ERGODIC CONVERGENCE IN SUBGRADIENT OPTIMIZATION

T. LARSSON[1], M. PATRIKSSON[2] and A.-B. STRÖMBERG[1]

[1] *Department of Mathematics, Linköping University, Linköping, Sweden*
[2] *Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden*

When nonsmooth, convex minimization problems are solved by subgradient optimization methods, the subgradients used will in general *not* accumulate to subgradients which verify the optimality of a solution obtained in the limit. It is therefore not a straightforward task to monitor the progress of a subgradient method in terms of the approximate fulfilment of optimality conditions. Further, certain supplementary information, such as convergent estimates of Lagrange multipliers and convergent lower bounds on the optimal objective value, is not directly available in subgradient schemes.

As a means of overcoming these weaknesses in subgradient methods, we introduce the computation of an ergodic (averaged) sequence of subgradients. Specifically, we consider a nonsmooth, convex program solved by a conditional subgradient optimization scheme with divergent series step lengths, and show that the elements of the ergodic sequence of subgradients in the limit fulfil the optimality conditions at the optimal solution, to which the sequence of iterates converges.

This result has three important implications. First, it enables the finite identification of active constraints at the solution obtained in the limit. Second, it is used to establish the ergodic convergence of sequences of Lagrange multipliers; this result enables us to carry out sensitivity analyses for solutions obtained by subgradient methods. The third implication is the convergence of a lower bounding procedure based on an ergodic sequence of affine underestimates of the objective function; this procedure provides a proper termination criterion for subgradient methods.

KEY WORDS: Nonsmooth minimization, conditional subgradient optimization, ergodic convergence, finite identification results, Lagrange multipliers, bounding procedure

1 INTRODUCTION

Subgradient optimization methods are popular practical tools for nonsmooth, convex minimization, especially in the context of Lagrangean relaxation; their simplicity has contributed to their success. This paper focuses on information which is lacking in the straightforward application of subgradient optimization methods and on how to access it through suitable modifications or extensions of the scheme. A technique which is both classical and modern for inducing properties which an original sequence lacks is the construction of an ergodic (averaged) sequence which smoothes out oscillations in the original sequence. We introduce the computation of ergodic sequences from sequences of subgradients, normal elements, and Lagrange multipliers, and establish that the ergodic sequences inherit important properties from analogous sequences associated with smooth functions. The ergodic sequences are then utilized in procedures generating the interesting information.

The basis for our construction of ergodic sequences is the *conditional subgradient optimization method* (Larsson et al., 1996a) — which includes traditional subgradient optimization as a special case — using divergent series step lengths. The work preceding the writing of this paper is basic theoretical research with the purpose of finding and analyzing inherent properties of subgradient optimization methods; it reveals new and interesting characteristics of these methods.

2 BACKGROUND and BASIC CONCEPTS

Let $X \subseteq \Re^n$ be a nonempty, closed and convex set, and the function $f : \Re^n \mapsto \Re$ be nonsmooth and convex and thus everywhere continuous and subdifferentiable. Further, assume that $\inf_{\mathbf{x}\in X} f(\mathbf{x})$ is attained at some $\mathbf{x} \in X$. These properties are assumed to hold throughout the paper. The program considered is

$$\text{[P]} \qquad f^* = \min_{\mathbf{x}\in X} f(\mathbf{x}),$$

with the nonempty, closed and convex *solution set* $X^* = \{\mathbf{x} \in X \mid f(\mathbf{x}) \leq f^*\}$.

To enable the statement of optimality conditions for this program, we first give some basic definitions. The *subdifferential* of f at $\mathbf{x} \in \Re^n$ is

$$\partial f(\mathbf{x}) = \left\{ \boldsymbol{\gamma} \in \Re^n \;\middle|\; f(\mathbf{y}) \geq f(\mathbf{x}) + \boldsymbol{\gamma}^T(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in \Re^n \right\}, \tag{1}$$

the elements of which are called *subgradients*. The *normal cone* to the set X at some $\mathbf{x} \in X$ is the set

$$N_X(\mathbf{x}) = \left\{ \boldsymbol{\nu} \in \Re^n \;\middle|\; \boldsymbol{\nu}^T(\mathbf{y} - \mathbf{x}) \leq 0, \quad \forall \mathbf{y} \in X \right\}, \tag{2}$$

and the *tangent cone* to the set X at $\mathbf{x} \in X$ is

$$T_X(\mathbf{x}) = \left\{ \boldsymbol{\tau} \in \Re^n \;\middle|\; \boldsymbol{\tau}^T\boldsymbol{\nu} \leq 0, \quad \forall \boldsymbol{\nu} \in N_X(\mathbf{x}) \right\}.$$

Further, we let

$$\mathtt{proj}(\mathbf{x}, S) = \arg\min_{\mathbf{y}\in S} \|\mathbf{y} - \mathbf{x}\| \qquad \text{and} \qquad \mathtt{dist}(\mathbf{x}, S) = \min_{\mathbf{y}\in S} \|\mathbf{y} - \mathbf{x}\|$$

denote the *Euclidean projection* of the vector $\mathbf{x}$ onto the set S, and the *Euclidean distance* from the point $\mathbf{x}$ to $\mathtt{proj}(\mathbf{x}, S)$, respectively. The function $\mathtt{dist}(\cdot, S) : \Re^n \mapsto \Re_+$ is convex and continuous.

Optimality conditions for the program [P] and properties of its solution set are given in the next proposition (Hiriart-Urruty and Lemaréchal, 1993, Thm. VII:1.1.1 and Cond. VII:(1.1.3), and Le. 2 in Burke and Ferris, 1991).

Proposition 2.1. (Optimality conditions) $\mathbf{x} \in X^*$ *if and only if there exists a* $\boldsymbol{\gamma} \in \partial f(\mathbf{x})$ *such that* $\mathtt{proj}(\boldsymbol{\gamma}, -T_X(\mathbf{x})) = \mathbf{0}$ *or, equivalently,* $\partial f(\mathbf{x}) \cap -N_X(\mathbf{x}) \neq \emptyset$. *Further, the composite mapping* $(\partial f) \cap (-N_X)$ *is constant on* X^* *and the set* $\partial f(\mathbf{x}) \cap -N_X(\mathbf{x})$, $\mathbf{x} \in X^*$, *is convex and compact.*

We consider *subgradient optimization methods* for solving the program [P]. These methods originate in research made by N.Z. Shor in the early 60's; see Shor (1991) for a review of the early history of nonsmooth optimization. These methods are conceptually simple, in general easy to implement, and have been shown to be practically useful. They have therefore been frequently applied, particularly in connection with Lagrangean duality (e.g., Fisher, 1985, and Beasley, 1993).

The ergodic convergence results to be presented are derived for the conditional subgradient optimization method (Larsson et al., 1996a), which generalizes the traditional subgradient optimization method in the sense that the feasible set X may be taken into account when determining the step direction. It was shown (ibid.) that the practical performance of the resulting method applied to three types of structured optimization models is better than that of the classical approach.

The *conditional subdifferential* of f with respect to the feasible set X at $\mathbf{x} \in X$ (see Dem'janov and Šomesova, 1978) is the nonempty, closed and convex set

$$\partial_X f(\mathbf{x}) = \left\{ \boldsymbol{\gamma} \in \Re^n \;\middle|\; f(\mathbf{y}) \geq f(\mathbf{x}) + \boldsymbol{\gamma}^T(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in X \right\},$$

the elements of which will be referred to as *conditional subgradients*. Clearly, for all $\mathbf{x} \in X$, $\partial_X f(\mathbf{x}) \supseteq \partial f(\mathbf{x})$ holds. Dem'janov and Šomesova show that $\partial_X f(\mathbf{x}) = \partial f(\mathbf{x}) + N_X(\mathbf{x})$ holds for all $\mathbf{x} \in X$.

Given an initial iterate $\mathbf{x}^0 \in X$, the *conditional subgradient optimization* method (Larsson et al., 1996a) is defined by the iteration formula

$$\mathbf{x}^{t+\frac{1}{2}} = \mathbf{x}^t - \alpha_t \boldsymbol{\gamma}_X^t, \qquad \mathbf{x}^{t+1} = \mathtt{proj}\left(\mathbf{x}^{t+\frac{1}{2}}, X\right), \qquad t = 0, 1, \ldots, \tag{3}$$

where $\alpha_t > 0$ and $\boldsymbol{\gamma}_X^t$ is a conditional subgradient of f at $\mathbf{x}^t$, that is, $\boldsymbol{\gamma}_X^t = \boldsymbol{\gamma}^t + \boldsymbol{\nu}^t$, with $\boldsymbol{\gamma}^t \in \partial f(\mathbf{x}^t)$ and $\boldsymbol{\nu}^t \in N_X(\mathbf{x}^t)$. Note that the vector $\mathbf{x}^{t+\frac{1}{2}} - \mathbf{x}^{t+1} \in N_X(\mathbf{x}^{t+1})$. Clearly, if $\{\boldsymbol{\nu}^t\} = \{\mathbf{0}\}$, then the method (3) reduces to the classical method.

Choosing the normal vector $\boldsymbol{\nu}^t = \mathtt{proj}(-\boldsymbol{\gamma}^t, N_X(\mathbf{x}^t))$ results in the conditional subgradient $\boldsymbol{\gamma}_X^t = \mathtt{proj}(\boldsymbol{\gamma}^t, -T_X(\mathbf{x}^t))$ (see Larsson et al., 1996a, Le. 3.2), defining the special case of the method (3) that is called the *subgradient projection* method, and for the case of a polyhedral feasible set, the step direction $-\boldsymbol{\gamma}_X^t$ from $\mathbf{x}^t$ is then feasible.

We consider step lengths satisfying the divergent series conditions

$$\alpha_t > 0, \ \forall t, \qquad \{\alpha_t\} \to 0, \qquad \sum_{t=0}^{\infty} \alpha_t = \infty, \qquad \text{and} \qquad \sum_{t=0}^{\infty} \alpha_t^2 < \infty. \tag{4}$$

The next result was established in Larsson et al. (1996a, Thm. 2.7).

Proposition 2.2. (Convergence of conditional subgradient optimization) *Let the method (3)–(4) be applied to the program* [P]. *If the sequences* $\{\boldsymbol{\gamma}^t\}$ *and* $\{\boldsymbol{\nu}^t\}$ *are bounded, then* $\{\mathbf{x}^t\} \to \mathbf{x}^\infty \in X^*$.

The boundedness of the sequence $\{\boldsymbol{\gamma}^t\}$ is ensured if the program [P] has some suitable additional property, for example that X is bounded or that f is piecewise linear with a finite number of pieces. In contrast, the elements of the sequence $\{\boldsymbol{\nu}^t\}$ may always be chosen to ensure its boundedness. If subgradient projection is used, then $\|\boldsymbol{\nu}^t\| \leq \|\boldsymbol{\gamma}^t\|$ so that the sequence $\{\boldsymbol{\nu}^t\}$ is bounded whenever this is the case for the sequence $\{\boldsymbol{\gamma}^t\}$. We make the blanket assumption that the sequences $\{\boldsymbol{\gamma}^t\}$ and $\{\boldsymbol{\nu}^t\}$ are bounded.

