Ergodic, primal convergence in dual subgradient schemes for convex programming, II: the case of inconsistent primal problems

(article starts on next page)

CrossMark

Ergodic, primal convergence in dual subgradient schemes for convex programming, II: the case of inconsistent primal problems

Magnus Önnheim[1] · **Emil Gustavsson[1]** ·
Ann-Brith Strömberg[1] · **Michael Patriksson[1]** ·
Torbjörn Larsson[2]

Abstract Consider the utilization of a Lagrangian dual method which is convergent for consistent convex optimization problems. When it is used to solve an infeasible optimization problem, its inconsistency will then manifest itself through the divergence of the sequence of dual iterates. Will then the sequence of primal subproblem solutions still yield relevant information regarding the primal program? We answer this question in the affirmative for a convex program and an associated subgradient algorithm for its Lagrange dual. We show that the primal–dual pair of programs corresponding to an associated homogeneous dual function is in turn associated with a saddle-point problem, in which—in the inconsistent case—the primal part amounts to finding a solution in the primal space such that the Euclidean norm of the infeasibility in the relaxed constraints is minimized; the dual part amounts to identifying a feasible steepest ascent direction for the Lagrangian dual function. We present convergence results for a conditional ε-subgradient optimization algorithm applied to the Lagrangian dual problem, and the construction of an ergodic sequence of primal subproblem solutions; this composite algorithm yields convergence of the primal–dual sequence to the set of saddle-points of the associated homogeneous Lagrangian function; for linear programs, convergence to the subset in which the primal objective is at minimum is also achieved.

✉ Ann-Brith Strömberg
anstr@chalmers.se

[1] Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden

[2] Division of Optimization, Department of Mathematics, Linköping University, 581 83 Linköping, Sweden

Mathematics Subject Classification 90C25 · 90C30 · 90C47 · 90C46

1 Introduction and motivation

Lagrangian relaxation—together with a search in the Lagrangian dual space of multipliers—has a long history as a popular means to attack complex mathematical optimization problems. Lagrangian relaxation is especially popular in cases when an inherent problem structure is present, such that a suitable relaxation is much easier to solve than the original problem, and where the result from optimizing the multipliers is acceptable even if the final primal solution is only near-feasible; examples are found, among others, among economics and logistics applications where the relaxed constraints are associated with capacity or budget constraints. Lagrangian relaxation is also frequently applied in combinatorial optimization, as a starting phase or as a heuristic. In the history of mathematical optimization, several classical works are built on the use of Lagrangian relaxation; see, e.g., the work by Held and Karp [15,16] on the traveling salesperson problem. For text book coverage and tutorials on Lagrangian relaxation, see, e.g., [2,3,28,32] and [10,11,13,29], respectively.

The convergence theory of Lagrangian relaxation is quite well developed for the cases in which the original, primal, problem has an optimal solution, or at least exhibits feasible solutions, even in cases when strong duality fails to hold. For the case when strong duality holds, several techniques have been developed in order to "translate" a dual optimal solution to a primal optimal one; this translation is supported by a consistent primal–dual system of equations and inequalities, sometimes referred to as characterizations of "saddle-point optimality" (cf. [28, Sect. 1.3.3] and [2, Thm. 6.2.5]).

In linear programming, decomposition–coordination techniques, like Dantzig–Wolfe decomposition and its dual equivalent Benders decomposition, ensure the convergence to a primal–dual optimal solution. In convex programming, ascent methods for the Lagrange dual, such as (proximal) bundle methods, can be equipped with the construction of an additional, convergent sequence of primal points which are provided by the optimality certificate of the ascent direction-finding quadratic subproblems (e.g., [19,20,33]). When utilizing classical subgradient methods from the "Russian school" (e.g., [9,35,36,40])—in which one subgradient, calculated at the current dual iterate, is utilized as a search direction and combined with a pre-defined step length rule—a convergent sequence of primal vectors can also be constructed as a convex combination of primal subproblem solutions (see, e.g., [40, pp. 116–118] and [39] for linear programs, and [1,14,25–27] for general convex programs). In the case where strong duality does *not* hold, the "translation" from an optimal Lagrangian dual solution to a primal optimal solution is much more involved, since the primal and dual optimal solutions may then violate both Lagrangian optimality and any complementarity conditions (cf. [22]).

What is hitherto an unsufficiently explored question is to *what* the sequence of above-mentioned simple convex combinations of primal subproblem solutions converges—if at all—when the original primal problem is *inconsistent*, in which case the Slater constraint qualification (CQ) assumed in [25] cannot hold. The purpose of

this article is to investigate this issue for convex programming; for the special case of linear programming quite strong results are obtained.

2 Preliminaries and main result

Consider the problem to

$$\text{minimize} \quad f(\mathbf{x}), \tag{2.1a}$$

$$\text{subject to} \quad g_i(\mathbf{x}) \le 0, \quad i \in \mathcal{I}, \tag{2.1b}$$

$$\mathbf{x} \in X, \tag{2.1c}$$

where the set $X \subset \mathbb{R}^n$ is nonempty, convex and compact, $\mathcal{I} = \{1, \ldots, m\}$, and the functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^n \mapsto \mathbb{R}$, $i \in \mathcal{I}$, are convex and, thus, continuous; these properties are assumed to hold throughout the article. The notation $\mathbf{g}(\mathbf{x})$ is in the sequel used for the vector $[g_i(\mathbf{x})]_{i \in \mathcal{I}}$. Moreover, whenever f and g_i, $i \in \mathcal{I}$, are affine functions, and X is polyhedral, we denote the program (2.1) as a *linear program*. The corresponding *Lagrange function* $\mathcal{L}_f : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ with respect to the relaxation of the constraints (2.1b) is defined by $\mathcal{L}_f(\mathbf{x}, \mathbf{u}) := f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x})$, $(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m$. The *Lagrangian dual objective function* $\theta_f : \mathbb{R}^m \mapsto \mathbb{R}$ is the concave function defined by

$$\theta_f(\mathbf{u}) := \min_{\mathbf{x} \in X} \mathcal{L}_f(\mathbf{x}, \mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^m. \tag{2.2}$$

With no further assumptions on the properties of the program (2.1), the minimization problem defined in (2.2) can be solved in finite time only to ε-optimality ([2, Ch. 7]). For any *approximation error* $\varepsilon \ge 0$, an ε-optimal solution, $\mathbf{x}_f^\varepsilon(\mathbf{u})$, to the minimization problem in (2.2) at $\mathbf{u} \in \mathbb{R}^m$ is denoted an *ε-optimal Lagrangian subproblem solution*, and fulfils the inclusion

$$\mathbf{x}_f^\varepsilon(\mathbf{u}) \in X_f^\varepsilon(\mathbf{u}) := \left\{ \mathbf{x} \in X \mid f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) \le \theta_f(\mathbf{u}) + \varepsilon \right\}. \tag{2.3}$$

The *Lagrange dual* to the program (2.1) with respect to the relaxation of the constraints (2.1b) is the convex program to find

$$\underset{\mathbf{u} \in \mathbb{R}_+^m}{\text{supremum}} \; \theta_f(\mathbf{u}). \tag{2.4}$$

2.1 Primal and dual convergence in the case of consistency

We first recall some convergence results for dual subgradient methods for the case when the feasible set of the program (2.1) is *nonempty*, while assuming a *constraint qualification* (e.g., Slater CQ, which for the problem (2.1) is stated as $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) < \mathbf{0}^m\} \ne \emptyset$; see [25]). Denote the optimal objective value of the program (2.1) by $\theta_f^* > -\infty$, and its solution set by

$$X_f^* := \left\{ \mathbf{x} \in X \,\big|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m, \ f(\mathbf{x}) \leq \theta_f^* \right\}. \tag{2.5}$$

By the continuity of f and g_i, $i \in \mathcal{I}$, and the compactness of X, we have, according to [37, Thm. 30.4(g)], that the primal optimal objective value equals the value obtained when solving the Lagrangian dual program, i.e., that

$$\theta_f^* := \sup_{\mathbf{u} \in \mathbb{R}_+^m} \theta_f(\mathbf{u}). \tag{2.6}$$

We denote the solution set to the Lagrange dual as

$$U_f^* := \left\{ \mathbf{u} \in \mathbb{R}_+^m \,\big|\, \theta_f(\mathbf{u}) \geq \theta_f^* \right\}, \tag{2.7}$$

the nonemptiness of which can be assured by presuming, e.g., Slater CQ or that the program (2.1) is linearly constrained ([2, Sect. 5]).

With respect to a convex set $U \subseteq \mathbb{R}^m$ and an $\varepsilon \geq 0$, the *conditional ε-subdifferential* ([2, Thms. 3.2.5 and 6.3.7], [26, Sect. 2], [8,23]) of the concave function θ_f at $\mathbf{u} \in U$ is given by

$$\partial_\varepsilon^U \theta_f(\mathbf{u}) := \left\{ \boldsymbol{\gamma} \in \mathbb{R}^m \,\big|\, \theta_f(\mathbf{v}) \leq \theta_f(\mathbf{u}) + \boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{v} - \mathbf{u}) + \varepsilon, \ \mathbf{v} \in U \right\}. \tag{2.8}$$

This definition implies the inclusions $\partial_\varepsilon^{\mathbb{R}^m} \theta_f(\mathbf{u}) \subseteq \partial_\varepsilon^U \theta_f(\mathbf{u}) \subseteq \partial_{\varepsilon'}^U \theta_f(\mathbf{u})$ for all $\mathbf{u} \in U$ and $0 \leq \varepsilon \leq \varepsilon' < \infty$. Further, from (2.2), (2.3), and (2.8), follow the inclusion

$$\mathbf{g}(\mathbf{x}_f^\varepsilon(\mathbf{u})) \in \partial_\varepsilon^{\mathbb{R}^m} \theta_f(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^m, \quad \varepsilon \geq 0. \tag{2.9}$$

The *normal cone* of a convex set $U \subseteq \mathbb{R}^m$ at $\mathbf{u} \in U$ is defined by

$$N_U(\mathbf{u}) := \left\{ \boldsymbol{\eta} \in \mathbb{R}^m \,\big|\, \boldsymbol{\eta}^{\mathrm{T}}(\mathbf{v} - \mathbf{u}) \leq 0, \ \mathbf{v} \in U \right\}. \tag{2.10}$$

This definition implies the equivalences $N_{\mathbb{R}_+^m}(\mathbf{u}) = \left\{ \boldsymbol{\eta} \in \mathbb{R}_-^m \,\big|\, \boldsymbol{\eta}^{\mathrm{T}}\mathbf{u} = 0 \right\}$ for all $\mathbf{u} \in \mathbb{R}_+^m$, and $\partial_\varepsilon^U \theta_f(\mathbf{u}) = \partial_\varepsilon^{\mathbb{R}^m} \theta_f(\mathbf{u}) - N_U(\mathbf{u})$ for all $\mathbf{u} \in U$ and $\varepsilon \geq 0$. Hence, the inclusion $\mathbf{g}(\mathbf{x}_f^\varepsilon(\mathbf{u})) - \boldsymbol{\eta} \in \partial_\varepsilon^{\mathbb{R}_+^m} \theta_f(\mathbf{u})$ holds whenever $\boldsymbol{\eta} \leq \mathbf{0}^m$, $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{u} = 0$, $\mathbf{u} \geq \mathbf{0}^m$, and $\varepsilon \geq 0$.

We consider solving the Lagrangian dual program (2.6) by the *conditional ε-subgradient optimization* algorithm[1] ([26, Sect. 2.1]). It starts at some initial vector $\mathbf{u}^0 \in \mathbb{R}_+^m$ and computes iterates $\mathbf{u}^t$ according to

$$\mathbf{u}^{t+\frac{1}{2}} := \mathbf{u}^t + \alpha_t \left[\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t \right], \quad \mathbf{u}^{t+1} := \left[\mathbf{u}^{t+\frac{1}{2}} \right]_+, \quad t = 0, 1, \ldots, \tag{2.11}$$

where $[\,\cdot\,]_+$ denotes the Euclidean projection onto the nonnegative orthant, the sequence $\{\boldsymbol{\eta}^t\}$ obeys the inclusion $\boldsymbol{\eta}^t \in N_{\mathbb{R}_+^m}(\mathbf{u}^t)$ for all t, and $\alpha_t > 0$ is the step

---

[1] The conditional subgradient algorithm was analyzed in [23]; its special case defined by the projection of the subgradient step direction onto the tangent cone of the feasible set yields a bounded sequence $\{\boldsymbol{\eta}^t\}$ in (2.11).

length chosen and $\epsilon_t \geq 0$ denotes the approximation error at iteration t. To simplify the presentation, the *cumulative step lengths* Λ_t are defined by

$$\Lambda_t := \sum_{s=0}^{t-1} \alpha_s, \quad t = 1, 2, \ldots.$$

For any closed and convex set $S \subseteq \mathbb{R}^r$ and any point $\mathbf{x} \in \mathbb{R}^r$, where $r \geq 1$, the convex Euclidean distance function dist $: \mathbb{R}^r \mapsto \mathbb{R}$ and the Euclidean projection mapping proj $: \mathbb{R}^r \mapsto S$ are defined as

$$\text{dist}(\mathbf{x}; S) := \min_{\mathbf{y} \in S} \{\|\mathbf{y} - \mathbf{x}\|\} \quad \text{and} \quad \text{proj}(\mathbf{x}; S) := \arg\min_{\mathbf{y} \in S} \{\|\mathbf{y} - \mathbf{x}\|\}, \quad (2.12)$$

respectively, where $\| \cdot \|$ denotes the Euclidean norm. For a sequence $\{\mathbf{x}^t\} \subset \mathbb{R}^n$ and a vector $\mathbf{y} \in \mathbb{R}^n$, the notation $\mathbf{x}^t \to \mathbf{y}$ means that the sequence $\{\mathbf{x}^t\}$ converges to the point $\mathbf{y}$.

The following proposition specializes [26, Thm. 8] to the setting at hand.

