. \quad (15.56)$$

To understand (at least partially) why conservative heuristics can produce very good feasible solutions, we first recall that $\tilde{\mathbf{x}}$ is *optimal* in a relaxation of the original problem. Further, in many applications a subproblem solution obtained using a near optimal Lagrangian dual solution is near feasible in the original problem. Moreover, if $\tilde{\mathbf{x}}$ is near feasible, the adjustments needed to reach feasibility are small, and a feasible solution $\mathbf{x}_{\text{heur}}(\tilde{\mathbf{x}})$, will indeed be close to the input $\tilde{\mathbf{x}}$. Hence, it appears likely that $\mathbf{x}_{\text{heur}}(\tilde{\mathbf{x}})$ can be near optimal in the original problem. The risk of failure in a conservative heuristic is due to the adjustments being only local, which may be insufficient to repair intricate infeasibilities.

As the dual sequence $\{\mathbf{u}^t\}$ approaches the set U^* , the ergodic sequence $\{\tilde{\mathbf{x}}^t\}$ will approach the set X_{conv}^* . If the Lagrangian lower bound h^* is strong, that is, if the duality gap $z^* - h^* \geq 0$ is small, the sets X_{conv}^* and X^* are expected to be close, in the sense that the value $\min\{\|\mathbf{x} - \mathbf{y}\| \mid \mathbf{x} \in X_{\text{conv}}^*, \mathbf{y} \in X^*\}$ is (relatively) small. Since the adjustments made to the ergodic primal iterates are quite small, a heuristic based on the inclusion (15.56) can thus be expected to produce solutions which are close to the set X^* . If, however, as in classical Lagrangian heuristics, the subproblem solutions $\mathbf{x}(\mathbf{u}^t)$ are used as inputs to the heuristic, the above arguments cannot be used to claim that the resulting point will be close to X_{conv}^* or to X^* . We thus propose to use the ergodic iterate $\tilde{\mathbf{x}}^t$ in place of the subproblem solution $\mathbf{x}(\mathbf{u}^t)$, as input to the Lagrangian heuristic of Algorithm 15.1; it is based on (15.56), that is, to find $\mathbf{x}_{\text{heur}}(\tilde{\mathbf{x}}^t)$.

Algorithm 15.1: Lagrangian heuristic for (15.2) utilizing ergodic sequences of subproblem solutions

Data: $\mathbf{u}^0 \in \mathbb{R}_+^m$; $t := 0$;

Result: a (heuristic) solution \mathbf{x}_{heur} to (15.2);

repeat

 Compute \mathbf{u}^{t+1} and $\mathbf{x}(\mathbf{u}^{t+1})$ according to (15.21), (15.23), (15.27);

$t := t + 1$;

 Update $\tilde{\mathbf{x}}^{t+1}$ according to (15.35) and (15.36) [or (15.38)];

 Compute a solution $\mathbf{x}_{\text{heur}}(\tilde{\mathbf{x}}^{t+1})$ according to (15.56);

until the best solution \mathbf{x}_{heur} found is satisfactory in (15.2);

15.6.2 Approximate Solutions to the Primal Problem via Core Problems

A common solution strategy for discrete optimization problems is to use some heuristic, problem adapted technique for predicting optimal values for a (relatively large) subset of the binary variables, and solve the (relatively small) restriction of the original problem to the nonfixed variables—the *core problem*—either exactly or approximately. The core problem should be constructed with the aim of making it feasible; an optimal solution to the core problem then yields a feasible and near optimal solution to the original problem. Whenever the core problem turns infeasible, previously fixed variables need to be relaxed and reinserted in the core.

Let $\mathcal{J}_0^* \subseteq \{1, \dots, n_b\}$ and $\mathcal{J}_1^* \subseteq \{1, \dots, n_b\} \setminus \mathcal{J}_0^*$ denote the sets of indices for the variables in \mathbf{x}_b which possess the value 0 and 1, respectively, in every optimal solution to the convexified problem (15.5). Further, let $\mathcal{J}_{\text{frac}}^* := \{1, \dots, n_b\} \setminus (\mathcal{J}_0^* \cup \mathcal{J}_1^*)$ denote the complementary index set, corresponding to the variables in \mathbf{x}_b which possess a fractional optimal value in *at least one* optimal solution to (15.5).

For each iteration t in the method (15.21), let $\tilde{\mathbf{x}}^t$ denote the weighted average of the solutions to the Lagrangian subproblem (15.8) as defined in (15.35), with step lengths according to (15.23), (15.27), and the convexity weights fulfilling the conditions (15.36) or (15.38). For each $j \in \{1, \dots, n_b\}$ the value $\tilde{x}_{b,j}^t \in [0, 1]$ can then be interpreted as the weighted relative frequency by which the variable $x_{b,j}$ attains the value 1 in an optimal solution to the subproblem. The following result is immediate.

Proposition 15.10 (On the Weighted Relative Frequency of Binary Solutions)

It holds that $\{\tilde{x}_{b,j}^t\} \rightarrow 0$ for all $j \in \mathcal{J}_0^$ and $\{\tilde{x}_{b,j}^t\} \rightarrow 1$ for all $j \in \mathcal{J}_1^*$. If the sequence $\{\tilde{x}_{b,j}^t\}$ accumulates at a point in the open interval $(0, 1)$, then $j \in \mathcal{J}_{\text{frac}}^*$.*

Proposition 15.10 motivates the use of the weighted relative frequency of the binary subproblem solutions as an indicator for the solution to the convexified optimization problem (15.5), as well as for the solution to the original problem (15.1) [or (15.2)]. For each variable $x_{b,j}$, $j = 1, \dots, n_b$, we thus define the two threshold values $\sigma_j^0, \sigma_j^1 \in (0, \frac{1}{2})$, which are used to define the two approximating sets

$$\mathcal{J}_0(\boldsymbol{\sigma}^0, \tilde{\mathbf{x}}^t) := \left\{ j \in \{1, \dots, n_b\} \mid \tilde{x}_{b,j}^t \leq \sigma_j^0 \right\}$$

and

$$\mathcal{J}_1(\boldsymbol{\sigma}^1, \tilde{\mathbf{x}}^t) := \left\{ j \in \{1, \dots, n_b\} \mid \tilde{x}_{b,j}^t \geq 1 - \sigma_j^1 \right\}.$$

A sequence of core problems—to be utilized in Algorithm 15.2—is then defined by

$$z_{\text{core}}^*(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1, \tilde{\mathbf{x}}^t) := \min_{\mathbf{x}_b, \mathbf{x}_c} \mathbf{c}_b^\top \mathbf{x}_b + \mathbf{c}_c^\top \mathbf{x}_c \tag{15.57a}$$

$$\text{subject to } (\mathbf{x}_b^\top, \mathbf{x}_c^\top)^\top \in \{ \mathbf{x} \in X \mid A\mathbf{x} \geq \mathbf{b} \}, \tag{15.57b}$$

$$x_{b,j} = 0, \quad j \in \mathcal{J}_0(\boldsymbol{\sigma}^0, \tilde{\mathbf{x}}^t), \tag{15.57c}$$

$$x_{b,j} = 1, \quad j \in \mathcal{J}_1(\boldsymbol{\sigma}^1, \tilde{\mathbf{x}}^t). \tag{15.57d}$$

