Working Paper

Proximal Minimization Methods with Generalized Bregman Functions

Krzysztof C. Kiwiel

WP-95-024
March 1995

IIASA
International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Proximal Minimization Methods with Generalized Bregman Functions

Krzysztof C. Kiwiel

WP-95-024
March 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Proximal Minimization Methods with Generalized Bregman Functions*

*Krzysztof C. Kiwiel***

Abstract

We consider methods for minimizing a convex function f that generate a sequence $\{x^k\}$ by taking x^{k+1} to be an approximate minimizer of $f(x) + D_h(x, x^k)/c_k$, where $c_k > 0$ and D_h is the D-function of a Bregman function h. Extensions are made to B-functions that generalize Bregman functions and cover more applications. Convergence is established under criteria amenable to implementation. Applications are made to nonquadratic multiplier methods for nonlinear programs.

Key words. Convex programming, nondifferentiable optimization, proximal methods, Bregman functions, B-functions.

1 Introduction

We consider the convex minimization problem

$$f_* = \inf\{\, f(x) : x \in X \,\}, \tag{1.1}$$

where $f : \mathbb{R}^n \to (-\infty, \infty]$ is a closed proper convex function and X is a nonempty closed convex set in $\mathbb{R}^n$. One method for solving (1.1) is the proximal point algorithm (PPA) [Mar70, Roc76b] which generates a sequence

$$x^{k+1} = \arg\min\{\, f(x) + |x - x^k|^2/2c_k : x \in X \,\} \quad \text{for } k = 1, 2, \ldots, \tag{1.2}$$

starting from any point $x^1 \in \mathbb{R}^n$, where $|\cdot|$ is the Euclidean norm and $\{c_k\}$ is a sequence of positive numbers. The convergence and applications of the PPA are discussed, e.g., in [Aus86, CoL93, EcB92, GoT89, Gül91, Lem89, Roc76a, Roc76b].

Several proposals have been made for replacing the quadratic term in (1.2) with other distance-like functions [BeT94, CeZ92, ChT93, Eck93, Egg90, Ius95, IuT93, Teb92, TsB93]. In [CeZ92], (1.2) is replaced by

$$x^{k+1} = \arg\min\{\, f(x) + D_h(x, x^k)/c_k : x \in X \,\}, \tag{1.3}$$

*Research supported by the State Committee for Scientific Research under Grant 8S50502206.
**Systems Research Institute, Newelska 6, 01–447 Warsaw, Poland (`kiwiel@ibspan.waw.pl`)

where $D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle$ is the D-function of a Bregman function h [Bre67, CeL81], which is continuous, strictly convex and differentiable in the interior of its domain (see §2 for a full definition); here $\langle \cdot, \cdot \rangle$ is the usual inner product and ∇h is the gradient of h. Accordingly, this is called *Bregman proximal minimization* (BPM). The convergence of the BPM method is discussed in [CeZ92, ChT93, Eck93, Ius95, TsB93], a generalization for finding zeros of monotone operators is given in [Eck93], and applications to convex programming are presented in [Cha94, Eck93, Ius95, NiZ92, NiZ93a, NiZ93b, Teb92, TsB93].

This paper discusses convergence of the BPM method using the *B-functions* of [Kiw94] that generalize Bregman functions, being possibly nondifferentiable and infinite on the boundary of their domains (cf. §2). Then (1.3) involves $D_h^k(x, x^k) = h(x) - h(x^k) - \left\langle \gamma^k, x - x^k \right\rangle$, where γ^k is a subgradient of h at x^k. We establish for the first time convergence of versions of the BPM method that relax the requirement for exact minimization in (1.3). (The alternative approach of [Flå94], being restricted to Bregman functions with Lipschitz continuous gradients, cannot handle the applications of §§7–9.) We note that in several important applications, strictly convex problems of the form (1.3) may be solved by dual ascent methods; cf. references in [Kiw94, Tse90].

The application of the BPM method to the dual functional of a convex program yields nonquadratic multiplier methods [Eck93, Teb92]. By allowing h to have singularities, we extend this class of methods to include, e.g., *shifted* Frish and Carroll barrier function methods [FiM68]. We show that our criteria for inexact minimization can be implemented similarly as in the nonquadratic multiplier methods of [Ber82, Chap. 5]. Our convergence results extend those in [Eck93, TsB93] to quite general *shifted penalty functions*, including twice continuously differentiable ones.

We add that the continuing interest in nonquadratic modified Lagrangians stems from the fact that, in contrast with the quadratic one, they are twice continuously differentiable, and this facilitates their minimization [Ber82, BTYZ92, BrS93, BrS94, CGT92, CGT94, GoT89, IST94, JeP94, Kiw96, NPS94, Pol92, PoT94, Teb92, TsB93]. By the way, our convergence results seem stronger than ones in [IST94, PoT94] for *modified barrier functions*, resulting from a dual application of (1.3) with $D_h^k(x, x^k)$ replaced by an entropy-like ϕ-divergence.

The paper is organized as follows. In §2 we recall the definitions of B-functions and Bregman functions and state their elementary properties. In §3 we present an inexact BPM method. Its global convergence under various conditions is established in §§4–5. In §6 we show that the exact BPM method converges finitely when (1.1) enjoys a sharp minimum property. Applications to multiplier methods are given in §7. Convergence of general multiplier methods is studied in §8, while §9 focuses on two classes of shifted penalty methods. Additional aspects of multiplier methods are discussed in §10. The Appendix contains proofs of certain technical results.

Our notation and terminology mostly follow [Roc70]. $\mathbb{R}^m_+$ and $\mathbb{R}^m_>$ are the nonnegative and positive orthants of $\mathbb{R}^m$ respectively. For any set C in $\mathbb{R}^n$, $\operatorname{cl} C$, $\mathring{C}$, $\operatorname{ri} C$ and $\operatorname{bd} C$ denote the closure, interior, relative interior and boundary of C respectively. δ_C is the *indicator* function of C ($\delta_C(x) = 0$ if $x \in C$, ∞ otherwise). $\sigma_C(\cdot) = \sup_{x \in C} \langle \cdot, x \rangle$ is the *support* function of C. For any closed proper convex function f on $\mathbb{R}^n$ and x in its *effective* *domain* $C_f = \{x : f(x) < \infty\}$, $\partial_\epsilon f(x) = \{p : f(y) \geq f(x) + \langle p, y - x\rangle - \epsilon\ \forall y\}$ is the *ϵ-subdifferential* of f at x for each $\epsilon \geq 0$, $\partial f(x) = \partial_0 f(x)$ is the ordinary *subdifferential* of f at x and $f'(x;d) = \lim_{t\downarrow 0}[f(x+td) - f(x)]/t$ denotes the *derivative* of f in any direction $d \in \mathbb{R}^n$. By [Roc70, Thms 23.1–23.2], $f'(x;d) \geq -f'(x;-d)$ and

$$f'(x;d) \geq \sigma_{\partial f(x)}(d) = \sup\{\langle \gamma, d\rangle : \gamma \in \partial f(x)\}. \tag{1.4}$$

The *domain* and *range* of ∂f are denoted by $C_{\partial f}$ and $\operatorname{im}\partial f$ respectively. By [Roc70, Thm 23.4], $\operatorname{ri} C_f \subset C_{\partial f} \subset C_f$. f is called *cofinite* when its *conjugate* $f^*(\cdot) = \sup_x \langle \cdot, x\rangle - f(x)$ is real-valued. A proper convex function f is called *essentially smooth* if $\mathring{C}_f \neq \emptyset$, f is differentiable on $\mathring{C}_f$, and $|\nabla f(x^k)| \to \infty$ if $x^k \to x \in \operatorname{bd} C_f$, $\{x^k\} \subset \mathring{C}_f$. If f is closed proper convex, its *recession* function $f0^+(\cdot) = \lim_{t\to\infty}[f(x + t\cdot) - f(x)]/t$ ($\forall x \in C_f$) is positively homogeneous [Roc70, Thm 8.5].

2 B-functions

We first recall the definitions of B-functions [Kiw94] and of Bregman functions [CeL81].

For any convex function h on $\mathbb{R}^n$, we define its *difference functions*

$$D_h^\flat(x,y) = h(x) - h(y) - \sigma_{\partial h(y)}(x - y) \quad \forall x, y \in C_h, \tag{2.1a}$$
$$D_h^\sharp(x,y) = h(x) - h(y) + \sigma_{\partial h(y)}(y - x) \quad \forall x, y \in C_h. \tag{2.1b}$$

By convexity (cf. (1.4)), $h(x) \geq h(y) + \sigma_{\partial h(y)}(x - y)$ and

$$0 \leq D_h^\flat(x,y) \leq h(x) - h(y) - \langle \gamma, x - y\rangle \leq D_h^\sharp(x,y) \quad \forall x, y \in C_h, \gamma \in \partial h(y). \tag{2.2}$$

$D_h^\flat$ and $D_h^\sharp$ generalize the usual *D-function* of h [Bre67, CeL81], defined by

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle \quad \forall x \in C_h, y \in C_{\nabla h}, \tag{2.3}$$

since

$$D_h(x,y) = D_h^\flat(x,y) = D_h^\sharp(x,y) \quad \forall x \in C_h, y \in C_{\nabla h}. \tag{2.4}$$

Definition 2.1. A closed proper (possibly nondifferentiable) convex function h is called a B-function (*generalized Bregman function*) if

(a) h is strictly convex on C_h.

(b) h is continuous on C_h.

(c) For every $\alpha \in \mathbb{R}$ and $x \in C_h$, the set $\mathcal{L}_h^1(x,\alpha) = \{y \in C_{\partial h} : D_h^\flat(x,y) \leq \alpha\}$ is bounded.

(d) For every $\alpha \in \mathbb{R}$ and $x \in C_h$, if $\{y^k\} \subset \mathcal{L}_h^1(x,\alpha)$ is a convergent sequence with limit $y^* \in C_h \setminus \{x\}$, then $D_h^\sharp(y^*, y^k) \to 0$.

Definition 2.2. Let S be a nonempty open convex set in $\mathbb{R}^n$. Then $h : \bar{S} \to \mathbb{R}$, where $\bar{S} = \operatorname{cl} S$, is called a *Bregman function* with *zone* S, denoted by $h \in \mathcal{B}(S)$, if

(i) h is continuously differentiable on S.

(ii) h is strictly convex on $\bar{S}$.

(iii) h is continuous on $\bar{S}$.

(iv) For every $\alpha \in \mathbb{R}$, $\tilde{y} \in S$ and $\tilde{x} \in \bar{S}$, the sets $\mathcal{L}_h^2(\tilde{y},\alpha) = \{x \in \bar{S} : D_h(x,\tilde{y}) \le \alpha\}$ and $\mathcal{L}_h^3(\tilde{x},\alpha) = \{y \in S : D_h(\tilde{x},y) \le \alpha\}$ are bounded.

(v) If $\{y^k\} \subset S$ is a convergent sequence with limit y^*, then $D_h(y^*,y^k) \to 0$.

(vi) If $\{y^k\} \subset S$ converges to y^*, $\{x^k\} \subset \bar{S}$ is bounded and $D_h(x^k,y^k) \to 0$ then $x^k \to y^*$.

(Note that the extension e of h to $\mathbb{R}^n$, defined by $e(x) = h(x)$ if $x \in \bar{S}$, $e(x) = \infty$ otherwise, is a B-function with $C_e = \bar{S}$, $\operatorname{ri} C_e = S$ and $D_e^\flat(\cdot,y) = D_e^\sharp(\cdot,y) = D_e(\cdot,y)$ $\forall y \in S$.)

$D_h^\flat$ and $D_h^\sharp$ are used like distances, because for $x,y \in C_h$, $0 \le D_h^\flat(x,y) \le D_h^\sharp(x,y)$, and $D_h^\flat(x,y) = 0 \iff D_h^\sharp(x,y) = 0 \iff x = y$ by strict convexity. Definition 2.2 (due to [CeL81]), which requires that h be *finite-valued* on $\bar{S}$, does not cover Burg's entropy [CDPI91]. Our Definition 2.1 captures features of h essential for algorithmic purposes. As shown in [Kiw94], condition (b) implies (c) if h is cofinite. Sometimes one may verify the following stronger version of condition (d)

$$C_{\partial h} \supset \{y^k\} \to y^* \in C_h \quad \Rightarrow \quad D_h^\sharp(y^*,y^k) \to 0 \tag{2.5}$$

by using the following three lemmas proven in [Kiw94].

Lemma 2.3. (a) *Let h be a closed proper convex function on $\mathbb{R}^n$, and let $S \neq \emptyset$ be a compact subset of $\operatorname{ri} C_h$. Then there exists $\alpha \in \mathbb{R}$ s.t. $|\sigma_{\partial h(y)}(x-z)| \le \alpha|x-z|$, $|h(x)-h(y)| \le \alpha|x-y|$ and $|D_h^\sharp(x,y)| \le 2\alpha|x-y|$ for all $x,y,z \in S$.*

(b) *Let $h = \delta_S$, where δ_S is the indicator function of a convex polyhedral set $S \neq \emptyset$ in $\mathbb{R}^n$. Then h satisfies condition* (2.5).

(c) *Let h be a proper polyhedral convex function on $\mathbb{R}^n$. Then h satisfies condition* (2.5).

(d) *Let h be a closed proper convex function on $\mathbb{R}$. Then h is continuous on C_h, and $D_h^\sharp(y^*,y^k) \to 0$ if $y^k \to y^* \in C_h$, $\{y^k\} \subset C_h$.*

Lemma 2.4. (a) *Let $h = \sum_{i=1}^k h_i$, where $h_1,\ldots,h_k$ are closed proper convex functions s.t. $h_{j+1},\ldots,h_k$ $(j \ge 0)$ are polyhedral and $\cap_{i=1}^j \operatorname{ri}(C_{h_i}) \cap_{i=j+1}^k C_{h_i} \neq \emptyset$. If h_1 satisfies condition* (c) *of Def.* 2.1, *then so does h. If $h_1,\ldots,h_j$ satisfy condition* (d) *of Def.* 2.1 *or* (2.5), *then so does h. If h_1 is a B-function, $h_2,\ldots,h_j$ are continuous on $C_h = \cap_{i=1}^k C_{h_i}$ and satisfy condition* (d) *of Def.* 2.1, *then h is a B-function. In particular, h is a B-function if so are $h_1,\ldots,h_j$.*

(b) *Let $h_1,\ldots,h_j$ be B-functions s.t. $\cap_{i=1}^j \operatorname{ri} C_{h_i} \neq \emptyset$. Then $h = \max_{i=1:j} h_i$ is a B-function.*

(c) *Let h_1 be a B-function and let h_2 be a closed proper convex function s.t. $C_{h_1} \subset \operatorname{ri} C_{h_2}$. Then $h = h_1 + h_2$ is a B-function.*

(d) *Let $h_1,\ldots,h_n$ be closed proper strictly convex functions on $\mathbb{R}$ s.t. $\mathcal{L}_{h_i}^1(t,\alpha)$ is bounded for any $t,\alpha \in \mathbb{R}$, $i = 1{:}n$. Then $h(x) = \sum_{i=1}^n h_i(x_i)$ is a B-function.*

Lemma 2.5. *Let h be a proper convex function on $\mathbb{R}$. Then $\mathcal{L}_h^1(x,\alpha)$ is bounded for each $x \in C_h$ and $\alpha \in \mathbb{R}$ iff $C_{h^*} = \mathring{C}_{h^*}$.*

Lemma 2.6. (a) *If ψ is a B-function on $\mathbb{R}$ then ψ^* is essentially smooth and $C_{\psi^*} = \mathring{C}_{\psi^*}$.*

(b) *If $\phi : \mathbb{R} \to (-\infty,\infty]$ is closed proper convex essentially smooth and $C_\phi = \mathring{C}_\phi$ then ϕ^* is a B-function with $\operatorname{ri} C_{\phi^*} \subset \operatorname{im} \nabla\phi \subset C_{\phi^*}$.*

Proof. (a): This follows from Def. 2.1, Lem. 2.5 and [Roc70, Thm 26.3]. (b): By [Roc70, Thms 23.4, 23.5 and 26.1], $\operatorname{ri} C_{\phi^*} \subset C_{\partial\phi^*} = \operatorname{im}\partial\phi = \operatorname{im}\nabla\phi \subset C_{\phi^*}$ and ϕ^* is strictly convex on $C_{\partial\phi^*}$, and hence on C_{ϕ^*} by an elementary argument. Since ϕ^* is closed proper convex and $\phi^{**} = \phi$ [Roc70, Thm 12.2], the conclusion follows from Lems. 2.3(d) and 2.5. □

Examples 2.7. Let $\psi : \mathbb{R} \to (-\infty, \infty]$ and $h(x) = \sum_{i=1}^n \psi(x_i)$. In each of the examples, it can be verified that h is an essentially smooth B-function.

1 [Eck93]. $\psi(t) = |t|^\alpha/\alpha$ for $t \in \mathbb{R}$ and $\alpha > 1$, i.e., $h(x) = \|x\|_\alpha^\alpha/\alpha$. Then $h^*(\cdot) = \|\cdot\|_\beta^\beta/\beta$ with $\alpha + \beta = \alpha\beta$ [Roc70, p. 106]. For $\alpha = 1/2$, $h(x) = |x|^2/2$ and $D_h(x, y) = |x - y|^2/2$.

2. $\psi(t) = -t^\alpha/\alpha$ if $t \geq 0$ and $\alpha \in (0, 1)$, $\psi(t) = \infty$ if $t < 0$, i.e., $h(x) = -\|x\|_\alpha^\alpha/\alpha$ if $x \geq 0$. Then $h^*(y) = -\|y\|_\beta^\beta/\beta$ if $y < 0$ and $\alpha + \beta = \alpha\beta$, $h^*(y) = \infty$ if $y \not< 0$ [Roc70, p. 106].

3 ('$x \log x$'-entropy) [Bre67]. $\psi(t) = t \ln t$ if $t \geq 0$ ($0 \ln 0 = 0$), $\psi(t) = \infty$ if $t < 0$. Then $h^*(y) = \sum_{i=1}^n \exp(y_i - 1)$ [Roc70, p. 105] and $D_h(x, y) = \sum_{i=1}^n x_i \ln(x_i/y_i) + y_i - x_i$ (the Kullback-Liebler entropy).

4 [Teb92]. $\psi(t) = t \ln t - t$ if $t \geq 0$, $\psi(t) = \infty$ if $t < 0$. Then $h^*(y) = \sum_{i=1}^n \exp(y_i)$ [Roc70, p. 105] and D_h is the Kullback-Liebler entropy.

5 [Teb92]. $\psi(t) = -(1 - t^2)^{1/2}$ if $t \in [-1, 1]$, $\psi(t) = \infty$ otherwise. Then $h^*(y) = \sum_{i=1}^n (1 + y_i^2)^{1/2}$ [Roc70, p. 106] and $D_h(x, y) = \sum_{i=1}^n \frac{1 - x_i y_i}{(1 - y_i^2)^{1/2}} - (1 - x_i^2)^{1/2}$ on $[-1, 1]^n \times (-1, 1)^n$. (If $\psi(t) = -[2t(1 - 2t)]^{1/2}$ for $t \in [0, 1]$, $\psi^*(t) = (1 + t^2/4)^{1/2} + \frac{1}{2}t$.)