Proposition 2.1 (Convergence to a dual optimal point) *Let the method* (2.11) *be applied to the program* (2.6), *the sequence* $\{\boldsymbol{\eta}^t\}$ *be bounded, and the sequences of step lengths,* $\{\alpha_t\}$, *and approximation errors,* $\{\epsilon_t\}$, *fulfil the conditions*

$$\alpha_t > 0, \quad t = 0, 1, \ldots; \quad \alpha_t \to 0 \quad and \quad \Lambda_t \to \infty \quad as \quad t \to \infty; \quad (2.13a)$$

$$\sum_{s=0}^{\infty} \alpha_s^2 < \infty; \quad (2.13b)$$

$$\epsilon_t \geq 0, \quad t = 0, 1, \ldots; \quad \epsilon_t \to 0 \quad as \quad t \to \infty; \quad \sum_{s=0}^{\infty} \alpha_s \epsilon_s < \infty. \quad (2.13c)$$

If $U_f^* \neq \emptyset$, *then* $\mathbf{u}^t \to \mathbf{u}^\infty \in U_f^*$.

Proof Assume that $U_f^* \neq \emptyset$. Then the primal problem (2.1) attains an optimal solution and the Lagrange function defined by $\mathcal{L}_f(\mathbf{x}, \mathbf{u}) := f(\mathbf{x}) + \mathbf{u}^{\mathsf{T}} \mathbf{g}(\mathbf{x})$ has a saddle-point over the set $X \times \mathbb{R}_+^m$ ([2, Thm. 6.2.5]); the result then follows from [26, Thm. 8]. $\square$

At each iteration of the method (2.11) an ϵ_t-optimal Lagrangian subproblem solution $\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)$ is computed; an *ergodic* (that is, averaged) sequence $\{\bar{\mathbf{x}}_f^t\}$ is then defined by

$$\bar{\mathbf{x}}_f^t := \frac{1}{\Lambda_t} \sum_{s=0}^{t-1} \alpha_s \mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s), \quad t = 1, 2, \ldots. \quad (2.14)$$

The following result is a special case of that in [26, Thm. 19].

Proposition 2.2 (Convergence to the primal optimal set) *Let the method* (2.11)*,* (2.13) *be applied to the program* (2.6)*, the sequence* $\{\boldsymbol{\eta}^t\}$ *be bounded, and the sequence* $\{\overline{\mathbf{x}}_f^t\}$ *be defined by* (2.14)*. If* $U_f^* \neq \emptyset$*, then it holds that*

$$\mathrm{dist}(\overline{\mathbf{x}}_f^t; X_f^*) \to 0 \quad as \quad t \to \infty.$$

Proof As in the proof of Proposition 2.1, the condition $U_f^* \neq \emptyset$ ensures the existence of a saddle-point of $\mathcal{L}_f$. The compactness assumptions (on the dual solution set U_f^*) in [26, Thm. 20] is here replaced by the conditions (2.13b)–(2.13c), under which the dual sequence $\{\mathbf{u}^t\}$ is bounded (see, e.g., the proof of [26, Thm. 8]), i.e., for all $t \geq 0$, $\|\mathbf{u}^t\| \leq M$ holds, where $M > \|\mathbf{u}^*\|$ for all $\mathbf{u}^* \in U_f^*$. By restricting the dual space to $\mathbb{R}_+^m \cap \{\mathbf{u} \,|\, \|\mathbf{u}\| \leq M\}$, the result in [26, Thm. 19] applies. □

Proposition 2.2 states that whenever the dual solution set U_f^* is nonempty, the sequence $\{\overline{\mathbf{x}}_f^t\}$ of primal iterates defined in (2.14) will converge towards the primal optimal set X_f^*, provided that the sequence $\{\alpha_t\}$ of step lengths and the sequence $\{\epsilon_t\}$ are chosen such that the assumptions (2.13) are fulfilled. For convergence results when more general constructions of the ergodic sequence (2.14) are employed, see [14].

2.2 Outline and main result

Section 2.1 considers the consistent case of the program (2.1) and presents convergence results for the primal and dual sequences ($\{\overline{\mathbf{x}}_f^t\}$ and $\{\mathbf{u}^t\}$, respectively) obtained when the method (2.11) is applied to its Lagrange dual. In the remainder of the article we will analyze the properties of these two sequences when the primal problem (2.1) is *inconsistent*, i.e., when $\{\mathbf{x} \in X \,|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$, in which case the Slater CQ cannot be assumed.

The remainder of the article is structured as follows. In Sect. 3 we show that, during the course of the iterative scheme (2.11) for solving the program (2.4), the sequence $\{\mathbf{u}^t\}$ of dual iterates diverges when employing step lengths (α_t) and approximation errors (ϵ_t) fulfilling (2.13). As the sequence diverges, i.e., as $\|\mathbf{u}^t\| \to \infty$, the term $f(\mathbf{x})$ of the Lagrange function $\mathcal{L}_f(\mathbf{x}, \mathbf{u}^t)$ loses significance in the definition (2.3) of the ϵ_t-optimal subproblem solution, $\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t) \in X_f^{\epsilon_t}(\mathbf{u}^t)$. In Sect. 4 we characterize the *homogeneous dual function*, which is the Lagrangian dual function obtained when $f \equiv 0$. We show that there is a primal–dual problem associated with the *homogeneous dual* in which the primal part amounts to finding the set X_0^* of points in X with minimum infeasibility with respect to the relaxed constraints, i.e.,

$$X_0^* := \arg\min_{\mathbf{x} \in X} \left\{ \left\| [\mathbf{g}(\mathbf{x})]_+ \right\| \right\}. \tag{2.15}$$

In Sect. 5 we show that a sequence of scaled dual iterates will in fact converge to the optimal set of the homogeneous dual problem.

Section 6.1 presents the corresponding primal convergence results, i.e., that the sequence of primal iterates $\{\overline{\mathbf{x}}_f^t\}$ converges to the set X_0^*. To simplify notation we

redefine the primal optimal set X_f^* (defined in 2.5) as the optimal set for the so–called *selection problem*, i.e.,

$$X_f^* := \arg\min_{\mathbf{x} \in X_0^*} \{ f(\mathbf{x}) \}. \tag{2.16}$$

Note that, when $\left\{ \mathbf{x} \in X \,\middle|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \right\} \neq \emptyset$, the equivalence $X_0^* = \left\{ \mathbf{x} \in X \,\middle|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \right\}$ holds, then implying that X_f^* equals the optimal set for the program (2.1). Here lies the main point of departure when differentiating the convex program (2.1) from its linear programming special case (i.e., when f and g_i, $i \in \mathcal{I}$, are affine functions and X is a polyhedral set), in which the selection problem (2.16) is a linear program (possessing Lagrange multipliers). For general convex programming, however, this may not be the case. The stronger convergence results achieved for the linear programming case are presented in Sect. 6.2.

Our analysis leads to the main contribution of this article, which is then formulated as the following generalization of Proposition 2.2 to hold also for inconsistent convex programs.

Theorem 2.3 (Main result) *Apply the method* (2.11), (2.13a), (2.13b) *to the program* (2.4), *let the sequence* $\{\boldsymbol{\eta}^t\}$ *be bounded, and let the sequence* $\{\bar{\mathbf{x}}_f^t\}$ *be defined by* (2.14).

(a) *Let X_0^* be defined by* (2.15). *If* (2.13c) *holds, then*

$$\text{dist}(\bar{\mathbf{x}}_f^t; X_0^*) \to 0 \quad as \quad t \to \infty.$$

(b) *Let X_f^* be defined by* (2.16) *and* $\{\epsilon_t\} = \{0\}$. *If the program* (2.1) *is linear, then*

$$\text{dist}(\bar{\mathbf{x}}_f^t; X_f^*) \to 0 \quad as \quad t \to \infty.$$

□

Within the context of mathematical optimization, studies of characterizations of inconsistent systems are of course as old as the history of theorems of the alternative and the associated theory of optimality in linear and nonlinear optimization.

An inconsistent system of linear inequalities is studied in [7], which establishes a primal-dual theory on a strictly convex quadratic *least–correction* problem in which the left-hand sides of the linear inequalities possess negative slacks, the sum of squares of which is minimized. The article [6] is related to ours, in that it studies the behaviour of an *augmented Lagrangian* algorithm applied to a convex *quadratic* optimization problem with an inconsistent system of linear inequalities. The algorithm converges— with a linear speed—to a primal vector that minimizes the original objective function over a set defined by *minimally shifted constraints* through negative slacks.

For optimization problems involving (twice) differentiable, possibly nonconvex functions the methods in [5,31] are able to detect infeasibility and find solutions in which the norm of the infeasibility is at minimum. Filter methods (see [12] and references therein, and—for the nondifferentiable case—[18,34]) employ *feasibility*

restoration steps to reduce the value of a constraint violation function. In [4] *dynamical steering of exact penalty methods* toward feasibility and optimality is reviewed and analysed. While these references consider infeasibility within traditional nonlinear programming (inspired) algorithms, our work is devoted to the study of corresponding issues within subgradient based methods applied to Lagrange duals.

As stated in Theorem 2.3(a), for the general convex case we can only establish convergence to the set of minimum infeasibility points, while for the case of linear programs we also show—in Theorem 2.3(b)—that all primal limit points minimize the objective over the set of minimum infeasiblity points.

Then, in Sect. 7 we make some further remarks and present an illustrative example. Finally, in Sect. 8, we draw conclusions and suggest further research.

3 Dual divergence in the case of inconsistency

Consider the inconsistent program (2.1) and its Lagrangian dual function θ_f defined in (2.2). We begin by establishing that the emptiness of the set $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\}$ implies the existence of a nonempty cone $C \subset \mathbb{R}_+^m$, such that the value of the function θ_f increases in every direction $\mathbf{v} \in C$. Consequently, for this case the Lagrangian dual solution set (defined in (2.7) for the consistent case) fulfils $U_f^* = \emptyset$, and $\theta_f^* = \infty$ holds.

Proposition 3.1 (A theorem of the alternative) *The set* $\left\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\right\}$ *is empty if and only if the cone* $C := \left\{\mathbf{w} \in \mathbb{R}_+^m \mid \min_{\mathbf{x} \in X}\{\mathbf{w}^T\mathbf{g}(\mathbf{x})\} > 0\right\}$ *is nonempty.*

Proof By convexity of X and $\mathbf{g}$, the set $\{(\mathbf{x}, \mathbf{z}) \in X \times \mathbb{R}^m \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{z}\}$ is convex. Hence, its projection defined by $K := \{\mathbf{z} \in \mathbb{R}^m \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{z}, \mathbf{x} \in X\} = \{\mathbf{g}(\mathbf{x}) + \mathbb{R}_+^m \mid \mathbf{x} \in X\}$ is also convex. Since $\mathbf{g}$ is continuous the set K is closed, and from its definition follows that $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$ if and only if K can be separated strictly from $\mathbf{0}^m$.

Assume that $C \neq \emptyset$ and let $\mathbf{w} \in C$. The inequality $\mathbf{w}^T\mathbf{g}(\mathbf{x}) > 0$ then holds for all $\mathbf{x} \in X$. Hence, for each $\mathbf{x} \in X$, $g_i(\mathbf{x}) > 0$ must hold for at least one $i \in \mathcal{I}$, implying that $\mathbf{0}^m \notin K$.

Assume then that $\mathbf{0}^m \notin K$. Then there exist a $\mathbf{w} \in \mathbb{R}^m$ and a $\delta \in \mathbb{R}$ such that $\mathbf{w}^T\mathbf{z} \geq \delta > 0 = \mathbf{w}^T\mathbf{0}^m$ holds for all $\mathbf{z} \in K$. By definition of the set K, letting $\mathbf{e}_i \in \mathbb{R}^m$ denote the ith unit vector, it follows that $\mathbf{g}(\mathbf{x}) + \mathbf{e}_i\gamma \in K$ for all $\mathbf{x} \in X$ and all $\gamma \in \mathbb{R}_+$. Hence, $\mathbf{w}^T\mathbf{g}(\mathbf{x}) + w_i\gamma > 0$ holds for all $\mathbf{x} \in X$ and $\gamma \in \mathbb{R}_+$. Letting $\gamma \to \infty$ yields that $w_i \geq 0$ for all $i \in \mathcal{I}$. Choosing $\gamma = 0$ yields that $\min_{\mathbf{x} \in X}\{\mathbf{w}^T\mathbf{g}(\mathbf{x})\} > 0$. It follows that $\mathbf{w} \in C \neq \emptyset$. The proposition follows. □

Proposition 3.2 (The cone of ascent directions of the dual function) *If* $\left\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\right\} = \emptyset$ *then* $\theta_f(\mathbf{u} + \mathbf{v}) > \theta_f(\mathbf{u})$ *holds for all* $\mathbf{u} \in \mathbb{R}^m$ *and all* $\mathbf{v} \in C$.

Proof The proposition follows by the definition (2.2) of the function θ_f, and the relations $\theta_f(\mathbf{u} + \mathbf{v}) = \min_{\mathbf{x} \in X}\{f(\mathbf{x}) + (\mathbf{u} + \mathbf{v})^T\mathbf{g}(\mathbf{x})\} \geq \min_{\mathbf{x} \in X}\{f(\mathbf{x}) + \mathbf{u}^T\mathbf{g}(\mathbf{x})\} + \min_{\mathbf{x} \in X}\{\mathbf{v}^T\mathbf{g}(\mathbf{x})\} > \theta_f(\mathbf{u})$, where the strict inequality follows from the definition of C in Proposition 3.1. □

We next utilize the fact that the cone C is independent of the objective function f to show that in the inconsistent case the sequence $\{\mathbf{u}^t\}$ diverges.