We let $\mathbf{x}_{\text{core}}^t$ denote a feasible (optimal or approximate) solution to the core problem (15.57), the value of which—whenever feasible—is an upper bound on the optimal value, that is, the inequalities $\mathbf{c}^\top \mathbf{x}_{\text{core}}^t \geq z_{\text{core}}^*(\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1, \tilde{\mathbf{x}}^t) \geq z^*$ hold. We define $z_{\text{core}}^t := \min_{s=0, \dots, t} \{ \mathbf{c}^\top \mathbf{x}_{\text{core}}^s \}$. Since lower bounds $h(\mathbf{u}^t) \leq h^*$ are given by the dual iterates \mathbf{u}^t , a termination criterion can be based on the differences

$$z_{\text{core}}^t - \bar{h}_t \geq z^* - h^* \geq 0, \tag{15.58}$$

where $\bar{h}_t := \max_{s=0, \dots, t} \{ h(\mathbf{u}^s) \}$, $t = 1, 2, \dots$

Algorithm 15.2: Approximate solution of (15.2) from a sequence of core problems

Data: $\tau \in \mathbb{Z}_+$; $\varepsilon > z^* - h^*$; $\boldsymbol{\sigma}^0, \boldsymbol{\sigma}^1 \in (0, \frac{1}{2})^{n_b}$; $\mathbf{u}^0 \in \mathbb{R}_+^m$; $t := 0$;

Result: a (approximate) solution \mathbf{x}_{core} to (15.2) [or (15.1)];

repeat

 Perform τ iterations of the method (15.21), (15.23), (15.27);

 Compute \bar{h}_τ as defined in (15.58);

 Compute $\tilde{\mathbf{x}}^\tau$ as defined in (15.35) and (15.36) [or (15.38)];

repeat

 Decrease the values of σ_j^0 and σ_j^1 , $j = 1, \dots, n_b$;

 Generate the sets $\mathcal{J}_0(\boldsymbol{\sigma}^0, \tilde{\mathbf{x}}^\tau)$ and $\mathcal{J}_1(\boldsymbol{\sigma}^1, \tilde{\mathbf{x}}^\tau)$;

until the core problem (15.57) is feasible;

 Compute a solution $\mathbf{x}_{\text{core}}^\tau$ (exact or approximate) to (15.57);

 Update z_{core}^τ ;

 Increase the values of σ_j^0 and σ_j^1 , $j = 1, \dots, n_b$;

$\mathbf{u}^0 := \mathbf{u}^\tau$, $t := 0$;

until $z_{\text{core}}^\tau - \bar{h}_\tau \leq \varepsilon$, or the solution $\mathbf{x}_{\text{core}}^\tau$ is satisfactory in (15.2);

15.6.3 Optimal Solutions via a Branch-and-Bound Framework

We next consider using ergodic sequences to obtain feasible solutions to the problem (15.2) within a branch-and-bound framework. A subgradient method (15.21) can be

applied to the Lagrange dual (15.10) of the local linear program corresponding to each branch-and-bound tree node, yielding

- (1) a *lower bound* on z^* from a lower estimate of h^* (that is, from the value of an approximate solution to the Lagrangian dual of the node problem),
- (2) an *upper bound* on z^* from feasible solutions to (15.2) constructed using the subproblem solutions obtained in (15.21), and
- (3) a *branching decision* based on the (approximate) *reduced costs* obtained in the dual sequence $\{\mathbf{u}^t\}$.

One drawback of this principle is that it seldom provides a fractional primal solution, since the subproblem solutions are integer valued. So when aiming for a breadth-first branching, deciding on which variable to branch on is nontrivial. We propose a procedure in which the upper bounding (2) and the branching rule (3) are replaced by measures based on the ergodic iterates $\tilde{\mathbf{x}}^t$. The upper bound is obtained by applying Algorithm 15.1, which provides feasible solutions to the problem (15.2). The branching rule is based on the ergodic iterate $\tilde{\mathbf{x}}^t$ obtained from the method (15.21), (15.23), (15.27), (15.35) and (15.36) [or (15.38)]. Branching can be done on the variables with values close to binary or close to $\frac{1}{2}$. The optimization problem addressed in node n of the branch-and-bound tree is then the problem (15.5), with the additional constraints $x_{b,j} = \ell, j \in \mathcal{I}_\ell^n, \ell \in \{0, 1\}$, where the set \mathcal{I}_ℓ^n contains the indices of the variables that have been fixed to ℓ in the parent node of node n . By defining the set $X_{\text{conv}}^n := \text{conv}\{\mathbf{x} \in X \mid x_{b,j} = \ell, j \in \mathcal{I}_\ell^n, \ell \in \{0, 1\}\} \subseteq X_{\text{conv}}$, this linear program can then be expressed as

$$z_n^* := \min_{\mathbf{x}} \mathbf{c}^\top \mathbf{x} \tag{15.59a}$$

$$\text{subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b}, \tag{15.59b}$$

$$\mathbf{x} \in X_{\text{conv}}^n. \tag{15.59c}$$

The branching procedure of Algorithm 15.3 is then applied, where the dual starting point \mathbf{u}^0 in step 1 is often chosen as the final point \mathbf{u}^τ obtained from the subgradient scheme for the parent node. The search strategy for the branch-and-bound tree can be defined as, for example, depth-, breadth-, or best lower bound-first.

Algorithm 15.3: Branching decision based on ergodic primal iterates

Data: $\tau \in \mathbb{Z}_+; \mathbf{u}^0 \in \mathbb{R}_+^m; t := 0.$

Step 1. Apply τ iterations of Algorithm 15.1 to the linear optimization problem (15.59); this yields lower and upper bounds on z_n^* .

Step 2. Based on the lower and upper bounds on z^* , decide whether or not branching should be performed.

Step 3. Perform a branching based on the ergodic iterate $\tilde{\mathbf{x}}^\tau$: branch on a variable $x_{b,j}$ with $\tilde{x}_{b,j}^\tau$ close to $\frac{1}{2}$, or close to 0 or 1.

15.7 Notes and Further Reading

In this section we collect notes regarding the various results presented in the former sections and review important background material, along with references to the literature. While mentioning a few of the classic articles in the field of Lagrangian duality—chosen for their eloquence and pioneer status rather than for representing the state of the art—we also include a selection of more recent articles in the field—that we think tie in well with the classics.

Notes on Sect. 15.2.1: The Lagrangian Dual That the Lagrangian dual function h is piecewise affine and concave is shown in, for example, [11, Proposition 5.1.12]. That the dual optimal set U^* is nonempty and polyhedral when the feasible set of the convexified problem (15.5) is nonempty is verified in, for example, [11, Section 5]. When the feasible set of (15.5) is empty, however, the Lagrangian dual (15.10) is unbounded; conditions for the existence of an optimal Lagrangian dual set are detailed in [74].