6 (Burg's entropy) [CDPI91]. $\psi(t) = -\ln t$ if $t > 0$, $\psi(t) = \infty$ if $t \leq 0$. Then $h^*(y) = -n - \sum_{i=1}^n \ln(-y_i)$ if $y < 0$, $h^*(y) = \infty$ if $y \not< 0$, and $D_h(x, y) = -\sum_{i=1}^n \{\ln(x_i/y_i) - x_i/y_i\} - n$.

7 [Teb92]. $\psi(t) = (\alpha t - t^\alpha)/(1 - \alpha)$ if $t \geq 0$ and $\alpha \in (0, 1)$, $\psi(t) = \infty$ if $t < 0$. Then $h^*(y) = \sum_{i=1}^n (1 - y_i/\beta)^{-\beta}$ for $y \in C_h^* = (-\infty, \beta)^n$, where $\beta = \alpha/(1 - \alpha)$. For $\alpha = \frac{1}{2}$, $D_h(x, y) = \sum_{i=1}^n (x_i^{1/2} - y_i^{1/2})^2/y_i^{1/2}$.

3 The BPM method

We make the following standing assumptions about problem (1.1) and the algorithm.

Assumption 3.1. (i) f is a closed proper convex function.
(ii) X is a nonempty closed convex set.
(iii) h is a (possibly nonsmooth) B-function.
(iv) $C_{f_X} \cap C_h \neq \emptyset$, where $f_X = f + \delta_X$ is the *essential objective* of (1.1).
(v) $\{c_k\}$ is a sequence of positive numbers satisfying $\sum_{k=1}^\infty c_k = \infty$.
(vi) $\{\epsilon_k\}$ is a sequence of nonnegative numbers satisfying $\lim_{l\to\infty} \sum_{k=1}^l c_k\epsilon_k / \sum_{k=1}^l c_k = 0$.

Consider the following *inexact BPM method.* At iteration $k \geq 1$, having

$$x^k \in C_{f_X} \cap C_{\partial h}, \tag{3.1}$$

$$\gamma^k \in \partial h(x^k), \tag{3.2}$$

$$D_h^k(x, x^k) = h(x) - h(x^k) - \left\langle \gamma^k, x - x^k \right\rangle \quad \forall x, \tag{3.3}$$

find x^{k+1}, γ^{k+1} and p^{k+1} satisfying

$$\gamma^{k+1} \in \partial h(x^{k+1}), \tag{3.4}$$

$$c_k p^{k+1} + \gamma^{k+1} - \gamma^k = 0, \tag{3.5}$$

$$p^{k+1} \in \partial_{\epsilon_k} f_X(x^{k+1}), \tag{3.6}$$

$$f_X(x^{k+1}) + D_h^k(x^{k+1}, x^k)/c_k \leq f_X(x^k). \tag{3.7}$$

We note that $x^{k+1} \approx \arg\min\{f_X + D_h^k(\cdot, x^k)/c_k\}$. By (2.1), (2.2), (3.2) and (3.3)

$$0 \leq D_h^\flat(x, x^k) \leq D_h^k(x, x^k) \leq D_h^\sharp(x, x^k) \quad \forall x, \tag{3.8}$$

so (cf. (3.7)) $x^{k+1} \in X$ and $f(x^{k+1}) \leq f(x^k)$. In fact x^{k+1} is an ϵ_k-minimizer of

$$\phi_k(x) = f_X(x) + D_h^k(x, x^k)/c_k, \tag{3.9}$$

as shown after the following (well-known) technical result (cf. [Roc70, Thm 27.1]).

Lemma 3.2. *A closed proper and strictly convex function ϕ on $\mathbb{R}^n$ has a unique minimizer iff ϕ is inf-compact, i.e., the α-level set $\mathcal{L}_\phi(\alpha) = \{x : \phi(x) \leq \alpha\}$ is bounded for any $\alpha \in \mathbb{R}$, and this holds iff $\mathcal{L}_\phi(\alpha)$ is nonempty and bounded for one $\alpha \in \mathbb{R}$.*

Proof. If $x \in \operatorname{Arg\,min} \phi$ then, by strict convexity of ϕ, $\mathcal{L}_\phi(\phi(x)) = \{x\}$ is bounded, so ϕ is inf-compact (cf. [Roc70, Cor. 8.7.1]). If for some $\alpha \in \mathbb{R}$, $\mathcal{L}_\phi(\alpha) \neq \emptyset$ is bounded then it is closed (cf. [Roc70, Thm 7.1]) and contains $\operatorname{Arg\,min} \phi \neq \emptyset$ because ϕ is closed. □

Lemma 3.3. *Under the above assumptions, we have:*

(i) *ϕ_k is closed proper and strictly convex.*

(ii) *$\phi_k(x^{k+1}) \leq \inf \phi_k + \epsilon_k$ (i.e., $0 \in \partial_{\epsilon_k} \phi_k(x^{k+1})$).*

(iii) *If $f_* = \inf_X f > -\infty$ then ϕ_k is inf-compact.*

(iv) *ϕ_k is inf-compact if $(\gamma^k - c_k \operatorname{im} \partial f_X) \cap \operatorname{im} \partial h \neq \emptyset$, where $\operatorname{im} \partial h = \mathring{C}_{h^*}$, so that $\operatorname{im} \partial h = \mathbb{R}^n$ iff h is cofinite. In particular, ϕ_k is inf-compact if $(\gamma^k - c_k \operatorname{ri} C_{f_X^*}) \cap \operatorname{ri} C_{h^*} \neq \emptyset$.*

(v) *If ϕ_k is inf-compact and either $\operatorname{ri} C_{f_X} \cap \operatorname{ri} C_h \neq \emptyset$, or $C_{f_X} \cap \operatorname{ri} C_h \neq \emptyset$ and f_X is polyhedral, then there exist $\hat{x}^{k+1} = \arg\min \phi_k$, $\hat{p}^{k+1} \in \partial f_X(\hat{x}^{k+1})$ and $\hat{\gamma}^{k+1} \in \partial h(\hat{x}^{k+1})$ s.t. $f_X(\hat{x}^{k+1}) + D_h^k(\hat{x}^{k+1}, x^k)/c_k \leq f_X(x^k)$ and $c_k \hat{p}^{k+1} + \hat{\gamma}^{k+1} - \gamma^k = 0$; also $\hat{x}^{k+1} \in \mathring{C}_h$ if $C_{\partial f_X} \subset \mathring{C}_h$ or $C_{\partial h} = \mathring{C}_h$, e.g., h is essentially smooth.*

(vi) *The assumptions of* (v) *hold if either $\operatorname{ri} C_{f_X} \subset \mathring{C}_h$ and $\inf_X f > -\infty$, or $C_{\partial f_X} \subset \mathring{C}_h$ and $\operatorname{im} \partial h = \mathbb{R}^n$.*

Proof. (i) Since f, δ_X and h are closed proper convex, so are $f_X = f + \delta_X$, $D_h^k(\cdot, x^k)$ and $\phi_k = f_X + D_h^k(\cdot, x^k)/c_k$ (cf. [Roc70, Thm 9.3]), having nonempty domains $C_f \cap X$, C_h and $C_{f_X} \cap C_h$ respectively (cf. Assumption 3.1(iv)). $D_h^k(\cdot, x^k)$ and ϕ_k are strictly convex, since so is h (cf. Def. 2.1(a)).

(ii) For any x, add the inequality $D_h^k(x, x^k) \geq D_h^k(x^{k+1}, x^k) + \left\langle \gamma^{k+1} - \gamma^k, x - x^{k+1} \right\rangle$ (cf. (3.3), (3.4)) divided by c_k to $f_X(x) \geq f_X(x^{k+1}) + \left\langle p^{k+1}, x - x^{k+1} \right\rangle - \epsilon_k$ (cf. (3.6)) and use (3.5) to get $\phi_k(x) \geq \phi_k(x^{k+1}) - \epsilon_k$.

(iii) By part (i), $\psi = D_h^k(\cdot, x^k)$ is closed proper strictly convex, and $\mathcal{L}_\psi(0) = \{x^k\}$ by strict convexity of h (cf. Def. 2.1(a), (2.2) and (1.4)), so ψ is inf-compact (cf. Lem. 3.2). Let $\beta = \inf \phi_k$. Since $\psi \geq 0$ (cf. (3.8)), $\beta \geq f_*$ and $\emptyset \neq \mathcal{L}_{\phi_k}(\beta + 1) \subset \mathcal{L}_\psi(c_k(\beta - f_* + 1))$ (cf. (3.9)). The last set is bounded, since ψ is inf-compact, so ϕ_k is inf-compact by part (i) and Lem. 3.2.

(iv) Let $\hat{y} \in C_{\partial f_X}$, $\hat{\gamma} \in \partial f_X(\hat{y})$, $\tilde{x} \in C_{\partial h}$ and $\tilde{\gamma} \in \partial h(\tilde{x})$ satisfy $\gamma^k - c_k\hat{\gamma} = \tilde{\gamma}$. Then $\tilde{\psi}(\cdot) = f_X(\hat{y}) + \langle \hat{\gamma}, \cdot - \hat{y}\rangle + D_h^k(\cdot, x^k)/c_k$ is closed proper and strictly convex (so is $D_h^k(\cdot, x^k)$; cf. part (i)), and $\tilde{x} = \arg\min \tilde{\psi}$ because $0 \in \partial\tilde{\psi}(\tilde{x}) = \hat{\gamma} + (\partial h(\tilde{x}) - \gamma^k)/c_k$ (cf. [Roc70, Thm 23.8]). Hence $\tilde{\psi}$ is inf-compact (cf. Lem. 3.2), and so is ϕ_k, since $\phi_k \geq \tilde{\psi}$ from $f_X(\cdot) \geq f_X(\hat{y}) + \langle \hat{\gamma}, \cdot - \hat{y}\rangle$. To see that strict convexity of h (cf. Def. 2.1(a)) implies $\operatorname{im} \partial h = \mathring{C}_{h^*}$, we note that $\mathring{C}_{h^*} = C_{\partial h^*}$ by [Roc70, Thms 26.3 and 26.1], and $\partial h^* = (\partial h)^{-1}$ by [Roc70, Thm 23.5], so that $C_{\partial h^*} = \operatorname{im} \partial h$. Of course, $\mathring{C}_{h^*} = \mathbb{R}^n$ iff $C_{h^*} = \mathbb{R}^n$, i.e., iff h is cofinite. The second assertion follows from $\operatorname{ri} C_{f_X^*} \subset C_{\partial f_X^*} = \operatorname{im} \partial f_X$.

(v) By part (i) and Lem. 3.2, $\hat{x}^{k+1} = \arg\min \phi_k$ is well defined. The rest follows from $D_h^k(\cdot, x^k) \geq 0$ (cf. (3.8)), the fact $0 \in \partial \phi_k(\hat{x}^{k+1}) = \partial f(\hat{x}^{k+1}) + c_k(\partial h(\hat{x}^{k+1}) - \gamma^k)$ due to our assumptions on C_{f_X} and $\operatorname{ri} C_h$ (cf. [Roc70, Thm 23.8]), and [Roc70, Thm 26.1].

(vi) If $\inf_X f > -\infty$ or $\operatorname{im} \partial h = \mathbb{R}^n$ then ϕ_k is inf-compact by parts (iii)–(iv). If $\operatorname{ri} C_{f_X} \subset \mathring{C}_h$ then $\operatorname{ri} C_{f_X} \cap \operatorname{ri} C_h = \operatorname{ri} C_{f_X} \neq \emptyset$, since $C_{f_X} \neq \emptyset$ (cf. Assumption 3.1(iv)). □

Remark 3.4. Lemma 3.3(v,vi) states conditions under which the *exact* BPM method (with $x^{k+1} = \hat{x}^{k+1} = \arg\min \phi_k$ and $\epsilon_k = 0$ in (3.6)) is well defined. Our conditions are slightly weaker than those in [Eck93, Thm 5], which correspond to $\operatorname{ri} C_{f_X} \subset \mathring{C}_h$, and either $\operatorname{cl} C_{f_X} \subset \mathring{C}_h$ and $\operatorname{im} \partial h = \mathbb{R}^n$, or f being finite, continuous and bounded below on X.

Example 3.5. Let $X = \{x \geq 0 : Ax = b\}$, $f = \langle \hat{c}, \cdot \rangle + \delta_X$ and $h(x) = -\sum_{i=1}^n \ln x_i$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $\hat{c} \in \mathbb{R}^n$. Suppose $f_* > -\infty$ and $Ax = b$ for some $x > 0$. Since $\mathring{C}_h = \{x : x > 0\}$, Lem. 3.3(iii,v) implies that $\hat{x}^{k+1}$ is well defined.

Example 3.6. Let $n = 1$, $X = \mathbb{R}$, $f(x) = -x$ and $h(x) = e^{-x} + x$. Then $f^* = \delta_{\{-1\}}$, $\operatorname{ri} C_{f^*} = \operatorname{im} \partial f = \{-1\}$, $\mathring{C}_{h^*} = \operatorname{im} \partial h = (-\infty, 1)$ and $\operatorname{ri} C_{f^*} \cap \mathring{C}_{h^*} \neq \emptyset$. Clearly, $\phi_k(x) = e^{-x} + x(e^{-x^k} - 1) + \text{const}$ for $c_k = 1$, so $\arg\min \phi_k \neq \emptyset$ iff $x^k < 0$. Although h is not a Bregman function, this is a counterexample to [Teb92, Thm 3.1].

4 Convergence of the BPM method

We first derive a global convergence rate estimate for the BPM method. We follow the analysis of [ChT93], which generalized that in [Gül91]. Let $s_k = \sum_{j=1}^k c_j$ for all k.

Lemma 4.1. *For all $x \in C_h$ and $k \leq l$, we have*

$$D_h^{k+1}(x, x^{k+1}) + D_h^k(x^{k+1}, x^k) - D_h^k(x, x^k) = \left\langle \gamma^k - \gamma^{k+1}, x - x^{k+1} \right\rangle \leq c_k[f_X(x) - f_X(x^{k+1})] + c_k\epsilon_k, \quad (4.1)$$

$$D_h^{k+1}(x, x^{k+1}) \leq D_h^k(x, x^k) - D_h^k(x^{k+1}, x^k) + c_k\epsilon_k \quad \text{if} \quad f_X(x) \leq f_X(x^{k+1}), \quad (4.2)$$

$$f_X(x^{k+1}) \le f_X(x^k), \tag{4.3}$$

$$s_l[f_X(x^{l+1}) - f_X(x)] \le D_h^1(x,x^1) - D_h^{l+1}(x,x^{l+1}) - \sum_{k=1}^{l}(s_k/c_k)D_h^k(x^{k+1},x^k) + \sum_{k=1}^{l} c_k\epsilon_k, \tag{4.4}$$

$$f_X(x^{l+1}) - f_X(x) \le D_h^1(x,x^1)/s_l + \sum_{k=1}^{l} c_k\epsilon_k/s_l. \tag{4.5}$$

Proof. The equality in (4.1) follows from (3.3), and the inequality from $\gamma^k - \gamma^{k+1} = c_k p^{k+1}$ (cf. (3.5)) and $p^{k+1} \in \partial_{\epsilon_k} f_X(x^{k+1})$ (cf. (3.6)), i.e., $\langle p^{k+1}, x - x^{k+1}\rangle \le f_X(x) - f_X(x^{k+1}) + \epsilon_k$, since $c_k > 0$. (4.2) is a consequence of (4.1). (4.3) follows from (cf. (3.7), (3.8)) $f_X(x^k) - f_X(x^{k+1}) \ge D_h^k(x^{k+1},x^k)/c_k \ge 0$. Multiplying the last inequality by $s_{k-1} = s_k - c_k$ (with $s_0 = 0$) and summing over $k = 1\colon l$ yields

$$- s_l f_X(x^{l+1}) + \sum_{k=1}^{l} c_k f_X(x^{k+1}) \ge \sum_{k=1}^{l}(s_{k-1}/c_k)D_h^k(x^{k+1},x^k). \tag{4.6}$$

Summing (4.1) over $k = 1\colon l$ we obtain

$$D_h^{l+1}(x,x^{l+1}) - D_h^1(x,x^1) + \sum_{k=1}^{l} D_h^k(x^{k+1},x^k) \le s_l f_X(x) - \sum_{k=1}^{l} c_k f_X(x^{k+1}) + \sum_{k=1}^{l} c_k\epsilon_k. \tag{4.7}$$

Subtract (4.6) from (4.7) and rearrange, using $1 + s_{k-1}/c_k = s_k/c_k$, to get (4.4). (4.5) follows from (4.4) and the fact $D_h^k(\cdot,x^k) \ge 0$ for all k (cf. (3.8)). □

We shall need the following two results proven in [TsB91].

Lemma 4.2 ([TsB91, Lem. 1]). *Let $h : \mathbb{R}^n \to (-\infty,\infty]$ be a closed proper convex function continuous on C_h. Then:*
(a) *For any $y \in C_h$, there exists $\epsilon > 0$ s.t. $\{x \in C_h : |x - y| \le \epsilon\}$ is closed.*
(b) *For any $y \in C_h$ and z s.t. $y + z \in C_h$, and any sequences $y^k \to y$ and $z^k \to z$ s.t. $y^k \in C_h$ and $y^k + z^k \in C_h$ for all k, we have $\limsup_{k\to\infty} h'(y^k;z^k) \le h'(y;z)$.*

Lemma 4.3. *Let $h : \mathbb{R}^n \to (-\infty,\infty]$ be a closed proper convex function continuous on C_h. If $\{y^k\} \subset C_h$ is a bounded sequence s.t., for some $y \in C_h$, $\{h(y^k) + h'(y^k; y - y^k)\}$ is bounded from below, then $\{h(y^k)\}$ is bounded and any limit point of $\{y^k\}$ is in C_h.*

Proof. Use the final paragraph of the proof of [TsB91, Lem. 2]. □

Lemmas 4.2–4.3 could be expressed in terms of the following analogue of (2.1)

$$D_h'(x,y) = h(x) - h(y) - h'(y;x-y) \quad \forall x,y \in C_h. \tag{4.8}$$

Lemma 4.4. *Let $h : \mathbb{R}^n \to (-\infty,\infty]$ be a closed proper strictly convex function continuous on C_h. If $y^* \in C_h$ and $\{y^k\}$ is a bounded sequence in C_h s.t. $D_h'(y^*,y^k) \to 0$ then $y^k \to y^*$.*

Proof. Let y^∞ be the limit of a subsequence $\{y^k\}_{k\in K}$. Since $h(y^k)+h'(y^k;y^*-y^k)=h(y^*)-D'_h(y^*,y^k)\to h(y^*)$, $y^\infty\in C_h$ by Lem. 4.3 and $h(y^k)\xrightarrow{K}h(y^\infty)$ by continuity of h on C_h. Then by Lem. 4.2(b), $0=\liminf_{k\in K}D'_h(y^*,y^k)\ge h(y^*)-h(y^\infty)-h'(y^\infty;y^*-y^\infty)$ yields $y^\infty=y^*$ by strict convexity of h. Hence $y^k\to y^*$. □

By (1.4), (3.2), (3.3), (2.2) and (4.8), for all k

$$0\le D'_h(x,x^k)\le D^\flat_h(x,x^k)\le D^k_h(x,x^k)\le D^\sharp_h(x,x^k)\quad\forall x. \tag{4.9}$$

Lemma 4.5. *If $\sum_{k=1}^\infty c_k\epsilon_k<\infty$ and $x\in C_h$ is s.t. $f_X(x^k)\ge f_X(x)$ for all k then:*
(i) *$\{x^k\}$ is bounded and $\{x^k\}\subset\mathcal{L}^1_h(x,\alpha)$, where $\alpha=D^1_h(x,x^1)+\sum_{k=1}^\infty c_k\epsilon_k$.*
(ii) *Every limit point of $\{x^k\}$ is in C_h.*
(iii) *$\{x^k\}$ converges to some $x^\infty\in C_{f_X}\cap C_h$ s.t. $f_X(x^k)\ge f_X(x^\infty)$ for all k.*

Proof. (i) We have $D^l_h(x,x^l)\le D^1_h(x,x^1)+\sum_{k=1}^{l-1}c_k\epsilon_k\le\alpha$ for all l (cf. (4.2), (3.8)) and $\{x^k\}\subset C_{\partial h}$ (cf. (3.1)), so $\{x^k\}\subset\mathcal{L}^1_h(x,\alpha)$, a bounded set (cf. Def. 2.1(c)).