Although the step length condition (4) may seem to limit the choices of step lengths, it in fact allows for the use of almost any technique for step length selection, such as line searches and formulas involving estimates of the optimal value. This possibility rests on the observation that *any* sequence of step lengths chosen satisfies (4), whenever there *exist* two sequences of step lengths satisfying (4) which (respectively) majorizes and minorizes this sequence. This proof technique is known as the *almost complete relaxation* strategy (Dem'yanov and Vasil'ev, 1985, Sec. 3.4; see also Cor. 2.8 in Larsson et al., 1996a), and is reminiscent of the celebrated Spacer Step Theorem (e.g., Luenberger, 1984, p. 231).

We define the sequence $\{\mathbf{g}^t\}$ of *weighted averages* of subgradients used up to iteration t as

$$\mathbf{g}^t = \sum_{s=0}^{t-1} \frac{\alpha_s}{\sum_{r=0}^{t-1} \alpha_r} \boldsymbol{\gamma}^s, \qquad t = 1, 2, \ldots.$$

The central result of the paper (Thm. 3.9) is that the elements of the *ergodic sequence* $\{\mathbf{g}^t\}$ *of subgradients* in the limit satisfy the optimality conditions stated in Prop. 2.1.

We note that the construction of the ergodic sequence of subgradients is an example of a widely used technique for inducing convergence properties which are lacking in an original sequence. The typical situation where an averaging process is necessary is when the original sequence oscillates or contains stochastic noise; the averaging of the elements then has the effect of stabilizing the process by smoothing out the oscillations and filtering out the stochastic noise, respectively. For example applications, see Baillon (1975), Ruszczyński and Syski (1983), Higle and Sen (1991), Kim and Ahn (1991), Larsson and Liu (1997), and Petersson and Patriksson (1997).

The rest of the paper is organized as follows. In Sec. 3 we define ergodic sequences of subgradients, normal cone elements, and objective values generated by the method (3)–(4), and establish the core result of this work that these sequences verify optimality. In Sec. 4 we utilize the ergodic sequence of subgradients to finitely identify the face (of smallest dimension) of the feasible set that contains this optimal solution. Sec. 5 contains convergence results for ergodic sequences of Lagrange multipliers and multipliers associated with pieces of the objective function. These results are then used to derive a sensitivity analysis result for nonsmooth programming and to show a convergence result for solutions of systems of linear inequalities in the case of inconsistency. In Sec. 6 we present a convergent lower bounding procedure based on the minimization of affine underestimates of the nonsmooth objective function; this procedure provides a proper termination criterion for subgradient methods. An application to systems of linear inequalities is also presented. Finally, in Sec. 7, we discuss briefly some opportunities for future research.

Earlier versions of Secs. 1, 3, 5, and 6 have been published in Larsson et al. (1996b and 1996c). An expanded version of this paper, including a larger number of applications, appears in Strömberg (1997).

3 MAIN THEORETICAL RESULTS

3.1 Preliminaries

We denote by A_t the cumulative step length up to iteration t, namely,

$$A_t = \sum_{s=0}^{t-1} \alpha_s, \qquad t = 1, 2, \ldots. \tag{5}$$

Definition 3.1. The *ergodic sequence of subgradients* is defined by

$$\mathbf{g}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \boldsymbol{\gamma}^s, \qquad t = 1, 2, \ldots.$$

The *ergodic sequences* $\{\mathbf{n}_c^t\}$ *and* $\{\mathbf{n}_p^t\}$ *of normal terms of the conditional subgradients and projection steps* are defined by $\mathbf{n}_c^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \boldsymbol{\nu}^s$ and $\mathbf{n}_p^t = A_t^{-1} \sum_{s=0}^{t-1} (\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1})$, respectively, and the *ergodic sequence of normal vectors* is defined by

$$\mathbf{n}^t = \mathbf{n}_c^t + \mathbf{n}_p^t, \qquad t = 1, 2, \ldots.$$

The behaviour of the sequences $\{\mathbf{g}^t\}$ and $\{\mathbf{n}^t\}$ play a central role in the derivation of the results of this paper. A main source of inspiration for the study of these sequences was Shor's (1985, pp. 116–118) study of Lagrangean relaxation applied to linear programs. Clearly, each vector $\mathbf{g}^t$ is a convex combination of the subgradients $\{\boldsymbol{\gamma}^s\}_{s=0}^{t-1}$, and $\mathbf{n}_c^t$ is a convex combination of the normal terms of the conditional subgradients used up to iteration t of the method (3), while $\mathbf{n}_p^t$ is a positive linear combination of its projection steps. If the traditional subgradient optimization method is employed, then for all t, $\mathbf{n}_c^t = \mathbf{0}$ and $\mathbf{n}^t = \mathbf{n}_p^t$. A recursive computation of $\mathbf{g}^t$ can be implemented using the formula

$$\mathbf{g}^1 = \boldsymbol{\gamma}^0, \qquad \mathbf{g}^t = \frac{A_{t-1}}{A_t}\mathbf{g}^{t-1} + \frac{\alpha_{t-1}}{A_t}\boldsymbol{\gamma}^{t-1}, \qquad t = 2, 3, \ldots.$$

Remark 3.2. We will utilize the properties in the *limit* of the sequence $\{\mathbf{g}^t\}$. Since the information from the initial iterations of a subgradient optimization scheme is usually of low quality, the initialization of this sequence is preferably delayed until the iterates $\mathbf{x}^t$ are near-optimal. Any of the results of this paper can easily be modified to the initialization of the ergodic sequences at any iteration $t_0 \geq 0$.

The next result, which is a consequence of Def. 3.1 and the inequality $\|\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1}\| \leq \alpha_s \|\boldsymbol{\gamma}_X^s\|$, is the first step towards a convergence result for the ergodic sequence $\{\mathbf{g}^t\}$.

Lemma 3.3. *The sequences* $\{\mathbf{g}^t\}$ *and* $\{\mathbf{n}^t\}$ *are bounded.*

The ergodic sequence of objective values obtained in the method (3) is defined as

$$\varphi_t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s f(\mathbf{x}^s), \qquad t = 1, 2, \dots. \tag{6}$$

It follows that $\varphi_t \geq f^*$ for all t, since, for all s, $\mathbf{x}^s \in X$.

Utilizing that, for all s, $\boldsymbol{\gamma}^s \in \partial f(\mathbf{x}^s)$, and the definition (1), we derive the *surrogate inequalities*

$$\sum_{s=0}^{t-1} \alpha_s f(\mathbf{x}) \geq \sum_{s=0}^{t-1} \alpha_s \left(f(\mathbf{x}^s) + (\boldsymbol{\gamma}^s)^T (\mathbf{x} - \mathbf{x}^s) \right), \qquad \mathbf{x} \in \Re^n, \qquad t = 1, 2, \dots,$$

and define the *affine* functions $f_t : \Re^n \mapsto \Re$, $t = 1, 2, \dots$, as

$$f_t(\mathbf{x}) = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \left(f(\mathbf{x}^s) + (\boldsymbol{\gamma}^s)^T (\mathbf{x} - \mathbf{x}^s) \right), \qquad \mathbf{x} \in \Re^n. \tag{7}$$

Then, $f_t(\mathbf{x}) \leq f(\mathbf{x})$ holds for all $\mathbf{x} \in \Re^n$ and all t, that is, each function f_t is an *affine underestimate* of the objective f (see Fig. 1).

[Place of Figure 1]

Clearly, the hyperplane defined by the function f_t is a weighted average of all supporting hyperplanes generated in the previous iterations of the conditional subgradient optimization method. From Def. 3.1 it follows that $\nabla f_t(\mathbf{x}) = \mathbf{g}^t$ and $f_t(\mathbf{x}) = f_t(\mathbf{y}) + (\mathbf{g}^t)^T (\mathbf{x} - \mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \Re^n$ and all t.

We let the convex functions $\delta_t : \Re^n \mapsto \Re$, $t = 1, 2, \dots$, describe the pointwise vertical distance between the functions f and f_t, that is,

$$\delta_t(\mathbf{x}) = f(\mathbf{x}) - f_t(\mathbf{x}), \qquad \mathbf{x} \in \Re^n. \tag{8}$$

It follows that $\delta_t(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \Re^n$ and all t.

Further, we define the affine functions $\varrho_t : \Re^n \mapsto \Re$, $t = 1, 2, \dots$, as

$$\varrho_t(\mathbf{x}) = A_t^{-1} \sum_{s=0}^{t-1} \left(\alpha_s (\boldsymbol{\nu}^s)^T (\mathbf{x}^s - \mathbf{x}) + \left(\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1} \right)^T (\mathbf{x}^{s+1} - \mathbf{x}) \right), \quad \mathbf{x} \in \Re^n. \tag{9}$$

Using, for all s, that $\boldsymbol{\nu}^s \in N_X(\mathbf{x}^s)$ and $\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1} \in N_X(\mathbf{x}^{s+1})$, and the definition (2), it follows that $\varrho_t(\mathbf{x}) \geq 0$ for $\mathbf{x} \in X$ and all t, that is, each of these inequalities is a *valid inequality* for the set X (see Fig. 2).

[Place of Figure 2]

From Def. 3.1 follows that $\nabla \varrho_t(\mathbf{x}) = -\mathbf{n}^t$ and $\varrho_t(\mathbf{x}) = \varrho_t(\mathbf{y}) - (\mathbf{n}^t)^T (\mathbf{x} - \mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \Re^n$ and all t.

The following lemma is crucial for our continued analysis.

Lemma 3.4. *$\{\delta_t(\mathbf{x}^\infty)\} \to 0$, $\{\varrho_t(\mathbf{x}^\infty)\} \to 0$, and $\{\varphi_t\} \to f^*$. Further, $\{\delta_t(\mathbf{x}^t)\} \to 0$ and $\{\varrho_t(\mathbf{x}^t)\} \to 0$.*

Proof. By Prop. 2.2, $\{\mathbf{x}^t\} \to \mathbf{x}^\infty$. From the iteration formula (3) it follows that

$$\begin{aligned} \left\|\mathbf{x}^{t+1} - \mathbf{x}^\infty\right\|^2 \quad &\leq \quad \left\|\mathbf{x}^t - \mathbf{x}^\infty\right\|^2 + \alpha_t^2 \left\|\boldsymbol{\gamma}_X^t\right\|^2 \\ &\quad -2\left(\alpha_t(\boldsymbol{\gamma}_X^t)^T\left(\mathbf{x}^t - \mathbf{x}^\infty\right) + \left(\mathbf{x}^{t+\frac{1}{2}} - \mathbf{x}^{t+1}\right)^T\left(\mathbf{x}^{t+1} - \mathbf{x}^\infty\right)\right). \end{aligned}$$

Repeated application of the inequality and utilization of the definitions (8), (9), and (6), result in $\|\mathbf{x}^0 - \mathbf{x}^\infty\|^2 + \sum_{s=0}^{t-1}\alpha_s^2\|\boldsymbol{\gamma}_X^s\|^2 - 2A_t(\delta_t(\mathbf{x}^\infty) + \varrho_t(\mathbf{x}^\infty) + \varphi_t - f^*) \geq 0$. Since $\delta_t(\mathbf{x}^\infty) \geq 0$, $\varrho_t(\mathbf{x}^\infty) \geq 0$, $\varphi_t \geq f^*$, and $A_t > 0$, the immediate result is that

$$0 \leq \delta_t(\mathbf{x}^\infty) + \varrho_t(\mathbf{x}^\infty) + \varphi_t - f^* \leq \frac{1}{2A_t}\left(\left\|\mathbf{x}^0 - \mathbf{x}^\infty\right\|^2 + \sum_{s=0}^{t-1}\alpha_s^2\left\|\boldsymbol{\gamma}^s + \boldsymbol{\nu}^s\right\|^2\right).$$

Let $t \to \infty$ and invoke the conditions (4).