In [31] Everett establishes a theorem for the Lagrangian function, which states that for a given vector of dual multipliers (not necessarily optimal), the minimization of the Lagrangian function over the primal variables yields a primal vector that is globally optimal for a primal problem whose available resources are identical to those utilized in the primal minimum of the Lagrangian function for the given multipliers. This result suggests that dual multipliers are near optimal when the minimum of the corresponding Lagrangian subproblem is near feasible. Brooks and Geoffrion extend in [15] the analysis in [31] to provide near optimal solutions also when the original problem may include, for example, integer variables.

In [46, 47] Held and Karp investigate approaches to the symmetric traveling salesperson problem, based on a 1-tree relaxation.¹⁰ One seeks prices (i.e., multipliers) on the nodes (which, however, appear on the adjacent links) such that the cheapest 1-tree equals an optimal Hamiltonian cycle. Optimal prices are derived from the solution of the corresponding Lagrangian dual problem, which is to maximize the value of the 1-tree. The lower bounds obtained from the dual procedure are also utilized in a branch-and-bound procedure.

Guignard provides in [39] an extensive introduction to Lagrangean approaches for the exact or approximate solution of difficult combinatorial optimization problems. The theme is similar to ours, while it is aimed at less experienced readers.

A quite little studied form of Lagrangian relaxation—not covered by this chapter—is known as *Lagrangian decomposition*, or variable splitting; see [40, 53]. It can be applied to mixed-integer linear optimization problems with (at least) two sets of explicit constraints [in (15.1b), (15.2b), or (15.7b)], such that two different Lagrangian relaxations are possible. The starting point for this approach is a problem reformulation in which copies of (a subset of) the primal variables are introduced, in one of the sets of constraints, together with additional constraints

¹⁰A 1-tree is the union of a spanning tree and one additional link.

that ensure consistency between the original variables and the copies. The consistency constraints are then Lagrangian relaxed. The corresponding Lagrangian dual problem can yield stronger bounds than both of the two possible straightforward Lagrangian relaxations of the original problem, provided that neither of the LP relaxations of the resulting Lagrangian subproblems possesses the integrality property.

Notes on Sect. 15.2.2: Optimality Conditions for the Convexified Problem

While some literature on nondifferentiable functions (naturally) draw a distinction between ‘supdifferentials of concave functions’ and ‘subdifferentials of convex functions’, we use the term subdifferential all through, relying on the readers’ familiarity with the concept.

The subdifferential of a concave function is defined in [11, Definition 3.2.3]. Proposition 15.1 follows from [11, Theorem 6.3.7], the convexity of the set X_{conv} , and [64, Theorem 11]. Further, the differentiability property of the dual function h relying on the singleton property of the set $X_{\text{conv}}(\mathbf{u})$ is depicted in [11, Theorem 6.3.3]. For a proof of Proposition 15.4, see [11, Theorem 3.4.3]. The saddle point optimality in (15.13) is proven in [11, Theorem 6.2.4], while Proposition 15.5 follows from [11, Theorem 6.2.5]. In [17, Lemma 2] it is shown that the composite mapping $\partial h \cap N_{\mathbb{R}_+^m}$ is constant on the solution set U^* . The definition of the subdifferential of an objective function is in [24, 25] generalized to take the feasible set into account; in our terminology it is called the *conditional subdifferential*. The *non-coordinability* phenomenon is a consequence of the non-differentiability of the Lagrangian dual function, which in turn is a consequence of the linearity of the primal problem. This phenomenon is of interest in economic systems which are described by linear models and where *constant prices* are used as a tool for steering decentralized decisions towards system optimality (e.g., [27]).

Turning to the question of how to exploit the Lagrangian dual problem (15.10) as a tool for finding an (near) optimal solution to the mixed-binary linear optimization problem (15.2), the fundamental observation is that the dual problem effectively solves the convexified primal problem (15.5), that is, finds an element of the solution set X_{conv}^* . Worth noting, however, is that whether such an element is directly available or not depends on the solution strategy employed for the dual problem.¹¹ If an (near) optimal, solution to the convexified problem (15.5) is at hand, a simple strategy to find a mixed-binary feasible solution is to employ *rounding*, that is, to systematically round the values of x_b in a solution to the convexified problem to binary values, and adjust the values of x_c accordingly, with the aim of reaching feasibility and near optimality in the original problem (15.2).

Simple rounding procedures are, however, in general inadequate for finding near optimal or even feasible solutions. A more general idea is *randomized rounding* (see

¹¹For example, a cutting-plane scheme for the dual problem identifies an element of X_{conv}^* while a subgradient optimization scheme does not. The key difference is that the former scheme provides, at termination, an optimal dual solution to the Lagrangian dual problem, that is, an optimal primal solution, while this is not the case for the latter scheme.

[78]), which for certain applications works as approximation algorithms. A generic such technique entails the solution of a continuous relaxation of the original problem, followed by a randomization scheme to decide whether to round up or down. The article [12] provides a framework for finding approximate solutions to covering problems through generic heuristics, all based on rounding (deterministic—using primal and dual information—or randomized—with nonlinear rounding functions) of an optimal solution to a LP relaxation. The roundings are applied to several known, as well as to new, results for the set covering, facility location, general covering, network design, and cut covering problems. The follow-up article [13] describes a class of structure exploiting methods to round fractional solutions, by introducing dependencies in the process. The technique improves approximation bounds for several problems, including the min- k -SAT problem.

The reader may note that rounding and randomized rounding strategies have in common with the core problem principle (see Sect. 15.6.2) that they identify and elaborate with a subset of the original variables, based on primal information (e.g., variable values) or dual information (e.g., Lagrangian reduced costs).

Notes on Sect. 15.2.3: Conditions for Optimality and Near Optimality of Mixed-Binary Linear Optimization Problems The *global primal–dual optimality conditions* (15.15) and the equivalent *near saddle point condition* (15.16) were introduced in [57], in which also computational results are reported. For examples of solution methods related to the enumeration principle discussed in Remark 15.2, see [45] and [21] (constrained shortest path problems), and [20] (train timetabling).

Notes on Sect. 15.3: Conditional Subgradient Optimization Subgradient optimization methods for minimizing non-differentiable convex functions originate in a work by Shor from 1962; [83] reviews the early history of nonsmooth optimization. For the case of unconstrained optimization, Ermol'ev established in [30] the convergence of the method using step lengths according to a *divergent series*. Polyak extended in [76, 77] the method to the case of constrained convex optimization and presented additional convergence results; see also [82, Section 2]. These methods have been frequently and often successfully applied, particularly in connection with Lagrangian duality; see, for example, [32, 33, 36, 48]. Worth noting is that subgradient optimization methods are closely related to *relaxation methods* for solving systems of linear inequalities; see [37].

The important *Polyak step length* rule, which has proved to be very useful in computational practice, was presented in [77]. For the case when the optimal objective value h^* is known a priori, convergence to an optimal solution by the method (15.21) using the step lengths (15.30) was established: the restrictions on the step length parameter θ_t guarantee that the distance between the current iterate and the solution set decreases in every iteration. Finite convergence to a near optimal point was also established for the case when the optimal value h^* in (15.30) is replaced by an estimate $\bar{h} \geq h^*$; this generalization, expressed in (15.31), is the most commonly used in practice.