(ii) $D'_h(x,x^k)\le D^k_h(x,x^k)\le\alpha$ implies $h(x^k)+h'(x^k;x-x^k)\ge h(x)-\alpha$ for all k (cf. (4.8), (4.9)), so the desired conclusion follows from continuity of h on C_h (cf. Def. 2.1(b)), $\{x^k\}$ being bounded in C_h (cf. (3.1) and part (i)) and Lem. 4.3.

(iii) By parts (i)–(ii), a subsequence $\{x^{l_j}\}$ converges to some $x^\infty\in C_h$. Suppose $x^\infty\ne x$. Since $\{x^k\}\subset\mathcal{L}^1_h(x,\alpha)$, $D^\sharp_h(x^\infty,x^{l_j})\to 0$ (cf. Def. 2.1(d)) and $D^{l_j}_h(x^\infty,x^{l_j})\to 0$ (cf. (3.8)). But $f_X(x^k)\ge f_X(x^\infty)$ for all k, since $x^{l_j}\to x^\infty$, $f_X(x^{k+1})\le f_X(x^k)$ (cf. (4.3)) and f_X is closed (cf. Assumption 3.1(i,ii)). Hence for $l>l_j$, $D^l_h(x^\infty,x^l)\le D^{l_j}_h(x^\infty,x^{l_j})+\sum_{k=l_j}^{l-1}c_k\epsilon_k$ (cf. (4.2)) with $\sum_{k=l_j}^\infty c_k\epsilon_k\to 0$ as $j\to\infty$ yield $D^l_h(x^\infty,x^l)\to 0$ as $l\to\infty$. Thus $D'_h(x^\infty,x^k)\to 0$ (cf. (4.9)) and $x^k\to x^\infty$ by Lem. 4.4. Finally, if $x^\infty=x$ but $\{x^k\}$ does not converge, it has a limit point $x'\ne x^\infty$ (cf. parts (i)–(ii)), and replacing x and x^∞ by x^∞ and x' respectively in the preceding argument yields a contradiction. □

We may now prove our main result for the inexact BPM descent method (3.1)–(3.7).

Theorem 4.6. *Suppose Assumption 3.1(i–ii,iv–v) holds with h closed proper convex.*
(a) *If $\lim_{l\to\infty}\sum_{k=1}^l c_k\epsilon_k/\sum_{k=1}^l c_k=0$ then $f_X(x^k)\downarrow\inf_{C_h}f_X=\inf_{\mathrm{cl}(C_h\cap C_{f_X})}f$. Hence $f_X(x^k)\downarrow\inf_X f$ if $C_{f_X}\subset C_h$. If $\mathrm{ri}\,C_h\cap\mathrm{ri}\,C_{f_X}\ne\emptyset$ (e.g., $\mathring{C}_h\cap C_{f_X}\ne\emptyset$) then $\inf_{C_h}f_X=\inf_{(\mathrm{cl}\,C_h)\cap(\mathrm{cl}\,C_{f_X})}f=\inf_{\mathrm{cl}\,C_h}f_X$. If $\mathrm{ri}\,C_{f_X}\subset\mathrm{cl}\,C_h$ (e.g., $C_{\partial f_X}\subset\mathrm{cl}\,C_h$) then $\mathrm{cl}\,C_h\supset\mathrm{cl}\,C_{f_X}$ and $\mathrm{Arg\,min}_X f\subset\mathrm{cl}\,C_h$.*
(b) *If h is a B-function, $f_X(x^k)\to\inf_{C_h}f_X$, $\sum_{k=1}^\infty c_k\epsilon_k<\infty$ and $X_*=\mathrm{Arg\,min}_{C_h}f_X$ is nonempty then $\{x^k\}$ converges to some $x^\infty\in X_*$, and $x^\infty\in\mathrm{Arg\,min}_X f$ if $C_{f_X}\subset C_h$.*
(c) *If $f_X(x^k)\to\inf_{C_h}f_X$, $C_{f_X}\subset C_h$ and $X_*=\emptyset$ then $|x^k|\to\infty$.*

Proof. (a) For any $x\in C_h$, taking the limit in (4.5) yields $\lim_{l\to\infty}f_X(x^l)\le f_X(x)$, using $f_X(x^{l+1})\le f_X(x^l)$ (cf. (4.3)), $s_l\to\infty$ (cf. Assumption 3.1(v)) and $\sum_{k=1}^l c_k\epsilon_k/s_l\to 0$. Hence $f_X(x^k)\to\inf_{C_h}f_X=\inf_{C_h\cap C_{f_X}}f=\inf_{\mathrm{cl}(C_h\cap C_{f_X})}f$ (cf. [Roc70, Cor. 7.3.2]). If $\mathrm{ri}\,C_h\cap\mathrm{ri}\,C_{f_X}\ne\emptyset$ (e.g, $\mathring{C}_h\cap C_{f_X}\ne\emptyset$; cf. [Roc70, Cor. 6.3.2]) then $\mathrm{cl}(C_h\cap C_{f_X})=\mathrm{cl}(C_h)\cap\mathrm{cl}(C_{f_X})$ (cf. [Roc70, Thm 6.5]) and $\inf_{C_h}f_X=\inf_{(\mathrm{cl}\,C_h)\cap(\mathrm{cl}\,C_{f_X})}f\le\inf_{C_{f_X}\cap\mathrm{cl}\,C_h}f=\inf_{\mathrm{cl}\,C_h}f_X$, so $\inf_{C_h}f_X=\inf_{\mathrm{cl}\,C_h}f_X$. If $\mathrm{ri}\,C_{f_X}\subset\mathrm{cl}\,C_h$ then $\mathrm{cl}\,C_{f_X}\subset\mathrm{cl}\,C_h$ (cf. [Roc70, Thm 6.5]).

(b) If $x \in X_*$ then $f_X(x^k) \to f_X(x)$. But $f_X(x^k) \geq f_X(x)$ for all k (cf. (3.1)), so $x^k \to x^\infty \in C_{f_X} \cap C_h$ and $\lim_{k\to\infty} f_X(x^k) \geq f_X(x^\infty)$ by Lem. 4.5, and thus $x^\infty \in X_*$.

(c) If $|x^k| \not\to \infty$, $\{x^k\}$ has a limit point x with $f_X(x) \leq \inf_{C_h} f_X \leftarrow f_X(x^k)$ (f_X is closed; cf. Assumption 3.1(i,ii)), so $C_{f_X} \subset C_h$ yields $x \in C_h \cap X_*$, i.e., $X_* \neq \emptyset$. □

Remark 4.7. For the exact BPM method (with $\epsilon_k \equiv 0$), Thm 4.6(a,b) subsumes [ChT93, Thm 3.4], which assumes $\operatorname{ri} C_{f_X} \subset \mathring{C}_h$ and $C_h = \operatorname{cl} C_h$. Thm 4.6(b,c) strengthens [Eck93, Thm 5], which only shows that $\{x^k\}$ is unbounded if $\operatorname{cl} C_{f_X} \subset \mathring{C}_h$ and $X_* = \emptyset$. Thm 4.6(a,b) and Lem. 3.3 subsume [Ius95, Thm 4.1], which assumes that h is essentially smooth, f is continuous on C_f, $C_f \cap \mathring{C}_h \neq \emptyset$, $X = \operatorname{cl} C_h$, $\operatorname{Arg\,min}_X f \neq \emptyset$ and $\inf_k c_k > 0$.

For choosing $\{\epsilon_k\}$ (cf. Assumption 3.1(vi)), one may use the following simple result.

Lemma 4.8. (i) *If $\epsilon_k \to 0$ then $\sum_{k=1}^l c_k\epsilon_k/s_l \to 0$ as $l \to \infty$.*
(ii) *If $\sum_{k=1}^\infty \epsilon_k < \infty$ and $\{c_k\} \subset (0, c_{\max}]$ for some $c_{\max} < \infty$ then $\sum_{k=1}^\infty c_k\epsilon_k < \infty$.*

Proof. (i) For any $\epsilon > 0$, pick $\bar{k}$ and $\bar{l} > \bar{k}$ s.t. $\epsilon_k \leq \epsilon$ for all $k \geq \bar{k}$ and $\sum_{k=1}^{\bar{k}} c_k\epsilon_k/s_l \leq \epsilon$ for all $l \geq \bar{l}$; then $\sum_{k=1}^l c_k\epsilon_k/s_l \leq \sum_{k=1}^{\bar{k}} c_k\epsilon_k/s_l + \epsilon\sum_{k=\bar{k}+1}^l c_k/\sum_{k=1}^l c_k \leq 2\epsilon$ for all $l \geq \bar{l}$.

(ii) We have $\sum_{k=1}^\infty c_k\epsilon_k \leq c_{\max}\sum_{k=1}^\infty \epsilon_k < \infty$. □

5 Convergence of a nondescent BPM method

In certain applications (cf. §7) it may be difficult to satisfy the descent requirement (3.7). Hence we now consider a *nondescent BPM method*, in which (3.7) is replaced by

$$f_X(x^{k+1}) + D_h^k(x^{k+1}, x^k)/c_k \leq f_X(x^k) + \epsilon_k. \tag{5.1}$$

By Lem. 3.3(ii), (5.1) holds *automatically*, since it means $\phi_k(x^{k+1}) \leq \phi_k(x^k) + \epsilon_k$.

Lemma 5.1. *For all $x \in C_h$ and $k \leq l$, we have*

$$f_X(x^{k+1}) \leq f_X(x^k) + \epsilon_k, \tag{5.2}$$

$$s_l[f_X(x^{l+1}) - f_X(x)] \leq D_h^1(x, x^1) - D_h^{l+1}(x, x^{l+1}) - \sum_{k=1}^l (s_k/c_k) D_h^k(x^{k+1}, x^k) + \sum_{k=1}^l s_k\epsilon_k, \tag{5.3}$$

$$f_X(x^{l+1}) - f_X(x) \leq D_h^1(x, x^1)/s_l + \sum_{k=1}^l s_k\epsilon_k/s_l. \tag{5.4}$$

Proof. (4.1)–(4.2) still hold. (5.2) follows from $D_h^k(x^{k+1}, x^k) \geq 0$ (cf. (3.8)) and (cf. (5.1)) $f_X(x^k) - f_X(x^{k+1}) \geq D_h^k(x^{k+1}, x^k)/c_k - \epsilon_k$. Multiplying this inequality by $s_{k-1} = s_k - c_k$ and summing over $k = 1\colon l$ yields

$$-s_l f_X(x^{l+1}) + \sum_{k=1}^l c_k f_X(x^{k+1}) \geq \sum_{k=1}^l (s_{k-1}/c_k) D_h^k(x^{k+1}, x^k) - \sum_{k=1}^l s_{k-1}\epsilon_k. \tag{5.5}$$

Subtract (5.5) from (4.7) and rearrange, using $s_k = s_{k-1} + c_k$, to get (5.3). (5.4) follows from (5.3) and the fact $D_h^k(\cdot, x^k) \geq 0$ for all k (cf. (3.8)). □

Theorem 5.2. *Suppose Assumption* 3.1(i–ii,iv–v) *holds with h closed proper convex.*

(a) *If $\sum_{k=1}^{l} s_k\epsilon_k/s_l \to 0$ (see Lem. 5.3 for sufficient conditions), then $f_X(x^k) \to \inf_{C_h} f_X$. Hence the assertions of Theorem* 4.6(a) *hold.*

(b) *If h is a B-function, $f_X(x^k) \to \inf_{C_h} f_X$, $\sum_{k=1}^{\infty} c_k\epsilon_k < \infty$ and $X_* = \operatorname{Arg\,min}_{C_h} f_X$ is nonempty then $\{x^k\}$ converges to some $x^\infty \in X_*$, and $x^\infty \in \operatorname{Arg\,min}_X f$ if $C_{f_X} \subset C_h$.*

(c) *If $f_X(x^k) \to \inf_{C_h} f_X$, $C_{f_X} \subset C_h$ and $X_* = \emptyset$ then $|x^k| \to \infty$.*

Proof. (a) The upper limit in (5.4) for any $x \in C_h$ yields $\limsup_{l\to\infty} f_X(x^l) \le \inf_{C_h} f_X$, using $\sum_{k=1}^{l} s_k\epsilon_k/s_l \to 0$. But $\{x^k\} \subset C_h$ (cf. (3.1)), so $\liminf_{l\to\infty} f_X(x^l) \ge \inf_{C_h} f_X$.

(b) If $x \in X_*$ then $f_X(x^k) \to f_X(x)$ and $f_X(x^k) \ge f_X(x)$ for all k (cf. (3.1)). Assertions (i)–(iii) of Lem. 4.5 still hold, since the proofs of (i)–(ii) remain valid, whereas in the proof of (iii) we have $x^\infty \in C_h$ and $f_X(x^\infty) \le \lim_{j\to\infty} f_X(x^{l_j}) = f_X(x)$ (f_X is closed), so $x^\infty \in X_*$ and $f_X(x^k) \ge f_X(x^\infty)$ for all k as before yield $x^k \to x^\infty$.

(c) Use the proof of Thm 4.6(c). □

Lemma 5.3. (i) *Let $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\varepsilon_k\}$ be sequences in $\mathbb{R}$ s.t. $0 \le \alpha_{k+1} \le (1-\beta_k)\alpha_k + \varepsilon_k$, $\alpha_1 \ge 0$, $0 < \beta_k \le 1$, $\varepsilon_k \ge 0$ for $k = 1, 2, \ldots$, $\sum_{k=1}^{\infty} \beta_k = \infty$ and $\lim_{k\to\infty} \varepsilon_k/\beta_k = 0$. Then $\lim_{k\to\infty} \alpha_k = 0$.*

(ii) *If $\sum_{l=1}^{\infty} c_l/s_l = \infty$ and $\lim_{k\to\infty} \epsilon_k s_k/c_k = 0$ then $\lim_{l\to\infty} \sum_{k=1}^{l} s_k\epsilon_k/s_l = 0$.*

(iii) *If $\{c_k\} \subset [c_{\min}, c_{\max}]$ for some $0 < c_{\min} \le c_{\max}$ and $k\epsilon_k \to 0$ then $\sum_{k=1}^{l} s_k\epsilon_k/s_l \to 0$.*

Proof. (i) See, e.g., [Pol83, Lem. 2.2.3].

(ii) Use part (i) with $\alpha_l = \sum_{k=1}^{l} s_k\epsilon_k/s_l$, $s_l = \sum_{k=1}^{l} c_k$ and $\alpha_{l+1} = (1 - c_{l+1}/s_{l+1})\alpha_l + \epsilon_{l+1}$.

(iii) Use part (ii) with $c_l/s_l \in [c_{\min}/lc_{\max}, c_{\max}/lc_{\min}]$ for all l. □

6 Finite termination for sharp minima

We now extend to the exact BPM method the finite convergence property of the PPA in the case of sharp minima (cf. [Fer91, Roc76b] and [BuF93]).

Theorem 6.1. *Let f have a sharp minimum on X, i.e., $X_* = \operatorname{Arg\,min}_X f \ne \emptyset$ and there exists $\alpha > 0$ s.t. $f_X(x) \ge \min_X f + \alpha \min_{y\in X_*} |x - y|$ for all x. Consider the exact BPM method applied to* (1.1) *with a B-function h s.t. $C_{f_X} \subset C_{\nabla h}$, $\epsilon_k \equiv 0$ and $\inf_k c_k > 0$. Then there exists k s.t. $p^k = 0$ and $x^k \in X_*$.*

Proof. By Thm 4.6, $x^k \to x^\infty \in X_*$, so $x^\infty \in C_{\nabla h}$, $\gamma^k = \nabla h(x^k) \to \nabla h(x^\infty)$ (cf. (3.2) and continuity of ∇h on $C_{\nabla h}$ [Roc70, Thm 25.5]) and $\partial f_X(x^k) \ni p^k = (\gamma^{k-1} - \gamma^k)/c_{k-1} \to 0$ (cf. (3.5)–(3.6)). But if $x \notin X_*$ and $\gamma \in \partial f_X(x)$ then $|\gamma| \ge \alpha$ (cf. [Ber82, §5.4]) (since for $y = \arg\min_{y\in X_*} |x - y|$, $\min_X f = f_X(y) \ge f_X(x) + \langle \gamma, y - x\rangle$ yields $|\gamma||x - y| \ge \langle \gamma, x - y\rangle \ge \alpha|x - y|$). Hence for some k, $|p^k| < \alpha$ implies $p^k = 0$ and $x^k \in X_*$. □

We note that piecewise linear programs have sharp minima, if any (cf. [Ber82, §5.4]).

7 Inexact multiplier methods

Following [Eck93, Teb92], this section considers the application of the BPM method to dual formulations of convex programs of the form presented in [Roc70, §28]:

$$\text{minimize} \quad f(x), \quad \text{subject to} \quad g_i(x) \leq 0, \ i = 1{:}\,m, \tag{7.1}$$

under the following

Assumption 7.1. f, $g_1, \ldots, g_m$ are closed proper convex functions on $\mathbb{R}^n$ with $C_f \subset \bigcap_{i=1}^m C_{g_i}$ and $\operatorname{ri} C_f \subset \bigcap_{i=1}^m \operatorname{ri} C_{g_i}$.