For the latter result, we note that Definition 3.1, and the definitions (7) and (8), yield

$$0 \leq \delta_t(\mathbf{x}^t) = \delta_t(\mathbf{x}^\infty) + f(\mathbf{x}^t) - f(\mathbf{x}^\infty) - \left(\mathbf{g}^t\right)^T\left(\mathbf{x}^t - \mathbf{x}^\infty\right), \quad t = 1, 2, \ldots.$$

From Definition 3.1, and the definition (9), follow that

$$0 \leq \varrho_t(\mathbf{x}^t) = \varrho_t(\mathbf{x}^\infty) - \left(\mathbf{n}^t\right)^T\left(\mathbf{x}^t - \mathbf{x}^\infty\right), \quad t = 1, 2, \ldots.$$

Proposition 2.2, Lemma 3.3, the continuity of f, and Lemma 3.4 yield that the right–hand–sides of both the above equations tend toward zero as t approaches infinity. The result follows. ◻

A small value of $\delta_t(\mathbf{x}^t) \geq 0$ implies that the affine underestimate f_t of the function f defines a hyperplane that is close to a support to $\mathtt{epi}\, f$ at $(\mathbf{x}^t, f(\mathbf{x}^t))$, where $\mathtt{epi}$ denotes epigraph. A small value of $\varrho_t(\mathbf{x}^t)$ means that the hyperplane $\{\mathbf{x} \in \Re^n \mid \varrho_t(\mathbf{x}) = 0\}$ is close to a support to X at $\mathbf{x}^t$ (while $X \subseteq \{\mathbf{x} \in \Re^n \mid \varrho_t(\mathbf{x}) \geq 0\}$ always holds).

Remark 3.5. Should, for example, $\delta_t(\mathbf{x}^t) = 0$ hold at some iteration t, one can, however, not infer the optimality of $\mathbf{x}^t$. An analogous statement applies to the quantity $\mathbf{g}^t + \mathbf{n}^t$, which is utilized next.

We will also utilize the following lemma in our continued analysis, when proving optimality fulfilment of the sequence $\{\mathbf{g}^t\}$ in the limit.

Lemma 3.6. *$\{\mathbf{g}^t + \mathbf{n}^t\} \to \mathbf{0}$.*

Proof. By Def. 3.1 and the iteration formula (3), $\mathbf{g}^t + \mathbf{n}^t = A_t^{-1}\sum_{s=0}^{t-1}(\mathbf{x}^s - \mathbf{x}^{s+\frac{1}{2}} + \mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1}) = A_t^{-1}(\mathbf{x}^0 - \mathbf{x}^t)$. Prop. 2.2 yields that $\{\mathbf{x}^t\} \to \mathbf{x}^\infty$. The result then follows from the definition (5) and the condition (4). ◻

3.2 Ergodic convergence of the sequence of subgradients

As concluded above, the affine functions f_t, $t = 1, 2, \ldots$, underestimate the convex function f everywhere, and from Le. 3.4 it follows that $\{f_t(\mathbf{x}^\infty)\} \to f(\mathbf{x}^\infty)$. Hence, in the limit, the sequence $\{f_t\}$ provides supporting hyperplanes to $\mathtt{epi}\, f$ at $(\mathbf{x}^\infty, f(\mathbf{x}^\infty))$; this observation is further developed in the following theorem.

Theorem 3.7 (Convergence of $\{\mathbf{g}^t\}$ to $\partial f(\mathbf{x}^\infty)$). $\{\mathtt{dist}(\mathbf{g}^t, \partial f(\mathbf{x}^\infty))\} \to 0$.

Proof. Def. 3.1 and the definitions (7) and (8) imply that $\mathbf{g}^t$ is a $\delta_t(\mathbf{x})$-subgradient of f at $\mathbf{x}$ (e.g., Hiriart-Urruty and Lemaréchal, 1993, Def. XI:1.1.1) for any $\mathbf{x} \in \Re^n$; applying this result to $\mathbf{x} = \mathbf{x}^t$ yields

$$f(\mathbf{y}) \geq f(\mathbf{x}^t) + (\mathbf{g}^t)^T(\mathbf{y} - \mathbf{x}^t) - \delta_t(\mathbf{x}^t), \quad \forall \mathbf{y} \in \Re^n.$$

By Le. 3.3, the sequence $\{\mathbf{g}^t\}$ is bounded. Let $\widetilde{\mathbf{g}}$ be an accumulation point of a convergent subsequence $\mathcal{T}$. Then, by Le. 3.4, in the limit of $\mathcal{T}$ of the above inequality, we obtain that $\widetilde{\mathbf{g}} \in \partial f(\mathbf{x}^\infty)$. The boundedness of $\partial f(\mathbf{x}^\infty)$ then yields the desired result. □

Theorem 3.8 (Convergence of $\{\mathbf{g}^t\}$ to $-N_X(\mathbf{x}^\infty)$). $\{\mathtt{dist}(\mathbf{g}^t, -N_X(\mathbf{x}^\infty))\} \to 0$.

Proof. From the definition (9) it follows that, for all t and any $\mathbf{y} \in X$, $(\mathbf{n}^t)^T(\mathbf{y} - \mathbf{x}^\infty) = -\varrho_t(\mathbf{y}) + \varrho_t(\mathbf{x}^\infty) \leq \varrho_t(\mathbf{x}^\infty)$. By Le. 3.3, the sequence $\{\mathbf{n}^t\}$ is bounded. Let $\widetilde{\mathbf{n}}$ be an accumulation point of the sequence $\{\mathbf{n}^t\}_{t \in \mathcal{T}}$ for some convergent subsequence $\mathcal{T}$. From Le. 3.4 it then follows that $\widetilde{\mathbf{n}}^T(\mathbf{y} - \mathbf{x}^\infty) \leq 0$ for any $\mathbf{y} \in X$. By the continuity of the function $\mathtt{dist}(\cdot, N_X(\mathbf{x}^\infty))$, for any $\varepsilon > 0$ and all t that are sufficiently large, $\mathtt{dist}(\mathbf{n}^t, N_X(\mathbf{x}^\infty)) < \varepsilon/2$; further, by Le. 3.6, $\|\mathbf{g}^t + \mathbf{n}^t\| < \varepsilon/2$. This yields that $\mathtt{dist}(\mathbf{g}^t, -N_X(\mathbf{x}^\infty)) \leq \mathtt{dist}(-\mathbf{n}^t, -N_X(\mathbf{x}^\infty)) + \|\mathbf{g}^t + \mathbf{n}^t\| < \varepsilon$, and the result follows. □

Applying the Moreau decomposition (Zarantonello, 1971, Le. 2.2) to $\mathbf{g}^t$, with respect to the polar cones $N_X(\mathbf{x}^\infty)$ and $T_X(\mathbf{x}^\infty)$, it can easily be established that $\{\mathtt{proj}(\mathbf{g}^t, -T_X(\mathbf{x}^\infty))\} \to \mathbf{0}$ holds.

We go on to our main result, which establishes that the sequence $\{\mathbf{g}^t\}$ in the limit provides subgradients that verify the optimality of the solution $\mathbf{x}^\infty$ to the program [P]; this follows from Thms. 3.7 and 3.8.

Theorem 3.9 (Optimality fulfilment). $\{\mathtt{dist}(\mathbf{g}^t, \partial f(\mathbf{x}^\infty) \cap -N_X(\mathbf{x}^\infty))\} \to 0$. *Further, if* $X = \Re^n$, *then,* $\{\mathbf{g}^t\} \to \mathbf{0}$.

This result and Le. 3.6 imply similar results for the sequence $\{\mathbf{n}^t\}$.

Thm. 3.9 shows that the ergodic sequence $\{\mathbf{g}^t\}$ of subgradients inherits important properties from sequences of gradients of smooth functions.

4 FINITENESS RESULTS

In this section we demonstrate that the ergodic sequence $\{\mathbf{g}^t\}$ can be used to finitely determine a nontrivial face of the feasible set containing the limit point $\mathbf{x}^\infty$ and in particular the constraints that are active at $\mathbf{x}^\infty$. This is achieved through the construction of an auxiliary sequence of feasible solutions that exhibits convergence properties that the original sequence $\{\mathbf{x}^t\}$ is lacking.

4.1 Preliminaries

The auxiliary sequence of feasible solutions to the program [P] is constructed in a manner akin to that of the gradient projection algorithm (see, for example, Bertsekas, 1976), the elements $\mathbf{g}^t$ taking the role of the gradient values.

Definition 4.1. Let the sequence $\{\sigma_t\}$ satisfy $\inf_t\{\sigma_t\} > 0$ and $\sup_t\{\sigma_t\} < \infty$. The *auxiliary sequence* $\{\overline{\mathbf{x}}^t\}$ is defined by $\overline{\mathbf{x}}^t = \mathtt{proj}\,(\mathbf{x}^t - \sigma_t \mathbf{g}^t, X)$.

Theorem 4.2. $\{\overline{\mathbf{x}}^t\} \to \mathbf{x}^\infty$.

Proof. The existence of a limit $\mathbf{x}^\infty$ of $\{\mathbf{x}^t\}$ that solves [P] is a consequence of Prop. 2.2. By the relations $\|\overline{\mathbf{x}}^t - \mathbf{x}^t\| = \|\mathtt{proj}(\mathbf{x}^t - \sigma_t\mathbf{g}^t, X) - \mathbf{x}^t\| \le \|\mathbf{x}^t - \sigma_t\mathbf{g}^t - \mathbf{x}^t\| = \sigma_t\|\mathbf{g}^t\|$, and the boundedness of the sequences $\{\mathbf{x}^t\}$, $\{\mathbf{g}^t\}$ and $\{\sigma_t\}$, the sequence $\{\overline{\mathbf{x}}^t\}$ is also bounded. The result then follows from Thm. 3.8 and the properties that $\mathtt{proj}(\cdot, X)$ is continuous and $\mathtt{proj}(\mathbf{x} + \mathbf{d}, X) = \mathbf{x}$ holds for any $\mathbf{d} \in N_X(\mathbf{x})$, $\mathbf{x} \in X$ (see also Equation (3.3) in Burke and Moré, 1994). □

The gradient projection algorithm for differentiable optimization enjoys finite identification properties; these have been investigated in Calamai and Moré (1987), Dunn (1987), and Burke and Moré (1988 and 1994). We will subsequently show that natural extensions of these results to the nonsmooth case hold for the auxiliary sequence $\{\overline{\mathbf{x}}^t\}$. Throughout this section we shall frequently utilize the notion of *exposed faces*.

Definition 4.3. The *face* of the closed and convex set $S \subseteq \Re^n$ that is *exposed* by a vector $\mathbf{d} \in \Re^n$ is $F_S(\mathbf{d}) = \arg\max_{\mathbf{y}\in S}\{\mathbf{d}^T\mathbf{y}\}$.

Exposed faces are related to normal cones through the equivalence relation

$$\mathbf{d} \in N_S(\mathbf{x}) \Longleftrightarrow \mathbf{x} \in F_S(\mathbf{d}). \tag{10}$$

Our first result establishes that for polyhedral feasible sets, given an accumulation point $\widetilde{\mathbf{g}}$ of the sequence $\{\mathbf{g}^t\}$, the auxiliary points $\overline{\mathbf{x}}^t$ are finitely in the face of X that is exposed by $-\widetilde{\mathbf{g}}$.