Proposition 3.3 (Divergence of the dual sequence) *Let the sequence $\{\mathbf{u}^t\}$ be generated by the method* (2.11)*,* (2.13a) *applied to the program* (2.4)*, the sequence $\{\boldsymbol{\eta}^t\}$ be bounded, and the sequence $\{\epsilon_t\} \subset \mathbb{R}_+$. If $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$ then $\|\mathbf{u}^t\| \to \infty$.*

Proof Let $\mathbf{w} \in C \neq \emptyset$ and define $\delta := \min_{\mathbf{x} \in X}\{\mathbf{w}^T\mathbf{g}(\mathbf{x})\} > 0$ and $\beta_t := \mathbf{w}^T\mathbf{u}^t$ for all t. Then

$$\beta_{t+1} = \mathbf{w}^T\big(\mathbf{u}^t + \alpha_t\big[\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t\big]\big)_+ \geq \mathbf{w}^T\big(\mathbf{u}^t + \alpha_t\big[\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t\big]\big)$$
$$\geq \beta_t + \alpha_t\mathbf{w}^T\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) \geq \beta_t + \alpha_t\delta,$$

where the first inequality holds since $\mathbf{w} \in \mathbb{R}_+^m$, the second since $\boldsymbol{\eta}^t \in \mathbb{R}_+^m$, and the third since $\mathbf{w}^T\mathbf{g}(\mathbf{x}) \geq \delta$ for all $\mathbf{x} \in X$. From (2.13a) then follows that $\beta_t \to \infty$, and hence $\|\mathbf{u}^t\| \to \infty$. □

4 A homogeneous dual and an associated saddle-point problem in the case of inconsistency

We next use the result of Proposition 3.3 to establish that for large dual variable values the dual objective function can be closely approximated by an associated *homogeneous dual* function. Associated with this homogeneous dual is a saddle-point problem, in which the primal part amounts to finding the points in the primal space having minimum total infeasibility in the relaxed constraints.

Consider the Lagrange function associated with the program (2.1), i.e.,

$$\mathcal{L}_f(\mathbf{x}, \mathbf{u}) = \|\mathbf{u}\|\left(\frac{f(\mathbf{x})}{\|\mathbf{u}\|} + \frac{\mathbf{u}^T\mathbf{g}(\mathbf{x})}{\|\mathbf{u}\|}\right), \quad (\mathbf{x}, \mathbf{u}) \in \mathbb{R}^n \times \{\mathbb{R}^m\backslash\{\mathbf{0}^m\}\}.$$

As the value of $\|\mathbf{u}\|$ increases, the term $f(\mathbf{x})$ in the computation of the subproblem solution $\mathbf{x}_f(\mathbf{u})$ in (2.3) loses significance. Hence, according to Proposition 3.3, for large values of t the method (2.11), (2.13a) will tackle an approximation of the *homogeneous dual* problem to maximize θ_0 over $\mathbb{R}_+^m$.

In what follows, unless otherwise stated, we will assume that $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$.

4.1 The homogeneous version of the Lagrange dual

Consider the problem to find an $\mathbf{x} \in X$ such that $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m$. To this feasibility problem we associate the *homogeneous Lagrangian dual problem* to find

$$\underset{\mathbf{u} \in \mathbb{R}_+^m}{\text{supremum}} \ \theta_0(\mathbf{u}), \tag{4.1}$$

where $\theta_0 : \mathbb{R}^m \mapsto \mathbb{R}$ is defined by (2.2), for $f \equiv 0$ (i.e., $\theta_0(\mathbf{u}) = \min_{\mathbf{x} \in X}\{\mathbf{u}^T\mathbf{g}(\mathbf{x})\}$). A corresponding (optimal) subproblem solution $\mathbf{x}_0^0(\mathbf{u})$ and the subdifferential $\partial_0^{\mathbb{R}^m}\theta_0(\mathbf{u})$ are analogously defined.

According to its definition, the function θ_0 is *superlinear* (e.g., [17, Proposition V:1.1.3]), meaning that its hypograph is a nonempty and convex cone in $\mathbb{R}^{m+1}$, and implying that $\theta_0(\delta\mathbf{u}) = \delta\theta_0(\mathbf{u})$ holds for all $(\delta, \mathbf{u}) \in \mathbb{R}_+ \times \mathbb{R}^m$. The definition of the directional derivative, $\theta_0'(\mathbf{u}; \mathbf{d})$, of θ_0 at $\mathbf{u}$ in the direction of $\mathbf{d}$ (e.g., [17, Rem. I:4.1.4]), then yields that $\theta_0'(\mathbf{0}^m; \mathbf{d}) = \theta_0(\mathbf{d})$ holds for all $\mathbf{d} \in \mathbb{R}^m$. The program (4.1) can thus be interpreted as the search for a steepest feasible ascent direction of θ_0. Such a search requires that the argument of θ_0 is restricted. Hence, we define

$$\theta_0^{V_0^*} := \max_{\mathbf{u} \in V} \left\{ \theta_0(\mathbf{u}) \right\} = \max_{\mathbf{d} \in V} \left\{ \theta_0'(\mathbf{0}^m; \mathbf{d}) \right\}, \quad \text{where} \quad V := \left\{ \mathbf{u} \in \mathbb{R}_+^m \mid \|\mathbf{u}\| \leq 1 \right\}. \tag{4.2}$$

Defining V using the unit ball is somewhat arbitrary. Owing to the homogeneity of θ_0, the unit ball—viewed as the convex hull of the projective space—is, however, a natural choice. As shown in Sect. 4.2, for this choice the dual mapping yields a *singleton* set.

4.2 An associated saddle-point problem

From the Definition (2.2) of the function θ_f and the Definition (4.2) of $\theta_0^{V_0^*}$ follow that

$$\theta_0^{V_0^*} = \max_{\mathbf{u} \in V} \left\{ \min_{\mathbf{x} \in X} \left\{ \mathbf{u}^\mathsf{T} \mathbf{g}(\mathbf{x}) \right\} \right\} = \min_{\mathbf{x} \in X} \left\{ \max_{\mathbf{u} \in V} \left\{ \mathbf{u}^\mathsf{T} \mathbf{g}(\mathbf{x}) \right\} \right\} \tag{4.3}$$

hold, since the Lagrange function, defined by $\mathcal{L}_0(\mathbf{x}, \mathbf{u}) = \mathbf{u}^\mathsf{T} \mathbf{g}(\mathbf{x})$, is convex with respect to $\mathbf{x}$, for $\mathbf{u} \in \mathbb{R}_+^m$, and linear with respect to $\mathbf{u}$, and since the sets X and $V \subset \mathbb{R}_+^m$ are convex and compact (see, e.g., [17, Thms. VII:4.2.5 and VII:4.3.1]).

Definition 4.1 *(The set of saddle-points of the homogeneous Lagrange function [17, Def. VII:4.1.1]) Let the mappings $X_0(\cdot) : V \mapsto 2^X$ and $V(\cdot) : X \mapsto 2^V$ be defined by*

$$X_0(\mathbf{v}) := \underset{\mathbf{x} \in X}{\arg\min} \, \mathcal{L}_0(\mathbf{x}, \mathbf{v}), \quad \mathbf{v} \in V, \quad and \quad V(\mathbf{x}) := \underset{\mathbf{v} \in V}{\arg\max} \, \mathcal{L}_0(\mathbf{x}, \mathbf{v}), \quad \mathbf{x} \in X,$$

where the homogeneous Lagrange function $\mathcal{L}_0 : X \times V \mapsto \mathbb{R}$ is defined as $\mathcal{L}_0(\mathbf{x}, \mathbf{v}) := \mathbf{v}^\mathsf{T} \mathbf{g}(\mathbf{x})$. A point $(\overline{\mathbf{x}}, \overline{\mathbf{v}}) \in X \times V$ is said to be a saddle-point *of the function $\mathcal{L}_0$ on $X \times V$ when the inclusions $\overline{\mathbf{x}} \in X_0(\overline{\mathbf{v}})$ and $\overline{\mathbf{v}} \in V(\overline{\mathbf{x}})$ hold. The set of saddle-points is denoted by $X_0^* \times V_0^*$.* □

By the definition of the set X_f^* in (2.5), for the case when the program (2.1) is *consistent*, $X_0^* = \{ \mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \} \neq \emptyset$ denotes its feasible set. For the *inconsistent* case, whenever $\mathbf{x} \in X$ it holds that $g_i(\mathbf{x}) > 0$ for at least one $i \in \mathcal{I}$, implying that $\left\| [\mathbf{g}(\mathbf{x})]_+ \right\| > 0$.

Lemma 4.2 (A homogeneous dual mapping) *If $\mathbf{x} \in X$ then the set $V(\mathbf{x})$ is a singleton, given by $V(\mathbf{x}) = \{ \|[\mathbf{g}(\mathbf{x})]_+\|^{-1} [\mathbf{g}(\mathbf{x})]_+ \}$.*

Proof Let $\mathbf{x} \in X$ and $\overline{\mathbf{v}} = \|[\mathbf{g}(\mathbf{x})]_+\|^{-1}[\mathbf{g}(\mathbf{x})]_+$. Then, for any $\mathbf{v} \in V$ it holds that

$$\mathbf{v}^{\mathrm{T}}\mathbf{g}(\mathbf{x}) \leq \mathbf{v}^{\mathrm{T}}[\mathbf{g}(\mathbf{x})]_+ \leq \|\mathbf{v}\| \cdot \|[\mathbf{g}(\mathbf{x})]_+\| \leq \|[\mathbf{g}(\mathbf{x})]_+\| = \overline{\mathbf{v}}^{\mathrm{T}}\mathbf{g}(\mathbf{x}), \qquad (4.4)$$

where the first inequality holds since $\mathbf{v} \geq \mathbf{0}^m$ and $\mathbf{g}(\mathbf{x}) \leq [\mathbf{g}(\mathbf{x})]_+$, the second follows from the Cauchy–Schwartz inequality, and the third holds since $\|\mathbf{v}\| \leq 1$. Since $\mathbf{x} \in X$ is arbitrary and $\overline{\mathbf{v}} \in V$, it follows that $\overline{\mathbf{v}} \in V(\mathbf{x})$. That the set $V(\mathbf{x})$ is a singleton follows from the fact that equality holds in each of the inequalities in (4.4) only when both $\|\mathbf{v}\| = 1$ holds and the vectors $\mathbf{v}$ and $[\mathbf{g}(\mathbf{x})]_+$ are parallel, in which case $\mathbf{v} = \overline{\mathbf{v}}$. The lemma follows. □

Since for all $\mathbf{x} \in X$ and $\{\overline{\mathbf{v}}\} = V(\mathbf{x})$ the equality $\overline{\mathbf{v}}^{\mathrm{T}}\mathbf{g}(\mathbf{x}) = \|[\mathbf{g}(\mathbf{x})]_+\|$ holds, the right-hand side of (4.3) may be interpreted as the *minimal total deviation from feasibility* in the constraints $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m$ over $\mathbf{x} \in X$, that is,

$$\theta_0^{V_0^*} = \min_{\mathbf{x} \in X} \left\{ \|[\mathbf{g}(\mathbf{x})]_+\| \right\} > 0. \qquad (4.5)$$

The set $X_0^* \times V_0^*$ of saddle-points of $\mathcal{L}_0$ on $X \times V$ is thus given by

$$X_0^* := \arg\min_{\mathbf{x} \in X} \left\{ \|[\mathbf{g}(\mathbf{x})]_+\| \right\} \quad \text{and} \quad V_0^* := \arg\max_{\mathbf{v} \in V} \left\{ \min_{\mathbf{x} \in X} \{\mathbf{v}^{\mathrm{T}}\mathbf{g}(\mathbf{x})\} \right\}. \qquad (4.6)$$

Note that this definition of X_0^* agrees with (2.5) and is valid regardless of the consistency or inconsistency of the program (2.1). Since V_0^* is a singleton, we define the vector $\mathbf{v}^*$ by

$$\{\mathbf{v}^*\} := V_0^*. \qquad (4.7)$$

Note the equivalence $X_0^* = \{\mathbf{x} \in X_0(\mathbf{v}^*) \mid V(\mathbf{x}) = \{\mathbf{v}^*\}\}$.

Proposition 4.3 (Characterization of the set of saddle-points) *The following hold:*

(a) *The primal optimal set X_0^* is nonempty and compact.*
(b) *The dual optimal set $V_0^* = V(\mathbf{x}^*)$, irrespective of $\mathbf{x}^* \in X_0^*$.*
(c) *The dual optimal point fulfils $\mathbf{v}^* \in C$.*
(d) *The dual optimal point fulfils $\|\mathbf{v}^*\| = 1$.*
(e) *If the program (2.1) is linear then X_0^* is polyhedral.*

Proof The statements (a) and (e) follow by identifying $X_0^* \times \{\mathbf{z}^*\} = \arg\min_{(\mathbf{x},\mathbf{z}) \in X \times \mathbb{R}^m} \{\|\mathbf{z}\|^2 \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{z}\}$ ([2, Thm. 2.3.1]); by assumption, then $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$ holds, implying that $\mathbf{z}^* \neq \mathbf{0}^m$. By Definition 4.1, $V_0^* \subseteq V(\mathbf{x}^*)$ holds for all $\mathbf{x}^* \in X_0^*$, and by Lemma 4.2, $V(\mathbf{x})$ is a singleton for any $\mathbf{x} \in X$; hence (b) holds. The statement (c) follows from the definition of the sets C and V_0^* in Proposition 3.1 and (4.6), respectively, while (d) follows from Lemma 4.2, Proposition 4.3(b), and (4.7). □

5 Convergence to the homogeneous dual optimal set in the inconsistent case

We have characterized the Cartesian product set $X_0^* \times \{ \mathbf{v}^* \}$ of saddle-points of the homogeneous Lagrange function $\mathcal{L}_0$ over $X \times V$. Next, we will show that a sequence of simply scaled dual iterates obtained from the subgradient scheme converges to the point $\mathbf{v}^*$.