Letting $g(\cdot) = (g_1(\cdot), \ldots, g_m(\cdot))$, we define the *Lagrangian* of (7.1)

$$L(x, \pi) = \begin{cases} f(x) + \langle \pi, g(x) \rangle & \text{if } x \in C_f \text{ and } \pi \in \mathbb{R}^m_+, \\ -\infty & \text{if } x \in C_f \text{ and } \pi \notin \mathbb{R}^m_+, \\ \infty & \text{if } x \notin C_f, \end{cases}$$

and the *dual functional* $d(\pi) = \inf_x L(x, \pi)$. Then $d(\pi) = -\infty$ if $\pi \notin \mathbb{R}^m_+$. Assume that $d(\pi) > -\infty$ for some π. The *dual problem* to (7.1) is to maximize d, or equivalently to minimize $q(\pi)$ over $\pi \geq 0$, where $q = -d$ is a closed proper convex function. We will apply the BPM method to this problem, using some B-function h on $\mathbb{R}^m$.

We assume that $\mathbb{R}^m_> \subset C_h$, so that $h_+ = h + \delta_{\mathbb{R}^m_+}$ is a B-function (cf. Lem. 2.4(a)). The *monotone conjugate* of h (cf. [Roc70, p. 111]) defined by $h^+(\cdot) = \sup_{\pi \geq 0}\{\langle \pi, \cdot \rangle - h(\pi)\}$ is *nondecreasing* (i.e., $h^+(u) \leq h^+(u')$ if $u \leq u'$, since $\langle \pi, u \rangle \leq \langle \pi, u' \rangle \ \forall \pi \geq 0$) and coincides with the convex conjugate h_+^* of h_+, since $h^+(\cdot) = \sup_\pi\{\langle \pi, \cdot \rangle - h_+(\pi)\} = h_+^*(\cdot)$. We need the following variation on [Eck93, Lem. A3]. Its proof is given in the Appendix.

Lemma 7.2. *If h is a closed proper essentially strictly convex function on $\mathbb{R}^m$ with $\mathbb{R}^m_+ \cap \operatorname{ri} C_h \neq \emptyset$, then h^+ is closed proper convex and essentially smooth, $\partial h^+(u) = \{\nabla h^+(u)\}$ for all $u \in C_{\partial h^+}$, $\partial h^+ = (\partial h_+)^{-1}$ and ∇h^+ is continuous on $C_{\partial h^+} = \mathring{C}_{h^+} = \operatorname{im} \partial h_+$. Further, $C_{h^+} = C_{h^+} - \mathbb{R}^m_+$, $\mathring{C}_{h^+} = \mathring{C}_{h^+} - \mathbb{R}^m_+$, $\partial h_+ = \partial h + N_{\mathbb{R}^m_+}$ and $\nabla h^+ = \nabla h^+ \circ (I + N_{\mathbb{R}^m_+} \circ \nabla h^+)$, where I is the identity operator and $N_{\mathbb{R}^m_+} = \partial \delta_{\mathbb{R}^m_+}$ is the normal cone operator of $\mathbb{R}^m_+$, i.e., $N_{\mathbb{R}^m_+}(\pi) = \{\gamma \leq 0 : \langle \gamma, \pi \rangle = 0\}$ if $\pi \geq 0$, $N_{\mathbb{R}^m_+}(\pi) = \emptyset$ if $\pi \ngeq 0$. If additionally $\operatorname{im} \partial h \supset \mathbb{R}^m_>$ then h_+ is cofinite, $C_{h^+} = \mathbb{R}^m$ and h^+ is continuously differentiable.*

Since $\mathbb{R}^m_> \subset C_{h_+} \subset \mathbb{R}^m_+$, to find $\inf_{\pi \geq 0} q(\pi)$ via the BPM method we replace in (3.1)–(3.6) f, X, h and x^k by q, $\mathbb{R}^m$, h_+ and π^k respectively. Given $\pi^k \in C_q \cap C_{\partial h_+}$ and $\gamma^k \in \partial h_+(\pi^k)$, our *inexact multiplier method* requires finding π^{k+1} and x^{k+1} s.t.

$$L(x^{k+1}, \pi^{k+1}) \leq \inf_x L(x, \pi^{k+1}) + \epsilon_k = d(\pi^{k+1}) + \epsilon_k, \tag{7.2}$$

$$\pi^{k+1} = \nabla h^+(\gamma^k + c_k g(x^{k+1})) \tag{7.3}$$

with

$$p^{k+1} \in \partial_{\epsilon_k} q(\pi^{k+1}), \tag{7.4}$$

$$\gamma^{k+1} = \gamma^k - c_k p^{k+1} \in \partial h_+(\pi^{k+1}) \tag{7.5}$$

for some p^{k+1} and γ^{k+1}. Note that (7.2) implies

$$-g(x^{k+1}) \in \partial_{\epsilon_k} q(\pi^{k+1}) = \partial_{\epsilon_k} q(\pi^{k+1}) + \partial \delta_{\mathbb{R}^m_+}(\pi^{k+1}), \tag{7.6}$$

since $-d = q \geq \tilde{q} := -f(x^{k+1}) - \langle \cdot, g(x^{k+1}) \rangle = \tilde{q}(\pi^{k+1}) + \langle -g(x^{k+1}), \cdot - \pi^{k+1} \rangle$ and $C_q \subset \mathbb{R}^m_+$ from $q = \sup_x -L(x, \cdot)$, and $\tilde{q}(\pi^{k+1}) \geq q(\pi^{k+1}) - \epsilon_k$ (cf. (7.2)). Next, (7.3) gives $\pi^{k+1} \in C_{\partial h_+} \subset C_{h_+} \subset \mathbb{R}^m_+$, whereas $q(\pi^{k+1}) \leq q(\pi^k) + \epsilon_k$ (cf. (5.1)) yields $\pi^{k+1} \in C_q$. By (7.6), (7.4)–(7.5) hold if we take $p^{k+1} = (\gamma^k - \gamma^{k+1})/c_k$ and

$$\gamma^{k+1} = \gamma^k + c_k g(x^{k+1}) - \tilde{\gamma}^{k+1} \in \partial h_+(\pi^{k+1}) \quad \text{with} \quad \tilde{\gamma}^{k+1} \in N_{\mathbb{R}^m_+}(\pi^{k+1}), \tag{7.7}$$

since then

$$p^{k+1} = -g(x^{k+1}) + \tilde{\gamma}^{k+1}/c_k \in \partial_{\epsilon_k} q(\pi^{k+1}). \tag{7.8}$$

Using (7.3) and $(\partial h_+)^{-1} = \nabla h^+$ (Lem. 7.2), we have

$$\gamma^k + c_k g(x^{k+1}) \in \partial h_+(\pi^{k+1}) = \partial h(\pi^{k+1}) + N_{\mathbb{R}_+}(\pi^{k+1}), \tag{7.9}$$

so we may take $\tilde{\gamma}^{k+1} = 0$; other choices will be discussed later.

Further insight may be gained as follows. Rewrite (7.3) as

$$\pi^{k+1} = \nabla P_k(g(x^{k+1})), \tag{7.10}$$

where

$$P_k(u) = h^+(\gamma^k + c_k u)/c_k \quad \forall u \in \mathbb{R}^m. \tag{7.11}$$

Let

$$L_k(x) = f(x) + \frac{1}{c_k}[h^+(\gamma^k + c_k g(x)) - h^+(\gamma^k)] \tag{7.12}$$

if $x \in C_f$ ($\subset C_g = \bigcap_{i=1}^m C_{g_i}$; cf. Assumption 7.1), $L_k(x) = \infty$ otherwise.

Lemma 7.3. *Suppose* $\inf_{C_f} \max_{i=1}^m g_i \leq 0$, *e.g., the* feasible set $C_0 = \{x \in C_f : g(x) \leq 0\}$ *of* (7.1) *is nonempty. Then L_k is a proper convex function and*

$$\partial L_k(x) = \partial f(x) + \sum_{i=1}^m [\nabla P_k(g(x))]_i \partial g_i(x) \quad \forall x \in C_{L_k} \supset C_{\partial L_k}. \tag{7.13}$$

If $\partial L_k(x) \neq \emptyset$ then $\pi = \nabla P_k(g(x))$ is well defined, $\pi \geq 0$ and $\partial L_k(x) = \partial_x L(x, \pi)$, where

$$\partial_x L(x, \pi) = \partial f(x) + \sum_{i=1}^m \pi_i \partial g_i(x) \quad \forall x \in \mathbb{R}^n, \forall \pi \in \mathbb{R}^m_+. \tag{7.14}$$

If $\hat{x} \in \operatorname{Arg\,min} L_k$ then $\hat{x} \in \operatorname{Arg\,min}_x L(x, \hat{\pi})$ for $\hat{\pi} = \nabla P_k(g(\hat{x}))$. The preceding assertions hold when $\inf_{C_f} \max_{i=1}^m g_i > 0$ but $C_{h^+} = \mathbb{R}^m$, e.g., if $\operatorname{im} \partial h \supset \mathbb{R}^m_>$ (cf. Lem. 7.2).

Proof. Using $\gamma^k \in \partial h_+(\pi^k) \subset \mathring{C}_{h^+}$ (cf. Lem. 7.2) and $\mathring{C}_{P_k} = (\mathring{C}_{h^+} - \gamma^k)/c_k$, pick $\tilde{u} \in C_{P_k} \cap \mathbb{R}^m_>$ and $\tilde{x} \in C_f$ s.t. $g(\tilde{x}) < \tilde{u}$. Then, since P_k is nondecreasing (so is h^+) and $\operatorname{ri} C_f \subset \bigcap_i \operatorname{ri} C_{g_i}$ (cf. Assumption 7.1), Lem. A.1 in the Appendix yields $\operatorname{im} \partial P_k \subset \mathbb{R}^m_+$ and (7.13), using $\partial P_k = \{\nabla P_k\}$ (cf. Lem. 7.2). Hence if $\partial L_k(x) \neq \emptyset$ then $\pi = \nabla P_k(g(x)) \geq 0$, so $\operatorname{ri} C_f \subset \bigcap_i \operatorname{ri} C_{g_i}$ implies (cf. [Roc70, Thm 23.8]) $\partial_x L(x, \pi) = \partial f(x) + \sum_i \pi_i \partial g_i(x) = \partial L_k(x)$. If $\hat{x} \in \operatorname{Arg\,min} L_k$ then $0 \in \partial L_k(\hat{x}) = \partial_x L(\hat{x}, \hat{\pi})$ for $\hat{\pi} = \nabla P_k(g(\hat{x}))$ yields $\hat{x} \in \operatorname{Arg\,min}_x L(x, \hat{\pi})$. Finally, when $C_{h^+} = \mathbb{R}^m$ then for any $\tilde{x} \in C_f$ we may pick $\tilde{u} \in C_{P_k}$ with $g(\tilde{x}) < \tilde{u}$, since $C_f \subset \bigcap_i C_{g_i}$ (Assumption 7.1) and $C_{P_k} = \mathbb{R}^m$. □

The *exact* multiplier method of [Eck93, Thm 7] takes $x^{k+1} \in \operatorname{Arg\,min} L_k$ and $\pi^{k+1} = \nabla P_k(g(x^{k+1}))$, assuming h is smooth, $\mathring{C}_h \supset \mathbb{R}^m_>$ and $\operatorname{im} \nabla h \supset \mathbb{R}^m_>$. Then (7.2) holds with $\epsilon_k = 0$ (cf. Lem. 7.3). Our inexact method only requires that $x^{k+1} \tilde{\in} \operatorname{Arg\,min} L_k$ in the sense that (7.2) holds for a given $\epsilon_k \geq 0$. Thus we have derived the following

Algorithm 7.4. *At iteration $k \geq 1$, having $\pi^k \in C_q$ and $\gamma^k \in \partial h_+(\pi^k)$, find*

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_{x \in C_f} \left\{ f(x) + \frac{1}{c_k} h^+(\gamma^k + c_k g(x)) \right\},$$
$$\pi^{k+1} = \nabla h^+(\gamma^k + c_k g(x^{k+1}))$$

s.t. (7.2) *holds, choose γ^{k+1} satisfying* (7.7) *and set $p^{k+1} = (\gamma^k - \gamma^{k+1})/c_k$.*

To find x^{k+1} as in [Ber82, §5.3], suppose f is *strongly convex*, i.e., for some $\breve{\alpha} > 0$

$$f(x) \geq f(\bar{x}) + \langle \gamma, x - \bar{x} \rangle + \breve{\alpha}|x - \bar{x}|^2/2 \quad \forall x, \bar{x}, \forall \gamma \in \partial f(\bar{x}). \tag{7.15}$$

Adding subgradient inequalities of g_i, $i = 1{:}\,m$, and using (7.14) yields for all x

$$L(x, \pi^{k+1}) \geq L(x^{k+1}, \pi^{k+1}) + \left\langle \gamma, x - x^{k+1} \right\rangle + \breve{\alpha}|x - x^{k+1}|^2/2 \quad \forall \gamma \in \partial_x L(x^{k+1}, \pi^{k+1}). \tag{7.16}$$

Let $\Delta_x L_k(x^{k+1}) = \arg\min_{\gamma \in \partial L_k(x^{k+1})} |\gamma|$, assuming $\partial L_k(x^{k+1}) \neq \emptyset$ and $\partial_x L(x^{k+1}, \pi^{k+1}) = \partial L_k(x^{k+1})$ (e.g., $C_0 \neq \emptyset$ or $C_{h^+} = \mathbb{R}^m$; cf. Lem. 7.3). Minimization in (7.16) yields

$$d(\pi^{k+1}) \geq L(x^{k+1}, \pi^{k+1}) - |\Delta_x L_k(x^{k+1})|^2/2\breve{\alpha}, \tag{7.17}$$

so (7.2) holds if

$$|\Delta_x L_k(x^{k+1})|^2/2\breve{\alpha} \leq \epsilon_k. \tag{7.18}$$

Thus, as in the multiplier methods of [Ber82, §5.3], one may use any algorithm for minimizing L_k that generates a sequence $\{z^j\}$ such that $\liminf_{j \to \infty} |\Delta_x L_k(z^j)| = 0$, setting $x^{k+1} = z^j$ when (7.18) occurs. (If $\breve{\alpha}$ is unknown, it may be replaced in (7.18) by any fixed $\bar{\alpha} > 0$; this only scales $\{\epsilon_k\}$.) Of course, the strong convexity assumption is not necessary if one can employ the direct criterion (7.2), i.e., $L(z^j, \pi) \leq d(\pi) + \epsilon_k$ with $\pi = \nabla P_k(g(z^j))$ (cf. (7.10)), where $d(\pi)$ may be computed with an *error* that can be absorbed in ϵ_k.

Some examples are now in order.

Example 7.5. Suppose $h(\pi) = \sum_{i=1}^m h_i(\pi_i)$, where h_i are B-functions on $\mathbb{R}$ with $C_{h_i} \supset \mathbb{R}_>$, $i = 1{:}\,m$ (cf. Lem. 2.4(d)). For each i, let $\bar{u}_i = h_i'(0; 1)$ if $0 \in C_{h_i}$, $\bar{u}_i = -\infty$ if $0 \notin C_{h_i}$, so that (cf. [Eck93, Ex. 6]) $h_i^+(u_i) = h_i^*(\max\{u_i, \bar{u}_i\})$ and $\nabla h_i^+(u_i) = \max\{0, \nabla h_i^*(u_i)\}$. Using (7.9) and "maximal" γ^{k+1} in (7.7), Alg. 7.4 may be written as

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^m h_i^*(\max\{\bar{u}_i, \gamma_i^k + c_k g_i(x)\}) \right\}, \tag{7.19a}$$

$$\pi_i^{k+1} = \max\{0, \nabla h_i^*(\gamma_i^k + c_k g_i(x^{k+1})\}, \quad i = 1{:}\,m, \tag{7.19b}$$

$$\gamma_i^{k+1} = \max\{\bar{u}_i, \gamma_i^k + c_k g_i(x^{k+1})\}, \quad i = 1{:}\,m. \tag{7.19c}$$

Remark 7.6. To justify (7.19c), note that if we had $\gamma^k < \bar{u} \in \mathbb{R}^m$, then (7.19a) would not penalize constraint violations $g_i(x) \in (0, (\bar{u}_i - \gamma_i^k)/c_k]$. An *ordinary* penalty method (cf. [Ber82, p. 354]) would use (7.19a,b) with $\gamma^k \equiv \bar{u}$ and $c_k \uparrow \infty$. Thus (7.19) is a *shifted* penalty method, in which the *shifts* γ^k should ensure convergence even for $\sup_k c_k < \infty$, thus avoiding the ill-conditioning of ordinary penalty methods.

Example 7.7. Suppose $C_{\partial h} \cap \mathbb{R}_+^m = C_{\nabla h} \cap \mathbb{R}_+^m$, so that $\partial h_+ = \nabla h + \partial \delta_{\mathbb{R}_+^m}$ from $\mathbb{R}_>^m \subset C_h$ (cf. [Roc70, Thms 23.8 and 25.1]). Then we may use $\gamma^k = \nabla h(\pi^k)$ for all k, since the maximal shift $\gamma^{k+1} = \nabla h(\pi^{k+1})$ satisfies (7.7) due to (7.9). Thus Alg. 7.4 becomes

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{c_k} h^+(\nabla h(\pi^k) + c_k g(x)) \right\},$$
$$\pi^{k+1} = \nabla h^+(\nabla h(\pi^k) + c_k g(x^{k+1})).$$

In the separable case of Ex. 7.5, the formulae specialize to

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^m h_i^*(\max\{\bar{u}_i, \nabla h_i(\pi_i^k) + c_k g_i(x)\}) \right\},$$
$$\pi_i^{k+1} = \max\{0, \nabla h_i^*(\nabla h_i(\pi_i^k) + c_k g_i(x^{k+1}))\}, \quad i = 1\colon m,$$

where $\bar{u}_i = \nabla h_i(0)$ if $0 \in C_{\partial h_i}$, $\bar{u}_i = -\infty$ if $0 \notin C_{\partial h_i}$, $i = 1\colon m$.

Example 7.8. Let $h(\pi) = \sum_{i=1}^m \psi(\pi_i)$, where ψ is a B-function on $\mathbb{R}$ with $C_{\nabla\psi} \supset \mathbb{R}_>$. Let $\bar{v} = \psi'(0;1)$ if $0 \in C_\psi$, $\bar{v} = -\infty$ if $0 \notin C_\psi$. Then $\partial\psi_+(t) = \{\psi'(t;1)\}$ for $t > 0$, $\partial\psi_+(0) = (-\infty, \bar{v}]$ if $\bar{v} > -\infty$, $\partial\psi_+(0) = \emptyset$ if $\bar{v} = -\infty$. Using (7.7) and (7.9) as in Ex. 7.5, we may let $\gamma_i^{k+1} = \psi'(\pi_i^{k+1}; 1)$, $i = 1\colon m$. Thus Alg. 7.4 becomes

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^m \psi^*(\max\{\bar{v}, \psi'(\pi_i^k; 1) + c_k g_i(x)\}) \right\}, \tag{7.20a}$$
$$\pi_i^{k+1} = \max\{0, \nabla\psi^*(\psi'(\pi_i^k; 1) + c_k g_i(x^{k+1}))\}, \quad i = 1\colon m. \tag{7.20b}$$

Example 7.9. For $\psi(t) = |t|^\alpha/\alpha$ with $\alpha > 1$ and $\beta = \alpha/(\alpha - 1)$ (cf. Ex. 2.7.1), (7.20) becomes

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{\beta c_k} \sum_{i=1}^m \max\{0, (\pi_i^k)^{1/(\beta-1)} + c_k g_i(x)\}^\beta \right\}, \tag{7.21a}$$
$$\pi_i^{k+1} = \max\{0, (\pi_i^k)^{1/(\beta-1)} + c_k g_i(x^{k+1})\}^{\beta-1}, \quad i = 1\colon m. \tag{7.21b}$$

Even if f and all g_i are smooth, for $\beta = 2$ the objective of (7.21a) is, in general, only once continuously differentiable. This is a well-known drawback of quadratic augmented Lagrangians (cf. [Ber82, TsB93]). However, for $\beta = 3$ we obtain a *cubic multiplier method* [Kiw96] with a *twice* continuously differentiable objective.