Notes on Sect. 15.3.1: Basic Convergence in the Lagrangian Dual Problem The *conditional subgradient* optimization method was presented in [58] (see also [24, 25]) and generalizes the subgradient optimization method; it includes as a special case the *subgradient projection* method, as given in (15.22), which has shown a better practical performance than traditional subgradient methods (see [58]). The convergence of the method (15.21) was established in [58] for the divergent series and the Polyak step length rules. The proofs of Theorems 15.1 and 15.2 are special cases of those given in [58, Theorems 2.6 and 2.7, respectively]. The special case of subgradient projection is detailed in [58, Section 3]. Proposition 15.7 can be proven analogously as [77, Theorem 1] adapted to a Lagrangian dual problem, while Proposition 15.8 follows from [77, Theorem 4]. The condition (15.32) is referred to as the *almost complete relaxation* strategy; see [26, Section 3.4] and [58, Corollary 2.8].

Notes on Sect. 15.3.2: Ergodic Convergence in the Primal Problem The principle of constructing *ergodic sequences* of primal subproblem solutions in subgradient optimization can be traced back a long time; see [82, pp. 117] and [2]. The results presented here are developed in a series of articles; see [56, 59, 60]. Proposition 15.9 and Theorem 15.3 are special cases of [60, Proposition 5 and Theorem 1, respectively]. The convergence of sequences of convex combinations in general relies on a result in [54] and which is described in [44, Lemma 3].

A relative to the principle of constructing ergodic primal solutions, and also to the two-phase method presented in Sect. 15.5, is the *volume algorithm* of Barahona and Anbil in [8]. Referencing the classic works of Held and Karp [46, 47] and Held et al. [48] on the search for good lower bounds for computationally challenging *large-scale optimization* problems, Barahona and Anbil identify the drawback of not considering the convergence characteristics in the primal space and of the lack of a natural stopping criterion. While at the outset admitting the lack of a complete convergence theory, the authors describe a dual scheme, which is similar to a classic conjugate subgradient method (see [93]), in tandem with a constructive primal heuristic that mimics the master problem in Dantzig–Wolfe decomposition. The volume algorithm is applied to large-scale linear optimization problems arising from continuous relaxations of set partitioning, set covering, airline crew scheduling, max-cut, and facility location problems. The conclusion from the numerical experiments is that the more favourable problem instances are those in which variables are bounded within the interval $[0, 1]$, constraint coefficients lie in the set $\{0, 1, -1\}$, and the pricing problem is solvable in linear time.

Notes on Sect. 15.3.3: Enhanced Primal Ergodic Convergence Ergodic sequences that exploit more information from later subproblem solutions (as compared to earlier ones) were first presented in 1996 by Sherali and Choi [81] for the case of LP; this construction of ergodic sequences was generalized in 2015 by Gustavsson et al. [44] to incorporate also general convex optimization, as well as refined to the so-called s^k -rule. Theorem 15.4 follows from [44, Theorem 1]. The result that the requirements (15.38) on the convexity weights μ_s^t combined with step

lengths α_t according to a modified harmonic series (15.37) imply the requirements (15.36b)–(15.36e) on the parameters γ_s^t is established in [44, Proposition 5]. Convergence analyses for the s^k -rule are presented in [44], while the delayed initialization of the ergodic sequences according to Remark 15.8 is detailed in [60, Section 3, Remark 1].

Notes on Sect. 15.3.4: Finite Primal Feasibility and Finite ε -Optimality Theorem 15.5 and the finiteness result in Corollary 15.2 are established in [60, Theorem 3 and Corollary 3, respectively]. Approximate solution of the Lagrangian subproblems yields directions being ε -subgradients; the resulting *conditional ε -subgradient* method is investigated and analysed in [61]. Finiteness results are also presented in [63], in which the ergodic sequences are generated within a *simplicial decomposition* framework for nondifferentiable convex optimization.

A characterization in the case of a possibly inconsistent primal problem (15.5) is carefully detailed in [74], for the more general case of convex optimization. Convergence is established of an ergodic sequence of subproblem solutions to a point in the primal space such that the Euclidean norm of the infeasibility in the relaxed primal constraints is minimized, while the sequence of dual iterates diverges along the feasible direction of steepest ascent for the Lagrangian dual function.

Notes on Sect. 15.4: Dual Cutting-Planes: Dantzig–Wolfe Decomposition The article “Decomposition principle for linear programs” published in 1960 by Dantzig and Wolfe [23] is a pioneering work on the subject of large-scale optimization and among the most influential publications in the history of operations research. The Dantzig–Wolfe decomposition principle is founded on the *representation theorem for polyhedra* (e.g., [65, Chapter 3, Theorem 3]), which states that a point belongs to a polyhedron if and only if it can be expressed as a convex combination of the polyhedron’s extreme points plus a nonnegative linear combination of its extreme directions.

Dantzig–Wolfe decomposition is derived from the structure of the constraints of the linear program. It is assumed that the constraints can be partitioned into two sets, of which one is (relatively) computationally tractable. The other set of constraints is deemed complicating; hence these are Lagrangian relaxed—giving the method’s subproblem—and instead dealt with in a separate linear program—the restricted master problem. For practical problems Dantzig–Wolfe decomposition typically exploits a block diagonal structure of the subproblem, which can then be solved as several separate subproblems. Assuming that the subproblems always have finite optima, the optimal solutions obtained for different price vectors are—in the restricted master problem—convex combined, such that the complicating constraints are optimally utilized, with respect to the problem’s objective function.¹² Subject to the usual non-degeneracy assumptions (or means for dealing with degen-

¹²The case of unbounded subproblem solutions can also be handled, by finding feasible directions along which the objective value is unbounded and—in the restricted master problem—considering nonnegative linear combinations of such directions.

eracy), the convergence is finite; this follows directly from the finite convergence of the simplex method, of which Dantzig–Wolfe decomposition is, in essence, an application.

Plenty of variations of this computational scheme are possible. For example, each subproblem can be terminated at a near optimal solution, as long as this solution is good enough to ensure some progress in the restricted master problem. Further, the solution of the restricted master problem can be truncated, as long as some progress has been made. Further, convergence is ensured even if only the current basis is maintained in the restricted master problem, in a revised simplex manner.

The extension of the Dantzig–Wolfe decomposition principle to the case of nonconvex [e.g., (mixed-) integer] or non-linear optimization problems is known as *generalized LP*. In [68], a fundamental property of generalized LP is established. It is shown that for nearly all problems of practical importance, any limit point of the sequence of dual solutions produced by the algorithm is optimal in the Lagrangian dual problem of the given, primal, problem. This result holds even if the generalized LP algorithm does not solve the primal problem, which is typically the case whenever this problem is nonconvex.