Theorem 4.4 ($\overline{\mathbf{x}}^t$ is finitely in the face exposed by $-\widetilde{\mathbf{g}}$). *Assume that the set X is polyhedral, let the assumptions of Thm. 4.2 hold and $\widetilde{\mathbf{g}}$ be the limit of the convergent subsequence $\{\mathbf{g}^t\}_{t\in\mathcal{T}}$. Then, for some t_0, $\overline{\mathbf{x}}^t \in F_X(-\widetilde{\mathbf{g}})$, for all $t \ge t_0$, $t \in \mathcal{T}$.*

Proof. Defining $\mathbf{d}^t = \mathbf{x}^t - \sigma_t\mathbf{g}^t - \overline{\mathbf{x}}^t$, it follows that $\mathbf{d}^t \in N_X(\overline{\mathbf{x}}^t)$ holds for all t. From the equivalence (10) it then follows that $\overline{\mathbf{x}}^t \in F_X(\mathbf{d}^t)$ for all t. By Thm. 4.2 we may assume that $\{\mathbf{d}^t\}_{t\in\mathcal{T}} \to -\widetilde{\sigma}\widetilde{\mathbf{g}}$ for some $\widetilde{\sigma} > 0$. Le. 3.5 in Robinson (1984) (see also Burke and Moré, 1994, Thm. 3.1) implies that $F_X(\mathbf{d}) \subseteq F_X(\mathbf{d}^*)$ for any $\mathbf{d}$ in a sufficiently small neighbourhood of $\mathbf{d}^* \in \Re^n$. Applying this result to the above yields the theorem. □

A direct consequence of Thm. 4.4 and the Moreau decomposition (Zarantonello, 1971, Le. 2.2) is the result that $\{\mathtt{proj}(\mathbf{g}^t, -T_X(\overline{\mathbf{x}}^t))\} \to \mathbf{0}$.

4.2 Non-degeneracy

Faces of polyhedral sets enjoy several properties that are not shared with faces of general convex sets; for example, every face of a polyhedral set is exposed by some vector in $\Re^n$, which is not the case for general convex sets. Establishing finite identification results for non-polyhedral sets therefore require certain non-degeneracy assumptions on the feasible set and the optimal solution obtained by the algorithm under consideration.

It is well known (e.g., Burke and Moré, 1988) that every point in a convex set X belongs to the relative interior of a *unique* face of X. We let the face F_X^∞ be defined by the relation

$$\texttt{rint}\ F_X^\infty \ni \mathbf{x}^\infty, \tag{11}$$

where $\mathbf{x}^\infty$ is the limit point of the sequence $\{\mathbf{x}^t\}$ generated by the conditional subgradient optimization algorithm. One of the assumptions employed in this section is that the face F_X^∞ is *quasi-polyhedral,* as defined below; this definition is due to Burke and Moré (1988). Here, $\texttt{aff}$ denotes the affine hull.

Definition 4.5. A face F_X of a convex set X is *quasi-polyhedral* if $\texttt{aff}\ F_X = \mathbf{x} + \{T_X(\mathbf{x}) \cap -T_X(\mathbf{x})\}$ for all $\mathbf{x} \in \texttt{rint}\ F_X$.

Note that every face of a polyhedral set is quasi-polyhedral. For detailed analyses of the geometry of faces, see Burke and Moré (1988 and 1994). We will further employ a non-degeneracy assumption, which extends the one introduced by Dunn (1981) for the differentiable case.

Assumption 4.6. $-\widetilde{G} \subseteq \texttt{rint}\ N_X(\mathbf{x}^\infty)$.

To discuss the meaning of this assumption, we represent the set X as $X = \{\mathbf{x} \in \Re^n \mid a_i(\mathbf{x}) \leq 0,\ i \in \mathcal{I}\}$, where the functions $a_i : \Re^n \mapsto \Re$, $i \in \mathcal{I}$, are convex and differentiable (and the set $\mathcal{I}$ is finite) and denote by $\mathcal{I}(\mathbf{x}) = \{\, i \in \mathcal{I} \mid a_i(\mathbf{x}) = 0 \,\}$ the set of constraints that are active at $\mathbf{x} \in X$. The assumption, together with the Guignard (1969) *constraint qualification* (CQ) for X, is equivalent to the existence of a set of strictly complementary Lagrange multipliers for any accumulation point of $\{\mathbf{g}^t\}$; to see this, we note that under the Guignard CQ,

$$N_X(\mathbf{x}) = \left\{ \boldsymbol{\nu} \in \Re^n \;\middle|\; \boldsymbol{\nu} = \sum_{i \in \mathcal{I}(\mathbf{x})} \mu_i \nabla a_i(\mathbf{x}); \quad \mu_i \geq 0,\ i \in \mathcal{I} \right\}.$$

An equivalent expression of Ass. 4.6 is then that for every accumulation point $\widetilde{\mathbf{g}}$ of the sequence $\{\mathbf{g}^t\}$ there is a vector $\boldsymbol{\mu} \in \Re_+^{|\mathcal{I}|}$ of Lagrange multipliers such that (cf. Le. 3.2 of Burke and Moré, 1988)

$$-\widetilde{\mathbf{g}} = \sum_{i \in \mathcal{I}(\mathbf{x}^\infty)} \mu_i \nabla a_i(\mathbf{x}^\infty) \qquad \text{and} \qquad \mu_i > 0,\ i \in \mathcal{I}(\mathbf{x}^\infty).$$

We now establish that the auxiliary sequence $\{\overline{\mathbf{x}}^t\}$ finitely enters and remains in the face F_X^∞, defined by the relation (11); this result strengthens that of Thm. 4.4.

Theorem 4.7 ($\overline{\mathbf{x}}^t$ is finitely in the optimal face of X). *Assume that the Guignard CQ holds for the set X, and that the face F_X^∞, given by the relation (11), is quasi-polyhedral. Further, let Ass. 4.6 and the assumptions of Thm. 4.2 hold. Then, for some t_0, $\overline{\mathbf{x}}^t \in \mathtt{rint}\ F_X^\infty$, $t \geq t_0$.*

Proof. Let the sequence $\{\mathbf{d}^t\}$ be defined as in the proof of Thm. 4.4, and let $\widetilde{\mathbf{g}}$ be the limit of the convergent subsequence $\{\mathbf{g}^t\}_{t\in\mathcal{T}}$. By Thm. 4.2 we may assume that $\{\mathbf{d}^t\}_{t\in\mathcal{T}} \to -\widetilde{\sigma}\widetilde{\mathbf{g}}$ for some $\widetilde{\sigma} > 0$. Thm. 4.2 and Ass. 4.6 then enable us to invoke Le. 3.3 of Burke and Moré (1988) to arrive at the result that $\overline{\mathbf{x}}^t \in \mathtt{rint}\, F_X^\infty$ for all $t \in \mathcal{T}$ that are sufficiently large. But since, by Ass. 4.6 and Thm. 2.5 of Burke and Moré (1994), all vectors $\widetilde{\mathbf{g}} \in \widetilde{G}$ expose the same face of X, the result follows. □

A direct consequence of this result is that $\{\mathtt{proj}(\mathbf{g}^t, -T_X(\overline{\mathbf{x}}^t))\} \to \mathbf{0}$.

For *linearly constrained* problems, where $X = \{\, \mathbf{x} \in \Re^n \mid \mathbf{a}_i^T\mathbf{x} \leq b_i,\ i \in \mathcal{I}\,\}$, the result of Thm. 4.7 can be expressed in terms of active constraints.

Theorem 4.8 ($\overline{\mathbf{x}}^t$ identifies finitely the active constraints at $\mathbf{x}^\infty$). *Let the set X be polyhedral and the assumptions of Thm. 4.7 hold. Then, for some t_0, $\mathcal{I}(\overline{\mathbf{x}}^t) = \mathcal{I}(\mathbf{x}^\infty)$, for all $t \geq t_0$.*

Proof. The result follows from Thm. 4.7 and the fact that [cf. Burke and Moré, 1988, equation (2.7)] $\mathtt{rint}\ F_X^\infty = \{\mathbf{x} \in X \mid \mathbf{a}_i^T\mathbf{x} = b_i,\ i \in \mathcal{I}(\mathbf{x}^\infty);\ \mathbf{a}_i^T\mathbf{x} < b_i,\ i \notin \mathcal{I}(\mathbf{x}^\infty)\}$. □

Under a stronger non-degeneracy assumption than Ass. 4.6, the auxiliary sequence $\{\overline{\mathbf{x}}^t\}$ exhibits finite convergence to a solution to [P].

Definition 4.9. An solution $\mathbf{x}^*$ to [P] is a *sharp minimum* of f on X if, for some constant $\delta_f > 0$, $f(\mathbf{x}) \geq f^* + \delta_f\, \|\mathbf{x} - \mathbf{x}^*\|$ for all $\mathbf{x} \in X$.

This definition is due to Polyak (1987, p. 205). We note that a sharp minimum is unique. A point $\mathbf{x}^*$ is a sharp minimum of f on X if and only if $\mathbf{0} \in \mathtt{int}(\partial f(\mathbf{x}^*) + N_X(\mathbf{x}^*))$; the assumption made below is that $-\widetilde{G} \subset \mathtt{int}\, N_X(\mathbf{x}^\infty)$, which is a strengthening of Ass. 4.6 in the sense that the interior of $N_X(\mathbf{x}^\infty)$ is presumed to be nonempty.

The following result strengthens those of Thms. 4.7 and 4.8.

Corollary 4.10 ($\overline{\mathbf{x}}^t$ identifies finitely the optimal solution). *Let the assumptions of Thm. 4.7 hold. If $\mathtt{int}\, N_X(\mathbf{x}^\infty) \neq \emptyset$, then the sequence $\{\overline{\mathbf{x}}^t\}$ converges finitely to $\mathbf{x}^\infty$.*

Proof. If $N_X(\mathbf{x}^\infty)$ has a nonempty interior, then the set $\{\mathbf{x}^\infty\}$ is a quasi-polyhedral face of X (Burke and Moré, 1988, proof of Cor. 3.5). Then, Thm. 4.7 shows that $\overline{\mathbf{x}}^t = \mathbf{x}^\infty$ for all sufficiently large t. □

5 ERGODIC CONVERGENCE OF SEQUENCES OF MULTIPLIERS

5.1 Multipliers of constraints

Throughout this subsection we assume that the feasible set X of the program [P] is explicitly described as

$$X = \left\{\, \mathbf{x} \in \Re^n \;\middle|\; a_i(\mathbf{x}) \leq 0,\ i \in \mathcal{I};\quad \mathbf{b}_j^T\mathbf{x} = e_j,\ j \in \mathcal{E}\,\right\}, \tag{12}$$

where the functions $a_i : \Re^n \mapsto \Re$, $i \in \mathcal{I}$, are convex and differentiable, $\mathbf{b}_j \in \Re^n$ and $e_j \in \Re$, $j \in \mathcal{E}$, and the sets $\mathcal{I}$ and $\mathcal{E}$ are finite and disjoint. The convergence analysis for the sequences of multipliers associated with the constraints defining the set X relies on the assumption that this set fulfils the following strong Slater CQ (e.g., Hiriart-Urruty and Lemaréchal, 1993, Def. VII:2.3.1).