Example 7.10 ([Eck93, Ex. 7]). For $\psi(t) = t \ln t - t$ (cf. Ex. 2.7.4), (7.20) reduces to

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^m \pi_i^k \exp[c_k g_i(x)] \right\}, \tag{7.22a}$$
$$\pi_i^{k+1} = \pi_i^k \exp[c_k g_i(x^{k+1})], \quad i = 1\colon m, \tag{7.22b}$$

i.e., to an inexact *exponential multiplier method* (cf. [Ber82, §5.1.2], [TsB93]).

Example 7.11. For $\psi(t) = -\ln t$ (cf. Ex. 2.7.6), (7.20) reduces to

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) - \frac{1}{c_k} \sum_{i=1}^m \ln[1/\pi_i^k - c_k g_i(x)] \right\},$$
$$\pi_i^{k+1} = \pi_i^k / [1 - c_k \pi_i^k g_i(x^{k+1})], \quad i = 1{:}\,m,$$

i.e., to an inexact *shifted logarithm barrier method* (which was also derived heuristically in [Cha94, Ex. 4.2]). This method is related, but not indentical, to ones in [CGT92, GMSW88]; cf. [CGT94].

Example 7.12. If $\psi(t) = -t^\alpha/\alpha$, $\alpha \in (0, 1)$ (cf. Ex. 2.7.2), (7.20) reduces to

$$x^{k+1} \tilde{\in} \operatorname{Arg\,min}_x \left\{ f(x) - \frac{1}{\beta c_k} \sum_{i=1}^m [(\pi_i^k)^{1/(\beta-1)} - c_k g_i(x)]^\beta \right\},$$
$$\pi_i^{k+1} = [(\pi_i^k)^{1/(\beta-1)} - c_k g_i(x^{k+1})]^{\beta-1}, \quad i = 1{:}\,m,$$

where $\beta = \alpha/(\alpha - 1)$; $\beta = -1$ corresponds to a *shifted Carroll barier method.*

8 Convergence of multiplier methods

In addition to Assumption 7.1, we make the following standing assumptions.

Assumption 8.1. (i) h_+ is a B-function s.t. $C_{h_+} \supset \mathbb{R}^m_>$ (e.g., so is h; cf. Lem. 2.4(a)).
(ii) Either $C_q \cap \mathbb{R}^m_> \neq \emptyset$ or $\emptyset \neq C_q \subset C_{h_+}$, where $-q = d = \inf_x L(x, \cdot)$.
(iii) $\{c_k\}$ is a sequence of positive numbers s.t. $s_k = \sum_{j=1}^k c_j \to \infty$.

Remark 8.2. Under Assumption 8.1, q is closed proper convex, $\mathring{C}_{h_+} = \mathbb{R}^m_> \subset C_{h_+} \subset \mathbb{R}^m_+$, $\operatorname{cl} C_{h_+} = \mathbb{R}^m_+ \supset C_q$, $C_q \cap \mathring{C}_{h_+} \neq \emptyset$ if $C_q \cap \mathbb{R}^m_> \neq \emptyset$, and $\inf_{C_{h_+}} q = \inf q = \inf_{\operatorname{cl} C_{h_+}} q$. Hence for the BPM method applied to the dual problem $\sup d = -\inf q$ with a B-function h_+ we may invoke the results of §§3–6 (replacing f, X and h by q, $\mathbb{R}^m$ and h_+ respectively).

Theorem 8.3. *If* $\sum_{j=1}^k s_j \epsilon_j / s_k \to 0$ *(cf. Lem.* 5.3*), then* $d(\pi^k) \to \sup d$. *If* $d(\pi^k) \to \sup d$, $C_{h_+} \cap \operatorname{Arg\,max} d \neq \emptyset$ *and* $\sum_{k=1}^\infty c_k \epsilon_k < \infty$ *then* $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$. *If* $d(\pi^k) \to \sup d$, $C_q \subset C_{h_+}$ *and* $\operatorname{Arg\,max}_{C_{h_+}} d = \emptyset$ *(e.g.,* $C_{h_+} = \mathbb{R}^m_+$ *and* $\operatorname{Arg\,max} d = \emptyset$*) then* $|\pi^k| \to \infty$.

Proof. This follows from Rem. 8.2 and Thm 5.2, since $C_{h_+} \cap \operatorname{Arg\,max} d \subset \operatorname{Arg\,max}_{C_{h_+}} d \subset \operatorname{Arg\,max} d$ if $C_{h_+} \cap \operatorname{Arg\,max} d \neq \emptyset$. □

Theorem 8.4. *Let* $C_{\nabla h} \supset \mathbb{R}^m_+$, $\gamma^k = \nabla h(\pi^k)$ *for all* k *(cf. Ex.* 7.7*) and* $\sum_{j=1}^k s_j \epsilon_j / s_k \to 0$. *Then* $d(\pi^k) \to \sup d$. *If* $\operatorname{Arg\,max} d \neq \emptyset$ *and* $\sum_{k=1}^\infty c_k \epsilon_k < \infty$ *then* $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$, *and if* $\inf_k c_k > 0$ *then*

$$\limsup_{k\to\infty} f(x^k) \le \sup_\pi d(\pi) \quad \textit{and} \quad \limsup_{k\to\infty} g_i(x^k) \le 0, \quad i = 1{:}\,m, \tag{8.1}$$

and every limit point of $\{x^k\}$ *solves* (7.1). *If* $\operatorname{Arg\,max} d = \emptyset$ *then* $|\pi^k| \to \infty$.

Proof. Since $C_h \supset C_{\nabla h} \supset \mathbb{R}^m_+$, the assertions about $\{\pi^k\}$ follow from Thm 8.3. Suppose $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$, $\inf_k c_k > 0$. Since $p^k = (\gamma^{k-1} - \gamma^k)/c_{k-1}$ with $p^k + g(x^k) \in N_{\mathbb{R}^m_+}(\pi^k)$ (cf. Ex. 7.7), we have (cf. Lem. 7.2) $\langle \pi^k, g(x^k)\rangle = -\langle \pi^k, p^k\rangle$ and $g(x^k) \le -p^k$ $\forall k > 1$, with $p^k \to 0$, since $\pi^k \to \pi^\infty$, ∇h is continuous on $\mathbb{R}^m_+$ and $c_k \ge c_{\min}$ $\forall k$. Hence $\langle \pi^k, g(x^k)\rangle \to 0$ and $\limsup_{k\to\infty} g_i(x^k) \le 0$ $\forall i$. Since $L(x^k, \pi^k) \le \inf_x L(x, \pi^k) + \epsilon_{k-1}$ (cf. (7.2)) means $f(x^k) + \langle \pi^k, g(x^k)\rangle \le f(x) + \langle \pi^k, g(x)\rangle + \epsilon_{k-1}$ for any x, in the limit $\limsup_k f(x^k) \le L(x, \pi^\infty)$ ($\epsilon_k \to 0$), so $\limsup_k f(x^k) \le d(\pi^\infty)$. Suppose $x^k \xrightarrow{K} x^\infty$ for some x^∞ and $K \subset \{1, 2, \ldots\}$. By (8.1), $f(x^\infty) \le \sup d$ and $g(x^\infty) \le 0$ (f and g are closed), so by weak duality, $f(x^\infty) \ge \sup d$, $f(x^\infty) = \max d$ and x^∞ solves (7.1). □

Remark 8.5. Let C_* denote the optimal solution set for (7.1). If (7.1) is consistent (i.e., $C_0 \neq \emptyset$), then C_* is nonempty and compact iff f and g_i, $i = 1{:}m$, have no common direction of recession [Ber82, §5.3], in which case (8.1) implies that $\{x^k\}$ is bounded, and hence has limit points. In particular, if $C_* = \{x^*\}$ then $x^k \to x^*$ in Thm 8.4.

Remark 8.6. Theorems 8.3–8.4 subsume [Eck93, Thm 7], which additionally requires that $\epsilon_k \equiv 0$, $\operatorname{im} \nabla h \supset \mathbb{R}^m_>$ and each g_i is continuous on C_f.

Theorem 8.7. *Let* (7.1) *be s.t.* $q = -d$ *has a sharp minimum. Let* $C_{\nabla h} \supset \mathbb{R}^m_+$, $\inf_k c_k > 0$, $\epsilon_k = 0$ *and* $\gamma^k = \nabla h(\pi^k)$ *(cf. Ex. 7.7) for all* k. *Then there exists* k *s.t.* $p^k = 0$, $\pi^k \in \operatorname{Arg\,max} d$ *and* x^k *solves* (7.1).

Proof. Using the proof of Thm 6.1 with $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d \subset C_{\nabla h}$ and $\gamma^k = \nabla h(\pi^k) \to \nabla h(\pi^\infty)$, we get k s.t. $\pi^k \in \operatorname{Arg\,max} d$ and $p^k = 0$; the conclusion follows from the proof of Thm 8.4. □

Remark 8.8. Results on finite convergence of other multiplier methods are restricted to only once continuously differentiable augmented Lagrangians [Ber82, §5.4], whereas Thm 8.7 covers Ex. 7.9 also with $\beta > 2$. Applications include polyhedral programs.

We shall need the following result, similar to ones in [Ber82, §5.3] and [TsB93].

Lemma 8.9. *With* $u^{k+1} := g(x^{k+1})$, *for each* k, *we have*

$$L(x^{k+1}, \pi^{k+1}) = L_k(x^{k+1}) + D_{h^+}(\gamma^k, \gamma^k + c_k u^{k+1})/c_k \ge L_k(x^{k+1}), \tag{8.2}$$

$$L_k(x^{k+1}) = L(x^{k+1}, \pi^k) + D_{h^+}(\gamma^k + c_k u^{k+1}, \gamma^k)/c_k \ge L(x^{k+1}, \pi^k), \tag{8.3}$$

$$L(x^{k+1}, \pi^{k+1}) - L(x^{k+1}, \pi^k) = \langle \pi^{k+1} - \pi^k, u^{k+1}\rangle \tag{8.4a}$$

$$= \langle \nabla h^+(\gamma^k + c_k u^{k+1}) - \nabla h^+(\gamma^k), u^{k+1}\rangle \ge 0, \tag{8.4b}$$

$$d(\pi^k) \le L(x^{k+1}, \pi^k) \le L_k(x^{k+1}) \le L(x^{k+1}, \pi^{k+1}) \le d(\pi^{k+1}) + \epsilon_k. \tag{8.5}$$

Proof. As for (8.2), use (7.12), (7.3), (2.3) and convexity of h^+ to develop

$$\begin{aligned} L(x^{k+1},\pi^{k+1}) - L_k(x^{k+1}) &= \left\langle \pi^{k+1}, u^{k+1} \right\rangle - [h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k)]/c_k \\ &= [h^+(\gamma^k) - h^+(\gamma^k + c_k u^{k+1}) \\ &\quad - \left\langle \nabla h^+(\gamma^k + c_k u^{k+1}), -c_k u^{k+1} \right\rangle]/c_k \\ &= D_{h^+}(\gamma^k, \gamma^k + c_k u^{k+1})/c_k \ge 0. \end{aligned}$$

Since $\nabla h^+ = (\partial h_+)^{-1}$ (cf. Lem. 7.2) and $\gamma^k \in \partial h_+(\pi^k)$ (cf. (7.5)), $\pi^k = \nabla h^+(\gamma^k)$, so

$$\begin{aligned} L_k(x^{k+1}) - L(x^{k+1},\pi^k) &= [h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k)]/c_k - \left\langle \pi^k, u^{k+1} \right\rangle \\ &= [h^+(\gamma^k + c_k u^{k+1}) - h^+(\gamma^k) - \left\langle \nabla h^+(\gamma^k), c_k u^{k+1} \right\rangle]/c_k \\ &= D_{h^+}(\gamma^k + c_k u^{k+1}, \gamma^k)/c_k \ge 0 \end{aligned}$$

yields (8.3), and (8.4) holds with $\left\langle \nabla h^+(\gamma^k + c_k u^{k+1}) - \nabla h^+(\gamma^k), c_k u^{k+1} \right\rangle / c_k \ge 0$ by the convexity of h^+. (8.5) follows from (8.2)–(8.4) and (7.2). □

Theorem 8.4 only covers methods with $C_{\nabla h} \supset \mathbb{R}^m_+$, such as Exs. 7.7 and 7.9. To handle other examples in §9, we shall use the following abstraction of the ergodic framework of [TsB93]. For each k, define the aggregate primal solution

$$\breve{x}^{k+1} = \sum_{j=1}^k c_j x^{j+1}/s_k, \quad \text{where} \quad s_k = \sum_{j=1}^k c_j. \tag{8.6}$$

Since g is convex and $c_j g(x^{j+1}) \le -c_j p^{j+1} = \gamma^{j+1} - \gamma^j$ for $j = 1{:}\,k$ by (7.7)–(7.8),

$$g(\breve{x}^{k+1}) \le \sum_{j=1}^k c_j g(x^{j+1})/s_k \le (\gamma^{k+1} - \gamma^1)/s_k. \tag{8.7}$$

Lemma 8.10. *Suppose* $\sup_{i,k} \gamma_i^k < \infty$, $\epsilon_k \to 0$, $\left\langle \pi^k, u^k \right\rangle \to 0$ *and* $d(\pi^k) \to d^\infty < \infty$. *Then*

$$\limsup_{k\to\infty} f(\breve{x}^k) \le d^\infty \quad \textit{and} \quad \limsup_{k\to\infty} g_i(\breve{x}^k) \le 0, \quad i = 1{:}\,m. \tag{8.8}$$

If $\{\breve{x}^k\}$ *has a limit point* x^∞ *(e.g.,* $C_* \ne \emptyset$ *is bounded; cf. Rem.* 8.11*), then* x^∞ *solves* (7.1)*,* $f(x^\infty) = d^\infty = \max d$ *and each limit point of* $\{\pi^k\}$ *maximizes* d.

Proof. By (8.7), $\limsup_k g_i(\breve{x}^k) \le 0$ $\forall i$, since $s_k \to \infty$. By (8.6) and convexity of f, $f(\breve{x}^{k+1}) \le \sum_{j=1}^k c_j f(x^{j+1})/s_k$, while $f(x^k) = L(x^k, \pi^k) - \left\langle \pi^k, u^k \right\rangle \to d^\infty$ from (8.5), so $\limsup_k f(\breve{x}^k) \le d^\infty$. Suppose $\breve{x}^k \xrightarrow{K} x^\infty$. By (8.8), $f(x^\infty) \le d^\infty$ and $g(x^\infty) \le 0$ (f and g are closed). Hence by weak duality, $f(x^\infty) \ge d^\infty \leftarrow d(\pi^k)$, $f(x^\infty) = d^\infty = \max d$ and x^∞ solves (7.1). Since $d(\pi^k) \to d^\infty$ and d is closed, each cluster of $\{\pi^k\}$ maximizes d. □

Remark 8.11. If $C_* \ne \emptyset$ is bounded then (8.8) implies that $\{\breve{x}^k\}$ is bounded (cf. Rem. 8.5). In particular, if $C_* = \{x^*\}$ then $\breve{x}^k \to x^*$ in Lem. 8.10.

9 Classes of penalty functions

Examples 7.10–7.12 stem from B-functions of the form $h(\pi) = \sum_{i=1}^m \psi(\pi_i)$, where ψ is a B-function on $\mathbb{R}$ s.t. $\psi_+ = \psi$. Since $\psi_+ = (\psi^+)^*$, such examples may also be derived by choosing suitable penalty functions ϕ on $\mathbb{R}$ and letting $\psi = \phi^*$ (cf. Lem. 2.6). We now define two classes of penalty functions and study their relations with B-functions.

Definition 9.1. We say $\phi \in \Phi$ iff $\phi : \mathbb{R} \to (-\infty, \infty]$ is closed proper convex essentially smooth, $\mathring{C}_\phi = C_\phi$ and $\mathbb{R}_> \subset \operatorname{im} \nabla\phi \subset \mathbb{R}_+$. Let $t_\phi = \sup_{t \in C_\phi} t$, $t_\phi^0 = \sup_{\nabla\phi(t)=0} t$, $\Phi_s = \{\phi \in \Phi : \phi \text{ is strictly convex}\}$ and $\Phi_0 = \{\phi \in \Phi : \phi \text{ is strictly convex on } (t_\phi^0, t_\phi), t_\phi^0 > -\infty\}$.

Remark 9.2. If $\phi \in \Phi$ then ϕ is nondecreasing ($\operatorname{im}\nabla\phi \subset \mathbb{R}_+$), $C_\phi = (-\infty, t_\phi)$, $t_\phi^0 = -\infty$ iff $\operatorname{im}\nabla\phi = \mathbb{R}_>$, $\phi \in \Phi_s$ iff $\nabla\phi$ is increasing, $\phi \in \Phi_0$ iff $\nabla\phi$ is increasing on (t_ϕ^0, t_ϕ) and $t_\phi^0 > -\infty$ (cf. [Roc70, p. 254]). Also $\phi \in \Phi$ iff ϕ is closed proper convex, $C_{\nabla\phi} = \mathring{C}_\phi = C_\phi$ and $\mathbb{R}_> \subset \operatorname{im}\nabla\phi \subset \mathbb{R}_+$. (For the "if" part, note that $\nabla\phi(t_k) \uparrow \infty$ if $t_k \uparrow t_\phi < \infty$, since $\mathbb{R}_> \subset \operatorname{im}\nabla\phi$ and $\nabla\phi$ is nondecreasing.)