Among the first applications of the Dantzig–Wolfe decomposition principle to a mixed-integer optimization problem is the work in [29] by Dzielinski and Gomory from 1965; see also [27, Section 7.2]. They consider a problem of production and inventory planning, which can be characterized as a time indexed, multi product economic lot size scheduling problem with common production resources. The problem is straightforward to model as a mixed-integer optimization problem, but they instead consider an approximate linear optimization model, which is based on the work [69] by Manne from 1958 and in which each variable corresponds to a complete production schedule for a single product. The approximate model can be regarded as a Dantzig–Wolfe master problem for a convexified version of a mixed-integer model of the problem. Since the number of possible production schedules can be huge, column generation is used. The column generation problem (i.e., the Dantzig–Wolfe subproblem) finds a production schedule for a single product and can be efficiently solved by the well-known Wagner–Whitin lot-sizing algorithm (which is a dynamic programming scheme), while the restricted master problem optimally combines the available production schedules with respect to overall cost and the common production resources.

In the pioneering two-part work [3, 4] on the application of column generation in the field of vehicle routing, Appelgren describes column generation approaches to a ship scheduling problem, obtained from a Swedish ship owning company. The first article applies the Dantzig–Wolfe decomposition method to the LP relaxation of the scheduling problem, which—probably thanks to the favourable matrix structure—achieves solutions that are near integer. In the second article, the column generation method is combined with a branch-and-bound algorithm, in which the branching is performed on one of the “essential” fractional variables and the bounds are obtained by the decomposition algorithm. This combined method was able to solve all problem instances tested, mostly with one branching only.

The work by Appelgren is an early predecessor to what is today known as *branch-and-price*. (The term price refers to the column generation, i.e., the pricing problem.) In the article [10] the authors first summarize the relations between *branch-and-cut* and branch-and-price for (mixed-) integer optimization problems. In branch-and-cut, classes of *valid inequalities*, preferably facets of the convex hull of the set of feasible solutions, are left out of the LP relaxation, because they comprise far too many constraints to generate and handle efficiently, and most of them will neither be binding in an optimal solution. Then, if an optimal solution to a LP relaxation is infeasible (with respect to integrality restrictions), a subproblem—called the separation problem—is solved in order to identify violated inequalities in a class. If one or more violated inequalities are found, some of them are added to the linear optimization problem, in order to cut off the infeasible solution, followed by its re-optimization. Branching occurs only when no violated inequalities can be found. Branch-and-cut is thus a generalization of branch-and-bound with LP relaxations, which allows cutting to be applied throughout the branch-and-bound tree.

The philosophy of branch-and-price resembles that of branch-and-cut, except that the focus is on column generation instead of row generation. In fact, pricing and cutting are complementary procedures for tightening a LP relaxation. In branch-and-price, sets of columns are left out of the LP relaxation because they are far too many to be generated and handled efficiently, and most of the associated variables will anyway be zero valued in an optimal solution. To check the optimality of a LP solution, a subproblem—the *pricing problem*, which is a separation problem for the dual linear program—is solved in order to identify columns to enter the basis, and the linear program is re-optimized. Branching occurs when no columns price out to enter the basis and the linear programming solution does not satisfy the integrality conditions. branch-and-price is a generalization of branch-and-bound with linear programming relaxations, which allows column generation to be applied throughout the branch-and-bound tree.

While appearing contradictory at first, there are several reasons (see [10]) for considering formulations with huge numbers of variables. Not infrequently a mixed-integer optimization formulation with many variables has a better LP relaxation (with respect to bound quality). Further, a compact formulation—which is a formulation *not* involving a huge number of variables—of a mixed-integer optimization problem may possess structural symmetries that allows solutions being mathematically different but having indifferent real-life interpretations; this causes branch-and-bound perform poorly as the problem barely changes after branching. A reformulation with a huge number of variables may eliminate such symmetries. Further, column generation provides a decomposition of the problem into a master problem and one or more subproblems. This decomposition may have a natural interpretation in the problem context, thus allowing for the incorporation of additional important constraints. Finally, a formulation with a huge number of variables may be the only choice.

At first glance, it may seem that branch-and-price involves nothing more than combining well-known ideas for solving linear programs by column generation and traditional branch-and-bound. This is, however, not that straightforward, as

observed already many years ago by Appelgren in [3, 4]. The most fundamental difficulty arising is that the traditional single variable branching is no longer viable, the reason being that it leads to *regeneration* of (already available) columns. Appelgren resolved this by finding a best column among those that are not already available.

Today, the common countermeasure is to use other branching techniques, which are compatible with column generation by allowing branching restrictions to be transferred to the column generation problem without leading to essential changes of its properties. Vanderbeck analyzes in [88] the challenges in combining a branching scheme with column generation. The article presents a generic branching scheme in which the pricing oracle of the root node remains of use after branching, and which does not require an extended formulation of the original problem. It then recursively partitions the subproblem solution set. Branching constraints are enforced in the pricing problem, which is solved approximately by a limited number of calls to the pricing oracle. The scheme is illustrated on the cutting stock and bin packing problems; it is the first branch-and-price algorithm capable of solving such problems to integrality without modifying the subproblem or expanding its variable space.

An early and tidy application of branch-and-price is to the generalized assignment problem [80], which is decomposed into a set partitioning master problem and knapsack column generation problems. Another tidy application of branch-and-price is given in [87], which considers a time-indexed (i.e., time discretized) formulation of a machine scheduling problem. Such formulations are known to provide strong LP bounds, but they tend to be extremely large. The authors show how to (partly) alleviate this difficulty by means of Dantzig–Wolfe decomposition, leading to a reformulation with many more variables, but far fewer constraints. The central pricing problem is solved by dynamic programming in $O(nT)$ time, with T and n being the number of time steps and jobs, respectively. To find an integer optimum, the decomposition approach is embedded in a branch-and-bound scheme.

For general surveys of column generation and branch-and-price, see [66, 92]. A recent research trend in mixed-integer optimization is to develop effective and efficient solution methods by combining decomposition approaches with heuristic or metaheuristic principles, in order to exploit their respective advantages. For a general overview of metaheuristic methods based on decomposition principles, see [79]. In a column generation context, such a combined method would extend a heuristic search beyond the columns necessary for solving the LP master problem. Classes of column generation based primal heuristics for mixed-integer linear optimization are reviewed in [51], with the aim to extract generic classes of column generation methods for use as black-box primal heuristics across applications. One such class consists of the so-called *diving heuristics*, which perform depth first searches in a branch-and-price tree, gradually obtaining integer solutions by variable fixings according to branchings which priorities columns that are part of LP optimal solutions in the nodes. To escape from local optima, partial backtracking can be used. Examples of applications of diving heuristics are found in, for example, [35, 41].

An interesting topic for further studies is to use an enumeration principle of the type outlined in Remark 15.2 with the aim to find favourable columns for inclusion in a restricted master problem. The goal is then to construct a restricted master problem capable of identifying an (near) optimal solution to the original problem (15.2), rather than to directly find an optimal solution through enumeration. This strategy is reasonable if the original problem has a Cartesian product structure—see Remarks 15.1 and 15.10—such that columns from different sets in the Cartesian product can be combined in order to achieve overall (near) optimality. Solution methods related to this strategy are found in [86] (a production planning problem) and [7] (a vehicle routing problem, not having a Cartesian product structure but a related suitable structure.)