To simplify the notation in the analysis to follow, we let

$$L_t := \max_{s=0,\ldots,t} \{ \|\mathbf{u}^s\|, 1 \}, \quad \bar{\varepsilon} := \max_{\mathbf{x},\mathbf{y} \in X} \{ f(\mathbf{x}) - f(\mathbf{y}) \},$$

$$\text{and} \quad \varepsilon_t := \frac{\bar{\varepsilon} + \epsilon_t}{L_t}, \quad t = 0, 1, \ldots, \tag{5.1a}$$

where $\epsilon_t \geq 0$, $t = 0, 1, \ldots$. In tandem with the iterations of the conditional ϵ_t-subgradient algorithm (2.11) we construct the *scaled* dual iterate

$$\mathbf{v}^t := \frac{\mathbf{u}^t}{L_t}, \quad t = 0, 1, \ldots. \tag{5.1b}$$

We next show that the *conditional* (with respect to $\mathbb{R}_+^m$) ϵ_t-*subgradients* $\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t$, used in the algorithm (2.11), are also *conditional* (with respect to V) ε_t-*subgradients* of the homogeneous dual function θ_0 at the scaled iterate $\mathbf{v}^t$, with $\varepsilon_t = L_t^{-1}(\bar{\varepsilon} + \epsilon_t)$.

Lemma 5.1 (Conditional ε_t-subgradients of the homogeneous dual function) *Let the sequence $\{ \mathbf{u}^t \}$ be generated by the method (2.11), (2.13a) applied to the program (2.4), let the sequence $\{ \boldsymbol{\eta}^t \}$ be bounded, and the sequences $\{ \varepsilon_t \}$ and $\{ \mathbf{v}^t \}$ be defined by (5.1). Then,*

$$\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t \in \partial_{\varepsilon_t}^V \theta_0(\mathbf{v}^t), \quad t = 0, 1, \ldots.$$

Proof From the definitions (2.2) and (2.3) (for $\varepsilon = 0$) follow that the relations

$$\theta_f(\mathbf{u}^t) \leq f(\mathbf{x}_0^0(\mathbf{u}^t)) + (\mathbf{u}^t)^{\mathrm{T}} \mathbf{g}(\mathbf{x}_0^0(\mathbf{u}^t)) = f(\mathbf{x}_0^0(\mathbf{u}^t)) + \theta_0(\mathbf{u}^t), \quad t \geq 0,$$

and

$$\theta_0(\mathbf{u}) \leq \mathbf{u}^{\mathrm{T}} \mathbf{g}(\mathbf{x}_f^0(\mathbf{u})) = \theta_f(\mathbf{u}) - f(\mathbf{x}_f^0(\mathbf{u})), \quad \mathbf{u} \in \mathbb{R}_+^m,$$

hold. The combination of these relations with (2.8)–(2.9) (for $\varepsilon = \epsilon_t$), (2.10)–(2.11), and the definition of $\bar{\varepsilon}$ in (5.1a) yield that the inequalities

$$\theta_0(\mathbf{u}) - \theta_0(\mathbf{u}^t) \leq \theta_f(\mathbf{u}) - f(\mathbf{x}_f^0(\mathbf{u})) - \left[\theta_f(\mathbf{u}^t) - f(\mathbf{x}_0^0(\mathbf{u}^t)) \right]$$

$$\leq \left[\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) - \boldsymbol{\eta}^t \right]^{\mathrm{T}} (\mathbf{u} - \mathbf{u}^t) + \bar{\varepsilon} + \epsilon_t$$

hold for all $t \geq 0$ and all $\mathbf{u} \in \mathbb{R}^m_+$, implying that $\mathbf{g}(\mathbf{x}^{\epsilon_t}_f(\mathbf{u}^t)) - \boldsymbol{\eta}^t \in \partial^{\mathbb{R}^m_+}_{\bar{\varepsilon}+\epsilon_t} \theta_0(\mathbf{u}^t)$. By (2.8) and (5.1), the superlinearity of the function θ_0, and since $V \subset \mathbb{R}^m_+$, it holds that $\partial^{\mathbb{R}^m_+}_{\bar{\varepsilon}+\epsilon_t} \theta_0(\mathbf{u}^t) = \partial^{\mathbb{R}^m_+}_{\varepsilon_t} \theta_0(\mathbf{v}^t) \subseteq \partial^V_{\varepsilon_t} \theta_0(\mathbf{v}^t)$, and the result follows. $\qquad \square$

The following two lemmas are needed for the analysis to follow.

Lemma 5.2 (Normalized divergent series step lengths form a divergent series]) *Assume that* $\{\alpha_t\}^\infty_{t=1} \subset \mathbb{R}_+$. *If* $\Lambda_t \to \infty$ *as* $t \to \infty$, *then* $\left\{ \sum^t_{r=1}(1 + \Lambda_r)^{-1}\alpha_r \right\} \to \infty$ *as* $t \to \infty$.

Proof Since $\log(1 + d) \leq d$ whenever $d > -1$, for any $r \geq 1$, the relations

$$\log(1+\Lambda_{r+1}) = \log(1+\Lambda_r) + \log\left[1 + (1+\Lambda_r)^{-1}\alpha_r\right] \leq \log(1+\Lambda_r) + (1+\Lambda_r)^{-1}\alpha_r$$

hold. Aggregating these inequalities for $r = 1, \dots, t$ then yields the inequality

$$\log(1 + \Lambda_{t+1}) \leq \log(1 + \alpha_0) + \sum^t_{r=1}\left[(1 + \Lambda_r)^{-1}\alpha_r\right],$$

and the lemma follows. $\qquad \square$

Lemma 5.3 (Projection onto V) *For any* $\mathbf{u} \in \mathbb{R}^m$ *the equalities* $\mathrm{proj}(\mathbf{u}; V) = \mathrm{proj}([\mathbf{u}]_+; V) = \left(\max\left\{1; \|[\mathbf{u}]_+\|\right\}\right)^{-1}[\mathbf{u}]_+$ *hold.*

Proof The result follows by applying the optimality conditions (e.g., [2, Thm. 4.2.13]) to the convex and differentiable optimization problems defined by the projection operator in (2.12) for $S = \mathbb{R}^m_+$ and $S = V$, respectively. $\qquad \square$

We now establish the convergence characteristics of the scaled dual sequence $\{\mathbf{v}^t\}$ defined in (5.1b) to the dual part of the set of saddle-points for $\mathcal{L}_0$.

Theorem 5.4 (Convergence of a scaled dual sequence) *Let the sequence* $\{\mathbf{u}^t\}$ *be generated by the method* (2.11), (2.13a), (2.13c) *applied to the program* (2.4), *let the sequence* $\{\boldsymbol{\eta}^t\}$ *be bounded, let the sequence* $\{\mathbf{v}^t\}$ *be defined by* (5.1b), *and let the optimal solution to the homogeneous dual,* $\mathbf{v}^*$, *be defined by* (4.7). *If* $\{\mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m\} = \emptyset$, *then it holds that* $\mathbf{v}^t \to \mathbf{v}^*$ *as* $t \to \infty$.

Proof Let $\boldsymbol{\gamma}^t := \mathbf{g}(\mathbf{x}^{\epsilon_t}_f(\mathbf{u}^t)) - \boldsymbol{\eta}^t$ for all $t \geq 0$. From the definition (2.11) and the triangle inequality it follows that

$$\|\mathbf{u}^t\| \leq \|\mathbf{u}^0\| + \sum^{t-1}_{r=0}\alpha_r\|\boldsymbol{\gamma}^r\|, \quad t \geq 1. \tag{5.2}$$

Since X is compact, $\mathbf{g}$ is continuous, and the sequence $\{\boldsymbol{\eta}^r\}$ is bounded, it holds that $\Gamma := L_0 + \sup_{r\geq 0}\left\{\|\boldsymbol{\gamma}^r\|\right\} < \infty$. From the definition (5.1a) of L_t then follows that

$$1 \leq L_t \leq \max_{s=1,\dots,t}\left\{\Gamma(1 + \Lambda_s)\right\} = \Gamma(1 + \Lambda_t), \quad t \geq 1,$$

which implies that $L_t^{-1}\alpha_t \geq \left[\Gamma(1+\Lambda_t)\right]^{-1}\alpha_t, t \geq 1$. It follows that $L_t^{-1}\alpha_t > 0$ for all t and, by Proposition 3.3, that $\left\{ L_t^{-1}\alpha_t \right\} \to 0$ as $t \to \infty$. Since, by Assumption (2.13a), $\Lambda_t \to \infty$ as $t \to \infty$ it follows from Lemma 5.2 that $\left\{ \sum_{s=0}^{t-1}(L_s^{-1}\alpha_s) \right\} \to \infty$ as $t \to \infty$. Consequently the sequence $\left\{ L_t^{-1}\alpha_t \right\}$ fulfils the conditions (2.13a).

From Lemma 5.3 follows that

$$\text{proj}(\mathbf{v}^t + L_t^{-1}\alpha_t\boldsymbol{\gamma}^{\,t}; V) = \text{proj}(L_t^{-1}[\mathbf{u}^t + \alpha_t\boldsymbol{\gamma}^{\,t}]_+; V) = \text{proj}(L_t^{-1}\mathbf{u}^{t+1}; V).$$

If $\|\mathbf{u}^{t+1}\| \leq L_t$, then $L_{t+1} = L_t$ and $\text{proj}(L_t^{-1}\mathbf{u}^{t+1}; V) = \text{proj}(\mathbf{v}^{t+1}; V) = \mathbf{v}^{t+1}$ hold. Otherwise, $\|\mathbf{u}^{t+1}\| = L_{t+1} > L_t$ and $\text{proj}(L_t^{-1}\mathbf{u}^{t+1}; V) = L_{t+1}^{-1}\mathbf{u}^{t+1} = \mathbf{v}^{t+1}$ hold. In both cases

$$\mathbf{v}^0 = L_0^{-1}\mathbf{u}^0 \quad \text{and} \quad \mathbf{v}^{t+\frac{1}{2}} = \mathbf{v}^t + L_t^{-1}\alpha_t\boldsymbol{\gamma}^{\,t}, \quad \mathbf{v}^{t+1} = \text{proj}(\mathbf{v}^{t+\frac{1}{2}}; V), \quad t = 0, 1, \ldots,$$

hold. Further, by Lemma 5.1 the inclusion $\boldsymbol{\gamma}^{\,t} \in \partial_{\varepsilon_t}^V \theta_0(\mathbf{v}^t)$ holds for $t \geq 1$. By (2.13c) the sequence $\{\epsilon_t\}$ is bounded; from Proposition 3.3 and (5.1a) it then follows that $\varepsilon_t \to 0$. The theorem then follows from [26, Thm. 3]. □

The main idea utilized in the proof of Theorem 5.4 is that the scaled sequence $\{\mathbf{v}^t\}$ obtained from the subgradient method defines a conditional (with respect to V) ε_t-subgradient algorithm, as applied to the homogeneous Lagrange dual (4.2). Hence, by tackling the Lagrange dual (2.4) by a subgradient method, we receive—in the case of an inconsistent primal problem—a solution to its homogeneous version (4.1).

Next follow two technical corollaries, to be used in the primal convergence analysis.

Corollary 5.5 (Convergence of a normalized dual sequence) *Under the assumptions of Theorem 5.4 it holds that $\left\{ \|\mathbf{u}^t\|^{-1}\mathbf{u}^t \right\} \to \mathbf{v}^*$ as $t \to \infty$.*

Proof By the superlinearity of the function θ_0, it holds that $\theta_0(\|\mathbf{u}^t\|^{-1}\mathbf{u}^t) \geq \theta_0(L_t^{-1}\mathbf{u}^t) = \theta_0(\mathbf{v}^t)$, $t \geq 0$. The corollary then follows since, by Theorem 5.4, $\theta_0(\mathbf{v}^t) \to \theta_0^{V_0^*}$. □

Corollary 5.6 (Convergence to the optimal value of the homogeneous dual) *Under the assumptions of Theorem 5.4 it holds that $\left\{ (\mathbf{v}^*)^{\mathrm{T}}\mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) \right\} \to \theta_0^{V_0^*}$ as $t \to \infty$.*

Proof For each $t \geq 0$ and each $\mathbf{x} \in X$, let $\rho_t(\mathbf{x}) := L_t^{-1}f(\mathbf{x}) + (\mathbf{v}^t - \mathbf{v}^*)^{\mathrm{T}}\mathbf{g}(\mathbf{x})$. Since $L_t \to \infty$, $\mathbf{v}^t \to \mathbf{v}^*$, and X is compact, it follows that $\rho_t(\mathbf{x}) \to 0$ for all $\mathbf{x} \in X$. Using the definition (2.2) and the equivalence (4.3), and by separating the minimization over $\mathbf{x} \in X$, it follows that

$$L_t^{-1}\theta_f(\mathbf{u}^t) = \min_{\mathbf{x}\in X} \left\{ (\mathbf{v}^*)^{\mathrm{T}}\mathbf{g}(\mathbf{x}) + \rho_t(\mathbf{x}) \right\} \geq \min_{\mathbf{x}\in X} \left\{ (\mathbf{v}^*)^{\mathrm{T}}\mathbf{g}(\mathbf{x}) \right\} + \min_{\mathbf{x}\in X}\{\rho_t(\mathbf{x})\}$$

$$= \theta_0^{V_0^*} + \min_{\mathbf{x}\in X}\{ \rho_t(\mathbf{x}) \}. \tag{5.3}$$

On the other hand, since $X_0(\mathbf{v}^*) \subseteq X$, and $(\mathbf{v}^*)^T \mathbf{g}(\mathbf{x}) = \theta_0^{V_0^*}$ for any $\mathbf{x} \in X_0(\mathbf{v}^*)$, we have that

$$L_t^{-1}\theta_f(\mathbf{u}^t) \leq \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ (\mathbf{v}^*)^T \mathbf{g}(\mathbf{x}) + \rho_t(\mathbf{x}) \right\} = \theta_0^{V_0^*} + \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \rho_t(\mathbf{x}) \right\}.$$

It follows that $\left\{ L_t^{-1}\theta_f(\mathbf{u}^t) \right\} \to \theta_0^{V_0^*}$ as $t \to \infty$. By the left-most equality in (5.3) and (2.3) the inequality $(\mathbf{v}^*)^T \mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) \leq L_t^{-1}\theta_f(\mathbf{u}^t) - \rho_t(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) + L_t^{-1}\epsilon_t$ holds. The corollary follows. □

6 Primal convergence in the case of inconsistency

We apply the conditional ε-subgradient scheme (2.11) to the Lagrange dual of the program (2.1). In each iteration we construct an ergodic primal iterate $\bar{\mathbf{x}}_f^t$, according to the scheme defined in (2.14). We here aim at analyzing the convergence of the ergodic sequence $\{\bar{\mathbf{x}}_f^t\}$ when the primal program (2.1) is inconsistent. In Sect. 6.1 we establish convergence of the ergodic sequence to the feasible set of the selection problem (2.16) for the case of convex programming [i.e., Theorem 2.3(a)]. In Sect. 6.2 we specialize this result to the case of linear programming, in which case the stronger result of convergence to *optimal* solutions to the selection problem (2.16) is obtained [i.e., Theorem 2.3(b)].