Assumption 5.1. *The vectors $\mathbf{b}_j$, $j \in \mathcal{E}$, are linearly independent, and there exists an $\mathbf{x} \in X$ such that $a_i(\mathbf{x}) < 0$, $i \in \mathcal{I}$.*

We let, for all $\mathbf{x} \in X$, $\mathcal{I}(\mathbf{x}) = \{ i \in \mathcal{I} \mid a_i(\mathbf{x}) = 0 \}$ denote the set of inequality constraints that are active at $\mathbf{x}$, and $\Pi_X(\mathbf{x})$ the set of multiplier values that satisfy the complementary slackness conditions for the program [P] at $\mathbf{x}$, that is,

$$\Pi_X(\mathbf{x}) = \left\{ \boldsymbol{\pi} \in \Re_+^{|\mathcal{I}|} \times \Re^{|\mathcal{E}|} \;\middle|\; \pi_i = 0,\ i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}) \right\} \ni \mathbf{0}.$$

Letting $\boldsymbol{\omega} \in \Re^{|\mathcal{I}|+|\mathcal{E}|}$, the normal cone of X at $\mathbf{x} \in X$ can then be expressed as

$$N_X(\mathbf{x}) = \left\{ \boldsymbol{\nu} \in \Re^n \;\middle|\; \boldsymbol{\nu} = \sum_{i \in \mathcal{I}} \omega_i \nabla a_i(\mathbf{x}) + \sum_{j \in \mathcal{E}} \omega_j \mathbf{b}_j; \quad \boldsymbol{\omega} \in \Pi_X(\mathbf{x}) \right\}.$$

For any $\mathbf{x} \in X$, let

$$M_X(\mathbf{x}) = \left\{ \boldsymbol{\mu} \in \Pi_X(\mathbf{x}) \;\middle|\; \mathbf{0} \in \partial f(\mathbf{x}) + \sum_{i \in \mathcal{I}} \mu_i \nabla a_i(\mathbf{x}) + \sum_{j \in \mathcal{E}} \mu_j \mathbf{b}_j \right\} \tag{13}$$

be the (possibly empty) set of Lagrange multipliers associated with the constraints defining the set X. The optimality conditions for the program [P] can be stated in terms of this set.

Proposition 5.2 (Optimality condition). *Let $\mathbf{x} \in X$. Then, $\mathbf{x} \in X^*$ if and only if $M_X(\mathbf{x}) \neq \emptyset$. Further, the mapping M_X is constant on X^* and the set $M_X(\mathbf{x})$, $\mathbf{x} \in X^*$, is convex and compact.*

Proof. The first conclusion follows from Hiriart-Urruty and Lemaréchal, 1993, Thms. VII:2.1.4 and VII:2.3.2. The invariance of M_X over X^* and its boundedness follow from Prop. VII:3.1.1 and Thm. VII:2.3.2, respectively, in ibid. The closedness and convexity of $M_X(\mathbf{x})$ follow from its definition (13). □

Our derivation of convergence results for the ergodic sequence of Lagrange multipliers starts with the following lemma.

Lemma 5.3. *Let $\mathbf{x} \in X$ and $\boldsymbol{\pi} \in \Re_+^{|\mathcal{I}|} \times \Re^{|\mathcal{E}|}$. Then,*

$$\sum_{i \in \mathcal{I}(\mathbf{x})} \pi_i \nabla a_i(\mathbf{x}) + \sum_{j \in \mathcal{E}} \pi_j \mathbf{b}_j = \mathbf{0}, \qquad \boldsymbol{\pi} \in \Pi_X(\mathbf{x}) \tag{14}$$

holds if and only if $\boldsymbol{\pi} = \mathbf{0}$.

Proof. Clearly, $\boldsymbol{\pi} = \mathbf{0}$ satisfies (14).

Conversely, suppose that $\boldsymbol{\pi}$ satisfies (14). From Ass. 5.1 follows that there is an $\widehat{\mathbf{x}} \neq \mathbf{x}$ such that $\widehat{\mathbf{x}} \in \mathtt{rint}\, X$. This fact together with the convexity and differentiability of the functions a_i and the fact that $a_i(\mathbf{x}) = 0$ for all $i \in \mathcal{I}(\mathbf{x})$ implies that $\nabla a_i(\mathbf{x})^T(\widehat{\mathbf{x}}-\mathbf{x}) \leq a_i(\widehat{\mathbf{x}}) < 0$ for all $i \in \mathcal{I}(\mathbf{x})$. Since $\mathbf{x}, \widehat{\mathbf{x}} \in X$ it then follows from (14) and (12) that

$$\begin{aligned} 0 &= \sum_{i\in\mathcal{I}(\mathbf{x})} \pi_i \nabla a_i(\mathbf{x})^T(\widehat{\mathbf{x}}-\mathbf{x}) + \sum_{j\in\mathcal{E}} \pi_j \mathbf{b}_j^T(\widehat{\mathbf{x}}-\mathbf{x}) \\ &= \sum_{i\in\mathcal{I}(\mathbf{x})} \pi_i \nabla a_i(\mathbf{x})^T(\widehat{\mathbf{x}}-\mathbf{x}) \leq \sum_{i\in\mathcal{I}(\mathbf{x})} \pi_i a_i(\widehat{\mathbf{x}}) \leq 0. \end{aligned}$$

Hence, $\sum_{i\in\mathcal{I}(\mathbf{x})} \pi_i a_i(\widehat{\mathbf{x}}) = 0$ and, since $\pi_i a_i(\widehat{\mathbf{x}}) \leq 0$ for all $i \in \mathcal{I}$, therefore, $\pi_i = 0$ holds for all $i \in \mathcal{I}(\mathbf{x})$. The system (14) then reduces to the equations $\sum_{j\in\mathcal{E}} \pi_j \mathbf{b}_j = \mathbf{0}$ and $\pi_i = 0$, $i \in \mathcal{I}$. By Ass. 5.1, the vectors $\mathbf{b}_j$, $j \in \mathcal{E}$, are linearly independent, thus implying that $\pi_j = 0$ for all $j \in \mathcal{E}$. Hence, $\boldsymbol{\pi} = \mathbf{0}$. ∎

Definition 5.4. The sequences $\{\boldsymbol{\omega}^s\}$ and $\{\boldsymbol{\lambda}^s\}$ are defined by

$$\boldsymbol{\omega}^s \in \left\{ \boldsymbol{\omega} \in \Pi_X(\mathbf{x}^s) \;\middle|\; \sum_{i\in\mathcal{I}} \omega_i \nabla a_i(\mathbf{x}^s) + \sum_{j\in\mathcal{E}} \omega_j \mathbf{b}_j = \boldsymbol{\nu}^s \right\}, \qquad s = 0, 1, \dots,$$

and

$$\boldsymbol{\lambda}^s \in \left\{ \boldsymbol{\lambda} \in \Pi_X(\mathbf{x}^{s+1}) \;\middle|\; \sum_{i\in\mathcal{I}} \lambda_i \nabla a_i(\mathbf{x}^{s+1}) + \sum_{j\in\mathcal{E}} \lambda_j \mathbf{b}_j = \mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1} \right\}, \; s = 0, 1, \dots,$$

where $\boldsymbol{\nu}^s \in N_X(\mathbf{x}^s)$ and $\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1} \in N_X(\mathbf{x}^{s+1})$ for all s. The *ergodic sequence* $\{\boldsymbol{\mu}^t\}$ *of Lagrange multipliers* is defined by

$$\boldsymbol{\mu}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \left(\boldsymbol{\omega}^s + \frac{\boldsymbol{\lambda}^s}{\alpha_s} \right), \qquad t = 1, 2, \dots.$$

Note that $\mu_i^t \geq 0$ for all $i \in \mathcal{I}$ and all t.

Clearly, the value of the vector $\boldsymbol{\omega}^s$ depends on the choice of the normal vector $\boldsymbol{\nu}^s$ in the construction of the conditional subgradients used in the method (3). In the traditional subgradient optimization method, $\boldsymbol{\nu}^s = \mathbf{0}$ for all s, and then, from Le. 5.3, $\boldsymbol{\omega}^s = \mathbf{0}$ for all s. In the special case of the method (3) where $\boldsymbol{\gamma}_X^s$ is a *projected subgradient*, $\boldsymbol{\nu}^s = \arg\min_{\boldsymbol{\nu}}\{\|\boldsymbol{\nu}\|^2 \mid \nabla a_i(\mathbf{x}^s)^T(\boldsymbol{\nu}+\boldsymbol{\gamma}^s) \geq 0,\ i \in \mathcal{I}(\mathbf{x}^s);\ \mathbf{b}_j^T(\boldsymbol{\nu}+\boldsymbol{\gamma}^s) = 0,\ j \in \mathcal{E}\}$, and the elements ω_i^s, $i \in \mathcal{I}(\mathbf{x}^s)$, and ω_j^s, $j \in \mathcal{E}$, are Lagrange multipliers associated with the constraints of this projection problem.

The vector $\boldsymbol{\lambda}^s$ always contains Lagrange multipliers associated with the constraints of the projection problem $\mathbf{x}^{s+1} = \arg\min_{\mathbf{x}}\{\|\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}\|^2 \mid a_i(\mathbf{x}) \leq 0,\ i \in \mathcal{I};\ \mathbf{b}_j^T\mathbf{x} = e_j,\ j \in \mathcal{E}\}$.

The following boundedness result is crucial for the convergence of the ergodic sequence $\{\boldsymbol{\mu}^t\}$ of multipliers.

Lemma 5.5. *The sequences* $\{\boldsymbol{\omega}^s\}$ *and* $\{\boldsymbol{\lambda}^s/\alpha_s\}$ *are bounded.*

Proof. We first prove by contradiction that the sequence $\{\boldsymbol{\omega}^s\}$ is bounded.

Suppose that this is not the case. Then there exists a subsequence $\mathcal{S}$ such that $\|\boldsymbol{\omega}^s\| > 0$ for all $s \in \mathcal{S}$ and $\{\|\boldsymbol{\omega}^s\|\}_{s\in\mathcal{S}} \to \infty$. Let $\boldsymbol{\pi}^s = \boldsymbol{\omega}^s/\|\boldsymbol{\omega}^s\|$, implying that $\boldsymbol{\pi}^s \geq \mathbf{0}$ and $\|\boldsymbol{\pi}^s\| = 1$ for all $s \in \mathcal{S}$. According to Prop. 2.2, $\{\mathbf{x}^s\} \to \mathbf{x}^\infty$. Suppose that, for some $i \in \mathcal{I}$, $a_i(\mathbf{x}^\infty) < 0$. By the continuity of the function a_i, it must then hold for all s which are sufficiently large that $a_i(\mathbf{x}^s) < 0$, $\omega_i^s = 0$, and $\pi_i^s = 0$; hence, for any $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}^\infty)$, $\pi_i^s = 0$ for all $s \in \mathcal{S}$ that are large enough. The sequence $\{\boldsymbol{\pi}^s\}_{s\in\mathcal{S}}$ is clearly bounded and thus has an accumulation point, say $\widetilde{\boldsymbol{\pi}}$. Then, from the preceding, $\widetilde{\boldsymbol{\pi}} \in \Pi_X(\mathbf{x}^\infty)$ and $\|\widetilde{\boldsymbol{\pi}}\| = 1$.