Another interesting topic for further studies within the field of column generation is the little studied *combination of Lagrangian decomposition* (see [40, 53]) and *Dantzig–Wolfe decomposition* (i.e., dual cutting-planes) for solving the Lagrangian dual problem; see [72, 75, 94] for examples of this combination. Both Lagrangian decomposition and Dantzig–Wolfe decomposition can separately provide strong lower bounds, and the synergy between these two bounding principles has the potential to provide even stronger lower bounds.

Notes on Sect. 15.5: A Two-Phase Method: Subgradient Optimization and Dantzig–Wolfe Decomposition The inherent instability of dual cutting-plane procedures is discussed in, for example, [49, Chapter 15]. The fundamental *dual box step stabilization* was introduced in [70]. Examples of applications of this type of stabilized dual cutting-plane method (i.e., stabilized Dantzig–Wolfe decomposition) are found in [62, 90]. More general stabilization techniques are given in [28]. The use of heuristically generated high quality starting columns in Dantzig–Wolfe decomposition is discussed in, for example, [66, Subsection 4.1.1].

The two-phase method was introduced in [91], which also reports successful computational experience from an application to large-scale multicommodity network flows. Some similar methods can be found in the literature; in contrast to the one presented, those methods are, however, not justified by theoretical results.

In [95], subgradient optimization is performed in a first phase, which—in each iteration—stores dual cuts and solves a Dantzig–Wolfe master problem; the objective value of the latter is used in the Polyak step length formula. If the same cut is found too many times, the method switches to a second phase: the Dantzig–Wolfe method. A drawback of this two-phase method is that a computationally demanding master problem is solved in each subgradient iteration of the first phase, although it is only the objective value of the master problem that is actually used. This two-phase method is further developed in [89], which studies different criteria for switching to the second phase, in which the effect of using a bundle method is also studied. Also [55] studies a combined subgradient optimization and bundle method.

In [9] a fixed number of subgradient iterations is run every few iterations of the Dantzig–Wolfe method, starting from the restricted master dual optimum. The columns found are then included in the master problem. Substantially shorter

computing times are reported, as compared to the standard Dantzig–Wolfe method. This line of research is continued in [50], which employs subgradient optimization also for finding approximate dual solutions to Dantzig–Wolfe master problems.

Notes on Sect. 15.6: Recovery of Primal Integer Solutions by Means of Lagrangian Dual Methods Besides our own stream of research, the recovery of primal solutions has been treated in, for example, [67, 81, 84].

Notes on Sect. 15.6.1: Primal Integer Solutions from Lagrangian Heuristics A classic reference on Lagrangian relaxation and *Lagrangian heuristics* in integer optimization is [32] by Fisher from 1981. In [57] the characteristics of the Lagrangian heuristic principle is described more formally and such heuristics are classified as *conservative* or *radical*, depending on their nature. The essential difference is that in a conservative Lagrangian heuristic, the goal is to keep the values of $\varepsilon(\mathbf{x}, \mathbf{u})$ and $\delta(\mathbf{x}, \mathbf{u})$ [defined in (15.19)] small, while in a radical heuristic they may be large. A conservative heuristic typically starts at $\mathbf{x}(\mathbf{u}) \in X(\mathbf{u})$ and makes small changes, while a radical often involves solving an auxiliary optimization problem over a subset of the original problem variables, while keeping the values of the other variables fixed. For examples of conservative heuristics for specific applications, see [18, 34, 38, 52]. Examples of radical heuristics are found in [16, 22, 73, 85]; all of these exploit auxiliary optimization problems, some of which, however, being trivially solved.

The presented Lagrangian heuristic methodology utilizing ergodic sequences is developed in [1, 43].

Notes on Sect. 15.6.2: Approximate Solutions to the Primal Problem via Core Problems The use of *core problems* was introduced by Balas and Zemel [6] in 1981, then applied to binary knapsack problems; a collection of improvements of their scheme is found in [71].

The core problem principle has also been applied to very large-scale set covering models arising in crew scheduling [19, 22]; the construction of the core problem is there based on near optimal LP reduced costs found by Lagrangian relaxation and subgradient optimization. Further applications of core problems include capacitated facility location [5] and fixed charge transportation [94].

Our procedure for constructing core problems using ergodic sequences of Lagrangian subproblem solutions is developed in [42, 43].

Notes on Sect. 15.6.3: Optimal Solutions via a Branch-and-Bound Framework The use of dual subgradient methods and Lagrangian heuristics as a means for obtaining branching rules and bounds in branch-and-bound schemes is well studied (e.g., [14, 32]). The utilization of ergodic sequences to guide the branching in a branch-and-bound framework is developed in [1, 43].

Acknowledgements This chapter relies heavily on research and earlier publications by the authors and their collaborators. The most important of these publications are—in chronological order— [56], [58], [60], [57], [44], [43], [91], and [1].

References

1. Aldenvik, P., Schierscher, M.: Recovery of primal solutions from dual subgradient methods for mixed binary linear programming; a branch-and-bound approach. Master's Thesis, University of Gothenburg, Göteborg (2015)
2. Anstreicher, K.M., Wolsey, L.A.: Two "well-known" properties of subgradient optimization. *Math. Program.* **120**(1), 213–220 (2009)
3. Appelgren, L.H.: A column generation algorithm for a ship scheduling problem. *Transp. Sci.* **3**(1), 53–68 (1969)
4. Appelgren, L.H.: Integer programming methods for a vessel scheduling problem. *Transp. Sci.* **5**(1), 64–78 (1971)
5. Avella, P., Boccia, M., Sforza, A., Vasil'ev, I.: An effective heuristic for large-scale capacitated facility location problems. *J. Heuristics* **15**(6), 597–615 (2009)
6. Balas, E., Zemel, E.: An algorithm for large zero-one knapsack problems. *Oper. Res.* **28**(5), 1130–1154 (1980)
7. Baldacci, R., Christofides, N., Mingozzi, A.: An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. *Math. Program.* **115**(2), 351–385 (2008)
8. Barahona, F., Anbil, R.: The volume algorithm: producing primal solutions with a subgradient method. *Math. Program.* **87**(3), 385–399 (2000)
9. Barahona, F., Jensen, D.: Plant location with minimum inventory. *Math. Program.* **83**(1–3), 101–111 (1998)
10. Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W.P., Vance, P.H.: Branch-and-price: column generation for huge integer programs. *Oper. Res.* **46**(3), 316–329 (1998)
11. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: *Nonlinear Programming: Theory and Algorithms*, 2nd edn. Wiley, New York (2006)
12. Bertsimas, D., Vohra, R.: Rounding algorithms for covering problems. *Math. Program.* **80**, 63–89 (1998)
13. Bertsimas, D., Teo, C., Vohra, R.: On dependent randomized rounding algorithms. *Oper. Res. Lett.* **24**(3), 105–114 (1999)
14. Borchers, B., Mitchell, J.E.: An improved branch and bound algorithm for mixed integer nonlinear programs. *Comput. Oper. Res.* **21**(4), 359–367 (1994)
15. Brooks, R., Geoffrion, A.: Finding Everett's Lagrange multipliers by linear programming. *Oper. Res.* **14**(6), 1149–1153 (1966)
16. Brown, G.G., Geoffrion, A.M., Bradley, G.H.: Production and sales planning with limited shared tooling at the key operation. *Manag. Sci.* **27**(3), 247–259 (1981)
17. Burke, J.V., Ferris, M.C.: Characterization of solution sets of convex programs. *Oper. Res. Lett.* **10**(1), 57–60 (1991)
18. Campbell, G.M., Mabert, V.A.: Cyclical schedules for capacitated lot sizing with dynamic demands. *Manag. Sci.* **37**(4), 409–427 (1991)
19. Caprara, A., Fischetti, M., Toth, P.: A heuristic method for the set covering problem. *Oper. Res.* **47**(5), 730–743 (1999)
20. Caprara, A., Fischetti, M., Toth, P.: Modeling and solving the train timetabling problem. *Oper. Res.* **50**(5), 851–861 (2002)
21. Carlyle, W.M., Royset, J.O., Wood, R.K.: Lagrangian relaxation and enumeration for solving constrained shortest-path problems. *Networks* **52**(4), 256–270 (2008)
22. Ceria, S., Nobili, P., Sassano, A.: A Lagrangian-based heuristic for large-scale set covering problems. *Math. Program.* **81**(2), 215–228 (1998)
23. Dantzig, G.B., Wolfe, P.: Decomposition principle for linear programs. *Oper. Res.* **8**(1), 101–111 (1960)
24. Dem'yanov, V.F., Shomesova, V.K.: Conditional subdifferentials of convex functions. *Soviet Math. Doklady* **19**, 1181–1185 (1978)