The set of indices of the strictly positive elements of the vector $\mathbf{v}$ is denoted by

$$\mathcal{I}_+(\mathbf{v}) := \left\{ i \in \mathcal{I} \mid v_i > 0 \right\} \subseteq \mathcal{I}, \quad \mathbf{v} \in V.$$

6.1 Convergence results for general convex programming

To simplify the notation, we let the ergodic sequence, $\{\bar{\boldsymbol{\gamma}}^t\}$, of conditional ε-subgradients be defined by

$$\bar{\boldsymbol{\gamma}}^t := \frac{1}{\Lambda_t} \sum_{s=0}^{t-1} \alpha_s \left[\mathbf{g}(\mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s)) - \boldsymbol{\eta}^s \right], \quad t = 1, 2, \ldots,$$

for some choices of step lengths $\alpha_s > 0$ and approximation errors $\epsilon_s \geq 0$, $s = 0, 1, \ldots, t - 1$. We will also need the following technical lemma (see [21, p. 35] for its proof).

Lemma 6.1 (A convergent sequence of convex combinations) *Let the sequences* $\{\mu_{ts}\} \subset \mathbb{R}_+$ *and* $\{\mathbf{e}^s\} \subset \mathbb{R}^r$, *where* $r \geq 1$, *satisfy the relations* $\sum_{s=0}^{t-1} \mu_{ts} = 1$ *for* $t = 1, 2, \ldots$, $\mu_{ts} \to 0$ *as* $t \to \infty$ *for* $s = 0, 1, \ldots$, *and* $\mathbf{e}^s \to \mathbf{e}$ *as* $s \to \infty$. *Then,* $\left\{ \sum_{s=0}^{t-1} \mu_{ts} \mathbf{e}^s \right\} \to \mathbf{e}$ *as* $t \to \infty$. □

We are now set up to establish the first part of the main result of this article.

Proof of Theorem 2.3(a) The case when $\{ \mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \} \neq \emptyset$ is treated in Proposition 2.2.

Consider the case when $\{ \mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \} = \emptyset$. We will show that the ergodic sequence $\{ \overline{\mathbf{x}}_f^t \}$ converges to the set $X_0^* = \arg\min_{\mathbf{x} \in X} \{ \| [\mathbf{g}(\mathbf{x})]_+ \| \}$.

Since X is convex and compact, any limit point $\overline{\mathbf{x}}_f^\infty$ of $\{ \overline{\mathbf{x}}_f^t \}$ fulfils $\overline{\mathbf{x}}_f^\infty \in X$. Then, by the continuity of $\mathbf{g}$ and $\| \cdot \|$, and the equivalence in (4.5), the relations

$$\theta_0^{V_0^*} = \min_{\mathbf{x} \in X} \left\{ \| [\mathbf{g}(\mathbf{x})]_+ \| \right\} \leq \| [\mathbf{g}(\overline{\mathbf{x}}_f^\infty)]_+ \| \leq \limsup_{t \to \infty} \left\{ \| [\mathbf{g}(\overline{\mathbf{x}}_f^t)]_+ \| \right\}$$

hold. From Lemma 4.2 and the definitions (4.6)–(4.7) follow that the vectors $[\mathbf{g}(\mathbf{x})]_+$ and $\mathbf{v}^*$ are parallel, and that $\| [\mathbf{g}(\mathbf{x})]_+ \| = \theta_0^{V_0^*}$ hold for all $\mathbf{x} \in X_0^*$. Further, by Proposition 4.3(d), $\| \mathbf{v}^* \| = 1$ holds. Hence, it suffices to show that

$$\limsup_{t \to \infty} \left\{ [\mathbf{g}(\overline{\mathbf{x}}_f^t)]_+ \right\} \leq \theta_0^{V_0^*} \mathbf{v}^*, \tag{6.1}$$

which will imply the equivalence $\| [\mathbf{g}(\overline{\mathbf{x}}_f^\infty)]_+ \| = \theta_0^{V_0^*}$ and thus the sought inclusion $\overline{\mathbf{x}}_f^\infty \in X_0^*$.

Since $\{ \mathbf{x} \in X \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \} = \emptyset$, it follows from Proposition 4.3(d) that $\mathcal{I}_+(\mathbf{v}^*) \neq \emptyset$. Consider any $i \in \mathcal{I}_+(\mathbf{v}^*)$. When $t \geq 0$ is large enough, by Corollary 5.5, $u_i^t > 0$ holds, implying that $u_i^{t+1} = u_i^{t+\frac{1}{2}}$ holds in the iteration formula (2.11). Hence, for $N \geq 0$ large enough, it holds that

$$u_i^t = u_i^N + \sum_{s=N}^{t-1} \alpha_s \big[g_i(\mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s)) - \eta_i^s \big] = u_i^N + \Lambda_t \overline{\gamma}_i^{\,t} - \Lambda_N \overline{\gamma}_i^{\,N}, \quad t \geq N+1.$$

By rearranging this equation and dividing the resulting terms by $\| \mathbf{u}^t \|$, it follows that

$$\left\{ \| \mathbf{u}^t \|^{-1} \Lambda_t \overline{\gamma}_i^{\,t} \right\} = \left\{ \| \mathbf{u}^t \|^{-1} \big(u_i^t - u_i^N + \Lambda_N \overline{\gamma}_i^{\,N} \big) \right\} \to v_i^* \quad \text{as} \quad t \to \infty, \tag{6.2}$$

since, by Proposition 3.3, $\{ \| \mathbf{u}^t \| \} \to \infty$ and, by Corollary 5.5, $\{ \| \mathbf{u}^t \|^{-1} u_i^t \} \to v_i^*$. By Corollary 5.6 it holds that $\left\{ (\mathbf{v}^*)^{\mathrm{T}} \mathbf{g}(\mathbf{x}_f^{\epsilon_t}(\mathbf{u}^t)) \right\} \to \theta_0^{V_0^*}$ as $t \to \infty$. Applying Lemma 6.1 with the identifications

$$r = 1, \quad \mu_{ts} = \Lambda_t^{-1} \alpha_s, \quad \mathbf{e}^s = (\mathbf{v}^*)^{\mathrm{T}} \mathbf{g}(\mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s)), \quad \text{and} \quad \mathbf{e} = \theta_0^{V_0^*}$$

then yields that $\left\{ (\mathbf{v}^*)^{\mathrm{T}} \overline{\gamma}^{\,t} \right\} \to \theta_0^{V_0^*}$ as $t \to \infty$. Since $v_j^* = 0$ for $j \in \mathcal{I} \backslash \mathcal{I}_+(\mathbf{v}^*)$, it follows by utilizing (6.2) that

$$\left\{ (\mathbf{v}^*)^{\mathrm{T}} \overline{\gamma}^{\,t} \| \mathbf{u}^t \|^{-1} \Lambda_t \right\} \to \| \mathbf{v}^* \|^2 = 1 \quad \text{as} \quad t \to \infty,$$

which implies that

$$\left\{ \|\mathbf{u}^t\|^{-1} \Lambda_t \right\} = \left\{ (\mathbf{v}^*)^{\mathsf{T}} \overline{\boldsymbol{\gamma}}^t \|\mathbf{u}^t\|^{-1} \Lambda_t \left[(\mathbf{v}^*)^{\mathsf{T}} \overline{\boldsymbol{\gamma}}^t \right]^{-1} \right\} \to \left(\theta_0^{V_0^*} \right)^{-1} \quad \text{as} \quad t \to \infty.$$

By combining this result with (6.2) it then follows that

$$\overline{\gamma}_i^t \to \theta_0^{V_0^*} v_i^* \quad \text{as} \quad t \to \infty, \quad i \in \mathcal{I}_+(\mathbf{v}^*).$$

By the convexity of the functions g_i, for each $i \in \mathcal{I}_+(\mathbf{v}^*)$ it then holds that

$$\limsup_{t \to \infty} \left\{ g_i(\overline{\mathbf{x}}_f^t) \right\} \le \limsup_{t \to \infty} \left\{ \frac{1}{\Lambda_t} \sum_{s=0}^{t-1} \alpha_s g_i(\mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s)) \right\} \le \limsup_{t \to \infty} \left\{ \overline{\gamma}_i^t \right\} = \theta_0^{V_0^*} v_i^*.$$

$$(6.3)$$

From (5.2) follows that the inequality $\|\mathbf{u}^t\| \le \|\mathbf{u}^0\| + \Lambda_t \Gamma$ holds for every $t \ge 1$, where $\Gamma = L_0 + \sup_{r \ge 0} \left\{ \|\mathbf{g}(\mathbf{x}_f^{\epsilon_r}(\mathbf{u}^r)) - \boldsymbol{\eta}^r\| \right\} < \infty$. By Proposition 3.3, for a large enough $N \ge 1$, $\|\mathbf{u}^t\| > \|\mathbf{u}^0\|$ holds for each $t \ge N$, which implies the inequalities

$$\Lambda_t^{-1} \le \Gamma \left(\|\mathbf{u}^t\| - \|\mathbf{u}^0\| \right)^{-1}, \quad t \ge N. \tag{6.4}$$

Then, for each $j \in \mathcal{I} \backslash \mathcal{I}_+(\mathbf{v}^*)$ and all $t \ge N$, the relations

$$g_j(\overline{\mathbf{x}}_f^t) \le \frac{1}{\Lambda_t} \sum_{s=0}^{t-1} \alpha_s g_j(\mathbf{x}_f^{\epsilon_s}(\mathbf{u}^s)) \le \frac{1}{\Lambda_t} \sum_{s=0}^{t-1} \left(u_j^{s+1} - u_j^s \right) = \frac{1}{\Lambda_t} \left(u_j^t - u_j^0 \right)$$

$$\le \Gamma \frac{u_j^t - u_j^0}{\|\mathbf{u}^t\| - \|\mathbf{u}^0\|}$$

hold, where the first inequality follows from the convexity of g_j, the second from (2.11) and the fact that $\eta_j^t \le 0$, the equality by telescoping, and the final inequality by (6.4). As $t \to \infty$, by Proposition 3.3, $\|\mathbf{u}^t\| \to \infty$, and by Corollary 5.5, $\left\{ \|\mathbf{u}^t\|^{-1} u_j^t \right\} \to v_j^* = 0$. It follows that

$$\limsup_{t \to \infty} \left\{ g_j(\overline{\mathbf{x}}_f^t) \right\} \le 0, \quad j \in \mathcal{I} \backslash \mathcal{I}_+(\mathbf{v}^*). \tag{6.5}$$

From (6.3) and (6.5) we then conclude (6.1). The theorem follows. □

6.2 Properties of and convergence results for the linear programming case

We now analyze the special case when the program (2.1) is a linear program, i.e., when the program can be formulated as the problem to

$$\text{minimize } \mathbf{c}^{\mathrm{T}}\mathbf{x}, \tag{6.6a}$$

$$\text{subject to } \mathbf{A}\mathbf{x} \geq \mathbf{b}, \tag{6.6b}$$

$$\mathbf{x} \in X, \tag{6.6c}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $X \subset \mathbb{R}^n$ is a nonempty and bounded polyhedron. The aim of this subsection is to provide a proof of Theorem 2.3(b), stating that the ergodic sequence $\{\bar{\mathbf{x}}_f^t\}$ [defined in (2.14)] converges to the optimal set of the selection problem (2.16).[2] For this linear case, the Lagrangian subproblems in (2.2) can be solved exactly in finite time; hereafter we thus let $\epsilon_t := 0$, $t \geq 0$.

Let $\mathbf{A}_i$ denote the ith row of the matrix $\mathbf{A}$ and let $\mathbf{x}_0 \in X_0(\mathbf{v}^*)$. The selection problem (2.16) can then be expressed as the linear program to

$$\text{minimize } \mathbf{c}^{\mathrm{T}}\mathbf{x}, \tag{6.7a}$$

$$\text{subject to } \mathbf{A}_i\mathbf{x} = b_i - [b_i - \mathbf{A}_i\mathbf{x}_0]_+, \qquad i \in \mathcal{I}_+(\mathbf{v}^*), \tag{6.7b}$$

$$\mathbf{A}_i\mathbf{x} \geq b_i, \qquad i \in \mathcal{I}\backslash\mathcal{I}_+(\mathbf{v}^*), \tag{6.7c}$$

$$\mathbf{x} \in X_0(\mathbf{v}^*). \tag{6.7d}$$

Using that $[b_i - \mathbf{A}_i\mathbf{x}_0]_+ = 0$ for all $i \in \mathcal{I}\backslash\mathcal{I}_+(\mathbf{v}^*)$, we define the (projected) Lagrangian dual function, $\theta_{\mathbf{c}}^+ : \mathbb{R}^m \mapsto \mathbb{R}$, to the program (6.7) with respect to the relaxation of the constraints (6.7b) and (6.7c), as

$$\theta_{\mathbf{c}}^+(\mathbf{u}) := \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \mathbf{u}^{\mathrm{T}}\big(\mathbf{b} - \mathbf{A}\mathbf{x} - [\mathbf{b} - \mathbf{A}\mathbf{x}_0]_+\big) \right\}, \quad \mathbf{u} \in \mathbb{R}^m. \tag{6.8}$$

Defining the *radial cone* to $\mathbb{R}_+^m$ at $\mathbf{v} \in \mathbb{R}_+^m$ as

$$R(\mathbf{v}) := \left\{ \mathbf{u} \in \mathbb{R}^m \,\middle|\, u_i \geq 0, \; i \in \mathcal{I}\backslash\mathcal{I}_+(\mathbf{v}) \right\}, \tag{6.9}$$

the corresponding Lagrange dual is then given by the problem to

$$\max_{\mathbf{u} \in R(\mathbf{v}^*)} \theta_{\mathbf{c}}^+(\mathbf{u}). \tag{6.10}$$

We will show that when applying the conditional ε-subgradient optimization algorithm (2.11) to the Lagrange dual (2.4) of the inconsistent linear program (6.6) with respect to the relaxation of the constraints (6.6b), a subgradient scheme is obtained for

---

[2] For the linear program (6.6), the selection problem (cf. 2.16) is defined as $\min_{\mathbf{x} \in X_0^*}\{\mathbf{c}^{\mathrm{T}}\mathbf{x}\}$, where the set X_0^* is the subset of X possessing minimum infeasibility in the relaxed constraints (6.6b), i.e., in mathematical notation, $X_0^* = \arg\min_{\mathbf{x} \in X}\{\,\|[\mathbf{b} - \mathbf{A}\mathbf{x}]_+\|\,\}$.

the Lagrange dual (6.10) of the selection problem (6.7), which is a consistent linear program. We will then deduce that the ergodic sequence $\{\bar{\mathbf{x}}_f^t\}$ converges to the set of optimal solutions to (6.7). But first we introduce some definitions needed for the analysis to follow.