By Def. 5.4,

$$\sum_{i\in\mathcal{I}} \pi_i^s \nabla a_i(\mathbf{x}^s) + \sum_{j\in\mathcal{E}} \pi_j^s \mathbf{b}_j = \frac{\boldsymbol{\nu}^s}{\|\boldsymbol{\omega}^s\|}, \qquad \forall s \in \mathcal{S}.$$

In the limit of the subsequence of $\mathcal{S}$ corresponding to $\widetilde{\boldsymbol{\pi}}$, $\sum_{i\in\mathcal{I}(\mathbf{x}^\infty)} \widetilde{\pi}_i \nabla a_i(\mathbf{x}^\infty) + \sum_{j\in\mathcal{E}} \widetilde{\pi}_j \mathbf{b}_j = \mathbf{0}$, since $\{\mathbf{x}^s\} \to \mathbf{x}^\infty$ and the functions ∇a_i, $i \in \mathcal{I}$, are continuous, and since the sequence $\{\boldsymbol{\nu}^t\}$ is bounded and the sequence $\{\|\boldsymbol{\omega}^s\|\}_{s\in\mathcal{S}}$ is presumed to tend to infinity. Then, by Le. 5.3, $\widetilde{\boldsymbol{\pi}} = \mathbf{0}$ must hold, and we have arrived at a contradiction. Hence, the sequence $\{\boldsymbol{\omega}^s\}$ is bounded.

To establish that the sequence $\{\boldsymbol{\lambda}^s/\alpha_s\}$ is bounded, we first observe that, from (3), $\|\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1}\|/\alpha_s \leq \|\boldsymbol{\gamma}^s + \boldsymbol{\nu}^s\|$ for all s, which implies that the sequence $\{(\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1})/\alpha_s\}$ is bounded. Assume then that $\{\boldsymbol{\lambda}^s/\alpha_s\}$ is not bounded and define the index set $\mathcal{S}$ analogously to the above. Letting $\boldsymbol{\pi}^s = \boldsymbol{\lambda}^s/\|\boldsymbol{\lambda}^s\|$ for all $s \in \mathcal{S}$, Def. 5.4 yields that

$$\sum_{i\in\mathcal{I}} \pi_i^s \nabla a_i(\mathbf{x}^{s+1}) + \sum_{j\in\mathcal{E}} \pi_j^s \mathbf{b}_j = \frac{(\mathbf{x}^{s+\frac{1}{2}} - \mathbf{x}^{s+1})/\alpha_s}{\|\boldsymbol{\lambda}^s\|/\alpha_s}, \qquad \forall s \in \mathcal{S}.$$

Clearly, the right–hand–side of this equality tends to the zero vector and we may derive a contradiction which is analogous with the first part of the proof. ◻

From Def. 5.4 and Le. 5.5 follow that under the conditions (4) the sequence $\{\boldsymbol{\mu}^t\}$ is bounded. We conclude this subsection with its main result and an application thereof.

Theorem 5.6 (Convergence of $\{\boldsymbol{\mu}^t\}$ to $M_X(\mathbf{x}^\infty)$). $\{\mathtt{dist}(\boldsymbol{\mu}^t, M_X(\mathbf{x}^\infty))\} \to 0$.

Proof. From the continuity of the functions a_i, $i \in \mathcal{I}$, it follows that $\mathcal{I}(\mathbf{x}^s) \subseteq \mathcal{I}(\mathbf{x}^\infty)$ for all s that are sufficiently large. From Def. 5.4 it then follows that $\omega_i^s = \lambda_i^s = 0$ for all $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}^\infty)$ and all s that are sufficiently large, so that $\{\mu_i^t\} \to 0$, $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{x}^\infty)$. Hence, $\{\mathtt{dist}(\boldsymbol{\mu}^t, \Pi_X(\mathbf{x}^\infty))\} \to 0$.

To simplify the notation, we define $\boldsymbol{\nu}^t = \sum_{i\in\mathcal{I}} \mu_i^t \nabla a_i(\mathbf{x}^\infty) + \sum_{j\in\mathcal{E}} \mu_j^t \mathbf{b}_j$. From Defs. 3.1 and 5.4 it then follows that

$$\boldsymbol{\nu}^t \quad = \quad \mathbf{n}^t - \sum_{i\in\mathcal{I}} A_t^{-1} \left(\sum_{s=0}^{t-1} \alpha_s \omega_i^s \left(\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty) \right) \right.$$

$$+\sum_{s=0}^{t-1} \lambda_i^s \left(\nabla a_i(\mathbf{x}^{s+1}) - \nabla a_i(\mathbf{x}^\infty)\right)\Bigg). \tag{15}$$

Let $\varepsilon > 0$. Le. 3.6 and Thm. 3.7 then yield that $\mathtt{dist}(\mathbf{n}^t, -\partial f(\mathbf{x}^\infty)) \leq \varepsilon/3$ for all t that are sufficiently large. We next show that the second term of the right–hand–side of equation (15) tends to the zero vector. Le. 5.5 implies that the sequence $\{\omega_i^s\}$ is bounded for each $i \in \mathcal{I}$. Hence, since the function ∇a_i is continuous, there is an iterate $t(\varepsilon)$ such that, for all $s \geq t(\varepsilon)$, $\omega_i^s \|\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty)\| \leq \varepsilon/(6|\mathcal{I}|)$. Therefore, for all $t > t(\varepsilon)$,

$$A_t^{-1} \sum_{s=t(\varepsilon)}^{t-1} \alpha_s \omega_i^s \|\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty)\| \leq A_t^{-1} \sum_{s=t(\varepsilon)}^{t-1} \alpha_s \frac{\varepsilon}{6|\mathcal{I}|} \leq \frac{\varepsilon}{6|\mathcal{I}|}.$$

Since $\{A_t\} \to \infty$ we have, for all t large enough, $A_t^{-1}\sum_{s=0}^{t(\varepsilon)-1} \alpha_s \omega_i^s \|\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty)\| \leq \varepsilon/(6|\mathcal{I}|)$. From the derivations above it follows that, for all $t > t(\varepsilon)$ large enough,

$$\begin{aligned}\left\|\sum_{i\in\mathcal{I}} A_t^{-1}\sum_{s=0}^{t-1} \alpha_s\omega_i^s \left(\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty)\right)\right\| &\leq \sum_{i\in\mathcal{I}} A_t^{-1}\sum_{s=0}^{t-1} \alpha_s\omega_i^s \|\nabla a_i(\mathbf{x}^s) - \nabla a_i(\mathbf{x}^\infty)\| \\ &\leq \frac{\varepsilon}{3}.\end{aligned}$$

Analogous to these derivations, it can be shown that the corresponding conclusion holds for the last term of the right–hand–side of equation (15). It then follows that $\mathtt{dist}(\boldsymbol{\nu}^t, -\partial f(\mathbf{x}^\infty)) \leq \varepsilon$ for all t that are sufficiently large and, consequently, we have that $\{\mathtt{dist}(\boldsymbol{\nu}^t, -\partial f(\mathbf{x}^\infty))\} \to 0$. Now, let $\mathcal{T}$ be a subsequence such that

$$\lim_{\substack{t\in\mathcal{T}\\ t\to\infty}} \mathtt{dist}\left(\boldsymbol{\mu}^t, M_X(\mathbf{x}^\infty)\right) = \limsup_{t\to\infty} \mathtt{dist}\left(\boldsymbol{\mu}^t, M_X(\mathbf{x}^\infty)\right),$$

and $\widetilde{\boldsymbol{\mu}}$ be an accumulation point of the sequence $\{\boldsymbol{\mu}^t\}_{t\in\mathcal{T}}$. From this analysis and the closedness of the sets $\Pi_X(\mathbf{x}^\infty)$ and $\partial f(\mathbf{x}^\infty)$ it follows that the inclusions $\widetilde{\boldsymbol{\mu}} \in \Pi_X(\mathbf{x}^\infty)$ and $\sum_{i\in\mathcal{I}} \widetilde{\mu}_i \nabla a_i(\mathbf{x}^\infty) + \sum_{i\in\mathcal{E}} \widetilde{\mu}_j \mathbf{b}_j \in -\partial f(\mathbf{x}^\infty)$ hold, thus implying that $\widetilde{\boldsymbol{\mu}} \in M_X(\mathbf{x}^\infty)$. ∎

We consider an application of this result to the post-optimal analysis of a special case of the program [P] of the form

$$\begin{aligned} f^* = \min \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & a_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{I}, \end{aligned} \tag{16}$$

where the objective function has the explicit form

$$f(\mathbf{x}) = \max_{k\in\mathcal{K}} c_k(\mathbf{x}), \qquad \mathbf{x} \in \Re^n, \tag{17}$$

where the functions $c_k : \Re^n \mapsto \Re$, $k \in \mathcal{K}$, are convex and differentiable and the set $\mathcal{K}$ is finite.

The *shadow prices* associated with the constraints of the program (16) are the rates of change of the optimal objective value caused by perturbations of the right–hand–sides. Hence, the shadow price, p_i, for the i:th constraint is defined as the directional derivative of the perturbation function, $v(\mathbf{u}) = \min\{\, f(\mathbf{x}) \mid a_i(\mathbf{x}) \leq u_i,\ i \in \mathcal{I}\,\}$, at the origin, along the i:th coordinate, that is, $p_i = v'(\mathbf{0}; \mathbf{e}_i)$, $i \in \mathcal{I}$. If v is differentiable at the origin, then $[p_i]_{i \in \mathcal{I}} = \nabla v(\mathbf{0})$ (Rockafellar, 1970, Cor. 29.1.3). In this case, the ergodic sequence of multipliers provides estimates of the shadow prices, as stated next.

Corollary 5.7 (Shadow prices from multiplier estimates). *If the perturbation function v is differentiable at $\mathbf{0}$, then $M_X(\mathbf{x}^\infty) = \{-\nabla v(\mathbf{0})\}$, and $\{\boldsymbol{\mu}^t\} \to -\nabla v(\mathbf{0})$.*

Proof. It follows by Cor. 29.1.3 in Rockafellar (1970) that the set $M_X(\mathbf{x}^\infty)$ is a singleton, which moreover equals $-\nabla v(\mathbf{0})$. Thm. 5.6 then yields the result. □

5.2 Multipliers of objective pieces

Throughout this subsection we assume that the nonsmooth objective function f of the program [P] has the form (17) and the feasible set X is given by (12).

We define the set of multipliers associated with the pieces c_k satisfying a complementary slackness condition at $\mathbf{x} \in \Re^n$ as

$$\Pi_f(\mathbf{x}) = \left\{ \boldsymbol{\pi} \in \Re_+^{|\mathcal{K}|} \;\middle|\; \pi_k\,(c_k(\mathbf{x}) - f(\mathbf{x})) = 0,\ k \in \mathcal{K};\quad \sum_{k \in \mathcal{K}} \pi_k = 1 \right\}. \tag{18}$$

Letting $\boldsymbol{\xi} \in \Re^{|\mathcal{K}|}$, the subdifferential of f at $\mathbf{x}$ can then be expressed as (e.g., Hiriart-Urruty and Lemaréchal, 1993, Cor. VI:4.4.4)

$$\partial f(\mathbf{x}) = \left\{ \boldsymbol{\gamma} \in \Re^n \;\middle|\; \boldsymbol{\gamma} = \sum_{k \in \mathcal{K}} \xi_k \nabla c_k(\mathbf{x});\quad \boldsymbol{\xi} \in \Pi_f(\mathbf{x}) \right\}. \tag{19}$$

For any $\mathbf{x} \in X$, let

$$M_f(\mathbf{x}) = \left\{ \boldsymbol{\eta} \in \Pi_f(\mathbf{x}) \;\middle|\; \mathbf{0} \in \sum_{k \in \mathcal{K}} \eta_k \nabla c_k(\mathbf{x}) + N_X(\mathbf{x}) \right\} \tag{20}$$

be the (possibly empty) set of multipliers associated with the pieces c_k defining the function f. The optimality conditions for the program [P] can be stated in terms of this set.