25. Dem'yanov, V.F., Shomesova, V.K.: Subdifferentials of functions on sets. *Cybernetics* **16**(1), 24–31 (1980)
26. Dem'yanov, V.F., Vasil'ev, L.V.: *Nondifferentiable Optimization*. Optimization Software, New York (1985)
27. Dirickx, Y.M.I., Jennergren, L.P.: *System Analysis by Multilevel Methods: with Applications to Economics and Management*. Wiley, Chichester (1979)
28. du Merle, O., Villeneuve, D., Desrosiers, J., Hansen, P.: Stabilized column generation. *Discret. Math.* **194**(1–3), 229–237 (1999)
29. Dzielinski, B.P., Gomory, R.E.: Optimal programming of lot sizes, inventory and labor allocations. *Manag. Sci.* **11**(9), 874–890 (1965)
30. Ermolev, Y.M.: Methods for solving nonlinear extremal problems. *Cybernetics* **2**(4), 1–14 (1966)
31. Everett, H.: Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Oper. Res.* **11**(3), 399–417 (1963)
32. Fisher, M.L.: The Lagrangian relaxation method for solving integer programming problems. *Manag. Sci.* **27**(1), 1–18 (1981)
33. Fisher, M.L.: An applications oriented guide to Lagrangian relaxation. *Interfaces* **15**(2), 10–21 (1985)
34. Fisher, M.L.: Optimal solution of vehicle-routing problems using minimum K -trees. *Oper. Res.* **42**(4), 626–642 (1994)
35. Gamache, M., Soumis, F., Marquis, G., Desrosiers, J.: A column generation approach for large-scale aircrew rostering problems. *Oper. Res.* **47**(2), 247–263 (1999)
36. Geoffrion, A.M.: Lagrangian relaxation for integer programming. In: M.L. Balinski (ed.) *Approaches to Integer Programming*. Mathematical Programming Study, vol. 2, pp. 82–114 Springer, Berlin (1974)
37. Goffin, J.L.: Nondifferentiable optimization and the relaxation method. In: Lemaréchal, C., Mifflin, R. (eds.) *Nonsmooth Optimization*. Proceedings of a IIASA Workshop 1977, pp. 31–49. Pergamon Press, Oxford (1978)
38. Graves, S.C.: Using Lagrangean techniques to solve hierarchical production planning problems. *Manag. Sci.* **28**(3), 260–275 (1982)
39. Guignard, M.: Lagrangean relaxation. *Top* **11**(2), 151–228 (2003)
40. Guignard, M., Kim, S.: Lagrangean decomposition: a model yielding stronger Lagrangean bounds. *Math. Program.* **39**(2), 215–228 (1987)
41. Günlük, O., Kimbrel, T., Ladanyi, L., Schieber, B., Sorkin, G.B.: Vehicle routing and staffing for sedan service. *Transp. Sci.* **40**(3), 313–326 (2006)
42. Gustavsson, E.: *Topics in convex and mixed binary linear optimization*. Ph.D. Thesis, University of Gothenburg, Göteborg (2015)
43. Gustavsson, E., Larsson, T., Patriksson, M., Strömberg, A.-B.: Recovery of primal solutions from dual subgradient methods for mixed binary linear programming (2015). Preprint. Chalmers University of Technology and University of Gothenburg
44. Gustavsson, E., Patriksson, M., Strömberg, A.-B.: Primal convergence from dual subgradient methods for convex optimization. *Math. Program.* **150**(2), 365–390 (2015)
45. Handler, G.Y., Zang, I.: A dual algorithm for the constrained shortest path problem. *Networks* **10**(4), 293–310 (1980)
46. Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees. *Oper. Res.* **18**(6), 1138–1162 (1970)
47. Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees: part II. *Math. Program.* **1**(1), 6–25 (1971)
48. Held, M., Wolfe, P., Crowder, H.P.: Validation of subgradient optimization. *Math. Program.* **6**(1), 62–88 (1974)
49. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. Springer, Berlin (1993)
50. Huisman, D., Jans, R., Peeters, M., Wagelmans, A.P.M.: Combining column generation and Lagrangian relaxation. In: Desaulniers, G., Desrosiers, J., Solomon, M.M. (eds.) *Column Generation*, Chap. 9, pp. 247–270. Springer, New York (2005)