A decomposition of any vector $\mathbf{u} \in \mathbb{R}^m$ into two vectors being parallel and orthogonal, respectively, to $\mathbf{v}^*$, is given by the maps $\beta : \mathbb{R}^m \mapsto \mathbb{R}$ and $\boldsymbol{\omega} : \mathbb{R}^m \mapsto \mathbb{R}^m$, according to

$$\mathbf{u} = \mathbf{v}^* \beta(\mathbf{u}) + \boldsymbol{\omega}(\mathbf{u}), \quad \beta(\mathbf{u}) := \mathbf{u}^{\mathrm{T}} \mathbf{v}^*, \quad \text{and} \quad \boldsymbol{\omega}(\mathbf{u}) := \mathbf{u} - \mathbf{v}^* \beta(\mathbf{u}). \qquad (6.11)$$

Here, $\beta(\mathbf{u})$ equals the length of the projection of $\mathbf{u} \in \mathbb{R}^m$ onto $\mathbf{v}^*$, while $\boldsymbol{\omega}(\mathbf{u})$ equals the projection of $\mathbf{u}$ onto the orthogonal complement to $\mathbf{v}^*$. Both maps β and $\boldsymbol{\omega}$ define projections onto linear subspaces.

Property 6.2 (Properties of maps) *The following properties of the maps β and $\boldsymbol{\omega}$ hold.*

(a) $\boldsymbol{\omega}(\mathbf{u})^{\mathrm{T}} \boldsymbol{\omega}(\mathbf{v}) = \boldsymbol{\omega}(\mathbf{u})^{\mathrm{T}} \mathbf{v} = \mathbf{u}^{\mathrm{T}} \boldsymbol{\omega}(\mathbf{v})$ *for all* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$,
(b) $\beta(\mathbf{u} + \varrho \mathbf{v}^*) = \beta(\mathbf{u}) + \varrho$ *for all* $\mathbf{u} \in \mathbb{R}^m$ *and all* $\varrho \in \mathbb{R}$, *and*
(c) $\boldsymbol{\omega}(\mathbf{u} + \varrho \mathbf{v}^*) = \boldsymbol{\omega}(\mathbf{u})$ *for all* $\mathbf{u} \in \mathbb{R}^m$ *and all* $\varrho \in \mathbb{R}$. $\qquad\qquad \square$

Using the fact that $\boldsymbol{\omega}(\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x} - [\mathbf{b} - \mathbf{A}\mathbf{x}_0]_+$ for any $\mathbf{x}_0 \in X_0^*$, we can rewrite the Lagrangian dual function, defined in (6.8), as

$$\theta_{\mathbf{c}}^+(\mathbf{u}) = \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \mathbf{c}^{\mathrm{T}} \mathbf{x} + \mathbf{u}^{\mathrm{T}} \boldsymbol{\omega}(\mathbf{b} - \mathbf{A}\mathbf{x}) \right\}.$$

The following lemma follows from Property 6.2(a) and establishes that the value of $\theta_{\mathbf{c}}^+$ at $\mathbf{u} \in \mathbb{R}^m$ depends solely on the component $\boldsymbol{\omega}(\mathbf{u})$ of $\mathbf{u}$ that is perpendicular to $\mathbf{v}^*$.

Lemma 6.3 (A characterization of a projected dual function) *For any* $\mathbf{u} \in \mathbb{R}^m$, *the equivalence* $\theta_{\mathbf{c}}^+(\boldsymbol{\omega}(\mathbf{u})) = \theta_{\mathbf{c}}^+(\mathbf{u})$ *holds.* $\qquad\qquad \square$

Given constants $\delta > 0$, $p > 1$, and $q > (p-1)^{-1}p$, we define the set

$$U_{pq}^\delta := \left\{ \mathbf{u} \in \mathbb{R}^m \mid \beta(\mathbf{u}) \geq q\delta^2, \ \beta(\mathbf{u}) \geq p\delta^2 \|\boldsymbol{\omega}(\mathbf{u})\| \right\} \qquad (6.12)$$

of vectors possessing a large enough norm and a small enough angle with the direction of $\mathbf{v}^*$. The following lemma ensures that after a finite number N of iterations, all of the dual iterates $\mathbf{u}^t$ are contained in the set U_{pq}^δ; it follows from Proposition 3.3 and Corollary 5.5.

Lemma 6.4 (The dual iterates are eventually in the set U_{pq}^δ) *Let the sequence* $\{\mathbf{u}^t\}$ *be generated by the method* (2.11), (2.13a) *applied to the program* (2.4), *let the sequence* $\{\eta^t\}$ *be bounded, and let* $\{\epsilon_t\} = \{0\}$. *Then, for any constants* $\delta > 0$, $p > 1$, *and* $q > (p-1)^{-1}p$ *there is an* $N \geq 0$ *such that* $\mathbf{u}^t \in U_{pq}^\delta$ *for all* $t \geq N$. $\qquad \square$

Propositions 6.5–6.7 below demonstrate that, for $p > 1$, $q > (p-1)^{-1}p$, and a large enough value of $\bar{\delta} > 0$, the condition $\mathbf{u} \in U^{\bar{\delta}}_{pq}$ implies certain relations between the function values $\theta_f(\mathbf{u})$ and $\theta_{\mathbf{c}}^+(\mathbf{u})$, as well as between their respective conditional subdifferentials. First we establish the inclusion $X^0_f(\mathbf{u}) \subseteq X_0(\mathbf{v}^*)$ whenever $\mathbf{u} \in U^{\bar{\delta}}_{pq}$. We then show that the value $\theta_f(\mathbf{u})$ of the Lagrangian dual function equals $\beta(\mathbf{u})\theta_0^{V_0^*} + \theta_{\mathbf{c}}^+(\mathbf{u})$ whenever $\mathbf{u} \in U^{\bar{\delta}}_{pq}$.

Proposition 6.5 (Inclusion of the solution set) *Let $p > 1$ and $q > (p-1)^{-1}p$. There exists a constant $\bar{\delta} > 0$ such that $X^0_f(\mathbf{u}) \subseteq X_0(\mathbf{v}^*)$ holds for all $\mathbf{u} \in U^{\bar{\delta}}_{pq}$.*

Proof For the case when $X_0(\mathbf{v}^*) = X$ the proposition is immediate. Consider the case when $X_0(\mathbf{v}^*) \subset X$. Denote by P_X, $P_{X^0_f(\mathbf{u})}$, and $P_{X_0(\mathbf{v}^*)}$ the (finite) sets of extreme points of X, $X^0_f(\mathbf{u})$, and $X_0(\mathbf{v}^*)$, respectively. From (2.3) and [32, Ch. I.4, Def. 3.1] follow that $X^0_f(\mathbf{u})$ and $X_0(\mathbf{v}^*)$ are faces of X, implying the relations $P_{X^0_f(\mathbf{u})} \subseteq P_X$, $\mathbf{u} \in \mathbb{R}^m$, and $P_{X_0(\mathbf{v}^*)} \subseteq P_X \subset X$. Hence, it suffices to show that $P_{X^0_f(\mathbf{u})} \subseteq P_{X_0(\mathbf{v}^*)}$ holds whenever $\mathbf{u} \in U^{\bar{\delta}}_{pq}$. Let $\mathbf{x}^*_0 \in P_{X_0(\mathbf{v}^*)}$ and $\bar{\mathbf{x}} \in P_X \backslash P_{X_0(\mathbf{v}^*)}$ be arbitrary. Since the set P_X is finite there exists a $\bar{\delta} > 0$ such that the relations

$$\mathbf{c}^{\mathrm{T}} \left(\bar{\mathbf{x}} - \mathbf{x}^*_0 \right) \geq -\bar{\delta}, \tag{6.13a}$$

$$-(\mathbf{v}^*)^{\mathrm{T}}\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}^*_0) \geq \bar{\delta}^{-1}, \tag{6.13b}$$

$$\|\mathbf{A} \left(\bar{\mathbf{x}} - \mathbf{x}^*_0 \right) \| \leq \bar{\delta} \tag{6.13c}$$

hold. For any $\mathbf{u} \in U^{\bar{\delta}}_{pq}$ it then follows that

$$\mathcal{L}_f(\bar{\mathbf{x}}, \mathbf{u}) - \mathcal{L}_f(\mathbf{x}^*_0, \mathbf{u}) = \mathbf{c}^{\mathrm{T}}(\bar{\mathbf{x}} - \mathbf{x}^*_0) - \beta(\mathbf{u})(\mathbf{v}^*)^{\mathrm{T}}\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}^*_0) - \boldsymbol{\omega}(\mathbf{u})^{\mathrm{T}}\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}^*_0) \tag{6.14a}$$

$$\geq -\bar{\delta} + \beta(\mathbf{u})\bar{\delta}^{-1} - \bar{\delta}\|\boldsymbol{\omega}(\mathbf{u})\| \tag{6.14b}$$

$$\geq \bar{\delta}\left(qp^{-1}[p - 1] - 1\right) > 0, \tag{6.14c}$$

where (6.14a) follows from (6.11), (6.14b) from (6.13) and Cauchy-Schwartz inequality, and (6.14c) from the definition (6.12) and the assumptions made. It follows that $\bar{\mathbf{x}} \notin X^0_f(\mathbf{u})$, which then implies that $X^0_f(\mathbf{u}) \subseteq X_0(\mathbf{v}^*)$. The proposition follows. □

In the analysis to follow we choose $\bar{\delta} > 0$ such that the inclusion in Proposition 6.5 holds.

Proposition 6.6 (A decomposition of the dual function) *Let $p > 1$ and $q > (p-1)^{-1}p$. For every $\mathbf{u} \in U^{\bar{\delta}}_{pq}$ the identity $\theta_f(\mathbf{u}) = \beta(\mathbf{u})\theta_0^{V_0^*} + \theta_{\mathbf{c}}^+(\mathbf{u})$ holds.*

Proof The result follows since Proposition 6.5, (6.11), (4.3), and Property 6.2(a) yield the equalities

$$
\begin{aligned}
\theta_f(\mathbf{u}) &= \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \mathbf{c}^\mathrm{T}\mathbf{x} + \mathbf{u}^\mathrm{T}(\mathbf{b} - \mathbf{A}\mathbf{x}) \right\} \\
&= \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \mathbf{c}^\mathrm{T}\mathbf{x} + \left(\mathbf{v}^*\beta(\mathbf{u}) + \omega(\mathbf{u})\right)^\mathrm{T}(\mathbf{b} - \mathbf{A}\mathbf{x}) \right\} \\
&= \theta_0^{V_0^*}\beta(\mathbf{u}) + \min_{\mathbf{x} \in X_0(\mathbf{v}^*)} \left\{ \mathbf{c}^\mathrm{T}\mathbf{x} + \omega(\mathbf{u})^\mathrm{T}(\mathbf{b} - \mathbf{A}\mathbf{x}) \right\} = \theta_0^{V_0^*}\beta(\mathbf{u}) + \theta_{\mathbf{c}}^+(\mathbf{u}).
\end{aligned}
$$

$\square$

We next establish that if $\boldsymbol{\gamma}$ is a conditional (with respect to $\mathbb{R}_+^m$) subgradient to θ_f at $\mathbf{u} \in \mathbb{R}_+^m$, where $\mathbf{u}$ has a sufficiently large norm and a sufficiently small component $\omega(\mathbf{u})$ (being orthogonal to $\mathbf{v}^*$), then $\omega(\boldsymbol{\gamma})$ is a conditional [with respect to $R(\mathbf{v}^*)$; see (6.9)] subgradient of $\theta_{\mathbf{c}}^+$ at $\omega(\mathbf{u}) \in R(\mathbf{v}^*)$.