Proposition 5.8 (Optimality condition). *Let $\mathbf{x} \in X$. Then, $\mathbf{x} \in X^*$ if and only if $M_f(\mathbf{x}) \neq \emptyset$. Further, the mapping M_f is constant on X^* and the set $M_f(\mathbf{x})$, $\mathbf{x} \in X^*$, is convex and compact.*

Proof. The first conclusion is a consequence of Prop. 2.1 and the expressions (19) and (20). The invariance of $M_f(\mathbf{x})$ over X^* follows from Prop. VII:3.1.1 in Hiriart-Urruty and Lemaréchal (1993) applied to the reformulation $\min\{\, z \mid z \geq c_k(\mathbf{x}),\ k \in$

$\mathcal{K}$; $\mathbf{x} \in X$ } of the program [P], and its convexity and compactness are immediate from its definition. ∎

Definition 5.9. The sequence $\{\boldsymbol{\xi}^s\}$ is defined by

$$\boldsymbol{\xi}^s \in \left\{ \boldsymbol{\xi} \in \Pi_f(\mathbf{x}^s) \;\middle|\; \sum_{k \in \mathcal{K}} \xi_k \nabla c_k(\mathbf{x}^s) = \boldsymbol{\gamma}^s \right\}, \qquad s = 0, 1, \ldots,$$

where $\boldsymbol{\gamma}^s \in \partial f(\mathbf{x}^s)$ for all s. The *ergodic sequence* $\{\boldsymbol{\eta}^t\}$ *of multipliers* is then defined by

$$\boldsymbol{\eta}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \boldsymbol{\xi}^s, \qquad t = 1, 2, \ldots.$$

Note that $\boldsymbol{\eta}^t \geq \mathbf{0}$ and $\sum_{k \in \mathcal{K}} \eta_k^t = 1$ for all t, so that the sequence $\{\boldsymbol{\eta}^t\}$ is bounded.

Theorem 5.10 (Convergence of $\{\boldsymbol{\eta}^t\}$ to $M_f(\mathbf{x}^\infty)$). $\{\mathtt{dist}(\boldsymbol{\eta}^t, M_f(\mathbf{x}^\infty))\} \to 0$.

Proof. As in the proof of Thm. 5.6 we can show that $\{\mathtt{dist}(\boldsymbol{\eta}^t, \Pi_f(\mathbf{x}^\infty))\} \to 0$. Defining $\boldsymbol{\gamma}^t = \sum_{k \in \mathcal{K}} \eta_k^t \nabla c_k(\mathbf{x}^\infty)$, it follows from Defs. 3.1 and 5.9 that

$$\boldsymbol{\gamma}^t = \mathbf{g}^t - \sum_{k \in \mathcal{K}} A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \xi_k^s \left(\nabla c_k(\mathbf{x}^s) - \nabla c_k(\mathbf{x}^\infty) \right).$$

Let $\varepsilon > 0$. Thm. 3.7 then yields that $\mathtt{dist}(\mathbf{g}^t, -N_X(\mathbf{x}^\infty)) \leq \varepsilon/2$ for all t that are sufficiently large. Using the same technique as in the proof of Thm. 5.6 and utilizing the boundedness of the sequence $\{\boldsymbol{\xi}^s\}$, we can show that

$$0 \leq \left\| \sum_{k \in \mathcal{K}} A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \xi_k^s \left(\nabla c_k(\mathbf{x}^s) - \nabla c_k(\mathbf{x}^\infty) \right) \right\| \leq \frac{\varepsilon}{2}.$$

It follows that $\mathtt{dist}(\boldsymbol{\gamma}^t, -N_X(\mathbf{x}^\infty)) \leq \varepsilon$ for all t that are sufficiently large, and hence that $\{\mathtt{dist}(\boldsymbol{\gamma}^t, -N_X(\mathbf{x}^\infty))\} \to 0$. Using the same technique as in the proof of Thm. 5.6, we can also show that all accumulation points, $\widetilde{\boldsymbol{\eta}}$, of the sequence $\{\boldsymbol{\eta}^t\}$ fulfil the inclusions $\widetilde{\boldsymbol{\eta}} \in \Pi_f(\mathbf{x}^\infty)$ and $\sum_{k \in \mathcal{K}} \widetilde{\eta}_k \nabla c_k(\mathbf{x}^\infty) \in -N_X(\mathbf{x}^\infty)$, thus implying $\widetilde{\boldsymbol{\eta}} \in M_f(\mathbf{x}^\infty)$. ∎

As a specialization of this result, let $\mathbf{a}_i \in \Re^n$ and $b_i \in \Re$ for all $i \in \mathcal{L}$ (presumed finite) and consider the system

$$\mathbf{a}_i^T \mathbf{x} \leq b_i, \qquad i \in \mathcal{L}, \tag{21}$$

of linear inequalities with solution set Y.

Letting the piece-wise affine function $f : \Re^n \mapsto \Re$ be defined as

$$f(\mathbf{x}) = \max_{i \in \mathcal{L}} \left\{ \mathbf{a}_i^T \mathbf{x} - b_i \right\}, \qquad \mathbf{x} \in \Re^n, \tag{22}$$

it follows that $\mathbf{x} \in Y$ if and only if $f(\mathbf{x}) \leq 0$.

By *Farkas' theorem of the alternative*, exactly one of the systems (21) and

$$\sum_{i\in\mathcal{L}} \eta_i \mathbf{a}_i = \mathbf{0}, \qquad \sum_{i\in\mathcal{L}} \eta_i b_i < 0, \qquad \eta_i \geq 0, \qquad \forall i \in \mathcal{L}, \tag{23}$$

has a solution. Therefore, the system (23) is *consistent* if and only if $f^* = \min_{\mathbf{x}\in\Re^n} f(\mathbf{x}) > 0$. For this case, the next result, which is a consequence of Thm. 5.10, states that an ergodic sequence of multipliers generated within the conditional subgradient method applied to the minimization of f in (22), provides solutions to the system (23) in the limit.

Corollary 5.11 (A Farkas theorem). $\{\sum_{i\in\mathcal{L}} \eta_i^t \mathbf{a}_i\} \to \mathbf{0}$, $\{\sum_{i\in\mathcal{L}} \eta_i^t b_i\} \to -f^*$, *and* $\eta_i^t \geq 0$, $i \in \mathcal{L}$, *for all* t. *Hence, if* $Y = \emptyset$, *then the sequence* $\{\boldsymbol{\eta}^t\}$ *tends towards the set of solutions to the system (23).*

Proof. The first result follows from Theorem 3.9, identifying $\{\sum_{i\in\mathcal{L}} \eta_i^t \mathbf{a}_i\}$ with $\{\mathbf{g}^t\}$. By Definition 5.9, the definitions (18) and (7), and Lemma 3.4,

$$\begin{aligned}
\sum_{i\in\mathcal{L}} \eta_i^t b_i &= \sum_{i\in\mathcal{L}} A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \xi_i^s b_i = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \sum_{i\in\mathcal{L}} \xi_i^s \left(\mathbf{a}_i^T \mathbf{x}^s - f(\mathbf{x}^s)\right) \\
&= A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \left((\boldsymbol{\gamma}^s)^T \mathbf{x}^s - f(\mathbf{x}^s)\right) = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s (\boldsymbol{\gamma}^s)^T \mathbf{x}^\infty - f_t(\mathbf{x}^\infty) \\
&= (\mathbf{g}^t)^T \mathbf{x}^\infty - f_t(\mathbf{x}^\infty) \to 0 - f(\mathbf{x}^\infty) = -f^*.
\end{aligned}$$

The last result follows from Definition 5.9 and the definition (18). □

Further aspects on this result are presented in the next section.

6 ERGODIC UNDERESTIMATING FUNCTIONS

The results established in Sec. 3 will now be utilized to derive a convergent sequence of lower bounds on the optimal value of the program [P]. The results obtained are then utilized to finitely detect the inconsistency and separation in a system of linear inequalities intersected by a convex set. We assume throughout this section that X is bounded.

Definition 6.1. The *sequence* $\{f_t^*\}$ *of lower bounds* on f^* is defined by $f_t^* = \min_{\mathbf{x}\in X} f_t(\mathbf{x})$.

We show that the sequence $\{f_t^*\}$ of lower bounds converges to the optimal value.

Theorem 6.2 (Convergence of lower bounds). $f_t^* \in (-\infty, f^*]$ *for all* t, *and* $\{f_t^*\} \to f^*$.

Proof. From Def. 6.1, and the definitions (7) and (1) it follows that $f_t^* \leq f^*$ for all t. Since X is bounded, f_t^* is finite for all t.

By the same definitions, $f_t^* = f_t(\mathbf{x}^\infty) + \min_{\mathbf{x}\in X} (\mathbf{g}^t)^T(\mathbf{x} - \mathbf{x}^\infty)$ for all t. From Thm. 3.8 follows that $\liminf_{t\to\infty} (\mathbf{g}^t)^T(\mathbf{x} - \mathbf{x}^\infty) \geq 0$ for all $\mathbf{x} \in X$, and from Le. 3.4 that $\{f_t(\mathbf{x}^\infty)\} \to f^*$. Hence, $\liminf_{t\to\infty} f_t^* \geq f^*$. Since $f_t^* \leq f^*$ for all t it follows that $\lim_{t\to\infty} f_t^* = f^*$. □

In an application context it is natural to repeatedly save the best upper and lower bounds on the optimal value found so far. In such a manner, the convergent sequence $\{f_t^*\} \to f^*$ of lower bounds can be utilized in a termination criterion for the conditional subgradient method (3)–(4). The following result, which is a consequence of Prop. 2.2 and Thm. 6.2, implies that an ε-optimal solution is found finitely.

Corollary 6.3 (Finite termination). *Define* $u_t = \min_{0 \leq s \leq t} f(\mathbf{x}^s)$ *and* $l_t = \max_{1 \leq s \leq t} f_s^*$ *for all* t. *Then,* $l_t \leq f^* \leq u_t$ *holds for all* t, *and for any* $\varepsilon > 0$ *there is a* $t(\varepsilon)$ *such that* $u_t - l_t \leq \varepsilon$ *holds for all* $t \geq t(\varepsilon)$.

In Larsson et al. (1996c) the upper and lower bounds defined in the corollary is utilized in a resource directive decomposition scheme for the linear minimum cost multicommodity network flow problem. The results presented in that report illustrate the practical usefulness of the lower bounding procedure derived in this section.

In a forthcoming paper, we will utilize the lower bounding procedure for assessing the quality of solutions obtained within Lagrangean dualization schemes.

As a specialization of the theorem, we consider the generalization of the system (21) to

$$\mathbf{x} \in X, \qquad \mathbf{a}_i^T \mathbf{x} \leq b_i, \quad i \in \mathcal{L}, \tag{24}$$

and the optimization setting of this system,

$$f^* = \min_{\mathbf{x} \in X} f(\mathbf{x}), \tag{25}$$

where $f : \Re^n \mapsto \Re$ is defined by (22). We define the set $Y = \{\mathbf{x} \in \Re^n \mid f(\mathbf{x}) \leq 0\}$, and consider solving the program (25) by the conditional subgradient optimization method (3)–(4) extended with the lower bounding procedure of Def. 6.1.