Lemma 9.3. *If $\phi \in \Phi$ then ϕ^* is a B-function with $\mathbb{R}_> \subset C_{\phi^*} \subset \mathbb{R}_+$, $(\phi^*)^+ = \phi^{**} = \phi$, $\lim_{t\downarrow-\infty} \nabla\phi(t) = 0$, $\lim_{t\uparrow t_\phi} \nabla\phi(t) = \lim_{t\uparrow t_\phi} \phi(t) = \infty$ and $\phi 0^+ = \sigma_{\mathbb{R}_+}$. If $\phi \in \Phi_s$ then ϕ^* is essentially smooth, $C_{\partial\phi^*} = C_{\nabla\phi^*} = \mathbb{R}_>$ and $\partial\phi^*(0) = \emptyset$. If $\phi \in \Phi_0$ then $C_{\nabla\phi^*} = \mathbb{R}_>$ and $\partial\phi^*(0) = (-\infty, t_\phi^0]$.*

Proof. By Def. 9.1 and Lem. 2.6, $\mathbb{R}_> \subset \operatorname{im}\nabla\phi \subset \mathbb{R}_+$ and ϕ^* is a B-function with $\operatorname{ri} C_{\phi^*} \subset \operatorname{im}\nabla\phi \subset C_{\phi^*}$, so $\mathbb{R}_> \subset C_{\phi^*} \subset \mathbb{R}_+$. $C_{\phi^*} \subset \mathbb{R}_+$ yields $(\phi^*)^+ = \phi^{**} = \phi$. Since $\mathbb{R}_> \subset \operatorname{im}\nabla\phi \subset \mathbb{R}_+$ and $\nabla\phi$ is nondecreasing, $\lim_{t\downarrow-\infty} \nabla\phi(t) = 0$ and $\lim_{t\uparrow t_\phi} \nabla\phi(t) = \infty$. Since ϕ is closed and proper, $\phi 0^+ = \sigma_{C_{\phi^*}}$ [Roc70, Thm 13.3] with $\sigma_{C_{\phi^*}} = \sigma_{\operatorname{cl} C_{\phi^*}}$ and $\operatorname{cl} C_{\phi^*} = \mathbb{R}_+$ from $\mathbb{R}_> \subset C_{\phi^*} \subset \mathbb{R}_+$. If $t_\phi < \infty$ then $\lim_{t\uparrow t_\phi} \phi(t) = \infty$ from $t_\phi \notin C_\phi$ and closedness of ϕ; otherwise $\lim_{t\uparrow t_\phi} \phi(t) = \infty$ from $\infty = \phi 0^+(1) = \lim_{t\uparrow\infty}[\phi(t) - \phi(0)]/t$ [Roc70, Thm 8.5]. By [Roc70, Thm 26.1], $\partial\phi = \{\nabla\phi\}$ and $C_{\partial\phi} = \mathring{C}_\phi$. If $\phi \in \Phi_s$ then ϕ^* is essentially smooth [Roc70, Thm 26.3], so $\partial\phi^* = \{\nabla\phi^*\}$ and $C_{\nabla\phi^*} = C_{\partial\phi^*} = \mathring{C}_{\phi^*} = \mathbb{R}_>$ [Roc70, Thm 26.1]. If $\phi \in \Phi_0$ then $\partial\phi^* = (\partial\phi)^{-1} = \{(\nabla\phi)^{-1}\}$ yields $\partial\phi^*(0) = \{t : \nabla\phi(t) = 0\} = (-\infty, t_\phi^0]$ ($0 \le \nabla\phi(t) \le \nabla\phi(t_\phi^0)\ \forall t \le t_\phi^0$), whereas $\nabla\phi$ is increasing on (t_ϕ^0, t_ϕ) (also if $\phi \in \Phi_s$; cf. Rem. 9.2), so $\partial\phi^* = \{(\nabla\phi)^{-1}\}$ is single-valued on $\mathbb{R}_> = (\nabla\phi(t_\phi^0), \infty) \subset \operatorname{im}\nabla\phi$, and hence $\partial\phi^* = \{\nabla\phi^*\}$ on $\mathbb{R}_>$ [Roc70, Thm 25.1]. □

Lemma 9.4. *Let ψ be a B-function on $\mathbb{R}$ s.t. $C_\psi \supset \mathbb{R}_>$. Then $\psi^+ \in \Phi$. Suppose $C_{\nabla\psi} \supset \mathbb{R}_>$. If $\partial\psi(0) = \emptyset$ (i.e., $0 \notin C_\psi$ or $\psi'(0;1) = -\infty$) then ψ_+ is essentially smooth and $\psi^+ \in \Phi_s$. If $\partial\psi(0) \neq \emptyset$ (i.e., $\psi'(0;1) > -\infty$) then $\psi^+ \in \Phi_0$ with $t_{\psi^+}^0 = \psi'(0;1)$, and there exists a B-function $\breve{\psi}$ s.t. $\psi_+ = \breve{\psi}_+$, $\psi^+ = \breve{\psi}^+$, $C_{\nabla\breve{\psi}} \supset \mathbb{R}_+$ and $\nabla\breve{\psi}(0) = t_{\psi^+}^0$.*

Proof. $\psi_+ = \psi + \delta_{\mathbb{R}_+}$ is a B-function (Lem. 2.4(a)) and $\psi^+ = \psi_+^*$, so $C_{\psi^+} = \mathring{C}_{\psi^+}$ (Lem. 2.6(a)). Also ψ^+ is nondecreasing and essentially smooth (Lem. 7.2), so $\operatorname{im}\nabla\psi^+ \subset \mathbb{R}_+$, whereas $\mathbb{R}_> \subset C_{\psi_+}$ yields $\mathbb{R}_> \subset \mathring{C}_{\psi_+} \subset C_{\partial\psi_+} = \operatorname{im}\partial\psi^+ = \operatorname{im}\nabla\psi^+$. Suppose $C_{\nabla\psi} \supset \mathbb{R}_>$. By strict convexity of ψ (cf. Def. 2.1(a)), $\nabla\psi_+ = \nabla\psi$ is increasing on $\mathbb{R}_>$, so $\nabla\psi^+ = (\nabla\psi_+)^{-1}$ is increasing on $(t^0, \infty) \cap C_{\psi^+}$ with $t^0 = \lim_{t\downarrow 0} \nabla\psi(t)$, and hence ψ^+ is strictly convex on (t^0,∞) (cf. [Roc70, p. 254]). If $\partial\psi(0)=\emptyset$, then $t^0=-\infty$, $\psi^+\in\Phi_s$ and ψ_+ is essentially smooth [Roc70, Thm 26.3]. Otherwise, $t^0=\psi'(0;1)=t^0_{\psi^+}$. Let $\breve\psi(t)=\psi(t)$ $\forall t\geq 0$, and let $\breve\psi(t)$ for $t\leq 0$ be a strictly convex quadratic function s.t. $\breve\psi(0)=\psi(0)$ and $\breve\psi'(0;-1)=-\psi'(0;1)$. Then $\breve\psi_+=\psi_+$ and $\nabla\breve\psi(0)=t^0_{\psi^+}$. □

Corollary 9.5. *If $\phi\in\Phi_0$ then the method of Ex. 7.8 with $\psi=\phi^*$ coincides with the method of Ex. 7.7 with $h(\pi)=\sum_{i=1}^m\breve\psi(\pi_i)$, where $\breve\psi$ is the smooth extension of ψ described in Lem. 9.4, so that $C_{\nabla h}\supset\mathbb{R}^m_+$ and Thms 8.4 and 8.7 apply.*

Proof. We have $C_{\nabla\psi}=\mathbb{R}_>$, $\partial\psi(0)=(-\infty,t^0_\phi]$ and $\psi^+=\psi^*=\phi$ for $\psi=\phi^*$ (Lem. 9.3), so $\psi'(0;1)=\nabla\breve\psi(0)=t^0_\phi$ and $\psi'(t;1)=\nabla\breve\psi(t)$ if $t>0$ (Lem. 9.4). □

Remark 9.6. In terms of $\phi\in\Phi_0$, the method of Ex. 7.8 with $\psi=\phi^*$ becomes

$$x^{k+1}\mathrel{\tilde\in}\operatorname{Arg\,min}_x\left\{f(x)+\frac{1}{c_k}\sum_{i=1}^m\phi\left(\phi^{*\prime}(\pi_i^k;1)+c_kg_i(x)\right)\right\},$$
$$\pi_i^{k+1}=\nabla\phi\left(\phi^{*\prime}(\pi_i^k;1)+c_kg_i(x^{k+1})\right),\quad i=1\colon m,$$

where $\phi^{*\prime}(\pi_i^k;1)=(\nabla\phi)^{-1}(\pi_i^k)$ if $\pi_i^k>0$, $\phi^{*\prime}(\pi_i^k;1)=t^0_\phi$ if $\pi_i^k=0$, $i=1\colon m$.

In view of Cor. 9.5, we restrict attention to methods generated by $\phi\in\Phi_s$.

Example 9.7. Choosing $\phi\in\Phi_s$ and $\psi=\phi^*$ in Ex. 7.8 yields the method

$$x^{k+1}\mathrel{\tilde\in}\operatorname{Arg\,min}_x\left\{f(x)+\frac{1}{c_k}\sum_{i=1}^m\phi\left((\nabla\phi)^{-1}(\pi_i^k)+c_kg_i(x)\right)\right\},$$
$$\pi_i^{k+1}=\nabla\phi\left((\nabla\phi)^{-1}(\pi_i^k)+c_kg_i(x^{k+1})\right),\quad i=1\colon m,$$

with $\gamma_i^k=(\nabla\phi)^{-1}(\pi_i^k)$, $\pi_i^k=\nabla\phi(\gamma_i^k)$, $i=1\colon m$, for all k. (Indeed, $C_{\partial\psi}=\mathbb{R}_>$, $\partial\psi(0)=\emptyset$ and $\psi^*=\phi$ by Lem. 9.3, $\bar v=-\infty$ by Lem. 9.4, $\psi'(t;1)=\nabla\psi(t)$ if $t>0$, and $\nabla\phi^*=(\nabla\phi)^{-1}$ by Def. 9.1 and [Roc70, Thms 26.3 and 26.5].) Note that $\phi(t)=e^t$ for Ex. 7.10, $\phi(t)=-1-\ln(-t)$ $(t<0)$ for Ex. 7.11, $\phi(t)=-(-t)^\beta/\beta$ $(t<0,\ \beta<0)$ for Ex. 7.12.

The following results will ensure that $\left\langle\pi^k,u^k\right\rangle\to 0$, as required in Lem. 8.10.

Definition 9.8. We say $\phi\in\Phi$ is forcing on $[t'_\phi,t''_\phi]$ if $[\phi'(t'_k)-\phi'(t''_k)](t'_k-t''_k)\to 0$ implies $\phi'(t''_k)(t'_k-t''_k)\to 0$ for any sequences $\{t'_k\},\{t''_k\}\subset[t'_\phi,t''_\phi]\cap C_\phi$, where $\phi'=\nabla\phi$.

Lemma 9.9. *If $\phi\in\Phi_s$, $\inf\phi>-\infty$ and $t''_\phi\in C_\phi$ then ϕ is forcing on $[-\infty,t''_\phi]$.*

Proof. Replace ϕ by $\phi-\inf\phi$, so that $\inf\phi=0$. Since $\phi'=\nabla\phi$ is positive and increasing (cf. Rem. 9.2), so is ϕ. Let $[\phi'(t'_k)-\phi'(t_k)]\tau_k\to 0$, $\tau_k>0$, $t'_k=t_k+\tau_k\leq t''_\phi$. If $\phi'(t_k)\tau_k\not\to 0$, there are $\epsilon>0$ and $K\subset\{1,2,...\}$ s.t. $\phi'(t_k)\tau_k\geq\epsilon$ $\forall k\in K$, so $\frac{\phi'(t'_k)}{\phi'(t_k)}\xrightarrow{K}1$. Since $\phi'(t_k)<\phi'(t''_\phi)$ and $\phi(t'_k)\geq\phi(t_k)+\phi'(t_k)\tau_k\geq\epsilon$, $\tau_k\geq\epsilon/\phi'(t''_\phi)$ and $t'_k\geq\phi^{-1}(\epsilon)$ $\forall k\in K$. Pick t_∞ and $K'\subset K$ s.t. $t'_k\xrightarrow{K'}t_\infty$. Then $t_k+\epsilon/2\phi'(t''_\phi)\leq t_\infty$ and $\phi'(t_k)\leq\phi'(t_\infty-\epsilon/2\phi'(t''_\phi))<\phi'(t_\infty)=\lim_{k\in K'}\phi'(t'_k)$ for large $k\in K'$ contradict $\frac{\phi'(t'_k)}{\phi'(t_k)}\xrightarrow{K}1$. Therefore, $\phi'(t_k)\tau_k\to 0$, i.e., ϕ is forcing. □

Lemma 9.10. *The following functions are forcing on* $[-\infty, t''_\phi]$*:* $\phi_1(t) = e^t$ *with* $t''_\phi \in \mathbb{R}$, $\phi_2(t) = -1 - \ln(-t)$ $(t < 0)$ *with* $t''_\phi \le 0$, $\phi_3(t) = -(-t)^\beta/\beta$ $(t < 0,\ \beta < 0)$ *with* $t''_\phi < 0$.

Proof. Let $\phi = \phi_2$. Suppose $\frac{\phi'(t_k+\tau_k)-\phi'(t_k)}{\phi'(t_k)}\phi'(t_k)\tau_k \to 0$. Since $\phi'(t_k)\tau_k = -\tau_k/t_k$ and $\frac{\phi'(t_k+\tau_k)-\phi'(t_k)}{\phi'(t_k)} = \frac{-1}{1+t_k/\tau_k}$, $\phi'(t_k)\tau_k \to 0$, i.e., ϕ is forcing. Invoke Lem. 9.9 for ϕ_1 and ϕ_3. □

Example 9.11. Let $\phi \in \Phi_s$ be s.t. $\phi(t) = -\frac{(-t)^\beta - 1}{\beta}$ for $t < -\frac{1}{2}$, $\beta \in (0,1)$. Let $t_k = -k$, $\tau_k = 1/\phi'(t'_k)$. Then $[\phi'(t_k + \tau_k) - \phi'(t_k)]\tau_k = (1 - k^{-\beta})^{\beta-1} - 1 \to 0$, but $\phi'(t_k)\tau_k \to 1$, i.e., ϕ is not forcing on $[-\infty, -1]$, although $\lim_{\beta\downarrow 0} -\frac{(-t)^\beta - 1}{\beta} = -\ln(-t)$ is; cf. Lem. 9.10.

Lemma 9.12. *Consider Ex.* 9.7 *with* $\phi \in \Phi_s$, $t_\phi = \sup_{t\in C_\phi} t$ *and* $t_\gamma = \sup_{i,k}\gamma_i^k$. *Then* $t_\gamma \le t_\phi$ *(so that* $t_\gamma < \infty$ *if* $t_\phi < \infty$*). In general,* $t_\gamma < t_\phi$ *iff* $\{\pi^k\}$ *is bounded.*

Proof. This follows from the facts $\pi_i^k = \nabla\phi(\gamma_i^k) \ge 0$, $\gamma_i^k \in C_\phi = (-\infty, t_\phi)$, $\lim_{t\uparrow t_\phi} \nabla\phi(t) = \infty$ and strict monotonicity of $\nabla\phi$; cf. Rem. 9.2, Lem. 9.3 and Ex. 9.7. □

Lemma 9.13. *Suppose in Ex.* 9.7 $\phi \in \Phi_s$ *is forcing on* $(-\infty, t_\gamma]$ *with* $t_\gamma = \sup_{i,k}\gamma_i^k$, $c_k \ge c_{\min} > 0$ *for all* k, *and* $\left\langle \pi^{k+1} - \pi^k, u^{k+1}\right\rangle \to 0$. *Then* $\left\langle \pi^k, u^k\right\rangle \to 0$.

Proof. Since $\nabla\phi$ is nondecreasing and $h^+(u) = \sum_i \phi(u_i)$, we deduce from (8.4) that

$$\begin{aligned} 0 \leftarrow \left\langle \pi^{k+1} - \pi^k, u^{k+1}\right\rangle &= \sum_{i=1}^m [\phi'(\gamma_i^k + c_k u_i^{k+1}) - \phi'(\gamma_i^k)]u_i^{k+1} \qquad (9.1)\\ &\ge \sum_{i=1}^m [\phi'(\gamma_i^k + c_{\min} u_i^{k+1}) - \phi'(\gamma_i^k)]u_i^{k+1} \ge 0 \end{aligned}$$

and $[\phi'(\gamma_i^k + c_{\min}u_i^{k+1}) - \phi'(\gamma_i^k)]c_{\min}u_i^{k+1} \to 0$, $i = 1{:}m$. But $\gamma^{k+1} = \gamma^k + c_k u^{k+1}$ for all k (cf. Ex. 9.7) yields $\sup_{i,k}\{\gamma_i^k + c_{\min}u_i^{k+1}\} \le t_\gamma$, so the preceding relation and the forcing property of ϕ give $\pi_i^k u_i^{k+1} = \phi'(\gamma_i^k)u_i^k \to 0$ $\forall i$; hence $\left\langle \pi^{k+1}, u^{k+1}\right\rangle \to 0$ by (9.1). □

Theorem 9.14. *Consider Ex.* 9.7 *with* $\phi \in \Phi_s$ *s.t.* $\inf\phi > -\infty$. *Suppose* $\operatorname{Arg\,max} d \ne \emptyset$, $\sum_{j=1}^k s_j\epsilon_j/s_k \to 0$, $\sum_{k=1}^\infty c_k\epsilon_k < \infty$ *and* $\inf_k c_k > 0$. *Then* $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$, $d(\pi^k) \to d^\infty = d(\pi^\infty)$ *and* (8.10) *holds. If* $\{\breve{x}^k\}$ *has a limit point* x^∞ *(e.g.,* $C_* \ne \emptyset$ *is bounded; cf. Rem.* 8.11*), then* x^∞ *solves* (7.1) *and* $f(x^\infty) = d^\infty$.

Proof. Let $\psi = \phi^*$. We have $\psi(0) = -\inf\phi < \infty$, $C_\psi = \mathbb{R}_+$ (cf. Lem. 9.3), $C_{\psi_+} = \mathbb{R}_+$ and $C_{h_+} = \mathbb{R}_+^m$, so the assertions about $\{\pi^k\}$ follow from Thm 8.3. Then $t_\gamma = \sup_{i,k}\gamma_i^k < t_\phi$ by Lem. 9.12 ($\{\pi^k\}$ is bounded), so ϕ is forcing on $[-\infty, t_\gamma]$ (Lem. 9.9). Since $d(\pi^k) \to d^\infty < \infty$ and $0 \le \epsilon_k \le \sum_{j=1}^k s_j\epsilon_j/s_k \to 0$, (8.4)–(8.5) yield $\left\langle \pi^{k+1} - \pi^k, u^{k+1}\right\rangle \to 0$. Then $\left\langle \pi^k, u^k\right\rangle \to 0$ by Lem. 9.13. The conclusion follows from Lem. 8.10. □

Remark 9.15. For the exponential multiplier method (Ex. 7.10 with $\phi(t) = e^t$), Thms 8.3 and 9.14 subsume [TsB93, Prop. 3.1] (in which $\operatorname{Arg\,max} d \ne \emptyset$, $C_* \ne \emptyset$ is bounded, $\epsilon_k \equiv 0$) and [IST94, Thm 7.3] (in which $x^k \to x^\infty$ implies $\breve{x}^k \to x^\infty$).