Proposition 6.7 (Conditional subgradients of a projected dual function) *Let $p > 1$ and $q > (p-1)^{-1}p$. For each $\mathbf{u} \in U_{pq}^{\overline{\delta}}$ and $\boldsymbol{\gamma} \in \partial_0^{\mathbb{R}_+^m}\theta_f(\mathbf{u})$ the inclusion $\omega(\boldsymbol{\gamma}) \in \partial_0^{R(\mathbf{v}^*)}\theta_{\mathbf{c}}^+(\omega(\mathbf{u}))$ holds.*

Proof Let $\mathbf{v} \in R(\mathbf{v}^*)$ and choose $\varrho \geq 0$ such that $\mathbf{v} + \varrho\mathbf{v}^* \in U_{pq}^{\overline{\delta}}$. From Lemma 6.3, Proposition 6.6, and Property 6.2(c) follow that the equalities

$$
\theta_f(\mathbf{v} + \varrho\mathbf{v}^*) - \theta_f(\mathbf{u}) = \left[\beta(\mathbf{v} + \varrho\mathbf{v}^*) - \beta(\mathbf{u}) \right]\theta_0^{V_0^*} + \theta_{\mathbf{c}}^+(\mathbf{v} + \varrho\mathbf{v}^*) - \theta_{\mathbf{c}}^+(\mathbf{u}) \tag{6.15a}
$$

$$
= \left[(\mathbf{v} - \mathbf{u})^\mathrm{T}\mathbf{v}^* + \varrho \right]\theta_0^{V_0^*} + \theta_{\mathbf{c}}^+(\omega(\mathbf{v})) - \theta_{\mathbf{c}}^+(\omega(\mathbf{u})) \tag{6.15b}
$$

hold. Since $\mathbf{u} \in U_{pq}^{\overline{\delta}} \subset \mathbb{R}_+^m$, $\mathbf{v} + \varrho\mathbf{v}^* \in U_{pq}^{\overline{\delta}}$, and $\boldsymbol{\gamma} \in \partial_0^{\mathbb{R}_+^m}\theta_f(\mathbf{u})$ it follows that

$$
\theta_f(\mathbf{v} + \varrho\mathbf{v}^*) - \theta_f(\mathbf{u}) \leq \boldsymbol{\gamma}^\mathrm{T}\left(\mathbf{v} + \varrho\mathbf{v}^* - \mathbf{u}\right) \tag{6.16a}
$$

$$
= \boldsymbol{\gamma}^\mathrm{T}\left[\omega(\mathbf{v} + \varrho\mathbf{v}^*) - \omega(\mathbf{u}) \right] + \boldsymbol{\gamma}^\mathrm{T}\left[\beta(\mathbf{v} + \varrho\mathbf{v}^*) - \beta(\mathbf{u}) \right]\mathbf{v}^* \tag{6.16b}
$$

$$
= \omega(\boldsymbol{\gamma})^\mathrm{T}\left[\omega(\mathbf{v}) - \omega(\mathbf{u}) \right] + \boldsymbol{\gamma}^\mathrm{T}\left[(\mathbf{v} - \mathbf{u})^\mathrm{T}\mathbf{v}^* + \varrho \right]\mathbf{v}^*, \tag{6.16c}
$$

where (6.16a) follows from (2.8), (6.16b) from (6.11), and (6.16c) from Property 6.2. Combining the relations in (6.15) and (6.16) yields the inequality

$$
\theta_{\mathbf{c}}^+(\omega(\mathbf{v})) - \theta_{\mathbf{c}}^+(\omega(\mathbf{u})) \leq \omega(\boldsymbol{\gamma})^\mathrm{T}\left[\omega(\mathbf{v}) - \omega(\mathbf{u}) \right] + \left[(\mathbf{v} - \mathbf{u})^\mathrm{T}\mathbf{v}^* + \varrho \right]\left(\boldsymbol{\gamma}^\mathrm{T}\mathbf{v}^* - \theta_0^{V_0^*} \right). \tag{6.17}
$$

Since $\boldsymbol{\gamma} \in \partial_0^{\mathbb{R}_+^m}\theta_f(\mathbf{u})$ and (according to Proposition 6.5) $X_f^0(\mathbf{u}) \subseteq X_0(\mathbf{v}^*)$, it then follows that $\theta_0^{V_0^*} = \min_{\mathbf{x} \in X}\left\{ (\mathbf{b} - \mathbf{A}\mathbf{x})^\mathrm{T}\mathbf{v}^* \right\} = \boldsymbol{\gamma}^\mathrm{T}\mathbf{v}^*$. By inserting this into (6.17),

and utilizing Property 6.2(a) and (c), and Lemma 6.3, we then receive the inequality

$$\theta_{\mathbf{c}}^+(\mathbf{v}) - \theta_{\mathbf{c}}^+(\boldsymbol{\omega}(\mathbf{u})) \leq \boldsymbol{\omega}(\boldsymbol{\gamma})^{\mathrm{T}}\big[\,\mathbf{v} - \boldsymbol{\omega}(\mathbf{u})\,\big].$$

The proposition then follows since $\mathbf{v} \in R(\mathbf{v}^*)$. □

We now define the sequences $\{\boldsymbol{\omega}^t\}$ and $\{\boldsymbol{\omega}^{t+\frac{1}{2}}\}$ according to

$$\boldsymbol{\omega}^t := \boldsymbol{\omega}(\mathbf{u}^t) \quad \text{and} \quad \boldsymbol{\omega}^{t+\frac{1}{2}} := \boldsymbol{\omega}^t + \alpha_t \boldsymbol{\omega}(\mathbf{b} - \mathbf{A}\mathbf{x}_f^0(\mathbf{u}^t) - \boldsymbol{\eta}^t), \quad t = 0, 1, \dots. \quad (6.18)$$

In each iteration, t, the intermediate iterate $\boldsymbol{\omega}^{t+\frac{1}{2}}$ is the result of a step in the direction of $\boldsymbol{\omega}(\mathbf{b} - \mathbf{A}\mathbf{x}_f^0(\mathbf{u}^t) - \boldsymbol{\eta}^t)$. The vector $\mathbf{b} - \mathbf{A}\mathbf{x}_f^0(\mathbf{u}^t) - \boldsymbol{\eta}^t \in \mathbb{R}^m$ is a conditional (with respect to $\mathbb{R}_+^m$) subgradient to the Lagrangian dual function (2.2), so by Proposition 6.7 the vector $\boldsymbol{\omega}(\mathbf{b} - \mathbf{A}\mathbf{x}_f^0(\mathbf{u}^t) - \boldsymbol{\eta}^t)$ is a conditional [with respect to $R(\mathbf{v}^*)$; see (6.9)] subgradient to the dual function (6.8) for large enough values of t. To show that the formula (6.18) actually defines a conditional [with respect to $R(\mathbf{v}^*)$] subgradient algorithm, we must also show that $\boldsymbol{\omega}^{t+1} = \mathrm{proj}(\boldsymbol{\omega}^{t+\frac{1}{2}}; R(\mathbf{v}^*))$.

Proposition 6.8 (A subgradient method for the projected dual function) *Let the sequence $\{\mathbf{u}^t\}$ be generated by the method (2.11), (2.13a) applied to (6.6), and the sequences $\{\boldsymbol{\omega}^t\}$ and $\{\boldsymbol{\omega}^{t+\frac{1}{2}}\}$ by (6.18); let the sequence $\{\boldsymbol{\eta}^t\}$ be bounded and let $\{\epsilon_t\} = \{0\}$. Then, there exists an $N \geq 0$ such that $\mathrm{proj}(\boldsymbol{\omega}^{t+\frac{1}{2}}; R(\mathbf{v}^*)) = \boldsymbol{\omega}^{t+1}$ for all $t \geq N$.*

Proof By (6.18), (6.11), and (2.11), $\boldsymbol{\omega}^{t+\frac{1}{2}} = \boldsymbol{\omega}(\mathbf{u}^{t+\frac{1}{2}})$ holds for all $t \geq 0$. Define $\overline{\boldsymbol{\omega}}^t := \mathrm{proj}(\boldsymbol{\omega}^{t+\frac{1}{2}}; R(\mathbf{v}^*))$ and note that $\omega_i(\mathbf{u}) = u_i - v_i^*(\mathbf{v}^*)^{\mathrm{T}}\mathbf{u}$ holds for all $i \in \mathcal{I}$ and all $\mathbf{u} \in \mathbb{R}^m$.

Consider $i \in \mathcal{I} \setminus \mathcal{I}_+(\mathbf{v}^*)$, so that $v_i^* = 0$. By (6.9), (6.11), (2.11), and (6.18) follow that

$$\overline{\omega}_i^t = \big[\omega_i^{t+\frac{1}{2}}\big]_+ = \big[\omega_i(\mathbf{u}^{t+\frac{1}{2}})\big]_+ = \big[u_i^{t+\frac{1}{2}}\big]_+ = u_i^{t+1} = \omega_i(\mathbf{u}^{t+1}) = \omega_i^{t+1}, \quad t \geq 0.$$

Consider $i \in \mathcal{I}_+(\mathbf{v}^*)$, so that $v_i^* > 0$. Due to (6.9) and (6.11) it then holds that

$$\overline{\omega}_i^t = \omega_i^{t+\frac{1}{2}} = \omega_i(\mathbf{u}^{t+\frac{1}{2}}) = u_i^{t+\frac{1}{2}} - v_i^*(\mathbf{v}^*)^{\mathrm{T}}\mathbf{u}^{t+\frac{1}{2}}$$
$$= u_i^{t+\frac{1}{2}} - v_i^* \sum_{j \in \mathcal{I}_+(\mathbf{v}^*)} v_j^* u_j^{t+\frac{1}{2}}, \quad t \geq 0.$$

For all $t \geq N$, where $N \geq 0$ is large enough, the relations $0 < u_i^{t+1} = u_i^{t+\frac{1}{2}}$ hold, implying that $\overline{\omega}_i^t = u_i^{t+1} - v_i^* \sum_{j \in \mathcal{I}_+(\mathbf{v}^*)} v_j^* u_j^{t+1} = \omega_i^{t+1}$, where the latter identity is due to (6.18).

We conclude that $\overline{\boldsymbol{\omega}}^t = \boldsymbol{\omega}^{t+1}$ for all $t \geq N$, and the proposition follows. □

We summarize the development made in this section. Associated with the sequence $\{\mathbf{u}^t\} \subset \mathbb{R}_+^m$ of dual iterates resulting from a conditional subgradient scheme for

maximizing θ_f over $\mathbb{R}^m_+$, we define in (6.18) a sequence $\{\boldsymbol{\omega}^t\} \equiv \{\boldsymbol{\omega}(\mathbf{u}^t)\} \subset R(\mathbf{v}^*)$ of iterates corresponding to the function $\theta_{\mathbf{c}}^+$. Proposition 6.7 shows that a conditional (with respect to $\mathbb{R}^m_+$) subgradient of θ_f at $\mathbf{u}^t \in \mathbb{R}^m_+$ can be mapped to a conditional [with respect to $R(\mathbf{v}^*)$] subgradient of $\theta_{\mathbf{c}}^+$ at $\boldsymbol{\omega}^t \in R(\mathbf{v}^*)$. Then, Proposition 6.8 shows that for a large enough value of t the projection of $\mathbf{u}^{t+\frac{1}{2}}$ onto $\mathbb{R}^m_+$ in (2.11) has a one-to-one correspondence with the projection of $\boldsymbol{\omega}^{t+\frac{1}{2}}$ onto the set $R(\mathbf{v}^*)$ [defined in (6.9)].

We are now prepared to establish the remaining part of the main result of this article.

Proof of Theorem 2.3(b) The case when $\{\mathbf{x} \in X \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\} \neq \emptyset$ is treated in Proposition 2.2.

Assume that $\{\mathbf{x} \in X \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\} = \emptyset$. By Lemma 6.4 and Propositions 6.5–6.8, there is an integer $N \geq 0$ such that the sequence $\{\boldsymbol{\omega}^t\}_{t \geq N}$ is the result of a conditional [with respect to $R(\mathbf{v}^*)$] subgradient method applied to the Lagrangian dual (6.10) of the linear program (6.7), which has a nonempty and bounded feasible set. It then follows from Proposition 2.2 that $\mathrm{dist}(\bar{\mathbf{x}}^t_f; X^*_f) \to 0$ as $t \to \infty$. The theorem follows. $\square$

This proof of Theorem 2.3(b) contains no explicit reference to any particular choice of the weights defining the ergodic sequence (2.14). Although this article is written with reference to the formula (2.14), the result of Theorem 2.3(b) will be valid for *any* ergodic sequence of primal iterates that is convergent for consistent programs (see, e.g., [14]), assuming that the corresponding version of Theorem 5.4 can be established.

7 Illustrations and a separation result

We next present an example which numerically illustrates the main findings made in this article and, finally, a separation result following from our analysis.

7.1 A numerical example

Consider the linear programming instance of the program (2.1) given by that to

$$\text{minimize} \quad 4x_1 + 2x_2, \tag{7.1a}$$
$$\text{subject to} \quad x_1 - x_2 \geq 2, \tag{7.1b}$$
$$-x_1 + 2x_2 \geq 4, \tag{7.1c}$$
$$\mathbf{x} \in X := [0, 4]^2. \tag{7.1d}$$

Here $\{\mathbf{x} \in X \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}\} = \emptyset$; see illustration in Fig. 1; the corresponding cone $C = \{\mathbf{w} \in \mathbb{R}^2_+ \mid 2w_2 < 3w_1 < 12w_2\}$ is illustrated in Fig. 2. It is straightforward to show that

$$X^*_0 = X^*_f = \left\{(4, \tfrac{18}{5})^{\mathrm{T}}\right\}, \quad \mathbf{v}^* = \tfrac{1}{\sqrt{5}}(2, 1)^{\mathrm{T}}, \quad \text{and} \quad \theta_0^{V^*_0} = \tfrac{4}{\sqrt{5}}.$$

We computed the sequence $\{\mathbf{u}^t\}$ by the method (2.11), with $\mathbf{u}^0 = \boldsymbol{\eta}^t = (0,0)^{\mathrm{T}}$ and $\alpha_t = \frac{10}{1+t}$.

Figure 2 illustrates the dual iterates $\mathbf{u}^t$ (*left*) and the scaled dual iterates $\mathbf{v}^t$ (*right*), for $t = 0, \ldots, 30$; the dashed lines indicate (the direction of) the vector $\mathbf{v}^* = \frac{1}{\sqrt{5}}(2,1)^{\mathrm{T}}$. The sequence $\{\mathbf{u}^t\}$ diverges in the direction of $\mathbf{v}^*$ and the sequence $\{\mathbf{v}^t\}$ converges to $\mathbf{v}^*$.

In Figure 1 the ergodic primal iterates $\bar{\mathbf{x}}_f^t$ [defined in (2.14)] are illustrated for $t = 1, \ldots, 30$. The sequence $\{\bar{\mathbf{x}}_f^t\}$ converges to the singleton set $X_0^* = \left\{ (4, \frac{18}{5})^{\mathrm{T}} \right\}$.