If the system (24) is *consistent*, then the conditional subgradient method either finds an $\mathbf{x}^s \in X \cap Y$, or the sequence $\{\mathbf{x}^t\} \to \overline{\mathbf{x}} \in X \cap \mathsf{bd}\, Y$. The following result, which is a consequence of Thm. 6.2, deals with the case of *inconsistency*.

Corollary 6.4 (Finite detection of inconsistency and finite separation). *If* $X \cap Y = \emptyset$, *then there is an iteration* t_0 *such that* $f_t^* > 0$ *for all* $t \geq t_0$. *If further* $Y \neq \emptyset$, *then for every* $t \geq t_0$ *the hyperplane* $\{\mathbf{x} \in \Re^n \mid f_t(\mathbf{x}) = f_t^*/2\}$ *strictly separates the sets* X *and* Y.

Proof. The first part of the result is immediate from Thm. 6.2.

Since $f_t(\mathbf{x}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \Re^n$, it follows that $Y \subseteq \{\mathbf{x} \in \Re^n \mid f_t(\mathbf{x}) \leq 0\}$ for all t, and by definition of the sequence $\{f_t^*\}$, $X \subseteq \{\mathbf{x} \in \Re^n \mid f_t(\mathbf{x}) \geq f_t^*\}$ for all t. Invoke the first part of the result. □

7 CONCLUSIONS AND FURTHER RESEARCH

The analysis of ergodic sequences of subgradients within conditional subgradient optimization schemes reveals new and interesting characteristics of subgradient methods. In particular, they are used to establish the ergodic convergence of sequences of multipliers, thus enabling the performance of sensitivity analyses of solutions obtained by subgradient methods. As a consequence of these results we prove the convergence of a lower bounding procedure, providing a new termination criterion for subgradient methods.

In a forthcoming paper (Larsson et al., 1996d) we apply a conditional subgradient method to a Lagrange dual of a *general convex program* and establish the ergodic convergence of the sequence of primal subproblem solutions obtained.

The lower bounding procedure of Sec. 6 may be employed to assess the quality of the dual solutions obtained from the method (3) applied to a Lagrange dual of any optimization problem.

Sherali and Choi (1996) apply subgradient methods to Lagrange duals of *linear programs* and establish ergodic convergence of sequences of subproblem solutions, employing more general choices of subgradient step lengths and convexity weights than those of this paper.

An interesting subject for further research is the study of ergodic sequences of subgradients, normal components, and multipliers under similar, more general rules, as well as for more efficient computational schemes for nondifferentiable optimization. The use of ergodic sequences of subgradients for accelerating subgradient algorithms is another subject for further study.

ACKNOWLEDGEMENTS

We thank an anonymous referee for valuable comments that resulted in an improved presentation. The research was financially supported by grants (Nos. 94-292 and 282-93-1195) from the Swedish Research Council for Engineering Sciences (TFR).

REFERENCES

1. J.-B. Baillon (1975), Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (Paris), Série A*, 280, pp. 1511–1514.
2. D.P. Bertsekas (1976), On the Goldstein–Levitin–Polyak gradient projection method, *IEEE Transactions on Automatic Control*, AC-21, pp. 174–184.
3. J.V. Burke and M.C. Ferris (1991), Characterization of solution sets of convex programs, *Operations Research Letters*, 10, pp. 57–60.
4. J.V. Burke and J.J. Moré (1988), On the identification of active constraints, *SIAM Journal on Numerical Analysis*, 25, pp. 1197–1211.
5. J.V. Burke and J.J. Moré (1994), Exposing constraints, *SIAM Journal on Optimization*, 3, pp. 573–595.
6. P.H. Calamai and J.J. Moré (1987), Projected gradient methods for linearly constrained problems, *Mathematical Programming*, 39, pp. 93–116.
7. V.F. Dem'janov and V.K. Šomesova (1978), Conditional subdifferentials of convex functions, *Soviet Mathematics Doklady*, 19, pp. 1181–1185.
8. V.F. Dem'yanov and L.V. Vasil'ev (1985), *Nondifferentiable Optimization*, Optimization Software, New York, NY.
9. J.C. Dunn (1981), Global and asymptotic convergence rate estimates for a class of projected gradient processes, *SIAM Journal on Control and Optimization*, 19, pp. 368–400.

10. J.C. Dunn (1987), On the convergence of projected gradient processes to singular critical points, *Journal of Optimization Theory and Applications*, 55, pp. 203–216.
11. M.L. Fisher (1985), An applications oriented guide to Lagrangian relaxation, *Interfaces*, 15, pp. 10–21.
12. M. Guignard (1969), Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space, *SIAM Journal on Control*, 7, pp. 232–241.
13. J.L. Higle and S. Sen (1991), Stochastic decomposition: An algorithm for two-stage linear programs with recourse, *Mathematics of Operations Research*, 16, pp. 650–669.
14. J.-B. Hiriart-Urruty and C. Lemaréchal (1993), *Convex Analysis and Minimization Algorithms I & II*, Springer-Verlag, Berlin.
15. S. Kim and H. Ahn (1991), Convergence of a generalized subgradient method for nondifferentiable convex optimization, *Mathematical Programming*, 50, pp. 75–80.
16. T. Larsson and Z. Liu (1997), A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows, *Optimization*, 40, pp. 247–284.
17. T. Larsson, M. Patriksson and A.-B. Strömberg (1996a), Conditional subgradient optimization: Theory and applications, *European Journal of Operational Research*, 88, pp. 382–403.
18. T. Larsson, M. Patriksson and A.-B. Strömberg (1996b), Ergodic results in subgradient optimization, in: *Nonlinear Optimization and Applications*, G. Di Pillo, F. Giannessi, eds., Proceedings of the International School of Mathematics "G. Stampacchia" 21st Workshop on Nonlinear Optimization and Applications, Erice, Italy, June, 1995, Plenum Press, New York, pp. 229–248.
19. T. Larsson, M. Patriksson and A.-B. Strömberg (1996c), Ergodic results and bounds on the optimal value in subgradient optimization, in: *Operations Research Proceedings 1995*, P. Kleinschmidt, A. Bachem, U. Derigs, D. Fischer, U. Leopold-Wildburger, R. Möhring, eds., Selected papers of the Symposium on Operations Research, Passau, Germany, September, 1995, Springer-Verlag, Berlin, pp. 30–35.
20. T. Larsson, M. Patriksson and A.-B. Strömberg (1996d), *Ergodic, primal convergence in dual subgradient schemes for convex programming*, Report LiTH-MAT-R-96-26, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden.
21. D. G. Luenberger (1984), *Linear and Nonlinear Programming*, second ed. Addison-Wesley, Reading, MA.
22. J. Petersson and M. Patriksson (1997), Topology optimization of sheets in contact by a subgradient method, *International Journal of Numerical Methods in Engineering*, 40, pp. 1295–1321.
23. B.T. Polyak (1987), *Introduction to optimization*, Optimization Software, New York, NY.
24. S.M. Robinson (1984), Local structure of feasible sets in nonlinear programming, part II: Nondegeneracy, *Mathematical Programming Study*, 22, pp. 217–230.
25. R.T. Rockafellar (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.
26. A. Ruszczyński and W. Syski (1983), Stochastic approximation method with gradient averaging for unconstrained problems, *IEEE Transactions on Automatic Control*, AC-28, pp. 1097–1105.
27. H.D. Sherali and G. Choi (1996), Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs, *Operations Research Letters*, 19, pp. 105–113.
28. N.Z. Shor (1985), *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin.
29. N.Z. Shor (1991), The development of numerical methods for nonsmooth optimization in the USSR, in: *History of Mathematical Programming: A Collection of Personal Reminiscences*, J.K. Lenstra, A.H.G. Rinnooy Kan, A. Schrijver, eds., North-Holland, Amsterdam, pp. 135–139.
30. A.-B. Strömberg (1997), *Conditional Subgradient Methods and Ergodic Convergence in Nonsmooth Optimization*, PhD Thesis, Linköping Studies in Science and Technology, Dissertations, No. 467, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden.

31. E.H. Zarantonello (1971), Projections on convex sets in Hilbert space and spectral theory, in: *Contributions to Nonlinear Functional Analysis*, E.H. Zarantonello, ed., Proceedings of a Symposium Conducted by the Mathematics Research Center, University of Wisconsin, Madison, April 1971, Academic Press, New York, NY, pp. 237–424.

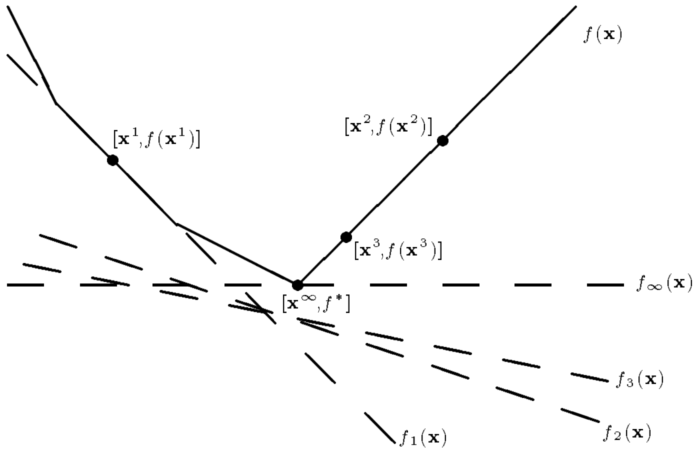
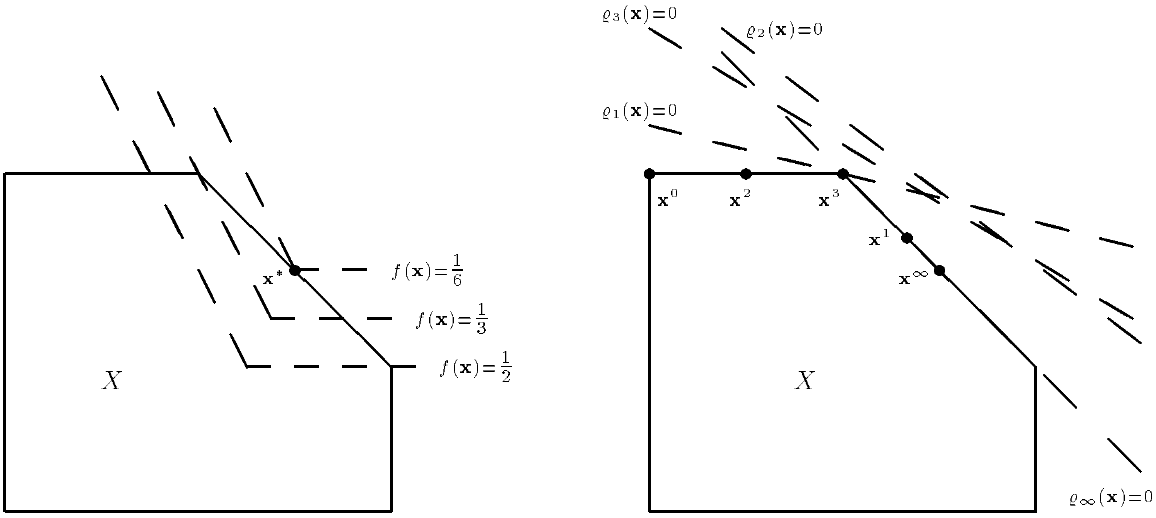


FIGURE 1: *Iteration points and affine underestimates of* f.



a) *b)*

FIGURE 2: *a) The set X and level curves of f. b) Iteration points and hyperplanes defining valid inequalities for X.*