Theorem 9.16. *Consider Ex.* 9.7 *with* $\phi \in \Phi_s$ *forcing on* $(-\infty, t_\phi) \neq \mathrm{IR}$ *(e.g.,* $\phi(t) = -1 - \ln(-t)$*; cf. Lem.* 9.10*). Suppose* $\epsilon_k \to 0$, $\inf_k c_k > 0$ *and* $d(\pi^k) \to d^\infty < \infty$*. Then* (8.10) *holds. If* $\{\breve{x}^k\}$ *has a limit point* x^∞ *(e.g.,* $C_* \neq \emptyset$ *is bounded; cf. Rem.* 8.11*), then* x^∞ *solves* (7.1)*,* $f(x^\infty) = d^\infty = \max d$ *and each limit point of* $\{\pi^k\}$ *maximizes* d*.*

Proof. By Lem. 9.12, $t_\gamma = \sup_{i,k} \gamma_i^k \le t_\phi$, so ϕ is forcing on $(-\infty, t_\gamma]$. Since $d(\pi^k) \to d^\infty < \infty$ and $\epsilon_k \to 0$, (8.4)–(8.5) yield $\left\langle \pi^{k+1} - \pi^k, u^{k+1} \right\rangle \to 0$. Then $\left\langle \pi^k, u^k \right\rangle \to 0$ by Lem. 9.13. Since $t_\gamma \le t_\phi < \infty$, the conclusion follows from Lem. 8.10. □

Remark 9.17. Suppose $\sum_{k=1}^\infty \epsilon_k < \infty$. Then $d^{k+1} \ge d^k - \epsilon_k \ \forall k$ (cf. (8.4)) yields $d(\pi^k) \to d^\infty \in (\infty, \infty]$ (cf. [Pol83, Lem. 2.2.3]). If $d^\infty = \infty$, then $C_0 = \emptyset$ by weak duality. If $d^\infty < \infty$, then $\{\pi^k\}$ is bounded iff so is $\operatorname{Arg\,max} d \neq \emptyset$ (cf. [Roc70, Cor. 8.7.1]), whereas if $C_0 \neq \emptyset$, then $\operatorname{Arg\,max} d \neq \emptyset$ is bounded iff Slater's condition holds, i.e., $g(x) < 0$ for some $x \in C_f$ [GoT89, Thm 1.3.4]. This observation may be used in Lem. 8.10 and Thm 9.16.

Theorem 9.18. *Consider Ex.* 9.7 *with* $\phi \in \Phi_s$ *s.t.* $\inf \phi > -\infty$*. Suppose* $g(x) < 0$ *for some* $x \in C_f$*,* $\sum_{k=1}^\infty \epsilon_k < \infty$ *and* $\inf_k c_k > 0$*. Then* $d(\pi^k) \to d^\infty < \infty$ *and* (8.10) *holds. If* $\{\breve{x}^k\}$ *has a limit point* x^∞ *(e.g.,* $C_* \neq \emptyset$ *is bounded; cf. Rem.* 8.11*), then* x^∞ *solves* (7.1)*,* $f(x^\infty) = d^\infty = \max d$ *and each limit point of* $\{\pi^k\}$ *maximizes* d*. If* $d^\infty = \sup d$ *and* $\sum_{k=1}^\infty c_k \epsilon_k < \infty$*, then* $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$*.*

Proof. Since $\epsilon_k \to 0$, $d(\pi^k) \to d^\infty < \infty$, $\{\pi^k\}$ and $\operatorname{Arg\,max} d \neq \emptyset$ are bounded (Rem. 9.17), we get, as in the proof of Thm 9.14, $C_{h_+} = \mathrm{IR}^m_+$, $t_\gamma < t_\phi$ and $\left\langle \pi^k, u^k \right\rangle \to 0$. Hence the first two assertions follow from Lem. 8.10, and the third one from Thm 8.3. □

Theorem 9.19. *Consider Ex.* 9.7 *with* $\phi \in \Phi_s$ *forcing on* $(-\infty, t''_\phi]$ $\forall t''_\phi \in \mathrm{IR}$*. Suppose* $g(x) < 0$ *for some* $x \in C_f$*,* $\sum_{k=1}^\infty \epsilon_k < \infty$ *and* $\inf_k c_k > 0$*. Then* $d(\pi^k) \to d^\infty < \infty$ *and* (8.10) *holds. If* $\{\breve{x}^k\}$ *has a limit point* x^∞ *(e.g.,* $C_* \neq \emptyset$ *is bounded; cf. Rem.* 8.11*), then* x^∞ *solves* (7.1)*,* $f(x^\infty) = d^\infty = \max d$ *and each limit point of* $\{\pi^k\}$ *maximizes* d*. If* $d^\infty = \sup d$*,* $\operatorname{Arg\,max} d \cap C_{h_+} \neq \emptyset$ *and* $\sum_{k=1}^\infty c_k \epsilon_k < \infty$*, then* $\pi^k \to \pi^\infty \in \operatorname{Arg\,max} d$*.*

Proof. Use the proof of Thm 9.18, without asserting that $C_{h_+} = \mathrm{IR}^m_+$. □

Remark 9.20. It is easy to see that we may replace $\phi \in \Phi_s$ by $\phi \in \Phi_0$ and Ex. 9.7 by Ex. 7.8 with $\psi = \phi^*$ in Lems. 9.9, 9.12, 9.13 and Thms 9.14, 9.16, 9.18, 9.19. (In the proof of Lem. 9.9, $t_\infty \ge \psi^{-1}(\epsilon) > t^0_\phi$, since ϕ' and ϕ are positive and increasing on (t^0_ϕ, t_ϕ); in proving Lem. 9.12, recall the proof of Cor. 9.5; in the proof of Lem. 9.13, use $\gamma^{k+1} \ge \gamma^k + c_k u^{k+1}$; cf. (7.7).) Such results complement Thms 8.4 and 8.7; cf. Cor. 9.5.

10 Additional aspects of multiplier methods

Modified barrier functions can be extrapolated quadratically to facilitate their minimization; cf. [BTYZ92, BrS93, BrS94, NPS94, PoT94]. We now extend such techniques to our penalty functions, starting with a technical result.

Lemma 10.1. *Let $\phi_1, \phi_2 \in \Phi$ be s.t. for some $t_s \in (t^0_{\phi_1}, t_{\phi_1})$, $\phi_1(t_s) = \phi_2(t_s)$, $\phi_1'(t_s) = \phi_2'(t_s)$, ϕ_1 is forcing on $(-\infty, t_s]$ and ϕ_2 is forcing on $[t_s, t''_{\phi_2}]$ with $t''_{\phi_2} \in [t_s, t_{\phi_2}]$. Let $\phi(t) = \phi_1(t)$ if $t \le t_s$, $\phi(t) = \phi_2(t)$ if $t > t_s$. Then ϕ is forcing on $(-\infty, t''_{\phi_2}]$. If $\phi_2 \in \Phi_s \cup \Phi_0$, then $\phi \in \Phi_s$ iff $\phi_1 \in \Phi_s$, $\phi \in \Phi_0$ iff $\phi_1 \in \Phi_0$.*

Proof. Suppose $[\phi'(t''_k) - \phi'(t'_k)](t''_k - t'_k) \to 0$ with $t'_k \le t_s \le t''_k \le t''_{\phi_2}$ (other cases being trivial). Since ϕ_1' and ϕ_2' are nondecreasing, so is ϕ'; therefore, all terms in

$$\begin{aligned}[\phi'(t''_k) - \phi'(t'_k)](t''_k - t'_k) &\ge [\phi'(t''_k) - \phi'(t_s)](t''_k - t_s) + [\phi'(t_s) - \phi'(t'_k)](t_s - t'_k) \\ &= [\phi_2'(t''_k) - \phi_2'(t_s)](t''_k - t_s) + [\phi_1'(t_s) - \phi_1'(t'_k)](t_s - t'_k)\end{aligned}$$

are nonnegative and tend to zero. Thus $\phi_2'(t_s)(t''_k - t_s) \to 0$ and $\phi_1'(t_s)(t_s - t'_k) \to 0$ (Def. 9.8). Hence $t'_k, t''_k \to t_s$ ($\phi_2'(t_s) = \phi_1'(t_s) > 0$), $\phi'(t''_k)(t''_k - t'_k) \to \phi'(t_s)0$ and $\phi'(t'_k)(t''_k - t'_k) \to 0$ yield the first assertion. For the second one, use Def. 9.1 and Rem. 9.2. □

Examples 10.2. Using the notation of Lem. 10.1, we add the condition $\phi_1''(t_s) = \phi_2''(t_s)$ to make ϕ twice continuously differentiable. In each example, $\phi \in \Phi_s \cup \Phi_0$ is forcing on $(-\infty, t''_\phi]$ $\forall t''_\phi \in \mathbb{R}$; cf. Rem. 9.2, Lems. 9.9–9.10 and Rem. 9.20.

1 (cubic-quadratic). $\phi(t) = \frac{\max\{0, t+t_s\}^3}{12 t_s} - \frac{t_s^2}{6}$ if $t \le t_s$, $\phi(t) = \frac{\max\{0,t\}^2}{2} = \phi_2(t)$ if $t > t_s$, $t_s > 0$. This ϕ only grows as fast as ϕ_2 in Ex. 7.9 with $\beta = 2$, but is smoother.

2 (exponential-quadratic). $\phi(t) = e^t$ if $t \le t_s > 0$, $\phi(t) = e^{t_s}(\frac{t^2}{2} + (1 - t_s)t + 1 - t_s - \frac{t_s^2}{2})$ if $t > t_s$, $\phi_2(\cdot) = a\max\{0, \cdot - t^0_{\phi_2}\}^2 + b$. This ϕ does not grow as fast as e^t in Ex. 7.10.

3 (log-quadratic). $\phi(t) = -\ln(-t) - 1 = \phi_1(t)$ if $t \le t_s < 0$, $\phi(t) = \frac{t^2}{2t_s^2} - \frac{2t}{t_s} + \frac{1}{2} - \ln(-t_s)$ if $t > t_s$. This ϕ allows arbitrarily large infeasibilities, in contrast to ϕ_1 in Ex. 7.11.

4 (hyperbolic-quadratic). $\phi(t) = -\frac{1}{t} = \phi_1(t)$ if $t \le t_s < 0$, $\phi(t) = \frac{t^2}{|t_s|^3} + \frac{3t}{t_s^2} - \frac{3}{t_s}$ if $t > t_s$. Again, this ϕ has $C_\phi = \mathbb{R}$, in contrast to ϕ_1 in Ex. 7.12.

5 (hyperbolic-log-quadratic). $\phi(t) = \frac{-4t'_s}{-t'_s - t} - 2 - \ln(-t'_s)$ if $t \le t'_s < 0$, $\phi(t) = -\ln(-t)$ if $t'_s \le t \le t_s < 0$, $\phi(t) = \frac{t^2}{2t_s^2} - \frac{2t}{t_s} + \frac{3}{2} - \ln(-t_s)$ if $t > t_s$.

Remark 10.3. Other smooth penalty functions (e.g., cubic-log-quadratic) are easy to derive. Such functions are covered by the various results of §9. Their properties, e.g., $\inf \phi > -\infty$, may also have practical significance; this should be verified experimentally.

The following result (inspired by [Ber82, Prop. 5.7]) shows that minimizing L_k (cf. (7.12)) in Alg. 7.4 is well posed under mild conditions (see the Appendix for its proof).

Lemma 10.4. *Let $h(\pi) = \sum_{i=1}^m \psi(\pi_i)$, where ψ is a B-function with $C_\psi \supset \mathbb{R}_>$. Suppose $L_k \not\equiv \infty$ (e.g., $\inf_{C_f} \max_{i=1}^m g_i \le 0$). Then* $\operatorname{Arg\,min} L_k$ *is nonempty and compact iff f and $g_1, \ldots, g_m$ have no common direction of recession, and if $C_0 \ne \emptyset$ then this is equivalent to* (7.1) *having a nonempty and compact set of solutions.*

We now consider a variant of condition (7.18), inspired by one in [Ber82, p. 328].

Lemma 10.5. *Under the strong convexity assumption* (7.15), *consider* (7.17) *with*

$$|\Delta_x L_k(x^{k+1})|^2 \le \eta_k [L(x^{k+1}, \pi^{k+1}) - L_k(x^{k+1})] \tag{10.1}$$

and $\epsilon_k = |\Delta_x L_k(x^{k+1})|^2/2\breve{\alpha}$ *replacing* (7.18), *where* $\eta_k \geq 0$. *Then*

$$L(x^{k+1}, \pi^{k+1}) - d(\pi^{k+1}) \leq \epsilon_k \leq \frac{\eta_k}{2\breve{\alpha}}[L(x^{k+1}, \pi^{k+1}) - L_k(x^{k+1})], \tag{10.2}$$

$$d(\pi^k) \leq L(x^{k+1}, \pi^k) \leq L_k(x^{k+1}) \leq d(\pi^{k+1}) \quad \text{if} \quad \eta_k \leq 2\breve{\alpha}, \tag{10.3}$$

$$\epsilon_k \leq \frac{\eta_k}{\breve{\alpha}}[d(\pi^{k+1}) - d(\pi^k)] \leq d(\pi^{k+1}) - d(\pi^k) \quad \text{if} \quad \eta_k \leq \breve{\alpha}. \tag{10.4}$$

Next, suppose $\eta_k \to 0$ *in* (10.1). *Then* $d(\pi^k) \to d^\infty \in (-\infty, \infty]$. *If* $d^\infty < \infty$ *then* $\sum_{k=1}^\infty \epsilon_k < \infty$, $\epsilon_k \to 0$, $\sum_{j=1}^k c_j\epsilon_j/s_k \to 0$*; further,* $\sum_{k=1}^\infty c_k\epsilon_k < \infty$ *if* $\{c_k\eta_k\}$ *is bounded.*

Proof. By (7.17) and (10.1), (10.2) holds with $L(x^{k+1}, \pi^{k+1}) \geq L_k(x^{k+1})$ by (8.2). Thus $\eta_k \leq 2\breve{\alpha}$ yields $L_k(x^{k+1}) \leq d(\pi^{k+1})$ and (10.3) follows from (8.5). Similarly, $L(x^{k+1}, \pi^{k+1}) - d(\pi^{k+1}) \leq \frac{1}{2}[L(x^{k+1}, \pi^{k+1}) - L_k(x^{k+1})]$ for $\eta_k \leq \breve{\alpha}$ yields $L(x^{k+1}, \pi^{k+1}) - L_k(x^{k+1}) \leq 2[d(\pi^{k+1}) - L_k(x^{k+1})]$, so (10.4) follows from (10.2) and $d(\pi^k) \leq L_k(x^{k+1})$ (cf. (10.3)). Next, let $\eta_k \to 0$. Pick $\bar{k}$ s.t. $\eta_k \leq \breve{\alpha}$ $\forall k \geq \bar{k}$. (10.3)-(10.4) yield $d(\pi^k) \to d^\infty$, $\sum_{k=\bar{k}}^\infty \epsilon_k \leq [d^\infty - d(\pi^{\bar{k}})]$, $\sum_{k=\bar{k}}^\infty c_k\epsilon_k \leq \sup_k \frac{c_k\eta_k}{\breve{\alpha}}[d^\infty - d(\pi^{\bar{k}})]$. If $d^\infty < \infty$ then $\epsilon_k \to 0$ gives $\sum_{j=1}^k c_j\epsilon_j/s_k \to 0$ (Lem. 4.8(i)). □

Remark 10.6. In view of Lem. 10.5, suppose in the strongly convex case of (7.15), (10.1) is used with $\eta_k \to 0$. Since $q(\pi^{k+1}) \leq q(\pi^k)$ for all large k (cf. (10.3)), the results of §§8–9 may invoke, instead of Thm 5.2 with $\sum_{j=1}^k s_j\epsilon_j/s_k \to 0$, Thm 4.6 with $\sum_{j=1}^k c_j\epsilon_j/s_k \to 0$. The latter condition holds automatically if $\lim_{k\to\infty} d(\pi^k) < \infty$, e.g., $\sup d < \infty$. Thus we may drop the conditions: $\sum_{j=1}^k s_j\epsilon_j/s_k \to 0$ from Thms 8.3, 8.4, 9.14, $\epsilon_k \to 0$ from Lem. 8.10 and Thm 9.16, and $\sum_{k=1}^\infty \epsilon_k < \infty$ from Thms 9.18–9.19. Instead of $\sum_{k=1}^\infty c_k\epsilon_k < \infty$, we may assume that $\{c_k\eta_k\}$ is bounded in Thms 8.3, 8.4, 9.14 and 9.18–9.19.

Condition (10.1) can be implemented as in [Ber82, Prop. 5.7(b)].

Lemma 10.7. *Suppose* f *is strongly convex,* $\inf_{C_f} \max_{i=1}^m g_i \leq 0$, *and* g *is continuous on* C_f. *Consider iteration* k *of Ex. 7.5 with* $h(\pi) = \sum_{i=1}^m \psi(\pi_i)$, *where* ψ *is a B-function s.t.* $C_{\nabla\psi} \supset \mathbb{R}_>$. *If* $\eta_k > 0$, π^k *is not a Lagrange multiplier of* (7.1), $\{z^j\}$ *is a sequence converging to* $\hat{x} = \arg\min L_k$, *and* $\Delta_x L_k(z^j) \to 0$, *then there exists* $x^{k+1} \in \{z^1, z^2, \ldots\}$ *satisfying the stopping criterion* (10.1).