51. Joncour, C., Michel, S., Sadykov, R., Sverdllov, D., Vanderbeck, F.: Column generation based primal heuristics. *Electron Notes Discrete Math.* **36**, 695–702 (2010)
52. Jonsson, Ö., Larsson, T., Värbrand, P.: A Lagrangean relaxation scheme for a scheduling problem. *Asia-Pacific J. Oper. Res.* **7**, 155–162 (1990)
53. Jörnsten, K., Näsberg, M.: A new Lagrangian relaxation approach to the generalized assignment problem. *Eur. J. Oper. Res.* **27**(3), 313–323 (1986)
54. Knopp, K.: *Infinite Sequences and Series*. Dover Publications, New York (1956)
55. Kohl, N., Madsen, O.B.G.: An optimization algorithm for the vehicle routing problem with time windows based on Lagrangian relaxation. *Oper. Res.* **45**(3), 395–406 (1997)
56. Larsson, T., Liu, Z.: A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows. *Optimization* **40**(3), 247–284 (1997)
57. Larsson, T., Patriksson, M.: Global optimality conditions for discrete and nonconvex optimization—with applications to Lagrangian heuristics and column generation. *Oper. Res.* **54**(3), 436–453 (2006)
58. Larsson, T., Patriksson, M., Strömberg, A.-B.: Conditional subgradient optimization—Theory and applications. *Eur. J. Oper. Res.* **88**(2), 382–403 (1996)
59. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic results and bounds on the optimal value in subgradient optimization. In: Kleinschmidt, P., Bachem, A., Derigs, U., Fischer, D., Leopold-Wildburger, U., Möhring, R. (eds.) *Operations Research Proceedings 1995*, pp. 30–35. Springer, Berlin (1996)
60. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic, primal convergence in dual subgradient schemes for convex programming. *Math. Program.* **86**(2), 283–312 (1999)
61. Larsson, T., Patriksson, M., Strömberg, A.-B.: On the convergence of conditional ε -subgradient methods for convex programs and convex–concave saddle-point problems. *Eur. J. Oper. Res.* **151**(3), 461–473 (2003)
62. Larsson, T., Patriksson, M., Rydergren, C.: A column generation procedure for the side constrained traffic equilibrium problem. *Transp. Res. B* **38**(1), 17–38 (2004)
63. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic convergence in subgradient optimization—with application to simplicial decomposition of convex programs. In: S. Reich, A.J. Zaslavski (eds.) *Optimization Theory and Related Topics. Contemporary Mathematics*, vol. 568, pp. 159–186 American Mathematical Society and Bar-Ilan University (2012)
64. Lasdon, L.S.: Duality and decomposition in mathematical programming. *IEEE Trans. Sys. Sci. Cybern.* **4**(2), 86–100 (1968)
65. Lasdon, L.S.: *Optimization Theory for Large Systems*. Dover Publications, Mineola (2002)
66. Lübbecke, M.E., Desrosiers, J.: Selected topics in column generation. *Oper. Res.* **53**(6), 1007–1023 (2005)
67. Ma, J.: Recovery of primal solution in dual subgradient schemes. Master’s Thesis, *Computation for Design and Optimization*, School of Engineering, Massachusetts Institute of Technology, Cambridge (2007)
68. Magnanti, T.L., Shapiro, J.F., Wagner, M.H.: Generalized linear programming solves the dual. *Manag. Sci.* **22**(11), 1195–1203 (1976)
69. Manne, A.S.: Programming of economic lot sizes. *Manag. Sci.* **4**(2), 115–135 (1958)
70. Marsten, R.E., Hogan, W.W., Blankenship, J.W.: The boxstep method for large-scale optimization. *Oper. Res.* **23**(3), 389–405 (1975)
71. Martello, S., Toth, P.: *Knapsack Problems: Algorithms and Computer Implementations*. Wiley, Chichester (1990)
72. Mingozzi, A., Roberti, R.: An exact algorithm for the fixed charge transportation problem based on matching source and sink patterns. *Transp. Sci.* **52**(2), 229–238 (2018)
73. Mulvey, J.M., Crowder, H.P.: Cluster analysis: an application of Lagrangian relaxation. *Manag. Sci.* **25**(4), 329–340 (1979)
74. Önnheim, M., Gustavsson, E., Strömberg, A.-B., Patriksson, M., Larsson, T.: Ergodic, primal convergence in dual subgradient schemes for convex programming, II: the case of inconsistent primal problems. *Math. Program.* **163**(1–2), 57–84 (2017)

75. Pimentel, C.M.O., Alvelos, F.P.e., Valério De Carvalho, J.M.: Comparing Dantzig–Wolfe decompositions and branch-and-price algorithms for the multi-item capacitated lotsizing problem. *Optim. Methods Softw.* **25**(2), 299–319 (2010)
76. Polyak, B.T.: A general method of solving extremum problems. *Soviet Math. Doklady* **8**(3), 593–597 (1967)
77. Polyak, B.T.: Minimization of unsmooth functionals. *USSR Comput. Math. Math. Phys.* **9**, 14–29 (1969)
78. Raghavan, P., Thompson, C.D.: Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica* **7**(4), 365–374 (1987)
79. Raidl, G.R.: Decomposition based hybrid metaheuristics. *Eur. J. Oper. Res.* **244**(1), 66–76 (2015)
80. Savelsbergh, M.: A branch-and-price algorithm for the generalized assignment problem. *Oper. Res.* **45**(6), 831–841 (1997)
81. Sherali, H.D., Choi, G.: Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs. *Oper. Res. Lett.* **19**(3), 105–113 (1996)
82. Shor, N.Z.: *Minimization Methods for Non-Differentiable Functions*. Springer, Berlin. Translated from the Russian by K.C. Kiwiel and A. Ruszczyński (1985)
83. Shor, N.Z.: The development of numerical methods for nonsmooth optimization in the USSR. In: Lenstra, J.K., Rinnoy Kan A.H.G., Schrijver, A. (eds.) *History of Mathematical Programming: A Collection of Personal Reminiscences*, pp. 135–139. North-Holland, Amsterdam (1991)
84. Simonetto, A., Jamali-Rad, H.: Primal recovery from consensus-based dual decomposition for distributed convex optimization. *J. Optim. Theory Appl.* **168**(1), 172–197 (2016)
85. Sridharan, R.: A Lagrangian heuristic for the capacitated plant location problem with single source constraints. *Eur. J. Oper. Res.* **66**(3), 305–312 (1993)
86. Sweeney, D.J., Murphy, R.A.: A method of decomposition for integer programs. *Oper. Res.* **27**(6), 1128–1141 (1979)
87. van den Akker, J.M., Hurkens, C.A.J., Savelsbergh, M.W.P.: Time-indexed formulations for machine scheduling problems: column generation. *INFORMS J. Comput.* **12**(2), 111–124 (2000)
88. Vanderbeck, F.: Branching in branch-and-price: a generic scheme. *Math. Program.* **130**(2), 249–294 (2011)
89. Vera, J.R., Weintraub, A., Koenig, M., Bravo, G., Guignard, M., Barahona, F.: A Lagrangian relaxation approach for a machinery location problem in forest harvesting. *Pesquisa Oper.* **23**(1), 111–128 (2003)
90. Westerlund, A., Göthe-Lundgren, M., Larsson, T.: A stabilized column generation scheme for the traveling salesman subtour problem. *Discret. Appl. Math.* **154**(15), 2212–2238 (2006)
91. Westerlund, A., Göthe-Lundgren, M., Larsson, T., Yuan, D.: Subgradient optimization prior to column generation—A means for predicting optimal columns. *Inter. J. Pure Appl. Math.* **103**(4), 797–818 (2015)
92. Wilhelm, W.E.: A technical review of column generation in integer programming. *Optim. Eng.* **2**(2), 159–200 (2001)
93. Wolfe, P.: A method of conjugate subgradients for minimizing nondifferentiable functions. In: M.L. Balinski, P. Wolfe (eds.) *Nondifferentiable Optimization*. Mathematical Programming Study, vol. 3, pp. 145–173 Springer, Berlin (1975)
94. Zhao, Y., Larsson, T., Rönnberg, E., Pardalos, P.M.: The fixed charge transportation problem: a strong formulation based on Lagrangian decomposition and column generation. *J. Glob. Optim.* **72**(3), 517–538 (2018)
95. Zhu, S.: *Hybrid methods for solving Lagrangean duals and applications in mixed integer programming*. Ph.D. Thesis, Operations and Information Management, University of Pennsylvania, Philadelphia (1995)