The appearance of the solution set $X_0^* = \arg\min_{\mathbf{x} \in X} \{ \|[\mathbf{g}(\mathbf{x})]_+\| \}$ depends on the scaling of the constraints in (2.1b), as demonstrated next. The set $X_0^* = \left\{ (4, \frac{18}{5})^{\mathrm{T}} \right\}$ is illustrated in Fig. 3 (*left*). Scaling the constraint (7.1c) to "$-\frac{1}{4}x_1 + \frac{1}{2}x_2 \geq 1$" yields the solution set $X_0^* = \left\{ (4, \frac{12}{5})^{\mathrm{T}} \right\}$, which is illustrated in Fig. 3 (*right*).

Fig. 1 Illustration in the primal space for the linear programming instance (7.1) of the program (2.1). The primal ergodic iterates $\bar{\mathbf{x}}_f^t$, $t = 1, \ldots, 30$, (*circle*) tend to the optimal set $X_0^* = \left\{ (4, \frac{18}{5})^{\mathrm{T}} \right\}$ (*bullet*)
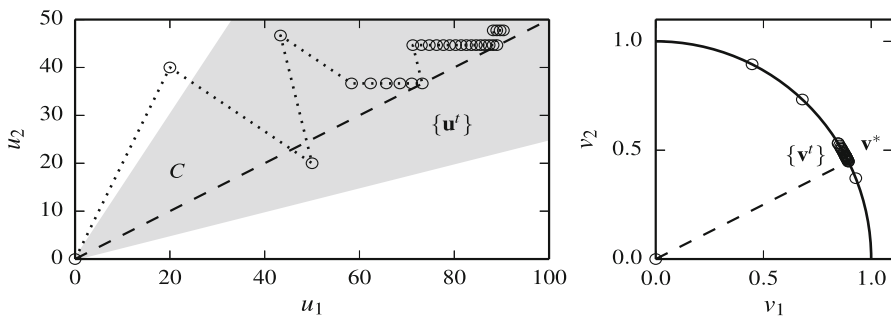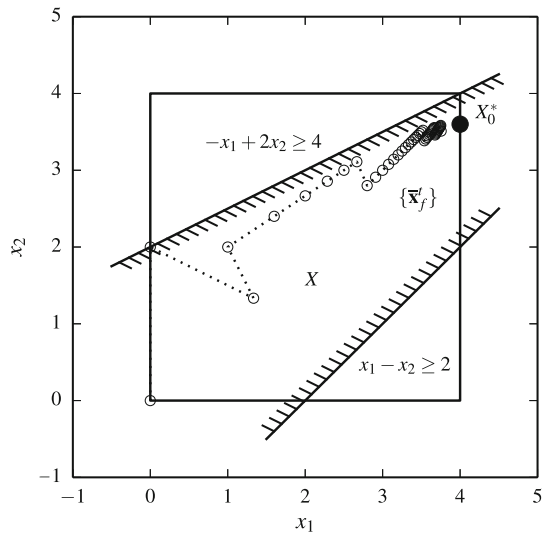
Fig. 2 Illustration in the Lagrangian dual space of the instance (7.1) of (2.1). *Left* the cone C (*shaded*), the dual iterates $\mathbf{u}^t$, $t = 1, \ldots, 30$, (*circle*) and the direction of $\mathbf{v}^*$ (*dashed line*). *Right* the scaled dual iterates $\mathbf{v}^t$, $t = 1, \ldots, 30$, (*circle*) and the vector $\mathbf{v}^*$ (*dashed line*)
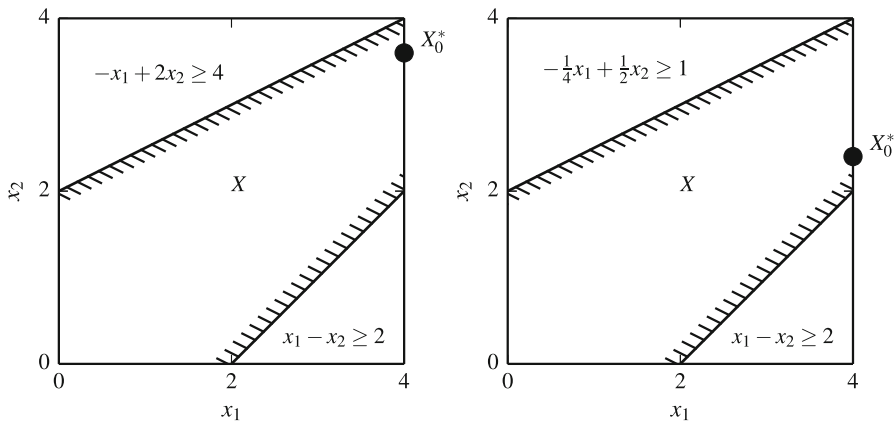
Fig. 3 Illustration of the set X_0^* (*bullet*) for two scalings of the constraint (7.1c) of the problem instance (7.1). *Left* $X_0^* = \left\{ (4, \frac{18}{5})^\mathrm{T} \right\}$. *Right* $X_0^* = \left\{ (4, \frac{12}{5})^\mathrm{T} \right\}$

7.2 Finite attainment of a separating hyper-surface

For the case when the set X is nonempty, closed, and convex, and all functions g_i are affine, we have previously, in [24, Cor. 6.4] (see also Cor. 3.24 in the survey [27]), utilized ergodic sequences of subgradient optimization based underestimating affine functions to finitely detect inconsistency and identify a separating hyper-plane.

We now return to our original setting of convex functions f and g_i, $i \in \mathcal{I}$, and a nonempty, convex and compact set X. Provided that the feasible set $\{ \mathbf{x} \in X \,|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \}$ is empty and that the sets X and $Y := \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \right\}$ are both nonempty, a hyper-surface that strongly separates the sets X and Y can be identified in a finite number of steps.

Theorem 7.1 (Finite attainment of a separating hyper-surface) *Let the sequence $\{ \mathbf{u}^t \}$ be generated by the method* (2.11), (2.13a), (2.13c) *applied to the program* (2.4), *the sequence $\{ \boldsymbol{\eta}^t \}$ be bounded, and the sequence $\{ \mathbf{v}^t \}$ be defined by* (5.1b). *Suppose that the sets X and Y are nonempty, but $X \cap Y = \emptyset$. Then there exists an integer $N \geq 0$ such that the hyper-surface $H(\mathbf{v}^t) := \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \mathbf{g}(\mathbf{x})^\mathrm{T} \mathbf{v}^t = \frac{1}{2} \theta_0^{V_0^*} \right\}$ strongly separates the sets X and Y for all $t \geq N$.*

Proof Since $\mathbf{v}^t \geq \mathbf{0}^m$ for all t, it holds that $\mathbf{g}(\mathbf{x})^\mathrm{T} \mathbf{v}^t \leq 0$ for all $\mathbf{x} \in Y$ and all t. From (2.4) and (2.2) it follows that $\mathbf{g}(\mathbf{x})^\mathrm{T} \mathbf{v}^t \geq \theta_0(\mathbf{v}^t)$ for all $\mathbf{x} \in X$. From Proposition 3.1 it follows that the relations $\emptyset \neq C \subset \mathbb{R}_+^m$ hold and also that the strict inequality $\theta_0(\mathbf{w}) > 0$ holds for all $\mathbf{w} \in C$. Since $C \cap V \neq \emptyset$, (4.2) yields that $\theta_0^{V_0^*} > 0$. Moreover, by Theorem 5.4, $\theta_0(\mathbf{v}^t) \to \theta_0^{V_0^*}$ as $t \to \infty$. Hence, there is an $N \geq 0$ such that $\theta_0(\mathbf{v}^t) > \frac{1}{2} \theta_0^{V_0^*}$ for all $t \geq N$, and it follows that $\mathbf{g}(\mathbf{x})^\mathrm{T} \mathbf{v}^t > \frac{1}{2} \theta_0^{V_0^*} > 0$ for all $\mathbf{x} \in X$ and all $t \geq N$. The theorem follows. $\qquad\square$

8 Conclusions and further research

In this work we apply a conditional subgradient optimization algorithm to the Lagrange dual of a (possibly) inconsistent convex program and compute an associated primal ergodic sequence. We establish the convergence of the resulting dual and primal sequences to the set of saddle-points for the Lagrange function; for the special case of linear programming the primal ergodic sequence converges to the minimum of the primal objective function over the primal part of the set of saddle-points. The stronger result for the case of linear programming is explained by the fact that the corresponding selection problem (6.7) is guaranteed to possess Lagrange multipliers.[3]

The convergence rate for both the primal and the dual sequences can probably be improved via more careful choices of weights for the sequence of weighted averages of primal subproblem solutions, e.g., as suggested in [14] for consistent convex programs.

Another interesting subject for further study is the behavior of the primal–dual sequences when applying other solution schemes, such as bundle methods (see, e.g., [30]) and augmented Lagrangian methods (see, e.g., [6,38]), to the Lagrangian dual problem.

References

1. Anstreicher, K.A., Wolsey, L.A.: Two "well-known" properties of subgradient optimization. Math. Program. **120**(1), 213–220 (2009)
2. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, 3rd edn. Wiley, New York (2006)
3. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont (1999)
4. Byrd, R.H., Nocedal, J., Waltz, R.A.: Steering exact penalty methods for nonlinear programming. Optim. Methods Soft. **23**(2), 197–213 (2008)
5. Chen, L., Goldfarb, D.: Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties. Math. Program. A **108**, 1–36 (2006)
6. Chiche, A., Gilbert, J.C.: How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. J. Convex Anal. **23**(2), 425–459 (2016)
7. Dax, A.: The smallest correction of an inconsistent system of linear inequalities. Optim. Eng. **2**(3), 349–359 (2001)
8. Dem'janov, V.F., Šomesova, V.K.: Conditional subdifferentials of convex functions. Sov. Math. Dokl. **19**(5), 1181–1185 (1978)

---

[3] Establishing the stronger result without assuming the existence of Lagrange multipliers for the selection problem would likely be akin to establishing convergence of primal–dual methods for *consistent* programs without Lagrange multipliers. Such a result is deemed by the authors to be more interesting in its own right than as applied to the current context.

9. Ermol'ev, Yu.M.: Methods for solving nonlinear extremal problems. Cybernetics **2**(4), 1–14 (1966)
10. Fisher, M.L.: The Lagrangian relaxation method for solving integer programming problems. Manag. Sci. **27**(1), 1–18 (1981)
11. Fisher, M.L.: An applications oriented guide to Lagrangian relaxation. Interfaces **15**(2), 10–21 (1985)
12. Fletcher, R., Leyffer, S., Toint, P.: A brief history of filter methods. SIAG/OPT Views News **18**(1), 2–12 (2007)
13. Geoffrion, A.M.: Lagrangian relaxation for integer programming. Math. Program. Study **2**, 82–114 (1974)
14. Gustavsson, E., Patriksson, M., Strömberg, A.-B.: Primal convergence from dual subgradient methods for convex optimization. Math. Program. **150**(2), 365–390 (2015)
15. Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees. Oper. Res. **18**(6), 1138–1162 (1970)
16. Held, M., Karp, R.M.: The traveling-salesman problem and minimum spanning trees: Part II. Math. Program. **1**(1), 6–25 (1971)
17. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I. Springer, Berlin (1993)
18. Karas, E., Ribeiro, A., Sagastizábal, C., Solodov, M.: A bundle-filter method for nonsmooth convex constrained optimization. Math. Program. B **116**, 297–320 (2009)
19. Kiwiel, K.C.: Approximations in proximal bundle methods and decomposition of convex programs. J. Optim. Theory Appl. **84**(3), 529–548 (1995)
20. Kiwiel, K.C., Larsson, T., Lindberg, P.O.: Lagrangian relaxation via ballstep subgradient methods. Math. Oper. Res. **32**(3), 669–686 (2007)
21. Knopp, K.: Infinite Sequences and Series. Dover Publications, New York (1956)
22. 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)
23. Larsson, T., Patriksson, M., Strömberg, A.-B.: Conditional subgradient optimization—theory and applications. Eur. J. Oper. Res. **88**(2), 382–403 (1996)
24. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic convergence in subgradient optimization. Optim. Methods Softw. **9**(1–3), 93–120 (1998)
25. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic, primal convergence in dual subgradient schemes for convex programming. Math. Program. **86**(2), 283–312 (1999)
26. 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)
27. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic convergence in subgradient optimization—with application to simplicial decomposition of convex programs. Contemp. Math. **568**, 159–189 (2012)
28. Lasdon, L.S.: Optimization Theory for Large Systems. MacMillan Series for Operations Research. MacMillan Publishing Co., Inc., New York (1970). Reprinted by Dover Publications, Mineola (2002)
29. Lemaréchal, C.: Lagrangian relaxation. In: Jünger, M., Naddef, D. (eds.) Computational Combinatorial Optimization: Optimal or Provably Near-Optimal Solutions, no. 2241 in Lecture Notes in Computer Science, pp. 112–156. Springer, Berlin (2001)
30. Lemaréchal, C., Nemirovski, A., Nesterov, Yu.: New variants of bundle methods. Math. Program. **69**(1–3), 111–147 (1995)
31. Liu, X., Sun, J.: A robust primal-dual interior-point algorithm for nonlinear programs. SIAM J. Optim. **14**(4), 1163–1186 (2004)
32. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York (1988)
33. Nesterov, Yu.: Primal-dual subgradient methods for convex programs. Math. Program. Ser. B **120**(1), 221–259 (2009)
34. Peng, Y., Feng, H., Li, Q.: A filter-variable-metric method for nonsmooth convex constrained optimization. Appl. Math. Comput. **208**, 119–128 (2009)
35. Polyak, B.T.: A general method of solving extremum problems. Sov. Math. Dokl. **8**(3), 593–597 (1967)
36. Polyak, B.T.: Minimization of unsmooth functionals. USSR Comput. Math. Math. Phys. **9**(3), 14–29 (1969)
37. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
38. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex optimization. Math. Oper. Res. **1**(2), 97–116 (1976)

39. 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**, 105–113 (1996)
40. Shor, N.Z.: Minimization Methods for Non-differentiable Functions. Springer, Berlin (1985)