Proof. By Lemmas 9.3–9.4, Ex. 7.5 has $\bar{u}_i = t_\phi^0$, $\pi_i^k = \nabla\phi(\gamma_i^k)$, $\gamma_i^k \geq t_\phi^0$, $i = 1{:}m$, $h^+(u) = \sum_{i=1}^m \phi(u_i)$, where $\phi = \psi^+ \in \Phi_s \cup \Phi_0$. Let $\hat{u} = g(\hat{x})$ and $\hat{\pi} = \nabla h^+(\gamma^k + c_k\hat{u})$. Then, as in (8.2),

$$L(\hat{x}, \hat{\pi}) - L_k(\hat{x}) = D_{h^+}(\gamma^k, \gamma^k + c_k\hat{u})/c_k \geq 0. \tag{10.5}$$

Suppose $L(\hat{x}, \hat{\pi}) = L_k(\hat{x})$. By (10.5), (2.3) and convexity of h^+, $\phi(\gamma_i^k) - \phi(\gamma_i^k + c_k\hat{u}_i) - \nabla\phi(\gamma_i^k + c_k\hat{u}_i)(-c_k\hat{u}_i) = 0$, $i = 1{:}m$. Therefore, since ϕ is strictly convex on $[t_\phi^0, t_\phi)$ with $\nabla\phi(t) = 0$ iff $t \leq t_\phi^0$ (Def. 9.1), and $\gamma_i^k \geq t_\phi^0$, for each i, either $\gamma_i^k + c_k\hat{u}_i = \gamma_i^k > t_\phi^0$ yields $\hat{u}_i = 0$ and $\hat{\pi}_i = \pi_i^k = \nabla\phi(\gamma_i^k)$, or $\gamma_i^k + c_k\hat{u}_i \leq t_\phi^0 = \gamma_i^k$ yields $\hat{u}_i \leq 0$ and $\hat{\pi}_i = \pi_i^k = \nabla\phi(\gamma_i^k) = 0$. Hence $\hat{\pi} = \pi^k$, $\hat{u} \leq 0$ and $\langle\hat{\pi}, \hat{u}\rangle = 0$. Combining this with $0 \in \partial L_k(\hat{x}) = \partial_x L(\hat{x}, \hat{\pi})$ (Lem. 7.3), we see (cf. [Roc70, Thm 28.3]) that π^k is a Lagrange multiplier, a contradiction. Therefore, we must have strict inequality in (10.5). Since $g(z^j) \to \hat{u}$ and $D_{h^+}(\gamma^k, \gamma^k + c_k g(z^j)) \to D_{h^+}(\gamma^k, \gamma^k + c_k\hat{u}) > 0$ by continuity, whereas $\eta_k > 0$ and $\Delta_x L_k(z^j) \to 0$, the stopping criterion will be satisfied for sufficiently large j. □

A Appendix

Proof of Lemma 7.2. $\mathbb{R}^m_+ \cap \operatorname{ri} C_h \neq \emptyset$ implies $\partial h_+ = \partial h + \partial \delta_{\mathbb{R}^m_+}$ (cf. [Roc70, Thm 23.8]), so $C_{\partial h_+} = C_{\partial h} \cap \mathbb{R}^m_+$ and h_+ is essentially strictly convex (cf. [Roc70, p. 253]). Hence (cf. [Roc70, Thm 26.3]) $h^+ = h^*_+$ is closed proper essentially smooth, so $\partial h^+(u) = \{\nabla h^+(u)\}$ $\forall u \in \mathring{C}_{h^+} = C_{\partial h^+}$ by [Roc70, Thm 26.1] and ∇h^+ is continuous on $\mathring{C}_{h^+}$ by [Roc70, Thm 25.5]. By [Roc70, Thm 23.5], $\partial h^*_+ = (\partial h_+)^{-1}$, so $\operatorname{im} \partial h_+ = C_{\partial h^+}$. Since h^+ is nondecreasing, $C_{h^+} = C_{h^+} - \mathbb{R}^m_+$, so $\mathring{C}_{h^+} = \mathring{C}_{h^+} - \mathbb{R}^m_+$ as the union of open sets. That $N_{\mathbb{R}^m_+}(\pi) = \{\gamma \leq 0 : \langle \gamma, \pi \rangle = 0\}$ for $\pi \geq 0$ is elementary (cf. [Roc70, p. 226]). If $\pi = \nabla h^+(\gamma)$ and $\tilde{\gamma} \in N_{\mathbb{R}^m_+}(\pi)$ then $\gamma \in \partial h_+(\pi)$ and $\gamma + \tilde{\gamma} \in \partial h_+(\pi)$, so $\pi = \nabla h^+(\gamma + \tilde{\gamma})$. If $\operatorname{im} \partial h \supset \mathbb{R}^m_>$ and $u \in \mathbb{R}^m$ then $-h^+(u) = \inf \phi$, where $\phi = h_+ - \langle u, \cdot \rangle$ is inf-compact. Indeed, pick $\tilde{\pi}$ and $\tilde{u} \in \partial h(\tilde{\pi})$ s.t. $\tilde{u} > u$. Then $\tilde{\phi}(\pi) = h(\tilde{\pi}) + \langle \tilde{u}, \pi - \tilde{\pi} \rangle - \langle u, \pi \rangle \leq \phi(\pi)$ for all $\pi \geq 0$ and if $\{\pi^k\} \subset \mathbb{R}^m_+$, $|\pi^k| \to \infty$ then $\tilde{\phi}(\pi^k) \to \infty$ since $\tilde{u} - u > 0$. Hence ϕ is inf-compact and $u \in C_{h^+}$, so $C_{h^+} = \mathbb{R}^m$. □

We need the following slightly sharpened version of [GoT89, Thm 1.5.4].

Lemma A.1 (subdifferential chain rule). *Let $f_1, \ldots, f_m$ be proper convex functions on $\mathbb{R}^n$ with $\bigcap_{i=1}^m \operatorname{ri} C_{f_i} \neq \emptyset$. Let $f(\cdot) = (f_1(\cdot), \ldots, f_m(\cdot))$ and $C_f = \bigcap_{i=1}^m C_{f_i}$. Let ϕ be a proper convex nondecreasing function on $\mathbb{R}^m$ s.t.*
/ $f(\tilde{x}) < \tilde{y}$ for some $\tilde{x} \in C_f$ and $\tilde{y} \in C_\phi$. Let $\psi(x) = \phi(f(x))$ if $x \in C_f$, $\psi(x) = \infty$ if $x \notin C_f$. Then ψ is proper convex, $\operatorname{im} \partial \phi \subset \mathbb{R}^m_+$, and for each $\bar{x} \in C_f$ and $\bar{y} = f(\bar{x})$

$$\partial \psi(\bar{x}) = \bigcup \{ \sum_{i=1}^m \gamma_i \partial f_i(\bar{x}) : \gamma \in \partial \phi(\bar{y}) \}. \tag{A.1}$$

Proof. For any $x^1, x^2 \in C_f$ and $\lambda \in [0,1]$, $f(\lambda x^1 + (1-\lambda)x^2) \leq \lambda f(x^1) + (1-\lambda) f(x^2)$ and hence $\psi(\lambda x^1 + (1-\lambda)x^2) \leq \phi(\lambda f(x^1) + (1-\lambda) f(x^2)) \leq \lambda \psi(x^1) + (1-\lambda)\psi(x^2)$, so ψ is convex. Since $\psi(x) > -\infty$ for all x, ψ is proper. Let $Q = \bigcup_{\gamma \in \partial \phi(\bar{y})} \sum_{i=1}^m \gamma_i \partial f_i(\bar{x})$. Let $\gamma \in \partial \phi(\bar{y})$, $\gamma^i \in \partial f_i(\bar{x})$, $i = 1{:}m$, $\Gamma = [\gamma^1, \ldots, \gamma^m]^T$. For any x, $f(x) \geq f(\bar{x}) + \Gamma(x - \bar{x})$ yields $\psi(x) \geq \phi(f(\bar{x}) + \Gamma(x - \bar{x})) \geq \psi(\bar{x}) + \gamma^T \Gamma (x - \bar{x})$, i.e., $\Gamma^T \gamma \in \partial \psi(\bar{x})$, so $Q \subset \partial \psi(\bar{x})$. To prove the opposite inclusion, let $\bar{\gamma} \in \partial \psi(\bar{x})$. Consider the convex program

$$\text{minimize} \quad \phi(y) - \langle \bar{\gamma}, x \rangle, \text{ s.t. } f(x) - y \leq 0, \ x \in C_f, \ y \in C_\phi. \tag{A.2}$$

By the monotonicity of ϕ and the definition of subdifferential, $(\bar{x}, \bar{y})$ solves (A.2), which satisfies Slater's condition (cf. $f(\tilde{x}) < \tilde{y}$), so (cf. [Roc70, Cor. 28.2.1]) it has a Kuhn-Tucker point $\bar{\pi} \in \mathbb{R}^m_+$ s.t. (cf. [Roc70, Thm 28.3])

$$\phi(y) - \langle \bar{\gamma}, x \rangle + \langle \bar{\pi}, f(x) - y \rangle \geq \phi(\bar{y}) - \langle \bar{\gamma}, \bar{x} \rangle + \langle \bar{\pi}, f(\bar{x}) - \bar{y} \rangle \quad \forall x \in C_f, y \in C_\phi.$$

Then $\phi(y) \geq \phi(\bar{y}) + \langle \bar{\pi}, y - \bar{y} \rangle$ $\forall y$ yields $\bar{\pi} \in \partial \phi(\bar{y})$, whereas $\langle \bar{\pi}, f(x) \rangle \geq \langle \bar{\pi}, f(\bar{x}) \rangle + \langle \bar{\gamma}, x - \bar{x} \rangle$ $\forall x$ yields $\bar{\gamma} \in \partial (\sum_{i=1}^m \bar{\pi}_i f_i)(\bar{x}) = \sum_{i=1}^m \bar{\pi}_i \partial f_i(\bar{x})$ from $\bigcap_{i=1}^m \operatorname{ri} C_{f_i} \neq \emptyset$ (cf. [Roc70, Thm 23.8]). Thus $\partial \psi(\bar{x}) \subset Q$, i.e., $\partial \psi(\bar{x}) = Q$. To see that $\operatorname{im} \partial \phi \subset \mathbb{R}^m_+$, note that if $\gamma \in \partial \phi(y^1)$ then $\phi(y^1) \geq \phi(y^2) \geq \phi(y^1) + \langle \gamma, y^2 - y^1 \rangle$ for all $y^2 \leq y^1$ implies $\gamma \geq 0$. □

Proof of Lemma 10.4. Let $\phi_i(x) = \psi^+(\gamma_i^k + c_k g_i(x))$ if $x \in C_{g_i}$, $\phi_i(x) = \infty$ if $x \notin C_{g_i}$, $i = 1{:}m$. Each ϕ_i is closed: for any $\alpha \in \mathbb{R}$, $\{t : \psi^+(t) \le \alpha\} = (-\infty, \beta]$ for some $\beta < \infty$ (ψ^+ is closed nondecreasing and $\lim_{t \uparrow t_{\psi^+}} \psi^+(t) = \infty$ by Lemmas 9.3–9.4) and $\{x : \phi_i(x) \le \alpha\} = \{x : g_i(x) \le (\beta - \gamma_i^k)/c_k\}$ is closed (so is g_i). We have $L_k = f + \frac{1}{c_k}\sum_{i=1}^m [\phi_i - \psi^+(\gamma_i^k)]$ with f and ϕ_i closed proper and $L_k \not\equiv \infty$, so L_k is closed and $L_k 0^+ = f0^+ + \frac{1}{c_k}\sum_{i=1}^m \phi_i 0^+$ [Roc70, Thm 9.3]. Suppose $g_i 0^+(y) \le 0$. Since $L_k \not\equiv \infty$, $C_{\psi^+} = (-\infty, t_{\psi^+})$ (cf. Lem. 9.4 and Def. 9.1) and g_i is closed, there is $x \in \operatorname{ri} C_{g_i}$ s.t. $\gamma_i^k + c_k g_i(x) \in C_{\psi^+}$. Let $\gamma \in \partial g_i(x)$. Then $g_i(x) + t\langle \gamma, y\rangle \le g_i(x + ty) \le g_i(x)$ $\forall t \ge 0$, so $\langle \gamma, y \rangle \le 0$ and, since ψ^+ is nondecreasing, $\psi^+(\gamma_i^k + c_k[g_i(x) + t\langle\gamma, y\rangle]) \le \psi^+(\gamma_k^k + c_k g_i(x + ty)) \le \psi^+(\gamma_i^k + c_k g_i(x))$ $\forall t \ge 0$. Hence $\psi^+ 0^+(c_k\langle\gamma, y\rangle) \le \phi_i 0^+(y) \le 0$, so $\langle \gamma, y\rangle \le 0$ and $\psi^+ 0^+ = \sigma_{\mathbb{R}_+}$ (cf. Lemmas 9.3–9.4) yield $\phi_i 0^+(y) = 0$. Now suppose $g_i 0^+(y) > 0$. Pick $\bar{t} > 0$ and $\bar{\alpha} > 0$ s.t. $[g_i(x + ty) - g_i(x)]/t \ge \bar{\alpha}$ $\forall t \ge \bar{t}$. Then

$$\begin{aligned}
\phi_i 0^+(y) &= \lim_{t\uparrow\infty}[\psi^+(\gamma_i^k + c_k g_i(x + ty)) - \psi^+(\gamma_i^k + c_k g_i(x))]/t \\
&\ge \lim_{t\uparrow\infty}[\psi^+(\gamma_i^k + c_k(g_i(x) + t\bar{\alpha})) - \psi^+(\gamma_i^k + c_k g_i(x))]/t \\
&= \psi^+ 0^+(c_k\bar{\alpha}) = \infty
\end{aligned}$$

from $\psi^+ 0^+ = \sigma_{\mathbb{R}_+}$. Thus $\phi_i 0^+(y) = 0$ if $g_i 0^+(y) \le 0$, $\phi_i 0^+(y) = \infty$ if $g_i 0^+(y) > 0$. Therefore, $L_k 0^+(y) = f0^+(y)$ if $g_i 0^+(y) \le 0$ for $i = 1{:}m$, $L_k 0^+(y) = \infty$ otherwise. The proof may be finished as in [Ber82, §5.3]. □

References

[Aus86] A. Auslender, *Numerical methods for nondifferentiable convex optimization*, Math. Programming Stud. **30** (1986) 102–126.

[Ber82] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[BeT94] D. P. Bertsekas and P. Tseng, *Partial proximal minimization algorithms for convex programming*, SIAM J. Optim. **4** (1994) 551–572.

[Bre67] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, Zh. Vychisl. Mat. i Mat. Fiz. **7** (1967) 620–631 (Russian). English transl. in U.S.S.R. Comput. Math. and Math. Phys. 7 (1967) 200–217.

[BrS93] M. G. Breitfeld and D. F. Shanno, *Computational experience with modified log-barrier methods for nonlinear programming*, Research Report RRR 17-93, RUTCOR, Rutgers Univ., New Brunswick, NJ 08903, 1993. Revised March 1994.

[BrS94] ———, *A globally convergent penalty-barrier algorithm for nonlinear programming and its computational performance*, Research Report RRR 12-94, RUTCOR, Rutgers Univ., New Brunswick, NJ 08903, 1994.

[BTYZ92] A. Ben-Tal, I. Yuzefovich and M. Zibulevsky, *Penalty/barrier multiplier methods for minimax and constrained smooth convex programs*, Research Report 9/92, Optimization Laboratory, Technion, Haifa, Israel, 1992.

[BuF93] J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim. **31** (1993) 1340–1359.

[CDPI91] Y. Censor, A. R. De Pierro and A. N. Iusem, *Optimization of Burg's entropy over linear constraints*, Appl. Numer. Math. **7** (1991) 151–165.

[CeL81] Y. Censor and A. Lent, *An iterative row action method for interval convex programming*, J. Optim. Theory Appl. **34** (1981) 321–353.

[CeZ92] Y. Censor and S. A. Zenios, *Proximal minimization algorithm with D-functions*, J. Optim. Theory Appl. **73** (1992) 451–464.

[CGT92] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds*, Report 92/07, Département de Mathématique, Facultés Universitaires de Namur, Namur, 1992.

[CGT94] ———, *Large-scale nonlinear constrained optimization: A current survey*, Report 94/0?, Département de Mathématique, Facultés Universitaires de Namur, Namur, 1994.

[Cha94] I. Chabini, *Nouvelles méthodes séquentielles et parallèles pour l'optimisation de réseaux à coûts linéaires et convexes*, Ph.D. thesis, Département d'informatique et de recherche opérationelle, Université de Montréal, Montréal, Québec, Canada, 1994.

[ChT93] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim. **3** (1993) 538-543.

[CoL93] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Math. Programming **62** (1993) 261–275.

[EcB92] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming **55** (1992) 293–318.

[Eck93] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res. **18** (1993) 202–226.

[Egg90] P. P. B. Eggermont, *Multiplicative iterative algorithms for convex programming*, Linear Algebra Appl. **130** (1990) 25–42.

[Fer91] M. C. Ferris, *Finite termination of the proximal point algorithm*, Math. Programming **50** (1991) 359–366.

[FiM68] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[Flå94] S. D. Flåm, *Equilibrium programming using proximal-like algorithms*, Working Paper ?, Dept. of Economics, Univ. of Bergen, Bergen, Norway, 1994.

[GMSW88] P. E. Gill, W. Murray, M. A. Saunders and M. H. Wright, *Shifted barrier methods for linear programming*, Report SOL 88-9, Department of Operations Research, Stanford Univ., Stanford, CA, 1988.

[GoT89] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrange Functions; Theory and Optimization Methods*, Nauka, Moscow, 1989 (Russian).

[Gül91] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim. **29** (1991) 403–419.

[IST94] A. N. Iusem, B. Svaiter and M. Teboulle, *Entropy-like proximal methods in convex programming*, Math. Oper. Res. ? (1994) ?–? To appear.

[Ius95] A. N. Iusem, *On some properties of generalized proximal point methods for quadratic and linear programming*, J. Optim. Theory Appl. ? (1995) ?–? To appear.

[IuT93] A. N. Iusem and M. Teboulle, *On the convergence rate of entropic proximal optimization algorithms*, Comp. Appl. Math. **12** (1993) 153–168.

[JeP94] D. L. Jensen and R. A. Polyak, *The convergence of a modified barrier method for convex programming*, IBM J. Res. Develop. **38** (1994) 307–321.

[Kiw94] K. C. Kiwiel, *Free-steering relaxation methods for problems with strictly convex costs and linear constraints*, WP-94-89, International Institute for Applied Systems Analysis, Laxenburg, Austria, September 1994.

[Kiw96] ———, *On the twice differentiable cubic augmented Lagrangian*, J. Optim. Theory Appl. **88** (1996). To appear.

[Lem89] B. Lemaire, *The proximal algorithm*, in New Methods in Optimization and Their Industrial Uses, J. P. Penot, ed., International Series of Numerical Mathematics 87, Birkhäuser, Basel, 1989, pp. 73–87.

[Mar70] B. Martinet, *Régularisation d'inéquations variationelles par approximations successives*, RAIRO Rech. Opér. **4(R3)** (1970) 154–158.

[NiZ92] S. S. Nielsen and S. A. Zenios, *Massively parallel algorithms for singly constrained convex programs*, ORSA J. Comput. **4** (1992) 166–181.

[NiZ93a] ———, *A massively parallel algorithm for nonlinear stochastic network problems*, Oper. Res. **41** (1993) 319–337.

[NiZ93b] ———, *Proximal minimizations with D-functions and the massively parallel solution of linear network programs*, Comput. Optim. Appl. **1** (1993) 375–398.

[NPS94] S. G. Nash, R. Polyak and A. Sofer, *A numerical comparison of barrier and modified barrier methods for large-scale bound-constrained optimization*, in Large Scale Optimization: State of the Art, W. W. Hager, D. W. Hearn and P. M. Pardalos, eds., Kluwer, Dordrecht, 1994, pp. 319–338.

[Pol83] B. T. Polyak, *Introduction to Optimization*, Nauka, Moscow, 1983. English transl., Optimization Software Inc., New York, 1987.

[Pol92] R. Polyak, *Modified barrier functions (theory and methods)*, Math. Programming **54** (1992) 177–222.

[PoT94] R. Polyak and M. Teboulle, *Nonlinear rescaling and proximal-like methods in convex optimization*, Tech. Report ORE-940720, Dept. of Operations Research and Applied Statistics, George Mason Univ., Fairfax, Virginia, 1994.

[Roc70] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[Roc76a] ———, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. **1** (1976) 97–116.

[Roc76b] ———, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976) 877–898.

[Teb92] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res. **17** (1992) 670–690.

[TsB91] P. Tseng and D. P. Bertsekas, *Relaxation methods for problems with strictly convex costs and linear constraints*, Math. Oper. Res. **16** (1991) 462–481.

[TsB93] ———, *On the convergence of the exponential multiplier method for convex programming*, Math. Programming **60** (1993) 1–19.

[Tse90] P. Tseng, *Dual ascent methods for problems with strictly convex costs and linear constraints: A unified approach*, SIAM J. Control Optim. **28** (1990) 214–242.